

# Contents

<b>1</b>	<b>Variables aleatorias bidimensionales o de mayor dimension</b>	<b>1</b>
1.1	Variables aleatorias bidimensionales . . . . .	1
1.2	Distribucion de probabilidades marginales y condicionales . . . . .	3
1.2.1	Probabilidad Condicional . . . . .	4
1.3	Variables aleatorias independientes . . . . .	5
1.4	Funciones de una variable aleatoria . . . . .	6
1.4.1	Esperanza matematica . . . . .	7
1.4.2	Varianza . . . . .	8
1.5	Coficiente de correlacion . . . . .	9
1.6	Desigualdad de Chebyshev . . . . .	10
1.7	Distribucion de la suma de variables aleatorias . . . . .	11
1.7.1	Propiedades reproductivas . . . . .	11
1.8	La ley de los grandes numeros . . . . .	12
1.9	Aproximación normal a la distribución binomial . . . . .	14
1.10	El teorema del límite central . . . . .	15

**Universidad Nacional de Rosario**

FACULTAD DE CIENCIAS EXACTAS, INGENIERÍA Y AGRIMENSURA

# PROBABILIDAD Y ESTADÍSTICA

## *Unidad 6*

Autor del resumen:

DEMAGISTRIS, Santiago Ignacio

Julio 2020

# 1 Variables aleatorias bidimensionales o de mayor dimension

## 1.1 Variables aleatorias bidimensionales

En nuestro estudio de las variables aleatorias hemos considerado, hasta aquí, sólo el caso unidimensional. Es decir, el resultado del experimento se podía registrar como un solo número  $x$ . En muchos casos, sin embargo, nos interesa observar simultáneamente dos o mas características numéricas. Por ejemplo, podríamos observar la cantidad de lluvia total,  $LL$ , y el promedio de temperatura,  $T$ , en cierta región durante un mes específico, que daría lugar al resultado  $(ll, t)$ .

**Definición.** Sea  $\epsilon$  un experimento y  $S$  un espacio muestral asociado con  $\epsilon$ . Sean  $X = X(s)$  y  $Y = Y(s)$  dos funciones que asignan un número real a cada uno de los resultados  $s \in S$ . Llamamos a  $(X, Y)$  **variable aleatoria bidimensional** (que también se denomina vector aleatorio).

Si  $X_1 = X_1(s), X_2 = X_2(s), \dots, X_n = X_n(s)$ , son  $n$  funciones, cada una de las cuales asigna un número real a cada resultado  $s \in S$ , entonces llamamos a  $(X_1, \dots, X_n)$  una **variable aleatoria n-dimensional** (o un vector aleatorio n-dimensional).

**Definición.**

- $(X, Y)$  es una **variable aleatoria bidimensional discreta** si los valores posibles de  $(X, Y)$  son finitos o infinitos numerables. Es decir, los valores posibles de  $(X, Y)$  se pueden representar como  $(x_i, y_j), i = 1, 2, \dots, n, \dots; j = l, 2, \dots, m, \dots$
- $(X, Y)$  es una **variable aleatoria bidimensional continua** si  $(X, Y)$  puede tomar todos los valores en un conjunto no numerable del plano euclidiano. [Por ejemplo, si  $(X, Y)$  toma todos los valores en el rectángulo  $\{(x, y) | a \leq x \leq b, c \leq y \leq d\}$ , diríamos que  $(X, Y)$  es una variable aleatoria bidimensional continua.]

En otras palabras,  $(X, Y)$  es una variable aleatoria bidimensional si representa el resultado de un experimento aleatorio en el cual hemos medido las dos características numéricas  $X$  y  $Y$ .

### Definición.

- a) Sea  $(X,Y)$  una variable aleatoria bidimensional discreta. Con cada resultado posible  $(x_i, y_j)$ , asociamos un número  $p(x_i, y_j)$  que representa  $P(X = x_i, Y = y_j)$  y que satisface las condiciones siguientes:

$$\forall_{x_i, y_j} p(x_i, y_j) \geq 0$$

$$\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} p(x_i, y_j) = 1$$

La función  $p$  definida para toda  $(x_i, y_j)$  en el recorrido de  $(X,Y)$  se llama **función de probabilidad** de  $(X,Y)$ . El conjunto de ternas  $(x_i, y_j, p(x_i, y_j))$ ,  $i, j = 1, 2, \dots$ , en algunos casos se denomina **distribución de probabilidades** de  $(X, Y)$ .

