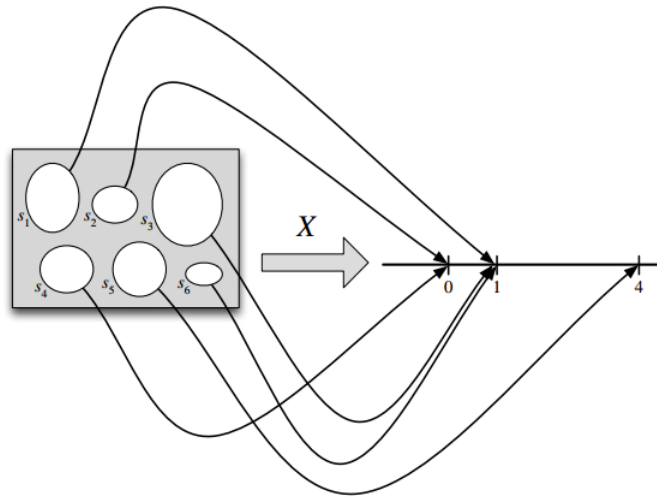


1 Variables aleatorias y funciones de distribución

1.1 Variables aleatorias discretas y continuas

En muchos modelos probabilísticos, las salidas o resultados son numéricos, por ejemplos, las lecturas de instrumentos o el precio de las acciones. En otros experimentos, las salidas no son numéricas, pero pueden estar asociadas a un valor numérico. Por ejemplo, si el experimento es la selección de estudiantes desde una población, podemos considerar su promedio de calificaciones. Cuando tratamos con esos valores numéricos, es a menudo útil asignarles una probabilidad. Esto se puede hacer con la noción de **variable aleatoria**.

Dado un experimento y el correspondiente conjunto de posibles respuestas o salidas (el espacio muestral), una variable aleatoria asocia un número particular con cada respuestas o salidas, como se indica en la figura siguiente, donde la variable aleatoria X representada, es definida sobre un espacio muestral con 6 elementos y los valores posibles 1, 2 y 4. La aleatoriedad viene de la elección de un punto aleatorio de acuerdo con la función de probabilidad \mathbb{P} para el espacio muestral.



Definición 1.1 Una variable aleatoria es una función $X : \Sigma \rightarrow \mathbb{R}$ que asigna un número real $X(\omega)$ a cada salida o respuesta ω .

Asumiendo $E \subset \mathbb{R}$, denotamos el evento $\{\omega \in \Omega : X(\omega) \in E\} \subset \Omega$ por $\{X \in E\}$ o sólo $X \in E$.

- Las variables aleatorias son denotadas por letras mayúsculas X, Y, Z .
- Algunas veces usamos la notación $\{X \in E\}$ o $X \in E$ de forma indistinta. Por ejemplo, la notación $X = x$ corresponde a $\{X \in \{x\}\}$ y la notación $a < X < b$ corresponde a $\{X \in (a, b)\}$.
- Mientras que X denota una función, la notación $X \in A$ y $X = x$ corresponde a subconjuntos de Ω .

Una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$, define una función de probabilidad \mathbb{P}' sobre un nuevo espacio muestral $\Omega' = \mathbb{R}$:

$$\mathbb{P}'(E) = \mathbb{P}(X \in E), \quad E \subset \mathbb{R} = \Omega'.$$

Las variables aleatorias que toman valores de un conjunto discreto de números (es decir, cuyo rango consiste de puntos aislados en la recta real) se denominan variables aleatorias discretas. El conjunto puede ser infinito pero contable. Los ejemplos a continuación son discretos y finitos. Las variables aleatorias discretas se utilizan para representar distintos elementos indivisibles, como personas, automóviles, árboles, etc.

Las variables aleatorias continuas toman valores en un intervalo continuo. Por ejemplo, una variable aleatoria que representa el tiempo entre dos llegadas sucesivas de un sistema de colas, o la que representa la temperatura en un reactor nuclear, es un ejemplo de una variable aleatoria continua.

El tiempo en el primer caso, o temperatura en el segundo, son reales pertenecientes a un subconjunto continuo del sistema de números reales. Las variables aleatorias que representan cantidades como tiempo, temperatura, distancia, área, volumen, etc, pueden discretizarse en minutos, horas, yardas, millas, etc., en cuyo caso es apropiado utilizar variables aleatorias discretas para representarlas.

Todas las variables aleatorias definidas en un espacio muestral discreto deben ser discretas. Sin embargo, las variables aleatorias definidas en un espacio muestral continuo pueden ser discretas o continuas.

Ejemplo 1.1 Sea el experimento de probabilidad que consiste en lanzar dos dados. Hay 36 resultados posibles que pueden ser representados por los pares $(i, j); i = 1, 2, \dots, 6, j = 1, 2, \dots, 6$. Estos son los 36 elementos del espacio muestral. Es posible asociar muchas variables aleatorias a este espacio muestral (dominio). Una posibilidad es la variable aleatoria X , definida por la siguiente enumeración como

$$\begin{aligned} X(1, 1) &= 2; \\ X(1, 2) &= X(2, 1) = 3; \\ X(1, 3) &= X(2, 2) = X(3, 1) = 4; \\ X(1, 4) &= X(2, 3) = X(3, 2) = X(4, 1) = 5; \\ X(1, 5) &= X(2, 4) = X(3, 3) = X(4, 2) = X(5, 1) = 6; \\ X(1, 6) &= X(2, 5) = X(3, 4) = X(4, 3) = X(5, 2) = X(6, 1) = 7; \\ X(2, 6) &= X(3, 5) = X(4, 4) = X(5, 3) = X(6, 2) = 8; \\ X(3, 6) &= X(4, 5) = X(5, 4) = X(6, 3) = 9; \\ X(4, 6) &= X(5, 5) = X(6, 4) = 10; \\ X(5, 6) &= X(6, 5) = 11; \\ X(6, 6) &= 12. \end{aligned}$$

Así, X asigna el resultado $(1,1)$ al número real 2, los resultados $(1,2)$ y $(2,1)$ al número real 3 y así sucesivamente. En este caso la variable aleatoria X representa la suma de los puntos obtenidos cuando se lanzan dos dados simultáneamente. El dominio de esta función son los 36 elementos del espacio muestral y su rango es el conjunto de enteros desde el 2 a 12.

Observa que la variable aleatoria X divide el espacio muestral en 11 subconjuntos; es decir, en 11 eventos que son mutuamente excluyentes y colectivamente exhaustivos: en otras palabras, estos 11 eventos forman un espacio de eventos. Los subconjuntos se distinguen de acuerdo con el valor particular tomado por la función. Así $\{(1,3), (2,2), (3,1)\}$ es el subconjunto que consiste de elementos del espacio muestral que se asignan al número real 4; $\{(5,6), (6,5)\}$ es el subconjunto que consiste de los elementos que se asignan al número real 11 y así sucesivamente.

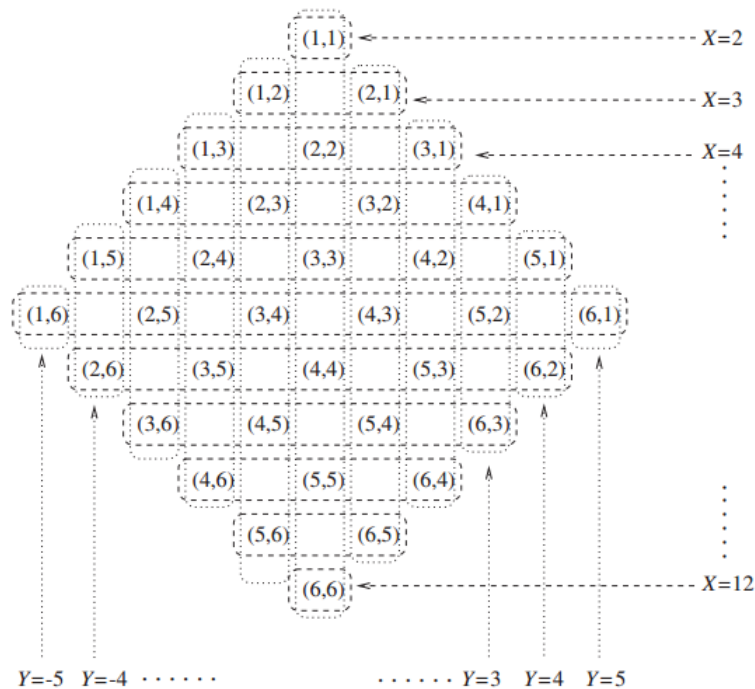
Cada resultado en el espacio muestral se asocia con un número real, varios se pueden asociar con el mismo número real: por ejemplo, $(1,3)$, $(2,2)$ y $(3,1)$ están todos asociados con el número entero 4, pero ninguno está asociado con más de un número real y esto es porque X se ha definido como una función.

