Curso Introductorio de Probabilidades

Autor: Dr. Juan Lucas Bali

Última fecha de actualización: 9 de marzo de 2022

Este apunte ha sido desarrollado en el contexto de la situación de pandemia COVID-19, a fin de complementar el material de aprendizaje de los estudiantes de una materia introductoria de Probabilidad y Estadística, tal como se daría en el curso de la materia correspondiente en la Escuela de Ciencia y Tecnología de la Universidad Nacional de San Martín (ECyT-UNSAM). Como tal, servirá como material primario y fundamental para el estudiante de esta materia y otras afines.

Creo que con este material se ha podido, parcialmente, resolver un problema importante que arrastraba la materia en diversas cursadas, que era la falta de un material oficial de estudio que sea autocontenido, incluyendo numerosa ejemplificación que pueda servir para la formación del estudiante. Por supuesto, no se espera que este material pueda reemplazar por completo la actividad docente propia de la profesión, sino como un complemento que ofrecería un primer acercamiento necesario al tema.

Algunas aclaraciones, por motivos de tiempo de cursada por supuesto no se han podido contemplar todos los temas que uno quisiera, en tal sentido se ha practicado una selección que consideramos apropiada para los cursantes de la misma, con un fuerte enfoque en temas de Probabilidad y con una introducción a los de Estadística. Soy de la firme creencia que es necesario, para un entendimiento cabal de la Estadística en sí, tener las bases de la teoría de Probabilidad bien firmes, de ahí el sesgo si se quiere que se podrá encontrar en este material.

Para una cursada normal de Probabilidad y Estadística, propongo el siguiente presupuesto de tiempo para los capítulos y sus temas:

- Capítulo 1 (Introducción a la Probabilidad): 1 semana.
- Capítulo 2 (Combinatoria): 1/2 semana.
- Capítulo 3 (Modelo de Laplaca): 1/2 semana.
- Capítulo 4 (Probabilidad Condicional): 1 semana.
- Capítulo 5 (Variable Aleatoria Discreta): 1 semana.
- Capítulo 6 (Familias de Distribuciones): 1 a 2 semanas.
- Capítulo 7 (Variable Aleatoria Continua): 1 semana.
- Capítulo 8 (Familias de Variables Continuas): 1 a 2 semanas.
- Capítulo 9 (Vectores Aleatorios): 1 a 2 semanas.
- Capítulo 10 (Teoremas Límite): 1 semana.
- Capítulo 11 (Introducción a la Estadística): 1 semana.

Índice

1. Introducción a la Probabilidad

1.1. ¿Qué es la "probabilidad"?

Consideremos estos experimentos:

1. Realizar un tiro de ruleta y registrar el resultado.
2. Arrojar una moneda y verificar de qué lado cayó.
3. Registrar la cantidad de abonados que levantan el tubo de teléfono entre las 11 y las 12.
4. Registrar en un kiosco dado el tiempo que transcurre desde el momento que abre hasta que llega el primer cliente.
5. Elegir una persona al azar y medir su peso.

Vamos a decir que en estos cinco problemas se realiza un **experimento aleatorio**. ¿Qué se entiende por experimento aleatorio? ¿Qué tienen en común?

1. Se conocen todos los resultados posibles.
2. Si realizo una sola vez el experimento no puedo saber de antemano qué resultado obtendré.

¿Son realmente aleatorios estos problemas? ¿No se podrían modelar determinísticamente? Es probable que varios de ellos sí, pero quizás presenten un modelo excesivamente complicado e impracticable, además de que en varios casos probablemente ni contemos con la información adecuada para poder alimentar las ecuaciones de ese modelo. En definitiva, estamos frente a presencia de o bien desinformación o bien azar implícito al problema. No vamos a ponernos a hilar fino en estos detalles, sobre el verdadero significado del "azar". Para un acercamiento a este tema, mas bien de tipo filosófico, recomiendo la lectura del siguiente libro :"Azar, Ciencia y Sociedad" (Jacovkis P.M. y Perazzo R., EUDEBA).

Pensemos en el problema más básico de probabilidad. Tiramos una moneda equilibrada, cuál es la probabilidad de que salga cara? La respuesta es 1/2 (para nosotros, la probabilidad será un número entre 0 y 1), no hace falta mucha teoría para afirmar eso. ¿Pero qué quiere decir ese "1/2"? ¿Que cada vez que arrojo la moneda la mitad de las veces sale cara y la otra mitad cruz? Está claro para todos que no es eso lo que estará ocurriendo. Hay que tratar de darle una conceptualización más formal a eso. Por otro lado, ¿qué pasa si tenemos situaciones más complicadas? Por ejemplo, se arrojan quince monedas, ¿cuál es la probabilidad de que exactamente en 8 de esas monedas salga cara? Claramente, esta pregunta es más difícil de contestar. Vamos a necesitar para eso un formalismo que nos permita responder este y otro tipo de preguntas.

Preguntas típicas:

¿Cuál se imaginan que será la probabilidad de sacar dos caras al arrojar dos monedas? ¿Y si son tres monedas y queremos tres caras? ¿Y si son n monedas?

Suponganse la siguiente apuesta: se paga un peso para participar y se arrojan 7 dados. Si sale 6 en los 7 dados les dan un millón de pesos. ¿Jugarían o no a este juego? ¿Cómo hago para darme cuenta si me conviene o no jugar? Estas cosas las vamos a ir viendo en el transcurrir del curso.

1.2. Espacio Muestral

Un **experimento aleatorio** es un mecanismo bien definido capaz de generar resultados. Lo daremos como algo que existe porque sí, no explicaremos formalmente la idea de "azar", como dijimos antes. El conjunto de los resultados posibles lo llamaremos **espacio muestral** y lo notaremos como Ω.

Por ejemplo, si estamos arrojando un dado, el espacio muestral para este experimento sería:

$$\Omega_1 = \{1, 2, 3, 4, 5, 6\} = \{i \in \mathbb{N}, 1 \leq i \leq 6\}.$$

Si estamos arrojando una moneda, podría ser

$$\Omega_2 = \{\text{ca}, \text{cr}\}.$$

Si son dos dados, tenemos

$$\Omega_3 = \{(1,1),(1,2),\ldots,(6,6)\} = \{(i,j) : i,j \in \mathbb{N}, 1 \leq i \leq 6, 1 \leq j \leq 6\} = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}.$$

Si arrojamos una moneda hasta que sale cara, tenemos

$$\Omega_4 = \{(\text{ca}), (\text{cr,ca}), (\text{cr,cr,ca}), \ldots\}.$$

Notemos que este espacio muestral es infinito (numerable), a diferencia de los anteriores.

Si estamos modelanzo el lanzamiento de una jabalina y medimos la distancia recorrida, podría ser:

$$\Omega_5 = \mathbb{R}_{>0}.$$

Este es también infinito pero nótese que a diferencia del anterior, es no numerable ("continuo"). Pero además, si tuviésemos que asignar una probabilidad a un elemento en particular, por ejemplo el 60.988, lo más razonable es que esta fuese cero. Justo la probabilidad de embocar a un "átomo" en particular tiene probabilidad cero, y así con cualquier número. Todos los resultados tienen probabilidad cero, veremos más adelante como trabajar con este tipo de experimentos "continuos".

1.3. Álgebra de Eventos

Un **evento** es un subconjunto cualquiera del espacio muestral. En rigor de verdad, no es "cualquier" subconjunto, pueden haber restricciones. A los fines prácticos de esta materia podríamos decir que casi cualquier subconjunto serviría. La definición más correcta sería un subconjunto del espacio muestral que tiene asignada una probabilidad.

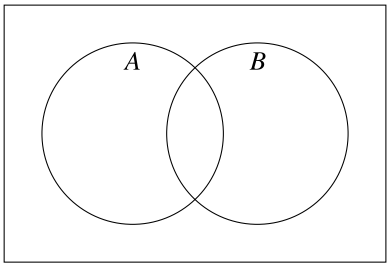
Por ejemplo, en el experimento de arrojar un dado, podemos considerar $A = \text{el dado salió par} = \{2,4,6\}$. Intuitivamente, aunque no la hayamos definido, su probabilidad en el caso en que el dado sea equilibrado es 1/2.

Tenemos las siguientes particularidades:

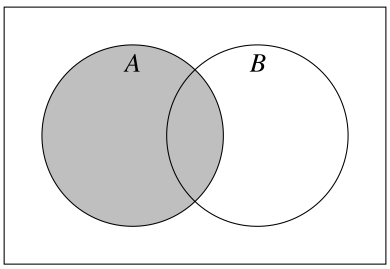
- Ω es un evento, conocido como el evento cierto o **seguro**.
- $\emptyset$ es el evento **vacío** o imposible.
- Si A y B son eventos, $A \cup B$ es el evento **unión** y ocurre cuando pasa cualquiera de los dos eventos, o los dos.
- Si A y B son eventos, $A \cap B$ es el evento **intersección** y ocurre cuando pasan ambos eventos a la vez.
- A^c es el evento complementario y ocurre cuando no pasa A.
- $A - B = A \cap B^c$ es el evento diferencia y ocurre cuando pasa A pero no pasa B.

Veamos gráficamente cómo serían estos casos.

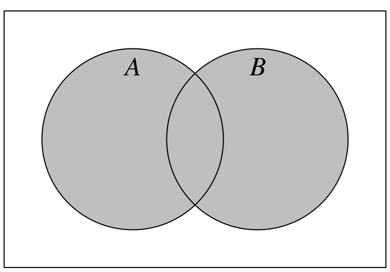
En general, si tenemos un espacio muestral Ω con dos eventos en su interios, A y B, podemos pensar el siguiente gráfico:



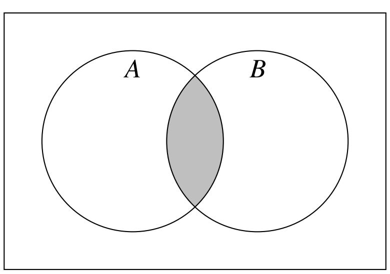
Con el evento A pintado, tenemos



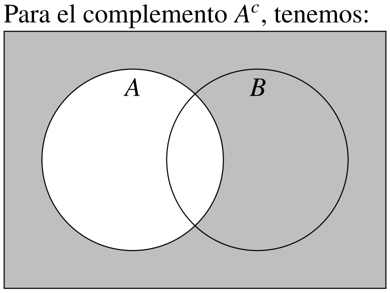
Vamos a ir rellenando cada una de las posibles regiones. Para el caso de la unión $A \cup B$, tendremos lo siguiente:



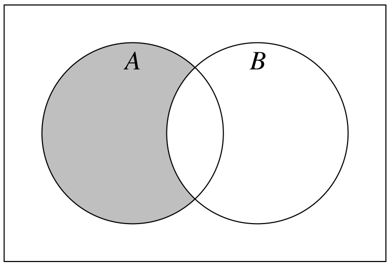
Para el caso de la intersección, $A \cap B$, nos queda:



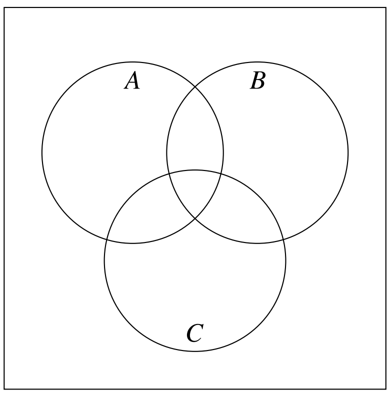
Para el complemento A^c, tenemos:



Notemos que la resta de conjuntos, es decir, $A - B$, se puede pensar como tomar el conjunto A e intersecarlo con lo que no está en B, es decir, con el complemento de B. Luego, $A - B = A \cap B^c$, lo cual puede resultar útil en algunos casos que ya iremos viendo.



Finalmente, mostremos un caso general con tres eventos:



Un ejemplo, con el experimento en donde se arroja un dado. Sea A el evento "el número es par" = $\{2, 4, 6\}$ y B el evento "el número es divisible por 3" = $\{3, 6\}$. Calculemos cada cosa:

- $A \cup B = \{2, 3, 4, 6\}$.
- $A \cap B = \{6\}$.
- A^c es el evento de los números impares, es decir $A^c = \{1, 3, 5\}$.
- $A - B = \{2, 4\}$.

Se dice que A está **contenido** en B y lo notamos como $A \subset B$ si cada vez que ocurre A ocurre B. Por ejemplo, en el caso de un dado, si $A =''$ el número es par$''$ y $B = \{4\}$ luego $B \subset A$.

Se dice que A y B son **disjuntos** (o mutuamente excluyentes) si $A \cap B = \emptyset$, es decir, no tienen nada en común ambos eventos. Por ejemplo, si A es obtener un número par y B uno impar, es evidente que ambos son disjuntos pues un número no puede ser a la vez par e impar. Es claro también que siempre A y A^c van a ser disjuntos.

Se dice que una colección de eventos $A_1, \dots, A_n$ conforma una **partición** si son entre sí disjuntos tomados de a dos, es decir, $A_i \cap A_j = \emptyset$ si $i \neq j$, y además $A_1 \cup \dots A_n = \Omega$. Es decir, los eventos $A_1, \dots, A_n$ particionan al espacio muestral. Notemos que siempre A y A^c formarán una partición.

Otro ejemplo, con el experimento de los dos dados, si A es "ambos números son iguales" y B ambos números son pares, luego $A \cap B = \{(2, 2), (4, 4), (6, 6)\}$.

1.3.1. Al menos, a lo sumo, exactamente

Supongamos que el experimento consiste en arrojar tres monedas y registrar los resultados. Un espacio muestral posible para esto sería:

$$\Omega = \{000, 001, 010, 011, 100, 101, 110, 111\}$$

donde 0 representa una cara y 1 una cruz.

Un posible evento sería A_1 =”salió al menos una cara”, notemos que $A_1 = \{000, 001, 010, 011, 100, 101, 110\}$. Definamos A_2 =”salió exactamente una cara”, tendremos que

$$A_2 = \{011, 101, 110\}$$

Es claro que $A_2 \subset A_1$. Otro evento posible, A_3 =”salió a lo sumo una cara”.

$$A_3 = \{011, 101, 110, 111\}.$$

En definitiva, el ”al menos” se asocia con la relación matemática mayor o igual, el ”exactamente” con el igual y el ”a lo sumo” con el menor o igual.

El complemento de A_1 se leería como “no salió ninguna cara”, y es claro que así visto tenemos que $A_1^c = \{111\}$.

Supongamos que B =”salió exactamente dos caras”. ¿Qué daría $A_3 \cup B$? Se leería como “saió a lo sumo dos caras”, y se describiría como

$$A_3 \cup B = \{001, 010, 011, 100, 101, 110, 111\}.$$

Finalmente, ¿cómo se leería $A_3 \cap B$? Debería dar vacío, pues es el evento en donde aparecen a lo sumo una cara y exactamente dos caras, no hay forma de cumplir estas condiciones simultáneamente, son contradictorias. Luego $A_3 \cap B = \emptyset$.

Otro evento, C =”las tres monedas dieron lo mismo”. Es claro que $C = \{000, 111\}$. Qué daría $A_1 \cap C$? Notar que es “al menos una cara” y “las tres monedas dieron lo mismo”. Como al menos hay una cara y todas las monedas dan lo mismo, es porque las tres son caras, luego $A_1 \cap C = \{000\}$.

Propiedades de conjuntos:

- Asociatividad: $(A \cup B) \cup C = A \cup (B \cup C)$ y $(A \cap B) \cap C = A \cap (B \cap C)$.
- Conmutatividad: $A \cup B = B \cup A$ y $A \cap B = B \cap A$.
- Distributiva: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ y $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$.
- Leyes de de Morgan: $(A \cup B)^c = A^c \cap B^c$ y $(A \cap B)^c = A^c \cup B^c$. Pensemos esto como decir lo siguiente: que NO ocurra A o B es lo mismo que decir que no ocurra A y que no ocurra B.

1.4. Axiomas de Probabilidad

En base a la noción de evento, uno podría dar una definición de probabilidad. Dado un evento $A \subset \Omega$, si se pudiera ejecutar el experimento n veces con $n \to \infty$, y n_A es la cantidad de veces que sucede A al ejecutar n veces el experimento, podemos entonces definir

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}.$$

El problema de esta definición es que es difícil de caracterizar este tipo de límite puesto que n_A es aleatorio, sería un límite “probabilístico”. La teoría matemática a partir de esta definición es complicada y por eso se prefiere el acercamiento aximático de Kolmogorov que es lo que daremos en este curso.

Ahora sí, vamos a caracterizar formalmente las propiedades que debe cumplir una **probabilidad**. La probabilidad es una función que dado un evento A devuelve un número $P(A) \in [0, 1]$. Los axiomas fundamentales son

A1 $P(A) \in [0, 1]$.

A2 $P(\Omega) = 1$.

A3 $P(\emptyset) = 0$.

A4 (aditividad) Si A y B son dos eventos disjuntos, es decir, con intersección nula, entonces $P(A \cup B) = P(A) + P(B)$.

En rigor de verdad, el último resultado no es formalmente un "axioma", un axioma más correcto sería el siguiente:

4A* (σ-aditividad). Si $A_1, A_2, \ldots$ es una sucesión de eventos disjuntos de a pares, luego $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Se puede probar, usando A3 y A4*, el axioma A4, aunque no al revés. Omitiremos estos detalles para no hacer más farragosa la lectura.

Algunos resultados.

De los axiomas se deducen estas propiedades. Las demostraciones son optativas pero seguirlas marcaría un buen punto para ver de entender la correcta comprensión de los axiomas.

Proposition 1.1. *(probabilidad del complemento)* $P(A^c) = 1 - P(A)$.

Demostración. $\Omega = A \cup A^c$, luego usando la aditividad $1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$ y de acá se despeja lo deseado. □

Proposition 1.2. *(monotonía de la probabilidad) Si* $A \subset B$, *luego* $P(A) \leq P(B)$ *y además* $P(B - A) = P(B) - P(A)$.

Demostración. $B = A \cup (B - A)$, de forma disjunta. Luego, $P(B) = P(A) + P(B - A)$ y de aquí se tiene lo primero. Lo segundo sale despejando. □

Proposition 1.3. *(principio de inclusión/exclusión) Sean A y B dos eventos no necesariamente disjuntos (o excluyentes), luego*

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

Demostración. $A \cup B = A \cup (B - A)$, siendo disjuntos así escritos. Luego, $P(A \cup B) = P(A \cup (B - A)) = P(A) + P(B - A)$. Ahora con cuidado, no puedo usar que $P(B - A) = P(B) - P(A)$ pues no es cierto que $A \subset B$. Sin embargo, podemos hacer el siguiente "truco". $B - A = B - (A \cap B)$ (no puedo restar "de más", convencerse haciendo el diagrama de Venn de esta igualdad), luego $P(B - A) = P(B - (A \cap B))$ y ahora sí tenemos que $A \cap B \subset B$, con lo cual tenemos que $P(B - A) = P(B - (A \cap B)) = P(B) - P(A \cap B)$. Juntando con lo de arriba, queda que $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ y listo. □

Este principio se puede generalizar a n eventos, nos limitaremos a exhibir aquí la expresión para tres eventos. Dejamos como curiosidad para el lector que postule la fórmula para cuatro eventos.

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

1.5. Ejemplos

1.5.1. Ejemplo 1

Se arroja un dado equilibrado. Sea E_i el evento "sale el número i". $\Omega = \{1, 2, 3, 4, 5, 6\}$ y $E_i = \{i\}$. Sea $A ='' \text{obtener un número par}'' = E_2 \cup E_4 \cup E_6$, que es una unión disjunta. Luego, $P(A) = P(E_2) + P(E_4) + P(E_6) = 1/6 + 1/6 + 1/6 = 1/2$.

Sea B el evento “sale un número divisible por 3”. Es fácil ver que $A \cap B = \{6\}$ y su probabilidad es 1/6. Consideremos el evento “sale un número par o divisible por 3”. Esto es $A \cup B$. Podemos caracterizarlo directamente, en este caso es $A \cup B = \{2, 3, 4, 6\}$ que tiene probabilidad $4/6$, o bien usamos las reglas de antes.

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1/2 + 1/3 - 1/6 = 4/6$.

Supongamos ahora que el dado está cargado y que la probabilidad de los números pares es el doble que el de los impares. Queremos calcular la probabilidad de que salga un número par. Ya no será más 1/2. Llamemos $p = P(E_1) = P(E_3) = P(E_5)$. Luego, $2p = P(E_2) = P(E_4) = P(E_6)$, con lo cual tenemos que $1 = P(E_1) + P(E_2) + P(E_3) + P(E_4) + P(E_5) + P(E_6) = 9p$, de acá se deduce que $p = 1/9$.

Entonces, $P(\text{salga par}) = P(E_2) + P(E_4) + P(E_6) = 6p = 6/9 = 2/3$.

1.5.2. Ejemplo 2

En un determinado barrio, el 60 % de los hogares tiene servicio de Internet, el 80 % tiene servicio de telefonía fija y el 50 % cuenta con ambos servicios. Si se escoge un hogar al azar:

1. ¿cuál es la probabilidad de que tenga al menos uno de los servicios?
2. ¿cuál es la probbabilidad de no tener ningún servicio?
3. ¿cuál es la probabilidad de tener exactamente servicio de Internet?
4. ¿cuál es la probabilidad de tener exactamente servicio de telefonía fija?
5. ¿cuál es la probabilidad de tener exactamente uno de los dos servicios?

Respuesta del primer inciso. Sea I el evento ”el hogar tiene Internet”, sea T el evento ”el hogar tiene telefonía fija”.

Nos dan como dato que $P(I) = 0{,}6$, que $P(T) = 0{,}8$ y, finalmente, nos dicen que los hogares que tienen **ambos** servicios son el 50 %, es decir, que $P(I \cap T) = 0{,}5$. Lo que nos piden es la probabilidad de **al menos** tener uno de los dos servicios, es decir, es $P(I \cup T)$. Usando el principio de inclusión/exclusión, nos sale que

$$P(I \cup T) = P(I) + P(T) - P(I \cap T) = 0{,}6 + 0{,}8 - 0{,}5 = 0{,}9.$$

Veamos la probabilidad de no tener ningún servicio. Eso es el complemento del anterior. En efecto, tendremos que

$$P((I \cup T)^c) = 1 - P(I \cup T) = 1 - 0{,}9 = 0{,}1.$$

Calculemos la probabilidad ahora la probabilidad de tener únicamente servicio de Internet.

Notemos lo siguiente, el mundo de Internet se divide entre aquellos que tienen Internet junto con telefonía, es decir, $I \cap T$, y por otro lado tenemos los que solamente tienen Internet y no tienen telefonía. Eso se refleja como $I \cap T^c$. Luego, I es una unión disjunta entre $I \cap T$ e $I \cap T^c$. Entonces, por la regla de aditividad de la probabilidad, tendremos así pues que

$$0{,}6 = P(I) = P(I \cap T) + P(I \cap T^c) = 0{,}5 + P(I \cap T^c)$$

de donde deducimos que la probabilidad pedida es $P(I \cap T^c) = 0{,}1$.

Un razonamiento similar lo podemos efectuar para ver la proporción de hogares que solamente tienen telefonía fija, obteniendo que

$$0{,}8 = P(T) = P(T \cap I) + P(T \cap I^c) = 0{,}5 + P(T \cap I^c)$$

de donde despejamos que $P(T \cap I^c) = 0{,}3$. O sea, el 30 % de los hogares tiene únicamente telefonía fija.

Finalmente, queremos los hogares que tienen exactamente un servicio. Es decir, o tienen telefonía fija nada más o tienen Internet nada más. Eso sería la unión disjunta entre los que tienen solamente Internet con los que tienen solamente teléfono.

Luego, lo que nos piden es

$$P(\text{exactamente un servicio}) = P((I \cap T^c) \cup_d (T \cap I^c)) = P(I \cap T^c) + P(T \cap I^c) = 0{,}1 + 0{,}3 = 0{,}4.$$

O sea, el 40 % de los hogares tiene únicamente un servicio.

Esto se podría haber deducido también de hacer $P(I \cup T) - P(I \cap T) = 0{,}9 - 0{,}5 = 0{,}4$. Es decir, a la proporción que tiene al menos uno de los dos servicios le resto la proporción de los que tienen los dos servicios.

1.6. Espacios Equiprobables

Supongamos que Ω tiene cardinal finito y que además cada elemento del espacio muestral tiene la misma probabilidad p. Tenemos que

$$1 = \sum_{\omega \in \Omega} P(\{\omega\}) = \sum_{\omega \in \Omega} p = |\Omega| p$$

con lo cual $p = 1/|\Omega| = 1/n$, si n es el cardinal del espacio muestral. Supongamos que queremos calcular la probabilidad de un evento A de cardinal n_A. Entonces,

$$P(A) = \sum_{\omega \in A} p = |A| p = n_A p = \frac{n_A}{n}.$$

Es decir, para calcular la probabilidad del evento A debemos hacer el cociente entre el cardinal de A ("casos favorables") sobre el cardinal del espacio muestral ("casos totales o posibles").

1.6.1. Ejemplo 1

Se tiene una urna que contiene dos bolillas blancas y tres rojas. Se extraen **con reposición** dos bolillas de la urna. Es decir, se extrae una bolita, ser la observa, y se la vuelve a colocar en la urna para volver a extraer.

1. ¿Cuál es la probabilidad de que ambas bolitas sean blancas?
2. ¿Cuál es la probabilidad de que haya al menos una bolita roja?
3. ¿Cuál es la probabilidad de que la primer bolita sea roja y la segunda blanca?

Vamos a responder la primer pregunta. Un posible espacio muestral sería

$$\Omega = \{(x_1, x_2) : \ x_i \in \{B, R\}\} = \{RR, BB, RB, BR\}.$$

Si bien técnicamente este espacio muestral es "válido", no nos resultará práctico puesto que no es equiprobable. Es intuitivo ver que RR tiene más probabilidad que BB.

El espacio muestral que consideraremos es

$$\Omega = \{(x_1, x_2) : \ x_i \in \{B_1, B_2, R_3, R_4, R_5\}\}.$$

El cardinal de Ω es 5x5 = 25, y las extracciones con cada par posible (numerado) tendrá la misma probabilidad (las bolitas se han numerado para tal fin, no hay nada que distinga un par de otro). Sea A = "ambas bolitas son blancas" = $\{(x_1, x_2) : \ x_i \in \{B_1, B_2\}\}$, que tiene cardinal 2x2 = 4. Luego, la probabilidad de A es $P(A) = 4/25$.

Para la segunda pregunta, notemos que lo que nos están pidiendo es el complemento del primer conjunto, con lo cual $P(A^c) = 1 - P(A) = 21/25$. Se podría haber calculado de forma directa, como sigue:

$$P(\text{al menos una roja}) = P(\text{exactamente una roja}) + P(\text{exactamente dos rojas}) = \frac{3 \cdot 2 + 2 \cdot 3}{5 \cdot 5} + \frac{3 \cdot 3}{5 \cdot 5} = 21/25.$$

El primer sumando expresa la idea de que podría salir primero roja y después blanca, o al revés. Tenemos que distinguir esas dos situaciones.

Veamos la tercer pregunta. Definimos el evento $B = \{(x_1, x_2) : \ x_1 \in \{R_3, R_4, R_5\}, \ x_2 \in \{B_1, B_2\}\}$. El cardinal de B es 3x2, luego la probabilidad de B es 6/25.

1.6.2. Ejemplo 2

Mismas preguntas que antes, pero ahora las bolitas se extraen *sin reposición*. Es decir, se extrae una bolita, se la observa y se la descarta, no vuelve a ingresar a la urna. Tenemos el siguiente espacio muestral $\Omega = \{(x_1, x_2) : x_i \in \{B_1, B_2, R_3, R_4, R_5\}, x_1 \neq x_2\}$. Notemos que ahora su cardinal es 5x4 = 20, sigue siendo equiprobable pues no hay ningún par numerado que se distinga sobre otro. Es decir, tengo 5 opciones para la primer bolita, y por cada una de las 5 opciones me quedan 4 para la segunda, pues no se puede repetir con la primera.

Veamos la primer pregunta, el evento es ahora $A = \{(x_1, x_2) : \ x_i \in \{B_1, B_2\}, x_1 \neq x_2\}$ que tiene cardinal 2x1 = 2, luego la probabilidad de A es 2/20.

La segunda pregunta se responde igual que en el primer ejemplo, y la probabilidad es 18/20.

Para la tercer pregunta, tenemos ahora el evento $B = \{(x_1, x_2) : \ x_1 \in \{R_3, R_4, R_5\}, \ x_2 \in \{B_1, B_2\}, x_1 \neq x_2\}$, con cardinal 3x2 = 6, luego la probabilidad es $6/20$.

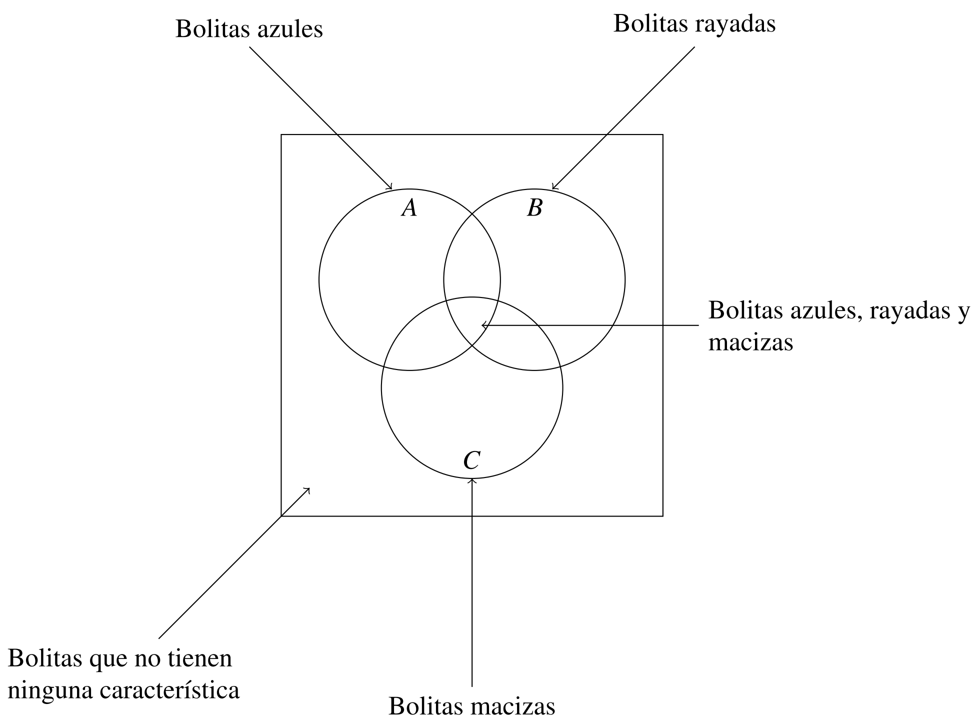
Agregamos una cuarta pregunta, calcular la probabilidad de que salga una bolita roja y una blanca, sin importar el orden.

Para eso, definimos $B_1 = \{(x_1, x_2) : \ x_1 \in \{R_3, R_4, R_5\}, \ x_2 \in \{B_1, B_2\}, x_1 \neq x_2\}$ y $B_2 = \{(x_1, x_2) : \ x_2 \in \{R_3, R_4, R_5\}, \ x_1 \in \{B_1, B_2\}, x_1 \neq x_2\}$. Ambos eventos son disjuntos y la unión es exactamente lo que nos piden. Luego, la probabilidad será $P(B_1) + P(B_2) = 6/20 + 6/20 = 12/20$.

Para problemas más complicados, tendremos que recurrir a algunas técnicas combinatorias de conteo, que es el próximo capítulo.

1.6.3. Ejemplo con tres eventos

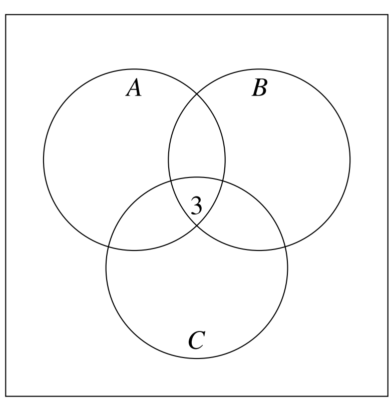
Supongamos que tenemos canicas que pueden tener tres características: ser azules o no, ser rayadas o no, ser macizas o no. Las condiciones pueden venir combinadas de cualquier forma, con lo cual tenemos entonces en general esta situación:



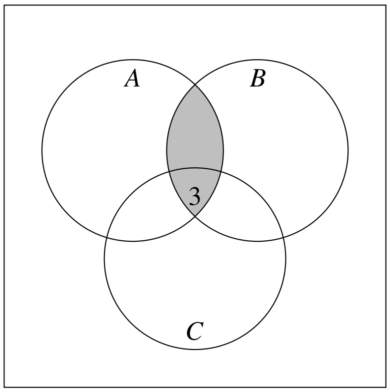
Nos dan como dato que hay 32 bolitas en total, de las cuales hay 3 azules, rayadas y macizas. Luego, tenemos 7 que son azules y rayadas, 5 azules y macizas y 6 que son rayadas y macizas. Por último, tenemos que hay 10 bolitas azules, 12 rayadas y 11 macizas. Nos piden lo siguiente:

1. Calcular la probabilidad de elegir una bolita al azar y que no responda a ninguna característica.
2. Calcular la probabilidad de elegir una bolita y que resulte azul, rayada pero no maciza.

