Tema 4: VECTORES ALEATORIOS

1 Concepto de variable aleatoria bidimensional

Sea Ω el espacio muestral de un experimento aleatorio. Definimos **variable aleatoria bidimensional**, como una aplicación

$$(X,Y):\Omega \longrightarrow \mathbb{R}^2$$

tal que $\forall w \in \Omega$,

$$\begin{array}{rccc} (X,Y): & \Omega & \longrightarrow & \mathbb{R}^2 \\ & w & \hookrightarrow & (X(w),Y(w)) \in \mathbb{R}^2 \end{array}$$

Las variables aleatorias bidimensionales se puede dividir en:

- **Discretas:** Cuando X e Y son v.a. discretas.
- **Continuas:** Cuando X e Y son v.a. continuas.
- **Mixtas:** Cuando una de las variables es discreta y la otra continua.

2 Variables aleatorias bidimensionales discretas

`Ejemplo:` Consideremos el experimento aleatorio de lanzar 3 veces una moneda trucada tal que

$$P(cara \equiv c) = 2/3$$
y
$$P(cruz \equiv x) = 1/3$$

El espacio muestral consta de 8 resultados:

$$\Omega = \{ccc, ccx, cxc, xcc, cxx, xcx, xxc, xxx\}$$

Definimos las variables aleatorias:

$$\begin{array}{rcl} X & = & \textbf{N}^o \textbf{ de caras obtenidas en el primer lanzamiento} \rightarrow \{0,1\} \\ Y & = & \textbf{N}^o \textbf{ total de cara obtenidas en los tres lanzamientos} \rightarrow \{0,1,2,3\} \end{array}$$

Escuela Técnica Superior de Ingeniería Agronómica.
Universidad Politécnica de Cartagena.

Podemos considerar la **variable aleatoria bidimensional** (X,Y) que toma los valores:

$$
\begin{array}{rcl}
(X,Y): \quad \Omega & \longrightarrow & \mathbb{R}^2 \\
ccc & \hookrightarrow & (1,3) \\
\left.\begin{array}{c} ccx \\ cxc \end{array}\right\} & \hookrightarrow & (1,2) \\
xcc & \hookrightarrow & (0,2) \\
cxx & \hookrightarrow & (1,1) \\
\left.\begin{array}{c} xcx \\ xxc \end{array}\right\} & \hookrightarrow & (0,1) \\
xxx & \hookrightarrow & (0,0)
\end{array}
$$

con las siguientes probabilidades:

$$
\begin{aligned}
f(0,0) &= P(X=0,Y=0) = p(\{xxx\}) = (p(x))^3 = 1/27 \\
f(0,1) &= P(X=0,Y=1) = p(\{xcx, xxc\}) = 2\,(p(x))^2\,p(c) = 4/27 \\
f(1,1) &= P(X=1,Y=1) = p(\{cxx\}) = P(c)\,(p(x))^2 = 2/27 \\
f(0,2) &= P(X=0,Y=2) = p(\{xcc\}) = p(x)\,(p(c))^2 = 4/27 \\
f(1,2) &= P(X=1,Y=2) = p(\{ccx, cxc\}) = 2\,(p(c))^2\,p(x) = 8/27 \\
f(1,3) &= P(X=1,Y=3) = p(\{ccc\}) = (p(c))^3 = 8/27
\end{aligned}
$$

que las podemos escribir en una tabla de doble entrada como sigue:

$Y \rightarrow$ / $X \downarrow$	0	1	2	3
0	$1/27$	$4/27$	$4/27$	0
1	0	$2/27$	$8/27$	$8/27$

▲ **Función Puntual de Probabilidad Conjunta de (X,Y)**

que es la función

$$f_{(X,Y)} = f : \mathbb{R}^2 \longrightarrow \mathbb{R}$$

definida por:

$$\boxed{f(x,y) = P(X=x, Y=y) = P\left(\{w \in \Omega \;/\; [(X(w), Y(w)] = (x,y)\}\right), \quad \forall x, y \in \mathbb{R}}$$

y verificando que:

