Variables Aleatorias (I). Introducción.

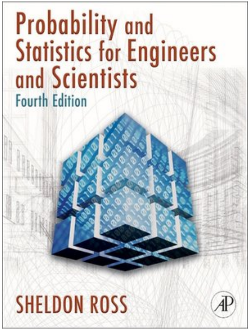
Estadística, Grado en Sistemas de Información

Constantino Antonio García Martínez

Universidad San Pablo Ceu

Contenidos

References



S. Ross. Introduction to Probability and Statistics for Engineers and Scientists. Chapter 4.

C.D. Barr, D.M. Diez, M. Çetinkaya-Rundel. OpenIntro Statistics. Chapters 2-3.

Variables aleatorias

Variables aleatorias

Variables aleatorias discretas

Cuando realizamos un experimento, generalmente no nos interesan los detalles del mismo, sino solo un valor numérico determinado por el resultado.

Ejemplo: Variables aleatorias

Después de 20 lanzamientos de una moneda, podemos haber obtenido el resultado:

$$CCCCCCCXCCXXCCCCCCXX$$

Sin embargo, la información puede sintetizarse con números:

- Hay 5 cruces.
- La frecuencia de las caras es 0.75.
- Hay 6 rachas.

En resumen, generalmente nos interesa **asociar a cada resultado de un experimento aleatorio, rasgos cuantitativamente medibles**.

Variable aleatoria
Una **variable aleatoria** X es una función de que asocia un resultado de un experimento, a un número:

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

Ejemplo: Variable aleatoria
Denotemos con X la variable aleatoria que es la suma del lanzamiento de dos dados. ¿Cuál es la probabilidad de X?

Variable aleatoria
Una **variable aleatoria** X es una función de que asocia un resultado de un experimento, a un número:

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

Ejemplo: Variable aleatoria
Denotemos con X la variable aleatoria que es la suma del lanzamiento de dos dados. ¿Cuál es la probabilidad de X?

$$P\{X=2\} = P\{(1,1)\} = \frac{1}{36}$$
$$P\{X=3\} = P\{(1,2),(2,1)\} = \frac{2}{36}$$
$$P\{X=4\} = P\{(1,3),(2,2),(3,1)\} = \frac{3}{36}$$
$$P\{X=5\} = P\{(1,4),(2,3),(3,2),(4,1)\} = \frac{4}{36}$$
$$P\{X=6\} = P\{(1,5),(2,4),(3,3),(4,2),(5,1)\} = \frac{5}{36}$$
$$P\{X=7\} = P\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} = \frac{6}{36}$$
$$P\{X=8\} = P\{(2,6),(3,5),(4,4),(5,3),(6,2)\} = \frac{5}{36}$$

$$P\{X=9\} = P\{(3,6),(4,5),(5,4),(6,3)\} = \frac{4}{36}$$
$$P\{X=10\} = P\{(4,6),(5,5),(6,4)\} = \frac{3}{36}$$
$$P\{X=11\} = P\{(5,6),(6,5)\} = \frac{2}{36}$$
$$P\{X=12\} = P\{(6,6)\} = \frac{1}{36}$$

Ejercicio: Variable aleatoria

Compramos 2 componentes electrónicos, cada uno de los cuales puede ser defectuoso o aceptable. La probabilidad de que sea defectuoso un componente es 0.3. Define una variable aleatoria para estudiar el número de componentes defectuosos en nuestra compra; define otra variable aleatoria para estudiar si al menos hay un componente aceptable.

Distribución de una VA
La colección $P(X \in B)$, siendo B cualquier subconjunto de $\mathbb{R}$, se denomina **la distribución de la VA X**. A la función que a cada posible número x_k asigna una probabilidad $P(X = x_k)$ se le denomina **función de probabilidad**.

En ocasiones, en lugar de trabajar con la distribución de una VA, es más útil emplear

Distribución acumulada de probabilidad (o función de distribución)

$$F(x) = P(X \leq x).$$

Muchas preguntas probabilísticas se pueden contestar empleando $F(x)$.

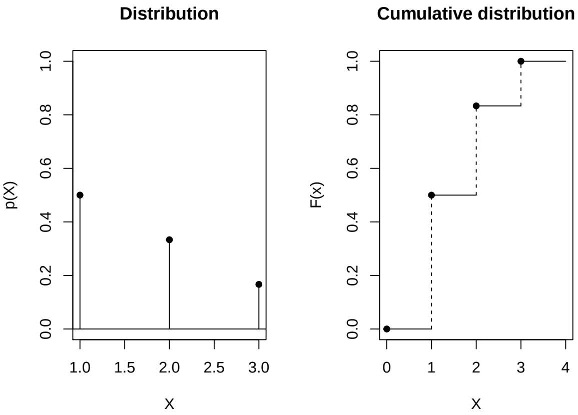
Ejemplo: Distribución acumulada de probabilidad
Para calcular $F(a < X \leq b)$ usamos que $\{X \leq b\}$ y $\{a < X \leq b\}$ son disjuntos:

$$P(\{X \leq b\}) = P(\{X \leq a\}) + P(\{a < X \leq b\}).$$

$$P(\{a < X \leq b\}) = F(b) - F(a).$$

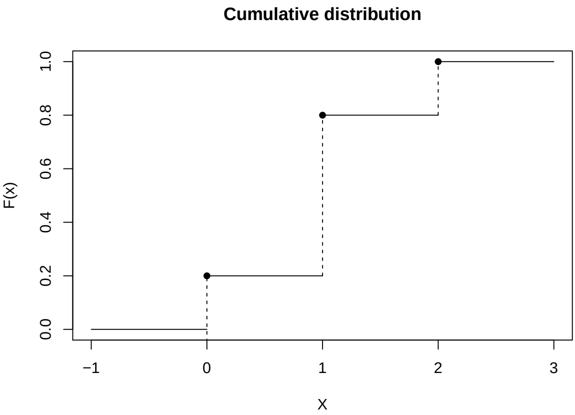
Ejemplo: Distribución de una Variable aleatoria

Considera una variable aleatoria X que es igual a $1, 2, 3$ con probabilidades $p(1) = 1/2$, $p(2) = 1/3$, $p(3) = 1/6$.



Ejercicio: Distribución de probabilidad

Halla la distribución de probabilidad a partir de la siguiente función de distribución:



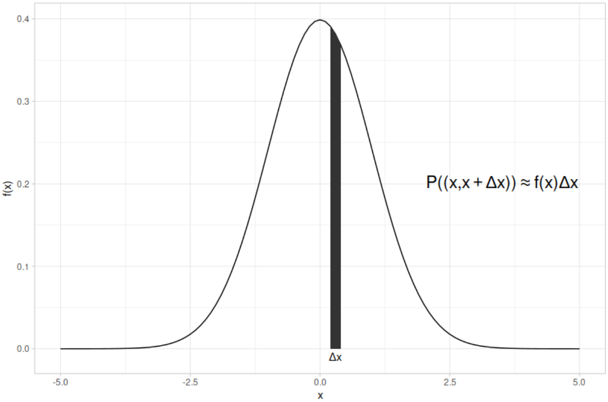
Variables aleatorias

Variables aleatorias continuas

En muchos casos, la observación de un fenómeno aleatorio consiste en realizar medidas de magnitudes continuas. En este caso lo adecuado es $\Omega = \mathbb{R}$. Sin embargo, este espacio muestral plantea dificultades teóricas:

- No es posible un modelo que asigne probabilidad a todos los subconjuntos de $\mathbb{R} \to$ Sólo se consideran como sucesos los intervalos de $\mathbb{R}$ y conjuntos que puedan formarse a partir de ellos.
- La probabilidad de un suceso no puede obtenerse sumando la probabilidad de sus puntos, ya que cada punto debe tener probabilidad 0.

Variables aleatorias continuas

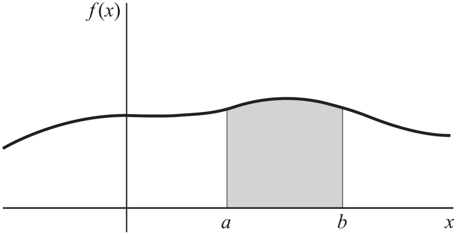


Función de densidad
Una función $f(x)$ se denomina **función de densidad de probabilidad (o función de masa)** si

1. $f(x) \geq 0 \qquad \forall\, x$
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1.$

Las probabilidades serán las áreas entre $f(x)$ y el eje x. P. ej.:

$$P(a < X < b) = \int_a^b f(x)\, dx.$$

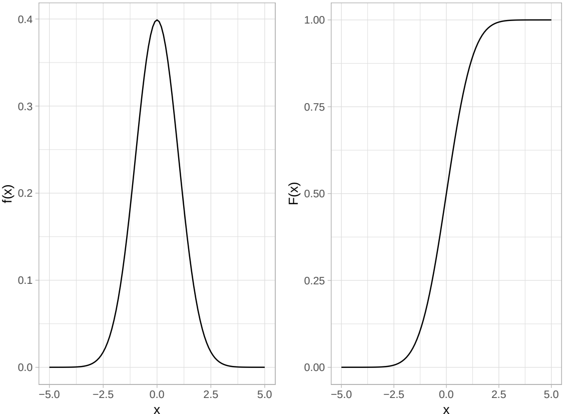


Función de densidad acumulada

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\, du.$$

Nótese que, por el teorema fundamental del cálculo:

$$\frac{dF(x)}{dx} = f(x).$$



Ejercicio:
Funciones de densidad

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

¿Cuál es el valor de C? ¿$P(X > 1)$?