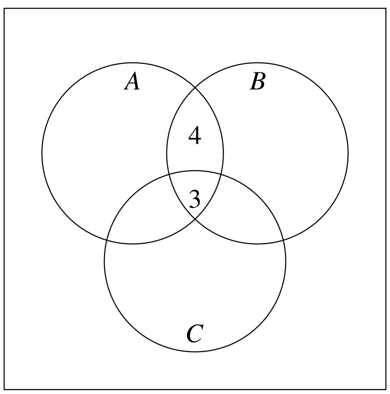
Vamos a proceder a ir, de adentro hacia fuera, ”pintando” las regiones con los datos que nos dan. Para empezar, tenemos que hay 3 bolitas que cumplen las tres características, irán a parar a la triple intersección:



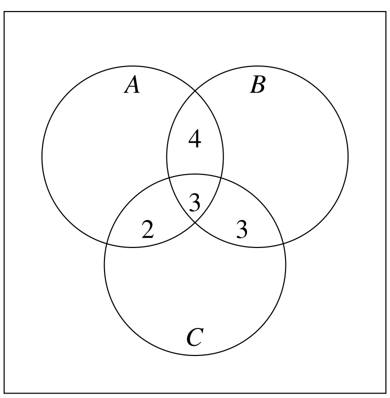
Ahora, veamos el hecho de que haya 7 que son azules y rayadas. Esto establece la siguiente región:



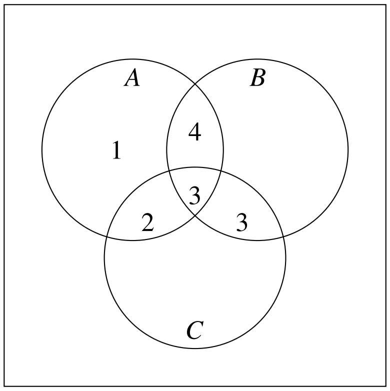
Toda esa región totaliza 7, contando los tres en la triple intersección. Así pues, nos queda



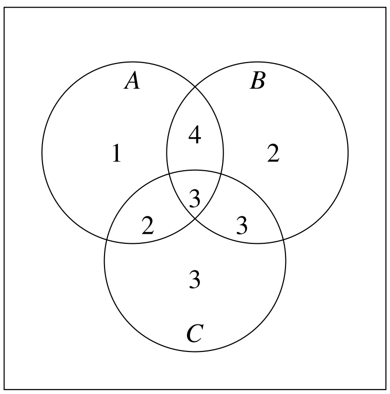
De la misma forma, vamos completando los otros datos de intersección de a pares:



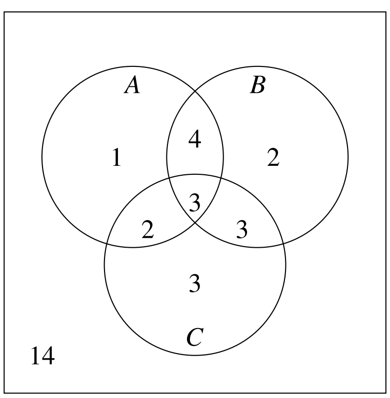
Luego, tenemos los datos de los conjuntos individuales. Tenemos 10 bolitas azules en total, a los cuales tenemos que tener en consideración 2, 3 y 4 que ya están pintadas en esa región. Queda pues 1 bolita que es exclusivamente azul:



Completando con los datos de 12 rayadas y 11 macizas tenemos pues que

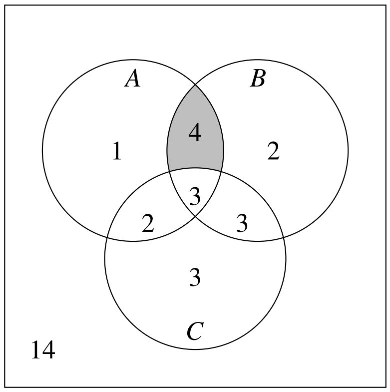


Finalmente, las bolitas que no responden a ninguna categoría serán 32 menos la suma de todos los números indicados



Así pues, respondiendo a la primera pregunta, la probabilidad de que una bolita no sea de ninguna de las tres clases será $14/32$.

Veamos la segunda pregunta, pintemos la región de las bolitas que son azules, rayadas y no macizas:



Así pues, lo pedido será $4/32$.

1.6.4. Ejemplo 4: jugando al Truco

El truco es un juego de cartas que se juega con 40 naipes de la baraja española, distribuidas en 10 cartas por palo (oros, bastos, espadas y copas). Se reparten tres cartas a cada jugador, y existen diversas combinaciones que pueden tener valor para la partida. Por ejemplo, una de ellas es lo que se conoce como ”flor”, sacar tres cartas del mismo palo.

Vamos a proceder a calcular eso, en un caso más simplificado primero, viendo la probabilidad de obtener flor de oros. Es decir, tres cartas de oro.

$$P(\text{flor de oros}) = \frac{10 \cdot 9 \cdot 8}{40 \cdot 39 \cdot 38} = 0{,}0121.$$

Las extracciones son sin reposición, por eso se va ”decrementando” el número tanto en el numerador como en el denominador. En el numerador, estamos considerando los casos favorables empezando desde 10 opciones de oros para la primer carta, a lo que le quedan 9 para la segunda y finalmente 8 para la tercera. Un razonamiento similar, pero ahora con la totalidad de las cartas, se aplica para el denominador.

Ahora sí, calculemos la probabilidad de obtener una flor, de cualquier palo.

$$P(\text{flor}) = P(\text{flor de oros}) + P(\text{flor de espadas}) + P(\text{flor de copas}) + P(\text{flor de bastos}).$$

Hemos usado aquí la aditividad. Notemos que las distintas flores con distintos palos son disjuntos entre sí, no puedo tener a la vez una flor de oros y una de espadas. Por otro lado, cada una de las 4 probabilidades debería ser igual pues no hay nada que distinga a un palo de otro en términos de probabilidades, la cuenta sería la misma en cada uno de ellos. Luego, nos queda

$$P(\text{flor}) = 4 \cdot P(\text{flor de oros}) = 4 \cdot 0{,}0121 = 0{,}0484.$$

Veamos ahora la probabilidad de sacar un envido de espadas. Interpretamos esto como obtener en la mano de tres cartas dos cartas que sean de espada y una remanente que sea de otro palo.

Luego, tenemos:

$$P(\text{envido de espadas}) = \frac{10 \cdot 9 \cdot 30 + 10 \cdot 30 \cdot 9 + 30 \cdot 10 \cdot 9}{40 \cdot 39 \cdot 38} = 3\frac{30 \cdot 10 \cdot 9}{40 \cdot 39 \cdot 38} = 0{,}1366.$$

Notemos que hemos debido descomponer los casos favorables en tres partes, en función de en qué momento surge la carta que no es de espada (el 30), son casos distintos.

De forma análoga, es fácil deducir que

$$P(\text{envido}) = 4 \cdot P(\text{envido de espadas}) = 4 \cdot 0{,}1366 = 0{,}5466.$$

Así pues, la probabilidad de obtener un envido o una flor se puede calcular por aditivad, teniendo que

$$P(\text{envido o flor}) = P(\text{envido}) + P(\text{flor}) = 0{,}5466 + 0{,}0484 = 0{,}595.$$

O sea, parece bastante probable obtener un buen juego (envido o flor) en la mano. Y si nos restringimos a envidos "buenos"? Por ejemplo, un envido de 33 que se consigue con un 7 y un 6 del mismo palo. Calculemos eso.

$$P(\text{envido de 33}) = \frac{4(30 \cdot 2 \cdot 1 + 2 \cdot 30 \cdot 1 + 2 \cdot 1 \cdot 30)}{40 \cdot 39 \cdot 38} = \frac{4 \cdot 3 \cdot 2 \cdot 30}{40 \cdot 39 \cdot 38} = 0,0122.$$

El primer 4 surge de elegir el palo entre los 4 posibles. Luego, el 30 es para marcar la carta que debe ser de otro palo (estamos excluyendo la flor), y los 2 y el 1 son para elegir el 7 y el 6 del palo ya escogido. No parece muy probable esa jugada...

Ejercicio 1.4. *Calcule la probabilidad de obtener en la mano el ancho de espada o el ancho de basto.*

1.7. Ejercitación

1. *a*) Defina un espacio muestral adecuado para cada uno de los siguientes experimentos aleatorios:
    1) Se arroja una moneda.
    2) Se arroja un dado.
    3) Se arrojan dos dados.
    4) Se arrojan n dados, siendo n fijo.
    5) Se arroja un dado hasta que salga un as.

   b) Calcule el cardinal de cada uno de ellos.

2. Supongamos que Ω es un espacio muestral con tres elementos, a, b y c donde tenemos definida una probabilidad P tal que $P(\{a\}) = 1/2$, $P(\{b\}) = 1/3$ y $P(\{c\}) = 1/6$.

   a) Explicite los 8 posibles subconjuntos (eventos) que puede tener este espacio muestral.

   b) Calcule la probabilidad de cada uno de ellos.

3. *a*) Considere el experimento de arrojar dos dados, describa explícitamente los elementos de cada uno de los siguientes eventos y calcule su probabilidad.
    - A: la suma de los números obtenidos es *exactamente* 5.
    - B: la suma de los números obtenidos es *a lo sumo* 5.
    - C: la suma es *al menos* 5.
    - D: el valor obtenido en el primer dado es superior al segundo.
    - E: el valor obtenido en el primer tiro es un 4.

   b) Describa los elementos de los siguientes conjuntos y calcule su probabilidad:
    - $A \cap D$.
    - $D \cup E$.
    - $A \cap (D \cup E)$.

   c) Si tuviese que elegir el número más probable para la suma de dos dados, ¿cuál elegiría?

4. Se sabe que el 75 % de las personas de una determinada localidad tiene Facebook, el 15 % tiene WhatsApp y hay un 15 % que no tiene ninguno de los dos aplicativos.

   a) ¿Cuál es la probabilidad de que una persona use ambos aplicativos?

b) ¿Cuál es la probabilidad de que una persona tenga Facebook pero no WhatsApp?

5. Se tienen dos barajas de cartas francesas (52 cartas distribuiídas en 4 palos).

   a) Se extrae una carta de una baraja. ¿Cuál es la probabilidad de que sea un as de corazones?
   b) Se extrae una carta de cada baraja. ¿Cuál es la probabilidad de que ambos sean el as de corazones?
   c) Se extrae una carta de cada baraja. ¿Cuál es la probabilidad de que haya exactamente un as de corazones?
   d) Se extrae una carta de cada baraja. ¿Cuál es la probabilidad de que por lo menos haya un as de corazones?
   e) Repita b), c) y d) suponiendo que ahora se extraen dos cartas de una misma baraja, sin reposición.

6. Una caja tiene 3 bolitas rojas, 2 azules y 4 blancas. Se extraen dos bolitas CON reposición.

   a) Calcule la probabilidad de obtener dos bolitas rojas.
   b) Calcule la probabilidad de obtener una bolita roja y una azul.
   c) Calcule la probabilidad de no obtener una bolita roja.
   d) Calcule la probabilidad de obtener al menos una bolita roja.
   e) Calcule la probabilidad de obtener a lo sumo una bolita roja.
   f) Calcule la probabilidad de que entre las bolitas extraídas haya una bolita azul o una roja.
   g) Calcule la probabilidad de obtener dos bolitas rojas.
   h) Calcule la probabilidad de obtener dos bolitas del mismo color.
   i) Repita lo anterior pero suponiendo extracciones SIN reposición.

7. Se arroja un dado 6 veces. ¿Cuál de las siguientes opciones ofrece la mejor chance de ganar? ¿O son equivalentes?

   a) Ganar 1$ si sale al menos un as.
   b) Ganar 1$ si sale un as todas las veces.
   c) Ganar 1$ si sale la secuencia 1,2,3,4,5,6 (en ese orden).
   d) Ganar 1$ si los dos primeros números son iguales.

8. Se tienen 28 residuos, de los cuales 19 son orgánicos, 20 hospitalarios y 13 industriales. Además, 7 de los residuos son orgánicos, hospitalarios e industriales, 10 son hospitalarios e industriales , 9 son orgánicos e industriales y 2 son inorgánicos, hospitalarios y de origen no industrial. Se extrae un residuo al azar, obtener la probabilidad de:

   a) que sea orgánico pero no hospitalario ni industrial.
   b) que sea hospotalario, orgánico y no industrual.

9. La urna A tiene 3 bolitas rojas y 5 azules, la urna B tiene 4 rojas y 7 azules. Se extrae una bolita de cada urna, calcule la probabilidad de que ambas sean del mismo color.

10. Se busca tener una idea del nivel de fragmentación que puede tener un parlamento determinado. Para eso, se define, dado un parlamento con una composición partidaria dada, el *índice de fragmentación* como la probabilidad de que al elegir dos parlamentarios distintos estos sean de distintos partidos. España al año 2020 dispone de un parlamento bicameral, el Senado y el Congreso de los Diputados, con la siguiente composición:

    Senado (total de 265 bancas)

Partido	# Representantes
Socialista (PSOE)	113
Popular (PP)	97
ERC-EH Bildu	15
Vasco	10
Ciudadanos	9
Izquierda Confederal	6
Nacionalista	6
Otros	8

Congreso de los Diputados (total de 350 bancas)

Partido	# Representantes
Socialista (PSOE)	120
Popular (PP)	88
Vox	52
Unidas Podemos	35
Republicano	13
Plural	12
Ciudadanos	10
Vasco	6
EH Bildu	5
Otros	9

Según el índice fragmentación, ¿cuál de las dos cámaras presenta mayor nivel de fragmentación? ¿Bajo qué situación puede una cámara parlamentaria tener un índice de 0? ¿Y un índice de 1?

11. ¿Supongamos que no hay años bisiestos, calcula la probabilidad de que en un grupo de 20 personas hayan al menos dos de ellos que cumplan el mismo día?

12. (*) (Nota: para este ejercicio se debe tener postulado una fórmula para el principio de inclusión/exclusión para 4 eventos) Se tiene 4 parejas de hombres y mujeres en una salón de baile. Se mezclan de forma tal que cada hombre puede llegar a bailar con cualquier mujer.
    - Calcule probabilidad de que el hombre 1 baile con su mujer.
    - Calcule la probabilidad de que el hombre 1 y el 2 bailen con sus respectivas mujeres.
    - Calcule la probabilidad de que ningún hombre baile con su mujer.

13. (Entropía) Dado un experimento aleatorio con n resultados posibles $r_1, \dots, r_n$ con probabilidades $p_1, \dots, p_n$, se define la **entropía** del experimento aleatorio como

$$H = -\sum_{i=1}^{n} p_i \log(p_i).$$

La entropía permite medir de alguna manera la incertidumbre de un mecanismo de generación aleatoria. Considerando la entropía binaria (es decir, con un logaritmo en base 2), permite dar una idea de compresibilidad de la información en bits. Eso es, la cantidad promedio de bits necesarios para codificar eficientemente los datos provenientes del experimento aleatorio.

a) Suponga un mecanismo binario, es decir, con dos resultados posibles, 0 y 1, con probabilidad p y $1 - p$ respectivamente. Calcule el valor de la entropía ante la situación de ”desconocimiento perfecto”, es decir, cuando $p = 1/2$, situación de azar perfecto. Compárelo con el caso $p = 1/3$, en donde ya sigue habiendo azar pero más reducido.

b) Bajo la misma situación, considere los casos $p = 0$ y $p = 1$, es decir, ausencia de azar (el mecanismo siempre devuelve el mismo valor). Para evaluar, informalmente, el caso ”$0 \cdot \log(0)$” recurra a un argumento de límite.

c) Demuestre que la entropía en el caso de un mecanismo binario (dos resultados posibles) se hace máxima cuando $p = 1/2$, es decir, bajo equiprobabilidad.

d) (*) Generalice el resultado al caso en donde el experimento puede generar n resultados posibles. Pista: considere el uso de multiplicadores de Lagrange.

14. (Nota: puede dejar este ejercicio para un poco después) (El problema de Monty Hall) Llega usted a un concurso televisivo, con la intención de ganarse un auto. Monty le ofrece tres puertas (1,2 y 3), en donde detrás de dos de ellas hay una cabra y en otra está el auto, que usted anhela. Comienza el juego, elige usted una puerta. Antes de abrirla, Monty lo detiene y le abre otra puerta, una que usted no eligió, mostrándole una cabra. Le ofrece la posibilidad de quedarse con la puerta originalmente elegida o bien cambiar a la otra puerta que sigue cerrada. ¿Qué elige, seguir en la misma puerta o cambiar?

2. Combinatoria

2.1. Argumento multiplicativo

Cuando trabajamos con espacios equiprobables, notamos que el problema de determinar una probabilidad en buena medida se reduce a uno de conteo: obtener la cantidad de elementos de algunos conjuntos y luego ensayar un cociente.

Por tal motivo, en este capítulo haremos una introducción a una disciplina matemática conocida como **análisis combinatorio**, que se centra en la cuestión de cómo contar de forma ”eficiente”.

Veremos estos temas a través de una serie de ejemplos guiados. A medida que avancemos iremos introduciendo la teoría.

Ejercicio 2.1. *¿Cuántas posibles patentes de auto existen?*

Resolución:

Una patente se compone de dos dígitos, tres letras y dos dígitos. Es decir, tendremos

$$\text{DDAAADD}$$

donde ”A” es una letra del abecedario (26 opciones) y D es un dígito decimal (10 opciones). Pensemos que por cada primer dígito (que son 10 opciones), tenemos a su vez 10 opciones para el segundo dígito, y por cada elección de $10 \cdot 10$ opciones para los dos primeros dígitos, tenemos 26 opciones para la primer letra... y así. En definitiva, nos quedan como cantidad de opciones:

$$10 \cdot 10 \cdot 26 \cdot 26 \cdot 26 \cdot 10 \cdot 10.$$

De alguna manera, estamos ensayando un argumento multiplicativo, puesto que si graficamos esto notaremos que hay una ramificación natural.

Ejercicio 2.2. *¿Cuántas posibles patentes de auto existen con todos símbolos distintos?*

Razonando de forma similar, pero la diferencia es que ahora habiendo elegido entre 10 primeros dígitos, tenemos 9 opciones para el segundo dígito, luego 26 letras, luego 25 y así. Nos queda pues

$$10 \cdot 9 \cdot 26 \cdot 25 \cdot 24 \cdot 8 \cdot 7.$$

Se va reduciendo el conteo puesto que no queremos que se repitan símbolos con respecto a los que venían antes.

Ejercicio 2.3. *¿Cuántas posibles patentes de auto existen con dígitos pares?*

Hay 5 dígitos pares, con lo cual tendremos

$$5 \cdot 5 \cdot 26 \cdot 26 \cdot 26 \cdot 5 \cdot 5.$$

Ejercicio 2.4. *¿Cuántas patentes de auto ”capicúa” existen?*

Recordemos que ser capicúa es que se lea igual de izquierda a derecha como de derecha a izquierda. Podemos pensar que el primer dígito es ”libre”, el segundo también, la primer letra también, la segunda también, pero ya la tercer letra no, debe coincidir con la primer letra. De la misma forma, el tercer dígito debe coincidir con el segundo, y el cuarto dígito debe coincidir con el primero. Así pues, nos queda

$$10 \cdot 10 \cdot 26 \cdot 26 \cdot 1 \cdot 1 \cdot 1.$$

2.2. Permutaciones

Supongamos que tenemos una ”palabra” compuesta por distintas letras $l_1, l_2, \ldots, l_n$. Queremos obtener la cantidad de permutaciones, es decir reordenamientos, de la palabra en cuestión. La forma de pensar esto es considerar que tenemos n celdillas.

$$\_\_\_\_\ldots\_$$

Para la primer celda, tenemos n opciones distintas. Para cada una de las n opciones de la primer celda, tenemos $n-1$ opciones para la segunda celda (puesto que no se debe repetir el elemento), con lo cual vamos teniendo $n \cdot (n-1)$ para las dos primeras celdas. Extendiendo naturalmente esto nos queda que el total de permutaciones será $n(n-1)\ldots 2 \cdot 1$. Es decir, $n!$, el **factorial** de n.

Veamos un ejemplo sencillo.

Ejercicio 2.5. *¿Cuántas permutaciones tiene la palabra CARBON?*

Dado que son 6 letras (distintas) es sencillamente

$$6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 6!.$$

¿Qué pasa si tenemos letras repetidas?

Cambia un poco el planteo, veamos un ejemplo:

Ejercicio 2.6. *¿Cuántas permutaciones tiene la palabra CARBONO?*

En principio, pensemos que las letras son todas distintas. Es decir, tenemos $CARBO_1NO_2$. Los posibles reordenamientos que ofrece esto en principio es 7!. Ahora bien, pensemos que estas dos permutaciones AO_1RCBNO_2 y AO_2RCBNO_1 responden en el fondo a la misma permutación, los estamos contando como dos casos distintos cuando en realidad son uno mismo. En ese sentido, pasará esto con cada una de las 7! permutaciones, estamos contando dos veces de más. Para arreglarlo, dividimos por 2, que son los posibles reordenamientos de las letras repetidas O. Así pues, el resultado de este problema es $7!/2$.

Supongamos que fuese tres las letras repetidas. Por ejemplo, CARBONOO. Entonces, tendríamos en principio 8! permutaciones suponiendo a todas las letras distinta. Pues bien, para arreglar el problema de las tres letras repetidas, debemos dividir ahora por 3!, que son las posibles permutaciones de $O_1O_2O_3$, que son los casos que estamos contando de más.

Luego, quedará $8!/3!$.

Ejercicio 2.7. *¿Cuántas permutaciones tiene la palabra ANANA?*

Son 5 letras, con 3 repeticiones de A y 2 de N, con lo cual nos queda en principio 5! para los reordenamientos distinguidos de las 5 letras, y anulamos las repeticiones de la letra A dividiendo por 3!, y para anular las repeticiones de las N dividimos por 2!.

Así pues, queda:

$$\frac{5!}{3!2!}.$$

2.3. Variaciones

Nuevamente, explicaremos el concepto a través de un ejemplo. Supongamos nuevamente que tenemos la palabra CARBON, todas las letras distintas. Queremos ahora contar todas las posibles palabras de cuatro letras distintas que se pueden formar con las letras de CARBON.

Tenemos un problema de 4 celdas, lo que conformaría una palabra de cuatro letras. La primer letra tiene 6 opciones, la segunda 5 (para no repetirse con la primera), la tercera 4 y la cuarta 3.

Es decir,

$$6 \cdot 5 \cdot 4 \cdot 3.$$

Notemos que en el fondo esto es hacer $6!/2!$. Es decir, a la cantidad total de permutaciones lo divido por las letras que faltarían para completar el total. Es decir, podríamos pensar que tenemos $6!/(6-4)!$.

En general, si tenemos un conjunto de n elementos distintos, la cantidad de palabras (donde por *palabra* se entiende un vector ordenado) de k elementos se lo llama *variaciones n* contados de a k, se lo escribe como V_k^n y vale

$$V_k^n = \frac{n!}{(n-k)!}.$$

En probabilidades no se usa mucho este concepto, pero a partir del mismo podemos definir entonces el que viene que sí lo usaremos bastante.

2.4. Combinaciones

Tomemos nuevamente la palabra CARBON, ahora lo que queremos es la cantidad de *subconjuntos* de 4 elementos distintos que podemos formar con las letras de CARBON. ¿Cuál es la diferencia con el problema anterior? Fijarse que no nos está pidiendo la cantidad de *palabras* sino *subconjuntos*. La diferencia es que, como habíamos dicho, una palabra tiene un orden determinado, mientras que en un subconjunto no existe tal noción. En efecto, el subconjunto $\{C, A, R, B\}$ y el subconjunto $\{R, A, B, C\}$ son el mismo subconjunto, por más que está "escrito" de distinta forma. Podríamos pensar que como palabras son distintas CARB y RABC, pero como subconjuntos son la misma cosa.

Entonces pues, ¿cuántos subconjuntos de 4 elementos admite el conjunto $\{C, A, R, B, O, N\}$? Empezemos con la cuenta de la sección anterior. La cantidad de palabras de 4 letras era $6!/(6-4)!$. Ahora bien, como dijimos, acá estamos considerando el orden interno y eso lo queremos de alguna manera anular. Es decir, queremos sacar el orden interno de las palabras. Notemos que si tomamos el subconjunto $\{C, A, R, B\}$, este admite un total de $4!$ palabras, que son las posibles permutaciones de las letras del subconjunto.

Así pues, estamos contando $4!$ veces de más. Luego, para "arreglarlo" lo que debemos hacer con $6!/(6-4)!$ es dividirlo por $4!$, quedándonos

$$\frac{6!}{4!(6-4)!}.$$

En general, a la cantidad de subconjuntos de k elementos que tiene un conjunto de n elementos se lo llama **combinatorio** de n elementos tomados de a k, lo notaremos como $\binom{n}{k}$ y vale

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

Algunas aclaraciones. En realidad, el número combinatorio $\binom{n}{k}$ se define como la cantidad de subconjuntos de k elementos que tiene un conjunto de n elementos. La diferencia fundamental con la permutación ordenada es que en el número combinatorio NO IMPORTA EL ORDEN. Por ejemplo, se quiero sacar una mano de truco, acá serían tres cartas de un mazo de 40. Si me importa contar en qué orden salen las cartas, debería hacer $40 \cdot 39 \cdot 38$. En cambio, si no me interesa el orden en que salen las cartas, será $\binom{40}{3} = \frac{40!}{(40-3)!3!} = \frac{40 \cdot 3938}{3 \cdot 2 \cdot 1}$. O sea, es 6 veces menor.

Ejercicio 2.8. *¿De cuántas maneras puedo armar una comisión de 5 alumnos a partir de un curso de 20?*

Respuesta: en el armado de la comisión no me importa "el orden" en el que son elegidas las personas, por lo cual directamente haríamos $\binom{20}{5}$

Ejercicio 2.9. *Como antes, pero supongamos que el aula tiene 12 chicos y 8 chicas y la comisión debería tener 3 chicos y 2 chicas?*

En este caso, debo primero armar una subcomisión, por ejemplo la masculina a través de $\binom{12}{3}$, y multiplicarlo por la cantidad de opciones de la segunda subcomisión, que sería $\binom{8}{2}$. Así, la respuesta sería

$$\binom{12}{3} \cdot \binom{8}{2}.$$

Notar que este número indefectiblemente debe ser menos que el anterior, ya que en el fondo estoy armando una comisión de 5 alumnos (como el caso anterior), pero imponiendo más restricciones respecto a la composición.

Algunas cuestiones.

Qué es más grande, $\binom{20}{5}$ o $\binom{20}{15}$? Rta, son iguales. O bien se puede verificar eso haciendo la cuenta trivialmente, o bien tratemos de razonarlo. Armar una comisión de 5 alumnos de un total de 20 implícitamente me está armando una "no-comisión" de 15 alumnos que no estarían en la comisión, y esta relación es biunívoca. Es decir, puedo definir un subconjunto por los elementos que pertenecen al mismo o bien, que es lo mismo, por los elementos que NO pertenecen al mismo. Se generaliza esto a lo siguiente:

$$\binom{n}{k} = \binom{n}{n-k}.$$

Cuánto vale $\binom{n}{0}$. Pista: NO es 0. En realidad, vale 1. Cuántos subconjuntos de cero elementos tiene un conjunto de n elementos? Uno, el conjunto vacío ($\emptyset$).

Cuánto vale $\binom{n}{n}$? Pensemos, cuántos conjuntos de n elementos tiene un conjunto de n elementos. Necesariamente es 1, pues es el propio conjunto.

Cuánto vale $\binom{n}{1}$? Ejercicio, razonar y ver que vale n.

Cuántos subconjuntos tiene un conjunto de n elementos? Nota: no estamos especificando cuántos elementos tiene el subconjunto. Así pues, en principio podríamos pensar que debería sumar todos los subconjuntos de 0 elementos, más los de 1, más los de 2, etc. En efecto, tendríamos

$$\sum_{k=0}^{n} \binom{n}{k}.$$

Otra forma de pensarlo, cómo puedo armar un subconjunto cualquiera de un conjunto de n elementos? Una forma, la "interrogativa", es colocar a todos los elementos en fila e ir preguntándole a cada uno si desea pertenecer al subconjunto. Hay pues dos opciones para cada elemento, y es multiplicativo el proceso pues habiendo elegido entre las dos opciones para el primero, tengo otras dos opciones para el segundo y así. Luego, queda

$$2 \cdot 2 \cdot \ldots \cdot 2 = 2^n.$$

De donde deducimos esta igualdad:

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

2.5. Ejercitación

1. *a*) ¿Cuántos números de dos dígitos (que se pueden repetir) se pueden formar a partir de los dígitos 1,2 y 3?

   b) ¿Cuántos números de dos dígitos repetibles se pueden formar a partir de 0,1,2 t 3? Considere que un número no puede empezar con 0.

   c) ¿Cuántos números de tres cifras repetibles se pueden formar a partir de 0,1,2 y 3?

d) ¿Cuántos números de 4 cifras repetibles se pueden formar a partir de todos los dígitos (de 0 al 9)?
e) Repita los incisos anteriores pero esta vez que no se repitan los dígitos.
f) ¿Cuántos números pares de cuatro dígitos se pueden formar?
g) ¿Cuántos números capicúa de cinco cifras se pueden formar?

2. Una habitación tiene 6 puertas, ¿de cuántas maneras se puede entrar por una puerta y salir por otra distinta?

3. Sean A,B y C tres ciudades distintas. Supongamos que existen 4 rutas que unen A con B y 5 que unen B con C. ¿Cuántas rutas existen de A hasta C pasando por B? ¿Cuántas rutas de ida y vuelta de A a C pasando por B? Vuelva a responder la pregunta de antes pero suponiendo que si se transitó por una ruta determinada entre dos ciudades durante la ida no se la puede elegir para la vuelta.

4. *a*) ¿Cuántos resultados distintos puedo obtener al arrojar dos dados de seis caras?
   b) ¿Cuántos se pueden obtener en donde la suma sea mayor que 9?

5. *a*) ¿De cuántas formas puede fotografiarse una familia de 6 personas puestas en hilera?
   b) ¿De cuántas maneras se puede lograr suponiendo que padre y madre están siempre juntos?
   c) ¿Qué pasaría si ahora la foto no es en hilera sino en una mesa circular? Note que aquí la diferencia está en que ”rotar” la mesa produciría la misma foto.

6. ¿De cuántas maneras se pueden sentar 8 matrimonios en una fila con la condición de que cada marido esté al lado de su esposa?

7. ¿Cuántas palabras (anagramas o permutaciones) puedo lograr a partir de ESCUDRIÑA? ¿Cuántas comenzando con Ñ? ¿Cuántas que comienzan con una consonante?

8. ¿De cuántas maneras pueden elegirse un grupo de tres docentes para un curso de combinatoria si hay disponibles un total de 40 docentes? ¿De cuántas maneras si uno de ellos debe tener un rol distinguido de titular de la cátedra?

9. ¿Cuántos exámenes con 10 preguntas de geometría, 6 de aritmética y 4 de probabilidades pueden formarse a partir de un archivo que contiene 50 preguntas de cada asignatura?

10. ¿De cuántas maneras puede formarse una comisión integrada por un presidente, dos secretarios y cinco vocales elegidos de un grupo de 20 personas? Suponga ahora que hay 12 mujeres en el grupo y que necesariamente tiene que haber exactamente tres vocales femeninas.

11. Se tiene una biblioteca con 40 libros de los cuales se quiere seleccionar 18. ¿De cuántas maneras se puede hacer si *Divina Comedia* debe estar siempre incluído? ¿De cuántas maneras que excluyan a ese libro?

12. Una persona tiene 10 amigos, de los cuales debe elegir 5 para invitar a una fiesta, pero resulta que 2 de sus 10 amigos están peleados entre sí y nunca irían juntos a la fiesta. ¿De cuántas maneras puede armar grupos de 5 personas cumpliendo esa condición?

13. *a*) Supongamos que tenemos 5 urnas etiquetadas y queremos distribuir 20 bolitas numeradas de 1 a 20 entre las 5 urnas. ¿De cuántas maneras puedo realizar esto?
    b) (*) Supongamos que sacamos la numeración a las bolitas de tal forma que ahora son indistinguibles entre sí. Vuelva a responder la pregunta anterior. *Nota*: este concepto se conoce como *bosones*.
    c) (*) Tengo un capital de 20$ que puedo invertir en 5 opciones de inversión, en valores enteros. ¿De cuántas maneras puedo efectuar la inversión?
    d) (*) Igual al anterior, pero suponga ahora que no es necesario que invierta todo el capital.

3. Espacios Equiprobables: Modelo de Laplace

3.1. Cálculo de probabilidades

El conocido Modelo de Laplace se aplica en casos de equiprobabilidad. Es decir, a modo de repaso, consideraremos un espacio muestral finito:

$$\Omega = \{\omega_1, \ldots, \omega_n\} \tag{1}$$

donde vale que $P(\omega_i) = 1/n$.

Ya vimos entonces que si tenemos un *evento* $A \subset \Omega$, entonces $P(A) = \frac{\#A}{\#\Omega}$, siendo el denominador los "casos totales" o "casos posibles", y el numerador los "casos favorables". En el caso equiprobable pues, tenemos que para calcular probabilidades necesitamos efectuar ese cociente, calculando la cantidad de casos favorables sobre casos totales.

Consideremos el juego del truco. Habíamos calculado en clase la probabilidad de obtener "flor de oros", es decir, las tres cartas en mano que sean del palo "oros". Nos había quedado:

$$P(\text{flor de oros}) = \frac{10 \cdot 9 \cdot 8}{40 \cdot 39 \cdot 38}.$$

Bajo esta modalidad, estamos suponiendo que las cartas, en algún sentido, se reciben con algún orden. Es decir, distinguimos el hecho de que la primer carta sea un as de espada o que sea la tercera. Qué pasa si calculamos la probabilidad de "flor de oros" suponiendo que nos llegan las tres cartas "de golpe"? Es decir, ya no hay un orden establecido interno entre las cartas. Con lo cual, tendremos que usar el combinatorio. La cantidad total posible de manos que tenemos es, entonces, $\binom{40}{3}$, y lo que nos piden nos queda:

$$P(\text{flor de oros}) = \frac{\binom{10}{3}}{\binom{40}{3}}.$$

Parece distinto a lo que nos había dado antes. Será mayor, menor, cómo será??? Bueno, la respuesta es que son iguales (verificarlo!). En efecto, estamos calculando la probabilidad sobre un evento que en sí no distingue cuestiones de orden. Luego, debiera intuitivamente dar lo mismo si recibo las tres cartas de golpe o me las van pasando una por una. En efecto, pasa eso. Y esto es una moraleja importante: si la probabilidad que me piden no involucra ningún tipo de "orden" (por ejemplo, no pregunta sobre la probabilidad de que la segunda carta sea de oros), entonces es indistinto hacer las cuentas CON o SIN orden, es decir, usando o no el número combinatorio, siempre y cuando seamos CONSISTENTES: si contamos CON orden en el denominador, deberemos hacer lo mismo en el numerador, y si lo hacemos SIN orden, luego todo deberá ser SIN orden.