1. $f(x,y) \geq 0, \quad \forall x,y \in \mathbb{R}$
2. $\sum_x \sum_y f(x,y) = \sum_y \sum_x f(x,y) = 1$

Además, mediante la f.p.p. conjunta podemos obtener probabilidades asociadas a la v.a. bidimensional (X,Y):

$$\boxed{\forall A \subset \mathbb{R}^2, \quad P\left[(X,Y) \in A\right] = \sum_{(x,y)\in A}\sum f(x,y)}$$

2.1 Distribuciones Marginales de X e Y

- La **función puntual de probabilidad marginal de** X:

$$\boxed{f_X(x) = P(X=x) = \sum_{y\in\mathbb{R}} f(x,y), \quad \forall x \in \mathbb{R}}$$

(implica sumar por filas)

- La **función puntual de probabilidad marginal de** Y:

$$\boxed{f_Y(y) = P(Y=y) = \sum_{x\in\mathbb{R}} f(x,y), \quad \forall y \in \mathbb{R}}$$

(implica sumar por columnas)

Y se pueden presentar en la tabla de doble entrada en la columna de la derecha y la fila inferior, respectivamente:

$Y \rightarrow$ / $X \downarrow$	0	1	2	3	$f_X \downarrow$
0	1/27	4/27	4/27	0	9/27
1	0	2/27	8/27	8/27	18/27
$f_Y \rightarrow$	1/27	6/27	12/27	8/27	1

2.2 Distribuciones Condicionadas de X e Y

- La **función puntual de probabilidad de** X **condicionada a que** $Y=b$ se define:

$$\boxed{\forall x \in \mathbb{R}, \qquad f_{X/b}(x/b) = P(X=x/Y=b) = \frac{P(X=x,Y=b)}{P(Y=b)} = \frac{f(x,b)}{f_Y(b)}}$$

siempre que $f_Y(b) > 0$.

- La **función puntual de probabilidad de** Y **condicionada a que** $X=a$ se define:

$$\boxed{\forall y \in \mathbb{R}, \qquad f_{Y/a}(y/a) = P(Y=y/X=a) = \frac{P(X=a,Y=y)}{P(X=a)} = \frac{f(a,y)}{f_X(a)}}$$

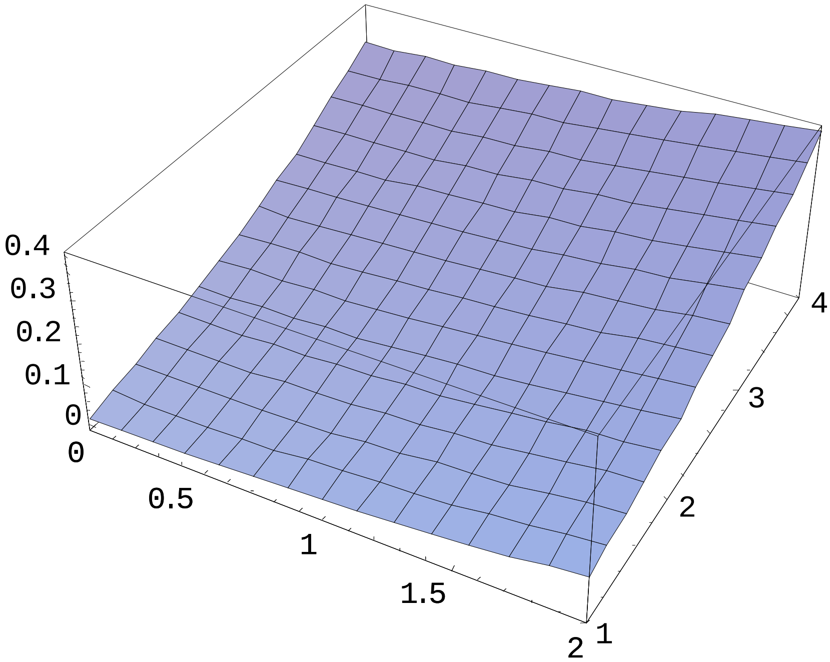
siempre que $f_X(a) > 0$.

