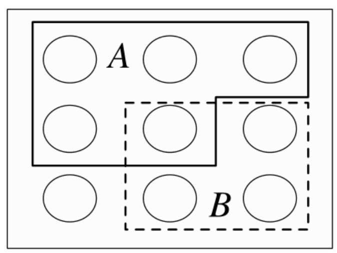
Curso: Inteligencia Artificial CM-421
Repaso

1 Probabilidad

La probabilidad es el lenguaje para cuantificar la incertidumbre.

1.1 Espacio muestral y eventos

El espacio muestral Ω de un experimento es el conjunto de todos los resultados o salidas del experimento. Puntos ω en Ω son llamados elementos, realizaciones o resultados muestrales. Subconjuntos de Ω son llamados **eventos**. El espacio muestral de un experimento puede ser finito, infinito contable o infinito no contable. En particular el conjunto vacio $\emptyset$ y el espacio muestra Ω son eventos. Cuando el espacio muestral es finito, se puede visualizar como la figura siguiente. Cada óvalo representa una salida y un evento es un conjunto de óvalos A o B.



Ejemplo 1.1 Si lanzamos una moneda dos veces, entonces $\Omega = \{SS, SC, CS, CC\}$. El evento que la primera cara es obtenida en los lanzamientos es $A = \{CS, CC\}$.

Ejemplo 1.2 Sea ω el resultado de la medición de alguna cantidad física, por ejemplo, la temperatura, entonces $\Omega = \mathbf{R} = (-\infty, \infty)$. El evento en que la medida es mayor que 10 pero menos o igual a 23 es $A = (10, 23]$.

Ejemplo 1.3 Si lanzamos una moneda por siempre, entonces el espacio muestral es infinito

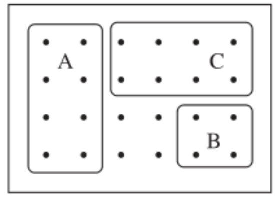
$$\Omega = \{\omega = (\omega_1, \omega_2, \dots,) : \omega_i \in \{C, S\}\}$$

Sea A el evento en que la primera cara aparece en el tercer lanzamiento. Entonces

$$A = \{(\omega_1, \omega_2, \dots,) : \omega_1 = S, \omega_2 = S, \omega_3 = S, \omega_i \in \{C, S\} \text{ para } i > 3\}.$$

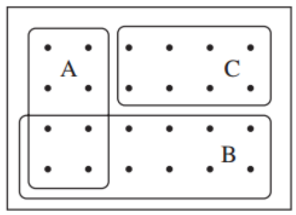
Ejemplo 1.4 Sea el evento A `lanzamiento de un número mayor que 3` y sea el evento B `lanzamiento de un número impar`. El evento A ocurre si un 4, 5 o 6 es lanzado y el evento B ocurre si 1, 3 o 5 es lanzado. Así, ambos eventos ocurren si se lanza un 5 (este es el evento que es la intersección de los eventos A y B) y ninguno de los eventos ocurre si se lanza un 2 (este es el evento que es el complemento de la unión de A y B). En este ejemplo se tiene

$$A^c = \{1, 2, 3\}; \quad A \cup B = \{1, 2, 3, 4, 5, 6\}; \quad A \cap B = \{5\}; \quad A - B = \{4, 6\}$$

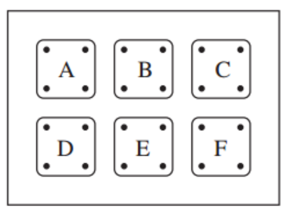


Definición 1.1 Una lista de eventos se dice mutualmente exclusivo si ningún elemento del espacio muestral pertenece a más de un evento. Como lo muestra la siguiente figura

Cuando todos los elementos en un espacio muestral pueden ser encontrados en al menos un evento en una lista de eventos, entonces la lista de eventos se dice colectivamente exhaustivos. En este caso ningún elemento del espacio muestral, es omitido y un único elemento puede pertenecer a más de un evento. Como se muestra en la siguiente figura



Se dice que los eventos que son mutuamente exclusivos y colectivamente exhaustivos, forman una partición del espacio muestral. Como se muestra en la figura.



Cualquier conjunto de eventos mutuamente excluyentes y colectivamente exhaustivos se denomina espacio de eventos.

Ejemplo 1.5 Secuencias de bits se transmiten a través de un canal de comunicación en grupos de cinco. Cada bit puede ser recibido correctamente o puede ser modificado en tránsito, lo que ocasiona un error. Consideremos un experimento que consiste en observar los valores de bits a medida que llegan e identificarlos con la letra c si el bit es correcto y con la letra e si el bit está en error. El espacio muestral consta de 32 resultados de $ccccc$ a $eeeee$, de cero bits transmitidos incorrectamente a los cinco bits que están en error.

Sea el evento $A_i, i = 0, 1, \ldots, 5$, consistente de todos los resultados en los cuales los i bits están en error. Así $A_0 = \{ccccc\}, A_1 = \{ecccc, ceccc, ccecc, cccec, cccce\}$ y así sucesivamente hasta $A_5 = \{eeeee\}$. Los sucesos $A_i, i = 0, 1, \ldots 5$, particionan el espacio muestral y por lo tanto constituya un espacio de eventos.

Puede ser mucho más fácil trabajar en este pequeño espacio de eventos que en el espacio muestral, especialmente si nuestro único interés es saber el número de bits transmitidos por error. Además, cuando los bits se transmiten en grupos más grandes, la diferencia es aun más importante. Con 16 bits por grupo en lugar de cinco, el espacio de eventos contiene ahora 17 eventos, mientras que el espacio muestral contiene 216 resultados.

1.2 Función de probabilidad

Asignamos un número $\mathbb{P}(A)$ para cada evento A, llamada probabilidad de A. También llamamos a $\mathbb{P}$ distribución de probabilidad o medida de probabilidad. Las probabilidades son números reales en el intervalo cerrado $[0, 1]$. Cuanto mayor sea el valor de la probabilidad, más probable es que ocurra el evento. Si un evento tiene probabilidad cero, ese evento no puede ocurrir; si tiene una probabilidad, entonces es seguro que ocurrirá.

Definición 1.2 Sea un espacio muestral Ω asociado a un experimento aleatorio. Una función probabilidad $\mathbb{P}$ es una función que asigna un número real a cada evento $A \subset \Omega$ y que satisface los siguientes tres axiomas

1. $\mathbb{P}(A) \geq 0, \quad \text{para todo} \quad A.$
2. $\mathbb{P}(\Omega) = 1.$
3. Sean $A_n, n \in \mathbb{N}$ es una secuencia de eventos disjuntos por pares $(A_i \cap A_j = \emptyset)$ siempre que $i \neq j$, entonces

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Podemos derivar algunas propiedades de $\mathbb{P}$ desde los anteriores axiomas.

1. $\mathbb{P}(\emptyset) = 0.$
2. $A \subset B \to \mathbb{P}(A) \leq \mathbb{P}(B).$
3. $0 \leq \mathbb{P}(A) \leq 1.$
4. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$
5. $A \cap B = \emptyset \to P(A \cup B) = P(A) + P(B).$

Ejemplo 1.6 En el experimento de lanzamiento de una moneda, la probabilidad de obtener una cara en un solo lanzamiento es 0.5, ya que tenemos la misma probabilidad de obtener una cara como lo es para obtener un sello. Esto se escribe como

$$\mathbb{P}(C) = 0.5. \text{ o } \mathbb{P}(A_1) = 0.5$$

donde A_1 es el evento $\{C\}$.

Del mismo modo, la probabilidad de lanzar un 6 con un dado es $1/6$ y la probabilidad de elegir la reina de corazones es $1/52$. En estos casos, los eventos de cada espacio muestral tienen la misma probabilidad de ser el resultado en cualquier experimento dado. Estos eventos se llaman equiprobables y se dice que el resultado del experimento es aleatorio, ya que cada evento tiene la misma probabilidad de ocurrir.

En un espacio muestral que contiene n resultados igualmente probables, la probabilidad de que se produzca cualquier resultado particular es $1/n$. Naturalmente podemos asignar probabilidades a eventos distintos de eventos elementales.

Ejemplo 1.7 Encuentre la probabilidad que debería estar asociada con el evento $A_2 = \{1, 2, 3\}$, es decir, lanzando un número menor que 4 usando un dado. Este evento ocurre si cualquiera de los números 1, 2 ó 3 es el resultado del lanzamiento. Puesto que cada uno tiene una probabilidad de $1/6$ y hay tres de ellos, la probabilidad del evento A_2 es la suma de las probabilidades de estos tres eventos elementales y por lo tanto es igual a 0.5.

1.3 Significado de la probabilidad

Asignar probabilidades a los eventos es una parte importante del desarrollo de modelos de probabilidad. En algunos casos, sabemos de antemano las probabilidades de asociarse con eventos elementales, mientras que en otros casos deben ser estimadas. Si asumimos que una moneda y un dado son imparciales y la baraja de cartas completamente mezclada, entonces es fácil asociar las probabilidades con los elementos del espacio muestral y posteriormente con los eventos descritos en estos espacios muestrales. En otros casos, las probabilidades deben ser estimadas.

Se han desarrollado dos enfoques para definir las probabilidades: el enfoque de frecuencia relativa y el enfoque axiomático. El primero, como su nombre lo indica, consiste en realizar el experimento de probabilidad muchas veces, digamos N y contar el número de veces que ocurre un cierto evento, digamos n. Entonces se puede obtener una estimación de la probabilidad del evento como la frecuencia relativa n/N con la que se produce el evento, ya que se espera que, en el límite (límite en un sentido probabilístico) cuando $N \to \infty$, la relación n/N tiende a la probabilidad correcta del evento.

En términos matemáticos, esto se afirma de la siguiente manera: Dado que la probabilidad de un evento es p, entonces:

$$\lim_{N\to\infty} \mathbb{P}\left\{ \left| \frac{n}{N} - p \right| > \epsilon \right\} = 0$$

Para cualquier $\epsilon > 0$. En otras palabras, no importa lo pequeño que elegimos ϵ, la probabilidad de que la diferencia entre n/N y p sea mayor que ϵ tiende a cero cuando $N \to \infty$. El uso de frecuencias relativas como estimaciones de probabilidad puede justificarse matemáticamente.

