

Distribuciones multivariada y probabilidad condicional

Gustavo Ahumada

Distribuciones de probabilidad multivariada

Distribución conjunta de dos variables aleatorias

Nuestro estudio de variables aleatorias y sus distribuciones de probabilidad en las secciones anteriores se restringen al espacio muestral unidimensional, en los que registramos los resultados de un experimento como los valores que toma una sola variable aleatoria. Habrá situaciones, sin embargo, donde podemos encontrar que es deseable registrar los resultados simultáneos de diversas variables aleatorias. Por ejemplo, podemos medir el tiempo de ensamblaje de un automovil \mathbf{T} y el número de trabajadores \mathbf{N} de un taller mecanico, que dan lugar a un espacio muestral bidimensional que consiste en los resultados (t, n) .

Distribución de probabilidad conjunta para variables aleatorias discretas

Si X y Y son dos variables aleatorias discretas, la distribución de probabilidad para sus ocurrencias simultáneas se puede representar mediante una función con valores $f(x, y)$ para cualquier par de valores (x, y) dentro del rango de valores de X y Y .

Definición. La función $f(x, y)$ es una **distribución de probabilidad conjunta** o **función de masa de probabilidad** de las variables aleatorias discretas X y Y si:

1. $f(x, y) \geq 0$ para toda (x, y) ,
2. $\sum_x \sum_y f(x, y) = 1$,
3. $P(X = x, Y = y) = f(x, y)$.

Para cualquier región A en el plano xy , $P[(X, Y) \in A] = \sum \sum_A f(x, y)$.

Ejemplo. Para graduarse con un pregrado en matemáticas, todos los estudiantes deberían pasar Calculo III y Álgebra Lineal con una nota igual o superior a 3. La notas de los estudiantes de matemáticas se encuentran en la siguiente tabla.

		Linear Algebra			
		4	3	2	Total
Calculus III	4	2	13	6	21
	3	5	85	40	130
	2	7	33	9	49
	Total	14	131	55	200

Figure 1: Notas Álgebra lineal y Calculo III

- ¿Cuál es la probabilidad de obtener una calificación igual o superior a 3 en Álgebra Lineal?
- ¿Cuál es la probabilidad de obtener una calificación igual o superior a 3 en Calculo III?
- ¿Cuál es la probabilidad de obtener una calificación igual o superior a 3 tanto en Álgebra Lineal como en Calculo III?

Solución

- Sea X y Y variables aleatorias que representan la nota obtenida (en una escala de 4) por los estudiantes de Calculo III y Álgebra lineal, respectivamente, de una población de 200 estudiantes. Si A representa el par ordenado de la notas de Calculo III y Álgebra lineal tales que la nota en Álgebra lineal es igual o mayor a 3, entonces la probabilidad de obtener una nota igual o superior a 3 en Álgebra lineal es escrita:

$$\mathbb{P}[(X, Y), \in A] = \sum_{(x, y) \in A} p_{X, Y}(x, y) = \frac{2 + 5 + 7 + 13 + 85 + 33}{200} = \frac{145}{200}.$$

- Sea X y Y variables aleatorias que representan la nota obtenida (en una escala de 4) por los estudiantes de Calculo III y Álgebra lineal, respectivamente, de una población de 200 estudiantes. Si A representa el par ordenado de la notas de Calculo III y Álgebra lineal tales que la nota en Álgebra lineal es igual o mayor a 3, entonces la probabilidad de obtener una nota igual o superior a 3 en Calculo III es escrita:

$$\mathbb{P}[(X, Y), \in A] = \sum_{(x, y) \in A} p_{X, Y}(x, y) = \frac{2 + 13 + 6 + 5 + 85 + 40}{200} = \frac{151}{200}.$$

- Sea X y Y variables aleatorias que representan la nota obtenida (en una escala de 4) por los estudiantes de Calculo III y Álgebra lineal, respectivamente, de una población de 200 estudiantes. Si A representa el par ordenado de la notas de Calculo III y Álgebra lineal tales que la nota en Álgebra lineal es igual o mayor a 3, entonces la probabilidad de obtener una nota igual o superior a 3 tanto en Calculo III como en Álgebra lineal es escrita:

$$\mathbb{P}[(X, Y), \in A] = \sum_{(x, y) \in A} p_{X, Y}(x, y) = \frac{2 + 5 + 13 + 85}{200} = \frac{105}{200}.$$

