1 Distribuciones continuas

Algunas de la más importantes variables aleatorias continuas, se listan a continuación:

1.1 La familia Normal

De acuerdo con el teorema del límite central, la distribución limitante de la suma estandarizada un gran número de variables aleatorias independientes,es la distribución normal.

La densidad de $N(\mu, \sigma^2)$ es:

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

La `distribución normal estándar` tiene media $\mu = 0$ y $\sigma^2 = 1$, así la función densidad:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

El CDF de esta variable aleatoria, se denota como Φ y se define como:

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

(Familia Normal) Si Z es una variable normal estándar, entonces:

$$\sigma Z + \mu \sim N(\mu, \sigma^2)$$

y si

$$X \sim N(\mu, \sigma^2), \text{ entonces } \frac{x-\mu}{\sigma} \sim N(0,1)$$

Si X y Z son normales e independientes, entonces $aX + bZ$ es también normal.

1.2 La familia Exponencial

La familia Exponencial se usa para modelar tiempos de espera. Una variable aleatoria exponencial `Exponencial (`λ`)` tiene una densidad:

$$f(t) = \lambda e^{-\lambda t}$$

y un CDF:

$$F(t) = 1 - e^{-\lambda t}.$$

Si definimos la función (Survival function):

$$G(t) = 1 - F(t) = e^{-\lambda t}.$$

Entonces la tasa de Hazard $f(t)/G(t)$ es igual a λ.

La familia Exponencial tiene la **Propiedad de Markov** o `memoryless`:

$$\mathbb{P}(T < t + s | T > t) = \mathbb{P}(T > s).$$

1.3 La familia Gamma

La familia de distribuciones `Gamma(r, λ)` es versátil. La distribución de la suma de r variables aleatorias `Exponencial(λ)` independientes, la distribución del r- ésimo tiempo de llegada en un proceso de Poisson con tasa de llegada λ es la distribución `Gamma(r, λ)`.

La distribución de la suma de cuadrados de n distribuciones normales estándar independientes, la distribución $\chi^2(n)$ es la distribución `Gamma(1/2, 1/2)`.

La distribución `Gamma(r, λ)` en general ($r > 0, \lambda > 0$), tiene una densidad dado por:

$$f(t) = \frac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t} \quad (t > 0).$$

1.4 Distribución de Cauchy

Si X e Y son normales estándar independientes, entonces Y/X tiene una distribución de Cauchy. La distribución de Cauchy, es la distribución de la tangente de un ángulo aleatoriamente seleccionado desde $[-\pi, \pi]$.

La densidad es:

$$f(x) = \frac{1}{\pi(1 + x^2)},$$

y el CDF:

$$F(t) = \frac{1}{\pi} \arctan(t) + \frac{1}{2}.$$

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{dx}{(1 + x^2)} &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{d \tan^{-1}(x)}{dx} \\ &= \frac{1}{\pi} [\tan^{-1}(\infty) - \tan^{-1}(-\infty)] = \frac{1}{\pi} [\frac{\pi}{2} - (-\frac{\pi}{2})] = 1.\end{aligned}$$

1.5 Distribución uniforme

Una variable aleatoria continua X que es igualmente probable que tome cualquier valor en un rango de valores (a, b), con $a < b$, da lugar a la distribución uniforme. Dicha distribución está uniformemente distribuida en su rango. Algunas veces se denota como $X \sim \text{Uniforme}(a, b)$.

La función densidad de la distribución es:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & x \in [a, b] \\ 0 & \text{en otros casos.} \end{cases}$$

El CDF es:

$$F(x) = \begin{cases} \dfrac{x-a}{b-a} & x \in [a, b] \\ 0 & \text{en otros casos.} \end{cases}$$

1.6 Distribución beta

X tiene una distribución Beta con parámetros $\alpha > 0$ y $\beta > 0$, denotado por $X \sim \text{Beta}(\alpha, \beta)$, si

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \ \ 0 < x < 1.$$

1.7 La distribución $\chi^2(n)$

Sea $Z_1, \ldots, Z_n$ variables aleatorias normales estándar. La distribución Chi-cuadrado con n grados de libertad, $\chi^2(n)$ es la distribución de:

$$R_n^2 = Z_1^2 + \cdots + Z_n^2.$$

La función densidad de esta distribución, está dado por:

$$f_{R_n^2}(t) = \frac{1}{2^{n/2}\Gamma(n/2)} t^{(n/2)-1} e^{-t/2}, \ \ (t > 0),$$

y escribimos:

$$R_n^2 \sim \chi^2(n).$$

Se sigue de la definición que si X e Y son independientes y $X \sim \chi^2(n)$ y $Y \sim \chi^2(m)$, entonces:

$$(X+Y) \sim \chi^2(n+m).$$

Esta es una distribución importante en estadística inferencial y es la base de la prueba de ajustes chi-cuadrado y el método de estimación de mínimos de chi-cuadrado.