Ejercicio: con el número combinatorio, calcular la probabilidad de obtener flor y de obtener envido.

Ejercicio 3.1. *Se elige una comisión al azar de 5 personas en un curso con 12 varones y 8 mujeres. Cuál es la probabilidad de que hayan exactamente tres varones y dos mujeres en la comisión?*

Rta:

$$\frac{\binom{12}{3} \cdot \binom{8}{2}}{\binom{20}{5}}$$

Ejercicio 3.2. *Bajo la misma situación de antes, cuál es la probabilidad de que la comisión no tenga varones?*

Rta:

$$\frac{\binom{12}{0} \cdot \binom{8}{5}}{\binom{20}{5}} = \frac{\binom{8}{5}}{\binom{20}{5}}.$$

Ejercicio 3.3. *Bajo la misma situación de antes, cuál es la probabilidad de que la comisión tenga al menos 1 varón?*

Rta:

$$P(\text{al menos un varón}) = 1 - P(\text{ningún varón}) = 1 - \frac{\binom{8}{5}}{\binom{20}{5}}.$$

Ejercicio 3.4. *Bajo la misma situación de antes, cuál es la probabilidad de que la comisión tenga al menos 4 varones?*

$$P(\text{al menos 4 varones}) = P(\text{exactamente 4 varones}) + P(\text{exactamente 5 varones}) = \frac{\binom{12}{4}\binom{8}{1}}{\binom{20}{5}} + \frac{\binom{12}{5}}{\binom{20}{5}}.$$

Ejercicio 3.5. *Bajo la misma situación de antes, cuál es la probabilidad de que en la comisión caiga José?*

Rta:
José ya lo colocamos ”de prepo” en la comisión. Nos falta elegir 4 personas, luego nos queda:

$$P(\text{José esté en la comisión}) = \frac{\binom{19}{4}}{\binom{20}{5}}.$$

3.2. Extracciones SIN reposición

Supongamos que tenemos una urna con 15 bolitas rojas y 8 bolitas azules (o sea, un total de 23 bolitas). Se extraen SIN reposición 5 bolitas. Cuál es la probabilidad de obtener 3 bolitas rojas y 2 azules?
Rta:
Razonando como en el caso de las comisiones, podemos pensar que tenemos que elegir un subconjunto (es sin repetición, pues las extracciones son SIN reposición) de 5 elementos de un conjunto de 20. Es decir, el denominador será $\binom{23}{5}$. Pensando el numerador, tenemos de las bolitas rojas que elegir 3, con lo cual será $\binom{15}{3}$, y luego tenemos un remanente de 8 bolitas azulen en donde debemos escoger 2, es decir $\binom{8}{2}$. Así pues, nos queda

$$P(\text{3 rojas y 2 azules}) = \frac{\binom{15}{3} \cdot \binom{8}{2}}{\binom{23}{5}}.$$

Cómo se resolvería el problema si fuesen CON reposición? Lo veremos más adelante esa cuestión.

3.3. Ejemplo

Se arroja una moneda equilibrada 15 veces. ¿Cuál es la probabilidad de obtener exactamente 8 caras?
Pensemos, como siempre por ahora, en casos favorables sobre casos totales. Esto se puede hacer puesto que la moneda está *equilibrada*, conduciendo a un espacio muestrar equiprobable. Caso contrario el problema sería más difícil y lo veremos más adelante.
En el denominador, tengo todas las formas posibles de obtener resultados con 15 arrojadas. Pensando que cada vez que arrojo la moneda tengo dos resultados posibles, no es difícil convencerse que en el denominador tendré 2^{15} opciones.

¿Y en el numerador? Bien, pensemos ahora que tenemos 15 celdas que representan cada una de las arrojadas. Debemos escoger 8 de ellas en donde establecemos que es ”cara”. Esto se puede lograr de $\binom{15}{8}$ formas, obteniendo así

$$\frac{\binom{15}{8}}{2^{15}}.$$

3.4. Ejercitación

1. De un grupo de siete estudiantes, entre los cuales están Santiago y Agustina, se escoge un comité de tres estudiantes al azar. ¿Cuál es la probabilidad de que:
   a) Santiago pertenezca al comité?
   b) Santiago y Agustina pertenezcan al comité?
   c) Santiago o Agustina pertenezcan al comité?
   d) Agustina no pertenezca al comité?
   e) Santiago pertenezca pero Agustina no?
2. Una caja contiene 8 bolitas rojas, 10 blancas y 6 azules. Se extraen 3 bolas sin reposición, hallar la probabilidad de que:
   a) las tres bolitas sean rojas.
   b) dos bolitas sean rojas y una blanca.
   c) no haya bolitas rojas.
   d) al menos una bolita sea blanca.
   e) salgan roja, blanca y azul en ese orden.
   f) Repita los ítems anteriores pero suponiendo que las extracciones son con reposición.
3. En un determinado edificio de 50 individuos hay 30 que apoyarían una moción para construir un Salón de Usos Múltiples (SUM). Se toma una muestra de 10 individuos. Calcula la probabilidad de que:
   a) exactamente 6 de ellos apoyen la iniciativa.
   b) ninguno apoye la iniciativa.
   c) al menos uno apoye la iniciativa.
   d) A lo sumo ocho individuos apoyen la iniciativa.
4. Se arroja un dado equilibrado de 6 caras 10 veces. Calcule la probabilidad de que:
   a) se obtenga exactamente 4 números pares.
   b) se obtenga a lo sumo un número par.
   c) se obtenga exactamente 8 números pares.
   d) se obtenga al menos 8 números pares.
5. Recibe usted una mano de póker, es decir, 5 cartas de una baraja francesa (13 cartas en 4 palos). Calcule la probabilidad de:
   a) obtener un póker de ases. Es decir, tener los cuatro ases.
   b) obtener un póker. Es decir, cuatro cartas con el mismo número.

c) obtener un full. Es decir, tres cartas del mismo número (una pierna) y las otras dos cartas también del mismo número (un par).

d) obtener una pierna. Es decir, tres cartas del mismo número. Nota: un full o un póker NO sería una pierna.

e) obtener un doble par. Es decir, dos pares. Nota: un full no sería un doble par.

f) (*) obtener un par. Nota, ninguno de los juegos anteriores se considera un par.

g) obtener cinco cartas del mismo palo.

h) obtener una escalera. Es decir, 5 cartas consecutivas del mismo palo. Nota: el as puede o bien venir después de la K o bien venir antes del 2.

6. Calcule la probabilidad que al colocar 8 torres al azar en un tablero de ajedrez de 8×8, ninguna torre amenace a otra. Eso es, no hay ninguna fila o columna con más de una torre.

7. Un armario tiene 10 pares de zapatos. Se eligen 8 zapatos al azar, calcule la probabilidad de que no se conforme ningún par completo.

4. Probabilidad Condicional

4.1. Definición, resultados y aplicaciones

Se arroja un dado equilibrado de 6 caras, cuál es la probabilidad de que salga un 4? Claramente, la respuesta será $1/6$. Ahora bien, supongamos que es otra persona la que arroja el dado por nosotros y nos informa que el número obtenido es par, cuál es entonces la probabilidad de haber obtenido un 4?

Tenemos un cambio de situación, disponemos de una información *a priori* que antes no estaba, que el dado salió par. Con esa información, podríamos "atinar" una respuesta, nos quedan tres casos nada más (2, 4 y 6) que en algún sentido "altera" nuestro espacio muestral, está más restringido, a esos valores nada más. Ahora bien, en ese caso podríamos pensar que la probabilidad debiera ser entonces $1/3$, y estaríamos en lo correcto.

En esta sencilla idea radica la idea de **probabilidad condicional**. Es decir, queremos calcular la probabilidad de que ocurra un evento A sabiendo que ocurrió (o que va a ocurrir) el evento B. Es decir, no hay azar impuesto sobre el evento B, ya sabemos que va a pasar o que pasó. Es información que conocemos, que B ocurrió, y nos preguntamos cuál es la probabilidad de que ocurra entonces A.

Formalmente, notaremos esta situación como $P(A|B)$.

La definición formal es la siguiente:

Definition 4.1. *Dado un evento B tal que* $P(B) > 0$*, se define*

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

La idea de esto es que como tenemos que B ya ocurrió, no está en discusión de azar esto. Queremos calcular la probabilidad de A, tomamos todos los elementos de A que también están en B (eso es $P(A \cap B)$) y renormalizamos considerando que ahora todo el espacio muestral pasaría a ser, en algún sentido, el conjunto B (eso es dividir por $P(B)$).

Siempre está bien definida la probabilidad condicional? No, necesitamos efectivamente que $P(B) > 0$. Tiene sentido, por un lado para no dividir por cero y por otro lado para representar una idea lógica de "condicional". No podemos condicionar respecto a un evento imposible, no tendría sentido.

La | se debe leer como "condicionado a" o "dado". Por ejemplo, $P(A|B)$ es la probabilidad de A dado que ocurrió B, o bien condicionado a que ocurrió B.

Vamos a operar con esta definición a través de algunos ejemplos.

Ejercicio 4.2. *Se arroja un dado de seis caras y nos dicen que salió un número par. Cuál es la probabilidad de haber obtenido un 4?*

Rta: nos piden $P(\text{obtener } 4|\text{dado es par})$. Pues bien, esto es $\frac{P(\text{obtener } 4 \cap \text{dado es par})}{P(\text{dado es par})}$. Fijarse que en numerador la restricción de "obtener 4" es más fuerte que "dado es par". Luego, nos queda:

$$\frac{P(\text{obtener } 4)}{P(\text{obtener par})} = \frac{1/6}{3/6} = 1/3$$

que es lo que habíamos intuido antes.

Ejercicio 4.3. *Se arroja un dado de seis caras y nos dicen que salió un número mayor o igual a 4. Cuál es la probabilidad de haber obtenido un número par?*

Rta:

$$P(\text{obtener par}|\text{obtener} \geq 4) = \frac{P(\text{obtener par} \cap \text{obtener } \geq 4)}{P(\text{obtener} \geq 4)} = \frac{P(\{4,6\})}{P(\{4,5,6\})} = \frac{2/6}{3/6} = 2/3.$$

Vamos a resolver ahora un problema completo, usando nociones de probabilidad condicional. Nótese que para este problema, rápidamente se notará que las herramientas que teníamos antes de este capítulo resultarán insuficientes. Vamos a ir introduciendo teoría intercalada en la resolución del problema.

Ejercicio 4.4. *Se tienen dos urnas. La urna A tiene 7 bolitas rojas y 5 azules, la urna B tiene 4 bolitas rojas y 9 azules. Se arroja una moneda NO EQUILIBRADA, con probabilidad $p = 1/3$ de que salga cara. Si sale cara, luego se extrae una bolilla de la urna A, en caso contrario se extrae de la urna B. Calcule la probabilidad de obtener una bolita roja.*

Iremos por etapas, nótese que la naturaleza del problema dificulta un planteo equiprobable como los que veníamos realizando. Vamos a considerar una serie de incisos que nos irán llevando al problema en cuestión.

Inciso a) Sabiendo que la moneda salió Cara, calcular la probabilidad de obtener una bolita roja.

Rta: este problema es sencillo, en este caso directamente se puede calcular la probabilidad condicional a través de su definición conceptual, sin necesidad de apelar a la definición. Sabemos que la moneda es Cara, luego en términos prácticos solamente exite la urna A, es decir, la que tiene 7 bolitas rojas y 5 azules. Ahí la probabilidad será $7/12$.

Es decir,

$$P(Roja|Cara) = 7/12.$$

Inciso b) Calcule la probabilidad de obtener una bolita roja Y cara en la moneda.

Es decir, nos pide

$$P(Roja \cap Cara).$$

Ahora, enunciaremos un resultado auxiliar que nos ayudará mucho en estas situaciones. Es un resultado trivial de demostrar pero muy práctica, que se conoce como la **regla multiplicativa**.

A partir de que

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

es inmediato despejar que

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

En esto sencillamente consiste la regla multiplicativa, permite expresar una **probabilidad conjunta** (una intersección) a partir de una probabilidad condicional.

Apliquémoslo en nuestro caso. Usando la regla multiplicativa, podríamos obtener lo siguiente:

$$P(Roja \cap Cara) = P(Roja|Cara)P(Cara)$$

aunque también, invirtiendo los roles de A y de B, podríamos obtener

$$P(Roja \cap Cara) = P(Cara|Roja)P(Roja).$$

Son las dos correctas? Sí. Son las dos igual de útiles? No. Deténgase acá un segundo para pensar cuál de las dos resultará más fácil calcular. Si optó por la primer opción, está en lo correcto. Una regla "heurística" dice que en general, cuando tenemos eventos ordenados en el tiempo (como es en este caso, la moneda antecede al dado), es conveniente condicionar el evento futuro con respecto al evento pasado. Así pues, tenemos que

$$P(Roja \cap Cara) = P(Roja|Cara)P(Cara) = \frac{7}{12} \cdot \frac{1}{3} = 7/36.$$

Inciso C Calcule la probabilidad de obtener una bolita roja.

Rta: esto es lo que nos pidió el ejercicio originalmente. Pues bien, la dificultad que tenemos es que ahora nos está pidiendo la probabilidad de un evento "futuro" sin saber el "pasado". Qué se hace en ese caso? Debemos en algún sentido "partir en casos" según el pasado. La forma correcta de hacer esto se conoce como **teorema de probabilidad total**.

Mire el gráfico 2.

Tenemos que la probabilidad de B se puede calcular a partir de cortes con los diversos eventos $A_1, \ldots, A_n$.

Tentemos así el siguiente resultado:

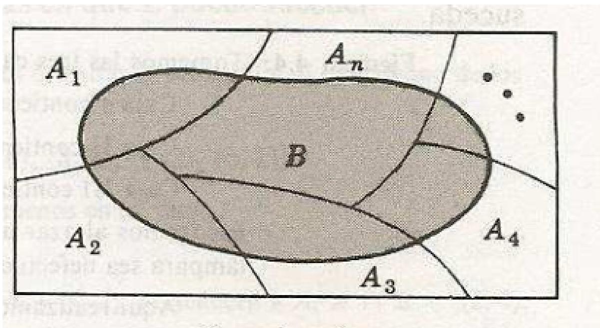


Figura 1: Probabilidad Total

Theorem 4.5. *Sean $A_1, \ldots, A_n$ una* ***partición*** *del espacio muestral Ω. Es decir, son disjuntos de a pares ($A_i \cap A_j = \emptyset$ si $i \neq j$) y tales que $\cup_{i=1}^{n} A_i = \Omega$ (o sea, la unión de los A_i es todo el espacio muestral). Luego,*

$$P(B) = P(A_1 \cap B) + \ldots + P(A_n \cap B).$$

Mire el gráfico fijamente para convencerse que fectviament este teorma tiene sentido. Y en el fondo, es lo que nos permite "partir en casos".

En efecto, nosotros tenemos que calcular la probabilidad de obtener una bolita roja. Es decir, $P(Roja)$. Pues bien, cómo partimos en casos? En función del resultado de la moneda. Sabemos que el espacio muestral se parte en función del resultado de la moneda, que puede ser cara o cruz. No hay otro caso posible, y no hay intersección (la moneda no puede obtener los dos resultados a la vez). Así pues:

$$P(Roja) = P(Roja \ \cap \ Cara) + P(Roja \ \cap \ Cruz).$$

Operando de forma similar a la que hacíamos en el inciso anterior, tendremos que esto a su vez es, por regla multiplicaltiva:

$$P(Roja|Cara)P(Cara) + P(Roja|Cruz)P(Cruz) = 7/12 \cdot 1/3 + 4/13 \cdot 2/3.$$

Habremos así resuelto el problema que originalmente nos habíamos planteado resolver. Agreguemos otro inciso más, que nos permitirá incorporar otro concepto importante: el **teorema de Bayes**.

Inciso d) Sabiendo que se obtuvo una bolita roja, calcule la probabilidad de que en la moneda haya salido cara.

Esto, en términos de probabilidad, resulta ser una probabilidad condicional. Tenemos como información a priori que la bolita es roja (algo en el "futuro") y nos preguntan con respecto al estado de la moneda (algo en el "pasado"). Es decir:

$$P(Cara|Roja).$$

Este es un tipo de probabilidad condicional "antipática", pues está el pasado dado el futuro. Cómo lo damos vuelta? No alcanza con intercambiar los dos conjuntos. Debemos hacer una pequeña operatoria:

$$P(Cara|Roja) = \frac{P(Cara \cap Roja)}{P(Roja)} = \frac{P(Roja|Cara)P(Cara)}{P(Roja)} = \frac{P(Roja|Cara)P(Cara)}{P(Roja \cap Cara) + P(Roja \cap Cruz)}$$

Este sencillo procedimiento es **Bayes**, que es la forma "correcta" de dar vuelta un condicional. Terminen el ejercicio, pues ya tenemos los datos para todos los términos de la expresión.

Ejercicio 4.6. *Una universidad de una determinada localidad tiene un examen de ingreso. La universidad recibe aspirantes provenientes de dos colegios secundarios, la A y la B. Se sabe que el 70 % de los aspirantes proviene de la escuela A, y de ellos aprueba el examen de ingreso el 25 %. El restante 30 % proviene de la escuela B, de los cuales aprueba el examen de ingreso el 50 %.*

1. *Se elige un aspirante al azar, calcular la probabilidad de que haya aprobado el examen de ingreso.*
2. *Se elige un aspirante y se sabe que aprobó el examen de ingreso, cuál es la probabilidad de que haya provenido del colegio B?*

Rta: estamos en una situación en donde nos es conveniente pensar también que es como una situación en dos etapas: primero una persona ”sale” de un colegio (A o B) y luego intenta efectuar el examen de ingreso. Nos piden $P(aprobar)$. Usando probabilidad total, tenemos que

$$P(aprobar) = P(aprobar \cap A) + P(aprobar \cap B) = P(aprobar|A)P(A) + P(aprobar|B)P(B) = 0{,}25 \cdot 0{,}70 + 0{,}50 \cdot 0{,}30.$$

Vamos al inciso b) Nos está pidiendo $P(B|aprobar)$. Suena también a un condicional de los que llamábamos ”antipáticos”, puesto que sería más natural por ejemplo tener el condicional al revés. Apelamos al procedimiento de Bayes.

Entonces,

$$P(B|aprobar) = \frac{P(B \cap aprobar)}{P(aprobar)} = \frac{P(aprobar|B)P(B)}{P(aprobar|A)P(A) + P(aprobar|B)P(B)} = \frac{0{,}50 \cdot 0{,}30}{0{,}25 \cdot 0{,}70 + 0{,}50 \cdot 0{,}30}.$$

Veamos otro problema.

Ejercicio 4.7. *Una urna tiene 3 monedas equilibradas y 2 tales que $P(cara) = 1/5$. Se elige una moneda al azar de la urna y se la arroja, obteniéndose cara. Sabiendo eso, calcule la probabilidad de haber elegido una moneda equilibrada.*

Nos está pidiendo $P(M_{eq}|Cara)$. Como está en algún sentido al revés, convendrá aplicar Bayes.

$$P(M_{eq}|Cara) = \frac{P(M_{eq} \cap Cara)}{P(Cara)} = \frac{P(Cara|M_{eq})P(M_{eq})}{P(Cara)}.$$

Bien, $P(M_{eq}) = 3/5$ pues hay tres monedas equilibradas entre el total de cinco. Luego, $P(Cara|M_{eq}) = 1/2$ ya que sabemos que se usa la moneda equilibrada.

Luego, para obtener $P(Cara)$ aplicamos Probabilidad Total, ya que es algo del futuro en donde no tenemos información del pasado.

Nos queda

$$P(Cara) = P(Cara \cap M_{eq}) + P(Cara \cap M_{noeq}) = P(Cara|M_{eq})P(M_{eq}) + P(Cara|M_{noeq})P(M_{noeq}) = 1/2 \cdot 3/5 + 1/5 \cdot 2/5.$$

Dejamos un ejercicio para practicar el tema:

Ejercicio 4.8. *Se tiene una urna con 5 bolitas rojas y 2 azules. Se arroja un dado de 4 caras equilibrado. Sea k el número obtenido en el dado, se agregan entonces k bolitas azules a la urna. Luego, se extraen 5 bolitas SIN reposición.*

1. *Calcule la probabilidad de obtener exactamente 3 bolitas rojas y 2 azules en la extracción.*
2. *Sabiendo que se obtuvieron 3 bolitas rojas y 2 azules en la extracción, cuál es la probabilidad de que haya salido un 2 en el dado?*

4.2. Independencia

Si tiramos dos dados, es razonable suponer que existe una suerte de independencia entre los resultados de cada dado. Es decir, saber el valor obtenido en un dado no da ningún tipo de información particular sobre el otro dado en cuestión. Si tuviésemos el evento A =”el primer dado salió un 4” y B=”el segundo dado salió un 2”, creo que es claro que en algún sentido, que habrá que definir, los dos eventos son ”independientes”.

Definition 4.9. *Dados dos eventos A y B, diremos que son **independientes** si*

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

O sea, la intersección se factoriza, en algún sentido.

Por qué esto representaría la noción de independencia? Notemos lo siguiente, supongamos que A y B son independientes, y supongamos también que $P(B) > 0$ (qué pasaría si $P(B) = 0$?).

Entonces, notemos que

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

Luego, bajo independencia, $P(A|B) = P(A)$. Esto es muy importante, está diciendo que cuando los dos eventos son independientes, el tener información a priori sobre uno de ellos no altera la probabilidad del otro evento. Eso sí es más natural con la idea de independencia que queremos reflejar.

En un ejercicio dado, la independencia puede ser una hipótesis (implícita o explícita), en cuyo caso podemos usar que $P(A \cap B) = P(A)P(B)$ (o bien, que $P(A|B) = P(A)$), o a veces nos pueden pedir que demostremos que son independientes.

En el contexto de esta materia, para probar que dos eventos son independientes, tendremos que verificar si $P(A \cap B)$ es igual (o no) a $P(A)P(B)$.

Hagamos un ejercicio.

Ejercicio 4.10. *Se arroja una moneda no equilibrada, con $P(cara) = 1/4$. Si sale cara, se extraen 4 bolitas SIN reposición de una urna con 7 rojas y 5 azules, caso contrario se extraen 6 bolitas SIN reposición de la misma urna. Son independientes los eventos ”salió cara” y ”se obtuvieron de color rojo exactamente 3 bolitas”?*

Rta: Debemos verificar si $P(cara \cap R = 3) = P(cara)P(R = 3)$, donde $R = 3$ representa el evento ”se obtuvieron de color rojo exactamente 3 bolitas”.

Sabemos que $P(cara) = 1/4$, eso ya lo tenemos. Calculemos las otras cosas. Para obtener $P(Cara \cap R = 3)$, podemos usar la regla multiplicativa (son dos eventos ordenados en el tiempo, suena ”natural” hacer eso). Luego,

$$P(Cara \cap R = 3) = P(R = 3|Cara)P(Cara) = \frac{\binom{7}{3}\binom{5}{1}}{\binom{12}{4}}\frac{1}{4} = 0{,}088.$$

Ahora, calculemos $P(R = 3)$. Como es un evento en el ”futuro” sin saber el ”pasado”, recurrimos a probabilidad total.

$$P(R = 3) = P(R = 3 \cap Cara) + P(R = 3 \cap Cruz) = P(R = 3|Cara)P(Cara) + P(R = 3|Cruz)P(Cruz) = \frac{\binom{7}{3}\binom{5}{1}}{\binom{12}{4}}\frac{1}{4} + \frac{\binom{7}{3}\binom{5}{3}}{\binom{12}{6}}\frac{3}{4} =$$

$$0{,}088 + 0{,}284 = 0{,}372.$$

Verifiquemos si vale la independencia:

$$P(Cara \cap R = 3) = 0{,}088$$

y
$$P(Cara)P(R = 3) = 0{,}25 \cdot 0{,}372 = 0{,}093$$

Como $0{,}088 \neq 0{,}093$, no son independientes los dos eventos.

Ejercicio 4.11. *Se arroja un dado de 6 caras hasta que salga un 1. Calcular la probabilidad de que se hagan n arrojadas.*

Rta: En este ejercicio, suponemos razonablemente que las arrojadas de los dados son independientes. Es decir, los valores obtenidos entre distintas arrojadas no afectarían, no es que se va ”gastando” el dado o algo así.

Definamos auxiliarmente el evento A_k =”sale un 1 en el evento k”.

Luego, pensemoque tener que arrojar exactamente n veces el dado significa que los n-1 primeras arrojadas NO salió un 1, y en la arrojada n salió un 1. O sea, es lo mismo que $P(A_1^c \cap A_2^c \cap \ldots \cap A_{n-1}^c \cap A_n)$. Ahora, como son independientes los eventos (los dados lo son), entonces nos queda

$$P(A_1^c \cap A_2^c \cap \ldots \cap A_{n-1}^c \cap A_n) = P(A_1^c)P(A_2^c)\ldots P(A_{n-1}^c)P(A_n) = (1-1/6)\cdot(1-1/6)\cdot\ldots\cdot(1-1/6)1/6 = (1-1/6)^{n-1}1/6.$$

Veremos más adelante una generalización de esto, que es lo que se conoce como un **experimento geométrico**.

4.3. Ejercicio Integrador

Ejercicio Ïntegrador”de todo lo visto hasta ahora.

Ejercicio 4.12. *Se tiene una urna con 10 bolitas azules y 8 rojas. Se arroja un dado de 6 caras, si sale de 2 a 6, se extraen esa cantidad de bolitas (lo que salió en el dado) de la urna SIN reposición. Si sale un 1, se extrae CON reposición bolitas de la urna hasta que salga una roja. Definamos los eventos: A = ”se extrajeron exactamente de color azul 3 bolitas”. B= ”salió un 4 en el dado”.*

La pregunta es, ¿son A y B independientes?

4.4. Procedimientos a n pasos

Hasta ahora, notemos que hemos trabajado en situaciones que se modelaban como lo que se conoce como un ”procedimiento a dos pasos”. Es decir, experimentos en donde existe una situación del pasado y otra del futuro. Por ejemplo, en un ejercicio en donde se arroja la moneda y en función de su resultado se extrae de una u otra urna, eso respondería a un procedimiento a dos pasos. Se puede generalizar esto a n pasos. Veamos esto con un ejemplo.

Ejercicio 4.13. *Se arroja una moneda equilibrada, si sale cara luego se arroja un dado equilibrado de seis caras. Si sale cruz, se lanza en vez uno de cuatro caras. Por último, en una urna que ya tiene 7 bolitas azules, se coloca el número obtenido en el dado en bolitas rojas. Se extrae una bolita y se la observa. Calcule la probabilidad de que la bolita extraída sea roja.*

En esta situación, tenemos tres etapas en el experimento: la moneda, el dado y la extracción de una urna. Hay una fuerte dependencia entre todos estos procesos. Antes de resolver el problema, veamos algunas cuentas. Supongamos que queremos la probabilidad de obtener una cara, sacar un dos en el dado y observar una bolita roja. Podemos representar eso como $P(Cara \cap D = 2 \cap R)$.

El ”truco” es el mismo de siempre: usar la regla multiplicativa siempre condicionando el futuro con respecto a todo el pasado. Es decir, primero hacemos

$$P(Cara \cap D = 2 \cap R) = P(R|Cara \cap D = 2)P(Cara \cap D = 2).$$

Volvemos a usar regla multiplicativa con el segundo factor, y nos queda:

$$P(R|Cara \cap D = 2)P(D = 2|Cara)P(Cara)$$

Fijarse que ahora cada uno de estos factores es sencillo y directo de calcular. $P(Cara) = 1/2$, $P(D = 2|Cara)$ será 1/6 pues si sale Cara en la moneda, arrojamos el dado de seis caras. Por último, $P(Roja|Cara \cap D = 2)$ es una situación de obtener una bolita roja en una situación en donde como salió dos en el dado, la urna tendrá 7 bolitas azules y dos rojas, con lo cual quedará 2/7. Juntando todo nos queda $2/7 \cdot 1/6 \cdot 1/2$.

Ahora, vamos a resolver el problema que nos pide, que es $P(Roja)$. Esto es bastante elaborado, vamos a tener que hacer probabilidad total considerando TODOS los posibles pasados. Es decir,

$$\begin{aligned} P(Roja) = P(Cara \cap D = 1 \cap R) + P(Cara \cap D = 2 \cap R) + P(Cara \cap D = 3 \cap R) + \\ P(Cara \cap D = 4 \cap R) + P(Cara \cap D = 5 \cap R) + P(Cara \cap D = 6 \cap R) + P(Cruz \cap D = 1 \cap R) + \\ P(Cruz \cap D = 2 \cap R) + P(Cruz \cap D = 3 \cap R) + P(Cruz \cap D = 4 \cap R). \end{aligned}$$

Por último, cada una de estas probabilidades se obtienen con el procedimiento visto antes. Es laborioso, pero no difícil.

4.5. Ejercitación

1. Se arrojan dos dados de seis caras equilibrados. Calcular la probabilidad de que la suma sea 7 dado que:
   a) la suma es impar.
   b) la suma es mayor que 6.
   c) el número del segundo dado es par.
   d) el número de alguno de los dados es impar.
   e) los números de los dados son iguales.
2. En días lluviosos, José llega tarde al trabajo con probabilidad 0.4. En días no lluviosos, eso ocurre con probabilidad 0.1. Se sabe que la probabilidad de que mañana llueve es de 0.7.
   a) Calcule la probabilidad de que José llegue tarde al trabajo mañana.
   b) Sabiendo que José llego temprano al trabajo, ¿cuál es la probabilidad de que haya llovido?
3. Un sistema de computación *on-line* tiene 4 líneas de entrada para comunicación. Cada línea cubre un porcentaje del tráfico de entrada y cada línea tiene un porcentaje de mensajes que ingresan con error. La tabla a continuación describe estos porcentajes:

Línea	% de mensajes que entra por la línea	% de mensajes sin error
1	40	99.8
2	30	99.9
3	10	99.7
4	20	77.2

   a) ¿Cuál es la probabilidad de que al elegir al azar un mensaje éste haya ingresado sin error?
   b) ¿Cuál es la probabilidad de que, si el mensaje entró con error, haya ingresado por la línea 1?
   c) ¿se tiene un mensaje que ingresó con error, ¿por cuál línea es más probable que haya ingresado?

4. Tres proveedores A,B y C producen respectivamente el 50 %, 30 % y 20 % del número total de artículos de una fábrica. Los porcentajes de desperfectos de producción de estos proveedores son 3 %, 4 % y 5 %. Si se selecciona al azar un artículo, ¿cuál es la probabilidad de resultar defectuoso? Supóngase ahora de que se selecciona un artículo al azar y resulta defectuoso, hallar la probabilidad de que el artículo haya provenido de A.

5. Hay tres monedas en una caja. Una de ellas tiene dos caras, otra es equilibrada y la tercera está desequilibrada, con probabilidad 0.7 de obtener cara. Se elige una moneda al azar y se la arroja, obteniéndose cara. Tomando en cuenta ese dato, hallar la probabilidad de que haya sido la moneda de dos caras.

6. Una caja tiene 10 monedas del 1 al 10, donde la moneda i-ésima tiene probabilidad $i/10$ de que salga cara. Se elige una moneda al azar y sale cara, calcule la probabilidad de que haya sido la quinta moneda.

   Ayuda: puede usar la siguiente fórmula:

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$$

7. Se tiene una urna con 12 bolitas rojas y 6 azules. Se arroja un dado de 4 caras equilibrado y se extraen, sin reposición, el número obtenido en el dado.

   a) Calcule la probabilidad de obtener exactamente 2 bolitas de color rojo, pudiendo o no haber además bolitas de color azul.

   b) Sabiendo que se obtuvieron de color rojo exactamente 2 bolitas, calcule la probabilidad de haber obtenido un 3 en el dado.

   c) Sabiendo que se obruvieron de color rojo exactamente 2 bolitas, ¿qué número del dado es el más probable?

8. Una enfermedad afecta a una de cada 500 personas de cierta población. Se usa un examen radiológico para detectar posibls enfermos, Se sabe que la probabilidad de que el examen aplicado a un enfermo lo muestre como tal es 0.90 y que la probabilidad de que el examen aplicado a una persona sana la muestre como enferma es 0.01 (falso positivo). Calcular la probabilidad de que una persona esté realmente enferma si su examen dio positivo.

9. Suponga que $P(C) > 0$.

   a) Demuestre que $P(\emptyset|C) = 0$.

   b) Demuestre que $P(\Omega|C) = 1$.

   c) Demuestre que $P(A|C) + P(A^c|C) = 1$ y deduzca que $P(A^c|C) = 1 - P(A|C)$.

   d) Más en general, suponga que A y B son dos eventos disjuntos, demuestre que $P(A \cup B|C) = P(A|C) + P(B|C)$.

   En definitiva, tenemos que si definimos una función $\tilde{P}(\cdot) = P(\cdot|C)$ esta resultará una probabilidad.

10. Se tiran dos dados. Sea A el evento ”la suma de los dados es 9” y B el evento ”la suma de los dados es 10”.

    a) Calcular P(A) y P(B).

    b) ¿Son A y B mutuamente excluyentes (disjuntos)? ¿Son independientes?

