

Probabilidad 1

Facultad de Ciencias. UNAM

4 de marzo de 2021

Introducción

Existen multitud de fenómenos en la realidad que difícilmente pueden ser modelados con certeza. Ya sea por la complejidad de su naturaleza o porque efectivamente es imposible establecer algún modelo determinístico para el mismo. Ejemplos que nos vienen a la mente provienen directamente de juegos de azar tales como,

- El lanzamiento de una moneda,
- el lanzamiento de un par de dados,
- el resultado de seleccionar 5 cartas a partir de un mazo de 52 cartas,
- y un largo etcétera...

Introducción

Si bien estos ejemplos son perfectamente válidos para ejemplificar un fenómeno aleatorio, la teoría de probabilidad puede ir mucho mas lejos. A decir verdad su rango de aplicabilidad solo esta sujeto a nuestra creatividad e imaginación.

Aplicaciones en economía, finanzas, biología, física... abundan.

También es importante notar que la Probabilidad es fundamento de muchas otras disciplinas, lo cual le confiere aun mayor relevancia.

Espacio Muestral

Empecemos por lo más básico... Al momento de modelar un fenómeno aleatorio, debemos empezar por delimitar las posibilidades. El ejemplo mas simple sería el lanzamiento de una moneda, típicamente podemos pensar que las únicas dos posibilidades es que la moneda caiga en una cara u otra...

Esto lo podemos traducir diciendo que la moneda cae en "sol" (S) o bien cae en "águila" (A). Es decir el posible resultado de este sencillo experimento es alguno de los elementos del conjunto:

$$\Omega = \{A, S\}$$

En el caso del lanzamiento de un par de dados, el conjunto Ω podría ser:

$$\Omega = \{(i, j), i, j = 1, \dots, 6\}$$

Espacio Muestral

En general. . .

Espacio Muestral

El espacio muestral de un fenómeno aleatorio, es el conjunto de resultados que consideramos **posibles**.

No es imperativo pero si muy común, denotar al espacio muestral con Ω , nosotros seguiremos esta costumbre.

De acuerdo con esta definición, cualquier cosa que este fuera de Ω es para todos los fines de nuestro modelo, algo completamente imposible. Si. . . en el conjunto Ω que propusimos para el lanzamiento de una moneda, dejamos fuera la posibilidad de que la moneda caiga de canto. . .

Espacio Muestral

Debemos hacer énfasis en que el espacio muestral que definamos es válido en un contexto muy específico. Es decir, estamos modelando y cualquier modelo tiene imperfecciones. Ningún modelo es un reflejo fiel de la realidad y en general no se espera que sea así. Lo importante es que cuando proponemos un modelo para describir algún fenómeno social, natural o de cualquier índole, debemos hacer supuestos simplificadores que permitan atacar el problema de manera eficiente. Mas aun, podemos proponer mas de un modelo perfectamente válido para atacar un problema desde enfoques diferentes.

Pensemos en la Ciudad de México, si queremos determinar una ruta aérea de la CDMX a Mérida, sería perfectamente válido y útil pensar que ambas ciudades son “puntos” en el mapa. Sin embargo seria una muy mala idea pensar del mismo modo si lo que deseamos es proponer un modelo que busque resolver el problema del tráfico en la ciudad.

Espacio Muestral

En un contexto de orden mas probabilístico; supongamos que un inversionista esta interesado en el comportamiento que tendrá el precio de un activo a lo largo de los tres días siguientes.

Es claro que al día de hoy conocemos el precio del activo. Pero desconocemos el precio que tendrá en los días subsecuentes. Este precio es **aleatorio**. Sin embargo aunque desconocemos los precios de los días venideros si podemos intentar definir algún modelo.

¿Que valores podrá tomar el activo? La primera observación es que el precio del activo es siempre no negativo. Además de ello ¿que podemos suponer con respecto del precio que tomara mañana o pasado mañana? Un modelo simple, pero de ningún modo despreciable supone que de un día a otro el precio solo puede tomar uno de dos valores. Uno de estos corresponde a un alza y el otro a una baja. Asi de simple

Espacio Muestral

Matemáticamente; si denotamos con S_t al precio del activo al tiempo t entonces proponemos el modelo:

$$S_{t+1} = \begin{cases} uS_t & u > 1 \\ dS_t & 0 < d < 1 \end{cases}$$

Si todo ‘empieza’ al tiempo t entonces S_{t+3} es desconocida al tiempo t pero sabemos que puede tomar cualquiera de los valores:

$$S_{t+3} = u^i d^{3-i} S_t, \quad i = 0, \dots, 3.$$

El problema es que no sabemos cual de ellos tomará. No importa, podemos modelar el precio del activo al tiempo $t + 3$ como un experimento aleatorio. ¿cual sería en este caso el espacio muestral?

Espacio Muestral

No es la única posibilidad, pero una buena idea sería definir el espacio muestral en términos de las alzas y bajas del activo:

$$\Omega = \{w = (x_1, x_2, x_3) : x_i \in \{u, d\}, i = 1, 2, 3\}$$

Así por ejemplo si $w = (u, u, d)$ entonces sabemos que en $t + 3$ el precio del activo es $u^2 d S_t$. Pero si $w = (d, d, d)$ entonces $S_{t+3} = d^3 S_t$.

En resumen, el conjunto Ω , reúne todos los escenarios posibles. De acuerdo con nuestro modelo, solo las alternativas que están en Ω pueden ocurrir. Cualquier otra cosa es imposible. Aun que no sabemos cual $w \in \Omega$ ocurrirá, al menos hemos delimitado las posibilidades.

Espacio Muestral

El modelo que hemos propuesto, si bien simple, es un modelo muy popular en el medio financiero. En la práctica se usa tal cual lo describimos, usualmente con un mayor número de periodos, pero fuera de ello sin cambio alguno. Por muy simple que parezca de primera instancia, es un modelo muy adecuado para resolver algunos tipos de problemas en finanzas. Claro esta que no es el único modelo de uso extendido para modelar precios de activos. Una alternativa quizás mas ambiciosa seria suponer que a lo largo de los tres días de operación, el activo puede tomar cualquier valor no negativo. En este caso:

$$\Omega = \mathbb{R}^+$$

Sirva esto para hacer notar que Ω puede tomar formas muy variadas, desde conjuntos sencillos, pequeños y finitos, hasta conjuntos infinitos no numerables.

Espacio Muestral

Ejemplo

En ciertas circunstancias una adecuada definición del espacio muestral, puede ser la diferencia entre resolver o no un problema. Consideremos el juego de la “Catafixia”.

Un concursante llega al final de un programa cargado de premios. Sin embargo, se le ofrece la posibilidades de entrar a la “catafixia”. Si entra entonces le son presentadas tres cortinas diferentes. Sabemos que detrás de una de ellas existe un premio sumamente deseable. . . el premio mayor digamos. Detrás de las dos restantes, existen premios indeseables. Si abrimos alguna de estas cortinas, perdemos los premios con los que llegamos al final del programa y nos vamos con cualquier cosa que hubiese detrás de la cortina seleccionada. ¿Entramos a la catafixia?

Pensemos por un momento. . . si entramos tenemos un tercio de probabilidad de obtener el “premio mayor”. ¿Estamos dispuestos a arriesgar los premios que ya ganamos?

Espacio Muestral

Decidimos entrar, seleccionamos una de la cortinas. . . pero antes de abrir la cortina seleccionada, Chabelo nos ofrece una alternativa. Abre una de las dos cortinas que no fueron seleccionadas, una detrás de la cual se encuentra uno de los premios “broma”. Nos encontramos entonces en nueva situación; el premio mayor esta detrás de la cortina que seleccionamos o bien detrás de la cortina que Chabelo dejó cerrada. Se nos ofrece entonces la posibilidad de cambiar nuestra decisión. ¿Lo hacemos? En realidad si llegamos a este punto cambiar de cortina es irrelevante. . . ahora la posibilidad que tenemos de ganar es de un medio. Problema resuelto. . .

¿o no?

Espacio Muestral

Hasta el momento el único concepto de probabilidad que conocemos es el de Espacio Muestral. Bueno, entonces seguramente, tomar una decisión adecuada dependerá de una correcta elección de nuestro espacio muestral. Para ello supongamos que las cortinas están numeradas del 1 al 3. Para simplificar la situación supongamos además que el premio mayor se encuentra detrás de la cortina uno. ¿Como sería una elección típica en este juego?

En realidad, depende. . . depende de la elección que tomemos, cuando llegue el momento ¿cambiamos de cortina?

Hasta el momento, la conclusión “obvia” es que el cambio es irrelevante. Bueno, pensemos que pasaría si hacemos el cambio. Pase lo que pase, llegado el momento, nuestra estrategia es la de cambiar la cortina seleccionada. En este caso ¿como se vería nuestro espacio muestral?

Espacio Muestral

Para encontrar una respuesta satisfactoria, partamos del hecho, de que pase lo que pase, hemos decidido cambiar nuestra selección llegado el momento. Luego, sabemos que en primera instancia, tenemos tres posibilidades a elegir, puerta 1, puerta 2 o puerta 3. Esta primera selección la hacemos de forma completamente aleatoria. En este momento no tenemos ninguna razón para seleccionar una cortina sobre otra. Después de esta primera selección, puede ocurrir una de las siguientes cosas

- ➊ Seleccionamos la cortina 1.
 - ➊ Chabelo abre la cortina 2. Cambiamos por la 3. Perdemos
 - ➋ Chabelo abre la cortina 3. Cambiamos por la 2. Perdemos
- ➋ Seleccionamos la cortina 2. Chabelo abre la cortina 3. Cambiamos por la 1. ganamos
- ➌ Seleccionamos la cortina 3. Chabelo abre la cortina 2. Cambiamos por la 1. ganamos

Espacio Muestral

Ante estas circunstancias, parece razonable el siguiente espacio muestral:

$$\Omega = \{(1, 2, 3, P), (1, 3, 2, P), (2, 3, 1, G), (3, 2, 1, G)\}$$

Efectivamente, si seleccionamos la cortina 2, es equivalente en nuestro espacio muestral a haber seleccionado:

$$\omega = (2, 3, 1, G)$$

Es decir seleccionamos la cortina 2, nos abren la 3, cambiamos a la 1 y en consecuencia ganamos! Notemos que si seleccionamos la uno, entonces se derivan dos posibilidades, en función de la cortina que nos abran en segunda instancia.

Espacio Muestral

Bueno pues el problema esta resuelto, si estamos convencidos del espacio muestral elegido, entonces ganamos en dos de cuatro resultados posibles. En consecuencia, la probabilidad de ganar siguiendo nuestra estrategia de cambio de elección, es igual a $2/4 = 1/2$. Lo cual confirma nuestra hipótesis inicial.

mmm

Un momento... regresemos al principio... Antes del cambio seleccionamos la puerta 1, 2 o 3 de forma completamente aleatoria. Asi que cada una tiene $1/3$ de probabilidad de ser seleccionada. Notemos también que de acuerdo con nuestra elección de espacio muestral, perdemos si el resultado del concurso es $(1, 2, 3, P)$ o $(1, 3, 2, P)$. Pero esto es equivalente al haber elegido la cortina 1 desde un principio. Lo cual lo hacemos con probabilidad $1/3$.

Espacio Muestral

Por otro lado, ganamos si el resultado del concurso es $(2, 3, 1, G)$ o $(3, 2, 1, G)$. Pero esto es equivalente al haber seleccionado la cortina 2 o la cortina 3 desde un principio. Seleccionamos alguna de estas dos cortinas con probabilidad $1/3 + 1/3 = 2/3$.

Conclusión: Si seguimos la estrategia de cambio, la probabilidad de ganar aumenta de uno a dos tercios!!! :-)

Fundamentos

Una vez que hemos comprendido la idea de espacio muestral, debemos trabajar en establecer cierto orden sobre el mismo. Un orden que nos permita extraer información valiosa. Información que nos conduzca a cálculos mas precisos.

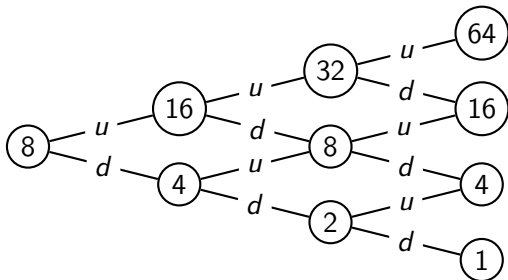
Para este fin retomemos el ejemplo de los precios de activos. Habíamos propuesto el espacio de estados:

$$\Omega = \{w = (x_1, x_2, x_3) : x_i \in \{u, d\}, i = 1, 2, 3\}$$

Para lo que intentamos exponer a continuación, conviene escribir este espacio de manera explícita:

$$\Omega = \{(uuu), (uud), (udu), (udd), \\ (ddd), (ddu), (dud), (duu)\}$$

Una representación gráfica de este problema la tenemos a continuación:



Hemos dado valores a las variables. Por lo pronto $S_t = 8$, mientras que $u = 2$ y $d = 1/2$.

Fundamentos

Lo primero que debemos notar es que el espacio de probabilidad que hemos propuesto describe cada una de las rutas en el árbol. Bien nos podríamos preguntar si es que es necesario tanto detalle. A final de cuentas solo estamos interesados en el resultado final, el precio al término de los tres días.

De acuerdo con los parámetros seleccionados, el precio del activo al tiempo $t + 3$ solo puede tomar uno de los cuatro valores: 64, 16, 4 o 1. En este sentido, podríamos proponer un espacio de estados alternativo:

$$\Omega = \{64, 16, 4, 1\}$$

¿Cuál es mejor? En realidad los dos pueden ser correctos, todo depende del problema que deseamos resolver. Sin embargo lo que si es claro es que el primer espacio que propusimos, contiene mayor información que el segundo.

Fundamentos

Pensemos por ejemplo, que ocurre si después de transcurridos los tres días, lo que ocurrió fueron tres alzas: $\omega = (u, u, u)$. En términos de los precios, estamos diciendo que $S_{t+3} = 64$. Si lo que ocurre es $\omega = (u, d, u)$ entonces $S_{t+3} = 16$.

De hecho observemos que si $\omega \in \{(u, u, d), (u, d, u), (d, u, u)\}$ entonces $S_{t+3} = 16$. De manera análoga, si $\omega \in \{(d, d, u), (d, u, d), (u, d, d)\}$ entonces $S_{t+3} = 4$.

Pero podemos hacer aún mas; Consideremos por ejemplo los conjuntos:

$$A_u = \{(u, u, u), (u, u, d), (u, d, u), (u, d, d)\}$$

$$A_d = \{(d, d, d), (d, d, u), (d, u, d), (d, u, u)\}$$

El conjunto A_u contiene la información del primer día. Si $\omega \in A_u$ sabemos que el primer día hubo una alza en el precio del activo, si $\omega \in A_d$ entonces hubo una baja. Observemos sin embargo que ni A_u , ni A_d proporcionan información de lo que ocurrirá en el segundo o tercer día, aunque si acotan las posibilidades. Si sabemos que $\omega \in A_u$ automáticamente descartamos la posibilidad de que $\omega = (d, d, d)$ por ejemplo.

Fundamentos

Continuando con el ejemplo; tomemos ahora los conjuntos:

$$A_{uu} = \{(u, u, u), (u, u, d)\}$$

$$A_{du} = \{(d, u, u), (d, u, d)\}$$

$$A_{ud} = \{(u, d, u), (u, d, d)\}$$

$$A_{dd} = \{(d, d, u), (d, d, d)\}$$

La información que obtenemos de estos conjuntos esta relacionada con lo que ocurre en los dos primeros días. Si $\omega \in A_{uu}$ podemos afirmar que en los dos primeros días se registraron alzas de precios. Notemos ademas que ninguno de estos conjuntos determinan con certeza lo que ocurrirá en el tercer día. Aunque si limitan las posibilidades.

Aquí viene algo interesante, notemos que decir que $S_{t+2} = 8$ es equivalente a decir que $\omega \in A_{ud} \cup A_{du}$. O también por ejemplo decir que $S_{t+2} > 2$ es equivalente a afirmar que $\omega \in A_{dd}^c$

Fundamentos

Tomemos ahora un ejemplo diferente, pesemos en el lanzamiento de dos dados. En este caso el espacio muestral parece muy natural; la colección de pares (i, j) donde $i, j = 1, \dots, 6$. Es decir:

$$\begin{aligned}\Omega = \{ & (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6) \\ & (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6) \\ & (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6) \\ & (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6) \\ & (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6) \\ & (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6) \}\end{aligned}$$

Fundamentos

¿Cuales son los casos donde obtuvimos un 3 en el lanzamiento del primer dado? Esto lo resumimos en el siguiente conjunto:

$$A = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$$

Mientras que el el conjunto

$$B = \{(1, 3), (2, 3), (3, 3), (4, 3), (5, 3), (6, 3)\}$$

representa los casos donde hubo un tres en el segundo dado. Algo un poco mas complejo, de forma coloquial decimos:

$$A \cup B = \text{sale 3 en primer dado o sale 3 en el segundo}$$

a diferencia de

$$A \cap B = \text{sale 3 en el primer dado y sale 3 en el segundo dado}$$

Las dos aseveraciones anteriores podemos escribirlas de manera equivalente como:

$$A \cup B = \text{obtuvimos al menos un 3}$$

a diferencia de

$$A \cap B = \text{obtuvimos 3 en ambos dados}$$

En términos de conjuntos ¿cómo podríamos expresar la frase “obtuvimos exactamente un 3”?

Observemos que $A \cup B$ no es útil pues contiene a $(3, 3)$, en este elemento tenemos mas ocurrencias del 3 de las que necesitamos. $A \cap B = \{(3, 3)\}$ nos sirve aun menos, En realidad es el elemento que desearíamos eliminar de $A \cup B$ para obtener lo que estamos buscando. Esto es:

$$(A \cup B) \setminus (A \cap B) = \text{Obtenemos exactamente un tres}$$

Fundamentos

Recordemos que la diferencia simétrica de A y B se define como:

$$A \triangle B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$$

Por lo tanto, en el ejemplo que estamos estudiando $A \triangle B$ es el conjunto donde “obtenemos exactamente un tres”. ¿Que podemos decir entonces de la diferencia usual de conjuntos?

