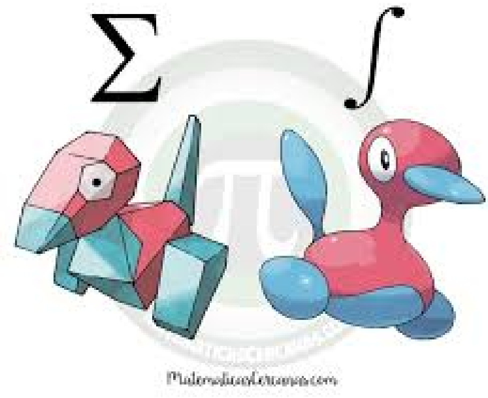
Lo discreto y lo continuo



Probabilidad - Clase 22
Suma de variables independientes. Ejemplos

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.

Independencia: caso discreto.

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 $k, j = 1, 2, \dots$.

Independencia: caso continuo

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)

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

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

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

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

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

$$p(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}\sigma_1 \cdots \sigma_n} e^{-\sum_{k=1}^{n}(x_k - a_k)^2/(2\sigma_k^2)}. \tag{3}$$

Además, cada X_k tiene distribución normal (a_k, σ_k). Su densidad es

$$p_k(x) = \frac{1}{\sigma_k\sqrt{2\pi}} e^{-(x-a_k)^2/(2\sigma_k^2)},$$

Factorizando

$$p(x_1, \dots, x_n) = \prod_{k=1}^{n} \frac{1}{\sigma_k\sqrt{2\pi}} e^{-(x_k-a_k)^2/(2\sigma_k^2)} = \prod_{k=1}^{n} p_k(x_k). \tag{4}$$

Conclusión: las $X_1, \dots, X_n$ son independientes.

Recíproco

- Sean las variables aleatorias $X_1, \ldots, X_n$ son independientes,
- Cada una X_k es normal con parámetros (a_k, σ_k) $(k = 1, \ldots, n)$, con densidad $p_k(x_k)$.
- Aplicando la proposición:

$$p(x_1, \ldots, x_n) = \prod p_k(x_k)$$

- De allí obtenemos la fórmula

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

- Luego la matriz B es diagonal.

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 2, 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 $u + v = x$:

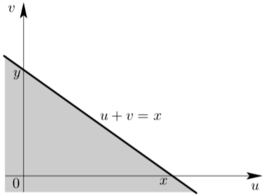


Figura 3.11: Región $\mathbf{D} = \{(u, v) \colon u + v \leq x\}$

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 π^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 uniformes

Sean U_1 y U_2 v.a. independientes uniformes en $[0, 1]$. Halla la distribución de

$$U = U_1 + U_2$$

- Los valores que toma están en $[0, 2]$.
- ¿tiene la misma probabilidad caer en

$$[0, 2\varepsilon] \quad \text{que en} \quad [1 - \varepsilon, 1 + \varepsilon]?$$

Veamos

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

Para el integrando:

$$\begin{aligned} p_1(u)p_2(x-u) &= \mathbf{1}_{\{0\le u\le 1\}}\mathbf{1}_{\{0\le x-u\le 1\}} \\ &= \mathbf{1}_{\{0\le u\le 1,\ x-1\le u\le x\}}. \end{aligned}$$

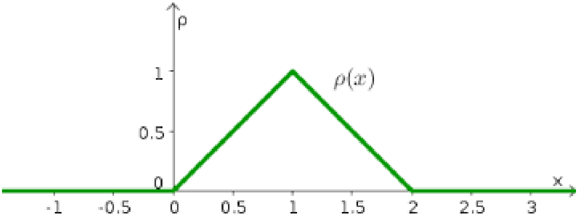
Por lo tanto tenemos tres casos:

- Si $x < 0$, también si $x > 2$ el integrando vale cero
- Si $0 < x < 1$ tenemos $0 < u < x$
- Si $1 < x < 2$ tenemos $x - 1 < u < 1$

Entonces

$$p_U(x) = \begin{cases} 0, & \text{si } x < 0, \\ x, & \text{si } 0 < x < 1, \\ 2x - 1, & \text{si } 1 < x < 2 \\ 0, & \text{si } x > 1 \end{cases}$$

Es lo que se denomina una *densidad triangular*.



