I

CUADERNOS DE TEORÍA DE LA PROBABILIDAD
PARA MODELOS ECONOMÉTRICOS Y ACTUARIALES

TEORÍA GENERAL DE LAS VARIABLES ALEATORIAS

ALBERTO H. LANDRO
MIRTA L. GONZALEZ

Universidad de Buenos Aires
Facultad de Ciencias Económicas

ISBN: 978-950-29-1736-8

Universidad de Buenos Aires
Facultad de Ciencias Económicas

CUADERNOS DE

TEORIA DE LA PROBABILIDAD

PARA MODELOS

ECONOMÉTRICOS Y ACTUARIALES

Alberto H. LANDRO **Mirta L. GONZALEZ**

Centro de Investigaciones en Econometría
Instituto de Investigaciones en Administración, Contabilidad
y Métodos Cuantitativos para la Gestión – IADCOM
Facultad de Ciencias Económicas
Universidad de Buenos Aires

Landro, Alberto
Teoría general de las variables aleatorias / Alberto Landro ; Mirta L. González. - 1a ed. - Ciudad Autónoma de Buenos Aires : Universidad de Buenos Aires. Facultad de Ciencias Económicas, 2018.
Libro digital, PDF - (Cuadernos de teoría de la probabilidad para modelos econométricos y actuariales / Alberto Landro ; Mirta L. González ; 1)

Archivo Digital: descarga y online
ISBN 978-950-29-1736-8

1. Teoría de las Probabilidades. 2. Econometría. 3. Análisis de Riesgo. I. González, Mirta L. II. Título
CDD 515

ISBN 978-950-29-1736-8

9 789502 917368

Imagen de tapa: *"El Nombre de la Rosa"*, xilografía de Carmen Gilardi.

Hecho el depósito legal que establece la Ley 11.723 – Registro de la Propiedad Intelectual.

INTRODUCCION A LA SERIE

El objetivo último de esta serie de Cuadernos sobre *Teoría de la Probabilidad para Modelos Econométricos y Actuariales* es el tratamiento de aquellos aspectos de la teoría de la probabilidad necesarios para “explicar” el comportamiento de fenómenos dinámicos de las ciencias económicas y actuariales. Fenómenos cuyo estado natural es de no-equilibrio. Un no-equilibrio constructivo que, como consecuencia de su propiedad fundamental de auto-organización, genera nuevos estados y nuevas estructuras infinitamente complejas en el ámbito de la irreversibilidad temporal y que, por tanto, permite asimilarlos a procesos estocásticos que evolucionan en el dominio del tiempo y cuya configuración varía en el dominio de los estados, definidos éstos por la realización simultánea de infinitas variables aleatorias. Esta interpretación de un proceso como una sucesión de variables aleatorias ordenadas en el tiempo implica, en consecuencia, la necesidad de analizar la naturaleza estocástica de sus características individuales, intentando una aproximación a las distribuciones de probabilidades que, sin resentir la rigurosidad del tratamiento de los complejos formalismos de su fundamento matemático, evite la distorsión de la esencia de los fenómenos analizados.

Mediante un riguroso tratamiento conceptual y aplicado se pretende no sólo ofrecer una fuente acabada de conocimiento para la enseñanza y aprendizaje sino brindar una herramienta integral para las aplicaciones profesionales en economía y finanzas, el análisis de riesgo y los problemas actuariales. Cada Cuaderno presentará un contenido preciso y autónomo de un capítulo de la teoría de la probabilidad, adecuado para su uso por profesores y alumnos de grado y posgrado, investigadores y profesionales de las ciencias económicas y actuariales. A todos les damos la bienvenida.

Alberto Landro y Mirta González

Universidad de Buenos Aires
Facultad de Ciencias Económicas

CUADERNO NÚMERO UNO

TEORIA GENERAL
DE LAS
VARIABLES ALEATORIAS

Alberto H. LANDRO **Mirta L. GONZALEZ**

Centro de Investigaciones en Econometría
Instituto de Investigaciones en Administración, Contabilidad
y Métodos Cuantitativos para la Gestión – IADCOM
Facultad de Ciencias Económicas
Universidad de Buenos Aires

INTRODUCCION AL PRIMER CUADERNO

Teniendo en cuenta que, el objetivo de esta serie es “explicar” el comportamiento de fenómenos dinámicos y que éstos son asimilables a procesos estocásticos cuyas características varían de acuerdo con el tipo de aplicación de que se trate, se consideró conveniente comenzar con un número inicial dedicado a la teoría general de las variables aleatorias.

El contenido de este Cuaderno incluye el tratamiento de las características de las variables continuas y discretas, unidimensionales y multidimensionales. En particular está dedicado al estudio de la convergencia estocástica y a las propiedades de las funciones características y generatrices de momentos y probabilidades.

Los números siguientes estarán dedicados al estudio de variables individuales. En particular, aquellas utilizadas más frecuentemente en aplicaciones econométricas y actuariales.

CONTENIDO

1.- Las variables aleatorias unidimensionales

1.1.- Definición

Una variable aleatoria se define como una función real de una partición (ε) del espacio muestral (Ω) asociado a un fenómeno de comportamiento no-determinístico, formada por los eventos incompatibles $\varepsilon = \{E_1, E_2, E_3, \dots\}$ que representan el conjunto exhaustivo de resultados posibles de dicho fenómeno: $X(w) = x_1|E_1| + x_2|E_2| + x_3|E_3| + \cdots (w\epsilon\Omega)$, donde los $|E_i|$ denotan indicadores que pueden asumir los valores 1 ó 0 según que el evento E_i ocurra o no y los x_i son números reales[1]. De esta forma, la definición de variable aleatoria traslada el ámbito del análisis del espacio Ω de los eventos al "dominio" (o "soporte") $\mathbb{R}_n$ al cual pertenecen los valores que puede asumir (este traslado implica una importante simplificación conceptual, en la medida que Ω es un espacio abstracto en tanto que $\mathbb{R}_n$ es un espacio Euclidiano sobre el cual las operaciones son más simples)[2].

La teoría de las variables aleatorias está dirigida fundamentalmente a la asignación de sus probabilidades asociadas. Es decir, dada una variable aleatoria X, su análisis está relacionado, en principio, con la asignación de las probabilidades de realización de cada uno de los valores de su dominio ($x\epsilon\Omega(X)$) o, en términos más generales, de las probabilidades de asumir valores pertenecientes a un conjunto $B\epsilon\Omega(X)$:

$$p(x\epsilon B) = p[X(w)\epsilon B] = p\{w/[X(w)\epsilon B]\} = p[X^{-1}(B)]$$

(donde B está definido por un intervalo o por un número finito o una infinidad numerable de operaciones y $X^{-1}(B)$ denota la imagen inversa de B)[3].

La partición ε que define a una variable aleatoria puede ser numerable o poseer la cardinalidad del continuo (es decir, ser biunívocamente correspondiente con el conjunto de los números reales, $card(\varepsilon) = card(\mathbb{R})$). En el primer caso se dice que la variable es discreta, en el segundo que es continua.

---

[1] En términos más intuitivos, una variable aleatoria (según la nomenclatura de de Finetti (1970)) es un número unívocamente definido pero no conocido (o no conocido aún).

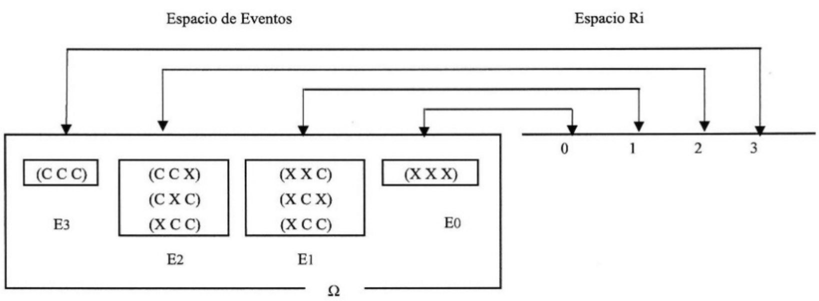
[2] De acuerdo con Slutsky (1925), la denominación de "variable aleatoria" fue introducida por Nekrasov (1901) y difundida por Chuprov. En realidad, ya en 1850 Chebychev había utilizado las expresiones *"cantidad"*, *"variable"* y *"cantidad variable"* al tratar la teoría de los errores de observación no-sistemáticos.

[3] Si los eventos x_i que forman el conjunto $X(w)$ son incompatibles, entonces los eventos $X^{-1}(B)$ también serán incompatibles.

Como una extensión de la definición anterior, se dice que $Z(w) = X(w) + iY(w)$ es una variable aleatoria compleja si $X(w)$ e $Y(w)$ son variables aleatorias reales.

Ejemplo n° 1:

Sea una prueba consistente en arrojar tres monedas "clásicas". La variable aleatoria que representa el número de veces que se puede obtener el resultado "cara" está definida de la siguiente forma:



De acuerdo con el supuesto de "perfección" de las monedas, la probabilidad a asignar al evento "*que la variable asuma el valor 0*", es decir la probabilidad de que ocurra el evento:

$$E_0 : no\ obtener\ ninguna\ vez\ el\ resultado\ "cara"$$

será $p(E_0) = \frac{1}{8}$. De la misma forma, las probabilidades a asignar a la ocurrencia de los eventos:

$$E_i: obtener\ i\ (= 1{,}2{,}3)\ veces\ "cara"$$

serán, respectivamente: $p(E_1) = \frac{3}{8}$, $p(E_2) = \frac{3}{8}$ y $p(E_3) = \frac{1}{8}$.

1.2.- Las variables aleatorias unidimensionales discretas

Sea $X(w)$ una variable aleatoria discreta y sea $\Omega(X) = \{x_i, i = 1{,}2, \dots\}$ su dominio y sea la sucesión:

$$p_i = p[X(w) = x_i] = p\{w/[X(w) = x_i]\} = p[X^{-1}(x_i)](i = 1{,}2{,}3, \dots)$$

la asignación de la probabilidad de que la variable asuma cada uno de estos valores, la cual define la "función de probabilidades" de la variable.

Teniendo en cuenta que los eventos $X(x_i)$ son incompatibles, de acuerdo con los axiomas de no-negatividad de la probabilidad y de aditividad numerable, se obtiene que la función de probabilidades posee las siguientes propiedades:

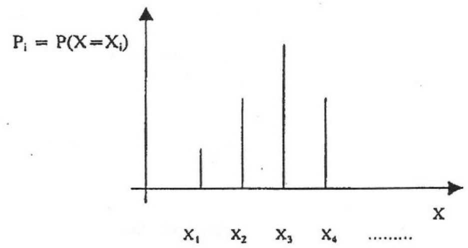
i) $p_i \geq 0 \qquad (i = 1,2,3, \dots)$

ii) $p[X(w)\epsilon\Omega(X)] = p\{X^{-1}[\Omega(X)]\} = p[(X_1 = x_1) \cup (X_2 = x_2) \cup \dots] =$

$= \sum_{x_i\epsilon\Omega(X)} p_i = 1$

Esta propiedad de "aditividad numerable" o "completa" o "monótona" o "σ-aditividad" fue introducida por Borel (1896)(1897)(1898), quien la convirtió en el tema central de la teoría de la probabilidad al demostrar su ley fuerte de los grandes números. Deben reconocerse, además, los aportes de Lebesgue (1901)(1904), Radon (1913)(1915), Fréchet (1915a)(1915b)(1930a)(1930b), Daniell (1918)(1919a)(1919b)(1920)(1921), Wiener (1920)(1921a)(1921b)(1923)(1924) y Steinhaus (1923)(1930a)(1930b). Si bien Kolmogorov (1933), a partir del teorema de continuidad, incorporó a su axiomática la condición de aditividad numerable, manifestó ciertas reservas sobre su validez y la justificó exclusivamente en virtud de su utilidad en ciertos ámbitos de la investigación con respecto a la interpretación frecuencista de la probabilidad (la aditividad numerable no siempre puede derivarse de colectivos que satisfacen los axiomas de aleatoriedad y convergencia)[4]. Por su parte, de Finetti (1937) considera que la aditividad numerable constituye un ejemplo del riesgo que traen aparejadas las conclusiones que se obtienen de un sistema axiomático basado exclusivamente en conveniencias matemáticas (ver Landro (2010)).

La función de probabilidades puede ser graficada en un plano cartesiano en el que los valores de la variable figuran en el eje de abcisas y las probabilidades respectivas están representadas por segmentos de longitud proporcional a cada probabilidad:



[4] Ver Landro, González (2016).

El dominio $\Omega(X) = \{x_i, i = 1,2,\dots\}$ de la variable X y la función de probabilidades $\{x_i, p_i, i = 1,2,\dots\}$ componen su "distribución de probabilidades".

En este caso particular de las variables unidimensionales, cualquier conjunto de valores de su dominio puede ser definido a partir de intervalos, utilizando operaciones de unión y negación. Se puede concluir, entonces, que los intervalos de la recta constituyen la base para la definición de la distribución de probabilidades de una variable unidimensional: el conjunto de posibles valores de la variable comprendidos en un intervalo $[x_h, x_k]$ (para $x_h < x_k$) puede ser expresado por la diferencia$[x_h, x_k] =]-\infty, x_k] -]-\infty, x_h]$ a partir de la cual se obtiene que:

$$p(x_h \leq X \leq x_k) = p(X \leq x_k) - p(X \leq x_h)$$

En resumen, para calcular la probabilidad de un evento $(X \epsilon B)$ basta conocer, para todo $x \epsilon \Omega(X)$, la probabilidad del intervalo $]-\infty, x]$; es decir, basta conocer la función:

$$F_X(x) = p(X \leq x) = p\big[X^{-1} \in]-\infty, x]\big] = \sum_{x_i \leq x} p(X = x_i) = \sum_{x_i \leq x} p_i$$

$$p(x_h \leq x \leq x_k) = F_X(x_k) - F_X(x_h) = \Delta_{x_h}^{x_k} F_X(x)$$

denominada "función de distribución" o "de probabilidades acumuladas" de la variable X, la cual posee las siguientes propiedades:

1) Es una función no-negativa: $F_X(x_i) \geq 0 \;\; (i = 1,2,3,\dots)$ y continua por la izquierda: $F_X(x_i -) = F_X(x_i)$ (los valores x_i se denominan "puntos de continuidad" de X).

2) Dados dos eventos: $A = \{X \leq x_i\}$ y $B = \{X \leq x_j\}$, donde $x_i \leq x_j$ (es decir, tales que $A \subseteq B$), de acuerdo con el axioma de aditividad, se verificará que $p(A) \leq p(B)$. Es decir que $F_X(\cdot)$ es una función no-decreciente, $F_X(x_i) \leq F_X(x_j)\big(\forall x_i \leq x_j\big)$[5].

3) Se verifica que:

$$\lim_{x \to -\infty} F_X(x) = F_X(-\infty) = 0 \qquad\qquad \lim_{x \to +\infty} F_X(x) = F_X(+\infty) = 1$$

---

[5] Dado que se trata de una variable aleatoria discreta y debido a las consiguientes discontinuidades de la función de probabilidades, en este caso la función de distribución crece "a saltos".

Estas condiciones son necesarias y suficientes para que una función pueda ser considerada una función de distribución. Es decir, si una función real $F_X(x)$ satisface las condiciones enunciadas, se puede asegurar que existe una única función de distribución de probabilidades correspondiente a la variable X que es igual a $F_X\,(x)$, para el intervalo $]-\infty, x]$.

4) Sean dos variables aleatorias, X e Y, con funciones de distribución $F_X(x)$ y $F_Y(y)$, respectivamente y tales que $p(|X-Y| \leq \eta) > 1-\varepsilon$ (donde η y ε denotan dos constantes positivas). De acuerdo con el teorema de la probabilidad total, se puede escribir:

$$p[(|X-Y| \leq \eta) \cup (Y \leq y)]$$

$$= p(|X-Y| \leq \eta) + p(Y \leq y) - p[(|X-Y| \leq \eta) \cap (Y \leq y)]$$

De las relaciones anteriores se obtiene que:

$$p[(|X-Y| \leq \eta) \cap (Y \leq y)] = p(X \leq Y+\eta) \geq p(Y \leq y) - \varepsilon$$

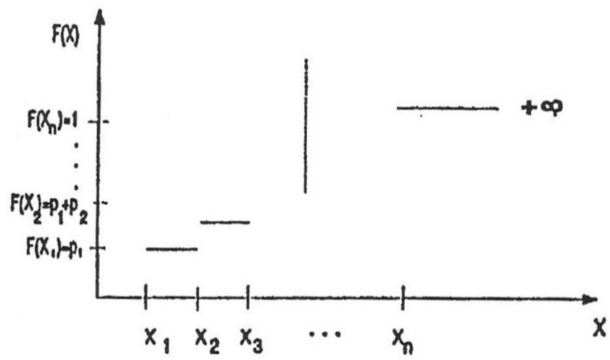
y, por lo tanto, que $p(X > y-\eta) \geq p(Y > y) - \varepsilon$. Es decir, resulta que:

$$F_X(y-\eta) - \varepsilon \leq F_Y(y) \leq F_X(y+\eta) + \varepsilon$$

$$F_X(y-\eta) - F_X(y+\eta) - \varepsilon \leq F_X(y) - F_Y(y) \leq F_X(y+\eta) - F_Y(y-\eta) + \varepsilon$$

$$|F_X(y) - F_Y(y)| \leq [F_X(y+\eta) - F_X(y-\eta)] + \varepsilon$$

La función de distribución puede ser graficada de la siguiente forma:



La función $F_X(x)$ asume el valor p_i para todos los valores de X comprendidos en el intervalo $[x_1, x_2[$. En x_2 y para todos los puntos incluidos en el intervalo $[x_2, x_3[$, será $F_X(x_2) = p_1 + p_2$, y así sucesivamente.

Ejemplo n° 2:

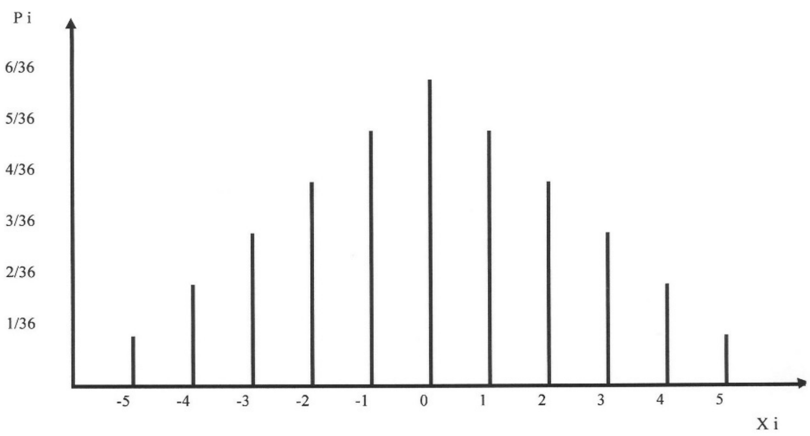
Sea X una variable aleatoria definida como la diferencia de los puntos a obtener entre la primera y la segunda tirada de un dado "clásico". Su dominio está definido de la siguiente forma:

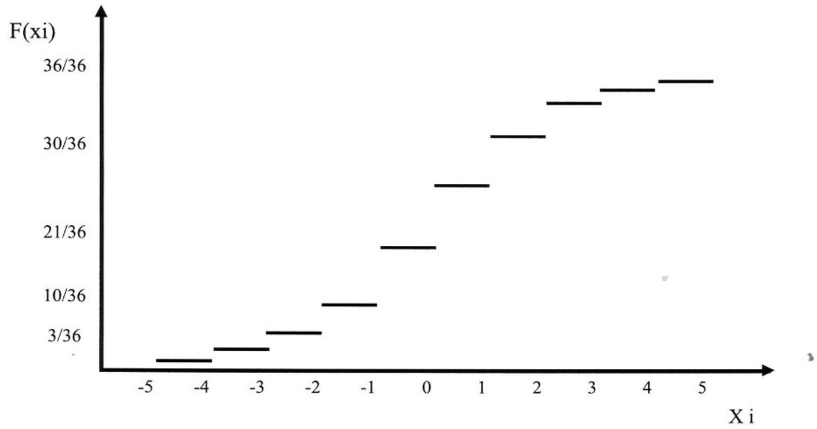
(1 - 1)	(2 - 1)	(3 - 1)	(4 - 1)	(5 - 1)	(6 - 1)
(1 - 2)	(2 - 2)	(3 - 2)	(4 - 2)	(5 - 2)	(6 - 2)
(1 - 3)	(2 - 3)	(3 - 3)	(4 - 3)	(5 - 3)	(6 - 3)
(1 - 4)	(2 - 4)	(3 - 4)	(4 - 4)	(5 - 4)	(6 - 4)
(1 - 5)	(2 - 5)	(3 - 5)	(4 - 5)	(5 - 5)	(6 - 5)
(1 - 6)	(2 - 6)	(3 - 6)	(4 - 6)	(5 - 6)	(6 - 6)

Ω

y las funciones de probabilidades y de distribución asumen los valores:

X	P (X=xi)	F(xi)	X	P (X=xi)	F(xi)
$X_1 = -5$	1/36	1/36	$X_7 = 1$	5/36	26/36
$X_2 = -4$	2/26	3/26	$X_8 = 2$	4/36	30/36
$X_3 = -3$	3/36	6/36	$X_9 = 3$	3/36	33/36
$X_4 = -2$	4/36	10/36	$X_{10} = 4$	2/36	35/36
$X_5 = -1$	5/36	15/36	$X_{11} = 5$	1/36	36/36
$X_6 = 0$	6/36	21/36			





Ejemplo n° 3: **La variable tiempo de espera**

Supóngase una serie de lanzamientos independientes de una moneda tal que la probabilidad de que en una prueba dada se produzca el resultado "cara" (C) es igual a p y la probabilidad de que se produzca el resultado "ceca" (X) es igual a $q = 1 - p$.

Sean los eventos:

$$C_n\text{: } que\ se\ produzca\ el\ resultado\ \text{"cara"}\ en\ el\ n - \text{é}simo\ lanzamiento$$

$$X_n\text{: } que\ se\ produzca\ el\ resultado\ \text{"ceca"}\ en\ el\ n - \text{é}simo\ lanzamiento$$

De modo que $p(C_n) = p$ y $p(X_n) = q$. La probabilidad de que en n lanzamientos de la moneda se produzca una sucesión dada de k resultados "cara" y $n - k$ resultados "ceca" (por ejemplo, la sucesión $C_1, C_2, \dots, C_k$, $X_{k+1}, X_{k+2}, \dots, X_n$) será, entonces:

$$p(C_1 \cap C_2 \cap \dots \cap C_k \cap X_{k+1} \cap X_{k+2} \cap \dots \cap X_n) = p^k q^{n-k}$$

Sea $Z^{(n)}$ la variable aleatoria que representa el número de resultados "cara" a obtener en una serie de n lanzamientos. Su distribución de probabilidades será de la forma:

$$p\left(Z^{(n)} = k\right) = p[(C_1 \cap C_2 \cap \dots \cap X_k \cap C_{k+1} \cap \dots \cap X_n) \cup$$

$$\cup\ (C_1 \cap C_2 \cap \dots \cap X_{k-1} \cap C_k \cap C_{k+1} \cap \dots \cap X_n) \cup$$

$$\cup \dots \cup (X_1 \cap X_2 \cap \dots \cap X_{n-k} \cap C_{n-k+1} \cap \dots \cap C_n)] =$$

$$= p[\cup(C_1 \cap C_2 \cap \dots \cap C_k \cap X_{k+1} \cap \dots \cap X_n)] =$$

$$= p^k q^{n-k} + p^k q^{n-k} + \cdots + p^k q^{n-k} = \binom{n}{k} p^k q^{n-k} \qquad (k = 0,1,2,\dots,n)$$

Sea T la variable aleatoria que representa el número de tiradas hasta la aparición por primera vez del resultado "cara":

$$(T = n) = (X_1 \cap X_2 \cap \dots \cap X_{n-1} \cap C_n) \qquad (n = 1,2,3,\dots)$$

Su distribución de probabilidades será de la forma:

$$p_T(n) = \begin{cases} pq^{n-1} & si \quad n = 1,2,\dots. \\ 0 & \quad\;\; n = 0 \end{cases}$$

Se comprueba fácilmente que:

$$\sum_{n=1}^{\infty} p_T(n) = \sum_{n=1}^{\infty} pq^{n-1} = p\sum_{n=1}^{\infty} q^{n-1} = \frac{p}{1-q} = 1$$

La función de distribución queda, entonces, definida de la siguiente forma:

$$F_T(n) = p(T \leq n) = 1 - p(T > n) = 1 - p(T \geq n+1)$$

$$= 1 - \sum_{j=n+1}^{\infty} p(T = j) =$$

$$= 1 - \sum_{j=n+1}^{\infty} pq^{j-1} = 1 - pq^n \sum_{j=0}^{\infty} \frac{pq^n}{1-q} = 1 - q^n$$

Un problema interesante relacionado con la variable T, es el referido a la distribución de la denominada "espera residual", es decir, a la probabilidad de que el resultado "cara" se produzca al cabo de n_2 lanzamientos de la moneda, si ya sufre un retraso de n_1 tiradas (en otros términos, la probabilidad de que el tiempo de espera del resultado "cara" sea $n_1 + n_2$, sabiendo que debe ser mayor que n_1). De acuerdo con el análisis precedente, se tiene que:

$$p[T = n_1 + n_2/(T > n_1)] = \frac{p[(T = n_1 + n_2) \cap (T > n_1)]}{p(T > n_1)} =$$

$$= \frac{p(T = n_1 + n_2)}{p(T > n_1)} =$$

$$= \frac{pq^{n_1+n_2-1}}{q^{n_1}} = pq^{n_2-1} = p(T > n_2)$$

De lo que se concluye que la variable espera residual tiene la misma distribución de probabilidades que la variable "espera desde el inicio". Un resultado que contradice la idea intuitiva que el tiempo de espera ya transcurrido debería de alguna forma reducir el tiempo de espera no transcurrido. En realidad, debido a la independencia postulada entre estos eventos, el tiempo de espera se reproduce con las mismas características que poseía al inicio, sin que se vea afectado por la extensión del tiempo de espera ya transcurrido[6].

Asociada a una variable aleatoria X es posible definir también una "función de supervivencia":

$$S_X(x_i) = 1 - F_X(x_i) = p(X > x_i) = \sum_{x > x_i} p_j \qquad \left(x_i \epsilon \Omega(X)\right)$$

Se denomina "espectro" de una variable aleatoria al conjunto de puntos con probabilidad no-nula; es decir, al conjunto de puntos de discontinuidad de $F_X(x)$. Se puede concluir en forma inmediata que, para una variable aleatoria discreta, el espectro coincide con su dominio ($\Omega(X)$) y, por lo tanto, tiene probabilidad igual a uno.

Como se verá en la sección siguiente, para una variable aleatoria continua el espectro es vacío y con probabilidad nula. Por lo que se puede concluir que como máximo el espectro puede ser un conjunto numerable. Sea una variable X tal que:

$$F_X(x) = I_{(x>a)=} \begin{cases} 0 & si \quad x \leq a \\ 1 & si \quad x > a \end{cases}$$

(donde $I_{(x>a)}$ denota una función indicadora del conjunto $[x > a]$). Su espectro está definido por el punto a, de modo que $F_X(x)$ representa la función de distribución del caso límite de una variable aleatoria cuyo dominio

[6] Es una creencia generalizada (e injustificada) suponer que, a pesar de la independencia de los eventos, el tiempo de espera de un resultado dado depende de un retardo acumulado. Una hipótesis que, obviamente, contradice la condición de coherencia de la interpretación deFinettiana (ver Landro (2010, p. 58)).

es $\Omega(X) = \{a\}$ y su función de probabilidades es $p(X = a) = 1$. Estas variables se denominan "constantes" o "quasi constantes" o "quasi constantes con certeza" o "degeneradas" o "concentradas en un punto".

1.3.- Las variables aleatorias unidimensionales continuas

Se dice que X es una variable aleatoria continua si su función de distribución, $F_X(x)$ (no-decreciente), es continua para todo $x\epsilon\mathbb{R}_1$ y si existe, además, una "función de densidad de probabilidad" ($f_X(\theta, x)$) no-negativa tal que:

$$F_X(x) = p(X \leq x) = \int_{-\infty}^{x} f_X(\theta, y)dy \qquad (\forall x\epsilon\mathbb{R})$$

(donde θ denota un vector de parámetros).

De modo que, si $F_X(x)$ es absolutamente continua, será[7]:

$$f_X(\theta, x) = \frac{d}{dx}F_X(x) =$$

$$= \lim_{\Delta x \to 0} \frac{F_X(x + \Delta x) - F_X(x)}{\Delta x} = \lim_{\Delta x \to 0} p(x < X < x + \Delta x) \quad (\Delta x > 0)$$

Si bien desempeña el mismo rol que la función $p_i(i = 1,2, \dots)$ para las variables aleatorias discretas, estrictamente hablando, $f_X(\theta, x)$ no es una función de probabilidades, ya que, para las variables continuas la probabilidad en cada punto es nula:

$$p(X = a) = \int_{a}^{a} f_X(x)dx = 0$$

Se puede concluir entonces que, en este caso, por razones puramente matemáticas, el concepto de probabilidad resulta en sí mismo insuficiente para atribuir distintos grados de confiabilidad a los diferentes eventos de la partición que define a la variable aleatoria, ya que los grados de creencia están todos identificados con la probabilidad del evento imposible. En consecuencia, se conviene que, si $f_X(x) \geq f_X(y)$, debe interpretarse que el grado de confiabilidad en la ocurrencia del evento $(X = x)$ es mayor o igual

[7] Que $\frac{d}{dx}F_X(x)$ esté definida sólo para casi todos los valores de X, no afecta a la determinación unívoca de la integral $F_X(x) = \int_{-\infty}^{x} \frac{d}{dx}F_X(x)$.

que el correspondiente al evento $(X = y)$ (aun sabiendo que a los dos eventos les corresponde la misma probabilidad, $p(X = x) = p(X = y) = 0)$[8].

De acuerdo con una interpretación de la probabilidad en términos de teoría de la medida, una probabilidad nula es asimilable a la probabilidad de acertar en un blanco definido por un punto con una flecha cuya punta tiene espesor cero. Se genera entonces una paradoja que podría ser expresada de la siguiente forma: si es imposible acertar en cada punto, ¿cómo se puede explicar que la probabilidad de acertar por lo menos en un punto sea igual a la unidad? En realidad, este planteo es ficticio y surge del error de intentar transformar la probabilidad en certeza, de ignorar que entre un evento "posible", aunque de probabilidad nula y un evento "imposible" existe una diferencia de carácter cualitativo.

Según la definición Boreliana, la probabilidad de que la variable X asuma valores incluidos en un conjunto cualquiera de su dominio queda definida por la siguiente integral de Lebesgue-Stieltjes[9]:

$$p(X \epsilon B) = \int_B dF_X(x) = \int_B f_X(\theta, x)dx \qquad (B \epsilon \mathbb{R}_1)$$

De acuerdo con el concepto de densidad y, teniendo en cuenta la propiedad de aditividad completa, será:

$$f_X(\theta, x)dx = \int_x^{x+dx} f_X(\theta, u)du = p(x \leq X \leq x + dx)$$

La función de densidad posee las siguientes propiedades:

1) $f_X(\theta, x) \geq 0 \qquad (\forall x \epsilon \mathbb{R}_1)$

2) $\int_{-\infty}^{\infty} f_X(\theta, x)\, dx = 1$

Estas condiciones son suficientes para poder asegurar que $f_X(\theta, x)$ es una función de densidad. Es decir, si una función $f_X(\theta, x)$ en $\mathbb{R}_1$cumple estas condiciones, entonces se puede asegurar que la función $F_X(x)$ definida en la página precedente, es una función de distribución de la variable X y, por lo

[8] Se puede concluir, entonces que, como se mencionó en la sección precedente, para una variable aleatoria continua el espectro es vacío y, por lo tanto, con probabilidad nula.
[9] Se concluye fácilmente que $p(X = x) \neq 0$ si y sólo si $F_X(x)$ es discontinua en x. Pero $F_X(x)$ posee puntos de discontinuidad en los puntos $\Omega(X) = \{x_1, x_2, \ldots\}$ que definen el dominio de una variable discreta.

tanto, que $f_X(\theta, x)$ define una función de densidad. Obsérvese que $f_X(\theta, x)dx$ es el valor del área de un rectángulo de base dx y altura proporcional a $f_X(\theta, x)$. En general, dada una variable aleatoria X que existe en el intervalo $\Omega(X) = [c, d]$, la probabilidad de que X asuma valores comprendidos en un intervalo $[a, b]$ $(c \leq a \leq b \leq d)$, puede ser representada como la suma de las áreas de n rectángulos $(n \to \infty)$ de base dx:

$$p(a \leq X \leq b) = \lim_{n \to \infty} \sum_{i=1}^{n} f_X(\xi_i)\, dx = p\left(\bigcup_{x \epsilon [a,b]} E_X\right) = \int_{a}^{b} f_X(x)dx =$$

$$= \int_{-\infty}^{b} f_X(x)dx - \int_{-\infty}^{a} f_X(x)\, dx = F_X(b) - F_X(a) =$$

$$= \Delta_a^b F_X(x) \qquad (-\infty < a < b < \infty)$$

Sea una variable aleatoria X con función de distribución $F_X(x)$ $\big(x\epsilon\Omega(X)\big)$, entonces toda potencia positiva $(\alpha > 0)$ $F_X^{\alpha}(x)$ es una función de distribución. Una propiedad similar se cumple para las funciones de supervivencia.

Ejemplo n°4:

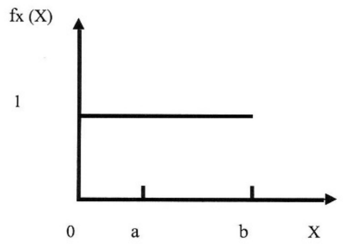
Como se vio en la página precedente, para variables continuas la condición $p(X = x) = 0$ (para cualquier valor de x) no deriva de ninguna hipótesis de equiprobabilidad. Para definir una distribución continua que atribuya el mismo grado de creencia a todos los puntos de su dominio o de cualquiera de sus subconjuntos, es natural imponer la condición que, para cualquier par de valores $x, y\epsilon\Omega(X)$, se verifique que $f_X(x) = f_X(y)$. Es decir, que la función de densidad sea constante sobre el conjunto de puntos que se suponen equiprobables. La selección de esta constante obviamente debe ser realizada de modo que se cumpla la condición $\int_{-\infty}^{\infty} f_X(x)dx = 1$. Luego, dado un dominio $\Omega(X) = [a, b]$, sea una variable aleatoria continua X con función de densidad de la forma[10]:

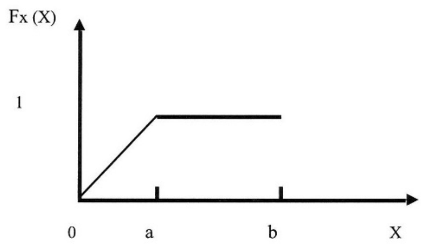
$$f_X(x) = \begin{cases} 0 & para \quad x < a \\ \dfrac{1}{b-a} & para \quad a \leq x \leq b \\ 0 & para \quad x > b \end{cases}$$

[10] La utilización de intervalos cerrados o abiertos es indiferente ya que, como se mencionó más arriba, dado que la distribución de probabilidades es continua, la probabilidad en un punto es igual a cero.

La función de distribución está definida por:

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy = \begin{cases} \int_{-\infty}^{a} 0du = 0 & (x < a) \\ \int_{-\infty}^{a} 0du + \int_{a}^{x} \frac{1}{b-a}du = \frac{x-a}{b-a} & (a \leq x \leq b) \\ \int_{-\infty}^{a} 0du + \int_{a}^{b} \frac{1}{b-a}du + \int_{b}^{x} 0du = 1 & (x > b) \end{cases}$$





De acuerdo con lo mencionado en las páginas precedentes, la densidad constante en el intervalo $[a, b]$ traduce la idea intuitiva de equiprobabilidad para todos los puntos del intervalo. De modo que la probabilidad de que la variable asuma valores comprendidos en el intervalo $[x_1, x_2]$, siendo $a \leq x_1 \leq x_2 \leq b$, es directamente proporcional a la longitud de dicho intervalo:

$$p(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} f_X(x)dx = \int_{x_1}^{x_2} \frac{1}{b-a}dx = \frac{x_2 - x_1}{b-a}$$

Luego, a intervalos de igual longitud les corresponderán iguales probabilidades. Esta propiedad es análoga a la de la distribución discreta que asigna probabilidades iguales a un número finito de puntos.

Esta distribución –a la que se conoce como "uniforme" y se denota por $U(a, b)$- es un ejemplo de los peligros que trae aparejado un planteo axiomático de la teoría de la probabilidad y la pretensión de extender conclusiones abstractas a aplicaciones concretas: la probabilidad unitaria del evento cierto surge como la suma de infinitas probabilidades nulas relativas a los puntos racionales, lo cual conduce a las consideraciones sobre la insuficiencia del concepto de aditividad completa realizadas en la Sec.1.2.

2.- Las variables aleatorias de más de una dimensión

2.1.- Las variables aleatorias multidimensionales continuas

Sea $\{X_1(w), X_2(w), \dots, X_n(w)\}(w \epsilon \Omega)$ un conjunto de variables aleatorias unidimensionales continuas. La combinación de estas variables permite definir una variable de n dimensiones $\boldsymbol{X(w)}$ cuyo dominio está dado por el producto cartesiano de los dominios de las "variables marginales" (o "variables componentes"). Se dice, entonces, que el vector:

$$\boldsymbol{X(w)} = [X_1(w) \quad X_2(w) \dots \quad X_n(w)]^T \quad (\forall w \epsilon \Omega)$$

puede ser considerado una variable aleatoria continua n-dimensional con "función de distribución de probabilidades conjunta":

$$F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) =$$

$$= p[(X_1 \leq x_1) \cap (X_2 \leq x_2) \cap \dots \cap (X_n \leq x_n)] \quad (x_1, x_2, \dots, x_n \epsilon B, B \epsilon \mathbb{R}_n)$$

si existe una función de densidad $f_{X_1,X_2,\dots,X_n}(\boldsymbol{\theta}, x_1, x_2, \dots, x_n)$ tal que:

$$F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) = F_{\boldsymbol{X}}(\boldsymbol{x}) = p(\boldsymbol{X} \leq \boldsymbol{x}) =$$

$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f_{X_1,X_2,\dots,X_n}(\boldsymbol{\theta}, y_1, y_2, \dots, y_n)\, dy_1 dy_2 \dots dy_n$$

donde $\boldsymbol{X} \leq \boldsymbol{x}$ debe ser entendida como la desigualdad entre dos vectores, de modo que cada componente del primer vector es menor o igual que el correspondiente componente del segundo.

Si se verifica que:

$$F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) =$$

$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} \frac{\partial^n F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n} dx_1 dx_2 \dots dx_n$$

la función $F_{\boldsymbol{X}}(\boldsymbol{x})$ es absolutamente continua.

Esta función de densidad posee las siguientes propiedades:

1) $f_{X_1,X_2,\dots,X_n}(\boldsymbol{\theta}, x_1, x_2, \dots, x_n) \geq 0$

2) $\int_{\Omega(X)} f_{X_1,X_2,\ldots,X_n}(\boldsymbol{\theta}, x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n = 1$

$(\Omega(X) = \Omega(X_1) \times \Omega(X_2) \times \ldots \times \Omega(X_n) \in \mathbb{R}_n)$.

La función de distribución conjunta para cualquier subconjunto $\{X_1, X_2, \ldots, X_k\}(k < n)$ del vector $\boldsymbol{X}(\boldsymbol{w})$ define la llamada "función de distribución de probabilidades marginal".

A partir de la definición de la función de densidad conjunta es posible calcular la probabilidad de que la variable $\boldsymbol{X}$ esté incluida en una región cualquiera de su dominio:

$$p(\boldsymbol{X} \epsilon B) = \int_B f_{X_1,X_2,\ldots,X_n}(\boldsymbol{\theta}, x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n \qquad (B \epsilon \mathbb{R}_n)$$

Sea, en particular, una variable aleatoria de dos dimensiones, $\boldsymbol{X}(\boldsymbol{w}) = [X_1(w), X_2(w)]$. Entonces será:

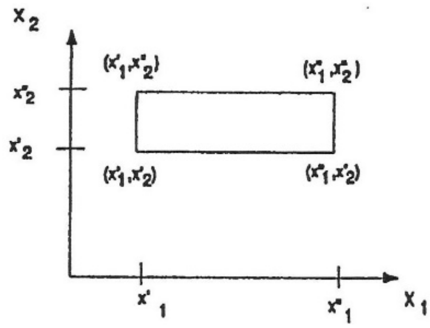
$$F_{X_1,X_2}(x_1, x_2) = p[(X_1 \leq x_1) \cap (X_2 \leq x_2)] =$$

$$= p\{w/[(X_1 \leq x_1) \cap (X_2 \leq x_2)]\} =$$

$$= p\{[w/(X_1 \leq x_1)] \cap [w/(X_2 \leq x_2)]\} =$$

$$= \int_{\substack{y_1 \leq x_1 \\ y_2 \leq x_2}} dF_{X_1,X_2}(y_1, y_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} dF_{X_1,X_2}(y_1, y_2) \quad (X_1, X_2 \epsilon B, B \epsilon \mathbb{R}_2)$$

La probabilidad de que la variable $\boldsymbol{X}$ asuma valores comprendidos en un rectángulo $(x_{1'} \leq X_1 \leq x_{1''}, x_{2'} \leq X_2 \leq x_{2''})$, para $x_{1'} \leq x_{1''}, x_{2'} \leq x_{2''}$ - definido por el producto Cartesiano $\Omega(X_1) \times \Omega(X_2)$, entonces está definida de la siguiente forma:



$$p[(x_{1'} \leq X_1 \leq x_{1''}) \cap (x_{2'} \leq X_2 \leq x_{2''})] = \int\limits_{\substack{x_{1'} \leq u_1 \leq x_{1''} \\ x_{2'} \leq u_2 \leq x_{2''}}} dF_{X_1,X_2}(u_1, u_2) =$$

$$= p[(x_{1'} \leq X_1 \leq x_{1''}) \cap (X_2 \leq x_{2''})] - p[(x_{1'} \leq X_1 \leq x_{1''}) \cap (X_2 \leq x_{2'})] =$$

$$= p[(X_1 \leq x_{1''}) \cap (X_2 \leq x_{2''})] - p[(X_1 \leq x_{1'}) \cap (X_2 \leq x_{2'})] -$$

$$-\{p[(X_1 \leq x_{1''}) \cap (X_2 \leq x_{2'})] - p[(X_1 \leq x_{1'}) \cap (X_2 \leq x_{2'})]\} =$$

$$= F_{X_1,X_2}(x_{1''}, x_{2''}) - F_{X_1,X_2}(x_{1'}, x_{2''}) - F_{X_1,X_2}(x_{1''}, x_{2'}) + F_{X_1,X_2}(x_{1'}, x_{2'})$$

Introduciendo un operador diferencia simple de la función F con respecto a la variable X_1, de $x_{1'}$ a $x_{1''}$, de la forma:

$$\Delta_{X_1}^{(x_{1'}, x_{1''})} F_{X_1,X_2}(x_1, x_2) = F_{X_1,X_2}(x_{1''}, x_2) - F_{X_1,X_2}(x_{1'}, x_2)$$

y, teniendo en cuenta que las diferencias múltiples se obtienen de la iteración de las diferencias simples con respecto a distintas variables, resulta que la definición anterior puede ser expresada de la siguiente forma:

$$p[(x_{1'} \leq X_1 \leq x_{1''}) \cap (x_{2'} \leq X_2 \leq x_{2''})] =$$

$$= \Delta_{X_1}^{(x_{1'}, x_{1''})} \Delta_{X_2}^{(x_{2'}, x_{2''})} F_{X_1,X_2}(x_1, x_2) =$$

$$= \Delta_{X_1}^{(x_{1'}, x_{1''})} \left[F_{X_1,X_2}(x_1, x_{2''}) - F_{X_1,X_2}(x_1, x_{2'}) \right]$$

La función de distribución $F_{X_1,X_2}(x_1, x_2)$ posee las mismas propiedades que las correspondientes a la variable unidimensional:

1) Es no-decreciente, es decir, para $x_{1'} < x_{1''}$ y $x_{2'} < x_{2''}$ todas las diferencias de segundo orden son no-negativas:

$$\Delta_{X_1}^{(x_{1'}, x_{1''})} \Delta_{X_2}^{(x_{2'}, x_{2''})} F_{X_1,X_2}(x_1, x_2) =$$

$$= p[(x_{1'} \leq X_1 \leq x_{1''}) \cap (x_{2'} \leq X_2 \leq x_{2''})] \geq 0$$

2) Es continua por la izquierda para cada una de las variables marginales.

3) Por definición $F_{X_1,X_2}(x_1, x_2) = p[(X_1 \leq x_1) \cap (X_2 \leq x_2)]$. Ahora bien, cuando $x_1 \to \infty$, el conjunto$(X_1 \leq x_1) \to \emptyset$. Luego, la intersección $(X_1 \leq$

$x_1) \cap (X_2 \leq x_2) \rightarrow \emptyset$ y, por lo tanto, $p[(X_1 \leq x_1) \cap (X_2 \leq x_2)] \rightarrow 0$. Es decir, se verifica que:

$$\lim_{x_1 \rightarrow -\infty} F_{X_1,X_2}(x_1, x_2) = \lim_{x_2 \rightarrow -\infty} F_{X_1,X_2}(x_1, x_2) = 0$$

4) Supóngase que $x_1 \rightarrow +\infty$ y $x_2 \rightarrow +\infty$, entonces $(X_1 \leq x_1) \rightarrow \Omega$ y $(X_2 \leq x_2) \rightarrow \Omega$. Luego, la intersección $(X_1 \leq x_1) \cap (X_2 \leq x_2) \rightarrow \Omega$ y, por lo tanto $p[(X_1 \leq x_1) \cap (X_2 \leq x_2)] \rightarrow 1$. Es decir, se verifica que:

$$\lim_{\substack{x_1 \rightarrow +\infty \\ x_2 \rightarrow +\infty}} F_{X_1,X_2}(x_1, x_2) = 1$$

5) Supóngase que sólo $x_1 \rightarrow \infty$, entonces se verifica: **i)** que $(X_1 \leq x_1) \rightarrow \Omega$ u **ii)** que la intersección $\Omega \cap (X_2 \leq x_2) = (X_2 \leq x_2)$. Luego, será:

$$\lim_{x_1 \rightarrow +\infty} F_{X_1,X_2}(x_1, x_2) = \lim_{x_1 \rightarrow +\infty} \int_{-\infty}^{x_2} \left[\int_{-\infty}^{x_1} f_{X_1,X_2}(\theta, y_1, y_2) dy_1 \right] dy_2 =$$

$$\int_{-\infty}^{x_2} f_{X_2}(y_2) dy_2 = p(X_2 \leq x_2) = F_{X_2}(x_2)$$

De la misma forma, resulta que:

$$\lim_{x_2 \rightarrow +\infty} F_{X_1,X_2}(x_1, x_2) = \lim_{x_2 \rightarrow +\infty} \int_{-\infty}^{x_1} \left[\int_{-\infty}^{x_2} f_{X_1,X_2}(\theta, y_1, y_2) dy_2 \right] dy_1 =$$

$$= \int_{-\infty}^{x_2} f_{X_1}(y_1) dy_1 = p(X_1 \leq x_1) = F_{X_1}(x_1)$$

donde $F_{X_1(\cdot)}$ y $F_{X_2(\cdot)}$ denotan las funciones de distribución marginales y

$$f_{X_1}(\theta, x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(\theta, x_1, x_2) dx_2$$

$$f_{X_2}(\theta, x_2) = \int_{-\infty}^{\infty} f_{X_1,X_2}(\theta, x_1, x_2) dx_1$$

denotan las funciones de densidad marginales (obsérvese que, a partir de la función de la función de distribución conjunta de una variable multidimensional $\{X_1, X_2, \dots, X_n\}$ es posible analizar las relaciones –estocásticas- entre las variables componentes).

6) A partir de las propiedades **1)** y **3)** se demuestra que:

$$si\ x_{1'} \leq x_{1''} \Rightarrow F_{X_1,X_2}(x_{1'}, x_2) \leq F_{X_1,X_2}(x_{1''}, x_2) \Rightarrow$$

$$\Rightarrow \Delta_{X_1}^{(x_{1'}, x_{1''})} F_{X_1,X_2}(x_1, x_2) \geq 0$$

$$si\ x_{2'} \leq x_{2''} \Rightarrow F_{X_1,X_2}(x_1, x_{2'}) \leq F_{X_1,X_2}(x_1, x_{2''}) \Rightarrow$$

$$\Rightarrow \Delta_{X_1}^{(x_{2'}, x_{2''})} F_{X_1,X_2}(x_1, x_2) \geq 0$$

Es decir que la función de distribución correspondiente a una variable bidimensional es no-decreciente con respecto a cada una de las variables marginales.

El conjunto de funciones de distribución $F_{X,Y}(x, y)$ que admite a $F_X(x)$ y $F_Y(y)$ como funciones de distribución marginales define la "clase de Fréchet" asociada a las funciones $F_X(x)$ y $F_Y(y)$", $\mathcal{F}\big(F_X(x), F_Y(y)\big)$[11]. De la misma forma, dadas las funciones de supervivencia $S_X(x)$ y $S_Y(y)$ marginales, es posible definir una clase de Fréchet asociada a $S_X(x)$ y $S_Y(y)$, $\mathcal{F}\big(S_X(x), S_Y(y)\big)$.

Dado que, como se mencionó más arriba, $F_{X,Y}(x, y)$ es una función no-decreciente en x e y, será $F_{X,Y}(x, y) \leq F_{X,Y}(x, \infty) = F_X(x)$ y $F_{X,Y}(x, y) \leq F_{X,Y}(\infty, y) \leq F_Y(y)$ y, por lo tanto, $F_{X,Y}(x, y) \leq min[F_X(x), F_Y(y)] = F_{X,Y}^{(m)}(x, y)$.

Por otra parte, dado que:

$$F_{X,Y}(x, y) = S_{X,Y}(x, y) - 1 + F_X(x) + F_Y(y)$$

y, teniendo en cuenta que $S_{X,Y}(x, y) \geq 0$, será:

[11] Fréchet (1935) (1951) (1957). Dada la importancia de las contribuciones realizadas por Hoeffding (1940) (1941) sobre este tema, a esta clase se la debería denominar como de "Fréchet-Hoeffding".

$$F_{X,Y}(x,y) \geq Max[0, F_X(x) + F_Y(y) - 1] = F_{X,Y}^{(M)}(x,y)$$

Estas distribuciones “mínima” y “máxima” definen las “cotas de Fréchet”.

En general, el conjunto de funciones de distribución $F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n)$ que admite a $F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)$ como funciones de distribución marginales, define la clase de Fréchet asociada a dichas funciones, $\mathcal{F}\left[F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)\right]$, cuyas cotas están dada por las desigualdades:

$$Max\left[0, F_{X_1}(x_1) + F_{X_2}(x_2) + \cdots + F_{X_n}(x_n) - (n-1)\right] \leq$$

$$\leq F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) \leq \min_j \left[F_{X_j}(x_j)\right]$$

Se demuestra que $F_{X_1,X_2,\dots,X_n}^{(M)}(x_1, x_2, \dots, x_n)$ es una función de distribución. Pero $F_{X_1,X_2,\dots,X_n}^{(M)}(x_1, x_2, \dots, x_n)$ será un función de distribución si y sólo si se verifica una de las siguientes condiciones: **i)** que $\sum_{j=1}^{n} F_{X_j}(x_j) \leq 1$ (siendo $0 < F_{X_j}(x_j) < 1,\ \ j = 1,2, \dots, n$) y **ii)** que $\sum_{j=1}^{n} F_{X_j}(x_j) \leq n-1$ (siendo $0 < F_{X_j}(x_j) < 1, j = 1,2, \dots, n$))[12].

Ejemplo n° 5:

Sea una variable aleatoria bidimensional (X, Y) con función de densidad de la forma:

$$f_{X,Y}(x,y) = \begin{cases} 4xy & para \quad 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & para\ cualquier\ otro\ valor\ de\ x\ y\ de\ y \end{cases}$$

El valor de la función de distribución $F_{X,Y}(x,y)$ depende de la región a la que pertenece el punto (X, Y). Efectuando los cálculos correspondientes, se obtiene que:

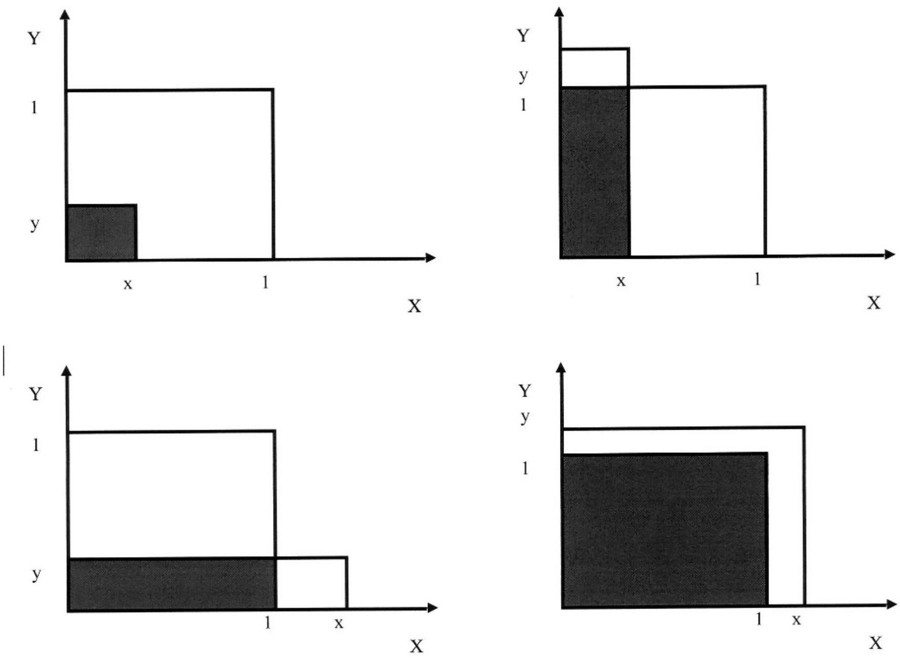
$$F_{X,Y}(x,y) = \int_0^y \int_0^x 4xydxdy = x^2y^2 \quad (0 \leq x \leq 1, 0 \leq y \leq 1)$$

[12] Ver Dall’Aglio (1972).

$$F_{X,Y}(x,y) = \int_0^1 \int_0^x 4xydxdy = 2x^2 \int_0^1 ydy = x^2 \qquad (0 \leq x \leq 1, y > 1)$$

$$F_{X,Y}(x,y) = \int_0^y \int_0^1 4xydxdy = 2y^2 \int_0^1 xdx = y^2 \qquad (x > 1, 0 \leq y \leq 1)$$

$$F_{X,Y}(x,y) = \int_0^1 4xydxdy = 1 \qquad (x > 1, y > 1)$$



El área sombreada representa el conjunto sobre el cual se integra la función $f_{X,Y}(x,y)$ para obtener $F_{X,Y}(x,y)$. En resumen, se puede escribir:

$$F_{X,Y}(x,y) = \begin{cases} x^2y^2 & para \quad 0 \leq x \leq 1; 0 \leq y \leq 1 \\ x^2 & 0 \leq x \leq 1; y > 1 \\ y^2 & x > 1; 0 \leq y \leq 1 \\ 1 & x > 1, y > 1 \\ 0 \ para\ cualesquiera\ otros\ valores\ de\ x\ e\ y & \end{cases}$$

Las funciones de probabilidades marginales están definidas por:

$$f_X(x) = \int_0^1 f_{X,Y}(x,y)dy = \int_0^1 4xydy = 2x \qquad (0 \leq x \leq 1)$$

$$f_Y(y) = \int_0^1 f_{X,Y}(x,y)dxdy = \int_0^1 4xydx = 2y \qquad (0 \leq y \leq 1)$$

Las propiedades enunciadas para las funciones de distribución correspondientes a variables aleatorias bidimensionales pueden ser generalizadas a variables aleatorias de n dimensiones. Es decir, se puede escribir:

$$p[(x_{1'} \leq X_1 \leq x_{1''}) \cap (x_{2'} \leq X_2 \leq x_{2''}) \cap \ldots \cap (x_{n'} \leq X_n \leq x_{n''})] =$$

$$= \Delta_{X_1}^{(x_{1'},x_{1''})} \Delta_{X_2}^{(x_{2'},x_{2''})} \ldots \Delta_{X_n}^{(x_{n'},x_{n''})} F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) =$$

$$= \Delta_{X_1}^{(x_{1'},x_{1''})} \Delta_{X_2}^{(x_{2'},x_{2''})} \ldots \Delta_{X_n}^{(x_{n'},x_{n''})} F_{\boldsymbol{X}}(\boldsymbol{x})$$

Con un razonamiento similar al del caso bidimensional, se demuestra que $F_{\boldsymbol{X}}(\boldsymbol{x})$ posee las siguientes propiedades:

1) Es una función no-decreciente, es decir que, para $x_{1'} < x_{1''}, x_{2'} < x_{2''}, \ldots, x_{n'} < x_{n''}$, se verifica que:

$$\Delta_{X_1}^{(x_{1'},x_{1''})} \Delta_{X_2}^{(x_{2'},x_{2''})} \ldots \Delta_{X_n}^{(x_{n'},x_{n''})} F_{\boldsymbol{X}}(\boldsymbol{x}) \geq 0$$

2) Es continua por la izquierda para cada una de las variables marginales.

3) Tiende a cero cuando alguna de las variables marginales tiende a $-\infty$:

$$\lim_{x_1 \to -\infty} F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) = \lim_{x_2 \to -\infty} F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) =$$

$$= \lim_{x_n \to -\infty} F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) = 0$$

4) Tiende a 1 cuando todas las variables marginales tienden a $+\infty$:

$$\lim_{\substack{x_1 \to +\infty \\ x_2 \to +\infty \\ \dots \\ x_n \to +\infty}} F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) = 1$$

5) Cuando $k\ (= 1,2, \dots, n-1)$ variables marginales tienden a $+\infty$ queda definida la función de distribución de la variable aleatoria formada por la combinación de las $n-k$ variables restantes:

$$\lim_{\substack{x_1 \to +\infty \\ x_2 \to +\infty \\ \dots \\ x_k \to +\infty}} F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) = F_{X_{k+1},X_{k+2},\dots,X_n}(x_{k+1}, x_{k+2}, \dots, x_n)$$

6) La diferencia de orden $k\ (= 1,2, \dots, n-1)$ respecto de k de las n variables marginales es no-negativa:

$$\Delta_{X_1}^{(x_{1'}, x_{1''})} \Delta_{X_2}^{(x_{2'}, x_{2''})} \dots \Delta_{X_n}^{(x_{k'}, x_{k''})} F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) \geq 0$$

$(x_{1'} < x_{1''}, x_{2'} < x_{2''}, \dots, x_{k'} < x_{k''})$.

7) Toda potencia $\alpha \geq n-1$ de la función de distribución conjunta, $F^{\alpha}_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n)$ es una función de distribución. Una propiedad similar se cumple para la función de supervivencia conjunta.

8) Si la potencia $F^{\alpha}_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n)$ define una función de distribución para todo $\alpha > 0$, entonces la función $F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n)$ es "max-infinitamente divisible".

9) Si la potencia $\left[1 - F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n)\right]^{\alpha} = S^{\alpha}_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n)$ define una función de supervivencia para todo $\alpha > 0$, entonces la función de distribución conjunta, $F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n)$ es "min-infinitamente divisible".

Sea una variable n-dimensional $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$. Su función de supervivencia será:

$$S_{\boldsymbol{X}}(\boldsymbol{x}) = S_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) = 1 - F_{\boldsymbol{X}}(\boldsymbol{x}) =$$

$$= p[(X_1 > x_1) \cap (X_2 > x_2) \cap \ldots \cap (X_n > x_n)]$$

Como ya se mencionó en la Sec. 1.2, para $n = 1$, $S_{X_1}(x_1) = 1 - F_{X_1}(x_1)$. Para $n \geq 2$, será:

$$S_{X_1,X_2}(x_1, x_2) = 1 - F_{X_1}(x_1) - F_{X_2}(x_2) + F_{X_1,X_2}(x_1, x_2)$$

$$S_{X_1,X_2,X_3}(x_1, x_2, x_3) = 1 - F_{X_1}(x_1) - F_{X_2}(x_2) - F_{X_3}(x_3) +$$

$$+F_{X_1,X_2}(x_1, x_2) + F_{X_1,X_3}(x_1, x_3) + F_{X_2,X_3}(x_2, x_3) - F_{X_1,X_2,X_3}(x_1, x_2, x_3)$$

$$\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$$

$$S_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = 1 + (-1)\sum_{i=1}^{n} F_{X_i}(x_i) +$$

$$+(-1)^2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} F_{X_i,X_j}(x_i, x_j) + (-1)^3 \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{h=j+1}^{n} F_{X_i,X_j,X_h}(x_i, x_j, x_h)$$

$$+ \cdots + (-1)^n F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$$

2.2.- Las variables multidimensionales discretas

Dado un conjunto S finito o infinito numerable cuyos puntos son vectores de n componentes, es posible definir una variable discreta n-dimensional tal que $p(\boldsymbol{X} \epsilon S) = 1$.

En particular, dada una variable aleatoria de dos dimensiones formada por la combinación de las variables marginales (X, Y) la probabilidad de que asuma valores comprendidos en un conjunto $B \epsilon S$ de su dominio, está definida por $p[(X, Y)\epsilon B] = \sum_{(x_i, y_j)\epsilon B} p_{ij}$, donde $p_{ij} = p[(X = x_i) \cap (Y = y_j)](i, j = 1,2, \ldots)$ representa la función de probabilidades conjunta, la cual posee las siguientes propiedades:

1) $p_{ij} \geq 0 \quad (i = 1,2, \ldots; j = 1,2, \ldots)$

2) $\sum_i \sum_j p_{ij} = 1$

A partir de la función de probabilidades conjunta es posible definir las funciones de probabilidades marginales de la siguiente forma:

$$p_{i.} = \sum_j p_{ij} = \sum_j p[(X = x_i) \cap (Y = y_j)] = p(X = x_i)$$

$$p_{.j} = \sum_i p_{ij} = \sum_i p[(X = x_i) \cap (Y = y_j)] = p(Y = y_j)$$

pero la definición de la función de probabilidades conjunta a partir de las probabilidades marginales sólo es posible si las variables son estocásticamente independientes entre sí.

En general de la combinación de un conjunto $\{X_1, X_2, \dots, X_n\}$ de variables aleatorias unidimensionales discretas, queda definida una variable aleatoria $\boldsymbol{X}_{(n)}$ n-dimensional, cuya función de probabilidades conjunta será de la forma:

$$p(\boldsymbol{X}) = p_{i,j,\dots,h} =$$

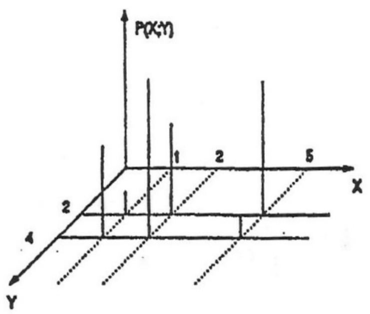
$$= p[(X_1 = x_i) \cap (X_2 = x_j) \cap \dots \cap (X_n = x_h)] \qquad (i, j, \dots, h = 1,2, \dots)$$

Las definiciones y propiedades de las funciones de distribución correspondiente a las variables aleatorias multidimensionales discretas pueden obtenerse como una extensión inmediata del caso continuo.

Ejemplo n° 6:

Sea (X, Y) una variable aleatoria bidimensional discreta, cuya función de probabilidades conjunta figura en la siguiente tabla:

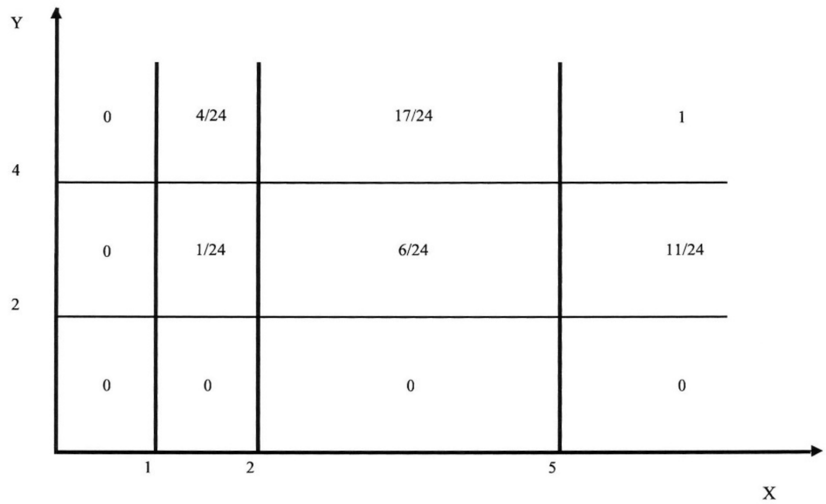
X \ Y	$Y_1 = 2$	$Y_2 = 4$	$P_{i.} = P(X = x_i)$
$X_1 = 1$	1/24	3/24	4/24
$X_2 = 2$	5/24	8/24	13/24
$X_3 = 5$	5/24	2/24	7/24
$P_{.j} = P(Y = y_j)$	11/24	13/24	1



En la última columna y la última fila de la tabla figuran los valores correspondientes a las distribuciones de probabilidades marginales de X e Y, respectivamente.

La función de distribución conjunta queda definida de la siguiente forma:

$$F_{X,Y}(x,y) = \begin{cases} 0 & para & x < 1, y < 2 \\ \frac{1}{24} & & 1 \leq x < 2, 2 \leq y < 4 \\ \frac{4}{24} & & 1 \leq x < 2, y \geq 4 \\ \frac{6}{24} & & 2 \leq x < 5, 2 \leq y < 4 \\ \frac{17}{24} & & 2 \leq x < 5, y \geq 4 \\ \frac{11}{24} & & x \geq 5, 2 \leq y < 4 \\ 1 & & x \geq 5, y \geq 4 \end{cases}$$



En el gráfico precedente los valores numéricos que figuran en los respectivos intervalos de X e Y deben ser interpretados como los volúmenes de los prismas representativos de los correspondientes valores de $F_{X,Y}(x,y)$.

3.- Igualdad y semejanza de variables aleatorias

Se dice que dos variables aleatorias, X e Y son "iguales" (o "equivalentes") cuando son iguales los valores que asumen las funciones de densidad para sus correspondientes valores sobre el dominio Ω que las definen (obviamente, en consecuencia, los valores a asumir por las variables son necesariamente los mismos, $X(w) = Y(w)(\forall w \epsilon \Omega)$).

Como corolario de la definición precedente se puede concluir que, si las variables X e Y son iguales, entonces la expresión $X = Y$ define un evento cierto, es decir $\{w: X(w) = Y(w)\} = \Omega$ y, por lo tanto, será $p(X = Y) = 1$ y $p(X \neq Y) = 0$. Esto significa que la distribución de probabilidades de la variable bidimensional (X, Y) está centrada sobre la recta $y = x$. Entonces, será:

$$F_X(u) = p(X \leq u) = p(Y \leq u) = F_Y(u)$$

Si las variables X e Y son tales que $p(X = Y) \approx 1$ (es decir, si el evento $(X = Y)$ es "quasi-cierto") entonces se dice que son "quasi-iguales"[13].

Se dice que la variable aleatoria X es "menor" que la variable aleatoria Y cuando se verifica que $X(w) < Y(w)(\forall w \epsilon \Omega)$. Como corolario de esta definición, si se verifican simultáneamente las condiciones $X_{i+1}(w) < X_i(w)$ $(i = 1,2, \dots)$, se dice que $\{X_i, i = 1,2, \dots\}$ constituye una sucesión decreciente de variables aleatorias.

Se dice que dos variables aleatorias X e Y son "semejantes" (o "marginalmente iguales") cuando poseen la misma distribución de probabilidades. Obviamente la propiedad de igualdad implica la de semejanza pero la implicación inversa no se verifica, la propiedad de semejanza no implica la de igualdad: para cada par de funciones $F_X(x)$ y $F_Y(y)$ existe más de una función conjunta, $F_{X,Y}(x, y)$, que las admite como funciones marginales.