3 Variables aleatorias bidimensionales continuas

`Ejemplo:` Sea (X, Y) una variable aleatoria continua con **Función de Densidad Conjunta** dada por:

$$\boxed{f(x,y) = \frac{1}{50}(x^2+y^2), \quad \text{si} \quad \begin{array}{c} 0 \le x \le 2 \\ 1 \le y \le 4 \end{array}}$$

y cuya gráfica es:



Notar que la **función de densidad conjunta** de una variable aleatoria bidimensional continua (X, Y) la definimos como la función

$$f_{(X,Y)} = f : \mathbb{R}^2 \longrightarrow \mathbb{R}$$

verificando que:

1. $f(x,y) \ge 0, \quad \forall x, y \in \mathbb{R}$
2. $\displaystyle\int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)dxdy = \int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)dydx = 1$

Además, mediante la f.d. conjunta podemos obtener probabilidades asociadas a la v.a. bidimensional (X, Y):

$$\boxed{\forall A \subset \mathbb{R}^2, \quad P\left[(X,Y) \in A\right] = \int\int_A f(x,y)dxdy}$$

3.1 Distribuciones Marginales de X e Y

- La **función de densidad marginal de** X:

$$\boxed{f_X(x) = \int_{\mathbb{R}} f(x,y)\cdot dy, \quad \forall x \in \mathbb{R}}$$

- La **función de densidad marginal de** Y:

$$\boxed{f_Y(y) = \int_{\mathbb{R}} f(x,y)\cdot dx, \quad \forall y \in \mathbb{R}}$$

3.2 Distribuciones Condicionadas de X e Y

- Sea $b \in \mathbb{R}$ tal que $f_Y(b) > 0$, entonces se define la **función de densidad de** X **condicionada a un valor concreto** b **de la v.a.** Y como:

$$\boxed{\forall x \in \mathbb{R}, \qquad f_{X/b}(x/b) = \frac{f(x,b)}{f_Y(b)}} \quad \text{siempre que } f_Y(b) > 0$$

- Sea $a \in \mathbb{R}$ tal que $f_X(a) > 0$, entonces se define la **función de densidad de** Y **condicionada a un valor concreto** a **de la v.a.** X como:

$$\boxed{\forall y \in \mathbb{R}, \qquad f_{Y/a}(y/a) = \frac{f(a,y)}{f_X(a)}} \quad \text{siempre que } f_X(a) > 0$$

4 Independencia de variables aleatorias

Las variables aleatorias X e Y **son independientes** si verifican:

- $f(x,y) = f_X(x)\cdot f_Y(y), \quad \forall (x,y) \in \mathbb{R}^2$
- $f_{X/b}(x/b) = f_X(x), \quad \forall x \in \mathbb{R}$
- $f_{Y/a}(y/a) = f_Y(y), \quad \forall y \in \mathbb{R}$

donde:

- $f(x,y)$ es la función puntual de probabilidad conjunta si (X,Y) es DISCRETA, o bien, la función de densidad conjunta si (X,Y) es CONTINUA.
- $f_X(x)$ y $f_Y(y)$ son las funciones puntuales de probabilidad marginal de X e Y si X e Y son DISCRETAS, o bien, las funciones de densidad marginal de X e Y si X e Y son CONTINUAS.
- $f_{X/b}(x/b)$ y $f_{Y/a}(y/a)$ son las distribuciones condicionadas de X e Y, respectivamente.