- b) Sea  $(X,Y)$  una variable aleatoria bidimensional continua que toma todos los valores en una región  $R$  del plano euclidiano. La **función de densidad de probabilidades conjuntas**  $f$  es una función que satisface las siguientes condiciones:

$$\forall_{(x,y) \in R} f(x, y) \geq 0$$

$$\int \int_R f(x, y) \, dx \, dy = 1$$

### Observaciones

- a) La ultima condicion indica que el volumen total bajo la superficie dada por la ecuación  $z = f(x, y)$  es igual a 1.
- b) Como en el caso unidimensional,  $f(x, y)$  no representa la probabilidad de nada. Sin embargo, para  $\Delta x$  y  $\Delta y$  positivos y suficientemente pequeñas,  $f(x, y)\Delta x\Delta y \sim P(x \leq X \leq x + \Delta x, y \leq Y \leq y + \Delta y)$ .
- c) Como en el caso unidimensional, adoptaremos la convención de que  $f(x, y) = 0$  si  $(x, y) \notin R$ . Por tanto, podemos considerar  $f$  definida para toda  $(x, y)$  en el plano y la ultima condición se convierte en  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1$ .
- d) Si  $B$  está en el recorrido de  $(X, Y)$ , tenemos

$$P(B) = P[(X(s), Y(s)) \in B] = P[s \in S | (X(s), Y(s)) \in B]$$

Esta ultima probabilidad se refiere a un evento en  $S$  y, por tanto, determina la probabilidad de  $B$ . En nuestra terminología previa,  $B$  y  $[s \in S | (X(s), Y(s)) \in B]$  son **eventos equivalentes**.

- Si  $B$  está en el recorrido de  $(X,Y)$ , siendo esta ultima discreta, tenemos

$$P(B) = \sum \sum_B p(x_i, y_j)$$

, la suma se toma con todos los índices  $(i, j)$  para los cuales  $(x_i, y_j) \in B$

- Si  $(X,Y)$  es continua

$$P(B) = \int \int_B f(x, y) \, dx \, dy$$

### Ejemplos p. 125-128

**Definición.** Sea  $(X,Y)$  una variable aleatoria bidimensional. Su función de distribución acumulada (fda)  $F$  está definida como

$$F(x,y) = P(X \leq x, Y \leq y)$$

**Observación.** Si  $F$  es fda de una variable aleatoria bidimensional con fdp conjunta  $f$ , entonces

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y)$$

dondequiera que  $F$  sea diferenciable.

## 1.2 Distribución de probabilidades marginales y condicionales

Con cada variable aleatoria bidimensional  $(X,Y)$  asociamos dos variables aleatorias unidimensionales llamadas  $X$  y  $Y$ , respectivamente. Es decir, podemos interesarnos por la distribución de probabilidades de  $X$  o por la distribución de probabilidades de  $Y$ . Dada una variable aleatoria bidimensional  $(X,Y)$  obtenemos las respectivas **distribuciones marginales** (de  $X$  y de  $Y$ ) de la siguiente manera:

- a) **Caso Discreto.** Puesto que  $X = x_i$  debe ocurrir con  $Y = y_j$ , para alguna  $j$ , y puede ocurrir solo con una  $j$  tenemos:

$$p_x(x_i) = P_x(X = x_i) = P_{(x,y)}(X = x_i, Y = y_1 \circ X = x_i, Y = y_2 \circ \dots) = \sum_{j=1}^{\infty} p_{(x,y)}(x_i, y_j).$$

La función  $p_x$  definida para  $x_1, x_2, \dots$ , representa la **distribución marginal de probabilidades de  $X$** . Análogamente definimos  $p_y(y_j) = P_y(Y = y_j) = \sum_{i=1}^{\infty} p_{(x,y)}(x_i, y_j)$ . como la **distribución marginal de probabilidades de  $Y$** .

- b) **Caso Continuo.** Sea  $f$  la fdp conjunta de la variable aleatoria bidimensional continua  $(X,Y)$ . Definimos  $f_x$  y  $f_y$ , las funciones de densidad de probabilidades marginales de  $X$  y  $Y$ , respectivamente, como sigue

$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)dy; \quad f_y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

Estas fdp corresponden a las fdp básicas de las variables aleatorias unidimensionales  $X$  y  $Y$ , respectivamente. Por ejemplo,

$$P_x(c \leq X \leq d) = P_{(x,y)}[c \leq X \leq d, -\infty \leq Y \leq \infty] = \int_c^d \int_{-\infty}^{\infty} f(x,y)dydx = \int_c^d f_x(x)dx$$

**Ejemplos p. 129-130**

**Definición.** Decimos que una variable aleatoria bidimensional continua está **distribuida uniformemente** en una región  $R$  en un plano euclidiano si

$$f(x, y) = \begin{cases} k & \text{si } (x, y) \in R \\ 0 & \text{si } \text{otherwise} \end{cases}$$

Debido a la condición  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$ , lo anterior implica que la constante es igual a  $\frac{1}{\text{area}(R)}$ . Suponemos que  $R$  es una región con Área finita distinta de cero.

## Ejemplos p. 131-132

### 1.2.1 Probabilidad Condicional

Distribución de probabilidades condicionales.

