

# Chapter 4

## VARIABLES ALEATORIAS MÚLTIPLES

### 4.1. Distribución conjunta y marginal

**Definición 4.1.1** Un *vector aleatorio  $n$ -dimensional* es una función que va de un espacio muestral  $S$  a un espacio euclideo  $n$ -dimensional  $\mathfrak{R}^n$ .

**Ejemplo 4.1.1** Considere el experimento de lanzar 2 veces un dado. El espacio muestral para este experimento tiene 36 puntos igualmente probables. Sean las variables aleatorias  $X =$  suma de los dados, y  $Y =$  |diferencia de los dados|, entonces el vector aleatorio  $(X, Y)$  es llamado *vector aleatorio discreto* ya que solo tiene un número finito de posibles valores.

**Definición 4.1.2** Sea  $(X, Y)$  un vector aleatorio discreto bivariado. Entonces la función  $f(X, Y)$ , que va de  $\mathfrak{R}^2$  hacia  $\mathfrak{R}$ , definido por  $f(x, y) = \Pr(X = x, Y = y)$  es llamado la *función de probabilidad conjunta* de  $(X, Y)$ .

Table 4.1: Valores de la función de probabilidad conjunta  $f(x, y)$

	$x$											
	2	3	4	5	6	7	8	9	10	11	12	
$y$	0	$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$
	1		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$	
	2			$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		
	3				$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$			
	4					$\frac{1}{18}$		$\frac{1}{18}$				
	5						$\frac{1}{18}$					

```

> set.seed(100)
> dado1 <- sample(1:6, size=50000, prob=rep(1/6, 6), replace=T)
> set.seed(200)
> dado2 <- sample(1:6, size=50000, prob=rep(1/6, 6), replace=T)
> x <- dado1 + dado2
> y <- abs(dado1 - dado2)
> fxy <- table(y, x)/50000
> round(fxy, 3)

```

	x											
y	2	3	4	5	6	7	8	9	10	11	12	
0	0.026	0.000	0.027	0.000	0.029	0.000	0.028	0.000	0.028	0.000	0.026	
1	0.000	0.056	0.000	0.055	0.000	0.055	0.000	0.056	0.000	0.056	0.000	
2	0.000	0.000	0.056	0.000	0.057	0.000	0.057	0.000	0.055	0.000	0.000	
3	0.000	0.000	0.000	0.056	0.000	0.054	0.000	0.057	0.000	0.000	0.000	
4	0.000	0.000	0.000	0.000	0.056	0.000	0.055	0.000	0.000	0.000	0.000	
5	0.000	0.000	0.000	0.000	0.000	0.055	0.000	0.000	0.000	0.000	0.000	

La función de probabilidad conjunta puede usarse en el cálculo de probabilidades para cualquier evento definido en términos de  $(X, Y)$ . Sea  $A$  un subconjunto de  $\mathfrak{R}^2$ . Entonces:

$$\Pr((X, Y) \in A) = \sum_{(X, Y) \in A} f(x, y)$$

**Ejemplo 4.1.2** Para  $(X, Y)$  cuya función de probabilidad conjunta se encuentra en la Tabla 4.1, suponga que  $A = \{(x, y) : x = 7, y \leq 4\}$ , entonces:

$$\Pr(X = 7, Y \leq 4) = \Pr((X, Y) \in A) = f(7, 1) + f(7, 3) = \frac{1}{18} + \frac{1}{18} = \frac{1}{9}$$

Sea  $g(x, y)$  una función de valor real definido para todos los posibles valores  $(x, y)$  de vector aleatorio discreto  $(X, Y)$ . Entonces  $g(X, Y)$  es también una variable aleatoria y su valor esperado se define por:

$$E[g(X, Y)] = \sum_{(x, y) \in \mathfrak{R}^2} g(x, y) f(x, y)$$

**Ejemplo 4.1.3** Usando la Tabla 4.1 el valor esperado de  $g(X, Y) = XY$ ,

$$E[g(X, Y)] = \sum xyf(x, y) = (2)(0)\frac{1}{36} + \cdots + (7)(5)\frac{1}{18} = 13\frac{11}{18}$$

Las propiedades vistas en el teorema 3.3.1 son válidas al reemplazar  $x$  por  $(x, y)$ . Por ejemplo, si  $g_1(x, y)$ ,  $g_2(x, y)$  son dos funciones;  $a$ ,  $b$  y  $c$  son constantes, entonces:

$$E[ag_1(X, Y) + bg_2(X, Y) + c] = aE[g_1(X, Y)] + bE[g_2(X, Y)] + c$$

La función de probabilidad conjunta del vector aleatorio  $(X, Y)$  debe cumplir con  $f(x, y) \geq 0$ , para todo  $(x, y)$ . Además:

$$\sum_{(x,y) \in \mathfrak{R}^2} f(x, y) = \Pr((X, Y) \in \mathfrak{R}^2) = 1.$$

**Teorema 4.1.1** Sea  $(X, Y)$  un vector aleatorio discreto cuya función de probabilidad conjunta es  $f(x, y)$ , entonces la función de probabilidad marginal de  $X$ ,  $f_X(x) = \Pr(X = x)$ , y de  $Y$ ,  $f_Y(y) = \Pr(Y = y)$ , están dadas por:

$$f_X(x) = \sum_{y \in \mathfrak{R}} f(x, y) \quad \text{y} \quad f_Y(y) = \sum_{x \in \mathfrak{R}} f(x, y)$$

**Ejemplo 4.1.4** Usando el teorema 4.1.1 se puede calcular las distribuciones marginales de  $X$  y  $Y$  a partir de la distribución conjunta de la tabla 4.1:

$$f_Y(0) = \frac{1}{6} \quad f_Y(1) = \frac{5}{18} \quad f_Y(2) = \frac{2}{9} \quad f_Y(3) = \frac{1}{6} \quad f_Y(4) = \frac{1}{9} \quad f_Y(5) = \frac{1}{18}$$

```
> fy <- apply(fxy, 1, sum)
> fx <- apply(fxy, 2, sum)
> fy
```

```
      0      1      2      3      4      5
0.16498 0.27756 0.22390 0.16754 0.11066 0.05536
```

**Ejemplo 4.1.5** Usando la función de probabilidad marginal de  $Y$  se puede calcular:

$$\Pr(Y < 3) = \frac{2}{3} \quad E[Y^3] = 20\frac{11}{18}$$

**Definición 4.1.3** Una función  $f(x, y)$  que va de  $\mathfrak{R}^2$  hacia  $\mathfrak{R}$  es llamada *función de densidad conjunta del vector aleatorio bivariado continuo*  $(X, Y)$  si, para todo  $A \subset \mathfrak{R}^2$ :

$$\Pr((X, Y) \in A) = \iint_A f(x, y) dx dy$$

Si  $g(x, y)$  es una función de valor real, entonces *el valor esperado* de  $g(X, Y)$  se define por:

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy \quad (4.1.1)$$

Las *funciones de densidad marginales* de  $X$  y  $Y$  son definidas, reemplazando las sumatorias por las integrales. Estas funciones pueden usarse para calcular probabilidades o valores esperados que involucran solo a  $X$  o  $Y$ . Simplificando, las funciones de densidad marginales de  $X$  y  $Y$  son definidas por:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy, \quad -\infty < x < \infty \\ f_Y(y) &= \int_{-\infty}^{\infty} f(x, y) dx, \quad -\infty < y < \infty \end{aligned} \quad (4.1.2)$$

Toda función  $f(x, y)$  que satisface  $f(x, y) \geq 0$ , para todo  $(X, Y) \in \mathfrak{R}^2$ , y:

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

se dice que es la función de densidad conjunta para algún vector aleatorio bivariado  $(X, Y)$ .