11. Se arroja un dado equilibrado hasta que sale un as. Dado un n, ¿cuál es la probabilidad de que se deban efectuar n tiradas?

12. Se tiene un circuito eléctrico. Cada vez que se lo mide tiene una probabilidad 0.2 de quemarse y dejar de andar. Se efectúan una serie de mediciones sobre el circuito de forma independiente hasta que se termina quemando, ¿cuál es la probabilidad de que se deban efectuar 17 mediciones?

13. *a*) Supongamos que A y B son mutuamente excluyentes, ¿bajo qué condiciones podrán ser independientes?

    b) Sean A y B eventos independientes. Probar que A y B^c son también independientes.

    c) Probar que si $P(A) = 0$ luego A es independiente de todo evento.

    d) Si A es independiente de todo evento entonces $P(A) = 0$ o bien $P(B) = 1$.

14. Una caldera tiene 5 válvulas de seguridad idénticas y que operan de forma independiente. La idea es que las válvulas se activen cuando hay demasiada presión en la caldera, para evitar una explosión. La probabilidad de que una válvula se active por la presión de la caldera es 0.96. Suponga que se acumula presión en la caldera, calcula la probabilidad de que al menos una válvula se activa.

15. Se reciben 100 telas, en donde exactamente 25 de ellas presentan defectos. Se eligen dos telas al azar, sin reposición, y se los examina. Sea A_i =”la tela i es defectuoso”. ¿Son independientes los eventos A_1 y A_2? ¿Qué pasa si en vez son 10000 telas totales y hay 2500 defectuosas? ¿Qué supone que pasará a medida que se tomen más telas?

16. Se tiene una urna con 4 bolitas rojas. Se arroja un dado de 4 caras equilibrado y se agrega a la urna el número obtenido en bolitas azules. A continuación, se extrae una bolita. Si esta es azul, se arroja una moneda equilibrada. Si la bolita extraída es roja, se arroja una moneda no equilibrada con $P(cara) = 0{,}3$.

    a) Calcule la probabilidad de sacar un 3 en el dado, sacar una bolita roja y obtener una cara.

    b) Calcula la probabilidad de obtener cara en la moneda.

    c) Sabiendo que la moneda resultó cara, calcule la probabilidad de haber sacado un 2 en el dado.

17. Se tiene una urna con 10 bolitas rojas y 8 azules. Se arroja una moneda, si sale cara se extraen dos bolitas con reposición. Si sale cruz, se extraen dos bolitas sin reposición. ¿Son independientes los eventos C=”la moneda es cara” y R_2=”la segunda bolita es roja”?

18. (esquema de Polya) Se tiene una urna que tiene una bolita negra y una blanca. Se extrae una bolita, se mira el color y se la repone, agregando a la urna otra bolita del mismo color. Se procede así indefinidamente.

    a) Calcule la probabilidad de que la tercer bolita sea negra.

    b) Sabiendo que la tercer bolita resultó negra, calcule la probabilidad de que la primer bolita sea blanca.

    c) Sabiendo que la tercer bolita resultó negra, calcule la probabilidad de que la segunda bolita sea blanca.

5. Variable Aleatoria Discreta

5.1. Ejemplos

Supongamos que estamos con el problema de arrojar dos dados. Teníamos que un (buen) espacio muestral podría ser

$$\Omega = \{(i, j),\ 1 \leq i, j \leq 6, i, j \in \mathbb{N}\}.$$

Ahora la pregunta, si estamos interesados en la suma, realmente nos importa si salió (5,3), (4,4) o (3,5)? No nos interesa en el fondo que la suma es 8?

Acá entra en juego la variable aleatoria. La idea de la misma es una suerte ”filtro” cuyo objetivo es, a partir del espacio muestral, obtener y quedarse con la información que realmente nos interesa. Por ejemplo, la suma.

Así pues, una **variable aleatoria** X es, formalmente, una función

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

En el caso del dado, entonces tendremos que $X((i, j)) = i + j$. Podría quizás interesarnos el valor más alto obtenido en el dato, así pues tendríamos una variable aleatoria $Y((i, j)) = \text{máx}(i, j)$.

Otro ejemplo. Supongamos que estamos haciendo el experimento de lanzamiento de un dardo a una diana. Teníamos que el espacio muestras podría pensarse como

$$\Omega = \mathbb{R}^2.$$

Ahora bien, realmente importa en qué punto cae del plano cartesiano? O bien es más relevante la distancia al origen. Podríamos así definir una variable aleatoria X tal que

$$X((x_1, x_2)) = \sqrt{x_1^2 + x_2^2}.$$

Podríamos incluso precisarlo aún más, imaginando la forma de una diana, con puntajes según la distancia al origen, podríamos definir una variable aleatoria D que sería la cantidad de dinero cobrado. Entonces,

$$D((x_1, x_2)) = \begin{cases} 100 & 0 \leq \sqrt{x_1^2 + x_2^2} < 1 \\ 50 & 1 \leq \sqrt{x_1^2 + x_2^2} < 2 \\ 20 & 2 \leq \sqrt{x_1^2 + x_2^2} < 4 \\ 0 & \sqrt{x_1^2 + x_2^2} \geq 4 \end{cases}$$

5.2. Formalismos

Abusaremos de la notación, diremos por ejemplo que si $A \subset \mathbb{R}$, luego

$$P(X \in A) = P(\{\omega \in \Omega, X(\omega) \in A\}).$$

A modo de ejemplo, si tiramos dos dados y consideramos como X la suma. Entonces,

$$P(X = 8) = P(\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}) = 5/36.$$

Definimos la **función de probabilidad puntual** de una variable aleatoria X a una función $p_X : \mathbb{R} \to [0, 1]$

$$p_X(x) = P(X = x).$$

Definimos el **rango** R_X al conjunto de valores ”alcanzables” por la variable aleatoria con una probabilidad positiva. Es decir,

$$R_X = \{x \in \mathbb{R}, p_X(x) > 0\}.$$

Ejemplo, en el caso de arrojar los dos dados, tenemos que

$$R_X = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

y p_X lo representaremos en la siguiente tabla:

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

Diremos que X es una variable aleatoria **discreta** si su imagen (el conjunto de posibles valores) conforma un conjunto ”discreto”, es decir, contable (finito o infinito pero discreto). Por ejemplo, podría ser $\{-1, +1\}$, $\{1, 2, 3, 4, 5, 6\}$ o bien, por ejemplo para tomar un caso infinito, $\{1, 2, 3, 4, \ldots\}$. Podría tomar valores no enteros, no hay limitación con eso, lo importante es que el total de valores que conforman su imagen forma un conjunto discreto.

Una propiedad fundamental de las variables aleatorias discretas es que

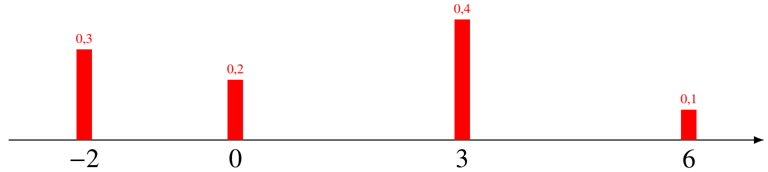
$$\sum_{x \in R_X} p_X(x) = 1.$$

Así pues, vale una suerte de regla del complemento. Si sabemos la probabilidad de todos los valores menos uno, el remanente debe ser tal que todo termine sumando 1.

Un ejemplo, supongamos X con la siguiente función de probabilidad puntual (algo que a veces llamaremos la ”distribución” de X):

x	$p_X(x)$
-2	0.3
0	0.2
3	0.4
6	0.1

Gráficamente, si se quiere, estaríamos teniendo:



Calculemos algunas probabilidades:

$$P(X = 4) = p_X(4) = 0.$$
$$P(X = 3) = p_X(3) = 0{,}4.$$
$$P(X \geq 1) = p_X(3) + p_X(6) = 0{,}4 + 0{,}1 = 0{,}5.$$
$$P(-3 \leq X \leq 1) = p_X(-2) + p_X(0) = 0{,}3 + 0{,}2 = 0{,}5.$$

5.3. Función de Distribución Acumulada

Dada una variable aleatoria X, definimos su **función de fistribución acumulada** a $F_X : \mathbb{R} \to [0, 1]$, definida como:

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

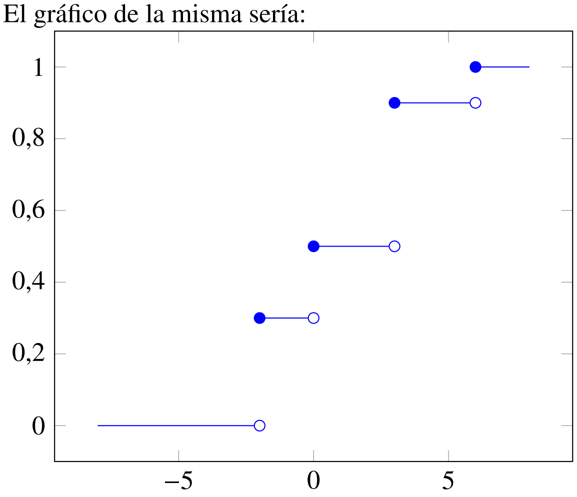
En el caso discreto su utilidad es, digamos, "relativa". Jugará un papel más relevante en el caso continuo. Veamos un ejemplo de cómputo y luego veremos algunas propiedades.

Para el ejemplo anterior, por ejemplo, $F_X(-5) = 0$, $F_X(0) = p_X(-2) + p_X(0) = 0{,}3 + 0{,}2 = 0{,}5$.

La función de distribución acumulada quedará:

$$F_X(x) = \begin{cases} 0 & x < -2 \\ 0{,}3 & -2 \leq x < 0 \\ 0{,}5 & 0 \leq x < 3 \\ 0{,}9 & 3 \leq x < 6 \\ 1 & x \geq 6 \end{cases}$$

El gráfico de la misma sería:



Interesantemente, podemos hacer el proceso inverso. Es decir, a partir de una expresión por $F_X(x)$, podemos obtener $p_X(x)$ considerando como R_X los puntos en donde hay discontinuidad en F_X, y como valor puntual de la probabilidad $p_X(x)$ como la magnitud del salto. Por ejemplo si

$$F_X(x) = \begin{cases} 0 & x < -3 \\ 0{,}1 & -3 \leq x < 1{,}5 \\ 0{,}4 & 1{,}5 \leq x < 5 \\ 0{,}6 & 5 \leq x < 9 \\ 1 & x \geq 9 \end{cases}$$

Así pues, $R_X = \{-3; 1{,}5; 5; 9\}$ y p_X será:

x	$p_X(x)$
-3	0.1 - 0.0 = 0.1
1.5	0.4 - 0.1 = 0.3
5	0.6 - 0.4 = 0.2
9	1.0 - 0.6 = 0.4

Algunas propiedades de F_X:

- Es una función continua a derecha, es decir,

$$\lim_{x \to x_0^+} F_X(x) = F(x_0).$$

- En el ”extremo izquiero”, vale 0.

$$\lim_{x \to -\infty} F_X(x) = 0$$

- En el ”extremo derecho”, vale 1.

$$\lim_{x \to \infty} F_X(x) = 1$$

- Es no decreciente. Es decir, si $x_1 < x_2$, luego $F_X(x_1) \leq F_X(x_2)$.

5.4. Esperanza

Supongamos que tenemos de vuelta la distribución dada por

x	$p_X(x)$
-2	0.3
0	0.2
3	0.4
6	0.1

Buscamos entonces ofrecer una suerte de idea de ”promedio”, de valor central. Una forma ”naive” de hacerlo es promediar los valores del rango, es decir, $\frac{-2+0+3+6}{4}$.

Sin embargo, claramente esta propuesta resulta insuficiente, pues no toma en cuenta el peso (probabilidad puntual) que tiene cada valor del rango. Por eso, es más adecuada (y utilizada) la siguiente definición:

$$E(X) = \sum_{x \in R_X} x \cdot p_X(x).$$

A esto se llama **esperanza** o también, **valor esperado**, o también **media**, e incluso lo podemos llamar **promedio poblacional** (que se debe distinguir del promedio muestral, que veremos más adelante). En nuestro caso, obtendríamos:

$$E(X) = (-2) \cdot 0{,}3 + 0 \cdot 0{,}2 + 3 \cdot 0{,}4 + 6 \cdot 0{,}1 = 1{,}2.$$

O sea, que el valor esperado, o esperanza, es 1.2.

La interpretación física de esto es que si tenemos una barra con pesas ubicadas en los puntos del rango, de magnitud proporcional en kilos a la probabilidad puntual, entonces el punto en donde debo colocar un tope para que la barra se mantenga perfectamente equilibrada (sin caerse por izquierda o derecha), será en la esperanza (en nuestro ejemplo, en 1.2). O sea, es en términos físicos el **centro de masa**.

Ejercicio 5.1. *Se arroja un dado equilibrado de 6 caras, obtenga el valor esperado.*

Rta: Tenemos que

x	$p_X(x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Luego, $E(X) = \sum_{n=1}^{6} n \cdot p_X(n) = \frac{1}{6}\sum_{n=1}^{6} n = \frac{1}{6}\frac{6(6+1)}{2} = 21/6 = 3{,}5$.

Ejercicio 5.2. *Se arrojan dos dados equilibrados, sea X la suma. Obtenga $E(X)$.*

Rta: Ya teníamos de antes p_X, haciendo una cuenta un poco más elaborada pero sencilla al fin obtendremos que $E(X) = 7$.

5.4.1. Propiedades

Valen las siguientes propiedades de la esperanza:

1. (linealidad) Sean X e Y dos variables aleatoria, y a, b dos constantes. Luego, $E(aX+bY) = aE(X)+bE(Y)$.
2. (constantes) Sea a una constante, luego $E(a) = a$.
3. (fórmula del estadístico inconsciente). Sea $h : \mathbb{R} \to \mathbb{R}$, luego $E(h(X)) = \sum_{x \in R_X} h(x)p_X(x)$.

Ejemplo, supongamos que tenemos el ejemplo del dado, siendo X el número obtenido. Deseamos obtener $E(X^2)$. Por la fórmula del estadístico inconsciente, esto es igual a $\sum_{n=1}^{6} n^2 p_X(n) = \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2)$.

Una propiedad que veremos más adelante en la materia se conoce como **Ley de los Grandes Números**(LGN). Supongamos que tenemos un experimento aleatorio cuyo resultado es una variable aleatoria X. Ahora bien, realizamos este experimento una cantidad "grande"de veces, que llamaremos n. Sea $\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$, lo que se conoce como el **promedio muestral** (para diferenciarlo del promedio poblacional). Es decir, el promedio del experimento luego de haberlo observado en una muestra de tamaño n, presumiblemente grande. Lo que valdrá es que si n es "grande" (a precisar...) luego $\overline{X}_n$ será "parecido" (también a precisar después) a $E(X)$. Es decir, con n grande el promedio muestral será parecido al promedio poblacional.

Cuestión notacional, muchas veces a la esperanza de una variable aleatoria X se la suele notar como μ_X, o directamente μ si no hay posibilidad de confusión con otra variable aleatoria.

5.4.2. La Ruleta

La ruleta es un juego de azar, con una composición variable según el tipo de "modelo".

1. La Ruleta Europea: consiste en 37 números. Un 0 de color Verde, 18 números de color Rojo y 18 números de color Negro.
2. La Ruleta Americana: consiste en 38 números. Un 0 y un doble-0, ambos de color Verde, 18 números de color Rojo y 18 números de color Negro.

Supongamos que jugamos a apostar 1$ al color Rojo. Si acertamos (con probabilidad 18/37 en el caso europeo y 18/38 en el americano), nos devuelven el peso apostado y nos darán otro más, con una ganancia neta de 1$. Si no acertamos (con probabilidad 19/37 en el caso europeo y 20/38 en el americano) , directamente perdemos el peso, y la ganancia neta será de -1$. Así pues, si definimos G la ganancia neta de apostar 1$ al color Rojo, obtenemos (suponiendo caso europe):

k	$p_G(k)$
-1	19/37
+1	18/37

Luego, $E(G) = (-1) \cdot 19/37 + 1 \cdot 18/37 = -1/37 = -0{,}027$.

Es decir, tenemos una pérdida neta de 2.7 centavos por cada peso apostado. Eso quiere decir que cada vez que juguemos perderemos 2.7 centavos? ¡NO! En una sola jugada, nunca podremos perder esa cantidad, lo que puede pasar es ganar o perder el peso. Pero lo que dice es que en promedio, si juego muchas, muchas veces, por la ley de los grandes números, mi ganancia promedio, que es $\overline{G}_n$, será parecido a la esperanza, que es -0.027. O sea, el promedio de jugar muchas veces es como si en cada juego perdiera 2.7 centavos. Así pues, si jugara 10000 veces, luego mi ganancia neta de esas 10000 jugadas probablemente sea un valor ”cercano” a $-0{,}027 \cdot 10000 = -270$. O sea, perdería algo así como 270 pesos.

Lo importante, para el casino, es que la esperanza es negativa. O sea, es un juego que tiene un sesgo favorable para el casino, y obviamente desfavorable para el jugador.

Qué pasa en el caso de la ruleta americana? Es cuestión de repetir las cuentas, pero saldrá que incluso la cosa es peor, el valor esperado será ahora 5.4 centavos negativo (el doble de antes, por la presencia de otro número que nos juega en contra, el doble-0).

¿Qué pasa si jugamos 10 veces apostando al rojo en la ruleta europea? ¿Cuál es el valor esperado de dinero que ”ganaré”?

Apelaremos a la linealidad de la esperanza. En efecto, sea X_i la ganancia de la jugada i-ésima, ya sabemos que $E(X_i) = -1/37$ por lo antes calculado. Luego, mi ganancia total será $X_1 + \ldots + X_{10}$, como la esperanza es lineal tendremos que

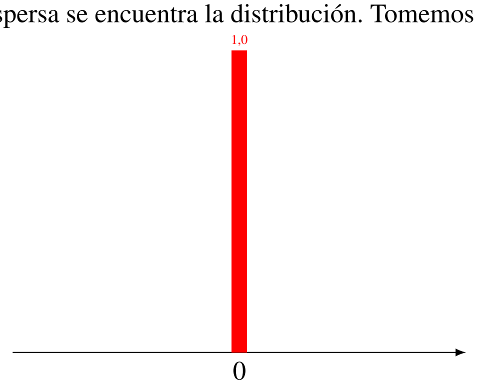
$$E(X_1 + \ldots + X_{10}) = E(X_1) + \ldots + E(X_10) = -1/37 + \ldots + -1/37 = -10/37,$$

o sea que en valor esperado perderé 10/37 pesos.

5.5. Varianza

Tenemos por ahora una medida de centralidad de una distribución, que es la esperanza. No es la única, pero es la más usualmente utilizada. Otras medidas son la mediana y la moda, pero no las veremos por el momento.

Lo que sí va a importar es tener una noción de dispersión, que en cierto sentido lo que mida es qué tan dispersa se encuentra la distribución. Tomemos dos ejemplos:



Ambos tienen la misma esperanza, que es cero. Es decir, ambas distribuciones orbitan alrededor del mismo punto. Sin embargo, es claro de observar los gráficos que la primera está mucho más concentrada (de hecho, completamente concentrada) en su valor medio, a diferencia del segundo.

Necesitamos definir entonces una medida de dispersión, algo que considere la dispersión respecto al valor esperado.

Sea X una variable aleatoria, podríamos considerar la medición de distancias respecto de su valor central tomando una distancia. Una alternativa es tomar $X - E(X)$, y tomarle a su vez el valor esperado a eso. Es decir, una tentativa definición de "dispersión" dada como $E(X - E(X))$. Sin embargo, lo malo que tiene este es que al ser signado, se produce un efecto de cancelación. Es un ejercicio sencillo, usando la linealidad, que esto valdrá siempre 0. No sirve como definición. Debemos corregir el tema de que estar por debajo o por encima debería ser lo mismo. Una alternativa sería tomar el valor absoluto, sin embargo, por razones históricas y por ciertas propiedades, se suele usar en vez la función cuadrado. Es decir, definimos la **varianza** como

$$V(X) = E((X - E(X))^2).$$

Antes de trabajar con esto, haremos una pequeña cuenta para simplificar su expresión.

En efecto, notemos que

$$V(X) = E((X - E(X))^2) = E(X^2 - 2XE(X) + E(X)^2) = E(X^2) - 2E(XE(X)) + E(E(X)^2).$$

No perdamos de vista que $E(X)$ es constante, luego sabemos que $E(cte) = cte$. Por ende, nos queda

$$V(X) = E(X^2) - 2E(X)E(X) + E(X)^2 = E(X^2) - E(X)^2.$$

Esta es la expresión que usaremos más frecuentemente. Hagamos un ejemplo, en el caso de la ruleta europea teníamos que

x	$p_X(x)$
-1	19/37
+1	18/37

Teníamos que $E(X) = -0{,}027$. Luego, $E(X)^2 = 0,000729$.

Además, $E(X^2) = (-1)^2 \frac{19}{37} + (1)^2 \frac{18}{37} = 1$.

Luego pues, $V(X) = E(X^2) - E(X)^2 = 1 - 0{,}000729 = 0,999271$.

Es alto o bajo esto? Bueno, no se puede decir nada en principio, lo importante de la varianza es en términos comparativos, para analizar distintas distribuciones y ver cuál podría ser la más dispersa. Ahí tiene más sentido este cómputo.

Un detalle importante, así como una variable aleatoria podría tener unidades (metros, kilos, pesos, etc), la esperanza tendrá las mismas unidades que la variable aleatoria. Por el cuadrado, la varianza tendrá las unidades al cuadrado. A fin de mejorar la comparabilidad, se suele usar el **desvío estándar**, que es la raíz cuadrada de la varianza.

Es decir, $SD(X) = \sqrt{V(X)}$.

En términos notacionales, a veces notaremos como σ_X^2 a la varianza y como σ_X al desvío estándar de X.

Algunas propiedades:

1. $V(X) \geq 0$, y si $V(X) = 0$ es porque es constante.
2. Si a es una constante, luego $V(aX) = a^2 V(X)$.
3. Si a es una constante, luego $SD(aX) = |a| SD(X)$.
4. Si X e Y son independientes (a definir mejor después...), luego $V(X + Y) = V(X) + V(Y)$.

5.6. Ejercitación

1. Obtener los rangos de las siguientes variables aleatorias discretas:
   a) $X =$"1 si el artículo elegido es defectuoso, 0 si no lo es"
   b) $X =$"cantidad de caras obtenidas al arrojar una moneda tres veces"
   c) Si se extrae simultáneamente dos bolas de una urna que contiene siete bolas numeradas del 0 al 6, siendo los números 0 y 1 blancas y las demás rojas, $X =$"suma de los números de las bolas extraídas" e $Y =$"cantidad de bolas blancas extraídas".
   d) Se arroja una moneda hasta que aparece cara por primera vez, sea $X =$"cantidad de tiros efectuados".
2. Armar la tabla de probabilidad puntual p_X en los siguientes casos:
   a) Se tiene un lote con 4 % de artículos defectuosos, sea $X = 1$ si el artículo es defectuoso y 0 si no.
   b) Se lanza una moneda equilibrada tres veces, con X la cantidad de caras obtenidas.
   c) sea p la probabilidad de que una moneda no necesariamente equilibrada salga cara. Se arroja la moneda hasta que salga cara, sea X cantidad de tiros necesarios.
3. Se eligen tres autos al azar y cada uno de ellos es clasificado N si es naftero o D si tiene motor diesel. Por ejemplo, un resultado posible sería NND.
   a) Definir un espacio muestral Ω apropiado para este experimento.
   b) Considere la variable aleatoria X como la cantidad de autos diesel entre los elegidos. Enumerar los elementos ω de Ω con su correspondiente valor X. Por ejemplo, a NND le correspondería $X = 1$.
   c) Suponga que Ω es equiprobable, calcula la función de probabilidad puntual de X.
4. De un lote que contiene 15 artículos hay 4 que son defectuosos. Se elige una muestra de 5 artículos distintos al azar, sea X la cantidad de artículos defectuosos en la muestra.
   a) Hallar la función de probabilidad puntual p_X.
   b) ¿Cuál es la probabilidad de que al menos 2 artículos sean defectuosos?
   c) Hallar la función de distribución acumulada F_X de X
5. Sea X una variable aleatoria con función de distribución acumulada

$$F_X(x) = \begin{cases} 0 & \text{si } x < 1 \\ 0{,}3 & \text{si } 1 \leq x < 3 \\ 0{,}4 & \text{si } 3 \leq x < 6 \\ 0{,}6 & \text{si } 6 \leq x < 12 \\ 1 & \text{si } x \geq 12 \end{cases}$$

   a) Obtenga la función de probabilidad puntual p_X.
   b) Calcular $P(3 < X \leq 6)$, $P(3 \leq X \leq 6)$, $P(x \leq 4)$, $P(X > 6)$ y $P(X \geq 6 | X \geq 3)$.

6. Sea X la variable aleatoria que mide la cantidad de caras obtenidas al arrojar tres veces una moneda equilibrada.

   a) Hallar el rango R_X.

   b) Hallar y graficar la función de distribución acumulada F_X.

   c) Hallar $E(X)$ y $V(X)$.

7. Se tienen dos urnas con 5 bolillas cada una. En la urna A hay dos bolillas blancas y tres negras, y en la urna B hay una bolilla blanca y cuatro negras. Se tira un dado equilibrado. Si el resultado es múltiplo de 3, se sacan dos bolillas sin reposición de la urna A, en caso contrario las extracciones se hacen de la urna B. Sea X el número de bolillas blancas extraídas.

   a) Hallar p_X.

   b) Calcular $E(X)$ y $V(X)$.

   c) Sabiendo que se extrajo al menos una bolilla blanca, ¿cuál es la probabilidad de que en el dado haya salido un 1?

8. Se compra una bolsa de semillas de una especie de orquídea muy particular. La cantidad de semillas en la bolsa es una variable aleatoria X con la siguiente función de probabilidad puntual:

x	1	2	3
$p_X(x)$	1/6	1/2	1/3

   Se plantan las semillas, la probabilidad de que cada una germine es 1/2, independientemente del resto. Sea Y la cantidad de semillas germinadas.

   a) Hallar la función de probabilidad puntual de Y.

   b) Hallar $E(Y)$ y $V(Y)$.

   c) Sabiendo que germinaron dos plantas, ¿cuál es la probabilidad de que originalmente se hayan obtenido dos semillas en la bolsa?

9. Se dice que X es una **variable de Bernoulli** con probabilidad p si su función de probabilidad puntual es como sigue:

x	0	1
$p_X(x)$	$1-p$	p

   a) Hallar $E(X)$ y $V(X)$.

   b) Hallar $E(X^{11})$.

10. Suponga que $E(X) = 5$ y que $V(X) = 10$. Obtenga:

    - $E(X^2)$.
    - $E(5X + 10)$.
    - $E((X + 1)^2)$.
    - $V(5x + 10)$.

11. Suponga que $E(X) = 6$ y que $E(X(X - 1)) = 29$. Hallar $V(X)$.

12. Un tipo de examen de laboratorio es capaz de detectar al 100 % la presencia de una determinada infección virósica en una muestra de sangre, aunque el mismo resulta bastante costoso. Se sabe que en una determinada población, una proporción p de individuos (la prevalencia) padecen esa infección. Se obtiene una muestra 10 individuos donde se quiere determinar quiénes padecen o no la enfermedad, si es que está presenta en la muestra. Consideramos estas dos estrategias de testeo.

    - Tradicional. Consiste en extraer una muestra de sangre a cada individuo y realizar individualmente a cada muestra el test.
    - Pool de testeos. Se juntan las 10 muestras en una sola y se aplica un test a esa muestra global. Si sale negativo, se sabe que los 10 individuos son sanos y no es necesario realizar más tests. Si sale positivo, luego se efectúan tests sobre cada individuo para poder determinar exactamente quién padece la infección.

    Calcule el valor esperado de testeos necearios para cada estrategia, que quedará en términos de la prevalencia p. ¿Para qué valores de p resulta más efectiva la estrategia de pool por sobre la tradicional, en valor esperado?

13. Una moneda con probabilidad p de obtener cara se arroja hasta que se acumulen o bien dos caras o bien dos cruces. Halle el valor esperado de veces que se debe arrojar la moneda.

14. (*, Lotería de San Petersburgo) En la localidad de San Petersburgo se practica un curioso juego. Se paga un monto M en rublos para poder participar del mismo. Una vez pagado el monto, se arroja una moneda equilibrada. Si sale cara, se recibe un rublo y se termina el juego. Si sale cruz, se vuelve a arrojar la moneda. Si en esta segunda instancia sale cara, se reciben dos rublos y se termina el juego. Caso contrario, se vuelve a arrojar la moneda una tercera vez, donde si sale cara ahora se cobran 4 rublos y si es cruz nuevamente se vuelve a arrojar. Así se sigue arrojando hasta que sale eventualmente cara, duplicando la recompensa por cada vez que se vuelve arrojar.

    a) Sin hacer ninguna cuenta, ¿cuánto pagaría para poder jugar a este juego?
    b) Calcule el valor esperado de dinero recaudado por este juego.
    c) El resultado del inciso anterior, ¿cambia en algo su respuesta en el ítem a)? ¿Pagaría 1.000.000 para jugar? ¿Por qué sí o por qué no?

6. Familias de Distribuciones

Veremos que en muchos casos, hay un patrón en los problemas probabilísticos que tiende a repetirse. En esos casos, puede ser bueno identificar lo que hay en común para ver de no repetir innecesariamente las cuentas. Para tal fin, veremos algunos modelos ”ejemplares” de distribuciones famosas.

6.1. Distribución Binomial

Supongamos los siguientes ejercicios:

Ejercicio 6.1. *Se tiene una urna con 8 bolitas rojas y 12 azules. Se extraen CON reposición 18 bolitas. Cuál es la probabilidad de obtener exactamente 9 bolitas rojas?*

Ejercicio 6.2. *Se tiene un examen ”multiple-choice” consistente en 20 preguntas con 4 respuestas cada una. Un alumno responde completamente al azar, cuál es la probabilidad de acertar exactamente a 10 respuestas?*

Ejercicio 6.3. *Se arroja una moneda no equilibrada 15 veces, con probabilidad de cara p=1/3. Cuál es la probabilidad de obtener exactamente 6 caras?*

Qué tienen en común todos estos problemas? Tratemos de listar las cuestiones que los identifican:

1. Todos tienen un experimento que se repite una y otra vez (sacar bolita, responder pregunta, tirar moneda).
2. Los experimentos son siempre el mismo en cada caso.
3. La cantidad de veces que se repite el experimento está fija de antemano (18 en el de las bolitas, 20 en el del examen, 15 en el de la moneda).
4. Los experimentos son independientes entre sí.
5. Cada experimento tiene dos resultados posibles en el fondo (bolita: roja o no roja, examen: acertar o no acertar, moneda: cara o no cara).
6. Me interesa calcular la probabilidad de veces que sale determinado resultado de cada experimento.

En el fondo, este listado de características son lo que en algún sentido definen lo que es un *experimento binomial*. Es interesante que, salvo por la prosa y los números, son todos en esencia el mismo problema. Al resolver uno, estaré resolviendo todos.

Vamos a apuntar a formalizar un poco esto, y de dar una estructura de solución a estos problemas.

Ahora bien, qué es en el fondo un experimento binomial? Consiste en un experimento que cumple los siguientes principios:

1. Se tiene una sucesión n de ensayos.
2. Los ensayos son independientes entre sí.
3. Cada ensayo tiene dos resultados posibles: Éxito o Fracaso.
4. Para cada ensayo, la probabilidad del Éxito es p, que no cambia con el ensayo.
5. Nos interesa X la cantidad de Éxitos obtenidos en los n ensayos.

Ante esta situación, diremos que X tiene una distrubución binomial de parámetros n y p, y lo notaremos como $X \sim Bi(n, p)$.

Veamos algunas características de X. Por ejemplo,

$$R_X = \{0, 1, 2, \ldots, n\}.$$

Calculemos la probabilidad puntual en un valor $x \in R_X$. O sea, queremos $p_X(x) = P(X = x)$, es decir, la probabilidad de obtener x éxitos en n ensayos.

Definamos E_i al evento "salió un Éxito en el ensayo i". Calculemos primero la probabilidad de que los primeros x ensayos sean un éxito, y el resto fracasos. Es decir,

$$P(E_1 \cap \ldots \cap E_x \cap E_{x+1}^c \cap \ldots \cap E_n^c)$$

Como los ensayos son independientes, luego los eventos lo serán, así pues esto quedará:

$$P(E_1)\ldots P(E_x)P(E_{x+1}^c)\ldots P(E_n^c).$$

Pero $P(E_i) = p$ y $P(E_i^c) = 1 - p$. Luego, nos termina quedando:

$$p \ldots p(1-p)\ldots(1-p) = p^x(1-p)^{n-x}.$$

Ahroa bien, esto fue hecho solamente en el caso en que los primeros x ensayos fuesen un éxito y el resto un fracaso. Los x éxitos podrían venir distribuidos de cualquier manera, podrían estar intercalados con los fracasos, estar al final, etc. De cuántas maneras puedo distribuir x éxitos en n ensayos? Pensemos que tenemos n celdas, de las cuales x de ellas deben ser "pintadas" con una E de Éxito, y el remanente $n - x$ serán con F de Fracaso.

De cuántas maneras puedo elegir x celdas entre un total de n. Sí, la respuesta es que se hace con un combinatorio, concretamente, con $\binom{n}{x}$. Para cada una de las $\binom{n}{x}$ opciones de distribuir los x éxitos, tengo una probabilidad de $p^x(1-p)^{n-x}$. Luego,

$$p_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Esta es la fórmula de la distribución puntual de la binomial.

