

Distribución conjunta de variables aleatorias

En muchos problemas prácticos, en el mismo experimento aleatorio, interesa estudiar no sólo una variable aleatoria sino dos o más. Por ejemplo:

Ejemplo 1: Se elige un hombre adulto al azar en una ciudad y se observa X = peso, Y =estatura.

Ejemplo 2: se elige una muestra de un mineral y se le mide su contenido de hierro por dos métodos distintos. Sea X = medición obtenida con el método 1, Y =idem con el método 2.

Ejemplo 3: Se elige un alumno al azar de 1er. año y se le pregunta X = nota en Matemáticas del curso de ingreso Y = número de horas por semana que trabaja fuera de la facultad.

En cada uno de estos ejemplos se puede definir un espacio muestral S , una función de probabilidad P y dos variables aleatorias

$$X:S \rightarrow \mathbf{R}, \quad Y:S \rightarrow \mathbf{R},$$

definidas en el **mismo espacio S** .

Hasta ahora para cada v.a. nos interesaba conocer su distribución, o sea poder calcular $P(X \in B)$ para todo $B \subset \mathbf{R}$

Con dos variables aleatorias nos puede interesar conocer también la distribución conjunta de las vs. X e Y . Conocer la distribución conjunta de X e Y quiere decir saber calcular

$$P((X,Y) \in B) \text{ para todo } B \subset \mathbf{R}^2.$$

En forma análoga a como se define función de distribución para una sola v.a. se define la función de distribución conjunta de dos vs. as.

Función de distribución conjunta de dos variables aleatorias. Definición

Sea X e Y dos vs. as. definidas en el mismo espacio de probabilidad. Su función de distribución conjunta es la función:

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

Puede observarse que $F: \mathbf{R}^2 \rightarrow \mathbf{R}$.

Distribución conjunta de dos variables aleatorias discretas.

Definición de función de probabilidad puntual conjunta.

Sean X e Y dos vs. as. discretas definidas en el mismo espacio muestral S . Su f.p.p. conjunta es:

$$p(x,y) = P(X=x, Y=y) \quad \forall (x,y) \in \mathbf{R}^2$$

Puede observarse que $p: \mathbf{R}^2 \rightarrow \mathbf{R}$

Ejemplo 4: Se tira 3 veces una moneda equilibrada. Sea X el número de caras, Y el nro. de caras en las dos primeras tiradas.

Evidentemente ambas variables son discretas, ya que $I_X = \{0,1,2,3\}$, $I_Y = \{0,1,2\}$

La siguiente tabla da la f.p.p. conjunta de estas dos variables aleatorias (verifíquelo):

$y \backslash x$	0	1	2	3	Total
0	1/8	1/8	0	0	
1	0	2/8	2/8	0	
2	0	0	1/8	1/8	
Total					1

Propiedades de las funciones de probabilidad puntual conjunta de 2 vs. as.

Son funciones $p: \mathbb{R}^2 \rightarrow \mathbb{R}$ que cuplen dos propiedades:

- a) $p(x,y) \geq 0$ para todo $(x,y) \in \mathbb{R}^2$
- b) $\sum_{x \in I_X} \sum_{y \in I_Y} p(x,y) = 1$

Conocida la f.p.p. conjunta de X e Y, se puede calcular $P((X,Y) \in B)$ del siguiente modo:

$$P((X,Y) \in B) = \sum_{\substack{x \in I_X \\ y \in I_Y \\ (x,y) \in B}} p(x,y)$$

¿Si conozco la f.p.p. conjunta de X e Y, puedo calcular la f.p.p. de X?

Calcularla en el ejemplo 4.

En general:

$$p_X(x) = \sum_{y \in I_Y} p(x,y)$$

$$p_Y(y) = \sum_{x \in I_X} p(x,y)$$

Un nombre: la función $p(x,y)$ se llama como dijimos f.p.p. conjunta de X e Y. Las funciones p_X y p_Y se suelen llamar "**funciones de probabilidad puntual marginales**", ¿de donde viene ese nombre?

Distribución conjunta continua de dos variables aleatorias.

En el caso de una variable aleatoria dimos la siguiente definición.