Se pueden definir diferentes variables aleatorias en el mismo espacio muestral. Con el mismo dominio, pero con un rango que puede ser diferente.

Ejemplo 1.2 Consideremos una variable aleatoria diferente Y cuyo dominio es el mismo espacio muestral que el de X , pero que se define como la diferencia en el número de puntos en los dos dados. En este caso tenemos:

$$\begin{aligned}
Y(1, 1) &= Y(2, 2) = Y(3, 3) = Y(4, 4) = Y(5, 5) = Y(6, 6) = 0; \\
Y(2, 1) &= Y(3, 2) = Y(4, 3) = Y(5, 4) = Y(6, 5) = 1; \\
Y(1, 2) &= Y(2, 3) = Y(3, 4) = Y(4, 5) = Y(5, 6) = -1; \\
Y(3, 1) &= Y(4, 2) = Y(5, 3) = Y(6, 4) = 2; \\
Y(1, 3) &= Y(2, 4) = Y(3, 5) = Y(4, 6) = -2; \\
Y(4, 1) &= Y(5, 2) = Y(6, 3) = 3; \\
Y(1, 4) &= Y(2, 5) = Y(3, 6) = -3; \\
Y(5, 1) &= Y(6, 2) = 4; \\
Y(1, 5) &= Y(2, 6) = -4; \\
Y(6, 1) &= 5; \\
Y(1, 6) &= -5.
\end{aligned}$$

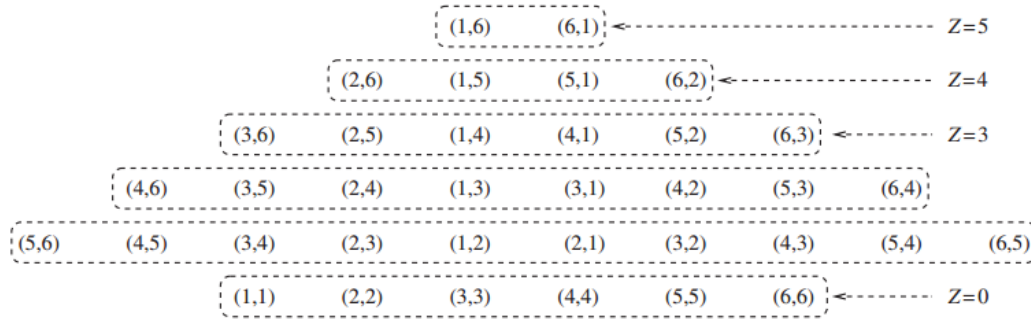
En este caso observa que la variable aleatoria Y tiene el mismo dominio que X , pero el rango de Y es el conjunto de enteros entre -5 y $+5$, como se ilustra en la siguiente figura que muestra las particiones según X (horizontal y discontinuo) e Y (vertical y punteada). Y también divide el espacio muestral en 11 subconjuntos, pero no los mismos 11 que los obtenidos con X .



Sin embargo, una tercera variable aleatoria Z , puede definirse como el valor absoluto de la diferencia entre los puntos obtenidos en los dos dados. De nuevo, Z tiene el mismo dominio que X e Y , pero ahora su rango es el conjunto de enteros entre 0 y 5.

Esta variable aleatoria divide el espacio muestral en seis subconjuntos: el primer subconjunto que contiene elementos del espacio muestral en el que el número de puntos difiere en 5, independientemente de cuál sea el primero y cuál es el segundo; el segundo subconjunto de la partición contiene elementos en los que el número de puntos difiere en 4 y así sucesivamente. Esta partición se muestra en la siguiente figura:

Esta es una propiedad importante de las variables aleatorias, que dividen el espacio muestral en representaciones más significativas del experimento de probabilidad. Si nuestro interés es sólo con la suma obtenida al tirar dos dados, tiene más sentido trabajar con los 11 eventos obtenidos por la partición generada por la variable aleatoria X , que los 36 eventos elementales del espacio muestral. De esta manera, se utilizan las variables aleatorias como medio de simplificación.



Ejemplo 1.3 Sea R la variable aleatoria que cuenta el número de caras obtenidas al lanzar tres monedas. Entonces R asigna el valor 3 al resultado HHH ; el valor 2 para cada uno de los resultados HHT, HTH y THH ; el valor 1 a cada uno de los resultados HTT, THT y TTH .

Finalmente, R asigna el valor 0 al resultado TTT . Si nuestro interés es únicamente el número de caras obtenidas, entonces es más fácil trabajar con R que con los ocho resultados del experimento de probabilidad.

Ejemplo 1.4 Consideremos un experimento aleatorio de lanzar un dardo en un tablero circular, sin salirse del tablero. Asumiendo que el radio del tablero es 1, el espacio muestral es el conjunto de pares ordenados dentro del círculo unitario

$$\Omega = \left\{ (x, y) : x, y \in \mathbb{R}, \sqrt{x^2 + y^2} < 1 \right\}.$$

y el evento que describe un golpe en el tablero es dado por:

$$E = \left\{ (x, y) : x, y \in \mathbb{R}, \sqrt{x^2 + y^2} < 0.1 \right\} \subset \Omega.$$

La variable aleatoria R definida como $R(x, y) = \sqrt{x^2 + y^2}$ representa la distancia entre la posición del dardo y el origen. Como $\mathbb{P}(R = r) = 0$, desde que $\{R = r\} = \emptyset$, para todo r . También tenemos que:

$$\begin{aligned} \mathbb{P}(R < r) &= \mathbb{P} \left(\left\{ (x, y) : x, y \in \mathbb{R}, \sqrt{x^2 + y^2} < r \right\} \right) = \frac{\pi r^2}{\pi 1^2} = r^2 \\ \mathbb{P}(0.2 < R < 0.5) &= \mathbb{P} \left(\left\{ (x, y) : x, y \in \mathbb{R}, 0.2 < \sqrt{x^2 + y^2} < 0.5 \right\} \right) \\ &= \frac{\pi(0.5^2 - 0.2^2)}{\pi 1^2}. \end{aligned}$$

Puesto que una variable aleatoria es una función, no es posible para dos o más valores en el rango ser asignados a la misma salida en el espacio muestral. Para una variable aleatoria X , el conjunto de elementos del espacio muestral al que se asigna el valor x es denotado por A_x o $X^{-1}(x)$.

Observa que este describe un evento, ya que consiste en un subconjunto del espacio muestral. Tenemos

$$A_x \equiv [X = x] \equiv \{\omega \in \Omega | X(\omega) = x\}$$

el cuál es el conjunto de todas las salidas (resultados) en el espacio muestral (dominio) para el que la variable aleatoria X tiene el valor x (un elemento del rango).

En el ejemplo de los dos dados lanzados simultáneamente y la variable aleatoria X que describe la suma de puntos obtenidos, encontramos, por ejemplo, que A_3 es el conjunto $\{(1, 2), (2, 1)\}$ y A_5 es el conjunto $\{(1, 4), (2, 3), (3, 2), (4, 1)\}$.

