Cambio de variables y generación de números al azar

Ximena Fernández

Probabilidades y Estadística (Computación)

19 de mayo de 2020

Ejercicio 1: Cambio de variables

(a) Sea X una variable aleatoria con distribución uniforme en $(0, 2)$. Hallar la función de densidad de $Y = \ln(X)$ y de $Z = (X - 1)^2$.

Solución

Sea $x \in \mathbb{R}$.

Calculamos la función de distribución de Y:

$$F_Y(x) = P(Y \leq x) = P(\ln(X) \leq x) = P(X \leq e^x) = F_X(e^x)$$

Derivando,

$$f_Y(x) = f_X(e^x) \cdot e^x = \frac{1}{2}\mathrm{I}_{(0,2)}(e^x) \cdot e^x = \frac{1}{2}\mathrm{I}_{(-\infty,\ln(2))}(x) \cdot e^x$$

Calculamos ahora la función de distribución de Z

$$F_Z(x) = P(Z \leq x) = P((X-1)^2 \leq x)$$

Aquí la función $(x-1)^2$ no es inversible en $[0,2]$. Para $x \geq 0$ podemos realizar el siguiente despeje:

$$P((X-1)^2 \leq x) = P(|X-1| \leq \sqrt{x}) = P(-\sqrt{x} \leq X-1 \leq \sqrt{X}) =$$
$$P(-\sqrt{x}+1 \leq X \leq \sqrt{x}+1) = F_X(\sqrt{x}+1) - F_X(-\sqrt{x}+1)$$

Luego,

$$F_Z(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ F_X(\sqrt{x}+1) - F_X(-\sqrt{x}+1) & \text{si } x \geq 0 \end{cases}$$

Derivando,

$$f_Z(x) = \left(f_X(\sqrt{x}+1) \cdot \frac{1}{2\sqrt{x}} - f_X(-\sqrt{x}+1) \cdot \frac{-1}{2\sqrt{x}} \right) \mathrm{I}_{[0,+\infty)}(x)$$

$$= \frac{1}{2}\mathrm{I}_{[0,1)}(x) \cdot \frac{1}{2\sqrt{x}} + \frac{1}{2}\mathrm{I}_{[0,1)}(x) \cdot \frac{1}{2\sqrt{x}} = \mathrm{I}_{[0,1)}(x) \cdot \frac{1}{2\sqrt{x}}$$

Ejercicio 1: Cambio de variables

(b) Sea X una variable aleatoria con función de densidad

$$f_X(x) = \begin{cases} 0 & \text{si } x < 9 \\ \dfrac{5e^{-5(\sqrt{t}-3)}}{2\sqrt{t}} & \text{si } x \geq 9 \end{cases}$$

Hallar la función de distribución acumulada de $Y = \sqrt{X} - 3$

Solución

Sea $x \in \mathbb{R}$.

Calculamos la función de distribución de Y:

$$F_Y(x) = P(Y \leq x) = P(\sqrt{X} - 3 \leq x) = P(\sqrt{X} \leq x + 3) = P(0 \leq X \leq (x+3)^2)$$

Usando la fórmula de $f_X(x)$:

$$F_Y(x) = \int_0^{(x+3)^2} \frac{5e^{-5(\sqrt{t}-3)}}{2\sqrt{t}} I_{[9,+\infty)}(t)dt$$

Si $(x+3)^2 \geq 9$ (es decir, $x \geq 0$)

$$F_Y(x) = \int_9^{(x+3)^2} \frac{5e^{-5(\sqrt{t}-3)}}{2\sqrt{t}} dt = -e^{-5(\sqrt{t}-3)}\Big|_9^{(x+3)^2} = 1 - e^{-5x}$$

Luego, $F_Y(x) = 1 - e^{-5x}\mathrm{I}_{[0,+\infty)}(x)$, es decir, $Y \sim \mathcal{E}(5)$.

Ejercicio 2: El running está de moda

El tiempo en horas que tarda Ximena en correr una media maratón es una variable aleatoria exponencial de parámetro $\lambda = 0.75$.
Si la carrera le insume un tiempo X, el tiempo en horas que necesita descansar para recuperarse es $Y = e^{0.5X} + 24$.

a) Calcular $E(Y)$ y $V(Y)$. Compararlos con $E(X)$ y $V(Y)$.
b) Calcular la probabilidad de que tarde más de 2 hs en terminar la carrera.
c) Hallar la densidad f_Y.