Veamos algunos ejemplos, tomando los ejercicios que teníamos antes. En el de las bolitas, tenemos que hay $n = 18$ ensayos, en donde la probabilidad de Éxito es sacar una bolita roja, con una probabilidad de $p = 8/20$. O sea que en este caso, $X \sim Bi(n = 18, p = 8/20)$. Nos piden la probabilidad de obtener exactamente 9 bolitas rojas. Es decir, nos piden $p_X(9) = \binom{18}{9}(8/20)^9(1 - 8/20)^{18-9}$.

En el ejercicio del examen multiple-choice tenemos que hay $n = 20$ ensayos, con probabilidad de éxito $p = 1/4$. $X \sim Bi(n = 20, p = 1/4)$. Nos piden 10 exactamente respuestas correctas, o sea que $p_X(10) = \binom{20}{10}(1/4)^1 0(1 - 1/4)^{20-10}$.

Finalmente, en el ejercicio de la moneda, tenemos que hay $n = 15$ ensayos, con probabilidad de éxito $p = 1/3$. $X \sim Bi(n = 15, p = 1/3)$. Nos piden que obtengamos exactamente 6 caras, o sea que $p_X(6) = \binom{15}{6}(1/3)^6(1 - 1/3)^{15-6}$.

6.1.1. El Ensayo de Bernoulli

Definamos la mínima variable aleatoria "concebible". Este caso recae en lo que se conoce como el **ensayo de Bernoulli**. Una variable aleatoria X es un ensayo de Bernoulli si solamente puede tomar dos valores, 0 o 1, con probabilidad $P(X = 1) = p$, y naturalmente $P(X = 0) = 1 - p$. Es decir, es un caso particular de una variable binomial para el caso en que $n = 1$. Lo notamos como $X \sim Be(p)$. Es fácil probar que si X es un ensayo de Bernoulli, luego $E(X) = p$ y $V(X) = p(1 - p)$.

Interesantemente, es fácil ver que si tenemos un experimento binomial con parámetros n y p, entonces este experimento se puede descomponer como una suma de n ensayos de Bernoulli. En efecto, sea $X \sim Bi(n, p)$. Luego, recordemos que por su definición, X mide la cantidad de éxitos al realizar n ensayos con dos resultados posibles, es decir, n ensayos de Bernoulli. Luego,

$$X = X_1 + \ldots + X_n$$

con $X_i \sim Be(p)$.

Por ejemplo, si X mide la cantidad de respuestas correctamente respondidas en el caso del examen multiple-choice, entonces X_i sería 0 si se responde mal la i-ésima preguna y 1 si se responde bien.

6.1.2. Esperanza y Varianza de la Binomial

Sea $X \sim Bi(n, p)$. Recordando que podemos pensar que $X = X_1 + \ldots + X_n$, entonces vale que $E(X) = E(X_1) + \ldots + E(X_n) = p + \ldots + p = np$. Idénticamente, sale que $V(X) = np(1 - p)$.

Por ejemplo, en el caso del examen multiple-choice, el valor esperado de preguntas correctamente respondidas al responder al azar será $X(X) = np = 20 \cdot 1/4 = 5$.

Otra manera de probarlo, sin usar la linealidad de la esperanza. Consideren como algo optativo (pero recomendado) leer esta parte, se usará un truco bastante usual en probabilidades.

$$E(X) = \sum_{x \in R_x} x p_X(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}.$$

La expresión que tenmos dentro de la sumatoria parece difícil de manejar, pero veamos algunas cosas que se puede hacer. Para empezar, desarrollemos el combinatorio. Nos queda:

$$\sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

Esto último es porque en $x = 0$ la expresión queda cero, podemos empezar sumando desde 1. Ahora bien:

$$\sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} = np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x}$$

Efectuamos la sustitución $y = x - 1$, nos queda:

$$np \sum_{y=0}^{n-1} \frac{(n-1)!}{(x-1)!(n-1-y)!} p^y (1-p)^{n-1-y} = np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y}$$

Lo que quedó dentro de la sumatoria es la expresión de la probabilidad puntual de una $Y \sim Bi(n-1, p)$, y como estoy sumando en todo su rango necesariamente deberá sumar 1. Con lo cual, nos queda:

$$np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} = np,$$

concluyendo la demostración.

6.2. Distribución Geométrica

Supongamos el siguiente problema: se arroja una moneda con probabilidad $0 < p < 1$ de que salga cara, hasta que salga la primer cara. Sea X la cantidad de veces que se arroja la moneda. Este caso ya lo habíamos analizado antes. Tenemos que $R_X = \{1, 2, \ldots\}$.

Tenemos que cada vez que se arroja una moneda, tenemos un ensayo de Bernoulli, con 1 si sale cara. Se va repitiendo el ensayo de Bernoulli hasta el primer éxito (obtener un 1). Sea E_i =”el ensayo i-ésimo es un Éxito”. Luego,

$$p_X(x) = P(X = x) = P(E_1^c \cap \ldots \cap E_{x-1}^c \cap E_x) = P(E_1^c) \ldots P(E_{x-1}^c)P(E_x) = (1 - p) \ldots (1 - p)p = (1 - p)^{x-1}p.$$

O sea, las características de una distribución geométrica son:

1. Se ejecuta un mismo ensayo una y otra vez.
2. Los ensayos son independientes entre sí.
3. Cada ensayo tiene dos resultados posibles: Éxito o Fracaso.
4. Para cada ensayo, la probabilidad del Éxito es p, que no cambia con el ensayo.
5. Se detiene el proceso ante el primer Éxito.
6. Nos interesa X, la cantidad de veces que se ejecutó el ensayo.

En este caso, diremos que X tiene una **distribución geométrica** de parámetro p y lo notamos como $X \sim Ge(p)$.

Por ejemplo, en el caso de arrojar un dado hasta que salga un 1, queremos calcular la probabilidad de tener que arrojar exactamente 7 veces el dado. Pues bien, es $p_X(7) = P(X = 7) = (1 - 1/6)^{7-1}1/6$.

6.2.1. Esperanza y Varianza

Queremos calcular $E(X) = \sum_{k=1}^{\infty} kp_X(k)$. Ahora bien,

$$\sum_{k=1}^{\infty} k(1 - p)^{k-1}p = p \sum_{k=1}^{\infty} k(1 - p)^{k-1}$$

Hacemos un ”truco” para resolver esta serie, integrando y derivando. Primero, llamamos $q = 1 - p$, con lo cual nos queda:

$$(1 - q) \sum_{k=1}^{\infty} kq^{k-1}.$$

Ahora sí, integramos y derivamos con respecto a q dentro de la serie, obteniendo:

$$(1 - q) \sum_{k=1}^{\infty} \frac{\partial q^k}{\partial q} = (1 - q)\frac{\partial \sum_{k=1}^{\infty} q^k}{\partial q} = (1 - q)\frac{\partial \left(\sum_{k=0}^{\infty} q^k - 1\right)}{\partial q}.$$

En el último paso completamos la serie geómetrica para que empiece desde cero, para eso compensamos restando el primer término de la seria, que sería 1. Recordemos que una serie geométrica con razón menor en valor absoluto a 1, converge, con lo cual tenemos que:

$$(1 - q)\frac{\partial(1/(1 - q) - 1)}{\partial q} = (1 - q)\frac{1}{(1 - q)^2} = \frac{1}{1 - q} = 1/p.$$

Así pues, $E(X) = 1/p$.

Una cuenta similar, pero algo más engorrosa, permite deducir que $V(X) = \frac{1-p}{p^2}$.

6.2.2. Función de distribución acumulada

Interesantemente, a diferencia de lo que pasará en el resto de las familias de distribuciones que veremos, tendremos una expresión cerrada para F_X cuando X es geométrica. Veamos, queremos calcular $F_X(x) = P(X \leq x)$. El truco que aplicaremos es complementar, pues resultará curiosamente que en este caso será más fácil de tratar. En efecto, al complementar tenemos que

$$F_X(x) = 1 - P(X \geq x+1).$$

Miremos conceptualmente lo que quiere decir $P(X \geq x+1)$. Nos está diciendo que el primer éxito ocurrió a partir del ensayo $x+1$, es decir, que no hubo éxitos antes de ese ensayo. Por lo tanto, todos los primeros x ensayos deberán ser fracasos. Así pues, $P(X \geq x+1) = P(E_1^c \cap \ldots \cap E_x^c) = P(E_1^c)\ldots P(E_x^c) = (1-p)\ldots(1-p) = (1-p)^x$.

Luego, $F_X(x) = 1 - P(X \geq x+1) = 1 - (1-p)^x$.

Tenemos así una fórmula cerrada para F_X en el caso geométrico.

6.3. Distribución Hipergeométrica

Este es un ejemplo que ya hemos visto cuando analizamos modelos equiprobables. Supongamos que tenemos una urna con N bolitas, de las cuales hay K bolitas rojas y $N-K$ de otro color. Extraemos una muestra (es decir, una extracción SIN reposición) de n bolitas. Nos interesa la probabilidad puntual de obtener exactamente k bolitas rojas dentro de la muestra. Tal como habíamos deducido antes, la fórmula será, usando equiprobabilidad:

$$p_X(x) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

siempre y cuando los combinatorios tengan sentido, en caso contrario será 0.

En esta situación, diremos que X tiene una **distribución hipergeométrica**, y se nota como $X \sim H(N, K, n)$.

Algunos resultados que daremos sin demostración:

$$E(X) = n\frac{K}{N}$$

y

$$V(X) = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{n-1}.$$

Cualquier problema que de alguna manera pueda pensarse como extracción de bolitas en una urna sin reposición, por ejemplo, todo lo que tenga que ver con extracciones de muestras de una población finita, se lo puede modelar a través de una hipergeométrica.

Un detalle interesante de la hipergeométrica es que si N y K son grandes (en especial en comparación con n), luego se se puede aproximar a la hipergeométrica por una binomial con $p = K/N$, siempre y cuando este p no quede muy cerca del 0 o del 1. Intuitivamente, la idea es que si N y K son grandes con respecto a n, las extracciones sin reposición no alterarán significativamente la urna y esta podría operar en términos prácticos como si fuese con reposición.

6.4. Ejercicios integradores

Ejercicio 6.4. *Se tiene una urna con 10 bolitas rojas y 6 azules. Se arroja un dado equilibrado de cuatro caras, si sale par se extrae el número obtenido en el dado de bolitas CON reposición. Si sale impar, se extrae el número obtenido en el dado de bolitas SIN reposición. Sea X la cantidad de bolitas rojas obtenidas al final, calcule $E(X)$.*

Rta: en este ejercicio, para calcular $E(X)$ será conveniente armar la tabla de la probabilidad puntual de X. Notemos que en los casos pares parece asomarse un comportamiento característico de una binomial, mientras que en los casos impares sería más bien algo del estilo hipergeométrico. La cuestión es cómo juntar todos estos resultados, para eso usaremos probabilidad total.

En general, tendremos que

$$p_X(x) = P(X = x) = P(X = x \cap D = 1) + P(X = x \cap D = 2) + P(X = x \cap D = 3) + P(X = x \cap D = 4)$$

donde D es el número obtenido en el dado.

Por la regla multiplicativa, obtenemos que

$$p_X(x) = P(X = x|D = 1)P(D = 1)+P(X = x|D = 2)P(D = 2)+P(X = x|D = 3)P(D = 3)+P(X = x|D = 4)P(D = 4).$$

Pues bien, $X|D = 1$ es una hipergeométrica con $N = 16$, $K = 10$ y $n = 1$. De la misma forma, $X|D = 3$ es una hipergeométrica con $N = 16$, $K = 10$ y $n = 3$. Ahora bien, $X|D = 2$ es una binomial con $n = 2$ y $p = 10/16$, y $X|D = 4$ es una binomial con $n = 4$ y $p = 10/16$. En cualquier caso, $P(D = d) = 1/4$.

Juntando todo, tenemos que

$$p_X(x) = \frac{\binom{K}{x}\binom{N-K}{1-x}}{\binom{n}{1}} 1/4 + \binom{2}{x}(10/16)^x(1 - 10/16)^{2-x} 1/4 + \frac{\binom{K}{x}\binom{N-K}{3-x}}{\binom{n}{3}} 1/4 + \binom{4}{x}(10/16)^x(1 - 10/16)^{4-x} 1/4.$$

Es cuestión de evaluar esto en x =0, 1, 2, 3 y 4 para armar la tabla de p_X y luego calcular "a mano" $E(X)$ como $0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2) + 3 \cdot p_X(3) + 4 \cdot p_X(4)$.

Nótese que podríamos haber omitido el cálculo de uno de los $p_X(x)$ ya que sabemos que la suma de todos ellos debe dar 1, con lo cual se lo puede obtener por el complemento.

Ejercicio 6.5. *Se arroja un dado de 4 caras hasta que salga 1. Se paga 10$ por cada vez que se arroja el dado, con un máximo de 50$. Halle el valor esperado de dinero recaudado.*

Rta: una respuesta INCORRECTA es usar la fórmula de la distribución geométrica (puesto que la cantidad de veces que se arroja el dado corresponde efectivamente a una distribución geométrica) y calcular así 10$ $\cdot \frac{1}{1/4}$ = 40$.

Por qué está mal esto? Pues porque no considera que en realidad el dinero recaudado está topado por 50$, así sea que se arroje 5 veces o 5000 veces el dado. La distribución de dinero recaudado no es geométrica, debido a este tope.

Conviene calcularla a mano. Sea D el dinero recaudado, el rango posible es $R_D = \{10, 20, 30, 40, 50\}$ (no llega a infinito). Debemos obtener las probabilidades de cada uno de estos valores. Sea X la cantidad de veces que se arroja el dado.

$$P(D = 10) = P(X = 1) = 1/4.$$

$$P(D = 20) = P(X = 2) = (1 - 1/4)1/4 = 3/16.$$

$$P(D = 30) = P(X = 3) = (1 - 1/4)^2 1/4 = 9/64.$$

$$P(D = 40) = P(X = 4) = (1 - 1/4)^3 1/4 = 27/256.$$

Ahora bien, ojo con cómo se calcula $P(D = 50)$. Es incorrecto decir que esto es $P(X = 5)$, ya que si arrojo por ejemplo 10 veces el dado, también se pagará 50$. Por eso, en realidad $P(D = 50) = P(X \geq 5)$. Esto se puede calcular o bien por complemento, o, como esto es una distribución geométrica, tenemos una buena fórmula para esto. $P(X \geq 5)$, como habíamos razonado antes, quiere decir que tenemos exactamente 4 fracasos en los primeros 4 tiros. Es decir, $P(D = 50) = P(X \geq 5) = (1 - 1/4)^4 = 81/256$.

Así, ya tenemos la tabla completa:

d	$p_D(d)$
10	1/4
20	3/16
30	9/64
40	27/256
50	81/256

Luego, el valor esperado pedido será $E(D) = 10 \cdot 1/4 + 20 \cdot 3/16 + 30 \cdot 9/64 + 40 \cdot 27/256 + 50 \cdot 81/256$.

Ejercicio 6.6. *Una fábrica produce tuercas de forma independiente, en donde hay una probabilidad de 0.1 de que una tuerca presente un defecto (es decir, en promedio el 10 % de las tuercas son defectuosas). Se embalan las tuercas en cajas de 12 tuercas. Un inspector revisa cajas hasta encontrar una caja con 2 o más tuercas defectuosas. Hallar el valor esperado de cajas que habrá de revisar.*

Rta: La cantidad de tuercas que son embaladas en una caja son 12, de las cuales existe una probabilidad de 0.1 para cada tuerca de ser defectuosa, independientemente del resto. Es decir, si defino X =cantidad de tuercas defectuosas en una caja, tendremos que $X \sim Bi(n = 12, p = 0{,}1)$.

Ahora bien, cuál es la distribución sobre la que debemos operar? Tiene que ver con el proceso de revisión de cajas. Sea Y=la cantidad de cajas revisadas hasta encontrar una caja con 2 o más tuercas defectuosas. Qué distribución tiene Y? Si usted respondió "geométrica", estará en lo cierto. Y tendrá una distribución geométrica, con parámetro p desconocido. En realidad, este parámetro p es la probabilidad de que una caja contenga dos o más tornillos defectuosos.

Es decir,

$$p = P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = 1 - \binom{12}{0} 0{,}1^0 0{,}9^1 2 - \binom{12}{1} 0{,}1^1 0{,}9^1 1 = 0{,}341.$$

Luego, $Y \sim Ge(p = 0{,}341)$, su valor esperado será $1/0{,}341 = 2{,}93$. Ese será el valor esperado de cajas que habrá de revisar el inspector hasta encontrar una caja con dos o más tuercas defectuosas.

Como pueden ver, este es un ejemplo de un problema en donde se combina más de una distribución, algo que es muy común de encontrar por ejemplo en ejercicios de parcial.

6.5. Distribución Binomial Negativa

Consiste en una generalización de la distribución Geométrica. Trabajamos exactamente bajo el mismo supuesto, tenemos una sucesión de ensayos de Bernoulli independientes con probabilidad de éxito $0 < p < 1$. La diferencia es que en vez de detenernos ante el primer éxito, lo hacemos cuando se acumulan un total de r éxitos. Sea X la cantidad de ensayos ejecutados, se notará como $X \sim BN(p, r)$. Por ejemplo, arrojo una moneda equilibrada hasta obtener tres caras, esto responde a una distribución binomial negativa con $p = 1/2$ y $r = 3$.

Veamos la expresión para la probabilidad puntual, haremos una deducción informal. Supongamos que ejecutamos x ensayos hasta acumular los r éxitos. Pensemos que tenemos x celdas en donde cada una puede ser un éxito o un fracaso. Necesariamente, la última celda (la celda x) deberá ser un éxito, pues determina la finalización del experimento.

El rango deberá ser cualquier natural a partir de r, pues necesitaré como mínimo r ensayos para acumular r éxitos.

Tendremos necesariamente que tener un éxito en el último ensayo. Luego, tenemos que distribuir $r-1$ éxitos entre $x-1$ celdas, eso lo hacemos con una selección de celdas, dándonos $\binom{x-1}{r-1}$. Por último, tenemos un total de r éxitos, con probabilidad p^r y $x-r$ fracasos, con probabilidad $(1-p)^{x-r}$.

Así pues, la probabilidad puntual queda, juntando todo:

$$p_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}.$$

6.5.1. Esperanza y varianza

Es interesante notar que la binomial negativa es una suma de geométricas. En efecto, si tenemos que acumular r éxitos, podemos pensar que sumamos r geométricas en donde cada una junta un éxito. Así pues, si $X \sim BN(p, r)$ y $X_i \sim Ge(p)$ luego

$$X = X_1 + \ldots + X_r$$

con lo cual

$$E(X) = E(X_1) + \ldots + E(X_r) = 1/p + \ldots + 1/p = r/p.$$

De la misma forma, se deduce que $V(X) = r(1-p)/p^2$.

6.5.2. Algunas cuentas con acumuladas

Supongamos que estamos arrojando un dado equilibrado de 6 caros y nos detenemos al juntar 3 ases. Nos preguntamos cuál es la probabilidad de que tenga que correr 12 ensayos o más. Haremos un truco parecido al usado para la geométrica. Tenemos que si X es la cantidad de ensayos ejecutados, luego $X \sim BN(p = 1/6, r = 3)$

Podríamos hacer el complemento, y calcular $1 - P(X \leq 11)$, pero es un poco laborioso esto, requiere muchas cuentas.

Pensemos que si necesitamos 12 ensayos o más, eso es equivalente a decir que en los primeros 11 ensayos hubo dos éxítos o menos.

Definamos Y como la cantidad de éxitos entre los primeros 11 ensayos. Es claro que Y será binomial, $Y \sim Bi(n = 11, p = 1/6)$. Lo que nos está pidiendo luego es que $P(Y \leq 2) = P(Y = 0) + P(Y = 1) + P(Y = 2)$ utilizando la fórmula de la binomial, que es mucho más realizable.

6.5.3. Ejercicio de Binomial Negativa

Ejercicio 6.7. *Un jugador juega en una ruleta europea (18 rojos, 18 negros y 1 verde). Apostará 5$ al rojo hasta que acumule un total de 4 apuestas ganadoras. Obtenga el valor esperado de ganancia acumulada al detenerse.*

Sea X la cantidad de veces que juega en la ruleta. Está claro que la distribución será una binomial negativa con $p = 18/37$ y $r = 4$. El valor esperado de X es $r/p = 8{,}22$ aproximadamente. Pero no nos interesa directamente esta variable aleatoria sino el dinero. Sea G la ganancia neta, vamos a relacionarla con X.

Cada vez que acierta, se genera una ganancia neta de 5$ y cada vez que no la ganancia será de -5$. Como se detiene al acumular 4 aciertos, entonces el dinero que el casino le pagará a él será $4 \times 5\$ = 20\$$. Por otro lado, tendrá una cantidad de fracasos en donde perderá 5$. ¿Cuántos fracasos habrán?

Aquí aparecerá la X. X es la cantidad de veces que juega, luego los fracasos serán $X - 4$, así pues tenemos que

$$D = 20 + (x-4) \cdot (-5) = -5X + 40.$$

Entonces, $E(D) = E(-5X + 40) = -5E(X) + 40 = -5 \cdot 8{,}22 + 40 = -1{,}11$ aproximadamente.

En valor esperado, perderá 1,11$.

6.6. Distribución de Poisson

Este es un ejemplo en donde bajaremos la definición de la variable ”por decreto” y justificaremos su utilización posteriormente, con algunos criterios heurísticos.

Dado un $\lambda > 0$ real, una variable aleatoria discreta con **distribución de Poisson** es una $X \sim \mathcal{P}(\lambda)$ tal que

$$R_X = \{0, 1, 2, 3, \dots\}$$

y

$$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}.$$

Lo primero que debemos preguntarnos es si esta expresión, que ”bajó por decreto”, corresponde a una distribución de probabilidad. Debemos ver que es positiva (lo que es obvio), y que suma 1, que veremos ahora mismo.

$$\sum_{x=0}^{+\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{+\infty} \frac{\lambda^x}{x!}.$$

Bien, $\sum_{x=0}^{+\infty} \frac{\lambda^x}{x!}$ es el desarrollo en serie de potencias de e^λ, luego nos queda $e^{-\lambda}e^\lambda = 1$. Así pues, es una distribución de probabilidad.

6.6.1. Aproximación de la Binomial por una Poisson

Uno de los primeros puntos que permiten explicar la existencia de esta variable es a través de un límite de la distribución Binomial. En efecto, veremos que si $X \sim Bi(n, p)$, con $n \to \infty$, $p \to 0$ de forma tal que $np \to \lambda$, entonces $p_X(x) \to \frac{e^{-\lambda}\lambda^x}{x!}$.

En efecto, tenemos que

$$p_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = \frac{1}{x!} \frac{n!}{(n-x)!} p^x (1-p)^{n-x}$$

$$= \frac{1}{x!}(n-x+1)(n-x+2) \cdot \ldots \cdot (n-1)np^x(1-p)^{n-x}.$$

Ahora bien, $p = \lambda/n$, haciendo este reemplazo tenemos que:

$$= \frac{1}{x!}(n-x+1)(n-x+2)\cdot\ldots\cdot(n-1)n(\lambda/n)^x(1-\lambda/n)^{n-x} = \frac{1}{x!}(n-x+1)(n-x+2)\cdot\ldots\cdot(n-1)n\frac{1}{n^x}\lambda^x(1-\lambda/n)^n(1-\lambda/n)^{-x}$$

$$= \frac{1}{x!}\lambda^x \frac{n-x+1}{n} \dots \frac{n}{n}(1-\lambda/n)^n(1-\lambda/n)^{-x}.$$

Ahora bien, $(1-\lambda/n)^{-x}$ tiende a 1 cuando $n \to \infty$. Luego, cada cociente $\frac{n-x+i}{n}$ también tiende a 1 cuando $n \to \infty$. Por último, usando el ”truco de e” de las indeterminaciones de tipo 1^1, sale que $(1-\lambda/n)^n$ tiende a $e^{-\lambda}$.

Así pues, tomando límite nos queda

$$\frac{e^{-\lambda}\lambda^x}{x!}$$

que es a lo que queríamos llegar.

Este resultado ofrece consecuencias prácticas, lo que está diciendo es que la distribución de Poisson puede surgir como un límite de una distribución binomial, con n ”grande” y p ”pequeño”, en cuyo caso podemos aproximar la distribución de una $Bi(n, p)$ por una $\mathcal{P}(\lambda = np)$.

Tienen un ejercicio en la práctica en donde pide comparar dos formas de calcular una probabilidad binomial, una de forma exacta usando la fórmula binomial y otra usando la Poisson, a modo de aproximación, para ver que la diferencia no es grande.

6.7. Esperanza y Varianza de la Poisson

Vale que si $X \sim \mathcal{P}(\lambda)$, luego $E(X) = \lambda$ y $V(X) = \lambda$ (es decir, coinciden ambos con el parámetro λ).
Hagamos la cuenta de la esperanza, es similar en cierto sentido a un ”truco” que se usó con la binomial.

$$E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^x}{(x-1)!}.$$

Hacemos la sustitución $y = x - 1$, nos queda

$$\sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^{y+1}}{y!} = \lambda \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!} = \lambda$$

pues $\sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!}$ suma 1 al ser la probabilidad puntual de una Poisson de parámetro λ.

6.8. Ejercicio

Supogamos que tenemos tres máquinas que emiten partículas, las tres emitiendo según una distribución de Poisson, con parámetros respectivos $\lambda_1 = 4$, $\lambda_2 = 3$ y $\lambda_3 = 2$. Se elige una máquina al azar y se miden las partículas emitidas. Se detectan 4 partículas. Calcule la probabilidad de que haya sido elegia la máquina 1.

Rta: este es un caso de probabilidad condicional. Sea $X =$”cantidad de partículas emitidas” y M el número de máquina. Nos están pidiendo

$$P(M = 1|X = 4)$$

Es un condicional de los que están ”al revés”, pues primero se elige la máquina y luego se emiten las partículas. Por lo tanto, lo tratamos de dar vuelta usando Bayes.

Así pues,

$$P(M = 1|X = 4) = \frac{P(M = 1 \cap X = 4)}{P(X = 4)} = \frac{P(X = 4|M = 1)P(M = 1)}{P(X = 4)}.$$

Bien, $P(M = 1) = 1/3$, pues cada máquina se elige completamente al azar. Ahora bien, $P(X = 4|M = 1)$ se calcula usando el hecho de que $X|M = 1$ es una Poisson de parámetro $\lambda_1 = 4$. Luego, nos queda

$$P(X = 4|M = 1) = \frac{e^{-4}4^4}{4!}.$$

Usamos probabilidad total para el denominador:

$$P(X = 4) = P(X = 4 \cap M = 1) + P(X = 4 \cap M = 2) + P(X = 4 \cap M = 3) =$$

$$P(X = 4|M = 1)P(M = 1) + P(X = 4|M = 2)P(M = 2) + P(X = 4|M = 3)P(M = 3) =$$

$$\frac{e^{-4}4^4}{4!}1/3 + \frac{e^{-3}3^4}{4!}1/3 + \frac{e^{-2}2^4}{4!}1/3.$$

Juntando todo, tenemos que

$$P(M = 1|X = 4) = \frac{\frac{e^{-4}4^4}{4!}\frac{1}{3}}{\frac{e^{-4}4^4}{4!}1/3 + \frac{e^{-3}3^4}{4!}1/3 + \frac{e^{-2}2^4}{4!}1/3}$$

Otro problema. Supongamos que queremos calcular $E(X)$. Vamos a proceder con eso. En principio, por definición,

$$E(X) = \sum_{x \in R_X} x p_X(x) = \sum_{x=0}^{\infty} x P(X = x) \sum_{x=0}^{\infty} x(P(X = x \cap M = 1) + P(X = x \cap M = 2) + P(X = x \cap M = 3))$$

$$= \sum_{x=0}^{\infty} x(P(X = x|M = 1)P(M = 1) + P(X = x|M = 2)P(M = 2) + P(X = x|M = 3)P(M = 3))$$

$$= \sum_{x=0}^{\infty} xP(X = x|M = 1)P(M = 1) + \sum_{x=0}^{\infty} xP(X = x|M = 2)P(M = 2) + \sum_{x=0}^{\infty} xP(X = x|M = 3)P(M = 3)$$

$$P(M = 1) \sum_{x=0}^{\infty} xP(X = x|M = 1) + P(M = 2) \sum_{x=0}^{\infty} xP(X = x|M = 2) + P(M = 3) \sum_{x=0}^{\infty} xP(X = x|M = 3).$$

Bien, $X|M = i$ es una Poisson de parámetro λ_i, luego notar que $\sum_{x=0}^{\infty} xP(X = x|M = i)$ es la esperanza de una Poisson de parámetro λ_i, que es a su vez λ_i. Luego, nos queda

$$P(M = 1)\lambda_1 + P(M = 2)\lambda_2 + P(M = 3)\lambda_3 = 1/3 \cdot 4 + 1/3 \cdot 3 + 1/3 \cdot 2 = 3.$$

Así pues, $E(X) = 3$.

6.9. Proceso de Poisson

A modo de ejemplo, supongamos que tenemos una tienda, que siempre se encuentra abierta. Esta tienda recibirá clientes en algunos determinados momentos. Supondremos que la transacción es instantánea, entonces podemos imaginar que lo que tenemos es una recta real, que refleja el tiempo, y vamos pintando ”marcas” en aquellos puntos en donde hay una visita de un cliente.

Supondremos lo siguiente:

- En un momento dado, no pueden llegar más de un cliente.
- Las compras son instantáneas.
- Si tenemos dos intervalos de tiempo disjuntos, entonces la cantidad de clientes que llega en un intervalo es disjunto con respecto a la cantidad que llega en el otro.
- La cantidad esperada de clientes que llegan en un intervalo de tiempo es proporcional a la longitud del intervalo. Es decir, en dos horas espero que lleguen en promedio el doble de clientes que en una hora.

Junto con algunos otros supuestos más, más bien de caracter técnico, se puede demostrar que si definimos la variable aleatoria X_t =”cantidad de clientes que llegan en el intervalo $[0, t]$, entonces esta será una distribución de Poisson.

Más aún, el valor esperado de X_t es proporcional a t, el tiempo que determina la longitud del intervalo.

En general, decimos que un **proceso de Poisson** con tasa $\lambda > 0$ es una colección de $\{X_t\}_{t>0}$ de variables aleatorias que tienen una distribución de Poisson cuyo parámetro es $t\lambda$.

Ejercicio 6.8. *Por ejemplo, supongamos que tenemos un aeropuerto que recibe aviones, en donde la cantidad de aviones que aterrizan sigue un proceso de Poisson de tasa λ= 4 aviones por hora. Nos preguntan cuál es la probabilidad de que en media hora aterricen 3 aviones.*

Rta: tenemos que un una hora llegan 4 aviones en promedio, luego en media hora llegarán en promedio 2 aviones, por ser un proceso de Poisson. Así pues, tenemos que $X_{1/2} \sim \mathcal{P}(2)$. Luego, nos piden que $P(X_{1/2} = 3) = e^{-2}2^3/3!$.

Como se ve, cuando tenemos un proceso de Poisson lo que debemos hacer es tratar de convertirlo a una variable de Poisson, al fijar la variable "tiempo". Ponemos entre comillas este indicativo del "tiempo", ya que el factor dimensional no necesariamente puede ser temporal. Podríamos tener un problema referido a que la cantidad de defectos en una tela que produce un telar sigue un proceso de Poisson de tasa 3 defectos por metro cuadrado. Pues ahí, el factor t serían los metros cuadrados.

Ejercicio 6.9. *Tenemos en una ciudad dada 8 locales de una cadena de comidas rápidas, en donde cada tienda recibe clientes de forma independiente, siguiendo un proceso de Poisson con una tasa $\lambda > 0$ desconocida de clientes por hora. Sabemos, sin embargo, que el valor esperado de locales que no reciben clientes en una horas es 1.2. Obtenga la probabilidad de que un local dado reciba al menos tres clientes en dos horas.*

Rta: Primero, para poder responder la pregunta, vamos a tener que saber el valor de λ. Sin eso, no hay cuenta posible. Usemos el dato que nos están dando, sabemos que en una hora, el valor esperado de locales sin clientes, es 1.2.

Definamos la variable $X =$"cantidad de locales que en una hora no reciben clientes". Qué distribución tiene X? Si respondiste "binomial", estás en lo correcto. Efectivamente, tenemos una cantidad fija de locales ("ensayos"), con dos resultados posibles (recibe o no recibe clientes), independientes y con el mismo comportamiento cada uno. Son todas las pautas indicadoras de una distribución binomial. El n será 8, el tema es el p. El p es la probabilidad de que un local dado no reciba clientes en una hora. Cómo calculamos p? Bueno, ahí entra en juego la Poisson.

p es la probabilidad de que en una hora hora no lleguen clientes en un local dado. Es decir, si Y_1="cantidad de clientes que llegan en una hora en una hora", entonces tenemos que Y_1 es una variable de Poisson de parámetro $1 \cdot \lambda$.

$$P(Y_1 = 0) = \frac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda}.$$

Luego, tenemos que $p = e^{-\lambda}$.

Ahora bien, sabemos que el valor esperado de X es 1.2, pero por otro lado, al ser binomial, será np. Luego $np = 1.2$, es decir, $8p = 1{,}2$, de donde obtenemos que $p = 1{,}2/8 = 0{,}15$.

Bien, $e^{-\lambda} = 0{,}15$, de donde despejamos que $\lambda = 1{,}90$. Ya tenemos el λ, podemos calcular lo que nos piden, que es en que en dos horas en un local determinado lleguen al menos tres clientes.

Eso es, $P(Y_2 >= 3) = 1 - P(Y_2 = 0) - P(Y_2 = 1) - P(Y_2 = 2) = 1 - e^{-1{,}90} - \frac{e^{-1{,}90}1{,}90^1}{1!} - \frac{e^{-1{,}90}1{,}90^2}{2!}$ y terminamos así el ejercicio.