Como se observó anteriormente, el conjunto de todos estos eventos se denomina un espacio de eventos y con frecuencia es más conveniente trabajar en este espacio de eventos que en el espacio muestral original. El espacio de eventos de nuestra variable aleatoria X define 11 eventos, cada uno correspondiente a un valor de $x = 2, 3, \dots, 12$. Para todos los demás valores de x , A_x es el conjunto nulo.

Dada una variable aleatoria, nos gustaría poder describir su comportamiento utilizando el lenguaje de las probabilidades. Por ejemplo, quizás quisiéramos responder las preguntas acerca de la probabilidad de que la variable aleatoria caiga dentro de un rango dado: si L es el ingreso vitalicio de un graduado universitario de los Estados Unidos elegido al azar, ¿cuál es la probabilidad de que supere el millón de dólares?. Si M es el número de terremotos mayores en Arequipa en los próximos cinco años, ¿cuál es la probabilidad de que M sea igual a 0?.

La **distribución** de una variable aleatoria proporciona las respuestas a estas preguntas; especifica las probabilidades de todos los eventos asociados con la variable aleatoria, tal como la probabilidad de que sea igual a 3 y la probabilidad de que sea al menos 110. Hay varias maneras equivalentes de expresar la distribución de una variable aleatoria. Para una variable aleatoria discreta la manera más natural de hacer esto es con la función de masa de probabilidad, que ahora definimos:

1.2 Función de masa de probabilidad para una variable aleatoria discreta

La función de masa de probabilidad (frecuentemente conocida como **PMF**) para una variable aleatoria discreta X , da la probabilidad de que el valor obtenido por X en el resultado de un experimento de probabilidad sea igual a x . Se denota por $p_X(x)$. A veces se utiliza el término función de densidad discreta en lugar de función de masa de probabilidad. Tenemos entonces:

$$p_X(x) \equiv \mathbb{P}([X = x]) = \mathbb{P}(\omega \in \Omega | X(\omega) = x) = \sum_{X(\omega)=x} \mathbb{P}(\omega)$$

La notación $[X = x]$ define un evento: el evento que contiene todos los resultados que corresponden con el número real x . Si ninguno de los resultados corresponden a x , entonces $[X = x]$ es el evento nulo y tiene probabilidad cero. La mayoría de las veces simplemente escribimos $\mathbb{P}([X = x])$ como $\mathbb{P}(X = x)$. Dado que $p_X(x)$ es una probabilidad, se tiene lo siguiente:

$$0 \leq p_X(x) \leq 1, \text{ para todo } x \in \mathbb{R}.$$

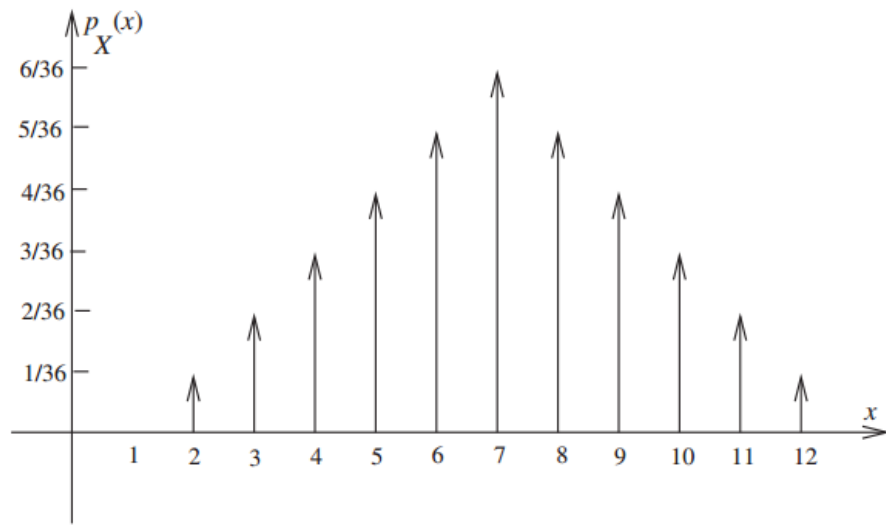
Ejemplo 1.5 En el caso de dos dados lanzados simultáneamente, y X la variable aleatoria que representa la suma de puntos obtenidos, tenemos $p_X(2) = 1/36$, ya que solo uno de los 36 resultados equiprobables en el espacio muestral da la suma 2 (cuando cada dado lanzado produce un solo punto). Del mismo modo, $p_X(3) = 2/36$ ya que hay dos resultados que dan un número total de puntos iguales a 3, a saber, $(1, 2)$ y $(2, 1)$. Podemos continuar de esta manera. Mientras x sea igual a uno de los enteros en el intervalo $[2, 12]$, entonces $p_X(x)$ es estrictamente positivo. Para todos los demás valores de x , $p_X(x) = 0$.

Las variables aleatorias discretas se caracterizan completamente por sus funciones de masa de probabilidad. Estas funciones y por lo tanto sus variables aleatorias discretas asociadas, se ilustran frecuentemente en forma tabular o en forma de gráficos de barras. En un gráfico de barras, la función de masa de probabilidad de la variable aleatoria X (suma de puntos obtenidos al tirar dos dados) se muestra en la siguiente figura:

La forma tabular correspondiente de este gráfico de barras se muestra a continuación:

Se debe notar que

$$\sum_{x \in \mathbb{R}} p_X(x) = 1 \tag{1}$$



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

ya que la variable aleatoria X asigna algún valor $x \in \mathbb{R}$ a cada punto de muestra $\omega \in \Omega$ manera que cuando sumamos sobre cada valor posible de x , sumamos sobre las probabilidades de todos los resultados en el espacio muestral y esto debe ser igual a 1.

Cuando (como es el caso aquí) la variable aleatoria es discreta, entonces el conjunto de valores asumidos por x es enumerable, y constituye un subconjunto de los reales. Denotaremos sucesivos elementos de este conjunto por $\{x_1, x_2, \dots\}$. El conjunto en si se llama imagen de X y es usual denotar a $\mathbb{P}(X = x_i)$ por p_i . De todo esto se tiene que

$$\sum_k p_X(x_k) = \sum_k p_k = 1. \quad (2)$$

Se puede demostrar que cualquier función de valor real definida en \mathbb{R} que satisface las ecuaciones (1) y (2) es la función de masa de probabilidad para alguna variable aleatoria X .

Ejemplo 1.6 Sea el experimento de lanzar una moneda tres veces. El espacio muestral está dado por:

$\{ CCC, CCS, CSC, CSS, SCC, SCS, SSC, SSS \}$

y cada uno de los elementos tiene una probabilidad de $1/8 = 0.125$. Sea Q la variable aleatoria definida como el número de caras obtenida durante esta secuencia de lanzamientos, esta variable aleatoria sólo tiene los valores 1, 2 y 3. Uno de los ocho eventos elementales da cero caras, así que la probabilidad que $Q = 0$ es $1/8$, es decir, $p_Q(0) = \mathbb{P}(Q = 0) = 0.125$. Tres de los eventos elementales produce, una cara $\{CSS, SCS, SSC\}$ y así $\mathbb{P}(Q = 1) = 0.375$ y así sucesivamente, como lo muestra la siguiente figura:

x_i	0	1	2	3
$p_R(x_i)$	0.125	0.375	0.375	0.125

1.3 Función de distribución acumulativa

Para cualquier variable aleatoria discreta X , hay un número numerable de reales que se asignan a X . Estos reales son denotados por el conjunto $\{x_1, x_2, \dots\}$ y es natural disponer estos reales en orden creciente en magnitud. Por ejemplo, se dispusieron de 2 a 12 para la variable aleatoria X y de 0 a 3 para la variable aleatoria Q respectivamente como se muestran en las tablas.