$A \setminus B$ = Obtuvimos 3 en el primer dado pero no en el segundo

Observemos que en esta última expresión estamos negando la ocurrencia de un 3 en el segundo dado, es decir:

B^c = No obtuvimos 3 en el segundo dado

Fundamentos

Todas estas divagaciones nos conducen a conclusiones interesantes. Por un lado como vimos, no existe una única forma de definir un espacio muestral, la forma adecuada dependerá en gran medida de las preguntas que necesitemos resolver. Por otro lado, debemos estar ya convencidos, que al reunir los elementos de Ω en subconjuntos, en realidad lo que logramos es organizar la información que eventualmente podamos extraer del experimento aleatorio. Estas ideas nos conducen a la siguiente definición.

σ -álgebra

Una colección \mathcal{F} de subconjuntos de Ω se conoce como σ -álgebra de Ω si:

- 1 $\Omega \in \mathcal{F}$
- 2 Si $A \in \mathcal{F}$ entonces $A^c \in \mathcal{F}$
- 3 Si $\{A_n\}$ es una colección numerable de elementos de \mathcal{F} entonces $\cup A_n \in \mathcal{F}$.

Fundamentos

Ejemplos

Si $A \in \mathcal{F}$ entonces se dice que A es un **evento**. Decimos que un evento A ocurre siempre y cuando el elemento $\omega \in \Omega$ que resulta después de ejecutado el experimento aleatorio es un elemento de A esto es $\omega \in A$.

Ejemplos de σ -álgebra. Dado un espacio muestral Ω , la σ -álgebra más simple y pequeña que podemos definir es $\mathcal{F} = \{\Omega, \emptyset\}$. Otro ejemplo simple lo conseguimos como sigue: si $A \subset \Omega$ entonces: $\mathcal{F} = \{\Omega, \emptyset, A, A^c\}$ es una σ -álgebra. Y claro está, la σ -álgebra mas grande de todas 2^Ω . El conjunto potencia es una σ -álgebra que es útil en algunas aplicaciones, por ejemplo en el caso del lanzamiento de dos dados bien podemos usar al conjunto potencia como la σ -álgebra en nuestro experimento.

En el ejemplo del precio de un activo también podríamos usar la potencia, sin embargo, podríamos jugar con algunas alternativas. Por ejemplo; para el tiempo t donde desconocemos cualquier cosa que pueda ocurrir en el futuro, podríamos seleccionar la σ -álgebra: $\mathcal{F} = \{\Omega, \emptyset\}$.

Fundamentos

Ejemplos

Ahora bien, si ya transcurrió un periodo, entonces tenemos conocimiento de lo ocurrido con el precio en el primer día. Es decir, ya sabemos si $\omega \in A_u$ o si $\omega \in A_d$. Podemos jugar con esta información y generar una σ -álgebra que denote el hecho que conocemos la historia del primer día:

$$\mathcal{F} = \{\emptyset, \Omega, A_u, A_d\}$$

Notemos que en este caso \mathcal{F} efectivamente es una σ -álgebra. Esto es fácil de ver dado que $A_u = A_d^c$.

Transcurridos dos días tenemos conocemos aún mas información, por lo pronto sabemos si ω pertenece a alguno de los cuatro conjuntos: A_{uu}, A_{ud}, A_{du} o A_{dd} . Sin embargo:

$$\mathcal{C} = \{\emptyset, \Omega, A_{uu}, A_{ud}, A_{du}, A_{dd}\}$$

no es una σ -álgebra.

Fundamentos

Ejemplos

Por lo pronto conjuntos como: $A_u = A_{uu} \cup A_{ud}$ o A_u^c no están considerados. De acuerdo con la definición de σ -álgebra ambos deberían estar incluidos. En ejemplos como este, no es difícil resolver la situación, basta con tomar al conjunto \mathcal{C} y enriquecerlo con los conjuntos que le hagan falta para completar una σ -álgebra. Añadimos todas las uniones y complementos que hagan falta y listo. Este mecanismo sin embargo sirve solo para casos sencillos, donde el conjunto Ω es pequeño y podemos ser exhaustivos al momento de enlistar los elementos de una σ -álgebra. Para concluir el ejemplo; transcurridos los tres días, conocemos la historia completa, en este caso si parece apropiada la potencia como la σ -álgebra que describa la situación de nuestro experimento.

Fundamentos

Ejemplos

Es importante hacer notar sin embargo que en general la potencia de Ω es demasiado grande y puede dar lugar a problemas serios, asimismo cuando Ω es muy grande, no es posible ser explícitos al momento de construir una σ -álgebra, en cambio, debemos optar por algún argumento mas descriptivo que constructivo al momento de determinar la σ -álgebra de nuestro experimento.

Fundamentos

El primer inciso de la definición, básicamente se asegura de que la σ -álgebra sea una colección no vacía. De hecho podríamos sustituirlo por el requerimiento de que \mathcal{F} sea una colección no vacía. Notemos además que decir que $\omega \in \Omega$ es equivalente a “no decir nada”, una afirmación así no nos provee de información adicional.

En cambio, decir que $\omega \in A^c$ es un tipo de negación, negamos el hecho de que A haya ocurrido.

Asimismo, podemos parafrasear $\omega \in A \cup B$ diciendo que “ocurrió A u ocurrió B ”. De manera parecida si $\omega \in A \cap B$, entonces ocurrió A y ocurrió B .

Fundamentos

En general, podemos concebir a una σ -álgebra como un mecanismo para ordenar la información contenida en el espacio muestral. Por lo pronto, el primer inciso en la definición de σ -álgebra garantiza que Ω es siempre un elemento de la misma, el segundo inciso implica que la σ -álgebra es cerrada bajo complementos, es decir; si sabemos que algo ocurrió también debemos saber si es que esto mismo, no ocurrió. El tercer inciso es el mas controvertido. La axiomática de la probabilidad que estudiaremos es la propuesta por Kolmogorov y no esta exenta de controversia. Hay probabilistas que afirman que una teoría de probabilidad debería considerar únicamente la cerradura bajo uniones finitas y no extenderlo al caso numerable. Lo cierto es que este último inciso es de vital importancia en la construcción de eventos que tienen que ver con algún comportamiento límite. No abundaremos más al respecto dado que es un tema correspondiente a un curso avanzado de probabilidad. Baste notar que el inciso tres implica no solo que una σ -álgebra es cerrada bajo uniones numerables, también es cerrada bajo uniones finitas.

Para probar la cerradura bajo uniones finitas antes observemos que de los dos primeros incisos se concluye que $\emptyset \in \Omega$. De este modo, si A y B son dos eventos en \mathcal{F} entonces $A \cup B = \cup_n A_n$ con $A = A_1$ y $B = A_2$, mientras que $A_i = \emptyset$ para $i \geq 3$. Usando el inciso tres se sigue que $A \cup B = \cup A_i \in \mathcal{F}$. Ahora bien, si $\{A_n\}$ es una colección numerable de eventos en \mathcal{F} entonces $\{A_n^c\}$ también es una colección numerable en \mathcal{F} . Usando de manera conjunta el inciso 2 y el inciso 3 y echando mano de la ley de Morgan tenemos que $\cap_n A_n = (\cup_n A_n^c)^c \in \mathcal{F}$. Esto es, una σ -álgebra también es cerrada bajo intersecciones numerables. Con un argumento similar al que usamos para probar la cerradura bajo uniones finitas, podemos probar la cerradura bajo intersecciones finitas.

Fundamentos

Podemos continuar con propiedades del estilo de las que acabamos de exponer, dejemos algunas para los ejercicios. Lo importante que debemos recordar es que si bien una σ -álgebra es cerrada bajo uniones e intersecciones numerables, no lo es bajo uniones o intersecciones arbitrarias.

Al par (Ω, \mathcal{F}) , se le conoce como **espacio medible**. En un espacio medible es que tiene sentido nuestra siguiente definición:

Medida de probabilidad

Dado un espacio medible (Ω, \mathcal{F}) , decimos que una función $P : \mathcal{F} \rightarrow [0, 1]$ es una **medida de probabilidad** si:

- 1 $P(\Omega) = 1$
- 2 Si $\{A_n\}$ es una sucesión de eventos disjuntos, entonces $P(\cup_n A_n) = \sum_n P(A_n)$.

Fundamentos

A la terna (Ω, \mathcal{F}, P) se le conoce como **espacio de probabilidad**. Observemos el importantísimo papel de la σ -álgebra, no sólo es útil para definir cierto orden a la información que podemos extraer en un experimento aleatorio, también funciona como el dominio de una medida de probabilidad. Debemos ser enfáticos en esto; seremos capaces de calcular probabilidades sobre elementos \mathcal{F} y nada mas. Si $A \subset \Omega$ pero $A \notin \mathcal{F}$ entonces P es incapaz de trabajar con A . Algunas propiedades de la probabilidad vienen bien en este momento:

$$P(\emptyset) = 0$$

Efectivamente, consideremos la sucesión $\{A_n\}$ donde $A_n = \emptyset$ para todo n . Evidentemente esta es una sucesión de eventos disjuntos, por lo tanto

$$P(\emptyset) = P(\cup_n A_n) = \sum_n P(A_n) = P(\emptyset) + \sum_{n \geq 2} P(A_n)$$

Fundamentos

De donde se sigue que:

$$0 = \sum_{n \geq 2} P(A_n) = P(\cup_{n \geq 2} A_n) = P(\emptyset)$$

Si A y B son dos elementos disjuntos de \mathcal{F} entonces:

$$P(A \cup B) = P(A) + P(B)$$

Básicamente estamos afirmando que una medida de probabilidad no solo es aditiva ante uniones numerables disjuntas, también lo es ante uniones finitas disjuntas. Para probar este resultado observemos que

$$A \cup B = \cup_n A_n$$

Donde $A_1 = A$, $A_2 = B$ y $A_n = \emptyset$ para $n \geq 3$. Claramente $\{A_n\}$ es una sucesión de eventos disjuntos.

Fundamentos

Por lo tanto

$$\begin{aligned}P(A \cup B) &= P(\cup_n A_n) \\&= \sum_n P(A_n) \\&= P(A) + P(B) + \sum_{n \geq 3} P(A_n) \\&= P(A) + P(B)\end{aligned}$$

Este resultado, nos conduce a la siguiente afirmación:

Si $A \in \mathcal{F}$ entonces:

$$P(A^c) = 1 - P(A)$$

Este resultado se sigue de que:

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

Fundamentos

Para dos eventos A y B tales que $A \subset B$ se tiene que:

$$P(A) \leq P(B)$$

Efectivamente; notemos que $B = A \cup (B \setminus A)$ por lo tanto:

$$P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A) \geq P(A)$$

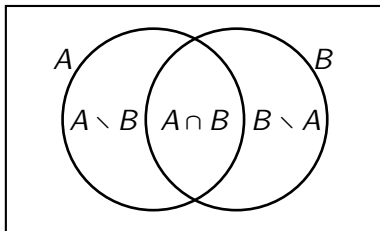
La última igualdad se debe al hecho de que $P(B \setminus A) \geq 0$ Finalmente, una de las relaciones mas útiles:

Dados los eventos A y B :

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

Fundamentos

En la igualdad anterior estamos tomando eventos A y B arbitrarios, no necesariamente disjuntos, de ahí la necesidad de restar la probabilidad de la intersección.



Observemos primero que $A \cup B = A \cup (B \setminus A)$ y además $B = (A \cap B) \cup (B \setminus A)$ por lo tanto

Fundamentos

$$P(A \cup B) = P(A) + P(B \setminus A)$$

y

$$P(B) = P(A \cap B) + P(B \setminus A)$$

Usando estas dos igualdades obtenemos la fórmula propuesta para la unión.

Retomemos el ejercicio que consiste en el lanzamiento de dos dados. Sabemos que

$$\begin{aligned}\Omega = \{ & (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6) \\ & (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6) \\ & (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6) \\ & (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6) \\ & (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6) \\ & (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6) \}\end{aligned}$$

Fundamentos

Si ambos dados son justos, si ninguno de ellos esta cargado y todas las caras son igualmente probables entonces cada cara tiene un sexto de probabilidad de aparecer. Derivado de Ω deducimos que cada combinación en el lanzamiento de los dados tiene $1/36$ de probabilidad de ocurrir.

Retomando los ejemplos anteriores, si

A = evento de obtener un 3 en el primer dado

entonces $P(A) = 6/36 = 1/6$. De manera análoga, si

B = evento de obtener un 3 en el segundo dado

entonces $P(B) = 1/6$.

Fundamentos

A su vez:

$$A \cap B = \text{Evento de obtener 3 en ambos dados} = \{(3, 3)\}$$

por lo cual $P(A \cap B) = 1/36$.

Con estos elementos, calculamos la probabilidad de:

$$A \cup B = \text{Evento de obtener al menos un tres}$$

esto es:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36}$$

Esto lo verificamos al notar que:

$$A \cup B = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), \\ (1, 3), (2, 3), (4, 3), (5, 3), (6, 3)\}$$

Fundamentos

Un ejemplo un poco mas complicado. Recordemos que

$A \triangle B$ = Evento de obtener exactamente un tres

Sabemos que:

$$A \triangle B = (A \setminus B) \cup (B \setminus A)$$

pero ademas:

$$A = (A \cap B) \cup (A \setminus B)$$

con lo cual:

$$P(A \setminus B) = P(A) - P(A \cap B)$$

análogamente:

$$P(B \setminus A) = P(B) - P(A \cap B)$$

Por lo tanto:

$$P(A \triangle B) = P(A) + P(B) - 2P(A \cap B) = \frac{1}{6} + \frac{1}{6} - 2\frac{1}{36} = \frac{10}{36}$$

Fundamentos

Lo cual se verifica al ver que:

$$A \triangle B = \{(3, 1), (3, 2), (3, 4), (3, 5), (3, 6), \\ (1, 3), (2, 3), (4, 3), (5, 3), (6, 3)\}$$

Fundamentos

La siguiente tabla puede ser de utilidad:

Ω	Espacio Muestral. Evento seguro
\emptyset	Evento imposible
ω	Resultado del experimento aleatorio
A	A ocurre si $\omega \in A$
A^c	Negación de A . Si $\omega \in A^c$ entonces A no ocurrió
$A \cap B$	A y B
$A \cup B$	A o B
$A \setminus B$	A pero no B
$A \triangle B$	A o B pero no ambos.
$A \subset B$	Si A entonces B

Fundamentos

A continuación vemos algunos ejemplos adicionales. En el ejemplo del lanzamiento de dos dados, sea

A = evento de obtener un par en el primer dado

y

B = evento de obtener un par en el segundo dado

Tenemos entonces que

$$A = \{(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), \\ (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), \\ (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$$

Por lo tanto: $P(A) = 18/36 = 1/2$. Análogamente deducimos que $P(B) = 1/2$.

Fundamentos

De lo anterior sigue además que

A^c = evento de obtener un impar en el primer dado

y

B^c = evento de obtener un impar en el primer dado

y claro: $P(A^c) = 1 - P(A) = 1/2$ y $P(B^c) = 1 - P(B) = 1/2$ Siguiendo con la misma idea tenemos que:

$A \cap B$ = Obtenemos par en ambos dados

Es decir:

$$A \cap B = \{(2, 2), (2, 4), (2, 6), \\ (4, 2), (4, 4), (4, 6), \\ (6, 2), (6, 4), (6, 6)\}$$

Fundamentos

De donde $P(A \cap B) = 9/36 = 1/4$. Con esto podemos calcular la probabilidad del evento:

$A \cup B$ = Obtenemos par en alguno de los dados

en efecto:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$$

Otro evento interesante:

$A \setminus B = A \cap B^c$ = El dado uno es par y el dado 2 es impar

Tenemos que:

$$P(A \setminus B) = P(A) - P(A \cap B) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

Fundamentos

Esto ultimo lo verificamos al ver que:

$$A \setminus B = \{(2, 1), (2, 3), (2, 5), \\ (4, 1), (4, 3), (4, 5), \\ (6, 1), (6, 3), (6, 5)\}$$

Osea, $P(A \setminus B) = 9/36 = 1/4$. Siguiendo en la misma linea:

$A \triangle B$ = obtenemos exactamente un par

en este caso $P(A \triangle B) = P(A \setminus B) + P(B \setminus A) = 1/4 + 1/4 = 1/2$.

Finalmente, un ejemplo un poco mas complejo:

C = La suma de los dados es par

Fundamentos

Podemos seguir dos rutas para encontrar la probabilidad de C . Una es la de “fuerza bruta” y enlistar todos los elementos de Ω que cumplen con la condición. La otra es tratar de razonar la lógica del conjunto y expresarlo en términos de otros mas simples.

Sigamos la segunda ruta y notemos que para que la suma de los dados sea par, es necesario que se cumpla que ambos dados sean pares o bien que ambos dados sean impares. Esto es:

$$\begin{aligned}C &= (A \cap B) \cup (A^c \cap B^c) \\ &= (A \cap B) \cup (A \cup B)^c\end{aligned}$$

Adicionalmente, observemos que como $(A \cap B) \subset A \cup B$, entonces $A \cap B$ y $(A \cup B)^c$ son disjuntos.

Fundamentos

En consecuencia:

$$\begin{aligned}P(C) &= P((A \cap B) \cup (A \cup B)^c) \\&= P(A \cap B) + P((A \cup B)^c) \\&= P(A \cap B) + (1 - P(A \cup B)) \\&= \frac{1}{4} + (1 - \frac{3}{4}) \\&= \frac{1}{2}\end{aligned}$$

Esto lo verificamos usando la “fuerza bruta” y enlistamos todos los pares cuya suma es par donde verificamos que: $P(C) = 18/36 = 1/2$.

$$C = \{(2, 2), (4, 2), (6, 2), (1, 1), (3, 1), (5, 1), \\(2, 4), (4, 4), (6, 4), (1, 3), (3, 3), (5, 3), \\(2, 6), (4, 6), (6, 6), (1, 5), (3, 5), (5, 5)\}$$

Combinatoria

Frecuentemente el cálculo de probabilidades involucra el conteo de elementos en un conjunto. Sobre todo cuando consideramos que los resultados de un posible experimento son igualmente probables.