El enfoque axiomático establece un pequeño número de leyes o axiomas sobre los cuales se basa toda la teoria de la probabilidad. Fundamental a este concepto es el hecho de que es posible manipular probabilidades usando la misma álgebra lógica con el cual los eventos mismos son manipulados, como se muestra en los axiomas mostrados anteriormente.

1.4 Propiedades importantes de la probabilidad

Proposición 1.1 (Aditividad finita de la probabilidad) Para cada secuencia $A_1, \ldots A_n$ de eventos disjuntos dos a dos (por pares) $(A_i \cap A_j) = \emptyset$ siempre que $i \neq j$), entonces

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \cdots \mathbb{P}(A_n).$$

Proposición 1.2 (Principio de inclusión-exclusión) Para dos eventos A y B,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

En general el principio de inclusión-exclusión se puede escribir como:

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i}^{n} A_i\right) &= \sum_i \mathbb{P}(A_i) \\
&\quad - \sum_{i<j} \mathbb{P}(A_i \cap A_j) \\
&\quad + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) \\
&\quad \vdots \\
&\quad + (-1)^n \mathbb{P}(A_i \cap A_2 \ldots A_n)
\end{aligned}$$

El signo que precede a una suma con la intersección de m conjuntos es $(-1)^{m+1}$. La razón para sumar sobre índices crecientes es evitar el doble conteo.)

Se debe notar que si los conjuntos son disjuntos dos a dos, las intersecciones anteriores son todas vacías y por lo tanto tienen probabilidad cero y esto se reduce a la aditividad finita.

Proposición 1.3 (Continuidad de la probabilidad) Si $A_n \to A$ entonces

$$\mathbb{P}(A_n) \to \mathbb{P}(A) \qquad \text{si } n \to \infty$$

Ejemplo 1.8 Dos lanzamiento de moneda. Sea H_1 el evento que sale cara en el primer lanzamiento y sea H_2 el evento que ocurra cara en el segundo lanzamiento. Si todas los lanzamientos son igual de probables, entonces $\mathbb{P}(H_1 \cup H_2) = \mathbb{P}(H_1) + P(H_2) - \mathbb{P}(H_1 \cap H_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = 3/4$.

Ejemplo 1.9 Un sistema de una computadora usa passwords que son 5 caracteres y cada caracter es una de las 26 letras (a-z) o 10 enteros (0-9). El primer caracter tiene que ser una letra. Determina la proporción de passwords que

1. Inician con una consonante.
2. Terminan con un número par $(0, 2, 4, 6, 8)$.
3. Inicia con una consonante o termina con un número par.

1. Suponiendo que el sistema de la computadora usa passwords que no distinguen las mayúsculas y las minúsculas. Sea A el evento en el cuál el password empieza con una consonante. Hay 21 consonantes, entonces $\mathbb{P}(A) = 21/26$. Aquí el denominador es 26, ya que todas las letras son posibles en el passwords.
2. Sea B el evento que los passwords terminan en un número par $(0, 2, 4, 6, 8)$. Entonces $\mathbb{P}(B) = 5/36$. Aquí el denominador es 36, ya que todas las letras del alfabeto son 26 y 10 enteros son posibles en el passwords.
3. La propiedad de inclusión e exclusión, dice que la probabilidad de A o B es

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Para calcular $\mathbb{P}(A \cap B)$, la probabilidad que un password, inicia con una consonante y termina en un número par, calculamos el número posible de passwords. El primer carácter puede ser cualquiera de las 26 letras del alfabeto, y cada uno de los siguientes caracteres puede ser cualquiera de las 26 letras, así como cualquiera de los 10 números enteros, esto es 36 posibilidades en total. Así que el número posible de passwords es $(26)(36^4)$.

Ahora hay 21 maneras de empezar con una consonante y 5 maneras de terminar en un número par, así que el número posible de passwords que empiezan en una vocal y terminan en un número impar es $(21)(36^3)(5)$.
Por tanto:

$$\mathbb{P}(A \cap B) = \frac{(21)(36^3)(5)}{(26)(36^4} = \left(\frac{21}{26}\right)\left(\frac{5}{36}\right).$$

Usando el principio de inclusión e inclusión que empieza con una vocal o termina con un número par

$$\mathbb{P}(A \cup B) = \frac{21}{26} + \frac{5}{36} - \left(\frac{21}{26}\right)\left(\frac{5}{36}\right).$$

esto significa que hay un 80% de posibilidad de seleccionar un passwords iniciando con una consonante y terminando en un número par.

Definición 1.3 Para un espacio muestral finito Ω un evento conteniendo un único elemento $A = \{\omega\}, \omega \in \Omega$ es llamado un evento elemental.

Si el espacio de muestra es finito $\omega = \{\omega_1, \ldots, \omega_n\}$ es relativamente sencillo definir la función de probabilidad definiendo n probabilidades de los eventos elementales. Más específicamente, para un espacio muestral con n elementos, supongamos que se nos da un conjunto de n números no negativos $\{p_\omega : \omega \in \Omega\}$ que suman a uno, entonces existe una función de probabilidad única $\mathbb{P}$ sobre esos eventos tales que $\mathbb{P}(\{\omega\}) = p_\omega$.

Esta probabilidad se define para eventos arbitrarios a través de la propiedad de aditividad finita

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \sum_{\omega \in A} p_\omega.$$

Un similar se cumple para espacios muestrales que son finitos contables.

En la clásica interpretación de probabilidad sobre espacios muestrales finitos, las probabilidades de todos los eventos elementales $\{\omega\}, \omega \in \Omega$ son iguales. Desde que la función de probabilidad debe satisfacer $\mathbb{P}(\Omega) = 1$, tenemos

$$\mathbb{P}(\{\omega\}) = |\Omega|^{-1}, \qquad \text{para todo} \quad \omega \in \Omega.$$

Esto implica que bajo el modelo clásico sobre un finito Ω, tenemos

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

Ejemplo 1.10 Consideremos el experimento de lanzar dos dados distintos y observar dos caras en orden. El espacio muestral es

$$\Omega = \{1, \ldots, 6\} \times \{1, \ldots, 6\} = \{(x, y) : x, y \in \{1, 2, \ldots, 6\}\}$$

Desde que Ω tiene 36 elementos, la probabilidad del evento elemental $A = \{(4, 4)\}$ es $\mathbb{P}(A) = 1/|\Omega| = 1/36$. La probabilidad de conseguir una suma de 9 en ambos dados es:

$$\begin{aligned}
\mathbb{P}(\text{suma} = 9) &= \mathbb{P}(\{(6,3), (3,6), (4,5), (5,4)\}) \\
&= \frac{|\{(6,3), (3,6), (4,5), (5,4)\}|}{36} \\
&= \frac{4}{36}.
\end{aligned}$$

El modelo clásico en este caso es razonable, suponiendo que los dados son lanzados de forma independiente y son imparciales.

Definición 1.4 Para un espacio muestral continuo de dimensión n (por ejemplo $\Omega = \mathbb{R}^n$), definimos la función de probabilidad clásica como

$$\mathbb{P}(A) = \frac{\text{vol}_n(A)}{\text{vol}_n(\Omega)},$$

$\text{vol}_n(S)$ es el volumen n-dimensional del conjunto S. El volumen 1-dimensional de un conjunto $S \subset \mathbb{R}$ es su longitud. El volumen 2-dimensional de un conjunto $S \subset \mathbb{R}^2$ es su area. El volumen 3-dimensional de un conjunto $S \subset \mathbb{R}^3$ es su volumen. En general el volumen n-dimensional de A es la n-ésima integral de la función constante 1 sobre el conjunto A.

Ejemplo 1.11 En un experimento que mide el peso de los residentes en una región geográfica determinada, el espacio muestral es $\Omega = (0, 1000) \subset \mathbb{R}^1$ (suponiendo que nuestras unidades de medida son libras y las personas pesan menos de 1000 libras). La probabilidad de obtener una medida entre 150 y 250 (en el modelo clásico) es la relación de de los volúmenes o longitudes 1-dimensionales

$$\mathbb{P}((150, 250)) = \frac{|250 - 150|}{|1000 - 0|} = 0.1.$$

El modelo clásico en este caso es muy impreciso y no es probable que sea útil.

1.5 Acerca de modelos de probabilidad

- Para que el modelo clásico se aplique, el espacio de muestra Ω debe ser finito o ser continuo con un volumen finito no nulo.
- El modelo clásico (en espacios finitos y continuos) satisface los tres axiomas que definen una función de probabilidad.
- Una consecuencia del modelo clásico en espacios continuos es que la probabilidad de un evento elemental es cero (el volumen de un solo elemento es 0).

2 Probabilidad condicional

Definición 2.1 Asumiendo $P(B) > 0$, definimos la probabilidad condicional de A, dado que B a ocurrido como sigue:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Para cualquier evento B fijo tal que $P(B) > 0, \mathbb{P}(\cdot|B)$ es una función de probabilidad (es decir, cumple los tres axiomas de probabilidad). En particular $\mathbb{P}(A|B) \geq 0, \mathbb{P}(\Omega|B) = 1$ y si $A_1, A_2, \ldots$ son disjuntos, entonces $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$. Pero no es cierto en general que $\mathbb{P}(A|B \cup C) = \mathbb{P}(A|B) + P(A|C)$.

Las reglas de la probabilidad se aplican a los eventos de la izquierda de la barra $|$. En general, no se cumple que $P(A|B) = P(B|A)$.

Hay mucha confusión con esto todo el tiempo. Por ejemplo, la probabilidad de que tengas puntos (manchas) dado que tienes sarampión es 1, pero la probabilidad de que tengas sarampión, dado que tienes puntos (manchas) no es 1. En este caso, la diferencia entre $\mathbb{P}(A|B)$ y $\mathbb{P}(B|A)$ es obvia, pero hay casos en los que es menos obvio. Este error sucede con bastante frecuencia en los casos legales que a veces se llaman *falacias del fiscal*.