Si hay m resultados posibles de un experimento y sea p_i la probabilidad de que el resultado i ocurre. Si el experimento se repite independientemente N veces, sea N_i el número de veces que se observa el resultado i, por lo que $N = N_1 + \cdots + N_m$. Entonces la estadística del chi-cuadrado:

$$\sum_{i=1}^{n} \frac{(N_i - Np_i)^2}{Np_i},$$

converge en distribución cuando $N \to$ a una distribución $\chi^2(m-1)$ con $m-1$ grados de libertad.

2 Funciones de variables aleatorias

Como se sabe, una variable aleatoria X es una función que asigna valores de $\mathbb{R}$ para cada resultado o salida de un espacio muestral. Dada una variable aleatoria X podemos definir otras variables aleatorias que son funciones de X. Por ejemplo, la variable aleatoria Y definida como $2X$ toma cada resultado en el espacio muestral y le asigna un real que es igual al doble del valor que X le asigna. La variable aleatoria $Y = X^2$ asigna un resultado que es el cuadrado del valor que X le asigna y así sucesivamente.

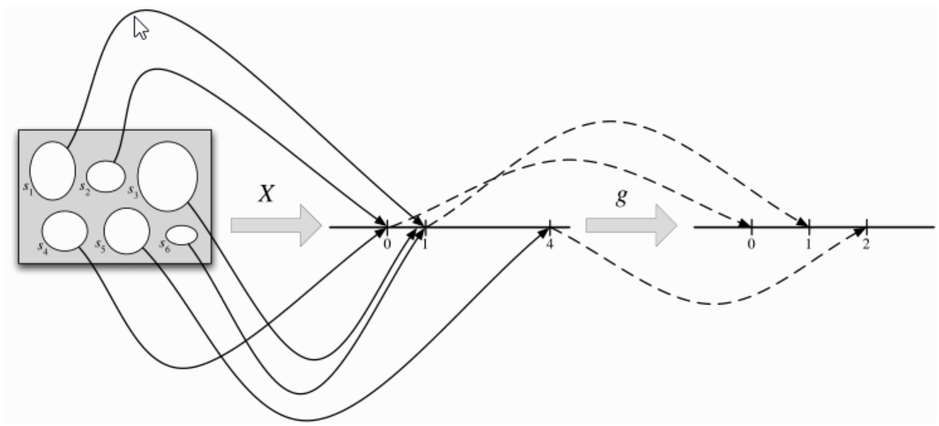
En general, si una variable aleatoria X asigna el valor x a un resultado y si $g(X)$ es una función de X, entonces $Y = g(X)$ es una variable aleatoria y se le asigna el valor $g(x)$ a ese resultado. Se dice que la variable aleatoria Y es una variable aleatoria derivada. Esto es, si X es una variable aleatoria, entonces X^2, e^X y $\sin(X)$ son variables aleatorias, como $g(X)$ para una función $g : \mathbb{R} \to \mathbb{R}$.

Por ejemplo, imaginemos que dos equipos de baloncesto (A y B) están jugando un partido de siete partidos, y que X sea el número de victorias para el equipo A (así que $X \sim \texttt{Binomial}(7, 1/2)$ si los equipos están igualados y los juegos son independientes). Sea $g(x) = 7$ y $h(x) = 1$ si $x \geq 4$ y $h(x) = 0$ si $x < 4$. Entonces $g(X) = 7 - X$ es el número de victorias para el equipo B y $h(X)$ es el indicador de que el equipo

A gana la mayoría de los juegos. Puesto que X es una variable aleatoria, $g(X)$ y $h(X)$ son también variables aleatorias.

Definición 2.1 Para un experimento con un espacio muestral Ω, una variable aleatoria X y una función $g : \mathbb{R} \to \mathbb{R}$, $g(X)$ es una variable aleatoria que mapea s a $g(X(s))$ para todo $s \in \Omega$.