Para cualquier variable aleatoria X y Y , el cdf conjunta se define como:

$$F_{X, Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y),$$

mientras la pdf marginal de X y Y , denotadas por $p_X(x)$ y $p_Y(y)$ se definen como:

$$p_X(x) = \sum_y p_{X, Y}(x, y)$$

$$p_Y(y) = \sum_x p_{X, Y}(x, y).$$

Distribución de probabilidad conjunta para variables aleatorias continuas

Cuando X y Y son variables aleatorias continuas, la **función de densidad conjunta** $f(x, y)$ es una superficie sobre el plano xy , y $\mathbb{P}[(X, Y) \in A]$, donde A es cualquier región en xy .

Definición. La función $f(x, y)$ es una **función de densidad conjunta** de las variables aleatorias X y Y si

1. $f(x, y) \geq 0$ para toda (x, y) ,
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$,
3. $\mathbb{P}[(X, Y) \in A] = \int_A \int f(x, y) dx dy$.

Para una variable aleatoria X y Y , la cdf conjunta se define:

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy, -\infty < x < \infty, -\infty < y < \infty,$$

mientras la pdf marginal de X y Y , denotadas por $f_X(x)$ y $f_Y(y)$, respectivamente se definen:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy, -\infty < x < \infty, -\infty < y < \infty,$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx, -\infty < y < \infty, -\infty < x < \infty,$$

Ejemplo. Dada la siguiente función de probabilidad conjunta (pdf) para variables aleatorias continuas

$$f_{X,Y}(x, y) = \begin{cases} 1 & \text{si } 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0 & \text{en otros casos.} \end{cases}$$

- a. Encontrar $F_{X,Y}(x = 0.6, y = 0.8)$.
- b. Encontrar $\mathbb{P}(0.25 \leq X \leq 0.75, 0.1 \leq Y \leq 0.9)$.
- c. Encontrar $f_X(x)$.

Solución.

a.

$$F_{X,Y}(x = 0.6, y = 0.8) = \int_0^{0.6} \int_0^{0.8} f_{X,Y}(x, y) dx dy = \int_0^{0.6} \int_0^{0.8} 1 dx dy = \int_0^{0.6} 0.8 dx = 0.48.$$

b.

$$\mathbb{P}(0.25 \leq X \leq 0.75, 0.1 \leq Y \leq 0.9) = \int_{0.25}^{0.75} \int_{0.1}^{0.9} f_{X,Y}(x, y) dx dy = \int_{0.25}^{0.75} \int_{0.1}^{0.9} 1 dx dy = \int_{0.25}^{0.75} 0.8 dx = 0.40.$$

c.

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

$f(x, y)$		x			Totales por renglón
		0	1	2	
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
Totales por columna		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Figure 2: Distribución de probabilidad conjunta

Distribución condicional de probabilidad e independencia estadística

Definición. Sean X y Y dos variables aleatorias, discretas o continuas. La **distribución condicional** de la variable aleatoria Y , dado que $X = x$, es

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

De manera similar se define la distribución condicional de X , dado que $Y = y$, como:

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

Si se desea encontrar la probabilidad de que la variable aleatoria discreta X se encuentra entre a y b cuando se sabe que la variable aleatoria discreta $Y = y$, evaluamos

$$P(a < X < b|Y = y) = \sum_x f(x|y),$$

donde la sumatoria se extiende a todos los valores de X entre a y b . Cuando X y Y son continuas, se tiene

$$P(a < X < b|Y = y) = \int_a^b f(x|y)dx.$$

Ejemplo. Se selecciona al azar dos repuestos para un automóvil de una caja que contiene tres repuestos azules, dos rojos y 3 verdes. Si X es el número de repuestos azules y Y es el número de repuestos rojos que se seleccionan, encontrar:

- La función de probabilidad conjunta $f(x, y)$.
- $\mathbb{P}[X, Y] \in A$, donde $A = (x, y)|x + y \leq 1$.
- Encontrar la distribución condicional de X , dado que $Y = 1$, y utilízela para determinar $\mathbb{P}(X = 0|Y = 1)$.