Si $\mathbb{P}(A) > 0$ y $\mathbb{P}(B) > 0$, tenemos:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A).$$

Intuitivamente, $\mathbb{P}(A|B)$ es la probabilidad de que A ocurra suponiendo que ocurrió el evento B. De acuerdo con esa intuición, la probabilidad condicional tiene las siguientes propiedades

- Si $B \subset A$ entonces $\mathbb{P}(A|B) = \mathbb{P}(B)/\mathbb{P}(B) = 1$.
- Si $A \cap B = \emptyset$ entonces $\mathbb{P}(A|B) = 0/\mathbb{P}(B) = 0$.
- $A \subset B$ entonces $\mathbb{P}(A|B) = \mathbb{P}(A)/\mathbb{P}(B)$.
- La probabilidad condicional, puede ser vista como una función de probabilidad

$$\mathbb{P}_A(E) = \mathbb{P}(E|A)$$

y todas las propiedades y intuiciones que se aplican a la función probabilidad se aplican a $\mathbb{P}_A$ también.

- Asumiendo que el evento A ocurrió, $\mathbb{P}_A$ generalmente tiene mejores pronósticos que $\mathbb{P}$.

Como se mencionó anteriormente, las probabilidades condicionales son usualmente intuitivas. El siguiente ejemplo de (Feller, 1968), sin embargo, muestra una situación contra-intuitiva que implica probabilidades condicionales. **Esto demuestra que la intuición no debe ser un sustituto de la computación rigurosa**.

Ejemplo 2.1 Consideramos dos familias con dos hijos donde la probabilidad de género de cada niño es simétrica $(1/2)$. Seleccionamos una familia al azar y consideramos el espacio muestral que describe el género de los niños $\Omega = \{MM, MF, FM, FF\}$. Asumimos un modelo clásico, implicando que las probabilidades de todos los eventos elementales son $1/4$.

Definimos el evento en que ambos hijos en la familia son niños como $A = \{MM\}$, el evento que una familia tiene un niño es $B = \{MF, FM, MM\}$ y el evento donde el primer hijo es un niño como $C = \{MF, MM\}$.

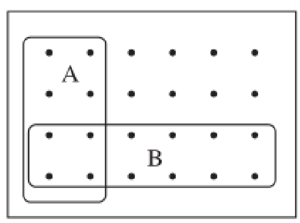
Dado que el primer hijo es un niño, la probabilidad de que ambos hijos sean niños es:

$$\mathbb{P}(A|C) = \mathbb{P}(A \cap C)/\mathbb{P}(C) = \mathbb{P}(A)/\mathbb{P}(C) = (1/4)/(1/2) = 1/2.$$

Esto coincide con nuestra intuición. Dado que la familia tiene un niño, la probabilidad de que ambos hijos sean niños es contraintuitiva:

$$\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = \mathbb{P}(A)/\mathbb{P}(B) = (1/4)/(3/4) = 1/3.$$

Ejemplo 2.2 Observamos en la siguiente figura, que $\mathbb{P}(A \cap B) = 1/6$, que $\mathbb{P}(B) = 1/2$ y que $\mathbb{P}(A|B) = (1/6)/(1/2) = 1/3$, como se esperaba. Sabemos que B ha ocurrido y que el evento A ocurrirá si uno de los cuatro resultados en $A \cap B$ se elige entre los 12 igualmente probable.



Podemos generalizar la ecuación de la probabilidad condicional a múltiples eventos. Sea $A_i, i = 1, 2, \ldots, k$ para k eventos para el cual $\mathbb{P}(A_1 \cap A_2 \cap \cdots A_k) > 0$. Entonces:

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots A_k) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1}).$$

Ejemplo 2.3 En una clase de nivel de posgrado de primer año de 60 estudiantes, diez estudiantes son estudiantes de pregrado.

Calculemos la probabilidad de que tres estudiantes elegidos al azar sean estudiantes de pregrado. Para ello sea A_1 el evento en el cual el primer estudiante elegido es un estudiante de pregrado, A_2 el evento en que el segundo estudiante sea de pregrado y así sucesivamente. Luego de la ecuación anterior tenemos

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)$$

reemplazando obtenemos:

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{10}{60} \times \frac{9}{59} \times \frac{8}{58} = 0.003507.$$

Ejemplo 2.4 Un test médico para una enfermedad D tiene salidas + y -

Salidas	D	D^c
+	.009	0.099
-	.001	0.891

Desde la definición de probabilidad condicional,

$$\mathbb{P}(+|D) = \frac{\mathbb{P}(+\cap D)}{\mathbb{P}(D)} = \frac{.009}{.009+.001} = .9$$

$$\mathbb{P}(-|D^c) = \frac{\mathbb{P}(-\cap D^c)}{\mathbb{P}(D^c)} = \frac{.891}{.891+.099} \approx .9$$

Al parecer, la prueba es bastante exacta. Las personas enfermas producen un positivo del 90 por ciento de la veces y las personas sanas producen una negativa sobre 90 por ciento de las veces . Supongamos que una persona se hace una prueba y obtiene un resultado positivo. ¿ Cuál es la probabilidad de que esa persona tiene la enfermedad? Muchas personas responden que $0,90$, sin embargo la respuesta correcta es:

$$\mathbb{P}(D|+) = \frac{\mathbb{P}(+\cap D)}{\mathbb{P}(+)} = \frac{.009}{.009+.099} \approx .08$$

3 Eventos independientes

Definición 3.1 Dos eventos A y B son independientes si $\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Para tres eventos, A, B y C definimos los eventos independientes si

$$\begin{aligned}\mathbb{P}(A\cap B) &= \mathbb{P}(A)\mathbb{P}(B)\\ \mathbb{P}(A\cap C) &= \mathbb{P}(A)\mathbb{P}(C)\\ \mathbb{P}(B\cap C) &= \mathbb{P}(A)\mathbb{P}(C) \text{ y}\\ \mathbb{P}(A\cap B\cap C) &= \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)\end{aligned}$$

Las tres primeras de estas condiciones establecen que los eventos son independientes dos a dos , por lo que llamamos a estos eventos que satisfacen estas tres condiciones como `independientes por pares`. El ejemplo siguiente mostrará que los eventos que satisfacen estas tres condiciones pueden no satisfacer la cuarta condición, de modo que la independencia por pares no determina la independencia.

Ejemplo 3.1 Una moneda es lanzada 4 veces. Consideremos los eventos A donde la primera moneda muestra cara; B la tercera moneda muestra sello; y C hay un número igual de caras y sellos. ¿ Esos son eventos independientes ?.

Supongamos que el espacio muestral consiste de 16 puntos que muestran los lanzamientos. El espacio muestral, indicando los eventos que ocurren en cada punto, es como sigue:

Entonces $\mathbb{P}(A) = 1/2$ y $\mathbb{P}(B) = 1/2$, mientras que C consiste de 6 puntos con exactamente dos caras y dos sellos. Así $P(C) = 6/16 = 3/8$. Ahora $\mathbb{P}(A\cap B) = \frac{4}{16} = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B); \mathbb{P}(A\cap C) = \frac{3}{16} = \mathbb{P}(A)\mathbb{P}(C)$ y $\mathbb{P}(B\cap C) = \frac{3}{16} = \mathbb{P}(B)\mathbb{P}(C)$. Por tanto los eventos A, B y C son independientes por pares.

El evento $A\cap B\cap C$ consiste de dos puntos $CSSC$ y $CCSS$ con probabilidad $2/16 = 1/8$. Así $\mathbb{P}(A\cap B\cap C) \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(B)$ y así A, B y C no son independientes.

Para un número finito de eventos $A_1, A_2, \ldots, A_n$ son independientes si:

$$\mathbb{P}(A_1\cap\cdots A_n) = \mathbb{P}(A_1)\cdots\mathbb{P}(A_n)$$

y son independientes por pares si cada par $A_i, A_j, i\neq j$ son independientes.

Si generalizamos un poco, dado un conjunto de eventos $\{A_i : i\in I\}$ estos son independientes si:

Puntos	Eventos
C C C C	A
C C C S	A
C C S C	A,B
S C C C	
C S C C	A
C C S S	A,B ,C
C S C S	A,C
S C C S	C
S C S C	B,C
C S S C	A,B,C
S S C C	C
S S S C	B
S S C S	
S C S S	B
C S S S	A,B
S S S S	B

$$\mathbb{P}\Big(\bigcap_{i\in J} A_i\Big) = \prod_{i\in J} \mathbb{P}(A_i)$$

para cada subconjunto J de I.

Existe cierta confusión entre eventos independientes y eventos mutuamente exclusivos. A menudo se habla de estos como `no tener ningún efecto sobre el otro`, pero eso no es una caracterización precisa en ambos casos. Debes tener en cuenta que aunque los eventos mutuamente exclusivos no pueden ocurrir juntos, los eventos independientes deben ser capaces de ocurrir juntos.

Supongamos que ni $\mathbb{P}(A)$ ni $\mathbb{P}(B)$ es 0 y que A y B son mutualmente exclusivos. Entonces $\mathbb{P}(A \cap B) = 0 \neq \mathbb{P}(A)\mathbb{P}(B)$. Así A y B no son independientes. Esto es equivalente a la declaración que si A y B no pueden ser mutualmente exclusivos.

La independencia puede surgir de dos maneras distintas. Por ejemplo al lanzar una moneda dos veces, asumimos que los lanzamientos son independientes, lo que se `traduce`, en que las monedas `no tienen memoria del primer lanzamiento`. En otro caso, derivamos la independencia verificando $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Por ejemplo, en el lanzamiento de un dado, sea $A = \{2,4,6\}$ y sea $B = \{1,2,3,4\}$, entonces $A \cap B = \{2,4\}, \mathbb{P}(A \cap B) = 2/6 = \mathbb{P}(A)\mathbb{P}(B) = (1/2)(2/3)$ y así A y B son independientes. En este caso no asumimos que A y B son independientes.

Supongamos que A y B son eventos disjuntos, cada uno con probabilidad positiva. ¿ Pueden ser independientes? No. Esto se debe a que $\mathbb{P}(A)\mathbb{P}(B) > 0$ y $\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$. Excepto en este caso especial, no hay manera de revisar la independencia mirando los conjuntos en un diagrama de Venn .

