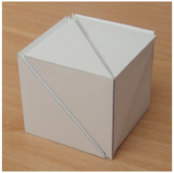
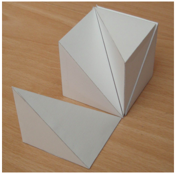
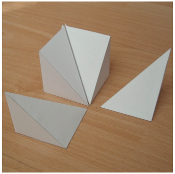
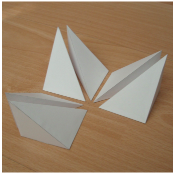








¡El cubo se parte en 6 partes isométricas![1]

[1] Gracias a Andrés Camargo

# Probabilidad - Clase 21
# Variables aleatorias independientes
# Suma de v.a.i.

Ernesto Mordecki

Facultad de Ciencias, Universidad de la República. Montevideo, Uruguay

Curso de la Licenciatura en Matemática - 2020

# Contenidos

# Variables aleatorias independientes (v.a.i)

- Sean $X_1, \dots, X_n$ v.a. en $(\Omega, \mathcal{A}, \mathbf{P})$,
- Sea $F_k(x)$ la distribución de la variable aleatoria $X_k$ $(k = 1, \dots, n)$,
- Sea $F(x_1, \dots, x_n)$ a la distribución del vector aleatorio $(X_1, \dots, X_n)$.
- Decimos que $X_1, \dots, X_n$ son *variables aleatorias independientes* cuando

$$F(x_1, \dots, x_n) = F_1(x_1) \dots F_n(x_n)$$

para reales $x_1, \dots, x_n$ arbitrarios.

## Lema

*Las variables aleatorias $X_1, \ldots, X_n$ son independientes si y solo si, para reales $x_1, \ldots x_n$ arbitrarios, son independientes los sucesos*

$$\{\omega\colon X_1(\omega) \leq x_1\}, \ldots, \{\omega\colon X_n(\omega) \leq x_n\}$$

## Lema (De los intervalos)

*Consideremos dos variables aleatorias $X, Y$ independientes. Para $a < b$, $c < d$, se verifica*

$$\mathbf{P}(a < X \leq b, c < Y \leq d) = \mathbf{P}(a < X \leq b)\,\mathbf{P}(c < Y \leq d).$$

## Proposición

- *Sean $X, Y$ v.a. discretas.*
- *Sean $x_1, x_2, \dots$ los valores que toma la variable $X$;*
- *$y_1, y_2, \dots$ los valores que toma la variable $Y$.*

*Entonces, $X$ e $Y$ son independientes, si y solo si*

$$\mathbf{P}(X = x_k, Y = y_j) = \mathbf{P}(X = x_k)\,\mathbf{P}(Y = y_j), \tag{1}$$

*para todos $k, j = 1, 2, \dots$.*

**Demostración.** Sean $X$ e $Y$ independientes. Para cada $k, n$ naturales, sea

$$\mathbf{A}_{k,n} = \{\omega \colon x_k - 1/n < X(\omega) \leq x_k\}.$$

Tenemos

$$\mathbf{A}_{k,1} \supset \mathbf{A}_{k,2} \supset \cdots ,$$

y además

$$\bigcap_{n=1}^{\infty} \mathbf{A}_{k,n} = \{\omega \colon X(\omega) = x_k\}.$$

Por la propiedad de continuidad, obtenemos

$$\lim_{n\to\infty} \mathbf{P}(\mathbf{A}_{k,n}) = \mathbf{P}(X = x_k).$$

En forma análoga, obtenemos

$$\lim_{n\to\infty} \mathbf{P}(y_j - 1/n < Y \leq y_j) = \mathbf{P}(Y = y_j), \quad \text{para } j \text{ arbitrario.}$$

Con la misma argumentación, obtenemos que

$$\lim_{n\to\infty} \mathbf{P}(x_k - 1/n < X \leq x_k, y_j - 1/n < Y \leq y_j) = \mathbf{P}(Y = y_j, X = x_k) \tag{2}$$

para $j, k$ arbitrarios.