En un caso así, la estrategia usual a seguir para el cálculo de las probabilidad de un evento es la de estimar el número de casos favorables y el número de casos totales en el experimento aleatorio.

Por ejemplo en el caso del lanzamiento de dos dados, el número de casos totales es 36. Podemos obtener 36 parejas diferentes. sin embargo, responder la pregunta: ¿Cuál es la probabilidad de que la suma de los dados sea 7? implica el conteo de todas aquellas parejas cuya suma es 7 a saber:

$$A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$$

de aquí, es fácil ver que

$$P(A) = \frac{6}{36} = \frac{1}{6}$$

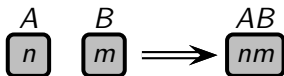
Combinatoria

Sin embargo, no siempre es fácil, enlistar todos los posibles resultados de un evento. Es aquí donde conviene analizar de mejor manera la situación y encontrar métodos que nos faciliten el conteo de elementos de un conjunto.

Principio básico de conteo.

Supongamos que dos experimentos son realizados. Si el experimento 1 puede resultar en una de n formas posibles y para cada resultado del experimento 1 existen m posibilidades para el experimento 2, entonces en conjunto existen nm resultados para los dos experimentos.

A menudo conviene ver esto con “casillas”. En una casilla ponemos el número de posibilidades al seleccionar un elemento de un conjunto A y en la otra el número de posibilidades al seleccionar un elemento de un conjunto B , si son n y m de forma correspondiente, entonces en total podremos hacer nm selecciones de los dos conjuntos.



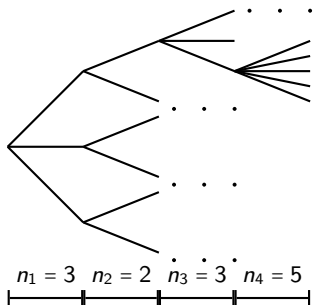
Combinatoria

En general, podemos pensar que tenemos m casillas. En la i -ésima casilla podemos depositar un elemento de un conjunto con n_i elementos.

Entonces el número de formas en que podemos 'llenar' estas casillas es igual a:

$$n_1 \cdot n_2 \cdots n_m$$

El siguiente diagrama esquematiza el caso para cuando tenemos solo 4 casillas, el caso general es similar.



Cada casilla puede tener uno elemento seleccionado de un conjunto con n_i elementos. En total tendremos

$$3 \cdot 2 \cdot 3 \cdot 5 = 90$$

selecciones posibles.

Combinatoria

Ejemplo.. La cantidad total de placas formadas con tres dígitos y tres letras es igual a:

$$10 \cdot 10 \cdot 10 \cdot 26 \cdot 26 \cdot 26 = 17,576,000$$

En este ejemplo seleccionamos las primeras tres entradas de un conjunto con 10 elementos las últimas 3 entradas fueron seleccionadas de un conjunto con 26 entradas. Un dato importante es que podemos repetir tanto números como letras

Un caso particular es cuando la selección se hace siempre del mismo conjunto.

Ejemplo.. El número total de números telefónicos de 8 dígitos es igual a:

$$10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 100,000,000$$

números.

Combinatoria

Este ejemplo, aunque simple de más tal vez sirve para ejemplificar dos cosas importantes. Por un lado al igual que en el caso de las placas, también en este caso es válido repetir números. Además el orden es importante. Los números:

$$55 - 21 - 28 - 35$$

y

$$55 - 82 - 12 - 35$$

son números diferentes, comparten la misma selección de números, sin embargo la selección no está en el mismo orden, por lo tanto son números diferentes. Adicionalmente, al realizar estas selecciones, las repeticiones están permitidas, el 5 aparece tres veces, mientras que el 2 aparece dos veces.

Combinatoria

El ejemplo anterior es un ejemplo claro de una **selección con orden y reemplazo** a partir de un conjunto con n elementos. En general cuando seleccionamos m elementos con orden y reemplazo de un conjunto con n elementos, el número total de selecciones posibles es igual a n^m .

Ejemplo. ¿Cuál es el número total de subconjuntos de un conjunto con k elementos?

Existen varias formas de responder esta pregunta, aquí ofrecemos una alternativa: supongamos establecemos un orden en los elementos del conjunto y formamos un vector:

$$(x_1, x_2, \dots, x_k)$$

donde $x_i = 1$ si el i -ésimo elemento es seleccionado y $x_i = 0$ si el i -ésimo elemento no es seleccionado.

Combinatoria

Entonces por ejemplo el vector compuesto por puros ceros significa que no hemos seleccionado ningún elemento, lo cual sería equivalente al conjunto vacío. Mientras que el vector compuesto por unos es que hemos seleccionado todos los elementos del conjunto, hemos seleccionado al conjunto original. Supongamos por un momento que el conjunto tiene 5 elementos. El vector:

$$(1, 1, 0, 0, 0)$$

indica que hemos seleccionado el subconjunto formado por los dos primeros elementos del conjunto original, mientras que el vector:

$$(0, 1, 1, 0, 0)$$

esta asociado al conjunto compuesto por el segundo y tercer elementos del conjunto original.

Combinatoria

Así el problema de contar el número de subconjuntos se traduce a encontrar el número posible de vectores

$$(x_1, x_2, \dots, x_k)$$

con $x_i \in \{0, 1\}$. Como cada x_i es seleccionado de un conjunto con dos elementos, entonces el número total de este tipo de vectores es igual a 2^k .

Ejemplo. Supongamos que tenemos el conjunto $A = \{1, 2, 3\}$. ¿De cuantas formas podemos ordenar al conjunto A ?

El conjunto A es lo suficientemente pequeño como para enlistar todos los posibles ordenamientos:

$$\begin{array}{lll} \{1, 2, 3\} & \{2, 1, 3\} & \{3, 2, 1\} \\ \{1, 3, 2\} & \{2, 3, 1\} & \{3, 1, 2\} \end{array}$$

Combinatoria

El conjunto A es lo suficientemente pequeño como para que estemos seguros de haber enlistado todos los posibles ordenamientos con lo cual afirmamos que el numero total de ordenamientos del conjunto A es igual a 6 . Sin embargo para un conjunto mas grande, esta estrategia quizás no sea la mas indicada. ¿Cómo llegamos a la misma conclusión sin necesidad de enlistar todos los posibles ordenamientos? En realidad estamos formando triadas del estilo:

$$(x_1, x_2, x_3)$$

donde $x_i \in A$. Con la peculiaridad de que x_1 puede ser cualquier elemento de A , mientras que x_2 , puede ser cualquier elemento de A excepto el que ya asignamos a x_1 . Asimismo x_3 puede ser cualquier elemento de A excepto los ya asignados a x_1 y x_2 . De este modo, x_1 puede seleccionarse de una de tres formas posibles, x_2 puede seleccionarse de una de dos formas posibles y x_3 solo puede seleccionarse de una única forma. En total podremos ordenar A de:

$$3! = 3 \cdot 2 \cdot 1 = 6$$

formas

Combinatoria

En general un conjunto con n elementos se puede ordenar de $n! = n \cdot (n-1) \cdot (n-2) \cdots 1$ formas posibles. El siguiente ejemplo se resuelve con una pequeña variación del argumento que hemos usado hasta este momento:

Ejemplo. ¿Cuántas palabras de cuatro letras podemos formar? ¿Cuántas si las letras no pueden repetirse?

Para resolver este problema primero debemos notar que una palabra de 4 letras será cualquier sucesión de cuatro letras, el resultado no tiene que ser necesariamente una palabra del diccionario. Así por ejemplo 'abcd' es una 'palabra'. En realidad buscamos el número de vectores del tipo.

$$(x_1, x_2, x_3, x_4)$$

donde cada x_i es alguna de las 26 letras del abecedario. En este caso el número de palabras será 26^4

Combinatoria

¿Que pasa si no podemos repetir letras? Obviamente en este caso palabras como *aabc* no son válidas. Ahora al componer los vectores

$$(x_1, x_2, x_3, x_4)$$

x_1 podrá seleccionarse de una de 26 formas posibles, x_2 de 1 de 25 posibles, x_3 de una de 24 formas posibles y x_4 de una de 23 formas posibles, eso con la finalidad de no repetir letras. En total tendremos

$$26 \cdot 25 \cdot 24 \cdot 23 = \frac{26!}{22!} = \frac{26!}{(26-4)!} = 358,800$$

formas de armar una palabra de cuatro letras que no repita letras.

Combinatoria

En general si queremos formar palabras de m letras tendremos un total de

$$26^m$$

palabras si es que podemos repetir letras.

Si no podemos repetir letras, entonces podremos formar un total de total de

$$\frac{26!}{(26 - m)!}$$

palabras. Observemos que este último caso solo tiene sentido solo si $m \leq 26$.

Por último, cuando permitimos repetición de letras decimos que hacemos una **selección con orden y con reemplazo**. Lo de reemplazo, creo que es claro, lo de orden también si consideramos que $aabc$ y $abac$ por ejemplo, son palabras diferentes aunque sean palabras formadas por las mismas letras.

Combinatoria

Por el contrario, cuando la repetición de letras no esta permitida entonces decimos que hicimos una **selección con orden y sin reemplazo**.

Nuevamente, creo que es claro la parte de 'sin reemplazo', pero también el orden si consideramos que las palabras *abcd* y *bacd* son diferentes a pesar de tener las mismas letras.

En general; si tenemos un conjunto con n elementos el número total de selecciones de m elementos, **con orden y con reemplazo** es igual a

$$n^m$$

en contraste; el número total de selecciones **con orden y sin reemplazo** es igual a:

$$\frac{n!}{(n-m)!}$$

Observemos que en este caso es necesario que $m \leq n$.

Combinatoria

Ejemplo. Tengo 6 discos de rock, 4 de música clásica y 3 de música regional. ¿De cuantas formas puedo ordenar mis discos de modo que los discos del mismo género estén en posiciones contiguas?

Los discos de rock pueden ordenarse de $6!$ formas, los de música clásica pueden ordenarse de $4!$ formas y los de música regional de $3!$ formas, por lo tanto, puedo ordenar mis discos de

$$6!4!3!$$

formas ¿cierto?

Falta un detalle. . . también podemos intercambiar los géneros! Podemos colocar primero los de rock, luego los regionales y finalmente los de música clásica. Otra posibilidad es poner primero los de música clásica, luego los regionales y al final los de rock. . . es decir falta contar las formas en que podemos ordenar los géneros. Esto lo hacemos de $3!$ formas. Por lo tanto el número que en realidad estoy buscando es:

$$3!6!4!3!$$

Combinatoria

Ejemplo. En un club con 25 miembros, deben ser seleccionados el presidente, el secretario y el tesorero. Cualquier miembro del club puede ser candidato a alguna de estas posiciones. ¿De cuantas formas podemos realizar la selección?

Tenemos 25 posibilidades para el puesto de presidente. Sin embargo una vez seleccionado este, solo tenemos 24 posibilidades para el puesto de secretario y una vez seleccionados presidente y secretario solo quedan 23 posibilidades para el tesorero, por lo tanto tenemos:

$$25 \cdot 24 \cdot 23 = \frac{25!}{22!} = 13,800$$

formas de seleccionar presidente, secretario y tesorero.

Ejemplo. Supongamos que queremos formar un equipo de super héroes. Están disponibles, Hulk, Ironman, Spiderman, Wolverine y Deadpool.



de estos super héroes queremos formar un equipo de solo tres super héroes. ¿Cuántos equipos de tres podemos formar?

Una primera aproximación sería pensar que para la primera selección tenemos 5 posibilidades, para la segunda tendríamos sólo 4 y para la tercera solo 2. Por lo tanto el número total de equipos que podemos formar es igual a:

$$5 \cdot 4 \cdot 3 = \frac{5!}{(5-3)!} = 60$$

¿Es correcto?

Combinatoria

Hemos hecho una selección de tres elementos tomados de un conjunto de cinco elementos. Sin embargo hemos realizado una selección sin reemplazo y con orden. Esta claro que la selección debe ser sin reemplazo, para formar un equipo, no podemos repetir a ningún superheroe, sin embargo ¿que pasa con el orden?

Supongamos que seleccionamos a Hulk a Ironman y a Wolverine. Notemos que en la forma en que realizamos la cuenta, dado que lo hicimos con orden, hemos considerado los siguientes equipos como si fueran diferentes:



Combinatoria

Observemos sin embargo que los seis equipos que enlistamos son en realidad el mismo. Son solo los 6 posibles ordenamientos de los 3 miembros seleccionados, es decir $3!$. Para evitar estas repeticiones debemos dividir nuestra estimación original entre $3!$, de modo que el número total de equipos es en realidad:

$$\frac{5!}{3!(5-3)!} = 10$$

Este ultimo cociente es tan importante que tenemos una notación especial para el mismo:

$$\binom{5}{3} = \frac{5!}{3!(5-3)!}$$

Combinatoria

En general para $m \leq n$:

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}$$

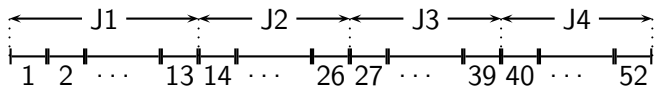
y se lee como las **combinaciones de n en m** . Del ejemplo que estudiamos, deberíamos llegar a la conclusión de que este cociente son las formas en las que podemos construir un subconjunto de m elementos a partir de un conjunto con n elementos **sin orden y sin reemplazo**.

Ejemplo. Ahora vamos a un ejemplo un poco mas retador. Supongamos que tenemos un mazo de 52 cartas que una vez que es barajado cuidadosamente es repartido entre cuatro jugadores. Cada jugador tendrá 13 cartas que le fueron repartidas aleatoriamente. Nos interesa conocer cual es la probabilidad de que cada jugador haya recibido exactamente un As.

Combinatoria

Dado que el mazo fue barajado con cuidado, podemos suponer que cualquier combinación posible de las 52 cartas es igualmente probable. En vista de ello calcularemos la probabilidad solicitada dividiendo el número de casos favorables entre el número de casos totales. El problema entonces se traduce en responder dos preguntas: ¿Cuál es el número de casos favorables? y ¿Cuál es el número de casos totales?

Comencemos por calcular el número de casos totales. ¿Que es lo que debemos contar? Notemos que en realidad lo que nos interesa es determinar la posición en que serán distribuidos los ases. Para ello consideremos el siguiente diagrama:



Combinatoria

Cada casilla representa una posible carta asignada. Las primeras 13 casillas son las correspondientes al jugador uno, las casillas 14 a 26 son las cartas asignadas al jugador dos, de la 27 a la 39 las asignadas al jugador 3 y de la 40 a la 52 son las cartas repartidas al jugador 4.

Esta repartición es completamente aleatoria y en principio una casilla cualquiera puede 'recibir' cualquier carta del mazo.

A nosotros lo que realmente nos interesa es la repartición de los ases, por lo tanto estamos interesados en la manera que estos ases son repartidos entre las 52 casillas. Una posible repartición sería la siguiente:



┌─ C5 ─┐



┌─ C14 ─┐┌─ C15 ─┐



┌─ C52 ─┐

Combinatoria

Notemos que esta asignación, si bien es posible, no es 'favorable'. El As de corazones fue asignado al jugador 1, el de tréboles al jugador 4 pero el as de espadas y el de diamantes fueron asignados al jugador 2, mientras que el jugador 3 no tiene ningún As

En cambio la asignación:



┌─ C5 ─┐



┌─ C14 ─┐



┌─ C27 ─┐



┌─ C52 ─┐

es una asignación 'favorable' ya que cada jugador tiene asignado exactamente un As.

En conclusión, para contar el número de casos totales, lo único que nos interesa es distribuir los cuatro ases en alguna de las 52 casillas posibles. Esto significa seleccionar 4 casillas de las 52 posibles, si importar a que jugador corresponden. Tenemos por tanto que el número de casos totales es igual a:

$$\binom{52}{4}$$

Para obtener el número de casos favorables, tenemos que limitarnos a aquellos en los cuales ya hay un as asignado en las posiciones del 1 al 13, otro en las posiciones del 14 al 26, uno más en las posiciones del 27 al 39 y el último en las posiciones del 40 al 52. Esto garantiza que cada jugador tiene asignado exactamente un As. Sin embargo notemos que cualquier As puede ser asignado a una de trece casillas posibles, de modo que el número casos favorables será

$$13^4$$

Combinatoria

Finalmente la probabilidad que buscamos es:

$$\frac{13^4}{\binom{52}{4}} \approx 0.1055$$

Como subproducto, observemos que si esta es la probabilidad de que cada jugador tenga exactamente un As, entonces el complemento es el evento de que algún jugador tenga al menos dos ases. La probabilidad de este segundo evento sería entonces

$$1 - \frac{13^4}{\binom{52}{4}} \approx 0.8945$$

Combinatoria

Hemos deducido varias formas de contar posibles selecciones de un conjunto, la siguiente tabla muestra un resumen de lo conseguido hasta el momento. El número de selecciones de m elementos a partir de un conjunto con n elementos es igual a:

	Con orden	Sin orden
Con reemplazo	n^m	?
Sin reemplazo	$\frac{n!}{(n-m)!}$	$\binom{n}{m} = \frac{n!}{m!(n-m)!}$

Aún nos hace falta el caso con reemplazo y sin orden, por el momento lo dejaremos para más tarde. Antes algunas observaciones primero; en ambos casos en los cuales la selección se realiza sin reemplazo, es necesario que $m \leq n$.

Combinatoria

En el caso en el que realizamos una selección sin orden y sin reemplazo, en realidad lo que estamos haciendo es contar las formas en que podemos construir un subconjunto de m elementos a partir de un conjunto con n elementos. Pero más aún dado que

$$\binom{n}{m} = \frac{n!}{m!(n-m)!} = \binom{n}{n-m}$$

Deducimos que el número de subconjuntos con m elementos es igual al número de subconjuntos con $n - m$ elementos. Dicho de otro modo; al hacer una selección sin orden y sin reemplazo, implícitamente estamos contando el número de formas en que podemos dividir al conjunto original en un par de conjuntos; uno con m elementos y otro con $n - m$ elementos. El siguiente ejemplo nos da una generalización de esta conclusión.