Dada esta situación, podemos ahora desear encontrar la probabilidad de que una variable aleatoria X asuma un valor que sea menor o igual que un x_i dado, es decir, la probabilidad $\mathbb{P}(X \leq x_i)$. Observa que esto corresponde a la probabilidad de un evento, el evento que consiste de todos los elementos ω del espacio muestral, para los cuales $X(\omega) = x_k$ con $x_k \leq x_i$.

Ejemplo 1.7 Consideremos una vez más la variable aleatoria X , definida como la suma de puntos obtenidos al tirar simultáneamente dos dados. Hemos examinado previamente la función de masa de probabilidad para X . Tal vez ahora nos interesa saber la probabilidad de que la variable aleatoria X tenga un valor menor que 6. Denotemos a este evento como A y observamos que su probabilidad es sólo la suma de las probabilidades:

$$\mathbb{P}(A) = \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \mathbb{P}(X = 4) + \mathbb{P}(X = 5),$$

que se puede denotar como $\mathbb{P}(X < 6)$, con lo cual se tiene:

$$\mathbb{P}(X < 6) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} = \frac{10}{36}.$$

El evento A es definido como el subconjunto consistiendo de 10 salidas

$$A = \{(1, 1), (1, 2), (2, 1), (1, 3), (2, 2), (3, 1), (1, 4), (2, 3), (3, 2), (4, 1)\}.$$

Supongamos que queremos saber la probabilidad de que la variable aleatoria X tenga un valor que esté entre 5 y 7 inclusive. En este caso, podemos escribir

$$\mathbb{P}(A) = \mathbb{P}(X = 5) + \mathbb{P}(X = 6) + \mathbb{P}(X = 7),$$

que de manera abreviada se escribe como $\mathbb{P}(5 \leq X \leq 7)$ y se calcula como sigue:

$$\mathbb{P}(5 \leq X \leq 7) = \frac{4}{36} + \frac{5}{36} + \frac{6}{36} = \frac{15}{36}.$$

Consideremos ahora un subconjunto S que contiene una colección arbitraria de valores que la variable aleatoria X puede asumir. Entonces, el conjunto $\{\omega | X(\omega) \in S\}$ contiene todos los resultados que la variable aleatoria X tiene un valor en S . Por definición, este constituye un evento que está denotado por $[X \in S]$ y por tanto tenemos:

$$[X \in S] = \{\omega | X(\omega) \in S\} = \bigcup_{x_k \in S} \{\omega | X(\omega) = x_k\}$$

Si $p_X(x)$ es la función masa de probabilidad para variables aleatorias discretas X , entonces

$$\mathbb{P}(X \in S) = \sum_{x_k \in S} p_k(x_k).$$

Ejemplo 1.8 Volvamos al ejemplo de dos dados y la variable aleatoria X que describe la suma de puntos obtenidos y sea S el conjunto que contiene solamente los dos valores $x_1 = 2$ y $x_{10} = 11$, es decir, $S = \{2, 11\}$. Entonces $\{\omega | X(\omega) = x_1\} = (1, 1)$ y $\{\omega | X(\omega) = x_{10}\} = \{(5, 6), (6, 5)\}$. Por lo tanto $[X \in S] = \{(1, 1), (5, 6), (6, 5)\}$ y

$$\mathbb{P}(X \in S) = p_X(2) + p_X(11) = 1/36 + 2/36 = 1/12.$$

Cuando el conjunto S contiene todos los valores posibles dentro de un rango de valores que X que podamos asumir, como $(a, b]$, entonces es más usual escribir $\mathbb{P}(X \in S)$ como $\mathbb{P}(a < X \leq b)$.

En el caso particular en que S contiene todos los valores que son menores o iguales a algún valor especificado x , entonces $\mathbb{P}(X \in S)$ se denomina *función de distribución acumulativa* (CDF) de la variable aleatoria X . Esta función se denota como $F_X(x)$ y también se conoce como la *función de distribución de probabilidad* (PDF).

Esto se define para valores reales de x en el rango $(-\infty < x < \infty)$, así tenemos:

$$F_X(x) \equiv \mathbb{P}(-\infty < X \leq x) = \mathbb{P}(X \leq x).$$

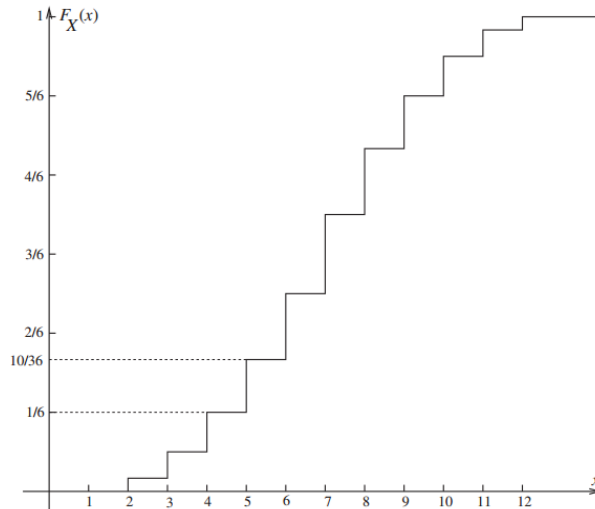
Observa que si X es una variable aleatoria discreta, entonces x no tiene que ser uno de los valores que X puede asumir, esto es, x no tiene por que ser uno de los x_i . Por ejemplo, cuando los x_i son enteros,

$$F_X(x) = \sum_{-\infty < x_i \leq \lfloor x \rfloor} p_X(x_i)$$

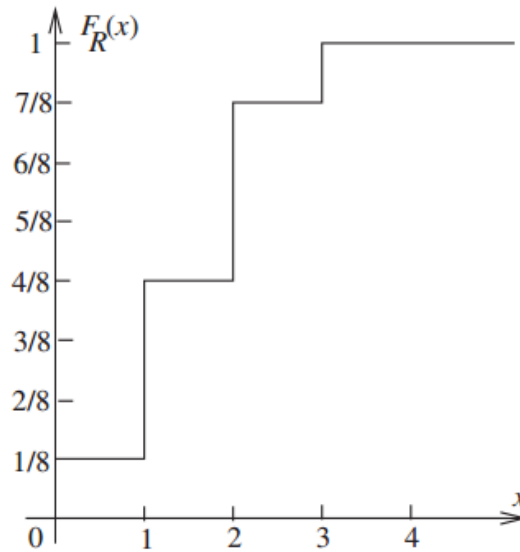
donde $\lfloor x \rfloor$ es la función que denota el mayor entero menor o igual a x .

Para ilustrar esto, consideremos la siguiente figura, que muestra la función de distribución acumulativa de la variable aleatoria X , definida como la suma de puntos encontrados en dos dados lanzados simultáneamente. De esta figura se puede observar que para cualquier valor de x en el intervalo $[4, 5)$, se tiene $F_X(x) = 1/6$.

Esta es la probabilidad de que el número de puntos obtenidos sea 4 o menos. Si x está en el intervalo $[5, 6)$ entonces $F_X(x) = 10/36$, la probabilidad de que el número de puntos sea 5 o menos.



Para la variable aleatoria R , que denota el número de caras obtenidas en tres lanzamientos de una moneda, la función de distribución acumulativa se muestra en la siguiente figura:



Si x está en el intervalo $[1, 2)$, entonces $F_R(x) = 1/2$, que es la probabilidad de lanzar uno o cero caras, es decir, el evento $A = \{SSS, CSS, SCS, SSC\}$.

En ambas de estas figuras, se han incluido los escalones simplemente para mostrar la característica escalonada de la función de distribución acumulativa de una variable aleatoria discreta; el valor asumido por la función de distribución acumulativa en x_i es el valor de la parte superior de la columna ascendente.

Se observa que en ambas figuras, el valor de la función de distribución acumulativa es cero para valores suficientemente pequeños de x . La función no decrece con valores crecientes de x y es igual a 1 para valores suficientemente grandes de x . Esto se debe al hecho de que la probabilidad de que la variable aleatoria asuma un valor menor que el menor x_k debe ser cero, mientras que la probabilidad de que tenga un valor menor o igual al mayor x_k deba ser 1. Estas propiedades son inherentes en todas las funciones de distribución acumulativa.