Solución.

- a. Los posibles pares de valores (x, y) son $(0, 0), (0, 1), (1, 0), (1, 1), (0, 2)$, y $(2, 0)$. Ahora bien, $(0, 1)$, por ejemplo, representa la probabilidad de que se seleccionen un repuesto rojo y uno verde. El número total de formas igualmente probables de seleccionar cualesquiera dos repuestos de los ocho es $\binom{8}{2} = 28$. El número de formas de seleccionar uno rojo de dos repuestos rojos y uno verde es $\binom{2}{1}\binom{3}{1} = 6$. De aquí $f(0, 1) = 6/28 = 3/14$. Cálculos similares dan las probabilidades para los otros casos, que se encuentran en la tabla de abajo. Notese que las probabilidades suman 1.

$$f(x, y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}},$$

para $x = 0, 1, 2; \quad y = 0, 1, 2; \quad 0 \leq x + y \leq 2$.

b.

$$\begin{aligned} f(x, y) &= \mathbb{P}[(X, Y) \in A] = P(X + Y \leq 1) \\ &= f(0, 0) + f(0, 1) + f(1, 0) \\ &= \frac{3}{28} + \frac{3}{14} + \frac{9}{28} \\ &= \frac{9}{14}. \end{aligned}$$

- c. Necesitamos encontrar $f(x, y)$, donde $y = 1$. Primero, encontramos que

$$h(1) = \sum_{x=0}^2 f(x, 1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}.$$

Ahora tenemos,

$$f(x|1) = \frac{f(x, 1)}{h(1)} = \frac{7}{3}f(x, 1), \quad x = 0, 1, 2.$$

Por lo tanto,

$$\begin{aligned} f(0|1) &= \frac{7}{3}f(0, 1) = \frac{\binom{7}{3}\binom{3}{14}}{\binom{7}{3}\binom{3}{14}} = \frac{1}{2}, \\ f(1|1) &= \frac{7}{3}f(1, 1) = \frac{\binom{7}{3}\binom{3}{14}}{\binom{7}{3}\binom{3}{14}} = \frac{1}{2}, \\ f(2|1) &= \frac{7}{3}f(2, 1) = \frac{\binom{7}{3}\binom{0}{0}}{\binom{7}{3}\binom{0}{0}} = 0 \end{aligned}$$

x	0	1	2
f(x)	$\frac{1}{2}$	$\frac{1}{2}$	0

Finalmente, $\mathbb{P}(X = 0|Y = 1) = f(0, 1) = \frac{1}{2}$.

Por lo tanto, si se sabe que uno de los dos repuestos seleccionados es de color rojo, tenemos una probabilidad igual a $\frac{1}{2}$ de que el otro repuesto no sea azul.

Definición. Sean X y Y dos variables aleatorias, discretas o continuas, con distribución de probabilidad conjunta $f(x, y)$ y distribuciones marginales $g(x)$ y $h(y)$, respectivamente. Se dice que las variables aleatorias X y Y son **estadísticamente independientes** si y sólo si

$$f(x, y) = g(x)h(y) \tag{1}$$

para todo (x, y) dentro de sus rangos.

Ejemplo. Demostrar que las variables aleatorias del ejemplo anterior no son estadísticamente independientes.

Solución. Considerar el punto $(0, 1)$. De la tabla anterior encontramos que las tres probabilidades $f(0, 1)$, $g(0)$ y $h(1)$ son

$$\begin{aligned} f(0, 1) &= \frac{3}{4}, \\ g(0) &= \sum_{y=0}^2 f(0, y) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14}, \\ h(1) &= \sum_{x=0}^2 f(x, 1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7} \end{aligned}$$

Claramente,

$$f(0, 1) \neq g(0)h(1),$$

y por lo tanto X y Y no son estadísticamente independientes.