**Caso discreto.** Sea  $(X, Y)$  una variable aleatoria bidimensional.

$$P_x(x_i|y_i) = P_x(X = x_i|Y = y_i) = \frac{p(x_i, y_i)}{p_y(y_i)}, \text{ si } p_y(y_i) > 0$$

$$P_y(y_i|x_i) = P_y(Y = y_i|X = x_i) = \frac{p(x_i, y_i)}{p_x(x_i)}, \text{ si } p_x(x_i) > 0$$

**Observación.** Para una  $j$  dada,  $p(x_i|y_j)$  satisface todas las condiciones de una distribución de probabilidades. Tenemos  $p(x_i|y_j) \geq 0$  y también

$$\sum_{i=1}^{\infty} p(x_i, y_j) = \sum_{i=1}^{\infty} \frac{p(x_i|y_j)}{p_y(y_j)} = \frac{p_y(y_j)}{p_y(y_j)} = 1$$

**Caso continuo. Definición.** Sea  $(X, Y)$  una variable aleatoria bidimensional continua con fdp conjunta  $f$ . Sean  $f_x$  y  $f_y$  las fdp marginales de  $X$  y  $Y$ , respectivamente

La fdp **condicional** de  $X$  para  $Y = y$  dada, está definida por

$$f_x(x|y) = \frac{f(x, y)}{f_y(y)}, \quad h(y) > 0$$

$$f_y(y|x) = \frac{f(x, y)}{f_x(x)}, \quad g(x) > 0$$

### Observaciones.

- a) Las fdp condicionales anteriores satisfacen todas las exigencias de una fdp unidimensional. Así, para y fija, tenemos  $f_x(x|y) \geq 0$  y

$$\int_{-\infty}^{\infty} f_x(x|y)dx = \int_{-\infty}^{\infty} \frac{f(x,y)}{f_y(y)}dx = \frac{1}{f_y(y)} \int_{-\infty}^{\infty} f(x,y)dx = \frac{f_y(y)}{f_y(y)} = 1$$

- b) Supongamos que (X, Y) representan la altura y el peso de una persona, respectivamente. Sea f la fdp conjunta de (X,Y) y sea  $f_x$  la fdp marginal de X (sin tomar en cuenta Y). Por tanto,  $\int_{5.8}^6 f_x(x)dx$  representaría la probabilidad del evento  $(5.8 \leq X \leq 6)$  sin tomar en cuenta el peso Y. Por su parte  $\int_{5.8}^6 f_x(x|150)dx$  se interpretaría como  $P(5.8 \leq X \leq 6|Y = 150)$ . Estrictamente hablando, esta probabilidad condicional no está definida en los términos de nuestra convención previa con la probabilidad condicional, puesto que  $P(Y = 150) = 0$ . Sin embargo, sólo usamos la integral anterior para definir esta probabilidad. Sobre una base intuitiva, ciertamente, éste debe ser el significado de dicho número.

### Ejemplo p. 134

## 1.3 Variables aleatorias independientes

Tal como definimos el concepto de independencia entre dos eventos A y B, ahora definiremos las variables aleatorias independientes. Lo que queremos decir intuitivamente es que X y Y son variables aleatorias independientes si el resultado de X, digamos, de ninguna manera influye en el resultado de Y. Ésta es una noción extremadamente importante y hay muchas situaciones en que dicha suposición se justifica.

### Definicion.

- a) Sea (X,Y) una variable aleatoria bidimensional discreta. Decimos que X y Y son variables aleatorias independientes si y sólo si  $p(x_i, y_j) = p_x(x_i)p_y(y_j)$  para toda i y j. Esto es,  $P_{(x,y)}(X = x_i, Y = y_j) = P_x(X = x_i)P_y(Y = y_j)$ , para toda i y j.
- b) Sea (X, Y) una variable aleatoria bidimensional continua. Decimos que X y Y son variables aleatorias independientes si y sólo si  $f(x, y) = f_x(x)f_y(y)$  para toda (x, y), donde f es la fdp conjunta, y  $f_x$  y  $f_y$  son las fdp marginales de X y Y, respectivamente.

### Teorema

- a) Sea (X, Y) una variable aleatoria bidimensional discreta. Entonces X y Y son independientes si y sólo si  $p_x(x_i|y_j) = p_x(x_i)$  para toda i y j (o lo que es equivalente, si y solo si  $p_y(y_j|x_i) = p_y(y_j)$  para toda i y j).
- b) Sea (X, Y) una variable aleatoria bidimensional continua. Entonces X y Y son independientes si y solo si  $f_x(x|y) = f_x(x)$ , o lo que es equivalente, si y sólo si  $f_y(y|x) = f_y(y)$  para toda (x,y)

### Ejemplos p. 135-136

**Observacion.** De la definición de la distribución marginal de probabilidades (en cualesquiera de los casos, discreto o continuo) es claro que la distribución conjunta de probabilidades determina, en forma única, la distribución marginal de probabilidades. Es decir, que conociendo la fdp  $f$  conjunta podemos obtener las fdp marginales  $f_x$  y  $f_y$ . Sin embargo, lo inverso no es verdadero. Es decir, en general, el conocimiento de las fdp marginales  $f_x$  y  $f_y$  no determina la fdp conjunta  $f$ . Esto sólo es verdadero cuando  $X$  y  $Y$  son independientes, porque en este caso tenemos  $f(x, y) = f_x(x)f_y(y)$ .

**Teorema.** Sea  $(X, Y)$  una variable aleatoria bidimensional. Sean  $A$  y  $B$  eventos cuya ocurrencia (o no ocurrencia) depende sólo de  $X$  y  $Y$  respectivamente. (Esto es,  $A$  es un subconjunto de  $R_x$ , el recorrido de  $X$ , mientras que  $B$  es un subconjunto de  $R_y$ , el recorrido de  $Y$ .) Entonces, si  $X$  y  $Y$  son variables aleatorias independientes, tenemos  $P(A \cap B) = P(A)P(B)$ .