Ejemplo 3.2 Lanzamos una moneda 10 veces. Sea el evento $A =$ `al menos una cara`. Sea T_j el evento que salga sello en el j-ésimo lanzamiento. Entonces

$$\begin{aligned}
\mathbb{P}(A) =& 1 - \mathbb{P}(A^C)\\
=& 1 - \mathbb{P}(\text{todos los sellos})\\
=& 1 - \mathbb{P}(T_1 T_2 \cdots T_{10})\\
=& 1 - \mathbb{P}(T_1)\mathbb{P}(T_2)\cdots\mathbb{P}(T_{10}) \text{ usando independencia}\\
=& 1 - \Big(\frac{1}{2}\Big)^{10} \approx .999.
\end{aligned}$$

Ejemplo 3.3 Sean dos personas que se turnan para encestar una pelota de baloncesto. La persona 1 tiene éxito con probabilidad $1/3$, la persona 2 tiene éxito con probabilidad $1/4$. ¿ Cuál es la probabilidad de que la persona 1 tenga éxito antes que la persona 2?. Sea E el evento de interés. Sea A_j el evento en que el primer éxito es de la persona 1 y que se produce en el lanzamiento j. Los conjuntos A_i son disjuntos y forman una partición para E, así:

$$\mathbb{P}(E) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$$

Ahora $\mathbb{P}(A_1) = 1/3$. A_2 ocurre si tenemos la secuencia persona 1 pierde, persona 2 pierde y la persona 1 tiene éxito. Luego $\mathbb{P}(A_2) = (2/3)(3/4)(1/3) = (1/2)(1/3)$. Siguiendo esa lógica tenemos que $\mathbb{P}(A_j) = (1/2)^{j-1}(1/3)$. Así

$$\mathbb{P}(E) = \sum_{j=1}^{\infty} \frac{1}{3}\left(\frac{1}{2}\right)^{j-1} = \frac{1}{3}\sum_{j=1}^{\infty}\left(\frac{1}{2}\right)^{j-1} = \frac{2}{3}.$$

Proposición 3.1 Si A y B son eventos independientes, entonces $\mathbb{P}(A|B) = \mathbb{P}(A)$. También para algún par de eventos A y B

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$$

Ejemplo 3.4 Se lanza dos cartas de una baraja, sin reemplazo. Sea A el evento de que el primera carta es el as de tréboles y sea B el evento de que la segunda corta es la reina de diamantes. Entonces $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A) = (1/52) \times (1/51)$.

Proposición 3.2 Si A, B son independientes, entonces son independientes los eventos A^c, B, los eventos A, B^c y los eventos A^c, B^c.

4 Teorema de Bayes

4.1 Ley de la probabilidad total

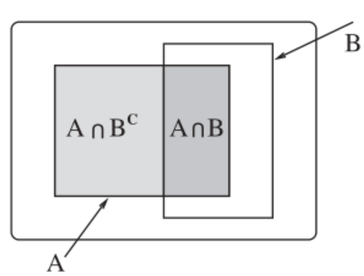
Sea A un evento, entonces se sabe que la intersección de A y el evento universal Ω es A. Se sabe además que B y su complemento B^c constituye una partición. Así:

$$A = A \cap \Omega \;\text{ y }\; B \cup B^c = \Omega$$

Sustituyendo el segundo resultado en el primero en la ecuación anterior y aplicando la Ley de Morgan, tenemos:

$$A = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c). \tag{1}$$

Los eventos $(A \cap B)$ y $(A \cap B^c)$ son mutualmente exclusivos y de acuerdo al siguiente gráfico, se muestra que B y B^c no pueden tener salidas en común, la intersección de A y B no puede tener salidas en común con la intersección de A y B^c



Usando el hecho que $(A \cap B)$ y $(A \cap B^c)$ son mutualmente exclusivos y por propiedad de la función probabilidad, se tiene:

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c).$$

Esto significa que para evaluar la probabilidad de un evento A, es suficiente encontrar las probabilidades de la intersección de A y B y A y B^c y sumarlos.

En general sean n eventos $B_i, i = 1, 2, \ldots, n$, una partición del espacio muestral Ω. Entonces para algún evento A, podemos escribir:

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i), \quad n \geq 1$$

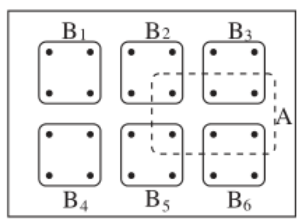
Esta es la ley de la probabilidad total. En efecto, los conjuntos $A \cap B_i, i = 1, 2, \ldots, n$ son mutualmente exclusivos (desde que los B_i lo son) y el hecho que $B_i, i = 1, 2, \ldots$ es una partición de Ω implica que:

$$A = \bigcup_{i=1}^{n} A \cap B_i, \; n \geq 1,$$

y por propiedad de la función probabilidad (tercer axioma), se tiene:

$$\mathbb{P}(A) = \mathbb{P}\left\{ \bigcup_{i=1}^{n} A \cap B_i \right\} = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i).$$

Ejemplo 4.1 Consideramos la siguiente figura que muestra una partición de un espacio muestral conteniendo 24 equiprobables salidas en seis eventos B_1 hasta B_6.



Se sigue entonces que la probabilidad del evento A es igual a $1/4$ ya que contiene seis de los puntos de muestreo. Debido a que los eventos B_i constituyen una partición, cada punto de A está en uno y sólo uno de los eventos B_i y la probabilidad del evento A se pueden encontrar sumando las probabilidades de los eventos $A \cap B_i$ para $i = 1, 2, \ldots, 6$.

Para este ejemplo particular se puede ver que estas seis probabilidades están dadas por $0, 1/24, 1/12, 0, 1/24$ y $1/12$ que cuando se suman juntas da $1/4$.

La Ley de la probabilidad total, es frecuentemente presentado en otro contexto, uno que involucra la probabilidad condicional

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

lo que significa que podemos encontrar $\mathbb{P}(A)$ al encontrar primero la probabilidad de A dado B_i para todo i y luego calcular su suma.

Ejemplo 4.2 Mañana habrá lluvia o nieve, pero no ambos, la probabilidad de lluvia es $2/5$ y la probabilidad de nieve es $3/5$. Si llueve, la probabilidad de que llegue tarde a mi conferencia es $1/5$ mientras que la probabilidad correspondiente en el caso de nieve es $3/5$. ¿Cuál es la probabilidad de que llegue tarde?

Sea A el evento en el que se llega tarde y B sea el evento que llueva. El par B y B^c es una partición del espacio muestral (ya que exactamente uno de ellos debe ocurrir). Por la ley de la probabilidad total tenemos:

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)\\ &= \frac{1}{5} \times \frac{2}{5} + \frac{3}{5} \times \frac{3}{5} = \frac{11}{25}.\end{aligned}$$

Ejemplo 4.3 Supongamos que tres cajas contienen bolas blancas y negras. La primera caja contiene 12 bolas blancas y tres negras, la segunda contiene cuatro bolas blancas y 16 negras y la tercera contiene seis bolas blancas y cuatro negras. Se selecciona una caja y se elige una sola bola. La elección de la caja se hace de acuerdo al lanzamiento de un dado. Si el número de puntos en el dado es 1, se selecciona la primera caja, si el número de puntos es 2 ó 3 se elige la segunda caja, de lo contrario (el número de puntos es igual a 4, 5 ó 6) se elige la tercera caja. Supongamos que queremos encontrar $\mathbb{P}(A)$ donde A es el evento en la que una bola blanca es extraida.

En este caso basaremos la partición en las tres cajas. Especificamente, sea $B_i, i = 1, 2, 3$ el evento en la que la caja i es escogida. Entonces $\mathbb{P}(B_1) = 1/6, \mathbb{P}(B_2) = 2/6$ y $\mathbb{P}(B_3) = 3/6$. Aplicando la ley de la probabilidad total, tenemos:

$$\mathbb{P}(A) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \mathbb{P}(A|B_3)\mathbb{P}(B_3),$$

que es fácil calcular, usando:

$$\mathbb{P}(A|B_1) = 12/15, \;\; \mathbb{P}(A|B_2) = 4/20, \;\; \mathbb{P}(A|B_3) = 6/10.$$

así:

$$\mathbb{P}(A) = 12/15 \times 1/6 + 4/20 \times 2/6 + 6/10 \times 3/6 = 1/2.$$

4.2 Regla de Bayes

Con frecuencia ocurre que se nos dice que ha ocurrido un cierto evento A y nos gustaría saber cuáles de los eventos mutuamente excluyentes y colectivamente exhaustivos B_j han ocurrido, al menos probabilísticamente. En otras palabras, nos gustaría conocer $\mathbb{P}(B_j|A)$ para cualquier j.

Obtenemos la regla de Bayes de resultados anteriores sobre la probabilidad condicional y el teorema de la probabilidad total, de la siguiente manera:

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A \cap B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

Aunque parezca que esto complica las cosas, lo que estamos haciendo es dividir el problema en piezas más simples. Esto se hace obvio en el ejemplo siguiente donde elegimos un espacio de muestra dividido en tres eventos.

Ejemplo 4.4 Considere un profesor universitario que observa a los estudiantes que entran en su oficina con preguntas. Este profesor determina que el 60% de los estudiantes son estudiantes de BSc, mientras que el 30% son estudiantes de MS, 10% son estudiantes de doctorado. El profesor también indica que puede manejar las preguntas del 80% de los estudiantes BSc en menos de cinco minutos, mientras que las preguntas de 50% de los estudiantes de MS y el 40% de los estudiantes de doctorado en cinco minutos o menos.

El próximo estudiante en ingresar a la oficina del profesor sólo necesitaba dos minutos del tiempo del profesor. ¿Cuál es la probabilidad de que el estudiante sea un estudiante de doctorado?.

Para responder, esta pregunta, sean los eventos $B_i, i = 1, 2, 3$ `el estudiante es un BCs, MS y doctorado` respectivamente y sea el evento A el evento `el estudiante requiere cinco minutos o menos`.