Combinatoria

Previamente dimos un argumento satisfactorio para calcular el número de subconjunto de un subconjunto con n elementos. Ahora damos un argumento alternativo pero usando combinaciones. Esto lo logramos contando uno por uno, el numero de subconjuntos con cero elementos, el numero de conjuntos con un elementos, el número de conjuntos con dos elementos etc. La suma de todos ellos es el número total de subconjuntos, esto es:

$$\sum_{i=0}^n \binom{n}{i}$$

Pero recordando el teorema del binomio:

$$(a + b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}$$

tenemos que

$$\sum_{i=0}^n \binom{n}{i} = \sum_{i=0}^n \binom{n}{i} 1^i 1^{n-i} = (1 + 1)^n = 2^n$$

Combinatoria

Ejemplo. ¿De cuantas formas podemos ordenar las letras de la palabra *PEPPER*?

Es claro que en esta palabra no es posible distinguir entre una P y otra o entre una E y otra. Sin embargo como primer paso, hagámos identificables a cada una de las letras:

$$PEPPER = P_1 E_1 P_2 P_3 E_2 R$$

En este caso tenemos 6! permutaciones de estas letras. Sin embargo observemos que las combinaciones:

$$PEPPER = P_2 E_2 P_1 P_3 E_1 R$$

y

$$PEPPER = P_3 E_1 P_1 P_2 E_2 R$$

son en realidad la misma combinación.

Combinatoria

Esto último es sólo un ejemplo, otro ejemplo podría ser el siguiente, las combinaciones:

$$PPPEER = P_1 P_2 P_3 E_1 E_2 R$$

y

$$PPPEER = P_2 P_1 P_3 E_2 E_1 R$$

son en realidad la misma combinación. Esto es al contar los ordenamientos con $6!$ estamos contando muchos casos repetidos. ¿Como calculamos los casos repetidos? Observemos que una vez que determinamos una combinación si intercambiamos de lugar una P con otra el resultado es la misma 'palabra'. Lo mismo ocurre con una E , si la cambio de lugar con otra E la palabra resultante no cambia. Solo tenemos una R , pero si hubiera mas de una, la conclusión sería idéntica.

Combinatoria

Es así que una vez que determinamos un ordenamiento de las letras, cualquier permutación de P s da lugar a la misma palabra, del mismo modo que cualquier permutación de E da lugar a la misma palabra y como dijimos previamente, lo mismo ocurriría de existir más de un R .

Ahora bien, el número total de ordenamientos que podemos realizar con las P s es igual a $3!$, de las E s es $2!$ y de las R s es $1!$. De este modo el número de repeticiones es:

$$3! \cdot 2! \cdot 1!$$

Por lo tanto, el número de ordenamientos que podemos realizar de la palabra *PEPPER* es igual a

$$\frac{6!}{3! \cdot 2! \cdot 1!}$$

Combinatoria

Este ejemplo es un caso particular del siguiente problema: si tenemos un conjunto con n elementos, ¿de cuantas formas podemos construir una partición al conjunto en k subconjuntos cada uno con n_i , $i = 1, \dots, k$ elementos? Siguiendo misma lógica del ejemplo anterior, podríamos realizar esta partición de

$$\frac{n!}{n_1!n_2!\cdots n_k!}$$

con $n_1 + n_2 + \cdots + n_k = n$. A este cociente se le conoce como **coeficiente multinomial** el cual es una generalización del coeficiente binomial:

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}$$

Ejemplo. El sorteo del Melate bolas numeradas del 1 al 56 son introducidas en una urna. Aleatoriamente son seleccionadas 6 de ellas, los números 'naturales' mas un 'adicional'. Los participantes compran un boleto en el cual hacen su propia selección de 6 números. Existen varias formas de ganar en el sorteo. Obtenemos el premio mayor si acertamos los 6 números naturales. El siguiente premio se consigue si acertamos exactamente 5 de los números naturales mas el adicional, el siguiente lo conseguimos si solo acertamos exactamente 5 naturales. La siguiente tabla resume las formas en que podemos ganar y la probabilidad de cada una.

Combinatoria

Categoría	Chances de ganar
6 Naturales	1 en 32,468,436
5 Naturales + Adicional	1 en 5,411,406
5 Naturales	1 en 110,437
4 Naturales + Adicional	1 en 44,175
4 Naturales	1 en 1,841
3 Naturales + Adicional	1 en 1,380
3 Naturales	1 en 88
2 Naturales + Adicional	1 en 117
2 Naturales	1 en 10

Combinatoria

¿Como verificamos que estas probabilidades son correctas? Empecemos por determinar el numero total de casos. Como la selección que hace un participante es de 6 números, debemos hacer una selección de 6 a partir de los 56 disponibles. El orden no importa, solo es relevante si acertamos o no los números. Por lo tanto el número total de casos es:

$$\binom{56}{5} = 32,468,436$$

¿Como ganamos el premio mayor? Este es el caso más fácil Sólo ganamos si acertamos exactamente los 6 números naturales, por lo tanto solo existe un caso favorable. Por lo tanto, la probabilidad de ganar el premio mayor es igual a

$$\frac{1}{\binom{56}{6}}$$

Combinatoria

El siguiente premio la ganamos si acertamos exactamente 5 naturales mas el adicional. ¿Como lo logramos? Pues debemos 'forzar' a que 5 de los números que seleccionamos provengan de los 6 naturales que fueron seleccionados. Es decir seleccionamos 5 de los 6 números naturales ganadores. El número de selecciones de este tipo es:

$$\binom{6}{5}$$

El número restante debe coincidir con el adicional. Como solo existe un adicional, esto se logra de una única manera. Por lo tanto la probabilidad de atinar 5 naturales mas el adicional es:

$$\frac{\binom{6}{5} \cdot 1}{\binom{56}{6}}$$

Combinatoria

El siguiente premio se obtiene si acertamos exactamente 5 naturales pero fallamos en el adicional. El acierto de los 5 naturales como ya vimos se puede obtener de

$$\binom{6}{5}$$

formas. ¿Que pasa con el que no acertamos? No puede ser ninguno de los naturales ganadores y tampoco puede ser el adicional, por lo tanto de los 56 naturales existen 49 perdedores. El número que en definitiva no acertamos debe ser uno de esos 49 perdedores. de modo que en este caso el número de casos favorables es:

$$\binom{6}{5} \cdot 49$$

y la probabilidad de obtener exactamente 5 naturales es:

$$\frac{\binom{6}{5} \cdot 49}{\binom{56}{6}}$$

Combinatoria

El siguiente premio lo ganamos si acertamos 4 naturales mas el adicional. Siguiendo la misma lógica, debemos entonces seleccionar 4 de los 6 naturales ganadores. Las formas de hacer esto es:

$$\binom{6}{4}$$

Pero ahora tenemos dos números restantes. Uno de ellos es el adicional y solo hay una forma de seleccionarlo. El número restante debe ser uno de los 49 perdedores, de modo que el número de casos favorables será:

$$\binom{6}{4} \cdot 49 \cdot 1$$

Finalmente, la probabilidad de obtener 4 naturales mas el adicional será

$$\frac{\binom{6}{4} \cdot 49 \cdot 1}{\binom{56}{6}}$$

Combinatoria

El ultimo caso que calcularemos es cuando acertamos exactamente 4 naturales. En este caso, nuevamente seleccionamos cuatro de los 6 naturales ganadores. Sabemos que esto podemos hacerlos de

$$\binom{6}{4}$$

formas. Nuevamente restan dos números, pero en este caso debe omitirse el adicional, debemos seleccionarlos de los 49 perdedores, esto podemos hacerlo de

$$\binom{49}{2}$$

formas. De modo que la probabilidad de obtener exactamente 4 naturales es:

$$\frac{\binom{6}{4} \cdot \binom{49}{2}}{\binom{56}{6}}$$

Combinatoria

El resto de los casos ganadores se obtienen con:

$$\text{Probabilidad de 3 naturales + adicional} = \frac{\binom{6}{3} \cdot \binom{49}{2} \cdot 1}{\binom{56}{6}}$$

$$\text{Probabilidad de 3 naturales} = \frac{\binom{6}{3} \cdot \binom{49}{3}}{\binom{56}{6}}$$

$$\text{Probabilidad de 2 naturales + adicional} = \frac{\binom{6}{2} \cdot \binom{49}{3} \cdot 1}{\binom{56}{6}}$$

$$\text{Probabilidad de 2 naturales} = \frac{\binom{6}{2} \cdot \binom{49}{4}}{\binom{56}{6}}$$

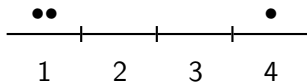
Combinatoria

Finalmente, para concluir el tema de combinatoria, determinaremos el número de selecciones con reemplazo y sin orden, de tamaño m a partir de un conjunto con n elementos.

Primero, notemos que dado que la selección se realiza con reemplazo, en este caso es posible que $m > n$ aunque no es necesario.

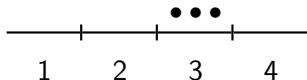
Pensemos en un ejemplo muy concreto. Supongamos que tenemos n bolas numeradas (del 1 al n claro esta). Seleccionamos una bola, tomamos registro de su número y la devolvemos con las demás. Esto lo repetimos m veces. Como devolvimos la bola en cada ocasión, pueden existir repeticiones sin ningún problema.

Este registro si tenemos cuatro bolas y hacemos tres selecciones, podría verse como sigue:

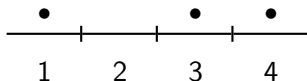


Combinatoria

El esquema previo ilustra el caso donde la bola uno fue seleccionada dos veces. El siguiente:



es el caso donde la bola tres fue seleccionada tres veces.



y este último esquematiza el caso cuando la bola uno, la tres y la cuatro son seleccionadas.

Combinatoria

En general podemos enlistar los 20 casos posibles como:

- | | |
|---------------|-----------------|
| 1. • • • | 11. • • • |
| 2. • • • | 12. • • • |
| 3. • • • | 13. • • • |
| 4. • • • | 14. • • • |
| 5. • • • | 15. • • • |
| 6. • • • | 16. • • • |
| 7. • • • | 17. • • • |
| 8. • • • | 18. • • • |
| 9. • • • | 19. • • • |
| 10. • • • | 20. • • • |

Combinatoria

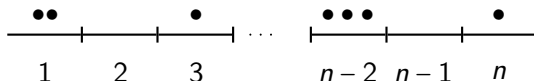
Notemos que este es el “registro completo”. En la lista anterior, el caso 4 coincide con el caso donde seleccionábamos dos veces la bola 1 y una vez la bola 4. El inciso 9 coincide con el caso donde seleccionábamos la bola 3, tres veces. El caso 19 es el caso donde seleccionábamos la bola 1, la 3 y la 4.

Ahora bien, observemos del listado anterior, que en realidad lo único que hicimos fue ordenar las bolas y las rayas. En total tenemos 6 objetos entre rayas y bolas, mismos que pueden ordenarse de $6!$ formas. Sin embargo, notemos que en cualquiera de los 20 arreglos podemos cambiar de lugar una bola por otra y el “registro” es el mismo. Lo mismo ocurre con las rayas, cambio una raya por otra y el “registro” es el mismo. Por lo tanto de los $6!$ posibles arreglos de bolas y rayas $3!3!$ están repetidos. En conclusión el número total de selecciones que podemos realizar es igual a:

$$\frac{6!}{3!3!} = \binom{6}{3} = 20$$

Combinatoria

En general si tenemos un total de n bolas y hago m selecciones, un “registro típico” seria como sigue:



Donde en total tenemos m bolas repartidas en las n casillas. Cada separación es un “raya” tenemos un total de $n - 1$ rayas. Con lo cual tenemos un total de $n - 1 + m$ objetos entre bolas y rayas. Que pueden arreglarse sin orden de

$$\binom{n + m - 1}{m} = \frac{(n - 1 + m)!}{(n - 1)!m!} = \binom{n + m - 1}{n - 1}$$

formas.

Combinatoria

Con esto completamos la tabla que resume la forma de seleccionar m elementos a partir de un conjunto con n elementos:

	Con orden	Sin orden
Con reemplazo	n^m	$\binom{n+m-1}{m}$
Sin reemplazo	$\frac{n!}{(n-m)!}$	$\binom{n}{m} = \frac{n!}{m!(n-m)!}$

Recordemos que en los casos sin reemplazo es necesario que $m \leq n$.

Probabilidad Condicional

Empecemos este tema resolviendo (si podemos) un problema muy bonito, sencillo pero bonito...



El problema es muy sencillo de plantear; sabemos que los padres del rey tuvieron dos hijos, ¿cuál es la probabilidad de que el otro hijo sea su hermana?

Nada mas sencillo... si los padres del rey tuvieron dos hijos y el rey es uno de ellos, entonces el otro hijo es su hermana con probabilidad $1/2$. ¿Cierto?

Probabilidad Condicional

Hay que pensar el tema con un poco de cuidado, no lleguemos tan pronto a conclusiones. Antes de dar una respuesta definitiva, pensemos en un problema un poco mas común. Supongamos que tiramos dos dados y estamos interesados en la suma de las caras. Si deseamos la probabilidad de que la suma sea 5.

En este experimento sabemos que el espacio muestral esta compuesto por todas las posibles parejas de resultados obtenidos al lanzar los dados. En total tenemos 36 casos posibles. Los casos favorables serían los contenidos en el conjunto:

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

= La suma de las caras es igual a 5

De aqui se sigue facilmente que

$$P(A) = \frac{4}{36} = \frac{1}{9}$$

Probabilidad Condicional

El ejemplo anterior fue simple y la conclusión satisfactoria pero, ¿que pasa si incluimos información adicional? que tal por ejemplo si sabemos que el resultado del primer dado fue uno? ¿cómo cambia esto nuestra conclusión? Primero, notemos que el evento donde obtenemos uno en el primer dado, es el evento:

$$B = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)\}$$

Es decir, si sabemos que en el primer dado salió uno, estamos diciendo que el evento B ocurrió. Lo que no sabemos es cual de los elementos de B es el que apareció después de lanzar los dados. Y claro, la pregunta persiste ¿cual es la probabilidad de que la suma sea 5?

Dado que sabemos que B ocurrió, cualquier resultado que involucre a ambos dados debe estar en B . Para hacer énfasis en esto introducimos un nuevo concepto:

Probabilidad Condicional

Probabilidad Condicional

Definimos la **Probabilidad Condicional** de A dado B como:

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

La idea de probabilidad condicional resuelve la pregunta: ¿cuál es la probabilidad de A dado que conozco B ? El experimento sigue siendo aleatorio, pero ahora tenemos información adicional, ahora sabemos que B ocurrió, lo cual puede ser sumamente relevante.

En el ejercicio de los dados:

$$A \cap B = \{(1, 4)\}$$

En consecuencia:

$$P(A \cap B) = \frac{1}{36}$$

y

$$P(B) = \frac{6}{36} = \frac{1}{6}$$

Por lo tanto:

$$P(A | B) = \frac{1/36}{1/6} = 1/6$$

Observemos que el conocimiento de B afecta considerablemente la probabilidad de que A ocurra.

Con esta idea a la mano, podemos regresar al problema del rey. originalmente sabemos que los padres del rey tuvieron dos hijos, con lo cual podemos pensar en el espacio muestral:

$$\Omega = \{(B, B), (B, G), (G, B), (G, G)\}$$

Probabilidad condicional

Por otro lado sabemos que uno de los hijos es rey, eso ocurre solo si los padres del rey tuvieron al menos un hijo varón. Es decir, si ocurrió el evento:

$$B = \{(B, B), (B, G), (G, B)\}$$

por otro lado queremos saber, cuál es la probabilidad de que el otro de los hijos sea la hermana del rey. Es decir que sus padres hayan tenido al menos una niña, este es el evento:

$$A = \{(B, G), (G, B), (G, G)\}$$

Finalmente, lo que en realidad nos interesa es la probabilidad:

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{2/4}{3/4} = 2/3$$

Probabilidad Condicional

La probabilidad condicional cumple con muchas propiedades de utilidad, por lo pronto notemos que

$$P(B \mid B) = 1$$

Esto tiene todo el sentido del mundo, al momento de saber que B ocurrió, cualquier resultado ω del experimento aleatorio deb estar en B , de manera un tanto informal podemos decir que el espacio muestral original se “redujo”, dado B , ahora todo lo que consideramos posible debe estar en B .

Probabilidad Condicional

Esto nos conduce a los siguiente

Probabilidad Condicional

La probabilidad condicional $P(\cdot | B)$ es una probabilidad. Esto es:

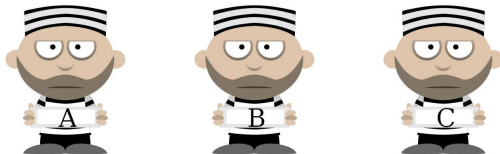
- 1 $P(\Omega | B) = 1$
- 2 Para todo evento A , $0 \leq P(A | B) \leq 1$
- 3 Si $\{A_n\}$ es una sucesión de eventos disjuntos entonces:

$$P(\cup_n A_n | B) = \sum_n P(A_n | B)$$

La prueba de estos tres incisos, quedará como tarea... :-)

Probabilidad Condicional

Va otro ejemplo interesante... El problema que veremos a continuación es tan popular que hasta nombre tiene; es conocido como “el dilema del prisionero”. Supongamos que tenemos tres prisioneros:



Dos de ellos serán liberados, por lo cual el prisionero A decide acercarse al guardia a preguntarle: *“Dime, ¿quién que no sea yo será liberado?”*, el guardia se rehúsa a contestar bajo el siguiente argumento: *“Mira, ahora mismo tienes $\frac{2}{3}$ de probabilidad de ser liberado, si te digo por ejemplo que B será liberado, entonces tu serás uno de los dos cuya destino sería incierto y en consecuencia tu probabilidad de ser liberado disminuiría a $\frac{1}{2}$.”* ¿Es correcto el razonamiento del guardia?