**Ejemplo 4.1.6** Se define la función de densidad conjunta por:

$$f(x, y) = \begin{cases} 6xy^2 & 0 < x < 1, \quad 0 < y < 1 \\ 0 & \text{de otro modo} \end{cases}$$

Si se desea calcular  $\Pr(X + Y \geq 1)$ , entonces  $A = \{(x, y) : x + y \geq 1\}$ :

$$\Pr(X + Y \geq 1) = \iint_A f(x, y) dx dy = \int_{y=0}^{y=1} \int_{x=1-y}^{x=1} 6xy^2 dx dy = \frac{9}{10}$$

Usando 4.1.2 puede obtenerse la función de densidad marginal de  $X$  y  $Y$ :

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dx dy = \int_{y=0}^{y=1} 6xy^2 dy = 2xy^3 \Big|_{y=0}^{y=1} = 2x$$

Esta función de densidad de  $X$  puede ser usada para calcular probabilidades, por ejemplo:

$$\Pr(1/2 < X < 3/4) = \int_{x=1/2}^{x=3/4} 2x dx = \frac{5}{10}$$

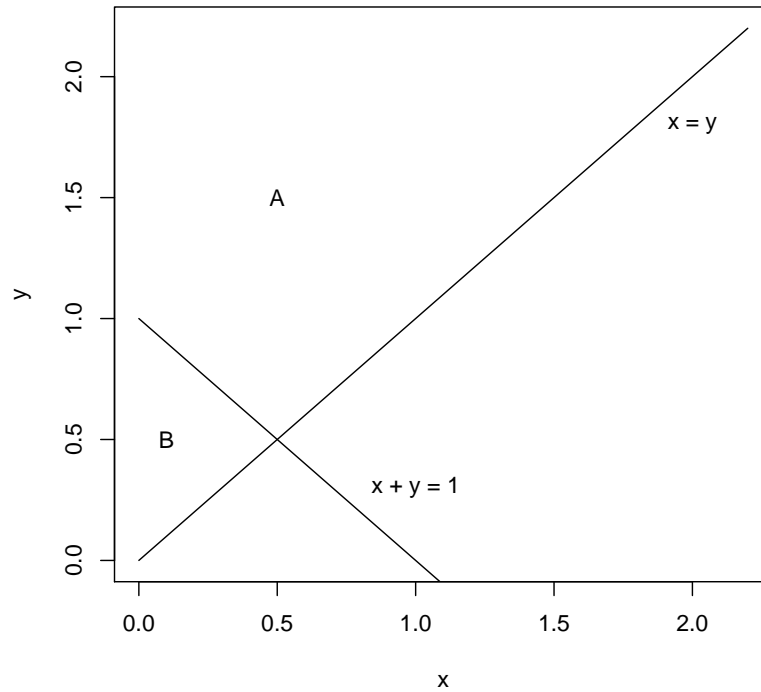
**Ejemplo 4.1.7** Sea  $f(x, y) = e^{-y}$ ,  $0 < x < y$ . Aparentemente  $f(x, y)$  no depende de  $X$  sin embargo:

$$f(x, y) = e^{-y} I_{\{0 < x < y\}}(x, y)$$

Para calcular  $\Pr(X + Y \geq 1)$  se define  $A = \{(x, y) : x + y \geq 1\}$ . El proceso de integración es mucho más sencillo si se hace sobre el conjunto  $B = \{(x, y) : x + y < 1\}$ :

$$\begin{aligned} \Pr(X + Y \geq 1) &= 1 - \Pr(X + Y < 1) \\ &= 1 - \int_{x=0}^{x=1/2} \int_{y=x}^{y=1-x} e^{-y} dy dx \\ &= 1 - \int_0^{1/2} (e^{-x} - e^{-(1-x)}) dx \\ &= 0,845182 \end{aligned}$$

Figura 4.1: Regiones de integración para el ejemplo 4.1.7



La función de densidad conjunta de  $(X, Y)$  puede describirse completamente usando la función de distribución acumulada conjunta  $F(x, y)$  definida por:

$$F(x, y) = \Pr(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) dt ds$$

para todo  $(x, y) \in \mathfrak{R}^2$ . Usando el teorema fundamental del cálculo:

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y) \quad (4.1.3)$$

para todos los puntos de continuidad de  $f(x, y)$ .

## 4.2. Distribuciones condicionales e independencia

**Definición 4.2.1** Sea  $(X, Y)$  un vector aleatorio bivariado discreto con función de probabilidad conjunta  $f(x, y)$  y funciones de probabilidad marginales  $f_X(x)$  y  $f_Y(y)$ . Para todo  $x$  tal que  $f_X(x) > 0$ , la *función de probabilidad condicional de  $Y$  dado  $X = x$*  es la función de  $y$  denotada por  $f(y|x)$  y definida por:

$$f(y|x) = \Pr(Y = y|X = x) = \frac{f(x, y)}{f_X(x)}$$

Para todo  $y$  tal que  $f_Y(y) > 0$ , la *función de probabilidad condicional de  $x$  dado que  $Y = y$*  es la función de  $x$  denotada por  $f(x|y)$  y definida por:

$$f(x|y) = \Pr(X = x|Y = y) = \frac{f(x, y)}{f_Y(y)}$$

**Definición 4.2.2** Sea  $(X, Y)$  un vector aleatorio continuo bivariado con función de densidad conjunta  $f(x, y)$  y funciones de densidad marginales  $f_X(x)$  y  $f_Y(y)$ . Para todo  $x$  tal que  $f_X(x) > 0$ , la *función de densidad condicional de  $Y$  dado que  $X = x$*  es la función de  $y$  denotada por  $f(y|x)$  y definida por:

$$f(y|x) = \frac{f(x, y)}{f_X(x)}$$

Para todo  $y$  tal que  $f_Y(y) > 0$ , la *función de densidad condicional de  $X$  dado que  $Y = y$*  es la función de  $x$  denotada por  $f(x|y)$  y definida por:

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}$$

Si  $g(Y)$  es una función de  $Y$ , entonces *el valor esperado condicional de  $g(Y)$  dado que  $X = x$*  se denota por  $E[g(Y)|x]$  y se define por:

$$E[g(Y)|x] = \sum_y g(y)f(y|x) \quad \text{y} \quad E[g(Y)|x] = \int_{-\infty}^{\infty} g(y)f(y|x)dy$$

El valor esperado condicional tiene todas las propiedades del valor esperado vistas en el teorema 3.3.1.