4.- La distribución de probabilidades condicionada

Sea (X, Y) una variable aleatoria discreta bidimensional con distribución de probabilidades $\{(x_i, y_j), p_{ij}, i = 1,2, \dots; j = 1,2, \dots\}$. La probabilidad de que la variable X asuma un valor particular x_i condicionado por el supuesto que Y asuma un valor dado, y_j, define la función de probabilidades condicionada por el evento condicionante $(Y = y_j)$:

---

[13] Esta condición puede ser expresada alternativamente como $p(|X - Y| > \varepsilon) < \eta$.

$$p[(X = x_i)/(Y = y_j)] = \frac{p[(X = x_i) \cap (Y = y_j)]}{p(Y = y_j)} = \frac{p_{ij}}{p_j} \quad (i, j = 1,2, \dots)$$

que posee las siguientes propiedades:

1) $p[(X = x_i)/(Y = y_j)] \geq 0 \qquad (i, j = 1,2, \dots)$

2) $\sum_i p[(X = x_i)/(Y = y_j)] = 1$

Obviamente, la probabilidad condicionada sólo puede ser definida si $p_j \neq 0$. Por lo tanto, la definición anterior es aplicable únicamente a variables aleatorias discretas. Dada una variable aleatoria continua (X, Y) con función de densidad $f_{X,Y}(\theta, x, y)$ continua, teniendo en cuenta que la función de densidad marginal $f_Y(\theta, y) = \int_{-\infty}^{\infty} f_{X,Y}(\theta, x, y)dy$ también será continua, si se verifica que $f_Y(\theta, y) \neq 0$, será:

$$p(y_0 \leq y \leq y_0 + h) = \int_{y_0}^{y_0+h} f_Y(\theta, u)du \neq 0 \qquad (h > 0)$$

De modo que la probabilidad condicionada puede ser definida de la siguiente forma:

$$p[(X \leq x)/(y_0 \leq Y \leq y_0 + h)] = \frac{p\big((X \leq x) \cap (y_0 \leq Y \leq y_0 + h)\big)}{p(y_0 \leq Y \leq y_0 + h)} =$$

$$= \frac{\int_{y_0}^{y_0+h}\left[\int_{-\infty}^{x} f_{X,Y}(\theta, u, v)du\right]dv}{\int_{y_0}^{y_0+h} f_Y(\theta, y)dy} = \frac{\frac{1}{h}\int_{y_0}^{y_0+h}\left[\int_{-\infty}^{x} f_{X,Y}(\theta, u, v)du\right]dv}{\frac{1}{h}\int_{y_0}^{y_0+h} f_Y(\theta, y)dy}$$

Luego, será:

$$\lim_{h \to 0} p[(X \leq x)/(y_0 \leq Y \leq y_0 + h)] = \int_{-\infty}^{x} \frac{f_{X,Y}(u, y)}{f_Y(\theta, y)} du$$

Donde la integral que figura en el segundo miembro define indudablemente una función de distribución de probabilidades correspondiente a una variable continua y el integrando la función de densidad de la variable aleatoria X, condicionada por el supuesto que $Y = y$, $f(x/y) = \frac{f_{X,Y}(\theta,x,y)}{f_Y(\theta,y)}$.

Si se verifica que:

$$p[(X > x) \;/\; (Y = y)] = 1 - p[(X \leq x) \;/\; (Y = y)] =$$
$$= 1 - F_{X/Y}(x \;/\; y) = S_{X/Y}(x \;/\; y) \uparrow y$$

es decir, si se verifica que es más probable que X asuma valores más grandes cuando Y crece, se dice que X es "estocásticamente creciente" respecto de Y ($X \uparrow_{est} Y$) o que $F_{X/Y}(x \;/\; y)\big(y \epsilon \Omega(Y)\big)$ es "estocásticamente creciente"[14].

En forma similar, si se verifica que $p[(X > x) \;/\; (Y = y)] = 1 - F_{X/Y}(x \;/\; y) \downarrow y$, se dice que X es "estocásticamente decreciente" respecto de Y ($X \downarrow_{est} Y$).

En general, dado un conjunto de variables aleatorias, $X_1, X_2, \dots, X_n$, si se verifica que X_i es estocásticamente creciente respecto de $X_1, X_2, \dots, X_{i-1}$ $(i = 1,2,\dots,n)$ es decir, si $p\big[(X_i > x_i) \;/\; \big(X_j = x_j\big)\big]$ $(i = 1,2,\dots,n$) es creciente respecto de $X_1, X_2, \dots, X_{i-1}$ $(i = 1,2,\dots,n)$, se dice que el vector es "condicionadamente creciente en sucesión"[15]. A partir de esta definición se puede concluir en forma inmediata que, para $n = 2$, la propiedad de condicionalidad creciente es equivalente a $X \uparrow_{est} Y$.

Sea una variable aleatoria bidimensional (X, Y) con función de distribución conjunta $F_{X,Y}(x, y)$, si se verifica que:

$$p[(X > x) \;/\; (Y = y)] = S_{X/Y}(x, y) = \frac{S_{X,Y}(x, y)}{S_Y(y)} \uparrow y$$

es decir, si es más probable que X asuma valores más grandes en la medida que Y crece, se dice que X es "creciente por la 'cola' de la derecha" respecto de Y (de acuerdo con la nomenclatura de Esary; Proschan (1972)). En forma similar, si se verifica que:

$$p[(X \leq x) \;/\; (Y = y)] = F_{X/Y}(x, y) = \frac{F_{X,Y}(x, y)}{F_Y(y)} \downarrow y$$

es decir, si es más probable que X asuma valores más pequeños en la medida que Y decrece, se dice que X es "decreciente por la 'cola' de la izquierda" respecto de Y[16].

---

[14] Esta condición –que es más fuerte que la de Kimeldorf-Sampson (ver Seccción 8.1) - se denomina también como de "dependencia de regresión positiva" (Lehmann (1966), Yanagimoto; Okamoto (1969), Schriever(1986)(1987), Fang;Joe(1992)).

[15] Como se verá en la Sección 10.2.5, se demuestra que esta condición implica la asociación positiva.

[16] Ver Caperaá; Genest (1990).

En general, dada una variable aleatoria n-dimensional $X_1, X_2, \dots, X_n$, si se verifica que $p\left[(X_i > x_i) \,/\, (X_j = x_j)\right] \uparrow x_k \ (i\epsilon\bar{A}, j, k\epsilon A)$ (donde A denota un subconjunto no-vacío del conjunto $\{1,2,\dots,n\}$) y $p\left[(X_i \le x_i) \,/\, (X_j = x_j)\right] \downarrow x_k \ (i\epsilon\bar{A}, j, k\epsilon A)$, se obtienen las respectivas extensiones de las condiciones anteriores.

Todas estas condiciones son invariantes con respecto a cualquier transformación estrictamente creciente de las variables marginales.

5.- Una introducción a las cópulas

Sean: **i)** un vector de variables aleatorias, $X = [X_1, X_2, \dots, X_n]$, todas con distribución $U(0,1)$; **ii)** $F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)$ sus correspondientes funciones de distribución y **iii)** $F(\cdot)$ una función de distribución n-dimensional para la que las variables marginales $F_{X_j}(x_j)$ $(j = 1,2,\dots,n)$, son tales que $F(\boldsymbol{x})\epsilon\mathcal{F}\left[F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)\right]$. Se denomina "cópula asociada con $F(\cdot)$" a la función de distribución $C(\cdot)$ que satisface en forma unívoca la relación $F(x) = C\left[F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)\right]$.

Ahora bien, se demuestra que: **i)** Si $H(\cdot)$ denota una función de distribución unidimensional con función de distribución inversa $H^{-1}(\cdot)$, entonces $H^{-1}(U)$ se distribuye como $H(U)$ y **ii)** Si $H(\cdot)$ es una función de distribución continua unidimensional y la variable X se distribuye como H, entonces $H(X)$ se distribuye de acuerdo con una función uniforme $U(0,1)$. Es decir, si $\boldsymbol{X}$ se distribuye como $F(\cdot)$ y $F(\cdot)$ es una función continua, entonces $\left[F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)\right]$ se distribuye como $C(\cdot)$ y si U se distribuye como $C(\cdot)$, entonces $\left[F_{X_1}^{-1}(x_1), F_{X_2}^{-1}(x_2), \dots, F_{X_n}^{-1}(x_n)\right]$ se distribuye como $F(\cdot)$ (donde $F^{-1}(\cdot)$ denota la inversa de la función de distribución $F(\cdot)$).

Luego, a partir de estas propiedades se puede concluir que, si $F(\cdot)$ es una función de distribución n-dimensional continua con variables marginales $F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)$ y funciones inversas $F_{X_1}^{-1}(x_1), F_{X_2}^{-1}(x_2), \dots,$ $F_{X_n}^{-1}(x_n)$, entonces será:

$$C(u) = C(u_1, u_2, \dots, u_n) = F\left[F_{X_1}^{-1}(x_1), F_{X_2}^{-1}(x_2), \dots, F_{X_n}^{-1}(x_n)\right]$$

Este teorema fundamental, debido a Sklar (1959), define el rol que juegan las cópulas en la relación entre las funciones de distribución multidimensionales y las correspondientes a sus variables marginales. Establece que en el contexto de vectores aleatorios continuos, es posible construir una función de distribución n-dimensional en términos de n funciones continuas unidimensionales y una cópula que permite relaciones de dependencia entre variables aleatorias individuales[17].

Dado que una cópula $C(\cdot)$ es la función de distribución de un vector aleatorio $U = [U_1, U_2, \dots, U_n]$ en el cual las variables marginales $U_j: U(0,1)$ $(j = 1,2, \dots, n)$, se puede concluir que es una función continua.

Sea en particular una variable bidimensional $Z = (X, Y)$ cuyas variables marginales se distribuyen de acuerdo a una función $U(0,1)$, entonces será $C(u, v)$=$F[F_X^{-1}(x), F_Y^{-1}(y)]$ $\big((u, v)\epsilon \times_2 [0,1]\big)$. Esta cópula asociada a la función $F_{X,Y}(x, y)$ posee las siguientes propiedades: **i)** $C(u, 0)$=$C(0, v) = 0$; **ii)** $C(u, 1) = u$ y $C(1, v) = v$; **iii)** para todo $u_1 \leq u_2$ y $v_1 \leq v_2$:

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

y **iv)** para todo $v\epsilon[0,1]$ y casi todo u se verifica que $0 \leq \frac{\partial}{\partial u} C(u, v) \leq 1$ y, para todo $u\epsilon[0,1]$ y casi todo v resulta que $0 \leq \frac{\partial}{\partial v} C(u, v) \leq 1$[18].

Ejemplo n° 7

Sea $Z = (X, Y)$ una variable bidimensional con función de distribución:

$$F_{X,Y}(x, y) = exp\left\{-\left[e^{-x} + e^{-y} - \left(e^{\delta x} + e^{\delta y}\right)^{-1/\delta}\right]\right\}$$

$(\delta > 0, -\infty < x, y < \infty)$.

Entonces, será:

$$\lim_{y \to \infty} F_{X,Y}(x, y) = F_X(x) = exp(-e^{-x})$$

$$\lim_{x \to \infty} F_{X,Y}(x, y) = F_Y(y) = exp(-e^{-y})$$

Aplicando la sustitución $u = F_X(x)$ y $v = F_Y(y)$ o $x = -log[-log(u)]$ e $y = -log[-log(v)]$, se obtiene la cópula:

$$C(u, v) = uv\ exp\left\{\left[\left(-log(u)\right)^{-\delta} + \left(-log(v)\right)^{-\delta}\right]^{-1/\delta}\right\}$$

Ejemplo n° 8

Sea $Z = (X, Y)$ una variable bidimensional con función de distribución:

---

[17] Ver Schweizer (1974) y Sklar (1973)(1996).
[18] Ver Seeley (1961).

$$F_{X,Y}(x,y) = \begin{cases} \dfrac{(x+1)(e^y-1)}{x+2e^y-1} & (x,y)\epsilon[-1,1]\times[0,\infty] \\ 1-e^{-y} & (x,y)\epsilon[-1,\infty]\times[0,\infty] \\ 0 \quad para\ cualesquiera\ otros\ valores\ de\ x\ e\ y \end{cases}$$

y funciones de distribución marginales:

$$F_X(x) = \begin{cases} 0 & (x<-1) \\ \dfrac{x+1}{2} & x\epsilon[-1,1] \\ 0 \quad para\ culauier\ otro\ valor\ de\ x \end{cases}$$

$$F_Y(y) = \begin{cases} 0 & (y<0) \\ 1--e^{-y} & (y\geq 0) \end{cases}$$

Las correspondientes inversas de estas funciones de distribución son de la forma $F_X^{-1}(u) = 2u-1$ y $F_Y^{-1}(v) = -ln(1-v)$ $(u,v\epsilon[,1])$. Como $Ran(F_X) = Ran(F_Y) = [0,1]$, será:

$$C(u,v) = \frac{uv}{u+v-uv}$$

Ejemplo n° 9

Sea $Z = (X,Y)$ una variable bidimensional con función de distribución:

$$F_{X,Y}(x,y) = \begin{cases} 1-e^{-x}-e^{-y}+e^{-(x+y+\theta xy)} & (x,y\geq 0, \theta\epsilon[0,1]) \\ 0 \quad para\ cualesquiera\ otros\ valores\ de\ x\ e\ y \end{cases}$$

y funciones de distribución marginales $F_Y(y) = 1-e^{-y}$ $(y\geq 0)$ y $F_X(x) = 1-e^{-x}$ $(x\geq 0)$. Las funciones inversas son de la forma $F_X^{-1}(u) = -ln(1-u)$ y $F_Y^{-1}(v) = -ln(1-v)$ (para $u,v\epsilon[0,1]$). Por lo tanto, la cópula correspondiente será:

$$C(u,v) = u+v-1+(1-u)(1-v)e^{-\theta ln(1-u)ln(1-v)}$$

Sea $\boldsymbol{X} = (X_1,X_2,\ldots,X_n)$ una variable n-dimensional cuyas variables marginales son discretas, $\Omega(X_i) = \{X_{i,1},X_{i,2},\ldots,X_{i,m_i}\}$ $(i=1,2,\ldots,n)$ y sean $p(X_i=x_i)$ y $F_{X_i}(x_i)$ las correspondientes funciones de probabilidades y de distribución de las variables marginales y $p(\boldsymbol{X})$ y $F_{\boldsymbol{X}}(\boldsymbol{x})$ las funciones de probabilidades y de distribución de la variable n-dimensional. Una cópula asociada a la función $p(\boldsymbol{X})$ satisface la siguiente relación:

$$p(\boldsymbol{X}) = p(X_1, X_2, \dots, X_n) = C\left[F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)\right]$$

Para que se verifique la unicidad de esta relación es condición necesaria que los restantes valores de $C(\cdot)$ se distribuyan uniformemente en cada rectángulo

$\times_{i=1}^{n}\left[F_{X_{i,j-1}}(x_{i,j-1}) - F_{X_{i,j}}(x_{i,j})\right]$.

6.- La independencia estocástica de las variables aleatorias

Se dice que las variables X e Y son independientes ($X \perp Y$ en la nomenclatura de Dawid (1979)) si y sólo si se verifica la relación $p\left[(X = x_i)/(Y = y_j)\right] = p_{i.}(i, j, = 1,2, \dots)$. Es decir, si el supuesto que Y asume el valor y_j no modifica la asignación de probabilidades de que la variable aleatoria X asuma uno cualquiera de sus posibles valores.

En general la condición de independencia puede ser definida como $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ y, dado que la función de distribución determina en forma unívoca la función de probabilidades de la variable aleatoria, a partir de esta expresión se puede escribir:

$$p[(X\epsilon A_X) \cap (Y\epsilon A_Y)] = p(X\epsilon A_X)p(Y\epsilon A_Y) \qquad \left(A_X\epsilon\Omega(X), A_Y\epsilon\Omega(Y)\right)$$

(es decir, que los eventos relativos a la variable X son independientes de los eventos relativos a la variable Y). Este concepto es generalizable a todo conjunto $\{X_1, X_2, \dots, X_n\}$ de variables aleatorias:

$$F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \dots F_{X_n}(x_n)$$

e, igual que en el caso bidimensional, esta expresión implica que:

$$p[(X_1\epsilon A_1) \cap (X_2\epsilon A_2) \cap \dots \cap (X_n\epsilon A_n)] =$$

$$= p(X_1\epsilon A_1)p(X_2\epsilon A_2) \dots p(X_n\epsilon A_n)$$

$(A_1\epsilon\Omega(X_1), A_2\epsilon\Omega(X_2), \dots, A_n\epsilon\Omega(X_n))$.

De esta relación se puede concluir que –contrariamente a lo que ocurre con la propiedad de independencia para n eventos- si n variables aleatorias son independientes, entonces $k\ (\leq n)$ cualesquiera de ellas también son independientes.

En particular, si la función de distribución de la variable $(X_1, X_2, \dots, X_n)$ es absolutamente continua, la condición necesaria y suficiente para la independencia de las n variables está dada por:

$$f(\theta, x_1, x_2, \dots, x_n) = f(\theta, x_1) f_{X_2}(\theta, x_2) \dots f_{X_n}(\theta, x_n)$$

Ejemplo n° 10:

Sea (X, Y) una variable aleatoria bidimensional cuya función de densidad conjunta está definida de la siguiente forma:

$$f_{X,Y}(x, y) =$$

$$= \begin{cases} \dfrac{2}{(e^3 - 1)(e^4 - 1)} e^{2x+y} \quad para \quad 0 \leq x \leq 2, 0 \leq y \leq 3 \\ 0 \quad para\ cualesquiera\ otros\ valores\ de\ x\ e\ y \end{cases}$$

Las correspondientes funciones de densidad marginales serán:

$$f_X(x) = \frac{2e^{2x}}{(e^3 - 1)(e^4 - 1)} \int_0^3 e^y dy = \frac{2}{e^4 - 1} e^{2x} \qquad (0 \leq x \leq 2)$$

$$f_Y(y) = \frac{2e^{2y}}{(e^3 - 1)(e^4 - 1)} \int_0^2 e^{2x} dx = \frac{1}{e^3 - 1} e^y \quad (0 \leq y \leq 3)$$

A partir de las cuales se obtiene que:

$$f_X(x) f_Y(y) = \frac{2}{(e^3 - 1)(e^4 - 1)} e^{2x+y} = f_{X,Y}(x, y)$$

De lo que se concluye que las variables X e Y son independientes.

Ejemplo n° 11:

Sea (X, Y) una variable aleatoria bidimensional cuya función de densidad conjunta está definida de la siguiente forma:

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1 - \rho}} =$$

$$exp\left\{-\frac{1}{2(1 - \rho^2)}\left[\left(\frac{x - m_X}{\sigma_X}\right)^2 - 2\rho\frac{x - m_X}{\sigma_X}\frac{y - m_Y}{\sigma_Y} + \left(\frac{y - m_Y}{\sigma_Y}\right)^2\right]\right\}$$

$(-\infty \leq x \leq +\infty, -\infty \leq y \leq +\infty,$ donde $m_X,\ m_Y,\ \sigma_X,\ \sigma_Y$ y $|\rho| \leq 1$ son constantes reales). El exponente de esta función puede ser expresado como:

$$\left(\frac{x-m_X}{\sigma_X}\right)^2 - 2\rho\frac{x-m_X}{\sigma_X}\frac{y-m_Y}{\sigma_Y} + \left(\frac{y-m_Y}{\sigma_Y}\right)^2 =$$

$$= \left(\frac{x-m_X}{\sigma_X}\right)^2 - 2\rho\frac{x-m_X}{\sigma_X}\frac{y-m_Y}{\sigma_Y} + \rho^2\left(\frac{y-m_Y}{\sigma_Y}\right)^2 - \rho^2\left(\frac{y-m_Y}{\sigma_Y}\right)^2 =$$

$$+\left(\frac{y-m_Y}{\sigma_Y}\right)^2 = \left(\frac{x-m_X}{\sigma_X}\frac{y-m_Y}{\sigma_Y}\right)^2 + (1-\rho^2)\frac{y-m_Y}{\sigma_Y}$$

Luego, la definición de la función de densidad marginal correspondiente a la variable Y asumirá la forma:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$

$$exp\left[-\frac{1}{2}\left(\frac{y-m_Y}{\sigma_Y}\right)^2\right]\int_{-\infty}^{\infty} exp\left[-\frac{1}{2(1-\rho^2)\sigma_X^2}\left(x-m_X-\rho\sigma_X\frac{y-m_Y}{\sigma_Y}\right)^2\right]dx$$

$$= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_X exp\left[-\frac{1}{2}\left(\frac{y-m_Y}{\sigma_Y}\right)^2\right] =$$

$$= \frac{1}{\sigma_Y\sqrt{2\pi}}\ exp\left[-\frac{1}{2}\left(\frac{y-m_Y}{\sigma_Y}\right)^2\right]$$

Con un razonamiento similar, se obtiene que:

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}}\ exp\left[-\frac{1}{2}\left(\frac{x-m_X}{\sigma_X}\right)^2\right]$$

A partir de las definiciones anteriores, se demuestra en forma inmediata que la función de densidad de la variable X condicionada por la variable Y será:

$$f(x\ /\ y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} =$$

$$= \frac{1}{\sigma_X\sqrt{2\pi}\sqrt{1-\rho^2}} exp\left[-\frac{1}{2\sigma_X^2(1-\rho^2)}\left(x-m_X-\rho\sigma_X\frac{y-m_Y}{\sigma_Y}\right)^2\right]$$

Expresión que permite concluir que las variables marginales X e Y pueden ser consideradas independientes sólo si $\rho = 0$.

Ejemplo n° 12:

Sea una urna de la que se sabe que contiene bolillas blancas y de otros colores en una proporción desconocida, sea $X\ (0 \leq X \leq 1)$ la variable aleatoria (continua) que representa la proporción de bolillas blancas y sea $f_X(x)$ su función de densidad.

Supóngase que de dicha urna se efectúen n extracciones al azar con reposición, considerando como éxito obtener una bolilla blanca y sea Z la variable aleatoria discreta que representa el número de éxitos a obtener en las n extracciones. Entonces, será:

$$p[(Z = z)/(X = x)] = \binom{n}{z} x^z (1-x)^{n-z} \qquad (z = 0,1,2,\dots,n)$$

Supóngase que se consideran más probables las composiciones de la urna con aproximadamente la mitad de bolillas blancas y poco probables (en forma simétrica) aquellas composiciones con casi todas o casi ninguna bolilla blanca y, en consecuencia, se asigne a la variable X (continua), por ejemplo, una función de densidad de la forma $f_X(x) = 6x(1-x)(0 \leq x \leq 1)$ (cuyo máximo se encuentra en $x = \frac{1}{2}$ y es simétrica con respecto a ese valor). La función de probabilidades conjunta de la variable bidimensional (X, Z) será, en este caso:

$$p[(X = x) \cap (Z = z)] = 6\binom{n}{z} x^{z+1}(1-x)^{n-z+1}$$

Esta variable (X, Z) posee una componente continua y una discreta. La distribución de probabilidades de la variable marginal Z queda definida, entonces, por la expresión:

$$p(Z = z) = \int_0^1 p[(X = x) \cap (Z = z)]dx = 6\binom{n}{z}\int_0^1 x^{z+1}(1-x)^{n-z+1}x =$$

$$= 6\binom{n}{z}\frac{(n-z+1)!\,(z+1)!}{(n+3)!} = \frac{6(z+1)(n-z+1)}{(n+3)(n+2)(n+1)}$$

$(z = 0,1,2,\dots,n)$.

De esta expresión se puede concluir que, a partir de la distribución de las posibles composiciones de la urna, la probabilidad de obtener, por ejemplo, una bolilla blanca de una extracción al azar, será $p(Z = 1) = \frac{1}{2}$. De la misma forma, si se supone, por ejemplo, que todas las composiciones de la urna son igualmente probables, es decir, si se supone una distribución para la variable X de la forma:

$$f_X(x) = \begin{cases} 1 & para \quad 0 \leq x \leq 1 \\ 0 & para\ cualquier\ otro\ valor\ de\ x \end{cases}$$

se obtiene, entonces, que los valores de la variable Z son equiprobables: $p(Z = z) = \frac{1}{n+1} (z = 0,1,2, \dots, n)$.

Sean X, Y y Z tres variables aleatorias. Se dice que X e Y son condicionadamente independientes supuesto un valor dado de Z ($X \perp Y/Z$), si se verifica que:

$$p\{(X = x_i)/[(Y = y_j) \cap (Z = z_h)]\} = p[(X = x_i)/(Z = z_h)]$$

Es decir, si dado el supuesto que Z ha asumido un valor z_h, cualquier supuesto sobre el comportamiento de Y no modifica la asignación de probabilidades de que la variable aleatoria X asuma uno cualquiera de sus posibles valores.

7.- Independencia condicionada y teoría de la información

De acuerdo con Dawid (1979) y Spohn (1980), la relación de independencia condicionada posee las siguientes propiedades:

1) Dado un estado condicionante Z, si un supuesto comportamiento de Y no agrega información sobre el comportamiento de X, entonces se puede asegurar que X no agregará información sobre el comportamiento de Y, es decir, si se verifica que ($X \perp Y/Z$), entonces también se verificará que ($Y \perp X/Z$).

2) Dadas cuatro variables aleatorias X, Y, Z y W, si se verifica que para un estado condicionante Z, el supuesto comportamiento de las variables Y y W no agrega información sobre el comportamiento de X, entonces se puede asegurar que cualquier supuesto sobre el comportamiento de cualquiera de las dos variables tomadas individualmente, no agregará información sobre el comportamiento de X, es decir, si se verifica que $[X \perp (Y \cap W)/Z]$, entonces se verificará que ($X \perp Y/Z$).

3) Dados un estado condicionante Z y dos variables, Y y W, ninguna de las cuales agrega información sobre el comportamiento de X, la incorporación de un supuesto sobre el comportamiento de una de estas variables, por ejemplo W, al conjunto original Z, no logrará que el supuesto sobre la otra variable agregue información sobre el comportamiento de X, es decir, si se verifica que $[X \perp (Y \cap W)/Z]$, entonces se verificará que $[X \perp Y/(Z \cap W)]$.

4) Dados un estado condicionante Z y una variable Y, la cual no agrega información sobre el comportamiento de X, si la incorporación de una variable W al conjunto $(Z \cap Y)$ no agrega información sobre el comportamiento de X, entonces se puede asegurar que la incorporación de cualquier supuesto sobre la variable W al conjunto original Z, tampoco agregará información sobre el comportamiento de X, es decir, si se verifican simultáneamente las condiciones $(X \perp Y/Z)$ y $[X \perp W/(X \cap W)]$, entonces se verificará que $[X \perp (Y \cap W)/Z]$.

5) Dado un estado de conocimiento Z, si la incorporación al conjunto (Z, W) de un supuesto sobre una variable Y no agrega información sobre el comportamiento de X y, si la incorporación al conjunto $(Z \cap Y)$ de un supuesto sobre la variable W no agrega información sobre el comportamiento de X, entonces se puede asegurar que ni la variable Y, ni la variable W, ni cualquier combinación de ellas, agregará información sobre el comportamiento de X, es decir, si se verifican simultáneamente las condiciones $[X \perp W/(Z \cap Y)]$ y $[X \perp Y/(Z \cap W)]$, entonces se verificará que $[X \perp (Y \cap W)\ /\ Z]$.

Obsérvese que las propiedades **3)** y **4)** establecen que la incorporación de una variable irrelevante a un conjunto de información dado no modifica el status del resto de las variables que componen el mismo. Es decir, si una variable de dicho conjunto es irrelevante, seguirá siendo irrelevante.

8.- Una introducción a la dependencia estocástica

8.1.- Definiciones

Sea una variable aleatoria bidimensional $Z = (X, Y)$ con funciones de distribución marginales, $F_X(x)$ y $F_Y(y)$, conocidas y función de distribución conjunta $F_{X,Y}(x, y)$ que posee cotas de Fréchet superior, $F_{X,Y}^{(m)}(x, y) = min[F_X(x), F_Y(y)]$, e inferior, $F_{X,Y}^{(M)}(x, y) = Max[0, F_X(x) + F_Y(y) - 1]$. Si se verifica que $F_{X,Y}(x, y) \geq F_X(x)F_Y(y)$ o, lo que es lo mismo, que:

$$\frac{F_{X,Y}(x, y)}{F_X(x)} = F[y\ /\ (X \leq x)] \geq F_Y(y)$$

es decir, si se verifica que la probabilidad de que las variables X e Y asuman sus valores más grandes –o los más pequeños- en forma conjunta, es mayor que la probabilidad de que se verifique la misma condición si las variables fueran independientes entre sí, se dice que las variables X e Y pertenecen a la familia de variables con "dependencia de cuadrante positiva" (en términos compactos $(X,Y) \in \mathcal{F}^{(+)}$), condición conocida como "de Kimeldorf-Sampson"[19].

Esta propiedad es equivalente a:

$$1 - F_{X,Y}(x,y) = S_{X,Y}(x,y) \geq [1 - F_X(x)]\big[[1 - F_Y(y)]\big] = S_X(x)S_Y(y)$$

Es decir, es equivalente a $S_{Y/X}(y \;/\; X \geq x) \geq S_Y(x)$.

A partir de estas definiciones se puede concluir que: **i)** como ya se vio en la Sec.6, dada una función $F_{X,Y}(x,y)$ que satisface la condición de Kimeldorf-Sampson, si $X \perp Y$, se verifica que $F_{X,Y}(x,y) = F_X(x)F_Y(y)$; **ii)** si $(X,Y) \in \mathcal{F}^{(+)}$ entonces, dada una función $g(\cdot)$ creciente, se verifica que $[g(X),Y]\epsilon\mathcal{F}^{(+)}$; **iii)** si $(X,Y)\epsilon\mathcal{F}^{(+)}$, entonces $(Y,X)\epsilon\mathcal{F}^{(+)}$; **iv)** si $(X,Y)\epsilon\mathcal{F}^{(+)}$, entonces $(-X,-Y)\epsilon\mathcal{F}^{(+)}$.

En forma similar, si se verifica que $F_{X,Y}(x,y) \leq F_X(x)F_Y(y)$ o $S_{X,Y}(x,y) \leq S_X(x)S_Y(y)$, se dice que X e Y pertenecen a la familia de variables con "dependencia de cuadrante negativa", $(X,Y)\epsilon\mathcal{F}^{(-)}$.

Generalizando estas definiciones al caso n-dimensional, sea una variable aleatoria $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ si se verifica que:

$$S_{X_1,X_2,\dots,X_n}(x_1,x_2,\dots,x_n) \geq S_{X_1}(x_1)S_{X_2}(x_2)\dots S_{X_n}(x_n)$$

(es decir, si se verifica que la probabilidad de que las variables $X_1, X_2, \dots, X_n$ asuman simultáneamente los valores más grandes es mayor que la correspondiente a un vector $X^{(I)} = \left[X_1^{(I)}, X_2^{(I)}, \dots, X_n^{(I)}\right]$ de variables iid), se dice que entre las variables marginales existe "dependencia positiva de ortante superior".

Si se verifica que:

$$F_{X_1,X_2,\dots,X_n}(x_1,x_2,\dots,x_n) \geq F_{X_1}(x_1)F_{X_2}(x_2)\dots F_{X_n}(x_n)$$

[19] Kimeldorf; Sampson (1975a)(1975b)(1978).

se dice que entre las variables X_i $(i = 1,2, \dots, n)$ existe "dependencia positiva de ortante inferior" y, si se verifican estas dos relaciones simultáneamente, se dice que las variables X_i $(i = 1,2, \dots, n)$ existe "dependencia de ortante positiva".

En forma similar, si se verifica que:

$$S_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) \leq S_{X_1}(x_1)S_{X_2}(x_2) \dots S_{X_n}(x_n)$$

se dice que entre las variables marginales existe "dependencia negativa de ortante inferior". Si:

$$F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) \geq F_{X_1}(x_1)F_{X_2}(x_2) \dots F_{X_n}(x_n)$$

se dice que entre las variables X_i $(i = 1,2, \dots, n)$ existe "dependencia negativa de ortante superior" y, si se verifican estas dos relaciones simultáneamente, se dice que entre las variables X_i $(i = 1,2, \dots, n)$ existe "dependencia de ortante negativa"[20].

Por otra parte, si se verifica que la variable $\left(X_i \,/\, X_j\right)$ $(i \neq j)$ es tal que $(X_i \uparrow_{est} x)(i = 1,2, \dots, n)$, se dice que la variable $(X_1, X_2, \dots, X_n)$ observa un "orden de dependencia estocástica positiva".

De esta definición se puede concluir en forma inmediata que, para $n = 2$, el orden de dependencia positiva es equivalente al cumplimiento de las condiciones $X \uparrow_{est} Y$ y $Y \uparrow_{est} X$.

Se demuestra que este orden de dependencia implica las dependencias de ortante positivas[21].

8.2.- Las medidas de dependencia

Hasta aquí el concepto de dependencia sobre dos variables se basó en la comparación de las funciones de distribución bidimensionales, $F_{X,Y}(x, y)$, con el producto $F_X(x)F_Y(y)$, obtenido a partir del supuesto que $X \perp Y$. Una forma de cuantificar la intensidad de la dependencia consiste en comparar funciones de distribución bidimensionales utilizando los coeficientes τ de Kendall (1938) y ϱ_S de Spearman (1904), los cuales son: **i)** invariantes con respecto a transformaciones estrictamente crecientes y **ii)** iguales a 1 para la cota

[20] Ver Block; Ting (1981), Block; Savits; Shaked (1985).

[21] Para un tratamiento más riguroso de estos conceptos, ver Joe (2001).

superior de Fréchet (es decir cuando una variable es una transformación creciente de la otra) e iguales a -1 para la cota inferior de Fréchet (es decir cuando una variable es una transformación decreciente de la otra): $-1 \leq \tau \leq 1$ y $-1 \leq \varrho_s \leq 1$[22]

Dadas dos variables aleatorias bidimensionales (X, Y) y (X^*, Y^*) independientes entre sí con funciones de distribución $F(\cdot)$ el coeficiente τ de Kendall considera la diferencia entre la probabilidad de dos pares aleatorios concordantes (es decir de aquellos pares (X, Y) y (X^*, Y^*) tales que uno de los dos posee el valor más grande para ambos componentes, en otros términos, tales que, $X > X^*$ e $Y > Y^*$ o $X < X^*$ e $Y < Y^*$o lo que es lo mismo, que, $(X - X^*)(Y - Y^*) > 0$) y la probabilidad de dos pares aleatorios discordantes (es decir de aquellos pares (X, Y) y (X^*, Y^*) en los que para cada par un componente es más grande que el correspondiente componente del otro par y uno es más chico, en otros términos, tales que $X > X^*$ e $Y < Y^*$ o $X < X^*$ e $Y > Y^*$ y, por lo tanto, que $(X - X^*)(Y - Y^*) < 0$). Como se mencionó más arriba, si una variable es una transformación creciente de la otra, la ocurrencia de un par concordante es cierta y la ocurrencia de un par discordante es imposible y, viceversa, si una variable es una transformación decreciente de la otra, la ocurrencia de un par concordante es imposible y la ocurrencia de un par discordante es cierta. Formalmente este criterio de medición asume la forma:

$$\tau = p[(X - X^*)(Y - Y^*) > 0] - p[(X - X^*)(Y - Y^*) < 0] =$$

$$= 2p[(X - X^*)(Y - Y^*) > 0] - 1 = 4 \int F(u)\, dF(u) - 1$$

Por otra parte, sea una variable bidimensional continua con función de distribución $F_{X,Y}(x, y)$. Bajo la condición de que las variables marginales se distribuyen de acuerdo con una función $U(0,1)$ con parámetros $E(X) = E(Y) = \frac{1}{2}$ y $\sigma^2(X) = \sigma^2(Y) = \frac{1}{12}$, el coeficiente ϱ_S de Spearman se define como la medida de correlación entre las variables $F_X(x)$ y $F_Y(y)$:

$$\varrho_S = 12 \int \int F_X(x) F_Y(y)\, dF_{X,Y}(x, y) - 3 =$$

$$= 12 \int \int S_{X,Y}(x, y)\, dF_X(x) dF_Y(y) - 3$$

[22] Ver Nelsen (2001).

9.- Funciones de variables aleatorias

Dada una variable aleatoria X, se puede definir una función $K(X)$ -que también es una variable aleatoria- obtenida de aplicar una transformación K a cada uno de los valores de su dominio. La distribución de probabilidades de la variable $Y = K(X)$ está dada por la función:

$$F_Y(y) = p(Y \leq y) = p[K(X) \leq y] = p[X \leq K^{-1}(y)] =$$

$$= \int_{K(X)\leq y} dF_X(x) = \int_{K(X)\leq y} f_X(x)dx$$

Supóngase el caso más simple en el que las variables aleatorias X e Y sean unidimensionales e Y esté definida como una transformación lineal de la forma $K(X) = aX + b (a > 0)$. Se verificará, entonces, que:

$$F_Y(y) = p(Y \leq y) = \int_{aX+b} dF_X(x) = P\left(x \leq \frac{y-b}{a}\right) =$$

$$= \int_{x\leq\frac{y-b}{a}} dF_X(x) = \int_{-\infty}^{\frac{y-b}{a}} dF_X(x) =$$

$$= F_X\left(\frac{y-b}{a}\right) = \int_{-\infty}^{\frac{y-b}{a}} f_X(x)dx = \frac{1}{a}\int_{-\infty}^{y} f_X\left(\frac{x-b}{a}\right)dx$$

Para $a < 0$, se tiene que:

$$F_Y(y) = 1 - F_X\left(\frac{y-b}{a}\right) = \int_{\frac{y-b}{a}}^{\infty} f_X(x)dx = \frac{1}{a}\int_{y}^{\infty} f_X\left(\frac{x-b}{a}\right)dx$$

$$= \frac{1}{|a|}\int_{-\infty}^{y} f_X\left(\frac{x-b}{a}\right)dx$$

Se puede concluir entonces que, si $F_X(x)$ es absolutamente continua y $a \neq 0$, entonces $F_Y(y)$ también será absolutamente continua, con función de densidad $f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$.

Esta propiedad puede ser generalizada para toda transformación $K(X)$ estrictamente monótona con derivada continua. Si $K(X)$ es creciente, se verificará que:

$$F_Y(y) = \int_{K(X)\leq y} f_X(x)dx = \int_{-\infty}^{K^{-1}(X)} f_X(x)dx = \frac{d}{dx}K^{-1}(x) \int_{-\infty}^{y} f_X[K^{-1}(x)]dx$$

Si $K(X)$ es decreciente, igual que en el caso lineal, será:

$$F_Y(y) = 1 - F_X[K^{-1}(y)] = \int_{K^{-1}(y)}^{\infty} f_X(u)du$$

$$= \left|\frac{d}{dy}K^{-1}(y)\right| \int_{-\infty}^{y} f_X[K^{-1}(u)]du$$

Es decir, se demuestra que, si X es absolutamente continua, entonces $Y = K(X)$ también es absolutamente continua con función de densidad de la forma:

$$f_Y(y) = \left|\frac{d}{dy}K^{-1}(y)\right| f_X[K^{-1}(y)]$$

Ejemplo n° 13:

Sea X una variable aleatoria unidimensional con función de densidad de la forma:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right\} \qquad (-\infty \leq x \leq \infty)$$

(donde m y σ son constantes) y sea una variable Y definida a partir de la aplicación de la transformación lineal $Y = K(X) = aX + b$ $(a > 0)$. La función de densidad de Y estará dada, entonces, por:

$$f_Y(y) = \frac{1}{a}\frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{y-b-am}{a\sigma}\right)^2\right] \qquad (-\infty \leq y \leq \infty)$$

Ejemplo n° 14:

Sea X una variable aleatoria unidimensional con función de densidad de la forma:

$$f_X(x) = \begin{cases} 2x & para \qquad 0 \leq x \leq 1 \\ 0 & para\ cualquier\ otro\ valor\ de\ x \end{cases}$$

y supóngase la transformación $Y = -X$. Se verificará, entonces, que:

$$F_Y(y) = p(-X \leq y) = p(X \geq -y) = 1 - p(X < -y) = 1 - F_X(-y)$$

donde:

$$F_X(-y) = \begin{cases} 0 & para \quad y > 0 \\ y^2 & \quad -1 \leq y \leq 0 \\ 1 & \quad y > -1 \end{cases}$$

Luego, será:

$$F_Y(y) = 1 - F_X(-y) = \begin{cases} 1 & para \quad y > 0 \\ 1 - y^2 & \quad -1 \leq y \leq 0 \\ 0 & \quad y < -1 \end{cases}$$

De modo que la función de densidad de la variable Y estará dada por:

$$f_Y(y) = \begin{cases} -2y & para \qquad -1 \leq y \leq 0 \\ 0 & para\ cualquier\ otro\ valor\ de\ y \end{cases}$$

Ejemplo n° 15:

Sea X una variable aleatoria unidimensional con función de densidad de la forma:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) \qquad (-\infty \leq x \leq \infty)$$

(donde σ denota una constante) y sea Y una variable aleatoria definida por la transformación $Y = X^2$. Se verificará, entonces que:

$$F_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{x^2 \leq y} exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) dx = \int\limits_{-\sqrt{y}}^{\sqrt{y}} exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) dx$$

Ahora bien, dado que el integrando es una función par, se puede escribir:

$$F_Y(Y) = \frac{2}{\sigma\sqrt{2\pi}} \int\limits_0^{\sqrt{y}} exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) dx$$

de modo que la función de densidad de la variable Y queda definida de la siguiente forma:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{2}{\sigma\sqrt{2\pi}} \frac{1}{2\sqrt{y}} exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) dx =$$

$$= \frac{1}{\sqrt{y}}\left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\sigma^2}\right] (y > 0)$$

Sea una variable aleatoria n-dimensional $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ y un conjunto de transformaciones $f_i(X_1, X_2, \dots, X_n)$ $(i = 1,2, \dots, m; 1 \leq m < n)$ tales que:

$$\begin{cases} Y_1 = f_1(x_1, x_2, \dots, x_n) \\ Y_2 = f_2(x_1, x_2, \dots, x_n) \\ \dots \dots \dots \dots \dots \dots \\ Y_m = f_m(x_1, x_2, \dots, x_n) \end{cases}$$

La función de distribución de probabilidades de la variable m-dimensional $(Y_1, Y_2, \dots, Y_m)$ está dada por:

$$p\left\{\bigcap_{i=1}^{m} (Y_i \leq y_i)\right\} = \int \int \dots \int dF_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$$

(la integral múltiple es sobre el dominio $\bigcap_{i=1}^{m}[f_i(x_1, x_2, \dots, x_n) \leq y_i]$). En particular, si $m = n$, $\boldsymbol{X}$ es una variable absolutamente continua y las transformaciones $f_i(\cdot)$ tienen derivadas parciales continuas y son tales que:

$$J^{-1}(x_1,x_2,\dots,x_n)=\begin{bmatrix}\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n}\\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n}\end{bmatrix}\neq 0$$

Denotando por $y=(y_1,y_2,\dots,y_n)$ el conjunto de puntos para los cuales el sistema $f_i(x_1,x_2,\dots,x_n)=y_i$ $(i=1,2,\dots,n)$ posee soluciones $(x_1,x_2,\dots,x_n)$ $\in \boldsymbol{X}$, entonces entre las variables X e Y existe una correspondencia biunívoca y bicontinua y las transformaciones inversas $x_i=f_i^{-1}(y_1,y_2,\dots,y_n)$ $(i=1,2,\dots,n;(y_1,y_2,\dots,y_n)\epsilon Y)$ existen y son continuas y la función de densidad conjunta de $(y_1,y_2,\dots,y_n)$ es de la forma:

$$f_{Y_1,Y_2,\dots,Y_n}(y_1,y_2,\dots,y_n)=\begin{cases} f_{X_1,X_2,\dots,X_n}[f_1^{-1}(X),f_2^{-1}(X),\dots,f_n^{-1}(X)]. & \\ .\left|J[f_1^{-1}(X),f_2^{-1}(X),\dots,f_n^{-1}(X)]\right| & (y_1-y_n)\epsilon Y\\ 0 & (y_1,y_2,\dots,y_n)\notin Y\end{cases}$$

Un caso particularmente importante por sus aplicaciones es aquél en el que X es una variable de n dimensiones, $\boldsymbol{X}=\{X_1,X_2,\dots,X_n\}$, cuyas variables marginales son independientes y la transformación $K(X)$ está definida por la suma de las mismas, $K(X_1,X_2,\dots,X_n)=X_1+X_2+\cdots+X_n$. En particular, de acuerdo a lo expresado en la Sec. 2.1, para $n=2$, la condición de independencia estará dada por la relación:

$$F_{X_1,X_2}(x_1,x_2)=p[(X_1\leq x_1)\cap(X_2\leq x_2)]=p(X_1\leq x_1)p(X_2\leq x_2)=$$
$$=F_{X_1}(x_1)F_{X_2}(x_2)$$

Luego, se puede escribir:

$$F_{X_1,X_2}(y)=\int\limits_{x_1+x_2\leq y} dF_{X_1,X_2}(x_1,x_2)=\int\limits_{x_1+x_2\leq y} dF_{X_1}(x_1)dF_{X_2}(x_2)$$

y, de acuerdo con los postulados del teorema de Fubini[23], se demuestra que:

---

[23] Propiedad de las integrales de Stieltjes. Denominado también teorema de inversión del orden de integración. Según el cual, si $F_{X,Y}(x,y)=F_X(x)F_Y(y)$, para $x,y\epsilon\mathbb{R}_n$, entonces, $\int_{\mathbb{R}_n}\left[\int_{\mathbb{R}_n} g(x,y)dF_X(x)\right]dF_Y(y)=\int_{\mathbb{R}_n}\left[\int_{\mathbb{R}_n} g(x,y)dF_Y(y)\right]dF_X(x)$.

$$F_{X_1,X_2}(y) = \int_{-\infty}^{\infty} \left[\int_{x_1<y-x_2} dF_{X_1}(x_1) \right] dF_{X_2}(x_2) =$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{y-x_2} dF_{X_1}(x_1) \right] dF_{X_2}(x_2) = \int_{-\infty}^{\infty} F_{X_1}(y-x) F_{X_2}(x)$$

Donde la operación que aparece en el segundo miembro se denomina convolución de las funciones de distribución $F_{X_1}(.)$ y $F_{X_2}(.)$ y se denota por $F_{X_1}(.) * F_{X_2}(.)$. Esta propiedad puede ser expresada, entonces, de la siguiente forma: la función de distribución de una suma de dos variables aleatorias independientes está dada por la convolución de las funciones de distribución de las variables sumando,$F_{X_1+X_2}(y) = F_{X_1}(y) * F_{X_2}(y)$.

Debe tenerse en cuenta que la igualdad $F_{X_1}(y) * F_X(y) = F_{X_2}(y) * F_X(y)$ no implica necesariamente que $F_{X_1}(y) = F_{X_2}(y)$. Es decir, debe tenerse en cuenta que es posible hallar funciones de distribución de probabilidades $F_{X_1}(y), F_{X_2}(y)$ y $F_X(y)$ tales que $F_{X_1}(y)$ sea distinta de $F_{X_2}(y)$ pero sus convoluciones con $F_X(y)$ sean iguales. Por otra parte, dado que la suma de dos variables aleatorias es conmutativa, la operación de convolución también es conmutativa.

En particular, si las funciones $F_{X_1}(y)$ y $F_{X_2}(y)$ son continuas, se verifica que:

$$F_{X_1+X_2}(y) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{y-x_2} f_{X_1(u)du} \right] f_{X_2}(x_2) dx_2$$

y aplicando la sustitución $u = v - x_2$, será:

$$F_{X_1+X_2}(y) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{y} f_{X_1}(v-x) dv \right] f_{X_2}(x) dx =$$

$$= \int_{-\infty}^{y} \left[\int_{-\infty}^{\infty} f_{X_1}(v-x) f_{X_2}(x) dx \right] dv$$

Lo que demuestra que la función de densidad de la suma de dos variables aleatorias independientes es igual a la convolución de las funciones de densidad de las variables sumando:

$$f_{X_1+X_2}(y)\int_{-\infty}^{\infty} f_{X_1}(v-x)f_{X_2}(x)dx = f_{X_1}(y) * f_{X_2}(y)$$

Como la convolución es asociativa, es fácil demostrar que, dadas tres variables aleatorias independientes, se verifica que:

$$F_{X_1+X_2+X_3} = F_{(X_1+X_2)+X_3} = F_{X_1+X_2} * F_{X_3} = F_{X_1} * F_{X_2} * F_{X_3}$$

y, en general, dadas n variables independientes, se verifica que $F_{X_1+X_2+\cdots+X_n} = F_{X_1} * F_{X_2} * \dots * F_{X_n}$. En particular, dada una variable aleatoria Y definida como la suma de n variables independientes, $\{X_1, X_2, \dots, X_n\}$, todas con la misma función de distribución de probabilidades, F_X, se verificará que $F_Y = F_{X_1} * F_{X_2} * \dots * F_{X_n} = \{F_X\}^{n*}$, denominada convolución de orden n de la función F_X.

Ejemplo n° 16:

Sea (X, Y) una variable aleatoria bidimensional con función de densidad de la forma:

$$f_{X,Y}(x,y) = \begin{cases} 1 & para \quad 0 \le x \le 1, 0 \le y \le 1 \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

La función de densidad de la variable $Z = X + Y$ está dada por la convolución de las funciones de densidad de las variables marginales, definidas de la siguiente forma:

$$f_X(x) = \begin{cases} 1 & para \quad 0 \le x \le 1 \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

$$f_{Z-X}(z-x) =$$

$$= \begin{cases} 1 & para \quad 0 \le z-x \le 1\ (es\ decir\ para\ z-1 \le x \le z) \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

Es decir, será:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_{Z-X}(z-x)dx =$$

$$= \int_{\substack{0 \le z-x \le 1 \\ 0 \le x \le 1}} dx = \int_{\substack{z-1 \le x \le 1 \\ 0 \le x \le 1}} dx = medida\ de\ I(z)$$

(donde $I(Z)$ denota el conjunto obtenido de la intersección de $[0,1]$ y $[Z - 1, Z]$).

Las igualdades segunda y tercera de la expresión anterior resultan del hecho que el integrando es no-nulo (igual a 1) sólo cuando ambos factores son distintos de cero y esto ocurre cuando los argumentos de las funciones $f_{Z-X}(z-x)$ y $f_X(x)$ asumen valores en el intervalo $[0,1]$. Por lo tanto, será:

$$f_Z(z) = \begin{cases} 0 & para \quad z > 0\ (I(z) = \emptyset) \\ z & 0 \le z \le 1 \\ 2 - z & 1 < z \le 2 \\ 0 & z > 2\ (I(z) = \emptyset) \end{cases}$$

Sea una tercera variable W (independiente de X y de Y) con distribución de probabilidades uniforme en el intervalo $[0,1]$.

La función de densidad de la variable $V = X + Y + W$ está dada por la convolución entre las funciones $f_Z(z)$, calculada más arriba y $f_W(w)$, $f_V(v) = \int_{-\infty}^{\infty} f_{V-Z}(v-z) f_Z(z) dz$. Donde:

$$f_{V-Z}(\eta - z) = \begin{cases} 1 & para \quad \eta - 1 \le z \le \eta\ (0 \le \eta - z \le 1) \\ 0 & para\ cualquier\ otro\ valor\ de\ \eta \end{cases}$$

$$f_Z(z) = \begin{cases} z & para \quad 0 \le z \le 1 \\ 2 - z & 1 < z \le 2 \\ 0 & para\ cualquier\ otro\ valor\ de\ z \end{cases}$$

Luego, será:

$$f_V(\eta) = \int_{\substack{\eta-1 \le z \le \eta \\ 0 \le z \le 1}} z dz + \int_{\substack{\eta-1 \le z \le \eta \\ 1 \le z \le 2}} (2 - z) dz$$

Es decir, se obtiene que:

$$f_V(\eta) = \begin{cases} 0 & para & \eta > 0 \\ \dfrac{\eta^2}{2} & & 0 \leq \eta \leq 0 \\ \dfrac{3}{4} - \left(\eta - \dfrac{3}{2}\right) & & 1 < \eta \leq 2 \\ \dfrac{1}{2}(3-\eta)^2 & & 2 < \eta \leq 3 \\ 0 & & \eta > 3 \end{cases}$$

Obsérvese que, si bien las funciones de densidad de las variables marginales son rectangulares, la función de densidad de la variable $X + Y$ es triangular y la correspondiente a la variable $X + Y + W$ asume una forma campanular. Estas consideraciones están relacionadas con los postulados del teorema central del límite según el cual, en condiciones bastante generales, la variable aleatoria definida como la suma de n variables independientes tiende a una distribución de probabilidades Normal, cuando el número de sumandos crece indefinidamente[24].

Ejemplo n° 17:

Sean X_1 y X_2 dos variables aleatorias independientes, ambas con distribución de probabilidades de la forma:

$$f_X(x) = \begin{cases} 0 & para & x \leq 0 \\ \lambda e^{-\lambda x} & & x > 0 \end{cases}$$

(siendo $\lambda > 0$). Entonces, la distribución de probabilidades de la variable Y_2=$X_1 + X_2$ estará dada por:

$$f_{Y_2}(y) = \int_{-\infty}^{\infty} f_{Y_2-X}(y-x)\,dx = \lambda^2 \int_{\substack{y-x>0 \\ x>0}} e^{-\lambda(y-x)} e^{-\lambda x} dx =$$

$$= \lambda^2 e^{-\lambda y} \int_0^y dx = \lambda^2 y e^{-\lambda y}$$

Es decir, será:

[24] Ver Sección 12.2.

$$f_{Y_2}(y) = \begin{cases} 0 & para & y \leq 0 \\ \lambda^2 y e^{-\lambda y} & & y > 0 \end{cases}$$

Sea, ahora, una tercera variable, X_3 (independiente de las anteriores) y con distribución de probabilidades similar a la de X_1 y X_2. La función de probabilidades de la variable $Y_3 = X_1 + X_2 + X_3$ estará dada por:

$$f_{Y_3}(y) = \lambda^3 \int_0^y e^{-\lambda(y-x)} x e^{-\lambda x} dx = \lambda^3 e^{-\lambda y} \int_0^y x dx = \frac{\lambda^3 y^2}{2} e^{-\lambda y}$$

En general la distribución de probabilidades de la variable $Y_n = X_1 + X_2 + \dots + X_n$, donde las X_j son variables aleatorias independientes con la misma distribución de probabilidades, será:

$$f_{Y_n}(y) = \frac{\lambda^n y^{n-1}}{(n-1)!} e^{-\lambda y} \qquad (y > 0)$$

Ejemplo n° 18:

Sea $\{X_1, X_2, \dots, X_n\}$ un conjunto de variables aleatorias independientes semejantes correspondientes a n ventanillas que comienzan a servir a n clientes en forma simultánea y sean las variables aleatorias $Y = \min_{1 \leq j \leq n} X_j$ y $Z = \max_{1 \leq j \leq n} X_j$ que representan, respectivamente, el momento en que se libera la primera ventanilla y el momento en el cual terminan de ser atendidos los n clientes. Teniendo en cuenta las condiciones de independencia y semejanza de las variables X_j, se tiene que:

$$F_Z(z) = p(Z \leq z) = p\left(\max_{1 \leq j \leq n} X_j \leq z\right) =$$

$$= p[(X_1 \leq z) \cap (X_2 \leq z) \cap \dots \cap (X_n \leq z)] =$$

$$= p(X_1 \leq z) p(X_2 \leq z) \dots p(X_n \leq z) = [F_X(z)]^n$$

Si la función F_X es absolutamente continua, se verifica que:

$$F_Z(z) = [F_X(z)]^n = \int_{-\infty}^{z} n[F_X(u)]^{n-1} f_X(u) du$$

De modo que la función de densidad queda definida como $f_Z(z) = n[F_X(z)]^n f_X(z)$. De la misma forma:

$$F_Y(y) = p(Y \leq y) = p\left(\min_{1\leq j\leq n} X_j \leq y\right) = 1 - p\left(\min_{1\leq j\leq n} X_j > y\right) =$$

$$= 1 - p[(X_1 > y) \cap (X_2 > y) \cap \dots \cap (X_n > y)] =$$

$$= 1 - p(X_1 > y)p(X_2 > y) \dots p(X_n > y) = 1 - [1 - F_X(y)]^n$$

Luego, la función de densidad será $f_Y(y) = n|1 - F_X(y)|^{n-1} f_X(y)$. En particular, si las variables X_j se distribuyen de acuerdo con una función de probabilidades de la forma $f_X(x) = 1(0 \leq x \leq 1)$. Es decir, si su función de distribución es de la forma $F_X(x) = x(0 \leq x \leq 1)$, se obtiene que $F_Y(x) = 1 - (1 - x)^n(0 \leq x \leq 1)$ y $f_Y(x) = n(1 - x)^{n-1}(0 \leq x \leq 1)$ y que $F_Z(x) = x^n(0 \leq x \leq 1)$ y $f_Z(x) = nx^{n-1}(0 \leq x \leq 1)$. Luego, para todo $x > 0$, se verifica que:

$$\lim_{n\to\infty} F_Y(x) = \lim_{n\to\infty} p\left(\min_{1\leq j\leq n} X_j \leq x\right) = 1$$

$$\lim_{n\to\infty} F_Z(x) = \lim_{n\to\infty} p\left(\min_{1\leq j\leq n} X_j \leq x\right) = 0$$

De lo cual se puede concluir que el mínimo tiende a concentrarse en el origen y el máximo tiende a aproximarse al valor 1.

De la misma forma, si las variables X_i se distribuyen de acuerdo con una función de probabilidades exponencial de la forma:

$$f_X(x) = \begin{cases} 0 & para & x \leq 0 \\ \lambda e^{-\lambda x} & & x > 0 \end{cases}$$

será $F_X(x) = 1 - e^{-\lambda x}$. De lo que se puede concluir que al mínimo también le corresponde una función de distribución $F_Y(x) = 1 - e^{-n\lambda x}$ y a Z le corresponde una función de distribución de la forma $F_Z(x) = \left(1 - e^{-\lambda x}\right)^n$. Igual que en el caso anterior, para $x > 0$, se verifica que $\lim_{n\to\infty} F_Y(x) = 1$ y $\lim_{n\to\infty} F_Z(x) = 0$, Es decir que el mínimo tiende a concentrarse en el entorno del origen y Z tiende a dispersarse en el infinito.

Ejemplo n° 19:

Sea $\{X_1, X_2, \ldots, X_n\}$ un conjunto de variables aleatorias iid con distribución de probabilidades $f_X(x)$ y función de distribución $F_X(x)$.

Sean $X^{(1)}$ la variable aleatoria que representa el menor de los valores a asumir por las n variables, $X^{(2)}$ la variable aleatoria que representa el segundo valor en orden de magnitud creciente y, en general $X^{(n)}$ la variable que denota el mayor de los valores a asumir por las n variables[25]. Entonces, será:

$$F_{(k)}(x) = p\left(X^{(k)} \leq x\right) =$$

$$= p\left(al\ menos\ k\ de\ las\ n\ variables\ X_j\ sean\ menores\ o\ iguales\ que\ x\right) =$$

$$= p\Big(al\ menos\ k\ de\ los\ eventos\ \Big\{\left(X^{(1)} \leq x\right), \left(X^{(2)} \leq x\right), \ldots,$$

$$, \left(\left(X^{(n)} \leq x\right)\right)\Big\}\ sean\ verdaderos\Big) =$$

$$= \sum_{j=k}^{n} \binom{n}{j} [F_X(x)]^j [1 - F_X(x)]^{n-j}$$

Si la función $F_X(x)$ es absolutamente continua, entonces $F_{(k)}(x)$ también lo es y, por lo tanto, su derivada proporciona la correspondiente función de densidad:

$$f_{(k)}(x) = \frac{d}{dx} F_{(k)}(x) =$$

$$= \sum_{j=k}^{n} \binom{n}{j} j [F_X(x)]^{j-1} [1 - F_X(x)]^{n-j} f_X(x) -$$

$$- \sum_{j=k}^{n-1} \binom{n}{j} [F_X(x)]^{j-1} (n - j) [1 - F_X(x)]^{n-j-1} f_X(x) =$$

$$= \binom{n}{k} k [F_X(x)]^{k-1} [1 - F_X(x)]^{n-k} f_X(x) +$$

---

[25] Nótese que $X^{(1)}$ y $X^{(n)}$ coinciden, respectivamente, con las variables Y y Z del ejemplo precedente.

$$+\sum_{j=k+1}^{n}\left[\binom{n}{j}j-\binom{n}{j-1}(n-j+1)\right][F_X(x)]^{j-1}[1-F_X(x)]^{n-j}f_X(x)=$$

$$=\binom{n}{k}k[F_X(x)]^{k-1}[1-F_X(x)]^{n-k}f_X(x)$$

(nótese que, para $k=1$ y $k=n$, quedan definidas las funciones de densidad del mínimo y máximo, $f_Y(\cdot)$ y $f_Z(\cdot)$, del ejemplo precedente).

En particular, si las variables X_j tienen distribución de probabilidades tal que $f_X(x)=1$ y $F_X(x)=x(0\leq x\leq 1)$, se obtiene que:

$$f_{(k)}(x)=\binom{n}{k}kx^{k-1}(1-x)^{n-k}$$

Sea una variable aleatoria X discreta con distribución de probabilidades $\{X,\Omega(X)=\{1,2,\ldots\},p(X=j)=p_j(j=1,2,\ldots)\}$ y sea la variable $Y=K(X)$ (que, obviamente, también será discreta). Entonces, será:

$$p(Y=y_j)=p[K(x_j)=y_j]=\sum_{K(x_j)=y_j}p_j$$

Se puede concluir en forma inmediata que cada valor y_j puede ser generado por distintos valores de X. En particular, si $K(X)$ es una función invertible, existe una correspondencia biunívoca entre los valores de Y y de X, es decir que Y asume valores distintos de X con las mismas probabilidades:

$$p(Y=y_j)=p[Y=K(x_j)]=p(X=x_j)$$

Ejemplo n° 20:

Sean X_1 y X_2 dos variables aleatorias independientes, cada una de las cuales representa los valores a obtener al arrojar un dado "clásico" de n caras numeradas de 1 a n. La distribución de probabilidades de ambas variables es de la forma $p(X)=\frac{1}{n}(X=1,2,\ldots,n)$. Sea, por otra parte, la variable aleatoria Y que representala suma de los puntos a obtener al arrojar los dos dados, $Y=X_1+X_2$. Se puede escribir, entonces:

$$p(Y=y)=p(X_1+X_2=y)=\sum_{j=1}^{n}p[(X_1=j)\cap(X_1+X_2=y)]=$$

$$=\sum_{j=1}^{n}p[(X_1=j)\cap(X_2=y-j)]=$$

$$=\sum_{j=1}^{n}p(X_1=j)p(X_2=y-j)=\sum_{\substack{1\le j\le n\\1\le y-j\le n}}\frac{1}{n^2}=\sum_{\substack{1\le j\le n\\y-n\le j\le y-1}}\frac{1}{n^2}$$

Es decir, la probabilidad $p(Y=y)$ estará dada por el factor $\frac{1}{n^2}$ multiplicado por el número de valores enteros contenidos en el intervalo definido por la intersección $\{(1\le j\le n)\cap(y-n\le j\le y-1)\}$. Esta intersección será vacía si $y-1<1$ (es decir, si $y<2$) o si $y-n>n$ (es decir, si $y>2n$)[26]. Si se verifica que $2\le y\le n+1$, entonces la intersección queda definida por el intervalo $1\le j\le y-1$, es decir, será:

$$p(Y=y)=\sum_{j=}^{y-1}\frac{1}{n^2}=\frac{y-1}{n^2}$$

Asimismo, si se verifica que $n+2\le y\le 2n$, la intersección queda definida por el intervalo $y-n\le j\le n$. Luego, será:

$$p(Y=y)=\sum_{j=y-n}^{n}\frac{1}{n^2}=\frac{2n-y+1}{n^2}$$

Resumiendo los resultados anteriores, se puede concluir que la función de probabilidades de Y queda definida de la siguiente forma:

$$p(Y=y)=\begin{cases}\dfrac{y-1}{n^2} & para \qquad 2\le y\le n+1\\ \dfrac{2n-y+1}{n^2} & \qquad n+1\le y\le 2n\end{cases}$$

---

[26] En otros términos, el dominio de la variable $Y=X_1+X_2$ será $\Omega(Y)=\{2,3,4,\dots,2n\}$.

Sea una variable aleatoria unidimensional Y con dominio $\Omega(Y) = \{y_1, y_2, \ldots, y_n\}$ y tal que sus valores observan el siguiente orden $y_i > y_j$, si $i > j$ y sean dos distribuciones de probabilidades $\pi_1(p_i)$ y $\pi_2(p_i)$ sobre dicho dominio tales que $\pi_1(p_i) > \pi_2(p_i)(i = 1,2, \ldots, k-1)$, $\pi_1(p_k) > \pi_2(p_k)$ y $\pi_1(p_i) = \pi_2(p_i)(i = k+1, k+2, \ldots, n)$, de modo que:

$$F_2(y_{k-1}) = \sum_{i=1}^{k-1} \pi_2(p_i) < \sum_{i=1}^{k-1} \pi_1(p_i) = F_1(y_{k-1})$$

y $\sum_{i=k+1}^{n} \pi_1(p_i) = \sum_{i=k+1}^{n} \pi_2(p_i)$ y, en consecuencia, que $F_2(y_h) \leq F_1(y_h)(h = 1,2, \ldots, n)$. Se dice, entonces que la distribución $\{\pi_1(p_i)\}$ es estocásticamente dominante de primer orden respecto de $\{\pi_2(p_i)\}$.

Con un razonamiento similar se demuestra que la condición necesaria y suficiente para la dominancia de segundo orden de la distribución $\{\pi_1(p_i)\}$ respecto de $\{\pi_2(p_i)\}$ es que $\sum_{i=1}^{k} F_2(y_i)\Delta y_i \leq \sum_{i=1}^{k} F_1(y_i)\Delta y_i$(para $k < n$), donde $\Delta y_i = y_{i+1} - y_i > 0$.

De la misma forma, dadas una variable Y continua en el dominio $\Omega(Y) = [y_1, y_n]$ y las funciones de densidad y de distribución $\phi_1[g_1(Y)]$, $\phi_2[g_2(Y)]$, $G_1(Y)$ y $G_2(Y)$, la condición:

$$\int_{y_1}^{y} G_2(u)du \leq \int_{y_1}^{y} G_1(u)du \left(y \epsilon \Omega(Y)\right)$$

es necesaria y suficiente para poder asegurar que $\phi_2[g(Y)]$ es estocásticamente dominante de segúndo orden respecto de $\phi_1[g(Y)]$. En general, la condición necesaria y suficiente para poder asegurar que $\phi_2[g(Y)]$es estocásticamente dominante de orden $j\ (= 2,3, \ldots)$ respecto de $\phi_1[g(Y)]$ es que:

$$I_j[G_2(y)] = \int_{y_1}^{y} \int_{y_1}^{v_1} \ldots \int_{y_1}^{v_{j-1}} G_2(u)dudv_1 \ldots dv_{j-1} \leq I_j[G_1(y)] =$$

$$= \int_{y_1}^{y} \int_{y_1}^{v_1} \ldots \int_{y_1}^{v_{j-1}} G_1(u)dudv_1 \ldots dv_{j-1}$$

Luego, se puede concluir en forma inmediata que la dominancia de orden j implica la de orden $j+1$, pero que la implicación inversa no tiene por qué verificarse.

10.- Valores esperados o momentos de una variable aleatoria

10.1.- Definición

Existen ciertos valores típicos de gran importancia en el estudio de las variables aleatorias denominados "momentos", que describen aspectos particulares de las mismas o, más precisamente, de sus distribuciones de probabilidades y permiten, en cierta forma, su caracterización.

En general un momento se define como el "valor esperado" o promedio aritmético ponderado de una función de la variable aleatoria elevada a un exponente real no-negativo, cuyo valor determina el orden del momento:

$$E\{[\varphi(X)]^s\} = m_s(X) = \int_{\Omega(X)} [\varphi(X)]^s dF_X(x) = \int_{\Omega(X)} [\varphi(X)]^s f_X(x)dx$$

En particular, los momentos llamados "naturales" o "con origen cero" están definidos como los valores esperados de potencias de la variable X. Según se trate de una variable discreta,$\{x_i, p_i, i = 1,2, \dots\}$o de una variable continua con función de distribución $F_X(x)$, su expresión será:

$$E(X^s) = m_s(X) = \sum_{i=1}^{\infty} x_i^s p_i$$

$$E(X^s) = m_s(X) = \int_{\Omega(X)} x^s dF_X(x) = \int_{\Omega(X)} x^s f_X(x)dx$$

Si $\varphi(X) = X - E(X)$, se generan los momentos denominados "centrados" o "con origen en el valor esperado" o "de los desvíos" de la variable X:

$$E\{[X - E(X)]^s\} = \mu_s(X) = \sum_i [x_i - E(X)]^s p_i = \sum_i [x_i - m(X)]^s p_i$$

$$E\{[X - E(X)]^s\} = \mu_s(X) = \int_{\Omega(X)} [X - E(X)]^s f_X(x)dx$$

Los momentos naturales y centrados están relacionados mediante las siguientes expresiones:

$$\mu_s(X) = \sum_{j=0}^{s} (-1)^j \binom{s}{j} [m(X)]^j m_{s-j}(X) \; (s = 0,1,2, \dots)$$

$$m_s(X) = \sum_{j=0}^{s} [m(X)]^j \mu_{s-j}(X) \; (s = 1,2, \dots)$$

Si $\varphi(X) = |X|$ quedan definidos los momentos "absolutos-naturales", $m_{|s|}(X) = E(|X|^s)$para los cuales se verifica que $\sqrt[s_1]{m_{|s_1|}(X)} \leq \sqrt[s_2]{m_{|s_2|}(X)}$ $(1 \leq s_1 < s_2)$. Asimismo, si $\varphi(X) = |X - E(X)|$, quedan definidos los momentos "absolutos-centrados", $\mu_{|s|}(X) = E(|X - E(X)|^s)$.

Otros valores típicos de una variable aleatoria son los denominados "momentos factoriales", definidos para todo s entero como $m_{[s]}(X) = E[X(X-1) \dots (X-s+1)]$[27]. En particular, será:

$$m_{[1]}(X) = E(X) = m(X)$$

$$m_{[2]}(X) = E[X(X-1)] = E(X^2) - E(X) = m_2(X) - m(X)$$

$$m_{[3]}(X) = E[X(X-1)(X-2)] = m_3(X) - 3m_2(X) + 2m(X)$$

… … … … … … … … … … … ….