Sea $g(X) = \sqrt{X}$, la siguiente figura muestra que $g(X)$ es la composición de las funciones X y g.En esta figura, la variable aleatoria X es definida sobre un espacio muestral con 6 elementos y tiene los posibles valores $0, 1$ y 4. La función g es la raíz cuadrada. Componiendo X y g nos da la variable aleatoria $g(X) = \sqrt{X}$ que tiene los valores $0, 1$ y 2.



Consideremos ahora el caso de una variable aleatoria discreta X cuya función de masa de probabilidad es $p_X(x)$. Dada una funci?n $g(X)$ de X, nos gustaría encontrar la función de masa de probabilidad de la variable aleatoria $Y = g(X)$, es decir, buscamos encontrar $p_Y(y)$.

Ejemplo 2.1 Sea X una variable aleatoria discreta con función de probabilidad, dada por,

$$p_X(x) = \begin{cases} 1/10 & x = 1, \\ 2/10 & x = 2, \\ 3/10 & x = 3, \\ 4/10 & x = 4, \\ 0 & \text{en otros casos.} \end{cases}$$

Con esta variable aleatoria, cada resultado en el espacio muestral es mapeado a uno de los cuatro números reales 1, 2, 3 o 4. Considere ahora la variable aleatoria $Y = 2X$. En este caso cada resultado se mapea en uno de los enteros 2, 4, 6 y 8. Pero ¿ qué pasa con la función de masa de probabilidad de Y?.

Es evidente que, si la probabilidad de que X asigne un resultado en 1 sea $1/10$, entonces la probabilidad de que Y asigne un resultado en 2 también debe ser $1/10$, si X otra vez asigna un resultado en 2 con probabilidad $2/10$, entonces, esta debe ser también la probabilidad de que Y asigne un resultado en 4 y así sucesivamente. En este ejemplo específico, tenemos:

$$p_Y(y) = \mathbb{P}(Y = g(x)) = \mathbb{P}(X = x) = p_X(x).$$

El caso anterior se da cuando g es una función uno a uno o inyectiva. El caso en el que $Y = g(X)$ con g inyectiva se ilustra en las siguientes tablas:

La idea es que si los distintos valores posibles de X son $x_1, x2, \ldots$ con probabilidades $p_1, p_2, \ldots$ (respectivamente), entonces los distintos valores posibles de Y son $g(x_1), g(x_2), \ldots$, con la misma lista $p_1, p_2, \ldots$, de probabilidades.

x	$P(X = x)$
x_1	p_1
x_2	p_2
x_3	p_3
$\vdots$	$\vdots$

(a) PMF de X

y	$P(Y = y)$
$g(x_1)$	p_1
$g(x_2)$	p_2
$g(x_3)$	p_3
$\vdots$	$\vdots$

(b) PMF de Y

Ejemplo 2.2 Una partícula se mueve n pasos en una línea numérica. La partícula comienza en 0 y en cada paso se mueve 1 unidad a la derecha o a la izquierda, con probabilidades iguales. Supongamos que todos los pasos son independientes. Sea Y la posición de la partícula después de n pasos. Hallemos el PMF de Y.

En efecto si consideramos cada paso como un ensayo de Bernoulli, donde ir por la derecha se considera un éxito e ir a la izquierda se considera un fracaso. Entonces el número de pasos que la partícula toma a la derecha es una variable aleatoria $\texttt{Binomial}(n, 1/2)$, que podemos nombrar como X. Si $X = j$, entonces la partícula ha tomado j pasos a la derecha y $n - j$ pasos a la izquierda, dando una posici?n final de $j - (n - j) = 2j - n$. Así podemos expresar Y como una función inyectiva de X, a saber $Y = 2X - n$ y desde que X toma valores en $\{0, 1, 2, \ldots, n\}$, Y toma valores en $\{-n, 2 - n, 4 - n, \ldots, n\}$.

El PMF de Y puede ser calculado desde el PMF de X:

$$\mathbb{P}(Y = k) = \mathbb{P}(2X - n = k) = \mathbb{P}(X = (n + k)/2) = \binom{n}{\frac{n+k}{2}} \left(\frac{1}{2}\right)^n.$$

si k es un entero entre $-n$ y n (inclusive) tal que $n + k$ es un número par.