**Ejemplo 4.2.1** Como en el ejemplo 4.1.7, sea un vector aleatorio continuo  $(X, Y)$  con función de densidad conjunta  $f(x, y) = e^{-y}, 0 < x < y$ . Suponga se desea calcular la función de densidad condicional de  $Y$  dado  $X = x$ . La función de densidad marginal de  $X$  se puede calcular como sigue:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_{y=x}^{y=\infty} e^{-y} dy = e^{-x}, \quad x > 0$$

entonces  $X \sim \mathcal{E}(\beta = 1)$ . Luego:

$$f(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{e^{-y}}{e^{-x}} = e^{-(y-x)}, \quad y > x$$

Dado  $X = x$ , la distribución de  $Y$  es exponencial donde  $x$  es el parámetro de locación y  $\beta = 1$  es el parámetro de escala. La distribución condicional de  $Y$  es diferente para cada valor de  $x$ . Además:

$$E[Y|X = x] = \int_{y=x}^{y=\infty} ye^{-(y-x)} dy = 1 + x$$

La variancia de la función de densidad  $f(y|x)$  es llamada *variancia condicional de  $Y$  dado  $X = x$* . Usando la notación  $\text{Var}(Y|X)$  se tiene:

$$\begin{aligned} \text{Var}(Y|X) &= E[Y^2|X] - (E[Y|X])^2 \\ &= \int_x^{\infty} y^2 e^{-(y-x)} dy - \left( \int_x^{\infty} ye^{-(y-x)} dy \right)^2 \\ &= 1 \end{aligned}$$

En este caso la variancia condicional de  $Y$  dado  $X = x$  es la misma para todos los valores de  $x$ . Esta variancia condicional puede compararse con la *variancia no condicional de  $Y$* . La distribución marginal de  $Y$  es  $\mathcal{G}(2, 1)$ , la cual tiene  $\text{Var}(Y) = 2$ . Dado el valor  $X = x$ , la variabilidad en  $Y$  se reduce considerablemente.

**Definición 4.2.3** Sea  $(X, Y)$  un vector aleatorio bivariado con función de probabilidad o función de densidad conjunta  $f(x, y)$  y funciones de probabilidad o densidad marginales  $f_X(x)$  y  $f_Y(y)$ . Entonces  $X$  y  $Y$  son llamadas *variables aleatorias independientes* si, para todo  $x \in \mathfrak{R}$  y  $y \in \mathfrak{R}$ ,

$$f(x, y) = f_X(x)f_Y(y) \tag{4.2.1}$$



Si  $X$  y  $Y$  son independientes, la función de probabilidad o densidad condicional  $Y$  dado  $X = x$  es:

$$f(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x)$$

para cualquier valor de  $x$ . Así, para todo  $A \subset \mathfrak{R}$  y  $x \in \mathfrak{R}$ ,

$$\Pr(Y \in A|x) = \int_A f(y|x)dy = \int_A f_Y(y)dy = \Pr(Y \in A)$$

El saber que  $X = x$  no brinda información adicional acerca de  $Y$ .

**Lema 4.2.1** Sea  $(X, Y)$  un vector aleatorio bivariado con función de probabilidad o densidad conjunta  $f(x, y)$ . Entonces  $X$  y  $Y$  son variables aleatorias independientes sí y solo si existen funciones  $g(x)$  y  $h(y)$  tales que, para todo  $x \in \mathfrak{R}$  y  $y \in \mathfrak{R}$ ,

$$f(x, y) = g(x)h(y)$$

**Prueba:** Si se define  $g(x) = f_X(x)$  y  $h(y) = f_Y(y)$  y usando 4.2.1 es fácil probar una de las direcciones. Para probar la otra dirección, suponga que  $f(x, y) = g(x)h(y)$ . Se define  $\int_{-\infty}^{\infty} g(x)dx = c$  y  $\int_{-\infty}^{\infty} h(y)dy = d$ , donde las constantes  $c$  y  $d$  satisfacen:

$$cd = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)dxdy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy \quad (4.2.2)$$

Además, las funciones de densidad marginales están dadas por:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} g(x)h(y)dy = g(x)d \\ f_Y(y) &= \int_{-\infty}^{\infty} g(x)h(y)dx = h(y)c \end{aligned} \quad (4.2.3)$$

Luego, usando 4.2.2 y 4.2.3, se tiene:

$$f(x, y) = g(x)h(y) = g(x)h(y)cd = f_X(x)f_Y(y)$$

demostrando que  $X$  y  $Y$  son independientes. Reemplazando las integrales por sumatorias se prueba el lema para vectores aleatorios discretos.

**Ejemplo 4.2.2** Considere la función de densidad conjunta:

$$f(x, y) = \frac{1}{384}x^2y^4e^{-y-(x/2)}, \quad x > 0, y > 0$$

Si se definen:

$$g(x) = \begin{cases} x^2e^{-x/2} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$h(y) = \begin{cases} y^4e^{-y}/384 & y > 0 \\ 0 & y \leq 0 \end{cases}$$

entonces  $f(x, y) = g(x)h(y)$  para todo  $x \in \mathfrak{R}$  y  $y \in \mathfrak{R}$ . Por el lema 4.2.1, se concluye que  $X$  y  $Y$  son variables aleatorias independientes. Notar que no fué necesario calcular las funciones de densidad marginales.

**Teorema 4.2.1** Sean  $X$  y  $Y$  variables aleatorias independientes:

- a. Para todo  $A \subset \mathfrak{R}$  y  $B \subset \mathfrak{R}$ ,  $\Pr(X \in A, Y \in B) = \Pr(X \in A)\Pr(Y \in B)$ , esto es, los eventos  $\{X \in A\}$  y  $\{Y \in B\}$  son independientes.
- b. Sea  $g(x)$  una función que depende sólo de  $x$  y  $h(y)$  una función que depende sólo de  $y$ . Entonces:

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

**Prueba:** Notar que:

$$\begin{aligned} E[g(X)h(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f(x, y)dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dx dy \\ &= \int_{-\infty}^{\infty} h(y)f_Y(y) \int_{-\infty}^{\infty} g(x)f_X(x)dx dy \\ &= \left( \int_{-\infty}^{\infty} g(x)f_X(x)dx \right) \left( \int_{-\infty}^{\infty} h(y)f_Y(y)dy \right) \\ &= E[g(X)]E[h(Y)] \end{aligned}$$

Sea  $g(x)$  la función indicadora del conjunto  $A$  y sea  $h(y)$  la función indicadora del conjunto  $B$ . Notar que  $g(x)h(y)$  es la función indicadora del

conjunto  $C \subset \mathfrak{R}^2$  definido por  $C = \{(x, y) : x \in A, y \in B\}$ . Notar que para una función indicadora como  $g(x)$ ,  $E[g(X)] = \Pr(X \in A)$ . Usando el resultado anterior se tiene:

$$\begin{aligned} \Pr(X \in A, Y \in B) &= \Pr((X, Y) \in C) = E[g(X)h(Y)] \\ &= E[g(X)]E[h(Y)] = \Pr(X \in A) \Pr(Y \in B) \end{aligned}$$

**Ejemplo 4.2.3** Sean  $X$  y  $Y$  variables aleatorias independientes con distribución  $\mathcal{E}(1)$ . Por el teorema 4.3.1 se tiene:

$$\Pr(X \geq 4, Y < 3) = \Pr(X \geq 4) \Pr(Y < 3)$$

Sean  $g(x) = x^2$  y  $h(y) = y$ , se tiene que:

$$E[X^2Y] = E[X^2]E[Y] = (\text{Var}(X) + E[X]^2) E[Y] = (1 + 1^2)1 = 2$$

**Teorema 4.2.2** Sean  $X$  y  $Y$  variables aleatorias independientes con funciones generatrices de momentos  $M_X(t)$  y  $M_Y(t)$  respectivamente. Entonces la función generatriz de momentos de la variable aleatoria  $Z = X + Y$  es:

$$M_Z(t) = M_X(t)M_Y(t)$$

**Prueba:** Usando la definición de función generatriz de momentos:

$$M_Z(t) = E[e^{tZ}] = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t)$$

**Ejemplo 4.2.4** Sean  $X \sim \mathcal{N}(\mu, \sigma^2)$  y  $Y \sim \mathcal{N}(\gamma, \tau^2)$  variables aleatorias independientes. Hallar la distribución de  $Z = X + Y$ .