5 Propiedades de la esperanza y la varianza

1. Sean X e Y variables aleatorias tal que existen la $E(X)$ y la $E(Y)$ entonces

   (a) $E(X \pm Y) = E(X) \pm E(Y)$

   (b) $E(aX \pm bY) = aE(X) \pm bE(Y)$, $\forall a,b \in \mathbb{R}$

2. Sean X e Y variables aleatorias **independientes** tal que existe la $Var(X)$ y la $Var(Y)$, entonces se verifica que:

   (a) $Var(X \pm Y) = Var(X) + Var(Y)$

   (b) $Var(aX \pm bY) = a^2 Var(X) + b^2 Var(Y)$, $\forall a,b \in \mathbb{R}$

Ambas propiedades se puede generalizar a n de variables aleatorias como sigue:

- Sean $X_1, X_2, ..., X_n$ variables aleatorias tal que existe la $E(X_i)$ $\forall i = 1, 2, ...n$, entonces se verifica que:

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

  o bien, $\forall a_i \in \mathbb{R}$

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

- Sean $X_1, X_2, ..., X_n$ variables aleatorias **independientes** tal que existe la $Var(X_i)$ $\forall i = 1, 2, ...n$, entonces se verifica que:

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

  o bien, $\forall a_i \in \mathbb{R}$

$$Var\left(\sum_{i=1}^{n} a_i \cdot X_i\right) = \sum_{i=1}^{n} a_i^2 \cdot Var(X_i)$$

6 La propiedad de Reproductividad/Aditividad para algunos modelos de distribuciones

Nos planteamos el caso de que si conocemos la distribución de probabilidad de las variables $X_1, X_2, \cdots, X_k$, podemos conocer la distribución de probabilidad de la suma, esto es, de $Y = \sum_{i=1}^{k} X_i$.

Teorema de la Aditividad de la distribución Binomial

Sean $X_1, X_2, \cdots, X_k$ variables aleatorias **independientes** tales que $X_i \sim B(n_i, p)$, $\forall i = 1, 2, \cdots, k$.

$\Longrightarrow$

$$\boxed{Y = \sum_{i=1}^{k} X_i \sim B(\sum_{i=1}^{k} n_i, p)}$$

(Observar que la probabilidad de éxito debe de ser la misma para todas las variables aleatorias que sumamos)

Teorema de la Aditividad de la distribución Poisson

Sean $X_1, X_2, \cdots, X_k$ variables aleatorias **independientes** tales que $X_i \sim P_o(\lambda_i)$, $\forall i = 1, 2, \cdots, k$.

$\Longrightarrow$

$$\boxed{Y = \sum_{i=1}^{k} X_i \sim P_o(\sum_{i=1}^{k} \lambda_i)}$$

Teorema de la Aditividad de la distribución Normal

Sean $X_1, X_2, \cdots, X_k$ variables aleatorias **independientes** tales que cada $X_i \sim N(\mu_i,\ \sigma_i)$, $\forall i = 1, 2, \cdots, k$.

$\Longrightarrow$ La v.a. $Y = \sum_{i=1}^{k} X_i$ sigue una distribución **normal** de media

$$E(Y) = \sum_{i=1}^{k} \mu_i$$

y de varianza

$$Var(Y) = \sum_{i=1}^{k} \sigma_i^2, \text{ de donde, } \sigma_Y = \sqrt{\sum_{i=1}^{k} \sigma_i^2},$$

Esto es,

$$\boxed{Y = \sum_{i=1}^{k} X_i \sim N\left(\mu_Y = \sum_{i=1}^{k} \mu_i, \sigma_Y = \sqrt{\sum_{i=1}^{k} \sigma_i^2}\right)}$$

EJERCICIO Una determinada empresa dedicada a la fabricación de abono ha adquirido una máquina de envasado. Según los datos que le suministra el fabricante de la envasadora, la cantidad que proporciona por bolsa es una variable aleatoria *normal* cuyo promedio es de 2000 grs. y su desviación estándar es de 50 grs.. Por otro lado, el fabricante de las bolsas en las que se empaqueta el producto le garantiza que el peso de las bolsas se distribuye según un modelo *normal* de media 50 grs. y de desviación estándar 5 grs.. Sabiendo que un saco de abono se considera defectuoso si su *peso final* (peso de la bolsa + cantidad de abono depositada en la bolsa) es inferior a 1950 grs..