Sin embargo, como muestra el siguiente ejemplo, no siempre se da el anterior caso:

Ejemplo 2.3 Sea X una variable aleatoria discreta, con una función de masa de probabilidad dada por:

$$p_X(x) = \begin{cases} 1/10 & x = -2, \\ 2/10 & x = -1, \\ 3/10 & x = 1, \\ 4/10 & x = 2, \\ 0 & \text{en otros casos.} \end{cases}$$

Con esta variable aleatoria X, cada resultado en el espacio muestral, es asignado a uno de los cuatro números reales $-2, -1, 1$ ó 2. Consideramos la variable aleatoria derivada $Y = X^2$. En este caso, cada resultado se le asigna 1 ó 4. Si X asigna un resultado -1 ó 1 , entonces Y asignará ese resultado a 1, mientras que si X asigna un resultado a -2 ó 2, entonces Y asignará ese resultado a 4.

La variable aleatoria Y tiene la siguiente función de masa de probabilidad, que no es la misma que la de X:

$$p_y(y) = \begin{cases} 2/10 + 3/10 & y = 1, \\ 1/10 + 4/10 & y = 4, \\ 0 & \text{en otros casos.} \end{cases}$$

Estos ejemplos ilustran la regla de que la función de masa de probabilidad de una variable aleatoria de Y , que es una función de una variable aleatoria X, es igual a la función de masa de probabilidad de X, si $g(x_1) \neq g(x_2)$ cuando $x_1 \neq x_2$. De lo contrario la función de masa de probabilidad de Y se obtiene de la relación general (general en el sentido de que también cubre el caso anterior):

$$p_Y(y) = \sum_{x:g(x)=y} p_X(x).$$

Aquí la suma es sobre todos los x, para el cuál $g(x) = y$.

Ejemplo 2.4 Sea X una variable aleatoria discreta, que es definida sobre los enteros en el intervalo $[-3, 4]$. Sea la función de masa de probabilidad de X:

$$p_X(x) = \begin{cases} 0.05 & x \in \{-3, 4\}, \\ 0.10 & x \in \{-2, 3\}, \\ 0.15 & x \in \{-1, 2\}, \\ 0.20 & x \in \{0, 1\}, \\ 0 & \text{en otros casos.} \end{cases}$$

Encontremos la función de masa de probabilidad de la variable aleatoria Y, definida como $Y = X^2 - |X|$.

Como los posibles valores de X son dados por $[-3, -2, -1, 0, 1, 2, 3, 4]$, se sigue que los valores de Y son $[6, 2, 0, 0, 0, 2, 6, 12]$ y así el rango de Y es $[0, 2, 6, 12]$. Esto permite calcular la función de masa de probabilidad de Y como:

$$p_Y(y) = \begin{cases} 0.15 + 0.20 + 0.20 = 0.55 & \text{si } y = 0 \quad (g(x) = y \text{ para } x = -1, 0, 1) \\ 0.10 + 0.15 = 0.25 & \text{si } y = 2 \quad (g(x) = y \text{ para } x = -2, 2), \\ 0.05 + 0.10 = 0.15 & \text{si } y = 6 \quad (g(x) = y \text{ para } x = -3, 3), \\ 0.05 = 0.05 & \text{si } y = 12 \quad (g(x) = y \text{ para } x = 4), \\ 0 & \text{en otros casos.} \end{cases}$$

Ejemplo 2.5 Continuando con el `Ejemplo 1.2`, sea D la distancia de la partícula, desde el origen después de n pasos. Asumiendo que n es par. Encontremos el PMF de D.

Podemos es escribir $D = |Y|$, esta es una función de Y, pero no es inyectiva. El evento $D = 0$, es el mismo que el evento $Y = 0$. Para $k = 2, 4, \ldots, n$, el evento $D = k$ es el mismo evento $\{Y = k\} \cup \{Y = -k\}$.

Así el PMF de D es:

$$\begin{aligned} \mathbb{P}(D = 0) &= \binom{n}{\frac{n}{2}} \left(\frac{1}{2}\right), \\ \mathbb{P}(D = k) &= \mathbb{P}(Y = k) + \mathbb{P}(Y = -k) = 2\binom{n}{\frac{n+k}{2}} \left(\frac{1}{2}\right), \end{aligned}$$

para $k = 2, 4, \ldots, n$. En el paso final, usamos la simetria para ver que $\mathbb{P}(Y = k) = \mathbb{P}(Y = -k)$.