### 4.3. Transformaciones bivariadas

Sea  $(X, Y)$  un vector aleatorio bivariado con una distribución conocida. Considere un nuevo vector aleatorio bivariado  $(U, V)$  definido por  $U = g_1(X, Y)$  y  $V = g_2(X, Y)$  donde  $g_1(x, y)$  y  $g_2(x, y)$  son funciones específicas. Si  $\mathcal{B}$  es cualquier subconjunto de  $\mathfrak{R}^2$ , entonces  $(U, V) \in \mathcal{B}$  sí y solo si  $(X, Y) \in \mathcal{A}$ , donde  $\mathcal{A} = \{(x, y) : (g_1(x, y), g_2(x, y)) \in \mathcal{B}\}$ . Luego  $\Pr((U, V) \in \mathcal{B}) = \Pr((X, Y) \in \mathcal{A})$  y la distribución conjunta de  $(U, V)$  se determina completamente usando la distribución conjunta de  $(X, Y)$ .

### 4.3.1. Caso discreto

Si  $(X, Y)$  es un vector aleatorio bivariado discreto, entonces existe solo un conjunto numerable de valores para los que la función de probabilidad conjunta de  $(X, Y)$  es positiva, digamos el conjunto  $\mathcal{A}$ . Se define el conjunto  $\mathcal{B} = \{(u, v) : u = g_1(x, y), v = g_2(x, y) \text{ para algún } (x, y) \in \mathcal{A}\}$ . Entonces  $\mathcal{B}$  es el conjunto numerable de posibles valores para el vector aleatorio discreto  $(U, V)$ . Si para todo  $(u, v) \in \mathcal{B}$ ,  $\mathcal{A}_{uv}$  se define como  $\{(x, y) \in \mathcal{A} : g_1(x, y) = u, g_2(x, y) = v\}$  entonces la función de probabilidad conjunta de  $(U, V)$ ,  $f_{U,V}(u, v)$ , puede calcularse a partir de la función de probabilidad conjunta de  $(X, Y)$  por:

$$f_{U,V}(u, v) = \Pr((x, y) \in \mathcal{A}_{uv}) = \sum_{(x,y) \in \mathcal{A}_{uv}} f_{X,Y}(x, y) \quad (4.3.1)$$

**Ejemplo 4.3.1** Sean  $X \sim \mathcal{P}(\theta)$  y  $Y \sim \mathcal{P}(\lambda)$  variables aleatorias independientes. Hallar la función de probabilidad de  $X + Y$ .

### 4.3.2. Caso continuo

Si  $(X, Y)$  es un vector aleatorio continuo con función de densidad conjunta  $f_{X,Y}(x, y)$ , entonces la función de densidad conjunta de  $(U, V)$  puede expresarse en términos de  $f_{X,Y}(x, y)$ , en forma análoga a 3.2.1. Sea  $\mathcal{A} = \{(x, y) : f_{X,Y}(x, y) > 0\}$  y  $\mathcal{B} = \{(u, v) : u = g_1(x, y), v = g_2(x, y) \text{ para todo } (x, y) \in \mathcal{A}\}$ . La función de densidad conjunta  $f_{U,V}(u, v)$  será positiva sobre el conjunto  $\mathcal{B}$ . Si se asume que  $u = g_1(x, y)$  y  $v = g_2(x, y)$  definen transformaciones uno a uno de  $\mathcal{A}$  hacia  $\mathcal{B}$  entonces dichas transformaciones serán sobreyectivas según la definición de  $\mathcal{B}$ . Entonces para todo  $(u, v) \in \mathcal{B}$  existe solo un  $(x, y) \in \mathcal{A}$  tal que  $(u, v) = (g_1(x, y), g_2(x, y))$ . Para cada transformación uno a uno y sobreyectiva, se pueden resolver las ecuaciones  $u = g_1(x, y)$  y  $v = g_2(x, y)$  para  $x$  y  $y$  en términos de  $u$  y  $v$ . Denotemos estas transformaciones inversas por  $x = h_1(u, v)$  y  $y = h_2(u, v)$ .

El rol que tuvo la derivada en el caso univariado ahora lo asume una cantidad llamada el *Jacobiano de la transformación*. Esta función de  $(u, v)$ , denotada por  $J$ , es el *determinante de la matriz* de derivadas parciales. Se define por:

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial y}{\partial u} \frac{\partial x}{\partial v}$$

donde:

$$\frac{\partial x}{\partial u} = \frac{\partial h_1(u, v)}{\partial u}, \quad \frac{\partial x}{\partial v} = \frac{\partial h_1(u, v)}{\partial v}, \quad \frac{\partial y}{\partial u} = \frac{\partial h_2(u, v)}{\partial u} \quad \text{y} \quad \frac{\partial y}{\partial v} = \frac{\partial h_2(u, v)}{\partial v}$$

Se asume que  $J$  es diferente de cero sobre  $\mathcal{B}$ . Entonces la función de densidad conjunta de  $(U, V)$  está dada por:

$$f_{U,V}(u, v) = f_{X,Y}(h_1(u, v), h_2(u, v)) |J| \quad (4.3.2)$$

donde  $|J|$  es el valor absoluto de  $J$ .

**Ejemplo 4.3.2** Sean  $X$  y  $Y$  variables aleatorias independientes con distribución normal estándar. Se definen las transformaciones  $U = X + Y$  y  $V = X - Y$ . Hallar  $f(u, v)$  y probar que  $U$  y  $V$  son variables aleatorias independientes.

**Teorema 4.3.1** Sean  $X$  y  $Y$  variables aleatorias independientes. Sea  $g(x)$  una función que depende sólo de  $x$  y  $h(y)$  una función que sólo depende de  $y$ . Entonces las variables aleatorias  $U = g(X)$  y  $V = h(Y)$  son independientes.