1. Si PF denota la variable aleatoria *peso final del saco de abono*, traducir los datos del enunciado sobre la distribución de PF.
2. Determinar la proporción de sacos defectuosos que se producen.
3. Si los sacos se almacenan en palés de 500 unidades, ¿cuál es el número de sacos defectuosos esperado en cada palé?

7 Distribución asociada al modelo normal: Distribución t de Student.

Definición:

Sean Z y X_1, X_2,....,X_n variables aleatorias independientes con $Z \sim N(0,1)$ y $X_i \sim N(0,1)$, $\forall i = 1,2,...,n \Longrightarrow$ La v.a. definida por

$T = \dfrac{Z}{\sqrt{\dfrac{\sum_{i=1}^{n} X_i^2}{n}}}$ sigue una **distribución t de Student con n grados de libertad**

Es decir, $\boxed{T \sim t_n}$

Propiedades:

- La **distribución t de Student con n grados de libertad** es una distribución continua con un único parámetro poblacional que es n, el número de grados de libertad y equivale al número de variables independientes X_i que intervine en la suma del denominador.
- (Opcional) Si $T \sim t_n$, su función de densidad está dada por la fórmula

$$f(t) = \begin{cases} \dfrac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\dfrac{t^2}{n})^{-(n+1)/2}, & \text{si } t > 0 \\ 0, & \text{en otro caso} \end{cases}$$

  donde $\Gamma(p)$ es la función Gamma dada por

$$\Gamma(p) = \int_0^{\infty} x^{p-1}e^{-x}dx \text{ cuando } p > 0$$

- Su rango es todo $\mathbb{R}$ y tiene una forma campaniforme como la distribución normal, sólo que tiene ramas más "gruesas", que significa que hay más probabilidad en las colas de la distribución t de Student que en la cola de la distribución $N(0,1)$.

- Si $T \sim t_n$, entonces

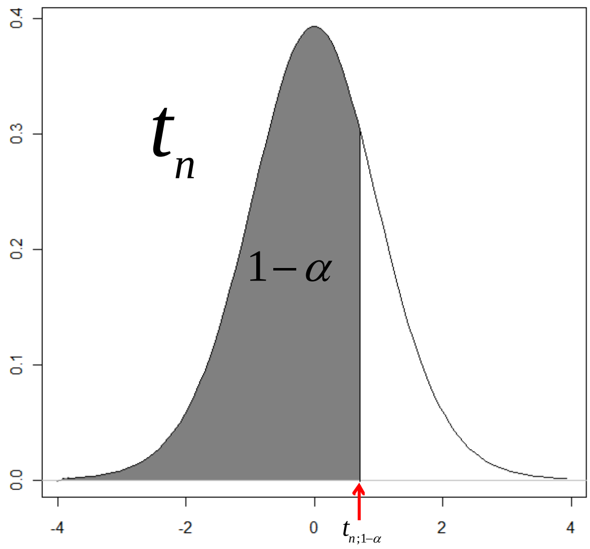
$$E(T) = 0$$

$$Var(T) = \frac{n}{n-2} \text{ para } n > 2$$

- La distribución t_n está tabulada y la tabla nos da el punto de abscisa x que en la *distribución t de Student* con n grados de libertad deja un área acumulada de valor $1-\alpha$. Este punto de abcisa se suele denotar por:

$$\boxed{x = t_{n;\ 1-\alpha}}$$

Gráficamente sería:



- Además, para tamaños muestrales grandes tenemos la siguiente aproximación:

$$\boxed{t_n \underset{n\geq 30}{\approx} N(0,1)}$$

Veámoslo gráficamente: