Clase 7 de Bioestadística

Variables aleatorias continuas

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Contenidos de la clase

Introducción

Variable aleatoria Uniforme

- Densidad de una variable aleatoria uniforme
- Distribución de una variable aleatoria uniforme
- Simulación de una variable uniforme

Variable aleatoria exponencial

- Simulación de una variable exponencial

Introducción

Nos interesa modelar un experimento cuyo resultado es un número real:

- La medición de una longitud, con un dispositivo sujeto a error,
- El tiempo hasta que se rompe una lamparita,
- La medición de la altura de una persona elegida al azar en una población,
- El peso de un bebé al nacer,
- El tiempo de desintegración radiactiva de un núcleo atómico inestable
- ¿más ejemplos?

Identificamos el conjunto de valores I que puede tomar la variable, que pude ser

- Todo $I = \mathbb{R} = \{x \colon -\infty < x < \infty\}$
- Los reales no-negativos $I = \mathbb{R}^{+} = [0, \infty) = \{x \colon x \geq 0\}$
- Un intervalo $I = [a, b] = \{a \leq x \leq b\}$

Estudiamos entonces variables aleatorias continuas

$$X : \Omega \to I.$$

Suponemos que hay un universo Ω en donde mediante un experimento se elige al azar un $\omega \in \Omega$ y que nosotros observamos

- $X(\omega)$ antes del experimento
- $X(\omega) = x \in I$ después del experimento

Lo que estudiamos a continuación es cómo se distribuyen las probabilidades en I

Variable aleatoria Uniforme

- Supongamos que $I = [0, 1]$ y elegimos un punto al azar X en I.
- Supongamos que queremos indiferencia para la ubicación del resultado en $[0, 1]$
- Para eso suponemos que los sucesos

$$X \in \left[0, \frac{1}{2}\right], \qquad X \in \left[\frac{1}{2}, 1\right]$$

  tienen la misma probabilidad.
- ¿Qué probabilidad tendría que tener cada intervalo?

- Si queremos que además los sucesos

$$X \in \left[0, \frac{1}{4}\right], \quad X \in \left[\frac{1}{4}, \frac{1}{2}\right], \quad X \in \left[\frac{1}{2}, \frac{3}{4}\right], \quad X \in \left[\frac{3}{4}, 1\right].$$

  tienen la misma probabilidad, deberían tener probabilidad $\frac{1}{4}$

- La distribución uniforme es la que asigna para $0 \leq x < y \leq 1$:

$$\boxed{P(X \in [x, y]) = y - x}$$

- Es decir, a cada intervalo le asignamos como probabilidad su longitud.

Ejercicio

- Se elige un número X al azar en $[0,1]$ con distribución uniforme.
- ¿Cuál es la probabilidad de que la primera cifra decimal sea un 3?
$$x = 0,\underline{3}\ \_\ \cdots$$
- ¿Cuál es la probabilidad de que la segunda cifra decimal sea un 5?
$$x = 0,\_\ \underline{5}\ \_\ \cdots$$

Solución

- Sean los sucesos

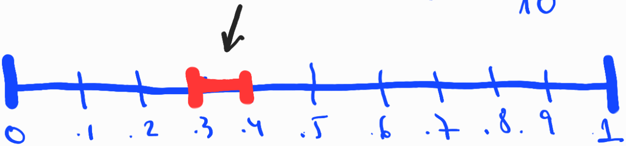
$$A = \{\text{la primer cifra es un 3}\}$$

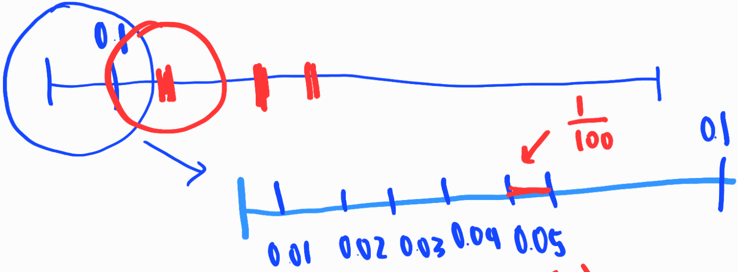
$$B = \{\text{la segunda cifra es un 5}\}$$

$$C = A \cap B = \{\text{primer cifra 3 } \textcolor{red}{\text{y}} \text{ segunda cifra 5}\}$$