Aplicando el lema de los intervalos, se tiene

$$\begin{aligned}\mathbf{P}(x_k - 1/n < X \leq x_k, y_j - 1/n < Y \leq y_j) \\ = \mathbf{P}(x_k - 1/n < X \leq x_k)\,\mathbf{P}(y_j - 1/n < Y \leq y_j),\end{aligned}$$

que sustituido en (2), da la igualdad (1).

# Dem del recíproco.

Supongamos que las variables aleatorias $X$ e $Y$ verifican

$$\mathbf{P}(X = x_k, Y = y_j) = \mathbf{P}(X = x_k)\,\mathbf{P}(Y = y_j),$$

Consideremos el vector aleatorio $(X, Y)$ y su función de distribución

$$F(x, y) = \mathbf{P}(X \leq x, Y \leq y).$$

Tenemos que ver

$$F(x, y) = \mathbf{P}(X \leq x)\,\mathbf{P}(Y \leq y) = F_X(x)F_Y(y).$$

para $x$ e $y$ arbitrarios.

Tenemos

$$
\begin{aligned}
F(x, y) &= \sum_{k,j:\ x_k \leq x, y_j \leq y} \mathbf{P}(X = x_k, Y = y_j) \\
&= \sum_{k,j:\ x_k \leq x, y_j \leq y} \mathbf{P}(X = x_k)\,\mathbf{P}(Y = y_j) \\
&= \sum_{k:\ x_k \leq x}\ \sum_{j:\ y_j \leq y} \mathbf{P}(X = x_k)\,\mathbf{P}(Y = y_j) \\
&= \sum_{k:\ x_k \leq x} \mathbf{P}(X = x_k) \sum_{j:\ y_j \leq y} \mathbf{P}(Y = y_j) \\
&= \mathbf{P}(X \leq x)\,\mathbf{P}(Y \leq y),
\end{aligned}
$$

LQQD

Utilizando un método similar, se obtiene la siguiente generalización.

## Proposición

*Consideremos n variables aleatorias $X_1, \dots, X_n$ con distribución discreta, siendo $x_{k1}, x_{k2}, \dots$ los valores que toma cada variable aleatoria $X_k$, para $k = 1, \dots, n$. Las variables aleatorias dadas son independientes, si y solo si se verifica*

$$\mathbf{P}(X_1 = x_{1m_1}, \dots, X_n = x_{nm_n}) = \prod_{k=1}^{n} \mathbf{P}(X_k = x_{km_k}),$$

*para naturales $m_1, \dots, m_n$ no nulos y arbitrarios.*

# Independencia para v.a. continuas

## Proposición

- *Consideremos un vector aleatorio* $X = (X_1, \ldots, X_n)$ *absolutamente continuo,*
- *Sea* $p_k(x)$ *la densidad de la variable aleatoria* $X_k (k = 1, \ldots, n)$.
- *Sea* $p(x_1, \ldots, x_n)$ *la densidad conjunta de* $X$

*Entonces* $X_1, \ldots, X_n$ *son independientes, si y solo*

$$p(x_1, \ldots, x_n) = p_1(x_1) \cdots p_n(x_n)$$

*(regla del producto para densidades)*

**Demostración** Si las variables aleatorias $X_1, \dots, X_n$ son independientes, la función de distribución se escribe de dos formas: como producto de las distribuciones $F_k(x_k)$ $(k = 1, \dots, n)$, y como integral múltiple de la densidad $p(u_1, \dots, u_n)$, es decir

$$\int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p(u_1, \dots, u_n) du_1 \cdots du_n = F_1(x_1) \cdots F_n(x_n).$$

Derivando $n$ veces en ambos miembros de esta identidad, primero respecto de $x_1$, luego respecto de $x_2$, ..., y finalmente respecto de $x_n$, obtenemos

$$p(x_1, \dots, x_n) = p_1(x_1) \cdots p_n(x_n)$$

# Recíproco

Supongamos

$$p(x_1, \ldots, x_n) = p_1(x_1) \cdots p_n(x_n)$$

Integramos $n$ veces en ambos lados de esta igualdad, primero con respecto de $x_1$ en el intervalo $(-\infty, y_1)$, ..., y por último con respecto de $x_n$ en el intervalo $(-\infty, y_n)$. Obtenemos

$$F(y_1, \ldots, y_n) = F_1(y_1) \cdots F_n(y_n).$$

y se verifica la definición de independencia.

LQQD

# Ejemplo

Sea nuevamente $(X, Y)$ normal bidimensional, con densidad

$$p(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$
$$\times \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-a_1}{\sigma_1}\right)^2 + \left(\frac{y-a_2}{\sigma_2}\right)^2 - 2\rho\frac{(x-a_1)}{\sigma_1}\frac{(y-a_2)}{\sigma_2}\right]\right\}.$$

Sabemos que

$$X \sim (a_1, \sigma_1^2) \qquad Y \sim (a_2, \sigma_2).$$

Si $\rho = 0$, la densidad se reduce a

$$p(x, y) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-a_1)^2/(2\sigma_1^2)} \times \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(y-a_2)^2/2\sigma_2^2} = p_1(x)p_2(y),$$

donde

- $p_1(x)$ es la densidad de $X$,
- $p_2(y)$ la densidad de $Y$.

Como consecuencia en el caso $\rho = 0$, las variables aleatorias $X$ e $Y$ son independientes.

# Ejemplo

Consideremos un vector aleatorio $(X_1, \ldots, X_n)$ con distribución normal $n$–dimensional. Sea $p(x)$ su densidad, con vector $a \in \mathbb{R}^n$ y matriz $B$ diagonal, dada por

$$B = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}. \tag{3}$$

Si $\sigma_k > 0$ $(k = 1, \ldots, n)$ la matriz $B$ verifica ser

- definida positiva,
- simétrica,
- no singular.

Sustituyendo la matriz $B$ en la fórmula de $p(x)$ obtenemos la densidad del vector considerado, que es

$$p(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}\sigma_1 \cdots \sigma_n} e^{-\sum_{k=1}^{n}(x_k - a_k)^2/(2\sigma_k^2)}. \qquad (4)$$

Como ocurre en el caso $n = 2$, se verifica que cada variable aleatoria $X_k$ tiene distribución normal, con parámetros $(a_k, \sigma_k)$, es decir, tiene densidad $p_k(x) = (\sigma_k\sqrt{2\pi})^{-1} e^{-(x-a_k)^2/(2\sigma_k^2)}$, para $k = 1, \dots, n$. Mas aún, factorizando la fórmula (4), tenemos

$$p(x_1, \dots, x_n) = \prod_{k=1}^{n} \frac{1}{\sigma_k\sqrt{2\pi}} e^{-(x_k - a_k)^2/(2\sigma_k^2)} = \prod_{k=1}^{n} p_k(x_k). \qquad (5)$$

Aplicando la proposición 3 se obtiene la independencia mutua de las variables aleatorias $X_1, \dots, X_n$.

Reciprocamente, si las variables aleatorias $X_1, \dots, X_n$ son independientes, y cada una de las variables $X_k$ tiene distribución normal con parámetros $(a_k, \sigma_k)$ $(k = 1, \dots, n)$, aplicando la proposición 3, obtenemos que la densidad $p(x_1, \dots, x_n)$ del vector $X = (X_1, \dots, X_n)$ verifica la fórmula (5), y en consecuencia, que el vector $X$ tiene distribución normal $n$–dimensional, con la densidad anterior, y matriz $B$ diagonal dada en (3).