Desde la ley de probabilidad total, tenemos:

$$
\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \mathbb{P}(A|B_3)\mathbb{P}(B_3) \\
&= 0.8 \times 0.6 + 0.5 \times 0.3 + 0.4 \times 0.1 = 0.6700
\end{aligned}
$$

Este cálculo nos da el denominador para la inserción en la regla de Bayes. Esto nos dice que aproximadamente dos tercios de todas las preguntas de los estudiantes se pueden manejar en cinco minutos o menos.

Calculemos $\mathbb{P}(B_3|A)$ usando la regla de Bayes, tenemos

$$
\mathbb{P}(B_3|A) = \frac{\mathbb{P}(A|B_3)\mathbb{P}(B_3)}{\mathbb{P}(A)} = \frac{0.4 \times 0.1}{0.6700} = 0.0597
$$

O alrededor del 6% y es la respuesta que buscamos.

El punto crítico en la respuesta a preguntas como estas es determinar qué conjunto de eventos constituye la partición del espacio muestral. Las pistas se encuentran generalmente en la pregunta planteada.

Si recordamos que se nos pide calcular $\mathbb{P}(B_j|A)$ y relacionar esto con las palabras en la pregunta, entonces se hace evidente que las palabras `era un estudiante de doctorado` sugiere una partición basada en el estatus del estudiante. El evento A se refiere al tiempo que toma el estudiante.

La clave está en entender que se nos da $\mathbb{P}(A|B_j)$ y se nos pide encontrar $\mathbb{P}(B_j|A)$. En esta forma simple, la ley de Bayes se escribe como:

$$
\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}.
$$

Ejemplo 4.5 Un experimentador tienes dos urnas A y B a su disposición. La urna A (B) tiene ocho (diez) canicas verdes y doce(ocho) canicas azules, todas del mismo tamaño y peso. El experimentador selecciona una urna con igual probabilidad y escoge una canica al azar. Se anuncia que la canica seleccionada es azul. Encontremos la probabilidad de que esta canica provenga de la urna B.

Definamos los eventos

A_i: La urna i es seleccionada, $i = 1, 2$.
B : La canica escogida desde la urna seleccionada es azul.

Entonces $\mathbb{P}(A_i) = \dfrac{1}{2}$ para $i = 1, 2$, mientras que $\mathbb{P}(B|A_1) = \dfrac{12}{20}$ y $\mathbb{P}(B|A_2) = \dfrac{8}{18}$.

Ahora apliquemos el teorema de Bayes,

$$
\mathbb{P}(A_2|B) = \mathbb{P}(A_2)P(B|A_2)/\{\mathbb{P}(A_1)P(B|A_1) + \mathbb{P}(A_2)\mathbb{P}(B|A_2)\}
$$

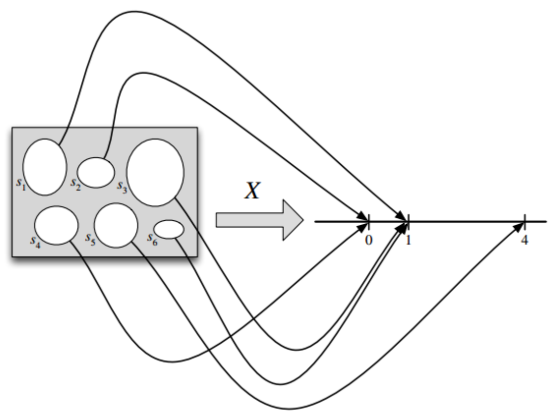
5 Variables aleatorias y funciones de distribución

5.1 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 $0, 1$ 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 5.1 Una variable aleatoria es una función $X : \Omega \to \mathbb{R}$ 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 $X \in E$.

- La 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 $X : \Omega \to \mathbb{R}$, 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 representalos.

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 5.1 Sea el experimento de probabilidad que consiste en lanzar dos dados. Hay 36 resultados posibles que pueden ser representados por los pares $(i, j); i = 1, 2, \ldots, 6, j = 1, 2, \ldots, 6$. 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

```
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.
```

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.

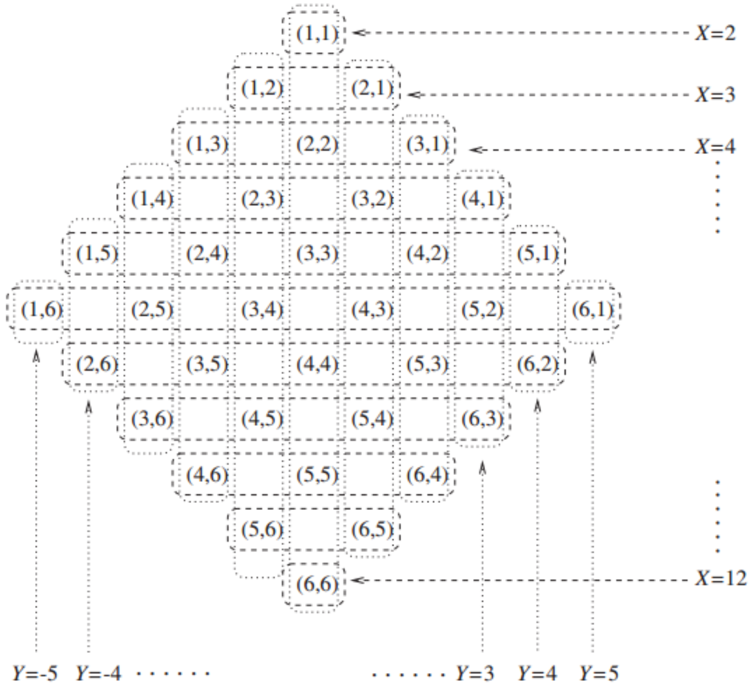
Ejemplo 5.2 Consideremos una variable aleatoria diferente Y cuyo dominio es el mismo espacio muestral que el de X , pero que se define como la diferencia en el número de puntos en los dos dados. En este caso tenemos:

```
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.
```

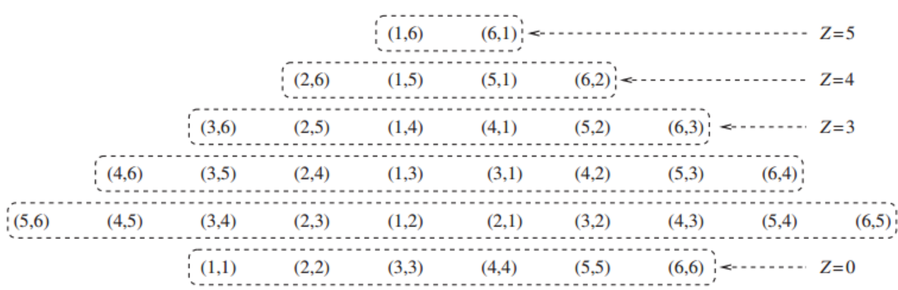
En este caso observa que la variable aleatoria Y tiene el mismo dominio que X, pero el rango de Y es el conjunto de enteros entre -5 y $+5$, como se ilustra en la siguiente figura que muestra las particiones según X (horizontal y discontinuo) e Y (vertical y punteada). Y también divide el espacio muestral 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 5.3 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 5.4 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 y el origen. Como $\mathbb{P}(R = r) = 0$, desde que $\{R = r\} = 0$, 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 se describe un evento, ya que consiste en un subconjunto del espacio muestral. Tenemos entonces:

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

el cuál es el conjunto de todos 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 $x = 2, 3, \ldots, 12$. 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 dolares?. 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 iguala 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:

5.2 Función de masa de probabilidad para una variable aleatoria discreta

La `función de masa de probabilidad` (f**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:

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

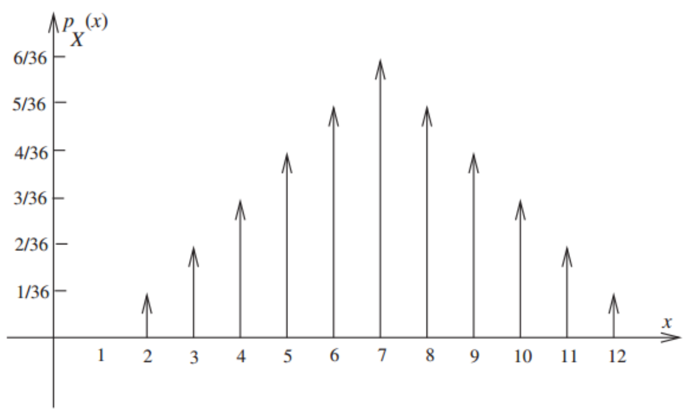
La notación $[X = x]$ define un evento: el evento que contiene todos los resultados que corresponden con el número real x. Si ninguno de los resultados corresponden a x, entonces $[X = x]$ es el evento nulo y tiene probabilidad cero. La mayoría de las veces simplemente escribimos $\mathbb{P}([X = x])$ 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 5.5 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:

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

Se debe notar que:

$$\sum_{x \in \mathbb{R}} p_X(x) = 1 \tag{2}$$

Ya que la variable aleatoria X asigna algún valor $x \in \mathbb{R}$ a cada punto de muestra $\omega \in \Omega$, de 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 numerable 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:

$$\sum_k p_X(x_k) = \sum_k p_k = 1. \tag{3}$$

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 5.6 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

5.3 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 5.7 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 5.8 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 de X que podamos asumir como $(a, b]$, entonces es 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, $\mathbb{P}(X \in S)$ 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,

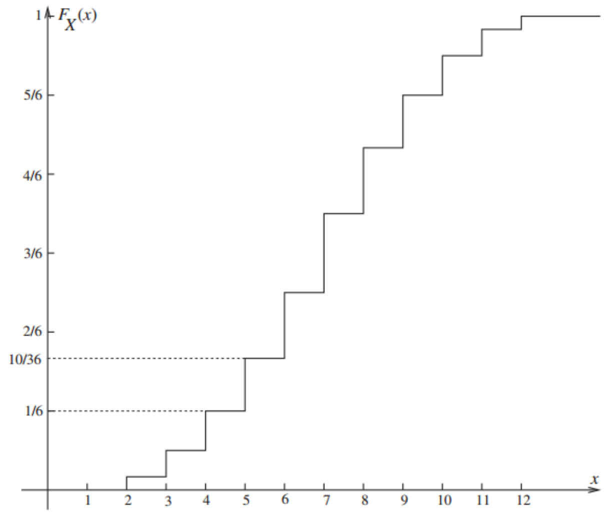
$$F_X(x) = \sum_{-\infty < x_i \leq \lfloor x \rfloor} p_X(x_i)$$

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.

