Probabilidad - Clase 17
Variables aleatorias continuas

Ernesto Mordecki

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

Curso de la Licenciatura en Matemática - 2020

Contenidos

Variables aleatorias (repaso)

- Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$.
- Llamamos *variable aleatoria* a una función

$$X : \Omega \to \mathbb{R}$$

es decir $X(\omega)$ que toma valores reales,

- y verifica la condición

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A} \tag{1}$$

para todo x real.

Función de distribución

- Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$
- una variable aleatoria $X = X(\omega), \ \omega \in \Omega$.
- El conjunto
$$\{\omega \in \Omega \colon X(\omega) \leq x\}$$
es un suceso (es decir, un conjunto de la σ–álgebra de sucesos $\mathcal{A}$),
- Está definida la probabilidad
$$\mathbf{P}(\{\omega \colon X(\omega) \leq x\})$$
para todo $x \in \mathbb{R}$;

- Esta probabilidad se designa

$$\mathbf{P}(X \leq x) = \mathbf{P}(\{\omega\colon X(\omega) \leq x\})$$

- Se lee: la probabilidad de que la variable aleatoria X tome un valor menor o igual que x.
- Se denomina *función de distribución* de la variable aleatoria X, a la función $F(x)$, definida para todos los valores x reales, mediante la fórmula

$$F(x) = \mathbf{P}(X \leq x). \tag{2}$$

Variables aleatorias discretas

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$ y una variable aleatoria $X = X(\omega),\ \omega \in \Omega$.

- X tiene *distribución discreta* si existe un conjunto B finito o numerable tal que $\mathbf{P}(X \in B) = 1$.
- Si X es una variable aleatoria discreta, y x verifica $p = \mathbf{P}(X = x) > 0$, decimos que la variable aleatoria X *toma el valor* x con probabilidad p.

Variables aleatorias continuas (Repaso)

X tiene distribución *absolutamente continua*, su distribución $F(x)$ puede representarse de la forma

$$F(x) = \int_{-\infty}^{x} p(u)du, \qquad \text{para todo } x \text{ real} \tag{3}$$

- $p(u)$ es una función no negativa e integrable
- En análisis real una función $F(x)$ que se representa así se denomina *absolutamente continua*.
- Cualquier función absolutamente continua es continua en todos los puntos.
- La afirmación recíproca, en general, es falsa.
- En nuestro curso asumimos, que $p(u)$ es continua salvo en una cantidad finita de puntos

- La función $p(u)$ se llama *densidad* de la distribución de la variable aleatoria X,
- Decimos, que X tiene densidad $p(x)$.
- Es claro que
$$\int_{-\infty}^{\infty} p(u)du = 1 \tag{4}$$
por (3) y la propiedad $\lim_{x\to\infty} F(x) = 1$.
- Si $p(u)$ es continua en un punto x, en este punto existe la derivada $F'(x) = p(x)$.
- En particular, si $p(u)$ es continua en todos los puntos, tenemos $F'(x) = p(x)$ para todo x.

Lema Si una variable aleatoria X tiene función de distribución absolutamente continua, entonces

$$\mathbf{P}(X = x_0) = 0 \qquad \text{para cualquier } x_0.$$

Corolario Para dos números $a < b$ arbitrarios, tenemos

$$\begin{aligned} \mathbf{P}(a \leq X < b) &= \mathbf{P}(a \leq X \leq b) = \mathbf{P}(a < X \leq b) \\ &= \mathbf{P}(a < X < b) = \int_a^b p(x)dx. \end{aligned} \tag{5}$$

Ejemplos

Ejemplo Una variable aleatoria X tiene *distribución uniforme* en el intervalo (a, b), donde $a < b$, si tiene densidad dada por

$$p(x) = \begin{cases} c, & \text{si } a < x < b, \\ 0, & \text{si } x \leq a \text{ ó } x \geq b \end{cases} \tag{6}$$

donde c es una constante.

¿cuánto vale c?

Tenemos

$$\int_a^b c dx = 1$$

luego

$$c = 1/(b-a)$$

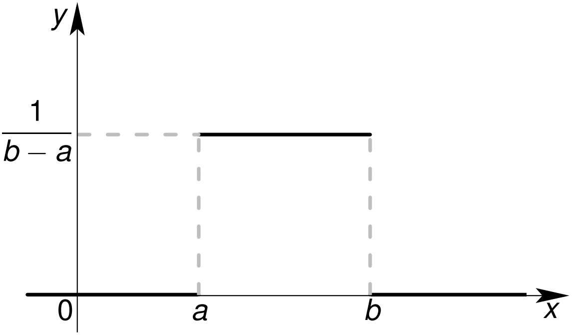


Figura: Gráfico de la densidad uniforme $p(x)$

Calculemos F

Distribución uniforme

Dada la densidad en (3), es fácil de obtener la función de distribución $F(x)$ de la variable aleatoria considerada. Tenemos

$$F(x) = \begin{cases} 0, & \text{si } x \leq a, \\ (x-a)/(b-a), & \text{si } a < x < b, \\ 1, & \text{si } x \geq b. \end{cases}$$

Esta función es continua para todo x, y es diferenciable en todos los puntos, con excepción de a y b (ver figura 2).