En muchas situaciones las transformaciones de interés no son uno a uno. Sea  $\mathcal{A} = \{(x, y) : f_{X,Y}(x, y) > 0\}$ . Suponga que  $A_0, A_1, \dots, A_k$  forma una partición de  $\mathcal{A}$ . El conjunto  $A_0$ , que podría ser vacío, satisface  $\Pr((X, Y) \in A_0) = 0$ . Las transformaciones  $U = g_1(X, Y)$  y  $V = g_2(X, Y)$  son uno a uno, desde  $A_i$  hacia  $\mathcal{B}$  para cada  $i = 1, 2, \dots, k$ . Entonces para cada  $i$  se pueden hallar las funciones inversas desde  $\mathcal{B}$  hacia  $A_i$ . Si se denotan las  $i$ -ésimas inversas por  $x = h_{1i}(u, v)$  y  $y = h_{2i}(u, v)$ , éstas dan para  $(u, v) \in \mathcal{B}$  un único  $(x, y) \in A_i$  tal que  $(u, v) = (g_1(x, y), g_2(x, y))$ . Sea  $J_i$  el Jacobiano calculado a partir de las  $i$ -ésimas inversas. Entonces se tiene la siguiente representación de la función de densidad conjunta  $f_{U,V}(u, v)$ :

$$f_{U,V}(u, v) = \sum_{i=1}^k f_{X,Y}(h_{1i}(u, v), h_{2i}(u, v)) |J_i| \quad (4.3.3)$$

**Ejemplo 4.3.3** Sean  $X$  y  $Y$  variables aleatorias independientes con distribución normal estándar. Considere las transformaciones  $U = X/Y$  y  $V = |Y|$ . Hallar la función de densidad conjunta de  $U$  y  $V$ . ¿Cuál es la distribución marginal de  $U$ ?

## 4.4. Modelos jerárquicos y mixtura de distribuciones

En las secciones anteriores se consideraron variables aleatorias cuya distribución dependía de un conjunto de parámetros. Sin embargo, se puede considerar la posibilidad de que algunos de estos parámetros sean también variables aleatorias con una determinada distribución.

**Ejemplo 4.4.1** Un insecto pone un número grande de huevos, cada uno con probabilidad de supervivencia  $p$ . En promedio, ¿cuántos huevos sobrevivirán?

Sean  $X =$  Número de huevos sobrevivientes, y  $Y =$  Número de huevos puestos. Se tiene el siguiente modelo *jerárquico*:

$$X|Y \sim \mathcal{BI}(Y, p) \quad Y \sim \mathcal{P}(\lambda)$$

La variable de interés tiene la siguiente distribución:

$$\begin{aligned} \Pr(X = x) &= \sum_{y=x}^{\infty} \Pr(X = x, Y = y) \\ &= \sum_{y=x}^{\infty} \Pr(X = x|Y = y) \Pr(Y = y) \\ &= \sum_{y=x}^{\infty} \left[ \binom{y}{x} p^x (1-p)^{y-x} \right] \left[ \frac{e^{-\lambda} \lambda^y}{y!} \right] \\ &= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{y=x}^{\infty} \frac{((1-p)\lambda)^{y-x}}{(y-x)!} \\ &= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{t=0}^{\infty} \frac{((1-p)\lambda)^t}{t!} \\ &= \frac{(\lambda p)^x e^{-\lambda}}{x!} e^{(1-p)\lambda} \\ &= \frac{(\lambda p)^x e^{-\lambda p}}{x!} \end{aligned}$$

es decir que  $X \sim \mathcal{P}(\lambda p)$ . Todo proceso de inferencia marginal sobre  $X$  se hace a través de la distribución de Poisson en la que  $Y$  no es parte del proceso.

Respondiendo a la pregunta inicial, se tiene que en promedio sobrevivirán  $E[X] = \lambda p$  huevos.

Los cálculos anteriores se pueden simplificar de manera importante usando el siguiente teorema.

**Teorema 4.4.1** Si  $X$  y  $Y$  son variables aleatorias, entonces:

$$E[X] = E_Y[E_X[X|Y]] \quad (4.4.1)$$

siempre que los esperados existan.

**Prueba:** Sea  $f(x, y)$  la función de densidad de  $X$  y  $Y$ . Por definición se tiene:

$$\begin{aligned} E[X] &= \int_y \int_x x f(x, y) dx dy \\ &= \int_y \left[ \int_x x f(x|y) dx \right] f_Y(y) dy \\ &= \int_y E[X|Y] f_Y(y) dy \\ &= E_Y[E_X[X|Y]] \end{aligned}$$

Volviendo al ejemplo anterior:

$$E[X] = E_Y[E_X[X|Y]] = E_Y[pY] = pE_Y[Y] = p\lambda$$

**Definición 4.4.1** Una variable aleatoria  $X$  se dice que tiene una *mezcla de distribuciones* si su distribución depende de una cantidad que también tiene distribución.

En el ejemplo 4.4.1 la distribución  $\mathcal{P}(\lambda p)$  es una *mezcla de distribuciones* ya que es el resultado de combinar una distribución  $\mathcal{BI}(Y, p)$  con  $Y \sim \mathcal{P}(\lambda)$ .

**Ejemplo 4.4.2** Considere una generalización del ejemplo 4.4.1:

$$X|Y \sim \mathcal{BI}(Y, p) \quad Y|\Lambda \sim \mathcal{P}(\Lambda) \quad \Lambda \sim \mathcal{E}(\beta)$$

entonces:

$$E[X] = E_Y[E_X[X|Y]] = E_Y[pY] = pE_\Lambda[E_Y[Y|\Lambda]] = pE_\Lambda[\Lambda] = p\beta$$

Este modelo de tres niveles también puede ser expresado como uno de dos jerarquías combinando los últimos dos estados.

Se tiene que:

$$\begin{aligned}
 \Pr(Y = y) &= \int_0^\infty f(y, \lambda) d\lambda \\
 &= \int_0^\infty f(y|\lambda) f(\lambda) d\lambda \\
 &= \int_0^\infty \frac{e^{-\lambda} \lambda^y}{y!} \frac{1}{\beta} e^{-\lambda/\beta} d\lambda \\
 &= \frac{1}{\beta y!} \int_0^\infty \lambda^y e^{-\lambda(1+\beta^{-1})} d\lambda \\
 &= \frac{1}{1+\beta} \left(1 - \frac{1}{1+\beta}\right)^y
 \end{aligned}$$

La expresión para la función de probabilidad de  $Y$  pertenece a la distribución binomial negativa.

La jerarquía de tres estados de este ejemplo es equivalente a la siguiente jerarquía de dos estados:

$$X|Y \sim \mathcal{BI}(Y, p) \quad Y \sim \mathcal{BN}(r = 1, p = (1 + \beta)^{-1})$$

**Ejemplo 4.4.3** Una generalización para ensayos de Bernoulli considera que la probabilidad de éxito no es constante, manteniendo los ensayos independientes. Un modelo estándar para esta situación es:

$$X_i|p_i \sim \mathcal{B}(p_i) \quad i = 1, 2, \dots, n \quad p_i \sim \mathcal{BE}(\alpha, \beta)$$

Este modelo puede ser apropiado, por ejemplo, si se mide el éxito de una droga en  $n$  pacientes y debido a que cada paciente es diferente no es posible asumir que la probabilidad de éxito sea constante.

Una variable aleatoria de interés es  $Y = \sum_{i=1}^n X_i$ , el número de éxitos, cuya media es:

$$E[Y] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n E_P[E_{X_i}[X_i|p_i]] = \sum_{i=1}^n E_P[p_i] = \sum_{i=1}^n \frac{\alpha}{\alpha + \beta} = \frac{n\alpha}{\alpha + \beta}$$

También es posible obtener una expresión para calcular  $\text{Var}(X)$  usando el valor esperado y la variancia condicional de  $X|Y$ .



**Teorema 4.4.2** Sean  $X$  y  $Y$  dos variables aleatorias, entonces:

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y]) \quad (4.4.2)$$

**Prueba:** Por definición:

$$\begin{aligned} \text{Var}[X] &= E[(X - E[X])^2] \\ &= E[(X - E[X|Y] + E[X|Y] - E[X])^2] \\ &= E[(X - E[X|Y])^2] + E[(E[X|Y] - E[X])^2] \\ &= E[\text{Var}(X|Y)] + \text{Var}(E[X|Y]) \end{aligned}$$

**Ejemplo 4.4.4** Calcular  $\text{Var}(Y)$  del ejemplo 4.4.3.

## 4.5. Covarianza y correlación

En las secciones anteriores se mencionó sobre la presencia o ausencia de relación entre dos variables aleatorias. En caso de existir relación esta podría ser fuerte o débil. En esta sección se presentan dos indicadores que miden de la fuerza de la relación existente entre dos variables aleatorias, la *covarianza* y la *correlación*.

**Definición 4.5.1** La *covarianza* de  $X$  y  $Y$  es el número definido por:

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

**Definición 4.5.2** La *correlación* de  $X$  y  $Y$  es el número definido por:

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

El valor  $\rho_{XY}$  también es llamado *coeficiente de correlación*.

**Teorema 4.5.1** Sean  $X$  y  $Y$  dos variables aleatorias, entonces:

$$\text{Cov}(X, Y) = E[XY] - \mu_X \mu_Y$$

**Teorema 4.5.2** Si  $X$  y  $Y$  son variables aleatorias independientes, entonces  $\text{Cov}(X, Y) = 0$  y  $\rho_{XY} = 0$ .

**Teorema 4.5.3** Si  $X$  y  $Y$  son variables aleatorias,  $a$  y  $b$  son constantes, entonces:

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$

**Teorema 4.5.4** Si  $X$  y  $Y$  son variables aleatorias entonces  $-1 \leq \rho_{XY} \leq 1$ .

**Definición 4.5.3** Sean  $-\infty < \mu_X < \infty$ ,  $-\infty < \mu_Y < \infty$ ,  $0 < \sigma_X$ ,  $0 < \sigma_Y$ , y  $-1 < \rho < 1$  números reales. La *función de densidad normal bivariada con medias  $\mu_X$ ,  $\mu_Y$  varianzas  $\sigma_X^2$ ,  $\sigma_Y^2$  y coeficiente de correlación  $\rho$*  esta dada por:

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \left( \frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho \left( \frac{x-\mu_X}{\sigma_X} \right) \left( \frac{y-\mu_Y}{\sigma_Y} \right) + \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\}$$

para  $-\infty < x < \infty, -\infty < y < \infty$ .

Algunas de las propiedades de la función de densidad conjunta anterior son:

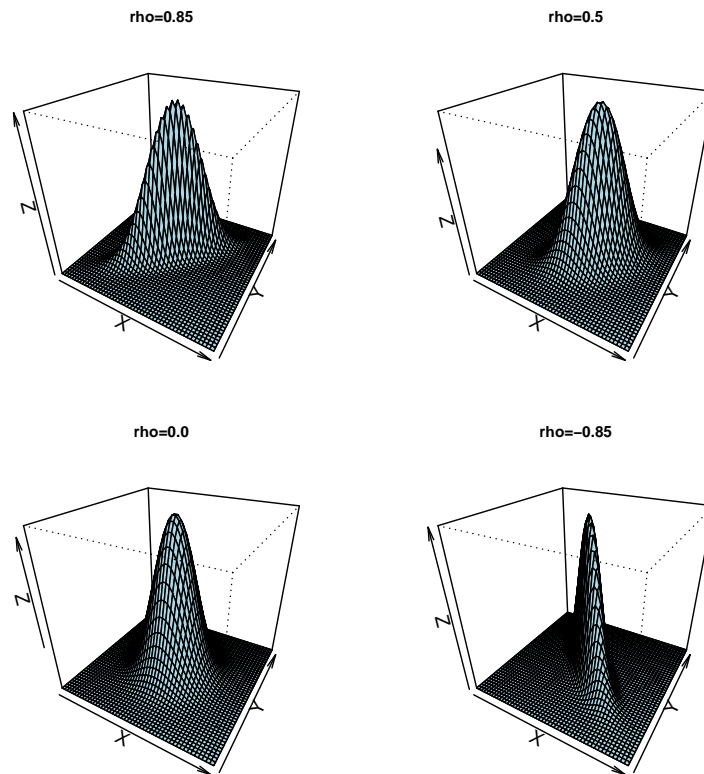
- a. La distribución marginal de  $X$  es  $\mathcal{N}(\mu_X, \sigma_X^2)$ .
- b. La distribución marginal de  $Y$  es  $\mathcal{N}(\mu_Y, \sigma_Y^2)$ .
- c. El coeficiente de correlación entre  $X$  y  $Y$  es  $\rho_{X,Y} = \rho$ .
- d. Para  $a$  y  $b$  constantes, la distribución de  $aX + bY$  es:

$$\mathcal{N}(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y)$$

- e. Todas las distribuciones condicionales también son normales. Por ejemplo:

$$f(y|x) \sim \mathcal{N} \left( \mu_Y + \rho \left( \frac{\sigma_Y}{\sigma_X} \right) (x - \mu_X), \sigma_Y^2(1 - \rho^2) \right)$$

Figura 4.2: Distribución normal bivariada



## 4.6. Distribuciones multivariadas

El vector aleatorio  $\mathbf{X} = (X_1, \dots, X_n)$  tiene un espacio muestral que es subconjunto de  $\mathfrak{R}^n$ . Si  $(X_1, \dots, X_n)$  es un vector aleatorio discreto (el espacio muestral es numerable) entonces la *función de probabilidad conjunta* de  $(X_1, \dots, X_n)$  es la función definida por  $f(\mathbf{x}) = f(x_1, \dots, x_n) = \Pr(X_1 = x_1, \dots, X_n = x_n)$  para cada  $(x_1, \dots, x_n) \in \mathfrak{R}^n$ . Entonces para todo  $A \subset \mathfrak{R}^n$ ,

$$\Pr(\mathbf{X} \in A) = \sum_{\mathbf{x} \in A} f(\mathbf{x}) \quad (4.6.1)$$

si  $(X_1, \dots, X_n)$  es un vector aleatorio continuo, entonces la *función de*

densidad conjunta de  $(x_1, \dots, x_n)$  es la función  $f(x_1, \dots, x_n)$  que satisface:

$$\Pr(\mathbf{X} \in A) = \int \cdots \int_A f(\mathbf{x}) d\mathbf{x} = \int \cdots \int_A f(x_1, \dots, x_n) dx_1 \cdots dx_n \quad (4.6.2)$$

Sea  $g(\mathbf{x}) = g(x_1, \dots, x_n)$  una función definida sobre el espacio muestral de  $\mathbf{X}$ . Entonces  $g(\mathbf{X})$  es una variable aleatoria y su *valor esperado* es:

$$E[g(\mathbf{X})] = \sum_{\mathbf{x} \in \mathfrak{R}^n} g(\mathbf{x}) f(\mathbf{x}) \quad \text{y} \quad E[g(\mathbf{X})] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \quad (4.6.3)$$

en el caso discreto y continuo respectivamente.