Ejercicio: Funciones de densidad

Un puente es bombardeado. Por ser muy estrecho, podemos aproximarlo por una línea unidimensional y, por tanto, describirlo mediante una sola coordenada. El puente se encuentra entre los puntos a y b. ¿Cuáles son las fuciones de densidad y de densidad acumulada del puente? ¿Cuál es la probabilidad de alcanzar un segmento del puente de tamaño l?

Distribuciones conjuntas de variables aleatorias

En un experimento, generalmente estamos interesados en la relación entre distintas VAs.

Ejemplo: Relaciones entre VAs

- En un experimento sobre cáncer, la relación entre la cantidad de cigarrillos fumados diariamente y la edad de aparición del cáncer.
- La relación entre la temperatura de operación de una granja de servidores y la vida media de los mismos.
- La relación entre las emisiones de CO_2 y la temperatura media en el planeta.

En las próximas transparencias, nos centraremos en el caso en el que tengamos dos VAs, aunque es posible generalizar los resultados a más de dos variables.

Distribuciones conjuntas de variables aleatorias

Distribuciones conjuntas: caso discreto

Distribuciones conjuntas: caso discreto

Para especificar la relación entre dos VAs discretas X e Y usamos la **distribución de probabilidad conjunta** $P(X = x, Y = y) = f(x, y)$. Se verifica:

1. $f(x, y) \geq 0$.
2. $\sum_x \sum_y f(x, y) = 1$.

También es posible caracterizar la relación entre X e Y a través de **distribución de probabilidad acumulada conjunta**

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{u \leq x, v \leq y} f(u, v).$$

El conocimiento de la probabilidad conjunta nos permite calcular cualquiera de las **distribuciones marginales**:

$$F(x) = P(X \leq x) = P(X \leq x, Y < \infty) = F(x, \infty)$$

Otra forma: mantenemos fija la x en $f(x, y)$ y sumamos todos los posibles valores de y:

$$P(X = x) = f(x) = \sum_y f(x, y).$$

Ejercicio: Distribuciones conjuntas

3 baterías se eligen al azar entre un grupo de 3 nuevas, 4 usadas pero que funcionan, y 5 baterías defectuosas. Sea X el número de baterías nuevas; sea Y el número de baterías usadas pero funcionando. ¿Cuál es la distribución conjunta de X e Y? ¿Y las marginales? Calcúlalas mediante razonamiento directo y marginalizando $p(x, y)$.

Distribuciones conjuntas de variables aleatorias

Distribuciones conjuntas: caso continuo

Distribuciones conjuntas: caso continuo

En el caso de dos VAs continuas X e Y usamos la **distribución de densidad conjunta** $f(x,y)$. Intuitivamente:

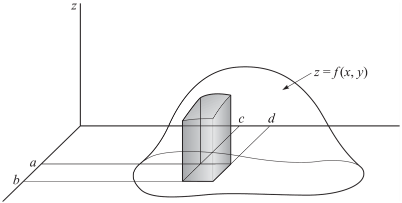
$$P\big(X \in (x, x+\Delta x), Y \in (y, y+\Delta y)\big) \approx f(x,y)\Delta x \Delta y$$

Se verifica que:

$$f(x,y) \geq 0 \qquad\qquad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1.$$

Gráficamente $z = f(x,y)$ represente una **superficie de probabilidad**. Los **volúmenes** entre esta superficie y el plano xy serán probabilidades. P.ej.:

$P(a < X < b, c < Y < d) =$
$\int_{x=a}^{b}\int_{y=c}^{d} f(x,y)\,dx\,dy.$



La **función de distribución conjunta** de X e Y se define como:

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, du\, dv.$$

De forma análoga al caso unidimensional, se deduce que

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y).$$

La **función de distribución conjunta** de X e Y se define como:

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, du\, dv.$$

De forma análoga al caso unidimensional, se deduce que

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y).$$

Las marginales se calculan de forma análoga al caso discreto:

$$F(x) = P(X \leq x) = P(X \leq x, Y < \infty) = F(x, \infty)$$

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\, dy. \qquad f(y) = \int_{-\infty}^{\infty} f(x, y)\, dx.$$