En el caso de variables aleatorias continuas, un enfoque es empezar con la función de distribución acumulativa de $g(X)$ y trasladar el evento $g(X) \leq y$ en un evento equivalente que implica X. Para una función general g, debemos expresar de forma acertada $g(X) \leq y$ en términos de X y no hay una fórmula fácil para eso. Pero cuando g es continua y estrictamente creciente, podemos encontrar una expresión conveniente, si $g(X) \leq y$ es lo mismo que $X \leq g^{-1}(y)$ y así tenemos que:

$$F_{g(X)}(y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}y).$$

Entonces podemos diferenciar con respecto a y para obtener el PDF de $g(X)$. Esto da una versión unidimensional de la fórmula de cambio de variables, que se generaliza a transformaciones invertibles en múltiples dimensiones.

Teorema 2.1 (Cambio de variable en una dimensión) Sea X una variable aleatoria continua con PDF f_X y sea $Y = g(X)$, donde g es diferenciable y estrictamente creciente (o estrictamente creciente). Entonces el PDF de Y es dado por:

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|,$$

donde $x = g^{-1}(y)$.

Sea g una función estrictamente creciente. El CDF de Y es:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) = F_X(x),$$

Así por la regla de la cadena, el PDF de Y es:

$$f_Y(y) = f_X(x)\frac{dx}{dy}.$$

La prueba para g estrictamente decreciente es análogo. En ese caso el PDF, termina como, $-f_X(x)\dfrac{dx}{dy}$, que no es negativa, desde que $\dfrac{dx}{dy} < 0$ si g es estrictamente decreciente. Usando $\left|\dfrac{dx}{dy}\right|$ como en la declaración del teorema, abarca ambos casos.

Definición 2.2 El soporte de una variable aleatoria continua X y su distribución, es el conjunto de todos los puntos x donde $f_X(x) > 0$.

Ejemplo 2.6 La función de distribución acumulativa de una variable aleatoria exponencial X con paramétro λ es dado por:

$$\mathbb{P}(X \leq x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0 & \text{en otro caso.} \end{cases}$$

Calculemos la función de distribución acumulativa y la función de densidad de una variable aleatoria Y, definida como $Y = X^2$. Nosotros tenemos:

$$\mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(X \leq \sqrt{y}) = 1 - e^{-\lambda\sqrt{y}} \text{ para } y \geq 0.$$

Podemos calcular la función densidad de Y, por diferenciación. Así tenemos:

$$f_Y(y) = \frac{d}{dy}\mathbb{P}(Y \leq y) = \lambda e^{-\lambda y^{1/2}} \times \frac{1}{2}y^{-1/2} = \frac{\lambda}{2\sqrt{y}}e^{-\lambda\sqrt{y}} \text{ para } y > 0.$$

En experimentos de simulación es frecuente que se tenga acceso a secuencias de números aleatorios que se distribuyen uniformemente en el intervalo $(0, 1)$, pero lo que realmente se requiere son números aleatorios de una distribución no uniforme. En particular, en teoria de colas, a menudo se necesitan números aleatorios que se distribuyen de acuerdo con una distribución exponencial (negativa). Este último ejemplo muestra cómo se pueden obtener tales números.

Ejemplo 2.7 Sea $X \sim N(0, 1), Y = e^X$. Usemos el teorema de cambio de variable para encontrar el PDF de Y, desde que $g(x) = e^X$ es estrictamente creciente. Sea $y = e^X$, Así $x = \log y$ y $dy/dx = e^x$. Entonces:

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = \varphi(x)\frac{1}{e^x} = \varphi(\log y)\frac{1}{y},\ y > 0.$$

Ten en cuenta que después de aplicar la fórmula de cambio de variables, escribimos todo a la derecha en términos de y y especificamos el soporte [1] de la distribución. Para determinar el soporte, sólo observamos que cuando x varía desde $-\infty$ a ∞, e^X varía desde 0 a ∞.

Podemos obtener el mismo resultado trabajando a partir de la definición de la función de distribución acumulativa, transladando el evento $Y \leq y$ en un evento equivalente involucrando X. Para $y > 0$,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(e^X \leq y) = \mathbb{P}(X \leq \log y) = \Phi(\log y),$$

y así el PDF, otra vez es:

$$f_Y(y) = \frac{d}{dy}\Phi(\log Y) = \varphi(\log y)\frac{1}{y},\ y > 0.$$

Ejemplo 2.8 Sea la función densidad de probabilidad de una variable aleatoria X, es dado por:

$$f_X(x) = \begin{cases} \alpha/x^5, & 1 \leq x < \infty \\ 0 & \text{en otros casos.} \end{cases}$$

Sea una nueva variable aleatoria definida como $Y = 1/X$. Encontremos la función densidad de Y.

Primero responderemos a la pregunta usando el enfoque estándar. Nuestra primera tarea es calcular el valor de α. Ya que debemos tener:

$$1 = \int_{-\infty}^{\infty} f_X(x)dx = \int_1^{\infty} \frac{\alpha}{x^5}dx = -\alpha\frac{x^{-4}}{4}\Big|_1^{\infty} = \frac{\alpha}{4},$$

y por propiedad $\alpha = 4$. La función distribución acumulativa de X, ahora se puede calcular como:

$$F_X(x) = \int_1^x f_X(t)dt = \int_1^x \frac{4}{t^5}dt = -t^{-4}\Big|_1^x = 1 - \frac{1}{x^4},\ x \geq 1.$$

Ahora, usando probabilidades y observando que el rango de Y es $(0, 1]$, podemos calcular la distribución acumulativa de Y como:

$$\mathbb{P}(Y \leq y) = \mathbb{P}(1/X \leq y) = \mathbb{P}(X \geq 1/y) = 1 - \mathbb{P}(X < 1/y) = 1 - F_X(1/y) = y^4,\ \text{para}\ \ 0 < Y \leq 1.$$

Finalmente, calculamos la función densidad de Y, tomando derivadas. Así obtenemos:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}y^4 = 4y^3,\ 0 < y \leq 1.$$

Utilizando el teorema de cambio de variables, cuando $y = 1/x$ y así $x = 1/y$, obtenemos

$$f_Y(y) = \frac{4}{(1/y)^5}\left|\frac{-1}{y^2}\right| = 4y^3$$

como antes.

[1] Si X es una variable aleatoria discreta, el conjunto finito o infinito contable de valores x tal que $\mathbb{P}(X = x) > 0$, es llamado soporte de X.

Ejemplo 2.9 Sea X una variable aleatoria continua con PDF, f_X y sea $Y = a + bX$, con $b \neq 0$. Sea $y = a + bx$, el espejo de la relación entre Y y X. Entonces $\frac{dy}{dx} = b$ y así el PDF de Y es:

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = f_X\left(\frac{y-a}{b}\right)\frac{1}{|b|}.$$

3 Independencia de variables aleatorias

Así como tuvimos la noción de independencia de eventos, podemos definir la independencia de las variables aleatorias. Intuitivamente, si dos variables aleatorias X e Y son independientes, entonces conocer el valor de X no da ninguna información sobre el valor de Y y viceversa. La siguiente definición, formaliza esta idea.

Definición 3.1 La variables aleatorias X e Y se dice que son independientes si:

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y),$$

para todo $y \in \mathbb{R}$. En el caso discreto, este caso es equivalente a la condición:

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y),$$

para todo x, y con x en el soporte de X y y en el soporte de Y.

La definición para más variables aleatorias es análogo:

Definición 3.2 La variables aleatorias $X_1, \ldots X_n$ son independientes si:

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1)\ldots\mathbb{P}(X_n \leq x_n)$$

para todo $x_1, \ldots, x_n \in \mathbb{R}$. Para infinitas variables aleatorias, decimos que son independientes si cada subconjunto finito de las variables aleatorias es independiente.

Comparando esto con el criterio de independencia de n eventos, puede parecer extraõ que la independencia de $X_1, \ldots, X_n$ variables aleatorias requiera sólo una igualdad, mientras que para la independencia de eventos necesitamos verificar la independencia de pares para todos los $\binom{n}{2}$ pares, independencia de a tres paras las $\binom{n}{3}$ triples y así sucesivamente.

Sin embargo, al examinar más detenidamente la definición, vemos que la independencia de variables aleatorias requiere que la igualdad se cumpla para todos los posibles $x_1, \ldots, x_n$ infinitamente muchas condiciones.Si podemos encontrar una única lista de valores $x_1, \ldots, x_n$ para el cuál la igualdad falla, entonces $X_1, \ldots, X_n$ no son independientes.