Probabilidad Condicional

Primero tratemos de razonar como lo hizo el guardia. Dado que de inicio afirmó que el prisionero A tenía $2/3$ de probabilidad de ser liberado, en realidad esta dando por hecho que el espacio muestral del experimento es:

$$\Omega = \{(A, B), (A, C), (B, C)\}$$

Donde las parejas en Ω son pares no ordenados. Luego supone que es igualmente probable la liberación de cualquiera de estas parejas. Su estimación inicial de $2/3$ en este contexto, es correcta.

Si segunda afirmación la escribiríamos como sigue:

$$P(A \text{ sea liberado} \mid \text{el guardia dijo que } B \text{ será liberado}) = \frac{1}{2}$$

Probabilidad Condicional

El evento “A es liberado” es el conjunto:

$$A \text{ es liberado} = \{(A, B), (A, C)\}$$

eso está claro, pero el evento: “el guardia dijo que B será liberado” ¿como lo escribimos?. Simplemente no hay forma dado el espacio muestral que seleccionó el guardia. Es necesario introducir la decision del guardia en la definición del espacio muestral. Que tal si componemos el espacio muestral con los siguientes resultados:

$$\omega_1 = \{A, B, \text{el guardia dice } B\}$$

$$\omega_2 = \{A, C, \text{el guardia dice } C\}$$

$$\omega_3 = \{B, C, \text{el guardia dice } B\}$$

$$\omega_4 = \{B, C, \text{el guardia dice } C\}$$

Con este nuevo espacio muestral, tenemos que:

$$E = A \text{ es liberado} = \{\omega_1, \omega_2\}$$

mientras que:

$$F = \text{El guardia dice B} = \{\omega_1, \omega_3\}$$

Ahora queda un poco mas claro como calcular:

$$P(E | F) = P(A \text{ sea liberado} | \text{el guardia dijo que B será liberado})$$

pero;

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

y tenemos que;

$$E \cap F = \{\omega_1\}$$

Probabilidad condicional

Antes de calcular $P(E \cap F)$ y $P(F)$ observemos lo siguiente:

$\{\omega_1\}$ = Evento de que A y B sean liberados

$\{\omega_2\}$ = Evento de que A y C sean liberados

$\{\omega_3, \omega_4\}$ = Evento de que A y B sean liberados

Esto se debe al hecho de que cuando A y B son liberados el guardia no tiene mas alternativa que revelar a B como el prisionero que será liberado, del mismo modo, cuando A y C son liberados el guardia no tiene opción, debe revelar a C como el prisionero a ser liberado. En cambio, cuando B y C son liberados, el guardia puede decidir revelar a B o revelar a C como el prisionero a ser liberado. Esto nos conduce a lo siguiente:

$$P(E \cap F) = P(\{\omega_1\}) = \frac{1}{3}$$

Probabilidad Condicional

¿Que podemos decir de $P(F)$?

$$P(F) = P(\{\omega_1, \omega_3\}) = P(\{\omega_1\}) + P(\{\omega_3\}) = \frac{1}{3} + P(\{\omega_3\})$$

¿que pasa con $P(\{\omega_3\})$? con los elementos que tenemos a la mano no tenemos forma clara de calcularlo, por lo pronto sabemos que;

$$\frac{1}{3} = P(\{\omega_3, \omega_4\}) = P(\{\omega_3\}) + P(\{\omega_4\})$$

en este caso una suposición natural sería suponer que

$$P(\{\omega_3\}) = P(\{\omega_4\}) = \frac{1}{6}$$

eso implicaría que si B y C son liberados el guardia decide decir que B o C serán liberados, “tirando una moneda” al aire, le es indistinto decir que B o que C sera liberado.

Probabilidad Condicional

En este caso:

$$P(F) = \frac{1}{3} + P(\{\omega_3\}) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}$$

y por lo tanto:

$$P(E | F) = \frac{P(E \cap F)}{P(F)} = \frac{1/3}{1/3 + 1/6} = \frac{2}{3}$$

Hay dos cosas muy importantes que debemos notar aquí, primero esta estimación surgió del supuesto de que ω_3 y ω_4 eran igualmente probables:

$$P(\{\omega_3\}) = P(\{\omega_4\}) = \frac{1}{6}$$

aquí dijimos que el guardia tiraba una moneda para decidir si decía B o decía C cuando B y C eran liberados. Pero lo que es cierto es que

$$\frac{1}{3} = P(\{\omega_3, \omega_4\}) = P(\{\omega_3\}) + P(\{\omega_4\})$$

Probabilidad Condicional

y esto se cumple siempre y cuando

$$0 \leq P(\{\omega_3\}) \leq \frac{1}{3}$$

y

$$P(\omega_4) = \frac{1}{3} - P(\{\omega_3\})$$

Esto nos conduciría a conclusiones muy diferentes en función del valor seleccionado, para $P(\{\omega_3\})$, por ejemplo si $P(\{\omega_3\}) = 1/3$ entonces:

$$P(E | F) = \frac{P(E \cap F)}{P(F)} = \frac{1/3}{1/3 + 1/3} = \frac{1}{2}$$

Esto confirmaría el argumento del guardia, sin embargo aunque es matemáticamente correcto, parece increíble que la decisión del guardia cambie de algún modo la probabilidades del prisionero A.

En contraste, si $P(\{\omega_3\}) = 0$ entonces:

$$P(E \mid F) = 1$$

lo cual aunque matemáticamente correcto, es en definitiva un sin sentido!
El segundo punto importante que debemos notar viene de nuestra elección original para $P(\{\omega_3\}) = 1/6$. En este caso concluimos que

$$P(E \mid F) = \frac{2}{3}$$

Es decir, en este caso la decisión del guardia no afecta para nada las probabilidades del prisionero A de ser liberado... esto en definitiva tiene mucho más sentido, aunque no sea la única solución posible para este problema. En este caso, la información adicional es irrelevante. Lo cual nos lleva al siguiente concepto:

Probabilidad Condicional

Eventos independientes

Decimos que dos eventos A y B son **independientes** si

$$P(A \cap B) = P(A)P(B)$$

Esto claro esta nos conduce inmediatamente al siguiente resultado; si B es un evento tal que $P(B) > 0$ y A y B son independientes entonces:

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

es de aquí de donde se comprende mejor la idea de “independencia”, al ser A y B independientes, el conocimiento de B no afecta para nada la probabilidad de A .

Probabilidad Condicional

Ejemplo. Un ejemplo sencillo de independencia lo encontramos en el lanzamiento de dos dados. Previamente calculamos la probabilidad que la suma de las caras fuera 5, esta probabilidad resulta ser igual a $1/9$, mientras que la misma probabilidad pero condicionada con el evento de que el primer dado mostró 1 cambiaba notablemente a $1/6$. Pensemos ahora en un problema casi idéntico, pero esta vez calculemos la probabilidad de que la suma sea 7. El evento e cuestión sería:

$$\begin{aligned} A &= \text{La suma de las caras es 7} \\ &= \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1), \} \end{aligned}$$

por lo tanto:

$$P(A) = \frac{1}{6}$$

Probabilida Condicional

B es el mismo evento que definimos previamente, el evento de que el primer dado muestre 1:

$$B = \{(1, 6), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), \}$$

de modo que:

$$P(B) = \frac{1}{6}$$

Por otro lado:

$$A \cap B = \{(1, 6)\}$$

y en consecuencia:

$$P(A \cap B) = \frac{1}{36}$$

Probabilidad Condicional

en conclusión:

$$P(A)P(B) = \frac{1}{36} = P(A \cap B)$$

es decir; en este caso A y B son independientes! y se sigue entonces que:

$$P(A | B) = P(A)$$

Sea C el evento de que salga un 6 en el primer dado. Un argumento análogo nos conduce a concluir que A y C son independientes. De hecho, también B y C son independientes. Por lo tanto:

$$P(A \cap B) = P(A)P(B) \quad P(A \cap C) = P(A)P(C) \quad P(B \cap C) = P(B)P(C)$$

En situaciones como esta se dice que A , B y C son **independientes dos a dos**.

Probabilidad Condicional

Sin embargo, debemos tener cuidado, observemos que

$$P(A \mid B \cap C) = 1$$

Por lo tanto A, B y C no son independientes en el sentido de que:

$$P(A \cap B \cap C) \neq P(A)P(B)P(C)$$

De aquí se sigue la definición:

Independencia

Decimos que A_1, A_2, \dots, A_n es una colección de eventos independientes si para toda $r \leq n$

$$P(A_1 \cap A_2 \cap \dots \cap A_r) = P(A_1)P(A_2) \dots P(A_r)$$

Probabilidad Condicional

Si $\{A_n\}$ es una colección infinita de eventos, decimos que ésta es independiente si cualquier subconjunto finito de esta colección es a su vez una colección de eventos independientes.

Si A y B son independientes entonces A y B^c son independientes.

No es difícil probar este hecho:

$$\begin{aligned}P(A) &= P(A \cap (B \cup B^c)) \\ &= P(A \cap B) + P(A \cap B^c)\end{aligned}$$

Con lo cual:

$$\begin{aligned}P(A \cap B^c) &= P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) \\ &= P(A)P(B^c)\end{aligned}$$

Probabilidad Condicional

Ejemplo. Supongamos que un experimento es repetido n veces. El resultado de un ensayo, no incide en el resultado de los demás, es decir se efectúan ensayos independientes. La probabilidad de éxito de cada ensayo es de p y la probabilidad de fracaso $1 - p$.

Si A_i es el evento de éxito del i -ésimo ensayo, entonces la colección $\{A_i\}$ es una colección de eventos independientes. Pero también la colección $\{A_i^c\}$ son independientes. Por lo tanto:

$$P(\cap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i) = p^n$$

y

$$P(\cap_{i=1}^n A_i^c) = \prod_{i=1}^n P(A_i^c) = (1 - p)^n$$

La última igualdad es la probabilidad de que todos los ensayos sean un fracaso. Por lo tanto, la probabilidad de que exista *al menos* un éxito en los n ensayos es igual a:

$$1 - (1 - p)^n$$

Este resultado está relacionado con la pregunta: ¿Cuál es la probabilidad de que existan exactamente i éxitos en los n ensayos? Esto sería equivalente a seleccionar i de los n ensayos. Estos i ensayos son un éxito con probabilidad p , el resto son fracaso con probabilidad $(1 - p)$. El número de selecciones de este estilo es igual a: $n!/i!(n - i)!$. Por lo tanto, la probabilidad de tener exactamente i éxitos en n ensayos es igual a:

$$\binom{n}{i} p^i (1 - p)^{n-i}$$

Probabilidad Condicional

El siguiente resultado, es de particular utilidad, de algún modo nos permite , calcular la probabilidad de una serie de eventos, de manera separada; “divide y vencerás” .

Regla de la multiplicación

Sean A_1, A_2, \dots, A_n eventos definidos en un espacio de probabilidad.

$$P(A_1, A_2, \dots, A_n) = P(A_1)P(A_2 | A_1) \cdots P(A_n | A_1, A_2, \dots, A_{n-1})$$

En el enunciado anterior hemos usado una nueva notación:

$$P(A_1, A_2, \dots, A_n) := P(\cap_1^n A_i)$$

solo es nueva notación, que en casos como este es de mayor utilidad. Esta regla de multiplicación es particularmente útil cuando se estudian temas donde la estructura de dependencia entre eventos es muy particular. Un ejemplo de ello es cuando se estudian Cadenas de Markov por ejemplo.

Probabilidad Condicional

La prueba de la regla de la multiplicación no es particularmente difícil, notemos que solo es el resultado de aplicar de forma recursiva la igualdad:

$$\begin{aligned}P(A_1, A_2, \dots, A_n) &= P(A_1, A_2, \dots, A_n) \frac{P(A_1, A_2, \dots, A_{n-1})}{P(A_1, A_2, \dots, A_{n-1})} \\&= P(A_n \mid A_1, A_2, \dots, A_{n-1}) P(A_1, A_2, \dots, A_{n-1})\end{aligned}$$

Por lo tanto:

$$P(A_1, A_2, \dots, A_n) = P(A_n \mid A_1, A_2, \dots, A_{n-1}) P(A_1, A_2, \dots, A_{n-1})$$

“Reciclando” este argumento obtenemos que:

$$P(A_1, A_2, \dots, A_{n-1}) = P(A_{n-1} \mid A_1, A_2, \dots, A_{n-2}) P(A_1, A_2, \dots, A_{n-2})$$

Probabilidad Condicional

por lo tanto:

$$P(A_1, A_2, \dots, A_n) = \\ P(A_n | A_1, A_2, \dots, A_{n-1})P(A_{n-1} | A_1, A_2, \dots, A_{n-2})P(A_1, A_2, \dots, A_{n-2})$$

repitiendo el argumento las veces necesarias, obtenemos el resultado.

Ejemplo. Un mazo convencional de 52 cartas es barajado con cuidado. Las 52 cartas son repartidas entre cuatro jugadores, de modo que cada jugador tenga 13 cartas. ¿Cuál es la probabilidad de que cada jugador tenga exactamente un As?

Para resolver este problema consideremos los eventos:

$A_1 = \{\text{El As de espadas lo tiene cualquier jugador}\}$

$A_2 = \{\text{El As de espadas y el de corazones los tienen jugadores diferentes}\}$

$A_3 = \{\text{Los Ases de espadas, corazones y diamantes
los tienen jugadores diferentes}\}$

$A_4 = \{\text{Los cuatro Ases los tienen jugadores diferentes}\}$

Dados estos eventos, la probabilidad que buscamos es:

$$P(A_1, A_2, A_3, A_4)$$

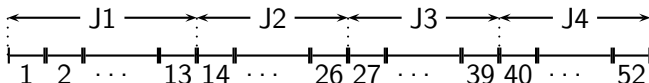
Para calcularla usaremos la regla de la multiplicación. Primero el más fácil:

$$P(A_1) = 1$$

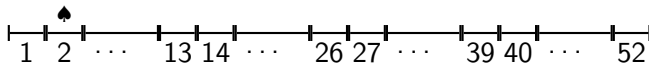
Creo que esto está claro, pues en realidad no hemos puesto ninguna restricción en la repartición del mazo. ¿Cómo calculamos $P(A_2 | A_1)$? en este caso debemos asegurarnos de que el As de espadas y el de corazones sean repartidos a jugadores diferentes:

Probabilidad Condicional

La siguiente figura nos puede ser útil:

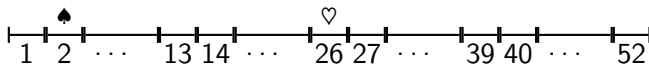


En la figura, las 13 primeras cartas le corresponden al jugador 1, las cartas de la 14 a la 26 al jugador 2, las cartas de la 27 al 39 al jugador 3 y las cartas de la 40 a la 53 al jugador 4. ¿Que es lo que pasa entonces con $P(A_2 | A_1)$? Supongamos que el As de espadas fue repartido al jugador 1. Una posibilidad es la siguiente:



Probabilidad Condicional

Calcularemos $P(A_2 | A_1)$ haciendo el cociente de casos favorables entre casos totales. Para ello observemos que el número de casos totales coincide con con las casillas en las cuales podemos colocar al As de corazones. Dado que el As de espadas ya fue asignado, el As de corazones solo tiene 51 casillas disponibles. Estas serian ahora los “casos totales”. Sin embargo los casos favorables son menos. Para que un caso sea favorable el As de corazones debe colocarse en alguna de las casillas de la 14 a la 52, una posibilidad seria la siguiente:



Es decir, los casos favorables serían 39., por lo tanto:

$$P(A_2 | A_1) = \frac{39}{51}$$

Probabilidad Condicional

El calculo de $P(A_3 | A_1, A_2)$ es similar, dado que el AS de espadas y el de corazones ya fueron asignados, ahora tenemos 50 casillas disponibles (casos totales) para colocar el As de diamantes, pero solo 26 dan lugar a una asignación favorable. Con lo cual:

$$P(A_3 | A_1, A_2) = \frac{26}{50}$$

de forma análoga concluimos que:

$$P(A_4 | A_1, A_2, A_3) = \frac{13}{49}$$

En conclusión, la probabilidad de que los Ases sean asignados a jugadores diferentes es igual a

$$\begin{aligned} P(A_1, A_2, A_3, A_4) &= P(A_4 | A_1, A_2, A_3)P(A_3 | A_1, A_2)P(A_2 | A_1)P(A_1) \\ &= \frac{13}{49} \times \frac{26}{50} \times \frac{39}{15} \times 1 \approx 0.105 \end{aligned}$$

Probabilidad Condicional

Ejemplo. En un jurado de tres personas, dos de los miembros toman cada uno, la decisión correcta, de manera independiente con probabilidad p , mientras que el tercero tira una moneda para dar su veredicto. Las discrepancias se resuelven por mayoría. Otro jurado, compuesto de una sola persona, toma la decisión correcta con probabilidad p . ¿Cuál de los dos jurados tiene una mayor probabilidad de tomar una decisión acertada?