6.10. Esperanza Condicional y Ley de la Esperanza Total

Para motivar lo que queremos explicar consideremos el siguiente problema: la cantidad de llamadas por hora que llega a una línea de reclamos es una variable de Poisson de tasa $\lambda > 0$ conocido. Un reclamo tendrá una atención exitosa con probabilidad $0 < p < 1$ también conocido. Calcule el valor esperado de reclamos exitosamente atendidos que llegan a la línea.

La dificultad radica en que si definimos a X como la cantidad de reclamos exitosamente atendidos, este dependerá bastante en primer lugar de la cantidad efectiva de llamados que lleguen a la línea de reclamos. Si fijamos el evento $A_n =$ "llegan n reclamos", entonces tendremos que $X|_{A_n}$, es decir, la variable aleatoria X condicionado a que sabemos que llegaron n llamados claramente será una binomial de parámetros n y p.

Por ende, podríamos calcular $E(X|A_n) = np$. Ahora bien, se puede demostrar que si tenemos una variable aleatoria X y consideramos $A_1, \ldots, A_n$ (potencialmente infinitos pero numerables) una partición del espacio muestral entonces vale que

$$E(X) = \sum_i E(X|A_i)P(A_i).$$

Esto se conoce como la **ley de la esperanza total**. Hagamos un esbozo de demostración para el caso en que X sea una variable discreta.

Tenemos que

$$E(X) = \sum_{x \in R_X} x p_X(x)$$

por definición.

Ahora bien, usando probabilidad total y regla multiplicativa tendremos que

$$p_X(x) = P(X = x) = \sum_i PX = x|A_i)P(A_i) = \sum_i p_{X|A_i}(x)P(A_i).$$

Luego,

$$E(X) = \sum_{x \in R_X} x \sum_i p_{X|A_i}(x)P(A_i) = \sum_i \sum_{x \in R_X} x p_{X|A_i}(x)P(A_i) = \sum_i P(A_i) \sum_{x \in R_X} x p_{X|A_i}(x) = \sum_i P(A_i)E(X|A_i)$$

lo cual concluiría la demostración. Volvamos ahora a nuestro problema. Tenemos que $E(X|A_n) = np$, donde recordemos que X era la cantidad de reclamos exitosamente atendidos y A_n era el evento que indica la llegada de n reclamos a la línea. Luego, $X|A_n$ es una $Bi(n, p)$ con valor esperado np.

Luego, por la ley de la esperanza total, tendremos que

$$E(X) = \sum_n P(A_n)E(X|A_i) = \sum_{n=0}^{\infty} P(A_n)np.$$

Recordemos que $P(A_n) = e^{-\lambda}\lambda^n/n!$, luego nos queda

$$E(X) = \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} np = p \sum_{n=0}^{\infty} n \frac{e^{-\lambda}\lambda^n}{n!}.$$

El lector "atento" notará que $\sum_{n=0}^{\infty} ne^{-\lambda}\lambda^n/n!$ es justamente la esperanza por definición de una Poisson de parámetro λ, donde ya sabíamos que su valor esperado es λ.

Luego, nos termina quedando que

$$E(X) = p\lambda.$$

6.11. Ejercitación

1. Considere las siguiente v.a.:
    a) Sea X el número de caras al arrojar 10 veces una moneda equilibrado. Hallar la probabilidad de obtener exactamente 6 caras.
    b) Sea X el número de ases al arrojar un dado 3 veces. Hallar la probabilidad de obtener exactamente 2 ases.
2. En un bolso vacío se colocan dos monedas de plata y tres monedas de oro. Se efectúan extracciones con reposición de la bolsa.
    a) ¿Cuál es la probabilidad de sacar por lo menos tres veces una moneda de plata si se realizan 5 extracciones?
    b) ¿Cuál es la probabilidad de obtener por lo menos una moneda de plata si se realizan 10 extracciones? Realizar la menor cantida de cuentas posible.

3. La probabilidad de que un estudiante que ingresa a una Universidad se licencia es 0.4. Hallar la probabilidad de que entre 5 estudiantes elegidos al azar:

   a) Exactamente uno se licencie.

   b) Al menos uno se licencie.

   c) Por lo menos dos se licencien.

   d) Todos se licencien.

4. El 70 % de las consultas de un sistema interactivo de computación requiere de acceso a una base de datos. Un sistema recibe 25 consultas independientes unas de otras.

   a) ¿Cuál es la probabilidad de que exactamente 20 consultas requieran acceso a una base de datos?

   b) ¿Cuál es la probabilidad de que el número se consultas que requieran acceso a una base de datos esté entre 20 y 24 inclusive?

   c) Calcular el valor esperado y la varianza del número de consultas que requieren acceso a una base de datos.

   d) Suponga que si la consulta no requiere acceso a una base de datos tarda 2 segundos en ejecutarse, mientras que si requiere acceso precisará 5 segundos. Calcular el valor esperado de tiempo necesario para procesar las 25 consultas.

5. Un comerciante sabe que hay una probabilidad $p = 0{,}2$ de que en un día cualquiera le sea pedido un televisor de una marca determinada. Supongamos además que en un día le es solicitado a lo sumo un televisor de esa marca, y que la demanda de un día es independiete de la de cualquier otro día.

   a) Hallar la probabilidad de que no le soliciten ningún televisor de esa marca en un período de 15 días.

   b) En el mismo período, ¿cuál es la probabilidad de que la demanda sea de a lo sumo 3 televisores?

   c) Si al comienzo de ese período cuenta con un stock de 6 televisores de esa marca, ¿cuál es la probabilidad de que no pueda satisfacer la demanda?

   d) Cada televisor que se vende se factura por 1000$. Obtenga el valor esperado, la varianza y el desvío estándar de dinero recaudado en el período.

6. Se arroja seis veces una moneda con $P(cara) = 2/3$ y luego cinco veces una moneda equilibrada. Obtenga el valor esperado de caras obtenidas.

7. Una variable aleatoria X tiene distribución binomial con $E(X) = 5$ y $V(X) = 2{,}4$. Hallar $p_X(5)$.

8. Un casino ofrece el siguiente juego. Un jugador decide un número del 1 al 6. Luego, se arrojan tres dados equilibrados y se recibe 1$ por cada acierto que haya en los dados. Ahora bien, si no se acierta ninguno se pierde 1$. ¿Es este juego justo en términos de valor esperado?

9. Se tiene un panel de 7 jurados que deben determinar una decisión binaria, es decir, con dos resultados posibles, referido a la culpabilidad de un acusado que es inocente en realidad. Cada jurado acertará a la decisión correcta (inocente) con probabilidad 0.7. El veredicto final se define por la mayoría. ¿Cuál es la probabilidad de que el jurado decida exculpar al acusado? ¿Cuál es la probabilidad de que exculpar al acusado dado que exactamente 4 de los jurados han decidio de la misma forma?

10. Se tienen dos dados, uno equilibrado y el otro cargado en el cual los números 1 y 2 tienen probabilidad 1/3 y el resto 1/12. Se elige un dado al azar y se lo arroja tres veces de forma independiente. Sea X el número de veces que salió 1 o 2.

a) ¿Cuál es la distribución de X condicional a que se eligió el dado cargado? (esto se puede responder expresando p_x o bien indicando la familia con sus parámetros)

b) Hallar una expresión general para p_X.

c) Sabiendo que salió tres veces un 1, ¿cuál es la probabilidad de que se haya elegido el dado cargado?

11. (bit de paridad) Se quieren enviar tres bits de información por un canal de comunicación imperfecto, con probabilidad $0 < p < 1$ de invertir el valor del bit, de forma independiente para cada bit. Por tal motivo, uno de los mecanismos que se usa para reducir la probabilidad de pérdida de información es la utilización de un mecanismo de detección de errores conocido como el **bit de paridad**.

    El principio del mismo consiste que por cada 3 bits que se manden, en realidad se mande cuatro bits, en donde el último de ellos se coloca de forma tal de garantizar que la cantidad de unos en el paquete sea par. Por ejemplo, si tenemos que enviar el mensaje 010, luego adjuntamos un cuarto bit para mandar 0101. Ahora bien, si el mensaje sufre una alteración solamente en uno de sus bits, luego llegará una cantidad impar de unos y por ende se puede concluir que hubo una perturbación en el mensaje y dictaminar que el paquete es incorrecto. En caso contrario, se decide que el paquete es correcto, se descarta el bit de paridad y se leen los primeros tres bits como el mensaje. Notemos que el bit de paridad tranquilamente puede estar sujeto al error del canal.

    Notemos además que este esquema con bit de paridad no es perfecto, convencerse de que si el mensaje es afectado por una cantidad par de errores, entonces se preservará la paridad y el mensaje no sería catalogado como defectuoso, colándose el error.

    a) Calcule la probabilidad que un paquete de tres bits enviados con el "método tradicional" (sin bit de paridad) llegue con algún error. Deje esto expresado en función de p.

    b) Calcule la probabilidad de que un paquete de tres bits, junto con un cuarto bit de paridad, sea perturbado y que esto no sea detectado por el receptor.

    c) ¿Para qué valores de p es mejor el esquema con bit de paridad sobre el tradicional? Es decir, compare las probabilidades de los incisos anteriores.

12. Una rueda de ruleta está dividida en 38 números de los cuales 18 son rojos, 18 son negros y 2 son verdes. Sea X el número necesario de juegos hasta obtener una sección verde en jugadas independientes.

    a) ¿Cuál es la probabilidad de que sean necesarias al menos 4 jugadas?

    b) Hallar la función de distribución acumulada de X.

    c) Si fueron necesarias 7 o más jugadas, ¿cuál es la probabilidad de q que se necesiten al menos 10 jugadas? Comparar con el a).

    d) ¿Cuál es la probabilidad de que sea necesario un número impar de jugadas?

    e) Hallar $E(X)$ y $V(X)$.

    f) Suponga ahora que se juega hasta juntar dos tiros en donde salió rojo. Hallar la probabilidad de que se necesiten exactamente 9 jugadas.

    g) Ídem anterior pero ahora se busca la probabilidad de que se precisen al menos 9 jugadas.

13. Se tienen dos monedas, una con probabilidad 0.3 de obtener cara y otra equilibrada. Se elige una moneda al azar y se la arroja hasta que salga cara. Sea X la cantidad de veces que se arroja la moneda.

    a) Hallar la función de probabilidad puntual de X.

    b) Sabiendo que se arrojó cuatro veces la moneda, ¿cuál es la probabilidad de que inicialmente se haya escogido la moneda equilibrada?

c) Obtenga el valor esperado de veces que se arrojará la moneda.

14. Tenemos las dos mismas monedas que en el inciso anterior, pero ahora en cada ronda se puede cambiar la moneda. Es decir, se elige la moneda y se la arroja, si es cara se detiene el proceso, sino se vuelve a escoger una moneda y se la arroja. se repite esto hasta obtener eventualmente cara. Rehaga los incisos del ejercicio anterior con este nuevo experimento.

15. Una urna contiene 4 bolitas blancas y 4 negras. Se extraen 4 bolitas sin reposición (es decir, una muestra de 4 bolitas). Si exactamente dos de ellas son blancas, se detiene el proceso. Caso contrario, se reponen las bolitas y se repite el proceso.

    a) ¿Cuál es el valor esperado de la cantidad de veces que se extraen bolitas?

    b) Se paga 1$ por cada vez que se efectúa la extracción, con un tope de 4$. ¿Cuál es el valor esperado de dinero recaudado?

16. Una fábrica produce heladeras a través de dos máquinas, A y B, que generan cada una lotes de 10 heladeras. La máquina A generará siempre 1 heladera defectuosa en cada lote de 10. La máquina B generará 4 heladeras defectuosas en cada lote de 10. El 70 % de la producción global de la empresa se realiza con A y el remanente con B. Recibe usted un lote de 10 heladeras, producidas por una máquina aunque sin saber cuál. Revisa al azar tres heladeras y si al menos dos de ellas son defectuosas rechaza la compra. ¿Cuál es la probabilidad de rechazar un lote?

17. Un minorista ha verificado que la demanda semanal de cajones de cierto producto es una v.a. con distribución de Poisson de parámetro $\lambda = 2$. Completa su stock los lunes por la mañana de manera de tener 4 cajones al principio de la semana. Al efectuar un análisis de la actividad de su negocio, se le plantean las siguientes preguntas:

    a) ¿Cuál es la probabilidad de vender todo su stock durante la semana?

    b) ¿Cuál es la probabilidad de que sea incapaz de cumplir con un pedido por lo menos?

    c) Hallar la distribución del número de cajones que efectivamente vende en una semana. Notar que esto no necesariamente es igual a la cantidad de cajones que le demandan.

    d) ¿Con cuántos cajones debería iniciar la semana a fin de que la probabilidad de cumplir con todos sus pedidos sea mayor o igual a 0.99?

18. Las tareas llegan a una cola de un sistema de computación con un solo serivdor de acuerdo con un proceso de Poisson de tasa 5 tareas por minuto.

    a) ¿Cuál es la probabilidad de que en 30 segundos lleguen menos de 4 tareas?

    b) ¿Cuál es el valor esperado de tareas que llegan en una hora?

19. Un bibliotecario ubica 1000 libros en un cierto día. Si la probabilidad de que un libro cualquiera sea mal ubicado es 0.001 y los libros se ubican de forma independiente, ¿cuál es la distribución exacta del número de libros mal ubicados en ese día? ¿Cuál sería la distribución aproximada?

    a) Calcule la probabilidad exacta y aproximada de que por lo menos un libro sea mal ubicado ese día.

    b) Calcule la probabilidad exacta y aproximada de que tres libros sean mal ubicados ese día.

20. En un concurso de pesca cada pescador paga 100$ por participar. La cantidad de peces obtenida por cada pescador durante el desarrollo del concurso es una variable aleatoria de Poisson de parámetro $\lambda = 4$. Hay un premio de 50$ por pieza recolectada, con un tope de 300$.

a) Calcular la función de probabilidad puntual de la ganancia neta de un pescador.

b) ¿Cuánto dinero espera ganar cada participante?

21. Una máquina produce telas siguiendo un proceso de Poisson de tasa desconocida, pero sabe que la probabilidad de que en 10 metros cuadrados de tela haya al menos un defecto es 0.8. Se arroja un dado de 4 caras y se produce el número obtenido en el dado en rollos independientes de un mentro cuadrado.

    a) Sabiendo que dos de los rollos tenían a lo sumo un defecto, calcule la probabilidad de haber obtenido un 3 en el dado.

    b) Calcule el valor esperado del total de defectos entre todas las telas.

    c) Calcule el valor esperado de rollos con a lo sumo 1 defecto.

22. Se tiene una urna de 10 bolitas rojas y 18 azules. Se arroja un dado equilibrado de 4 caras, en función del resultado se hace una de las siguientea acciones:

    - Si sale 1 se extrae una bolita CON reposición hasta que salga azul.
    - Si sale 2, se extraen 5 bolitas SIN reposición.
    - Si sale 3, se extraen 8 bolitas CON reposición.
    - Si sale 4, se vacía la urna de bolitas azules y se extraen CON reposición 4 bolitas (que naturalmente deberán ser rojas).

    a) Son independientes los eventos ”salió tres en el dado” con ”se obtuvieron exactamente 4 bolitas de color rojo” (pueden haber bolitas azules también, pero de color rojo serán exactamente 4)?

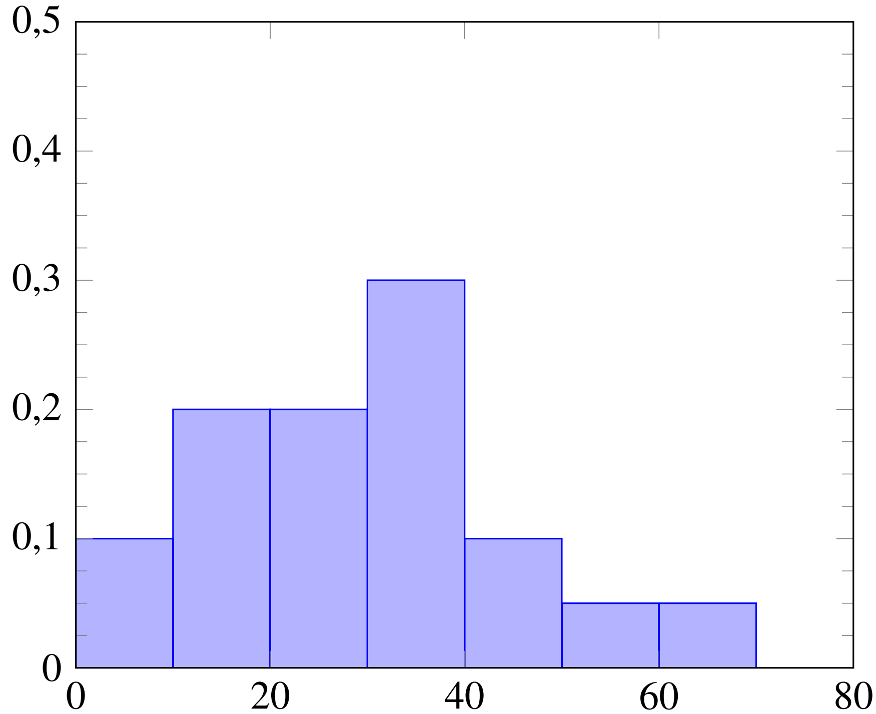
    b) Sea X la cantidad de bolitas rojas obtenidas, obtenga $E(X)$.

7. Variable Aleatoria Continua

7.1. Introducción

Ciertas situaciones ”aleatorias” no se pueden representar con la descripción discreta que estuvimos viendo hasta recién. En efecto, el problema de lanzamiento de una jabalina toma naturalmente valores en un conjunto continuo de la recta real. Si quisiéramos tomar una probabilidad puntual al estilo p_X dejaría de tener sentido puesto que para cada valor $x \in \mathbb{R}$ tendríamos que $p_X(x) = 0$. O sea, $R_X = \emptyset$. Así pues, necesitamos otro tipo de elemento descriptor para este tipo de situaciones.

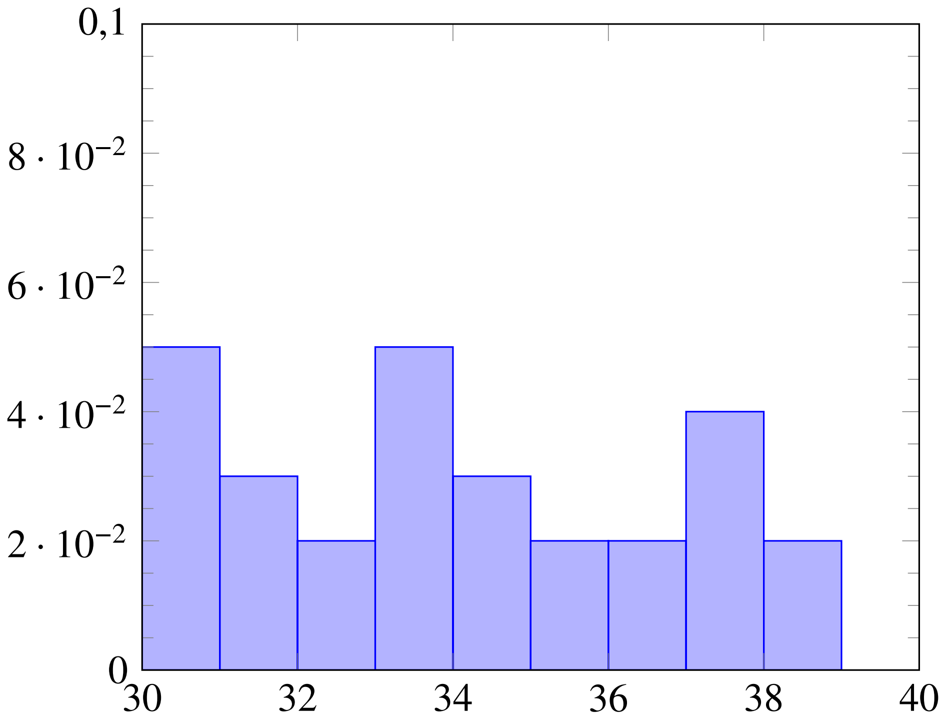
Pensemos en el problema de la jabalina, es cierto que no tiene sentido en principio analizar el comportamiento de cada valor puntual de la recta real, pero sí tiene sentido preguntarnos qué pasa con respecto a intervalos. Por ejemplo, podríamos armar lo que se conoce como un **histograma**, es decir, algo de esta pinta:



Cómo interpretamos este gráfico? Nos está diciendo que $P(0 \leq X < 10) = 0{,}1$, y que por ejemplo $P(30 \leq X < 40) = 0.3$.

Queremos ver qué pasa en el interior del subintervalo $[30, 40)$. Hacemos pues un ”zoom’ en esa región.

Mirando en su interior, obtenemos que



Tenemos que $P(33 \leq X < 34) = 0{,}05$, y que todo suma $P(30 \leq X < 40) = 0{,}3$, teniendo una descripción más fina. Este proceso de refinamiento se puede iterar infinitamente, teniendo así cada vez más una partición más fina de la recta. El cálculo de probabilidades se sigue operando como suma, hasta llegar a un punto en donde se "suman diferenciales", lo que en el sentido matemático se termina conociendo como integrar.

Tenemos pues así que el objeto que debemos integrar para poder calcular probabilidades se llamará **función de densidad de probabilidad**, que recibe su nombre en virtud al mismo concepto físico de densidad de masa. Podemos pensar que la recta real es una barra infinita, con una densidad lineal de masa sobre la misma. Integrando la densidad, obtendríamos masa o, en nuestro caso, probabilidad.

Tenemos así que f_X es la función de densidad de la variable aleatoria continua X. Si A es un intervalo (o incluso un conjunto más general) de la recta real, podemos calcular su probabilidad evaluando

$$P(X \in A) = \int_A f_X(x)dx.$$

Por ejemplo, si quisiéramos calcular en el ejemplo de la jabalina la probabilidad de que la misma caiga entre 25 y 48 metros, debemos hacer

$$P(25 < X < 48) = \int_{25}^{48} f_X(x)dx.$$

Nótese que en el caso continuo somos a propósito "laxos" con las desigualdades estrictas o amplias. Ya que la probabilidad puntual vale 0, pues es integrar en un punto, tendremos que $P(25 < X < 48) = P(25 \leq X < 48) = P(25 < X \leq 48) = P(25 \leq X \leq 48) = \int_{25}^{48} f_X(x)dx$. Es decir, podemos tomar los intervalos abiertos, cerrados, semiabiertos, etc. Es indistinto pues la probabilidad será la misma en todos esos casos, algo que NO pasaba en el caso discreto.

Notemos además que

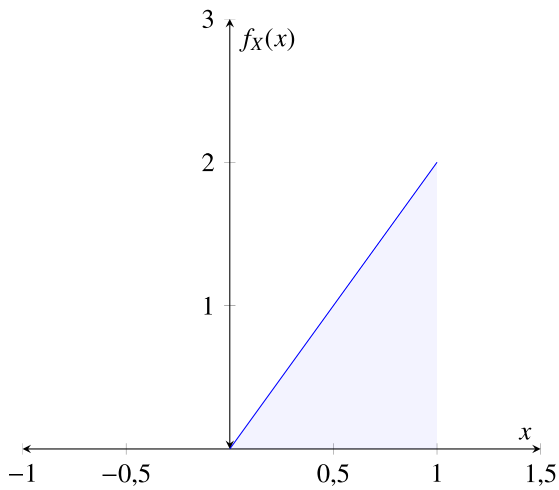
$$1 = P(X \in \mathbb{R}) = \int_{\mathbb{R}} f_X(x)dx = \int_{-\infty}^{+\infty} f_X(x)dx,$$

pues la probabilidad de toda la recta real es 1.

Consideremos este ejemplo de variable aleatoria:

$$f_X(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Gráficamente, tenemos algo de la pinta:



Tenemos así una variable aleatoria X con una densidad cuyo soporte (región del dominio donde $f_X(x) \neq 0$) es $(0, 1)$.

A fin de simplificar la notación, se introduce un concepto en probabilidades que es la llamada **función indicadora**. Permite explicitar el soporte de una función y reduce la escritura. Por definición, la función indicadora en el conjunto A se define como:

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

Luego, podemos escribir nuestra densidad como sigue:

$$f_X(x) = 2x \cdot I_{(0,1)}(x).$$

Ya de esa forma estamos estableciendo directamente el soporte de la densidad f_X.

Primero, veamos que efectivamente es una densidad. Debemos tener que f_X no sea negativa (lo cual es obvio) y además, debemos tener que integre 1. En efecto,

$$\int_{\mathbb{R}} f_X(x)dx = \int_{\mathbb{R}} 2xI_{(0,1)}(x)dx = \int_0^1 2xdx = x^2\Big|_0^1 = 1^2 - 0^2 = 1,$$

de donde sacamos que la densidad está bien definida.

Calculemos algunas probabilidades, por ejemplo, $P(X > 1/2)$.

$$P(X > 1/2) = \int_{1/2}^{\infty} f_X(x)dx = \int_{1/2}^{\infty} 2xI_{(0,1)}(x)dx = \int_{1/2}^{1} 2xdx = x^2\Big|_{1/2}^{1} = 1^2 - (1/2)^2 = 3/4.$$

Otro, $P(|X - 1/2| < 1/4) = P(-1/4 < X - 1/2 < 1/4) = P(1/4 < X < 3/4)$:

$$P(1/4 < X < 3/4) = \int_{1/4}^{3/4} f_X(x)dx = \int_{1/4}^{3/4} 2xdx = x^2\Big|_{1/4}^{3/4} = (3/4)^2 - (1/2)^2 = 1/2.$$

7.2. Esperanza y Varianza

Conceptualmente, tanto la Esperanza y la Varianza representan lo mismo de lo que hacían en el caso discreto. La esperanza sería el punto en donde debo colocar un tope para que una barra con densidad lineal f_X se mantenga equilibrada. Su definición, para el caso continuo, será:

$$E(X) = \int_{\mathbb{R}} x \cdot f_X(x)dx.$$

El caso de la varianza, se define en términos de la esperanza tal como lo hicimos en el caso discreto.

$$V(X) = E((X - E(X))^2) = E(X^2) - E^2(X) = \int_{\mathbb{R}} x^2 \cdot f_X(x)dx - \left(\int_{\mathbb{R}} x \cdot f_X(x)dx\right)^2.$$

Calculemos la $E(X)$ para el ejemplo que venimos desarrollando:

$$E(X) = \int_{\mathbb{R}} x \cdot f_X(x)dx = \int_0^1 x \cdot 2xdx = \int_0^1 2x^2dx = \left.\frac{2x^3}{3}\right|_0^1 = 2/3.$$

El punto de equilibrio de la barra está en $x = 2/3$.

Para la varianza, necesitamos primero $E(X^2)$. Vale, como en el caso discreto, la fórmula del "estadístico inconsciente", que afirma que $E(h(X)) = \int_{\mathbb{R}} h(x)f_X(x)dx$.

$$E(X^2) = \int_{\mathbb{R}} x^2 \cdot f_X(x)dx = \int_0^1 x^2 \cdot 2xdx = \int_0^1 2x^3dx = \left.\frac{2x^4}{4}\right|_0^1 = 1/2.$$

Luego, $V(X) = E(X^2) - E^2(X) = 1/2 - (2/3)^2 = 1/18$.

7.3. Función de Distribución Acumulada

Tal como en el caso discreto, definimos la **función de distribución acumulada** como $F_X(x) = P(X \leq x)$. En principio, esta definición coincide con la que dimos para el caso discreto. Su forma de cálculo es como sigue:

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

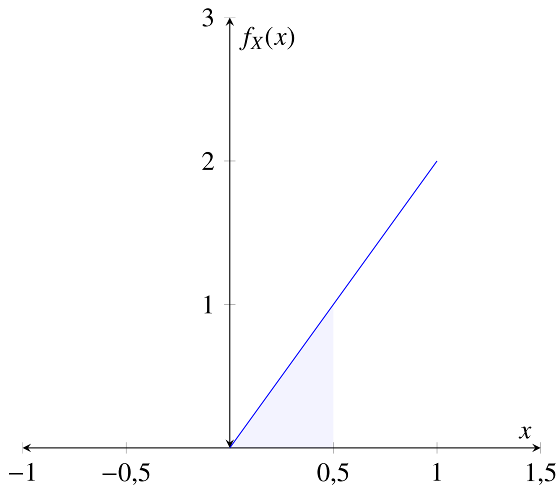
Nótese que hemos reemplazado la variable de integración, de x a t, para que no se confunda con el límite superior de integración.

Vamos a calcularla para nuestro ejemplo. Por la forma en que está definida la densidad, tenemos naturalmente tres regiones, o casos.

En el caso en que $x < 0$, luego es claro que la $P(X \leq x) = 0$, es decir, el área acumulada a izquierda de x será 0.

En el caso en que $x \geq 1$, luego tendremos que $P(X \leq x)$ cubrirá a todo el soporte, por lo tanto esta probabilidad será 1.

Veamos el caso intermedio, $x < 0 < 1$. Gráficamente, tenemos que



Entonces, para este caso tendremos que

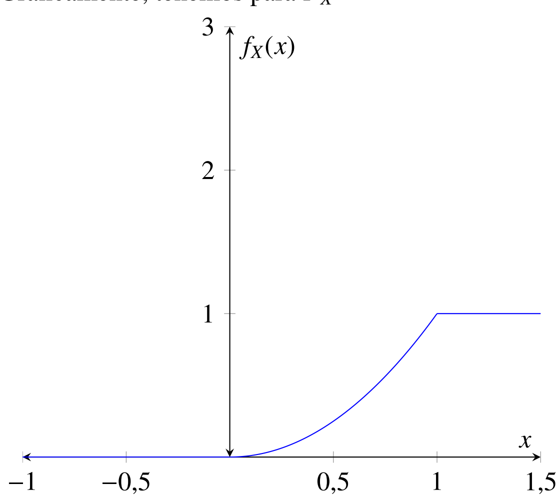
$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{0}^{x} 2tdt = t^2\Big|_0^x = x^2.$$

Tendremos, en resumen, lo siguiente:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Un punto importante, en el caso de X variable aleatoria continua, se tiene que necesariamente F_X deberá ser una función continua. Es una buena forma de verificar por si hay problemas.

Gráficamente, tenemos para F_X



Otro punto interesante de F_X para el caso continuo, permite calcular de forma relativamente sencilla probabilidades, puesto que el trabajo ”duro” de integrar de alguna manera ya está hecho por el cómputo de F_X.

Entonces, si quisiéramos calcular por ejemplo $P(X \leq 0{,}5)$ esto será $F_X(0{,}5) = 0{,}5^2 = 0{,}25$.

Otra cuenta, $P(0{,}1 < C < 0{,}6) = F_X(0{,}6) - F_X(0{,}1) = 0{,}6^2 - 0{,}1^2 = 0{,}36 - 0{,}001 = 0{,}35$. Otro ejemplo, $P(X > 0{,}2) = 1 - F_X(0{,}2) = 1 - 0{,}2^2 = 1 - 0{,}04 = 0{,}96$.

Finalmente, $P(-0{,}5 < X < 0{,}7) = 0{,}7^2 - 0 = 0{,}49$.
Así pues, teniendo F_X, podemos obtener cualquier probabilidad con evaluar F_X, sin tener que integrar.

7.4. Percentiles

Otro concepto útil que sirve para ordenar los datos es lo que se conoce como el **percentil**. Dado un $0 < p < 1$, definimos el percentil p a x_p tal que $P(X \leq x_p) = F_X(x_p) = p$. O sea, es un valor de la recta real tal que deja a su izquierda un probabilidad p.

Así pues, $x_{0,5}$ es el percentil 0,5, que se llama la **mediana** y se nota como $\tilde{x}$, que representa un valor de centralidad de la distribución, que "compite" con la media (esperanza). Hay casos en donde convendrá usar uno en lugar de otro, dejaremos esta discusión para después.

Calculemos por ejemplo la mediana de X. Necesitamos hallar x tal que $F_X(x) = 0{,}5$. Esto representa un despeje de una función partida, la F_X. Pues bien, vamos rama por rama.

Caso 1: para $x < 0$, existe x tal que $F_X(x) = 0{,}5$? En este caso, $F_X(x) = 0$, y $0 \neq 0{,}5$ para cualquier valor de x, en particular para valores de $x < 0$. Luego, la mediana no se encuentra en este caso.

Caso 2: $0 \leq x < 1$, existe x tal que $F_X(x) = 0{,}5$? En este caso, $F_X(x) = 0{,}5$, es decir, $x^2 = 0{,}5$, despejando así que $|x| = 1/\sqrt{2}$, ofreciéndonos dos soluciones ($-1/\sqrt{2}$ y $1/\sqrt{2}$). Nos quedamos con $x = 1/\sqrt{2}$ puesto que es el que cumple $0 <\leq x < 1$. Así pues, la mediana será $\tilde{x} = 1/\sqrt{2}$.

Es decir, a la izquierda de $1/\sqrt{2}$ acumulamos un área de 0.5. O sea, la mitad de la población se encuentra a la izquierda de $1/\sqrt{2}$.

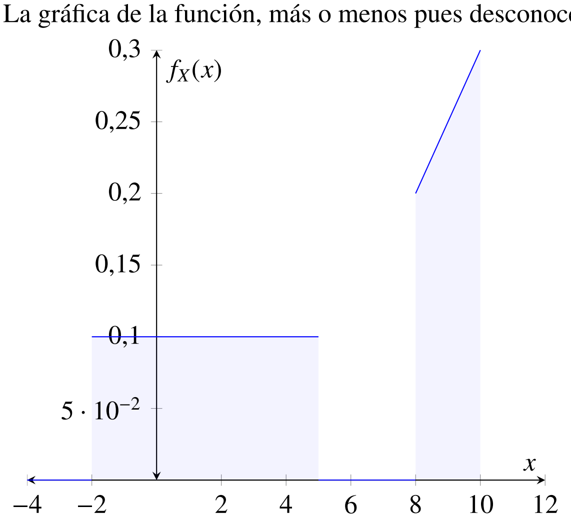
Otros que percentiles "famosos". El **primer cuartil** es el percentil 0,25, el **tercer cuartil** es el percentil 0,75. El segundo cuartil es naturalmente la mediana, el percentil 0,5.