Definición: Una variable aleatoria X es continua si existe una función $f:R \rightarrow R^+$ tal que

$$F(x) = \int_{-\infty}^x f(t)dt \quad \forall x \in R$$

En forma similar definimos:

Definición: Dos variables aleatorias X e Y tienen distribución conjunta continua si existe una función $f:R^2 \rightarrow R^+$ tal que si F es la función de distribución conjunta de X e Y entonces:

$$F(x,y) = \int_{-\infty}^x \int_{-\infty}^y f(u,v)dv du \quad \forall (x,y) \in R^2$$

La función f se llama **función de densidad conjunta** de X e Y.

Propiedades de las funciones de densidad conjunta

Son funciones $f: R^2 \rightarrow R$ que cumplen dos propiedades:

a) $f(x,y) \geq 0$ para todo $(x,y) \in R^2$

b) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)dx dy = 1$

Al igual que en el caso de una sola variable, también la función de densidad conjunta es un modelo para un histograma. Para el histograma “conjunto” de las variables X e Y (¿se entiende que quiero decir con histograma conjunto de las dos variables? En general se necesitan muchos datos para poder representar y modelar un histograma conjunto!

Proposición: A partir de la definición de distribución conjunta continua se puede demostrar que si X e Y tienen función de densidad conjunta $f(x,y)$ entonces:

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x,y)dy dx$$

Más aún, esto no sólo es cierto para rectángulos sino para cualquier $B \subset R^2$

$$P((X,Y) \in B) = \iint_B f(x,y) dx dy$$

Conocida la función de densidad conjunta ¿se puede calcular la función de densidad de X y la de Y?

La respuesta es sí y la forma de calcularlas es similar al caso discreto, con integrales en lugar de sumatorias:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y) dy$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y) dx$$

$f(x,y)$ se llama función de densidad conjunta, f_X y f_Y se llaman funciones de densidad marginales.

Variables aleatorias independientes:

Hemos definido la independencia entre dos variables aleatorias del siguiente modo:

Definición: Dos variables aleatorias X e Y son **independientes** si

$$P((X \in A) \cap (Y \in B)) = P(X \in A) P(Y \in B)$$

para todo $A \subset \mathbb{R}$, $B \subset \mathbb{R}$.

Se puede demostrar que esta definición es equivalente a que la función de distribución conjunta se escriba como producto de las marginales. Esto se enuncia en la siguiente proposición.

Proposición: Sean X e Y variables aleatorias. Sea F_X la función de distribución de X, F_Y la función de distribución de Y y F la función de distribución conjunta de X e Y. Entonces:

$$X \text{ e } Y \text{ son independientes} \Leftrightarrow F(x,y) = F_X(x)F_Y(y) \quad \forall (x,y) \in \mathbb{R}^2$$

La siguiente proposición dice que, para el caso particular de variables discretas, la independencia es equivalente a que la f.p.p. conjunta sea el producto de las marginales.

Proposición: Sean X e Y dos variables aleatorias discretas. Entonces:

$$X \text{ e } Y \text{ son independientes} \Leftrightarrow p(x,y) = p_X(x)p_Y(y) \quad \forall (x,y) \in \mathbb{R}^2$$

Para variables continuas la independencia es equivalente a que tengan una función de densidad conjunta que se escriba como producto de las marginales:

Proposición: Sean X e Y dos variables aleatorias continuas. Entonces:

$$X \text{ e } Y \text{ son independientes} \Leftrightarrow f(x,y) = f_X(x) f_Y(y) \quad \forall (x,y) \in \mathbb{R}^2$$

Distribución conjunta de n variables aleatorias.

A pesar de que el título de esta sección es “distribución conjunta de variables aleatorias” en todas las definiciones y propiedades nos restringimos a dos variables. Pero esto ha sido sólo para simplificar la notación, todo se extiende a n variables en forma natural. Veamos por ejemplo la extensión de la definición de f.p.p. conjunta y de la independencia para variables discretas y continuas.