# Distribución de la suma de variables aleatorias independientes

Demostremos la siguiente proposición.

## Proposición

*Sean $X_1$ y $X_2$ variables aleatorias independientes, con densidades $p_1(x)$ y $p_2(x)$ respectivamente. La suma $X_1 + X_2$ tiene densidad dada por*

$$p(x) = \int_{-\infty}^{\infty} p_1(u) p_2(x-u) du. \tag{6}$$

# Demostración

Sea $p(u, v)$ la densidad del vector $(X_1, X_2)$. Como las variables $X_1$ y $X_2$ son independientes, en vista de la proposición 3, tenemos $p(u, v) = p_1(u)p_2(v)$. Consideremos el valor de la función de distribución de la suma $X_1 + X_2$ en un punto $x$ arbitrario:

$$\mathbf{P}(X_1 + X_2 \leq x) = \mathbf{P}(X_1 + X_2 \in \mathbf{D}) = \int\int_{\mathbf{D}} p(u, v)dudv,$$

donde $\mathbf{D} = \{(u, v) \colon u + v \leq x\}$ es la región bajo la recta de ecuación $u + v = x$:

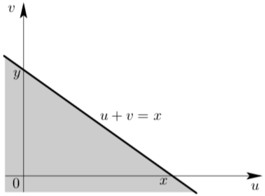



Figura 3.11: Región $\mathbf{D} = \{(u, v) \colon u + v \leq x\}$

De aquí obtenemos

$$\begin{aligned}
\mathbf{P}(X_1 + X_2 \leq x) &= \int_{-\infty}^{\infty} \Big( \int_{-\infty}^{x-u} p_2(v)dv \Big) p_1(u)du = \\
&= \int_{-\infty}^{\infty} \Big( \int_{-\infty}^{x} p_2(y-u)dy \Big) p_1(u)du \\
&= \int_{-\infty}^{x} \Big( \int_{-\infty}^{\infty} p_1(u)p_2(y-u)du \Big) dy
\end{aligned}$$

para $x$ arbitrario. En consecuencia la distribución de la suma $X_1 + X_2$ es absolutamente continua, y tiene densidad $p(x) = \int_{-\infty}^{\infty} p_1(u)p_2(x-u)du$.

LQQD

Observemos, que intercambiando los roles de $X_1$ y $X_2$, obtenemos también la fórmula

$$p(x) = \int_{-\infty}^{\infty} p_1(x-u)p_2(u)du \tag{7}$$

- Fórmulas análogas tienen lugar para la distribución de una suma de dos variables aleatorias independientes, que tienen distribuciones discretas.
- En este caso, en lugar de integrales aparecen sumas.
- Es posible obtener fórmulas mas generales (que incluyan ambos tipos de distribuciones) para la función de distribución $F(x)$ de una suma de variables aleatorias independientes $X_1$ y $X_2$, con funciones de distribución $F_1(x)$ y $F_2(x)$.

Mas precisamente, se trata de la fórmula

$$F(x) = \int_{-\infty}^{+\infty} F_1(x-y)dF_2(y) = \int_{-\infty}^{+\infty} F_2(x-y)dF_1(y), \quad (8)$$

en las cuales se utiliza la integral de Stieltjes. Las integrales anteriores se denominan *convolución* o *composición* de las distribuciones.

La definición de integral de Stieltjes es

$$\int_a^b h(x)dF(x) = \lim_{\|\pi^n\|\to 0} \sum_{k=1}^{k=k_n} h(\hat{x}_k^n)\left(F(x_k^n) - F(x_{k-1}^n)\right)$$

con $\pi^n$ partición con norma que tiende a cero, dada por

$$\pi^n = \{a = x_0^n, \dots, x_{k_n}^n = b\}$$

y $\hat{x}_k^n$ un punto arbitrario del intervalo $[x_{k-1}^n, x_k^n]$.

