## Función de Probabilidad Conjunta

Dadas dos v.a. discretas, X e Y, definidas sobre el mismo espacio de probabilidad de un experimento, la función de probabilidad *conjunta* para cada par de números (x,y) es

$$p_{XY}(x,y) = P(X = x, Y = y)$$

**Ejemplo**: De una urna que contiene 4 bolitas rojas y 4 bolitas blancas se extraen 3 sin reposición. Sean X = cantidad de bolitas rojas en las dos primeras extracciones, Y = cantidad de bolitas rojas en la 3ra. extracción.

| $\varpi$ | $P(\varpi)$ | X | Y |
|---|---|---|---|
| **bbb** | **(4/8)(3/7)(2/6)** | 0 | 0 |
| **rbb** | **(4/8)(4/7)(3/6)** | 1 | 0 |
| **brb** | **(4/8)(4/7)(3/6)** | 1 | 0 |
| **bbr** | **(4/8)(3/7)(4/6)** | 0 | 1 |
| **rrb** | **(4/8)(3/7)(4/6)** | 2 | 0 |
| **rbr** | **(4/8)(4/7)(3/6)** | 1 | 1 |
| **brr** | **(4/8)(4/7)(3/6)** | 1 | 1 |
| **rrr** | **(4/8)(3/7)(2/6)** | 2 | 1 |

¿Es el espacio muestral de equiprobabilidad?

P(X = 0, Y = 0) = P(bbb) = **(4/8)(3/7)(2/6)= 1/14**

Completamos la tabla de probabilidad conjunta de X e Y:

| y \ x | 0 | 1 | 2 | $p_Y(y)$ |
|---|---|---|---|---|
| 0 | 1/14 | 4/14 | 2/14 | 1/2 |
| 1 | 2/14 | 4/14 | 1/14 | 1/2 |
| $p_X(x)$ | 3/14 | 8/14 | 3/14 | 1 |

El par de v.a. (X , Y) se denomina *vector aleatorio.*

Las funciones $p_X$ y $p_Y$ se llaman <u>funciones de probabilidad *marginales*</u> de X y de Y respectivamente. Se obtienen a partir de la función de probabilidad conjunta por:

$$p_X(x) = \sum_{y \in R_Y} p_{XY}(x,y) \qquad p_Y(y) = \sum_{x \in R_X} p_{XY}(x,y)$$

## <u>Independencia</u>

<u>Caso discreto.</u> Diremos que X e Y son variables aleatorias *independientes* (v.a.i.) si $p_{XY}(x,y) = p_X(x)\ p_Y(y)$ para todo (x,y), es decir cuando *la función de probabilidad conjunta se factoriza como el producto de las funciones de probabilidad marginales.*

Siguiendo con el Ejemplo: Como $p_{XY}(0,0) = 1/14 \neq 3/4 * 1/2$
X e Y **no** son v.a.i

<u>Caso continuo.</u> Diremos que X e Y son v.a.i. si
$f_{XY}(x,y) = f_X(x)\ f_Y(y)$ para todo (x,y) es decir cuando *la función de densidad conjunta se factoriza como el producto de las funciones de densidad marginales.*

<u>Más de dos variables aleatorias</u>.

Diremos que una colección de n varibles aleatorias, $X_1, ..., X_n$, definidas sobre el mismo espacio de probabilidad de un experimento, son *independientes* si cumplen:

$$p_{X_1,\dots,X_k}(x_1,\dots,x_k) = p_{X_1}(x_1)\dots p_{X_k}(x_k) \quad \forall (x_1,\dots,x_k) \text{ en el caso discreto}$$

$$f_{X_1,\dots,X_k}(x_1,\dots,x_k) = f_{X_1}(x_1)\dots.f_{X_k}(x_k) \quad \forall (x_1,\dots,x_k) \text{ en el caso continuo}$$

## Funciones de variables y vectores aleatorios

Dado un vector aleatorio (X , Y) puede interesar el problema de determinar la distribución de una función g(X,Y). No estudiaremos las condiciones generales para que esto sea posible.

Las funciones que veremos en este curso serán principalmente sumas, promedios y productos.

Sumas de variables aleatorias

Sigamos con el ejemplo:
X = cantidad de bolitas rojas en las dos primeras extracciones,
Y = cantidad de bolitas rojas en la 3ra. extracción
Z = cantidad de bolitas rojas en las 3 extracciones = X + Y

| $\varpi$ | $P(\varpi)$ | X | Y | X + Y |
|---|---|---|---|---|
| **bbb** | **(4/8)(3/7)(2/6)** | 0 | 0 | 0 |
| **rbb** | **(4/8)(4/7)(3/6)** | 1 | 0 | 1 |
| **brb** | **(4/8)(4/7)(3/6)** | 1 | 0 | 1 |
| **bbr** | **(4/8)(3/7)(4/6)** | 0 | 1 | 1 |
| **rrb** | **(4/8)(3/7)(4/6)** | 2 | 0 | 2 |
| **rbr** | **(4/8)(4/7)(3/6)** | 1 | 1 | 2 |
| **brr** | **(4/8)(4/7)(3/6)** | 1 | 1 | 2 |
| **rrr** | **(4/8)(3/7)(2/6)** | 2 | 1 | 3 |

La función de probabilidad puntual de Z , $p_Z$, puede obtenerse a partir de la tabla de probabilidad conjunta de X e Y.

| z | 0 | 1 | 2 | 3 | |
|---|---|---|---|---|---|
| $p_Z(z)$ | 1/14 | 6/14 | 6/14 | 1/14 | 1 |

## Propiedades de la Esperanza y la Varianza

H emos visto

a) Esperanza de una función de una variable aleatoria.

Sea X una v.a. con f.p.p. $p_X(x)$ y sea Y = h( X ) entonces E(Y) puede calcularse como

$$E(Y) = \begin{cases} \sum_{i \geq 1} h(x_i)\, p_X(x_i) & \text{si X es discreta} \\ \int h(x)\; f_X(x)\, dx & \text{si X es continua} \end{cases}$$

Podemos generalizar

b) Esperanza de una función de dos variables aleatorias.
Sean X e Y v.a. con f.p.p. conjunta $p_{XY}(x,y)$ entonces E(g(X,Y)) puede calcularse de la siguiente manera

$$E(g(X,\; Y)\,) = \begin{cases} \sum \; \sum g(x,y)\, p_{XY}(x,y) & \text{caso discreto} \\ \int \; \int g(x,y)\; f_{XY}(x,y)\, dx dy & \text{caso continuo} \end{cases}$$

Consecuencias útiles

b) E(c) = c
c) E( aX + b) = aE(X) + b
d) Var (aX + b) = $a^2$ Var(X)
e) E( X + Y ) = E( X ) + E( Y )
f) Var (X) = E ( $X^2$ ) - E( X )]$^2$
g) Var( X + Y ) = E [ (X+Y - E(X+Y) )$^2$] = ............................

## Covarianza

Sean X e Y dos v.a. con esperanzas $\mu_X$ y $\mu_Y$ respectivamente, la **covarianza** entre X e Y se define como

$$\mathrm{Cov}(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right] = \begin{cases} \sum_x \sum_y (x-\mu_X)(y-\mu_Y)p_{XY}(x,y) \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)f_{XY}(x,y)\,dx\,dy \end{cases}$$

según sean *X* e *Y* discretas o continuas.

<u>Idea Intuitiva</u>
Si **X e Y son v.a. discretas** y tienen una fuerte **asociación positiva**, serán más probables los pares (x, y) en los que tanto x como y son grandes (mayores que sus respectivas medias) (o ambos pequeños). Los productos $(x-\mu_X)(y-\mu_Y)$ que en esos casos son positivos contribuirán a la suma con probabilidades más altas. Por lo tanto la suma mediante la que se calcula la covarianza resultará positiva.
Por otra parte, si X e Y tienen una fuerte asociación negativa, en el sentido que valores grandes de *X* aparecen con mayor probabilidad junto a valores pequeños de *Y* y valores pequeños de *X* aparecen con valores grandes de *Y*, entonces serán los productos $(x-\mu_X)(y-\mu_Y)$ negativos (un factor será positivo y el otro negativo, dando un producto negativo) los que tendrán mayores probabilidades y por lo tanto la covarianza será negativa.

### Propiedades

a) $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.
b) $\mathrm{Cov}(X,Y) = E(X\,Y) - E(X)E(Y)$.
c) Si X e Y son v.a. independientes, Cov (X,Y) = 0. La recíproca no es cierta en general.

<u>Ejemplo</u>. Consideremos dos variables aleatorias con función de probabilidad conjunta dada por la tabla siguiente

| | | X | | | | | $p_Y(y)$ |
|---|---|---|---|---|---|---|---|
| | | **0** | **1** | **2** | **3** | **4** | |
| **Y** | **0** | 1/5 | 0 | 0 | 0 | 1/5 | 2/5 |
| | **3** | 0 | 1/5 | 0 | 1/5 | 0 | 2/5 |
| | **4** | 0 | 0 | 1/5 | 0 | 0 | 1/5 |
| $p_X(x)$ | | 1/5 | 1/5 | 1/5 | 1/5 | 1/5 | 1 |

Vemos que
$p_{XY}(2, 3) = 0 \neq p_X(2)\, p_Y(3) = (1/5)\,(2/5)$

Luego X e Y no son v.a. independientes y sin embargo
*Cov* (X,Y) = 0

En efecto

$$E(XY) = 1 \cdot 3 \cdot \frac{1}{5} + 2 \cdot 4 \cdot \frac{1}{5} + 3 \cdot 3 \cdot \frac{1}{5} = 4$$

$$E(X) = 0 \cdot \frac{1}{5} + 1 \cdot \frac{1}{5} + 2 \cdot \frac{1}{5} + 3 \cdot \frac{1}{5} + 4 \cdot \frac{1}{5} = 2$$

$$E(Y) = 0 \cdot \frac{2}{5} + 3 \cdot \frac{2}{5} + 4 \cdot \frac{1}{5} = 2$$

Luego, $\mathrm{Cov}(X, Y) = 4 - 2 \cdot 2 = 0.$

Observación: La covarianza depende de las unidades.

**Coeficiente de correlación**

Sean *X* e *Y* dos v.a. con esperanzas $\mu_X$ y $\mu_Y$, desvíos estándar $\sigma_X > 0$ y $\sigma_Y > 0$, respectivamente. El **coeficiente de correlación** entre X e Y se define como

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \, \sigma_Y} \, .$$

Propiedades

1) Sean a, b, c y d números reales y *X* e *Y* dos v.a. cualesquiera con varianza positiva, entonces

$$\rho(aX + b, cY + d) = sg(ac)\, \rho(X, Y)$$

sg indica la función signo.

Observación: esta propiedad garantiza que el coefiente de correlación no se modifica al realizar un cambio en las unidades con que se expresan las variables.

2) $-1 \leq \rho(X,Y) \leq 1$

3) $|\rho(X,Y)| = 1 \Leftrightarrow Y = aX + b$ con probabilidad 1, para ciertos valores reales a y *b,* a $\neq$ 0. Es decir: si el coeficiente de correlacion es 1, en valor absoluto, los pares (x,y) estarán sobre una recta

El coeficiente de correlación mide
el grado de asociación *lineal* entre las variables aleatorias.