Es una aplicación real [1]

RECORDS ARGENTINOS / OTRAS ESPECIALIDADES

PRUEBAS DE RUTA

Varones					
10 km.	27,52	Antonio Fabián Silio	FAM	Copenhague	19.08.1990
15 km.	42,59	Antonio Fabián Silio	FAM	Uster	27.09.1998
10 millas	48,55	Carlos Paradelo	ER	Park Forest	06.09.1999
20 km.	57,28	Antonio Fabián Silio	FAM	Uster	27.09.1998
Medio maratón	1,00,45	Antonio Fabián Silio	FAM	Uster	27.09.1998
25 km.	1,16,13	Antonio Fabián Silio	FAM	Otsu	03.03.1996
30 km.	1,31,30	Antonio Fabián Silio	FAM	Otsu	03.03.1996
Maratón	2,09,57	Antonio Fabián Silio	FAM	Hamburgo	30.04.1995
Varones u23					
10 km	29,14	Federico Bruno	ER	Viedma	11.10.2015
10 millas	49,47	Leonardo Malgor	BA	Santa Clara	21.01.1990
15 km	45.08	Antonio Fabián Silio	FAM	Bell Ville	09.07.1988
Medio maratón	1,03,50	Antonio Fabián Silio	FAM	Bell Ville	09.07.1988
25 km	1,21,57	Raimundo Manquel	CHU	Buenos Aires	03.06.1973
30 km	1,45,43	Alejandro Giménez	LPA	Santa Rosa	07.04.1996
Maratón	2,21,22	Raúl Víctor Llusá	BA	Mar del Plata	30.09.1979
Varones u20					
10 km	30,20	Federico Bruno	ER	Buenos Aires	01.12.2012
15 km	49,29	Martín Ñancucheo	CHU	Com.Rivadavia	13.12.2006
Medio maratón	1,08,07	Juan Gabriel Gómez	ER	Gualeguaychú	29.04.2001
Mujeres					
10 km.	33,06	Elisa Noemí Cobanea	BA	Bolívar	12.05.2001
15 km	50,36	Daiana Alejandra Ocampo	FAM	Buenos Aires	04.08.2019
10 millas	56,55	Griselda González	FAM	Newry IRL	12.10.1997
20 km.	1,09,37	Rosa Liliana Godoy	COR	Rosario	09.05.2010
Medio maratón	1,11,58	Florencia Borelli	BA	Buenos Aires	10.09.2017
25 km.	1,28,48	Marcela Cristina Gómez	CHA	Sevilla ESP	23.02.2020
30 km.	1,45,55	Marcela Cristina Gómez	CHA	Sevilla ESP	23.02.2020
Maratón	2,28,58	Marcela Cristina Gómez	CHA	Sevilla ESP	23.02.2020
Mujeres u23					
10 km	34,07	Lucrecia Mendiburu	FAM	Buenos Aires	25.05.2000
15 km	53,32	Stella Maris Selles	FAM	Lisboa POR	09.11.1986
Medio maratón	1,17,36	Sofía Eva Luna	BA	Buenos Aires	05.06.2016
Maratón	2,54,20	Norma Fernández	FAM	Los Angeles US	06.03.1988
Mujeres u20					
10 km	34,07	Lucrecia Mendiburu	FAM	Buenos Aires	25.05.2000
Medio maratón	1,21,30	Laura Ximena Fernández	FAM	Ezeiza	22.08.2004

PRUEBAS DE MARCHA

records argentinos absolutos

Varones					
10.000 m pista	40:05.03	Juan Manuel Cano	SDE	Rosario	20.06.2014
10 km. ruta	40,35	Juan Manuel Cano	SDE	Londres	04.08.2012
20.000 m pista	1,22,18,5	Juan Manuel Cano	SDE	Buenos Aires	25.01.2014
20 km. ruta	1,22,10	Juan Manuel Cano	SDE	Londres	04.08.2012
50.000 m. pista	4,14,28,5	Jorge Loréfice	BA	Buenos Aires	09.05.1993
		Benjamín Loréfice	BA	Buenos Aires	09.05.1993
50 km. ruta	4,17,03	Jorge Loréfice	BA	Mar del Plata	03.11.1991
		Benjamín Loréfice	BA	Mar del Plata	03.11.1991