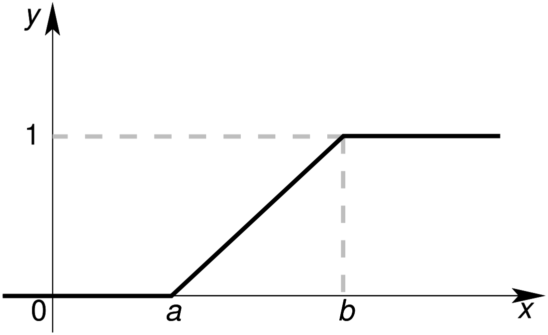


Figura: Gráfico de la función de distribución uniforme $F(x)$

Uniformes en distintos intervalos

Lema Si U es uniforme en $[0, 1]$, entonces, para $a - b$

$$X = a + (b - a)U$$

tiene distribución uniforme en $[a, b]$,
En efecto, si $a \leq x \leq b$:

$$\begin{aligned} P(X \leq x) &= \mathbf{P}(a + (b - a)U \leq x) \\ &= \mathbf{P}((b - a)U \leq x - a) = \mathbf{P}\left(U \leq \frac{x - a}{b - a}\right) = \frac{x - a}{b - a} \end{aligned}$$

Mientras que si $x < a$ tenemos $P(X \leq x) = 0$, y si $x > b$ tenemos $P(X \leq x) = 1$, porque $a + (b - a)U$ varia en $[a, b]$

Simulación de una variable uniforme

Un generador se define por la relación de recurrencia

$$X_{n+1} = (aX_n + c) \bmod m$$

donde X es una sucesión (pseudo) aleatoria[1], y

- m con $0 < m$ es el módulo (ej.: 2^{32})
- a tal que $0 < a < m$ es el multiplicador (ej: 1664525)
- c tal que $0 \leq c < m$ es le incremento, (ej: 1013904223)
- X_0, tal que $0 \leq X_0 < m$ es la semilla (la elige el usuario)
- En general m es primo, o c, m son co-primos, u otras variantes

[1] Tomado de Wikipedia
https://en.wikipedia.org/wiki/Linear_congruential_generator

Cómo saber si una variable es uniforme

- Generamos n valores de una variable supuesta uniforme en $[0, 1]$, en `vector`
- Exploración visual: `hist(vector)` genera un *histograma*
- Hacemos una escalera con salto $1/n$ en cada punto, y la comparamos con x en $[0, 1]$ la distribución teórica (Test de Kolmogorov)
- Partimos en P intervalos iguales y sumamos, para f_k la frecuencia observada:

  $$\sum_{k=1}^{P} \left(f_k - \frac{1}{P} \right)^2$$

  que no debe ser muy grande (Test Chi-cuadrado de Pearson)

Distribución exponencial

Ejemplo Una variable aleatoria X tiene *distribución exponencial* con parámetro $\alpha > 0$, si tiene densidad dada por

$$p(x) = \begin{cases} \alpha e^{-\alpha x}, & \text{si } x \geq 0, \\ 0, & \text{si } x < 0. \end{cases} \tag{7}$$

Integrando, encontramos la distribución:

$$F(x) = 1 - e^{-\alpha x}.$$

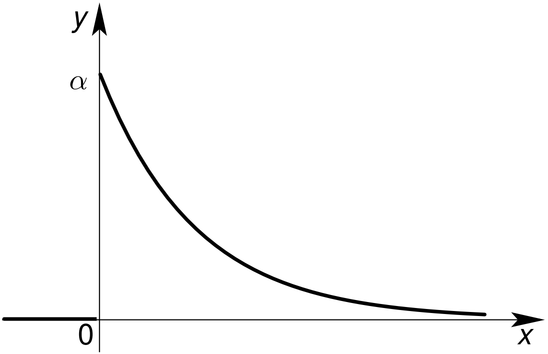


Figura: Gráfico de la densidad exponencial con parámetro α

Lema Si U tiene distribución uniforme en $[0,1]$, entonces

$$X = -\frac{1}{\alpha}\log U$$

tiene distribución exponencial de parámetro α. En efecto:

$$\begin{aligned}\mathbf{P}(T > x) = \mathbf{P}\left(-\frac{1}{\alpha}\log U > x\right) \\ = \mathbf{P}\left(\log U < -\alpha x\right) = \mathbf{P}\left(U < e^{-\alpha x}\right) = e^{-\alpha x}.\end{aligned}$$

Esto permite simular variables exponenciales a partir de la simulación de variables uniformes (claro, R hace eso para nosotros).

Ejemplo Decimos que una variable aleatoria tiene *distribución de Cauchy*, si tiene densidad dada por

$$p(x) = \frac{1}{\pi(1+x^2)}, \text{ para } x \text{ real.}$$

Integremos[2].

$$F(x) = \int_{-\infty}^{x} \frac{1}{\pi(1+t^2)} dt = \frac{1}{\pi}(\arctan(t) + \frac{\pi}{2}) = \frac{1}{\pi}\arctan(t) + \frac{1}{2}$$

[2] Se trata del arcotangente principal.