Su utilidad radica fundamentalmente en la posibilidad de dar, en algunos casos, una forma más simple a ciertas expresiones que involucran a los momentos. De acuerdo con las relaciones anteriores, su valor puede ser obtenido a partir de los momentos naturales y viceversa, los momentos naturales pueden ser calculados a partir de los momentos factoriales.

De la misma forma que para las variables aleatorias unidimensionales, es posible obtener momentos para variables de más de una dimensión. Dada, por ejemplo, una variable bidimensional (X, Y), se puede definir un momento "mixto natural" de orden (s_1, s_2) como $E(X^{s_1} Y^{s_2}) = m_{s_1,s_2}(X, Y)$.

[27] Obviamente, si X asume valores enteros, $[X(X-1) \dots (X-s+1)] = \frac{X!}{(X-s)!}$.

Obviamente los momentos correspondientes a las variables marginales pueden ser expresados de la siguiente forma: $m_{s_1,0}(X,Y) = E(X^{s_1})$ y $m_{0,s_2}(X,Y) = E(Y^{s_2})$.

Aplicando un criterio similar al utilizado en las variables unidimensionales, se pueden definir los momentos "mixtos centrados" de orden (s_1, s_2) como:

$$E\{[X - E(X)]^{s_1}[Y - E(Y)]^{s_2}\} = \mu_{s_1,s_2}(X,Y)$$

Entre estos momentos tiene particular importancia el momento centrado mixto de orden $(1,1)$, denominado comúnmente "covarianza":

$$cov(X,Y) \equiv \gamma(X,Y) \equiv \mu_{1,1}(X,Y) \equiv \mu(X,Y) = E\{[X - E(X)][Y - E(Y)]\} =$$

$$= E(XY) - E[YE(Y)] - E[XE(X)] + E(X)E(Y) =$$

$$E(XY) - E(X)E(Y) = m_{1,1}(X,Y) - m(X)m(Y)$$

La covarianza proporciona una medida del grado de relación lineal entre las variables X e Y[28]. Como corolario de su definición se obtiene en forma inmediata que $\gamma(aX, bY) = ab\gamma(X,Y)$ (donde a y b son constantes).

A partir de este corolario, se puede concluir que, dadas dos variables aleatorias, U y V, definidas como combinaciones lineales de sendos conjuntos de variables aleatorias unidimensionales, $\{X_1, X_2, \dots, X_m\}$ y $\{Y_1, Y_2, \dots, Y_n\}$:

$$U = a_1X_1 + a_2X_2 + \cdots + a_mX_m$$

$$V = b_1Y_1 + b_2Y_2 + \cdots + b_nY_n$$

se verifica que $\gamma(U,V) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \gamma(X_i, Y_j)$.

En general, dada una variable aleatoria n-dimensional, $\boldsymbol{X} = \{X_1, X_2, \dots, X_n\}$, los correspondientes momentos mixtos naturales y centrados de orden $\boldsymbol{s} = (s_1, s_2, \dots, s_n)$ quedan definidos, respectivamente, como:

$$m_{\boldsymbol{s}}(\boldsymbol{X}) = m_{s_1,s_2,\dots,s_n}(X_1, X_2, \dots, X_n) = E\left(\prod_{j=1}^{n} X_j^{s_j}\right)$$

$$\mu_{\boldsymbol{s}}(\boldsymbol{X}) = \mu_{s_1,s_2,\dots,s_n}(X_1, X_2, \dots, X_n) = E\left\{\prod_{j=1}^{n}\left[X_j - E(X_j)\right]^{s_j}\right\}$$

---

[28] Para un estudio más detallado de esta medida, ver Sec. 10.5.

Estos momentos se relacionan de la siguiente forma:

$$\mu_s(\boldsymbol{X}) = \sum_{j_1}^{s_1}\sum_{j_2}^{s_2}\dots\sum_{j_n}^{s_n}(-1)^{j_1+j_2+\cdots+j_n}\binom{s_1}{j_1}\binom{s_2}{j_2}\dots\binom{s_n}{j_n}.$$

$$.\,[m(X_1)]^{j_1}[m(X_2)]^{j_2}\dots[m(X_n)]^{j_n}m_{s-j}(\boldsymbol{X})$$

$$m_s(\boldsymbol{X}) = \sum_{j_1}^{s_1}\sum_{j_2}^{s_2}\dots\sum_{j_n}^{s_n}\binom{s_1}{j_1}\binom{s_2}{j_2}\dots\binom{s_n}{j_n}.$$

$$.\,[m(X_1)]^{j_1}[m(X_2)]^{j_2}\dots[m(X_n)]^{j_n}\mu_{s-j}(\boldsymbol{X})$$

El conocimiento de algunos momentos de la sucesión completa correspondiente a una variable obviamente no es suficiente para caracterizar completamente a su distribución de probabilidades. En realidad, ni siquiera es suficiente el conocimiento de la sucesión completa de momentos de orden entero: dos variables, X e Y, tales que $E(X^s) = E(Y^s) = m_s$ (para todo s entero), pueden distribuirse de acuerdo con funciones de probabilidades diferentes. Para poder asegurar el cumplimiento de la condición de unicidad (es decir, para poder asegurar que si dos variables tienen los mismos momentos son idénticamente distribuidas) es suficiente que se verifiquen las siguientes condiciones: **i)** que $E(X^s) = E(Y^s) < +\infty$ $(s = 1,2,\dots)$[29] y **ii)** que exista un valor $w > 0$, tal que la serie $\sum_{j=0}^{\infty} m_j(X)\frac{w^j}{j!}$ sea convergente.

Además del problema de la unicidad se debe considerar el de la existencia. Las condiciones necesarias para poder asegurar que $\{m_s, s = 1,2,\dots\}$ constituye la sucesión de momentos de una variable aleatoria son: **i)** que $m_s > 0$ para todo s par y **ii)** que se verifique la desigualdad $\left(m_{s_1}\right)^{1/s_1} < \left(m_{s_2}\right)^{1/s_2}(0 < s_1 < s_2)$[30].

Dada una variable bidimensional (X, Y), se pueden definir los "momentos factoriales mixtos":

$$m_{[s_1,s_2]}(X,Y) = E\{[X(X-1)\dots(X-s_1+1)][Y(Y-1)\dots(Y-s_2+1)]\}$$

[29] La demostración de esta propiedad será desarrollada en la Sec. 13.2, al estudiar la función característica.

[30] Esta expresión se obtiene inmediatamente haciendo $g_1(X) = X^{s_1}$, $g_2(X) = 1$ y $\alpha = \frac{s_2}{s_1}$ (para $0 < s_1 < s_2$) en la desigualdad de Schwarz-Hölder (ver Sec. 10.2.5).

Los “momentos mixtos absolutos-naturales”:

$$m_{|s_1,s_2|}(X,Y) = E(|X|^{s_1}|Y|^{s_2})$$

y los “momentos mixtos absolutos-centrados”:

$$\mu_{|s_1,s_2|}(X,Y) = E\left(\left|X - m_{1,0}(X)\right|^{s_1}\left|Y - m_{0,1}(Y)\right|^{s_2}\right)$$

Ejemplo n° 21:

Sea una variable aleatoria X con función de densidad de la forma:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] (-\infty \leq x \leq \infty)$$

y sea la variable $Y = e^x$, que puede asumir sólo valores positivos.

La función de distribución de la variable Y será de la forma:

$$F_Y(y) = p(Y \leq y) = p(e^x \leq y) = p[x \leq ln(y)] =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{ln(y)} exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] dx \; (y > 0)$$

y la correspondiente función de densidad será:

$$f_Y(y) = \frac{1}{\sigma y\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{ln(y)-m}{\sigma}\right)^2\right] (y > 0)$$

Realizando la sustitución $\frac{ln(y)-m}{\sigma} = z$ de modo que $y = e^{\sigma z + m}$, se obtienen los momentos naturales:

$$m_s(Y) = E(Y^s) = \int_0^{\infty} y^s \, f_Y(y) dy =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_0^{\infty} y^{s-1} exp\left[-\frac{1}{2}\left(\frac{ln(y)-m}{\sigma}\right)^2\right] dy =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} exp[(\sigma z + m)(s-1)]exp\left(-\frac{z^2}{2}\right)exp(\sigma z + m)dz =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} exp\left(s\sigma z + ms - \frac{z^2}{2}\right)dz$$

y eliminando del exponente el término lineal mediante la traslación $z = v + s\sigma$, se puede escribir:

$$m_s(Y) = E(Y^s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} exp\left(\frac{1}{2}s^2\sigma^2 + ms - \frac{v^2}{2}\right)dv =$$

$$= exp\left(\frac{1}{2}s^2\sigma^2 + ms\right) \int_{-\infty}^{\infty} \frac{e^{-v^2/2}}{\sqrt{2\pi}} dv = exp\left(\frac{1}{2}s^2\sigma^2 + ms\right)$$

Supóngase, en particular, que $m = 0$ y $\sigma = 1$, es decir, que:

$$f_Y(y) = \frac{1}{y\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(ln(y)\right)^2\right]$$

los momentos naturales quedan definidos de la siguiente forma:

$$m_s(Y) = E(Y^s) = \int_0^{\infty} y^s f_Y(y)dy = \frac{1}{\sqrt{2\pi}}\int_0^{\infty} y^{s-1} exp\left[-\frac{1}{2}\left(ln(y)\right)^2\right]dy$$

Realizando la sustitución $ln(y) = u$, se obtiene que:

$$m_s(Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} exp\left(us - \frac{u^2}{2}\right)du =$$

$$= \frac{e^{s^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} exp\left[-\frac{1}{2}(u-s)^2\right]du = e^{s^2/2}$$

Como se vio, la condición suficiente para que la sucesión de momentos identifique a la variable es que exista un $w > 0$ tal que la serie $\sum_{s=0}^{\infty} m_s(X)\frac{w^s}{s!}$

sea convergente. En este caso el término genérico de la sucesión es de la forma:

$$e^{s^2/2}\frac{w^s}{s!}=\frac{e^{s^2/2}w^s}{s^s e^{-s}\sqrt{2\pi s}}=\frac{\left(e^{s/2}ws^{-1}e\right)^s}{\sqrt{2\pi s}}$$

Para s suficientemente grande, cualquiera sea el valor de $w>0$ que se elija, se verificará que $e^{s/2}ws^{-1}\geq 2$ y, por lo tanto, resulta que:

$$\lim_{s\to\infty} m_s(Y)\frac{w^s}{s!}\geq\frac{2^s}{\sqrt{2\pi s}}=+\infty$$

De lo que se concluye que, para la variable transformada Y, la condición de unicidad no se cumple.

10.2.- La esperanza matemática

10.2.1.- Definición

El momento natural de orden uno define una medida denominada "esperanza matemática" o "valor esperado" de X:

$$E(X)=\sum_{i=1}^{\infty} x_i p_i = m_1(X)=m(X)$$

$$E(X)=\int_{-\infty}^{\infty} x dF_X(x)=\int_{-\infty}^{\infty} x f_X(x)dx$$

que caracteriza la posición de la variable (dado que la integrabilidad y la integrabilidad absoluta son equivalentes, se puede asegurar que $E(X)$ implica la existencia de $E(|X|)$).

Dada una variable aleatoria compleja, $Z=X+iY$, tal que $E(X)$ y $E(Y)$ existan, será $E(Z)=E(X)+iE(Y)$.

10.2.2.- Una interpretación subjetivista

De acuerdo con la interpretación subjetivista la probabilidad de ocurrencia de un evento A ($p(A)$) puede ser definida como el precio equitativo a pagar para recibir 1 si el resultado A se verifica y 0 si no se verifica. Obviamente se puede apostar por un importe x, distinto de 1, pagando una cantidad $xp(A)$ para recibir x si el resultado A se verifica[31].

Sean ahora n eventos incompatibles, $\{A_j, j = 1,2, \dots, n\}$, que definen una partición del dominio Ω, de modo que $A_i \cap A_j (i \neq j)$ y $\bigcup_{j=1}^{n} A_j = \Omega$ y sean $p_j = p(A_j)(j = 1,2, \dots, n)$ sus correspondientes probabilidades de ocurrencia, de modo que $\sum_{j=1}^{n} p_j = 1$. Si se supone que en caso de ocurrencia de un evento A_j un apostador recibirá la cantidad variable x_j, entonces las n apuestas sobre los eventos A_j sumarán $\sum_{j=1}^{n} x_j p_j$. La cantidad a recibir por dicho apostador según cuál sea el resultado que se verifique, puede ser interpretada como una variable aleatoria X con distribución de probabilidades $\{x_j, p_j, j = 1,2, \dots, n\}$. Definiendo una función indicador (la cual es en sí misma una variable aleatoria que puede asumir solamente los valores 0 ó 1 con probabilidades $p(\bar{A})$ y $p(A)$, respectivamente):

$$I_{A_j}(w) = \begin{cases} 1 & para \quad w \epsilon A_j \\ 0 & \quad\quad\quad\; w \epsilon \bar{A}_j \end{cases}$$

la variable aleatoria X puede ser expresada como $X(w) = \sum_{j=1}^{n} x_j I_{A_j}(w)$. La suma de las apuestas representa, entonces, el valor esperado (o la "previsión" en la nomenclatura subjetivista) de la variable X, $\wp\big(X(w)\big) = E\big(X(w)\big) = \sum_{j=1}^{n} x_j p_j$.

De estas consideraciones se puede concluir con Huygens que, en la interpretación subjetivista, el valor esperado representa el precio equitativo a pagar para recibir una cantidad aleatoria X[32]. En particular, si la variable aleatoria puede asumir sólo los valores 0 y 1 y utilizando la función indicador I_A (donde A denota el conjunto de puntos del dominio Ω para los cuales la variable aleatoria asume el valor 1), se verifica que $p(A) = E(I_A)$. De modo que la probabilidad puede ser considerada como el valor esperado para una variable aleatoria correspondiente a una apuesta simple y, a su vez, el valor esperado puede ser considerado como la extensión del concepto de probabilidad para una apuesta múltiple.

---

[31] La cantidad x puede ser negativa. Los términos "pagar" y "recibir" deben ser interpretados en sentido algebraico.

[32] En realidad, esta apreciación tiene un alcance muy limitado en la medida que el concepto de equidad de la apuesta está afectado por la actitud del individuo frente al riesgo y, por lo tanto, no se puede hablar de equivalencia en sentido absoluto entre una cantidad aleatoria y una cantidad cierta.

Sea ahora $A_j = X^{-1}(x_j) = \{w, X(w) = x_j\}(j = 1,2, \dots)$ un conjunto de eventos incompatibles que definen una partición del dominio Ω. La variable aleatoria $X(w)$ asumirá el valor x_j cuando se verifique que $w \epsilon A_j$ y, por lo tanto, cuando admita una representación de la forma $X(w) = \sum_{j=1}^{\infty\forall} x_j I_{A_j}(w)$ con distribución de probabilidades:

$$p_j = p(X = x_j) = p\left(X^{-1}(x_j)\right) = p(A_j)$$

De esta forma se demuestra que las consideraciones realizadas para la definición del valor esperado en el caso de n apuestas, son generalizables a una variable aleatoria discreta con distribución de probabilidades $\{x_j, p_j, j = 1,2, \dots\}$, de modo que se puede escribir $E(X) = \sum_{j=1}^{\infty} x_j p_j$, siempre que la sucesión sea absolutamente convergente, es decir siempre que la suma se mantenga finita y no varíe ante modificaciones en el orden de los sumandos; en otros términos, siempre que la cantidad a pagar no dependa del orden en el cual se realicen las apuestas.

10.2.3.- La apuesta de Pascal o paradoja de la salvación eterna

La conocida como apuesta de Pascal es un ejemplo típico de cómo el razonamiento mediante expectativas constituyó un sinónimo del nuevo racionalismo imperante en la primera mitad del siglo XVII. Un ejemplo en el cual el espíritu religioso parece basarse más "*...en el interés temporal individual que en un dogmatismo racional o un entusiasmo místico*" (Daston (1988, p. 62)).

Pascal –quien, basado en la supuesta superioridad de la fe sobre la razón, había rechazado por innecesarios todos los intentos para probar la existencia de Dios- incluye en "*Les pensées*" (1669) un diálogo en el que propone un argumento bajo la forma de un juego que admite sólo dos alternativas posibles: aceptar o rechazar la existencia de Dios, de modo que la aceptación no implicaría ninguna pérdida en esta vida y produciría la ganancia infinita de la salvación eterna y el rechazo podría significar una equivocación que conduciría a la condenación eterna.

La propuesta de Pascal consiste en considerar "*...las proposiciones: 'Dios existe' y 'Dios no existe' ¿Hacia cuál de las dos nos inclinaremos? La razón no puede determinar nada acerca de esta cuestión. Con respecto a esas afirmaciones, separadas por una distancia infinita, se plantea un juego cuyos resultados posibles serán 'cara' o 'cruz' ¿A qué apostaremos? (...) Dos cosas se pueden ganar: la verdad y la beatitud, y tres cosas se pueden comprometer: nuestra razón, nuestra voluntad y nuestro conocimiento; y de* *dos cosas debe huir nuestra naturaleza: el error y la miseria (...) Evaluemos la ganancia y la pérdida posibles considerando 'cara' que 'Dios existe': si ganamos, ganamos todo; si perdemos, no perdemos nada. Apostemos, pues, sin vacilar a que Dios existe (...) Y tendremos como ganancia posible una vida infinitamente feliz, una posibilidad de triunfar contra un número finito de posibilidades de perder, no hay que vacilar, hay que arriesgarlo todo (...) Pues de nada sirve decir que la ganancia es incierta y que el riesgo es cierto, y que la infinita distancia que media entre la certeza de lo que se arriesga y la incertidumbre de la ganancia iguala el bien finito que se arriesga en forma cierta con el infinito que es incierto. No es así. Todo jugador arriesga con certeza para ganar con incertidumbre y, sin embargo, sin pecar por ello contra la razón, arriesga con certeza lo finito para ganar con incertidumbre lo finito. No hay una distancia infinita entre esa certeza de lo que arriesga y la incertidumbre del triunfo (...) Hay, en verdad, una distancia infinita entre la certeza de ganar y la certeza de perder. Pero la incertidumbre de ganar es proporcional a la certeza de lo que se arriesga, según la proporción de las probabilidades de ganar y perder. Y de esto resulta que, si hay probabilidades iguales de un lado y del otro, la partida consiste en jugar de igual a igual y, entonces la certeza de lo que se arriesga es igual a la incertidumbre de la ganancia (...) En la medida que se arriesga lo finito, en un juego en el que hay iguales probabilidades de ganar y perder y la ganancia es infinita, nuestra propuesta posee una fuerza infinita*"[33]

De acuerdo con este razonamiento, si el jugador A ("*nosotros*") apostara a la ocurrencia del resultado "*cara*" (es decir, si confiara en que "*Dios existe*"), su posible ganancia (compuesta por infinitas "unidades de beatitud" que garantizan la "*salvación eterna*") sería $g_1 = \infty$, con una probabilidad $p_1 > 0$ y su pérdida posible en esta vida (de acuerdo con la hipótesis Pascaliana" sería $g_2 = 0$, con una probabilidad $p_2 > 0$ (dado que se supone que existe "*...un número finito de posibilidades de perder*"). Luego, según la definición de Huygens, la esperanza matemática del juego para A está dada por $E(Y) = g_1(1 - p_2) + g_2 p_2 = \infty$.

Desde un punto de vista estrictamente probabilístico, el hecho que la esperanza matemática de un juego sea infinita y que, en consecuencia, se deba integrar una suma ideal infinita (formada por infinitas "unidades de fe") para participar en él no constituye ningún contraejemplo a la definición de Huygens. La que se podría considerar como "la paradoja de la salvación eterna" surge del mismo planteo del juego en el que Pascal postula "*...que la*

[33] Pasaje n° 418 en la numeración de Lafuma y 233 en la de Brunschvicg. En realidad, Pascal no fue el creador de este famoso ejemplo de apologética religiosa de la teología natural. Un argumento bastante parecido se encuentra en *"La logique oul'art de penser"* (1662) de Antoine Arnauld y Pierre Nicole -los lógicos de Port Royal- y en "*Wisdom of being religious"* (1664) de John Tillotson.

ganancia es infinita" y, a partir de una confusa e inexplicable interpretación de la "*fortuna*" de A ("*...nuestra razón, nuestra voluntad y nuestro conocimiento*") como una magnitud medible en unidades físicas que sólo puede asumir un monto finito ("*...lo que comprometemos (...) es finito (...) y no hay una infinidad de posibilidades de perder contra la posibilidad de triunfo*") provoca una violación de la condición de equidad al considerar como natural la discrepancia entre la definición de esperanza matemática del juego y la presunta finitud de la "*fortuna*" del jugador A.

10.2.4.- La paradoja de San Petersburgo y la definición de la esperanza moral

En una carta fechada el 9 de setiembre de 1713 N. Bernoulli planteó a Pierre de Rémond Marqués de Montmort una serie de problemas relacionados con juegos de azar[34]. En particular, el quinto y último problema se refiere a un juego en el que un jugador A arroja sucesivamente un dado "clásico" hasta obtener el resultado "6". Si esta circunstancia se produce en la primera "tirada", entonces A le debe entregar a su contrincante, el jugador B, una moneda; si el resultado "6" se presenta por primera vez en la segunda tirada, A le debe entregar a B dos monedas; si este resultado se presenta en la tercera "tirada", A le debe entregar a B 2^2 monedas. En general, si el resultado "6" se presenta en la n-ésima "tirada", A le debe entregar a B $2^{n-1} (n = 12, \ldots)$ monedas. De modo que la esperanza matemática del juego para el jugador B está dada por:

$$E(Y) = \frac{1}{6} \sum_{j=1}^{\infty} 2^{j-1} \left(\frac{5}{6}\right)^{-1}$$

Este problema, que considera el incremento de la ganancia de B de acuerdo con una progresión geométrica es, en realidad, una modificación del cuarto problema en el cual el jugador A arroja sucesivamente un dado clásico hasta obtener el resultado "6". Si esta circunstancia se produce en la n-ésima "tirada" A le debe entregar al jugador B $n(= 1,2,3, \ldots)$monedas[35]. En este

[34] La correspondencia de Montmort-Bernoulli se publicó por primera vez, en la segunda edición del "*Essaid'analyse sur les jeux d'azard*" de de Montmort.

[35] En Cardano, G. (1564) figura un juego (una de cuyas versiones se conoce hoy como martingala) que podría ser considerado como un antecedente de estos problemas: Sean dos jugadores: A, cuya fortuna consta de c monedas y B, que posee una moneda. Suponiendo que ambos arriesguen apuestas iguales, si B gana, las apuestas se duplicarán en el juego siguiente y si A gana, el juego finaliza.

caso N. Bernoulli, aplicando los métodos de suma de series convergentes desarrollados por su tío Jakob, obtuvo que la esperanza matemática del juego para el jugador B está definida por:

$$E(Y) = \frac{1}{6}\sum_{j=1}^{\infty} j\left(\frac{5}{6}\right)^{j-1} = \frac{1}{6}\left(1-\frac{5}{6}\right)^{-2} = 6$$

(en general se demuestra que, en un juego de este tipo, la esperanza matemática es igual a la inversa de la probabilidad de que ocurra un resultado favorable al jugador B). Una solución que, obviamente, no es aplicable al quinto problema, en el que la serie es divergente[36].

Posteriormente Cramer (1728), a fin de simplificar el tratamiento del problema, sustituyó el dado por una moneda "clásica", de modo que, denotando por Y a la variable aleatoria que representa las cantidades a recibir por el jugador B, su función de probabilidades y su función de distribución quedan definidas, respectivamente, de la siguiente forma:

$$p(Y = 2^{y-1}) = \frac{1}{2^y} \qquad (y = 1,2,\dots)$$

$$F_Y(y) = p(Y \leq 2^{y-1}) = \sum_{j=1}^{y} \frac{1}{2^j} = 1 - \frac{1}{2^y}$$

Lo cual implica que, dado que la función de pago del jugador A ($f(y)$) crece en forma proporcional al decrecimiento de las probabilidades de ocurrencia de una sucesión ininterrumpida de resultados "ceca", la esperanza matemática del juego para el jugador B queda definida por una serie no-sumable de la forma:

$$E(Y) = \sum_{y=1}^{\infty} f(y)p(y) = \sum_{y=1}^{\infty} 2^{y-1}\frac{1}{2^y} = \frac{1}{2} + \frac{1}{2} + \cdots = \infty$$

[36] De aplicar la misma solución al quinto problema se obtendría que $E(Y) = \frac{1}{6}\sum_{j=1}^{\infty}\left(\frac{10}{6}\right)^{j-1} = -\frac{1}{4}$, un resultado que innegablemente implica una contradicción entre la definición de Huygens de esperanza matemática y la ganancia que B esperaría obtener del juego Es probable que los trabajos de Guido Grandi y la correspondencia entre Leibniz y N. Bernoulli sobre la convergencia y divergencia de series hayan influido en el planteo de estos problemas.

Este resultado obligaría al jugador B a entregar a A un número infinito de monedas antes de comenzar el juego para asegurar el cumplimiento de la condición de equidad pero, dado que "*...es prácticamente imposible que no ocurra el resultado 'cara' al cabo de una sucesión de infinitas tiradas del dado*" (Cramer (1728, p. 558)), es posible suponer que ningún jugador estaría dispuesto a pagar más de una cantidad finita (en general pequeña, si se tiene en cuenta que la duración esperada del juego es $E(n) = \sum_{j=1}^{n} \frac{j}{2^j} = \frac{1}{2}\left(1 + \frac{1}{2} + \frac{3}{2^2} + \cdots\right) = \frac{1}{2}\left(1 - \frac{1}{2}\right)^{-2} = 2$) para participar en él.

Como en el caso de la apuesta de Pascal, desde un punto de vista estrictamente probabilístico se puede concluir que el hecho que la esperanza matemática del juego de N. Bernoulli sea infinita y que, en consecuencia, se deba integrar idealmente una suma infinita para participar en él no constituye ninguna paradoja[37]. En realidad, la conocida como paradoja de San Petersburgo -igual que la de Pascal- surge de la interpretación operacional del juego, la cual implica una violación en la práctica, de la condición fundamental de equidad provocada, en este caso, por la discrepancia entre la definición de esperanza matemática y el presunto comportamiento racional del jugador.

Para resolver esta diferencia entre la interpretación matemática de la expectativa y lo que indica el sentido común, los probabilistas del siglo XVIII y la primera mitad del siglo XIX intentaron numerosas soluciones a este problema (todas concebidas a partir de definiciones objetivistas de la probabilidad). Unas, centradas en la decisión del jugador, estuvieron dirigidas a la evaluación de una apuesta aceptable de acuerdo con la definición de esperanza matemática; otras, vinculadas con la estructura del juego, se basaron en redefiniciones de la función de pagos, reinterpretaciones de la distribución de probabilidades y consideraciones acerca de la duración del juego y estuvieron dirigidas a lograr la finitud de su esperanza matemática. Un análisis comparado permite conjeturar que la naturaleza de estas soluciones es una consecuencia de la falta de una definición formal para las nociones de

[37] La calificación de aparentes atribuida a las paradojas de Pascal y de Bernoulli se basa en la definición canónica de Saisbury (1995), según la cual una paradoja es una argumento que, a partir de principios intuitivamente plausibles y a través de deducciones intuitivamente aceptables, conduce a una conclusión fuertemente contra-intuitiva absurda o a una verdadera contradicción y en la consideración que los ejemplos tratados aquí constituyen solamente afirmaciones no plausibes o contrarias al sentido común y no verdaderas contradicciones.

valor esperado y espacios de probabilidad infinitos, que fueron introducidos por Kolmogorov recién en 1933.

Si bien, como se verá, ninguno de estos intentos condujo a una solución de la versión original del problema ni a una respuesta completa y universalmente aceptada a la cuestión de por qué un jugador con sentido común decide arriesgar una apuesta menor (o mayor) que la esperanza matemática del juego, cabe destacar las propuestas debidas a N. Bernoulli, G. Cramer, D. Bernoulli, A. Fontaine, J.-B. Le-Rond d'Alembert, G.-L. Leclerc de Buffon, M.-J.A.N. Caritat de Condorcet, P.S. de Laplace, S.-F. Lacroix, S.-D. Poisson, E. Czuber, W.A. Whitworth, H.E. Timerding, K. Menger, W. Feller y A. Martin-Löf[38].

Cramer (1728) postuló que la solución del problema de N. Bernoulli consistía en redefinir la función de pagos $(f(y))$ a fin de lograr que la sucesión $\sum_{y=1}^{\infty} f(y)p(y)$ fuera convergente. A partir de la idea que "*Los matemáticos evalúan el dinero en proporción a su monto y las personas razonables en proporción al uso que puedan hacer de él*" (pp. 560-561), que la esperanza matemática identificada con el contexto legal de los contratos aleatorios no toma en consideración las "*circunstancias financieras*" individuales de los jugadores (tomadores de riesgo) y que, en consecuencia, es necesario interpretar el concepto de expectativa en un contexto económico afectado por la incertidumbre, propuso establecer una cota superior (f^*) para la función de pagos, de modo que:

$$E^*(Y) = \sum_{j=0}^{\infty} \frac{1}{2^{j+1}} f\left(b + 2^j\right) - \left(\sum_{j=1}^{\infty} \frac{1}{2^j}\right) f(b) \leq$$

$$\leq f^* \sum_{j=1}^{\infty} \frac{1}{2^j} - \left(\sum_{j=1}^{\infty} \frac{1}{2^j}\right) f(b) = f^* + f(b) < \infty$$

En particular, consideró razonable (sin ningún tipo de justificación) fijar una cota en $f^* = 2^{24}$ dado que, a partir de ese valor se podía suponer que el dinero poseía para la gente "*de sentido común*" un "*valor moral*" constante. Es decir que, estableciendo los fundamentos de una "*teoría matemática de la*

[38] Para una crónica más completa sobre la numerosísima literatura referida al análisis en términos de probabilidad clásica de esta paradoja, ver Todhunter (1865).

utilidad" (Starmer (2000, p. 332) postuló que, a partir de 2^{24}el incremento de la utilidad de la fortuna de B era nula y, en consecuencia, su utilidad marginal era decreciente.

Posteriormente abandonó la hipótesis de la existencia de una cota superior para la función de pagos y estableció (sin ninguna justificación conceptual) que el valor moral del dinero podía ser considerado como equivalente a la raíz cuadrada de su valor matemático y propuso la definición de una versión de la esperanza matemática del juego para B en función de dicho valor moral de la ganancia, a partir de la cual obtuvo que:

$$E^*(Y) = \sum_{y=1}^{\infty} \sqrt{2^{y-1}} \frac{1}{2^y} = \frac{1}{2} \sum_{y=1}^{\infty} \frac{1}{\sqrt{2^{y-1}}} = \frac{1}{2} \frac{\sqrt{2}}{2\sqrt{2}-1}$$

De acuerdo con la condición de equidad, según la cual la esperanza matemática de ganar de B debe ser igual a su esperanza matemática de perder, concluyó que el valor de la apuesta de debía ser:

$$\left[\frac{\sqrt{2}}{2\left(\sqrt{2}-1\right)}\right]^2 = \frac{2}{12-8\sqrt{2}} = 2{,}9 < 3$$

Posteriormente, como una alternativa a esta propuesta, consideró (sin intentar tampoco en este caso ningún tipo de justificación) la posibilidad de asignar al dinero un valor moral definido como la inversa de su valor matemático.

Debe tenerse en cuenta que la pretensión de Cramer no fue dar una definición general de la función que representa el valor moral del dinero, sino demostrar que, en ciertos casos, una función de pagos que cumpliera la condición que la utilidad marginal de la fortuna de B fuera decreciente, podría conducir a una esperanza matemática finita[39]. La falta de una justificación racional y su consiguiente dosis de subjetividad motivaron el rechazo de N. Bernoulli de estas propuestas.

A partir de la consideración de la diferencia entre el "valor matemático" y el "valor moral" del dinero planteada por Cramer, D. Bernoulli (1728)(1738) propuso sustituir, en la definición de esperanza matemática de un juego, el

[39] Menger (1934) demostró rigurosamente que la condición necesaria y suficiente para poder asegurar la finitud de la esperanza matemática de un juego construida utilizando el valor moral del dinero, es que la función $f(y)$ sea acotada.

valor de la ganancia en términos absolutos por su valor relativo o utilidad e introdujo el concepto de "esperanza moral"[40]. Basándose (sin justificación) en la hipótesis que considera que los incrementos infinitesimales en la utilidad de un jugador son directamente proporcionales a los incrementos infinitesimales en su fortuna e inversamente proporcionales al monto de la fortuna inicial, es decir suponiendo que $df(y) = k\frac{dy}{y}$ (donde $f(y)$ denota la función de utilidad o "*fortuna moral*" o "*emolumentum*", y denota la "*fortuna física*" o "*summa bonorum*"[41] del jugador y $k > 0$ es un factor de proporcionalidad que depende del comportamiento del jugador) y, por lo tanto, que la utilidad marginal es tal que $\frac{df(y)}{dy} = \frac{k}{y}$, definió la utilidad de la fortuna del jugador como proporcional a su logaritmo, $f(y) = \int_b^y \frac{dx}{x} k[ln(y) - ln(b)]$ (donde b, representa la fortuna inicial de B tal que $f(b) = 0$)[42].

Que $f''(y) < 0$ implica que la utilidad marginal de y es decreciente[43]. Esta concavidad estricta de la curva le permitió a D. Bernoulli concluir que a todo juego equitativo le corresponde una esperanza moral negativa para ambos

[40] Esta intención de reconciliar el significado matemático del concepto de expectativa con la interpretación que sugiere el sentido común, condujo a los probabilistas del siglo XVIII a formular diferentes definiciones matemáticas del concepto de expectativa de acuerdo con distintas interpretaciones de la noción de razonabilidad. Contrariamente a la posición de su primo Niklaus basada en la equidad, D. Bernoulli asimiló el sentido común a la prudencia, generando la primera expresión del concepto económico de utilidad (ver Arrow, K. (1951), Samuelson, P. (1977)).

[41] Ambas denominaciones de acuerdo con las nomenclaturas de Laplace (1812) y D. Bernoulli (1728).

[42] En términos de la teoría del valor asociada a su interpretación de expectativa, D. Bernoulli (1738) definió la fortuna como "*... todo lo que pueda contribuir a la adecuada satisfacción de todos los deseos del jugador*" (p. 24). Para que la función de utilidad esté definida en todo punto se debe verificar que *y >0* ("*Nadie puede alegar que no posee absolutamente nada (...) Para la gran mayoría la más importante de sus posesiones debe ser su capacidad productiva*" (p. 25)).

[43]Fechner, G.I. (1866) obtuvo una justificación de la hipótesis logarítmica de D. Bernoulli que culminó en la conocida como ley de Weber-Fechner. A partir de la propuesta de Weber, según la cual el incremento de un estímulo es proporcional al estímulo, $\frac{dR}{R} = k$, Fechner interpretó a la constante k como la "*unidad de la sensación*", de modo que $dS = c\frac{dR}{R}$ o $S = c\, ln\frac{R}{R_0}$ (donde R_0 representa el "*umbral inferior de la sensación*"). Para un análisis de los alcances de la fórmula de Fechner, ver Guilford (1936), Stigler, G.J. (1965).

jugadores, debido a que la utilidad del monto de la ganancia posible es menor que la utilidad del monto de la pérdida posible[44].

Suponiendo que a las ganancias sucesivas, $g_1, g_2, g_3, \ldots$, les correspondan, respectivamente, las probabilidades $p_1, p_2, p_3, \ldots$, la utilidad esperada o esperanza moral del juego para B queda definida por:

$$E^*(Y) = kp_1 ln(b + g_1) + kp_2 ln(b + g_2) + kp_3 ln(b + g_3) + \cdots - ln(b)$$

De esta expresión se obtiene que la "*fortuna física*" que genera dicha utilidad esperada está dada por:

$$y(b) = E[f(y) - b] = k^*(b + g_1)^{p_1}(b + g_2)^{p_2} \ldots b^*$$

(donde $b^* = k\ ln(b) \sum_{j=1}^{\infty} p_j = k\ ln(b)$).

En consecuencia, en el problema de N. Bernoulli la "*fortuna física*" del jugador B queda definida como:

$$y(b) = (b + 1)^{\frac{1}{2}}(b + 2)^{\frac{1}{4}}(b + 2^2)^{\frac{1}{8}} \ldots - b^*$$

y la esperanza moral del juego para el jugador B como:

$$E^*(Y) = ln[y(b)] = \frac{1}{2} ln(b + 1) + \frac{1}{2^2} ln(b + 2) + \frac{1}{2^3} ln(b + 2^2) + \cdots - b^*$$

De acuerdo con el criterio de d'Alembert (según el cual, dada una sucesión $\{a_j\}(a_j > 0, j = 1,2, \ldots)$ tal que $\lim_{j \to \infty} \frac{a_{j+1}}{a_j} < 1$, se verifica que $\lim_{n \to \infty} \sum_{j=1}^{n} a_j < \infty$) y teniendo en cuenta que:

$$\lim_{n \to \infty} \frac{\frac{1}{2^{n+1}} ln(b + 2^n)}{\frac{1}{2^n} ln(b + 2^n)} = \lim_{n \to \infty} \frac{1 + \frac{ln\left(\frac{b}{2^n} + 1\right)}{ln(2^n)}}{1 + \frac{ln\left(\frac{b}{2^n} + \frac{1}{2}\right)}{ln(2^n)}} = \frac{1}{2}$$

se puede concluir que $E^*(Y) < \infty$. En particular, para $b = 0$ resulta que[45]:

[44] La denominación de juego o paradoja de San Petersburgo -atribuida a d'Alembert- se debe a que la solución propuesta por D.Bernoulli ("*Specimen theoriæ novæ de mensura sortis*" (1730-31)) fue publicada en *Commentarii Academiæ Scientiarum Imperialis Pertropolitanæ (1738).*

$$E^*(Y) = ln[y(0)] = ln(2) \sum_{j=1}^{\infty} \frac{j-1}{2} = \frac{ln(2)}{2^2\left(1-\frac{1}{2}\right)^2} = ln(2)$$

y, por lo tanto, que $y(0) = 2$. Asimismo, se demuestra en forma inmediata que:

$$\frac{y(b)}{y(0)} = \prod_{j=1}^{\infty} \left(1 + \frac{b}{2^{j-1}}\right)^{\frac{1}{2^j}} - \frac{b}{2} < \prod_{j=1}^{\infty} (1+b)^{\frac{1}{2^j}} - \frac{b}{2} < 1 + \frac{b}{2}$$

Lo que permite concluir que, para $b < \infty$, el producto converge y que, para valores pequeños de b, $y(b)$ crece lentamente.

Como corolario de la propuesta de D. Bernoulli se puede concluir entonces: **i)** que, si b es infinita las esperanzas matemática y moral son equivalentes[46]; **ii)** que la fortuna de B define una cota superior a su esperanza moral y que, en consecuencia, si es infinita, entonces $E^*(Y) = \infty$; **iii)** que, a partir del principio de equidad en términos de utilidad, según el cual la utilidad negativa de la pérdida debe ser igual a la utilidad positiva de la ganancia, la apuesta de B debe ser menor que la esperanza matemática del juego[47]; **iv)** que la utilidad logarítmica es una condición suficiente, pero no necesaria, para explicar una tasa de aversión al riesgo mayor que la postulada por Cramer y está afectada por el mismo grado de subjetividad[48] y **v)** que este análisis puede

[45] Una revisión del famoso “problema de los puntos” planteado por A. Gombaud, Caballero de Méré, permite concluir que la solución propuesta por Pascal y Fermat admite una interpretación en términos del problema de N. Bernoulli finito, suponiendo la linealidad de la función de utilidad de A o de B.

[46] Samuelson (1934) demuestra que, para apuestas suficientemente pequeñas, la concavidad de la función de utilidad, considerada en la esperanza moral, se asimila al de la función lineal de la esperanza matemática.

[47] A partir de esta conclusión D. Bernoulli demostró sus tres teoremas fundacionales sobre la teoría del riesgo.

[48]. Con respecto a la actitud de aversión al riesgo, Condorcet postula que, en contradicción con el principio de equidad, los jugadores tienen tendencia a subvaluar circunstancias con probabilidades altas. Tienden a preferir una ganancia relativamente grande con una pequeña probabilidad a una ganancia pequeña con un alto grado de probabilidad. Esta conjetura, que admite una justificación formal sólo si se verifica que $f^{(3)}(y) > 0$ en el entorno de b, no se verifica para una función de utilidad cuadrática. En la moderna teoría de optimización de carteras, se considera que una función de utilidad de la forma $\frac{1}{a}[f(y)]^a$ representa, en general, una versión del principio de aversión al riesgo más realista que las de D. Bernoulli y Cramer, pero igualmente subjetiva.

generalizarse a funciones de utilidad lineales y aún convexas que indican una actitud neutral y de atracción al riesgo, respectivamente.

N. Bernoulli consideró que la solución de D. Bernoulli implicaba una simplificación excesiva de la actitud del jugador frente al riesgo, que la introducción del concepto de esperanza moral era un argumento "ad hoc" y que la utilización de la función de utilidad como representación de la aversión al riesgo no era adecuada para explicar la discrepancia entre la interpretación matemática de la expectativa y el comportamiento que indica el sentido común y sostuvo, en consecuencia, que la solución debía plantearse, inevitablemente, a partir de la definición de esperanza matemática.

En coincidencia con N. Bernoulli respecto del retorno al concepto de esperanza matemática, Fontaine (1764), basó su propuesta en el principio que la ganancia del jugador B no puede superar la fortuna (a) de A. Suponiendo que $2^n < a + S_B$, es decir, que $n < \log_2(a + S_B)$ (donde S_B denota la apuesta de B), la esperanza matemática del juego para B será:

$$E(Y) = \sum_{j=1}^{n} \frac{1}{2^j}\left(2^{j-1} - S_B\right) \qquad (j = 1,2,\dots,n)$$

$$E(Y) = \frac{a}{2^{j+1}} \qquad (j = n+1, n+2, \dots)$$

De modo que, cuando n aumenta indefinidamente, la esperanza matemática que garantizala equidad del juego será tal que:

$$\lim_{n\to\infty}\left[\frac{n}{2} - \left(1 - \frac{1}{2^n}\right)S_B + \frac{1}{2^{n+1}}2a\right] = 0$$

y la apuesta de B será $S_B = \frac{n2^{n-1}+a}{2^n-1}$.

Esta solución, basada en el supuesto de finitud del capital del jugador A (seguida, entre otros, por Poisson, Catalan y von Mises) fue rechazada por la interpretación Platónica de Bertrand (1889), en parte por Keynes (1921) y replanteada por Shapley, L.S. (1972) y Samuelson (1977).

Si bien admitió que el enfoque probabilístico convencional no consideraba adecuadamente los aspectos del comportamiento psicológico del jugador ante la incertidumbre, d'Alembert (1768) también coincidió con N. Bernoulli en que la solución del problema debería basarse exclusivamente en la definición de la esperanza matemática.

Cuestionó, en general, todas las soluciones que dependían de las fortunas de los jugadores y postuló que la componente más importante del concepto de esperanza matemática del juego no era la ganancia sino la probabilidad de obtenerla. Que la sobreestimación de la esperanza matemática, que en general, conducía a una interpretación errónea del comportamiento del jugador se debía fundamentalmente a la sobrevaluación de las probabilidades[49].

Sus críticas se refirieron fundamentalmente a la injustificable pretensión de extender las propiedades matemáticas de la probabilidad al mundo físico, en particular, las hipótesis de equiprobabilidad y de independencia de las "tiradas" sucesivas, y a la utilización de las mismas ponderaciones para las probabilidades y para los resultados del juego, en vez de subordinar estos a los valores de probabilidad, empleando ponderaciones no-lineales[50]. Este planteo según el cual "*... en la naturaleza nunca se verifica que los efectos sean siempre y constantemente los mismos*" (d'Alembert (1773, p. 301), lo condujo a proponer en un juego de "cara" y "ceca" la siguiente (injustificada) definición de las probabilidades de obtener "cara" en la n-ésima "tirada" $\frac{1}{2^n(1+cn^2)}$ (donde c denota una constante arbitraria) en vez de $\frac{1}{2^n}$ y, en consecuencia, una esperanza matemática del juego para el jugador B de la forma[51]:

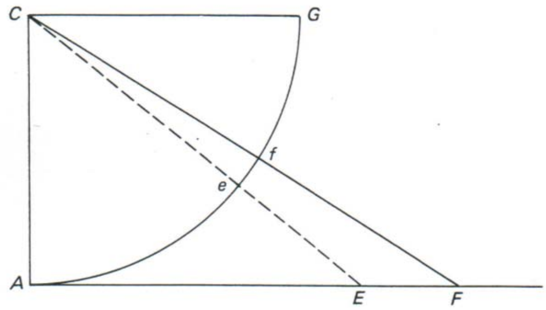
---

[49] d'Alembert (1768) (para quien el cálculo de probabilidades constituía fundamentalmente una descripción de la psicología común a todos los individuos frente a la incertidumbre) propuso varias expresiones alternativas para definir la expectativa y, en consecuencia, la apuesta según la situación financiera del jugador utilizando el concepto de esperanza matemática, pero no realizó ningún análisis comparativo entre ellas. Probablemente esto se debió a su conclusión que la evaluación del riesgo por un jugador según fuera de período corto o largo era totalmente subjetiva ("*¿Cómo comparar un riesgo presente con una ganancia remota desconocida? A este respecto el análisis de las correspondientes aleatoriedades no nos proporciona ninguna enseñanza*" (pp. 33-34)). Una conjetura que Buffon y Condorcet creyeron resolver a partir de la propuesta de definición de una "*unidad de riesgo psicológico mínimo*".

[50] Yaary (1987) demostró que la concavidad de la función de utilidad y la no-linealidad de las ponderaciones de las probabilidades constituían explicaciones igualmente válidas de la actitud de aversión al riesgo.

[51] d'Alembert propuso tres variantes (igualmente injustificadas) de la función de probabilidades de obtener "cara": **i)** $\frac{1}{2^{n(1+c)}}$ (que permite calcular el valor de la constante para cualquier apuesta); **ii)** $\frac{1}{2^{n+c(n-1)}}$ (que admite la equiprobabilidad de los resultados para $n = 1$) y **iii)** $\frac{1}{2^n\left[1+\frac{c}{(k-n)^2}\right]}$ (que tiende a cero cuando n aumenta indefinidamente).

$$E(Y) = \sum_{j=1}^{\infty} \frac{2^{j-1}}{2^j(1+cj^2)} = \frac{1}{2}\sum_{j=1}^{\infty} \frac{1}{1+cj^2} < \infty$$



Esta curiosa definición arbitraria de la probabilidad lo condujo a una (más curiosa) demostración que surge de representar la suma del segundo miembro de $E(Y)$ mediante un cuarto de circunferencia (ver Fig. 2), donde $CA = \sqrt{\frac{1}{c}}$ y los segmentos sobre la tangente $AE = n$ y $AF = n + 1$. Luego, denotando por u y v a los ángulos ACE y ACF, respectivamente y, dado que $tg(u) = n\sqrt{c}$ y $tg(v) = (n + 1)\sqrt{c}$, se obtiene que $tg(v) - tg(u) = \sqrt{c}$ y, dado que $CEcos(u) = CFcos(v) = c^{-\frac{1}{2}}$y, por lo tanto que $\frac{CF}{CE} = \frac{cos(u)}{cos(v)}$, se obtiene que:

$$sen(ef)\frac{CF}{CE} = sen(u - v)\frac{cos(u)}{cos(v)} =$$

$$= [sen(v)cos(u) - sen(u)cos(v)]\frac{cos(u)}{cos(v)} =$$

$$= [tg(v) - tg(u)]cos^2(u) = \sqrt{c}cos^2(u)$$

Dividiendo por $EF = AF - AE = tg(v) - tg(u) = \sqrt{c}$, se obtiene que $\frac{1}{EF}sen(ef)\frac{CF}{CE} = cos^2(u)$. Ahora bien, dado que $\frac{1}{1+cn^2} = \frac{1}{EF}sen(ef)\frac{CF}{CE}$, $EF = 1$ y $\lim_{n\to\infty}\frac{CF}{CE} = 1$, se puede concluir que, a partir de un cierto rango, la serie $\sum_{j=1}^{\infty}\frac{1}{1+cj^2}$ admite una aproximación por una función seno y que, reemplazando estos senos por los arcos correspondientes, se comprueba la validez del supuesto que la serie $\sum_{j=1}^{\infty} \frac{1}{1+cj^2}$ puede ser representada por un cuarto de circunferencia y, en consecuencia que, considerando una probabilidad de obtener "cara" en la n-ésima "tirada" igual a $\frac{1}{2^j(1+cj^2)}$, la esperanza matemática del juego para B es finita.

El aporte de Buffon a la solución del problema de San Petersburgo incluye un conjunto de propuestas heterodoxas desarrolladas en el período que va desde sus discusiones con Cramer (1728), hasta la publicación del "*Essai d'arithmétique morale*" (1777). Con la intención de obtener una expresión general para la función de utilidad, Buffon propuso su propia interpretación de la esperanza moral en la que la relación logarítmica entre la utilidad y la fortuna original del jugador planteada por Bernoulli es sustituida por una relación de la forma $f(y) = \frac{g}{b+g}$ (donde $f(y)$ denota la función de utilidad, g denota la ganancia y b representa la fortuna original de B). A partir de la cual obtuvo que $E^* = q\frac{g}{b} - p\frac{g}{g+b}$ (donde p y q denotan, respectivamente, las probabilidades de ganar o perder el juego)[52].

Alternativamente al planteo precedente y, a fin de obtener una interpretación más adecuada del principio de aversión al riesgo, Buffon propuso (en forma absolutamente subjetiva) una representación de la función de utilidad que incluía un umbral, al que denominó "*el nivel de vida habitual*" del jugador. De modo que arriesgar una parte del ingreso por encima de dicho umbral era asimilable a arriesgar un monto infinito para el cual ninguna expectativa podía significar una compensación adecuada[53]. Independientemente de D. Bernoulli, Buffon aplicó esta función de utilidad para demostrar los beneficios que

---

[52] Si bien Buffon fue más conocido como naturalista que como matemático, es necesario destacar sus trabajos sobre las aplicaciones de la teoría de la probabilidad en su monumental "*Histoire naturelle*", de la cual el "*Essai d'arithmétique morale*" es un suplemento. Esta obra marcó el comienzo de una transición (completada posteriormente por Laplace y Poisson) entre la interpretación netamente deductiva de los probabilistas del siglo XVIII y la orientación frecuencicista de Quetelet

[53] Mediante un controvertido desarrollo sumamente complejo, vinculado a la incorporación del principio de aversión al riesgo en el cálculo de la esperanza moral, Weirich (1984) obtuvo una demostración de este supuesto de Buffon de existencia de una cota superior en el monto de la apuesta inicial (ver Martin (2008)).

acarrea la atomización del riesgo y concluir, como aquél, que todo juego equitativo es un "*lost game*"[54].

En otro intento de resolución del problema de la convergencia de la esperanza matemática definida por una progresión geométrica, Buffon (1777) introdujo la noción de "*ganancia moral*" según la cual, si en el problema de N. Bernoulli se consideraba una función de pagos definida como un múltiplo de $2 - \varepsilon$ (donde $0 < \varepsilon < 2$) entonces, para el jugador B, $E(Y) < \infty$.

Asimismo, en coincidencia con la opinión de d'Alembert respecto de la sobrevaluación de las probabilidades en el cálculo de la esperanza matemática del juego, propuso una definición de imposibilidad moral o imposibilidad por analogía ("*...un punto medio entre la duda y la certeza física*") basada en la hipótesis (innegablemente subjetiva, a pesar de su intento de justificación biométrica) que la significatividad de una probabilidad poseía una cota inferior igual a $\left(\frac{1}{2}\right)^{10}$. Lo cual le permitió considerar como nulas a todas las probabilidades menores que 1/10.000 y, por lo tanto, concluir que el valor esperado del juego de San Petersburgo era igual a 6,5[55].

Más allá de las mencionadas propuestas acerca del comportamiento de la función de utilidad y la función de probabilidades, la importancia de Buffon en esta crónica de los intentos de solución del problema de San Petersburgo radica en la adopción del principio que el conocimiento se adquiere exclusivamente por la observación o por la analogía, el cual lo condujo a la conclusión que, sólo si el juego era repetido efectivamente una gran cantidad de veces era posible asegurar una aproximación entre la cantidad ganada por B y el monto de su apuesta y, por lo tanto, aproximar la apuesta de B a su esperanza matemática.

---

[54] Posteriormente, Laplace (1795), Lacroix (1816), Öttinger (1848) y Liagre (1852) demostraron rigurosamente la equivalencia de los resultados de D. Bernoulli y Buffon.

[55] Bufon (1777): "*A partir de los datos que proporcionan las tablas de mortalidad, se puede deducir que hay una contra 10249 posibilidades que un hombre de 56 años fallezca dentro de las siguientes 24 horas. Ahora bien, a pesar de esto, los hombres de dicha edad, en la cual la razón asume toda su madurez y la experiencia toda su fuerza, no teme morir al día siguiente. Luego, se puede concluir que toda probabilidad igual o menor a 1/10000 puede ser considerada nula y que todo temor o toda expectativa que se encuentre por debajo de este valor no debe afectar ni preocupar el corazón ni el pensamiento"* (p. 106)). Ver Landro (2010).

A partir de este tratamiento netamente empirista, basado en la información proporcionada por sucesiones de observaciones repetidas en igualdad de condiciones del juego propuesto por N. Bernoulli, rechazó por irrealizable la posibilidad de la duración infinita del juego y comparó el resultado de sus observaciones con la distribución binomial según la cual, dada una sucesión de 2^j juegos de San Petersburgo, el número esperado de juegos en los que el resultado "cara" se obtiene en la j-ésima "tirada" está dado por $\sum_{i=1}^{j} 2^{j-i} = 2^j - 1$ y la ganancia esperada de B es $\sum_{i=1}^{j} 2^{j-i} 2^{j-1} = 2^j \frac{j}{2}$. Si bien el experimento no puede considerarse concluyente, de acuerdo con una extensión de este razonamiento y suponiendo un número n^* suficientemente grande de juegos, se demuestra que la ganancia esperada de B en un juego equitativo puede ser aproximada por $E(B/n^*) = \frac{n}{2}\log_2(n^*)$ (ver Lupton (1890))[56].

Una expresión límite rigurosa de este argumento de Buffon (que contrariamente a los otros intentos mencionados aquí, constituye el único análisis de un juego de San Petersburgo no modificado) puede ser obtenida a partir de una ley débil de los grandes números debida a Feller (1936)(1945) según la cual, dada una sucesión $\{Y_i\}$ de variables aleatorias iid, que representan las posibles ganancias de B, con distribución de probabilidades de la forma $p(Y_i = 2^{i-1}) = 2^{-i}$ $(i = 1,2,3, \dots)$, se verifica que, cuando n^* aumenta indefinidamente:

$$\frac{\sum_{i=1}^{n^*} Y_i}{E(B/n^*)} \xrightarrow{p} 1$$

es decir que, cuando $n^* \to \infty$, la posible ventaja de un jugador sobre el otro tiende a desaparecer. Como corolario de este teorema se puede concluir que el valor esperado $E(B/n^*)$ converge a una función asintóticamente equivalente de la forma:

$$E^*(B/n^*) = \frac{n^*}{2}\log_2[n^* \log_2(n^*)]$$

[56] Como una generalización de la definición asintótica de Buffon, a partir de la aplicación de técnicas de simulación, algunos autores infirieron: **i)** que los resultados observados presentan una asimetría infinitamente positiva y, en consecuencia, una distribución no-Normal; **ii)** que su varianza infinita revela su naturaleza fractal e implica el no cumplimiento de los postulados de la ley de los grandes números; **iii)** que la esperanza (matemática o moral) no es infinita, sino indefinida; **iv)** que el promedio ponderado de las ganancias constituye un estimador pobre del valor esperado del juego y **v)** que una solución de la paradoja puede ser obtenida mediante la aplicación de la mediana como estimador robusto de la tendencia central.

Como una extensión del teorema de Feller, Martin-Löf (1985) demostró que la variable $\frac{1}{n}\left[\sum_{i=1}^{n^*} Y_i - E(B/n^*)\right]$ se distribuye de acuerdo a una función infinitamente divisible $g(u)$ tal que, para $n^* = 2^n$ y $n \to \infty$, asume la forma $\lim_{m\to\infty}[1 - g(2^m)] \approx (1{,}8)2^{-m}$. Lo cual implica que la probabilidad de que la apuesta $S_B = 2^m + n$ resulte insuficiente para pagar la ganancia $\sum_{i=1}^{n^*} Y_i$ (obtenida por B al cabo de $n^* = 2^n$ juegos) es igual a $(1{,}8)2^{-n}$ (≈ 0).

De los teoremas anteriores se puede concluir que, bajo el supuesto que $\sigma^2(Y) < \infty$, la condición necesaria y suficiente para el cumplimiento de la equidad asintótica de Buffon para un número suficientemente grande de repeticiones independientes, es que la apuesta de B no sea constante, sino una función del número de partidas ya jugadas.

En la segunda parte de su "*Mémoire sur le calcul des probabilités*" (1784) Condorcet plantea una severa crítica a la propuesta de Buffon sobre el concepto de certeza moral considerando que "*...es esencialmente incorrecta ya que conduce a una mezcla de dos cosas de naturaleza diferente - probabilidad y certeza; una idea similar a confundir una asíntota a una curva con una tangente en un punto muy distante. Tales propuestas no pueden ser aceptadas en las ciencias exactas sin destruir su rigor*" (Todhunter (1865), p. 347).

El fundamento de su propuesta ((1784)(1785a)(1785b)(1789)(1805)) -que, en cierta forma resumió las contribuciones de sus predecesores- se basa, en coincidencia con d'Alembert, en el concepto de esperanza matemática interpretada como una representación cuantitativa común de los diferentes estados en que se encuentran los individuos que enfrentan riesgos diferentes, involucrando en su definición consideraciones de carácter legal, económico y psicológico.

Igual que Buffon, Condorcet adoptó el concepto de esperanza matemática como promedio obtenido a partir de una sucesión de repeticiones del evento en igualdad de condiciones y la equivalencia asintótica entre la apuesta equitativa y la esperanza matemática del juego, y dedicó sus esfuerzos a la determinación del número finito de "tiradas" necesario para lograr la máxima aproximación a la equidad del juego, es decir, la máxima aproximación entre las ventajas de los jugadores.

Denotando por n al número máximo de "tiradas", la esperanza matemática del juego para B queda definida por $E(Y) = \sum_{j=1}^{n} 2^{j-1}\frac{1}{2^j} = \frac{n}{2} + 1$. De acuerdo con la propuesta de Condorcet, el jugador B comienza a ganar cuando el resultado "cara" ocurre en la n^*-ésima "tirada", tal que $2^{n^*-1} > \frac{n}{2} + 1$, es decir, tal que $n^* > \log_2\left(\frac{n}{2}+1\right)$. La equivalencia entre las ventajas de los jugadores se produce cuando el resultado "cara" ocurre en la tirada $n^* = \log_2\left(\frac{n}{2}+1\right)$. En cualquier otro caso, B pierde. De modo que la probabilidad de Bde ganar (es decir, de perder A) está dada por $\sum_{j=n^*+1}^{n}\frac{1}{2^j}$, la probabilidad de equilibrio (es decir, la probabilidad de que ni A ni B ganen o pierdan) está dada por $\frac{1}{2^{n^*}}$ y la probabilidad de que B pierda está dada por $\sum_{j=1}^{n^*}\frac{1}{2^j}$. Luego resulta que la ganancia máxima de A será $S_B - 1 = \frac{n}{2}$, si el resultado "cara" ocurre en la primera "tirada" y su pérdida máxima será $2^n - S_B$. Este resultado le permitió a Condorcet demostrar su tesis que la definición convencional de esperanza matemática de un juego, interpretada como un promedio finito de las ganancias de B, consideradas como los valores posibles de una variable aleatoria discreta, es la única que permite lograr "*...la mayor aproximación posible entre dos condiciones esencialmente diferentes de los jugadores*" (p. 654).

Contrariamente a los probabilistas del siglo XIX -quienes, siguiendo los trabajos de Buffon y Condorcet, basaron sus propuestas en el concepto de esperanza matemática y en un comportamiento asintótico de acuerdo con los postulados de la ley de los grandes números y consideraron a la paradoja de San Petersburgo como una confirmación de que la única interpretación válida del concepto de probabilidad era la frecuencista, Laplace, retomando -y generalizando- la propuesta sobre la teoría del riesgo de D. Bernoulli, demostró que la esperanza matemática puede ser considerada como el límite de la esperanza moral no sólo cuando se particiona infinitamente la fortuna, sino también cuando la división del riesgo resulta infinita.

Laplace (1795) analizó la cuestión de la determinación de la apuesta en el problema de N. Bernoulli a partir de la definición clásica de equidad, de una forma vinculada estrechamente con el espíritu de la propuesta de d'Alembert, en particular, en lo referido a la supuesta incorrecta aplicación del principio de indiferencia ante la falta de información acerca del comportamiento de las frecuencias relativas de los resultados[57]. No obstante, en su "*Théorie analitique des probabilités*" (1812) Laplace abandonó esta adhesión a la definición Huygensiana de esperanza matemática por la Bernoulliana de esperanza moral.

Denotando por S_B al valor de la apuesta de B, considerando una función de pago de la forma 2^n si se obtiene "cara" en la n-ésima "tirada" y 0 si este resultado no ocurre al cabo de las n "tiradas"' y denotando por b a la fortuna

[57] Esta postura Bayesiana de Laplace ejerció una notable influencia en las interpretaciones del concepto de probabilidad de d'Alembert y Condorcet.

mínima de utilidad nula de B, Laplace obtuvo que la utilidad esperada de B estaba dada por:

$$E^{*}(Y)=\frac{1}{2}k\,ln(b-S_B+2)+\frac{1}{2^2}k\,ln(b-S_B+2^2)+\cdots+$$

$$+\frac{1}{2^n}k\,ln(b-S_B+2^n)+\frac{1}{2^n}k\,ln(b-S_B)+ln(h)$$

De acuerdo con la condición de equidad, la utilidad esperada de la fortuna de B debe ser la misma antes o después del juego, es decir, la expresión precedente debe ser igual a $k\,ln(b)+ln(h)$. De modo que, suprimiendo los términos k y $ln(y)$ y dividiendo m. a m. por $b-S_B$, debe verificarse que:

$$1+\frac{S_B}{b-S_B}=\left(1+\frac{2}{b-S_B}\right)^{\frac{1}{2}}\left(1+\frac{2^2}{b-S_B}\right)^{\frac{1}{2^2}}\ldots\left(1+\frac{2^n}{b-S_B}\right)^{\frac{1}{2^n}}$$

Esta serie satisface el criterio de d'Alembert, es decir, se verifica que:

$$\left(1+\frac{2^j}{b-S_B}\right)^{\frac{1}{2^j}}<\left(1+\frac{2^{j-1}}{b-S_B}\right)^{\frac{1}{2^{j-1}}}$$

o, lo que es lo mismo, que:

$$1+\frac{2^{j+1}}{b-S_B}+\frac{2^j}{(b-S_B)^2}>1+\frac{2^{j+1}}{b-S_B}$$

Por otra parte, dado que:

$$\lim_{n\to\infty} ln\left(1+\frac{2^n}{b-S_B}\right)^{\frac{1}{2^n}}=\lim_{n\to\infty}\frac{n\,ln(2)}{2^n}+\frac{1}{2^n}ln\left(\frac{1}{b-S_B}+\frac{1}{2^n}\right)=0$$

se puede concluir que $\lim_{n\to\infty}\left(1+\frac{2^n}{b-S_B}\right)^{\frac{1}{2^n}}=1$. Asimismo, dado que la serie $1+\frac{S_B}{b-S_B}$ puede ser expresada de la siguiente forma:

$$ln\left(1+\frac{S_B}{b-S_B}\right)=\sum_{j=1}^{n-1}\frac{1}{2^j}ln\left(1+\frac{2^j}{b-S_B}\right)+\sum_{j=n}^{\infty}\frac{1}{2^j}ln\left(1+\frac{2^j}{b-S_B}\right)=$$

$$=\sum_{j=1}^{n-1}\frac{1}{2^j}ln\left(1+\frac{2^j}{b-S_B}\right)+\sum_{j=n}^{\infty}\frac{1}{2^j}ln\left(1+\frac{1}{b-S_B}\right)$$

$$+\sum_{j=n}^{\infty}\frac{n\,ln(2)}{2^j}+\sum_{j=n}^{\infty}\frac{1}{2^j}ln\left(1+\frac{b-S_B}{2^j}\right)=$$

$$=\sum_{j=1}^{n-1}\frac{1}{2^j}ln\left(1+\frac{2^j}{b-S_B}\right)+$$

$$+ln\left(\frac{1}{b-S_B}\right)\sum_{j=n}^{\infty}\frac{1}{2^j}+ln(2)\left[\frac{1}{2}\sum_{j=1}^{\infty}j\left(\frac{1}{2}\right)^{j-1}-\frac{1}{2}\sum_{j=1}^{n-1}j\left(\frac{1}{2}\right)^{j-1}\right]+$$

$$+\sum_{j=n}^{\infty}\frac{0{,}4342945}{2^j}ln\left(1+\frac{b-S_B}{2^j}\right)=$$

$$=\sum_{j=1}^{n-1}\frac{1}{2^j}ln\left(1+\frac{2^j}{b-z}\right)+\frac{1}{2^{n-1}}ln\left(\frac{1}{b-z}\right)+\frac{(n+1)ln(2)}{2^{n-1}}+$$

$$+\frac{0{,}4342945}{3}\frac{b-z}{2^{n-2}}$$

y, teniendo en cuenta que:

$$\sum_{j=n}^{\infty}\frac{0{,}4342945}{2^j}ln\left(1+\frac{b-S_B}{2^j}\right)\approx\sum_{j=n}^{\infty}\frac{0{,}4342945}{2^j}\frac{b-S_B}{2^j}=$$

$$=(0{,}4342945)(b-S_B)\sum_{j=n}^{\infty}\frac{1}{2^j}=$$

$$=(0{,}4342945)(b-S_B)\left(\sum_{j=1}^{\infty}\frac{1}{2^j}-\sum_{j=1}^{n-1}\frac{1}{2^j}\right)=\frac{0{,}4342945}{3}\frac{b-S_B}{2^{n-2}}$$

se obtiene la siguiente aproximación:

$$ln\left(1+\frac{S_B}{b-S_B}\right)$$

$$\approx\sum_{j=1}^{n-1}\frac{1}{2^j}ln\left(1+\frac{2^j}{b-z}\right)+\frac{1}{2^{n-1}}ln\left(\frac{1}{b-z}\right)+$$

$$+\frac{(n+1)ln(2)}{2^{n-1}}+\frac{0,4342945}{3}\frac{b-z}{2^{n-2}}$$

A partir de esta demostración Laplace obtuvo que, para $n = 11$ y $b - S_B = 100$, B debería poseer una fortuna $b = 107,89$ y apostar $S_B = 7,89$; para $b - S_B = 200$, B debería poseer 208,78 y apostar 8,78 y que, si la fortuna de B fuera infinita, entonces su apuesta debería ser infinita.

Posteriormente, Menger (1934) demostró rigurosamente que esta conclusión de Laplace se verifica si y, sólo si, la función $f(y)$ es acotada y que, en caso contrario, siempre será posible definir un juego de San Petersburgo para el cual la esperanza moral es infinita aún en el caso en que B poseyera una fortuna finita[58]. Debe tenerse en cuenta que, si bien estos resultados sobre las funciones de utilidad acotadas permitieron asegurar la finitud de la esperanza moral, no lograron justificar la diferencia entre la esperanza matemática de un juego y el comportamiento habitual de los jugadores[59].

A partir de su teoría de la esperanza moral como complemento de la esperanza matemática, Laplace (1812) obtuvo la siguiente versión más rigurosa de los tres teoremas de D. Bernoulli ya mencionados:

1.- Sea un jugador A, sea a su fortuna, p su probabilidad de ganar en un juego, S_A su apuesta y $S_B = \frac{q}{p}S_A$ (donde $q = 1 - p$) la apuesta de su contrincante (B). Si A gana, su fortuna será $a + \frac{q}{p}S_A$, con probabilidad p, si pierde, se transformará en $a - S_A$, con probabilidad q. De modo que la esperanza moral del juego para A será $E^*(Y) = \left(a + \frac{q}{p}S_A\right)^p (a - S_A)^q$ y la esperanza matemática $E(Y) = \left(a + \frac{q}{p}S_A\right)p + (a - S_A)q = a$. Dividiendo m. a m. por a

[58] A este respecto propone el siguiente ejemplo: Supóngase que el jugador B posea una fortuna b, al cabo de la n-ésima "tirada" su ganancia sería $(be^{2n} - b)$ y su esperanza matemática $E(Y) = \sum_{j=1}^{\infty}\frac{1}{2^j}\left(be^{2j} - b\right) = \infty$. De acuerdo con la hipótesis de D. Bernoulli, la utilidad de cada ganancia será $k\ ln\left(\frac{b+b2^{2j}-b}{2!}\right) = k\ ln\left(e^{2j}\right)$. De donde se obtiene que, dado que la función de utilidad es lineal, la esperanza matemática de las ganancias será $E(Y) = \sum_{j=1}^{\infty}\frac{1}{2^j}k2^j = k + k + k + \cdots = \infty$, que la esperanza moral, $E^*(Y) = (b.e.e.e\ \dots) - b$, será infinita aún cuando b sea finita, generando lo que Menger denominó la "*súper paradoja de San Petersburgo*" (ver Aumann (1977) , Samuelson (1977), Martin (2008)).