Ejemplo 3.1 En el lanzamiento de dos dados, si X es el número del primer dado y Y es el número del segundo dado, entonces $X + Y$ no es independiente de $X - Y$. Para ver esto, debes notar que:

$$0 = \mathbb{P}(X + Y = 12, X - Y = -1) \neq \mathbb{P}(X + Y = 12)\mathbb{P}(X - Y = 1) = \frac{1}{36} \cdot \frac{5}{36}.$$

Desde que hemos encontrado un par de valores (s, d), para el cual:

$$\mathbb{P}(X + Y = s, X - Y = -d) \neq \mathbb{P}(X + Y = s)\mathbb{P}(X - Y = d)$$

$X + Y$ y $X - Y$ son dependientes. Esto también tiene sentido intuitivamente: sabiendo que la suma de los dados es 12 nos dice que su diferencia debe ser 0, por lo que los variables aleatorias proporcionan información de cada una de ellas.

Definición 3.3 Las variables que son independientes y tienen las misma distribución, se les llama independientes y idénticamente distribuidas o `i.i.d` o `I.I.D`.

"Independiente" e "idénticamente distribuidos" son dos conceptos a menudo confundidos pero completamente diferentes. Las variables aleatorias son independientes si no proporcionan informaci?n entre sí; se distribuyen de forma idéntica si tienen el mismo PMF (o equivalentemente, el mismo CDF). Si dos variables aleatorias son independientes no tiene nada que ver con que si tienen o no la misma distribución. Podemos tener variables aleatorias:

- Independientes e idénticamente distribuidas: Sea X el resultado del lanzamiento de un dado, y sea Y el resultado del lanzamiento independiente de un segundo de dado . Entonces X e Y son `i.i.d`.
- Independientes y no idénticamente distribuidas: Sea X el resultado del lanzamiento de un dado y sea Y el precio de cierre de la bolsa de valores de Lima (un índice bursátil) en un mes a partir de ahora. Entonces X e Y no proporcionan ninguna información sobre cada una y X e Y no tienen la misma distribución.
- Dependientes e idénticamente distribuidas: Sea X el número de caras en n lanzamientos independientes de una moneda y sea Y el número de sellos en esos mismos n lanzamientos. Entonces X e Y son ambas distribuidas por $\texttt{Binomial}(n, 1/2)$, pero son altamente dependientes: si sabemos X, entonces conocemos Y perfectamente.
- Dependientes y no idénticamente distribuidas: Sea X el indicador de si el partido mayoritario retiene el control de la cámara de Representantes en los Estados Unidos despu?s de las próximas elecciones y sea Y la calificación de favorabilidad promedio del partido mayoritario en las encuestas tomadas a un mes de la elección. Entonces X e Y son dependientes y X e Y no tienen la misma distribución.

4 Variable aleatoria condicionada

Un evento A puede ser condicionado por un evento B diferente y la probabilidad asignada a A puede cambiar, permanecer igual o incluso llegar a cero como resultado de saber que el evento B ocurre. Escribimos anteriormento $\mathbb{P}(A|B)$ como la probabilidad del evento A dado el evento B. Debido a que una variable aleatoria X define eventos en un espacio muestral, se sigue que las probabilidades asignadas a variables aleatorias también pueden cambiar al saber que un cierto evento B ha ocurrido.

El evento $[X = x]$ contiene todos los resultados que son asignados por la variable aleatoria X al número real x y su probabilidad, $\mathbb{P}(X = x)$, es igual a la suma de las probabilidades de estos resultados. Sabiendo que un evento B ha ocurrido se puede alterar $\mathbb{P}(X = x)$, para todos los valores posibles de x. La probabilidad condicional $\mathbb{P}(X = x|B)$, se define cuando $\mathbb{P}(B) > 0$.

Cuando la variable aleatoria X es discreta, $\mathbb{P}(X = x|B)$ se denomina función de masa de probabilidad condicional de X y se denota como $p_{X|B}(x)$. De nuestra definición previa de probabilidad condicional para dos eventos, podemos escribir:

$$p_{X|B}(x) = \frac{\mathbb{P}([X = x] \cap B)}{\mathbb{P}(B)}$$

En muchos ejemplos, el evento $[X = x]$ está contenido en el evento B o su intersección de $[X = x]$ y B es el evento nulo. En el primer caso tenemos:

$$p_{X|B}(x) = \frac{p_X(x)}{\mathbb{P}(B)} \text{ si } [X = x] \subset B$$

Mientras que el segundo caso tenemos : $[X = x] \cap B = \emptyset$ y así $p_{X|B}(x) = 0$.