Lema Si U es uniforme en $[-\pi/2, \pi/2]$ entonces

$$X = \tan U,$$

tiene distribución de Cauchy. En efecto

$$\mathbf{P}(X \leq x) = \mathbf{P}(\tan U \leq x) = \mathbf{P}(U \leq \arctan x) = \frac{\arctan x + \pi/2}{\pi}$$

Distribución t de student[3]

Ejemplo Una variable aleatoria X tiene *distribución t* con n grados de libertad, donde $n \geq 1$ es un natural, si tiene densidad dada por

$$p(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \text{ para } x \text{ real}, \tag{8}$$

donde la función

$$\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} e^{-x} dx, \qquad \lambda > 0. \tag{9}$$

se denomina *función Gama*. La densidad dada en (8) es simétrica respecto del eje Oy, y tiene un único máximo en el punto $x = 0$, mientras que el eje Ox resulta ser asíntota, si $|x| \to \infty$. Si $n = 1$ obtenemos la distribución de Cauchy.

[3] William Sealy Gosset era el “estudiante”, recomiendo Wikipedia

Distribución normal

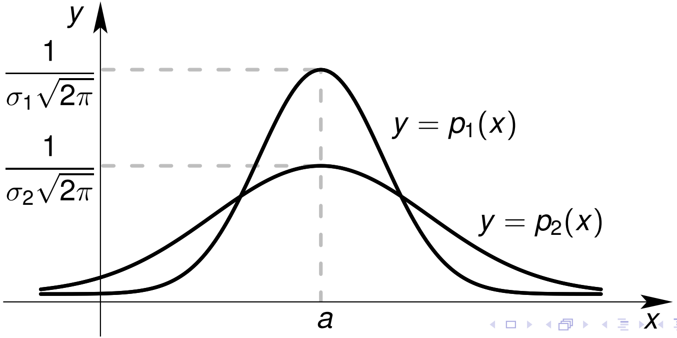
La distribución de probabilidades absolutamente continua más importante es la *distribución normal*.

Ejemplo Decimos que una variable aleatoria X tiene *distribución normal* con parámetros (a, σ), donde a y $\sigma > 0$ son números reales, si tiene densidad dada por

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/(2\sigma^2)} \text{ para } x \text{ real.} \tag{10}$$

Para verificar la fórmula (4) hay que hacer el cambio de variable $u = (x - a)/\sigma$ en la integral, y utilizar la identidad $\int_{-\infty}^{\infty} e^{-u^2/2} du = \sqrt{2\pi}$.
Es sencillo de ver que la función $p(x)$ tiene un único máximo en el punto $x = a$ (en donde $p'(x) = 0$), y toma el valor $p(a) = 1/(\sigma\sqrt{2\pi})$; que se verifica $p(x) \neq 0$ para todo x real, y que $\lim_{x\to-\infty} p(x) = \lim_{x\to+\infty} p(x) = 0$. La recta $x = a$ es eje de simetría de la curva $y = p(x)$. Se puede ver, que en los dos valores $x = a \pm \sigma$, el gráfico de $p(x)$ presenta puntos de inflexión. El gráfico de la función $p(x)$ se representa en la

Si variamos el valor de a, manteniendo σ constante, el gráfico de la función $p(x)$ se traslada, sin cambiar de forma. Si fijamos a y tomamos dos valores de σ, por ejemplo, $\sigma_1 < \sigma_2$, los gráficos de las densidades normales $p_1(x)$ y $p_2(x)$ con parámetros (a, σ_1) y (a, σ_2) respectivamente, presentan máximo en el mismo punto $x = a$, con valores máximos diferentes $1/(\sigma_1\sqrt{2\pi}) > 1/(\sigma_2\sqrt{2\pi})$. Teniendo en cuenta, que el área bajo cada uno de los gráficos de las densidades $p_1(x)$ y $p_2(x)$ es igual a 1, (por (4)), estos gráficos se representan como en la figura 5.



La distribución $F(x)$ de la variable aleatoria considerada, es

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(t-a)^2/2\sigma^2} dt,$$

en vista de (3).
El caso particular en el que $a = 0$ y $\sigma = 1$ corresponde a la denominada *distribución normal estándar*. Una variable aleatoria tiene distribución normal estándar, si tiene densidad dada por

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{11}$$

La distribución correspondiente a esta variable aleatoria se denota mediante $\Phi(x)$. De esta forma, la distribución normal estándar se define por la fórmula

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$

Las funciones $\varphi(x)$ y $\Phi(x)$, relacionadas por la igualdad $\Phi'(x) = \varphi(x)$, fueron introducidas en el capítulo 2 y están tabuladas. Se puede demostrar que la densidad de una variable aleatoria con distribución t con n grados de libertad definida en (8), tiene como límite, cuando $n \to \infty$, a la densidad normal $\varphi(x)$, de forma que para valores grandes de n, la densidad (8) es similar a $\varphi(x)$.