[59] Este trabajo de Menger despertó el interés de von Neumann en la teoría de la utilidad e influyó, a su vez, sobre el desarrollo del sistema axiomático de von Neumann-Morgenstern (1944), en el que el principio de la esperanza moral surge como un teorema, como corolario del cual se puede demostrar que las funciones de utilidad consistentes con dicho sistema forman una clase en la que cualquier función de utilidad puede ser expresada como una función lineal de cualquier otra de la misma clase. Los continuadores inmediatos de esta teoría de la utilidad fueron, entre otros, Strotz (1953), Ellsberg (1954), Savage (1954), Chipman (1969).

y aplicando m. a m. el operador logaritmo en la ecuación de la esperanza moral, se obtiene que:

$$ln\left(\frac{1}{a}E^*(Y)\right) = p\ ln\left(1+\frac{q}{p}\frac{S_A}{a}\right) + q\ ln\left(1-\frac{S_A}{a}\right) < 0$$

Lo cual permite concluir que, en un juego equitativo, $E^*(Y) < a$, es decir que $E^*(Y) < E(Y)$ o, lo que es lo mismo, que la apuesta de B debe ser menor que su ganancia esperada.

2.- Sea un comerciante A que posee una fortuna igual a 1 y que desea despachar una suma de dinero ε en un barco cuya probabilidad de arribar a destino es igual a p. De acuerdo con lo demostrado más arriba, su esperanza matemática y su esperanza moral serán tales que $(1+p\varepsilon) > (1+\varepsilon)^p$. Supóngase que el comerciante divida la suma ε en partes iguales para ser despachadas en n barcos. Aplicando el esquema de pruebas repetidas de Bernoulli, la esperanza moral de la operación puede ser expresada como:

$$E_n^*(Y) = k\left[p^n ln(1+\varepsilon) + np^{n-1}q\ ln\left(1+\frac{n-1}{n}\varepsilon\right) + \right.$$

$$\left. + \frac{n(n-1)}{2}p^{n-2}q^2 ln\left(1+\frac{n-2}{n}\varepsilon\right)\right] + ln(h) =$$

$$= kp\int\left[\frac{p^{n-1}}{1+\varepsilon} + \frac{(n-1)p^{n-2}q}{1+\frac{n-1}{n}\varepsilon} + \frac{(n-1)(n-2)p^{n-3}q^2}{(1)(2)\left(1+\frac{n-2}{n}\varepsilon\right)} + \cdots\right]d\varepsilon + ln(h)$$

(donde $q = 1 - p$). En particular, para $n = 1$, será:

$$E_1^*(Y) = kp\int\frac{d\varepsilon}{1+\varepsilon} + ln(h) = kp\int\frac{d\varepsilon}{1+\varepsilon}(p+q)^{n-1} + ln(h) =$$

$$= kp\int\left[\frac{p^{n-1}}{1+\varepsilon} + \frac{(n-1)p^{n-2}q}{1+\frac{n-1}{n}\varepsilon} + \frac{(n-1)(n-2)p^{n-3}q^2}{(1)(2)\left(1+\frac{n-2}{n}\varepsilon\right)} + \cdots\right]d\varepsilon + ln(h)$$

De modo que:

$$E_n^*(Y) - E_1^*(Y) = kpq\frac{n-1}{n}\int\frac{\varepsilon}{1+\varepsilon}\left[\frac{p^{n-2}}{1+\frac{n-1}{n}\varepsilon} + \frac{(n-2)p^{n-3}q}{1+\frac{n-2}{n}\varepsilon}\right]d\varepsilon > 0$$

Lo cual demuestra que la partición del riesgo implica un aumento en la utilidad esperada para el comerciante. Es decir que "*Para el comerciante es preferible exponer su fortuna por partes a riesgos independientes entre sí, que exponerla en su totalidad a un único riesgo*" (Laplace (1812, p. 442)).

Por otra parte, se puede escribir:

$$E_n^*(Y) = kp \int d\varepsilon \int_0^{\infty} e^{-\left(1+\frac{\varepsilon}{n}\right)x} \left[pe^{-\frac{\varepsilon}{n}x} + q\right]^{n-1} dx + ln(h) =$$

$$= k \int \frac{p}{1 + p\varepsilon + q\frac{\varepsilon}{n}} \left[1 + \frac{pq\varepsilon^2\left(1 - \frac{1}{n}\right)}{n\left(1 + p\varepsilon + q\frac{\varepsilon}{n}\right)} + \cdots \right] d\varepsilon + ln(h)$$

Luego, para valores de $n$suficientemente grandes, será:

$$E^*(Y) = k \int \frac{p}{1 + p\varepsilon} d\varepsilon + ln(h) = k\ ln(1 + p\varepsilon) + ln(h)$$

Lo que demuestra que la esperanza moral converge a la esperanza matemática $1 + p\varepsilon$.

3.- Si se supone que el comerciante asegura la suma ε despachada en un barco mediante el pago de una prima $(1 - p)\varepsilon$, su esperanza moral (utilidad esperada) aumentará de $(1 + \varepsilon)^p$ a $(1 + p\varepsilon)$, por lo que se puede concluir que el seguro produce lo que D. Bernoulli denomina una ventaja moral[60].

En coincidencia con el planteo de Buffon y Condorcet -basado en una interpretación frecuencista y, en consecuencia, en una definición de probabilidad concebible solamente en términos de eventos repetibles-, Lacroix (1816) consideró que el cálculo de la esperanza matemática como el promedio ponderado de las pérdidas y las ganancias sólo tenía sentido en un juego en el cual el número de resultados fuera finito. El hecho que esta condición no se cumpliera en el problema de San Petersburgo, lo condujo a retornar a la cuestión de la validez de la esperanza moral para eventos individuales[61].

---

[60] Esta teoría de la aseguración de Laplace fue continuada por B.J. Fourier (1819) y, fundamentalmente por Barrois (1835) quien, dada la asimilación de las esperanzas matemática y moral cuando los riesgos están infinitamente particionados, calculó la esperanza del asegurador como matemática y la esperanza del asegurado como moral.

[61] Con respecto a la controversia sobre si la esperanza moral podía ser considerada como un argumento respecto de eventos no repetibles fue aceptada por Lacroix (1816), Czuber (1878) y Timerding (1902) y rechazada por Fries (1842).

A partir de la definición de la esperanza moral y de la relación entre la utilidad y la fortuna original del jugador, Lacroix dedicó una parte importante de su trabajo al análisis comparativo de las fórmulas de Buffon y D. Bernoulli para la evaluación de las pérdidas y las ganancias relativas. De lo ya comentado, se puede concluir que la diferencia entre ambas propuestas radica en el supuesto sobre el comportamiento discreto o continuo de la variación de la riqueza. De acuerdo con la propuesta de Buffon, el valor relativo de una pérdida $\bar{g}$ con respecto a una fortuna b está dado por $\frac{\bar{g}}{b}$ y el valor relativo de una ganancia g por $\frac{g}{g+b}$, obteniendo el siguiente valor relativo de la pérdida equivalente a la ganancia g: $\bar{g} = \frac{gb}{g+b}$. Por otra parte, en la propuesta de D. Bernoulli, los valores relativos de una pérdida y una ganancia quedan expresados, respectivamente, por:

$$k\ ln(b - \bar{g}) - k\ ln(b) = -k\ ln\left(\frac{b - \bar{g}}{b}\right)$$

y,

$$k\ ln(b + \bar{g}) - k\ ln(b) = -k\ ln\left(\frac{b + \bar{g}}{b}\right)$$

Luego, a partir de la ecuación $\frac{b}{b-\bar{g}} = \frac{b+g}{b}$, que proporciona el valor relativo de una pérdida equivalente a una cierta ganancia, Lacroix concluyó que las propuestas de Buffon y D. Bernoulli conducen al mismo resultado[62].

Fries (1842) continuando la propuesta de d'Alembert y a diferencia de Lacroix, expresó el valor subjetivo de una pérdida relativa como $\frac{\bar{g}}{b-\bar{g}}$ y, a partir de la ecuación de equivalencia mencionada más arriba, obtuvo la siguiente evaluación de la pérdida equivalente a la ganancia: $\bar{g} = \frac{gb}{b+2g}$, menor que la de D. Bernoulli.

Czuber (1878) comparó esta relación con la propuesta de Buffon y concluyó que, si bien las fórmulas de Buffon, $\frac{\bar{g}}{b-\bar{g}}$ y $\frac{g}{b+g}$ no conducían a los mismos resultados numéricos obtenidos por D. Bernoulli, la diferencia respecto de la evaluación del valor de la pérdida equivalente a una cierta ganancia, podía considerarse no significativa.

Posteriormente, incorporando las fórmulas de Buffon en las ecuaciones de Laplace, Czuber demostró los teoremas de aversión al riesgo de D. Bernoulli y

[62] Esta demostración de Lacroix fue continuada por Öttinger (1848) y Liagre (1852).

concluyó que las fórmulas de Buffon eran preferibles a las de Bernoulli, debido a su independencia del supuesto de continuidad de la variación de la riqueza.

Si bien como se vio en las páginas precedentes -excepto por las críticas de d'Alembert y Condorcet-, los probabilistas del siglo XVIII y comienzos del siglo XIX aceptaron la teoría de la esperanza moral, los probabilistas franceses de la segunda parte del siglo XIX -entre los que cabe mencionar a Poisson (1837), Bertrand (1889) y Poincaré (1896)- minimizaron su importancia o simplemente la rechazaron.

En particular, Poisson adhirió a los trabajos de Laplace pero consideró, como Condorcet, que la esperanza moral sólo podía considerarse un complemento de la esperanza matemática, especialmente útil en las aplicaciones referidas a cuestiones económicas[63].

Como un caso particular de la solución propuesta por Fontaine (1764), Poisson (1837) propuso considerar una función de pago de la forma 2^n y expresar la fortuna del jugador A como $a = 2^k(1 + h)$ (donde $k(\leq n)$ era un número entero y h podía asumir cualquier valor real en el intervalo $(0,1)$). De modo que, para las primeras k "tiradas" la ganancia esperada de B sería igual a k y, para las $n - k$ "tiradas" restantes sería:

$$2^k(1 + h)\left(\frac{1}{2^{k+1}} + \cdots + \frac{1}{2^n}\right)$$

Es decir que, en total, la ganancia esperada de B sería:

$$k + 2^k(1 + h)\sum_{j=k+1}^{n} \frac{1}{2^j} = k + (1 + h)\left(1 - \frac{1}{2^{n-k}}\right)$$

Un resultado que le permitió concluir que cuando n aumenta indefinidamente, la ganancia esperada de B será menor o igual a $k + 1 + h$ y que, en consecuencia, su apuesta debería ser $k + 1 \leq S_B \leq k + 2$.

Whitworth (1901) arribó a los mismos resultados que D. Bernoulli a partir de un planteo del problema de San Petersburgo completamente diferente, independiente de la definición de la función de utilidad. Su propuesta se basó en el principio que todo jugador que posea una fortuna finita tiene la certeza de arruinarse al cabo de un número infinito de jugadas y que, en consecuencia,

[63] Debe tenerse en cuenta que Laplace y Poisson constituyen la transición entre la obra de los probabilistas del siglo XVIII y la interpretación estadística de Quetelet (1835). Para los probabilistas del siglo XIX el núcleo de la teoría de la probabilidad se trasladó de la cuantificación de la expectativa al estudio de las distribuciones de probabilidades.

un comportamiento racional le indicaría no apostar una cantidad fija, sino un porcentaje fijo de su fortuna previa a cada jugada[64].

A diferencia de las propuestas anteriores dirigidas a definir las condiciones que permitieran asegurar un valor finito para el juego, el objetivo de Whitworth consistió en hallar la proporción entre la apuesta y la fortuna inicial que, en el largo plazo, condujera al jugador a una posición de equilibrio (es decir, sin pérdida ni ganancia)[65].

Sea un jugador B con una fortuna inicial b, involucrado en una serie de partidas repetidas en las que: **i)** su apuesta (S_B) está dada por una proporción fija de su capital al inicio de cada juego; **ii)** su ganancia posible en cada juego es g y **iii)** las probabilidades de ganar o perder en cada juego son p y $q(p+q=1)$, respectivamente. De acuerdo con estas premisas, cada vez que B gana una partida su fortuna asume un valor igual a la fortuna inicial multiplicada por $\frac{b+g-S_B}{b}$ y, cada vez que pierde una partida su fortuna asume un valor igual a la fortuna inicial multiplicada por $b-\frac{S_B}{b}$. Luego, para garantizar una situación de equilibrio de largo plazo se debe verificar que $\left(1+\frac{g}{b}-\frac{S_B}{b}\right)^p\left(1-\frac{S_B}{b}\right)^q=1$.

En general, suponiendo que las cantidades a ganar por B sean g_1, g_2, $g_3, \dots$ $\left(g_j \geq 0; j=1,2,\dots,n\right)$ con probabilidades p_1, $p_2, \dots$, p_n $\left(\sum_{j=1}^{n} p_j = 1\right)$, el correspondiente multiplicador del capital inicial (b) será de la forma $\prod_{j=1}^{n}\left(1+\frac{g_j}{b}-\frac{S_B}{b}\right)^{p_j}=1$ ó $\prod_{j=1}^{n}\left(1+\frac{g_j}{b}\right)^{p_j}=\prod_{j=1}^{n}\left(1-\frac{S_B}{b+g_j}\right)^{p_j}$. A partir de la hipótesis que $S_B \ll b$, de modo que los términos de orden superior de los cocientes $\frac{S_B}{b+g_j}$ sean despreciables, el desarrollo de la expresión precedente puede ser aproximado de la siguiente forma:

$$\prod_{j=1}^{n}\left(1+\frac{g_j}{b}\right)^{p_j} \approx 1+S_B\sum_{j=1}^{n}\frac{p_j}{b+g_j}$$

---

64 Este principio -sin otra justificación que la que proporcionan los postulados del teorema de eponimia de Stigler- hoy se conoce como la “estrategia de apuestas de Kelly” (Kelly (1956) (ver Irwin (1967), Tijms (2004), Székely; Richards (2005)).

65 Whitworth (1901): “*No hemos asignado un nuevo valor a la esperanza matemática (...) Simplemente hemos determinado los términos de acuerdo con los cuales un individuo puede comprar la expectativa de obtener una ganancia, de modo que, al cabo de un número suficientemente grande de repeticiones -cada vez sobre una escala proporcional al monto de su fortuna previa a cada repetición-, no resulte ni más rico ni más pobre*” (p.246-247).

Dado que, para $n \to \infty$ esta sucesión es convergente, Whitworth obtuvo la siguiente aproximación:

$$S_B \approx \frac{\prod_{j=1}^{\infty}\left(1+\frac{g_j}{b}\right)^{p_j}-1}{\sum_{j=1}^{\infty}\frac{p_j}{b+g_j}}$$

para valores de n suficientemente grandes. Aplicando este resultado al problema de San Petersburgo obtuvo la apuesta:

$$S_B = \frac{\prod_{j=1}^{\infty}\left(1+\frac{2^{j-1}}{b}\right)^{\frac{1}{2^j}}-1}{\sum_{j=1}^{\infty}\frac{2^j}{b+2^{j-1}}}$$

que garantiza un equilibrio de largo plazo.

Timerding (1902), retornando a la teoría de Cramer acerca de la existencia de una cota superior (f^*) para $f(y)$, propuso una función de utilidad hiperbólica de la forma $f(y)=\frac{f^*y}{f^*+y}$, que combina las propiedades de la función acotada con las de la función logarítmica de Bernoulli.

Por otra parte, a fin de contrastar el postulado de Czuber sobre la independencia de la condición de continuidad en la variación de la riqueza de las fórmulas de Buffon, Timerding consideró la desagregación de una ganancia en dos ganancias sucesivas, $g=g_1+g_2$, cuyos valores relativos son $\frac{g_1}{b+g_1}$y $\frac{g_2}{b+g_1+g_2}$, respectivamente. En el ámbito de la interpretación discreta, este procedimiento lo condujo a la incongruencia que el valor relativo de la suma de las ganancias, $\frac{g_1+g_2}{b+g_1+g_2}=\frac{g_1}{b+g_1+g_2}+\frac{g_2}{b+g_1+g_2}$ era menor que la suma de los valores relativos, $\frac{g}{b+g}=\frac{g_1}{b+g_1}+\frac{g_2}{b+g_1+g_2}$. Por el contrario, de acuerdo con la definición de esperanza moral (sin tomar en consideración a la constante) en el ámbito de la interpretación continua de Bernoulli, obtuvo que:

$$ln\left(\frac{b+g}{b}\right)=ln\left(\frac{b+g_1}{b}\right)+ln\left(\frac{b+g_1+g_2}{b+g_1}\right)=ln\left(\frac{b+g_1+g_2}{b}\right)$$

Este resultado le permitió a Timerding demostrar que la aceptación del supuesto de continuidad de D. Bernoulli implica la solución de la contradicción y, en consecuencia, que el postulado de independencia de Czuber respecto de las fórmulas de Buffon era incorrecto.

Desde un punto de vista estrictamente probabilístico la sustitución del principio de la razón insuficiente por la ley de los grandes números como fundamento del concepto de valor esperado se continúa con el análisis de las soluciones dirigidas a especular acerca de la posibilidad de interpretar a la esperanza matemática como una expresión cuantitativa que involucra definiciones (neo-Niklausianas) de la probabilidad diferentes de la definición clásica[66].

A partir de la definición logicista Keynes (1921) (adoptando una posición neo-Niklausiana) intentó refutar "*...la doctrina que las esperanzas matemáticas de cursos de acción alternativos constituyen las medidas de nuestros grados de preferencia*" (p. 344). La aproximación de Keynes consistió, precisamente, en definir una teoría del vínculo parcial como una generalización de la teoría del vínculo total de la lógica deductiva[67] y considerar a la probabilidad como un grado de ese vínculo parcial[68], de modo que no es posible hablar de la probabilidad de una hipótesis sino solamente de su probabilidad condicionada por una evidencia relacionada parcialmente con ella[69].

Esta propuesta lo condujo a un modelo en el que la probabilidad $p(A/B)$ se traduce en un grado de creencia racional representativo del grado de vínculo parcial, concebido como una relación (indefinida) entre una proposición (A) y un cuerpo de conocimiento (B) condicionada por la verdad de dicha evidencia. Donde el evento A puede, en consecuencia, ser representado mediante un subconjunto de su espacio muestral $A \subset \Omega(\omega)$ formado por los elementos ω para los cuales una propiedad S es verdadera, $A = \{\omega/S(\omega)\ es\ verdadera\}$, de modo que a cada proposición $S(\omega)$ del espacio proposicional le corresponde un conjunto A en el espacio de eventos y viceversa.

---

66 Las soluciones neo-Danielianas, basadas exclusivamente en la teoría del valor, no serán tratadas en este texto.

67 Keynes (1921): *"Así como en ciertas circunstancias podemos juzgar directamente que una conclusión 'se sigue' de una premisa, la hipótesis que sostiene que a veces podemos reconocer que una conclusión 'se sigue parcialmente' de una premisa o se basa en una relación de probabilidad con la misma, constituye una extensión del supuesto original"* (p. 52).

68 Keynes (1921): *"Si tenemos en cuenta que para conocer correctamente una conexión lógica entre un conjunto de proposiciones, a las que denominamos nuestra evidencia y que suponemos conocidas y otro conjunto formado por las que denominamos conclusiones, debemos asignar a estas ponderaciones mayores o menores de acuerdo con los fundamentos proporcionados por las primeras (...) no resulta forzado describir este vínculo entre evidencias y conclusión como una relación de probabilidad"* (pp. 5-6).

69 Keynes (1921): "*Así como ningún lugar puede ser intrínsecamente distante, ninguna proposición es en sí misma ni probable, ni improbable. La probabilidad de dicha proposición varía de acuerdo con la evidencia presentada, la cual actúa como si fuera su origen de referencia*" (p. 7).

En la concepción Keynesiana las relaciones de probabilidad entre un cuerpo de conocimiento B y una conclusión consistente en una propiedad (o conjunto de propiedades) A no siempre son cuantificables y en muchos casos ni siquiera son comparables[70]. Estas consideraciones lo condujeron a la conclusión que, aun cuando se supusiera que las cantidades de bienestar fueran medibles, las asignaciones de las probabilidades no siempre son cuantificables y no toman en consideración el estado del conocimiento, ni la evidencia con que cuenta el observador, ni el riesgo asociado al juego. Keynes postuló que, precisamente, esta no consideración de que "*…el gran riesgo de la apuesta es lo que nos determina*" (p. 352) es la razón del fracaso de la definición de esperanza matemática en la paradoja de San Petersburgo.

Sean G_B la posible ganancia del jugador B, p su probabilidad de ganarla y, en consecuencia, $S_B = E(Y) = G_B p$ su apuesta para participar en el juego, Keynes definió el riesgo absoluto de B como $R_B = qS_B$, de modo que el riesgo relativo, q, queda definido como la probabilidad de perder la apuesta. Luego, denotando por k la "tirada" en la cual la posible ganancia de B es $E(y) = \frac{n}{2}$, el riesgo absoluto asociado en el juego de San Petersburgo es $R_B = \left(1 - \frac{1}{2^k}\right)\frac{n}{2}$ el cual, si se supone un número infinito de "tiradas", asume un valor infinito. En realidad, la apreciación de Keynes no es del todo correcta en la medida que en el juego de San Petersburgo el riesgo relativo es próximo a la unidad.

Keynes asimila la probabilidad p a un grado de creencia racional (igual para todos los individuos) no simplemente a un grado de creencia individual. Es decir, considera a las probabilidades como valores fijados objetivamente por el observador los cuales son asimilables a las relaciones lógicas conocidas por intuición, pero utilizando un concepto Platónico (metafísico) del término "objetivo", es decir, no referido a "cosas" del mundo material, sino a "algo" en un supuesto mundo formado por ideas abstractas, similar al postulado por los filósofos de Cambridge, que incluía ideas objetivas, cualidades éticas (con la idea de virtud ocupando un lugar prominente) y entes matemáticos.

Ramsey (1931) propuso una interpretación de la probabilidad como grado de creencia personal que difiere de la de Keynes en el supuesto de la existencia de relaciones lógicas de probabilidad como pares de proposiciones perceptibles por el observador. Considera que sólo es posible percibir las relaciones lógicas en casos de cierta complejidad, pero que habitualmente su

70 Keynes (1921): *"Existen pares de probabilidades para los cuales no existe ninguna comparación posible entre sus magnitudes, que respecto a ciertos pares de relaciones de probabilidad aunque no podemos medir la diferencia entre ellas, podemos asegurar que una es mayor que la otra, y que en un tipo muy especial de casos puede adjudicarse un significado a una comparación numérica de magnitudes"* (p. 34).

percepción en casos simples es imposible, lo que conduce a postular que la intuición lógica no es adecuada para establecer la existencia de grados de vinculación parcial ni para demostrar la satisfacción por parte de la teoría Keynesiana del sistema axiomático de Kolmogorov.

La conclusión de Ramsey es que "*...el grado de una creencia es una propiedad causal de ésta, que puede ser expresada vagamente como el límite superior a partir del cual estamos preparados para actuar sobre dicha creencia*" (p. 169).

Debe tenerse en cuenta que esta definición subjetivista considera que la probabilidad p corresponde siempre a eventos individuales y, cada vez que se asigne una probabilidad, es necesario pensarla como subordinada a la interpretación que hace cada observador de un conjunto de información particular. En este contexto, la condición de equidad implica la indiferencia entre ser uno u otro jugador, entre pagar o cobrar p para cobrar o pagar 1 ante la ocurrencia de un evento E. Se dice en ese caso que la evaluación de p es "*coherente*" en cuanto que no coloca a ninguno de los dos jugadores en la situación de ganar con seguridad. Una restricción que permite concluir que los grados de creencia son personalmente racionales (al menos hasta el punto de satisfacer dicha restricción).

Completando esta definición, de acuerdo con el teorema de Ramsey-de Finetti, se denomina probabilidad de ocurrencia de un evento E, para un individuo dado, al número real $p(E) = p$ que representa la medida de su grado de creencia personal en E y que puede ser interpretado como la esperanza matemática de un juego con una ganancia unitaria posible en caso de ocurrencia de E; es decir, como una apuesta de valor p sobre la ocurrencia de E, que cumple la condición de coherencia.

Este teorema y sus generalizaciones demuestran que, dado un grupo social $B = \{B_1, B_2, \dots, B_n\}$, la posición más favorable en cuanto al cumplimiento de la condición de coherencia sería converger en una probabilidad intersubjetiva o de consenso. No obstante, podría suceder que, en un grupo social, un subconjunto mayoritario de sus integrantes conviniera en adoptar una probabilidad de consenso y otro subconjunto (minoritario) asumiera probabilidades subjetivas sobre la ocurrencia de un evento distintas de la probabilidad intersubjetiva. Esta circunstancia no modifica el razonamiento precedente y permite concluir, en primer lugar, que las probabilidades intersubjetivas no deben reemplazar completamente a las probabilidades personalistas deFinettianas, que ambas son necesarias para una evaluación del grado de creencia de los individuos[71].

Estos resultados conducen a la conclusión que, dado que inevitablemente en todo grupo social existen individuos cuyos razonamientos son eminentemente intuitivos o cuyos grados de creencia se basan en patrones lógicos no-estándar, el consenso Keynesiano nunca será completo. De modo que se puede concluir que ese grado de creencia singular compartido a nivel universal es de existencia puramente ideal y que, en consecuencia, la interpretación intersubjetivista de la probabilidad puede ser considerada como una definición intermedia entre la interpretación logicista del "joven" Keynes y la personalista-crítica de Ramsey-de Finetti.

Fechner tomó de Quetelet la idea de la analogía entre el modelo de Laplace sobre mecánica celeste y una física de la sociedad humana, es decir, de la asimilación de las leyes gravitacionales a las causas constantes que gobiernan la sociedad en la que, al igual que las perturbaciones que afectan a los fenómenos físicos, las diferencias entre los individuos que la componen (las cuales no deben ser interpretadas como desvíos aleatorios del verdadero valor, sino como "*...la forma libre de reacción del individuo ante un estímulo*"[72] se distribuyen de acuerdo con una función de probabilidades y se vuelven despreciables si se considera un número suficientemente grande de individuos. Estos postulados condujeron a Fechner al concepto de "*objeto colectivo*" o "*serie colectiva*" al que definió como un conjunto de individuos heterogéneos que varían aleatoriamente con respecto a un atributo común (en particular, un atributo cuantificable), regido por leyes de naturaleza probabilística. En su expresión más simple el colectivo puede ser considerado como una sucesión de resultados obtenidos de una serie de observaciones repetidas en igualdad de condiciones, cada una de las cuales admite sólo dos alternativas posibles. Es decir, matemáticamente el colectivo puede ser considerado como aproximable por una serie binaria infinita.

El concepto de objeto colectivo tuvo su culminación en la reformulación realizada por Reichenbach (1935) y R. von Mises del concepto de probabilidad "*...a fin de reemplazar o complementar la rígida estructura causal de la teoría clásica*" (von Mises (1921, p. 427)).

---

[71] de Finetti (1937): *"...es posible demostrar que existen profundas razones psicológicas que hacen muy natural la concordancia exacta o aproximada que se observa entre las opiniones de distintos individuos, pero no hay ningún argumento racional, positivo o metafísico que pueda dar a este hecho un significado que vaya más allá de una simple concordancia de opiniones subjetivas"* (p. 152).
[72] Heidelberger (1987, p.137)).

Von Mises (1928) considera la necesidad de distinguir entre "*colectivos empíricos*" (que pueden ser hallados en el mundo real que, por lo tanto son observables y que, en consecuencia, están formados por un número finito de elementos) y "*colectivos matemáticos*" (formados por una sucesión infinita de elementos). De acuerdo con von Mises (1928), los colectivos empíricos obedecen a dos principios fundamentales: **i)** La "*ley de estabilidad de las frecuencias estadísticas*" y **ii)** la "*ley de irregularidad*" que postula la aleatoriedad absoluta en la proyección de los atributos sobre los elementos del colectivo.

Sea $E \subseteq \Omega$ un atributo asociado a un colectivo. Supóngase que en los primeros n elementos del colectivo el atributo E se presentó $n(E)$ veces, de modo que la frecuencia relativa del atributo E al cabo de n observaciones queda definida por $\frac{n(E)}{n}$. La ley de estabilidad de las frecuencias estadísticas establece que, a media que n crece, la frecuencia relativa se aproximará indefinidamente a un valor fijo asimilable a la probabilidad de ocurrencia de E.

Una de las principales objeciones a esta definición frecuencista radica en lo restringido de sus aplicaciones, dado que en muchos casos prácticos es imposible definir un colectivo empírico. Entre las numerosas modificaciones a que se sometió a la interpretación frecuencista a fin de eliminar esta restricción, la más importante fue, indudablemente, la debida a Reichenbach (1935), quien procuró obtener una definición de probabilidad por una vía axiomática y justificar el significado intuitivo de la misma. Con respecto a la primera cuestión Reichenbach intentó una solución basada exclusivamente en la teoría de conjuntos y en las operaciones lógicas, obteniendo una definición de probabilidad (puramente formal) expresada como una relación entre dos clases de proposiciones.

En lo que hace al papel que desempeña la intuición en el análisis filosófico de la probabilidad objetiva, tomando como punto de partida el supuesto (contrario a la propuesta de von Mises) de que la interpretación frecuencista de la probabilidad podía ser generalizable a todas las aplicaciones del término probable, Reichenbach intentó ampliar sus alcances a eventos no-repetibles, mediante la definición de las que se denominaron clases de referencia, formadas por eventos similares al analizado y consideró a la teoría de la probabilidad como una disciplina que calcula probabilidades desconocidas de colectivos derivados a partir de probabilidades conocidas de colectivos originarios. Pero esta generalización tropezó con la dificultad insalvable que significa la imposibilidad de la determinación de reglas de selección objetivas universalmente aceptadas de los eventos que deben integrar dichas clases de referencia.

De todo lo anterior se deduce que, tanto la definición de von Mises como la de Reichenbach, a fin de evitar cualquier otro tipo de regularidad en las sucesiones de eventos que constituyen su fundamento, intentaron proporcionar a la relación de probabilidad un contenido estrictamente matemático mediante la imposición de complicadas condiciones que, inevitablemente, restringían el concepto de aleatoriedad total y demostraban que era imposible dar un carácter matemáticamente preciso a la noción de "*irregularidad absoluta*".

Reichenbach aplicó estos conceptos al problema de N. Bernoulli pero, de acuerdo con el método de series convergentes, propuso que la divergencia de la esperanza matemática no es atribuible necesariamente a la forma de asignar los valores a p, sino a la definición de la función de pagos (que puede hacer que el juego se transforme en no-equitativo cuando el número de jugadas aumenta indefinidamente)[73].

10.2.5.- Propiedades

La esperanza matemática posee las siguientes propiedades:

1) Sea X una variable "degenerada" cuyo dominio es $\Omega(X) = a$ y su función de probabilidades es $p(X = a) = 1$. Entonces, será $E(X) = ap(X = a) = a$.

2) Sea una variable $Z = g(X)$ (donde $g(X) \epsilon \mathbb{R}_1$, X es una variable aleatoria con dominio $\Omega(X)$ y función de distribución de probabilidades $F_X(x)$). Entonces será:

$$E(Z) = E[g(X)] = \int_{-\infty}^{\infty} z dF_Z(z) = \int_{-\infty}^{\infty} g(x) dF_X(x) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

En el caso particular en que X sea una variable discreta, será $E(Z) = \sum_j g(x_j) p_j$.

En general, para una variable X de n dimensiones, se verifica que:

$$E[g(X_1, X_2, \dots, X_n)] = \sum_{j_1} \sum_{j_2} \dots \sum_{j_n} g(x_{j_1}, x_{j_2}, \dots, x_{j_n}) p_{j_1, j_2, \dots, j_n}$$

$$E[g(X_1, X_2, \dots, X_n)] =$$

[73] Ver Borel (1939)(1949a)(1949b)(1950).

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(X_1, X_2, \dots, X_n)\, f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

3) Como corolario de la propiedad anterior, dada una variable aleatoria $Z = kX$ (donde k denota una constante) se puede concluir que:

$$E(Z) = E(kX) = \int_{\Omega(X)} kx dF_X(x) = kE(X)$$

4) Sean dos variables aleatorias, $\{X, \Omega(X), F_X(x)\}$ y $\{Y, \Omega(Y), F_Y(y)\}$, y sea la variable bidimensional $Z = X + Y$. Entonces será:

$$E(Z) = E(X + Y) = \int_{\Omega(X)} \int_{\Omega(Y)} (x + y) dF_Z(x + y) =$$

$$= \int_{\Omega(X)} xdx \int_{\Omega(Y)} f_{X,Y}(x, y) dxdy + \int_{\Omega(Y)} ydy \int_{\Omega(X)} f_{X,Y}(x, y) dxdy =$$

$$= \int_{\Omega(X)} xdF_X(x) + \int_{\Omega(Y)} ydF_Y(y) = E(X) + E(Y)$$

Esta propiedad es generalizable a cualquier conjunto $\{X_1, X_2, \dots, X_n\}$ de variables aleatorias:

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

A partir de estas propiedades se obtiene el siguiente corolario: Sea la variable n-dimensional Z definida por $Z = \varphi(X_1, X_2, \dots, X_n)$ (donde φ es una función lineal). Se verificará que:

$$E(Z) = E[\varphi(X_1, X_2, \dots, X_n)] = \varphi[E(X_1), E(X_2), \dots, E(X_n)]$$

Ejemplo n° 22:

Sea una variable aleatoria X que sólo puede asumir valores positivos. Entonces se puede escribir:

$$X = E(X) \left[1 + \frac{X - E(X)}{E(X)}\right] = E(X)(1 + Y)$$

De modo que $E(Y) = E\left(\frac{X-E(X)}{E(X)}\right) = 0$. Luego, dada una variable aleatoria $Z = \varphi(X) = \frac{1}{X}$, se verificará que:

$$E(Z) = E\left(\frac{1}{X}\right) = E\left[\frac{1}{E(X)(1+Y)}\right] = \frac{1}{E(X)}E\left(\frac{1}{1+Y}\right) =$$

$$= \frac{1}{E(X)}E\left(1 - Y + \frac{Y^2}{1+Y}\right) = \frac{1}{E(X)}\left[1 + E\left(\frac{Y^2}{1+Y}\right)\right]$$

y, dado que las variables Y^2 y $(1+Y)$ pueden asumir solamente valores positivos, será $E\left(\frac{Y^2}{1+Y}\right) > 0$. Lo que permite concluir que el valor esperado de la variable Z, definida como la transformación no-lineal φ de la variable X será tal que $E(Z) = E\left(\frac{1}{X}\right) > \frac{1}{E(X)}$.

---

5) Sean dos variables aleatorias independientes $\{X, \Omega(X), F_X(x)\}$ y $\{Y, \Omega(Y), F_Y(y)\}$ y sea la variable bidimensional $W = XY$. Se verificará, entonces, que:

$$E(W) = E(XY) = \int\limits_{\Omega(X)}\int\limits_{\Omega(Y)} xy dF_W(x,y) = \int\limits_{\Omega(\mathrm{X})}\int\limits_{\Omega(Y)} (xy) f_X(x) f_Y(y) dx dy =$$

$$= \int\limits_{\Omega(X)} x f_X(x) dx \int\limits_{\Omega(Y)} y f_Y(y) dy = E(X)E(Y)$$

Esta propiedad es inmediatamente generalizable para cualquier conjunto $(X_1, X_2, \dots, X_n)$ de variables aleatorias independientes:

$$E(X_1.X_2.\dots.X_n) = E(X_1)E(X_2)\dots E(X_n)$$

Retornando al momento centrado mixto de orden (1,1), denominado comúnmente covarianza:

$$cov(X,Y) \equiv \gamma(X,Y) \equiv \mu_{1,1}(X,Y) = E\{[X - E(X)][Y - E(Y)]\} =$$

$$= E(XY) - E[YE(Y)] - E[XE(X)] + E(X)E(Y) = E(XY) - E(X)E(Y)$$

De la propiedad **5)** se puede concluir que, si las variables son independientes, la covarianza es nula. Ahora bien, que la covarianza sea nula no implica que las variables sean independientes, sólo implica que no están relacionadas linealmente (es decir, que no están relacionadas en los momentos de primer orden solamente). Cuando $\gamma(X,Y) > 0$ se dice que la relación entre las variables es positiva o directa. Esto ocurre cuando en el valor esperado $E\{[X - E(X)][Y - E(Y)]\}$ prevalecen las diferencias $X - E(X)$ e $Y - E(Y)$ del mismo signo; en otras palabras, cuando tendencialmente, a valores grandes de X corresponden valores grandes de Y. Viceversa, cuando $\gamma(X,Y) < 0$, se dice que la relación entre las variables es negativa o inversa.

Ejemplo n° 23:

Sea una variable aleatoria bidimensional (X,Y) con función de densidad conjunta de la forma:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{3}(x+y) & para \quad 0 \le x \le 2, 0 \le y \le 1 \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

Las funciones de densidad marginales están definidas por:

$$f_X(x) = \int_0^1 f_{X,Y}(x,y)dy = \frac{1}{3}\int_0^1 (x+y)dy = \frac{2x+1}{6}$$

$$f_Y(y) = \int_0^2 f_{X,Y}(x,y)dx = \frac{1}{3}\int_0^2 (x+y)dx = \frac{2}{3}(y+1)$$

Las correspondientes esperanzas matemáticas son:

$$E(X) = \int_0^2 x f_X(x)dx = \frac{1}{6}\int_0^2 x(2x+1)dx = \frac{11}{9}$$

$$E(Y) = \int_0^1 y f_Y(y)dy = \frac{3}{2}\int_0^1 y(y+1)dy = \frac{5}{9}$$

y la covarianza está dada por:

$$\gamma(X,Y) = E\left[\left(X - \frac{11}{9}\right)\left(Y - \frac{5}{9}\right)\right] = \frac{1}{3}\int_0^1 \left(y - \frac{5}{9}\right) dy \int_0^2 \left(x - \frac{11}{9}\right)(x+y)dx$$

$$= \frac{2}{27}\int_0^1 \left(y - \frac{5}{9}\right)(-2y+1)dy = -\frac{1}{81}$$

Es decir que entre las variables X e Y existe una correlación lineal inversa.

---

6) Sean X e Y dos variables aleatorias id, con función de distribución $F(u)$, sea una variable bidimensional $Z = (X, Y)$ con función de distribución $F_Z(z)$y sea $g(\cdot)$ una función real. Si se verifica que $\gamma[g(X), g(Y)] \geq 0$ (suponiendo que la covarianza exista), se dice que entre las variables X e Y existe "dependencia funcional positiva".

En general, dada una variable aleatoria n-dimensional $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ con funciones de distribución $F_{X_1}(x_1) = F_{X_2}(x_2) = \cdots = F_{X_n}(x_n) = F(u)$ y $F_{\boldsymbol{X}}(x)$ y una función real$h(\cdot)$, si se verifica que:

$$E\left[\prod_{j=1}^{n} h(X_j)\right] \geq \prod_{j=1}^{n} E[h(X_j)]$$

se dice que entre las variables marginales existe "dependencia funcional positiva" (en particular, si n es impar la función $h(\cdot)$ deber observar la condición de ser no-negativa).

7) Sea una variable aleatoria n-dimensional $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ y sean $g_1(\boldsymbol{X})$ y $g_2(\boldsymbol{X})$ dos funciones reales crecientes respecto de cada componente del vector $\boldsymbol{X}$, si se verifica que:

$$E[g_1(\boldsymbol{X})g_2(\boldsymbol{X})] \geq E[g_1(\boldsymbol{X})]E[g_2(\boldsymbol{X})]$$

es decir, que $\gamma[g(X), g(Y)] \geq 0$ (suponiendo que los valores esperados existan), se dice que $\boldsymbol{X}$ es una variable "asociada positivamente"[74].

---

[74] Debe tenerse en cuenta que, para distribuciones no-Normales multidimensionales, las medidas de asociación más adecuadas son los coeficientesτde Kendall y ρ_s de Spearman.

Como corolario se puede concluir en forma inmediata que esta condición también se verifica para $n = 1$, siempre que los valores esperados existan.

Esta condición de asociación implica la condición de dependencia de ortante positiva. Se demuestra, además, que cualquier subconjunto de $\boldsymbol{X}$ también está asociado positivamente[75]

8) Tomando como punto de partida la desigualdad:

$$\frac{a^{\alpha}}{\alpha}+\frac{b^{\beta}}{\beta}-ab \geq 0 \qquad \left(a,b>0;\alpha,\beta>0;\frac{1}{a}+\frac{1}{b}=1\right)$$

Supónganse dos funciones, g_1 y g_2, tales que $E[|g_1(X)|^{\alpha}] = E\left[|g_2(X)|^{\beta}\right] = 1$[76]. Haciendo en la desigualdad anterior $a = g_1(X)$ y $b = g_2(X)$, se obtiene que:

$$E[|g_1(X)g_2(X)|] \leq E\left[\frac{|g_1(X)|^{\alpha}}{\alpha}+\frac{|g_2(X)|^{\beta}}{\beta}\right] =$$

$$= \frac{1}{\alpha}E[|g_1(X)|^{\alpha}] + \frac{1}{\beta}E\left[|g_2(X)|^{\beta}\right] =$$

$$= \frac{1}{\alpha}+\frac{1}{\beta} = 1 = \{E[|g_1(X)|]^{\alpha}\}^{1/\alpha} + \left\{E[|g_2(X)|]^{\beta}\right\}^{1/\beta}$$

En particular, para $\alpha = \beta = 2$, se obtiene la desigualdad de “Schwarz-Hölder”:

$$\{E[g_1(X)g_2(X)]\}^2 \leq E\{[g_1(X)]^2\}E\{[g_2(X)]^2\}$$

En esta demostración se ha supuesto implícitamente que los valores esperados que figuran en el segundo miembro son finitos y distintos de cero. En el caso en que uno de los dos sea nulo –por ejemplo, el primero-, entonces $g_1(X) = 0$ y, por lo tanto, se anulará también el primer miembro. Si uno de los dos fuera infinito y el otro distinto de cero, el segundo miembro será infinito y, entonces, se verificaría la desigualdad.

---

75 Ver Esary; Proschan (1972), Yanagimoto (1990), Schriever (1987).

76 Si esta condición no se verifica se puede utilizar una normalización $g^*(X)$ de la forma $g^*(X) = \frac{g(X)}{\{E[|g(X)|^{\alpha}]\}^{1/\alpha}}$.

Sea una distribución de probabilidades discreta $\{x_j, p_j\}$ y sean las transformaciones $g_1(x_j) = y_j p_j^{-1/\alpha}$ y $g_2(x_j) = z_j p_j^{-1/\beta}$ (donde $\frac{1}{\alpha} + \frac{1}{\beta} = 1$). Entonces, como corolario de la demostración anterior, será:

$$\sum_j |y_j z_j| = \sum_j |g_1(x_j) g_2(x_j)| = E[|g_1(X) g_2(X)|] \leq$$

$$\leq \{|g_1(X)|^\alpha\}^{1/\alpha} \{|g_2(X)|^\beta\}^{1/\beta} =$$

$$= \left[\sum_j |g_1(x_j)|^\alpha p_j\right]^{1/\alpha} \left[\sum_j |g_2(x_j)|^\beta p_j\right]^{1/\beta}$$

Es decir, se demuestra que:

$$\sum_j |y_j z_j| \leq \left[\sum_j |y_j|^\alpha\right]^{1/\alpha} \left[\sum_j |z_j|^\beta\right]^{1/\beta}$$

Con un razonamiento análogo, para distribuciones continuas, se demuestra que:

$$\int |f_{X_1}(x) f_{X_2}(x)| dx \leq$$

$$\leq \left[\int |f_{X_1}(x)|^\alpha dx\right]^{1/\alpha} \left[\int |f_{X_1}(x)|^\alpha dx\right]^{1/\alpha} \left[\int |f_{X_2}(x)|^\beta dx\right]^{1/\beta}$$

9) La desigualdad de Minkowsky:

Sean dos variables aleatorias, X e Y, para $\alpha > 1$, se verificará que:

$$E(|X+Y|^\alpha) = E(|X+Y||X+Y|^{\alpha-1}) \leq E(|X||X+Y|^{\alpha-1} + |Y||X+Y|^{\alpha-1})$$

$$= E(|X||X+Y|^{\alpha-1}) + E(|Y||X+Y|^{\alpha-1})$$

Aplicando la desigualdad de Schwarz-Hölder a las dos esperanzas matemáticas que figuran en el segundo miembro, dado que $\frac{1}{\alpha} + \frac{1}{\beta} = 1$ (es decir que $\beta = \frac{\alpha}{\alpha - 1}$), resulta que:

$$E(|X+Y|^{\alpha}) \leq [E(|X|^{\alpha})]^{1/\alpha}\left[E\left(|X+Y|^{(\alpha-1)\beta}\right)\right]^{1/\beta}+$$

$$+[E(|Y|^{\alpha})]^{1/\alpha}\left[E\left(|X+Y|^{(\alpha-1)\beta}\right)\right]^{1/\beta}=$$

$$=\left\{[E(|X|^{\alpha})]^{1/\alpha}+[E(|Y|^{\alpha})]^{1/\alpha}\right\} \quad [E(|X+Y|^{\alpha})]^{(\alpha-1)/\alpha}$$

Multiplicando miembro a miembro en esta expresión por $[E(|X+Y|^{\alpha})]^{(\alpha-1)/\alpha}$ se obtiene que:

$$E\left[(|X+Y|^{\alpha})^{1/\alpha}\right] \leq [E(|X|^{\alpha})]^{1/\alpha}+[E(|Y|^{\alpha})]^{1/\alpha}$$

9) Sea una variable aleatoria que puede asumir valores no-negativos y sea s un número real positivo, aplicando, entonces el teorema de Fubini, se puede escribir[77]:

$$E(|X|^{s})=E(X^{s})=\int_{0}^{\infty} x^{s}dF_X(x)=\int_{0}^{\infty}\left(\int_{0}^{x+} su^{s-1}du\right)dF_X(x)=$$

$$=\int_{0}^{\infty} su^{s-1}\left(\int_{0}^{\infty} dF_X(x)\right)du=\int_{0}^{\infty} sx^{s-1}\,[1-F_X(x)]dx$$

A partir de esta expresión se obtiene, en particular, que:

$$E(X)=-\int_{-\infty}^{0} F_X(x)dx+\int_{0}^{\infty}[1-F_X(x)]dx$$

siempre que las integrales que figuran en el segundo miembro no sean ambas infinitas.

Asimismo, si la variable X es tal que sólo puede asumir valores enteros no-negativos, para $s=1$ y repitiendo para la suma el procedimiento utilizado en la demostración precedente, se puede escribir:

[77] La inversión del orden de integración que postula el teorema de Fubini es posible dado que la función integrando en no-negativa (ver Sec. 9).

$$E(X) = \sum_{j=0}^{\infty} jp_j = \sum_{j=1}^{\infty} \sum_{i=1}^{j} p_j = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} p_j = \sum_{j=1}^{\infty} p(x \geq j)$$

Como corolario de esta expresión, dadas dos variables aleatorias, X e Y, que sólo pueden asumir valores enteros no-negativos y tales que $F_X(j) \geq F_Y(j)$. Si los correspondientes valores esperados existen, se obtiene que $E(X) \leq E(Y)$. Asimismo, si las variables X e Y son tales que pueden asumir cualquier valor real no-negativo, entonces se verifica que $E(X^s) \leq E(Y^s)(s > 0)$.

Ejemplo n° 24:

Sea una variable aleatoria X con función de densidad de la forma:

$$f_X(x) = \begin{cases} k & para \quad -1 < x < 1 \\ \dfrac{k}{x^4} & para\ cualquier\ otro\ valor\ de\ x \end{cases}$$

Dado que la función de densidad es simétrica, se puede escribir:

$$\int_{-\infty}^{\infty} f_X(x)dx = 2\int_{1}^{\infty} \frac{k}{x^4}dx + \int_{-1}^{1} kdx = \left[\frac{2k}{3}\frac{1}{x^3}\right]_1^{\infty} + 2k = \frac{2k}{3} + 2k = \frac{8k}{3} = 1$$

y como por definición se debe verificar que $\int_{-\infty}^{\infty} f_X(x)dx = 1$, debe ser $k = \frac{3}{8}$.

Se desea determinar el intervalo $(-a, a)$ para el cual se verifique que $\int_{-a}^{a} f_X(x)dx = 0{,}90$. Dado que $\int_{-1}^{1} f_X(x)dx = \frac{3}{8}\int_{-1}^{1} dx = \frac{3}{4}$, se debe elegir un valor de b tal que:

$$\frac{3}{8}\int_{-b}^{b} \frac{1}{x^4}dx = \frac{6}{8}\int_{1}^{b} \frac{1}{x^4}dx = 0{,}90 - 0{,}75 = 0{,}15$$

$$\left[-\frac{1}{4x^3}\right]_1^{b} = \frac{1}{4} - \frac{1}{4b^3} = 0{,}15$$

Es decir, será $b = 1{,}357$.

Teniendo en cuenta la simetría de la función de densidad, el valor esperado de X puede ser definido de la siguiente forma:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_{-\infty}^{0} x f_X(x)dx + \int_{0}^{\infty} x f_X(x)dx$$

$$= \int_{-\infty}^{0} -x f_X(-x)d(-x) + \int_{0}^{\infty} x f_X(x)dx = \int_{\infty}^{0} x f_X(x)dx + \int_{0}^{\infty} x f_X(x)dx = 0$$

Nótese que este resultado depende sólo de la propiedad de simetría, $f_X(x) = f_X(-x)$, es decir que puede ser considerado válido para toda distribución de probabilidades continua simétrica con respecto al valor $x = 0$.

Ejemplo n° 25:

Sea una variable aleatoria X con función de densidad de la forma:

$$f_X(x) = \begin{cases} \frac{1}{2} & para \qquad 0 < x < 1 \\ \frac{1}{4} & \qquad 2 < x < 4 \\ 0 & para\ cualquier\ otro\ valor\ de\ x \end{cases}$$

Se verifica, entonces, que:

$$p(x > 1) = p(x > 2) = \int_{2}^{4} \frac{1}{4} dx = \left[\frac{x}{4}\right]_2^4 = \frac{1}{2}$$

$$p(x > 3) = \int_{3}^{4} \frac{1}{4} dx = \left[\frac{x}{4}\right]_3^4 = \frac{1}{4}$$

La esperanza matemática de X será entonces:

$$E(X) = \frac{1}{2}\int_{0}^{1} x dx + \int_{2}^{4} x dx = \frac{7}{4}$$

10.2.6.- La dominancia estocástica en la controversia entre la axiomática de von Neumann-Morgenstern y el planteo de Allais

A partir del concepto que el orden de preferencia entre distintas alternativas en condiciones de incertidumbre se basa en la maximización de la esperanza moral, von Neumann; Morgenstern (1953) plantearon un fundamento axiomático de la teoría del valor.

Tomando como punto de partida la noción de "*conjunto de indiferencia*" como aquél formado por las combinaciones de utilidades posibles que no generan preferencias entre sí, definieron (a partir de la función de utilidad ordinal de los neoclásicos) una función de utilidad cardinal que preserva el orden (lineal) de las preferencias, válida sólo bajo ciertas condiciones resumidas en los siguientes axiomas:

Axioma 1: Los conjuntos de indiferencia están ordenados de acuerdo con una relación de preferencias, de modo que ésta satisface las propiedades reflexiva, antisimétrica y transitiva.

Axioma 2: El conjunto de utilidades es continuo en-probabilidad. Es decir, si existe una preferencia entre las utilidades y_1 e y_2, entonces existe la misma preferencia entre la apuesta a integrar para participar en un juego cuyos resultados posibles son y_1 (con probabilidad p) e y_2 (con probabilidad $1-p$).

Axioma 3 (de aditividad): Si existe una preferencia entre y_1 e y_2, entonces existe la misma preferencia entre la apuesta a integrar para participar en un juego cuyos resultados posibles son y_1 (con probabilidad p) e y_2 (con probabilidad $1-p$) y otra apuesta para participar en un juego cuyos resultados posibles son y_1 (con probabilidad p) e y (con probabilidad $1-p$).

Axioma 4 (de independencia fuerte): Si existe una preferencia entre y_2 y una apuesta (r_1) para participar en un juego cuyos resultados posibles son y_1 (con probabilidad p_1) e y_2 (con probabilidad $1-p_1$), entonces existe la misma preferencia entre una apuesta r_2 para participar en un juegos cuyos resultados posibles son y_2 (con probabilidad p_2) e y (con probabilidad $1-p_2$) y otra apuesta para participar en una juego cuyos resultados posibles son un billete de lotería (r), que ofrece la posibilidad de ganar y_1, con probabilidad p_2 y de ganar y con probabilidad $1-p_2$[78].

La función de los tres primeros axiomas, los cuales están vinculados con los supuestos de completitud y continuidad de las preferencias, consiste en establecer la existencia de una función de preferencias continua sobre las distribuciones de probabilidades. El último axioma, que es condición necesaria para lograr la no-condicionalidad de la probabilidad, está vinculado

[78] Si bien este axioma no figura en von Neumann; Morgenstern (1953), de acuerdo con Malinvaud (1952), se encuentra implícito en su formulación preaxiomática. Una versión previa de este axioma puede ser hallado en Marshak (1950), Manne (1952) y Samuelson (1952).

con los supuestos de monotonicidad y linealidad de la función de preferencias sobre un conjunto de distribuciones de probabilidades.

La condición de linealidad determina una doble implicación entre el cumplimiento del axioma de independencia y el criterio de decisión basado en la maximización de la esperanza moral. En otros términos, implica que el axioma de independencia constituye una condición necesaria y suficiente para justificar el criterio de selección de las preferencias mediante la maximización de la utilidad esperada. En la primera parte del axioma 4 la comparación será entre $E_1^*(Y)$ y:

$$E_2^*(Y) = r_1 = y_1p_1 + y_2(1-p_1) = (y_1 - y_2)p_1 + y_2$$

En la segunda, la comparación será entre:

$$E_3^*(Y) = y_2p_2 + y(1-p_2) = (y_2 - y)p_2 + y$$

y,

$$E_4^*(Y) = rp_2 + y(1-p_2) = [y_1p_1 + y(1-p_1)]p_2 + y(1-p_2) =$$

$$= (y_1 - y)p_1p_2 + y$$

De modo que si el jugador prefiriera $E_1^*(Y)$ a $E_2^*(Y)$ ($E_2^*(Y)$ a $E_1^*(Y)$), debería preferir $E_3^*(Y)$ a $E_4^*(Y)$ ($E_4^*(Y)$ a $E_3^*(Y)$). Es decir, si se verifica que $y_2 > y_1p_1 + y(1-p_1)$, entonces se debería verificar que $y_2 > y_1p_1 + y_2(1-p_1)(y_1 < y_2 < y)$.

Esta relación entre el orden de preferencias y las utilidades esperadas que impone el axioma de independencia haciendo abstracción de la forma de la función de utilidad, hizo que la selección de este orden se basara habitualmente en comparaciones entre los valores esperados y los momentos centrados de segundo orden de las distribuciones de probabilidades[79]. Sin embargo, debe tenerse en cuenta que, dado que en general la esperanza moral es una función de la sucesión completa de momentos de la distribución de probabilidades, la definición de una regla de selección a partir de los dos primeros momentos, exclusivamente, sólo es válida para un conjunto limitado de funciones de utilidad o para clases muy especiales de distribuciones de probabilidades. Por ejemplo, si la función de utilidad es cuadrática, el cálculo de la utilidad esperada involucra a la esperanza matemática y al momento centrado de segundo orden. En términos más generales, Markowitz (1959) y Richter (1960) demostraron que, si la función de utilidad está definida por un

[79] Ver Markowitz (1959), Tobin (1958)(1965).

polinomio de grado k, entonces deben ser considerados los k primeros momentos naturales. Ahora bien, si dos distribuciones difieren en todos esos k momentos, entonces la definición del orden de preferencias requiere el conocimiento de las magnitudes y de los signos de las k derivadas de la función de utilidad, lo cual implica numerosas restricciones sobre dicha función con la consiguiente pérdida de generalidad de los resultados[80].

La mal llamada paradoja de Allais constituye una prueba de que el cuarto axioma, en la práctica, tiende a ser violado sistemáticamente, generando un comportamiento de los jugadores frente al riesgo inconsistente con la ya mencionada hipótesis de maximización de la utilidad esperada y la linealidad de las preferencias sobre las probabilidades[81].

La cuestión fundamental inherente a la controversia entre la axiomática de von Neumann-Morgenstern y el planteo de Allais se refiere a la cuestión de la preservación de orden de preferencias entre un evento cierto y uno incierto cuando ambos se transforman en inciertos.

De acuerdo con la propuesta de Friedman; Savage (1948), dadas una variable aleatoria unidimensional Y con dominio $\Omega(Y) = \{y_1, y_2, \dots, y_n\}$, que representa el conjunto de utilidades posibles, tal que sus valores observan el siguiente orden $y_i > y_j$, si $i > j$ y dos asignaciones de probabilidades sobre dichos resultados, $\pi_1(p_i)$ y $\pi_2(p_i)$, tales que $\pi_1(p_i)$ es estocásticamente dominante respecto de $\pi_2(p_i)$, entonces se demuestra que es posible definir un orden de preferencias basado en las esperanzas morales:

$$E_2^*(Y) = \sum_{i=1}^{n} y_i \pi_2(p_i) > \sum_{i=1}^{n} y_i \pi_1(p_i) = E_1^*(Y)$$

sin importar la forma que asuma la función de utilidad[82].

Supóngase, en particular, una función de utilidad $y(x)$, continua y con derivada primera continua y positiva en el intervalo $[x_1, x_n]$ y dos distribuciones de probabilidades, $\varphi_1[g(y)]$ y $\varphi_2[g(y)]$ sobre un intervalo finito $\Omega(X)$, se demuestra que, si $\varphi_1[g(y)]$ es estocásticamente dominante de primer orden respecto de $\varphi_2[g(y)]$, entonces $\varphi_2[g(y)]$ será débilmente

[80] La debilidad de los métodos que involucran comparaciones de valores esperados y momentos centrados de segundo orden exclusivamente se manifiesta, por ejemplo, en la definición de la frontera eficiente de Markowitz.
[81] Que constituya una prueba de que el cuarto axioma, en la práctica, tiende a ser violado sistemáticamente, implica reconocer que la paradoja de Allais no es una paradoja, sino que representa un razonable sistema de preferencias.
[82] Ver Sec. 9.

preferida a $\varphi_1[g(y)]$. Es decir, si se verifica que las funciones de distribución $G_1(x)$ y $G_2(x)$ son tales que $G_2(x) \leq G_1(x)$ (para todo $x \epsilon \Omega(X)$), entonces será $E_2^*(Y) \geq E_1^*(Y)$. De la misma forma se demuestra que la condición $G_2(x) < G_1(x)$ (para al menos un valor de X) es necesaria y suficiente para poder asegurar que $\varphi_2[g(y)]$ será estrictamente preferida a $\varphi_1[g(y)]$ y, por lo tanto, que $E_2^*(Y) > E_1^*(Y)$. Viceversa, si $\varphi_2[g(y)]$ es preferida a $\varphi_1[g(y)]$ para toda función de utilidad, entonces $\varphi_1[g(y)]$ es estocásticamente dominante de primer orden respecto de $\varphi_2[g(y)]$ y, en consecuencia $E_2^*(Y) > E_1^*(Y)$. La condición necesaria y suficiente para la preferencia estricta es que las distribuciones $\varphi_1[g(y)]$ y $\varphi_2[g(y)]$ no sean idénticas.

Dado que las variables $Y^s\big(s = 1,3,5, \dots ; Y \epsilon \Omega(X)\big)$ son funciones monótonas, se puede concluir en forma inmediata que, si $\varphi_1[g(y)]$ es estocásticamente dominante de primer orden respecto de $\varphi_2[g(y)]$, entonces todos los momentos naturales de orden impar de la distribución $\varphi_2[g(y)]$ son mayores o iguales que los respectivos momentos de $\varphi_1[g(y)]$. En particular, si el dominio $\Omega(X)$ incluye solamente valores no-negativos, entonces todos los momentos naturales de $\varphi_2[g(y)]$ serán mayores que los correspondientes momentos de $\varphi_1[g(y)]$.

Asimismo, si se considera una función de utilidad marginal no-creciente, más precisamente una función de utilidad con derivada de primer orden continua y positiva y derivada de segundo orden continua y menor o igual que cero en un intervalo $[x_1, x_n]$ (es decir una función representativa de una aversión débil al riesgo), se demuestra que la condición necesaria y suficiente para poder asegurar que $\varphi_2[g(y)]$ será, al menos, tan preferida como $\varphi_1[g(y)]$ -es decir que $E_2^*(Y) \geq E_1^*(Y)$- es que $\pi_1(p_i)$ sea estocásticamente dominante de segundo orden respecto de $\pi_2(p_i)$, es decir, que:

$$\sum_{i=1}^{k} F_2(y_i)\Delta y_i \leq \sum_{i=1}^{k} F_1(y_i)\Delta y_i.$$

Como corolario de este resultado se obtiene entonces que, contrariamente a lo que ocurre en la dominancia de primer orden, la condición de dominancia de segundo orden no impone ninguna restricción sobre las magnitudes relativas de los momentos de orden superior.

A partir de estos resultados, se puede concluir que: **i)** la condición de dominancia de segundo orden permite un ordenamiento de las preferencias de alcances más generales que la dominancia de primer orden; **ii)** la dominancia de primer orden es la condición más débil para garantizar un orden de preferencias entre dos opciones alternativas para toda función de utilidad continua con derivada de primer orden continua y positiva en el dominio $\Omega(Y)$; **iii)** dado que la experiencia indica que los jugadores tienden a sobrevaluar las probabilidades pequeñas y a subvaluar las probabilidades próximas a la unidad[83] para valores de p_i por debajo de una cierta cota, $\pi(p_i)$, es una función cóncava de p_i; **iv)** es de esperar que todo jugador cuya estrategia de decisión se base en la maximización de la esperanza moral subjetiva prefiera una distribución de probabilidades estocásticamente dominada; **v)** dado que las preferencias no son lineales en las probabilidades $\pi(p)$[84], la esperanza moral subjetiva no satisface la condición de monotonicidad en el sentido que cualquier maximización individual de la esperanza moral subjetiva generará necesariamente una preferencia por distribuciones dominadas y **vi)** la no-linealidad de $\pi(p)$ no está caracterizada por la forma de la función de utilidad exclusivamente y, en consecuencia los principales resultados de la teoría de la utilidad esperada (como por ejemplo, la caracterización de la aversión al riesgo por la concavidad de la función de utilidad) no pueden ser aplicados.

De acuerdo con la teoría de la dominancia estocástica, la paradoja de Allais se genera: **i)** por la tendencia de los individuos a evaluar la preferencia del par de ganancias posibles estocásticamente dominante de acuerdo con una función de utilidad más adversa al riesgo que la utilizada para evaluar la preferencia del par dominado y **ii)** por la aplicación del "*efecto certeza*" de Kahneman; Tversky (1979) y del efecto "*coeficiente común*" de MacCrimmon; Larson (1979)[85] según el cual, si un individuo con una fortuna inicial dada es indiferente entre una probabilidad $\pi_1(p)$ de ganar un monto x_1 y una probabilidad $\pi_1(p)\pi_2(p)$ de ganar x_2, entonces preferirá débilmente una probabilidad $\pi_1(p)\pi_2(p)\pi_3(p)$ de ganar x_2 a una probabilidad $\pi_1(p)\pi_3(p)$ de ganar x.

10.3.- Otras medidas de posición

Otro valor típico de la posición de una variable aleatoria es la "mediana", definida como el valor de la variable (x_{me}) que divide a la función de probabilidades en dos partes iguales. Es decir, el valor de x tal que:

$$p(X < x_{me}) \leq \frac{1}{2} \leq p(X \leq x_{me})$$

[83] *"Los jugadores, cuando apuestan, prefieren algunas probabilidades a otras (...) Las preferencias no pueden ser explicadas mediante la hipótesis de la utilidad esperada"* (Edwards (1962, p. 110)). Ver Preston; Baratta (1948), Griffith (1949), Sprowls (1953), Nogee; Lieberman (1960).

[84] Una afirmación basada en estudios empíricos sobre el comportamiento en condiciones de incertidumbre que comenzaron con los experimentos de Preston; Baratta (1948), Griffith (1949), Edwards (1953)(1954), Sprowls (1953).

[85] La *"Paradoja de Bergen"* que figura en Hagen (1979, pags. 278-279/290-292), es un caso particular de este efecto (ver también Allais (1979), Tversky (1975)).

Si la función $F_X(x)$ es continua, siempre existirá un valor $x = x_{me}$ tal que $F_X(x_{me}) = \frac{1}{2}$.

De la misma forma que se define la mediana es posible definir otros valores típicos de la posición de la variable denominados "cuantiles", que dividen a la función de probabilidades en un cierto número de partes iguales. Interesan, en particular, los denominados "cuartiles" $(x_{Q_1}, x_{Q_2}, x_{Q_3})$, que dividen a la distribución en cuatro partes iguales:

$$p(X < x_{Q_1}) \leq \frac{1}{4} \leq p(X \leq x_{Q_1})$$

$$p(X < x_{Q_2}) \leq \frac{1}{2} \leq p(X \leq x_{Q_2})$$

$$p(X < x_{Q_3}) \leq \frac{3}{4} \leq p(X \leq x_{Q_3})$$

y los denominados deciles $(x_{d_j}, j = 1,2, \dots ,9)$ que dividen a la distribución en diez partes iguales, $p\left(X < x_{d_j}\right) \leq \frac{j}{10} \leq p\left(X \leq x_{d_j}\right) (j = 1,2, \dots ,9)$.

Otra medida de posición es el valor central llamado "modo", "moda" o "valor modal", definido por el valor de la variable (x_{mo}) al cual le corresponde la mayor probabilidad.

Ejemplo n° 26:

Sea X una variable aleatoria discreta con distribución de probabilidades de la forma $p(X = x) = \frac{1}{n}(x = 1,2, \dots , n)$. Para n impar se tiene que:

$$p\left(x < \frac{n+1}{2}\right) = p\left(x \leq \frac{n+1}{2} - 1\right) = \frac{n-1}{2}\frac{1}{n} < \frac{1}{2}$$

$$p\left(x \leq \frac{n+1}{2}\right) = \frac{n+1}{2}\frac{1}{n} > \frac{1}{2}$$

Es decir, existe un único valor, $x_{me} = \frac{n+1}{2}$, que satisface la definición de mediana.

Si n es par, todos los valores de X comprendidos en el intervalo $\frac{n}{2} \leq x < \frac{n+1}{2}$ satisfacen la condición de la mediana.

10.4.- La varianza

10.4.1.- Definición

Una clase de momentos particularmente apropiada para caracterizar la "forma" de la concentración de los valores de la variable alrededor de su valor central es la de los momentos centrados o con origen en la esperanza matemática. En especial el momento centrado de segundo orden define una medida del grado de dispersión de los valores de la variable con respecto a su valor esperado denominada "varianza" o "desvío medio cuadrático", $\mu_2(X) \equiv \sigma^2(x) = E\{[X - E(X)]^2\}$. Desarrollando en esta ecuación el binomio al cuadrado se obtiene la siguiente expresión de la varianza en función de los momentos naturales:

$$\sigma^2(X) = E(X^2) + [E(X)]^2 - 2E(X)E(X) = E(X^2) - [E(X)]^2$$

$$= m_2(X) - [m(X)]^2$$

De esta relación se concluye que $E(X^2) \geq [E(X)]^2$.

Esta medida, por ser un momento de segundo orden, posee una dimensión distinta a la de la variable. Para solucionar este problema se define, entonces, otra medida –denominada "coeficiente de dispersión" o "desvío estándar"- igual a la raíz cuadrada positiva de la varianza, $\sigma(X) = +\sqrt{\mu_2(X)}$.

Tanto la esperanza matemática como el coeficiente de dispersión están estrechamente vinculados al sistema cartesiano de referencia de la variable aleatoria. Así dada una traslación de la variable, su esperanza matemática se verá afectada en la misma medida y, dada una modificación de su unidad de medida, su coeficiente de dispersión se verá afectado en forma correspondiente. Sea, por ejemplo, una variable aleatoria X con valor esperado $E(X)$ y varianza $\sigma^2(X)$. La variable:

$$u = \frac{X - E(X)}{\sigma(X)} = \frac{X}{\sigma(X)} - \frac{E(X)}{\sigma(X)}$$

-denominada "estandarizada" o "tipificada"- posee valor esperado nulo y varianza igual a uno. La importancia práctica de esta transformación radica en la posibilidad que brinda de comparar la forma de las distribuciones de probabilidad correspondientes a variables diferentes.