La estructura discontinua de las dos funciones mostradas es una característica distintiva de variables aleatorias discretas. Estos ejemplos ilustran el hecho de que la función de distribución de una variable aleatoria discreta no es continua. Estas funciones tienen al menos un punto en el que la aproximación de la función desde la izquierda y la derecha no se cumplen: la aproximación desde la derecha es estrictamente mayor que la aproximación desde la izquierda. En los ejemplos, estos son los puntos x_i que pueden ser asumidos por la variable aleatoria.

Para las variables aleatorias continuas, se mantiene la propiedad no decreciente monótona de la función de distribución acumulativa, pero el incremento es continuo. La función de distribución acumulativa de una variable aleatoria continua es una función continua de x para todo $-\infty < x < \infty$.

Mientras que la función de masa de probabilidad se define sólo para variables aleatorias discretas, la función de distribución acumulativa se aplica tanto a variables aleatorias discretas como continuas. La siguiente definición abarca ambos casos:

Definición 1.2 La función de distribución acumulativa F_X de una variable aleatoria X es definida como la función:

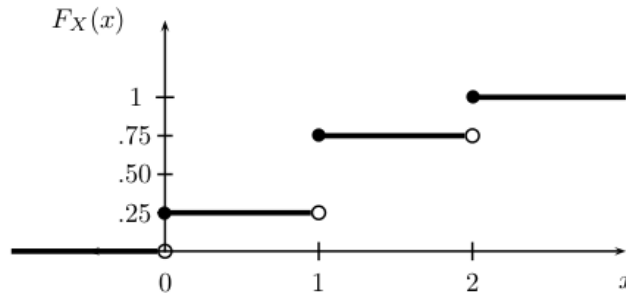
$$F_X(x) = \mathbb{P}(X \leq x) \text{ para } -\infty < x < \infty$$

En pocas palabras, es sólo la probabilidad de que la variable aleatoria X no exceda un número real dado x .

Ejemplo 1.9 Lanzamos una moneda dos veces y sea X el número de caras. Entonces $\mathbb{P}(X = 0) = \mathbb{P}(X = 2) = 1/4$ y $\mathbb{P}(X = 1) = 1/2$. La función distribución acumulativa es

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \leq x < 1 \\ 3/4 & 1 \leq x < 2 \\ 1 & x \geq 2. \end{cases}$$

y el CDF es mostrado en la siguiente figura. Se debe notar que la función es continua por la derecha, no decreciente y es definida para todo x , incluso si la variable aleatoria toma los valores 0, 1 y 2.



Ejemplo 1.10 La variable aleatoria X , cuya función de distribución tiene el valor cero, cuando $x < 0$ y el valor de uno, cuando $x \geq 1$ y que crece uniformemente desde el valor cero en $x = 0$ al valor 1 es llamada variable aleatoria continua uniforme sobre el intervalo $[0, 1]$. Esta distribución se escribe como:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x, \\ 1 & x \leq 1 \end{cases}$$

La función de distribución acumulada, no tienen derivada, excepto en $x = 0$ y $x = 1$. Las funciones de distribución de variables aleatorias deben ser absolutamente continuas, esto es, que son continuas y que sus derivadas existen salvo por un conjunto finito de puntos.

Ejemplo 1.11 Consideramos una variable aleatoria X definida como:

$$F_X(x) = \begin{cases} 1 - e^{-x} & \text{si } x \geq 0, \\ 0 & \text{en otros casos} \end{cases}$$

Una variable aleatoria teniendo esta función de distribución se dice ser una variable aleatoria exponencial con parámetro igual a 1.

1.4 Propiedades de la función de distribución

La función de distribución tiene muchas propiedades matemáticas. Entre las más importantes tenemos:

1. $0 \leq F_X(x) \leq 1$ para $-\infty < x < \infty$, desde que $F_X(x)$ es una probabilidad.
2. Se cumple que $\lim_{x \rightarrow -\infty} F_X(x) = 0$ y $\lim_{x \rightarrow \infty} F_X(x) = 1$.
3. Si la variable aleatoria X tiene una imagen finita, entonces
 - $F_X(x) = 0$ para todo x suficientemente pequeño.
 - $F_X(x) = 1$ para todo x suficientemente grande.
4. $F_X(x)$ es una función no decreciente en x . Como $(-\infty, x_1] \subset (-\infty, x_2]$ para $x_1 \leq x_2$, se sigue que $F_X(x_1) \leq F_X(x_2)$ para $x_1 \leq x_2$.

5. $F_X(x)$ es continua por la derecha. Esto significa que para algún x y una secuencia decreciente x_k con $k \geq 1$ que converga a x , se cumple que $\lim_{k \rightarrow \infty} F_X(x_k) = F_X(x)$.

Para alguna variable aleatoria X discreta o continua, tenemos:

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

Sea que $a \rightarrow b$ y observamos que ocurre. Obtenemos:

$$\mathbb{P}(X = b) = F_X(b) - F_X(b^-)$$

donde $F_X(b^-)$ es el límite de $F_X(x)$ desde la izquierda al punto b . Ahora hay dos posibilidades:

- $F_X(b^-) = F_X(b)$. En este caso $F(x)$ es continuo en el punto b y así la probabilidad $\mathbb{P}(X = b) = 0$. El evento tiene probabilidad cero, pero incluso un evento ocurre.
- $F_X(b^-) \neq F_X(b)$. Aquí F tiene una discontinuidad (un salto) en el punto b y $\mathbb{P}(X = b) > 0$.

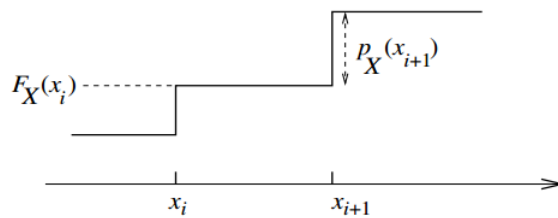
La función de distribución acumulativa de una variable aleatoria discreta crece sólo por saltos, mientras que la función de distribución acumulativa de una variable aleatoria continua no tiene saltos, pero crece continuamente. Cuando X es una variable aleatoria continua, la probabilidad de que X tenga un valor dado debe ser cero.

Todo lo que podemos hacer es asignar una probabilidad positiva de que X cae en algún intervalo finito, tal como $[a, b]$, en el eje real.

Para una variable aleatoria discreta, $F_X(x)$ tiene la apariencia de escalera. En cada uno de los puntos x_i , tiene un salto positivo igual a $p_X(x_i)$. Entre estos puntos, es decir, en el intervalo $[x_i, x_{i+1})$, tiene un valor constante. En otras palabras, tenemos lo siguiente:

$$\begin{aligned} F_X(x) &= F_X(x_i) \text{ para } x_i \leq x < x_{i+1} \\ F_X(x_{i+1}) &= F_X(x_i) + p_X(x_{i+1}). \end{aligned}$$

Como se muestra en la siguiente figura:



Finalmente, señalamos que cualquier función continua, monotóna no decreciente $\Phi(x)$ que cumpla con lo siguiente:

$$\lim_{x \rightarrow -\infty} \Phi_X(x) = 0 \text{ y } \lim_{x \rightarrow \infty} \Phi_X(x) = 1$$

puede servir como función de distribución de una variable aleatoria continua X . Como anteriormente se mencionó la función de masa de probabilidad proporciona una especificación completa de una variable aleatoria discreta.

Lo mismo puede decirse de una función de distribución acumulativa, que especifica completamente una variable aleatoria. Las dos distribuciones están estrechamente relacionadas. Dada una la otra se puede determinar fácilmente.