En esta figura, se ha 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 en la figura que 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 para 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 la funcion mostrada es una característica distintiva de variables aleatorias discretas. Este ejemplo ilustra 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 5.2 La función de distribución acumulativa F_X de una variable aleatoria X es definida como la función:

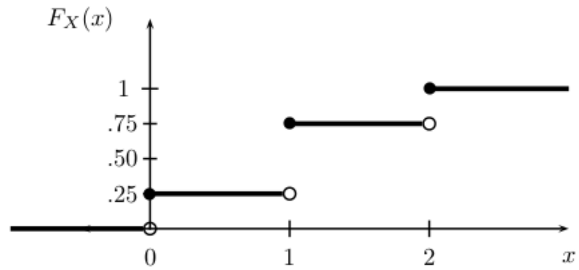
$$F_X(x) = \mathbb{P}(X \leq x) \text{ para } -\infty < x < \infty$$

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

Ejemplo 5.9 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 5.10 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:

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

La función de distribución acumulativa, tienen derivada en todas partes, 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.

5.4 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 $\lim_{x \to -\infty} F_X(x) = 0$ y $\lim_{x \to \infty} F_X(x) = 1$.
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 $\lim_{k \to \infty} F_X(x_k) = F_X(x)$.

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 $a \to b$ 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 $\mathbb{P}(X = b) > 0$.

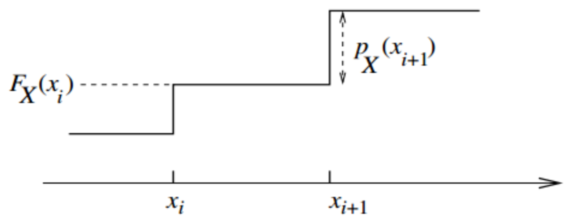
La función de distribución acumulativa de una variable aleatoria discreta crece sólo por saltos, mientras que la función de distribución acumulativa de una variable aleatoria continua no tiene saltos, pero crece continuamente. Cuando X es una variable aleatoria continua, la probabilidad de que X tenga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

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

Como se muestra en la siguiente figura:



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

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

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 5.11 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.

Para este ejemplo tenemos la siguiente función de masa de probabilidad de X:

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

5.5 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:

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

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:

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

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:

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

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:

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

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

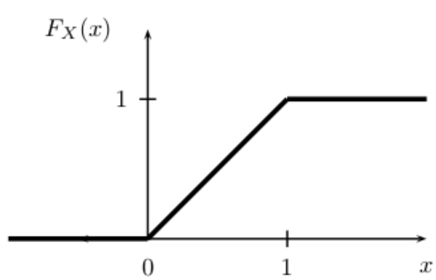
Ejemplo 5.12 Supóngase que X tiene `PDF`

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

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



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 $\displaystyle\int f_X(x)dx = 1$, 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 $\displaystyle\int f_X(x)dx = 1$ incluso si f_X no es acotada.

Ejemplo 5.13 Sea

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

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

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

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`, $F_X^{-}(1/2)$ 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`.

5.6 Varianza y covarianza

La varianza de una variable aleatoria X es definida como:

$$\mathrm{Var(X)} = \mathbb{E}([X - \mu]^2),$$

donde $\mu = \mathbb{E}(X)$. La varianza es una medida de dispersión con respecto a μ, en el sentido de que si X toma valores que difieren considerablemente de μ, entonces el valor de $|X - \mu|$ es grande y por lo tanto $\mathbb{E}([X-\mu]^2)$ tiene un valor grande, mientras que si X es cercano al valor de μ, entonces $|X - \mu|$ tiene un valor que pequeño igual que $\mathbb{E}([X-\mu]^2)$.

En un caso extremo cuando X es `concentrado` en algún punto, para una variable Y tenemos:

$$\mathbb{E}(Y^2) = 0 \text{ si y sólo si } \mathbb{P}(Y = 0) = 1$$

si aplicamos la ecuación anterior a $Y = X - \mu$, tenemos el siguiente resultado:

$$\text{Var}(X) = 0 \text{ si sólo si } \mathbb{P}(X = \mu) = 1$$

de esta forma la varianza cero significa que `no hay dispersión en lo absoluto`. Algunas propiedades de la varianza son las siguientes:

- $\text{Var}(X) = \mathbb{E}([X-\mu]^2) = \mathbb{E}(X^2) - \mu^2$ donde $\mathbb{E}(X) = \mu$.
- $\text{Var}(aX + b) = a^2\text{Var}(X)$.

La varianza como medida de dispersión tiene una propiedad no deseable: no es lineal en el sentido de que la varianza de aX es a^2 veces la varianza de X, por esta razón es preferible trabajar con la `desviación estándar` de X, definido como $\sqrt{\text{Var}(X)}$.

En términos de $\text{Var}(X)$ y $\text{Var}(Y)$, la expresión de $\text{Var}(X+Y)$ es de la forma:

$$\text{Var}(X+Y) = \text{Var}(X) + 2\mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]) + \text{Var}(Y)$$

Es conveniente tener una palabra especial para el término medio en la última expresión y para ello definimos la `covarianza` de X e Y.

Definición 5.4 La `covarianza` de las variables aleatorias X y Y es la cantidad denotada por $cov(X,Y)$ y es dada por:

$$cov(X,Y) = \mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

siempre que las esperanzas existan.

Podemos agregar algunas propiedades más, de la varianza y covarianza:

- $\text{Var}(X+Y) = \text{Var}(X) + 2cov(X,Y) + \text{Var}(Y)$
- Si X y Y son independientes : $cov(X,Y) = 0$ y $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$.

La covarianza se utiliza a menudo como una medida de la dependencia de X e Y y la razón de esto es que $cov(X,Y)$ es un sólo un número (en lugar de un objeto complicado como la función densidad conjunta) que contiene información útil sobre el comportamiento conjunto de X e Y. Por ejemplo, si $cov(X,Y) > 0$, entonces $X - \mathbb{E}(X)$ y $Y - \mathbb{E}(Y)$ pueden tener una buena chance (en algún sentido) de tener el mismo signo.

La principal desventaja de la covarianza como medida de dependencia de X e Y es que no es invariante en la escala: Si X y Y estaán medidas en pulgadas y U y V tienen las misma medida en centímetros (tal que $U = \alpha X$ y $V = \alpha Y$, donde $\alpha \sim 2.54$), entonces $cov(U,V) \sim 6cov(X,Y)$ a pesar del hecho de que el par (X,Y) y (U,V) miden la misma cantidad. Para tratar con esta `re-escala` en la covarianza se define lo siguiente:

Definición 5.5 `El coeficiente de correlación` de las variables aleatorias X y Y es la cantidad $\rho(X,Y)$ definida por:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

siempre que la última cantidad exista y $\text{Var}(X)\text{Var}(Y) \neq 0$.

Una propiedad del coeficiente de correlación que se pueden deducir de la definición es: $\rho(aX+b, cY+d) = \rho(X,Y)$, para $a,b,c,d \in \mathbb{R}$ tal que $ac \neq 0$, es decir el coeficiente de correlación es invariante a escala.

La correlación tiene una propiedad importante como medida de dependencia:

Teorema 5.1 Si X y Y son variables aleatorias, entonces $-1 \leq \rho(X,Y) \leq 1$, siempre que la correlación exista.

Un tratamiento más exahustivo distingue los valores de ± 1:

- $\rho(X,Y) = 1$ si y sólo si $P(Y = \alpha X + \beta)$ = 1 para algún real α y β con $\alpha > 0$.
- $\rho(X,Y) = -1$ si y sólo si $P(Y = \alpha X + \beta)$ = 1 para algún real α y β con $\alpha < 0$.

Se usa $\rho(X,Y)$ como una medida de las dependencias de X y Y. Si X y Y tienen varianza distintas de cero, entonces $\rho(X,Y)$ toma algún valor en el intervalo $[-1,1]$ y este valor debe ser interpretado de la forma en que los valores $-1, 0, 1$ surgen:

- Si X y Y son independientes entonces $\rho(X,Y) = 0$.
- Y es una función creciente de X si y sólo si $\rho(X,Y) = 1$.
- Y es una función decreciente de X si y sólo si $\rho(X,Y) = -1$.

Si $\rho(X,Y) = 0$ se dice que X y Y son `no correlacionados`.

6 Variables aleatorias multivariadas

6.1 Funciones de masa de probabilidad conjunta

La función de masa de probabilidad para una variable aleatoria X, es denotada por $p_X(x)$ y es la probabilidad de los eventos que consisten de todas las salidas que son mapeadas en el valor x. Si X y Y son variables aleatorias, una salida es un punto en el plano (x,y) y un evento es un subconjunto de esos puntos del plano: $[X = x, Y = y]$ es el evento que consiste de todas las salidas para el cual X es mapeada a x y Y es mapeada a y. La `función de masa de probabilidad conjunta` denotada como $p_{X,Y}$ es igual a la probabilidad del evento $[X = x, Y = y]$ y es la función $p_{X,Y} : \mathbb{R}^2 \to [0,1]$ definida por

$$p_{X,Y}(x,y) = \mathbb{P}(\omega \in \Omega : X(\omega) = x, Y(\omega) = y)$$

o de forma abreviada :

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

La función de masa de probabilidad conjunta de dos variables aleatorias discretas puede representarse convenientemente en forma tabular. Por ejemplo, podemos colocar valores realizables de X en la parte superior de una tabla y los de Y en el lado izquierdo. El elemento de la tabla correspondiente a la columna $X = x$ y la fila $Y = y$ denota la probabilidad del evento $[X = x, Y = y]$.