[1] https://www.cada-atletismo.org/estadisticas/

Solución

(a)

$$E(Y) = E(e^{0.5X}+24) = E(e^{0.5X})+24 = \int_{-\infty}^{+\infty} e^{0.5x} 0.75 e^{-0.75x} I_{[0,+\infty)(x)} dx + 24$$

$$= 0.75 \int_0^{+\infty} e^{-0.25x} dx + 24 = \frac{0.75}{0.25} \int_0^{+\infty} 0.25 e^{-0.25x} dx + 24 = \frac{0.75}{0.25} + 24 = 27$$

(b)

$$P(Y > 48) = P(e^{0.5X} + 24 > 48) = P(e^{0.5X} > 24) = P(0.5X > \ln(24))$$

$$= P(X > 2\ln(24)) = 1 - F_X(2\ln(24)) = 1 - (1 - 0.75e^{-0.75 \cdot 2\ln(24)}) \sim 0.00638$$

(c)

$$F_Y(x) = P(Y \leq x) = P(e^{0.5X} + 24 \leq x) = P(e^{0.5X} \leq x - 24)$$

Si $x > 24$,

$$F_Y(x) = P(0.5X \leq \ln(x-24)) = P(X \leq 2\ln(x-24)) = F_X(2\ln(x-24))$$

Luego, $F_Y(x) = F_X(2\ln(x-24))\mathrm{I}_{(24,+\infty)}(x)$

Derivando,

$$\begin{aligned} f_Y(x) &= f_X\left(2\ln(x-24)\right)\frac{2}{x-24}\mathrm{I}_{(24,+\infty)}(x) \\ &= 0.75e^{-0.75\cdot 2\ln(x-24)}\mathrm{I}_{[0,+\infty)}\left(2\ln(x-24)\right)\frac{-2}{x-24}\mathrm{I}_{(24,+\infty)}(x) \\ &= 0.75e^{-0.75\cdot 2\ln(x-24)}\mathrm{I}_{[25,\infty)}(x)\frac{2}{x-24}\mathrm{I}_{(24,+\infty)}(x) \end{aligned}$$

Luego, $f_Y(x) = 0.75e^{-1.5}(x-24)\frac{2}{x-24}\mathrm{I}_{[25,\infty)}(x)$

Ejercicio 2: El running está de moda

d) En una carrera de 21 km, el tiempo T que tardan los primeros 3 corredores en arribar a la meta es una variable aleatoria Gama $T \sim \Gamma(3, 1)$. Calcular la probabilidad de que se complete el podio en menos de 1.3 hs.

Solución

$f_T(x) = \frac{1^3}{\Gamma(3)} e^{-x} x^{3-1} \mathrm{I}_{(0,+\infty)}(x)$

$$P(T < 1.3) = \int_{-\infty}^{1.3} f_T(x)dx = \int_0^{1.3} \frac{1}{2!} e^{-x} x^2 dx$$

Calculamos la primitiva de $e^{-x}x^2$ usando partes:

$$\int e^{-x} x^2 dx = -x^2 e^{-x} - \int 2xe^{-x} dx = -x^2 e^{-x} - 2 \int xe^{-x} dx$$

$$= -x^2 e^{-x} - 2 \left(-xe^{-x} - \int e^{-x} dx \right) = -x^2 e^{-x} - 2 \left(-xe^{-x} + e^{-x} \right) + C$$

$$= e^{-x}(-x^2 + 2x - 2) + C$$

Luego,

$$P(T < 1.3) = \frac{1}{2!} e^{-x}(-x^2 + 2x - 2) \Big|_0^{1.3} \sim 0.85$$

Ejercicio 3: Tiro oblicuo

Si se dispara un proyectil desde tierra con velocidad v_0 a un ńangulo α respecto del suelo, entonces el alcance R (punto en el que retorna a tierra)[2] es

$$R = \frac{v_0^2}{g}\sin(2\alpha)$$

donde g es la cte. gravitacional $(9,8m/s^2)$. Asumiendo que v_0 está fijo, calcular la densidad de R para las siguientes distribuciones de α:

(a) $\alpha \sim U(0, \pi/4)$
(b) $\alpha \sim U(0, \pi/2)$

Solució n
(a)

$$F_R(x) = P(R \leq x) = P\left(\frac{v_0^2}{g}\sin(2\alpha) \leq x\right) = P\left(\sin(2\alpha) \leq x\frac{g}{v_0^2}\right)$$