Ejemplo n° 27:

Sea X una variable aleatoria discreta con función de probabilidades de la forma:

X	P(X)
-1	1/20
0	7/20
5	3/20
7	3/10
9	3/20

Sus momentos naturales de primero y segundo orden están dados por:

$$m(X) = E(X) = \sum_i x_i p_i =$$

$$= (-1)\left(\frac{1}{20}\right) + 0\left(\frac{7}{20}\right) + 5\left(\frac{3}{20}\right) + 7\left(\frac{6}{20}\right) + 8\left(\frac{3}{20}\right) = \frac{80}{20} = 4$$

$$m_2(X) = \sum_i x_i^2 p_i = (-1)^2\left(\frac{1}{20}\right) + 0^2\left(\frac{7}{20}\right) + 5\left(\frac{3}{20}\right) + 7\left(\frac{6}{20}\right) + 8\left(\frac{3}{20}\right)$$

$$= \frac{562}{20} = 28{,}1$$

Los correspondientes momentos centrados serán:

$$\mu_2(X) = \sigma^2(X) = \sum_i [x_i - E(X)]^2 p_i =$$

$$= (-1-4)^2\frac{1}{20} + (0-4)^2\frac{3}{20} + (5-4)^2\frac{3}{20} + (7-4)^2\frac{6}{20} + (8-4)^2\frac{3}{20}$$

$$= \frac{242}{20} = 12{,}1$$

$$\mu_3(X) = (-1-4)^3\frac{1}{20} + (0-4)^3\frac{3}{20} + (5-4)^3\frac{3}{20} + (7-4)^3\frac{6}{20} +$$

$$+(8-4)^3\frac{3}{20} = \frac{166}{20} = -8{,}3$$

$$\mu_4(X) = (-1-4)^4 \frac{1}{20} + (0-4)^4 \frac{3}{20} + (5-4)^4 \frac{3}{20} + (7-4)^4 \frac{6}{20} +$$

$$+(8-4)^4 \frac{3}{20} = \frac{3424}{20} = 171{,}20$$

Ejemplo n° 28:

Sea una variable aleatoria X con distribución de probabilidades de la forma:

$$f_X(x) = \begin{cases} 0 & para \quad x \leq a \\ \dfrac{1}{b-a} & \quad a < x < b \\ 0 & \quad x \geq b \end{cases}$$

Se obtiene en forma inmediata que:

$$E(X^s) = \int_{-\infty}^{\infty} x^s f_X(x) dx = \frac{1}{b-a} \int_a^b x^s dx = \frac{b^{s+1} - a^{s+1}}{(s+1)(b-a)}$$

En particular, será:

$$m(X) = E(X) = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

$$m_2(X) = \frac{b^2 - a^2}{3(b-a)} = \frac{b^2 + ab + a^2}{3}$$

Por lo tanto, será $\sigma^2(X) = m_2(X) - [m(X)]^2 = \frac{(b-a)^2}{12}$.

Ejemplo n° 29:

Sea una variable aleatoria X con distribución de probabilidades de la forma:

$$f_X(x) = \frac{1}{b\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2} (-\infty \leq x \leq \infty)$$

La esperanza matemática está definida por:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \frac{1}{b\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2} dx$$

Aplicando en esta expresión la sustitución $\frac{x-a}{b} = u$, se obtiene que:

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (ub + a)\, e^{-\frac{u^2}{2}} du = \frac{b}{\sqrt{2\pi}} \left[e^{-\frac{u^2}{2}} \right]_{\infty}^{-\infty} + a = a$$

Asimismo, el momento natural de segundo orden está definido por:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \frac{1}{b\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2} dx$$

Aplicando la misma sustitución que en el caso anterior, se obtiene que:

$$E(X^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (ub + a)^2\, e^{-\frac{u^2}{2}} du =$$

$$= \frac{1}{\sqrt{2\pi}} \left[b^2 \int_{-\infty}^{\infty} u^2 e^{-\frac{u^2}{2}} du + a^2 \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du + 2ab \int_{-\infty}^{\infty} ue^{-\frac{u^2}{2}} du \right]$$

e integrando por partes resulta que:

$$\int_{-\infty}^{\infty} ue^{-\frac{u^2}{2}} du = \left[e^{-\frac{u^2}{2}} \right]_{-\infty}^{\infty} = 0$$

$$\int_{-\infty}^{\infty} u^2 e^{-\frac{u^2}{2}} du = \left[-ue^{-\frac{u^2}{2}} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du = \sqrt{2\pi}$$

Es decir, que $E(X^2) = m_2(X) = b^2 + a^2$. Luego, será $\sigma^2(X) = m_2(X) - [m(X)]^2 = b^2$.

10.4.2.- Propiedades

A partir de su definición, es posible enumerar las siguientes propiedades de la varianza:

1) La igualdad $E\{[X - E(X)]^2\} = E\{|X - E(X)|^2\} = 0$ se verifica si y sólo si $X - E(X) = 0$, es decir, si $X = E(X)$ o, lo que es lo mismo, solamente si X es una variable que puede asumir un valor único que, obviamente, coincidirá con su valor esperado.

2) Sea una variable aleatoria X tal que su momento natural de segundo orden sea finito, $m_2(X) = E(X^2) < \infty$, y por lo tanto, tal que $E[(X-k)^2] < \infty$. Entonces, se puede escribir:

$$E[(X-k)^2] = E\left\{\left[\left(X - E(X)\right) + \left(E(X) - k\right)\right]^2\right\} =$$

$$= E\{[X - E(X)]^2\} + E\{[E(X) - k]^2\} + 2E\{[X - E(X)][E(X) - k]\} =$$

$$= \sigma^2(X) + [E(X-k)]^2 + 2[E(X) - k]E[X - E(X)] =$$

$$= \sigma^2(X) + [E(X) - k]^2 \geq \sigma^2(X)$$

Lo cual demuestra que los momentos de segundo orden de la forma $E[(X - k)^2]$ asumen su valor mínimo cuando los desvíos están calculados respecto de la esperanza matemática de la variable ($k = E(X)$). Se dice, entonces, que $E(X)$ define el centro de segundo grado de la variable X. En general, se denomina centro de grado s ($s > 0$) de la variable X al valor k_s que minimiza la expresión $E(|X-k|^s)$.

De la misma forma que se demostró que, con respecto a la varianza, el valor esperado es un centro de grado dos, se puede demostrar que la mediana es un centro de grado uno: Sea $k > x_{me}$, dividiendo la recta en tres intervalos, se tiene que:

$$E(|X-k|) - E(|X - x_{me}|) =$$

$$= \int_{-\infty}^{x_{me}} (|y-k| - |y - x_{me}|)\, dF_Y(y) + \int_{x_{me}}^{k} (|y-k| - |y - x_{me}|)\, dF_Y(y) +$$

$$+ \int_{k}^{\infty} (|y-k| - |y - x_{me}|)\, dF_Y(y)$$

Ahora bien, se verifica que:

$$para\ y \leq x_{me}: \quad |y-k| - |y - x_{me}| = (y-k) - (x_{me} - y) = k - x_{me}$$

$$para\ x_{me} < y < k: \quad |y-k| - |y - x_{me}| = (y-k) - (y - x_{me}) =$$

$$= k - x_{me} - 2y \geq x_{me} - k$$

$$para\ y \leq k: \quad |y - k| - |y - x_{me}| = (y - k) - (y - x_{me}) = x_{me} - k$$

Luego, se puede escribir:

$$E(|X - k|) - E(|X - x_{me}|) \geq [k - x_{me}]p(X \leq x_{me}) +$$

$$+ [x_{me} - k]p(x_{me} < X < k) + [x_{me} - k]p(X \geq k) =$$

$$= [k - x_{me}]p(X \leq x_{me}) = [k - x_{me}]\{p(X \leq x_{me}) - 1 + p(X \leq x_{me})\} =$$

$$= [k - x_{me}][1 - 2p(X \leq x_{me})]$$

De acuerdo con la definición de mediana que figura en la Sec. 10.3, se verifica que $1 - 2p(X \leq x_{me}) \geq 0$. Es decir, que $E(|X - k|) \geq E(|X - x_{me}|)$. Una demostración similar se obtiene para $k < x_{me}$. Luego, se puede concluir que la expresión $E(|X - k|)$ asume su valor mínimo para $k = x_{me}$.

Las propiedades **1)** y **2)** explican por qué la varianza ocupa un lugar preponderante en el análisis de las variables aleatorias. El hecho que se anule solamente si X es una variable degenerada ha motivado su adopción entre todos los momentos centrados como la medida óptima del grado de dispersión. Además, es el momento de menor orden que posee la propiedad **2)**. Debe agregarse a todo esto la simplicidad de su tratamiento la cual, como se verá oportunamente, se deriva de la propiedad de aditividad para variables aleatorias independientes. En realidad, todo momento de la forma $E(|X - E(X)|^s)$ es apto como medida del grado de dispersión de una variable aleatoria. La varianza tiene la ventaja de evitar las complicaciones que surgen de la utilización del valor absoluto.

3) Sea una variable aleatoria X y una constante k. De acuerdo con la definición de varianza, será:

$$\sigma^2(kX) = E\{[kX - E(kX)]^2\} = E\{k^2[X - E(X)]^2\} = k^2\sigma^2(X)$$

4) Propiedad de aditividad para variables aleatorias independientes

Sea una variable aleatoria bidimensional $Z = X + Y$. Por definición será:

$$\sigma^2(Z) = \sigma^2(X + Y) = E\{[(X + Y) - E(X + Y)]^2\} =$$

$$= E\left\{\left[\left(X - E(X)\right) + \left(Y - E(Y)\right)\right]^2\right\} =$$

$$= E\{[X - E(X)]^2\} + E\{[Y - E(Y)]^2\} + 2E\{[X - E(X)][Y - E(Y)]\}$$

Es decir:

$$\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y) + 2\gamma(X, Y)$$

Generalizando esta demostración, sea una variable aleatoria de n dimensiones, $Z = X_1 \pm X_2 \pm \cdots \pm X_n$, se verificará, entonces, que:

$$\sigma^2(Z) = \sigma^2(X_1 \pm X_2 \pm \cdots \pm X_n) = \sum_{i=1}^{n} \sigma^2(X_i) \pm \sum_{i \neq j} \gamma(X_i, X_j)$$

En consecuencia, si las n variables marginales son independientes, resulta que:

$$\sigma^2(X_1 \pm X_2 \pm \cdots \pm X_n) = \sum\nolimits_{i=1}^{n} \sigma^2(X_i)$$

Según lo expresada al tratar la covarianza (Sec. 10.2.5), la independencia es una condición suficiente, pero no necesaria para la verificación de esta propiedad de aditividad de la varianza.

5) La desigualdad de Bienaymé-Chebychev-Medolaghi[86]:

Sea una variable aleatoria X con función de distribución de probabilidades $F_X(x)$. Se verifica, entonces, que:

$$E(|X - E(X)|^s) = \int_{\Omega(X)} |X - E(X)|^s dF_X(x) =$$

$$= \int_{|X - E(X)| \leq a} |X - E(X)|^s dF_X(x) + \int_{|X - E(X)| > a} |X - E(X)|^s dF_X(x) \geq$$

[86] Habitualmente esta propiedad, relacionada con el resultado obtenido por el matemático ruso P. L. Chebychev (1867) (o Tchébycheff) para probar la generalización del teorema de Bernoulli, figura en la literatura como "desigualdad de Chebychev", pero como una evidencia irrefutable de la vigencia de esa ley de la sociología de la ciencia conocida como "ley de la eponimia de Stigler", esta relación había sido publicada por I.J. Bienaymé (1853)). Esta desigualdad fue generalizada por Medolaghi (1908), en la presentación de una teoría hiper-Normal del riesgo (ver Insolera (1923).

$$\geq \int_{|X-E(X)|\leq a} |X-E(X)|^s dF_X(x)$$

$$\geq a^s \int_{|X-E(X)|\geq a} dF_X(x) = a^s p(|X-E(X)| \geq a)$$

Luego, será:

$$p(|X-E(X)| \geq a) = 1 - p(|X-E(X)| < a) \leq \frac{E(|X-E(X)|^s)}{a^s}$$

Es decir:

$$p(|X-E(X)| < a) \geq 1 - \frac{E(|X-E(X)|^s)}{a^s}$$

Haciendo en esta expresión $s = 2$ y $a = t\sigma(X)(t > 0)$, se obtiene que:

$$p\big(|X-E(X)| < t\sigma(X)\big) \geq 1 - \frac{E\{[X-E(X)]^2\}}{t^2\sigma^2(X)} = 1 - \frac{1}{t^2}$$

La importancia de esta propiedad radica fundamentalmente en su gran generalidad: se verifica cualquiera sea la forma de la distribución de probabilidades de la variable[87].

6) La desigualdad de Kolmogorov

Sea $\{X_1, X_2, \dots, X_n\}$ una sucesión de variables aleatorias independientes, todas con valor esperado nulo y varianzas σ_i^2, sea la variable aleatoria $Y_n = X_1 + X_2 + \cdots + X_n$ y sean los eventos:

$$E_j\text{: } la\ suma\ parcial\ Y_j(j = 1,2, \dots, n)$$

$$es\ la\ primera\ que\ asume\ un\ valor\ absoluto$$

$$mayor\ o\ igual\ que\ una\ constante\ "a"$$

[87] Esta desigualdad puede ser considerada también como un corolario de la desigualdad de Markov según la cual, dada una variable aleatoria $X > 0$, se verifica que $p(X \geq a) < \frac{E(X)}{\alpha}\big(\alpha > E(X)\big)$.

Es decir:

$$E_1 = \{w/|Y_1| \geq a\}$$

$$E_j = \{w/[(|Y_1| < a) \cap \dots \cap \left(\left|Y_{j-1}\right| < a\right) \cap \left(\left|Y_j\right| \geq a\right)]\} \quad (j = 1,2, \dots, n)$$

$$E_0 = \{w/[(|Y_1| < a) \cap (|Y_2| < a) \cap \dots \cap (|Y_n| < a)]\}$$

Obviamente, se verificará que $E_1 \cup E_2 \cup \dots \cup E_n = \left(\max_{1 \leq j \leq n} \left|Y_j\right| \geq a\right)$ y $E_0 = \left(\max_{1 \leq j \leq n} Y_j < a\right)$. Por otra parte, $E_i \cap E_h = \emptyset$ y $\cup_{j=0}^{n} E_j = \Omega$. A partir de estas relaciones y utilizando una función indicador de los conjuntos E_j, se obtiene que $\sum_{j=0}^{n} I_{E_j} = 1$ y $\sum_{j=1}^{n} I_{E_j} \leq 1$. Esta última desigualdad permite escribir:

$$E(Y_n^2) \geq E\left[Y_n^2\left(\sum_{j=1}^{n} I_{E_j}\right)\right] = \sum_{j=1}^{n} E\left(Y_n^2 I_{E_j}\right) = \sum_{j=1}^{n} E\left\{\left[Y_j + \left(Y_n - Y_j\right)\right]^2 I_{E_j}\right\}$$

Ahora bien, se verifica que:

$$E\left(Y_n^2 I_{E_j}\right) = E\left(Y_n^2 I_{E_j}\right) + E\left[\left(Y_n - Y_j\right)^2 I_{E_j}\right] + 2E\left[\left(Y_n - Y_j\right) Y_j I_{E_j}\right] =$$

$$= E\left(Y_n^2 I_{E_j}\right) + E\left[\left(Y_n - Y_j\right)^2 I_{E_j}\right] +$$

$$+2E\left[\left(X_{j+1} + X_{j+2} + \dots + X_n\right)\left(X_1 + X_2 + \dots + X_j\right) I_{E_j}\right] =$$

$$= E\left(Y_n^2 I_{E_j}\right) + E\left[\left(Y_n - Y_j\right)^2 I_{E_j}\right] + 2E\left(Y_j I_{E_j}\right) E\left(Y_n - Y_j\right)$$

y, teniendo en cuenta que $\left(Y_n - Y_j\right)^2 \geq 0$, de la ecuación anterior se obtiene que:

$$E\left(Y_n^2 I_{E_j}\right) \geq E\left[\left(Y_n - Y_j\right)^2 I_{E_j}\right] + E\left(Y_j^2 I_{E_j}\right) \geq E\left(Y_j^2 I_{E_j}\right) \geq$$

$$\geq a^2 E\left(I_{E_j}\right) = a^2 p\left(A_j\right)$$

Luego, la desigualdad original queda expresada de la siguiente forma:

$$E(Y^2) \geq \sum_{j=1}^{n} E\left(Y_j^2 I_{E_j}\right) \geq a^2 p(A_j) = a^2 p\left(\bigcup_{j=1}^{n} A_j\right) = a^2\left(\max_{1\leq j\leq n}|Y_j| \geq a\right)$$

De lo que se concluye que $p\left(\max_{1\leq j\leq n}|Y_j| \geq a\right) \leq \frac{E(Y_n^2)}{a^2}$ (obsérvese que esta condición es más fuerte que la desigualdad de Bienaymé-Chebychev).

7) Sea una variable aleatoria bidimensional $Z = XY$, en la que las variables marginales que intervienen como factores son independientes entre sí.

De acuerdo con la definición de varianza, se puede escribir:

$$\sigma^2(XY) = E[(XY)^2] - [E(XY)]^2 = E(X^2)E(Y^2) - [E(X)]^2[E(Y)]^2 =$$

$$= m_2(X)m_2(Y) - [m(X)]^2[m(Y)]^2 =$$

$$= \{\sigma^2(X) + [m(X)]^2\}\{\sigma^2(Y) + [m(Y)]^2\} - [m(X)]^2[m(Y)]^2$$

Es decir,

$$\sigma^2(XY) = \sigma^2(X)\sigma^2(Y) + [m(Y)]^2\sigma^2(X) + [m(X)]^2\sigma^2(Y)$$

Generalizando este resultado, sea una variable aleatoria n-dimensional definida por $Z = X_1X_2 \dots X_n$, donde las n variables marginales son independientes entre sí. Se verificará, entonces, que:

$$\sigma^2(X_1X_2 \dots X_n) = \prod_{i=1}^{n}\{\sigma^2(X_i) + [m(X_i)]^2\} - \prod_{i=1}^{n}[m(X_i)]^2$$

La condición necesaria para obtener una expresión exacta de la varianza de un producto de variables en términos de las varianzas y covarianzas de las variables marginales es que estas sean independientes entre sí.

En general, dada una función $\varphi(X_1, X_2, \dots, X_n)$ de las variables aleatorias marginales, los dos primeros términos de su desarrollo en serie de Taylor en el entorno de su esperanza matemática, $\left(E(X_1), E(X_2), \dots, E(X_n)\right)$, permiten escribir:

$$\varphi(X_1, X_2, \dots, X_n) \approx \varphi[E(X_1), E(X_2), \dots, E(X_n)] +$$

$$+\sum_{i=1}^{n}[X_i - E(X_i)]\frac{\partial}{\partial X_i}\varphi(X_1, X_2, \dots, X_n)|_{X_1=E(X_1),\dots,X_n=E(X_n)} =$$

$$= \varphi[E(X_1), E(X_2), \dots, E(X_n)]$$

$$+\sum_{i=1}^{n}[X_i - E(X_i)]\,\varphi_i[E(X_1), E(X_2), \dots, E(X_n)]$$

Aplicando entonces la propiedad de aditividad de la varianza, se obtiene la siguiente aproximación a la varianza de una función arbitraria de variables aleatorias unidimensionales:

$$\sigma^2[\varphi(X_1, X_2, \dots, X_n)] \approx \sum_{i=1}^{n}\{\varphi_i[E(X_1), E(X_2), \dots, E(X_n)]\}^2\sigma^2(X_i) +$$

$$+\sum_{i=1}^{n}\sum_{j=1}^{n}\{\varphi_i[E(X_1), E(X_2), \dots, E(X_n)]\}\{\varphi_j[E(X_1), E(X_2), \dots, E(X_n)]\}\gamma(X_i, X_j)$$

10.5.- El coeficiente de correlación lineal de Pearson

A partir del resultado que proporciona la desigualdad de Schwarz, se tiene que:

$$[\gamma(X,Y)]^2 = (E\{[X - E(X)][Y - E(Y)]\})^2 \leq$$

$$\leq (E\{|[X - E(X)][Y - E(Y)]|\})^2 \leq$$

$$\leq E\{[X - E(X)]^2\}E\{[Y - E(Y)]^2\} = \sigma^2(X)\sigma^2(Y)$$

Con lo que queda demostrado que $\left|\frac{\gamma(X,Y)}{\sigma(X)\sigma(Y)}\right| \leq 1$. El cociente $\rho(X,Y) = \frac{\gamma(X,Y)}{\sigma(X)\sigma(Y)}$ define el coeficiente de correlación lineal entre las variables X e Y.

De acuerdo con la propiedad de aditividad de la varianza se puede escribir:

$$\sigma^2\left[\frac{X}{\sigma(X)} + \frac{Y}{\sigma(Y)}\right] = \sigma^2\left[\frac{X}{\sigma(X)}\right] + \sigma^2\left[\frac{Y}{\sigma(Y)}\right] + 2\gamma\left[\frac{X}{\sigma(X)} + \frac{Y}{\sigma(Y)}\right] =$$

$$= \frac{1}{\sigma^2(X)}\sigma^2(X) + \frac{1}{\sigma^2(Y)}\sigma^2(Y) +$$

$$+2E\left(\left[\frac{X}{\sigma(X)}-E\left(\frac{X}{\sigma(X)}\right)\right]\left[\frac{Y}{\sigma(Y)}-E\left(\frac{Y}{\sigma(Y)}\right)\right]\right)=$$

$$=2+\frac{2}{\sigma(X)\sigma(Y)}E\{[X-E(X)][Y-E(Y)]\}=2[1+\rho(X,Y)]$$

Como el primer miembro de esta ecuación es no-negativo, se verifica que $\rho(X,Y)\geq -1$. Será igual a -1 solamente si $\sigma^2\left[\frac{X}{\sigma(X)}+\frac{Y}{\sigma(Y)}\right]=0$. Es decir, si se verifica que $\frac{X}{\sigma(X)}+\frac{Y}{\sigma(Y)}=k$ o, lo que es lo mismo, si las variables X e Y están vinculadas por una relación lineal con pendiente negativa:

$$Y=-\frac{\sigma(X)}{\sigma(Y)}X+k\sigma(Y)$$

De la misma forma, $\sigma^2\left[\frac{X}{\sigma(X)}-\frac{Y}{\sigma(Y)}\right]=2[1-\rho(X,Y)]$. Por lo tanto, se verificará que $\rho(X,Y)\leq 1$. Será igual a la unidad cuando las variables estén vinculadas mediante una relación lineal con pendiente positiva $Y=\frac{\sigma(X)}{\sigma(Y)}X+k\sigma(Y)$. En general, para $\rho(X,Y)\neq 0$, se puede definir una recta que, utilizando una expresión muy poco rigurosa, sea "muy probable" que pase lo más próximo que sea posible a los puntos (x,y), $Y=a_1X+a_0$, conocida como recta de regresión de Y con respecto a X. Donde:

$$a_0=m(Y)-\rho(X,Y)\frac{\sigma(Y)}{\sigma(X)}m(X)$$

$$a_1=\rho(X,Y)\frac{\sigma(Y)}{\sigma(X)}$$

Se demuestra fácilmente que los coeficientes a_0 y a_1 son tales que minimizan el valor esperado $E[(Y-a_1X-a_0)^2]$. Este valor mínimo, $\sigma^2(\varepsilon)=\sigma^2(Y)[1-\rho^2(X,Y)]\leq\sigma^2(Y)$, será tanto menor cuanto más se aproxime a uno el coeficiente $\rho^2(X,Y)$ (es decir, cuanto más se aproxime a la forma lineal la distribución de (X,Y)), o cuanto menor sea la varianza marginal de Y (debe tenerse en cuenta que la recta de regresión se determina a partir de la minimización del valor esperado del cuadrado de la distancia, medida en la dirección del eje Y, entre el punto aleatorio (X,Y) y el punto de la recta correspondiente a la misma abscisa).

Todas las consideraciones precedentes conservan su validez en la definición de la recta de regresión de X con respecto a Y, $X=b_1Y+b_0$. Donde $b_0=$

$m(X) - \rho(X,Y)\frac{\sigma(X)}{\sigma(Y)}m(Y)$ y $b_1 = \rho(X,Y)\frac{\sigma(X)}{\sigma(Y)}$. La intersección de las dos rectas de regresión se produce en el baricentro de la distribución $\big(m(X), m(Y)\big)$. Si se verifica que $\rho(X,Y) = 0$, las rectas serán ortogonales, paralelas a los ejes de coordenadas y, si $|\rho(X,Y)| = 1$, las dos rectas coincidirán con los puntos (X,Y)[88].

Ejemplo n° 30:

Sea una variable aleatoria bidimensional $Z = X + Y$ con distribución de probabilidades de la forma:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-m_X}{\sigma_X}\right)^2 - \right.\right.$$

$$\left.\left. -2\rho\frac{x-m_X}{\sigma_X}\frac{y-m_Y}{\sigma_Y} + \left(\frac{y-m_Y}{\sigma_Y}\right)^2\right]\right\}$$

(para $-\infty \leq x \leq \infty$, $-\infty \leq y \leq \infty$, donde m_X, m_Y, ρ, σ_X y σ_Y son constantes reales, tales que $|\rho| \leq 1$ y $\sigma_X, \sigma_Y > 0$).

El exponente de esta función de densidad conjunta puede ser escrito de la siguiente forma:

$$\left(\frac{x-m_X}{\sigma_X}\right)^2 - 2\rho\frac{x-m_X}{\sigma_X}\frac{y-m_Y}{\sigma_Y} + \left(\frac{y-m_Y}{\sigma_Y}\right)^2 =$$

$$= u^2 - 2\rho uv + v^2 = (u^2 - 2\rho uv + \rho^2 v^2) - \rho^2 v^2 + v^2 =$$

$$= (u - \rho v)^2 - v^2(1-\rho^2)$$

De modo que ésta puede ser expresada como:

$$f_{X,Y}(x,y) =$$

$$= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} exp\left[-\frac{1}{2}\left(\frac{x-m_X}{\sigma_X}\right)^2\right] exp\left(-\frac{1}{2}\left[\frac{x-\left(m_X + \rho\sigma_X\frac{y-m_Y}{\sigma_Y}\right)}{\sigma_X\sqrt{1-\rho^2}}\right]^2\right)$$

[88] El concepto de correlación se debe a F. Galton (1888)(1890) (ver Stigler, S.M.(1986)(1989)).

La función de densidad de la variable marginal Y está dada por:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} exp\left[-\frac{1}{2}\left(\frac{x-m_X}{\sigma_X}\right)^2\right].$$

$$.\int_{-\infty}^{\infty} exp\left(-\frac{1}{2}\left[\frac{x-\left(m_X+\rho\sigma_X\frac{y-m_Y}{\sigma_Y}\right)}{\sigma_X\sqrt{1-\rho^2}}\right]^2\right)dx$$

Aplicando la sustitución $\frac{x-\left(m_X+\rho\sigma_X\frac{y-m_Y}{\sigma_Y}\right)}{\sigma_X\sqrt{1-\rho^2}} = u$ se obtiene que:

$$f_Y(y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} exp\left[-\frac{1}{2}\left(\frac{x-m_X}{\sigma_X}\right)^2\right]\sigma_X\sqrt{1-\rho^2}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}}du =$$

$$= \frac{1}{\sigma_Y\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{y-m_Y}{\sigma_Y}\right)^2\right]$$

De la misma forma se demuestra que la función de densidad para la variable marginal X está definida por:

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-m_X}{\sigma_X}\right)^2\right]$$

A partir de los resultados anteriores se puede definir la función de densidad condicionada:

$$f(X/Y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} =$$

$$= \frac{1}{\sigma_X\sqrt{2\pi}\sqrt{1-\rho^2}} exp\left(-\frac{1}{2\sigma_X^2(1-\rho^2)}\right)\left\{\left[x-\left(m_X+\rho\sigma_X\frac{y-m_Y}{\sigma_Y}\right)\right]^2\right\}$$

Obsérvese que la condición necesaria y suficiente para que las variables X e Y sean independientes, es decir, para que se verifique la igualdad:

$$f(X/Y) = f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-m_X}{\sigma_X}\right)^2\right]$$

es que $\rho = 0$.

La covarianza entre las variables marginales está dada por:

$$\gamma(X,Y) = E[(X - m_X)(Y - m_Y)] = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - m_X)(y - m_Y).$$

$$exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-m_X}{\sigma_X}\right)^2 - 2\rho\frac{x-m_X}{\sigma_X}\frac{y-m_Y}{\sigma_Y} + \left(\frac{y-m_Y}{\sigma_Y}\right)^2\right]\right)dxdy$$

$$= \frac{\sigma_X\sigma_Y}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} uv\ exp\left(-\frac{1}{2(1-\rho^2)}[(u-\rho v)^2 + v^2(1-\rho^2)]\right)dudv$$

$$= \frac{\sigma_X\sigma_Y}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty} ve^{-\frac{v^2}{2}}\left(\frac{1}{2\pi}\sqrt{1-\rho^2}\int_{-\infty}^{\infty} u\ exp\left[-\frac{1}{2}\frac{(u-v\rho)^2}{1-\rho^2}\right]du\right)dv$$

Si en la integral que figura entre paréntesis se realiza la sustitución $\frac{u-\rho v}{\sqrt{1-\rho^2}} = z$ se obtiene que:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(z\sqrt{1-\rho^2} + v\rho\right)e^{-\frac{z^2}{2}}dz = v\rho$$

Luego, será:

$$\gamma(X,Y) = \frac{\sigma_X\sigma_Y}{\sqrt{2\pi}}\int_{-\infty}^{\infty} v^2e^{-\frac{v^2}{2}}dv = \frac{\sigma_X\sigma_Y}{\sqrt{2\pi}}\rho\left(\left[-ve^{-\frac{v^2}{2}}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{v^2}{2}}dv\right)$$

$$= \rho\sigma_X\sigma_Y$$

De esta expresión se concluye fácilmente que el coeficiente ρ que formaba parte de la función de densidad conjunta, es el coeficiente de correlación lineal entre las variables marginales, $\rho = \frac{\gamma(X,Y)}{\sigma_X\sigma_Y} = \rho(X,Y)$.

10.6.- La asimetría

Los momentos centrados de orden impar son especialmente útiles para medir el grado de concentración de los valores de la variable a la izquierda o a la derecha de su valor central; es decir, el grado de distorsión de la función de densidad con relación a la distribución simétrica respecto de su valor esperado.

Sea una variable aleatoria X con distribución de probabilidades simétrica, es decir tal que las variables $[X - E(X)]$ y $-[X - E(X)]$ tengan la misma función de densidad:

$$F_{X-E(X)}(x) = p[X - E(X) \leq x] = p[X - E(X) > -x] = 1 - F_{X-E(X)}(-x)$$

Si la función $F_X(.)$ es absolutamente continua, la condición necesaria y suficiente para poder asegurar la asimetría de la distribución es que $f_X(x) = f_X(-x)$. Asimismo, si una variable $[X - E(X)]$ es simétrica con respecto al origen, entonces la variable $[X - E(X)]^{2s+1}(s = 1,2, \dots)$ también es simétrica con respecto al origen, tiene la misma distribución de probabilidades y, por lo tanto, el mismo valor esperado:

$$\mu_{2s+1}(X) = E\{[X - E(X)]^{2s+1}\} = E\{-[X - E(X)]^{2s+1}\} =$$

$$= E\{[X - E(X)]^{2s+1}\} = 0 \qquad (s = 1,2, \dots)$$

La medida de asimetría utilizada habitualmente es el momento estandarizado de tercer orden, $As(X) = \frac{\mu_3(X)}{\sigma^3(X)}$. Cuanto más se aleje de cero en valor absoluto este coeficiente, indicará que más asimétrica es la variable en cuestión.

10.7.- La kurtosis

El momento centrado de cuarto orden de la distribución Normal cumple la siguiente condición: $\mu_4 = 3\sigma^4$. Se puede definir, entonces, una medida que permita comparar una variable cualquiera con la variable Normal en lo que hace a su “kurtosis”, es decir al valor de la probabilidad del modo de la forma:

$$K = \frac{\mu_4(X)}{\sigma^4(X)} - 3$$

Según que este coeficiente sea mayor, igual o menor que cero, indicará que esta probabilidad modal es mayor, igual o menor a la de la función Normal; es decir, que la distribución de probabilidades dada es más, igual o menos "apuntada" que la distribución Normal[89].

10.8.- La esperanza matemática condicionada

10.8.1.- Definición

De acuerdo con lo expresado en la Sec. 8, dada una variable aleatoria bidimensional, (X, Y), es posible definir la función de probabilidades para la variable marginal X, condicionada por el supuesto de la ocurrencia de un valor particular y de la variable Y:

$$p\big[(X = x_i)/\big(Y = y_j\big)\big] = \frac{p_{ij}}{p_{.j}}\big(p_{.j} = p\big(Y = y_j\big) \neq 0\big)$$

(en el caso de variables aleatorias discretas) y:

$$f(x/y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}\,(f_Y(y) \neq 0)$$

(en el caso de variables aleatorias continuas). Se puede definir, entonces, la esperanza matemática condicionada, $E[X/(Y = y)]$, de la siguiente forma:

$$E\big[X/\big(Y = y_j\big)\big] = \sum_i x_i p\big[(X = x_i)/\big(Y = y_j\big)\big] = \frac{1}{p_{.j}}\sum_i x_i p_{ij}$$

$$E[X/(Y = y)] = \int_{\Omega(X)} xf(x/y)dx = \frac{1}{f_Y(y)}\int_{\Omega(X)} xf_{X,Y}(x, y)dx$$

Siempre que la suma o la integral existan.

Se puede concluir, entonces, que suponiendo que $E(X < \infty)$, la esperanza matemática $E(X/Y)$ es una función de la variable Y y, como tal, es una variable aleatoria.

Ejemplo n° 31:

Sea una variable aleatoria bidimensional (X, Y) con función de densidad conjunta de la forma:

---

[89] Si es más "apuntada" que la función Normal se dice que la variable es "leptokúrtica", si es menos "apuntada" que la Normal, se dice que es "platokúrtica".

$$f_{X,Y}(x,y)=\begin{cases}\dfrac{3}{11}(x^2+y) & para \quad 0\leq x\leq 2, 0\leq y\leq 1\\ 0\ para\ cualesquiera\ otros\ valores\ de\ x\ o\ de\ y\end{cases}$$

Las funciones de densidad marginales son:

$$f_X(x)=\frac{3}{11}\int_0^1(x^2+y)dy=\frac{3(2x^2+1)}{22} \qquad (0\leq x\leq 2)$$

$$f_Y(Y)=\frac{3}{11}\int_0^2(x^2+y)dx=\frac{2(3y+4)}{11} \qquad (0\leq y\leq 1)$$

El valor esperado de la variable X condicionado por el supuesto acerca de la ocurrencia de Y asume la forma:

$$E(X/Y)=\int_0^2 xf(x/y)\,dx=\frac{3}{2(3y+4)}\int_0^2 x(x+y)dx=\frac{3}{2}\frac{2y+4}{3y+4}$$

10.8.2.- Propiedades

Las definiciones anteriores permiten concluir que, dado un valor particular, y, que se supone ha sido asumido por la variable Y, la esperanza matemática de una variable Xcondicionada por este supuesto define el valor esperado de una distribución de probabilidades específica y, por lo tanto, posee todas las propiedades de la esperanza matemática.

1) Sean X e Y dos variables aleatorias independientes. Será, entonces:

$$E\left[X\ /\ \left(Y=y_j\right)\right]=\frac{1}{p_{.j}}\sum_i x_i p_i=\frac{1}{p_{.j}}\sum_i x_i p_{i.}=\sum_i x_i p_{i.}=E(X)$$

$$E[X\ /\ (Y=y)]=\frac{1}{f_Y(y)}\int_{\Omega(X)} xf_X(x)f_Y(y)dx=E(X)$$

2) De acuerdo con la desigualdad propuesta en la Sec. 10.4.1, dadas dos variables aleatorias, X e Y, se verifica que $\{E[X/(Y=y)]\}^2\leq E[X^2/(Y=y)]$.

Como corolario de esta propiedad se demuestra, además, que[90]:

$$E[(X-a)^2/(Y=y)] \geq E\big(\{X-E[X/(Y=y)]\}^2/(Y=y)\big)$$

De lo que se deduce que $E[(X-a)^2/(Y=y)]$ asume su valor mínimo cuando a coincide con la esperanza matemática condicionada. Es decir que $E[X/(Y=y)]$ define un centro de segundo grado de la variable aleatoria condicionada $[X/(Y=y)]$.

3) Dadas dos variables aleatorias, X e Y, la esperanza matemática condicionada de la variable X es una función de Y, por lo tanto, suponiendo que $E(X)$ exista, su valor esperado con respecto a la distribución de probabilidades de la variable Y será de la forma:

$$E[E(X/Y)] = \sum_j E\big[X/\big(Y=y_j\big)\big]p_{.j} = \sum_i x_i \sum_j p_{ij} = \sum_i x_i p_{i.} = E(X)$$

Para el caso de variables aleatorias continuas, será:

$$E[E(X/Y)] = \int\limits_{\Omega(Y)} E[X/(Y=y)]f_Y(y)dy = \int\limits_{\Omega(Y)} \left[\int\limits_{\Omega(X)} x f_{X,Y}(x,y)dx\right] dy =$$

$$= \int\limits_{\Omega(X)} x \left[\int\limits_{\Omega(Y)} f_{X,Y}(x,y)dy\right] dx = \int\limits_{\Omega(X)} x f_X(x)dx = E(X)$$

4) Propiedad de linealidad: Sea una variable aleatoria discreta bidimensional, $Z = aX + bY$ (donde a y b denotan constantes). Por definición, se tiene que:

$$E\left[\frac{(aX+bY)}{(Z=z_h)}\right] =$$

$$= \sum_i \sum_j \big(ax_i + by_j\big)p\big\{\big[(X=x_i) \cap \big(Y=y_j\big)\big]/(Z=z_h)\big\}$$

---

[90] Ver Sec. 10.4.2.

Donde:

$$p\left\{\frac{\left[(X = x_i) \cap \left(Y = y_j\right)\right]}{(Z = z_h)}\right\} =$$

$$= \frac{p\left[(X = x_i) \cap \left(Y = y_j\right) \cap (Z = z_h)\right]}{p(Z = z_h)} = \frac{p_{ijh}}{p_{..h}}$$

Luego, será:

$$E[(aX + bY)/(Z = z_h)] = \frac{1}{p_{..h}} \sum_i \sum_j \left(ax_i + by_j\right) p_{ijh} =$$

$$= \frac{1}{p_{..h}} \left(a \sum_i x_i \sum_j p_{ijh} + b \sum_j y_j \sum_i p_{ijh} \right)$$

$$= \frac{1}{p_{..h}} \left(a \sum_i x_i p_{i.h} + b \sum_j y_j p_{.jh} \right) =$$

$$= aE[X/(Z = z_h)] + bE[Y/(Z = z_h)]$$

5) Sea Z una variable aleatoria independiente de la variable bidimensional (X, Y). Se verifica que:

$$E\left[ZX/\left(Y = y_j\right)\right] = \sum_i \sum_h x_i z_h p\left\{[(X = x_i) \cap (Z = z_h)]/\left(Y = y_j\right)\right\} =$$

$$= \frac{1}{p_{.j}} \sum_i \sum_h x_i z_h p_{ijh} =$$

$$= \frac{1}{p_{.j}} \sum_i \sum_h x_i z_h p_{ij.} p_{..h} = \left(\sum_h z_h p_{..h} \right) \left(\frac{1}{p_{.j.}} \sum_i x_i p_{ij.} \right) =$$

$$= E(Z) E\left[X/\left(Y = y_j\right)\right]$$

Un resultado similar se obtiene para variables aleatorias continuas.

10.9.- La varianza condicionada

10.9.1.- Definición

Sea una variable aleatoria bidimensional (X, Y). Con un razonamiento similar al empleado en la definición de la esperanza matemática condicionada, es posible definir la varianza condicionada, $\sigma^2\left[X/\left(Y=y_j\right)\right]$, de la siguiente forma:

$$\sigma^2\left[X/\left(Y=y_j\right)\right] = E\left\{\left[X - E\left(X/\left(Y=y_j\right)\right)\right]^2/\left(Y=y_j\right)\right\} =$$

$$= \sum_i \left[x_i - E\left(X / \left(Y=y_j\right)\right)\right]^2 p\left[(X=x_i) / \left(Y=y_j\right)\right] =$$

$$= \frac{1}{p_{.j}} \sum_i \left[x_i - E\left(X/\left(Y=y_j\right)\right)\right]^2 p_{ij}$$

$$\sigma^e[X/(Y=y)] = \int_{\Omega(X)} \left[x - E\big(X/(Y=y)\big)\right]^2 f(x/y)dx =$$

$$= \frac{1}{f_Y(y)} \int_{\Omega(X)} \left[x - E\big(X/(Y=y)\big)\right]^2 f_{XY}(x,y)dx$$

según se trate de variables discretas o continuas.

10.9.2.- Una propiedad

Sean las variables aleatorias X, Y y Z. A partir de la definición de varianza dada en la Sec. 10.4.1, se puede escribir:

$$\sigma^2(X) = E\left\{\left[X_{E(X)}\right]^2\right\} =$$

$$= E\left(\left\{(X - E[X/(Z=z)]) + \left(E[X/(Z=z)] - E(X)\right)\right\}^2\right) =$$

$$= E\left(\left\{X - E\big(X/(Z=z)\big)\right\}^2\right) + E(\{E[X/(Z=z)] - E(X)\}^2) +$$

$$+2E(\{X - E[X/(Z=z)]\}\{E[X/(Z=z)] - E(X)\})$$

Ahora bien, de acuerdo con la propiedad **3)** de la esperanza matemática condicionada, se verifica que:

$$E(\{X - E(X/(Z = z))\}^2) = E\{E(\{X - E(X/(Z = z))\}^2/(Z = z))\} =$$

$$= E\{\sigma^2[X \,/\, (Z = z)]\}$$

Por otra parte, será:

$$E(\{E[X/(Z = z)] - E(X)\}^2) = \sigma^2\{E[X/(Z = z)]\}$$

y,

$$E(\{X - E[X/(Z = z)]\}\{E[X/(Z = z)] - E(X)\}) =$$

$$= E\big\{E\big(\{X - E[X/(Z = z)]\}\{E[X/(Z = z)] - E(X)\}/(Z = z)\big)\big\} =$$

$$= E\big(\{E[X/(Z = z)] - E(X)\}E\big\{\big[X - E[X/(Z = z)]\big]/(Z = z)\big\}\big) = 0$$

(debe tenerse en cuenta que $E\big\{\big[X - E\big(X/(Z = z)\big)\big]/(Z = z)\big\} = 0$).

Luego, se puede concluir que:

$$\sigma^2(X) = E\{\sigma^2[X/(Z = z)]\} + \sigma^2\{E[X/(Z = z)]\}$$

10.10.- La covarianza condicionada

10.10.1.- Definición

Dadas tres variables aleatorias, X, Y y Z, se puede definir la covarianza entre X e Y, condicionada por Z, de la siguiente forma:

$$\gamma[X, Y \,/\, (Z = z_h)] = E\big\{\big[X - E\big(X \,/\, (Z = z_h)\big)\big]\big[Y - E\big(Y \,/\, (Z = z_h)\big)\big]\big\} =$$

$$= \sum_i \sum_j \big[x_i - E\big(X/(Z = z_h)\big)\big]\big[y_j - E\big(Y/(Z = z_h)\big)\big].$$

$$.\, p\big[(X = x_i) \cap \big(Y = y_j\big)/(Z = z_h)\big]$$

$$\gamma[X, Y/(Z = z)] =$$

$$= \int_{\Omega(X)} \int_{\Omega(Y)} \left[x - E\big(X \,/\, (Z = z)\big)\right]\left[y - E\big(Y \,/\, (Z = z)\big)\right] f_{X,Y}(x,y)\,dxdy$$

según se trate de variables discretas o continuas.

10.10.2.- Una propiedad

A partir de la definición de covarianza dada en la Sec. 10.1, se puede escribir:

$$\gamma(X,Y) = E\{[X - E(X)][Y - E(Y)]\} =$$

$$= E\big\{\big[\big(X - E(X/Z)\big) + E(X/Z) - E(X)\big].$$

$$.\big[\big(Y - E(Y\,/\,Z)\big) + E(Y\,/\,Z) - E(Y)\big]\big\} =$$

$$= E\{[X - E(X/Z)][Y - E(Y/Z)]\} +$$

$$+E\{[E(X\,/\,Z) - E(X)][E(Y\,/\,Z) - E(Y)]\} +$$

$$+E\{[E(X\,/\,Z)][E(Y\,/\,Z)] - E(Y)\} + E\{[E(X\,/\,Z)][E(Y\,/\,Z)]\}$$

Ahora bien, de acuerdo con la propiedad **3)** de la esperanza matemática condicionada, se verifica que:

$$E\{[X - E(X/Z)][Y - E(Y/Z)]\} = E(E\{[X - E(X/Z)][Y - E(Y/Z)]\}) =$$

$$= E[\gamma(X,Y\,/\,Z)]$$

$$E\{[E(X\,/\,Z) - E(X)][E(Y\,/\,Z) - E(Y)]\} = \gamma[E(X\,/\,Z), E(Y\,/\,Z)]$$

y,

$$E\{[X - E(X/Z)][Y - E(Y/Z)]\} =$$

$$= E\{[E(X\,/\,Z) - E(X)][E(Y\,/\,Z) - E(Y)]\} = 0$$

Luego, se puede concluir que:

$$\gamma(X,Y) = E[\gamma(X,Y\,/\,Z)] + \gamma[E(X\,/\,Z), E(Y\,/\,Z)]$$

Ejemplo n° 32:

Sean X e Y dos variables aleatorias que representan los resultados a obtener al lanzar en forma independiente dos dados "clásicos" y sea la variable aleatoria bidimensional $Z = X + Y$. Si se supone que $Z = 2$, entonces:

$$E[X / (Z = 2)] = E[Y / (Z = 2)] = E\big[\big(XY / (Z = 2)\big)\big] = 1$$

Si se supone que $Z = 4$, será:

$$p\{[(X = i) \cap (Y = j)]/(Z = 4)\} = \frac{p[(X = i) \cap (Y = j) \cap (Z = 4)]}{p(Z = 4)} =$$

$$= \frac{p[(X = i) \cap (Y = j)]}{\frac{3}{36}}$$

Es decir,

$$p\{[(X = i) \cap (Y = j)] / (Z = 4)\} = \begin{cases} \frac{1}{3} & para \qquad i + j = 4 \\ 0 & \qquad\qquad i + j \neq 4 \end{cases}$$

Además, se verifica que:

$$p[(X = i) / (Z = 4)] = \frac{1}{3} \qquad (i = 1,2,3)$$

$$p[(Y = j) / (Z = 4)] = \frac{1}{3} \qquad (j = 1,2,3)$$

De lo que se concluye que, suponiendo la condición $Z = 4$, las variables aleatorias X e Y no son independientes; es más, están vinculadas por la relación lineal $X + Y = 4$. En consecuencia, quedan definidos los siguientes valores esperados:

$$E(X/Z = 4) = E(Y/Z = 4) = 1\frac{1}{3} + 2\frac{1}{3} + 3\frac{1}{3} = 2$$

$$E[X^2 / (Z = 4)] = E[Y^2 / (Z = 4)] = 1^2\frac{1}{3} + 2^2\frac{1}{3} + 3^2\frac{1}{3} = \frac{14}{3}$$

$$E[(XY) / (Z = 4)] = 1{,}3\frac{1}{3} + 2{,}2\frac{1}{3} + 3{,}1\frac{1}{3} = \frac{10}{3}$$

De modo que:

$$\sigma^2[X/(Z=4)] = \sigma^2[Y/(Z=4)] =$$

$$= E[X^2 \ / \ (Z=4)] - \{E[X \ / \ (Z=4)]\}^2 = \frac{2}{3}$$

$$\gamma[(X,Y)/(Z=4)] =$$

$$= E[(XY) \ / \ (Z=4)] - E[X \ / \ (Z=4)]E[Y \ / \ (Z=4)] = -\frac{2}{3}$$

$$\rho[(X,Y) \ / \ (Z=4)] = \frac{\gamma\big[\big(X, Y \ / \ (Z=4)\big)\big]}{\sigma[X \ / \ (Z=4)]\sigma[Y \ / \ (Z=4)]} = -1$$

11. -Acerca de la introducción del concepto de variable aleatoria en los modelos econométricos

En el transcurso del determinismo al aleatorismo en la representación del comportamiento de los fenómenos económicos la componente aleatoria fue interpretada originalmente como un producto de la combinación de los errores de medición de las variables y de los errores en las ecuaciones de las representaciones. Posteriormente Slutsky y Yule introdujeron la idea de "shock" aleatorio formado por la agregación de los errores y perturbaciones y Frisch, incorporando en forma parcial la teoría de la probabilidad a la econometría, asimiló las perturbaciones a "estímulos" entendidos como innovaciones Fechnerianas. La culminación de la aproximación probabilística se produjo con la "revolución Haavelmiana", la cual evolucionó hacia el teorema de descomposición predictiva de Wold, un hito fundamental en la solución de las deficiencias de las representaciones estructuralistas y en el tratamiento causal moderno de los fenómenos dinámicos.

Todo fenómeno dinámico $Y(t,w)$ es asimilable a un proceso estocástico que evoluciona en el dominio del tiempo ($t \in T$) y cuya configuración varía en el dominio de los "estados" (o de las "fases" o de las "variables") ($\omega \in \Omega(Y)$).

La interpretación determinística del comportamiento de $Y(t,w)$ (al menos a nivel macroscópico) se basa en ciertas premisas de orden metafísico[91]: **i)** que el ámbito al que pertenecen los fenómenos es real; **ii)** que existen leyes objetivas que rigen su comportamiento y **iii)** que estas leyes son inherentes a los fenómenos, racionales y asintóticamente cognoscibles. En otros términos, cada estado $\omega(t)$ se supone definido por la realización simultánea, en el momento t, de las infinitas variables aleatorias que forman su estructura causal:

$$\Omega\big(Y(t)\big) = \{Y(t-j), X_1(t-h_1), X_2(t-h_2), \dots\}$$

$(j, h_1, h_2, \dots \in \mathbb{R}\,; j > 0; h_i \geq 0; i = 1,2, \dots)$ y la sucesión temporal de los estados determina su trayectoria. De modo que el fenómeno $Y(t,w)$ queda definido como una entidad que evoluciona en un ámbito espacio-temporal caracterizado por el principio de causalidad.

Este supuesto principio de solidaridad universal que relaciona causalmente a los fenómenos y que hace que la naturaleza de $Y(t)$ aparezca como infinitamente complicada, permite concluir que la información con que cuenta el observador $(\Omega^*\big(Y(t)\big) \subset \Omega\big(Y(t)\big))$ siempre será insuficiente y que, en consecuencia, una parte importante de su comportamiento permanecerá ignorada para sí, de modo que, en ciertas condiciones de estacionariedad, se puede escribir $Y(t) = f\big[\Omega^*\big(Y(t)\big)\big] + \varepsilon(t)$, donde: **i)** $f\big[\Omega^*\big(Y(t)\big)\big]$ denota la representación del comportamiento de $Y(t)$ a partir del conjunto de información $\Omega^*\big(Y(t)\big)$, es decir, el comportamiento que debería observar $Y(t)$ si los factores incluidos en $\Omega^*\big(Y(t)\big)$ fueran sus únicas causas y no estuvieran afectados por errores de medición, y $f[\cdot]$ fuera la función que representara la verdadera relación causal invariante en el tiempo entre estos factores e $Y(t)$; **ii)** $\lim_{\Omega^*(Y(t)) \to \Omega(Y(t))} f\big[\Omega^*\big(Y(t)\big)\big] = Y(t)$ y **iii)** $\varepsilon(t)$ denota el componente azar-ignorancia $(\varepsilon(t) = Y(t) - f\big[\Omega^*\big(Y(t)\big)\big])$ y, en consecuencia, es tal que $\lim_{\Omega^*(Y(t)) \to \Omega(Y(t))} \varepsilon(t) = 0$.

En este paradigma, para un observador ideal que contara con un conjunto de información no afectado por errores de medición, que abarcara la totalidad de

---

[91] A las cuales no es posible atribuir ningún fundamento (ni inductivo, ni deductivo) y que, según Daston (1988), en la interpretación de Hume (1718), constituyen "*...una necesidad psicológica, un precepto casi involuntario instituido por la caritativa naturaleza para compensar las deficiencias de la razón humana*" (p. 202).

la estructura causal de un fenómeno, su trayectoria admitiría una representación en términos de mecánica clásica reversible en el dominio del tiempo, en la que: **i)** $Y(t)$ transcurriría uniformemente sin ninguna relación con ningún elemento externo y, en consecuencia, la diferencia entre el pasado y el futuro no poseería ningún significado y **ii)** conocido su estado presente, se podría reproducir su pasado y calcular su futuro en forma determinística. Es decir, una representación en la que el presente no sería sino un punto que separa el pasado del futuro[92].

La insuficiencia de este modelo en términos de mecánica clásica para explicar "*...un mundo inestable que conocemos a través de una ventana finita*" (Prigogine; Nicolis (1977, p. 16)), en el que el estado natural de los fenómenos es de no-equilibrio -un no-equilibrio constructivo que, como consecuencia de su propiedad fundamental de auto-organización, genera nuevos estados y nuevas estructuras complejas que sólo son imaginables en el ámbito de la irreversibilidad temporal- dio origen a una nueva formulación termodinámica -aleatorista-, cuya diferencia con la dinámica clásica radicó esencialmente: **i)** en la postulación del concepto de estado del proceso en un instante dado como resultante de una evolución orientada en el tiempo en la que, a diferencia del pasado y del presente, el futuro está formado por una sucesión de variables aleatorias no-observables vinculadas causalmente; **ii)** en la concepción de $f\left[\Omega^*\left(Y(t)\right)\right]$ como la representación de ciertas regularidades locales observadas y **iii)** en la sustitución de la interpretación clásica de $\varepsilon(t)$ como azar-ignorancia (epistemológico), generado por los "errores" en la medición de las variables por la interpretación como azar-absoluto (ontológico), generado por dichos errores más las innovaciones a que están sometidos los factores incluidos en el sistema $\Omega\left(Y(t)\right)$.

La econometría surge formalmente como una disciplina autónoma en el ámbito de la economía a fines de 1930 y comienzos de 1940, y se basó inicialmente en una interpretación determinística del comportamiento de los fenómenos económicos. Su hipótesis fundamental consistió en suponer que un modelo económico ($f\left[\Omega^*\left(Y(t)\right)\right](t = 1,2, \dots)$) constituye una representación incompleta del comportamiento de un fenómeno y que $\varepsilon(t)$ representa una

[92] Si se tiene en cuenta que la información infinita es indiscernible de ininteligibilidad, se puede concluir que, aún en una situación ideal en la que el observador contara con un conjunto de información de tamaño infinito, el problema de la inexplicabilidad del comportamiento del comportamiento de los fenómenos permanecería.

variable aleatoria no-observable definida por la agregación de los errores de medición que afectan a los factores incluidos en $\Omega^*\big(Y(t)\big)$ y de las influencias que ejercen sobre dicho fenómeno los infinitos factores de su entorno económico no incluidos en $\Omega^*\big(Y(t)\big)$[93].

Este punto de partida condujo con el tiempo a un proceso de formalización basado en el principio que la descripción de los aspectos cuantitativos de los fenómenos económicos a partir de conjuntos finitos de observaciones requería de la representación de sistemas estocásticos de variables interrelacionadas y, por lo tanto, de la teoría de la probabilidad para la estimación de dichas representaciones; es decir, para la construcción de modelos econométricos entendidos como el "*...vehículo útil para comparar la teoría económica con las observaciones, especificado de una forma medible y testeable*" (Qin (1993, p. 37) (ver Staehle (1933))[94].

Este proceso de formalización dio origen a varios problemas referidos: **i)** a la posibilidad de asegurar la adecuacidad de la representación de las relaciones teóricas; **ii)** a su identificación, es decir a la determinación de las condiciones que permitieran asegurar la unicidad de los coeficientes estimados y su interpretación, independientemente de las distribuciones de probabilidades consideradas y **iii)** a la modificación de la interpretación del término $\varepsilon(t)$ asociado a la representación $f\big[\Omega^*\big(Y(t)\big)\big]$ y, en consecuencia, de la componente residual ($\hat{\varepsilon}(t)$) asociada con el modelo correspondiente.

Durante el período pre-modelístico las aplicaciones econométricas consistieron en el cálculo y la predicción de relaciones económicas, sin considerar la confiabilidad estadística de los resultados. Dada la autorregresividad de las variables involucradas en la representación y la imposibilidad de asegurar la invariancia temporal de las condiciones que afectan el comportamiento del sistema, la teoría de la probabilidad era considerada como un argumento inapropiado para el análisis de los fenómenos económicos, de modo que, hasta la implementación del proceso de modelización, se utilizaron métodos de estadística matemática "sin

[93] Dado el carácter discreto de la información, las representaciones deben ser entendidas como aproximaciones discretas al comportamiento de los fenómenos en un espacio-tiempo continuo-continuo.
[94] De acuerdo con la tradición el término "modelo" fue introducido por Frisch en el Congreso de Econometric Society en 1931 y fue utilizado por primera vez por Tinbergen, en 1936 (Magnus; Morgan (1987), Morgan (1989)).

probabilidad”.

De acuerdo con esta concepción no-probabilística, las primeras interpretaciones atribuyeron la presencia de la componente aleatoria exclusivamente a “errores” en la medición de los factores incluidos en la representación (“*errores en las variables*”, según la denominación de Schultz (1925) y Frisch (1934))[95]. Ahora bien, esta interpretación resultaba insuficiente para explicar los errores provenientes de la omisión de variables (en general, de mayor importancia que los errores en las variables) que requería el agregado de otra componente aleatoria, lo cual implicaba la adopción de métodos estadísticos más avanzados, cuyo desarrollo implicaba una identificación desagregada del origen y las características de los errores. En principio, en el análisis de los ciclos económicos, esta insuficiencia fue resuelta mediante la introducción del concepto de “*errores en las ecuaciones*” y una mayor profundización en el estudio de la naturaleza de la aleatoriedad inherente al comportamiento de los fenómenos.

Originalmente el análisis de los ciclos se debe a Moore (1914)(1923), quien propuso una representación utilizando análisis harmónico y series de Fourier. Alternativamente Persons (1916)(1922-23) desarrolló un método conocido como el “*barómetro de Harvard*”, basado en la correlación entre series desfasadas un período (un modelo conocido hoy como “*indicador líder*”). Estos métodos perdieron vigencia ante las críticas de Slutsky (1927) y Yule (1926)(1927) referidas a la posible falta de significado conceptual de la correlación entre series cronológicas autorregresivas.

La propuesta de Yule se basó en el estudio de series perturbadas. Demostró que un esquema autorregresivo definido por un péndulo sometido a perturbaciones estocásticas genera un proceso “*armónico irregular*” y propuso una forma de representación utilizando cuadrados mínimos en base a una ecuación en-diferencias de la forma $Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2}$[96] más un “shock” aleatorio formado por la combinación de los errores y las perturbaciones que afectaban al comportamiento de Y_t. El hecho que los componentes de esta combinación fueran no-separables permite concluir que la solución de Yule no resolvía las cuestiones referidas a la forma en que las perturbaciones eran “absorbidas” por el sistema, ni a su interpretación conceptual.

[95] La idea de la posibilidad de existencia de errores de medición en las variables fue introducida por Gini (1921).
[96] Ecuación que puede ser considerada como el origen de las representaciones *AR*.

Basándose en los trabajos de Wicksell (1907), Hotelling (1927) y Walker (1931), Frisch (1933)(1936) propuso el método de descomposición estructural[97], cuyo objetivo era sistematizar, en ciertas aplicaciones, las investigaciones sobre la estimación de representaciones utilizando series cronológicas mediante métodos de estadística matemática basados en argumentos probabilísticos. Introdujo el concepto de "*shocks erráticos*" para completar la caracterización de las variables ε_t. Las consideró como "*...desvíos del comportamiento estructural (...) como algo nuevo espontáneo agregado a la estructura del proceso*" (p. 408) que dan origen a la modificación aleatoria de las regularidades locales observables en su comportamiento y constituyen "*...la fuente de energía que generan los ciclos económicos*", incorporando de esta forma, parcialmente, la teoría de la probabilidad al pensamiento econométrico[98]. Propuso, además, una clasificación de los "shocks" -de acuerdo con la forma en que son absorbidos por el sistema- en "*aberraciones*" y "*estímulos*", definiendo como estímulo a "*... una perturbación cuyo efecto perdura sobre estados sucesivos del sistema*" y como una aberración a "*... una perturbación que influye sobre el proceso solamente en el momento que se produce*" (p. 410), generada por el método de estimación utilizado[99].

De acuerdo con la tesis Fechneriana de la "*indeterminación por novedad*"[100] se puede considerar a la postulación de Frisch como el primer intento de

---

[97] Esta propuesta fue planteada por Frisch como una alternativa crítica al método utilizado habitualmente de "*descomposición mecánica*" (en las componentes tendencia, estacionalidad, ciclo y variaciones no-sistemáticas).

[98] Parcialmente en la medida que Frisch no estaba de acuerdo con la aplicación indiscriminada de la teoría de la probabilidad en economía.

[99] En particular, en este caso, como generadas por su método de estimación-identificación, al que denominó "*análisis de confluencia*", cuya finalidad era intentar resolver los problemas generados por la no consideración de relaciones ocultas para el observador, debido a los errores de medición en todas las variables (no solamente en la variable "dependiente", como postulaba el método de regresión utilizado habitualmente) y por la multicolinealidad (ver Hendry; Morgan (1989)).

[100] Fechner (1866)(1871)(1906). Esta tesis se basa en el principio que en la evolución de los fenómenos se van generando nuevas condiciones iniciales las cuales, de acuerdo con una ley general, conducen a efectos no ocurridos previamente y, por lo tanto, no permiten la recurrencia de comportamientos idénticos. Pero que, no obstante, cada conjunto de dichas condiciones conserva alguna similitud con los anteriores por lo que, si bien todo fenómeno posee cierta dosis de aleatoriedad objetiva, su comportamiento no es completamente libre. Ver Landro (2010).

interpretación de los estímulos como verdaderas innovaciones[101].

Por otra parte, a partir de los resultados de Slutsky acerca de la capacidad de los procesos aleatorios puros de generar procesos autorregresivos, concluyó *"...que las regularidades observadas en el comportamiento de Y_t podían ser consideradas como derivadas de un caos generado por elementos inconexos, por su propia inconexión*" (Frisch (1936, p. 107) y clasificó a dichos elementos (a los cuales denominó en forma poco feliz como "*factores caóticamente aleatorios*")[102] en "*coherentes*" e "*incoherentes*" según que fueran o no autorregresivos y demostró -generalizando los trabajos de Yule (1921) sobre "*sumas ponderadas móviles*" de variables aleatorias independientes- que un proceso coherente podía ser aproximado como una combinación lineal de variables incoherentes de la forma:

$$Y_t = \theta_0\varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \qquad (t = 1,2,\ldots)$$

(donde las variables Y_t son centradas).

Tinbergen (1935)(1937)(1938)(1939) realizó una notable síntesis de los métodos propuestos por R.A. Fisher y Koopmans y puede ser considerado el continuador inmediato de la obra de Frisch. Si bien coincidió con éste en las propiedades de las representaciones formadas por ecuaciones lineales ponderadas, abandonó su forma estructural combinada, compuesta por ecuaciones diferenciales y en-diferencias, a favor de ecuaciones discretas del tipo de las denominadas "*de rezagos distribuidos*" de la forma[103]:

$$Y_t = a_0 + b_0X_t + b_1X_{t-1} + \varepsilon_t^*$$

(donde los "shocks" ε_t^* representan los estímulos que preservan la dinámica del sistema).

La generalización del pensamiento probabilístico que generó la propuesta pionera de Frisch -la cual influyó más en los aspectos metodológicos que en los epistemológicos de la teoría de modelos- se vio impulsada por los avances de la teoría de la inferencia, la axiomatización de la teoría de la probabilidad

[101] Frisch (1938): "*...los estímulos pueden ser considerados como condicionantes de la evolución posterior del fenómeno, es decir como causantes de una suerte de cambio permanente de sus condiciones iniciales*" (p. 408).
[102] La calificación de "poco feliz" se basa en la aparente asimilación de los conceptos de caótico y aleatorio, de complejidad y aleatoriedad.
[103] El concepto de "*rezagos distribuidos*" fue introducido por I. Fisher en 1920, en un trabajo sobre oferta monetaria (ver Alt (1942)).

por obra de Kolmogorov (1933) y, posteriormente, por el desarrollo de la teoría de la decisión en condiciones de incertidumbre, la teoría del riesgo, el criterio de optimización mediante la maximización de la utilidad esperada y sus extensiones y por la aparición de la interpretación personalista del concepto de probabilidad por obra de Ramsey y de Finetti (ver Rowley; Hamouda (1987), Hamouda; Rowley (1988))[104]. En este sentido, merecen ser destacados los trabajos de Marschak (1937) -quien propuso la consideración del comportamiento de los operadores económicos como parte de la teoría de los juegos y, en consecuencia, la asimilación de los conceptos de función de preferencias y utilidad- y Hagstroem (1938) -quien propuso su descripción.

La culminación de la aproximación probabilística a la econometría se produjo con la llamada "*revolución Haavelmiana*" (Morgan (1987)(1989)). A partir de la propuesta esencialmente estocástica de Slutsky, de la postulación de Moore (1914) -reconsiderada por Schultz (1930)(1939)- sobre la utilización del análisis harmónico y de la aplicación rigurosa de la teoría de la probabilidad por Koopmans (1937), Haavelmo (1938a)(1938b)(1939)(1940a)(1940b) (1943)(1944) postuló que las ecuaciones estructurales debían ser interpretadas como leyes de comportamiento sólo en un sentido estadístico y que los modelos econométricos debían asumir la forma de las "*ecuaciones estructurales estocásticas*" de Frisch, incluyendo "*coeficientes estructurales estocásticos*". Esto lo condujo a la conclusión que la formalización de un procedimiento de modelización estocástica consistente y generalizado (basado en observaciones no-repetibles) requería inevitablemente de los métodos de máxima verosimilitud de Fisher y de la teoría de los tests de hipótesis de Neyman-Pearson, que permitieran decidir acerca de la correspondencia entre la interdependencia de las variables económicas y su tratamiento estadístico[105]. Haavelmo recurrió así al concepto de distribución conjunta de todas las variables observables como un argumento para justificar la forma de dicha interdependencia estadística y concluyó que para que esta solución fuera posible, era necesario atribuirle a los términos representativos de los errores ciertas distribuciones de probabilidades.

[104] Ver Landro; González (2013)(2014).

[105] En general la literatura considera que el punto de partida de la revolución Haavelmiana se encuentra en la respuesta de Haavelmo (1943) (fundamentalmente en Schumpeter (1939) y en el debate Keynes-Tinbergen) a los prejuicios planteados con respecto a la naturaleza no-experimental de las observaciones económicas y, en consecuencia, a la capacidad de la inferencia estadística basada en la probabilidad para verificar las supuestas verdaderas relaciones propuestas por la teoría económica.

Esta forma de especificación mediante la definición de la distribución de probabilidades conjunta es considerada, en general, como el fundamento de la revolución probabilística Haavelmiana y de los modelos de ecuaciones simultáneas y determinó la diferencia "*...entre la construcción de modelos econométricos y la formulación matemática de la teoría económica*" (Qin (1993, p. 58)). Es decir, determinó la diferencia entre los modelos econométricos y los modelos económicos.

Inmediatamente Mann; Wald (1943) aplicaron la especificación por la definición de la distribución conjunta y demostraron la consistencia y la Normalidad asintótica de los estimadores máximo-verosímiles en las ecuaciones lineales de la forma:

$$a_0Y_t + a_1Y_{t-1} + \cdots + a_jY_{t-j} = \varepsilon_t$$

(donde $\{Y_t\}$ denota un proceso débilmente estacionario de segundo orden y ε_t denota un vector de perturbaciones no-autocorrelacionadas, idénticamente distribuidas y con momentos de orden superior finitos).

Como corolario de esta aplicación surgió el problema práctico de cómo asegurar la completitud del sistema de ecuaciones desde un punto de vista estadístico. Con este fin Koopmans (1950) redefinió el concepto de completitud de Tinbergen y, a partir de la clasificación de las variables en exógenas y endógenas, postuló que un modelo se puede considerar completo cuando el número de ecuaciones coincide con el número de variables endógenas del sistema. Definió como variable exógena aquella que no está afectada por las variables endógenas ni por las perturbaciones e introdujo el concepto de variable predeterminada que incluye a las variables desfasadas y a las variables exógenas correspondientes al período t. A partir de estas definiciones propuso una transformación del método de identificación basado en la distribución conjunta de las perturbaciones, en otro basado en el producto de la distribución de las perturbaciones condicionada por las variables exógenas multiplicada por la distribución marginal de dichas variables exógenas.

Como una excepción a esa corriente de pensamiento que consideraba a la teoría de la probabilidad exclusivamente como el fundamento de los métodos estadísticos a aplicar en la construcción de los modelos estructurales, entendidos estos como representaciones (incompletas) de las verdaderas trayectorias de los fenómenos económicos más una componente representativa de los errores aleatorios, cabe mencionar los trabajos de Wold y Tintner.