Ejemplo 1.12 Sea X una variable aleatoria discreta con una función distribución de probabilidad dado por:

$$F_X(x) = \begin{cases} 0 & x \in (-\infty, -2), \\ 1/10 & x \in [-2, -1), \\ 3/10, & x \in [-1, 1), \\ 6/10, & x \in [1, 2), \\ 1, & x \in [2, \infty). \end{cases}$$

Calculemos la función de masa de probabilidad de esta variable aleatoria. Para ello primero debes observar que esta variable aleatoria tiene discontinuidades en los puntos $x = -2, -1, 1$ y 2 . Estos son los valores que X puede asumir y son los únicos puntos en los que $p_X(x)$ puede tener un valor distinto de cero. Restando el valor de la función anterior a uno de estos puntos del valor de la función en el punto mismo, obtenemos la siguiente función de masa de probabilidad de X :

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

1.5 Función de densidad de probabilidad para una variable aleatoria continua

Para una variable aleatoria continua X , la función definida como:

$$f_X(x) = dF_X(x)/dx$$

Es llamada función de densidad de probabilidad de X (PDF). Esta función cumple el mismo papel para una variable aleatoria continua que el papel desempeñado por la función de masa de probabilidad de una variable aleatoria discreta. En este caso, la suma utilizada anteriormente se sustituye por la integración. Vemos este caso:

$$f_X(x) = \lim_{\Delta x \rightarrow 0} \frac{F_X(x + \Delta x) - F_X(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\mathbb{P}(x < X \leq x + \Delta x)}{\Delta x}.$$

Así, cuando x es pequeño, tenemos

$$f_X(x)\Delta x \approx \mathbb{P}(x < X \leq x + \Delta x)$$

Esto es $f_X(x)\Delta x$ es aproximadamente igual a la probabilidad de que X pertenece al intervalo, $(x, x + \Delta x]$. Si Δx tiende al infinitesimal dx , podemos escribir:

$$\mathbb{P}(x < X \leq x + dx) = f_X(x)dx.$$

Se sigue que $f_X(x) \geq 0$ para todo x .

Si conocemos la función densidad de una variable aleatoria continua, podemos obtener la función de distribución acumulativa por integración:

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(t)dt \quad \text{para } -\infty < x < \infty.$$

Desde que $F_X(\infty) = 1$ se sigue que:

$$\int_{-\infty}^{\infty} f_X(x) dx = 1.$$

A partir de estas consideraciones, podemos afirmar que una función $f_X(x)$ es la función de densidad de probabilidad para alguna variable aleatoria continua si y sólo si satisface las dos propiedades:

$$f_X(x) \geq 0 \quad \text{para todo } x \in \mathbb{R}.$$

y

$$\int_{-\infty}^{\infty} f_X(x) dx = 1.$$

La probabilidad que X se encuentra en el intervalo $(a, b]$, es obtenido desde:

$$\begin{aligned} \mathbb{P}(X \in (a, b]) &= \mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) \\ &= \int_{-\infty}^b f_X(t) dt - \int_{-\infty}^a f_X(t) dt = \int_a^b f_X(t) dt. \end{aligned}$$

Por lo tanto, la probabilidad de que la variable aleatoria se encuentre en un intervalo dado en el eje real es igual al área bajo la curva de densidad de probabilidad en este intervalo.

La relación entre el CDF y el PDF se muestra en las siguientes ecuaciones:

$$\begin{aligned} f_X(r) &= \begin{cases} F'_X(r) & F'_X(r) \text{ existe} \\ 0 & \text{en otros casos} \end{cases} \\ F_X(x) &= \int_{-\infty}^x f_X(r) dr. \end{aligned}$$

donde la primera ecuación se sigue de la definición de un PDF y la segunda línea se sigue del teorema fundamental del cálculo.

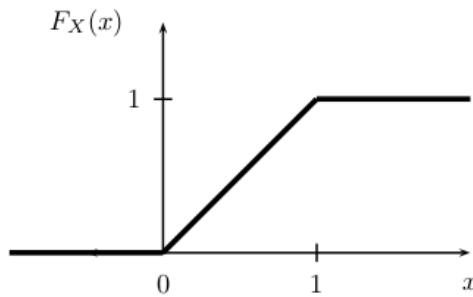
Ejemplo 1.13 Supóngase que X tiene PDF

$$f_X(x) = \begin{cases} 1 & \text{para } 0 \leq x < 1 \\ 0 & \text{en otros casos} \end{cases}$$

Claramente $f_X \geq 0$ y $\int f_X(x) dx = 1$. Una variable aleatoria con esta densidad se dice que tiene una distribución uniforme (Uniform(0,1)), lo que captura la idea, de escoger un punto entre 0 y 1. El CDF está dado por

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1. \end{cases}$$

y cuyo gráfico es



Ejemplo 1.14 Supongamos que X tiene PDF

$$f_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{(1+x)^2} & \text{en otros casos.} \end{cases}$$

Desde que $\int f_X(x)dx = 1$ esta expresión es una bien definida PDF.

Las variables aleatorias pueden traer confusión. Primero debes notar que si X es continua entonces $\mathbb{P}(X = x) = 0$, para cada x . No debes tratar de pensar que $f(x)$ es $\mathbb{P}(X = x)$. Esto sólo es válido para variables aleatorias discretas. Obtenemos probabilidades de un PDF integrando. Un PDF puede ser mayor que 1 (a diferencia de una función de masa). Por ejemplo, si $f_X(x) = 5$ para $x \in [0, 1/5]$ y 0 en otros casos,

entonces $f_X(x) \geq 0$ y $\int f_X(x)dx = 1$, por lo que este es un PDF bien definido aunque $f_X(x) = 5$ en algunos lugares.

En efecto un PDF puede ser no acotado. Por ejemplo si $f_X(x) = (2/3)x^{-1/3}$ para $0 < x < 1$ y $f(x) = 0$ en otros casos, entonces $\int f_X(x)dx = 1$ incluso si f_X no es acotada.

Ejemplo 1.15 Sea

$$f_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{(1+x)} & \text{en otros casos.} \end{cases}$$

Esto no es un PDF, desde que $\int f_X(x)dx = \log \infty = \infty$.

Sea F el CDF una variable aleatoria X . Entonces:

- $\mathbb{P}(x < X \leq y) = F_X(y) - F_X(x)$
- $\mathbb{P}(X > x) = 1 - F_X(x)$
- Si X es continua, entonces

$$\begin{aligned} F_X(b) - F_X(a) &= \mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) \\ &= \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b) \end{aligned}$$

Definición 1.3 Sea X una variable aleatoria con CDF F_X . La inversa CDF o función cuantil es definida como:

$$F_X^{-1}(q) = \inf\{x : F_X(x) \geq q\}$$

para $q \in [0, 1]$. Si F_X es estrictamente creciente y continua, entonces $F_X^{-1}(q)$ es el único número real x tal que $F_X(x) = q$.

Llamamos $F_X^{-1}(1/4)$ el primer cuartil, $F_X^{-1}(1/2)$ la mediana o segundo cuartil y $F_X^{-1}(3/4)$, el tercer cuartil.

Dos variables aleatorias X y Y son **iguales en distribución** si $F_X(x) = F_Y(y)$ para todo x . Esto significa que las aseveraciones de probabilidad acerca de X y Y deben ser las mismas, pero de ninguna manera que X e Y sean iguales. Supongamos por ejemplo $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$. Sea $Y = -X$. Entonces $\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = 1/2$ y así, X e Y son iguales en distribución, pero no son iguales. En efecto $\mathbb{P}(X = Y) = 0$.

Es tradicional escribir $X \sim F$ para indicar que X tiene una distribución F . Esto es una notación lamentable ya que el símbolo \sim también se usa para denotar una aproximación. Léase $X \sim F$ como X tiene una distribución F y no como X es aproximadamente F .

2 Ejemplos

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

2.1 Distribución de masa puntual

X tiene una distribución de masa puntual en a , $X \sim \delta_a$, si $\mathbb{P}(X = a) = 1$, en el caso que

$$F_X(x) = \begin{cases} 0 & x < a \\ 1 & x \geq a. \end{cases}$$

La función de masa de probabilidad es $p_X(x) = 1$ para $x = a$ y 0 en otros casos.