**Dem. p. 137**

## 1.4 Funciones de una variable aleatoria

Al definir una variable aleatoria  $X$  señalamos enfáticamente que  $X$  es una función definida del espacio muestral  $S$  a los números reales. Al definir una variable aleatoria bidimensional  $(X, Y)$ , tratamos un par de funciones,  $X = X(s)$ ,  $Y = Y(s)$ , cada una de las cuales está definida en el espacio muestral de algún experimento y cada una de las cuales asigna un número real a cada  $s \in S$ , produciendo así el vector bidimensional  $[X(s), Y(s)]$ .

Consideremos ahora  $Z = H_1(X, Y)$ , una función de las dos variables aleatorias  $X$  y  $Y$ . Debe ser claro que  $Z = Z(s)$  es otra vez una variable aleatoria. Consideremos la siguiente sucesión de pasos.

- a) Se realiza el experimento  $\epsilon$  y se obtiene el resultado  $s$
- b) Se evalúan los números  $X(s)$  y  $Y(s)$ .
- c) Se evalúa el número  $Z = H_1[X(s), Y(s)]$

El valor de  $Z$  depende claramente de  $s$ , el resultado original del experimento. Esto es,  $Z = Z(s)$  es una función que asigna a cada resultado  $s \in S$  un número real,  $Z(s)$ . Por lo tanto,  $Z$  es una variable aleatoria. Algunas variables aleatorias importantes que nos van a interesar son  $X + Y$ ,  $XY$ ,  $X/Y$ ,  $\min(X, Y)$ ,  $\max(X, Y)$ , etcétera.

**Ejemplo ilustrador p. 138**



### 1.4.1 Esperanza matematica

Sea  $(X_1, X_2, \dots, X_n)$  una variable aleatoria n-dimensional, se define a  $Z = H(X_1, X_2, \dots, X_n)$  como una funcion real de dicha variable, donde H es una funcion real definida en  $D \subset R^n$ . Por lo tanto, Z es una variable aleatoria (unidimensional) y definimos  $E(Z)$  como sigue:

- a) **Caso Discreto.** Si Z es una variable aleatoria discreta con valores posibles  $z_1, z_2, \dots$  y con  $p(z_i) = P(Z = z_i)$

$$E(Z) = \sum_{i=1}^{\infty} z_i p(z_i)$$

- b) **Caso Continuo.** Si z es una variable aleatoria continua con fdp f, tenemos

$$E(Z) = \int_{-\infty}^{\infty} z f(z) dx$$

**Teorema.** Sea  $(X_1, X_2, \dots, X_n)$  una variable aleatoria n-dimensional, y sea  $Z = H(X_1, X_2, \dots, X_n)$

- a) Si  $(X_1, X_2, \dots, X_n)$  es una variable aleatoria discreta y si  $p(x_{1i}, \dots, x_{nk}) = P(X_1 = x_1, \dots, X_n = x_n)$  tenemos

$$E(Z) = \sum_{(x_1, x_2, \dots, x_n) \in R(X_1 \times \dots \times X_n)} H(x_1, x_2, \dots, x_n) p(x_1, x_2, \dots, x_n)$$

- b) Si  $(X_1, X_2, \dots, X_n)$  es una variable aleatoria continua con fdp conjunta f, tenemos

$$E(Z) = \left( \int_{-\infty}^{\infty} \right)^n H(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n) dx_1 \dots dx_n$$

**Ejemplo bidimensional p. 168.**

**Propiedades.**

- a) Sea  $(X, Y)$  una variable aleatoria bidimensional con una distribución de probabilidades conjunta. Sean  $Z = H_1(X, Y)$  y  $W = H_2(X, Y)$ . Entonces,  $E(Z + W) = E(Z) + E(W)$ . **D/ p. 169**
- b) Sean X y Y dos variables aleatorias cualesquiera. Entonces,  $E(X + Y) = E(X) + E(Y)$ . (se deduce de la propiedad (a) al hacer  $H_1(X, Y) = X$  y  $H_2(X, Y) = Y$ )
- c) Si  $X = C$ , donde C es una constante, entonces  $E(X) = C$ . **D/ p. 169**
- d) Supongamos que C es una constante y X es una variable aleatoria. Entonces,  $E(CX) = CE(X)$ . **D/ p. 169**
- e) Sean  $X_1, X_2, \dots, X_n$  n variables aleatorias. Entonces,  $E(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n E(X_i)$
- f) Sea  $(X, Y)$  una variable aleatoria bidimensional y supongamos que X y Y son **independientes**. Entonces  $E(XY) = E(X)E(Y)$ . **D/ p.171**

### Observaciones

- a) Combinando las propiedades b,c,d, se observa que si  $Y = aX + b$ , donde  $a$  y  $b$  son constantes, entonces  $E(Y) = aE(X) + b$ . En palabras, la esperanza de una función lineal es esa misma función lineal de las esperanzas. Esto no es cierto a menos que esté implicada una función lineal
- b) Al combinar la observación a, con la propiedad a, obtenemos:  $E(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i E(X_i)$ , donde las  $a_i$  son constantes.