El teorema de la descomposición predictiva de Wold (1938) -su tesis doctoral- resultó el hito fundamental en el análisis del tratamiento causal moderno de los fenómenos dinámicos. Consideró a las series cronológicas como la porción observable de un proceso estocástico, es decir, como "*una realización de una distribución de infinitas dimensiones*" (p. 4) y completó las propuestas de Yule y Slutsky demostrando que todo proceso $\{Y_t\}$ débilmente estacionario de segundo orden puede ser desagregado en una componente estructural aproximable por un $AR(p)(p \gg 0)$ y una combinación lineal de infinitos "*errores*" incorrelacionados, asimilables a variables aleatorias (latentes) no-observables:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

(donde las variables Y_t son centradas).

Cabe destacar que más allá que la condición de aleatoriedad pura implicaba suponer que las variables ε_t se distribuían de acuerdo con una función de densidad simétrica con valor esperado nulo y la condición de estacionariedad implicaba suponer que su varianza era finita e, ignorando la aproximación asintótica a la Normalidad que proporcionaba la versión de Laplace del teorema central del límite, Wold consideró "*insoportablemente arbitrario*" (Epstein (1987), p. 161) asignar una distribución de probabilidades particular a las variables ε_t.

Como corolario de este trabajo liminar sobre la teoría pura de los procesos de parámetro discreto y del análisis de regresión de H. Cramér y en oposición a los modelos de ecuaciones simultáneas, Wold desarrolló una aproximación recursiva de la forma[106]:

$$Y_t^{(i)} = F\left(Y_{t-1}^{(i)}, \dots, Y_{t-1}^{(i-1)}, \dots, Y_{t-2}^{(i)}, \dots, Y_{t-1}^{(n)}, Y_{t-2}^{(n)}, \dots\right) + \varepsilon_t^{(i)}$$

(Bentzel; Wold (1946)) y demostró (Wold (1949)(1954)) que este modelo proporcionaba una interpretación causal y que, en general, todo conjunto de

[106] Con respecto a esta ecuación, debe tenerse en cuenta que, en el planteo dinámico de Wold las relaciones simultáneas no son justificables en la medida que no permiten establecer el sentido de la causalidad.

series cronológicas puede ser representado formalmente como un sistema de encadenamiento causal ((1948)(1951))[107].

La importancia fundamental de estos resultados está evidenciada en el hecho que todo el desarrollo posterior de la teoría econométrica de los fenómenos dinámicos se basó en corolarios y generalizaciones del teorema de Wold.

Contrariamente a la posición de Wold -quien justificó la utilización de la teoría pura de la probabilidad en las representaciones de fenómenos económicos basándose en elementos abstractos inherentes a la definición clásica-, Tintner (1938a)(1938b) planteó la posibilidad de utilizar la probabilidad de acuerdo con una interpretación más amplia y más flexible, como un argumento para vincular "*...la teoría económica de las expectativas y la teoría estadística de los errores*" (1938b, p. 154)[108]. Consideró a los errores aleatorios como generados por la imposibilidad de obtener "*...las representaciones óptimas de los factores controlables*" y de predecir "*...los factores no-controlables*", clasificándolos como errores en las variables y errores debidos a la influencia de fluctuaciones aleatorias en la descomposición de los sistemas de series cronológicas y postuló que los primeros debían ser tratados por el método de la diferencia variable y los segundos por los métodos postulados por Wold. En particular, su propuesta se basó en la interpretación Carnapiana, la cual admite dos conceptos posibles de probabilidad: P_{r_1}, relacionado con el "*grado de confirmación*" y P_{r_2}, relacionado con la "*frecuencia empírica*" y concluyó que la primera definición podía constituir el fundamento de una teoría de los tests de hipótesis apropiada para la econometría[109].

En desacuerdo con la propuesta de Wold, según la cual $\{\varepsilon_t\}: WN - débil$, Marschak (1953) justificó la posible presencia de autocorrelaciones entre las perturbaciones y postuló que "*...deben ser consideraradas como parte del comportamiento estructural, en la medida que no hay ninguna razón*

[107] Wold (como un producto de la posición adoptada por la escuela sueca de economía) rechazó el paradigma Mashalliano-Walrasiano del equilibrio general como un estado persistente de la dinámica económica.

[108] Debe tenerse en cuenta que, dado que Wold no consideró la necesidad de testear las representaciones propuestas respecto de su adecuacidad a las correspondientes posiciones teóricas, no reparó en las condiciónes de aplicabilidad del concepto de probabilidad. Su propuesta postulaba que los sistemas recursivos no estaban afectados por los problemas de identificación.

[109] Ver Landro (2010).

económica para descartar la posibilidad de que formen un proceso estocástico en el que cada shock depende de uno o más de sus predecesores" (p. 21). Esta propuesta hizo que la investigación econométrica se dirigiera (fundamentalmente por obra de Cochran; Orcutt (1949), Durbin (1960a) (1960b), Sargan (1959)(1961)) al estudio de las implicaciones producidas por las autocorrelaciones entre la perturbaciones sobre las propiedades de los estimadores de los coeficientes estructurales.

Muth, en su trabajo fundamental de 1961, propuso un retorno a la interpretación de Frisch (1933) de los estímulos en el comportamiento estructural de un sistema como una caracterización de los efectos dinámicos de los "shocks" aleatorios autocorrelacionados (pero económicamente no especificados), débilmente estacionarios de segundo orden, asimilables a variables exógenas y representables como una combinación lineal de variables ε_t no-autocorrelacionadas con valor esperado nulo y varianza constante ($Y(t) = f\left[\Omega^*\left(Y(t)\right)\right] + \varepsilon^*(t)$, donde $\varepsilon^*(t) = \Theta(B)\varepsilon_t$ y $\{\varepsilon_t\}: WN$)[110].

En este punto la hipótesis de la ortodoxia econométrica que sostenía que siempre era posible especificar un modelo teórico que explicara la verdadera ley que rige el comportamiento de un fenómeno[111], comenzó a debilitarse y surgió la necesidad de modificaciones "ad hoc" en las especificaciones de los modelos estructurales para representar las eventuales regularidades locales, de interpretar a los modelos como hipótesis generales sobre las principales relaciones causales sugeridas por la teoría económica, testeables a partir de las observaciones. En este contexto de un naciente paradigma aleatorista, Theil (1957)(1958), a partir del supuesto que "*...los modelistas en general no conocen la verdadera especificación de una representación*" (Theil (1958, p. 215)) (cabría agregar, suponiendo que ésta exista), propuso una definición alternativa de "*especificación*" y, por lo tanto, de capacidad de predicción, basada en la minimización de la varianza residual como criterio de

[110] A diferencia de Frisch (1938) -quien utilizó una representación MA para descomponer una variable económica- Muth utilizó una representación MA para descomponer un "shock" aleatorio.
[111] Entendida la especificación como "*...la selección de la forma matemática de la población*" (Koopmans (1937, p. 3)).

optimización para la selección de un modelo[112], pero sin asignarle ninguna interpretación conceptual a los residuos.

En general en el paradigma estructuralista los principales esfuerzos estuvieron dirigidos a la identificación y estimación de los coeficientes estructurales a partir de las formas reducidas, dedicándole muy poca atención a la verificación del cumplimiento de las condiciones impuestas a los "shocks" exógenos (Granger; Newbold (1977): "*Los residuos son tratados habitualmente por los econometristas como meras molestias de poca importancia. Muchos textos de econometría introducen a los modelos como un conjunto de relaciones determinísticas entre variables y adicionan descuidadamente términos representativos del 'error' a las ecuaciones para explicar cosas como errores en la especificación del modelo y en la medición de las variables. La utilización de denominaciones como 'residuos' y 'errores' implica juicios de valor sobre la importancia de dichos términos. Si bien 'error' es una denominación adecuada en el contexto de la predicción, una denominación más apropiada podría ser 'innovaciones*'" (p. 8)).

Un breve retorno al determinismo en este transcurso hacia el aleatorismo lo constituye la obra de Leamer (1978), quien reinterpretó la propuesta de Theil sobre el problema de la especificación desde una perspectiva Bayesiana[113]. Consideró a las leyes económicas como el fundamento teórico de las representaciones estructurales y propuso considerar a los errores como independientes de esa "verdadera" formulación teórica generados por la agregación de elementos "*no-observables*" pero "*manejables*", debidos exclusivamente a deficiencias en la especificación de la representación[114] y concluyó, en consecuencia, que una representación "completa" implicaría la eliminación de las perturbaciones[115].

Dadas las debilidades evidenciadas por el estructuralismo en la formulación dinámica de las representaciones y a partir de extensiones del teorema de

---

[112] Es necesario tener en cuenta que "*...dado que este criterio no siempre es concluyente ni factible, deben considerarse dos criterios adicionales de naturaleza más subjetiva: plausibilidad y simplicidad*" (Theil (1958, p. 208)).
[113] La diferencia fundamental entre Theil y los econometristas clásicos y Leamer y los econometristas Bayesianos radica en que éstos estaban más interesados en el análisis de las propiedades de los estimadores de los coeficientes que en la minimización de la varianza residual de los clásicos.
[114] Ver Qin; Gilbert (2001).
[115] Una interpretación que puede considerarse resumida en la expresión atribuida a Tukey, según la cual "*El hombre construye* $f\left[\Omega^{*}\left(Y(t)\right)\right]$ *y Dios nos proporciona* ε_t".

Wold, Sargent (1976)(1977), Sargent; Sims (1977) y Sims (1980) impulsaron una aproximación a la explicación del comportamiento de los fenómenos económicos mediante modelos puros de series cronológicas. En particular, propusieron la utilización de representaciones de la forma $\Phi(B)Y_t = \varepsilon_t$, en las cuales el orden p del operador $\Phi(B)$ fuera tal que $\{\varepsilon_t\}$ pudiera ser considerado un proceso de innovaciones no-autocorrelacionadas, de modo que esta condición de $AR(0)$ del proceso $\{\varepsilon_t\}$ se verificara por construcción y no por una hipótesis "a priori" y la representación $\Phi(B)Y_t = \varepsilon_t$ constituyera una caracterización económicamente válida del comportamiento de Y_t.

La representación inversa, $Y_t = [\Phi(\mathrm{B})]^{-1}\varepsilon_t$ retorna, en una interpretación aleatorista, a la propuesta de Slutsky-Wold de representación lineal de un proceso "*coherente*" ($\{Y_t\}$) en términos de un proceso "*incoherente*" ($\{\varepsilon_t\}$). En la que el operador $[\Phi(\mathrm{B})]^{-1}$, como función que describe la medida del impacto de los "shocks" aleatorios, constituye el argumento más importante para considerar a las perturbaciones como innovaciones y comprender cómo estas innovaciones son asimiladas por el proceso $\{Y_t\}$.

12.- La convergencia estocástica

El concepto de convergencia fue introducido en el análisis matemático respecto al comportamiento de sucesiones de números y posteriormente generalizado a sucesiones de funciones. Como se vio en la Sec. 1.1, una sucesión de variables aleatorias $\{X_n, n = 1,2, \ldots\}$ constituye una sucesión de funciones $\{X_n(w)\}(w\epsilon\Omega)$.

Las condiciones para la convergencia de la sucesión $\{X_n(w)\}$ constituyen el ámbito de la convergencia estocástica, cuya diferencia con la convergencia en el sentido del análisis radica en que aquélla se verifica a través del operador probabilidades y da origen a diferentes definiciones. Aquí se tratarán sólo aquellos tipos de convergencia que se presentan más frecuentemente en las aplicaciones a la inferencia y a la teoría de modelos.

12.1.- La convergencia casi-con-certeza

12.1.1.- Definición

Sea una sucesión de variables aleatorias $\{X_n, n = 1,2, \ldots\}$, todas con valores en $\mathbb{R}_k (k \geq 1)$. Si se verifica que $p\left(\lim_{n\to\infty} X_n = 0\right) = 1$ se dice que $\{X_n\}$ converge casi-con-certeza a cero, $X_n \xrightarrow{ccc} 0$. De la misma forma, dadas una sucesión de variables aleatorias $\{X_n, n = 1,2, \dots\}$ y una variable aleatoria X, todas con valores en $\mathbb{R}_k (k \geq 1)$, si se verifica que $(X_n - X) \xrightarrow{ccc} 0$ entonces se dice que $\{X_n\}$ converge casi-con-certeza (o con probabilidad 1) a X, $X_n \xrightarrow{ccc} X$.

12.1.2.- Un teorema de existencia

La condición necesaria y suficiente para que se verifique la convergencia casi-con-certeza es que:

$$\lim_{n\to\infty} p\left[\bigcap_{j=n}^{\infty}\left(\left|X_j - X\right| \leq \varepsilon\right)\right] = 1 \qquad (\varepsilon > 0)$$

o, lo que es lo mismo, que:

$$\lim_{n\to\infty} p\left[\bigcup_{j=n}^{\infty}\left(\left|X_j - X\right| > \varepsilon\right)\right] = 0$$

Teniendo en cuenta que

$$p\left[\cup_{j=n}^{\infty}\left(\left|X_j - X\right| > \varepsilon\right)\right] \leq \textstyle\sum_{j=n}^{\infty} p\left(\left|X_j - X\right| > \varepsilon\right)$$

se puede concluir en forma inmediata que una condición suficiente para la convergencia casi-con-certeza es que $\sum_{j=1}^{\infty} p\left(\left|X_j - X\right| > \varepsilon\right) < \infty$ (para todo $0 < \varepsilon < \varepsilon_0$).

Como corolario de este resultado se demuestra que, dadas una sucesión de variables aleatorias $\{X_n\}$ y una variable aleatoria X, si se verifica que $E(|X_n - X|^s) < \infty$ $(s = 1,2, \dots)$ y que $\sum_{n=1}^{\infty}(|X_n - X|^s) < \infty$ $(s = 1,2, \dots)$, entonces se puede asegurar que $X_n \xrightarrow{ccc} X$.

12.2.- La convergencia en-distribución

12.2.1.- Definición

Sea una sucesión de variables aleatorias $\{X_n, n = 1,2, \dots\}$ y sea la variable aleatoria X, todas con valores en $\mathbb{R}_k (k \geq 1)$ y con funciones de distribución de probabilidades $F_{X_n}(x) = p(X_n \leq x)$ $(n = 1,2, \dots)$ y $F_X(x)$ (continua en $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$). Si se verifica que $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ $(\forall x \epsilon \mathbb{R}_k)$, para todo x para el cual $F_X(x)$ sea continua, se dice que $\{X_n\}$ converge en-distribución a la variable X (o que existe una semejanza asintótica entre X_n y X), $X_n \xrightarrow{d} X$ [116].

De acuerdo con la definición anterior, para que exista convergencia en-distribución $F_{X_n}(x)$ debe converger a $F_X(x)$ en todo punto x de continuidad de $F_X(x)$. Asimismo, si se verifica que $F_{X_n}(x) \to F_X(x)$ y $F_X(x)$ es una función de distribución de probabilidades continua para todo x, se demuestra que la convergencia es uniforme (teorema de Pólya).

Aplicando el concepto de límite (en el sentido del análisis matemático), la definición de convergencia en-distribución para sucesiones de variables aleatorias es generalizable a cualquier conjunto de variables aleatorias con funciones de distribución dependientes de un parámetro real: Sea $\{Y_t, t\epsilon T\}$ $(T\epsilon\mathbb{R}_k)$ un conjunto de variables aleatorias con funciones de distribución de probabilidades $F_{Y_t}(y) = p(Y_t \leq y)$ y sea una variable aleatoria Y con función de distribución de probabilidades $F_Y(y)$. Si se verifica la convergencia $\lim_{t\to t_0} F_{Y_t}(y) = F_Y(y)$ para todo punto de continuidad de $F_Y(\cdot)$, se dice que Y_t converge en-distribución a la variable Y, $Y_t \xrightarrow{d} Y$ $(t \to t_0)$.

Sea, ahora, una variable aleatoria discreta unidimensional X con distribución de probabilidades de la forma $p(X = x_i) = p_i (i = 1,2, \dots, k)$ y sea una sucesión de variables aleatorias $\{X_n\}$ cada una de las cuales puede asumir el mismo número de valores que X, con distribución de probabilidades de la forma $p(X_n = x_{n,i}) = p_{n,i} (i = 1,2, \dots, k)$ y tales que $x_{n,i} \to x_i$ y $p_{n,i} \to p_i$ cuando $n \to \infty$. Sea x un punto de continuidad de la función de distribución de probabilidades $F_X(x)$ (distinto de todos los valores x_i) y sea δ la distancia mínima entre x y los x_i. En virtud de la convergencia expresada más arriba, es posible hallar un valor N tal que, para todo $n \geq N$, se verifique que $|x_{n,i} - x_i| < \delta$ y, por lo tanto, que $x_{n,i} \leq x$ para todos (y solamente) los valores de i tales que $x_i \leq x$. Se puede escribir, entonces:

$$F_{X_n}(x) = p(X_n \leq x) = \sum_{x_{n,i} \leq x_i} p_{n,i} = \sum_{x_i \leq x} p_{n,i}$$

[116] La convergencia en-distribución ha sido definida aquí como una convergencia de variables aleatorias pero, en realidad, se refiere a las funciones de distribución de probabilidades, no a las variables aleatorias. Obsérvese que se puede hablar de convergencia en-distribución o, más simplemente, de convergencia de funciones de distribución sin necesidad de individualizar los posibles conjuntos de variables aleatorias a los que corresponden dichas funciones. La introducción de la variable aleatoria sólo logra que algunas propiedades se vuelvan intuitivamente más comprensibles.

Teniendo en cuenta, además, que $p_{n,i} \to p_i$, se tiene que:

$$F_{X_n}(x) = \sum_{x_i \leq x} p_{n,i} \to \sum_{x_i \leq x} p_i = F_X(x)$$

Con lo que se demuestra que $X_n \xrightarrow{d} X$.

Supóngase ahora que el dominio de la variable X sea infinito. Dado un $\varepsilon > 0$, es posible hallar un valor K tal que $\sum_{i=1}^{K} p_i > 1 - \varepsilon$. Es decir, tal que $\sum_{i=K+1}^{\infty} p_i < \varepsilon$. Luego, para todo $n \geq N$, se tiene que $\sum_{i=1}^{K} p_{n,i} > 1 - 2\varepsilon$. Es decir, $\sum_{i=K+1}^{\infty} p_i < 2\varepsilon$. Ahora bien, se verifica que:

$$\left|F_{X_n}(x) - F_X(x)\right| = \left|\sum_{x_{n,i} \leq x} p_{n,i} - \sum_{x_i \leq x} p_i\right| =$$

$$= \left|\sum_{\substack{x_{n,i} \leq x \\ i \leq K}} p_{n,i} + \sum_{\substack{x_{n,i} \leq x \\ i > K}} p_{n,i} - \sum_{\substack{x_i \leq x \\ i \leq K}} p_i - \sum_{\substack{x_i \leq x \\ i > K}} p_i\right| \leq$$

$$\leq \left|\sum_{\substack{x_{n,i} \leq x \\ i \leq K}} p_{n,i} - \sum_{\substack{x_i \leq x \\ i \leq K}} p_i\right| + \left|\sum_{i=K+1}^{\infty} p_{n,i} - \sum_{i=K+1}^{\infty} p_i\right|$$

Por lo tanto, de acuerdo con lo demostrado más arriba, seleccionando adecuadamente el valor de K, para todo n suficientemente grande, se verifica que $\left|\sum_{i=K+1}^{\infty} p_{n,i} - \sum_{i=K+1}^{\infty} p_i\right| < \varepsilon$. Asimismo, con respecto al primer término del segundo miembro de la ecuación que figura más arriba, según lo expresado en la primera parte de esta demostración, se tiene que $\lim_{n \to \infty} \left|\sum_{\substack{x_{n,i} \leq x \\ i \leq K}} p_{n,i} - \sum_{\substack{x_i \leq x \\ i \leq K}} p_i\right| = 0$. Con lo que queda demostrado que $F_{X_n}(x) \to F_X(x)$. Es decir, que $X_n \xrightarrow{d} X$.

Ejemplo n° 33:

Sea $\{X_1, X_2, \dots, X_n\}$ una sucesión de variables aleatorias con función de distribución de probabilidades de la forma:

$$F_n(x) = \begin{cases} 0 & para \qquad x \leq \dfrac{1}{n} \\ \dfrac{n}{2}\left(x + \dfrac{1}{n}\right) & -\dfrac{1}{n} < x \leq \dfrac{1}{n} \\ 1 & x > \dfrac{1}{n} \end{cases}$$

(es decir que la función $F_n(x)$ varía linealmente de 0 a 1 en el intervalo $\left(-\frac{1}{n}, \frac{1}{n}\right)$) y sea $F_X(x)$ la función de distribución de probabilidades correspondiente a una variable constantemente nula:

$$F_X(x) = \begin{cases} 0 & para \qquad x \leq 0 \\ 1 & x > 0 \end{cases}$$

Se obtiene en forma inmediata que $\lim\limits_{n \to \infty} F_n(x) = F_X(x)$. Obsérvese, por otra parte, que $\lim\limits_{n \to \infty} F_n(0) = \frac{1}{2} \neq 0 = F_X(x)$. De lo que se concluye que, en este caso, la convergencia puntual de $F_n(x)$ a $F_X(x)$ se verifica sólo en los puntos de continuidad de la función $F_X(x)$[117].

Ejemplo n° 34:

Sea $\{X_1, X_2, \dots, X_n\}$ una sucesión de variables aleatorias cada una de las cuales puede asumir con la misma probabilidad cualquiera de los valores del dominio $\Omega = \left\{\frac{1}{\mathrm{n}}, \frac{2}{\mathrm{n}}, \dots, \frac{\mathrm{n}-1}{\mathrm{n}}, 1\right\}$. La función de distribución de probabilidades de la variable $X_i (i = 1,2, \dots, n)$ será, entonces, de la forma:

$$F_{X_i}(x) = \begin{cases} 0 & para \qquad x \leq \dfrac{1}{n} \\ \dfrac{1}{n} & \dfrac{1}{n} < x \leq \dfrac{2}{n} \\ & \dots\dots\dots\dots\dots\dots. \\ \dfrac{k}{n} & \dfrac{k}{n} < x \leq \dfrac{k+1}{n} \\ & \dots\dots\dots\dots\dots\dots \\ \dfrac{n-1}{n} & \dfrac{n-1}{n} < x \leq 1 \\ 1 & x > 1 \end{cases}$$

[117] Un resultado al cual se hubiera podido arribar en forma totalmente intuitiva, teniendo en cuenta que la distribución de probabilidades de la variable X_n está concentrada en el intervalo $\left(-\frac{1}{n}, \frac{1}{n}\right)$ el cual, al aumentar n indefinidamente, se reduce al origen.

Luego, para $x = \frac{k+1}{n}$, $\left|F_{X_i}(x) - x\right| = x - F_{X_i}(x) = \frac{k+1}{n} - \frac{k}{n} = \frac{1}{n}$. Es decir, para todo x en el intervalo [0,1], se verifica que $\lim_{n\to\infty} F_{X_n}(x) = x$ y, por lo tanto, que la función de distribución de probabilidades $F_{X_n}(x)$ tiende a una distribución uniforme en el intervalo [0,1] (la convergencia es puntual para todo punto x).

Obsérvese que las variables X_n son discretas y todas con distribución de probabilidades uniforme sobre n puntos equiespaciados del intervalo [0,1]. Luego, al tender n a infinito, la distribución de los valores de X_n tiende necesariamente a la distribución uniforme.

Ejemplo n° 35:

Sea una variable aleatoria X_n con distribución de probabilidades de la forma $p_{X_n}(x) = p_n q_n^{x-1} (x = 1,2, \dots)$ y sea la variable $Z_n = \frac{X_n}{n}$, cuya función de distribución será, en consecuencia, de la forma:

$$F_{Z_n(z)} = \sum_{j=1}^{nz} p_n q_n^{j-1} = p_n \sum_{j=0}^{nz-1} q_n^j = p_n \frac{1 - q_n^{nz}}{1 - q_n} = 1 - q^{nz} = 1 - (q_n^n)^z$$

(si bien, en general, el valor nz no es entero, aquí por razones de simplicidad se ha considerado solamente su parte entera)[118]. De esta expresión se puede concluir que el comportamiento en el límite cuando $n \to \infty$ de esta función de distribución depende del comportamiento de la sucesión $\{q_n^n\}$: si ésta no tiene límite, entonces $F_{X_n}(x)$ tampoco lo tiene. Si se verifica que $\lim_{n\to\infty}\{q_n^n\} = q$ (donde, dado que $0 \leq q_n \leq 1$, será $0 \leq q \leq 1$), entonces: **i)** si $q = 1$, será $\lim_{n\to\infty} F_{Z_n}(z) = 0$, es decir no existe límite en-distribución (Z_n tiende a ser infinitamente grande); **ii)** si $q = 0$ entonces, será $\lim_{n\to\infty} F_{Z_n}(z) = 1$, es decir Z_n tiende en-distribución a la variable aleatoria constantemente nula; **iii)** si $0 < q < 1$, entonces será $\lim_{n\to\infty} F_{Z_n}(z) = 1 - q^z$ y, haciendo en esta expresión $\lambda = -ln(q)$, se obtiene la distribución exponencial $1 - e^{\lambda z}$.

[118] Sea j_{nz} el máximo entero inferior a nz, $nz - 1 \leq j_{nz} < nz$. Será, entonces $F_{Z_n}(z) = \sum_{j\leq nz} p_n q_n^{j-1} = p_n \sum_{j=1}^{j_{nz}} q_n^{j-1} = p_n \frac{1-q^{j_{nz}}}{1-q_n} = 1 - q_n^{j_{nz}}$. La condición para que $F_{Z_n}(z)$ tienda a una función no-degenerada es que $q_n \to 1$. Si esta condición se cumple, se verificará que $\frac{q_n^{nz}}{q_n} \leq q_n^{j_{nz}} \leq q_n^{nz}$, de modo que $\lim_{n\to\infty} q_n^{j_{nz}} = \lim_{n\to\infty} q_n^{nz}$. De lo que se concluye que $\lim_{n\to\infty} F_{Z_n}(z) = \lim_{n\to\infty}\left(1 - q_n^{j_{nz}}\right) = \lim_{n\to\infty}(1 - q_n^{nz})$.

Por otra parte, considerando el valor esperado de la variable Z_n, se obtiene que:

$$E(Z_n) = \frac{1}{n}E(X_n) = \frac{1}{n}\sum_{j=1}^{\infty} j p_n q_n^{j-1} = \frac{1}{n}\sum_{j=1}^{\infty}\sum_{i=1}^{j} p_n q_n^{j-1} = \frac{1}{n}\sum_{i=1}^{\infty}\sum_{j=i}^{\infty} p_n q_n^{j-1} =$$

$$= \frac{1}{n}\sum_{j=1}^{\infty} p_n q_n^{j-1} \sum_{i=j}^{\infty} q_n^{j-i} = \frac{1}{n}\frac{1}{1-q_n}\sum_{j=1}^{\infty} p_n q_n^{j-1} = \frac{1}{n}\sum_{j=1}^{\infty} q_n^{j-1} =$$

$$= \frac{1}{n}\frac{1}{1-q_n} = \frac{1}{np_n}$$

Con lo que queda demostrado que la condición necesaria y suficiente para que este valor esperado tenga un límite finito es que $\lim_{n\to\infty} np_n = \lambda > 0$ y esta es condición necesaria y suficiente para que $\{q_n^n\}$ tienda a un límite $q = e^{-\lambda} < 1$. De todo esto se puede concluir que, dada una variable aleatoria X_n con distribución de probabilidades de la forma:

$$p_{X_n}(x) = \frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right)^{x-1} \qquad (x = 1,2,\dots)$$

la variable $Z_n = \frac{X_n}{n}$ tiende en-distribución a una variable exponencial con parámetro λ.

La sucesión de distribuciones consideradas en este ejemplo y en el anterior pueden ser interpretadas, mediante una división en intervalos, como originadas por la distribución límite: Dada la variable X con distribución de probabilidades uniforme en el intervalo$[0,1]$, la variable $X_n = \frac{j}{n}\left(\frac{j-1}{n} \le X \le \frac{j}{n}, j = 1,2,\dots,n\right)$ está uniformemente distribuida sobre los n puntos equiespaciados del intervalo $[0,1]$. De la misma forma, dada una variable aleatoria Z con distribución de probabilidades exponencial, la variable $Z_n = \frac{j}{n}\left(\frac{j-1}{n} \le X \le \frac{j}{n}, j = 1,2,\dots\right)$ es tal que:

$$p\left(Z_n = \frac{j}{n}\right) = F_Z\left(\frac{j}{n}\right) - F_Z\left(\frac{j-1}{n}\right) = 1 - e^{-\lambda\frac{j}{n}} - 1 + e^{-\lambda\frac{j-1}{n}} =$$

$$= e^{-\lambda\frac{j-1}{n}}\left(1 - e^{-\frac{\lambda}{n}}\right) = p_n q_n^{j-1}$$

donde $p_n = 1 - e^{-\frac{\lambda}{n}}$. Retornando así a la distribución de probabilidades de la variable Z_n. En ambos casos la distribución de probabilidades dividida en intervalos tiende, cuando la longitud de los mismos tiende a cero, a la distribución de origen (no dividida en intervalos).

Para distribuciones de probabilidades infinitas, como alternativa al método analizado en los ejemplos anteriores, se puede subdividir la recta en un número finito de intervalos (el primero y el último, obviamente, ilimitados) y concentrar la probabilidad de cada intervalo en un solo punto, de modo de obtener una distribución discreta.

Sea una variable aleatoria X con función de distribución de probabilidades $F_X(x)$. Dado un cierto valor n, se determinan los valores $C_{n,1}, C_{n,2}, \dots, C_{n,n}$ $\left(C_{n,j} < C_{n,j-1}, C_{n,j-1} - C_{n,j} = \varepsilon_n, j = 1,2, \dots, n-1\right)$ que dividen a la recta en $n+1$ intervalos y se selecciona, para cada intervalo, un punto $x_{n,j}$ de la siguiente forma: $x_{n,0} \le C_{n,1}$, $C_{n,j} \le x_{n,j} \le C_{n,j+1} (j = 1,2, \dots, n-1)$, $x_{n,n} \ge C_{n,n}$. La variable X_n, dividida en intervalos, queda expresada de la siguiente forma:

$$X_n = \begin{cases} x_{n,0} & si \quad X < C_{n,1} \\ x_{n,j} & \quad C_{n,j} \le X < C_{n,j-1} \; (j = 1,2, \dots, n-1) \\ x_{n,n} & \quad X > C_{n,n} \end{cases}$$

y su función de distribución puede ser definida como:

$$p\left(X_n = x_{n,j}\right) = F\left(C_{n,j+1}\right) - F\left(C_{n,j}\right) \qquad (j = 0,1,2, \dots, n)$$

$$F_{X_n}(x) = p(X_n \le x) = \sum_{x_{n,j} \le x} \left[F\left(C_{n,j+1}\right) - F\left(C_{n,j}\right)\right]$$

con $F\left(C_{n,0}\right) = 0$, $F\left(C_{n,n+1}\right) = 1$ y la suma que se extiende hasta el máximo valor de j tal que $x_{n,j} \le x$, es decir hasta el entero j' tal que $x_{n,j'} \le x < x_{n,j'+1}$. Luego, se puede escribir:

$$F_{X_n}(x) = \left[F\left(C_{n,1}\right) - 0\right] + \left[F\left(C_{n,2}\right) - F\left(C_{n,1}\right)\right] + \cdots +$$

$$+\left[F\left(C_{n,j'+1}\right) - F\left(C_{n,j'}\right)\right] = F\left(C_{n,j'+1}\right)$$

De las definiciones de j' y $x_{n,j}$ y tomando la máxima longitud del intervalo ε_n, se obtiene que:

$$C_{n,j'+1} - \varepsilon_n \le C_{n,j'} \le x_{n,j'} \le x < x_{n,j'+1} \le C_{n,j'+2} \le C_{n,j'+1} + \varepsilon_n$$

De lo que se puede concluir que $x - \varepsilon_n < C_{n,j'+1} \le x + \varepsilon_n$ y, por lo tanto, que $F(x - \varepsilon_n) \le F_{X_n}(x) \le F(x + \varepsilon_n)$. Esta desigualdad demuestra que, para todo punto de continuidad de $F_X(\cdot)$, se verifica que $F_{X_n}(x) \to F(x)$.

12.2.2.- Algunos teoremas fundamentales

1.- Sea $\{X_1, X_2, \dots, X_n, \dots\}$ una sucesión de variables aleatorias con función de distribución de probabilidades $F_{X_n}(x)$ $(n = 1,2, \dots)$, sea una variable aleatoria X con función de distribución $F_X(x)$ tal que $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ y sea $g(x)$ una función continua y acotada ($|g(x)| \le H$). Fijado un $\varepsilon > 0$, es posible hallar un valor α tal que α y $-\alpha$ sean puntos de continuidad de $F_X(x)$ y que se verifique la relación:

$$p(X \le -\alpha) + p(X > \alpha) = F_X(-\alpha) + [1 - F_X(\alpha)] < \varepsilon$$

Se puede escribir, entonces:

$$\left| \int_{\Omega(X)} g(x) dF_{X_n}(x) - \int_{\Omega(X)} g(x) dF_X(x) \right| =$$

$$= \left| \int_{\infty}^{-\alpha} g(x) dF_{X_n}(x) + \int_{-\alpha}^{\alpha} g(x) dF_{X_n}(x) + \int_{\alpha}^{\infty} g(x) dF_{X_n}(x) - \right.$$

$$\left. - \int_{-\infty}^{-\alpha} g(x) dF_X(x) - \int_{-\alpha}^{\alpha} g(x) dF_X(x) - \int_{\alpha}^{\infty} g(x) dF_X(x) \right| \le$$

$$\le \left| \int_{-\alpha}^{\alpha} g(x) dF_{X_n}(x) - \int_{-\alpha}^{\alpha} g(x) dF_X(x) \right| +$$

$$+H\left\{F_{X_n}(-\alpha) + \left[1 - F_{X_n}(\alpha)\right] + F_X(-\alpha) + [1 - F_X(\alpha)]\right\}$$

Como α y $-\alpha$ son puntos de continuidad, se verificará que $\lim_{n\to\infty} F_{X_n}(-\alpha) = F_X(-\alpha)$ y que $\lim_{n\to\infty} F_{X_n}(\alpha) = F_X(\alpha)$. De modo que cuando n aumenta indefinidamente en la expresión anterior, se obtiene que:

$$\lim_{n\to\infty}\left|\int_{-\infty}^{\infty} g(x)dF_{X_n}(x) - \int_{-\infty}^{\infty} g(x)dF_X(x)\right| \leq$$

$$\leq 2H[F_X(-\alpha) + 1 - F_X(\alpha)] < 2H\varepsilon$$

Con lo que se demuestra que la convergencia de la integral $\int_{\Omega(X)} g(x)dF_{X_n}(x) \to \int_{\Omega(X)} g(x)dF_X(x)$, es decir, la convergencia del valor esperado, $E[g(X_n)] \to E\big(g(X)\big)$ (donde, como se mencionó más arriba, $g(X)$ es una función continua y acotada) es condición necesaria y suficiente para poder asegurar que $X_n \xrightarrow{d} X$.

2.- Sea $\{X_1, X_2, \dots, X_n, \dots\}$ una sucesión de variables aleatorias que convergen en-distribución a la variable X y sean $h(X)$ una función continua y $g(Y)$ una función continua y acotada. Entonces la función $Z(X) = g[h(X)]$ también será continua y acotada. Por lo tanto, de acuerdo a lo demostrado en el teorema anterior:

$$E[Z(X_n)] = E\{g[h(X_n)]\} \to E[Z(X)] = E\{g[h(X)]\}$$

Luego, se puede asegurar que, si $X_n \xrightarrow{d} X$, entonces la variable $Y_n = h(X_n) \xrightarrow{d} Y = h(X)$.

3.- Sea $\{Z_n\} = \{X_n, Y_n\}$ una sucesión de variables aleatorias bidimensionales que converge en-distribución a la variable (X, Y), es decir, tal que $\lim_{n\to\infty} F_{X_n,Y_n}(x,y) = F_{X,Y}(x,y)$. A partir de la definición de la función de distribución de probabilidades, resulta que:

$$p(X_n \leq x) = 1 - p(X_n > x) \leq p[(X_n > x) \cap (Y_n \leq y)] =$$

$$= 1 - p(Y_n \leq y_m) + p[(X_n > x) \cap (Y_n \leq y_m)]$$

(donde (x, y_m) es un punto de continuidad para la función de probabilidades de la variable (X, Y). Ahora bien, teniendo en cuenta que:

$$\lim_{n\to\infty} infer\, p(X_n \leq x) \geq p(X \leq x)$$

$$\underset{n\to\infty}{lim}\, super\, p(X_n \leq x) = -\underset{n\to\infty}{lim}\, infer\, p(X_n \leq x) \leq p(X \leq x)$$

Se puede escribir:

$$\underset{n\to\infty}{lim}\, super\, p(X_n \leq x) \leq$$

$$\leq 1 - \underset{n\to\infty}{lim}\, infer\, p(Y_n \leq y_m) + \underset{n\to\infty}{lim}\, p[(X_n \leq x) \cap (Y_n \leq y_m)] \leq$$

$$\leq 1 - p(Y \leq y_m) + p[(X_n \leq x) \cap (Y_n \leq y_m)]$$

$$= 1 - p[(X_n > x) \cap (Y_n \leq y)]$$

Cuando en esta expresión m aumenta indefinidamente, se obtiene que:

$$\underset{n\to\infty}{lim}\, super\, p(X_n \leq x) \leq 1 - p(X > x) = p(X \leq x)$$

De modo que:

$$\lim_{n\to\infty} p(X_n \leq x) = \lim_{n\to\infty} F_{X_n}(x) = p(X \leq x) = F_X(x)$$

De la misma forma, se demuestra que:

$$\lim_{n\to\infty} p(Y_n \leq y) = \lim_{n\to\infty} F_{Y_n}(y) = p(Y \leq y) = F_Y(y)$$

Luego, se puede concluir que, si $(X_n, Y_n) \xrightarrow{d} (X, Y)$, entonces las variables marginales convergen respectivamente a X e Y, $X_n \xrightarrow{d} X$ e $Y_n \xrightarrow{d} Y$. Debe tenerse en cuenta que la implicación inversa no siempre se verifica[119]. La convergencia en-distribución de las variables marginales determinará la convergencia de la variable conjunta sólo si aquellas son independientes: Sean x e y puntos de continuidad de las variables X e Y, respectivamente. Se verificará, entonces, que:

$$\lim_{n\to\infty} p[(X_n \leq x) \cap (Y_n \leq y)] = \lim_{n\to\infty} F_{X_n,Y_n}(x, y) =$$

$$\lim_{n\to\infty} p(X_n \leq x)p(Y_n \leq y) = p(X \leq x)p(Y \leq y) = F_X(x)F_Y(y)$$

[119] Para cada par de funciones $F_X(x)$ y $F_Y(y)$ existe más de una función de distribución conjunta que las admite como funciones marginales (ver Sec. 2.1). Como se ve, la convergencia en-distribución no conserva las propiedades de la convergencia en el sentido del análisis según la cual la convergencia de las componentes de un vector implica la convergencia del vector.

Como corolario de este resultado se puede afirmar que, dada la sucesión de variables aleatorias $\{X_n, Y_n\}$ en la cual las variables marginales son independientes y tales que $X_n \xrightarrow{d} X \left(\lim_{n\to\infty} F_{X_n}(x) = F_X(x)\right)$ y $Y_n \xrightarrow{d} Y \left(\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y)\right)$, entonces $\{X_n + Y_n\} \xrightarrow{d} (X + Y)$. Es decir, la convolución de las funciones de distribución de las variables marginales converge a la convolución $F_X(x) * F_Y(y)$, cuando dichas variables marginales son independientes, $\lim_{n\to\infty} F_{X_n}(x) * F_{Y_n}(y) = F_X(x) * F_Y(y)$.

También como corolario de las demostraciones anteriores se puede demostrar que, dadas una sucesión de variables aleatorias $\{X_n\}$ y una variable aleatoria X tales que las variables X_n sean independientes de X y, además se cumpla la condición $X_n \xrightarrow{d} X$, si X' es una variable aleatoria semejante a X e independiente de X, entonces –dado que $\{X_n\}$, al converger en-distribución a X, converge también a X' (que tiene la misma distribución de probabilidades) y que, trivialmente, converge en-distribución a X'- se verifica que $(X_n, X) \xrightarrow{d} (X', X)$.

Ejemplo n° 36:

Sea $X_n (n > 0)$ una variable aleatoria con distribución de probabilidades de la forma:

$$f_{X_n}(x) = \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} \qquad (x > 0)$$

cuyos momentos de primero y segundo orden son $E(X_n) = \frac{n}{\lambda}$ y $\sigma^2(X_n) = \frac{n}{\lambda^2}$. La correspondiente variable estandarizada será, por lo tanto, de la forma:

$$U_n = \frac{X_n - E(X_n)}{\sigma(X_n)} = \frac{\lambda X_n - n}{\sqrt{\lambda}}$$

Entonces se verificará que:

$$f_{U_n}(u) = f_{X_n}(x) \frac{\sqrt{n}}{\lambda} = f_{X_n}\left(\frac{u\sqrt{n} + n}{\lambda}\right) =$$

$$= \frac{\sqrt{n}}{\lambda} \left\{ \frac{\lambda^n \left(\frac{u\sqrt{n}+n}{\lambda}\right)^{n-1}}{(n-1)!} exp\left[-\lambda\left(\frac{u\sqrt{n} + n}{\lambda}\right)\right] \right\} =$$

$$= \frac{\lambda^n}{(n-1)!}\left(\frac{u\sqrt{n}+n}{\lambda}\right)^{n-1} e^{-(u\sqrt{n}+n)}\frac{\sqrt{n}}{\lambda} = \frac{n^n\sqrt{n}}{n!}e^{-(\sqrt{n}+n)}\left(1+\frac{u}{\sqrt{n}}\right)^{n-1}$$

De modo que, de acuerdo con la fórmula de de Moivre-Stirling, según la cual $\lim_{n\to\infty}(n!) = e^{-n}n^n\sqrt{2\pi n}$, será:

$$\lim_{n\to\infty} f_{U_n}(u) = \lim_{n\to\infty}\frac{e^{-n}n^n\sqrt{n}}{n!}e^{-u\sqrt{n}}\left(1+\frac{u}{\sqrt{n}}\right)^{n-1} = \lim_{n\to\infty}\frac{e^{-u\sqrt{n}}}{\sqrt{2\pi}}\left(1+\frac{u}{\sqrt{n}}\right)^{n}$$

Ahora bien, teniendo en cuenta que:

$$\lim_{n\to\infty} ln\left[e^{-u\sqrt{n}}\left(1+\frac{u}{\sqrt{n}}\right)^{n}\right] =$$

$$= \lim_{n\to\infty}\left[-u\sqrt{n} + n\, ln\left(1+\frac{u}{\sqrt{n}}\right)\right] = \lim_{n\to\infty}\frac{ln\left(1+\frac{u}{\sqrt{n}}\right)-\frac{u}{\sqrt{n}}}{n^{-1}} =$$

$$= \lim_{n\to\infty}\frac{\frac{\sqrt{n}}{\sqrt{n}+u}\left(-\frac{u}{2n\sqrt{n}}\right)+\frac{u}{2n\sqrt{n}}}{n^{-1}} = \lim_{n\to\infty}\left(\frac{u^2\sqrt{n}}{u+\sqrt{n}}\right) = -\frac{u^2}{2}$$

queda demostrado que la variable U_n converge en-distribución a una variable con función de densidad de la forma:

$$\lim_{n\to\infty} f_{U_n}(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$$

12.3.- La convergencia en-probabilidad

12.3.1.- Definición

Sea una sucesión de variables aleatorias $\{X_n, n = 1,2,\dots\}$ y sea una variable aleatoria X, todas con valores en $\mathbb{R}_k (k \geq 1)$. Se dice que $\{X_n\}$ converge en-probabilidad a la variable X (o que existe una igualdad asintótica entre X_n y X), $X_n \xrightarrow{p} X$, si se verifica que $\lim_{n\to\infty} p(|X_n - X| \leq \varepsilon) = 1\,(\varepsilon > 0)$. Esta convergencia también suele ser expresada de la siguiente forma: $\operatorname{plim}_{n\to\infty}(X_n) = X$ [120].

[120] Teniendo en cuenta que X puede ser una variable degenerada, resulta admisible que la sucesión $\{X_n\}$ converja en-probabilidad a un número (no-aleatorio).

12.3.2.- Algunos teoremas fundamentales

1) A partir de esta definición se puede concluir en forma inmediata que, si la sucesión $\{X_n\}$ converge en-probabilidad a X, entonces, para valores de n suficientemente grandes, se verificará que $p(|X_n - X| \leq \varepsilon) \geq 1 - \eta \ (\varepsilon, \eta > 0)$. Luego, de acuerdo con lo demostrado en la Sec. 1.2, será:

$$\left|F_X(x) - F_{X_n}(x)\right| \leq F_X(x + \varepsilon) - F_X(x - \varepsilon) + \eta$$

de modo que si x es un punto de continuidad de $F_X(\cdot)$, seleccionando un ε arbitrariamente pequeño, el segundo miembro de esta inecuación se puede hacer tan pequeño como se desee. Es decir, se puede asegurar que $\lim\limits_{n\to\infty} F_{X_n}(x) = F_X(x)$. De este resultado se puede concluir, entonces, que la convergencia en-probabilidad implica la convergencia en-distribución, pero que la implicación inversa no se verifica: la convergencia en-distribución es más débil que la convergencia en-probabilidad. Es decir, que se verifique que $X_n \xrightarrow{d} X$ no implica que $X_n \xrightarrow{p} X$.

Ejemplo n° 37:

Sea una variable aleatoria X con distribución de probabilidades de la forma:

$$p(X = x) = \begin{cases} \dfrac{1}{2} & para \qquad x = 0 \\ \dfrac{1}{2} & \qquad\qquad x = 1 \\ 0 & para\ cualquier\ otro\ valor\ de\ x \end{cases}$$

y sea una variable aleatoria $Y = 1 - X$ (obviamente las variables X e Y tienen la misma distribución de probabilidades).

Sea, ahora, una sucesión de variables aleatorias $\{X_n = Y, n = 1,2, \dots\}$. Dado que a todas las X_n y a X les corresponde la misma función de distribución de probabilidades, $F_X(\cdot)$, se concluye en forma inmediata que $X_n \xrightarrow{d} X$. Sin embargo, como $|X_n - X| = |Y - X| = 1$, se puede asegurar que $\{X_n\}$ no converge en-probabilidad a X. Lo que demuestra que la convergencia en-distribución no implica necesariamente la convergencia en-probabilidad.

---

2) Supóngase que la sucesión $\{X_n\}$ y la variable aleatoria X sean tales que $X_n \xrightarrow{ccc} X$. Entonces, de acuerdo con lo expresado en la Sec. 12.1.2, se verificará que:

$$\lim_{n\to\infty} p\left[\bigcap_{j=n}^{\infty}\left(\left|X_j - X\right| \le \varepsilon\right)\right] = 1$$

Luego, teniendo en cuenta que $\left[\bigcap_{j=n}^{\infty}\left(\left|X_j - X\right| \le \varepsilon\right)\right] \subseteq |X_n - X| \le \varepsilon$, será:

$$p\left[\bigcap_{j=n}^{\infty}\left(\left|X_j - X\right| \le \varepsilon\right)\right] \le p(|X_n - X| \le \varepsilon) \le 1$$

Con lo que queda demostrado que $\lim_{n\to\infty} p(|X_n - X| \le \varepsilon) = 1$. Es decir, que $X_n \xrightarrow{p} X$. Luego, se puede concluir que la convergencia casi-con-certeza implica la convergencia en-probabilidad, pero que la implicación inversa no se verifica siempre. La implicación en ambos sentidos se verifica sólo en el caso particular en que $\{X_n\}$ sea una sucesión decreciente de variables aleatorias positivas.

3) Sea $\{X_n\}$ una sucesión de variables aleatorias con funciones de distribución tales que $\lim_{n\to\infty} F_{X_n}(x) = \varepsilon(x - c)$. Es decir, tales que, para valores de n suficientemente grandes, se verifica que $F_{X_n}(c - \varepsilon) \le \frac{\eta}{2}$ y $F_{X_n}(c + \varepsilon) \ge 1 - \frac{\eta}{2}\,(\varepsilon, \eta > 0)$. Entonces, será:

$$p(X_n \ge c - \varepsilon) \ge p(X_n > c - \varepsilon) \ge 1 - \frac{\eta}{2}$$

$$p(X_n \ge c + \varepsilon) \ge 1 - \frac{\eta}{2}$$

$$p(|X_n - X| \ge \varepsilon) \ge 1 - \eta$$

Lo que demuestra que $\lim_{n\to\infty} p(|X_n - X| \le \varepsilon) = 1$. O, lo que es lo mismo, que $X_n \xrightarrow{p} c$. Luego se puede concluir que la condición necesaria y suficiente para que la sucesión $\{X_n\}$ converja en-probabilidad a una constante c es que la correspondiente sucesión de funciones de distribución de probabilidades converja a una distribución degenerada $\varepsilon(x - c)$.

4) Sea una sucesión de variables aleatorias $\{X_n\}$ y sean las variables aleatorias X e Y tales que $X_n \overset{p}{\rightarrow} X$ y $Y_n \overset{p}{\rightarrow} Y$. Considerando los eventos $(|X-Y| \leq \varepsilon)$, $\left(|X-X_n| \leq \frac{\varepsilon}{2}\right)$ y $\left(|Y-X_n| \leq \frac{\varepsilon}{2}\right)$ se pueden escribir las siguientes implicaciones:

$$(|X-Y| \leq \varepsilon) \supset \left(|X-X_n| \leq \frac{\varepsilon}{2}\right) \cap \left(|Y-X_n| \leq \frac{\varepsilon}{2}\right) (\varepsilon > 0)$$

$$(|X-Y| > \varepsilon) \subset \left(|X-X_n| > \frac{\varepsilon}{2}\right) \cup \left(|Y-X_n| > \frac{\varepsilon}{2}\right) (\varepsilon > 0)$$

de las que resulta que:

$$p(|X-Y| > \varepsilon) \leq p\left(|X-X_n| > \frac{\varepsilon}{2}\right) + p\left(|Y-X_n| > \frac{\varepsilon}{2}\right) (\varepsilon > 0)$$

Ahora bien, a partir de las convergencias propuestas inicialmente, el segundo miembro de esta expresión se puede hacer tan pequeño como se desee, de modo que $p(|X-Y| > \varepsilon) < \eta\ (\varepsilon, \eta > 0)$. Lo que demuestra que si la sucesión $\{X_n\}$ converge en-probabilidad simultáneamente a X y a Y, entonces X e Y son variables iguales. De la misma forma se demuestra que, dadas dos variables aleatorias iguales y una sucesión de variables aleatorias $\{X_n\}$, si se verifica que $X_n \overset{p}{\rightarrow} X$, entonces se verifica también que $X_n \overset{p}{\rightarrow} Y$.

Dado que, según se demostró, la convergencia casi-con-certeza implica la convergencia en-probabilidad, se obtiene como corolario del resultado anterior que, dada una sucesión de variables aleatorias $\{X_n\}$, si se verifica que $X_n \overset{ccc}{\longrightarrow} X$ y $X_n \overset{ccc}{\longrightarrow} Y$, entonces X e Y son iguales.

5) Sean una sucesión de variables aleatorias $\{X_n\}$ y una variable aleatoria X tales que $X_n \overset{p}{\rightarrow} X$. Sea $g(\cdot)$ una función continua y sean $F_{|X_n|}(x)$ y $F_{|X|}(x)$ las funciones de distribución de probabilidades de las variables $|X_n|$ y $|X|$, respectivamente. Dado que $|X_n - X| \geq |X_n| - |X|$, se verifica que $|X_n| \overset{p}{\rightarrow} |X|$ y, por lo tanto, que $\lim\limits_{n\to\infty} F_{|X_n|}(x) = F_{|X|}(x)$.

Sea x^* un punto de continuidad de $F_{|X|}(x)$ tal que $F_{|X|}(x^*) > 1 - \frac{\varepsilon}{2} (\varepsilon > 0)$, es posible hallar un valor de n suficientemente grande $(n \geq N)$ tal que $F_{|X_n|}(x^*) - F_{|X|}(x^*) > -\frac{\varepsilon}{2}$ y $F_{|X_n|}(x^*) > F_{|X|}(x^*) - \frac{\varepsilon}{2} > 1 - \varepsilon$ y sea $x' = max(x_0, x_1, x_2, \dots, x_N)$. Se obtiene, entonces que, para todo n, $p(|X_n| \leq x') = F_{|X|}(x') > 1 - \varepsilon$ y de acuerdo con lo expresado más arriba, que $p(|X| \leq x') = F_{|X|}(x') > 1 - \varepsilon$. Ahora bien, como la función $g(y)$ es uniformemente continua en el intervalo $|y| \leq x'$, es posible hallar un $\delta = \delta(\eta)(\eta > 0)$ tal que, si se verifican las condiciones $|y_1| \leq x', |y_2| \leq x'$, $|y_1 - y_2| \leq \eta$, se verifica, también, que $|g(y_1) - g(y_2)| \leq \eta(\eta > 0)$. Es decir, que:

$$(|X_n| \leq x') \cap (|X| \leq x') \cap (|X_n - X| \leq \delta) \supset [|g(X) - g(X_n)| \leq \eta]$$

A partir de esta implicación y de las desigualdades anteriores se puede escribir:

$$p(|g(X) - g(X_n)| \leq \eta) > 1 - 3\varepsilon \qquad (\varepsilon, \eta > 0)$$

Con lo que queda demostrado que, si la sucesión $\{X_n\}$ converge en-probabilidad a la variable X, entonces, dada una función continua $g(\cdot)$, se verificará que $g(X_n) \overset{p}{\rightarrow} g(X)$.

Como corolario de este teorema se obtiene que, dadas dos sucesiones de variables aleatorias $\{X_n\}$ e $\{Y_n\}$ tales que $X_n \overset{p}{\rightarrow} X$ e $Y_n \overset{p}{\rightarrow} Y$ y una función continua $g(X, Y)$, se verifica que $g(X_n, y_n) \overset{p}{\rightarrow} g(X, Y)$.

6) Sean una sucesión de variables aleatorias $\{X_n\}$ y una variable aleatoria X con funciones de distribución de probabilidades $F_{X_n}(x)$ y $F_X(x)$, respectivamente y tales que $X_n \overset{p}{\rightarrow} X$ y sea $g(\cdot)$ una función continua no-negativa. Supóngase que $E[g(X_n)] < \infty$ $(n = 1,2, \dots)$ y que $E[g(X)] < \infty$ y que las integrales $\int_{x_1}^{x_2} g(x) dF_X(x)$ $(n = 1,2, \dots)$ (donde x_1 y x_2 son puntos de continuidad de $F_X(x)$) sean uniformemente convergentes, es decir, que existan dos constantes x_1^* y x_2^* independientes de n, tales que:

$$\left| \int_{x_1}^{x_2} g(x) dF_{X_n}(x) - \int_{-\infty}^{\infty} g(x) dF_{X_n}(x) \right| < \varepsilon \qquad (x_1 < x_1^*, x_2 < x_2^*, \varepsilon > 0)$$

Dado que se ha supuesto que $E[g(X)] < \infty$, los valores de x_1^* y x_2^* pueden ser elegidos de modo que también se verifique la relación:

$$\left| \int_{-\infty}^{\infty} g(x) dF_{X_n}(x) - \int_{x_1}^{x_2} g(x) dF_X(x) \right| < \varepsilon$$

Por otra parte, de acuerdo con el teorema de Helly[121], para valores de n suficientemente grandes ($n > N$), se puede escribir:

$$\left| \int_{x_1}^{x_2} g(x) dF_X(x) - \int_{x_1}^{x_2} g(x) dF_{X_n}(x) \right| < \varepsilon$$

De las tres ecuaciones anteriores se obtiene que:

$$\left| \int_{x_1}^{x_2} g(x) dF_{X_n}(x) - \int_{-\infty}^{\infty} g(x) dF_{X_n}(x) \right| < \varepsilon$$

$$\left| \int_{-\infty}^{\infty} g(x) dF_X(x) - \int_{-\infty}^{\infty} g(x) dF_{X_n}(x) \right| < 3\varepsilon \qquad (n > N)$$

Es decir, que $\lim_{n\to\infty} E[g(X_n)] = E[g(X)]$. Lo que demuestra que la convergencia en-probabilidad de $\{X_n\}$ a la variable X es una condición suficiente para la convergencia de las esperanzas matemáticas de dichas variables.

Supóngase ahora que se verifique la convergencia en los valores esperados. Entonces, se puede escribir:

$$\left| \int_{-\infty}^{\infty} g(x) dF_X(x) - \int_{x_1}^{x_2} g(x) dF_{X_n}(x) \right| \leq$$

$$\leq \left| \int_{-\infty}^{\infty} g(x) dF_{X_n}(x) - \int_{-\infty}^{\infty} g(x) dF_X(x) \right| +$$

$$+ \left| \int_{-\infty}^{\infty} g(x) dF_X(x) - \int_{x_1}^{x_2} g(x) dF_X(x) \right| +$$

$$+ \left| \int_{x_1}^{x_2} g(x) dF_X(x) - \int_{x_1}^{x_2} g(x) dF_{X_n}(x) \right|$$

[121] Ver Luckacs (1975).

donde x_1 y x_2 pueden ser seleccionados de forma tal que el segundo término del segundo miembro resulte tan pequeño como se desee. Asimismo, para valores de n suficientemente grandes, los otros términos del segundo miembro se harán tan pequeños como se desee. Es decir, es posible hallar un valor N tal que para todo $n > N$, se verifique que:

$$\left| \int_{-\infty}^{\infty} g(x) dF_{X_n}(x) - \int_{x_1}^{x_2} g(x) dF_{X_n}(x) \right| \leq \varepsilon$$

Como se ha supuesto que $E[g(X_n)] < \infty$, es posible hallar dos números, $x^{(j)}$ y $x^{*(j)}(j = 1,2, \dots, N)$ tales que:

$$\left| \int_{-\infty}^{\infty} g(x) dF_{X_n}(x) - \int_{x^{(j)}}^{x^{*(j)}} g(x) dF_{X_n}(x) \right| \leq \varepsilon$$

De las desigualdades anteriores se puede concluir que:

$$\left| \int_{-\infty}^{\infty} g(x) dF_{X_n}(x) - \int_{x_1}^{x_2} g(x) dF_{X_n}(x) \right| \leq \varepsilon$$

(donde $x_1 < min\left(x^{(1)}, x^{(2)}, \dots, x^{(N)}\right)$ y $x_2 > max\left(x^{*(1)}, x^{*(2)}, \dots, x^{*(N)}\right)$). Lo que demuestra que $\int_{x_1}^{x_2} g(x) dF_{X_n}(x)$ converge uniformemente a $\int_{-\infty}^{\infty} g(x) dF_X(x)$. Esa decir que la convergencia $X_n \overset{p}{\rightarrow} X$ no sólo es condición suficiente sino también necesaria para la convergencia de las esperanzas matemáticas.

12.4.- La convergencia en-los-momentos

12.4.1.- Definición

Sean una sucesión de variables aleatorias $\{X_n, n = 1,2, \dots\}$ y una variable aleatoria X, todas con valores en $\mathbb{R}_k (k \geq 1)$ y tales que $E(|X_n|^s) < \infty (n = 1,2, \dots, s \geq 1)$ y $E(|X|^s) < \infty (s \geq 1)$. Se dice que $\{X_n\}$ converge en el momento natural de orden s a la variable X, $X_n \overset{m_s}{\rightarrow} X$, si se verifica que $\lim\limits_{n \to \infty} E(|X_n - X|^s) = 0$.

En particular, si se verifica que $E(|X_n|^2) < \infty (n = 1,2,\dots)$, $E(|X|^2) < \infty$ y que $\lim_{n\to\infty} E(|X_n - X|^2) = 0$, se dice que $\{X_n\}$ converge a X en-valor medio cuadrático.

12.4.2.- Algunos teoremas fundamentales

1) Supóngase que la sucesión $\{X_n, n = 1,2,\dots\}$ y la variable X sean tales que $X_n \xrightarrow{m_s} X$. Entonces se verificará que[122]:

$$0 \leq p(|X_n - X| \geq \varepsilon) \leq \frac{E(|X_n - X|^s)}{\varepsilon^s} \qquad (\varepsilon > 0, s = 1,2,\dots)$$

Luego, será $\lim_{n\to\infty} p(|X_n - X| \geq \varepsilon) = 0$. Es decir, se demuestra que la convergencia en-los-momentos implica la convergencia en-probabilidad. Pero la implicación inversa no se verifica necesariamente. La implicación en ambos sentidos se verifica sólo en el caso particular en que $\{X_n\}$ sea una sucesión de variables aleatorias uniformemente acotadas.

2) A partir de la propiedad precedente y del teorema 4 de la Sec. 12.3.2 según los cuales, si se verifica que $X_n \xrightarrow{ccc} X$ y $X_n \xrightarrow{ccc} Y$, entonces X e Y son iguales, se puede concluir que, si se verifica que $X_n \xrightarrow{m_s} X$ y que $X_n \xrightarrow{m_s} Y$, entonces X e Y son iguales.

3) Supóngase que la sucesión $\{X_n, n = 1,2,\dots\}$ y la variable X sean tales que $X_n \xrightarrow{m_{s_2}} X$. Teniendo en cuenta que, para una variable aleatoria Y tal que $m_{s_2}(Y) < \infty$, la función $E(|Y|^{s_2})$ es no-decreciente con respecto a s_2, para todo $s_1 < s_2$, se puede escribir $E(|X_n - X|^{s_1}) \leq [E(|X_n - X|^{s_2})]^{s_1/s_2}$ y, dado que $\lim_{n\to\infty} E(|X_n - X|^{s_2}) = 0$, se verificará que $X_n \xrightarrow{m_{s_1}} X$. Es decir, se demuestra que, para $s_1 < s_2$, la convergencia en el momento de orden s_2 implica la convergencia en el momento de orden s_1. Ahora bien, dado que la existencia del momento de orden s_1 no implica la existencia del momento de orden s_2, no se puede asegurar el cumplimiento de la implicación inversa. En otros términos, que $X_n \xrightarrow{m_{s_1}} X$ no implica que $X_n \xrightarrow{m_{s_2}} X$.

Supóngase que $X_n \xrightarrow{d} X$ y que exista un $s \geq 1$ tal que $E(|X_n|^s) < a$. Entonces, para todo $b > 0$, se puede escribir:

---

[122] Ver Sec. 10.4.2.

$$\int_{|x|\geq b} x^j dF_{X_n}(x) = \int_{|x|\geq b} \frac{|x|^s}{|x|^{s-j}} dF_{X_n}(x) \leq$$

$$\leq \frac{1}{b^{s-j}} \int_{|x|\geq b} |x|^s dF_{X_n}(x) \leq \frac{1}{b^{s-j}} \int_{\Omega(X_n)} |x|^s dF_{X_n}(x) \leq \frac{a}{b^{s-j}}$$

De modo que, si $j < s$, entonces $\lim_{n\to\infty} \frac{a}{b^{s-j}} = 0$. Con lo que queda demostrada la convergencia uniforme, $\int_{|x|\geq b} x^j dF_{X_n}(x) \to 0$ y, por lo tanto, la convergencia, $\int_{-\infty}^{\infty} x^s dF_{X_n}(x) \to \int_{-\infty}^{\infty} x^s dF_X(x)$. Por lo que se puede concluir que la convergencia en-distribución implica la convergencia en los momentos. La implicación inversa se verifica sólo si la sucesión de momentos $\{m_s(X), s = 1,2,\dots\}$ identifica en forma unívoca a la distribución de probabilidades de la variable X. Es decir, dadas una sucesión de variables aleatorias $\{X_n, n = 1,2,\dots\}$ y una variable aleatoria X tal que su distribución de probabilidades resulta caracterizada en forma unívoca por la sucesión de momentos $\{m_s(X), s = 1,2,\dots\}$ y tal que $X_n \xrightarrow{m_s} X$, entonces se puede asegurar que $X_n \xrightarrow{d} X$.

4) Sean una sucesión de variables aleatorias $\{X_n, n = 1,2,\dots\}$ y una variable aleatoria X tales que $X_n \xrightarrow{m_2} X$. Para valores de n suficientemente grandes se puede escribir:

$$[E(X_n) - E(X)]^2 \leq \sigma^2(X_n - X) + [E(X_n) - E(X)]^2 =$$

$$= E[(X_n - X)^2] \leq \varepsilon \qquad (\varepsilon > 0)$$

Luego, es posible hallar un valor de n suficientemente grande como para que la diferencia $|E(X_n) - E(X)|$ sea arbitrariamente pequeña, es decir, para que se verifique que $\lim_{n\to\infty} E(X_n) = E(X)$. De la misma forma se puede escribir:

$$E(X_n^2) = E[(X_n - X + X)^2] = E[(X_n - X)^2] + E(X^2) + 2E[X(X_n - X)]$$

De acuerdo con la desigualdad de Schwartz-Hölder[123], se verifica que $\{E[X(X_n - X)]\}^2 \leq E(X^2)\, E[(X_n - X)^2]$. De modo que:

---

[123] Ver Sec. 10.2.5.

$$\left\{\sqrt{E(X^2)} - \sqrt{E[(X_n - X)^2]}\right\}^2 \leq E(X_n^2) \leq \left\{\sqrt{E(X^2)} + \sqrt{E[(X_n - X)^2]}\right\}^2$$

y, teniendo en cuenta que, de acuerdo con la hipótesis inicial $\lim_{n\to\infty} E[(X_n - X)^2] = 0$, se obtiene que $\lim_{n\to\infty} E(X_n^2) = E(X^2)$.

De la misma forma, dadas una sucesión de variables aleatorias $\{X_n, n = 1,2,\ldots\}$ y una variable aleatoria X tales que $X_n \overset{m_s}{\rightarrow} X$, se puede asegurar que $\lim_{n\to\infty} E(|X_n|^r) = E(|X|^r)\ (r \leq s)$.

Como corolario del teorema anterior se demuestra que, dadas dos sucesiones de variables aleatorias $\{X_n\}$ y $\{Y_{n'}\}$ y dos variables aleatorias X e Y, tales que $X_n \overset{m_2}{\rightarrow} X$ y $Y_{n'} \overset{m_2}{\rightarrow} Y$, se verifica que $\lim_{\substack{n\to\infty \\ n'\to\infty}} E(X_n, Y_{n'}) = E(X, Y)$.

Ejemplo n° 38:

Sea X_n una variable aleatoria con función de densidad de la forma:

$$f_{X_n}(x) = \begin{cases} 1 - \dfrac{1}{n} & para \quad x = 0 \\ \dfrac{1}{n} & \quad x = n \end{cases}$$

Su función de distribución de probabilidades será, en consecuencia, de la forma:

$$F_{X_n}(x) = \begin{cases} 0 & para \quad x < 0 \\ 1 - \dfrac{1}{n} & \quad 0 \leq x \leq n \\ 1 & \quad x > n \end{cases}$$

Sea, por otra parte, una variable X que puede asumir solamente el valor 0 (con probabilidad igual a 1) y cuya función de distribución será, en consecuencia, de la forma:

$$F_X(x) = \begin{cases} 0 & para \quad x < 0 \\ 1 & \quad x \geq 0 \end{cases}$$

Se puede concluir inmediatamente que $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$. Por otra parte, se observa que $E(X_n) = 0\left(1 - \frac{1}{n}\right) + n\frac{1}{n}$ y $E(X) = 0$. Es decir, que $\lim_{n\to\infty} E(X_n) = 1 \neq 0 = E(X)$.

Asimismo, se verifica que $E(X_n^s) = 0\left(1-\frac{1}{n}\right) + n^s\frac{1}{n} = n^{s-1}$. Suponiendo que s pueda asumir cualquier valor real positivo, se tiene que:

$$\lim_{n\to\infty} E(X_n^s) = \begin{cases} E(X^s) & para \quad s < 1 \\ 1 \neq E(X^s) & \quad s = 1 \\ +\infty \neq E(X^s) & \quad s > 1 \end{cases}$$

Esto demuestra que, para $s = 1$, existe lo que se podría denominar un "valor crítico". Por debajo de ese valor los momentos convergen a $E(X^s)$. Para $s > 1$ los momentos tienden a infinito y, obviamente no convergen a $E(X^s)$.

Ejemplo n° 39:

Sea $\{Y_n\}$ una sucesión de variables aleatorias independientes, todas con distribución de probabilidades de la forma:

$$p(Y_n = y_n) = \begin{cases} \dfrac{1}{2n^{1/4}} & para \quad y_n = -1 \\ 1 - \dfrac{1}{n^{1/4}} & \quad y_n = 0 \\ \dfrac{1}{2n^{1/4}} & \quad y_n = 1 \\ 0 \quad para\ cualquier\ otro\ valor\ de\ Y_n & \end{cases}$$

(para $n = 1,2,\dots$).