2.2 Distribución uniforme discreta

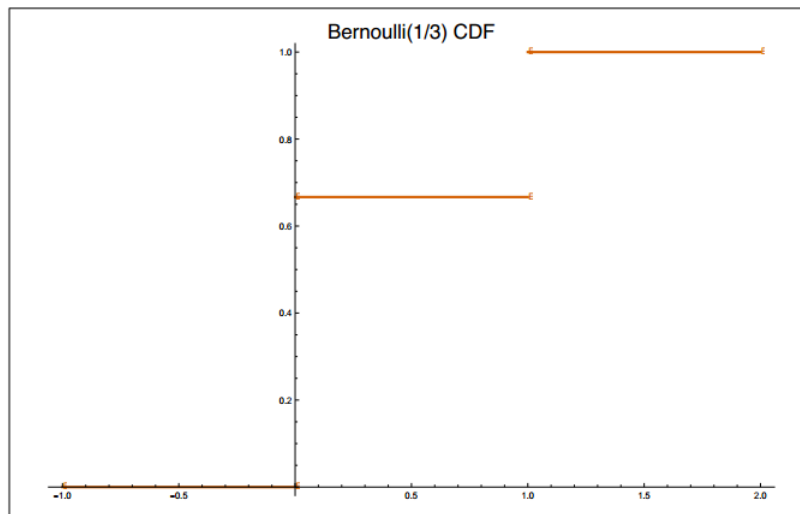
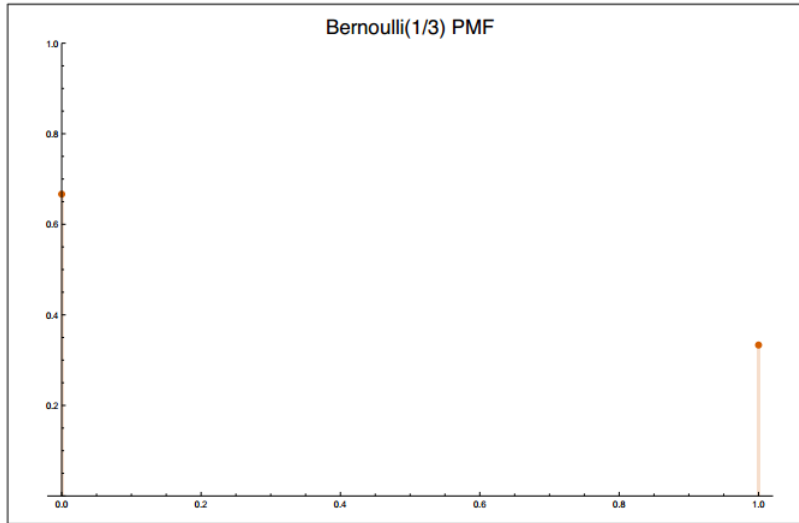
Sea $k > 1$ un número entero. Supongamos que X tiene PMF dado por

$$p_X(x) = \begin{cases} 1/k & \text{para } 1, \dots, k \\ 0 & \text{en otros casos.} \end{cases}$$

Decimos que X tiene una distribución uniforme sobre $\{1, \dots, k\}$.

2.3 Distribución de Bernoulli

Sea X que representa un lanzamiento de una moneda. Entonces $\mathbb{P}(X = 1) = p$ y $\mathbb{P}(X = 0) = 1 - p$ para algún $p \in [0, 1]$, entonces decimos que X tiene una distribución de Bernoulli, escrita como $X \sim \text{Bernoulli}(p)$. La función probabilidad es $p_X(x) = p^x(1 - p)^{1-x}$ para $x \in \{0, 1\}$.



2.4 Distribución binomial

Supongamos que tenemos una moneda, que cae cara con una probabilidad p para algún $0 \leq p \leq 1$. Lanzamos la moneda n veces y sea X el número de caras. Asumimos que estos lanzamientos son independientes. El PMF de X , $p_X(x) = \mathbb{P}(X = x)$ para esta distribución es:

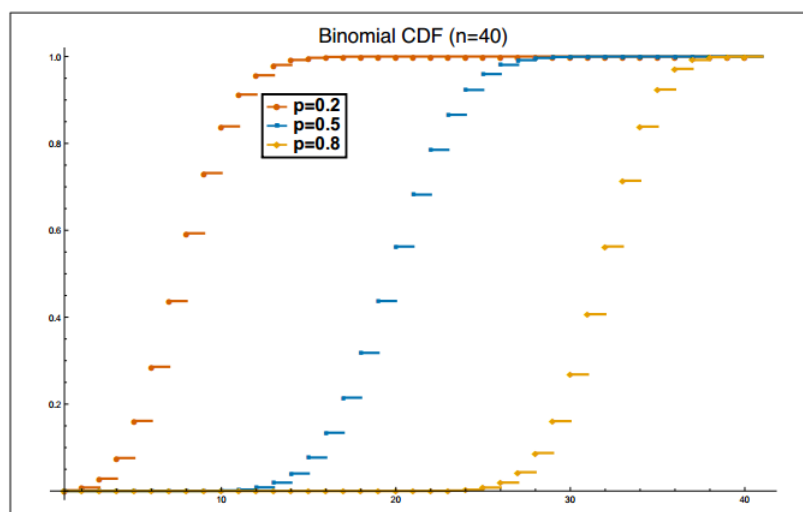
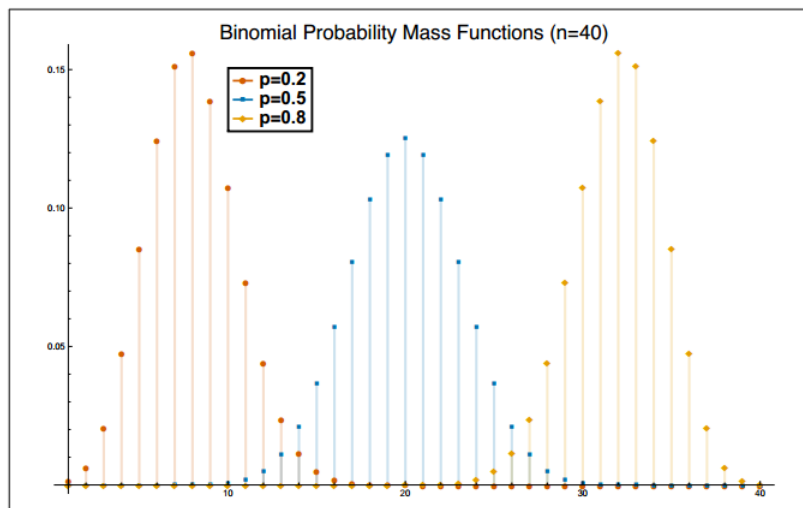
$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{para } x = 0, \dots, n \\ 0 & \text{en otros casos.} \end{cases}$$

Esta es la suma de n variables aleatorias independientes de Bernoulli. Se denota como $X \sim \text{Binomial}(n, p)$.

2.5 Distribución binomial negativa

Replicar un experimento de Bernoulli (p) de forma independiente hasta que ocurre el r -ésimo éxito ($r \geq 1$). Sea X el número de pruebas en el cual se produce el r -ésimo éxito. Entonces se dice que X tiene una distribución binomial negativa(r, p).

Existe otra definición de distribución binomial negativa, a saber, la distribución del número de fallos antes del r -ésimo éxito. (Esta es la definición empleada por R, en los cálculos).



La relación entre las dos definiciones es bastante simple. Si X es binomial negativo en el sentido usual y F es binomial negativo en el sentido de R, entonces:

$$F = X - r$$

¿Cuál es la probabilidad de que el r -ésimo éxito ocurra en la prueba t , para $t \geq r$?. Para que esto suceda, debe haber $t - r$ fracasos y $r - 1$ éxitos, en las primeras $t - 1$ pruebas, con un éxito en la prueba t .