### Ejemplos p. 171-175

#### 1.4.2 Varianza

- a) **Caso discreto.**

$$V(Z) = \sum_{(x_1, x_2, \dots, x_n) \in R(X_1, X_2, \dots, X_n)} [H(X_1, X_2, \dots, X_n) - E(Z)]^2 p(x_1, x_2, \dots, x_n)$$

- b) **Caso continuo.**

$$V(Z) = \int_{R^n} [H(X_1, X_2, \dots, X_n) - E(Z)]^2 f(x_1, x_2, \dots, x_n) dx_1 \dots dx_n$$

#### Propiedades.

- a) Si  $(X, Y)$  es una variable aleatoria bidimensional, y si  $X$  y  $Y$  son independientes, entonces  $V(X + Y) = V(X) + V(Y)$ . **D/p.180.**
- b) Sean  $X_1, \dots, X_n$   $n$  variables aleatorias independientes. Entonces,  $V(X_1 + \dots + X_n) = V(X_1) + \dots + V(X_n)$ . Se demuestra con el ítem (a) e inducción matemática

**Observación.** El ítem a, es solo válido para cuando las variables aleatorias son independientes. La varianza no posee la propiedad de linealidad que dimos para la esperanza, es decir,  $V(aX + b) \neq aV(X) + b$ . En su lugar tenemos  $V(aX + b) = a^2 V(X)$ . De aquí obtenemos que si  $X_1, \dots, X_n$  son variables aleatorias independientes dos a dos. Entonces

$$V(a_1 X_1 + \dots + a_n X_n) = \sum_{i=1}^n a_i^2 V(X_i)$$

## 1.5 Coeficiente de correlacion

Hasta ahora nos hemos interesado en asociar parámetros como  $E(X)$  y  $V(X)$  con la distribución de variables aleatorias unidimensionales. Estos parámetros miden, en el sentido antes descrito, ciertas características de la distribución. Si tenemos una variable aleatoria bidimensional  $(X, Y)$ , se encuentra un problema análogo. Por supuesto, podemos presentar nuevamente las variables aleatorias unidimensionales  $X$  y  $Y$  asociadas con  $(X, Y)$ . Sin embargo, surge la pregunta de si hay un parámetro significativo que mida de alguna manera el “grado de asociación” entre  $X$  y  $Y$ . Esta es una noción vaga que se precisará más adelante. Demos la siguiente definición formal.

**Definición.** Sea  $(X, Y)$  una variable aleatoria bidimensional. Definimos  $\rho_{xy}$ , el **coeficiente de correlación** entre  $X$  e  $Y$ , como sigue:

$$\rho_{xy} = \frac{E\{[X - E(X)][Y - E(Y)]\}}{\sqrt{V(X)V(Y)}}$$

### Observaciones

- a) Suponemos que todas las esperanzas existen y que  $V(X)$  y  $V(Y)$  son distintas de cero. Cuando no hay dudas de cuáles variables aleatorias están implicadas escribiremos simplemente  $\rho$  en vez de  $\rho_{xy}$ .
- b) El numerador de  $\rho$ ,  $E[X - E(X)][Y - E(Y)]$ , se llama **covarianza** de  $X$  e  $Y$ , y algunas veces se denota con  $\sigma_{xy}$ .
- c) El coeficiente de correlación es una cantidad adimensional.

### Teorema

$$\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)V(Y)}}$$

D/p. 190

**Teorema** Si  $X$  y  $Y$  son independientes, entonces  $\rho = 0$ . Se deduce del teorema anterior.  $E(XY) = E(X)E(Y) \Rightarrow \rho = 0$

**Observación.** El recíproco del teorema anterior no es verdadero en general. Esto es, podemos tener  $\rho = 0$ , y aun así  $X$  y  $Y$  no necesitan ser independientes. Si  $\rho = 0$  decimos que  $X$  y  $Y$  son no correlacionados. Así, siendo no correlacionados e independientes no son, en general, equivalentes. Que el coeficiente de correlación sea 0, indica que las variables no están linealmente relacionadas, pero pueden presentar otro patrón de relación. El ejemplo siguiente ilustra este punto (p. 190)

**Teorema.**  $-1 \leq \rho \leq 1$ . D/p.191

**Teorema.** Supongamos que  $\rho^2 = 1$ . Luego  $Y = AX + B$ , donde  $A$  y  $B$  son constantes. En palabras, si el coeficiente de correlación  $\rho$  es  $\pm 1$ , entonces  $Y$  es una función lineal de  $X$  (con probabilidad 1). D/p.192

**Teorema.** Supongamos que  $X$  y  $Y$  son dos variables aleatorias para las cuales  $Y = AX + B$ , donde  $A$  y  $B$  son constantes. Entonces,  $\rho^2 = 1$ . Si  $A > 0$ ,  $\rho = 1$ ; si  $A < 0$ ,  $\rho = -1$ . D/p.192

**Observacion.** Los teoremas anteriores establecen las siguientes características importantes del coeficiente de correlación: el coeficiente de correlación es una medida del grado de linealidad entre  $X$  y  $Y$ . Los valores de  $\rho$  próximos a  $+1$  o  $-1$  indican un alto grado de linealidad, mientras que los valores de  $\rho$  próximos a  $0$  indican una ausencia de tal linealidad. Los valores positivos de  $\rho$  muestran que  $Y$  tiende a aumentar con valores crecientes de  $X$ , mientras que los valores negativos de  $\rho$  muestran que  $Y$  tiende a disminuir con valores crecientes de  $X$ . Existe una considerable confusión acerca de la interpretación del coeficiente de correlación. Un valor de  $\rho$  cercano a cero sólo indica la ausencia de una relación lineal entre  $X$  y  $Y$ . No impide la posibilidad de alguna relación no lineal.