0.3_ _ _ _

$$P(A) = \frac{1}{10}$$





Total: $10 \times \frac{1}{100} = \frac{1}{10} = P(B)$

Observación

- Tenemos
$$P(C) = \frac{1}{100}$$

- Entonces
$$P(A \cap C) = P(A)P(B)$$
Los sucesos A y B son independientes

- Pero . . . ¿cómo? ¿independientes?
- ¡Sí! Cumplen la definición
- La primer y segunda cifra de un mismo número aleatorio en $[0, 1]$ son independientes.
- Ocurre que el conocimiento de uno no tiene información sobe el resultado del otro

Variable uniforme en $[a, b]$

- ¿Cómo tenemos que modificar el resultado anterior si queremos un resultado en un intervalo cualquiera $[a, b]$?
- Tomamos la longitud relativa al intervalo total
- Si X es tiene distribución uniforme en $[a, b]$, se verifica

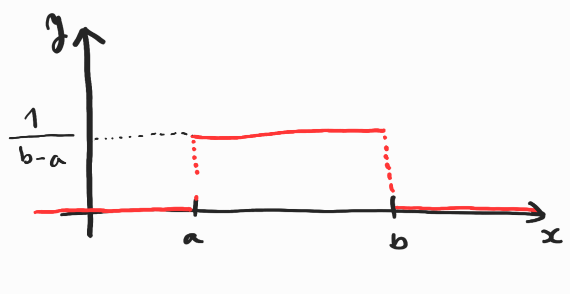
$$P(X \in [x, y]) = \frac{y - x}{b - a} \qquad a \leq x < y \leq b.$$

Densidad de la variable aleatoria uniforme en $[a, b]$

- La densidad de la variable aleatoria X uniforme en $[a, b$ es la función

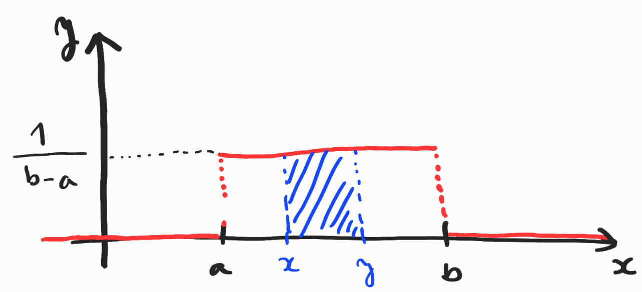
$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{cuando } a \leq x \leq b \\ 0, & \text{en otro caso} \end{cases}$$

se llama densidad de la variable aleatoria X, uniforme en $[a, b]$



- Las probabilidades se calculan como áreas:
- Para calcular áreas usamos integrales:

$$P(X \in [x, y]) = \int_x^y f(t)dt = \int_x^y \frac{1}{b-a}dt = \frac{y-x}{b-a}$$



Distribución de una variable aleatoria uniforme

- Más en general nos interesa la siguiente función:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

llamada función de distribución o también función de distribución de probabilidades e la variable aleatoria X

Calculamos

- Primero observamos que el cálculo es entre a y x (porque la $f(t)$ vale cero fuera de $[a, b]$:

$$F(x) = \int_a^x f(t)dt = \int_a^x \frac{1}{b-a}dt$$

$$= \frac{1}{b-a}\int_a^x 1dt = \frac{x-a}{b-a}$$

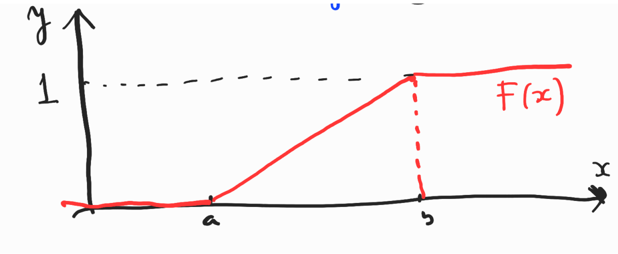
- Eso si $x \leq b$. En el caso en que $x > b$, tenemos

$$F(x) = \int_a^b f(t)dt = \int_a^b \frac{1}{b-a}dt$$

$$= \frac{1}{b-a}\int_a^b 1dt = \frac{b-a}{b-a} = 1$$

Obtenemos

$$F(x) = \begin{cases} 0, & \text{cuando } x < a \\ \frac{x-a}{b-a} & \text{cuando } a \leq x \leq b \\ 1, & \text{cuando } b \leq x \end{cases}$$



- También podemos obtener la densidad f a partir de la distribución F.
- Sabemos que se verifica

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

- Si derivamos con respecto a x obtenemos

$$F'(x) = f(x)$$

- La densidad f y la distribución F tienen la misma información: se puede obtener una a partir de la otra