Ejercicio:
La distribución de X e Y viene dada por:

$$f(x) = \begin{cases} 2e^{(-x)}e^{(-2y)} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Calcula (a) $P(X > 1, Y < 1)$; (b) $P(X < Y)$; y (c) $P(X < a)$

Variables aleatorias independientes

Variables aleatorias independientes

Variables discretas independientes
Dos variables discretas son independientes si, y solo si, $\forall x, y$:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

Variables continuas independientes
Dos variables continuas son independientes si, y solo si, $\forall x, y$:

$$P(X \leq x, Y \leq y) = F_{x,y}(x, y) = F_x(x)F_y(y) = P(X \leq x)P(Y \leq y),$$

o de manera equivalente

$$f_{x,y}(x, y) = f_x(x)f_y(y).$$

Ejercicio: Variables independientes
Sean X e Y VAs con

$$f(x, y) = \begin{cases} c(2x + y) & x = 0, 1, 2; y = 0, 1, 2, 3 \\ 0 & \text{en otro caso} \end{cases}$$

Halla c y determina si las variables X e Y son independientes.

Distribuciones condicionales

Ya sabemos que

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}.$$

Análogamente, para **VAs discretas**, tendremos que la **función de probabilidad condicional de Y dado X** es

$$P(Y = y \mid X = x) = \frac{f_{x,y}(x, y)}{f_x(x)}.$$

Para **VAs continuas**, **función de densidad condicional de Y dado X** se

define como

$$f(y \mid x) = \frac{f_{x,y}(x, y)}{f_x(x)}.$$

Cambios de variable

Considera el cambio de variable $Y = g(X)$. Entonces,

$$F(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \in g^{-1}(-\infty, y)).$$

Ejemplo: Cambios de variable
Si $Y = aX + b$ y la distribución de X es conocida, calcula la distribución de Y.

En general, podemos simplificar el tratamiento del problema si la variable es discreta o absolutamente continua.

Si X es una variable discreta, tomando valores en un conjunto numerable D, podemos calcular directamente la función de probabilidad:

$$P(Y = y) = P(g(X) = y) = P(X \in g^{-1}(y)) = \underset{x \in D \cap g^{-1}(y)}{p(X = x)}.$$

Ejercicio: Cambio de variable discreta

Se X una VA con **distribución geométrica**:

$$P(X = n) = p(1-p)^{n-1} \qquad \text{para } n = 1, 2, 3, \ldots$$

¿Cuál es la distribución de $Y = (X-5)^2$?

Nos centraremos en casos en los que g es una función monótona. En tal caso, Si X es una VA continua, $Y = g(X)$ también lo es, y

Cambio de variable continua

$$f_x(x)|dx| = f_y(y)|dy|.$$

$$f_y(y) = f_x(x)\left|\frac{dx}{dy}\right| = f_x(g^{-1}(y)) \cdot |(g^{-1})'(y)| = \frac{f_x(g^{-1}(y))}{|g'(g^{-1}(y))|}.$$

Cambio de variable continua (variables variables)
Sean X e Y VAs continuas y sean $U = \phi_1(X, Y)$, $V = \phi_2(X, Y)$ donde a cada par de valores de X e Y le corresponde uno y sólo un par de valores de U y V, y viceversa. Entonces

$$f_{x,y}(x, y)\,|dx\,dy| = f_{u,v}(u, v)\,|du\,dv|\,.$$

$$f_{u,v} = f_{x,y}(x, y)\left|\frac{\partial(x, y)}{\partial(u, v)}\right| = f_{u,v}(\phi_1^{-1}(u, v), \phi_2^{-1}(u, v))\,|J|\,,$$

donde J denota el determinante Jacobiano

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix}$$

Ejercicio: Convoluciones

Sean X e Y VAs continuas con función de densidad $f(x, y)$. Demuestrese que la función de densidad de $U = X + Y$ es:

$$f_u(u) = \int_{-\infty}^{\infty} f(v, u - v)\, dv$$

Ejercicio: Convoluciones

Sean X e Y VAs continuas con función de densidad $f(x, y)$. Demuestrese que la función de densidad de $U = X + Y$ es:

$$f_u(u) = \int_{-\infty}^{\infty} f(v, u - v)\, dv$$

Para el caso especial donde X e Y son independientes, obtenemos que

$$f_u(u) = \int_{-\infty}^{\infty} f_x(x) f_y(u - x)\, dx,$$

que recibe el nombre de **convolución** de f_x y f_y. Lo notamos como

$$(f_x * f_y)(u) = \int_{-\infty}^{\infty} f_x(x) f_y(u - x)\, dx.$$

Es fácil comprobar que esta operación es conmutativa $(f_x * f_y)(u) = (f_y * f_x)(u)$.