Sea $E_n (n = 2,3,\dots)$ el evento "*que todas las Y_j, para $n - \sqrt{n} \leq j \leq n$, sean nulas*". De modo que:

$$p(E_n) = \prod_{n-\sqrt{n}\leq j\leq n} p(Y_j = 0) = \prod_{n-\sqrt{n}\leq j\leq n}\left(1 - \frac{1}{j^{1/4}}\right) \quad (n = 2,3,\dots)$$[124]

y sea una sucesión de variables aleatorias $\{X_n\}$ tales que $X_1 = Y_1$ y

$$X_n = \begin{cases} Y_n, si\ ocurre\ \overline{E_n} \\ 1\ ó -1, con\ igual\ probabilidad, si\ ocurre\ E_n \end{cases}$$

[124] Obsérvese que $\prod_{n-\sqrt{n}\leq j\leq n}\left(1-\frac{1}{j^{1/4}}\right) \leq \prod_{n-\sqrt{n}\leq j\leq n}\left(e^{-1/4}\right)^j$. De modo que se verifica que $\lim_{n\to\infty} p(E_n) = 0$.

(para $n = 2,3, ...$). Se puede concluir en forma inmediata que $E(X_n) = 0$. Por otra parte, se verifica que:

$$E(X_n^2) = p(E_n) + [1 - p(E_n)]n^{-1/4}$$

y, por lo tanto, que $\lim_{n\to\infty} E(X_n^2) = 0$. Es decir, que $X_n \xrightarrow{m_2} 0$. De la misma forma, se demuestra que $X_n \xrightarrow{m_s} 0$ $(s = 3,4, ...)$.

Supóngase ahora que $X_n \xrightarrow{ccc} 0$. Para $\eta < 1$, será:

$$p\left[\bigcap_{j=n}^{\infty}\left(|X_j| \le \eta\right)\right] = p\left[\bigcap_{j=n}^{\infty}(X_j = 0)\right] = p\left\{\bigcap_{j=n}^{\infty}\left[\overline{E_j} \cap (Y_j = 0)\right]\right\}$$

Ahora bien, dado que $\left(|X_j| \le \eta\right) \subset (Y_j = 0)(\eta < 1)$, será:

$$\bigcap_{j=[n^*-\sqrt{n^*}]}^{\infty}\left(|X_j| \le \eta\right) \subset \bigcap_{j=[n^*-\sqrt{n^*}]}^{\infty}(Y_j = 0) \subset E_n$$

(donde $[n^* - \sqrt{n^*}]$ denota el mayor entero menor o igual a $n^* - \sqrt{n^*}$) y, dado que $\left(|X_j| \le \eta\right) \subset \overline{E_j}(\eta < 1)$, será $\bigcap_{j=n}^{\infty}\left(|X_j| \le \eta\right) \subset \bigcap_{j=n}^{\infty}\overline{E_j} \subset \overline{E_n}$. Ahora bien, de acuerdo con la definición, si $\{X_n\}$ converge casi-con-certeza a 0, es posible, para todo $\varepsilon > 0$, hallar un valor $N = N(\varepsilon)$ tal que, para $n \ge N$, se verifique que:

$$p\left[\bigcap_{j=n}^{\infty}\left(|X_j| \le \eta\right)\right] > 1 - \varepsilon$$

De las implicaciones definidas más arriba se puede concluir que, para todo $n^* - \sqrt{n^*} \ge N$, será $p(E_{n^*}) > 1 - \varepsilon$ pero, también se puede concluir que $p(\overline{E_{n^*}}) > 1 - \varepsilon$. Esta contradicción indica que el supuesto de convergencia casi-con-certeza en este caso no se verifica. Lo cual permite comprobar que la convergencia en-los-momentos no implica la convergencia casi-con-certeza.

Dado que, según se demostró, la convergencia en-los-momentos implica la convergencia en-probabilidad, este ejemplo también permite comprobar que la convergencia en-probabilidad no implica la convergencia casi-con-certeza.

Ejemplo n° 40:

Sea una sucesión de variables aleatorias $\{X_n\}$ todas con distribución de probabilidades de la forma:

$$p(X_n = x_n) = \begin{cases} \dfrac{1}{2n^2} & para \qquad x_n = -n^{2/s} \\ 1 - \dfrac{1}{n^2} & \qquad\qquad x_n = 0 \\ \dfrac{1}{2n^2} & \qquad\qquad x_n = n^{2/s} \\ 0 & para\ cualquier\ otro\ valor\ de\ x \end{cases}$$

(para $n, s = 1,2, \dots$). Para $\varepsilon > 0$ se obtiene, entonces, que:

$$p\left[\bigcup_{j=n}^{\infty}\left(|X_j| > \varepsilon\right)\right] \leq \sum_{j=n}^{\infty}\left(|X_j| > \varepsilon\right) = \sum_{j=n}^{\infty}\frac{1}{j^2}$$

De modo que se verificará $\lim_{n\to\infty}\left[\bigcup_{j=n}^{\infty}\left(|X_j| > \varepsilon\right)\right] = 0$. Es decir, que $\lim_{n\to\infty}\left[\bigcap_{j=n}^{\infty}\left(|X_j| \leq \varepsilon\right)\right] = 1$ y, por lo tanto, que $X_n \xrightarrow{ccc} 0$.

Por otra parte, se puede concluir fácilmente que $E(|X_n|^s) = 1$ y, en consecuencia, que $\lim_{n\to\infty} E(|X_n|^s) = 1$. Lo cual demuestra que la sucesión $\{X_n\}$ converge casi-con-certeza a cero, pero no converge en el momento de orden s. Asimismo, como según se demostró en la sección precedente, la convergencia casi-con-certeza implica la convergencia en-probabilidad, la sucesión $\{X_n\}$ aquí definida es un ejemplo de convergencia en-probabilidad y no convergencia en el momento de orden s.

Ejemplo n° 41:

Sea una sucesión de variables aleatorias $\{X_n\}$ todas con distribución de probabilidades de la forma:

$$p(X_n = x_n) = \begin{cases} \dfrac{1}{2} & para \qquad x_n = -\dfrac{1}{n} \\ \dfrac{1}{2} & \qquad\qquad x_n = \dfrac{1}{n} \\ 0 & para\ cualquier\ otor\ valor\ de\ x_n \end{cases}$$

Se obtiene en forma inmediata que $E(|X_n|^s) = \frac{1}{n^s}(s = 1,2,\dots)$ y, en consecuencia, que $\lim_{n\to\infty} E(|X_n|^s) = 0$. Es decir, que $X_n \xrightarrow{m_s} 0$.

Sea, por otra parte, el evento $(|X_j| \le \varepsilon)$. Teniendo en cuenta que, para $j < k$, $|X_j| > |X_k|$ y, por lo tanto, que $(|X_j| \le \varepsilon) \subset (|X_k| \le \varepsilon)$, se obtiene que $\bigcap_{j=n}^{\infty}(|X_j| \le \varepsilon) = (|X_n| \le \varepsilon)$. De modo que, para un $\varepsilon > 0$ y $n > \frac{1}{\varepsilon}$, resulta que:

$$p\left[\bigcap_{j=n}^{\infty}(|X_j| \le \varepsilon)\right] = p(|X_n| \le \varepsilon) = 1$$

Es decir, que $X_n \xrightarrow{ccc} 0$.

Obsérvese que los ejemplos 39, 40 y 41 permiten concluir que, si bien no existe relación de implicación entre las convergencias casi-con-certeza y en-el-momento de orden s, dichas convergencias no son incompatibles.

12.5.- La convergencia completa

Se dice que una sucesión de variables aleatorias $\{X_n, n = 1,2,\dots\}$ converge completamente a cero, $X_n \xrightarrow{compl} 0$, si se verifica que $\lim_{n\to\infty} \sum_{j=n}^{\infty} p(|X_j| > \varepsilon) = 0\ (\varepsilon > 0)$.

A partir de esta definición se obtiene en forma inmediata que $\{X_n, n = 1,2,\dots\}$ converge completamente a la variable X, $X_n \xrightarrow{compl} X$, si se verifica que $X_n - X \xrightarrow{compl} 0$.

13.- La función característica

13.1.- Definición

Además de las ya mencionadas funciones de probabilidades, de densidad, de distribución y de supervivencia existen otras funciones útiles para el estudio de las variables aleatorias, las cuales forman parte de un conjunto de transformaciones del tipo:

$$H_X(w) = \int_{\Omega(X)} h(x,w)\,dF_X(x)$$

donde w es una variable real (sin significado probabilístico).

Entre todas estas funciones es particularmente importante la "función característica"[125] , que se obtiene haciendo $h(x,w) = e^{iwx}$ (donde $i = \sqrt{-1}$ es la unidad imaginaria) y que es la trasformación de Fourier de la función $F_X(x)$:

$$C_X(w) = E\left(e^{iwx}\right) =$$

$$= \int_{-\infty}^{\infty} e^{iwx} dF_X(x) = \int_{-\infty}^{\infty} cos(wx)\, dF_X(x) + i \int_{-\infty}^{\infty} sen(wx)\, dF_X(x)$$

En general, dada una variable aleatoria k-dimensional, $Z = \{X_1, X_2, \dots, X_k\}$, su función característica queda definida de la siguiente forma:

$$C_{X_1,X_2,\dots,X_k}(w_1, w_2, \dots, w_k) = E\left[e^{i(w_1X_1+w_2X_2+\cdots+w_kX_k)}\right] =$$

$$= \int_{\Omega(Z)} e^{i(w_1X_1+w_2X_2+\cdots+w_kX_k)} dF_{X_1,\dots,X_k}(x_1, \dots, x_k)$$

La función característica correspondiente a la variable marginal X_i se obtiene haciendo $w_j = 0$ para todo $j \neq i$ en la definición anterior, $C_{X_i}(w_i) = C_{X_1,X_2,\dots,X_k}(0, \dots, w_i, \dots, 0)$.

13.2.- Propiedades

La función característica, independientemente de la forma que asuma la distribución de probabilidades de la variable, posee las siguientes propiedades:

1) $|C_X(w)| = \left|\int_{-\infty}^{\infty} e^{iwx} dF_X(x)\right| = \left|\int_{-\infty}^{\infty} [cos(wx) + i\, sen\,(wx)] dF_X(x)\right| \leq$

$$\leq \int_{-\infty}^{\infty} |cos(wx) + i\, sen\,(wx)|\, dF_X(x) = 1$$

2) $C_X(0) = \int_{-\infty}^{\infty} dF_X(x) = \int_{-\infty}^{\infty} f_X(x) dx = 1$

3) $|C_X(w+h) - C_X(w)| = \left|\int_{-\infty}^{\infty} e^{ix(w+h)} dF_X(x) - \int_{-\infty}^{\infty} e^{ixw} dF_X(x)\right| =$

[125] Denominación debida a Poincaré (1896).

$$= \left| \int_{-\infty}^{\infty} \left[e^{ix(w+h)} - e^{ixw}\right] dF_X(x) \right| \leq$$

$$\leq \int_{-\infty}^{\infty} \left|e^{ix(w+h)} - e^{ixw}\right| \, dF_X(x) = \int_{-\infty}^{\infty} \left|e^{ixw}\right|\left|e^{ixw} - 1\right| dF_X(x) =$$

$$= \int_{-\infty}^{\infty} |cos(wx) + i \, sen \, (wx)| \left|e^{ixw} - 1\right| \, dF_X(x) = \int_{-\infty}^{\infty} \left|e^{ixw} - 1\right| dF_X(x)$$

Lo que demuestra que la función $C_X(w)$ es uniformemente continua en $\mathbb{R}_k$, $\lim_{h \to 0}[C_X(w+h) - C_X(w)] = 0$.

4) Existe correspondencia biunívoca entre las funciones características y las funciones de distribución de probabilidades: a cada función de distribución le corresponde una función característica y viceversa, a cada función característica le corresponde una y sólo una función de distribución. La primera parte de esta propiedad es una consecuencia inmediata de la definición de función característica. En cuanto a la segunda, se obtiene mediante la utilización de la denominada fórmula de inversión[126].

5) Dadas dos constantes, a y b, reales, se tiene que:

$$C_{ax+b}(w) = E\left[e^{iw(aX+b)}\right] = E\left(e^{iwaX}e^{iwb}\right) =$$

$$= e^{iwb}E\left[e^{iw(aX)}\right] = e^{iwb}C_x(aw)$$

6) Propiedad multiplicativa

Sea $\{X_1, X_2, \ldots, X_n\}$ una sucesión de variables aleatorias independientes, se verificará entonces que:

$$C_{X_1+X_2+\cdots+X_n}(w) = E\left[e^{iw(X_1+\cdots+X_n)}\right] = E\left(e^{iwX_1}\right)E\left(e^{iwX_2}\right) \ldots E\left(e^{iwX_n}\right) =$$

$$= C_{X_1}(w)C_{X_2}(w) \ldots C_{X_n}(w)^{127}$$

---

126 Ver Sec. 13.4.

127 Obsérvese que, en base a este teorema y a la propiedad de unicidad, es posible transformar un problema de convolución de funciones de densidad en un simple producto de funciones características.

7) Sean X e Y dos variables aleatorias con funciones características $C_X(w_1)$ y $C_Y(w_2)$, respectivamente. La condición necesaria y suficiente para que estas variables puedan ser consideradas independientes es que $C_{(X,Y)}(w_1, w_2) = C_X(w_1)C_Y(w_2)$. Pero la condición de independencia de las variables no es necesaria para que se verifique la propiedad multiplicativa.

Ejemplo n° 42:

Sea una variable aleatoria X_1 con distribución de probabilidades de la forma:

$$f_{X_1}(x_1) = \frac{1}{\pi}\frac{1}{1 + x_1^2} \qquad (-\infty \leq x_1 \leq \infty)$$

y función característica definida, en consecuencia, por $C_{X_1}(w) = e^{-|w|}$ y sea una variable aleatoria $X_2 = X_1$. Estas dos variables están obviamente relacionadas. De acuerdo con la definición de función característica será:

$$C_{X_1+X_2}(w) = C_{2X_1}(w) = C_{X_1}(2w) = e^{-2|w|}$$

Por otra parte, se verifica que:

$$C_{X_1}(w)C_{X_2}(w) = e^{-|w|}e^{-|w|} = e^{-2|w|}$$

Es decir, a pesar de que X_1 y X_2 son variables relacionadas, se cumple la propiedad multiplicativa, $C_{X_1+X_2}(w) = C_{X_1}(w)C_{X_2}(w)$.

8) Sea una variable aleatoria X con función característica $C_X(w)$ y sea X' una variable aleatoria semejante a X e independiente de ella. Entonces la variable aleatoria $Z = X - X'$ será simétrica[128], con función característica:

$$C_{X-X'}(w) = C_X(w)C_{-X'}(w) = C_X(w)C_X(-w) = C_X(w)\overline{C_X(w)} = |C_X(w)|^2$$

Con lo que queda demostrado que la función característica de una variable $(X - X')$ simétrica, es real.

9) Dada una variable aleatoria unidimensional X y un entero positivo s, si se admite la posibilidad del pasaje al límite bajo el signo integral, se verifica que:

---

[128] Esta condición implica que las variables (X, X') y (X', X) poseen la misma distribución de probabilidades. Dadas dos variables aleatorias semejantes e independientes, X y X', a la operación $(X - X')$ se la denomina simetrización.

$$\frac{d^s}{dw^s}C_X(w) = \frac{d^s}{dw^s}\int_{-\infty}^{\infty} e^{iwx}dF_X(x) = \int_{-\infty}^{\infty} \frac{d^s}{dw^s}e^{iwx}dF_X(x) =$$

$$= \int_{-\infty}^{\infty}(ix)^s e^{iwx}dF_X(x)$$

Luego, será $\frac{d^s}{dw^s}C_X(w)|_{w=0} = i^s E(X^s)$. Es decir, se obtiene que[129]:

$$m_s(X) = E(X^s) = i^{-s}\left[\frac{d^s}{dw^s}C_X(w)\right]_{w=0} \quad (s = 1,2,\dots)$$

10) Sea $\{X_1, X_2, \dots, X_n\}$ una sucesión de variables aleatorias con funciones de distribución de probabilidades $F_{X_n}(x) = p(X_n \leq x)(n = 1,2,\dots)$ y funciones características $C_{X_n}(w) = \int_{\Omega(X_n)} e^{iwx_n}\, dF_{X_n}(x_n)(n = 1,2,\dots, w\epsilon\mathbb{R}_1)$. Como se vio en la Sec. 12.2.2, la condición necesaria y suficiente para que $F_{X_n}(x)$ converja a la función $F_X(x)$ es que, para toda función $g(x)$ continua y acotada, se verifique que:

$$\int_{-\infty}^{\infty} g(x)F_{X_n}(x) \to \int_{-\infty}^{\infty} g(x)F_X(x)$$

Dado que e^{iwX} es continua y acotada en módulo, se concluye que, si $F_{X_n}(x) \to F_X(x)$, entonces $C_{X_n}(w) \to C_X(w)$ para todo w real.

11) Propiedad de continuidad de la correspondencia entre funciones características y funciones de distribución de probabilidades (teorema de Lévy-Cramér)

Sea $\{X_1, X_2, \dots, X_n\}$ una sucesión de variables aleatorias con funciones de distribución de probabilidades $F_{X_n}(x)(n = 1,2,\dots)$y funciones características $C_{X_n}(w)$ y sea una variable aleatoria X con función de distribución $F_X(x)$ y función característica $C_X(w)$. Supóngase que $C_{X_n}(w) \to C_X(w)$ para todo punto de $C_X(w)$ y supóngase que $F_{X_n}(x)$ no converje a $F_X(x)$ (es decir, que no tiende a $F_X(x)$ en un punto x^* de continuidad de $F_X(x)$), existirá entonces

[129] Si se verifica que $|E(X^s)| \leq \infty$ (o $E(|X|^s) < +\infty$), entonces esta derivada existe y es uniformemente continua para cualquier valor de w y, viceversa, si para un s positivo, la derivada de $C_X(w)$ existe en el origen, entonces se verifica que $|E(X^s)| \leq \infty$.

una sub-sucesión de $\{F_{X_n}(x)\}$ que converje en el punto x^* a un valor distinto de $F_X(x^*)$. Esta sub-sucesión tendrá, a su vez, una sub-sucesión tal que $\{F_{n_j}(x)\} \to H$ en todos los puntos. Donde $H(+\infty) - H(-\infty) \geq 1$; es decir, es tal que puede ser considerada como una función de distribución de probabilidades. Entonces, de acuerdo a lo expresado en la propiedad precedente, se verificará que $C_{n_j}(w) \to C_H(w)$. Además, por hipótesis $C_{n_j}(w) \to C_X(w)$. Luego, se puede concluir que $C_H(w)$ y $C_X(w)$ deben coincidir y, en consecuencia, que H y $F_X(x)$ también deben coincidir en los puntos de continuidad. Esto demuestra que la hipótesis:

$$H(x^*) = \lim_{j\to\infty} F_{n_j}(x^*) = c \neq F_X(x)$$

es falsa.

Por otra parte, se puede escribir:

$$\int_0^u C_X(w)dw = \int_0^u \left[\int_{-\infty}^{\infty} e^{iwx} dF_X(x)dw\right] =$$

$$= \int_{-\infty}^{\infty} \left[\int_0^u e^{iwx} dw\right] dF_X(x) = \int_{-\infty}^{\infty} \frac{e^{iwx}-1}{ix} dF_X(x)$$

Teniendo en cuenta que $\left|e^{iwx}-1\right| \leq \left|e^{iwx}\right| + 1 = 2$ y que $\left|e^{iwx}-1\right| \leq |ux|$, para $a > 0$, se obtiene que:

$$\left|\int_0^u C_{n_j}(w)dw\right| \leq \int_{|x|\leq a} \frac{e^{iwx}-1}{ix} dF_{n_j}(x) + \int_{|x|>a} \frac{e^{iwx}-1}{ix} dF_{n_j}(x) \leq$$

$$\leq \int_{|x|\leq a} |u| \; dF_{n_j}(x) + \int_{|x|>a} \frac{2}{|x|} dF_{n_j}(x) \leq |u| \left[F_{n_j}(a) - F_{n_j}(-a)\right] + \frac{2}{a}$$

y, haciendo $j \to \infty$ y $a \to +\infty$, se obtiene que:

$$H(a) - H(-a) \geq \left|\frac{1}{u}\int_0^u C_X(w)dw\right| + \frac{2}{a}$$

$$H(\infty) - H(-\infty) \geq \left| \frac{1}{u} \int_{0}^{u} C_X(w)dw \right| (u > 0)$$

Dado que $C_X(w)$ es una función continua, se verifica que:

$$H(+\infty) - H(-\infty) \geq \lim_{u \to 0} H(a) - H(-a) \geq \left| \frac{1}{u} \int_{0}^{u} C_X(w)dw \right| = C_X(0) = 1$$

Es decir que H puede ser considerada una función de probabilidades. Queda entonces demostrado que, si $C_{X_n}(w) \to C_X(w)$ en todo punto de $C_X(w)$, entonces $F_{X_n}(x) \to F_X(x)$. Tomando en consideración la implicación demostrada en la propiedad anterior, se puede concluir que:

$$F_{X_n}(x) \to F_X(x) \Leftrightarrow C_{X_n}(w) \to C_X(w)$$

Es decir que la correspondencia entre las funciones de distribución de probabilidades y las funciones características no sólo es biunívoca sino que es bicontinua, es decir que si $X_n \xrightarrow{d} X$, entonces $C_{X_n}(w) \to C_X(w)$ (para todo w) y, viceversa, si $C_{X_n}(w) \to C_X(w)$, entonces X_n converge en-distribución a X[130].

Ejemplo n° 43:

Sea X una variable aleatoria con distribución de probabilidades de la forma:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & para \quad a \leq x \leq b \\ 0 & para\ cualquier\ otro\ valor\ de\ x \end{cases}$$

La función característica será de la forma:

$$C_X(w) = \int_{-\infty}^{\infty} e^{iwx} f_X(x)dx = \frac{1}{b-a} \int_{-\infty}^{\infty} e^{iwx}dx = \frac{e^{ibw} - e^{iaw}}{iw(b-a)} =$$

$$= \frac{1}{iw(b-a)} \left\{ \left[1 + ibw + \frac{(ibw)^2}{2!} + \frac{(ibw)^3}{3!} + \cdots \right] - \right.$$

[130] Todas estas propiedades constituyen, en realidad, las condiciones necesarias para que una función compleja de una variable real pueda ser considerada una función característica.

$$-\left[\left[1+iaw+\frac{(iaw)^2}{2!}+\frac{(iaw)^3}{3!}+\cdots\right]\right]\Bigg\}=$$

$$=\frac{1}{iw(b-a)}\sum_{j=1}^{\infty}\left(b^j-a^j\right)\frac{(iw)^j}{j!}=\frac{1}{b-a}\sum_{j=1}^{\infty}\left(b^j-a^j\right)\frac{(iw)^{j-1}}{j!}=$$

$$=\frac{1}{b-a}\sum_{s=0}^{\infty}\left(b^{s+1}-a^{s+1}\right)\frac{(iw)^s}{(s+1)!}=\sum_{s=0}^{\infty}\left[\frac{b^{s+1}-a^{s+1}}{b-a}\frac{i^s}{s+1}\right]\frac{w^s}{s!}$$

En el desarrollo de Taylor de la función $C_X(w)$ el coeficiente que acompaña al factor $\frac{w^s}{s!}$ es $C_X^{(s)}(0)$, es decir $i^s m_s(X)$:

$$\frac{b^{s+1}-a^{s+1}}{b-a}\frac{i^s}{s+1}=\frac{\partial^s}{\partial w^s}C_X(w)|_{w=0}=i^s m_s(X)$$

De modo que los momentos de la variable X quedan definidos de la siguiente forma:

$$m_s(X)=\frac{b^{s+1}-a^{s+1}}{b-a}\frac{1}{s+1}\qquad (s=0,1,2,\dots)$$

Ejemplo n° 44:

Sean X e Y dos variables aleatorias independientes con distribuciones de probabilidades de la forma:

$$f_X(x)=\begin{cases}1 & para \quad 0\le x\le 1\\ 0 & para\ cualquier\ otro\ valor\ de\ x\end{cases}$$

$$f_Y(y)=\begin{cases}1 & para \quad 0\le y\le 1\\ 0 & para\ cualquier\ otro\ valor\ de\ x\end{cases}$$

Sus funciones características serán, respectivamente:

$$C_X(w)=E\left(e^{iwX}\right)=\int_0^1 e^{iwx}dx=\frac{e^{iw}-1}{iw}$$

$$C_Y(w)=E\left(e^{iwY}\right)=\frac{e^{iw}-1}{iw}$$

De acuerdo con la propiedad multiplicativa la función característica de la variable $Z = X + Y$ será de la forma:

$$C_Z(w) = C_{X+Y}(w) = E\left(e^{iw(X+Y)}\right) = C_X(w)C_Y(w) = \left(\frac{e^{iw} - 1}{iw}\right)^2 =$$

$$= \frac{e^{2iw} - 2e^{iw} + 1}{-w^2}$$

Haciendo ahora el proceso inverso, si se tiene en cuenta que la función de densidad de la variable Z está definida por:

$$f_Z(z) = \begin{cases} z & para \quad 0 \leq z < 1 \\ 2 - z & \quad 1 < z \leq 2 \\ 0 & para\ cualquier\ otro\ valor\ de\ z \end{cases}$$

Su función característica será de la forma:

$$C_Z(w) = E\left(e^{iwZ}\right) = \int_{-\infty}^{\infty} e^{iwz} f_Z(z)dz = \int_0^1 e^{iwz} z dz + \int_1^2 e^{iwz}(2 - z)dz =$$

$$= \left\{\frac{1}{iw}\left[ze^{iwz}\right]_0^1 - \frac{1}{iw}\int_0^1 e^{iwz}dz\right\} +$$

$$+2\int_1^2 e^{iwz}dz - \left\{\frac{1}{iw}\left[ze^{iwz}\right]_1^2 - \frac{1}{iw}\int_1^2 e^{iwz}dz\right\} =$$

$$= \frac{e^{2iw} - 2e^{iw} + 1}{(iw)^2} = C_X(w)C_Y(w)$$

Igualdad que confirma la definición de la función característica obtenida de la aplicación de la propiedad multiplicativa.

Ejemplo n° 44:

Sea U una variable aleatoria con distribución de probabilidades de la forma:

$$f_U(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \qquad (-\infty \leq u \leq \infty)$$

Su función característica, considerando a la unidad imaginaria como una constante real, está definida por:

$$C_U(w) = E\left(e^{iwU}\right)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{iwU} e^{-u^2/2} d \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(u^2-2iuw)} du =$$

$$= \frac{e^{-\frac{w^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(u-iw)^2} du = e^{-w^2/2}$$

y, desarrollando en serie la función exponencial, se obtiene que:

$$C_U(w) = 1 - \frac{1}{2} w^2 + \frac{1}{2^2 2!} w^4 - \frac{1}{2^3 3!} w^6 + \cdots$$

De esta expresión se puede concluir que los momentos de orden impar son todos nulos (la distribución de probabilidades es simétrica) y que para los de orden par se verifica la siguiente relación:

$$\frac{(-1)^s}{2^s s!} = \frac{i^{2s}}{(2s)!} m_{2s}(U) = \frac{(-1)^s}{(2s)!} m_{2s}(U)$$

Es decir, que $m_{2s}(U) = \frac{(2s)!}{2^s s!} (s = 0,1,2, \dots)$.

Ejemplo n° 46:

Sean X e Y dos variables aleatorias independientes con distribuciones de probabilidades de la forma:

$$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x - m_X}{\sigma_X}\right)^2\right] \qquad (-\infty \leq x \leq \infty)$$

$$f_Y(y) = \frac{1}{\sigma_Y \sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{y - m_Y}{\sigma_Y}\right)^2\right] \qquad (-\infty \leq y \leq \infty)$$

La función característica de la variable X queda definida de la siguiente forma:

$$C_X(w) = E\left(e^{iwX}\right) = \frac{1}{\sigma_X\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{iwX} exp\left[-\frac{1}{2}\left(\frac{x-m_X}{\sigma_X}\right)^2\right] dx =$$

$$= \frac{1}{\sigma_X\sqrt{2\pi}} \int_{-\infty}^{\infty} exp\left\{-\frac{1}{2\sigma_X^2}\left[(x-m_X)^2 - 2iwx\sigma_X^2\right]\right\} dx =$$

$$= \frac{1}{\sigma_X\sqrt{2\pi}} \int_{-\infty}^{\infty} exp\left\{-\frac{1}{2\sigma_X^2}(x-m_X-iw\sigma_X^2)^2 + \frac{1}{2}(2iwm_X - w^2\sigma_X^2)\right\} dx =$$

$$= \frac{exp\left(iwm_X - \frac{1}{2}w^2\sigma_X^2\right)}{\sigma_X\sqrt{2\pi}} \int_{-\infty}^{\infty} exp\left[-\frac{1}{2\sigma_X^2}(x-m_X-iw\sigma_X^2)^2\right] dx$$

Aplicando en esta integral la transformación $\frac{1}{\sigma_X}(x-m_X-iw\sigma_X^2) = u$ se obtiene que:

$$C_X(w) = \frac{exp\left(iwm_X - \frac{1}{2}w^2\sigma_X^2\right)}{\sigma_X\sqrt{2\pi}} \int_{-\infty}^{-u^2/2} e^{-u^2/2} du = exp\left(iwm_X - \frac{1}{2}w^2\sigma_X^2\right)$$

De la misma forma, se demuestra que $C_Y(w) = \left(iwm_Y - \frac{1}{2}w^2\sigma_Y^2\right)$.

Por otra parte, de acuerdo con la propiedad multiplicativa, la función característica de la variable bidimensional $Z = X + Y$ está dada por:

$$C_Z(w) = C_{X+Y}(w) = E\left[e^{i(X+Y)w}\right] = C_X(w)C_Y(w) =$$

$$= exp\left[i(m_X+m_Y)w - \frac{1}{2}(\sigma_X^2+\sigma_Y^2)w^2\right]$$

De lo que se concluye que la variable $(X+Y)$ también posee distribución de probabilidades de la forma:

$$f_Z(z) = \frac{1}{\sigma_Z\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{z-m_Z}{\sigma_Z}\right)^2\right] \quad (-\infty \leq z \leq \infty)$$

Con medidas $m_Z = m_X + m_Y$ y $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$.

Por inducción se puede demostrar que la variable definida por la suma de un conjunto finito de variables aleatorias independientes, $\{X_1, X_2, \dots, X_n\}$, con distribuciones de probabilidades de la forma:

$$f_{X_j}(x_j) = \frac{1}{\sigma_j\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x_j - m_j}{\sigma_j}\right)^2\right] \quad (-\infty \leq x_j \leq \infty, j = 1,2, \dots, n)$$

también posee una distribución de la forma:

$$f_Z(z) = \frac{1}{\sigma_Z\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{z - m_Z}{\sigma_Z}\right)^2\right] \qquad (-\infty \leq z \leq \infty)$$

con medidas $m_Z = \sum_{j=1}^{n} m_j$ y $\sigma_Z^2 = \sum_{j=1}^{n} \sigma_j^2$. Este resultado puede ser invertido concluyéndose que, si la variable aleatoria definida por un conjunto finito de variables aleatorias independientes posee distribución de probabilidades $f_Z(z)$, entonces las variables marginales tendrán distribuciones de probabilidades de la forma:

$$f_{X_j}(x_j) = \frac{1}{\sigma_j\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x_j - m_j}{\sigma_j}\right)^2\right] \quad (-\infty \leq x_j \leq \infty, j = 1,2, \dots, n)$$

Ejemplo n° 47:

Sea $\{X_1, X_2, \dots, X_n\}$ una sucesión de variables aleatorias cada una de las cuales puede asumir con la misma probabilidad cualquiera de los valores de su dominio, $\Omega = \left\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\right\}$, Las funciones de distribución de probabilidades de las variables X_j $(j = 1,2, \dots, n)$ serán, entonces, de la forma:

$$F_{X_j}(x) = \begin{cases} 0 & para \quad x \leq \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} < x \leq \frac{2}{n} \\ & \dots\dots\dots\dots\dots\dots. \\ \frac{k}{n} & \frac{k}{n} < x \leq \frac{k+1}{n} \\ & \dots\dots\dots\dots\dots\dots.. \\ \frac{n-1}{n} & \frac{n-1}{n} < x \leq 1 \\ 1 & x > 1 \end{cases}$$

y su función característica será de la forma:

$$C_{X_j}(w) = \sum_{h=1}^{n} e^{iw\frac{h}{n}} \frac{1}{n} = \frac{1}{n} e^{i\frac{w}{n}} \frac{e^{i\frac{w}{n}n} - 1}{e^{i\frac{w}{n}} - 1} = \frac{1}{n} \frac{e^{i\frac{w}{n}}}{e^{i\frac{w}{n}} - 1} \left(e^{iw} - 1\right)$$

$(j = 1,2, \dots, n)$.

Luego, se verificará que:

$$\lim_{j\to\infty} C_{X_j}(w) = \left(e^{iw} - 1\right) \lim_{j\to\infty} \left(\frac{e^{iw/n} - 1}{1/_n}\right)^{-1} e^{iw/n} =$$

$$= \frac{e^{iw} - 1}{iw} \qquad (j = 1,2, \dots, n)$$

De acuerdo con lo demostrado en el ejemplo n° 43, ésta es la función característica correspondiente a una distribución de probabilidades uniforme en el intervalo $[0,1]$. Es decir, este resultado confirma, a partir de la definición de función característica, la convergencia en-distribución –ya analizada en el ejemplo n° 34 - de la sucesión $\{X_1, X_2, \dots, X_n\}$ a una variable uniforme en el intervalo $[0,1]$.

13.3.- Análisis de las condiciones de unicidad de los momentos

Sean X_1 y X_2 dos variables aleatorias con funciones de distribución de probabilidades $F_{X_1}(x)$ y $F_{X_2}(x)$, respectivamente, con todos sus momentos finitos y tales que $m_s = E(X_1^s) = E(X_2^s) (s = 0,1,2, \dots)$. Supóngase que exista por lo menos un valor de $t > 0$ tal que $\sum_{s=0}^{\infty} m_s\left(X_j\right) \frac{t^s}{s!} < +\infty$ $(j = 1,2)$. A partir de la relación $e^{iz} - 1 = \int_{-\infty}^{z} ie^{iy} dy$ $(z \in \mathbb{R})$, se obtiene inmediatamente que $\left|e^{iz} - 1\right| \leq |z|$. Con un razonamiento similar, por inducción, se demuestra que:

$$\left| e^{iz} - \sum_{j=0}^{n} \frac{(iz)^j}{j!} \right| = \int_0^{|z|} \left(e^{iy} - \sum_{j=0}^{n-1} \frac{(iy)^j}{j!} \right) dy \leq$$

$$\leq \int_0^{|z|} \frac{y^n}{n!} dy = \frac{|z|^{n+1}}{(n+1)!} \qquad (n > 0)$$

De modo que, para z real y $n = 0,1,2, \dots$, se puede escribir:

$$\left| e^{iz} - \sum_{j=0}^{n} \frac{(iz)^j}{j!} \right| \leq \frac{|z|^{n+1}}{(n+1)!}$$

Luego, para ambas variables se verificará que:

$$\left| C_{X_j}(t) - \sum_{s=0}^{2n-1} m_s \frac{(it)^s}{s!} \right| \leq$$

$$\leq \int_{-\infty}^{\infty} \left| e^{itx} - \sum_{s=0}^{2n-1} \frac{(itx)^s}{s!} \right| dF_{X_j}(x) \leq \int_{-\infty}^{\infty} \frac{|tx|^{2n}}{(2n)!} dF_{X_j}(x) = m_{2n} \frac{t^{2n}}{(2n)!}$$

(para $j = 1,2,; n = 0,1,2, \dots$). Teniendo en cuenta la convergencia supuesta más arriba, se puede concluir que $\lim_{n\to\infty} m_{2n} \frac{w^{2n}}{(2n)!} = 0$. De modo que, para todo $|w| \leq t$, se verificará que en un entorno del origen, las funciones características son iguales:

$$C_{X_1}(w) = \lim_{n\to\infty} \sum_{j=0}^{2n-1} m_j \frac{(iw)^j}{j!} = C_{X_2}(w)$$

Por otra parte, dado que $C_{X_1}(w)$ y $C_{X_2}(w)$ poseen derivadas de cualquier orden, utilizando la desigualdad anterior, será:

$$\left| C_{X_j}(w) - \sum_{s=0}^{2n-1} C_{X_j}(z) \frac{(w-z)^s}{s!} \right| \leq$$

$$\leq \int_{-\infty}^{\infty} \left| e^{iwx} - \sum_{s=0}^{2n-1} e^{izx} (ix)^s \frac{(w-z)^s}{s!} \right| dF_{X_j}(x) =$$

$$= \int_{-\infty}^{\infty} \left| e^{izx} \right| \left| e^{i(w-z)x} - \sum_{s=0}^{2n-1} \frac{[ix(w-z)]^s}{s!} \right| dF_{X_j}(x) \leq$$

$$\leq \int_{-\infty}^{\infty} \frac{|x|^{2n} |w-z|^{2n}}{(2n)!} dF_{X_j}(x) = m_{2n} \frac{|w-z|^{2n}}{(2n)!} \quad (j = 1,2; n = 0,1,2, \dots)$$

Luego, de acuerdo con lo demostrado más arriba, para $|w-z|<t$, resulta que:

$$C_{X_1}(w)=\lim_{n\to\infty}\sum_{s=0}^{2n-1} C_{X_j}(w)\frac{(w-z)^s}{s!}$$

Ahora bien, para $|z|<t$, se verifica que $C_{X_1}(z)=C_{X_2}(z)$ y, por lo tanto, que las derivadas también coinciden. En consecuencia, si $|w|<2t$, para $z=\frac{w}{2}$, se tiene que:

$$C_{X_1}(w)=\lim_{n\to\infty}\sum_{s=0}^{2n-1} C_{X_1}^{(s)}(z)\frac{(w-z)^s}{s!}=\lim_{n\to\infty}\sum_{s=0}^{2n-1} C_{X_2}^{(s)}(z)\frac{(w-z)^s}{s!}=C_{X_2}(w)$$

Por inducción, haciendo $z=\frac{k-1}{k}w$ $(k=2,3,\dots)$, se obtiene la igualdad de las funciones características para todo w real.

De la demostración anterior se puede concluir que, dadas dos variables aleatorias, X_1 y X_2 con funciones de distribución $F_{X_1}(x)$ y $F_{X_2}(x)$ y los correspondientes momentos finitos e iguales, $m_s(X_1)=m_s(X_2)$ $(s=0,1,2,\dots)$, si se verifica que $\sum_{s=0}^{\infty} m_s\frac{w^s}{s!}$ $(\exists w>0)$, entonces por la propiedad de correspondencia biunívoca entre la función característica y la función de distribución de probabilidades, se verificará que $F_{X_1}(x)$=$F_{X_2}(x)$. Es decir que la sucesión de momentos, $m_s(X)$, define unívocamente a la variable X[131].

Asimismo, se puede demostrar que las condiciones de convergencia que se mencionan a continuación son equivalentes a la demostrada precedentemente para poder asegurar la unicidad de los momentos de una variable:

i) $\sum_{s=0}^{\infty} m_{2s}\frac{w^{2s}}{(2s)!}<+\infty$ $(\exists w>0)$.

ii) $m_{2s}\frac{w^{2s}}{(2s)!}\to 0$ $(\exists w>0)$

iii) $\underset{s\to\infty}{limsuper}\frac{(m_{2s})^{1/2s}}{2s}<+\infty$

iv) Criterio de Riesz: $\underset{s\to\infty}{liminfer}\frac{(m_{2s})^{1/2s}}{2s}<+\infty$

[131] Esta condición fue enunciada (sin demostración) en la Sec. 10.1.

v) Criterio de Carleman: $\sum_{s=0}^{\infty}(m_{2s})^{-1/2s} = +\infty$

vi) Criterio de Ghizzetti: $\underset{s\to\infty}{liminfer}\frac{(m_s)^{1/s}}{s^2} < +\infty$

vii) $C(z) = \sum_{s=0}^{\infty} m_s \frac{(iz)^s}{s!} \, (\exists w > 0, |z| < w)$

viii) $\sum_{s=0}^{\infty} m_s \frac{z^s}{s!} = \int_{-\infty}^{\infty} e^{zx} dF_Z(z) \, (\exists w > 0, |z| < w)$

13.4.- Algunos teoremas particulares acerca de la inversión de la función característica

Como ya se mencionó al analizar las propiedades de la correspondencia biunívoca entre la función característica y la distribución de probabilidades de una variable, la utilización de la fórmula de inversión:

$$f_X(x) = \frac{1}{2\pi}\lim_{a\to\infty}\int_{-a}^{a} e^{iwx} C_X(w)dw$$

suele conducir a expresiones cuya resolución resulta extremadamente complicada. Aquí se tratarán algunos casos particulares intentando utilizar en su resolución fundamentalmente argumentos probabilísticos[132].

1) Sea X una variable aleatoria con distribución de probabilidades de la forma:

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} exp\left(-\frac{x^2}{2\sigma_X^2}\right) \qquad (-\infty \leq x \leq \infty)$$

Su función característica será de la forma $C_X(w) = exp\left(-\frac{w^2\sigma_X^2}{2}\right)$. Si se toma como punto de partida la expresión:

$$\int_{-\infty}^{\infty} e^{-iwx} C_X(w)g(w)dw = \int_{-\infty}^{\infty} e^{-iwx}\left[\int_{-\infty}^{\infty} e^{iwy} dF_X(y)\right] g(w)dw =$$

[132] Cabe destacar aquí los trabajos de Fourier (1811), que fueron los que probablemente condujeron a Gauss a la expresión de las relaciones recíprocas $G(u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{iwu} g(w)dw$ y $g(w) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-iwu} G(u)du$ (esta demostración data probablemente de 1813, pero fue descubierta recién a fines del siglo XIX).

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} e^{-iwx} C_X(w) g(w) dw \right] dF_X(y) = \int_{-\infty}^{\infty} C(y-x) dF_X(y) dF_X(y) =$$

$$= \int_{-\infty}^{\infty} exp\left[-\frac{1}{2} \sigma_X^2 (y-x)^2 \right] dF_X(y)$$

y, teniendo en cuenta que:

$$e^{-z^2c^2/2} = \frac{\sqrt{2\pi}}{c} \left(\frac{c}{\sqrt{2\pi}} e^{-z^2c^2/2} \right) = \frac{\sqrt{2\pi}}{\sigma_Z} h(z)$$

(donde $h(z)$ denota una función de densidad similar a $f_X(x)$ de una variable z, con $\sigma_z = \frac{1}{c}$), se puede escribir:

$$\int_{-\infty}^{\infty} e^{iwx} C_X(w) g(w) dw = \frac{\sqrt{2\pi}}{\sigma_X} \int_{-\infty}^{\infty} h(y-x) dF_X(y)$$

La integral que figura en el segundo miembro define la convolución de las funciones $f_X(x)$ y $h(y)$, es decir, la función de distribución de probabilidades de la variable $Z = X + Y$, siendo X e Y variables independientes:

$$\int_{-\infty}^{\infty} h(y-x) dF_X(y) =$$

$$= \frac{\sigma_X}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-iwx} C_X(w) g(w) dw = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iwx} C_X(w)\, e^{-w^2c^2/2} dw$$

Luego, será:

$$F_Z(z_1) - F_Z(z_0) = \int_{z_0}^{z_1} f_{X+Y}(x) dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{z_0}^{z_1} e^{-iwx} dx \right] C_X(w) e^{-w^2c^2/2} d$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-iwz_0} - e^{-iwz_1}}{iw} C_X(w) e^{-w^2c^2/2} dw$$

Aplicando a ambos miembros de esta igualdad el operador límite cuando $c \to 0$, se obtiene que

$$\lim_{c\to 0}[F_Z(z_1) - F_Z(z_0)] = F_X(z_1) - F_X(z_0) =$$

$$= \lim_{c\to 0}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-iwz_0} - e^{-iwz_1}}{iw}C_X(w)e^{-w^2c^2/2}dw$$

Si se supone, en particular, que $C_X(w)$ cumple la condición $\int_{-\infty}^{\infty}|C_X(w)|dw = K < +\infty$, se verificará que:

$$\left|\int_{-\infty}^{\infty} e^{iwx}\, C_X(w)e^{-w^2c^2/2}dw\right| \leq \int_{-\infty}^{\infty}|C_X(w)|dw < +\infty$$

de modo que es posible el pasaje al límite bajo el signo integral:

$$F_X(z_1) - F_X(z_0) = \frac{1}{2\pi}\int_b^a\left[\lim_{c\to 0}\int_{-\infty}^{\infty} e^{iwx}\, C_X(w)e^{-w^2c^2/2}dw\right]dx =$$

$$= \frac{1}{2\pi}\int_b^a\left[\int_{-\infty}^{\infty} e^{iwx}\, C_X(w)dw\right]dx$$

Haciendo $z_0 \to -\infty$, se obtiene que:

$$F_X(z_1) = \int_{-\infty}^{\mathrm{b}}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iwx}\, C_X(w)dw\right]dx$$

Es decir, se demuestra que la definición de la función de densidad de la variable X, a partir de su función característica, será de la forma:

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iwx}\, C_X(w)dw$$

2) Sea $(k + hX)$ (donde k es un número real, $h > 0$ y X denota una variable aleatoria que puede asumir sólo valores enteros) una variable aleatoria con función de probabilidades $p(X_j) = p_j$. De acuerdo con la definición de función característica, se tiene que:

$$\int_{-\pi/2}^{\pi/2} e^{-iw(k+hX)} C_X(w)dw = \int_{-\pi/2}^{\pi/2} e^{-iw(k+hX)} \left[\sum_j e^{iw(k+hj)} p_j\right] dw =$$

$$= \int_{-\pi/2}^{\pi/2} \left[\sum_j e^{iw(j-x)h} p_j\right] dw = \sum_j p_j \int_{-\pi/2}^{\pi/2} e^{iw(j-x)h} dw =$$

$$= \frac{1}{h} \sum_j p_j \int_{-\pi}^{\pi} [cos(j-x)y + i\, sen(-x)y] dy$$

Esta integral es nula para todo $j \neq x$ y es igual a 2π para $j = x$. Luego, será:

$$\int_{-\pi/2}^{\pi/2} e^{-iw(k+hX)} C_X(w)dw = \frac{\sqrt{2\pi}}{h} p_x$$

Queda demostrado entonces que:

$$p_x = \frac{h}{\sqrt{2\pi}} \int_{-\pi/2}^{\pi/2} e^{-iw(k+hX)} C_X(w)dw$$

3) La fórmula de inversión de Lévy

Sea una variable aleatoria X y sean x_0 y x_1 dos puntos de continuidad de su función de distribución de probabilidades $F_X(x)$. De acuerdo a lo demostrado más arriba:

$$F_X(x_1) - F_X(x_0) = \lim_{c \to 0} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-iwx_0} - e^{-iwx_1}}{iw} C_X(w) e^{-w^2c^2/2} dw$$

Para un valor $A > 0$ esta integral puede ser expresada como la suma de dos integrales sobre los intervalos $|w| \leq A$ y $|w| > A$, respectivamente. En el primer intervalo se puede pasar al límite bajo el signo integral, de modo que se puede escribir:

$$F_X(x_1) - F_X(x_0) =$$

$$= \frac{1}{2\pi} \int_{-A}^{A} \frac{e^{-iwx_0} - e^{-iwx_1}}{iw} C_X(w) dw +$$

$$+ \lim_{c \to 0} \frac{1}{2\pi} \int_{|w|>A} \frac{e^{-iwx_0} - e^{-iwx_1}}{iw} C_X(w) e^{-w^2c^2/2} dw =$$

$$= \frac{1}{2\pi} \int_{-A}^{A} \frac{e^{-iwx_0} - e^{-iwx_1}}{iw} C_X(w) dw +$$

$$+ \frac{1}{2\pi} \lim_{c \to 0} \int_{|w|>A} \frac{e^{-iwx_0}}{iw} C_X(w) e^{-w^2c^2/2} dw -$$

$$- \frac{1}{2\pi} \lim_{c \to 0} \int_{|w|>A} \frac{e^{-iwx_1}}{iw} C_X(w) e^{-w^2c^2/2} dw$$

Ahora bien, tomando la segunda integral del segundo miembro, se obtiene que:

$$\int_{|w|>A} \frac{e^{-iwx_0}}{iw} C_X(w) e^{-w^2c^2/2} dw$$

$$= \int_{|w|>A} \frac{e^{-iwx_0}}{iw} \left[\int_{-\infty}^{\infty} e^{iwx} dF_X(x) \right] e^{-w^2c^2/2} dw =$$

$$= \int_{-\infty}^{\infty} \left[\int_{|w|>A} \frac{e^{-iw(x-x_0)}}{iw} e^{-w^2c^2/2} dw \right] dF_X(x)$$

Aplicando la sustitución $y = w(x - x_0)$[133] y teniendo en cuenta que la función $\frac{cos(x)}{x}$ es impar y la función $\frac{sen(x)}{x}$ es par, resulta que:

[133] Como se ha supuesto que x_0 es un punto de continuidad de $F_X(\cdot)$, será $x - x_0 \neq 0$.

$$\int_{|w|>A} \frac{e^{-iw(x-x_0)}}{iw} e^{-w^2c^2/2} dw = \int_{|y|>A|x-x_0|} \frac{e^{iy}}{iy} exp\left[-\frac{1}{2}\frac{y^2c^2}{(x-x_0)^2}\right] dy =$$

$$= \int_{|y|>A|x-x_0|} \frac{cos(y) + i\, sen(y)}{iy}\ exp\left[-\frac{1}{2}\frac{y^2c^2}{(x-x_0)^2}\right] dy =$$

$$= 2 \int_{|y|>A|x-x_0|} \frac{sen(y)}{y} exp\left[-\frac{1}{2}\frac{y^2c^2}{(x-x_0)^2}\right] dy$$

Sea A^* el menor entero mayor que $\frac{A|x-x_0|}{2\pi}$, es decir, tal que $\frac{A|x-x_0|}{2\pi} < A^* \leq \frac{A|x-x_0|}{2\pi} + 1$. Entonces se puede escribir:

$$\left| \int_{|y|>A|x-x_0|} \frac{sen(y)}{y} exp\left[-\frac{1}{2}\frac{y^2c^2}{(x-x_0)^2}\right] dy \right| \leq$$

$$\leq \left| \int_{A|x-x_0|}^{2A^*\pi} \frac{sen(y)}{y} exp\left[-\frac{1}{2}\frac{y^2c^2}{(x-x_0)^2}\right] dy \right| +$$

$$+ \left| \sum_{j=2A^*\pi}^{\infty} \int_{j\pi}^{(j+1)\pi} \frac{sen(y)}{y} exp\left[-\frac{1}{2}\frac{y^2c^2}{(x-x_0)^2}\right] dy \right|$$

Dado que la serie que figura en el segundo miembro está formada por términos de signo alternado, siendo su primer término positivo, su suma será positiva e inferior al primer término. Luego, será:

$$\left| \int_{|y|>A|x-x_0|} \frac{sen(y)}{y} exp\left[-\frac{1}{2}\frac{y^2c^2}{(x-x_0)^2}\right] dy \right| \leq$$

$$\leq \left| \int_{A|x-x_0|}^{2A^*\pi} \frac{sen(y)}{y} exp\left[-\frac{1}{2}\frac{y^2c^2}{(x-x_0)^2} \right] dy \right| +$$

$$+ \int_{2A^*\pi}^{(2A^*+1)\pi} \frac{sen(y)}{y} exp\left[-\frac{1}{2}\frac{y^2c^2}{(x-x_0)^2} \right] dy \leq$$

$$\leq \int_{A|x-x_0|}^{(2A^*+1)\pi} \frac{sen(y)}{y} dy \leq \int_{A|x-x_0|}^{A|x-x_0|+3\pi} \frac{|sen(y)|}{y} dy$$

La última integral del segundo miembro es menor que 3π y tiende a cero cuando $A \to \infty$. Aplicando entonces el teorema de la convergencia dominada se obtiene que:

$$\lim_{A\to\infty} \left| \lim_{c\to 0} \int_{|w|>A} \frac{e^{-iw(x-x_0)}}{iw} e^{-w^2c^2/2} dw \right| \leq$$

$$\leq \lim_{A\to\infty} \int_{-\infty}^{\infty} 2 \left[\int_{A|x-x_0|}^{A|x-x_0|+3\pi} \frac{|sen(y)|}{y} dy \right] dF_X(x) =$$

$$= \int_{-\infty}^{\infty} 2 \left[\lim_{A\to\infty} \int_{A|x-x_0|}^{A|x-x_0|+3\pi} \frac{|sen(y)|}{y} dy \right] dF_X(x) = 0$$

Obviamente resulta también que:

$$\lim_{A\to\infty} \left[\frac{1}{2\pi} \lim_{c\to 0} \int_{|w|>A} \frac{e^{-iwx_1}}{iw} C_X(w) e^{-w^2c^2/2} dw \right] = 0$$

Luego, queda demostrado que:

$$F_X(x_1) - F_X(x_0) = \lim_{A\to\infty} \frac{1}{2\pi} \int_{-A}^{A} \frac{e^{-iwx_0} - e^{-iwx_1}}{iw} C_X(w) dw$$

14.- La función generatriz de momentos

14.1.- Definición

Íntimamente vinculada con la función característica, existe otra función que se obtiene haciendo $h(x,w) = e^{wx}$ en la función $H_X(w) = \int_{\Omega(X)} h(x,w)$ $dF_X(x)$, denominada "función generatriz de momentos" de la variable X:

$$M_X(w) = E(e^{wX}) = \int\limits_{\Omega(X)} e^{wX} dF_X(x)$$

Con respecto a la función característica, la función $M_X(w)$ permite la simplificación que significa pasar del campo complejo al campo real. Como contrapartida, si bien la integral que define la transformación existe siempre ya que la función integrando es positiva, no es posible asegurar su convergencia para cualquier valor de w. La función $M(\cdot)$ existe si la integral es finita por lo menos para un w en el entorno del origen ($|w| \leq w_0$)[134].

La función $M_X(\cdot)$ siempre existe en el punto $w = 0$, $M_X(0) = E(1) = 1$[135]

14.2.- Propiedades

La función generatriz de momentos posee las siguientes propiedades:

1) Sea X una variable aleatoria con función generatriz de momentos $M_X(w)$ y sean las constantes reales a y b. Se tiene, entonces, que:

$$M_{aX+b}(w) = E\left[e^{w(aX+b)}\right] = e^{wb} M_X(aw)$$

(nótese que, si se verifica que $E(e^{wX}) < +\infty$, para $|w| \leq w_0$, entonces $E(e^{awX}) < +\infty$, para $|aw| \leq w_0$, es decir, para $|w| \leq \frac{w_0}{a}$).

2) Propiedad multiplicativa

Sea $\{X_1, X_2, \dots, X_n\}$ una sucesión de variables aleatorias independientes con funciones generatrices de momentos $M_{X_1}(w), M_{X_2}(w), \dots, M_{X_n}(w)$, respectivamente. Entonces la función generatriz de momentos de la variable

---

[134] La integral $\int e^{wx} dF_X(x)$ está definida para todo w, pero su utilización sólo interesa para aquellos valores para los cuales se mantiene finita. Si la variable X es acotada, es finita para todo w real. Si X no es acotada puede existir para algunos valores de w y para otros no.

[135] Se puede concluir en formas inmediata que $M_X(w) = C_X(-iw)$.

n-dimensional definida como la suma de las variables X_j, Z=$X_1 + X_2 + \cdots + X_n$, existirá y estará definida por:

$$M_{X_1+X_2+\cdots+X_n}(w) = E\left[e^{(X_1+X_2+\cdots+X_n)w}\right] = E(e^{X_1 w})\, E(e^{X_2 w}) \ldots E(e^{X_n w}) =$$

$$= M_{X_1}(w) M_{X_2}(w) \ldots M_{X_n}(w)$$

3) Sea una variable aleatoria X con función generatriz de momentos finita para $|w| \leq w_0$. Dado que para todo (wX) real resulta $e^{wX} > 0$, se verifica que:

$$e^{|wX|} = e^{wX} \leq e^{wX} + e^{-wX} \qquad (para\ wX > 0)$$

$$e^{|wX|} = e^{-wX} \leq e^{wX} + e^{-wX} \qquad (para\ wX < 0)$$

Por lo tanto, se cumple la relación:

$$E(e^{wX}) + E(e^{-wX}) = M_X(w) + M_X(-w) > E\left(e^{|wX|}\right)$$

Pero $e^{|wX|} = \sum_{j=0}^{\infty} \frac{|wX|^j}{j!} \geq \frac{|wX|^k}{k!} (k = 0,1,2, \ldots)$. De modo que:

$$\frac{|w|^k}{k!} E\left(|X|^k\right) < M_X(w) + M_X(-w) < +\infty$$

Es decir que, si la función generatriz de momentos es finita para todo w en un intervalo alrededor del punto $w = 0$, entonces sus momentos $m_s(X) = E(X^s)$ existen y son finitos[136].

4) Sea una variable aleatoria X con función generatriz de momentos finita para todo $|w| \leq w_0$. Entonces será:

$$M_X(w) = \int_{\Omega(X)} e^{wx} dF_X(x) = \int_{\Omega(X)} \sum_{j=0}^{\infty} \frac{w^j x^j}{j!} dF_X(x) =$$

$$= \sum_{j=0}^{\infty} \frac{w^j}{j!} \int_{\Omega(X)} x^j dF_X(x) = \sum_{j=0}^{\infty} m_j(X) \frac{w^j}{j!}$$

---

[136] Como se ve la existencia de la función generatriz de momentos puede considerarse una condición restringida para la existencia de los momentos: muchas distribuciones no la poseen, en particular aquellas cuyos momentos no son todos finitos.

5) Derivando la serie de potencias anterior término a término, se obtiene que:

$$\frac{d^s}{dw^s}M_X(w) = \sum_{j=0}^{\infty}\frac{1}{j!}m_j(X)\frac{d^s}{dw^s}w^j =$$

$$= m_0(x)\frac{d^s}{dw^s}w^0 + m_1(X)\frac{d^s}{dw^s}w + \frac{1}{2!}m_2(X)\frac{d^s}{dw^s}w^2 + \cdots =$$

$$= \frac{1}{s!}m_s(X)\frac{d^s}{dw^s}w^s + \frac{1}{(s+1)!}m_{s+1}(X)\frac{d^s}{dw^s}w^{s+1} + \cdots$$

Luego, será $m_s(X) = \frac{d^s}{dw^s}M_X(w)|_{w=0}$.

6) De acuerdo con lo demostrado en la propiedad **4)**, de la existencia de la función generatriz de momentos se deriva la expresión $\sum_{j=0}^{\infty} m_j(X)\frac{w^j}{j!}$ $(\exists w > 0)$. La cual permite asegurar que la sucesión de momentos identifica a la distribución de probabilidades[137]. Como se demostró en la propiedad **3)**, esta es condición necesaria y suficiente para la existencia de la función generatriz de momentos (la existencia de esta función, para todo w en el entorno del origen, garantiza la existencia de todos los momentos[138]), pero sólo es condición suficiente para que la sucesión de momentos identifique a la distribución: puede suceder que una variable aleatoria posea todos sus momentos finitos, que sea identificada por los momentos y que no posea función generatriz de momentos.

7) Sean $\{X_1, X_2, \ldots, X_n\}$ una sucesión de variables aleatorias y una variable X, con funciones generatrices de momentos $M_{X_1}(w), M_{X_2}(w), \ldots, M_{X_n}(w)$ y $M_X(w)$, respectivamente. Si se verifica que $\lim_{n\to\infty} M_{X_n}(w) = M_X(w)$, entonces se puede asegurar que X_n converge en-distribución a la variable X.

Ejemplo n° 48:

Sea X una variable aleatoria con distribución de probabilidades de la forma:

---

137 Ver Sec. 10.1.

138 Nótese que la existencia de la función característica no garantiza la finitud de los momentos: la relación $C_X(w) = i^{-s}m_s(X)$ (para s entero) se verifica sólo si $|E(X^s)| < \infty$ (ver Sec. 12.4.1).

$$f_X(x) = \begin{cases} \dfrac{e^{-\sqrt{x}}}{2} & para \quad x > 0 \\ 0 & \quad\quad\quad x < 0 \end{cases}$$

De acuerdo con su definición, el momento de orden s de la variable X está dado por:

$$m_s(X) = \int_{\Omega(X)} x^s f_X(x)dx = \int_0^\infty x^s \frac{e^{-\sqrt{x}}}{2} dx$$

Aplicando la transformación $\sqrt{x} = y$, será:

$$m_s(X) = \int_0^\infty y^{2s+1} e^{-y} dy = (2s+1)!$$

Por otra parte, se observa el cumplimiento de la condición de Ghizzetti[139]:

$$\lim_{s\to\infty} \frac{[m_s(X)]^{1/s}}{s^2} = \lim_{s\to\infty} \frac{[(2s+1)!]^{1/s}}{s^2} =$$

$$= \lim_{s\to\infty} \frac{(2s+1)^{2+1/s} e^{-2-1/s} [2\pi(2s+1)]^{1/s}}{s^2} = 4e^{-2}$$

lo que permite asegurar que la sucesión de momentos caracteriza unívocamente a la variable. Pero se verifica también que:

$$M_X(w) = E(e^{wX}) = \frac{1}{2}\int_0^\infty e^{wx} e^{-\sqrt{x}} dx = \frac{1}{2}\int_0^\infty e^{wx-\sqrt{x}} dx = +\infty \qquad (w > 0)$$

Es decir que, a pesar de cumplirse la condición de Ghizzetti, la función generatriz de momentos no existe.

Ejemplo n° 49:

Sea X una variable aleatoria con distribución de probabilidades de la forma $f_X(x) = \lambda e^{-\lambda x} (x > 0)$. Se verificará que:

[139] Ver Sec. 13.3.

$$M_X(w) = E(e^{wX}) = \lambda \int_0^{\infty} e^{wx} e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{-(\lambda - w)x} dx =$$

$$= \frac{\lambda}{\lambda - w} \int_0^{\infty} e^{-y} dy = \frac{\lambda}{\lambda - w}$$

(la integral es finita para $|w| < \lambda$). La derivada de orden s de la función generatriz será de la forma:

$$M_X^{(s)}(w) = \frac{s!\,\lambda}{(\lambda - w)^{s+1}} \qquad (s = 0,1,2,\dots)$$

Luego, el momento de orden s será $m_s(x) = M_X^{(s)}(0) = \frac{s!}{\lambda^s} (s = 0,1,2,\dots)$.

Sea, por otra parte, una variable aleatoria Y, independiente de X y con la misma distribución de probabilidades, $f_Y(y) = \lambda e^{-\lambda y} (y > 0)$. La distribución de probabilidades de la variable $Z = X + Y$ estará dada por la convolución:

$$f_Z(z) = \int_{-\infty}^{z} f(z - y) f_Y(y) dy = \lambda^2 \int_0^{z} e^{-\lambda(z-y)} e^{-\lambda y} dy =$$

$$= \lambda^2 e^{-\lambda z} \int_0^{z} dy = \lambda^2 z e^{-\lambda z}$$

y, como se demuestra en forma inmediata, su función generatriz de momentos:

$$M_Z(w) = E(e^{wZ}) = \int_{\Omega(Z)} e^{wz} f_Z(z) dz = \lambda^2 \int_0^{\infty} z e^{-(\lambda - w)z} dz =$$

$$= \frac{\lambda^2}{(\lambda - w)^2} \int_0^{\infty} y e^{-y} dy = \frac{\lambda^2}{(\lambda - w)^2} =$$

$$= \frac{\lambda}{\lambda - w} \frac{\lambda}{\lambda - w} = M_X(w) M_Y(w) \qquad (|w| < \lambda)$$

cumple la propiedad multiplicativa.

Ejemplo n° 50:

Sea X una variable aleatoria con distribución de probabilidades de la forma:

$$f_X(x) = \begin{cases} \dfrac{1}{\pi} & para \quad -\dfrac{\pi}{2} < x < \dfrac{\pi}{2} \\ 0 & para\ cualquier\ otro\ valor\ de\ x \end{cases}$$

y sea la variable aleatoria $Y = tg(X)$. Se verificará, entonces, que:

$$F_Y(y) = p(Y \leq y) = p[tg(x) \leq y] = p[X \leq arc\ tg(y)] = \frac{1}{\pi} arc\ tg(y)$$

Derivando se obtiene que:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{\pi} \frac{1}{1 + y^2}$$

(función conocida como distribución de Cauchy). La integral $\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{wy}}{1+y^2} dy$ es infinita para todo $w > 0$, de modo que no existe la función generatriz de momentos para la variable Y. Por otra parte, se puede demostrar que esta distribución no posee ningún momento finito.

14.3.- La función generatriz de momentos factoriales

Se dice que una variable aleatoria X posee "función generatriz de momentos factoriales" si existe un valor de w en el entorno del origen para el cual el valor esperado $N_X(w) = E[(1 + w)^X]$ es finito.

A fin de evitar los problemas que se pueden presentar al tratar con una potencia de exponente real, se puede limitar la definición a los valores de $w > -1$. Se puede escribir, entonces:

$$N_X(w) = E\left\{e^{\ln[(1+w)^X]}\right\} = E\left[e^{X \ln(1+w)}\right]$$

Ahora bien, dada la desigualdad $a < w < b$, para $-1 < a < 0 < b$, se verifica que $\ln(1 + a) < \ln(1 + w) < \ln(1 + b)$, siendo $\ln(1 + a) < 0 < \ln(1 + b)$. Se puede asegurar entonces, que la función generatriz de momentos factoriales de una variable existe si y sólo si existe su función generatriz de momentos. Esta implicación permite asegurar que, si existe la función generatriz de momentos factoriales de una variable, entonces todos sus momentos existen y todos sus momentos factoriales (que son combinaciones lineales de los momentos)[140] también existen. La relación entre ambas funciones es de la forma:

$$N_X(w) = E\left[e^{X\ln(1+w)}\right] = M_X[\ln(1+w)]$$

$$M_X(w) = E(e^{wX}) = E[(1+e^w-1)^X] = N_X(e^w-1)$$

De estas expresiones se puede concluir que las propiedades de ambas funciones están íntimamente vinculadas.

En particular, dada una sucesión de variables aleatorias independientes, $\{X_1, X_2, \dots, X_n\}$, se tiene que:

$$N_{X_1,X_2,\dots,X_n}(w) = N_{X_1}(w)N_{X_2}(w)\dots N_{X_n}(w) = \prod_{i=1}^{n} E[(1+w)^{X_i}]$$

Por otra parte, a partir de su definición, se obtiene que:

$$N_X^{(s)}(w) = \frac{d^s}{dw^s}E[(1+w)^X] = E\left[\frac{d^s}{dw^s}(1+w)^X\right] =$$

$$= E[X(X-1)\dots(X-s+1)(1+w)^{X-s}]$$

De modo que los momentos factoriales quedan definidos de la siguiente forma:

$$N_X^{(s)}(0) = E[X(X-1)\dots(X-s+1)] = m_{[s]}(X)$$

Sea una variable aleatoria X con función generatriz de momentos factoriales $N_X(w)$. Dadas las constantes reales a y b, se tiene que:

$$N_{aX+b}(w) = E\left[e^{(aX+b)ln(1+w)}\right] = (1+w)^b N_X[(1+w)^a - 1]$$

Ejemplo n° 51:

Sea X una variable aleatoria con distribución de probabilidades de la forma:

$$p(X) = \binom{n}{X}p^X q^{n-X} \qquad (X = 0,1,2,\dots,n; n = 1,2,\dots; p+q = 1)$$

---

[140] Ver Sec. 10.1.

Su función generatriz de momentos está dada por:

$$M_X(w) = E(e^{wX}) = \sum_{X=0}^{n} e^{wX} \binom{n}{X} p^X q^{n-X} = \sum_{X=0}^{n} \binom{n}{X} (pe^w)^X q^{n-X} =$$

$$= (q + pe^w)^n$$

De la misma forma, su función generatriz de momentos factoriales queda definida por la expresión:

$$N_X(w) = E[(1+w)^X] = \sum_{X=0}^{n} (1+w)^X \binom{n}{X} p^X q^{n-X} =$$

$$= \sum_{X=0}^{n} \binom{n}{X} [(1+w)p]^X q^{n-X} =$$

$$= [(1+w)p + q]^n = (1+pw)^n$$

de la cual se obtiene que:

$$N_X^{(s)}(w) = \frac{d^s}{dw^s}(1+pw)^n = \frac{n!}{(n-s)!}(1+pw)^{n-s}p^s \qquad (s = 1,2,\dots)$$

De modo que sus momentos factoriales quedan definidos de la siguiente forma:

$$m_{[s]}(X) = \frac{n!}{(n-s)!}p^s \qquad (s = 1,2,\dots)$$

14.4.- La función generatriz de cumulantes o semi-invariantes

Esta función puede ser definida a partir de la función característica, como $L_X(w) = \ln[C_X(w)]$. Su utilidad radica en su particular aplicabilidad a la muy importante clase de las distribuciones de probabilidades infinitamente divisibles, cuyas funciones características son de la forma e^{cw}[141] .

De la relación anterior se obtiene, obviamente, que $C_X(w) = e^{L_X(w)}$, expresión de la cual se puede concluir que, dada una sucesión de variables

[141] La definición de la función $L(w)$ se complica cuando $C(w)$ asume valores complejos o reales-negativos. Es necesario en estos casos aislar, para cada w, cada uno de dichos valores.

aleatorias, $\{X_1, X_2, \dots, X_n\}$, con funciones características $C_{X_j}(w)$ $(j = 1,2,\dots,n)$ y funciones generatrices de semi-invariantes $L_{X_j}(w)(j = 1,2,\dots,n)$ y una variable aleatoria X con función característica $C_X(w)$ y función generatriz de semi-invariantes $L_X(w)$, si se verifica que $L_{X_n}(w) \to L_X(w)$, entonces se verificará que $C_{X_n}(w) \to C_X(w)$ y, por lo tanto, que $X_n \xrightarrow{d} X$.