Suma de v.a.i. normales

- Consideremos dos v.a. independientes X_1 y X_2,
- La densidad de X_k para $(k = 1, 2)$ es:

$$p_k(x) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-(x-a_k)^2/(2\sigma_k^2)}.$$

- Veamos que $X_1 + X_2$ tiene distribución normal con parámetros

$$(a_1 + a_2, \sigma_1^2 + \sigma_2^2).$$

- Es decir, su densidad es

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

Demostración

Consideramos primero $a_1 = a_2 = 0$. Aplicando la fórmula de convolución

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

El exponente es:

$$-\frac{(x-u)^2}{2\sigma_1^2} - \frac{u^2}{2\sigma_2^2} = -\frac{x^2}{2\sigma_1^2} + \frac{2xu}{2\sigma_1^2} - \frac{u^2}{2}\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)$$

Cambiamos entonces de variable: $v = u\frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2} - \frac{\sigma_2 x}{\sigma_1\sqrt{\sigma_1^2+\sigma_2^2}}$, resulta

$$v^2 = u^2\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right) - \frac{2xu}{\sigma_1^2} + \frac{\sigma_2^2}{\sigma_1^2}\frac{x^2}{\sigma_1^2+\sigma_2^2}.$$

$$-\frac{v^2}{2} = -\frac{u^2}{2}\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right) + \frac{xu}{\sigma_1^2} - \frac{\sigma_2^2}{2\sigma_1^2}\frac{x^2}{\sigma_1^2+\sigma_2^2}.$$

Para completar el exponente sumamos el término violeta

$$-\frac{v^2}{2} - \frac{x^2}{2\sigma_1^2} = -\frac{u^2}{2}\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right) + \frac{xu}{\sigma_1^2} - \frac{\sigma_2^2}{2\sigma_1^2}\frac{x^2}{\sigma_1^2+\sigma_2^2} - \frac{x^2}{2\sigma_1^2}$$

Para completar el exponente tenemos

$$-\frac{v^2}{2} - \frac{x^2}{2\sigma_1^2} + \frac{x^2}{2\sigma_1^2}\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2} = -\frac{u^2}{2}\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right) + \frac{xu}{\sigma_1^2} - \frac{x^2}{2\sigma_1^2}$$

El segundo sumando es

$$\frac{-x^2}{2\sigma_1^2}\left(1 - \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right) = \frac{-x^2}{2\sigma_1^2}\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2} = \frac{-x^2}{2(\sigma_1^2+\sigma_2^2)}$$

Entonces

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

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

Parámetros generales

- Si a_1 y a_2 son arbitrarios se reduce al caso anterior ($a_1 = a_2 = 0$),
- Consideramos $Y_k = X_k - a_k \quad (k = 1, 2)$.
- Estas variables son independientes, por serlo X_1 y X_2,
- tienen función de distribución

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

- Tienen densidad
$$\frac{d}{dx}P(Y_k \leq x) = p_k(x+a_k) = \frac{1}{\sigma_k\sqrt{2\pi}}e^{-x^2/(2\sigma_k^2)} \quad (k = 1, 2),$$
que es la densidad normal con parámetros $(0, \sigma_k)$.
- Conclusión: sumar una constante cambia el parámetro de la normal.
- 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 $(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$.

Algunas frases

- No te preocupes por tus dificultades en matemáticas. Te puedo asegurar que las mías son aún mayores.-Albert Einstein.
- La esencia de las matemáticas no es hacer las cosas simples complicadas, sino hacer las cosas complicadas simples.-S. Gudder.
- Las matemáticas son la creación más poderosa y bella del espíritu humano.-Stefan Banach.