[2] https://en.wikipedia.org/wiki/Projectile_motion

Como $\alpha \sim U(0, \pi/4)$, $\sin(2\alpha)$ es inversible y luego

$$F_R(x) = P(R \leq x) = P\left(2\alpha \leq \arcsin(x\frac{g}{v_0^2})\right) = P\left(\alpha \leq \frac{1}{2}\arcsin(x\frac{g}{v_0^2})\right)$$

En conclusión

$$F_R(x) = F_\alpha\left(\frac{1}{2}\arcsin(x\frac{g}{v_0^2})\right)$$

Tenemos entonces que,

$$f_R(x) = f_\alpha\left(\frac{1}{2}\arcsin(x\frac{g}{v_0^2})\right) \cdot \left(\frac{1}{2}\frac{1}{\sqrt{1-(\frac{g}{xv_0^2})^2}}\frac{g}{v_0^2}\right)$$

$$= \frac{4}{\pi}\mathrm{I}_{(0,\pi/4)}\left(\frac{1}{2}\arcsin(x\frac{g}{v_0^2})\right) \cdot \left(\frac{1}{2}\frac{1}{\sqrt{1-(\frac{g}{xv_0^2})^2}}\frac{g}{v_0^2}\right)$$

Reescribiendo la indicadora, queda

$$f_R(x) = \frac{4}{\pi}\mathrm{I}_{(0,v_0^2/g)}(x) \cdot \left(\frac{1}{2}\frac{1}{\sqrt{1-(\frac{g}{xv_0^2})^2}}\frac{g}{v_0^2}\right)$$

(b) Queda de ejercicio. Hay que tener algo más de cuidado porque $\sin(2\alpha)$ no es inyectiva en $(0, \frac{\pi}{2})$.

Generación de números al azar (motivación)

La idea es generar computacionalmente números aleatorios siguiendo distintas distribuciones.

- Existen algoritmos[3] para generar números aleatorios con distribución uniforme en $[0, 1]$.
- Ahora, supongamos que queremos programar un jueguito en el cual se dispara un proyectil de manera aleatoria con ángulo α entre $[0, \frac{\pi}{4}]$. Tenemos que saber generar números al azar entre 0 y $\frac{\pi}{4}$. Más aún, queremos simular cuál es el alcance de la bala (en qué punto de la tierra aterriza)[4]. Entonces en el fondo, necesitamos generar números aleatorios con distribución $R\frac{v_0^2}{g}sin(2\alpha)$.
- El señor/la señora que programó las librerías de R, que tienen incorporados generadores aleatorios de las distintas variables aleatorias famosas, también tuvo que hacer este tipo procedimiento.
- Si les interesan las competencias de programación, este problema del Round 3 de la Code Jam 2016 (spoiler alert) sale simulando una distribución a partir de la uniforme.

---

[3]https://www.geeksforgeeks.org/pseudo-random-number-generator-prng/
[4]Ver ejercicio 3

Ejercicio 4: Generación de números al azar

Generar a partir de una variable aleatoria con distribución $\mathcal{U}(0,1)$:

(a) una variable aleatoria con distribución $Be(p)$;

(b) una variable aleatoria discreta X de rango $R_X = \{x_1, x_2, \cdots, x_n\}$, cuya función de probabilidad puntual es $p_X(x_i) = p_i \ \forall 1 \leq i \leq n$.

(c) una variable aleatoria con distribución $\mathcal{U}(a,b)$.

Solución

(a)

$$X = \begin{cases} 1 & \text{si } U < p \\ 0 & \text{si } U \geq p \end{cases}$$

(b)

$$X = \begin{cases} x_1 & \text{si } U \in [0, p_1) \\ x_2 & \text{si } U \in [p_1, p_1 + p_2) \\ x_3 & \text{si } U \in [p_1 + p_2, p_1 + p_2 + p_3) \\ \dots & \\ x_n & \text{si } U \in [\sum_{i=0}^{n-1} p_i, 1] \end{cases}$$

(c) La función de distribución acumulada de $X \sim \mathcal{U}(a, b)$ es

$$F_X(x) = \frac{x-a}{b-a} \mathrm{I}_{(a,b)}$$

Para generar la variable aleatoria X a partir de U,

$$X \sim F_X|_{(a,b)}^{-1}(U) = (b-a)U + a$$