Sea una variable aleatoria X tal que $m_2(X) < \infty$, su función característica admite derivadas de segundo orden para todo w, de modo que $C_X'(0) = im(X)$ y $C_X''(0) = i^2 m_2(X)$. Luego, la función $L_X(w)$ es derivable en torno del origen (donde $C_X(w)$, dado que es continua y tal que $C_X(0) = 1$, es distinta de cero):

$$L_X'(w) = \frac{1}{C_X(w)} C_X'(w)$$

$$L_X''(w) = \frac{1}{C_X(w)} C_X''(w) = \frac{1}{[C_X(w)]^2} [C_X'(w)]^2$$

De estas expresiones se obtiene fácilmente que $L_X'(0) = im(X)$ y que $L_X''(0) = -\mu_2(X)$. Si la variable X posee momentos de orden superior finitos, entonces la función $L_X(w)$ admitirá derivadas de orden superior. El valor de la derivada de orden s, para $w = 0$, dejando de lado el factor i^s, define el llamado semi-invariante o cumulante de orden s.

Por otra parte, de la definición se obtiene inmediatamente que la propiedad multiplicativa para la suma de variables aleatorias independientes se transforma en una relación aditiva de la forma:

$$L_{X_1+X_2+\cdots+X_n}(w) = L_{X_1}(w) + L_{X_2}(w) + \cdots + L_{X_n}(w) =$$

$$= ln\left[M_{X_1,X_2,\dots,X_n}(w)\right]$$

La variable Normal, cuya distribución de probabilidades es de la forma:

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x - m_X}{\sigma_X}\right)^2\right] \qquad (-\infty \leq x \leq \infty)$$

Se caracteriza por poseer semi-invariantes de orden mayor o igual que 3 nulos. De esta forma, los semi-invariantes pueden ser utilizados como índices de la diferencia entre una distribución de probabilidades dada y la distribución Normal. En particular, el coeficiente:

$$\frac{L_X^{(3)}(0)}{i^3\sigma_X^3} = \frac{m_3(X) - 3m_2(X)m(X) + 2m^3(X)}{\sigma_X^3}$$

que se anula para distribuciones simétricas, puede ser utilizado como índice de asimetría y el coeficiente $\frac{L_X^{(4)}(0)}{\sigma_X^4}$ puede ser utilizado como medida de la kurtosis de la variable X.

Sea una variable aleatoria X con función generatriz de semi-invariantes $L_X(w)$. Dadas las constantes reales a y b, se tiene que $L_{aX+b}(w) = wb + L_X(wa)$.

15.- La función generatriz de probabilidades

15.1.- Definición

Sea una variable aleatoria X finita o infinita numerable que asume sólo valores enteros no-negativos y sea $p_j = p(X = j)(j = 0,1,2,\dots)$ su función de probabilidades. Para $|w| \leq 1$ se puede definir la denominada "función generatriz de probabilidades" $G_X(w) = E(w^X) = \sum_{j=0}^{\infty} w^j p_j$. Esta serie es absolutamente convergente, verificándose que $\sum_j |w|^j p_j \leq \sum_j p_j = 1$ [142].

La función generatriz de probabilidades puede ser generalizada sustituyendo las probabilidades p_j por una sucesión de números no-negativos $\{a_j, j = 0,1,2,\dots\}$, quedando definida de esta forma, la función generatriz de la sucesión $\{a_j\}$, $G(w) = \sum_{j=0}^{\infty} a_j w^j$, siempre que sea finita en un entorno del origen ($|w| < w_0$). En ese caso, si se verifica que $\sum_j a_j < +\infty$, puede decirse que queda definida una función generatriz de probabilidades, dado que los valores $\frac{a_k}{\sum_j a_j} = p_k$ constituyen formalmente una distribución de probabilidades[143].

Generalizando la definición anterior, sea $X = \{X_1, X_2, \dots, X_n\}$ una variable aleatoria n-dimensional tal que sus variables marginales sólo pueden asumir valores enteros no-negativos y cuya distribución de probabilidades conjunta es de la forma $p_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n)$ $(x_i = 0,1,2,\dots; i = 1,2,\dots,n)$. La función generatriz de probabilidades de la variable X está definida por:

---

[142] Se puede concluir en forma inmediata que $G_X(w) = N_X(w-1) = M_X[ln(w)]$.

[143] Aún en el caso en que la suma $\sum a_j$ no sea finita, si la sucesión $\{a_j, j = 0,1,2,\dots\}$ es acotada, es posible definir la función generatriz de probabilidades de la sucesión $\{a_j\}$. Si esta condición se cumple, $G_a(w)$ converge absolutamente en el intervalo $|w| \leq 1$.

$$G_{\boldsymbol{X}}(w) = G_{X_1,X_2,\ldots,X_n}(w_1, w_2, \ldots, w_n) = E\left(\prod_{j=1}^{n} w_j^{x_j}\right) =$$

$$= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(\prod_{j=1}^{n} w_j^{x_j}\right) p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$$

y la función generatriz de probabilidades de la distribución de la variable marginal $(X_1, X_2, \ldots, X_k)(k < n)$ está dada por:

$$G_{X_1,X_2,\ldots,X_n}(w_1, w_2, \ldots, w_k, 1,1, \ldots ,1) =$$

$$= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\left(\prod_{j=1}^{n} w_j^{x_j}\right) p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$$

(para $w_{k+1} = w_{k+2} = \cdots = w_n = 1$).

15.2.- Propiedades

La función $G_X(w)$ posee propiedades análogas a las de la función característica:

1) Sea una variable aleatoria X con función generatriz de probabilidades $G_X(w)$. Dadas las constantes reales a y b, se tiene que $G_{aX+b}(w) = w^b G_X(w^a)$.

2) Sean X e Y dos variables aleatorias independientes que sólo pueden asumir valores enteros no-negativos, con distribuciones de probabilidades $p(X = i) = f_i (i = 0,1,2, \ldots)$ y $p(Y = j) = g_j (j = 0,1,2, \ldots)$ y funciones generatrices de probabilidades $G_X(w)$ y $G_Y(w)$, respectivamente. Sea la variable $Z = X + Y$. Su función de probabilidades será:

$$p(Z = k) = p(X + Y = k) = \{h_k\} = \{f_i\} * \{g_j\}(k = 0,1,2, \ldots)$$

De acuerdo con la definición de función generatriz de probabilidades, se puede escribir:

$$G_X(w)G_Y(w) = \left(\sum_i w^i f_i\right)\left(\sum_j w^j g_j\right) =$$

$$= f_0 g_0 + (f_0 g_1 + f_1 g_0)w + (f_0 g_2 + f_1 g_1 + f_2 g_0)w^2 + \cdots =$$

$$= h_0 + h_1 w + h_2 w^2 \dots = G_Z(w)$$

lo que demuestra que la función generatriz de probabilidades de una suma de variables aleatorias independientes es igual al producto de las funciones generatrices de probabilidades de las variables sumando, $G_Z(w) = G_X(w)G_Y(w)$. En general, se puede demostrar que, dada una sucesión de variables aleatorias independientes, $\{X_1, X_2, \dots, X_n\}$, se verifica que $G_{X_1, +X_2+\cdots+X_n}(w) = G_{X_1}(w)G_{X_2}(w) \dots G_{X_n}(w)$.

3) Sea $\{X_1, X_2, \dots, X_n\}$ una sucesión de variables aleatorias independientes que sólo pueden asumir valores enteros no-negativos, sea una variable aleatoria X que sólo puede asumir valores enteros no-negativos y sean $G_{X_1}(w), G_{X_2}(w), \dots, G_{X_n}(w)$ y $G_X(w)$ sus respectivas funciones generatrices de probabilidades. Si X_n converge en-distribución a X, entonces se puede afirmar que $G_{X_n}(w) \to G_X(w)$ para $|w| \leq 1$ y viceversa, si la función generatriz de probabilidades $G_{X_n}(w)$ converge a $G_X(w)$, para $|w| \leq 1$, entonces la variable X_n converge en-distribución a X.

4) Derivando término a término la serie de potencias $\sum_j w^j p_j$ ($|w| \leq 1$) se obtiene que $\frac{d}{dw} G_X(w) = \sum_j j w^{j-1} p_j$. Luego:

$$\lim_{w \to 1} \frac{d}{dw} G_X(w) = \lim_{w \to 1} \sum_j j\, w^{j-1} p_j = \sum_j j p_j = E(X)$$

Este valor esperado existe siempre, pero puede ser infinito. Ahora bien, si $G'_X(w)$ es finita y continua para $w = 1$ entonces, será $\frac{d}{dw} G_X(1) = E(X) < +\infty$. Las derivadas sucesivas de $G_X(w)$ para $|w| \leq 1$, están dadas por:

$$\frac{d^s}{dw^s} G_X(w) = G_X^{(s)}(w) = \sum_{j=1}^{\infty} j(j-1) \dots (j-s+1) w^{j-s} p_j$$

Si $G_X^{(s)}(1)$ existe y es continua, entonces proporcionará el momento factorial de orden s[144], $G_X^{(s)}(1) = \sum_{j=1}^{\infty} j(j-1) \dots (j-s+1) p_j$. En particular, para $s = 2$, se tiene que:

[144] Ver Sec. 10.1.

$$G_X^{(2)}(1) = \sum_{j=2}^{\infty} j(j-1)p_j = \sum_j j^2 p_j - \sum_j j p_j = E(X^2) - E(X)$$

Luego la varianza de la variable puede ser expresada de la siguiente forma:

$$\sigma^2(X) = E(X^2) - [E(X)]^2 = [E(X^2) - E(X)] + E(X) - [E(X)]^2 =$$

$$= G_X^{(2)}(1) + G_X^{(1)}(1) - \left[G_X^{(1)}(1)\right]^2$$

5) Sea $\boldsymbol{X} = \{X_1, X_2, \dots, X_n\}$ una variable aleatoria n-dimensional tal que las variables marginales sólo pueden asumir valores enteros no-negativos y sea:

$$p_{j_1,j_2,\dots,j_n} = p[(X_1 = j_1) \cap (X_2 = j_2) \cap \dots \cap (X_n = j_n)]$$

$(j_i = 0,1,2, \dots; i = 1,2, \dots, n)$.

su función de probabilidades conjunta. De acuerdo con su definición, la función generatriz de probabilidades de la variable X está definida por:

$$G_{\boldsymbol{X}}(w) = G_{X_1,X_2,\dots,X_n}(w_1, w_2, \dots, w_n) = \sum_{j_1} \sum_{j_2} \dots \sum_{j_n} w_1^{j_1} w_2^{j_2} \dots w_n^{j_n} \, p_{j_1,j_2,\dots,j_n}$$

$(|w_1| \leq 1, |w_2| \leq 1, \dots, |w_n| \leq 1)$. De modo que:

$$\lim_{w_1,w_2,\dots,w_n \to 1} \frac{\partial}{\partial w_i} G_X(w) =$$

$$= \lim_{w_1,w_2,\dots,w_n \to 1} \sum_{j_1} \sum_{j_2} \dots \sum_{j_n} j_i w_1^{j_1} w_2^{j_2} \dots w_i^{j_i - 1} w_{i+1}^{j_{i+1}} \dots w_n^{j_n} \, p_{j_1,j_2,\dots,j_n} = E(X_i)$$

$(i = 1,2, \dots, n)$. En general, si la derivada de orden s existe y es continua para $w = 1$, se obtiene que:

$$\lim_{w_1,w_2,\dots,w_n \to 1} \frac{\partial^s}{\partial w_i^s} G_X(w) = \sum_{j_1} \sum_{j_2} \dots \sum_{j_n} j_i (j_i - 1) \dots (j_i - s + 1) \, p_{j_1,j_2,\dots,j_n} =$$

$$= m_s(X_i) \qquad (i = 1,2, \dots, n)$$

6) Sea $\boldsymbol{X} = \{X_1, X_2, \dots, X_n\}$ una variable aleatoria n-dimensional tal que las variables marginales sólo pueden asumir valores enteros no-negativos y sea $G_{\boldsymbol{X}}(w) = G_{X_1,X_2,\dots,X_n}(w_1, w_2, \dots, w_n)$ su función generatriz de probabilidades.

La función generatriz de probabilidades de la variable condicionada $[X_2, X_3 \dots, X_n \,/\, (X_1 = x_1)]$ estará dada, entonces, por:

$$G_{[X_2,X_3\dots,X_n/(X_1=x_1)]}(w_1, w_2, \dots, w_n) = \frac{\lim\limits_{w_1 \to 0} \frac{\partial^{x_1}}{\partial w_1^{x_1}} G_{X_1,X_2,\dots,X_n}(w_1, w_2, \dots, w_n)}{\lim\limits_{\substack{w_1 \to 0 \\ w_2 \to 1 \\ \dots \\ w_n \to 1}} \frac{\partial^{x_1}}{\partial w_1^{x_1}} G_{X_1,X_2,\dots,X_n}(w_1, w_2, \dots, w_n)}$$

Ejemplo n° 52:

Sea una sucesión de n fenómenos de naturaleza no-determinística, independientes entre sí y tales que cada uno de ellos admita solamente dos resultados posibles: E (éxito) –con probabilidad p- y $\bar{E}$ (fracaso) –con probabilidad $q = 1 - p$. Sea $X^{(i)}$ la variable aleatoria que representa el número de éxitos que se pueden producir al "realizarse" el i-ésimo fenómeno (obviamente esta variable puede asumir solamente los valores "1", si de verifica el resultado E ó "0", si se verifica el resultado $\bar{E}$). La función generatriz de probabilidades de esta variable será de la forma:

$$G_{X^{(i)}}(w) = E\left[w^{X^{(i)}}\right] = w^0 q + wp = q + wp$$

Por otra parte, de acuerdo con su definición serán $E\left[X^{(i)}\right] = 0q + 1p = p$ y $E\left[\left(X^{(i)}\right)^2\right] = 0q + 1^2 p = p$. Es decir, resulta que:

$$\sigma^2(X) = E\left[\left(X^{(i)}\right)^2\right] - \left[E\left[X^{(i)}\right]\right]^2 = p - p^2 = pq$$

Sea X_n la variable aleatoria que representa el número de veces que se puede presentar el resultado "éxito" al cabo de n realizaciones del evento. Esta variable puede ser definida, entonces, como una suma de variables aleatorias independientes de la forma $X_n = X^{(1)} + X^{(2)} + \dots + X^{(n)}$. Luego, su función generatriz de probabilidades será:

$$G_{X_n}(w) = G_{X^{(1)}}(w) G_{X^{(2)}}(w) \dots G_{X^{(n)}}(w) = (q + wp)^n =$$

$$= \sum_{j=0}^{n} \left[\binom{n}{j} p^j q^{n-j}\right] w^j = \sum_{j=0}^{n} p(X_n = j) w^j$$

De lo que se deduce que la distribución de probabilidades de la variable X_n será de la forma:

$$p(X_n = j) = \binom{n}{j} p^j q^{n-j} (j = 0,1,2, \dots, n)$$

El valor de las derivadas primera y segunda de la función $G_{X_n}(w)$ en el punto $w = 1$, será:

$$\frac{d}{dw} G_{X_n}(w)\Big|_{w=1} = E(X_n) = np = \sum_{i=1}^{n} E\left(X^{(i)}\right)$$

$$\frac{d^2}{dw^2} G_{X_n}(w)\Big|_{w=1} = m_{[2]}(X_n) = E[X_n(X_n - 1)] = n(n-1)p$$

De modo que la varianza queda definida de la siguiente forma:

$$\sigma^2(X_n) = G_{X_n}^{(2)}(1) + G_{X_n}^{(1)}(1) - \left[G_{X_n}^{(1)}(1)\right]^2 = npq = \sum_{i=1}^{n} \sigma^2\left(X^{(i)}\right)$$

Ejemplo n° 53:

Sea T la variable aleatoria que representa el tiempo a transcurrir hasta la aparición por primera vez de un resultado E en una sucesión de "realizaciones" independientes de un fenómeno de comportamiento eventual. La variable T asumirá el valor n si el resultado E se produce por primera vez en la n-ésima realización del proceso. De modo que su distribución de probabilidades será de la forma:

$$p(T = n) = p(\overline{E_1} \cap \overline{E_2} \cap \dots \overline{E_{n-1}} \cap E_n) = p(\overline{E_1})p(\overline{E_2}) \dots p(\overline{E_{n-1}})p(E_n) =$$

$$= q^{n-1}p \qquad (n = 1,2, \dots)$$

y su función generatriz de probabilidades está definida por:

$$G_T(w) = E(w^T) = \sum_{j=1}^{\infty} w^j p(T = j) = \sum_{j=1}^{\infty} w^j q^{j-1} p =$$

$$= wp \sum_{j=1}^{\infty} (wq)^{j-1} = \frac{wp}{1 - wq}$$

Sea ahora la variable T_n, que representa el tiempo a transcurrir hasta la aparición por n-ésima vez del resultado E. Esta variable asumirá el valor x cuando el resultado E ocurra $n - 1$ veces en las primeras $x - 1$ realizaciones y se presente por n-ésima vez en la x-ésima realización. Luego, puede ser definida como la agregación del tiempo a transcurrir hasta la primera ocurrencia del evento E ($T^{(1)}$), más el tiempo a trasnscurrir entre la primera y la segunda ocurrencia de E ($T^{(2)}$) y así sucesivamente hasta el tiempo a transcurrir entre la $(n-1)$-ésima y la n-ésima ocurrencia de E, $T_n = T^{(1)} + T^{(2)} + \cdots + T^{(n)}$. Como se ha supuesto que las realizaciones sucesivas del fenómeno son independientes, las variables $T^{(i)}$ resultan todas independientes entre sí e idénticamente distribuidas, por lo que se verificará que:

$$G_{T_n}(w) = G_{T^{(1)}}(w)G_{T^{(2)}}(w)\ldots G_{T^{(n)}}(w) = [G_T(w)]^n = \left(\frac{wp}{1-wp}\right)^n =$$

$$= (wp)^n \sum_{j=0}^{\infty} \binom{n+1-j}{j} q^j\, w^j =$$

$$= \sum_{j=0}^{\infty} \left[\binom{n+1-j}{j} p^n q^j\right] w^{n+j} = \sum_{j=n}^{\infty} \left[\binom{j-1}{n-1} p^n q^{j-n}\right] w^j$$

De lo que se concluye que la distribución de probabilidades de la variable T_n es de la forma:

$$p(T_n) = \binom{x-1}{n-1} p^n q^{x-n} \qquad (x = n, n+1, n+2, \ldots; n = 1,2,\ldots)$$

16.- Las variables aleatorias compuestas

Sea $F_j(x_1, x_2, \ldots, x_n)$ $(j = \cdots, -2, -1, 0, 1, 2, \ldots)$ una sucesión de funciones de distribución de probabilidades y sea la sucesión $\{\pi_j \geq 0\}$ tal que $\sum_{j=-\infty}^{\infty} \pi_j = 1$, entonces se dice que la operación:

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \sum_{j=-\infty}^{\infty} \pi_j F_j(x_1, x_2, \ldots, x_n)$$

(donde $F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$ también es una función de distribución de probabilidades) define una "composición" de las distribuciones $F_j(\cdot)$. Según se vio, estas funciones dependen de k coeficientes, $\theta_1, \theta_2, \ldots, \theta_k$, $F_{X_1,X_2,\ldots,X_n}[(x_1, x_2, \ldots, x_n) \,/\, (\theta_1, \theta_2, \ldots, \theta_k)]$. Ahora bien, si los coeficientes $\theta_1, \theta_2, \ldots, \theta_k$ definen una variable aleatoria k-dimensional y considerando a las π_j como probabilidades marginales de la distribución conjunta de la variable $(\theta_1, \theta_2, \ldots, \theta_k)$, la composición anterior puede ser definida como un valor esperado de la forma:

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) =$$

$$= E_{\theta_1,\theta_2,\dots,\theta_k}\{F_{X_1,X_2,\dots,X_n}[(x_1, x_2, \dots, x_n) \,/\, (\theta_1, \theta_2, \dots, \theta_k)]\} =$$

$$= \sum_{j}^{(k)} p(\theta_{1j}, \theta_{2j}, \dots, \theta_{kj})\, F_{X_1,X_2,\dots,X_n}[(x_1, x_2, \dots, x_n) \,/\, (\theta_1, \theta_2, \dots, \theta_k)]$$

(donde los subíndices que acompañan al operador E indican que el valor esperado está calculado con respecto a las variables $\theta_1, \theta_2, \dots, \theta_k$)[145]. En forma simbólica esta composición puede ser expresada como $F_X(x, \theta) \wedge_\theta p(\theta)$, donde $F_X(\cdot,\cdot)$ denota la "función compuesta" y $p(\theta)$ la "función componente".

Este concepto de composición puede ser extendido al caso en que la variable $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ se distribuya de acuerdo con una función conjunta continua $g(\theta_1, \theta_2, \dots, \theta_k)$. La función de distribución que se obtiene de la composición será, entonces, de la forma:

$$F_X(x) = E_{\theta_1,\theta_2,\dots,\theta_k}[F_X(x, \theta)] =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(\theta_1, \theta_2, \dots, \theta_k)\, F_X[x \,/\, (\theta_1, \theta_2, \dots, \theta_k)] d\theta_1 d\theta_2 \dots d\theta_k$$

Simbólicamente $F_X(x, \theta) \wedge_\theta g(\theta)$.

17.- Las variables aleatorias generalizadas

Sea una variable aleatoria X y sea $p_X(x)$ y $G_X(w)$ su función de probabilidades y su función generatriz de probabilidades, respectivamente. Supóngase que en esta última el argumento w sea reemplazado por la función generatriz $G_Y(w)$ de otra distribución de probabilidades $p_Y(y)$, entonces la función resultante, $G_X[G_Y(w)]$, constituye un desarrollo en serie de potencias de w con coeficientes no-negativos que define, a su vez, la función generatriz de una distribución de probabilidades[146] a la cual se denomina "función p_X generalizada por la distribución (generalizadora) p_Y"[147] y se denota como $p_X \vee p_Y$. En particular, si p_Y es una distribución que depende de un parámetro θ, entonces $[G_Y(w/\theta)]^j$ es su función generatriz de probabilidades con parámetro $j\theta$. Es decir $[G_Y(w/\theta)]^j = G_Y(w/j\theta)$. Entonces, se puede escribir:

---

[145] Debe tenerse en cuenta que, si el valor esperado está calculado con respecto a la totalidad de las variables marginales $\theta_1, \theta_2, \dots, \theta_k$, éstas no aparecerán en la definición de la función de distribución compuesta. En realidad, en la definición de la función compuesta aparecerán sólo los valores de θ que no varían y las variables marginales a las cuales no se les haya asignado una distribución de probabilidades.

[146] Debe tenerse en cuenta que $G_Y(1) = 1$ y que $G_X[G_Y(1)] = G_X(1) = 1$.

[147] Esta clasificación como "distribución generalizada" se debe a Feller (1968).

$$G_X[G_Y(w/\theta)] = \sum_{j=0}^{\infty} p(X=j)[G_Y(w/\theta)]^j = \sum_j p(X=j)\, G_Y(w/j\theta)$$

Expresión que define la función generatriz de probabilidades de la variable compuesta $p_Y \wedge_{\theta^*/\theta} p_X$.

Ejemplo n° 54:

Sea $X = \{X_1, X_2, \dots, X_n\}$ una variable aleatoria n-dimensional tal que las variables marginales sólo pueden asumir valores enteros no-negativos, todas con la misma distribución de probabilidades, $p(X_j = i) = f_i (j = 1,2, \dots)$ y sea $G_{X_j}(w) = G_X(w) = \sum w^i f_i\ (j = 1,2, \dots)$. Sea Y_N la variable definida por la suma $Y_N = X_1 + X_2 + \cdots + X_N$, donde N es, a su vez, una variable aleatoria independiente de las X_j con distribución de probabilidades $p(N = n) = h_n(n = 1,2, \dots)$ y cuya función generatriz de probabilidades está definida por $G_N(w) = \sum w^i h_i$. La función de probabilidades de la variable Y_n será de la forma:

$$p(Y_n = i) = p(N=1)p(X_1 = i) + p(N=2)p(X_1 + X_2 = i) =$$

$$= \sum_{n=1}^{\infty} p(N=n)p(X_1 + X_2 + \cdots + X_n = i) = q_i$$

Para un valor fijo de $N = n$ la distribución de la variable $Y_n = X_1 + X_2 + \cdots + X_n$ está dada por la convolución de orden n de la sucesión $\{f_i\}$:

$$p(Y_n = i) = p(N=n)p(X_1 + X_2 + \cdots + X_n = i) = \{h_n\}\{f_i\}^{n*}$$

Luego, será:

$$G_{Y_N}(w) = \sum_i w^i g_i = \sum_i w^i \left(\sum_{n=0}^{\infty} \{h_n\}\{f_i\}^{n*} \right) = \sum_n h_n \sum_i w^i \{f_i\}^{n*} =$$

$$= \sum_n [G_X(w)]^n h_n = G_N[G_X(w)]$$

17.1.- Una aplicación biométrica: el proceso ramificado simple o proceso de Galton-Watson

Supóngase que un individuo (que forma la generación-0 de una población) sea capaz de dar origen a X_1 $(= 0,1,2, \dots)$ descendientes (que formarán la generación-1), cada uno de los cuales sea capaz de dar origen a X_2 descendientes (que formarán la generación-2) y así sucesivamente.

Supóngase: **i)** que cada individuo de una generación dé origen a su descendencia en forma independiente de todos los demás individuos de la misma generación y **ii)** que las variables aleatorias $X_n(n = 1,2, \dots)$ sean iid con función de probabilidades:

$$p(X_1 = x) = p(X_2 = x) = \cdots = p(X = x) = p_X \quad (x = 0,1,2, \dots)$$

y función generatriz de probabilidades $G_{X_i}(w) = G_X(w) = \sum_{x=0}^{\infty} p_x w^x$. Sea $Z_n(n = 1,2, \dots)$ la variable aleatoria que representa el número de individuos que integran la n-ésima generación. Dado que se ha supuesto que la generación-0 está formada por un individuo ($Z_0 = 1$), se verificará que $p(Z_1 = x) = p(X = x) = p_x$. Teniendo en cuenta que la variable Z_2 representa el número de descendientes directos de los Z_1 individuos de la generación-1, puede ser considerada como el resultado de una suma de Z_1 variables aleatorias iid, todas con función de probabilidades p_X y cuya función generatriz de probabilidades será de la forma $G_{Z_2}(w) = G_{Z_1}[G_X(w)]$. De la misma forma, se demuestra que $G_{Z_n}(w) = G_{Z_{n-1}}[G_X(w)]$ y, considerando que la generación-n está formada por los descendientes de los Z_1 integrantes de la generación-1, se puede escribir $G_{Z_n}(w) = G_X\left[G_{Z_{n-1}}(w)\right]$. En particular, si se supone que cada individuo de la población está sometido exclusivamente al riesgo de división en dos nuevos individuos (con probabilidad p) y de fallecimiento (con probabilidad $1 - p$), entonces la distribución de probabilidades de la variable X queda definida de la siguiente forma:

$$p_x = \begin{cases} 1-p & para \quad x = 0 \\ 0 & \quad x = 1 \\ p & \quad x = 2 \\ 0 & \quad x = 3,4, \dots \end{cases}$$

y su función generatriz de probabilidades será $G_X(w) = (1-p) + pw^2$. Luego, la función generatriz de probabilidades de la variable Z_2 será $G_{Z_1}(w) = G_X(w) = (1-p) + pw^2$. Lo que confirma que:

$$p(Z_1 = x) = p_x = \begin{cases} 1-p & para \quad x = 0 \\ 0 & \quad x = 1 \\ p & \quad x = 2 \\ 0 & \quad x = 3,4, \dots \end{cases}$$

La función generatriz de probabilidades de la variable Z_2 será:

$$G_{Z_2}(w) = G_{Z_1}[(1-p) + pw^2] = (1-p) + p[(1-p) + pw^2]^2 =$$

$$= (1-p)[1+p(1-p)] + 2p^2(1-p)w^2 + p^3w^4$$

De los que se puede concluir que:

$$p(Z_2 = x) = \begin{cases} (1-p)[1+p(1-p)] & para \quad x = 0 \\ 2p^2(1-p) & x = 2 \\ p^3 & x = 4 \\ 0 & x = 5,6,\dots \end{cases}$$

De la misma forma será:

$$G_{Z_3}(w) = G_{Z_2}[G_X(w)] = (1-p) + p\{(1-p) + p[(1-p) + pw^2]^2\}^2$$

y así sucesivamente. En particular, la probabilidad de que la población –con origen en un único ancestro- en la generación-n se encuentre extinguida puede ser definida como $p(Z_n = 0) = G_{Z_n}(0)$. Ahora bien, suponiendo que $0 < p_0 = 1 - p < 1$, se verifica que $G_X(w)$ es una función monótona estrictamente creciente de w, con $0 < G_X(0) = p_0 < 1$. Luego, a partir de estas propiedades de la función $G_X(w)$ y de la relación $G_{Z_n}(w) = G_{Z_{n-1}}[G_X(w)]$ se puede escribir $G_{Z_n}(0) = G_{Z_{n-1}}[G_X(0)] > G_{Z_{n-1}}(0)$. Lo que demuestra que $p(Z_n = 0) > p(Z_{n-1} = 0)$ y, por lo tanto, que la sucesión $\{p(Z_n = 0)\}$ es monótonamente creciente con una cota superior igual a la unidad y que, en consecuencia, la probabilidad de que la población se extinga alguna vez será $\lim_{n\to\infty} p(Z_n = 0) = q$. Asimismo, a partir de la relación ya demostrada, se puede escribir:

$$G_{Z_n}(0) = p(Z_n = 0) = G_X\left[G_{Z_{n-1}}(0)\right] = G_X[p(Z_{n-1} = 0)]$$

de modo que:

$$\lim_{n\to\infty} p(Z_n = 0) = q = \lim_{n\to\infty} G_X[p(Z_{n-1} = 0)] = G_X(q)$$

Ahora bien, un breve análisis acerca del comportamiento de la función $y = G_X(w)(0 \leq w \leq 1)$ permite concluir que:

i) Si $p_0 + p_1 = 1$ entonces $G_X(w)$ es una función lineal y como $p_0 > 0$, tendrá una intersección con la recta $y = w$ en el punto $(1,1)$ y, por lo tanto, será $q = 1$.

ii) Si $p_0 + p_1 < 1$, entonces $G'_X(w)$ es una función creciente de w, la función $y = G_X(w)$ es convexa y, por lo tanto, tendrá a los sumo dos intersecciones con la recta $y = w$, una en el punto $(1,1)$. Si $G'_X(w) > 1$, existirá una segunda intersección en $q < 1$ (eventualmente la intersección podría verificarse en el punto $(0, p_0)$. Si $G'_X(w) \leq 1$, la curva $G_X(w)$ se mantendrá siempre por encima de la recta $y = w$ y, por lo tanto, no existirá ninguna intersección entre 0 y 1, es decir que $q = 1$.

Si se tiene en cuenta que $G'_X(1) = E(X)$ define el número esperado de descendientes para un individuo cualquiera de la población, las consideraciones anteriores pueden ser expresadas de la siguiente forma: **i)** si el número esperado de descendientes es mayor que uno, la probabilidad de extinción tiende a una cantidad $q < 1$; **ii)** si este valor esperado es menor o igual a la unidad, entonces la probabilidad de extinción tiende a uno.

Referencias

Allais, M. (1953a). Fondements d'une théorie positive des choix comportant un risque. Econométrie. *Colloques Internationaux du Centre National de la Recherche Scientifique.* Vol. 45, pp. 257-332.

Alt, F.L. (1942). Distributed lags. *Econometrica*, Vol. 10, pp 113-128.

Arnould, A., Nicole, P. (1662). *La logique ou l'art de penser*. Clair & Girbal.

Arrow, K. (1951). Alternative approaches to the theory of choice in risk-taking situations. *Econometrica*, Vol. 19, pp. 404-437.

Aumann, R.J. (1977). The St. Petersburg paradox. A discussion of some recent comments. *Journal of Economic Theory*, Vol.14, pp. 443-445.

Barlow, R.; Proschan, F. (1975). *Statistical theory of reliability and life testing*. Holt-Rinehart.

Barrois, Th. (1835). *Essai sur l'application du calcul des probabilités aux assurances contre l'incendie.* L. Douel.

Bauny, R.P.E. (1646). *Somme des péchéz qui se commettent en tous les états*. París.

Bentzel, R.; Wold, H. (1946). On statistical demand analysis from the viewpoint of simultaneous equations. *Skandinavisk Aktuarietidskrift*, Vol. 29, pp. 95-114.

Bernoulli,D.(1728). *Correspondencia con N. Bernoulli*. En Die Werke von Jakob Bernoulli (1975).

Bernoulli, D. (1738). Specimen theoriæ novæ de mensura sortis. *Commentarii Academiæ Scientiæ Imperialis Petropolitanæ*, Vol.5. pp. 175-192. Traducción al inglés en *Econometrica,* Vol.22, pp. 23-36, 1954.

Bernoulli, J. (1713). Ars conjectandi. En Translations from James Bernoulli, *Harvard University Department of Statistics Technical Report*. Nº2 (1966).

Bernoulli, J. (1975). *Die Werke von Jakob Bernoulli*. Birkhäuser.

Bertrand, J. (1889). *Calcul des probabilités*. París.

Bienaymé, I.J. (1853). Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrées. *Comptes Rendus de l´Académie des Sciences*, Vol. 37, pp. 309-324.

Bienaymé, I.J. (1853). Remarques sur les différences qui distinguent l'interpolation de M. Cauchy de la méthode des moindres carrées, et qui assurent la superiorité de cette méthode. *Comptes Rendus de l´Académie des Sciences*. Vol. 37, pp.5-13.

Block, H.W. (ed.) (1990). *Topics in statistical dependence*. Institute of mathematical statistics Lectures Notes.

Block, H.W.; Sampson, A.R.; Savits, T. (eds) (1990). *Topics in statistical independence*, Institute of mathematical statistics Lectures Notes. Monograph Series, Hayward, CA.

Block, H.W.; Savits, T.H.; Shaked, M. (1985). A concept of negative dependence using stochastic ordering. *Statistics and Probability Letters*, Vol. 3, pp. 81-86.

Block, H.W.; Ting, M-L. (1981). Some concepts of multivariate dependence. *Comments in Statistics A- Theory Methods*. Vol 10, pp. 749-762.

Borel, E. (1939). *Valeur pratique et philosophie des probabilités*. Gauthier-Villars.

Borel, E. (1949). Probabilité et Certitude. *Dialectica*, Vol 3, pp. 24-27.

Borel, E. (1950). *Éléments de la théorie des probabilités*. Albin Michel.

Borel, E. (1949a). Le paradoxe de Saint Pétersbourg. *Comptes Rendus de l'Académie des Sciences*. Vol. 229, pp. 404-405.

Borel, E. (1949b). Sur une propriété singulière de la limite d'une espérance mathématique. *Comptes Rendus de l'Académie des Sciences*. Vol. 229, pp. 429-431.

Borel, E. (1949c). Sur une martingale mineure. *Comptes Rendus de l'Académie des Sciences*. Vol. 229 pp. 1181-1183.

Borel, E. (1950). *Probabilité et certitude*. PUF, Paris.

Braithwaite, R.B. (ed.) (1931). *Ramsey's the foundations of mathematics and other logical essays*.

Buffon, G.L. (1777). Essai d'arithmétique moral. Publicado como suplemento del 4to.tomo de *Histoire naturelle, générale et particulière, servant de suite à l'stoire naturelle del'homme*. Imprenta Real.

Caperàa, P.; Genest, C. (1990). Concepts de dépendance et ordres stochastiques pour des lois bidimensionelles. *Canadian Journal of Statistics*, Vol. 18, Nro. 4, pp. 315-326.

Cardano, G. (1564). Liber de ludo aleæ. En *Opera omnia* (1663). Reeditado en *The book on games of chance*. Holt-Rinehart-Winston (1964).

Caritat-Condorcet, M.-J.A.N. (1784). Mémoire sur le calcul des probabilités, *Histoire de l'Académie Royale des Sciences*.

Caritat-Condorcet, M.-J.A.N. (1785a). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. París.

Caritat-Condorcet, M.-J.A.N. (1785b). Probabilité, *Encyclopédie méthodique*, Vol. 2, pp. 652-666. Panckouke.

Caritat-Condorcet, M.-J.A.N. (1789). Probabilité. En *Dictionnaire Encyclopédique des Mathématiques*. Vol. 2, pp. 649-663.

Caritat-Condorcet, M-J.A.N. (1805). *Élements du calcul des probabilités*. Royez.

Chebychev, P.L. (1867). Des valeurs moyennes. *Journal de Mathématiques Pures et Appliqués*. Vol. 12, pp.117-184.

Chebychev, P.L. (1887). Sur deux théorèmes relatives aux probabilités. *Acta Mathematica*. Vol. 14, pp. 98-116.

Chipman, J. S. (1696). The foundation of utility. *Econometrica*. Vol.28, No. 193-224.

Czuber, E. (1878). Vergleichungzweier Annamenüber die moralische Bedeutung von Geldsummen, *Archiv der Mathematik und Physic.* Vol. 62, pp. 267-284.

Czuber, E. (1882). Das Petersburger Problema. *Archiv der Mathematik un Physik*. Vol 67, pp. 1-28.

Chuprov, A.A. (1909). *Topics in the Theory of Statistics*. Moscú (original en ruso).

Cochrane, D.; Orcutt, G. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association*, Vol. 44, pp. 32-61.

Condorcet, J.N. (1784). Mémoire sur le calcul des probabilités. *Mémoires de l'Académie Royal des Sciences.* Año 1781, pp. 707-720.

Condorcet, J.N. (1785a). *Essai sur l'application de l'analyse a la probabilité des décisions rendues a la pluralité des voix*. Imprimerie Royale.

Condorcet, J.N. (1785b). *Probabilité en Encyclopédie Méthodique-Diderot & D'Alembert*. Premiers Editeurs de l'Encyclopédie.

Condorcet, J. N. (1789). *Foundations of Social Choice and Political Theory*. Traducido y editado por Iain McLean y Fioran Hewit (1994).

Condorcet, J. N. (1805). *Éléments du calcul des probabilités et son application aux jeux de hasard, a la loterie, et aux jugements des homes, avec un discours sur les avantages des mathématiques socials*. Royez.

Cramer, G. (1728). *Correspondencia con N. Bernoulli*. En Bernoulli, J. (1975), pp. 560-563.

Czuber, E. (1878). Vergleichungzweier Annamenüber die moralische Bedeutung von Geldsummen, *Archiv on Mathematik und Physic*. Vol. 62, pp. 267-284.

Czuber, E. (1882). Das Petersburger Problem. *Archiv der Mathematik un Physik*. Vol 67, pp. 1-28.

d'Alembert, J.-B.L.-R. (1768). *Opuscules mathématiques*, París.

d'Alembert, J.-B.L.-R. (1773). Doutes et questions sur le calcul des probabilités. En *Œuvres philosophiques, historiques, et littéraires de d'Alembert*, Vol. IV, Bastien (1805).

Dall'Aglio, G. (1972). Fréchet classes and compatibility of distribution functions. *Symposium of Mathematics*, Vol. IX, pp. 131-150.

Dall'Aglio, G., Kotz S., Salinetti G. (eds.) (1991). *Advances in Probability Distributions with given Marginals*. Kluwer.

Daniell, P.J. (1918). A general form of integral*. Annals of Mathematics*, Vol. 19, pp. 279-294.

Daniell, P.J. (1919). Functions of limited variation in an infinite number of dimensions. *Annals of Mathematics*, Vol. 20, pp. 30-38.

Daniell, P.J. (1920). Further Properties of the General Integral. *Annals of Mathematics*, Vol 21, pp. 203-220.

Daniell, P.J. (1921). Integral products and probability, *American Journal of Mathematics*, Vol 43, pp. 143-162.

Daston, L. (1988). *Classical probability in the enlightenment*. Princeton University Press.

David, F.N. (1962). *Games, gods and gambling*, Griffin.

Dawid, A.P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society*, Serie B, Vol. 41, pp. 278-292.

de Finetti, B. (1937). La prévision, ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, Vol. 7, pp. 1-68.

de Finetti, B. (1970). *Teoria della probabilitá*, Torino. Traducción al inglés como *Theory of probability*. Cambidge University Press.

de Montmort, P.-R. (1708). *Essai d'analyse sur les jeux de hasard*. Quillau.
de Witt, J. (1671). *Waerdie van Lyf-Renten, naer Poportie van Los-Renten*. S'Graven-Hage. Reeditado en Die Werke von Johann de Witt (1975).
Durbin, J. (1960a). The fitting of time-series models. *Review of the International Statistical Institute*, Vol. 28, pp. 233-243.
Durbin, J. (1960b). Estimation of parameters in time-series regression models. *Journal of Royal Statistical Society*, Serie B. Vol. 22, pp.139-153.
Dutka, J. (1988). On the St. Petersburg paradox. *Archive in History of Exact Sciences*, Vol. 39, pp. 13-39.
Edwards, W. (1953). Experiments on economic decision-making in gambling situations. *Econometrica*. Vol. 21, pp. 349 350.
Edwards, W. (1954a). The reliability of probability preferences. *American Journal of Psychology*. Vol. 67, pp. 68-95.
Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review*. Vol. 69, pp. 109-135.
Ellsberg, D. (1954). Classic and current notions of measurable utility. *Economic Journal*. Vol. 64, pp. 528-556.
Epstein, R.A. (1977). *The theory of gambling and statistical logic*. Academic Press.
Epstein, R.J. (1987*). A history of econometrics*. North-Holland.
Esary, J.D.; Proschan, F. (1972). Relationships among some concepts of bivariate dependence. *Annals of Mathematical Statistics*. Vol. 43 (1466-1474).
Euler, L. (1753). Vera æstimatio sortes in ludis. En *Opera omnia commentationes algebricæ ad theoriam combination numet probabilitatum pertinentes*, Vol. 7, 1923.
Fang, Z.; Joe, H. (1992). Further developments on some dependence ordering for continuous bivariate distributions. *Annals of the Institute of Statistical Mathematics*, Vol. 44, pp. 501-551.
Fechner, G.T. (1866). *Elemente der Psycophysics*. Versión en inglés, Holt-Rinehart-Winston (1966).
Fechner, G.T. (1871). Zúr experimentalen Ästhetik. *Abhandlungen der Känglich säschischen gesellchaft der Wissenschaften*. Vol. 9, pp. 555-635.
Fechner, G.T. (1906). *Send-Avesta, oder über die Dinge des Himmels und des Jenseits*. 3ra. edición, Leopold Voss.
Feller, W. (1936). Über das Gesetz der Grossen Zahlen. *Acta Litterarum ac Scientiarum Regiæ Universitatis Hungaricæ Francisco-Iosephinæ*, Vol. 8, pp. 191-201. Reeditado en *An introduction to probability theory and its applications*, Vol.1, Wiley (3ra. Edición, 1968).
Feller, W. (1945). Note on the law of large numbers and 'fair' games. *Annals of Mathematical Statistics*. Vol.16, pp. 301-304.
Feller, W. (1968). *An Introduction to Probability Theory and its applications-Volumen I*. Wiley.
Fontaine, A. (1764). Solution d'un problème sur les jeux de hasard. *Mémoires donnés à l'Académie Royales des Sciences*, pp.429-431.
Fourier, J.J. (1819). Extrait d'une mémoire sur la théorie analytique des assurances. *Annales de Chimie et de Physique*, Vol.10, pp. 177-189.
Fourier, Joseph (1822). *Théorie analytique de la chaleur*. Paris. Firmin Didot Père et Fils.

Fréchet, M. (1915). Sur l'intégrale d'une fonctionelle étendue à un ensemble abstrait. *Bulletin de la societé mathématique de France*, Vol 43, pp. 248-265.

Fréchet, M. (1930a). Sur l'extension du théorème des probabilités totales au cas d'une suite infinite d'événements. *Rendiconti del Reale Istituto Lombardo de Scienze e Letteres*. Vol. 6, pp. 899-900.

Fréchet, M. (1930b. Sur l'extension du théorème des probabilités totales au cas d'une suite infinite d'événements. *Rendiconti del Reale Istituto Lombardo de Scienze e Letteres*, Vol. 6, pp. 1059-1062.

Fréchet, M. (1935). Généralisations du théorème des probabilités totales. *Fundamenta Mathematica*, Vol. 25, pp. 379-387.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Annales Université de Lyon*, Serie Sci, Vol. 14, pp. 53-77.

Fréchet, M. (1957). Sur les tableaux de corrélation dont les marges et des bornes sont données. *Annales Université de Lyon*, Serie Sci, Vol. 20, pp.13-31.

Friedman, M.; Savage, L.J. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*. Vol. 56 pp. 279-304

Fries, J.F. (1842). *Versucheiner Kritik der Principien der Wahrscheinlichkeitsrec hnung*. F.Vieweg.

Frisch, R. (1933). *Propagation problems and impulse problems in dynamic economics*. En Economic essays in honour of Gustav Cassel. Allen-Unwin.

Frisch, R. (1934). *Statistical confluence analysis by means of complete regression systems*. Universitets Økonomiske Institut.

Frisch, R. (1936). *Time series and business cycle analysis. Economic macro dynamics*. En Report of the work done under the direction of Professor I. Wedervang, at the U. Institute of Economics, Oslo. Enero 1932-Junio 1936, Rockefeller Archive Centre.

Frisch, R. (1938). *Statistical versus theoretical relations in economic macrodynamics*. Memorandum, Oslo. Reproducido en Hendry; Morgan (1995).

Galton, F. (1888). Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society*, Vol. 45, pp.135-145.

Galton, F. (1890). Kindship and correlation, *North American Review*, Vol. 150, pp. 419-431.

Gauss, C.F. (1813). Commentationes societatis regiae scientiarum Gottenhensis resentiores 2, en *Werke 3*, pp. 123-162.

Gini, C. (1921). Sull' interpolazione di una retta quando i valori della variable indipendente sono affetti da errori accidentali. *Metron,* Vol. 1, pp. 63-82.

González, M.L.; Landro, A.H. (2016). Acerca de la interpretación del concepto de perturbación en los procesos discretos de parámetro finito. *Revista de Investigación en Modelos Matemáticos Aplicados a la Gestión y la Economía* Año 3 – N° 3, pp. 141-169.

Granger, C.W.J.; Newbold, P. (1977). *Forecasting economic time series*. Academic Press.

Guilford, J.P. (1936). *Psychometric methods*. McGraw-Hill.

Griffith, R.M. (1949). Odds adjustment by American horserace bettors. *American Journal of Psychology*. Vol. 62, pp. 290 – 294.

Hagen, O. (1979). *Towards a positive theory of preferences under risk*. En Allais, M.; Hagen, O. (eds.) (1979). Expected utility hypotheses and the Allais paradox. Reidel, Dordretch.

Haavelmo, T. (1938). The method of supplementary confluent relations. Illustrated by a study of stock prices. *Econometrica*, Vol. 6, pp. 203-218.
Haavelmo, T. (1939). Statistical testing of dynamic systems if the series observed are shock cumulants. Report of the 5th Annual Research Conference on Economics and Statistics.
Haavelmo, T. (1940a). The inadequacy of testing dynamic theory by comparing the theoretical solutions and observed cycles. *Econometrica*, Vol. 8, pp. 312-321.
Haavelmo, T. (1940b). The problem of testing economic theories by means of passive observations. *Report of the 6th Annual Research Conference on Economics and Statistics*.
Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica,* Vol. 11, pp. 1-12.
Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica, Supplement*, Vol. 12, pp. 1-115.
Hagstroem, K, -G. (1938). Pure economics as a stochastic theory. *Econometrica*, Vol. 6, p.p. 40-47.
Hahn, F.H.; Brechling, E.P.R. (eds.) (1965). *The theory of interest rates*. Oxford University Press.
Hamouda, O.; Rowley, J.C.R. (1988). *Expectation, equilibrium and dynamics.* Hemel Hempstead. St. Martin's Press.
Hazenwinkel, M. (2001). *Encyclopedia of mathematics*. Springer.
Heidelberger, M. (1987). *Fechner's determinism. From freedom to laws of chance*. En Krüger; Daston; Heidelberger (eds.).
Helly, E. (1912). Über lineare Funktionaloperationen. *Sitz. Nat. Akad. Wiss*. Vol. 121. 265-297.
Hendry, D.F.; Morgan, M.S. (1989). A re-analysis of confluence analysis. *Oxford Economic Journal*, Vol. 41, pp. 35-52.
Heyde, C.C.; Seneta, E. (1972). The simple branching process, a turning point test and a fundamental inequality. A historical note on I.J. Bienaymé. *Biometrika*, Vol. 59, n° 3, pp. 680-683.
Hoeffding, W. (1940). Nasstabinvariante Korrelationstheorie. *Schriften den Matematischen Institut, Universität Berlin,* Vol. 5, pp. 179-233. Reeditado como Scale-invariant correlation theory, en Collected works of Wassily Hoeffding.
Hoeffding, W. (1941). Nasstabinvariante Korrelationmasse für diskontinuierliche Verteilungen. *Arkiv für Matematischen Wirtschaften und Sozialforschung*, Vol. 7, pp. 49-70. Reeditado como Scale-invariant correlation measures for discontinuous distributions en Collected works of Wassily Hoeffding.
Hoeffding, W. (1941). *Collected works*. Springer-Verlag.
Hotelling, H. (1927). Differential equations subject to error. *Journal of American Statistical Association*, Vol. 22, pp. 283-314.
Huygens, C. (1657). Van Rekeningh in Seplen van Geuk. Traducido al latín como De ratiotiniis in ludo aleæ. En *Exercitationes mathematicum de F. van Schoten* (Elsevier). Reeditado en Oeuvres complètes (Nijhoff, 1920). Traducción al italiano (Dupont-Roero,1984).
Irwin, R. D. (ed) (1952). *Readings in price theory*. Chicago.
Irwin, J.O. (1967). William Allen Whitworth and a hundred years of probability. *Journal of the Royal Statistical Society*, SerieA, Vol. 130 (147-175).

Joe, H. (2001). *Multivariate models and dependence concepts.* Chapman & Hall/CRC.

Jorland, G. (1987). *The St. Petersburg paradox*. En Krüger,l.; Daston, L.J.; Heidelberger, M. (eds.), 1987.

Hume, D. (1718). *An inquiry concerning human understanding.* Reeditado por Handel (1955).

Insolera, F. (1923). *Corso di matematica finanziaria*. Società Reale Mutua di Assicurazione di Torino.

Kelly, J. (1956). A new interpretation of information rate. *The Bell System Technical Journal,* Vol. 35 (917-926).

Kendall, M.G. (1938). A new measure of rank correlation. *Biometrika,* Vol. 30 (81-93).

Keynes, J.M. (1921). *A treatise on probability*. Reeditado por MacMillan (1963).

Kimeldorf, G.; Sampson, A. (1975a). One parameter families of bivariate distributions with fixed marginals. *Communications in Statistics, Serie A. Theory and Methods*, Vol. 4, pp. 293-301.

Kimeldorf, G.; Sampson, A. (1975b). Uniform representation of bivariate distributions. *Communications in Statistics, Serie A. Theory and Methods*, Vol. 4, pp. 617-627.

Kimeldorf, G.; Sampson, A. (1978). Monotone dependence. *Annals of Statistics*, Vol. 6, pp. 895-903.

Kolmogorov, A.N. (1933). *Grundbegriffe der Wahrcheinlichkeitsrechnung*. Springer. Traducción al inglés. Foundations of the theory of probability, Chelsea, 1950.

Kolmogorov, A.N. (1941). Stationary sequences in Hilbert space. *Bulletin Moscow State University, Mathematics,* Vol. 2, pp. 1-40. (original en ruso).

Koopmans, T.C. (1937). *Linear regression analysis of economic time series*. Netherlands Economic Institute.

Koopmans, T.C. (1950). *Statistical inference in dynamic economic models*. Cowles Commission, Monografía nº 10.

Krüger, l.; Daston, L.J.; Heidelberger, M. (eds.) (1987). *The probabilistic revolution*. MIT Press.

Lacroix, S.F. (1816). *Traité élémentaire du calcul des probabilités*, Courcier.

Landro, A.H. (2010). *Acerca de la probabilidad*. Ediciones Cooperativas.

Landro, A.H.; González, M.L. (2016). *Acerca del problema de Bernoulli y la determinación del verdadero valor de una probabilidad*. Ediciones Cooperativas.

Laplace, P.S. (1774). Mémoire sur la probabilité des causes par les événements. En *Oeuvres complètes*, Gauthier- Villars (1912).

Laplace, P.S. (1795). Leçons de mathématiques données à l'Ecole Normale. En *Oeuvres complètes*, Gauthier-Villars (1912).

Laplace, P.S. (1812). *Théorie analytique des probabilités*. Reeditado por Gauthier-Villars. También en Oeuvres complètes, Gauthier-Villars (1912).

Laurent, H. (1873). *Traité du calcul des probabilités*, Gauthier-Villars.

Laurent, H. (1893). *Théorie des jeux de hasard*, Gauthier-Villars.

Leamer, E.E. (1978). *Specification searches. Ad hoc inference with non-experimental data.* Wiley.

Lebesgue, H. (1901). Sur une généralisation de l'intégrale définié. *Comptes Rendus Académie des Sciences*, Vol 132, pp. 1025-1028.
Lebesgue, H. (1904). *Leçons sur l'intégration et la recherche des fonctions primitives*. Gauthie-Villars.
Lehmann, E.L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics*, Vol. 37, pp. 1137-1153.
Liagre, J.-B.J. (1852). *Calcul des probabilités et théorie des erreurs*, Merzbach.
Luckacs, E. (1975). *Stochastic convergence*. Griffin.
Lupton, S. (1889-1890). The St. Petersburg problem. *Nature,* Vol. 41, pp. 168-169.
Magnus, J. R.; Morgan, M.S. (1987). The ET interview. Professor J.Tinbergen . *Econometric Theory*, Vol. 4 (187-209).
Maistrov, L.E. (1966). *Probability theory: A historical sketch.* Segunda edición, Academic Press (1974).
Malivaud, E. (1952). Note on von Neuman-Morgenstern's string independence axiom. *Econometrica*, Vol. 20, pp. 277-323.
Mann, H.B.; Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica*, Vol. 11, pp. 173-220.
Manne, A.S. (1952). The story independence assumption-gasoline blends and probability mixtures. *Econometrica*, Vol. 20, pp. 665-669.
Markowitz, H. (1959). *Portfolio selection*. Wiley.
Marschak, J. (1937). Utility and probability in human choice. En Report of 3rd. *Annual Research Conference on Economics and Statistics*. Cowles Commission.
Marschak, J. (1950). Rational behavior. Uncertain prospects and measurable utility. *Econometrica*, Vol. 18, pp. 604-615.
Marschak, J. (1953). *Economic measurements for policy and prediction*. En Hood; Koopmans (eds.).
Martin, R.M. (2008). The St. Petersburg paradox. *The Stanford encyclopedia of philosophy.* Http.//plato.stanford.edu/entriesparadox-stpetersburg.
Martin-Löf, A. (1985). A limit theorem which clarifies the Petersburg paradox. *Journal of Applied Probability*, Vol. 22, pp. 634-643.
Medolaghi, P. (1908). Di una nuova teoria del rischio. *Bulletin dell'Associazione degli Attuari Italiani*, Vol. 18 (20-40).
Menger, K. (1934). Das Unsicherheitsmoment in der Ertlehre. *Zeitschrift für Nationalökonomie,* Vol.4, pp. 459-485. Traducido al inglés en Shubik, M.(ed.) (1967).
Moore, H.L. (1914). *Economic cycles. Their law and cause.* Nueva York.
Moore, H.L. (1923). *Generating economic cycles*. Nueva York.
Morgan, M.S. (1987). *Statistics without probability and Haavelmo´s revolution in econometrics*. En Krüger; Gingerezer; Morgan (eds.).
Morgan, M.S. (1989). *The history of econometric ideas*. Cambridge University Press.
Nelsen, R.B. (2001). *Kendall tau metric*. En Hazebwinkel, M.
Nogee, P.; Lieberman, B. (1960). The auction value of certain risky situation. *Journal of Psychology*. Vol. 49, pp. 167-179.
Öttinger, L. (1848). Untrersuchunfgenüber die Wahrscheinlichkeitsrechnung . *Journal für die reineund Angewandte Mathematik*, Vol. 36, pp. 301-306.
Pearson, K. (1978). *The history of statistician the 17th and 18th Centuries*. Griffin.

Persons, W.M. (1916). Construction of a business barometer. *American Economic Review*, Vol. 6, pp. 739-769.

Persons, W.M. (1922-23). Correlation of time series. *Journal of American Statistical Association*. Vol. 18, pp.713-726.

Poincaré, H. (1896). *Calcul des probabilités. Leçons professées pendant le deuxième semestre*. Carré.

Poisson, S.D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile précédés des règles générales du calcul des probabilités.* Bachelier.

Preston, M.G.; Baratta, P. (1948). An experimental study of the auction-value of an uncertain outcome. *American Journal of Psychology*. Vol. 61, pp. 183-193.

Prigogine, I.; Nicolis, G. (1977). *Self-organization in non-equilibrium system, from dissipative structures to order to fluctuations*. Wiley.

Qin, D. (1993). *The formation of econometrics. A historical perspective.* Clarendon.

Qin, D.; Gilbert, C.L. (2001). The error term in the history of time series econometric. *Econometric Theory*, Vol. 17, pp. 424-450.

Quetelet, A.J. (1835). *Sur l'homme et le développement des facultés*. Bachelier.

Radon, J. (1913). Theorie und Anwendungen der absolute additive Mengenfunktionen. *Sitzungsberiche der kaiserlichen Akaderie der Wissenschaften*, Vol 122, pp. 1295-1438.

Ramsey, F. (1931). *Truth and probability*. En Braithwaite (ed.).

Reichenbach, H. (1935). *Wahrscheinlichkeitslehre*. Sijthoff's.

Rowley, J.C.R.; Hamouda, O. (1987). Troublesome probability and economics. *Journal of Post Keynesian Economics*, Vol. 10, pp. 44-64.

Ruscheidorf, L.; Schweizer, B.; Taylor, M.D. (eds.) (1996). *Distributions with fixed marginals and related topics.* Institute of Mathematical Statistics.

Sainsbury, R.M. (1995*). Paradoxes*. Cambridge University Press.

Samueli, J.-J.; Boudenot, J.-C. (2009). *Une histoire des probabilités des origines à 1900*. Ellipses.

Samuelson, P. (1942). *Constancy of the marginal utility of income*. En Lange, O. McIntyre, F. Yntema, T. (eds.), Studies in Mathematical Economics and Econometrics. University of Chicago Press. pp. 75-91.

Samuelson, P.A. (1952). *Utility, preference and probability*. En Conference sur 'Les fondements et applications de la théorie du risque en econométrie'. Reeditado en Stiglitz, J. (ed.) (1966).

Samuelson, P. (1977). St. Petersbug paradoxes. Defanged, dissected and historically described*. Journal of Economic Literature*. Vol. 15, pp. 24-55.

Sargan, J.D. (1959). The estimation of relationships with autocorrelated residuals by the use of instrumental variables. *Journal of Royal Statistical Society, Serie B*, Vol. 21, pp. 91-105.

Sargan, J.D. (1961). The maximum likelihood estimation of economic relationships with autorregressive residuals. *Econometrica,* Vol. 29, pp. 414-426.

Sargent, T.J. (1976). A classical macroeconomic model for the United States. *Journal of Political Economy*, Vol. 84, pp. 207-237.

Sargent, T.J. (1977). The demand for money during hyperinflations under rational expectations. *International Economic Review*, Vol. 18, pp. 59-82.

Sargent, T.J.; Sims, C.A. (1977). *Business cycle modelling without pretending to have too much 'a priori' economic theory*. En Sims, C.A. (Ed.).

Savage, L.J. (1954). *The foundations of statistic*. Wiley.
Schriever, B.F. (1986). *Order dependence*. CWI Tract 20, Centre for Mathematics and Computer Sciences.
Schriever, B.F. (1987). An ordering for positive dependence. *Annals of Statistics*, Vol. 15, pp. 1208-1214.
Schultz, H. (1925). The statistical law of demand. *Journal of Political Economy*, Vol. 33, pp. 481-504 y 577-637.
Schultz, H. (1930). *The meaning of statistical demand curves.* Veröffentlichungen der Frankfurter Gessellschaft für Konjunkturforaschung, Leipzig.
Schultz, H. (1939). *The theory and measurement of demand*. Chicago.
Schumpeter, J. (1939). *Business cycles.* Nueva York.
Schweizer, B. (1991). *Thirty years of copulas*. En Dall Aglio G, Kotz S. Salinetti G. (eds.).
Schweizer, B., Sklar, A. (1974). Operations on distribution functions not derivable from operations on random variables. *Studia Math*. Vol 52, pp. 43–52.
Seeley, R.T. (1961). Fubini implies Leibniz implies $F_{(y,X)}=F_{(X,Y)}$. *American Mathematical Monthly*, Vol. 68, pp. 56-57.
Sent, E.-M. (1998). *The evolving rationality of rational expectations. An assessment of Thomas Sargent's achievements*. Cambridge University Press.
Shapley, L.S. (1972). The Petersburg paradox. *Rand Report.* P4940.
Sheynin, O.B. (1972). Daniel Bernoulli's work on probability. *Rete*, Vol. 1, pp. 273-300.
Shubik, M. (ed.) (1967). *Essays in mathematical economics in honor of Oskar Morgenstern*. Princeton University Press.
Sims, C.A. (ed.) (1977). *New methods in business cycle research.* Federal Reserve Bank of Minneapolis.
Sims, C.A. (1980). Macroeconomic and reality. *Econometrica*, Vol. 48, pp. 1-48.
Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Institut de Statistique de l'Université de Paris*, Vol. 8 (229-231).
Sklar, A. (1973). Random variables, joint distributions and copulas. *Kybernetica,* Vol. 9, pp. 449-460.
Sklar ,A. (1996). *Random variables, distribution functions, and copulas*. En Rüscherdorf; Schweizer; Taylor (eds.).
Slutsky, E. (1925). Über stochastich Asymptoten und Grezenzwerte. *Metron*, Vol. 5, pp. 3-89.
Slutsky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica,* Vol. 5, pp.105-146 (original en ruso, 1927).
Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, Vol. 15 (72-101).
Spohn, W. (1980). Stochastic independence, causal independence and shieldability. *Journal of Philosophical Logic*, Vol 9, pp. 73-79.
Sprowls, R.C. (1953). Psychological-mathematical probability in relationships of lottery gambles. *American Journal of Psychology*. Vol. 66, pp. 126-130.
Staehle, H. (1933). Henry L. Moore and statistical economics. *Econometrica*, Vol. 1, pp.73-86.
Starmer, C. (2000). Developments in non-expected utility theory. The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, Vol. 38, pp. 332-341.

Steinhaus, H. (1923). Les probabilités dénombrables et leur rapport à la théorie de la mesure. *Fundamenta Mathematicae*, Vol. 4, pp. 286-310.
Steinhaus, H. (1930a). Über die Wahrscheinlichkeit Dafür Da Der Konnergenskreis Einer Potenzreibe Ihre natüralische Grenze Ist. *Mathematische Zeitschrift*, Vol. 31, 408-416.
Steinhaus, H. (1930b). Sur la probabilité de la convergence de séries. *Studia Mathematica*. Vol 2, pp. 21-39.
Stigler, G.J. (1965). *Essays in the history of economics*. University of Chicago Press.
Stigler, S.M. (1980). *American contributions to mathematical statistics in the 19th century*. Cambridge University Press.
Stigler, S.M. (1986). *The history of statistics. The measurement of uncertainty before 1900*. Harvard University Press.
Stigler, S.M. (1989). Francis Galton's account of the invention of correlation. *Statistical Science*, Vol. 4, pp. 73-86.
Stiglitz, J.E. (ed.) (1966). *Collected scientific papers of Paul A. Samuelson*. MIT Press.
Strotz, R.H. (1953). Cardinal utility. *American Economic Review*. Vol. 43, pp. 384-397.
Székely, G.J.; Richards, D. St. P. (2005). Remain steadfast with the St. Petersburg paradox to quantify irrational exuberance. *The American Statistician*, Vol. 59, pp. 235-239.
Theil, H. (1957). Specification errors and the estimation of economic relationships. *Review of International Statistical Institute*, Vol. 25, pp.41-51.
Theil, H. (1958). *Economic forecast and policy*. North-Holland.
Tijms, H.C. (2004). *Understanding probability. Chance, rules in everyday life*. Cambridge University Press.
Tillotson, J. (1664). Wisdom of being religious. En *The works of the Most Reverend John Tillotson*, Edimburgo- Glasgow (1748).
Timerding, H.E. (1902). Die Bernoullische Wertheorie. *Zeitschrift für Mathematik und Physik*, Vol. 47, pp. 321-354.
Tinbergen, J. (1935). Annual survey. Suggestion on quantitative business cycle theory. *Econometrica*, Vol. 3, pp. 241-308.
Tinbergen, J. (1936). *Grondproblemen der Theoretische Statistiek*. Bohn.
Tinbergen, J. (1937). *Econometric approach to business cycles*. París.
Tinbergen, J. (1938). On the theory of business-cycle control. *Econometrica*, Vol. 6, pp. 22-29.
Tinbergen, J. (1939). *Statistical testing of business-cycle theories*. Ginebra.
Tintner, G. (1938a). A note on economic aspects to the theory of errors in time series. *Quarterly Journal of Economics*, Vol. 53, pp.141-149.
Tintner, G. (1938b). The maximization of utility over time. *Econometrica*, Vol. 6, pp.154-158.
Tobin, J. (1958). Liquidity preference as behavior toward risk. Review of Economic Studies, Vol. 25, pp.65-86.
Tobin, J. (1965). *The theory of portfolio selection*. En Hahn, F.H.; Brechling, E.P.R. (eds.) (1965).
Todhunter, I. (1865). *History of the mathematical theory of probability*. Cambridge University Press. Reeditado por Chelsea (1961).

Tversky, A. (1975). A critique of expected utility theory: Descriptive and normative considerations, *Erkennis*. Vol. 9, pp. 163-173.

von Kries, J. (1886). *Die Principien der Wahrscheinlichkeitsrechnung*, Mohr.

von Mises, R. (1928). *Wahrscheinlichkeit, Statistik und Wahrhei*. Springer.

von Neumann, J.; Morgenstern, O. (1944). *Theory of games and economic behavior*. Tercera edición, Princeton University Press (1953).

Walker, G. (1931). On periodicity in series related terms. *Philosophical Transactions of the Royal Society, Serie A*, Vol. 131, pp.519-532.

Weirich, P. (1984). The St. Petersburg gamble and risk. *Theory and decision.* Vol. 17, pp.193-202.

Whitworth, W.A. (1901). *Choice and chance*. 5ta. edición, Deighton-Bell.

Wicksell, K. (1907). *Krisernas Gåta*. Statøkonomist Tedsskrift.

Wiener, N. (1920). The mean of a functional of arbitrary elements. *Annals of Mathematics*, Vol 22, pp. 66-72.

Wiener, N. (1921a). The average of an analytical functional. *Proceedings of the National Academy of Science USA*. Vol 7, pp. 253-260.

Wiener, N. (1921b). The Average of an Analytic Functional and the Brownian Movement. *Proceedings of the National Academy of Science USA*. Vol 7, pp, 294-298.

Wiener, N. (1923). Differential-space. *Studies in Applied Mathematics*. Vol. 2, pp. 171-174.

Wiener, N. (1924). The average value of a functional. *Proceedings. London Mathematical society*. Vol.22, pp. 454-467.

Wold, H. (1938). *A study in the analysis of stationary time series.* Almqvist-Wiksells.

Wold, H. (1948). One prediction in stationary time series. *Annals of Mathematical Statistics*, Vol. 19, pp. 558-567.

Wold, H. (1949). Statistical estimation of economic relationships. *Econometrica*. Vol. 17, pp.1-22.

Wold, H. (1951). Dynamic systems of the recursive type-Economic and statistical aspects. *Sankhyá*, Vol. 11, pp. 205-216.

Wold, H. (1954). Causality and econometrics. *Econometrica*. Vol. 22, pp.162-177.

Yaary, M.E. (1987). The dual theory of choice under risk. *Econometrica*, Vol. 55, pp. 95-115.

Yanagimoto, T. (1990). *Dependence ordering in statistical models and other notions*. En Block; Sampson; Savits (eds).

Yanagimoto, T.; Okamoto, M. (1969). Partial ordering of permutations and monotonicity of a rank correlation statistic. *Annals of the Institute of Statistical Mathematics,* Vol. 21, pp. 489-506.

Yule, G.U. (1921). On the time-correlation problems, with special reference to the cvariate-difference correlation method*. Journal of Royal Statistical Society*. Vol. 84, pp. 495-526.

Yule, G. U. (1926). Why do we sometimes get nonsense correlations between time-series? A study in sampling and the nature of time-series. *Journal of Royal Statistical Society.* Vol. 89, pp. 1-64. Reeditado en Stuart; Kendall (1971).

Yule, G. U. (1927). On a method investigating periodicities in disturbed series, with special application to Wolfert's sun spot numbers. *Philosophical Transactions of the Royal Society, Serie A*. Vol. 226, pp. 267-298.

Índice

D

E

F