La *función de probabilidad marginal* o *función de densidad marginal* para algún subconjunto de coordenadas de  $(X_1, \dots, X_n)$  puede calcularse sumando o integrando la función de probabilidad o función de densidad conjunta sobre las otras coordenadas. Por ejemplo, la distribución marginal de  $(X_1, \dots, X_k)$ , las primeras  $k$ -coordenadas de  $(X_1, \dots, X_n)$ , está dada por la función de probabilidad o función de densidad:

$$f(x_1, \dots, x_k) = \sum_{(x_{k+1}, \dots, x_n) \in \mathfrak{R}^{n-k}} f(x_1, \dots, x_n) \quad (4.6.4)$$

$$f(x_1, \dots, x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_{k+1} \cdots dx_n \quad (4.6.5)$$

para todo  $(x_1, \dots, x_k) \in \mathfrak{R}^k$ . La *función de probabilidad* o *función de densidad condicional* de un subconjunto de coordenadas de  $(x_1, \dots, x_n)$ , dados los valores de las coordenadas restantes, se obtiene dividiendo la función de probabilidad o función de densidad conjunta por la función de probabilidad o función de densidad marginal de las coordenadas restantes. Así, por ejemplo, si  $f(x_1, \dots, x_n) > 0$ , la función de probabilidad o función de densidad condicional de  $(x_{k+1}, \dots, x_n)$  dados  $X_1 = x_1, \dots, X_k = x_k$  es la función de  $(x_{k+1}, \dots, x_n)$  definida por:

$$f(x_{k+1}, \dots, x_n | x_1, \dots, x_k) = \frac{f(x_1, \dots, x_n)}{f(x_1, \dots, x_k)} \quad (4.6.6)$$

**Ejemplo 4.6.1** Sea la función de densidad conjunta:

$$f(x_1, x_2, x_3, x_4) = \begin{cases} \frac{3}{4}(x_1^2 + x_2^2 + x_3^2 + x_4^2) & 0 < x_i < 1 \\ 0 & \text{de otro modo} \end{cases}$$

del vector aleatorio  $(X_1, \dots, X_4)$ . Se puede obtener la función de densidad marginal de  $(X_1, X_2)$  integrando las variables  $X_3$  y  $X_4$ :

$$\begin{aligned} f(x_1, x_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, \dots, x_4) dx_3 dx_4 \\ &= \int_0^1 \int_0^1 \frac{3}{4}(x_1^2 + x_2^2 + x_3^2 + x_4^2) dx_3 dx_4 \\ &= \frac{3}{4}(x_1^2 + x_2^2) + \frac{1}{2} \end{aligned}$$

para  $0 < x_1 < 1, 0 < x_2 < 1$ . Cualquier probabilidad o valor esperado que incluya solo  $X_1$  y  $X_2$  puede calcularse usando esta función de distribución marginal. Por ejemplo:

$$\begin{aligned} E[X_1 X_2] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2) dx_1 dx_2 \\ &= \int_0^1 \int_0^1 x_1 x_2 \left( \frac{3}{4}(x_1^2 + x_2^2) + \frac{1}{2} \right) dx_1 dx_2 \\ &= \frac{5}{16} \end{aligned}$$

Para todo  $(x_1, x_2)$  con  $0 < x_1 < 1, 0 < x_2 < 1, f(x_1, x_2) > 0$  y la función de densidad condicional de  $(X_3, X_4)$  dados  $X_1 = x_1$  y  $X_2 = x_2$  puede obtenerse usando 4.6.6:

$$\begin{aligned} f(x_3, x_4 | x_1, x_2) &= \frac{f(x_1, x_2, x_3, x_4)}{f(x_1, x_2)} \\ &= \frac{\frac{3}{4}(x_1^2 + x_2^2 + x_3^2 + x_4^2)}{\frac{3}{4}(x_1^2 + x_2^2) + \frac{1}{2}} \\ &= \frac{x_1^2 + x_2^2 + x_3^2 + x_4^2}{x_1^2 + x_2^2 + \frac{2}{3}} \end{aligned}$$

**Definición 4.6.1** Sean  $n$  y  $m$  enteros positivos y sean  $p_1, \dots, p_n$  números tales que  $0 \leq p_i \leq 1, i = 1, \dots, n$  y  $\sum_{i=1}^n p_i = 1$ . Entonces el vector aleatorio  $(X_1, \dots, X_n)$  tiene *distribución multinomial con  $m$  ensayos y probabilidades de celda  $p_1, \dots, p_n$*  si su función de probabilidad conjunta es:

$$f(x_1, \dots, x_n) = \frac{m!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n} = m! \prod_{i=1}^n \frac{p_i^{x_i}}{x_i!}$$

sobre el conjunto de  $(x_1, \dots, x_n)$  tal que cada  $x_i$  es un entero no negativo y  $\sum_{i=1}^n x_i = m$ . El factor  $m!/(x_1! \cdots x_n!)$  es llamado *coeficiente multinomial*.

**Ejemplo 4.6.2** Considere el experimento aleatorio que consiste en lanzar 10 veces un dado. Suponga que el dado no se encuentra balanceado, tal que la probabilidad de observar  $i$  es  $\frac{i}{21}$ . Sea el vector aleatorio  $(X_1, \dots, X_6)$  tal que  $X_i$  representa el número de lanzamientos en los que se observó el número  $i$ , entonces su distribución es multinomial con  $m = 10$  lanzamientos,  $n = 6$  posibles resultados y probabilidades de celda  $p_1 = \frac{1}{21}, p_2 = \frac{2}{21}, \dots, p_6 = \frac{6}{21}$ . La fórmula anterior puede usarse para calcular la probabilidad de obtener el número 6 en cuatro lanzamientos, el número 5 en tres lanzamientos, el número 4 en dos lanzamientos y el número 3 en solo un lanzamiento:

$$\begin{aligned} f(0, 0, 1, 2, 3, 4) &= \frac{10!}{0!0!1!2!3!4!} \left(\frac{1}{21}\right)^0 \left(\frac{2}{21}\right)^0 \left(\frac{3}{21}\right)^1 \left(\frac{4}{21}\right)^2 \left(\frac{5}{21}\right)^3 \left(\frac{6}{21}\right)^4 \\ &= \frac{59}{10000} \end{aligned}$$

**Teorema 4.6.1 (Teorema Multinomial)** Sean  $m$  y  $n$  enteros positivos. Sea  $\mathcal{A}$  el conjunto de vectores  $\mathbf{x} = (x_1, \dots, x_n)$  tal que cada  $x_i$  es un entero no negativo y  $\sum_{i=1}^n x_i = m$ . Entonces, para números reales  $p_1, p_2, \dots, p_n$ :

$$(p_1 + \cdots + p_n)^m = \sum_{\mathbf{x} \in \mathcal{A}} \frac{m!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n}$$

**Definición 4.6.2** Sean  $\mathbf{X}_1, \dots, \mathbf{X}_n$  vectores aleatorios con función de probabilidad o función de densidad conjunta  $f(\mathbf{x}_1, \dots, \mathbf{x}_n)$ . Sea  $f_{\mathbf{X}_i}(\mathbf{x}_i)$  la función de probabilidad o función de densidad marginal de  $\mathbf{X}_i$ . Entonces  $\mathbf{X}_1, \dots, \mathbf{X}_n$  son *vectores aleatorios mutuamente independientes* si, para todo  $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ :

$$f(\mathbf{x}_1, \dots, \mathbf{x}_n) = f_{\mathbf{X}_1}(\mathbf{x}_1) \cdots f_{\mathbf{X}_n}(\mathbf{x}_n) = \prod_{i=1}^n f_{\mathbf{X}_i}(\mathbf{x}_i)$$

si todas las  $\mathbf{X}_i$ 's son unidimensionales, entonces  $X_1, \dots, X_n$  son llamadas *variables aleatorias mutuamente independientes*.