7.5. Ejemplo más elaborado

Consideremos otro ejemplo, X una variable aleatoria continua con la siguiente densidad:

$$f_X(x) = \begin{cases} 0 & x \leq -2 \\ 1/10 & -2 < x < 5 \\ 0 & 5 \leq x < 8 \\ cx & 8 \leq x < 10 \\ 0 & x \geq 10 \end{cases}$$

con $c > 0$ una constante desconocida.
La gráfica de la función, más o menos pues desconocemos c, será:



Primer desafío: obtener el valor de c. Esto se puede despejar usando la condición

$$1 = \int_{-\infty}^{+\infty} f_X(x)dx = \int_{-2}^{5} 1/10dx + \int_{8}^{10} cx = 1/10x\Big|_{-2}^{5} + \frac{cx^2}{2}\Big|_{8}^{10} = 7/10 + c\frac{10^2 - 8^2}{2} = 7/10 + c18.$$

De aquí, despejamos $c = 3/(10 \cdot 18) = 1/60$.

Vamos a calcular el valor esperado de esta distribución.

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)dx = \int_{-2}^{5} x1/10dx + \int_{8}^{10} x(1/60)xdx = \frac{x^2}{20}\Big|_{-2}^{5} + \frac{(1/60)x^3}{3}\Big|_{8}^{10} = \frac{5^2 - (-2)^2}{20} + \frac{(1/60)(10^3 - 8^3)}{3} = 3{,}7611.$$

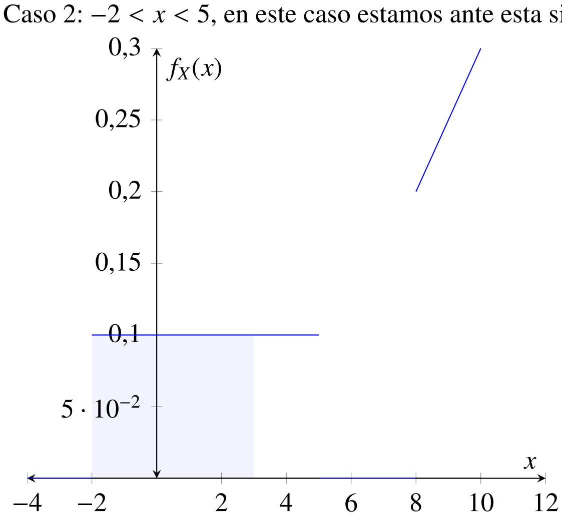
Notar que $E(X)$ tiene que caer siempre entre el mínimo valor (-2) y el máximo valor (10) que puede tomar la variable aleatoria, algo que se está cumpliendo en este caso.

Dejamos como ejercicio el cálculo de $V(X)$.

Calculemos la función de distribución acumulada. Como f_X está partido en 5 casos, similarmente ocurrirá con F_X, lo iremos haciendo por partes, graficando cada situación.

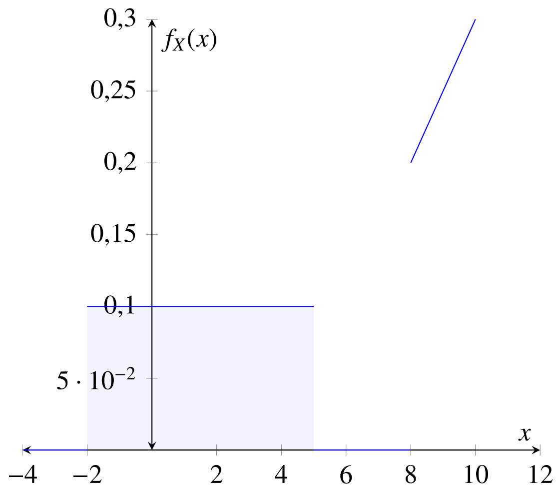
Caso 1: si $x < -2$, luego $F_X(x) = 0$, y no hay mucho más que hacer.

Caso 2: $-2 < x < 5$, en este caso estamos ante esta situación:

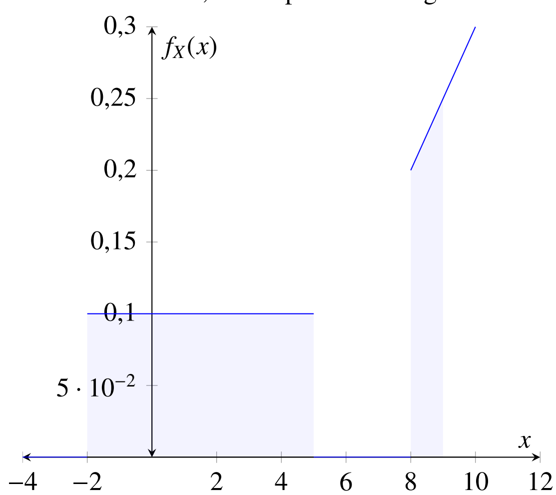


Entonces pues, $F_X(x) = \int_{-2}^{x} 1/10dt = (x + 2)/10$.

Caso 3: $5 < x < 8$. En esta región, la densidad es cero, por lo tanto no se acumulará más probabilidad, necesariamente tendremos el caso anterior evaluado en el extremo derecho ($x = 5$). Mirando el gráfico sería:

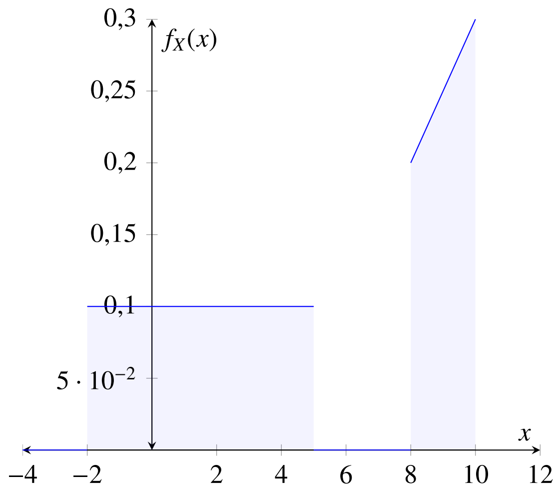


Con lo cual, nos queda en este caso $F_X(x) = (5+2)/10 = 0{,}7$.
Caso 4: $8 < x < 10$, esto representa la siguiente situación:



Es decir, a todo lo que venía acumulando de antes se le agrega $\int_8^x (1/60)tdt = (1/120)t^2\Big|_8^x = \frac{x^2-8^2}{120}$. Entonces, esta parte nos quedará $F_X(x) = 0{,}7 + \frac{x^2-8^2}{120}$.

Caso 5: $x > 10$, acumulamos todo el soporte de la distribución, necesariamnte deberá ser 1.



Así pues, tenemos finalmente la siguiente expresión:

$$F_X(x) = \begin{cases} 0 & x < -2 \\ (x+2)/10 & -2 \le x < 5 \\ 0{,}7 & 5 \le x < 8 \\ 0{,}7 + \frac{x^2-8^2}{120} & 8 \le x < 10 \\ 1 & x \ge 10 \end{cases}$$

Vamos ahora a calcular un percentil, consideremos obtener $x_{0,90}$, el percentil 0,90. Haremos como hicimos antes, caso por caso, rama por rama de la F_X.

Caso 1: $x < -2$, buscamos si $0 = F_X(x) = 0{,}9$, lo cual claramente no es posible, se descarta este caso y vamos al siguiente.

Caso 2: $-2 \le x < 5$, buscamos si $(x+2)/10 = F_X(x) = 0{,}9$. Despejamos y nos sale que $x = 7$, pero esto está fuera del rango $-2 \le x < 5$, con lo cual tampoco hay solución. Vamos al siguiente caso.

Caso 3: $5 \le x < 8$, buscamos si $0{,}7 = F_X(x) = 0{,}9$, lo cual es absurdo. Pasamos al próximo caso.

Caso 4: $8 \le x < 10$, buscamos si $0{,}7 + \frac{x^2-8^2}{120} = F_X(x) = 0{,}9$, despejamos y sale que $x = -9{,}38$ o $x = 9{,}38$, el único que cae dentro del rango es $x = 9{,}38$, es ese el percentil 0.90 que estábamos buscando.

7.6. Ejercitacion

1. Considere la siguiente función de densidad de la variable aleatoria Z que mide la distancia al origen del punto de impacto de un dardo en una diana:

$$f_Z(z) = \begin{cases} 2/7 & 0 \le z < 2 \\ 10/21 - (2/21)z & 2 \le z < 5 \\ 0 & \text{en otro caso} \end{cases}$$

   a) Graficar f_Z y verificar que sea una función de densidad.

   b) Calcular la probabilidad de que el impacto se produzca en la corona de radios 1 y 4. Es decir, obtener $P(1 < Z < 4)$.

2. Sea X una v.a. continua con densidad

$$f_X(x) = \begin{cases} 0{,}75(1 - x^2) & -1 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

a) Verificar que f_X sea una densidad.

b) Calcular $P(X > 0)$, $P(-0{,}5 < X < 0{,}5)$, $P(|X| > 0{,}25)$ y $P(X < 0{,}3|X > -0{,}5)$.

3. La temperatura T de solidificación del contenido líquido de ciertos frascos puede considerarse una variable aleatoria continua cuya función de densidad es

$$f_T(t) = \begin{cases} 1+t & -1 < t < 0 \\ 1-t & 0 \leq t < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Determinar un valor t_0 para que la probabilidad de que la temperatura de solidificación sea por lo menos t_0 resulte ser 0.08.

4. Sea X una variable aleatoria continua con función de densidad

$$f_X(x) = \begin{cases} (1/6)x + k & 0 < x < 3 \\ 0 & \text{en otro caso} \end{cases}$$

a) Calcular k para que f_X sea efectivamente una densidad.

b) Hallar $P(1 \leq X \leq 2)$.

c) Hallar F_X, $E(X)$ y la mediana.

5. Sea X con una función de distribución acumulada

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

a) ¿Cuál es el valor de a?

b) Calcular $P(X \leq 1$ y $P(0{,}5 \leq X \leq 1$.

c) Hallar f_X.

d) Hallar $E(X)$ y el percentil 0.9.

6. Consideremos X una v.a. con densidad $f_X(x) = \frac{1+\alpha x}{2} I_{(-1,1)}(x)$, siendo $-1 < \alpha < 1$.

a) Probar que f_X es efectivamente una densidad para todo $-1 < \alpha < 1$.

b) Hallar F_X.

c) Hallar $E(X)$ y $V(X)$, que pueden quedar en función de α.

d) Hallar la mediana y los cuartiles de la distribución (los percentiles 0.25 y 0.75).

7. Consideremos X una v.a. con densidad $f_X(x) = (a + bx^2)I_{(0,1)}(x)$.

a) Se sabe que $E(X) = 0{,}6$, hallar a y b.

b) Hallar la función de distribución acumulada F_X.

8. Consideremos una v.a. Y con densidad

$$f_Y(y) = \begin{cases} \frac{1}{25}y & 0 \leq y < 5 \\ \frac{2}{5} - \frac{1}{25}y & y \leq y < 10 \\ 0 & \text{en otro caso} \end{cases}$$

a) Hallar F_Y.

b) Hallar la mediana de Y.

c) Hallar $E(Y)$ y $V(Y)$.

d) Calcular $E(1/Y)$. ¿Coincide con $1/E(Y)$?

9. La biblioteca de una facultad dispone de una red de computadoras al alcance de los estudiantes. La proporción de tiempo (en porcentaje entre 0 y 100) que un usuario destina a búsqueda bibliográfica es una variable aleatoria T con función de densidad

$$f_T(t) = c(100 - t)I_{[0,100]}(t).$$

a) Hallar el valor de la constante c.

b) Supóngase que de acuerdo con el porcentaje de tiempo destinado a la búsqueda bibliográfica el usuario es clasificado dentro de cuatro categorías: 1 si $T < 25$, 2 si $25 < T < 50$, 3 si $50 < T < 75$ y 4 si $T > 75$. Hallar la distribución de la categoría asignada a un usuario. Notar que la categoría conformaría una variable discreta.

10. El diámetro (expresado en dm) del tronco de cierta especia de árboles es una v.a. continua con densidad

$$f_D(x) = kxI_{(0,10)}(x).$$

a) Hallar el valor de k.

b) ¿Cuál es la probabilidad de que un árbol elegido al azar tenga un tronco con diámetro comprendido entre 4 dm y 6 dm?

c) Ídem anterior pero sabiendo de antemano que el diámetro es mayor a 5 dm.

d) En un área del bosque hay 6 árboles de esa especie. ¿Cuál es la probabilidad de que exactamente 4 de ellos tengan troncos con diámetro comprendido entre 4 dm y 6 dm?

11. Sean f y g dos funciones de densidad, demostrar que si $\alpha in(0, 1)$ luego $h(x) = \alpha f(X)+(1-\alpha)g(x)$ también es una densidad. ¿Cuál es el valor esperado de esta nueva distribución?

8. Familias de Variables Continuas

Así como teníamos las familias de variables discretas más ”clásicas” (Binomial, Geométrica, Hipergeométrica, Poisson, entre otras), lo mismo pasará en el caso continuo. Tendremos una serie de variables aleatorias arquetípicas, que están clasificadas así por su utilización frecuente.

Empezaremos con la más sencilla de esa, la distribución uniforme.

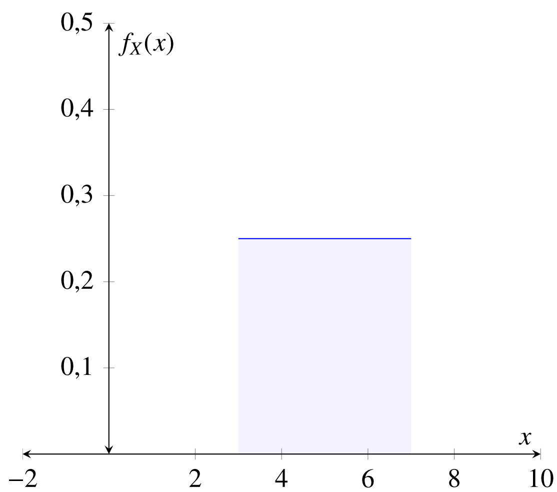
8.1. Distribución Uniforme

Diremos que X tiene una distribución **uniforme** con parámetros $a < b$ reales, y lo notaremos como $X \sim U(a, b)$, si su densidad es como sigue:

$$f_X(x) = \frac{1}{b-a} I_{(a,b)}(x).$$

Gráficamente, sería lo siguiente:

Gráficamente, tenemos algo de la pinta:



Es decir, es una variable aleatoria que tiene la misma densidad en todo su soporte (a, b). Representa la idea de una variable con densidad uniforme (como su nombre lo indica) dentro de todo su soporte (a, b).

Obtengamos la esperanza.

$$E(X) = \int_{\mathbb{R}} x f_X(x) dx = \int_a^b x \frac{1}{b-a} dx = \left. \frac{x^2}{2(b-a)} \right|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(a+b)}{2(b-a)} = \frac{a+b}{2}.$$

Una cuenta similar permite obtener que

$$V(X) = \frac{(b-a)^2}{12}.$$

Así mismo, la función de distribución acumulada es

$$F_X(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$

8.1.1. Ejemplo

Se tiene una vara de un metro, se pinta un punto $0 < p < 1$ fijo en la vara. Luego, se quiebra a la vara en un punto elegido al azar, quedando entonces dos pedazos. Uno de esos pedazos contiene al punto p y tendrá una longitud dada. Sea L esa longitud, calcule el valor esperado de L.

Rta:

Definimos la variable aleatoria U que es el punto en donde se corta la vara. Como todo se elige completamente al azar, tendremos que U será uniforme entre 0 y 1.

La longitud de la vara que contiene a p será U si si el punto p está debajo de U, y en caso contrario será $1 - U$. O sea,

$$L = \begin{cases} 1 - U & U < p \\ U & U > p \end{cases}$$

L como variable aleatoria es función de U ($L = L(U)$), podemos entonces apelar a la fórmula del estadístico inconsciente. Así pues,

$$E(L) = E(L(U)) = \int_{\mathbb{R}} L(u) f_U(u) du = \int_0^1 L(u) \cdot 1 du = \int_0^1 L(y) du = \int_0^p (1 - u) du + \int_p^1 u du$$

y esto se resuelve como

$$= 1/2 + p(1 - p).$$

Si quisiéramos elegir un p de forma tal que esto se haga máximo, es fácil verificar que con $p = 1/2$ maximizaremos lo deseado.

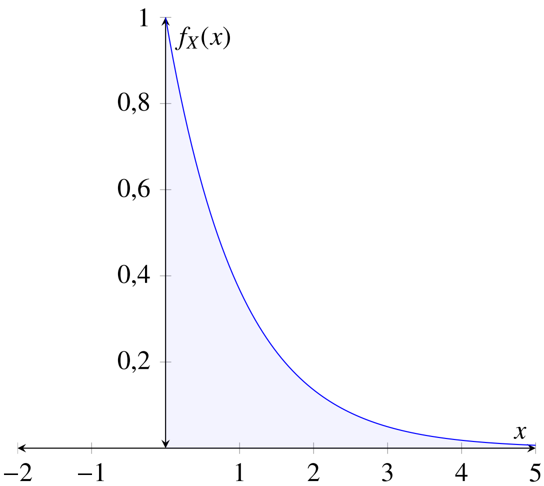
8.2. Distribución Exponencial

La distribución **exponencial** sirve para modelar situaciones en donde por ejemplo tenemos la duración o vida útil de un objeto (lamparita, por ejemplo) que empieza en 0 y que tiene lo que se conoce como ”falta de desgaste”. Es decir, es un objeto que se lo ”arranca” y empieza a durar en el tiempo, hasta que eventualmente en algún momento se rompe y deja de funcionar. Ahora bien, en el caso exponencial se supone además que la probabilidad de que deje de funcionar no depende del tiempo ya vivido hasta entonces. Formalizaremos mejor eso más adelante.

Formalmente, tenemos que fijado un $\lambda > 0$ real, $X \sim \varepsilon(\lambda)$ tiene la siguiente densidad:

$$f_X(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x).$$

Gráficamente, será algo de esta pinta:



Calculemos la distribución acumulada. Para el caso $x \leq 0$ es claro que $F_X(x) = 0$. Ahora bien, si $x > 0$, luego

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_0^x \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_0^x = 1 - e^{-\lambda x}.$$

Así mismo, obtengamos la esperanza.

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^{\infty} x\lambda e^{-\lambda x} dx \lambda \int_0^{\infty} xe^{-\lambda x} dx$$

Este tipo de integrales se resuelve bien utilizando el método de integración por partes. Recordemos brevemente el método, tenemos que:

$$f(x)g(x)\Big|_a^b = \int_a^b (fg)' dx = \int_a^b f'g dx + \int_a^b fg' dx$$

de donde sale que

$$\int_a^b f'g dx = f(x)g(x)\Big|_a^b - \int_a^b fg' dx.$$

Elegimos x para derivar (es decir, g) y $e^{-\lambda x}$ para integrar (es decir, f'). Nos queda:

$$E(X) = \lambda \left[xe^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} \frac{e^{-\lambda x}}{\lambda} dx \right]$$

Un argumento sencillo de L'Hospital nos permite ver que $xe^{-\lambda x}\Big|_0^{\infty} = 0$. Luego, resolviendo la integral (y tachando el λ) nos queda que

$$E(X) = \frac{-e^{-\lambda x}}{\lambda}\Big|_0^{\infty} = \frac{1}{\lambda}.$$

Una cuenta similar permite deducir que $V(X) = 1/\lambda^2$.

8.2.1. Falta de Desgaste

Observemos una propiedad ”curiosa” de la distribución exponencial, la que en el fondo marca su grado de utilidad. Supongamos que X mide la duración de un determinado objeto, y sabemos que esa duración responde a una distribución exponencial de parámetro λ. Sabemos que el objeto en cuestión viene durando al menos s unidades de tiempo, nos preguntamos la probabilidad de que dure al menos otras t unidades de tiempo. Es decir, formalmente, queremos $P(X > s + t | X > s)$. Calculemos esto:

$$P(X > s + t | X > s) = \frac{P(X > s + t \cap X > s)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)} = \frac{1 - F_X(s + t)}{1 - F_X(s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}.$$

Es fácil verificar que $P(X > s) = e^{-\lambda s}$. De esto, deducimos entonces que

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

O sea, la probabilidad de que el artefacto dure t unidades adicionales a las s, sabiendo que ya duró s, es lo mismo que considerar un artefacto nuevo y preguntarnos la probabilidad de que dure t unidades. En resumidas cuentas, todo tiempo pasado del artefacto no importa, no hay efecto de ”desgaste”, que eso sería que cuánto más viejo el artefacto, más probabilidad tiene de dejar de funcionar. También se lo conoce como la propiedad de ”falta de memoria”, pues se ”olvida” el tiempo pasado.

Esta propiedad es privativa de la distribución exponencial, no existe otra distribución continua con soporte en los reales positivos con la propiedad de falta de memoria.

Por ahora, lo que tenemos es que si X es una distribución exponencial. Dejamos como material optativo, a continuación, la recíproca. Eso es, si X es una variable aleatoria continua con la propiedad de falta de memoria, entonces necesariamente deberá ser la distribución exponencial. La demostración no es complicada, pero usa algunos argumentos técnicos que requieren cierta atención. Así pues, vamos con eso:

La propiedad de falta de memoria significa que para todo $s, t > 0$, tenemos que

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

Pues bien, desarrollando el condicional tenemos que

$$\frac{P(X > s + t)}{P(X > s)} = P(X > t).$$

De aquí despejamos y tenemos que

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

Definamos ahora, para simplificar la notación, $g(x) = P(X > x)$. Tenemos pues que g verifica la siguiente relación funcional:

$$g(s + t) = g(s)g(t).$$

Pequeña observación, notemos que

$$g(2) = g(1 + 1) = g(1)g(1) = g(1)^2.$$

En general, sale fácil que si $n \in \mathbb{N}$ entonces

$$g(n) = g(1)^n.$$

Por otro lado,

$$g(1) = g(1/2 + 1/2) = g(1/2)^2$$

de donde sale que

$$g(1/2) = \sqrt{g(1)}.$$

De la misma forma, se deduce que si $n \in \mathbb{N}$ entonces

$$g(1/n) = \sqrt[n]{g(1)} = g(1)^{1/n}.$$

Pues bien, luego tendremos que si $m, n \in \mathbb{N}$, sale que

$$g(m/n) = g(1/n + \ldots + 1/n) = g(1/n) \ldots g(1/n) = g(1/n)^m = g(1)^{m/n}.$$

O sea, que si a es un número racional positivo, luego

$$g(a) = g(1)^a$$

El próximo paso es ver que esto se verifica para $x \in \mathbb{R}$. Esto se deduce por el hecho de que X es una variable aleatoria continua, de donde sale que F_X, como habíamos dicho, conforma una función continua. Luego, g será una función continua. Definamos $h : \mathbb{R}_{>0} \to \mathbb{R}$ como

$$h(x) = g(1)^x.$$

Tenemos pues que h y g coinciden en lo que se conoce como un conjunto "denso" de los reales positivos, que son los racionales positivos. Como ambas funciones son continuas, necesariamente deben coincidir en todo su dominio. En efecto, sea $x \in \mathbb{R}_{>0}$. Luego, existe una sucesión $\{q_n\}_{n\in\mathbb{N}}$ de racionales tales que $q_n \to x$. Así pues, usando que ambas funciones son continuas (luego pueden sacar y meter límites):

$$h(x) = h(\lim_{n\to\infty} q_n) = \lim_{n\to\infty} h(q_n) = \lim_{n\to\infty} g(q_n) = g(\lim_{n\to\infty} q_n) = g(x).$$

Así pues, $g(x) = h(x)$ en todos los reales positivos. Luego, en conclusión:

$$g(x) = g(1)^x = e^{\ln(g(1))x}.$$

Pero

$$e^{\ln(g(1))x} = g(x) = P(X > x) = 1 - F_X(x).$$

Derivando, nos sale que $F_X(x)' = f_X(x)$, y luego tendremos que

$$f_X(x) = -\ln(g(1))e^{\ln(g(1))x}.$$

Llamando $\lambda = -\ln(g(1))$ nos sale que

$$f_X(x) = \lambda e^{-\lambda x}$$

que es la densidad de una distribución exponencial.

8.2.2. Relación con el proceso de Poisson (*)

Este material es optativo, pero consideramos que es de interés pues permite relacionar conceptos de procesos aleatorios discretos que habíamos visto antes.

Supongamos que tenemos un proceso de Poisson $\{X_t\}_{t>0}$ de tasa λ. Esto puede representar por ejemplo la cantidad de aviones que llamadas que llegan a una central telefónica en el tiempo. Queremos obtener una idea del tiempo de espera entre dos eventos (llamadas) instantáneos. Supongamos T una variable aleatoria continua que mide el tiempo entre un evento y el próximo. Queremos obtener una idea de la distribución de T.

Calculemos primero $F_T(t) = P(T \leq t) = 1 - P(T > t)$. Conceptualmente, el evento $T > t$ representa una situación en donde en un intervalo de tiempo de longitud t no hubo ningún evento. O sea, que si tomamos nuestro proceso de Poisson fijado en un tiempo t, entonces no ocurrió ningún evento. Es decir, $P(T > t) = P(X_t = 0) = e^{-t\lambda}(\lambda t)^0/0! = e^{-\lambda t}$ (esto sale de la expresión de la fórmula de Poisson).

Luego pues, $F_T(t) = 1 - e^{\lambda t}$ que es exactamente la función de distribución acumulada de una exponencial de parámetro λ.

Así, el tiempo de espera entre dos eventos de un proceso de Poisson de tasa λ será una exponencial de parámetro λ.

8.3. Ejercicios

8.3.1. Resueltos

Ejercicio 8.1. *Una caja tiene 10 lamparitas de tipo A, cuya duración tiene una distribución exponencial de media* 5 *meses, y 5 lamparitas de tipo B, cuya duración tiene una distribución uniforme* $U(3, 9)$ *meses. Se extraen una lamparita al azar y resulta que estuvo al menos 6 meses prendida, cuál es el tipo de lamparita más probable?*

Rta. Supongamos $X =$"duración en meses de la lamparita elegida". Lo que pide este problema es en esencia calcular $P(A|X \geq 6)$ y $P(B|X \geq 6)$. En realidad, como la suma de los dos debe dar 1 (si no es de tipo A será de tipo B) entonces basta calcular $P(A|X \geq 6)$ y compararlo con 0.5. Si es mayor a 0.5, luego lo más probable es que la lamparita sea de tipo A, caso contrario será de tipo B.

Es un caso clásico de Bayes, pues está condicionado el pasado dado el futuro. Entonces:

$$P(A|X \geq 6) = \frac{P(A \cap X \geq 6)}{P(X \geq 6)} = \frac{P(A \cap X \geq 6)}{P(X \geq 6 \cap A) + P(X \geq 6 \cap B)} = \frac{P(X \geq 6|A)P(A)}{P(X \geq 6|A)P(A) + P(X \geq 6|B)P(B)}$$

Calculemos cada una de las cosas que aparecen. Claramente:

$$P(A) = 10/15$$

y

$$P(B) = 5/15.$$

Ahora bien, para obtener $P(X \geq 6|A)$ necesitamos saber la distribución exacta de las lamparitas de tipo A. Como nos dicen que la media (es decir, la esperanza) es de 5 meses, luego recordando que la esperanza de una exponencial de parámetro $\lambda > 0$ es $1/\lambda$, deducimos que $X|_A \sim \varepsilon(\lambda = 1/5)$.

Luego, $P(X \geq 6|A) = 1 - F_{X_A}(6) = 1 - (1 - e^{-6/5}) = e^{-6/5} = 0,3023$.

En el caso de la uniforme, $P(X \geq 6|B) = \int_6^9 \frac{1}{9-3} dx = 0{,}5$.

Luego, juntando todo tendremos que

$$P(A|X \geq 6) = 0{,}5473.$$

O sea, que lo más probable es que sea una lamparita de tipo A.

Ejercicio 8.2. *Sea X una variable aleatoria con distribución* $U(0, 1)$*. Definimos* $Y = -\ln(X)$*, demuestre que su distribución es exponencial. Pista: expresar primero* F_Y

Rta.

El truco para este tipo de problemas que consisten en una transformación de otra variable aleatoria se basa en usar la función de distribución acumulada. La ventaja que tiene la misma es que expresa una probabilidad, y por ende es posible aplicar las propiedades de probabilidad. Veamos:

$$F_Y(y) = P(Y \leq y) = P(-\ln(X) \leq y) = P(X \geq e^{-y}) = 1 - F_X(e^{-y}).$$

Derivemos F_Y con respecto a y. Vamos a tener que usar regla de la cadena y recordar que la derivada de F_X respecto de x es f_X, la densidad. Nos queda pues:

$$f_Y(y) = (F_Y(y))' = f_X(e^{-y})e^{-y}.$$

Pero $f_X(e^{-y)} = I_{(0,1)}(e^{-y})$, recordando que X es una $U(0,1)$. Es decir, quedará e^{-y} si $0 < e^{-y} < 1$ y 0 sino. Ahora bien, esta condición se traduce, despejando, en que $0 < y < \infty$. Luego, nos queda que

$$f_Y(y) = e^{-y} I_{(0,\infty)}(y)$$

que se ajusta exactamente a una exponencial de parámetro $\lambda = 1$.

8.3.2. Extras

Ejercicio 8.3. *Se tiene X una distribución exponencial. Compare cuál es más grande, la media o la mediana de X.*

Ejercicio 8.4. *Se tiene un sistema electrónico compuesto de 5 partes que funcionan independientemente y donde la duración de cada una es una distribución exponencial de media 10 meses. El sistema solamente funciona en la medida de que funcionen las 5 partes a la vez. Cuál es la probabilidad de que el sistema funcione al menos 8 meses? Y qué funcione a lo sumo 12 meses?*

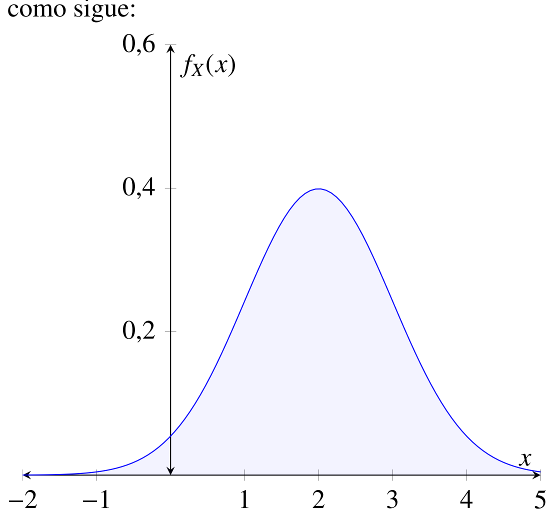
8.4. Distribución Normal

Esta probablemente sea la distribución más importante de todas, incluso entre discretas y continuas. La razón de su utilización no es tan fácil, por ahora, de dar a cuenta. Lo que haremos en esta sección es definir esta distribución, ver sus propiedades y cómo operar con ella. La razón de ser de su existencia la dejaremos para cuando veamos un resultado que se conoce como el Teorema Central del Límite (TCL).

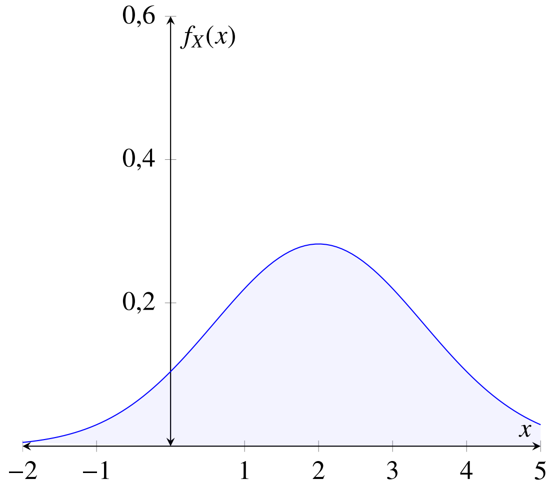
Diremos que X variable aleatoria continua tiene distribución **normal** de parámetros $\mu \in \mathbb{R}$ y $\sigma^2 \in \mathbb{R}_{>0}$, y se notará como $X \sim N(\mu, \sigma^2)$ si su densidad tiene la siguiente expresión:

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

con $x \in \mathbb{R}$. Es decir, el soporte son todos los reales. Su gráfica es lo que se conoce como la ”campana de Gauss”, y es como sigue:



El eje de simetría de la campana está ubicado exactamente en μ, el σ^2 determina la dispersión de la distribución con respecto al eje de simetría. Veamos un gráfico similar con un σ^2 mayor:



Como primera cuenta, veamos que efectivamente esto es una densidad y que integra 1. Tenemos una dificultad, no es fácil caracterizar la distribución acumulada de una densidad normal. En efecto, no existe una primitiva expresable en términos de composición y operatorias algebraicas de "funciones simples", lo cual conlleva una dificultad a la hora de operar con la distribución normal. Sin embargo, para el caso de integrar en toda la recta real, podemos resolverlo con un "truco" que pasaremos a ver, consistente a convertir el problema de integración en $\mathbb{R}$ en uno de integrales múltiples en el plano cartesiano.