- Con la función de distribución podemos calcular probabilidades.
- Si A es el suceso de obtener un 3 como primer dígito de X

$$
\begin{aligned}
P(A) = P(X \in [0{,}3, 0{,}4]) &= \int_{0{,}3}^{0{,}4} f(t)dt \\
&= F(0{,}4) - F(0{,}3) = 0{,}1.
\end{aligned}
$$

Simulación de una variable uniforme

Cada variable aleatoria tiene en R cuatro comandos asociados, que en el caso de la uniforme son

- `dunif` me da la densidad uniforme $f(x)$
- `punif` me da la distribución de probabilidades $F(x)$
- `qunif` me da la función inversa F^{-1} de la distribución F
- `runif` me numeros aleatorios (random) con distribución uniforme

Por ejemplo `runif(10,1,2)` me da 10 numeros aleatorios con distribución uniforme en $[1, 2]$:

```
> runif(10,1,2)
 [1] 1.306220 1.863714 1.076289 1.583812
 [5] 1.015811 1.149963 1.841357 1.572209
 [9] 1.837909 1.232794
```

Ahora . . . ¿cómo genera el R los números aleatorios?

- En primer lugar genera una variable aleatoria discreta con un N muy grande y divide por N.
- De esa forma obtiene un número en $[0, 1]$ que es casi continuo (tiene muchas cifras pero no infinitas)
- En segundo lugar, se genera una cantidad n (para ver como se distribuyen)
- Eso lo hace mediante un algoritmo
- Los números no son realmente aleatorios: si sabemos el algoritmo podemos “adivinarlos”.
- Se llaman pseudo-aleatorios

- El número $n+1$ se obtiene a partir del X_n:

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

donde

  - $m > 0$ es el módulo
  - $(x) \operatorname{mód} m$ es el resto de dividir x por m:
  - $(100) \operatorname{mód} 7 = 2$: siempre es un número entre 0 y $m = 6$
  - a tal que $0 < a < m$ es el multiplicador
  - c tal que $0 \leq c < m$ es el incremento
  - X_0 es la semilla

Variable aleatoria exponencial

- El segundo modelo continuo que consideramos es una variable aleatoria[1] T con distribución exponencial
- Es una variable que toma valores positivos: $I = [0, \infty)$
- Se puede definir a partir de su distribución:

$$F(t) = P(T \leq t) = 1 - e^{-\lambda t} \quad (t \geq 0)$$

- Derivando tenemos la densidad exponencial:

$$f(t) = F'(t) = \lambda e^{-\lambda t} \quad (t \geq 0)$$

[1] usamos la T porque podría tratarse de un tiempo aleatorio

- El modelo exponencial se utiliza para medir duraciones de componentes que no envejecen
- Por ejemplo, que se rompen por motivos externos a su desgaste
- Se puede usar para modelar el tiempo que dura una lámpara, que se quema por un pico de corriente externo
- Eso es porque cumple la propiedad de pérdida de memoria.
- Para eso usamos la sobrevida de T, que es

$$P(T > t) = e^{-\lambda t}.$$

Pérdida de memoria

- La pérdida de memoria es una propiedad matemática:

$$P(T > t + h \mid T > t) = P(T > h)$$

- Primero verificamos la propiedad:

$$\begin{aligned}
P(T > t + h \mid T > t) &= \frac{P(\{T > t + h\} \cap \{T > t\})}{P(T > t)} \\
&= \frac{P(\{T > t + h\})}{P(T > t)} \\
&= \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda(t+h)+\lambda t} \\
&= e^{-\lambda h} = P(T > h)
\end{aligned}$$

Ahora la interpretamos:

$$P(T > t + h \mid T > t) = P(T > h)$$

- Si T no ocurrió en tiempo t, esperar h más (hasta $t + h$) es equivalente a esperar h a partir de 0
- Por eso no importa cuanto tiempo ocurrió, si la lámpara no se rompió, es como que fuese nueva

Simulación de una variable exponencial

En R los cuatro comandos de una variable exponencial son:

- `dexp` me da la densidad exponencial
- `pexp` me da la distribución de probabilidades expnencial
- `qexp` me da la función inversa F^{-1} de la distribución F
- `rexp` me numeros aleatorios (random) con distribución exponencial
- Por ejemplo:

```
> rexp(3,1)
[1] 1.1611435 2.1357145 0.3589149
```

simula 3 variables exponenciales con parámetro $\lambda = 1$.