Ejemplo p. 193-194

## 1.6 Desigualdad de Chebyshev

Si conocemos la distribución de probabilidades de una variable aleatoria  $X$  (la fdp en el caso continuo o la probabilidad puntual en el caso discreto), podemos calcular  $E(X)$  y  $V(X)$ , si existen. Sin embargo, lo recíproco no es verdadero. Esto es, conociendo  $E(X)$  y  $V(X)$  no podemos reconstruir la distribución de probabilidades de  $X$  y, por tanto, no podemos calcular cantidades tales como  $P[|X - E(X)| \leq C]$ . Sin embargo, resulta que aunque no podemos evaluar tales probabilidades [a partir de un conocimiento de  $E(X)$  y  $V(X)$ ], es posible dar una cota superior (o inferior) muy útil para tales probabilidades. Este resultado está contenido en lo que se conoce como la desigualdad de Chebyshev

**Desigualdad de Chebyshev.** Sea  $X$  una variable aleatoria con  $E(X) = \mu$  y sea  $c$  un número real cualquiera. Entonces, si  $E(X - c)^2$  es finita y  $\varepsilon$  es cualquier número positivo, tenemos

$$P[|X - c| \geq \varepsilon] \leq \frac{1}{\varepsilon^2} E(X - c)^2$$

Consideremos las siguientes formas equivalentes

a) Al considerar el evento complementario tenemos (D/p. 187)

$$P[|X - c| < \varepsilon] \geq 1 - \frac{1}{\varepsilon^2} E(X - c)^2$$

b) Al elegir  $c = \mu$  tenemos

$$P[|X - \mu| \geq \varepsilon] \leq \frac{Var X}{\varepsilon^2}$$

c) Al elegir  $c = \mu$  y  $\varepsilon = k\sigma$ , donde  $\sigma^2 = Var X > 0$ , obtenemos

$$P[|X - \mu| \geq k\sigma] \leq k^{-2}$$

Esta última forma indica especialmente cómo la varianza mide el “grado de concentración” de la probabilidad próxima a  $E(X) = \mu$ .

**Observacion.** Es importante darse cuenta de que el resultado anterior es notable debido a lo poco que se presupone acerca de la conducta probabilística de la variable aleatoria  $X$ .

Ejemplo observacion b) p.188 -Recomendable leer-

## 1.7 Distribucion de la suma de variables aleatorias

### 1.7.1 Propiedades reproductivas

Hay varias distribuciones de probabilidades que tienen la notable y útil propiedad siguiente: si dos (o mas) variables aleatorias independientes que tienen cierta distribución se suman, la variable aleatoria que resulta tiene una distribución del mismo tipo que la de los sumandos. Esta propiedad se llama propiedad reproductiva.

Ejemplos p.287

**Teorema.** La propiedad reproductiva de la distribucion normal

Sean  $X_1, \dots, X_n$ ,  $n$  variables aleatorias independientes con distribución  $N(\mu_i, \sigma_i^2)$ ,  $i = 1, 2, \dots, n$ . Sea  $Z = X_1 + \dots + X_n$ . Entonces,  $Z$  tiene distribución

$$N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

**Observacion.** Esta propiedad vale para cualquier combinación lineal. Es decir bajo las mismas condiciones se cumple

$$a_1 X_1 + \dots + a_n X_n \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

**Teorema.** La propiedad reproductiva de la distribucion de Poisson

Sean  $X_1, \dots, X_n$ ,  $n$  variables aleatorias independientes con distribución de Poisson con parametro  $\alpha_i$ ,  $i = 1, 2, \dots, n$ . Sea  $Z = X_1 + \dots + X_n$ . Luego,  $Z$  tiene distribución de Poisson con parametro  $\alpha = \sum_{i=1}^n \alpha_i$ .

D/p.288

## 1.8 La ley de los grandes numeros

Con la ayuda de la desigualdad de Chebyshev podemos derivar el resultado antes citado. Consideremos otra vez un ejemplo. Supóngase que un cohete dirigido tiene una probabilidad de 0.95 de funcionar correctamente durante cierto periodo de operación. Así, si disparamos  $N$  cohetes que tienen la confiabilidad anterior, y si  $X$  es el número de cohetes que no funcionan en forma correcta, tenemos  $E(X) = 0.05N$ , puesto que podemos suponer que  $X$  está distribuida binomialmente. Es decir, esperaríamos que fallara alrededor de un cohete entre 20. Cuando aumenta  $N$ , el número de cohetes lanzados  $X$ , el número total de cohetes que fallan dividido entre  $N$ , debería converger de algún modo con el número 0.05. Este importante resultado puede indicarse con más precisión como la ley de los grandes números.