Probabilidad condicional

La solución de este problema la encontramos al calcular la probabilidad de que el jurado compuesto de tres miembros tome una buena decisión. Supongamos que A es el evento de que el jurado de tres miembros toma la decisión correcta. Por otro lado denotemos con J_i , $i = 1, 2, 3$, al evento de que el miembro i toma una decisión acertada. Tenemos entonces que:

$$P(J_1) = P(J_2) = p \quad \text{y} \quad P(J_3) = 1/2$$

Pero:

$$P(A \mid J_1, J_2) = 1$$

y

$$P(A \mid J_1^c, J_2) = P(A \mid J_1, J_2^c) = 1/2$$

además de que:

$$P(A \mid J_1^c, J_2^c) = 0$$

Esto nos permite usar una estrategia muy útil, notemos que:

$$\Omega = (J_1 \cap J_2) \cup (J_1^c \cap J_2) \cup (J_1 \cap J_2^c) \cup (J_1^c \cap J_2^c)$$

pero además obsérvese que esta unión es una unión de disjuntos. Por lo tanto:

$$\begin{aligned} P(A) = & P(A \cap (J_1 \cap J_2)) + P(A \cap (J_1^c \cap J_2)) \\ & + P(A \cap (J_1 \cap J_2^c)) + P(A \cap (J_1^c \cap J_2^c)) \end{aligned}$$

Condicionando las probabilidades anteriores obtenemos:

$$\begin{aligned} P(A) = & P(A \mid J_1, J_2)P(J_1 \cap J_2) + P(A \mid J_1^c, J_2)P(J_1^c \cap J_2) \\ & + P(A \mid J_1, J_2^c)P(J_1 \cap J_2^c) + P(A \mid J_1^c, J_2^c)P(J_1^c \cap J_2^c) \end{aligned}$$

Probabilidad Condicional

Usando los valores obtenidos para las condicionales y el supuesto de independencia entre los eventos J_i , concluimos que:

$$\begin{aligned} P(A) &= 1 \times p^2 + \frac{1}{2} \times (1-p)p + \frac{1}{2} \times p(1-p) + 0 \times (1-p)^2 \\ &= p^2 + (1-p)p \\ &= p \end{aligned}$$

por lo tanto, ambos jurados tiene la misma probabilidad de llegar a una decisión acertada. La solución de este problema y la solución del problema del dilema del prisionero, son ambas, aplicaciones de uno de los resultados más útiles del curso, la fórmula de **probabilidad total**.

Probabilidad Condicional

Fórmula de Probabilidad Total

Sea B_i una partición del espacio de probabilidad Ω , entonces:

$$P(A) = \sum_i P(A \cap B_i)$$

Si además $P(B_i) > 0$ para toda i , entonces

$$P(A) = \sum_i P(A | B_i)P(B_i)$$

Son en verdad muchos, los problemas que se simplifican notablemente mediante el uso de esta relación.

Probabilidad Condicional

Ejemplo. Un laboratorio esta diseñando una prueba de sangre para el Covid 19. La prueba es buena si detecta el virus con una probabilidad alta en un paciente que efectivamente tiene el virus. Es decir, si la probabilidad:

$$P(\text{la prueba es positiva} \mid \text{la enfermedad esta presente})$$

es alta.

Existen sin embargo dos casos donde la prueba nos puede conducir a conclusiones erróneas:

$P(\text{la prueba es negativa} \mid \text{la enfermedad esta presente})$ falso negativo

$P(\text{la prueba es positiva} \mid \text{la enfermedad esta ausente})$ falso positivo

Probabilidad condicional

Si la prueba identifica el virus con un 95 % de efectividad entonces:

$$P(\text{la prueba es positiva} \mid \text{la enfermedad esta presente}) = 0.95$$

por lo tanto

$$P(\text{la prueba es negativa} \mid \text{la enfermedad esta presente}) = 0.05$$

Esto tiene sentido, si la prueba es efectiva entonces la probabilidad de un falso negativo debe ser baja. Lo mismo debe ocurrir con un falso positivo, la probabilidad debe ser baja si es que la prueba es efectiva. Supongamos que la probabilidad de este falso positivo es también 0.05:

$$P(\text{la prueba es positiva} \mid \text{la enfermedad esta ausente}) = 0.05$$

Probabilidad Condicional

Si definimos :

$$A = \{\text{La prueba es positiva}\}$$

y

$$B = \{\text{La enfermedad esta presente}\}$$

Entonces:

$$P(A \mid B) = 0.95 \quad P(A^c \mid B) = 0.05 = P(A \mid B^c)$$

Hasta aqui todo bien... si la prueba es efectiva y nos detecta Covid... casi seguro que tenemos la enfermedad. ¿o no?

Somos muy desconfiados y nos preguntamos por la probabilidad:

$$P(B \mid A) = P(\text{La enfermedad esta presente} \mid \text{La prueba es positiva})$$

Probabilidad Condicional

Para calcular esta probabilidad echaremos mano de otro resultado de enorme importancia la **Fórmula de Bayes**

Fórmula de Bayes

Dados los eventos A y B ;

$$P(B | A) = \frac{P(A | B)P(B)}{P(A | B)P(B) + P(A | B^c)P(B^c)}$$

La prueba de este resultado no es difícil, observemos que:

$$\begin{aligned} P(B | A) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{P(A \cap B)}{P(B)} \frac{P(B)}{P(A)} \\ &= \frac{P(A | B)P(B)}{P(A)} \end{aligned}$$

Probabilidad Condicional

Es decir

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

Esta igualdad es en si misma de mucha utilidad. Para concluir, aplicamos probabilidad total sobre $P(A)$:

$$P(A) = P(A | B)P(B) + P(A | B^c)P(B^c)$$

de modo que

$$P(B | A) = \frac{P(A | B)P(B)}{P(A | B)P(B) + P(A | B^c)P(B^c)}$$

Probabilidad Condicional

Regresando al problema, busquemos la probabilidad;

$$P(B | A) = P(\text{La enfermedad esta presente} | \text{La prueba es positiva})$$

sabiendo que:

$$P(A | B) = 0.95 \quad P(A^c | B) = 0.05 = P(A | B^c)$$

Para usar Bayes, solo nos falta $P(A)$. Supongamos que $P(A) = 0.003$, tenemos entonces que:

$$P(B | A) = \frac{(0.95)(0.003)}{(0.95)(0.003) + (0.05)(0.997)} \approx 0.05$$

Es decir, la probabilidad de que efectivamente estemos enfermos dado que la prueba es positiva es 0.05! 🤔

Probabilidad Condicional

Ejemplo. Supongamos que en un examen de opción múltiple, un estudiante conoce la respuesta a una pregunta con probabilidad p . Si no conoce la respuesta, la adivina. Si la pregunta tiene n opciones, tendrá una probabilidad de $1/n$ de responder correctamente. ¿Cual es la probabilidad de que el estudiante realmente sepa la respuesta a la pregunta dado que respondió correctamente?

Supongamos que A es el evento donde el alumno responde correctamente la pregunta y B el evento donde el estudiante conoce la respuesta a la pregunta, deseamos encontrar:

$$P(B | A)$$

Probabilidad Condicional

Usando Bayes:

$$\begin{aligned}P(B | A) &= \frac{P(A | B)P(B)}{P(A | B)P(B) + P(A | B^c)P(B^c)} \\&= \frac{1 \times p}{1 \times p + (1/n) \times (1 - p)} \\&= \frac{np}{1 + (n - 1)p}\end{aligned}$$

Por ejemplo; si $n = 5$ y $p = 1/2$, entonces $P(B | A) = 5/6$.

Variables Aleatorias

La teoría que hemos desarrollado hasta el momento, tiene como fundamento, la construcción de un espacio de probabilidad (Ω, \mathcal{F}, P) . Sabemos bien que el conjunto Ω contiene todos los posibles resultados del experimento. Esto, claro esta, es útil por si mismo, sin embargo lo mas común en la práctica es que estemos interesados no en elementos de Ω directamente, si no en alguna función de estos. Esto nos conduce a la siguiente definición¹:

Variable Aleatoria

Dado un espacio de probabilidad (Ω, \mathcal{F}, P) , una función:

$$X : \Omega \rightarrow \mathbb{R}$$

Es conocida como **variable aleatoria**.

¹Esta es una definición adecuada para un primer curso, de manera mas formal, debemos exigir que la función sea *medible*.

Variables Aleatorias

Consideremos por ejemplo el lanzamiento de tres monedas. El espacio muestral sería:

$$\Omega = \{\omega_1 = (AAA), \omega_2 = (AAS), \omega_3 = (ASA), \omega_4 = (ASS) \\ \omega_5 = (SSS), \omega_6 = (SSA), \omega_7 = (SAS), \omega_8 = (SAA)\}$$

Esto es correcto y describe satisfactoriamente al experimento aleatorio. Supongamos sin embargo que en realidad estamos interesados en observar el numero X de "Águilas" que obtenemos después de arrojar la moneda. La variable X es una variable aleatoria:

$$X(\omega_1) = 3 \quad X(\omega_5) = 0$$

$$X(\omega_2) = 2 \quad X(\omega_6) = 1$$

$$X(\omega_3) = 2 \quad X(\omega_7) = 1$$

$$X(\omega_4) = 1 \quad X(\omega_8) = 2$$

Variables Aleatorias

Observemos como X mapea cada elemento de Ω en algún real. Notemos de hecho que:

$$X(\omega) = 0 \text{ si y solo si } \omega \in \{\omega_5\}$$

$$X(\omega) = 1 \text{ si y solo si } \omega \in \{\omega_4, \omega_6, \omega_7\}$$

$$X(\omega) = 2 \text{ si y solo si } \omega \in \{\omega_2, \omega_3, \omega_8\}$$

$$X(\omega) = 3 \text{ si y solo si } \omega \in \{\omega_1\}$$

Este ejemplo nos conduce a nueva notación:

$$X^{-1}(x) = \{\omega \in \Omega : X(\omega) = x\}$$

$X^{-1}(x)$ es conocida como la **preimagen** de x bajo X . Notemos que esta preimagen **no** es una función.

Variables Aleatorias

Así, en el ejemplo anterior:

$$X^{-1}(2) = \{\omega_2, \omega_3, \omega_8\}$$

y

$$X^{-1}(0) = \{\omega_5\}$$

Una notación más compacta y de uso mas frecuente es la siguiente²:

$$\{X = x\} = \{\omega \in \Omega : X(\omega) = x\}$$

Siguiendo el ejemplo anterior:

$$\{X = 1\} = \{\omega_4, \omega_6, \omega_7\}$$

y

$$\{X = 3\} = \{\omega_1\}$$

²Es una costumbre extendida usar mayúsculas para denotar variables aleatorias (la función) y minúsculas para denotar escalares.

Variables Aleatorias

Ejemplo. En el lanzamiento de dos dados, estamos interesados en la variable X que denota al máximo de las dos caras. Observemos que $X : \Omega \rightarrow \{1, 2, 3, 4, 5, 6\} \subset \mathbb{R}$ y tenemos por ejemplo que:

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

o

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

usando nuestra nueva notación, podemos calcular probabilidades como:

$$P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

Variables Aleatorias

En el ejemplo de los dados:

$$\begin{aligned} P(X = 1) &= \frac{1}{36} & P(X = 4) &= \frac{7}{36} \\ P(X = 2) &= \frac{3}{36} & P(X = 5) &= \frac{9}{36} \\ P(X = 3) &= \frac{5}{36} & P(X = 6) &= \frac{11}{36} \end{aligned}$$

En el ejemplo del lanzamiento de tres monedas, podemos exhibir una formula especifica, si p es la probabilidad de que la moneda muestre Águila, entonces es claro que:

$$P(X = 0) = (1 - p)^3 \quad \text{y} \quad P(X = 3) = p^3$$

Variables Aleatorias

Cuando $i = 1$ o $i = 2$, debemos tener un poco de mayor cuidado, ya que por ejemplo:

$$\{X = 1\} = \{(ASS), (SAS), (SSA)\}$$

De donde:

$$P(X = 1) = 3p(1 - p)^2$$

y

$$P(X = 2) = 3p^2(1 - p)$$

De hecho en general:

$$P(X = i) = \binom{3}{i} p^i (1 - p)^{3-i} \text{ para } i = 0, 1, 2, 3.$$

Variables Aleatorias

Ejemplo. Seleccionamos aleatoriamente tres bolas sin reemplazo extrayéndolas de una urna con 20 bolas numeradas del 1 al 20. Si apostamos a que al menos una de las bolas seleccionadas tiene un número mayor o igual que 17, ¿cuál es la probabilidad de que ganemos la apuesta? La clave en este problema esta en notar que si X es la variable aleatoria que registra el número máximo obtenido y X es mayor o igual a 17, entonces ganaremos la apuesta. Para ello observemos que X puede tomar uno de los valores: 3, 4, ..., 20. Por otro lado, si cada una de las extracciones es igualmente probable, entonces:

$$P(X = i) = \frac{\binom{i-1}{2}}{\binom{20}{3}}, \quad i = 3, \dots, 20.$$

Variables Aleatorias

En efecto, el denominador simplemente cuenta el número total de extracciones posibles de tres bolas a partir de una colección. El numerador, hace algo similar, pero ahora solo seleccionamos dos bolas de entre aquellas que sabemos que son menores a i . La probabilidad que buscamos sería:

$$P(X \geq 17) = \sum_{i=17}^{20} P(X = i) \approx 0.508$$

Ejemplo. Consideremos el experimento, donde lanzamos una moneda de forma repetida hasta que obtenemos águila o bien un total de n lanzamientos fueron realizados. Si X denota el número total de lanzamientos efectuados entonces X toma alguno de los valores: $1, 2, \dots, n$.

Mas aún si obtenemos águila con probabilidad p :

$$P(X = 1) = p$$

$$P(X = 2) = (1 - p)p$$

$$P(X = 3) = (1 - p)^2 p$$

$$\vdots$$

$$P(X = n - 1) = (1 - p)^{n-2} p$$

$$P(X = n) = (1 - p)^{n-1}$$

Notemos que:

$$\begin{aligned} P(\cup_{i=1}^n \{X = i\}) &= \sum_{i=1}^n P(X = i) \\ &= \sum_{i=1}^{n-1} (1 - p)^{i-1} p + (1 - p)^{n-1} \end{aligned}$$

Variables Aleatorias

de donde

$$\begin{aligned}P\left(\cup_{i=1}^n \{X = i\}\right) &= p \left[\frac{1 - (1-p)^{n-1}}{1 - (1-p)} \right] + (1-p)^{n-1} \\&= 1 - (1-p)^{n-1} + (1-p)^{n-1} \\&= 1\end{aligned}$$

El desarrollo anterior sugiere la idea de que una variable aleatoria induce cierta medida de probabilidad. La idea de manera superficial es como sigue³; Sabemos que una variable aleatoria es una función que mapea al conjunto de Ω en los reales, esto es:

$$X : \Omega \rightarrow \mathbb{R}$$

³El desarrollo formal requiere de conceptos de teoría de la medida, temas que exceden el nivel de este curso, sin embargo se ofrece aquí al menos un tratamiento superficial del tema.

Variables Aleatorias

Pero no olvidemos que todo modelo de probabilidad esta definido en un espacio de probabilidad (Ω, \mathcal{F}, P) , de algún modo, la variable X , también mapea a la σ -álgebra \mathcal{F} en una nueva σ -álgebra \mathcal{B} :

$$\mathcal{F} \xrightarrow{X} \mathcal{B}$$

\mathcal{B} es una σ -álgebra de subconjuntos de \mathbb{R} , no profundizaremos, pero esta suele coincidir con la σ -álgebra de Borel o la σ -álgebra de Lebesgue. Del mismo modo, podemos definir una nueva medida de probabilidad, inducida por la misma variable aleatoria:

$$P \xrightarrow{X} P_X$$

Esto lo logramos del siguiente modo; si $A \in \mathcal{B}$ entonces:

$$\begin{aligned} P_X(A) &= P(X \in A) \\ &= P(\{\omega \in \Omega : X(\omega) \in A\}) \end{aligned}$$

Variables Aleatorias

En este punto, es pertinente, hacer una distinción. Si una variable aleatoria X toma valores en algún conjunto numerable, decimos que la variable aleatoria es una **variable aleatoria discreta**. Si la variable aleatoria toma valores en un conjunto no numerable, entonces la variable aleatoria es una **variable aleatoria continua**.

Las variables aleatorias que hemos visto hasta el momento, toman valores en conjuntos no solo numerables si no finitos, por lo tanto son variables aleatorias discretas. Sin embargo pensemos por ejemplo en una variable aleatoria que registre el tiempo que un pasajero debe esperar en una estación del metro hasta que éste llega... Si concedemos que la llegada entre un metro y otro no debe exceder de 3 minutos⁴ y medimos el tiempo en minutos, entonces la variable aleatoria puede tomar cualquier valor en el intervalo $[0, 3]$. X es por tanto una variable aleatoria continua. Mas tarde veremos mas ejemplos.

⁴aha!!!

Variables Aleatorias

Esta observación es pertinente en este punto, si una variable aleatoria es discreta entonces podemos ir mas lejos con respecto a P_X , en efecto, si X es discreta:

$$\begin{aligned}P_X(A) &= P(X \in A) \\&= P(\{\omega \in \Omega : X(\omega) \in A\}) \\&= \sum_{a \in A} P(\{\omega \in \Omega : X(\omega) = a\}) \\&= \sum_{a \in A} P(X = a)\end{aligned}$$

En resumen, si X es discreta, entonces, para cualquier $A \in \mathcal{B}$:

$$P_X(A) = \sum_{a \in A} P(X = a)$$

Variables Aleatorias

Si S el rango de una variable aleatoria X , es decir $X : \Omega \rightarrow S \subset \mathcal{R}$ y X es una variable aleatoria discreta, definimos la **función de probabilidad** o **función de probabilidad de masa** $p(a)$ de X como:

$$p(a) = \begin{cases} P(X = a) & a \in S \\ 0 & a \notin S \end{cases}$$

Observemos que $p(a) \geq 0$ para todo $a \in S$, además:

$$P(X \in S) = \sum_{a \in S} p(a) = 1$$

Ejemplo. La función de probabilidad de una variable aleatoria X esta dada por

$$p(i) = \begin{cases} \frac{e^{-\lambda} \lambda^i}{i!} & i = 0, 1, \dots \\ 0 & \text{en otro caso.} \end{cases}$$

para $\lambda > 0$.

Variables Aleatorias

Se nos pide encontrar:

- 1 $P(X = 0)$
- 2 $P(X > 2)$.