**Teorema 4.6.2** Sean  $X_1, \dots, X_n$  variables aleatorias mutuamente independientes. Sean  $g_1, \dots, g_n$  funciones tales que  $g_i(x_i)$  es una función solo de  $x_i$ ,  $i = 1, 2, \dots, n$ , entonces:

$$E[g_1(X_1) \cdots g_n(X_n)] = E[g_1(X_1)] \cdots E[g_n(X_n)]$$

**Teorema 4.6.3** Sean  $X_1, \dots, X_n$  variables aleatorias mutuamente independientes con funciones generatrices de momentos  $M_{X_1}(t), \dots, M_{X_n}(t)$ . Si  $Z = X_1 + \dots + X_n$ , entonces:

$$M_Z(t) = M_{X_1}(t) \cdots M_{X_n}(t)$$

En particular, si las variables aleatorias  $X_1, \dots, X_n$  tienen la misma distribución, con función generatriz de momentos  $M_X(t)$ , entonces:

$$M_Z(t) = [M_X(t)]^n$$

**Ejemplo 4.6.3** Suponga  $X_1, \dots, X_n$  son variables aleatorias mutuamente independientes y la distribución de  $X_i \sim \mathcal{G}(\alpha_i, \beta)$ . Hallar la distribución de  $Z = X_1 + \dots + X_n$ .

**Corolario 4.6.1** Sean  $X_1, \dots, X_n$  variables aleatorias mutuamente independientes con funciones generatrices de momentos  $M_{X_1}(t), \dots, M_{X_n}(t)$ . Si  $a_1, \dots, a_n$  y  $b_1, \dots, b_n$  son constantes, entonces la función generatriz de momentos de  $Z = \sum_{i=1}^n (a_i X_i + b_i)$  es:

$$M_Z(t) = e^{t \sum b_i} M_{X_1}(a_1 t) \cdots M_{X_n}(a_n t)$$

**Ejemplo 4.6.4** Sean  $X_1, \dots, X_n$  variables aleatorias mutuamente independientes con  $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ . Si  $a_1, \dots, a_n$  y  $b_1, \dots, b_n$  son constantes, hallar la distribución de  $Z = \sum_{i=1}^n (a_i X_i + b_i)$ .

**Teorema 4.6.4 (Generalización del teorema 4.3.1)** Sean  $\mathbf{X}_1, \dots, \mathbf{X}_n$  vectores aleatorios independientes. Sea  $g_i(\mathbf{x}_i)$  una función que solo depende de  $\mathbf{x}_i$ ,  $i = 1, \dots, n$ . Entonces las variables aleatorias  $U_i = g_i(\mathbf{X}_i)$  son mutuamente independientes.

## 4.7. Transformaciones sobre un vector aleatorio

Sea  $(X_1, \dots, X_n)$  un vector aleatorio con función de densidad  $f_{\mathbf{X}}(\mathbf{x})$ . Sea  $\mathcal{A} = \{\mathbf{x} : f_{\mathbf{X}}(\mathbf{x}) > 0\}$ . Considere un nuevo vector aleatorio  $(U_1, \dots, U_n)$  definido por  $U_i = g_i(X_1, \dots, X_n)$ ,  $i = 1, \dots, n$ . Suponga que  $A_0, A_1, \dots, A_k$  forman una partición de  $\mathcal{A}$ . El conjunto  $A_0$ , el cual podría ser vacío, satisface  $\Pr((X_1, \dots, X_n) \in A_0) = 0$ . La transformación  $(U_1, \dots, U_n)$  es una transformación 1 - 1 desde  $A_i$  hacia  $\mathcal{B}$ . Entonces para cada  $i$ , se puede obtener las funciones inversas desde  $\mathcal{B}$  hacia  $A_i$ .

Denote la  $i$ -ésima inversa por  $x_1 = h_{1i}(u_1, \dots, u_n)$ ,  $x_2 = h_{2i}(u_1, \dots, u_n)$ ,  $\dots$ ,  $x_n = h_{ni}(u_1, \dots, u_n)$ . Estas inversas dan un único  $(x_1, \dots, x_n) \in A_i$  tal que  $(u_1, \dots, u_n) = (g_1(x_1, \dots, x_n), \dots, g_n(x_1, \dots, x_n))$ . Sea  $J_i$  el jacobiano calculado desde la  $i$ -ésima inversa, es decir:

$$J_i = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} & \dots & \frac{\partial x_1}{\partial u_n} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} & \dots & \frac{\partial x_2}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial u_1} & \frac{\partial x_n}{\partial u_2} & \dots & \frac{\partial x_n}{\partial u_n} \end{vmatrix} = \begin{vmatrix} \frac{\partial h_{1i}(\mathbf{u})}{\partial u_1} & \frac{\partial h_{1i}(\mathbf{u})}{\partial u_2} & \dots & \frac{\partial h_{1i}(\mathbf{u})}{\partial u_n} \\ \frac{\partial h_{2i}(\mathbf{u})}{\partial u_1} & \frac{\partial h_{2i}(\mathbf{u})}{\partial u_2} & \dots & \frac{\partial h_{2i}(\mathbf{u})}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_{ni}(\mathbf{u})}{\partial u_1} & \frac{\partial h_{ni}(\mathbf{u})}{\partial u_2} & \dots & \frac{\partial h_{ni}(\mathbf{u})}{\partial u_n} \end{vmatrix}$$

es el determinante de la matriz  $n \times n$ . Luego la función de densidad conjunta para  $\mathbf{u} \in \mathcal{B}$  es:

$$f_{\mathbf{U}}(u_1, \dots, u_n) = \sum_{i=1}^k f_{\mathbf{X}}(h_{1i}(u_1, \dots, u_n), \dots, h_{ni}(u_1, \dots, u_n)) |J_i| \quad (4.7.1)$$

**Ejemplo 4.7.1** Sean  $X_1, X_2$  y  $X_3$  variables aleatorias independientes con distribución  $\mathcal{E}(\beta = 1)$ . Probar que:

$$U_1 = \frac{X_1}{X_1 + X_2} \quad U_2 = \frac{X_1 + X_2}{X_1 + X_2 + X_3} \quad U_3 = X_1 + X_2 + X_3$$

son variables aleatorias mutuamente independientes.