La ley de los grandes numeros. (forma de Bernoulli). Sean  $\varepsilon$  un experimento y  $A$  un evento asociado a  $\varepsilon$ . Considerando  $n$  repeticiones independientes de  $\varepsilon$ , sea  $n_A$  el número de veces que ocurre  $A$  en las  $n$  repeticiones, y sea  $f_A = n_A/n$ . Sea  $P(A) = p$  (que se supone es igual para todas las repeticiones). Entonces para cualquier número positivo  $\epsilon$ , tenemos

$$Prob[|f_A - p| \geq \epsilon] \leq \frac{p(1-p)}{n\epsilon^2}$$

o, en forma equivalente

$$Prob[|f_A - p| < \epsilon] \geq 1 - \frac{p(1-p)}{n\epsilon^2}$$

D/p.324-325

### Observaciones

a) Lo anterior implica inmediatamente que

$$\forall \epsilon > 0 \quad \lim_{n \rightarrow \infty} P[|f_A - p| < \epsilon] = 1$$

En este sentido decimos que la frecuencia relativa “converge” a  $P(A)$

b) Es importante observar la diferencia entre la convergencia antes mencionada (llamada convergencia en probabilidad) y el tipo de convergencia citada a menudo en cálculo. Cuando decimos que  $2^{-n}$  converge a cero cuando  $n \rightarrow \infty$  significa que para una  $n$  suficientemente grande,  $2^{-n}$  se transforma y permanece arbitrariamente cercana a cero. Cuando decimos que  $f_A = n_A/n$  converge a  $P(A)$  indicamos que la probabilidad del evento

$$\{|n_A/n - P(A)| < \epsilon\}$$

puede hacerse arbitrariamente cercana uno para cualquier  $\epsilon > 0$  tomando una  $n$  suficientemente grande.

- c) Otra forma de la ley de los grandes números se obtiene cuando formulamos la siguiente pregunta. ¿Cuántas repeticiones de  $\varepsilon$  deberían hacerse para tener una probabilidad, digamos de 0.95, de que la frecuencia relativa difiera de  $p = P(A)$  en menos de 0.01? Es decir, para  $\epsilon = 0.01$  deseamos escoger  $n$ , de modo que  $1 - \frac{p(1-p)}{n(0.01)^2} = 0.95$ . Resolviendo para  $n$  obtenemos  $n = \frac{p(1-p)}{(0.01)^2(0.05)}$ . Sustituyendo los valores específicos de 0.05 y 0.01 por  $\delta$  y  $\epsilon$ , respectivamente, tenemos

$$P[|n_A/n - P(A)| < \epsilon] \geq 1 - \delta \quad \text{cuando} \quad n \geq \frac{p(1-p)}{\epsilon^2 \delta}$$

Nuevamente debería insistirse que tomar  $n \geq \frac{p(1-p)}{\epsilon^2 \delta}$  no garantiza nada acerca de  $|f_A - p|$ . Sólo hace probable que  $|f_A - p|$  sea muy pequeña.

### Ejemplo p. 326

#### Observaciones

- Recordemos que la  $f_A$  es una variable aleatoria y no precisamente un valor observado. Si lanzáramos ahora 27.778 veces un dado y luego calculamos la frecuencia relativa que salga un seis, este número dista o no 0.01 de  $\frac{1}{6}$ . Lo importante del ejemplo anterior es que si lanzáramos 27.778 veces un dado en cada una de 100 habitaciones, aproximadamente en 95 de ellas la frecuencia relativa observada estaría dentro de 0.01 de  $\frac{1}{6}$ .
- En muchos problemas no conocemos el valor de  $p = P(A)$  y, por lo tanto, no podemos usar el límite anterior de  $n$ . En ese caso podemos usar el hecho de que  $p(1-p)$  toma su valor máximo cuando  $p = 1/2$ .

### Ejemplos y observaciones p.326-327

## 1.9 Aproximación normal a la distribución binomial

Como se estableció antes, la Ley de los grandes números se relaciona esencialmente con la variable aleatoria  $X$  distribuida binomialmente.  $X$  se definió como el número de éxitos en  $n$  repeticiones independientes de un experimento, y necesitamos asociar simplemente “éxito” con la ocurrencia del evento  $A$  para reconocer esta relación. Así, el resultado anterior puede establecerse informalmente afirmando que cuando el número de repeticiones de un experimento se aumenta, la frecuencia relativa de éxito,  $X/n$ , converge a la probabilidad de éxito  $p$ , en el sentido indicado previamente.

Sin embargo, saber que  $X/n$  está “cercana” a  $p$  para una  $n$  grande no nos indica cómo se obtiene esta “cercanía”. Para investigar esto debemos estudiar la distribución de probabilidades de  $X$  cuando  $n$  es grande.

Ejemplo introductorio a la formula de Stirling para terminar obteniendo la aproximación de DeMoivre-Laplace para la distribución binomial (P.328-329)

Aproximación normal a la distribución binomial. Si  $X$  tiene una distribución binomial con parámetros  $n$  y  $p$  y si

$$Y = \frac{X - np}{[np(1-p)]^{1/2}}$$

luego, para una  $n$  grande,  $Y$  tiene aproximadamente una distribución  $N(0,1)$  en el sentido de que  $\lim_{n \rightarrow \infty} P(Y \leq y) = \Phi(y)$ . Esta aproximación es válida para valores de  $n > 10$  suponiendo que  $p$  está cercana a  $1/2$ . Si  $p$  está cercana a 0 o 1,  $n$  debería ser algo mayor para asegurar una buena aproximación.