Para ello debemos determinar previamente el valor de c :

$$\sum_{i=0}^{\infty} \frac{c\lambda^i}{i!} = 1$$

pero sabemos que $e^x = \sum_{i=0}^{\infty} x^i / i!$, lo cual implica que:

$$ce^{\lambda} = 1 \quad \text{o bien} \quad c = e^{-\lambda}$$

de modo que

$$p(i) = \begin{cases} \frac{e^{-\lambda}\lambda^i}{i!} & i = 0, 1, \dots \\ 0 & \text{en otro caso.} \end{cases}$$

Variables Aleatorias

De aquí se sigue que:

$$P(X = 0) = \frac{e^{-\lambda} \lambda^0}{0!} = e^{-\lambda}$$

por otro lado:

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) \\ &= 1 - \{P(X = 0) + P(X = 1) + P(X = 2)\} \\ &= 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2 e^{-\lambda}}{2} \end{aligned}$$

Variables Aleatorias

Si X es una variable aleatoria discreta definimos, la **funcion de probabilidad acumulada** F de X como:

$$F(a) = \sum_{x \leq a} p(x) = P(X \leq a)$$

Ejemplo. Supongamos que X es una variable aleatoria con función de probabilidad definida por:

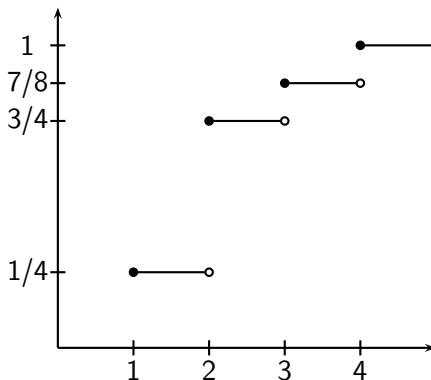
$$p(1) = \frac{1}{4} \quad p(2) = \frac{1}{2} \quad p(3) = \frac{1}{8} \quad p(4) = \frac{1}{8}$$

la función acumulada esta dada por:

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{4} & 1 \leq a < 2 \\ \frac{3}{4} & 2 \leq a < 3 \\ \frac{7}{8} & 3 \leq a < 4 \\ 1 & 4 < a \end{cases}$$

Variables Aleatorias

Conviene echar un ojo a la representación gráfica de la función acumulada:



Variables aleatorias

Como vemos, en el caso de variables aleatorias discretas, la función de acumulación es una función no decreciente con saltos. Es continua por la derecha y cada salto coincide con una evaluación de $p(x)$ de la función de probabilidad.

Variables Aleatorias

Valor Esperado

El **Valor Esperado**, **Esperanza** o **media** de una variable aleatoria es uno de los conceptos fundamentales en la teoría de probabilidad. Si X es una variable aleatoria discreta con función de probabilidad $p(x)$, definimos la *esperanza* de X o *valor esperado* de X como:

$$E[X] = \sum_{x:p(x)>0} xp(x) = \sum_{x \in S} xp(x)$$

de manera informal: el valor esperado de X es un promedio ponderado de los posibles valores que puede tomar la variable X .

Ejemplo. Encuentra el valor esperado $E[X]$ donde X es el resultado obtenido al lanzar un dado justo.

Dado que $p(x) = 1/6$ para $x \in S = \{1, 2, 3, 4, 5, 6\}$, tenemos que:

$$E[X] = \sum_{i=1}^6 i \cdot \frac{1}{6} = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}$$

Ejemplo. En el ejemplo del lanzamiento de las tres monedas. Teníamos que:

$$P(X = i) = \binom{3}{i} p^i (1 - p)^{3-i} \quad i = 0, 1, 2, 3.$$

Tenemos entonces que:

$$\begin{aligned} E[X] &= \sum_{i=0}^3 i \binom{3}{i} p^i (1 - p)^{3-i} \\ &= \sum_{i=1}^3 \frac{3!}{(i-1)!(3-i)!} p^i (1 - p)^{3-i} \\ &= 3p \cdot \sum_{i=1}^3 \frac{2!}{(i-1)!(2-(i-1))!} p^{i-1} (1 - p)^{2-(i-1)} \end{aligned}$$

Variables Aleatorias

Haciendo $j = i - 1$ tenemos que:

$$\begin{aligned} E[X] &= 3p \sum_{j=0}^2 \binom{2}{j} p^j (1-p)^{2-j} \\ &= 3p \cdot \{(1-p)^2 + 2p(1-p) + p^2\} \\ &= 3p \cdot \{(1-p) + p\}^2 \\ &= 3p \end{aligned}$$

El siguiente, es un resultado de gran utilidad para lo sucesivo; Si X es una variable aleatoria discreta que toma valores $x_i, i \geq 1$ con probabilidades $p(x_i)$ entonces, para cualquier función g se tiene que:

$$E[g(X)] = \sum_i g(x_i) p(x_i)$$

Variables aleatorias

Este resultado, aparentemente inocente. . . es muy útil y relevante, notemos que si X es una variable aleatoria, entonces

$$Y = g(X)$$

es también una variable aleatoria. Por definición:

$$E[g(X)] = E[Y] = \sum_y yP(Y = y)$$

Es decir, la esperanza de $g(X)$ tendríamos que calcularla en función de las probabilidades $P(Y = y)$, esto es; en función de la función de probabilidad de Y . Sin embargo el resultado que estamos proponiendo afirma que:

$$E[Y] = \sum_i g(x_i)p(x_i)$$

es decir, estamos usando directamente la función de probabilidad de X en lugar de la de Y .

Variables Aleatorias

Esto puede ser sumamente relevante, pues nos “ahorramos” el trabajo de estimar la función de probabilidad de Y . A continuación damos la prueba del resultado.

$$\begin{aligned}\sum_i g(x_i)p(x_i) &= \sum_j \sum_{i:g(x_i)=y_j} g(x_i)p(x_i) \\ &= \sum_j \sum_{i:g(x_i)=y_j} y_j p(x_i) \\ &= \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i) \\ &= \sum_j y_j P(g(X) = y_j) \\ &= \sum_j y_j P(Y = y_j) \\ &= E[Y] \\ &= E[g(X)]\end{aligned}$$

Variables Aleatorias

Un corolario inmediato e importantísimo lo encontramos a continuación, dados a y b escalares:

$$\begin{aligned} E[aX + b] &= \sum_{x:p(x)>0} (ax + b)p(x) \\ &= a \sum_{x:p(x)>0} xp(x) + b \sum_{x:p(x)>0} p(x) \\ &= aE[X] + b \end{aligned}$$

Variables Aleatorias

Otra aplicación de suma importancia la encontramos en la siguiente definición

Varianza de una variable Aleatoria

Si X es una variable aleatoria con media $E[X] = \mu$, definimos la **varianza** de X como:

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

Se dice que la varianza es una **medida de dispersión**. La varianza de una variable aleatoria, es una estimación de que tanto se despega la misma de su media. Aplicando el resultado sobre esperanzas de funciones de variables aleatorias mas la linealidad de la esperanza, tenemos el siguiente resultado:

$$\begin{aligned}\text{Var}(x) &= E[(X - \mu)^2] \\ &= E[X^2 - 2X\mu + \mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - \mu^2\end{aligned}$$

Variables Aleatorias

Esta nueva versión:

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

suele ser mas utilizado en la práctica.

La esperanza $E[X^k]$ se conoce como **k -ésimo momento muestral**. En este sentido $E[X^2]$ es el **segundo momento muestral**. Esperanzas de este tipo suelen tener aplicaciones en distintos ámbitos.

Ejemplo. Decimos que una variable aleatoria X **se distribuye Bernoulli** con parámetro p y se denota $X \sim \text{Bernoulli}(p)$ si tiene función de probabilidad:

$$P(X = i) = \begin{cases} p & \text{si } i = 1 \\ 1 - p & \text{si } i = 0 \end{cases}$$

Variables Aleatorias

Una variable aleatoria Bernoulli, modelo un fenómeno que ocurre o no, que solo tiene una de dos posibilidades de ocurrir. El ejemplo mas claro sería el lanzamiento de una moneda, si ésta es justa $p = 1/2$.

Tenemos entonces que:

$$E[X] = p \cdot 1 + (1 - p) \cdot 0 = p$$

Por otro lado:

$$E[X^2] = 1 \cdot p + (1 - p) \cdot 0 = p$$

Por lo tanto:

$$\text{Var}(X) = E[X^2] - E[X]^2 = p - p^2 = p(1 - p)$$

Variables Aleatorias

Decimos que una variable aleatoria se distribuye **Binomial** con parámetros n y p y se denota $X \sim \text{Binom}(n, p)$ si tiene función de probabilidad:

$$P(X = i) = \begin{cases} \binom{n}{i} p^i (1-p)^{n-i} & \text{si } i = 0, 1, \dots, n \\ 0 & \text{en otro caso} \end{cases}$$

Una variable aleatoria binomial cuanta el número de éxitos en n ensayos Bernoulli, cada uno de ellos con probabilidad p . Calcular la media y la varianza de una variable aleatoria Binomial es mas fácil si calculamos de manera genérica el k -ésimo momento muestral:

$$\begin{aligned} E[X^k] &= \sum_{i=0}^n i^k \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=1}^n i^k \binom{n}{i} p^i (1-p)^{n-i} \end{aligned}$$

Variables Aleatorias

Usando la igualdad:

$$i \binom{n}{i} = n \binom{n-1}{i-1}$$

tenemos que

$$\begin{aligned} E[X^k] &= np \sum_{i=1}^n i^{k-1} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \\ &= np \sum_{j=0}^n (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \\ &= np E[(Y+1)^{k-1}] \end{aligned}$$

donde $Y \sim \text{Binom}(n-1, p)$.

De aquí se sigue fácilmente que, si $k = 1$ entonces:

$$E[X] = np$$

Variables Aleatorias

Por otro lado si $k = 2$:

$$\begin{aligned} E[X^2] &= npE[Y + 1] \\ &= np[(n - 1)p + 1] \end{aligned}$$

por lo tanto:

$$\begin{aligned} \text{Var}(x) &= E[X^2] - E[X]^2 \\ &= np[(n - 1)p + 1] - (np)^2 \\ &= np(1 - p) \end{aligned}$$

Variables Aleatorias

Ejemplo. Decimos que una variable aleatoria X se distribuye **Poisson** con parámetro $\lambda > 0$ y se denota $X \sim Poi(\lambda)$ si tiene función de probabilidad:

$$P(X = i) = \begin{cases} e^{-\lambda} \frac{\lambda^i}{i!} & \text{para } i = 0, 1, \dots \\ 0 & \text{en otro caso} \end{cases}$$

Notemos que dado que

$$e^{\lambda} = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!}$$

entonces $\sum_{i=0}^{\infty} P(X = i) = 1$

Variables Aleatorias

La esperanza de una variable aleatoria Poisson la calculamos como sigue:

$$\begin{aligned} E[X] &= \sum_{i=0}^{\infty} i \cdot e^{-\lambda} \frac{\lambda^i}{i!} \\ &= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{(j)!} \\ &= \lambda \end{aligned}$$

Para estimar la varianza, calculamos primero el segundo momento:

Variables Aleatorias

$$\begin{aligned} E[X^2] &= \sum_{i=0}^{\infty} i^2 \cdot e^{-\lambda} \frac{\lambda^i}{i!} \\ &= \lambda \sum_{i=1}^{\infty} i \cdot e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} (j+1) \cdot \frac{\lambda^j}{j!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} (j+1) \cdot \frac{\lambda^j}{j!} \\ &= \lambda \left\{ e^{-\lambda} \sum_{j=0}^{\infty} j \frac{\lambda^j}{j!} + e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \right\} \\ &= \lambda \{ \lambda + 1 \} \\ &= \lambda^2 + \lambda \end{aligned}$$

Variables Aleatorias

Por lo tanto

$$\begin{aligned} \text{Var}(X) &= E[X^2] - E^2[X] \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda \end{aligned}$$

Este es un hallazgo curioso, la esperanza y la varianza de una variable aleatoria Poisson son iguales entre si y son iguales a su parámetro.

Variables Aleatorias

Por último, si $X \sim \text{Binomial}(n, p)$ y definimos; $\lambda = np$, tenemos que:

$$\begin{aligned} P(X = i) &= \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} \\ &= \frac{n!}{(n-i)!i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1) \cdot (n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{(1 - \lambda/n)^n}{(1 - \lambda/n)^i} \end{aligned}$$

de aquí se sigue que si n es grande y p es pequeño, de modo que λ es moderado:

$$\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda} \quad \frac{n(n-1) \cdot (n-i+1)}{n^i} \approx 1 \quad (1 - \lambda/n)^i \approx 1$$

Variables Aleatorias

De donde se concluye que si n es grande y p es pequeño de modo que λ es moderado, entonces:

$$P(X = i) \approx e^{-\lambda} \frac{\lambda^i}{i!}$$

Ejemplo. Un sistema de comunicación, consiste de n componentes, los cuales funcionan de manera independiente con probabilidad p . El sistema completo funcionara correctamente si al menos la mitad de los componentes se encuentran en funcionamiento. ¿Para que valores de p un sistema de 5 componentes, es mas probable que se encuentre en funcionamiento, que un sistema de 3 componentes?

Variables Aleatorias

La probabilidad de que el sistema con 5 componentes se encuentre en funcionamiento, coincide con la probabilidad de que el número de componentes en funcionamiento sean 3 o mas, esto se logra con probabilidad:

$$\binom{5}{3}p^3(1-p)^2 + \binom{5}{4}p^4(1-p) + p^5$$

de manera análoga, el sistema de 3 componentes se encuentra en funcionamiento con probabilidad:

$$\binom{3}{2}p^2(1-p) + p^3$$

de este modo, el sistema de 5 componentes será mejor que el sistema de 3 componentes si:

$$10p^3(1-p)^2 + 5p^4(1-p) + p^5 > 3p^2(1-p) + p^3$$

Variables Aleatorias

lo cual se reduce a:

$$3(p - 1)^2(2p - 1) > 0$$

o bien:

$$p > \frac{1}{2}$$

Ejemplo. Supongamos que el número de errores tipográficos en un página de un libro se distribuye Poisson con parámetro $\lambda = 1/2$. Calcula la probabilidad de que exista al menos un error en la página.

La probabilidad que buscamos se calcula como sigue:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1/2} \approx 0.393$$

Variables Aleatorias

Ejemplo. Supongamos que la probabilidad de que cierto artículo, producido en una máquina, sea defectuoso es igual a 0.1. Encuentra la probabilidad de que en una muestra de 10 artículos, contenga a lo mas un artículo defectuoso.

La probabilidad que buscamos es:

$$P(X \leq 1) = \binom{10}{0}(0.1)^0(0.9)^{10} + \binom{10}{1}(0.1)^1(0.9)^9 = 0.7361$$

Esta probabilidad podemos aproximarla a través de una variable aleatoria Poisson con parámetro $\lambda = np = 10 \cdot 0.1 = 1$. Asi es la probabilidad usando Poisson sería:

$$e^{-1} + e^{-1} \approx 0.7358$$

Observemos la precisión de esta aproximación, aun cuando n y p son valores no muy extremos. En otros contextos, esta aproximación puede ser de importancia fundamental, para n muy grande, el cálculo de los coeficientes binomiales, puede ser realmente complicado computacionalmente.

Variables Aleatorias

Ejemplo. Consideremos un experimento en el cual contamos el número de partículas α que emite un gramo de un material radioactivo a lo largo de un segundo. Si de la experiencia pasada sabemos que en promedio son emitidas 3.2 partículas α , da una aproximación de la probabilidad de que no mas de 2 partículas α sean emitidas.

Podemos pensar que el gramo de material radioactivo consiste de n átomos, cada uno susceptible de degradarse y emitir una partícula α . En este caso n es un número muy grande, sin embargo podemos considerar que un átomo emite una partícula α con probabilidad $p = 3.2/n$. En este sentido podemos suponer que el número de partículas α emitidas a lo largo de un segundo es una variable aleatoria Poisson con parámetro $\lambda = 3.2$. De modo que la probabilidad que buscamos es igual a:

$$P(X \leq 2) = e^{-3.2} + 3.2e^{-3.2} + \frac{(3.2)^2}{2}e^{-3.2} \approx 0.3799$$

Variables Aleatorias

Del ejemplo anterior, vemos que una variable aleatoria Poisson “*cuenta*”. Cuenta ocurrencias de algún experimento aleatorio, de hecho es la variable aleatoria de conteo por excelencia.

Ejemplo. Decimos que una variable aleatoria se distribuye **Geométrica** con parámetro p , $X \sim \text{Geom}(p)$ si tiene función de probabilidad:

$$P(X = i) = \begin{cases} (1 - p)^{i-1}p & \text{para } i = 1, 2, \dots \\ 0 & \text{en otro caso.} \end{cases}$$

Si p es la probabilidad de éxito en una variable aleatoria Bernoulli, entonces podemos decir que una variable aleatoria geométrica “*cuenta*” el **número de ensayos** Bernoulli, hasta obtener un éxito. Esto si consideramos la definición que recién exhibimos. Sin embargo, no siempre se define de este modo, también es frecuente definirla de un modo alternativo.

Variables Aleatorias

Decimos que una variable aleatoria se distribuye **Geométrica** con parámetro p , $X \sim \text{Geom}(p)$ si tiene función de probabilidad:

$$P(X = i) = \begin{cases} (1 - p)^i p & \text{para } i = 0, 1, 2, \dots \\ 0 & \text{en otro caso.} \end{cases}$$

Observemos las diferencias, en este caso la variable puede valer 0, mientras que en la definición original, la variable podía tomar solo enteros positivos. Notemos que en este caso la probabilidad de fracaso: $1 - p$ esta elevada a la i en lugar de $i - 1$. Todo es cuestión de enfoque. en este nuevo caso contamos el **número de fracasos** hasta el primer éxito, en este sentido *no contamos* el último ensayo, el cual finalmente resulta en éxito.

Variables Aleatorias Continuas

Empezamos con un concepto de gran importancia:

Función de distribución

Una función F es una **función de distribución** si F toma valores no negativos y

- 1 F es una función no decreciente, es decir, si $a < b$ entonces $F(a) \leq F(b)$
- 2 $\lim_{b \rightarrow \infty} F(b) = 1$
- 3 $\lim_{b \rightarrow -\infty} F(b) = 0$
- 4 F es continua por la derecha. Esto es, si b_n es una sucesión decreciente tal que $b_n \rightarrow b$, entonces $\lim_{n \rightarrow \infty} F(b_n) = F(b)$

Variables Aleatorias Continuas


El concepto de función de distribución no es privativo de las variables aleatorias continuas. De hecho, para toda variable aleatoria X podemos demostrar que:

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

es una función de distribución⁵.