Estos conceptos se trasladan de manera natural a variables aleatorias que son continuas. Para una variable aleatoria continua X y un evento B, podemos escribir:

$$f_{X|B}(x)dx = \mathbb{P}(x < X \leq x + dx|B) = \frac{\mathbb{P}([x < X \leq x + dx] \cap B)}{\mathbb{P}(B)}.$$

Así

$$f_{X|B}(x) = \begin{cases} f_X(x)/\mathbb{P}(B) & \text{si } [X = x] \subset B \\ 0 & \text{si } [X = x] \cap B = \emptyset \end{cases}$$

Ejemplo 4.1 Sea X la variable aleatoria que cuenta el número de puntos obtenidos cuando se lanzan dos dados. La función de masa de probabilidad para X es:

x_i	2	3	4	5	6	7	8	9	10	11	12
$p_X(x_i)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Sea B el evento de que el lanzamiento da un número par de puntos en uno de los dados y un número impar en el otro dado. Se sigue que el evento B contiene 18 resultados. Cada resultado es un par con la propiedad de que el primero puede ser cualquiera de los seis números, pero el segundo debe ser uno de los tres números pares (si el primero es impar) o uno de tres números impares (si el primero es par). Como todos los resultados son igualmente probables, tenemos que $\mathbb{P}(B) = 1/2$.

Dado este evento calculamos la función de masa de probabilidad condicional $p_{X|B}$. Desde que la intersección de B y $[X = x]$, cuando x es un número par es vacío, se sigue que la probabilidad de que la suma siendo un número par es cero.

Para valores impares de x, el evento $[X = x]$ está enteramente contenido dentro del evento B. Por tanto si resumimos, todos estos resultados, tenemos que la función de masa de probabilidad condicional es obtenida como:

$$p_{X|B}(x) = \begin{cases} p_X(x)/\mathbb{P}(B), & x = 3, 5, 7, 9, 11, \\ 0 & \text{en otros casos,} \end{cases}$$

y es presentada en forma tabular de la siguiente manera:

x_i	2	3	4	5	6	7	8	9	10	11	12
$p_{X|B}(x_i)$	0	2/18	0	4/18	0	6/18	0	4/18	0	2/18	0

Si $B_i, i = 1, 2, \ldots, n$ es un conjunto mutualmente exclusivos y colectivamente exhaustivos (espacio de eventos), entonces se cumple:

$$p_X(x) = \sum_{i=1}^{n} p_{X|B_i}(x)\mathbb{P}(B_i)$$

El resultado se sigue, de aplicar la ley de Probabilidad de Total al evento $[X = x]$.

Ejemplo 4.2 Sea X una variable aleatoria con función densidad dada por:

$$f_X(x) = \begin{cases} (1/2)e^{-x/2} & \text{si } x \geq 0 \\ 0 & \text{en otros casos.} \end{cases}$$

Sea B el evento que $X < 1$ y deseamos encontrar $f_{X|X<1}(x)$. Para ello, calculemos $\mathbb{P}(X < 1)$ como:

$$\mathbb{P}(X < 1) = \int_0^1 (1/2)e^{-x/2}dx = -e^{-x/2}\Big|_0^1 = 1 - e^{-1/2}.$$

La función densidad de probabilidad condicional de X es entonces dado por:

$$\begin{aligned} f_{X|X<1}(x) &= \begin{cases} f_X(x)/\mathbb{P}(X < 1) & \text{para } 0 \leq x < 1, \\ 0 & \text{en otro caso,} \end{cases} \\ &= \begin{cases} (1/2)e^{-x/2}/(1 - e^{-1/2}) & \text{para } 0 \leq x < 1 \\ 0 & \text{en otro caso,} \end{cases} \end{aligned}$$

La función de distribución acumulativa condicional, $F_{X|B}(x|B)$, de una variable aleatoria X, dado que B ha ocurrido, es definida como:

$$F_{X|B}(x|B) = \frac{\mathbb{P}(X \leq x, B)}{\mathbb{P}(B)}$$

donde $(X \leq x, B)$ es la intersección de los eventos $[X \leq x]$ y B. Además se tiene que:

$$F_{X|B}(-\infty|B) = 0 \;\text{ y }\; F_{X|B}(\infty|B) = 1$$

También:

$$\mathbb{P}(x_1 < X \leq x_2|B) = F_{X|B}(x_2|B) - F_{X|B}(x_1|B) = \frac{\mathbb{P}([x_1 < X \leq x_2], B)}{\mathbb{P}(B)}.$$

La función densidad condicional, puede ser obtenida como la derivada:

$$f_{X|B}(x|B) = \frac{d}{dx}F_{X|B}(x|B)$$

que debe ser no negativa y debe tener un área igual a 1.