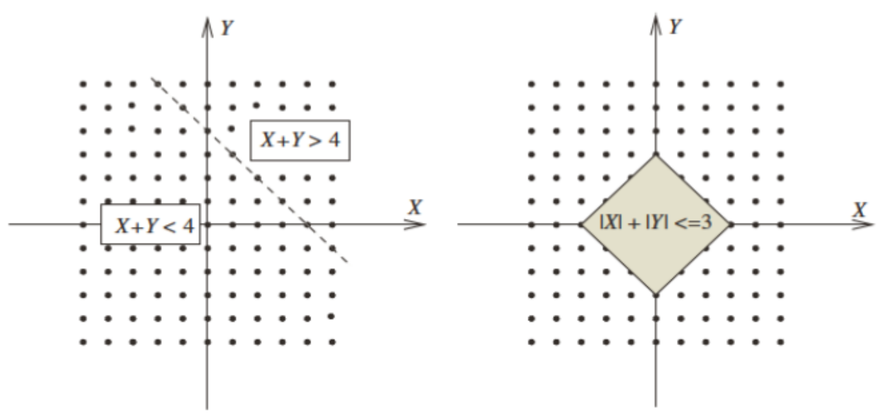
Ejemplo 6.1 Sea X una variable aleatoria que tiene los valores en $\{1,2,3,4,5\}$ y Y una variable aleatoria con valores en $\{-1,0,1,2\}$ y sea la función de masa de probabilidad conjunta dada por $p_{X,Y}(x,y) = \alpha$, para todos los posibles valores de x e y. Para calcular los valores de α, observamos que sólo 20 puntos en el plano (x,y) tiene probabilidad positiva y además la probabilidad de cada uno es la misma. Esto nos dá $\alpha = 1/20 = 0.05$. La tabla muestra la función de masa de probabilidad conjunta de X y Y.

	x = 1	x = 2	x=3	x =4	x =5
y = -1	1/20	1/20	1/20	1/20	1/20
y = 0	1/20	1/20	1/20	1/20	1/20
y = 1	1/20	1/20	1/20	1/20	1/20
y = 2	1/20	1/20	1/20	1/20	1/20

Cada evento A es un subconjunto de puntos en el plano (x,y), por lo que podemos escribir como:

$$\mathbb{P}(A) = \sum_{(x,y)\in A} p_{X,Y}(x,y).$$

Donde la suma es sobre todos los puntos $(x,y) \in A$. La siguiente figura, muestra dos eventos $A = [X+Y<4]$ y $B = [|X|+|Y| \leq 3]$, donde todos los posibles valores de X e Y están marcados por puntos.



Ejemplo 6.2 Sean X e Y variables aleatorias continuas cuya función de masa de probabilidad conjunta es dada por el anterior ejemplo. Sea A el evento que X no tiene valores mayores que Y escrito como $[X \leq Y]$. Los puntos en el plano, teniendo una probabilidad positiva, para el cual, el valor de X es menor o igual al valor de Y son $(1,1), (1,2)$ y $(2,2)$. Por tanto la probabilidad de este evento es dado por $\mathbb{P}(A) = \mathbb{P}(X \leq Y) = 3/20$.

Sea ahora el evento $[XY = 0]$. Este evento contiene cinco salidas: $(1,0), (2,0), (3,0), (4,0)$ y $(5,0)$ y así $\mathbb{P}(B) = \mathbb{P}(XY = 0) = 5/20$.

La función de masa de probabilidad conjunta, especifica la probabilidad de eventos conjuntos, denotados por $[X = x, Y = y]$. Es posible considerar la probabilidad de eventos que se refiere a una de las dos variables aleatorias, tales como $[X = x]$. En efecto considerando el ejemplo anterior, el evento $[X = 2]$ consiste de cuatro salidas $(2,-1), (2,0), (2,1)$ y $(2,2)$ y tienen una probabilidad igual a $4/20$. En este caso, usamos la función de masa de probabilidad conjunta de las dos variables aleatorias para construir la función de masa de probabilidad para cada una de las variables aleatorias. Esas funciones son las `funciones de masa probabilidad marginal`. La función de masa de probabilidad marginal de X es obtenida como:

$$p_X(x) = \mathbb{P}(X = x) = \sum_k \mathbb{P}(X = x, Y = y_k) = \sum_k p_{X,Y}(x, y_k). \tag{4}$$

Y es una consecuencia directa, del hecho que para un evento A, $\mathbb{P}(A) = \sum_{(x,y)\in A} p_{X,Y}(x,y)$. La distribución marginal de Y es obtenida de manera similar.

Definición 6.1 Las variables aleatorias X e Y se dicen que son independientes si se cumple :

$$\mathbb{P}(X = x_i, Y = y_i) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_i)$$

Para todo x_i y y_i. En otras palabras, X e Y son independientes si su función de masa conjunta, se factoriza en el producto de sus funciones de masa de probabilidad individualmente.

Ejemplo 6.3 Sea una distribución para dos variables aleatorias X y Y que toman los valores 0 - 1:

	Y = 0	Y = 1	
X = 0	1/9	2/9	1/3
X = 1	2/9	4/9	2/3
	1/3	2/3	1

Entonces $p_{X,Y}(1,1) = \mathbb{P}(X = 1, y = 1) = 4/9$.

Ejemplo 6.4 Sean X e Y dos variables aleatorias discretas con una función de masa de probabilidad conjunta :

	X = -1	X = 0	X=1
Y = -1	1/12	3/12	1/12
Y = 0	1/12	0/12	1/12
Y = 1	1/12	3/12	1/12

Así, ambas variables aleatorias X y Y toman los valores, desde el conjunto $\{-1, 0, 1\}$. La tabla indica que $\mathbb{P}(X = 0, Y = 0) = 0$, que $\mathbb{P}(X = 0, Y = 1) = 1/4$ y así sucesivamente. La función de masa de probabilidad marginal de X, $p_X(x)$ es obtenida sumando las entradas en cada columna de la tabla, la de Y, $p_Y(y)$ se encuentra sumando las entradas en cada fila de la tabla. Por lo que se tiene la siguiente tabla:

	X = -1	X = 0	X=1	$p_Y(y)$
Y = -1	1/12	3/12	1/12	5/12
Y = 0	1/12	0/12	1/12	1/6
Y = 1	1/12	3/12	1/12	5/12
$p_X(x)$	1/4	1/2	1/4	1

En este ejemplo, X e Y no son independientes, desde que :

$$1/12 = \mathbb{P}(X = 1, Y = 1) \neq \mathbb{P}(X = 1)\mathbb{P}(Y = 1) = 3/12 \times 5/12.$$

De manera similar, se aplica a $X = (X_1, X_2, \ldots, X_n)$ un vector de variables aleatorias discretas. Para este caso la función de masa de probabilidad conjunta de X es definido como la función p_X definida por:

$$p_X(x) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

para $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

6.2 Función de distribución acumulativa conjunta

La función de distribución acumulativa conjunta de dos variables X e Y es dada por:

$$F_X, Y(x, y) = \mathbb{P}(X \leq x, Y \leq y); \quad -\infty < x < \infty, \ -\infty < y < \infty.$$

Para ser un CDF de buena fe, una función debe poseer las siguientes propiedades, similares a las encontradas en el CDF de una sola variable aleatoria:

- $0 \leq F_{X,Y}(x, y) \leq 1$ para $-\infty < x < \infty, \ -\infty < y < \infty$, desde que $F_{X,Y}(x, y)$ es una probabilidad.
- $F_{X,Y}(x, \infty) = F_{X,Y}(-\infty, y) = 0$
  $F_{X,Y}(-\infty, -\infty) = 0$, $F_{X,Y}(+\infty, +\infty) = 1$.
- $F_{X,Y}$ debe ser una función no decreciente de ambos x e y, es decir:

$$F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2) \text{ si } x_1 \leq x_2 \text{ si } y_1 \leq y_2.$$

Ejemplo 6.5 La función distribución acumulativa de una variable aleatoria X dado en forma tabular, es dada por:

	$-2 < x$	$-2 \leq x < -1$	$-1 \leq x < 0$	$0 \leq x < 1$	$1 \leq x$
$F_x(x)$	0	1/8	1/4	1/2	1

Esto nos dice por ejemplo, que $\mathbb{P}(X < 1.0) = 1/2$. También $\mathbb{P}(X \leq 0.5) = 1/2, \mathbb{P}(X \leq 0.0) = 1/2$ y así. De manera similar $\mathbb{P}(X < -1.0) = 1/8, \mathbb{P}(X \leq -1.5) = 1/8$. Usando el mismo enfoque, podemos encontrar la función de distribución acumulativa conjunta de dos variables aleatorias discretas X e Y en forma tabular. Un ejemplo es el siguiente:

$y \geq 2$	0	1/8	1/4	1/2	1
$0 \leq y < 2$	0	3/32	3/16	3/8	3/4
$-2 \leq y < 0$	0	1/32	1/16	1/8	1/4
$y < 2$	0	0	0	0	0
$F_{X,Y}(x, y)$	$-2 < x$	$-2 \leq x < -1$	$-1 \leq x < 0$	$0 \leq x < 1$	$1 \leq x$

De esta tabla, podemos encontrar por ejemplo que $\mathbb{P}(X < 1, Y < 0) = 1/8$, que $\mathbb{P}(X \leq 0.5, Y < 0) = 1/8 = \mathbb{P}(X \leq 0.0, Y < 0)$, que $\mathbb{P}(X \leq -1.3, Y \leq 1.3) = 3/32$, que $\mathbb{P}(X \geq 1, Y < 1) = 3/4$ y así.

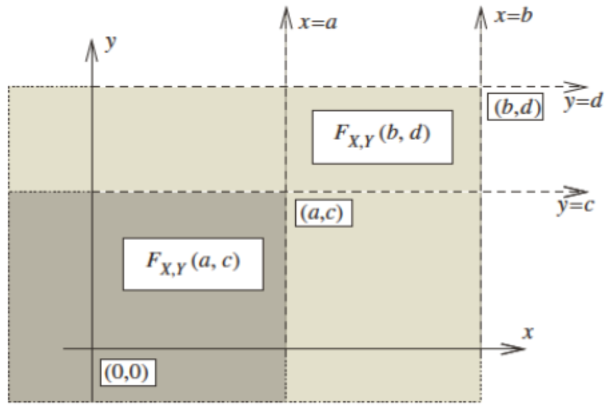
Se observa que las condiciones detalladas anteriormente para que una función sea un CDF de buena fe se mantienen en este ejemplo.