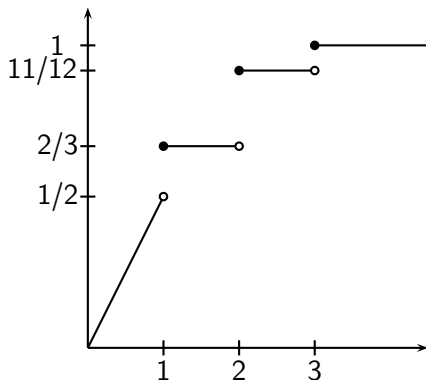
Ejemplo. La función de distribución de una variable aleatoria esta dada por:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2} & 0 \leq x < 1 \\ \frac{2}{3} & 1 \leq x < 2 \\ \frac{11}{12} & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$

⁵De hecho en un tratamiento mas formal de la probabilidad, la lógica es al revés. Definimos probabilidades en términos de funciones de distribución, quizás aquí no se percibe, pero de aquí surge la gran importancia de las funciones de probabilidad. 

Variables Aleatorias Continuas

Una representación gráfica de esta función es la siguiente:



Variables Aleatorias Continuas

Observemos por ejemplo que:

$$\begin{aligned}P(X < 3) &= \lim_n P\left(X \leq 3 - \frac{1}{n}\right) \\&= \lim_n F\left(3 - \frac{1}{n}\right) \\&= \frac{11}{12}\end{aligned}$$

Otro ejemplo:

$$\begin{aligned}P(X = 1) &= P(X \leq 1) - P(X < 1) \\&= F(1) - \lim_n F\left(1 - \frac{1}{n}\right) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}\end{aligned}$$

Sin embargo notemos por ejemplo que $P(X = 1.5) = 0$, esto debido a que F es continua en 1.5

Variables Aleatorias Continuas

Otro ejemplo

$$\begin{aligned}P\left(X > \frac{1}{2}\right) &= 1 - P\left(X \leq \frac{1}{2}\right) \\&= 1 - F\left(\frac{1}{2}\right) = \frac{3}{4}\end{aligned}$$

Finalmente:

$$\begin{aligned}P(2 < X \leq 4) &= P(X \leq 4) - P(X \leq 2) \\&= F(4) - F(2) = \frac{1}{12}\end{aligned}$$

Variables Aleatorias Continuas

Ahora si entramos en temas continuos. . .

Variables Aleatorias Continuas

Una variable aleatoria X se dice que tiene distribución continua con **función de densidad** f si para todo $a \leq b$ tenemos que:

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

De hecho para $B \subset \mathbb{R}$, tendríamos que⁶:

$$P(X \in B) = \int_B f(x) dx$$

⁶De manera formal deberíamos pedir que B fuera un conjunto *medible*, sin embargo no entraremos en esas complicaciones en este curso

Variables Aleatorias continuas

Sabemos que si X es continua:

$$\begin{aligned}P(X = x) &= P(X \leq x) - P(X < x) \\&= P(X \leq x) - \lim_n P(X < x - \frac{1}{n}) \\&= F(x) - \lim_n F(x - \frac{1}{n}) = 0\end{aligned}$$

La última igualdad se sigue de la continuidad de F . Por esta razón no tiene mucho sentido hablar de función de probabilidad cuando trabajamos con variables aleatorias continuas, sin embargo a manera de aproximación:

$$P(x \leq X \leq x + \Delta x) = \int_x^{x+\Delta x} f(y) dy \approx f(x) dx$$

Variables Aleatorias Continuas

Dada la definición de f , concluimos además que:

$$F(a) = P(X \leq a) = \int_{-\infty}^a f(y) dy$$

Esto tambien nos conduce a la conclusión de que:

$$\lim_{a \rightarrow \infty} F(a) = \int_{-\infty}^{\infty} f(y) dy = 1$$

Otra relación que suele ser de mucha utilidad es la siguiente:

$$\int_a^b f(y) dy = P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

lo cual era de esperarse.

Variables Aleatorias Continuas

Ejemplo. Supongamos que X es una variable aleatoria con función de densidad:

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

❶ ¿Cuál es el valor de C ?

❷ Encuentra $P(X > 1)$

Para encontrar el valor de C notemos que:

$$C \int_0^2 (4x - 2x^2) dx = 1$$

es decir:

$$C \left[2x^2 - \frac{2x^3}{3} \right] \Big|_0^2 = 1$$

Variables Aleatorias Continuas

Con lo cual concluimos que:

$$C = \frac{3}{8}$$

Por otro lado:

$$P(X > 1) = \int_1^{\infty} f(x) dx = \frac{3}{8} \int_1^2 (4x - 2x^2) dx = \frac{1}{2}$$

Ejemplo. Supongamos que la variable aleatoria X tiene función de distribución F_X y tiene densidad f_X . Encuentra la densidad de $Y = 2X$. Calculamos la densidad de Y como sigue:

$$\begin{aligned} F_Y(a) &= P(Y \leq a) \\ &= P(2X \leq a) \end{aligned}$$

Variables Aleatorias Continuas

por lo tanto:

$$\begin{aligned}F_Y(a) &= P(X \leq a/2) \\ &= F_X(a/2)\end{aligned}$$

de aqui tenemos que:

$$\begin{aligned}f_Y(a) &= \frac{d}{da}F_Y(a) \\ &= \frac{d}{da}F_X(a/2) \\ &= \frac{1}{2}f_X(a/2)\end{aligned}$$

Variables Aleatorias Continuas

La esperanza de una variable aleatoria continua con función de densidad la calculamos de manera parecida al caso discreto:

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

considerando esta igualdad, la varianza la definimos del mismo modo que en el caso discreto.

Ejemplo. Encuentra la esperanza y la varianza de una variable aleatoria con función de densidad:

$$f(x) = \begin{cases} 2x & \text{para } 0 \leq x \leq 1 \\ 0 & \text{en otro caso.} \end{cases}$$

Primero notemos que:

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^1 2x dx = 1$$

Por otro lado:

$$\begin{aligned} E[x] &= \int_{-\infty}^{\infty} xf(x) dx \\ &= \int_0^1 x \cdot 2x dx \\ &= \frac{2}{3} x^3 \Big|_0^1 = \frac{2}{3} \end{aligned}$$

Variables Aleatorias Continuas

Para la varianza calculamos:

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx \\ &= \int_0^1 x^2 \cdot 2x dx \\ &= \int_0^1 2x^3 dx \\ &= \frac{1}{2} x^4 \Big|_0^1 = \frac{1}{2} \end{aligned}$$

Por lo tanto

$$\text{Var}(X) = E[X^2] - E^2[X] = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

Variables Aleatorias Continuas

Ejemplo. Decimos que una variable aleatoria X se distribuye uniforme en el intervalo $(0, 1)$, $X \sim Unif(0, 1)$ si tiene función de densidad:

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Una variable aleatoria uniforme es sin duda la mas simple de las variables aleatorias continuas. Sin embargo su simplicidad no le resta importancia, por lo pronto es la piedra fundamental en todos los algoritmos de simulación.

Claramente $f(x) > 0$ para $x \in (0, 1)$ y $\int_0^1 f(x)dx = 1$. Por otro lado:

$$\begin{aligned} E[X] &= \int_0^1 xf(x)dx \\ &= \int_0^1 xdx \\ &= \frac{x^2}{2} \Big|_0^1 = \frac{1}{2} \end{aligned}$$

Variables Aleatorias Continuas

Para calcular la varianza, calculamos primero el segundo momento:

$$\begin{aligned} E[X^2] &= \int_0^1 x^2 f(x) dx \\ &= \int_0^1 x^2 dx \\ &= \left. \frac{x^3}{3} \right|_0^1 = \frac{1}{3} \end{aligned}$$

por lo tanto:

$$\text{Var}(x) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

Variables Aleatorias

Una versión mas general, seria una variable aleatoria uniforme en el intervalo (a, b) . En este caso, la función de densidad esta dada por:

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

Observemos que en este caso:

$$\begin{aligned} F(x) &= \int_a^x f(y) dy \\ &= \int_a^x \frac{1}{b-a} dy \\ &= \frac{x-a}{b-a} \end{aligned}$$

De donde se sigue inmediatamente que $F(x) = x$ para $X \sim Unif(0, 1)$

Variables Aleatorias Continuas

Ejemplo. Si $X \sim \text{Unif}(0, 1)$, demuestra que $Y = (b - a)X + a$ se distribuye uniforme en el intervalo (a, b) .

$$\begin{aligned}F_Y(x) &= P(Y \leq x) \\&= P((b - a)X + a \leq x) \\&= P\left(X \leq \frac{x - a}{b - a}\right) \\&= F_X\left(\frac{x - a}{b - a}\right) \\&= \frac{x - a}{b - a}\end{aligned}$$

Adicionalmente:

$$\begin{aligned}E[Y] &= E[(b - a)X + a] \\&= (b - a)E[X] + a \\&= \frac{b + a}{2}\end{aligned}$$

Variables Aleatorias Continuas

Un ejercicio sencillo que resulta de utilidad es el probar que:

$$\text{Var}(aX + b) = a^2 \text{Var}(x)$$

Usando este resultado tenemos que:

$$\text{Var}(Y) = \text{Var}((b - a)X + a) = (b - a)^2 \text{Var}(x) = \frac{(b - a)^2}{12}$$

Ejemplo. Decimos que una variable aleatoria X se distribuye **Exponencial** con parámetro $\lambda > 0$ si tiene función de densidad:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Variables Aleatorias Continuas

Algunas propiedades de la exponencial:

$$\begin{aligned} F(x) &= \int_0^x \lambda e^{-\lambda y} dy \\ &= 1 - e^{-\lambda x} \end{aligned}$$

Observemos que $\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$. Por otro lado:

$$\begin{aligned} E[X^k] &= \int_0^{\infty} x^k f(x) dx \\ &= \int_0^{\infty} x^k \lambda e^{-\lambda x} dx \\ &= -x^k e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} kx^{k-1} e^{-\lambda x} dx \\ &= \frac{k}{\lambda} \int_0^{\infty} x^{k-1} \lambda e^{-\lambda x} dx \\ &= \frac{k}{\lambda} E[X^{k-1}] \end{aligned}$$

Variables Aleatorias Continuas

Si hacemos $K = 1$ obtenemos que:

$$E[X] = \frac{1}{\lambda}$$

pero también, cuando $k = 2$:

$$E[X^2] = \frac{2}{\lambda} E[X] = \frac{2}{\lambda^2}$$

Por lo tanto

$$Var(x) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Variables Aleatorias Continuas

Ejemplo. Supongamos que la duración en minutos de una llamada telefónica se distribuye exponencial con parámetro $\lambda = 1/10$. Encuentra la probabilidad de que la duración de la llamada:

- 1 Dure mas de 10 minutos
- 2 Dure entre 10 y 20 minutos

Para responder el primer inciso calculamos:

$$P(X \geq 10) = 1 - F(10) = e^{-\lambda 10} = e^{-1} \approx 0.368$$

Para el segundo inciso:

$$P(10 \leq x \leq 20) = F(20) - F(10) = (1 - e^{-2}) - (1 - e^{-1}) \approx 0.233$$

Variables Aleatorias Continuas

Decimos que una variable aleatoria *no tiene memoria* si:

$$P(X > t + s \mid X > t) = P(X > s)$$

Esta igualdad es equivalente a

$$P(X > t + s) = P(X > s)P(X > t)$$

Esta propiedad de falta de memoria, puede interpretarse como sigue: si X es el tiempo de espera de cierto servicio, y ya hemos esperado t horas, la probabilidad de que que espemos $t + s$ horas es equivalente al evento donde esperamos s horas desde el principio. Es decir, es irrelevante que ya hayamos esperado t horas.

Distribución Normal

Decimos que una variable aleatoria X se distribuye **Normal** con parámetros μ y σ^2 , $X \sim N(\mu, \sigma^2)$ si tiene densidad:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty$$

La distribución Normal es sin duda una de las piedras fundamentales de la probabilidad y de aplicaciones de la probabilidad. Es tan útil que a menudo se abusa del uso del supuesto de normalidad, aun así en ciertos casos puede tolerarse cierto grado de imprecisión. Tantas son las ventajas de la normalidad que bien vale la pena sacrificar un modelo más sofisticado a favor de la flexibilidad que la normalidad otorga.

Empezamos por mostrar que:

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$

Distribución Normal

Esto no es tan directo como puede parecer a primera instancia. La densidad de la Normal no puede ser integrada de forma cerrada. Esto nos obliga a proceder de forma indirecta. Por lo pronto, empezamos por el cambio de variable: $y = (x - \mu)/\sigma$. Con este cambio de variable obtenemos:

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy$$

por lo tanto, basta que demostremos que:

$$I = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}$$

de hecho lo que vamos a demostrar es que:

$$I^2 = 2\pi$$

Distribución Normal

Para ello:

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy \end{aligned}$$

Esta integral se resuelve con mayor facilidad si transformamos las variables a coordenadas polares (es decir $x = r\cos\theta$ y $y = r\sin\theta$) de modo que:

$$\begin{aligned} I^2 &= \int_0^{\infty} \int_0^{2\pi} e^{-\frac{r^2}{2}} r d\theta dr \\ &= \int_0^{2\pi} d\theta \cdot \int_0^{\infty} e^{-\frac{r^2}{2}} r dr \\ &= 2\pi \int_0^{\infty} r e^{-\frac{r^2}{2}} dr \\ &= -2\pi e^{-\frac{r^2}{2}} \Big|_0^{\infty} = 2\pi \end{aligned}$$

Distribución Normal

Supongamos que $Y \sim N(\mu, \sigma^2)$ y $X = (Y - \mu)/\sigma$, tenemos entonces que:

$$\begin{aligned}F_X(x) &= P(X \leq x) \\&= P\left(\frac{Y - \mu}{\sigma} \leq x\right) \\&= P(Y \leq \sigma x + \mu) \\&= F_Y(\sigma x + \mu)\end{aligned}$$

Por lo tanto

$$\begin{aligned}f_X(x) &= \frac{\partial}{\partial x} F_Y(\sigma x + \mu) \\&= \sigma f_Y(\sigma x + \mu) \\&= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\end{aligned}$$

Es decir $X \sim N(0, 1)$.

Distribución Normal

Es tan importante una variable $X \sim N(0, 1)$, que amerita un nombre especial. Si $X \sim N(0, 1)$, decimos que X se distribuye **Normal Estándar**. Al proceso de sustituir una variable $Y \sim N(\mu, \sigma^2)$ por $X = (Y - \mu)/\sigma$ se le conoce como **estandarización**.

Un razonamiento análogo pero “inverso”, nos lleva a concluir que si $X \sim N(0, 1)$, entonces $Y = \sigma X + \mu \sim N(\mu, \sigma^2)$. Debemos recordar que en general es más fácil trabajar con variables normales estándar, razón por la cual es costumbre, estandarizar primero y luego de ser necesario, realizar la transformación $Y = \sigma X + \mu$.

Supongamos ahora que $Z \sim N(0, 1)$, entonces:

$$\begin{aligned} E[Z] &= \int_{-\infty}^{\infty} xf(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{x^2}{2}} dx \\ &= -\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} = 0 \end{aligned}$$

Distribución Normal

Por lo tanto:

$$\begin{aligned} \text{Var}(Z) &= E[Z^2] \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \left(-xe^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \\ &= 1 \end{aligned}$$

En resumen, si $Z \sim N(0, 1)$ entonces $E[Z] = 0$ y $\text{Var}(Z) = 1$. Ahora bien, sabemos que $Z = \sigma Z + \mu \sim N(\mu, \sigma^2)$ pero:

$$E[X] = \sigma E[Z] + \mu = \mu$$

y

$$\text{Var}[X] = \sigma^2 \text{Var}[Z] = \sigma^2$$

En resumen, si $X \sim N(\mu, \sigma^2)$ entonces $E[X] = \mu$ y $Var(X) = \sigma^2$. A σ se le conoce como **desviación estándar**.

También sabemos que si $Y \sim N(\mu, \sigma^2)$ entonces $X = (Y - \mu)/\sigma \sim N(0, 1)$. Es tan útil esta estrategia de estandarización que es usual denotar $F_X(x) = \Phi(x)$. Observemos porque es relevante; supongamos que $Y \sim N(\mu, \sigma^2)$ entonces:

$$F_Y(y) = P(Y \leq y) = P\left(X \leq \frac{y - \mu}{\sigma}\right) = \Phi\left(\frac{y - \mu}{\sigma}\right)$$

Esto es muy importante, pues nos permite encontrar la acumulada de Y en términos de la acumulada de una normal estándar.

Distribución Normal

Ejemplo. Sea $X \sim N(3, 9)$. Calcula $P(2 < x < 5)$, $P(X > 0)$ y $P(|X - 3| > 6)$

$$\begin{aligned}P(2 < X < 5) &= P\left(\frac{2-3}{3} < \frac{X-3}{3} < \frac{5-3}{3}\right) \\&= P\left(-\frac{1}{3} < \frac{X-3}{3} < \frac{2}{3}\right) \\&= \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{1}{3}\right) \approx 0.3779\end{aligned}$$

$$\begin{aligned}P(X > 0) &= P\left(\frac{X-3}{3} > \frac{0-3}{3}\right) \\&= P(Z > -1) \\&= P(-Z < 1) \\&= \Phi(1) \approx 0.8413\end{aligned}$$

En este último cálculo hemos usado el hecho de que si $Z \sim N(0, 1)$ entonces $-Z \sim N(0, 1)$

Distribución Normal

Por último:

$$\begin{aligned}P(|X - 3| > 6) &= P(X > 9) + P(X < -3) \\&= P(Z > 2) + P(Z < -2) \\&= 2P(Z < -2) \\&\approx 0.0456\end{aligned}$$