### Observaciones

- a) El resultado anterior no es sólo de considerable interés teórico sino también de gran importancia práctica. Indica que podemos usar la distribución normal, muy tabulada, para calcular probabilidades que provienen de la distribución binomial.
- b) Al usar la aproximación normal a la distribución binomial, estamos aproximando la distribución de una variable aleatoria discreta con la distribución de una variable aleatoria continua. Por tanto, se debe tener cierto cuidado con los puntos extremos del intervalo considerado. Por ejemplo, para una variable aleatoria continua,  $P(X = 3) = 0$ , mientras que para una variable aleatoria discreta esta probabilidad puede ser positiva.

Se ha encontrado que las siguientes correcciones para continuidad mejoran la aproximación anterior:

- a)  $P(X = k) \simeq P(k - \frac{1}{2} \leq X \leq k + \frac{1}{2})$
- b)  $P(a \leq X \leq b) \simeq P(a - \frac{1}{2} \leq X \leq b + \frac{1}{2})$

Ver todos los ejemplos y observaciones 328-331



## 1.10 El teorema del límite central

Recordemos que la variable aleatoria  $X$  distribuida binomialmente se puede representar como la suma de las siguientes variables aleatorias independientes:

$$X_i = \begin{cases} 1 & \text{si Ocorre exito en la } i\text{-ésima repetición} \\ 0 & \text{si } \quad \quad \quad \text{no} \end{cases}$$

Por lo tanto,  $X = X_1 + X_2 + \dots + X_n$ . Para esta variable aleatoria hemos demostrado que  $E(X) = np$ ,  $V(X) = np(1 - p)$  y, además, que para una  $n$  grande,  $(X - np)/\sqrt{np(1 - p)}$  tiene la distribución aproximada  $N(0,1)$ .

Si una variable aleatoria  $X$  se puede representar como una suma de  $n$  variables aleatorias independientes cualesquiera (satisfaciendo ciertas condiciones que son válidas en la mayor parte de las aplicaciones que veremos mas adelante), entonces esta suma, para una  $n$  suficientemente grande, está distribuida en forma aproximadamente normal. Este resultado notable se conoce como el teorema del límite central. Una forma de este teorema se puede establecer como sigue

**Teorema del limite central.** Sea  $X_1, X_2, \dots, X_n, \dots$  una sucesión de variables aleatorias independientes con  $E(X_i) = \mu_i$  y  $V(X_i) = \sigma_i^2, i = 1, 2, \dots$  Sea  $X = X_1 + X_2 + \dots + X_n$  entonces, en ciertas condiciones generales (que no se indicarán explícitamente aquí),

$$Z_n = \frac{X - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$$

tiene aproximadamente la distribución  $N(0,1)$ . Es decir, si  $G_n$  es la fda de la variable aleatoria  $Z_n$ , tenemos  $\lim_{n \rightarrow \infty} G_n(z) = \Phi(z)$ .

### Observaciones

- Este teorema representa una generalización obvia de la aproximación de DeMoivre-Laplace. Las variables aleatorias independientes  $X_i$  que toman sólo los valores 1 y 0 han sido sustituidas por variables aleatorias que poseen cualquier clase de distribución (mientras tengan esperanza y varianza finitas). El hecho de que las  $X_i$  pueden tener (esencialmente) cualquier clase de distribución y aun así la suma  $X = X_1 + \dots + X_n$ , puede ser aproximada por una variable aleatoria distribuida normalmente, representa la razón básica de la importancia de la distribución normal en la teoría de probabilidad. En muchos problemas, la variable aleatoria que se considera se puede representar con la suma de  $n$  variables aleatorias independientes y, por tanto, su distribución puede aproximarse con la distribución normal.
- En otras palabras, los términos individuales en la suma contribuyen con una cantidad despreciable a la variación de la suma y es muy improbable que cualquier tamaño individual haga una gran contribución a la suma. (Parece que los errores de medida tienen esta característica. El error final puede ser representado como una suma de varias pequeñas contribuciones, ninguna de las cuales contribuye mucho al error completo.)
- El teorema del límite central establece que los sumandos no necesitan ser distribuidos normalmente para aproximarse a la suma con una distribución normal.

**Teorema.** Sean  $X_1, \dots, X_n$   $n$  variables aleatorias independientes que tienen la misma distribución. Sean  $\mu = E(X_i)$  y  $\sigma^2 = V(X_i)$ , la esperanza y la varianza común. Sea  $S = \sum_{i=1}^n X_i$ . Entonces,  $E(S) = n\mu$  y  $V(S) = n\sigma^2$ , y para una gran  $n$  tenemos que  $T_n = (S - n\mu)/\sqrt{n}\sigma$  tiene aproximadamente la distribución  $N(0, 1)$  considerando que  $\lim_{n \rightarrow \infty} P(T_n \leq t) = \Phi(t)$ . (D/p.334)

**Ejemplos p.335-338**