Por independencia, esto sucede con la probabilidad binomial para $r - 1$ éxitos en las primeras $t - 1$ pruebas, por la probabilidad p de éxito en la prueba t :

$$\mathbb{P}(X = t) = \binom{t-1}{r-1} p^r (1-p)^{t-r} \quad (t \geq r).$$

de esta manera la probabilidad es cero, cuando $t < r$. El caso especial $r = 1$ es llamada distribución geométrica. Es decir, X tiene una distribución geométrica con parámetro $p \in (0,1)$, denotada como $X \sim \text{Geom}(p)$, si

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k \geq 1.$$

Tenemos a partir de este resultado que

$$\sum_{k=1}^{\infty} \mathbb{P}(X = k) = p \sum_{k=1}^{\infty} (1-p)^k = \frac{p}{1-(1-p)} = 1.$$

X es el número de lanzamientos necesarios hasta que la primera cara salga, cuando una moneda es lanzada.

2.6 Distribución multinomial

En un sentido, la distribución multinomial generaliza la distribución binomial a experimentos aleatorios independientes con más de dos resultados. Es la distribución de un vector que cuenta cuántas veces se produce cada resultado. Se denota como $X \sim \text{Multinomial}(n, \mathbf{p})$.

Un vector aleatorio multinomial es un m -vector \mathbf{X} de resultados en una secuencia de n repeticiones independientes de un experimento aleatorio con m resultados distintos.

Si el experimento tiene m resultados posibles y el i -ésimo resultado tiene probabilidad p_i , entonces la función de masa de probabilidad $\text{Multinomial}(n, \mathbf{p})$ viene dada por:

$$\mathbb{P}(\mathbf{X} = (k_1, \dots, k_m)) = \frac{n!}{k_1! k_2! \dots k_m!} p_1^{k_1} \cdot p_2^{k_2} \dots p_m^{k_m}.$$

donde $k_1 + k_2 + \dots + k_m = n$.

Si contamos el resultado k como éxito, entonces es obvio que cada X_k es simplemente una variable aleatoria binomial (n, p_k) . Pero los componentes no son independientes, ya que suman a n .

2.7 Distribución de Rademacher

La distribución $\text{Rademacher}(p)$ es una grabación de la distribución de Bernoulli, 1 todavía indica éxito, pero el fracaso se codifica como -1 . Si Y es una variable aleatoria Bernoulli (p) , entonces $X = 2Y - 1$ es una variable aleatoria $\text{Rademacher}(p)$. La función de masa de probabilidad es:

$$\mathbb{P}(X = x) = \begin{cases} p & x = 1 \\ 1 - p & x = -1 \end{cases}$$

Una secuencia de sucesivas sumas de variables aleatoria independientes de $\text{Rademacher}(p)$ es llamado **camino aleatorio**.

2.8 Distribución de Poisson

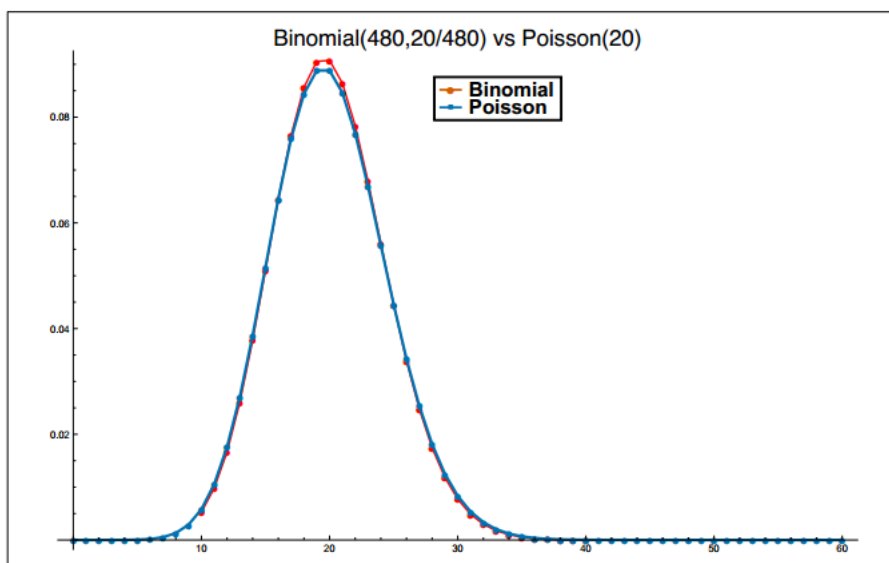
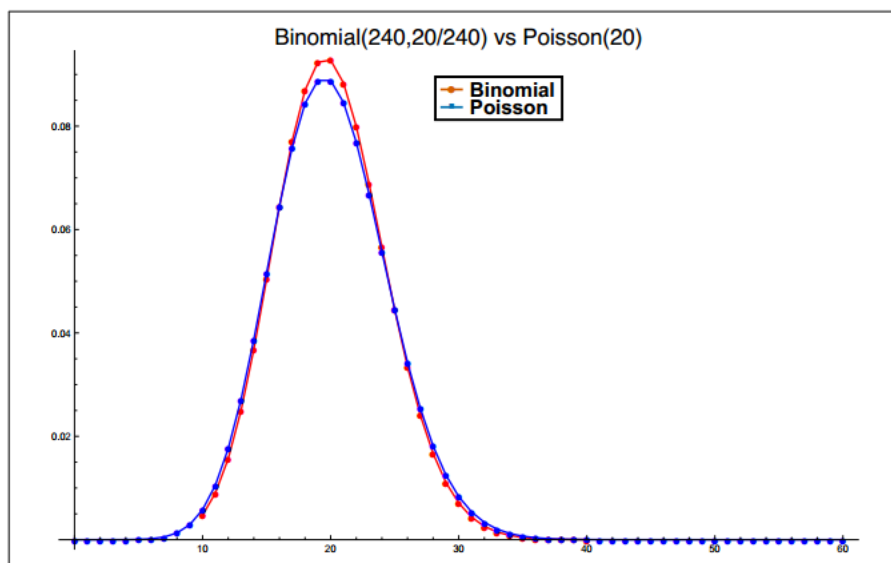
Una variable aleatoria de Poisson N modela el recuento de éxitos cuando la probabilidad de éxitos es pequeña y el número de pruebas independientes es grande, de modo que la tasa de éxito promedio μ . Se denota esta variable aleatoria como $X \sim \text{Poisson}(\mu)$ si:

$$\mathbb{P}(N = k) = e^{-\mu} \frac{\mu^k}{k!} \quad k = 0, 1, \dots$$

La distribución de Poisson es usado a menudo como un modelo para eventos, como decaimiento radiactivo y accidentes de tráfico.

Proposición 2.1 Para cada k :

$$\lim_{n \rightarrow \infty} \text{Binomial}(n, \mu/k)(k) = \text{Poisson}(\mu)(k).$$



Definimos las variables aleatorias como una aplicación de un espacio muestral Ω a \mathbb{R} pero no mencionamos el espacio muestral en cualquiera de las distribuciones anteriores. El espacio muestral a menudo "desaparece", pero este está realmente allí.

Vamos a construir un espacio muestral, de una variable aleatoria de Bernoulli. Sea $\Omega = [0, 1]$ y definamos \mathbb{P} de manera que $\mathbb{P}([a, b]) = b - a$ para $0 \leq a \leq b \leq 1$. Fijemos $p \in [0, 1]$ y definamos:

$$X(\omega) = \begin{cases} 1 & \omega \leq p \\ 0 & \omega > p \end{cases}$$

Entonces $\mathbb{P}(X = 1) = \mathbb{P}(\omega \leq p) = \mathbb{P}([0, p]) = p$ y $\mathbb{P}(X = 0) = 1 - p$. Así $X \sim \text{Bernoulli}(p)$.