# Aplicación: suma de normales

Apliquemos la proposición para demostrar que la suma de variables aleatorias independientes, cada una de las cuales tiene distribución normal, tiene distribución normal. Consideremos entonces dos variables aleatorias independientes $X_1$ y $X_2$, tales que $X_k$ tiene densidad $p_k(x) = \frac{1}{\sigma_k\sqrt{2\pi}}e^{-(x-a_k)^2/(2\sigma_k^2)}$ $(k = 1, 2)$. Demostremos que $X_1 + X_2$ tiene distribución normal con parámetros $(a_1 + a_2, \sqrt{\sigma_1^2 + \sigma_2^2})$, es decir, que tiene densidad dada por

$$p(x) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}}e^{-(x-a_1-a_2)^2/(2(\sigma_1^2+\sigma_2^2))}.$$

Demostremos primero esta afirmación en el caso particular $a_1 = a_2 = 0$. Aplicando la fórmula de convolución (7), obtenemos

$$p(x) = \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} e^{-(x-u)^2/(2\sigma_1^2)-u^2/(2\sigma_2^2)}du.$$

Introduciendo una nueva variable de integración, según la fórmula $v = u\frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2} - \frac{\sigma_2 x}{\sigma_1\sqrt{\sigma_1^2+\sigma_2^2}}$, resulta

$$\begin{aligned} p(x) &= \frac{1}{2\pi\sqrt{\sigma_1^2+\sigma_2^2}} \int_{-\infty}^{\infty} e^{-v^2/2 - x^2/(2(\sigma_1^2+\sigma_2^2))} dv \\ &= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} e^{-x^2/(2(\sigma_1^2+\sigma_2^2))}. \end{aligned}$$

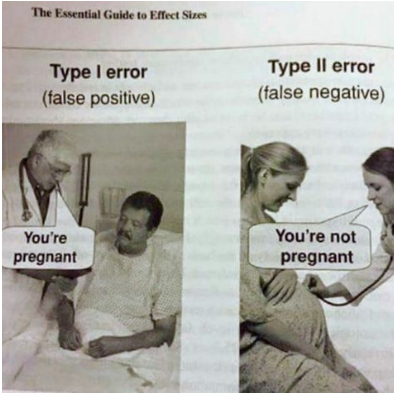
De esta forma, la densidad $p(x)$ de la suma $X_1 + X_2$ es normal con parámetros $(0, \sqrt{\sigma_1^2+\sigma_2^2})$.

La demostración cuando los parámetros $a_1$ y $a_2$ son arbitrarios se reduce al caso anterior ($a_1 = a_2 = 0$), mediante la consideración de nuevas variables aleatorias $Y_k = X_k - a_k \quad (k = 1, 2)$. Estas variables son independientes, por serlo las variables $X_1$ y $X_2$, tienen función de distribución

$$\mathbf{P}(Y_k \leq x) = \mathbf{P}(X_k \leq x + a_k) = F_k(x + a_k) = \int_{-\infty}^{x+a_k} p_k(u)du,$$

y densidad dada por $\frac{d}{dx}P(Y_k \leq x) = p_k(x + a_k) = \frac{1}{\sigma_k\sqrt{2\pi}}e^{-x^2/(2\sigma_k^2)} \quad (k = 1, 2)$, que es la densidad normal con parámetros $(0, \sigma_k)$. Según vimos, la suma $Y_1 + Y_2$ tiene distribución normal con parámetros $(0, \sqrt{\sigma_1^2 + \sigma_2^2})$. En consecuencia, la variable aleatoria $X_1 + X_2 = Y_1 + Y_2 + a_1 + a_2$ tiene distribución normal con parámetros $(a_1 + a_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

# Ayudamemoria




Para acordarse de los tipos de error:

- Tipo I (falso positivo): que el hombre **esté** embarazado
- Tipo II (falso negativo): que la mujer **no esté** embarazada