Función de probabilidad conjunta de n variables aleatorias discretas. Definición Sean X_1, X_2, \dots, X_n vs. as. discretas definidas en el mismo espacio de probabilidad. Su función de probabilidad puntual conjunta es la función:

$$p(x_1, x_2, \dots, x_n) = P(X_1=x_1, X_2=x_2, \dots, X_n=x_n)$$

Proposición: Las vs. as. discretas X_1, X_2, \dots, X_n son independientes si su f.p.p. conjunta es el producto de las f.p.p. marginales:

$$p(x_1, x_2, \dots, x_n) = p_{X_1}(x_1) p_{X_2}(x_2) \dots p_{X_n}(x_n) \quad \forall (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$$

Proposición: Las vs. as. continuas X_1, X_2, \dots, X_n son independientes si su función de densidad conjunta es el producto de las funciones de densidad marginales:

$$f(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \dots f_{X_n}(x_n) \quad \forall (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$$

Esperanza de una función de dos vs. aleatorias.

Enunciamos un teorema que dice que para una v.a. X discreta se cumple

$$E(g(X)) = \sum_{x \in I_x} g(x) p(x)$$

y que para una v.a. X continua es

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

Este teorema se generaliza para poder calcular la esperanza de una g(X,Y):

Teorema: Sean X e Y variables aleatorias y $g: \mathbb{R}^2 \rightarrow \mathbb{R}$,

a) Si X e Y son **discretas** con f.p.p. conjunta p(x,y) entonces:

$$E(g(X, Y)) = \sum_{x \in I_x} \sum_{y \in I_y} g(x, y) p(x, y)$$

b) Si X e Y son vs. as. con función de densidad conjunta f(x,y), entonces:

$$E(g(X, Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy$$

Usando el teorema anterior, se prueba fácilmente la propiedad que hemos usado tantas veces y que nunca dijimos como se podía demostrar:

Proposición: Si X e Y son vs. as. Si existe E(X) y E(Y) entonces existe también E(X+Y) y vale:

$$E(X + Y) = E(X) + E(Y)$$

¿Cuánto vale E(X.Y)? En general hay que calcularla usando el teorema que permite calcular E(g(X,Y)). Pero para el caso particular de variables independientes se puede demostrar fácilmente la siguiente proposición (la demostración fácil es para el caso discreto o continuo, aunque la proposición vale siempre):

Proposición: Si X e Y son vs. as. independientes con esperanza entonces:

$$E(X.Y) = E(X).E(Y)$$

Covarianza entre dos vs. as.

Se definió varianza de una v.a. X:

$$\text{Var}(X) = E(X - E(X))^2$$

En forma parecida se **define covarianza** entre dos vs.as. X e Y :

$$\text{cov}(X, Y) = E[(X - E(X)).(Y - E(Y))]$$

Observar que la varianza es un caso particular de la covarianza, ya que $\text{cov}(X, X) = \text{Var}(X)$.

Cómo se calcula? Usando el teorema que permite calcular E(g(X,Y)) se ve que:

a) Caso discreto

$$\text{cov}(X, Y) = \sum_{x \in I_x} \sum_{y \in I_y} (x - E(X))(y - E(Y)) p(x, y)$$

b) Caso continuo

$$\text{cov}(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - E(X))(y - E(Y)) f(x, y) dx dy$$

Interpretación intuitiva del signo de la covarianza: si cuando X toma valores “grandes”, Y también tiende a tomar valores grandes, el signo de la cov(X,Y) es positivo. Si cuando X toma valores grandes, Y tiende a tomar valores pequeños $\text{sg}(\text{cov}(X, Y)) < 0$ (justificar intuitivamente).

Si se elige un alumno al azar y X = nota en Matemáticas, Y = numero de horas por semana que trabaja fuera de la facultad, cual es el signo de cov(X,Y)?

Hemos visto que otra forma de calcular varianza es usar que: $\text{Var}(X) = E(X^2) - (E(X))^2$

La covarianza tiene una expresión semejante:

Proposición:

$$\text{cov}(X, Y) = E(X.Y) - E(X).E(Y)$$

Esta proposición permite calcular **covarianza en forma más simple, calculando primero por separado E(X), E(Y) y E(X.Y).**

Ejercicio: calcular cov(X,Y) para el ejemplo 4 (caso discreto).

Proposición: Si X e Y son independientes entonces $\text{cov}(X,Y)=0$

Vimos que $E(X+Y) = E(X) + E(Y)$ ¿Cuánto valdrá $\text{Var}(X+Y)$?

Proposición

- a) $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{cov}(X,Y)$
- b) $\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y) - 2 \text{cov}(X,Y)$
- c) Si X e Y son vs. as. independientes entonces

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$$

Hemos visto que $\text{Var}(aX+b) = a^2 \text{Var}(X)$. Una propiedad similar para la covarianza está dada por la siguiente proposición (fácil demostrar, usando propiedades ya conocidas de la esperanza)

Proposición:

$$\text{Cov}(aX+b,cY+d) = a.c.\text{cov}(X,Y) \tag{1}$$

Puede observarse que, como consecuencia de la proposición anterior, la covarianza de dos variables cambia al cambiar de unidades las variables. No vale lo mismo la covarianza entre peso (en kg.) y estatura (en metros) que si las unidades son gramos y cm. ¿Cómo cambia la cov en este ejemplo?

Una medida de la relación entre dos variables, que no cambia al cambiar de unidades es el coeficiente de correlación definido del modo siguiente.

Coefficiente de correlación (o coeficiente de correlación lineal). Definición

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Observación: Como es evidente que $\text{sg}(\rho(X,Y)) = \text{sg}(\text{cov}(X,Y))$, el signo del coeficiente de correlación tiene el mismo significado intuitivo que el signo de la covarianza.

Proposición

- a) $\rho(aX+b,cY+d) = \text{sg}(a.c) \rho(X,Y)$ (de esta propiedad surge que ρ no varía al cambiar las unidades de las variables X y/o Y)
- b) $-1 \leq \rho(X,Y) \leq 1$
- c) $|\rho(X,Y)|=1 \Leftrightarrow$ existen números reales a y b, con $a \neq 0$ tales que $Y=aX+b$ con probabilidad 1.

La demostración de a) sale fácilmente de las propiedades análogas para la varianza y la covarianza (1). Demostraremos propiedades similares a b) y c) cuando estudiemos el coeficiente de correlación muestral r, en inferencia estadística

**Suma y promedio de variables aleatorias independientes.
Ley de los grandes números. Teorema central del Limite.**

Teorema (Desigualdad de Tchebycheff).

Sea X una v.a. con cualquier distribución, con esperanza y varianza finitas. Sea $a > 0$. Entonces:

$$P(|X - E(X)| \geq a) \leq \frac{\text{Var}(X)}{a^2}$$

Comentario: observar que la desigualdad de Tchebycheff es equivalente a:

$$P(|X - E(X)| < a) \geq 1 - \frac{\text{Var}(X)}{a^2}$$

Ejemplo de aplicación: Sea X una v.a. con cualquier distribución con esperanza y varianza finitas. Llamemos μ a su esperanza y σ^2 a su varianza. Dar una cota para

a) $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$

Si sabemos que $X \sim \text{Normal}$, ¿cuánto vale esta probabilidad?

b) ídem a) para $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$

c) ídem a) para $P(\mu - \sigma \leq X \leq \mu + \sigma)$

Esperanza y varianza de sumas de vs. as. independientes

Dijimos que si X_1, \dots, X_n son vs. as.

$$E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i)$$

Si X_1, \dots, X_n son vs. as. independientes

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$$

Proposición : Sean X_1, \dots, X_n son vs. as. independientes, todas con la misma esperanza y la misma varianza finita. Llamemos μ a $E(X_i)$ y σ^2 a la $\text{Var}(X_i)$. Sea

$$S = \sum_{i=1}^n X_i \quad \text{y} \quad \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

entonces

$$E(S) = n\mu \quad , \quad \text{Var}(S) = n\sigma^2$$

$$E(\bar{X}) = \mu \quad , \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

Nota: obsérvese que el caso X_1, \dots, X_n vs. as. independientes igualmente distribuidas (i.i.d.) con esperanza y varianza finita cumple las condiciones de la proposición.

Convergencia de una sucesión de variables aleatorias.

Una definición (no es la única que se usa) es la siguiente:

Definición de convergencia en probabilidad.

Sea Y_1, \dots, Y_n, \dots una sucesión de variables aleatorias. Se dice que

$$Y_n \xrightarrow[n \rightarrow \infty]{} Y \quad \text{en probabilidad}$$

si para cualquier $\varepsilon > 0$ se cumple que

$$\lim_{n \rightarrow \infty} P(|Y_n - Y| > \varepsilon) = 0$$

Teorema (Ley de los Grandes Números, versión de Tchebycheff)

Sean X_1, \dots, X_n son vs. as. independientes, todas con la misma esperanza y la misma varianza finita. Llamemos μ a $E(X_i)$ y σ^2 a la $\text{Var}(X_i)$. Sea

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$$

Entonces

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{} \mu \quad \text{en probabilidad}$$

La demostración del teorema anterior sale fácil de la desigualdad de Tchebycheff. Luego se probó que no es necesario suponer varianza finita, como lo muestra el teorema siguiente.

Teorema (Ley de los Grandes Números, versión de Khintchine, 1929)

Sean X_1, \dots, X_n son vs. as. i.i.d., todas con esperanza finita. Llamemos μ a $E(X_i)$

Entonces

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{} \mu \quad \text{en probabilidad}$$

Distribución de la suma de variables independientes.

Proposición :

Sean X e Y variables aleatorias independientes.

- Si $X \sim \text{Bin}(n_1, p)$, $Y \sim \text{Bin}(n_2, p)$, entonces $X+Y \sim \text{Bin}(n_1+n_2, p)$.
- Si $X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2)$, entonces $X+Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$
- Si $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, entonces $X+Y \sim N(?, ?)$ (rellenar los ?)

d) Si $X \sim \Gamma(\alpha_1, \lambda)$, $Y \sim \Gamma(\alpha_2, \lambda)$ entonces $X+Y \sim \Gamma(\alpha_1+\alpha_2, \lambda)$

El inciso a) es intuitivamente razonable, verdad? Los otros no son intuitivos.

Comentario: Esta propiedad de que la distribución de la suma de 2 vs as independiente “pertenece a la misma familia” no es general. Por ejemplo, si X e Y son vs. as. independientes, cada una con distribución uniforme en el intervalo [a,b], X+Y no tiene distribución uniforme.

Proposición (generalización de la proposición anterior)

Sean X_1, \dots, X_m variables aleatorias independientes.

a) Si $X_i \sim \text{Bin}(n_i, p)$ entonces $\sum_{i=1}^m X_i \sim \text{Bin}(\sum_{i=1}^m n_i, p)$.

b) Si $X_i \sim \text{Poisson}(\lambda_i)$ entonces $\sum_{i=1}^m X_i \sim \text{Poisson}(\sum_{i=1}^m \lambda_i)$.

c) Si $X_i \sim N(\mu_i, \sigma_i^2)$, entonces $\sum_{i=1}^m X_i \sim N(\sum_{i=1}^m \mu_i, \sum_{i=1}^m \sigma_i^2)$.

d) Si $X_i \sim \Gamma(\alpha_i, \lambda)$ entonces $\sum_{i=1}^m X_i \sim \Gamma(\sum_{i=1}^m \alpha_i, \lambda)$

Vamos a enunciar dos resultados sobre la distribución de la suma y el promedio de variables i.i.d.

Teorema 1:

Sean X_1, \dots, X_n son vs. as. i.i.d. $N(\mu, \sigma^2)$. Sea

$$S = \sum_{i=1}^n X_i \quad \text{y} \quad \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

entonces

$$S \sim N(n\mu, n\sigma^2) \quad , \quad \bar{X} \sim N(\mu, \sigma^2/n)$$

Este teorema es consecuencia del inciso b) de la proposición anterior.

Teorema 2 (con la tesis escrita en forma no rigurosa)

Sean X_1, \dots, X_n son vs. as. i.i.d. **con cualquier distribución**, con esperanza y varianza finitas. Llamemos μ a la $E(X_i)$ y $\sigma^2 = \text{Var}(X_i)$. Sea

$$S_n = \sum_{i=1}^n X_i \quad \text{y} \quad \bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$$

entonces