Mientras que una sola variable aleatoria mapea los resultados en puntos de la línea real, la variable aleatoria conjunta mapea los resultados en puntos en un espacio de mayor dimensión. Por ejemplo, dos variables aleatorias asignan resultados en el plano (x, y), tres variables aleatorias mapean los resultados en el espacio tridimensional (x, y, z) y así sucesivamente. Para una sola variable aleatoria, el valor de la función de distribución acumulativa en un punto x es igual a la probabilidad del evento que contiene todos los resultados que son mapeadas en puntos menores o iguales a x.

El valor de la distribución acumulativa de dos variables aleatorias X e Y en el punto (x, y) es la probabilidad del evento que contiene todos los resultados que se asignan en el área infinita que se encuentra a la izquierda de una línea vertical a través de $X = x$ y por debajo de la línea horizontal a través de $Y = y$.

Puede observarse de la siguiente figura, que $F_{X,Y}(a, b) = \mathbb{P}(X \leq b, Y \leq d)$ es igual a la probabilidad del evento que contiene todas las salidas dentro de las áreas sombreadas (sombreado ligero y fuerte), mientras que $F_{X,Y}(a, c) = \mathbb{P}(X \leq a, Y \leq c)$ es igual a la probabilidad del evento que contiene todas las salidas dentro del área de sombreado fuerte.

Otros resultado siguen, por ejemplo:



$$\mathbb{P}(a < X \leq b, Y \leq c) = F_{X,Y}(b,c) - F_{X,Y}(a,c)$$

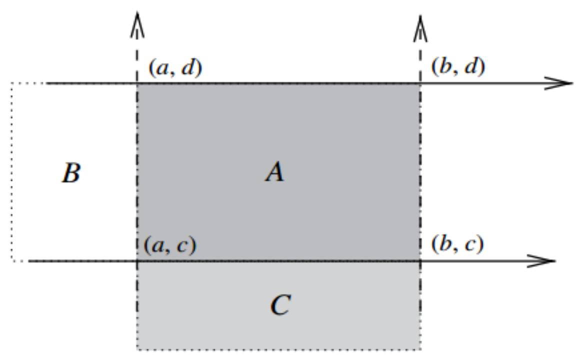
$$\mathbb{P}(X \leq a, c < Y \leq d) = F_{X,Y}(a,d) - F_{X,Y}(a,c)$$

El uso de la función de distribución acumulativa conjunta de dos variables aleatorias para calcular la probabilidad que (X, Y) esté en un rectángulo en el plano no es excesivamente complicado. Podemos usar el resultado:

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = F_{X,Y}(b,d) - F_{X,Y}(a,d) - F_{X,Y}(b,c) + F_{X,Y}(a,c).$$

Lo que requiere que la función de distribución acumulativa conjunta se evalúe en cada uno de los cuatro puntos del rectángulo. Desafortunadamente, cuando la probabilidad requerida se encuentra en una región no rectangular, el cálculo del CDF conjunto se vuelve mucho más difícil de usar.

Procedemos de la siguiente manera: Sea A que representa un rectángulo con vértices $(a,c), (b,c), (b,d)$ y (a,d) (es decir, el rectángulo dentro del cual buscamos la probabilidad que X, Y pueden ser encontradas). Entonces $\mathbb{P}(A) = \mathbb{P}(a < X \leq b, c < Y \leq d)$. Sea B el rectángulo semi-infinito que se encuentra a la izquierda de A y sea C que representa el rectángulo semi-infinito que está debajo de A, como se ilustra en la en la siguiente figura.



Se debe observar que A, B y C son eventos que no tienen puntos en común, es decir, son mutuamente excluyentes, así:

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C),$$

Lo que da como resultado:

$$\mathbb{P}(A) = \mathbb{P}(A \cup B \cup C) - \mathbb{P}(B) - \mathbb{P}(C)$$

Además de la definicón de CDF conjunta, se tienen los siguientes resultados:

$$\begin{aligned}
\mathbb{P}(B) &= F_{X,Y}(a,d) - F_{X,Y}(a,c), \\
\mathbb{P}(C) &= F_{X,Y}(b,c) - F_{X,Y}(a,c), \\
\mathbb{P}(A \cup B \cup C) &= F_{X,Y}(b,d) - F_{X,Y}(a,c),
\end{aligned}$$

y podemos concluir que :

$$\begin{aligned}
\mathbb{P}(A) &= F_{X,Y}(b,d) - F_{X,Y}(a,c) - [F_{X,Y}(a,d) - F_{X,Y}(a,c) + F_{X,Y}(b,c) - F_{X,Y}(a,c)] \\
&= F_{X,Y}(b,d) - F_{X,Y}(a,d) - F_{X,Y}(b,c) + F_{X,Y}(a,c).
\end{aligned}$$

Ejemplo 6.6 La función de distribución acumulativa conjunta de dos variables aleatorias continuas es dada por:

$$F_{X,Y}(x,y) = (1 - e^{-\lambda x})(1 - e^{-\mu y}), \ 0 \leq x < \infty, 0 \leq y < \infty,$$

y es cero en otros partes del plano, donde ambos λ y μ son estrictamente positivo. Calculemos la probabilidad $\mathbb{P}(0 \leq x \leq 1, 0 \leq y \leq 1)$. Por propiedad tenemos:

$$\begin{aligned}
\mathbb{P}(0 \leq x \leq 1, 0 \leq y \leq 1) &= F_{X,Y}(1,1) - F_{X,Y}(0,1) - F_{X,Y}(1,0) + F_{X,Y}(0,0) \\
&= (1 - e^{-\lambda})(1 - e^{-\mu}) - (1 - e^{-\lambda})(1 - e^{0}) - (1 - e^{0})(1 - e^{-\mu}) + (1 - e^{0})(1 - e^{0}) \\
&= (1 - e^{-\lambda})(1 - e^{-\mu}).
\end{aligned}$$

Ahora volvemos nuestra atención a las funciones de distribución acumulativa marginal derivadas de las CDF conjuntas considerando lo que ocurre cuando a una , pero no a ambas, de las variables aleatorias X e Y se hace tender al infinito. En el límite cuando $y \to \infty$, el evento $[X \leq x, Y \leq y]$ tiende al evento $[X \leq x, Y < \infty] = [X \leq x]$ (se ha utilizado que $Y < \infty$ para indicar que Y puede tomar cualquier valor). Esto implica que la única restricción es $[X \leq x]$, es decir, el evento $[X \leq x, Y < \infty]$ se reduce al evento $[X \leq x]$. Pero ahora, la probabilidad del evento $[X \leq x]$ es simplemente la distribución acumulativa de X, así:

$$\lim_{y \to \infty} F_{X,Y}(x,y) = F_X(x).$$

De manera similar:

$$\lim_{x \to \infty} F_{X,Y}(x,y) = F_Y(y).$$

Estas funciones son llamadas `distribuciones marginales` y pueden ser calculadas desde la distribución conjunta.

Ejemplo 6.7 Del ejemplo (6.5), podemos calcular las distribuciones marginales como:

	$-2 < x$	$-2 \leq x < -1$	$-1 \leq x < 0$	$0 \leq x < 1$	$1 \leq x$
$F_x(x)$	0	1/8	1/4	1/2	1

para X y para Y:

	$-2 < y$	$-2 \leq y < 0$	$0 \leq y < 2$	$y \geq 2$
$F_y(y)$	0	1/4	3/4	1

De manera similar, las CDF marginales de X y Y, derivadas desde la CDF conjunta del ejemplo (6.6), son calculadas fácilmente y son:

$$F_X(x) = (1 - e^{-\lambda x})(1 - 0) = (1 - e^{-\lambda x})$$
$$F_Y(y) = (1 - 0)(1 - e^{-\mu y}) = (1 - e^{-\mu y})$$

Definición 6.2 Dos variables aleatorias X e Y son independientes si:

$$F_{X,Y}(x,y) = F_X(x)F_Y(y), \quad -\infty < x < \infty, \ -\infty < y < \infty.$$

En otras palabras, si X e Y son variables aleatorias independientes, la función de distribución acumulativa conjunta se factoriza en el producto de sus funciones de distribución acumulativa marginal. Se puede comprobar que las variables aleatorias X e Y de los ejemplos (6.5) y (6.6) son independientes. Si, además, X e Y tienen la misma función de distribución (por ejemplo, cuando $\lambda = \mu$ en el ejemplo (6.6)), entonces el término `independiente e idénticamente distribuido` se le aplica a cada uno de ellos . Normalmente se abreviará como **IID**.

Definición 6.3 Dos variables aleatorias discretas X e Y son independientes si el par de eventos $\{X = x\}$ y $\{Y = y\}$ son independientes para todo $x, y \in \mathbb{R}$. Se escribe esta condición como:

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \quad \text{para} \ \ x, y \in \mathbb{R}.$$

Las variables aleatorias que no son independientes, se llaman `dependientes`.

Teorema 6.1 Las variables aleatorias discretas X e Y son independientes si y sólo si existen las funciones $f, g : \mathbb{R} \to \mathbb{R}$ tal que se cumple que las funciones de masa de probabilidad conjunta de X e Y satisface:

$$p_{X,Y}(x,y) = f(x)g(y) \quad \text{para} \ \ x, y \in \mathbb{R}.$$

Ejemplo 6.8 Supongamos que X e Y son variables aleatorias tomando valores enteros no negativos, con función de masa de probabilidad conjunta:

$$p_{X,Y}(i,j) = \mathbb{P}(X = i, Y = j) = \frac{1}{i!j!}\lambda^i \mu^j e^{-(\mu+\lambda)} \ \ \text{para} \ i, j = 0, 1, 2, \ldots$$

Por el teorema anterior, X e Y son independientes.