Queremos ver que

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

Primer paso, hagamos la sustitución $z = \frac{x-\mu}{\sigma}$. Es fácil ver que después de esta sustitución, nos quedará:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz.$$

Pues bien, queremos ver que esto integra 1. Llamemos L a esta integral, el objetivo es ver que L vale 1. Vamos a calcular en su lugar L^2, si probamos que $L^2 = 1$ ya estará pues L es positivo al ser una integral de una función positiva. Ahora bien, lo que tenemos es que

$$1 = L^2 = \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz\right)^2 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_1^2}{2}}\, dz_1 \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_2^2}{2}}\, dz_2.$$

Usando Fubini, podemos escribir esto como:

$$\int_{\mathbb{R}^2} \frac{1}{2\pi} e^{-\frac{z_1^2+z_2^2}{2}}\, dz_1 dz_2.$$

Realizamos un pasaje a coordenadas polares de esto. Nos quedará:

$$\int_0^{\infty} \int_0^{2\pi} \frac{1}{2\pi} \rho e^{-\rho^2/2} d\phi d\rho = \int_0^{\infty} \rho e^{-\rho^2/2} d\rho.$$

Estamos en una mejor situación, esta integral sí se puede resolver, admite una primitiva que sale haciendo sustitución. En efecto, tomamos $u = \rho^2/2$, luego $du = \rho d\rho$, con lo cual nos queda:

$$\int_0^\infty e^{-u} du = -e^{-u}\Big|_0^\infty = 1,$$

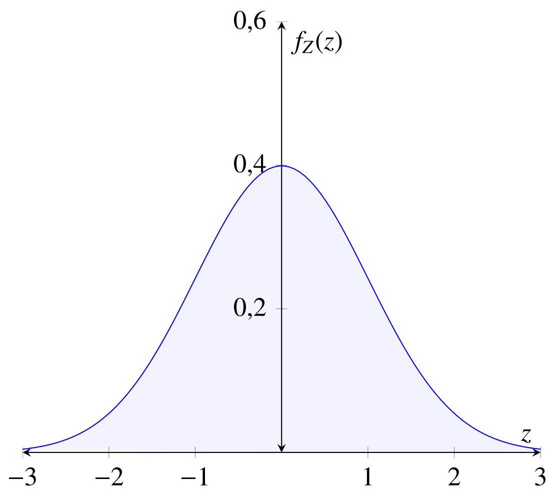
concluyendo así la demostración. Tenemos pues que f_X así definida efectivamente conforma una densidad.

8.4.1. Distribución Normal Estándar

Un caso particular, y muy importante, de la distribución normal es lo que se conoce como la distribución **normal estándar**. Esto típicamente lo notaremos como Z, y es el caso cuando $Z \sim N(\mu = 0, \sigma^2 = 1)$. Su densidad será, pues:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

Su gráfico es en esencia la de la campana de Gauss centrada en el cero, como sigue:



Veamos algunas propiedades de esta distribución particular.

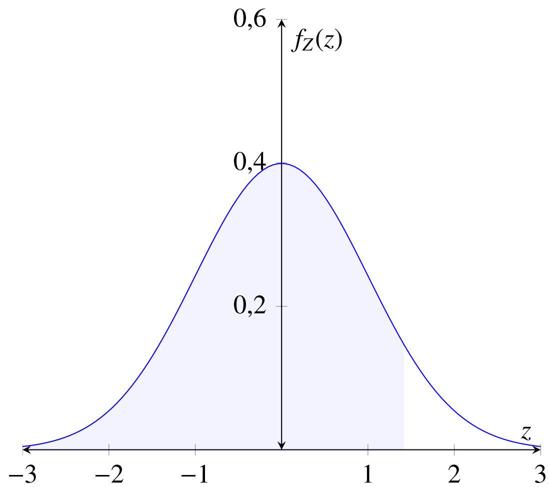
$$E(Z) = \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz.$$

Estamos integrando una función impar (es decir, anti-simétrica respecto del eje y) en todo el dominio, lo que esto nos devuelve es $E(Z) = 0$, lo cual es esperable por la simetría de la distribución. Argumentos similares permiten obtener también que $V(Z) = 1$.

Cómo se calculan probabiliadades con esta distribución? Supongamos que quisiéramos obtener $P(Z < 1{,}42)$, por poner un ejemplo. Formalmente, esto sería

$$P(Z < 1{,}42) = \int_{-\infty}^{1{,}42} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz.$$

Gráficamente, tenemos que encontrar el siguiente área:



No es fácil, por no decir imposible, obtener analíticamente esta integral. Se debe aproximar la integral numéricamente. Nosotros no nos meteremos en los detalles ”escabrosos” de lo que eso significa. Para simplificar la labor, trabajaremos con una tabla que tiene ya precalculada la distribución acumulada de la normal estándar.

La tabla que tienen disponible en la página cuenta con la acumulada para muchos valores (positivos) de la normal. Veamos cómo operar con la misma.

Queremos calcular $P(Z < 1{,}42) = F_Z(1{,}42)$. La función de distribución acumulada de la normal estándar, F_Z, se la suele notar como Φ. Tenemos tabulados varios valores de la Φ. Como queremos calcular $\Phi(1{,}42)$, lo que hacemos es lo siguiente:

Paso 1: vamos a la tabla de la distribución acumulada de la normal estándar. Buscamos la fila 1.4. Es decir, unidad ”1” y primer decimal ”4”.

Paso 2: vamos a la columna ”0.02”, el segundo decimal. Con lo cual, tenemos conformado así el número 1.42.

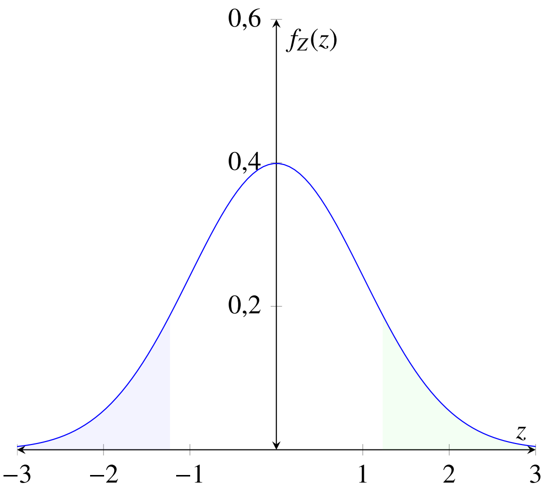
Paso 3: buscamos la celda ubicada en la intersección entre la fila y la columna y consultamos su valor, que en nuestro caso resulta ser: 0.9222. Esa será la probabilidad buscada.

Tenemos así que $P(Z < 1{,}42) = F_Z(1{,}42) = \Phi(1{,}42) = 0{,}9222$ (aproximadamente).

Calculemos otras probabilidades con la distribución normal estándar, es decir, usando la tabla.

Calculemos $P(Z > 0{,}41) = 1 - F_Z(0{,}41) = 1 - \Phi(0{,}41) = 1 - 0{,}6591 = 0{,}3409$ según la tabla, que se obtuve mirando la fila 0,4 con la columna 0,01.

Supongamos ahora que deseamos $P(Z < -1{,}23) = \Phi(-1{,}23)$. Esto es un poco más complicado, la tabla normal solamente tiene valores de Φ para casos positivos. Para obtenerla para valores negativos, miremos el siguiente gráfico:



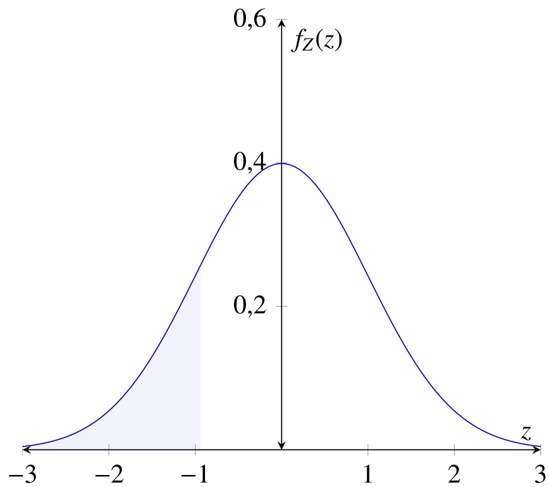
Por la simetría de la campana de Gauss, el área azul coincide con el área verde. O sea, $P(Z < -1{,}23) = P(Z > 1{,}23) = 1 - \Phi(1{,}23) = 1 - 0{,}8907 = 0{,}1093$ (sacándolo de la tabla).

Así pues, tenemos cómo calcular probabilidades con la variable normal estándar, usando su tabla de distribución acumulada. Un problema inverso a este es el de obtener percentiles. Recordemos, en general un percentil $0 < p < 1$ es un x_p tal que $F_X(x_p) = p$. En el caso de la distribución normal estándar, eso es $\Phi(z_p) = p$, o sea, $z_p = \Phi^{-1}(p)$. Necesitamos la inversa de Φ, y eso se obtiene a través de la tabla de la distribución normal estándar, pero esta vez mirando de adentro hacia afuera. Veamos un ejemplo.

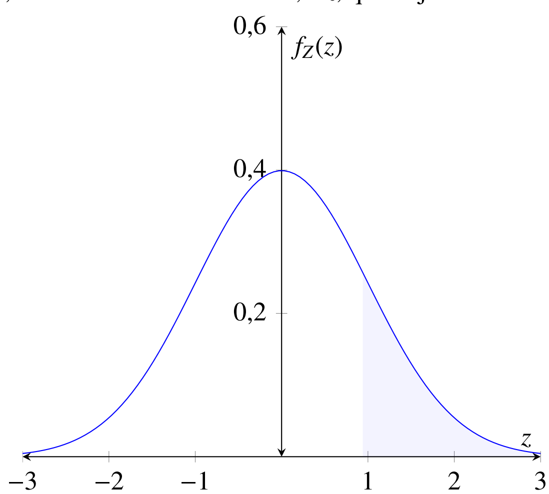
Supongamos que queremos el percentil 0,95. Es decir, un z tal que $\Phi(z) = 0{,}95$. O sea, $z_p = \Phi^{-1}(0{,}95)$. Para encontrar este percentil, debemos buscar ”de adentro hacia afuera” en la tabla. Es decir, miramos el interior de la tabla de la acumulada y buscamos el valor 0,95. En general, como puede pasar en este caso, es raro que exista exactamente ese valor dentro de la tabla, en cuyo caso tomamos el más parecido posible. Se puede mejorar esto con mecanismos de interpolación, pero nos conformaremos en términos prácticos con esta política.

Así pues, mirando la tabla vemos que el más parecido es 0.9495 (que corresponde, yendo a los márgenes, con 1.64) y 0.9505 (que corresponde con 1.65). Tomamos cualquiera en este caso, por ejemplo 1.64. Entonces pues, el percentil 0.95 es $z_p = 1{,}64$. Es decir, el valor 1.64 deja a su izquierda un área de 0.95.

Tenemos un problema parecido con los negativos, fijarse por ejemplo si queremos el percentil 0.4. Ese valor no lo encontraremos en el interior de la tabla, pues el mismo empieza con 0.5 (correspondiente a $z = 0{,}0$). Recurrimos a un truco parecido al que hicimos antes, de simetrización. Grafiquemos aproximadamnte la situación:



El z buscado es un valor tal que a su izquierda deja el área pintada de azul (que es 0.40) en el gráfico. Pues bien, consideremos el simétrico, $-z$, que dejará un área de 0.40 pero ahora a su derecha. Es decir:



O sea, tenemos que $0{,}4 = P(Z > -z) = 1 - \Phi(-z)$. Luego, $\Phi(-z) = 1 - 0{,}4 = 0{,}6$. Entonces, $-z = \Phi^{-1}(0{,}6)$. Así, $z = -\Phi^{-1}(0{,}6)$. Buscamos en el interior de la tabla 0.6, que sí es aproximadamente identificable, se realiza con 0.25. O sea, $-z = 0{,}25$. Finalmente, el percentil buscado será entonces $z_p = -0{,}25$.

8.4.2. Operatoria con la normal general: estandarización

Supongamos que tenemos problemas similares a los de antes, pero la diferencia es que ahora $X \sim N(\mu = 5, \sigma^2 = 9)$. Es decir, hemos dejado de tener una normal estándar. Nos piden calcular $P(X \leq 7)$, por ejemplo. Qué se hace? Buscamos una tabla para esos parámetros de μ y σ^2? La respuesta es negativa. Solamente tendremos una tabla que es la de la normal estándar.

Entonces pues, habremos de hacer algún mecanismo que permite convertir una normal general $X \sim N(\mu, \sigma^2)$ en una $Z \sim N(0, 1)$. Por suerte, esto es posible a través de un proceso que denominaremos ”estandarización”.

Theorem 8.5. *Sea $X \sim N(\mu, \sigma^2)$, luego si definimos $Z = \frac{X-\mu}{\sigma}$, entonces $Z \sim N(0, 1)$.*

Es importante marcar que se divide por σ y no σ^2, cuidado con eso pues es fuente de errores frecuente!

Vamos a probar este resultado. Tenemos Z definido como $\frac{X-\mu}{\sigma}$, obtengamos su distribución.

Para eso, consideremos F_Z, la distribución acumulada de esta nueva variable Z. Haremos algunas cuentas con ella.

$$F_Z(z) = P(Z \leq z) = P\left(\frac{X-\mu}{\sigma} \leq z\right) = P(X \leq \mu + \sigma z) = F_X(\mu + \sigma z).$$

Así pues, $F_Z(z) = F_X(\mu + \sigma z)$. Derivemos esto con respecto a z, lo debemos hacer según la regla de la cadena. Recordemos que la derivada de $F_X(x)$ con respecto a x es $f_X(x)$, la densidad. Entonces, nos sale que

$$f_Z(z) = f_X(\mu + \sigma z)\sigma = \sigma \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2} = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

y esto corresponde a la densidad de una normal estándar, con lo cual hemos deducido que Z tiene una distribución normal estándar. Es decir, con $\mu = 0$ y $\sigma^2 = 1$, concluyendo la prueba.

Algunas observaciones, de esto se deduce que si $Z \sim N(0, 1)$ entonces $X = \mu+\sigma Z \sim N(\mu, \sigma^2)$. Así, podemos calcular la esperanza y la varianza de una normal en general.

$$E(X) = E(\mu + \sigma Z) = \mu + \sigma E(Z) = \mu.$$

$$V(X) = V(\mu + \sigma Z) = V(\mu) + \sigma^2 V(Z) = 0 + \sigma^2 = \sigma^2.$$

Así pues, los parámetros de una normal estándar corresponden precisamente con la esperanza y la varianza de una normal.

Vamos ahora a hacer algunos ejercicios con $X \sim N(\mu = 5, \sigma^2 = 9)$. Calculemos por ejemplo $P(X \leq 7)$. El truco a tener en cuenta es que el primer paso, siempre, es estandarizar. En efecto, hacemos:

$$P(X \leq 7) = P\left(\frac{X-5}{3} \leq \frac{7-5}{3}\right) = P(Z \leq 2/3) = F_Z(2/3) = \Phi(0{,}67)$$

y buscamos este último en la tabla de la normal estándar, obteniendo 0.7486.

Calculemos un percentil, por ejemplo, el percentil $x_{0,90}$. Es decir, queremos x tal que $P(X \leq x) = 0{,}90$. Estandarizamos, como siempre hay que hacer ante un problema con una normal general, y obtenemos que

$$0{,}90 = P\left(\frac{X-5}{3} \leq \frac{x-5}{3}\right) = P\left(Z \leq \frac{x-5}{3}\right) = \Phi\left(\frac{x-5}{3}\right).$$

O sea, que $\frac{x-5}{3} = \Phi^{-1}(0{,}90)$. Buscamos este percentil (de la normal estándar) en la tabla, desde el interior a los márgenes, obtenemos que $\Phi^{-1}(0{,}90)$ es aproximadamente 1.28.

Luego, $\frac{x-5}{3} = 1{,}28$ de donde despejamos que el percentil original que buscábamos es $x = 3 \cdot 1{,}28 + 5 = 8{,}84$.

Veamos otro percentil, a fin de seguir repasando casos más complicados. Obtengamos ahora el percentil $x_{0,15}$. Procedemos como siempre, queremos que $P(X \leq x) = 0{,}15$. Estandarizamos, nos queda que

$$0{,}15 = P\left(\frac{X-5}{3} \leq \frac{x-5}{3}\right) = P\left(Z \leq \frac{x-5}{3}\right) = \Phi\left(\frac{x-5}{3}\right).$$

O sea, que $\frac{x-5}{3} = \Phi^{-1}(0{,}15)$. El problema es encontrar este percentil, el 0.15, en la tabla, que no estará pues corresponde un valor z negativo. Hacemos el truco de la simetrización, que vimos antes en el caso de la normal estándar. Tendremos que $\Phi^{-1}(0{,}15) = -\Phi^{-1}(0{,}85) = -1{,}04$. Luego, $x = 3 \cdot (-1{,}04) + 5 = 1{,}88$.

8.5. Ejercitación

1. Se llega a una parada de ómnibus a las 10AM en punto. El bus llegará en algún punto entre las 10AM y las 10:30 AM, de forma totalmente aleatoria.

   a) Calcule la probabilidad de tener que esperar 10 minutos o más hasta que llegue el bus.

   b) Son las 10:15 y todavía no llegó el bus, calcule la probabilidad de tener que esperar aún unos 10 minutos más al menos.

2. Se eligen n puntos al azar en el intervalo $[0, 1]$ de forma independiente. Sea X la cantidad de puntos que caen en el intervalor $[0, p]$ con $0 < p < 1$. ¿Qué distribución tiene X?

3. Se tiene una vara de longitud L. Se elige un punto al azar y se quiebra la vara en ese punto.

   a) Calcule la probabilidad de que el cociente entre la longitud de la pieza izquierda y la pieza derecha sea menor a 1/4.

   b) Calcule la probabilidad de que el cociente entre la longitud de la pieza más corta y la pieza más larga sea menor a 1/4.

4. Se tienen dos ciudades distanciadas por 100 kilómetros. Un colectivo realiza el recorrida entre A y B, pero en general tiene un desperfecto en un punto completamente al azar entre las dos ciudades. Precisa entonces ir hacia la estación de servicio más cercana, y se busca en general que el recorrido a efectuar sea lo menor posible. Se consideran estas dos estrategias para colocar tres estaciones de servicio:

   - Colocar una estación en A, otra a mitad camino entre A y B, y la última en B.
   - Colocar una estación a 25 km de A, otra a 50 km de A (mitad camino) y la última a 75 km de A (25 km de B).

   Obtenga el valor esperado de la distancia entre el bus y la estación más cercana según las dos estrategias y determine en función de eso la mejor estrategia.

5. Sea Y una variable aleatoria con distribución $U(0, 5)$. Consideremos el polinomio $Q(x) = 4x^2 + 4Yx + Y + 2$, calcule la probabilidad de que tenga dos raíces reales.

6. La vida útil de un cierto componente electrónico en meses es una v.a. V con distribución $\varepsilon(\lambda)$ con $P(V > 20) = 0{,}449$.

   a) Hallar $E(V)$ y $Var(V)$.

   b) Hallar la probabilidad de que la vida útil de la componente sea mayor a 10 meses.

   c) Sabiendo que la componente duró 20 meses o más, ¿cuál es la probabilidad de que en todal dure 30 meses o más? ¿Si se sabe que una componente duró al menos t meses, ¿cuál es la probabilidad de que dure al menos $s + t$ meses?

7. Un programa tiene 5 bloques, el tiempo de compilación en segundos de cada bloque es una variable aleatoria exponencial de parámetro $\lambda = 1$ y es independiente del tiempo de compilación de los otros bloques. Calcule la probabilidad de que exactamente tres bloques tengán un tiempo de compilación superior a los 2 segundos.

8. Una caja tiene 10 baterías de tipo A, con una duración en meses que sigue una distribución exponencial con $\lambda_1 = 1/5$. Además, hay 5 baterías de tipo B con una duración en meses que sigue una distribución también exponencial pero con $\lambda_2 = 1/10$.

a) Se elige una batería al azar y se sabe que a los 5 meses sigue funcionando, ¿cuál es la probabilidad de que dure más de 10 meses?

b) Calcule el valor esperado de la duración de la batería elegida al azar.

9. Sea Z una v.a. con distribución $N(0,1)$. Calcular:

   a) $P(0 \leq Z \leq 2)$.

   b) $P(|Z| \leq 2{,}5)$.

   c) $P(Z \geq -1{,}37)$.

   d) c tal que $P(Z \leq c) = 0{,}98$.

   e) c tal que $P(Z \leq c) = 0{,}32$.

   f) c tal que $P(|Z| \leq c) = 0{,}90$.

   g) Valor z_α para $\alpha = 0{,}05$, $0{,}025$, $0{,}01$ y $0{,}005$ tal que $P(Z > z_\alpha) = \alpha$.

10. Sea X una v.a. con distribución $N(\mu = 5, \sigma^2 = 0{,}25)$. Calcular:

    a) $P(4{,}75 \leq X \leq 5{,}50)$.

    b) $P(|X| > 5{,}75)$.

    c) c tal que $P(|X - 5| \leq c) = 0{,}90$.

11. Se supone que en cierta población humana, el índice cefálico I (anchura del cráneo expresada como porcentaje de la longitud) es una v.a. con distribución $N(\mu, \sigma^2)$. Se sabe que hay un 58 % de individuos con $I \leq 75$, un 38 % con $65 < I \leq 80$ y un 4 % con $I \geq 80$, hallar los parámetros μ y σ^2 y calcule $P(78 \leq I \leq 82)$.

12. La resistencia aeróbica de las personas jóvenes no fumadoras que no se entrenan para una maratón tiene distribución $N(80{,}9^2)$, mientras que la resistencia de las personas fumadoras tiene distribución $N(60, 8^2)$.

    a) Hallar el valor de la resistencia aeróbica que es superado exactamente por el 80 % de las personas no fumadoras.

    b) En cierta maratón, hay inscripta una persona fumadora por cada tres que no fuman. Se elige un corredor al azar, ¿cuál es la probabilidad de que su resistencia aeróbica sea mayor que 75?

    c) Sabiendo que la resistencia aeróbica del participante elegido es mayor que 75, ¿cuál es la probabilidad de que sea fumador?

13. Existen tres especies de nueces, las propias del cantón de St. Gallen (Suiza) cuyo peso en gramos es una variable aleatoria cuya función de densidad es como sigue:

$$f_A(x) = \begin{cases} c & 0 \leq x \leq 9 \\ kx^2 & 10 \leq x \leq 15 \\ 0 & \text{en otro caso} \end{cases}$$

siendo k y c constantes positivas desconocidas. Se sabe que exactamente 90 % de las nueces de esta especie pesan menos de 12 gramos. La otra especie proviene de Bavaria (Alemania), cuyo peso en gramos tiene una distribución uniforme con valor mínimo de 5 gramos y mediana de 9 gramos. Por último, tenemos las nueces de Gmunden (Austria), cuyo peso sigue una distribución normal con media desconocida y desvío estándar 3 gramos. Se sabe que la probabilidad de que una nuez de Gmunden pese más de 11 gramos es 0.3.

Se juntan en una caja 30 nueces de St. Gallen, 20 de Bavaria y 15 de Gmunden.

a) Se extrae una nuez y se determinó que pesaba más de 10 gramos, ¿cuál de los tres tipos de nueces sería el más probable?

b) Calcule el valor esperado del peso neto de esta caja.

9. Vectores Aleatorios

9.1. Introducción

En este capítulo, empezaremos a trabajar de forma conjunta con varias variables aleatorias. Estudiaremos la manera en que se deberá trabajar cuando se consideran más de una variable aleatoria a la vez.

Inicialmente, consideraremos solamente dos variables aleatorias, X e Y, como partes de un mismo experimento aleatorio.

Por ejemplo, podríamos estar muestreando personas de una determinada población, entonces definimos a X como la altura del individuo muestreado e Y su altura. Nos interesaría saber cómo varía por ejemplo una variable en función de otra. Hay alguna ”relación” entre ambas? Es razonable suponer independencia?

Vamos a atacar este tipo de preguntas en este capítulo.

9.2. Caso Discreto

Vamos a trabajar con X e Y dos variables aleatorias discretas, provenientes de un mismo experimento aleatorio. En base al mismo, construimos el vector aleatorio (X, Y) discreto.

Se define la **función de probabilidad puntual conjunta** a $p_{XY}(x, y)$, definido como

$$p_{XY}(x, y) = P(X = x, Y = y).$$

Así como en el caso de una única variable aleatoria, tendremos la definición de **rango del vector aleatorio**, $R_{XY} = \{(x, y) \mid p_{XY}(x, y) > 0\}$. Es decir, es la región en donde la probabilidad puntual es estrictamente positiva.

Algunas propiedades:

Proposition 9.1. *Sea* (X, Y) *un vector aleatorio discreto y* p_{XY} *su probabilidad puntual. Vale que:*

1. $p_{XY}(x, y) \geq 0$.
2. $R_{XY} \subset R_X \times R_Y$.
3. $\sum_{(x,y)\in R_{XY}} p_{XY}(x, y) = 1$

Demostración. 1) $p_{XY}(x, y) = P(X = x, Y = y) \geq 0$ pues es una probabilidad.

2) Supoongamos que $(x, y) \in R_{XY}$, luego tendremos que $p_{XY}(x, y) > 0$. Ahora bien, $0 < p_{XY}(x, y) = P(X = x, Y = y) = P(X = x \cap Y = y) \leq P(X = x)$ ya que el evento $\{X = x, Y = y\}$ está contenido en el evento $\{X = x\}$.

Pero $P(X = x) = p_X(x)$. Entonces, tenemos que $p_X(x) > 0$, es decir, $x \in R_X$. Análogamente, tendremos que $y \in R_Y$. Luego, $(x, y) \in R_X \times R_Y$.

Es decir, que $R_{XY} \subset R_X \times R_Y$.

3) Por el axioma de aditividad, ya que trabajamos sobre eventos disjuntos, tendremos que

$$\sum_{(x,y)\in R_{XY}} p_{XY}(x, y) = \sum_{(x,y)\in R_{XY}} P(X = x, Y = y) = P\left(\bigcup_{(x,y)\in R_{XY}} X = x, Y = y\right) = 1$$

por ser P una probabilidad y estar considerando un conjunto que abarca todos los posibles casos del espacio muestral.

□

9.2.1. Ejemplo

Veamos un ejemplo a fin de poder fijar algunas de estas ideas. Supongamos que tenemos una urna con 3 bolitas rojas, 4 blancas y 5 azules. Se obtiene una muestra (extracciones sin reposición) de tres bolitas, sea X la cantidad de bolitas rojas e Y la cantidad de bolitas blancas.

Obtengamos la distribución del vector aleatorio discreto (X, Y), representado por p_{XY}. Vamos a ir elemento por elemento.

$$p_{XY}(0,0) = \binom{5}{3}/\binom{12}{3} = \frac{10}{220}.$$

$$p_{XY}(0,1) = \binom{5}{2}\binom{4}{1}/\binom{12}{3} = \frac{40}{220}.$$

$$p_{XY}(0,2) = \binom{5}{1}\binom{4}{2}/\binom{12}{3} = \frac{30}{220}.$$

$$p_{XY}(0,3) = \binom{4}{3}/\binom{12}{3} = \frac{4}{220}.$$

$$p_{XY}(1,0) = \binom{3}{1}\binom{5}{2}/\binom{12}{3} = \frac{30}{220}.$$

$$p_{XY}(1,1) = \binom{3}{1}\binom{4}{1}\binom{5}{1}/\binom{12}{3} = \frac{60}{220}.$$

$$p_{XY}(1,2) = \binom{3}{1}\binom{4}{2}/\binom{12}{3} = \frac{18}{220}.$$

$$p_{XY}(2,0) = \binom{3}{2}\binom{5}{1}/\binom{12}{3} = \frac{15}{220}.$$

$$p_{XY}(2,1) = \binom{3}{2}\binom{4}{1}/\binom{12}{3} = \frac{12}{220}.$$

$$p_{XY}(3,0) = \binom{3}{3}/\binom{12}{3} = \frac{1}{220}.$$

Es fácil convencerse que cualquier otra posibilidad tendrá probabilidad 0. Expresado en forma de tabla de doble entrada, tendremos la siguiente estructura con

$X \downarrow \mid Y \rightarrow$	0	1	2	3	$p_X(x)$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
$p_Y(y)$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

En cada celda interios, tenemos entonces $p_{XY}(x, y)$, mientras que en los márgenes nótese que tenemos, precisamente, lo que se harán de llamar las **probabilidades marginales**.

Es relativamente sencillo probar, usando probabilidad total, que

$$p_X(x) = \sum_{y \in R_Y} p_{XY}(x, y)$$

y que

$$p_Y(y) = \sum_{x \in R_X} p_{XY}(x, y).$$

Por otro lado, la suma de todas las celdas interiores deberá ser, como habíamos visto antes, 1.

9.2.2. Representación de condicional

Tenemos así construidas $p_{XY}(x, y)$, vamos a dar una idea de definición para lo que se llamará la **distribución condicional** $p_{X|Y=y}(x)$. Su definición es bastante natural, la idea es que estamos representando la distribución ahora de la variable aleatoria X condicionado a un valor de $Y = y$.

Tenemos así que

$$p_{X|Y=y}(x) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{XY}(x, y)}{p_Y(y)}.$$

Análogamente, tenemos así que

$$p_{Y|X=x}(y) = P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)} = \frac{p_{XY}(x, y)}{p_X(x)}.$$

Siguiendo el caso de nuestro ejemplo, podemos supongamos que tenemos observado $Y = 1$, queremos obtener la distribución condicional $X|Y = 1$, lo que se reduce a la siguiente tabla:

X	$p_{X\|Y=1}(x)$
0	40/112
1	60/112
2	12/112

Nuevamente, es importante destacar que la distribución condicional es a su vez una distribución, cumpliendo todas las definiciones. Por ejemplo, la suma de las probabilidades puntuales condicionales deberá dar 1, cosa que por ejemplo podemos ver que se verifica en nuestro caso.

9.2.3. Esperanza, Covarianza, Independencia y Correlación

Nuevamente, en el afán de recuperar las construcciones que teníamos de antes, tenemos así una definición de esperanza. Dada una función $h : \mathbb{R}^2 \to \mathbb{R}$ entonces podemos definir

$$E(h(X, Y)) = \sum_{(x,y) \in R_{XY}} h(x, y) \cdot p_{XY}(x, y),$$

tal como hacíamos con la fórmula del ”estadístico inconsciente” en el caso univariado.

Por ejemplo,

$$E(X) = \sum_{(x,y) \in R_{XY}} x \cdot p_{XY}(x, y) = \sum_{x \in R_X} \sum_{y \in R_Y} x \cdot p_{XY}(x, y) = \sum_{x \in R_X} x \sum_{y \in R_Y} p_{XY}(x, y) =$$

Ahora bien, por la marginalización teníamos que $\sum_{y \in R_Y} p_{XY}(x, y) = p_X(x)$. Luego, nos queda que

$$E(X) = \sum_{x \in R_X} x \cdot p_X(x)$$

que es exactamente la fórmula de valor esperado que teníamos para el caso univariado, como habíamos visto antes. Esto nos dice que teniendo un vector aleatorio (X, Y), para calcular el valor esperado de alguna marginal, por ejemplo X, podemos recurrir a usar la fórmula de valor esperado partiendo desde el vector aleatorio o bien considerando a X de forma univariado. Es indistinto, y es sólo a conveniencia del practicante sobre qué usar y que no.

El siguiente es un concepto de crucial importancia cuando manejamos varias variables aleatorias:

Definition 9.2. *Dados (X, Y) vector aleatoria, definimos la* ***covarianza*** *entre X e Y como*

$$Cov(X, Y) = E((X - E(X))(Y - E(Y)).$$

Prontamente veremos la interpretación que tiene. Base decir que, en general, la covarianza tenderá a ser (más) positiva en tanto y en cuanto las dos variables tiendan a incrementarse de forma coordinada. Por ejemplo, es esperable que si X representa la altura e Y el peso de un individuo, exista una covarianza positiva entre ambas.

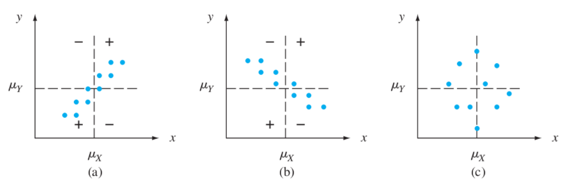


Figura 2: Los tres escenarios de la covarianza

La situación a) representa un caso de covarianza positiva, la b) de covarianza negativa y la c) de covarianza cero.

Un punto interesante, así como teníamos una forma ”fácil” de calcular la varianza, también nos ocurrirá con la covarianza:

Proposition 9.3. *$Cov(X, Y) = E(XY) - E(X)E(Y)$.*

Demostración.

$$Cov(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY - E(X)Y - XE(Y) + E(X)E(Y)) =$$

$$E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) = E(XY) - E(X)E(Y)$$

donde hemos usado que la esperanza es aditiva y saca constantes (por ejemplo, otras esperanzas) afuera.

□

Calculemos para el ejemplo que armamos de las bolitas la covarianza.

$$Cov(X, Y) = E(XY) - E(X)E(Y).$$

Obtengamos primero $E(X)$.

$$E(X) = 0 \cdot 84/220 + 1 \cdot 108/220 + 2 \cdot 27/220 + 3 \cdot 1/220 = 0{,}75.$$

Veamos $E(Y)$.

$$E(Y) = 0 \cdot 56/220 + 1 \cdot 112/220 + 2 \cdot 48/220 + 3 \cdot 4/220 = 1.$$

Luego, $E(X)E(Y) = 0{,}75 \cdot 1 = 0{,}75$.

Calculemos finalmente $E(XY)$.

$$E(XY) = 0 \cdot 0 \cdot 10/220 + 0 \cdot 1 \cdot 40/220 + 0 \cdot 2 \cdot 30/220 + 0 \cdot 3 \cdot 4/220+$$

$$1 \cdot 0 \cdot 30/220 + 1 \cdot 1 \cdot 60/220 + 1 \cdot 2 \cdot 18/220+$$
$$2 \cdot 0 \cdot 15/220 + 2 \cdot 1 \cdot 12/220+$$
$$3 \cdot 0 \cdot 1/220 =$$
$$0{,}55$$

Así pues, $Cov(X, Y) = 0{,}55 - 0 - 75 = -0{,}20$. La covarianza obtenida es negativa, indicando que en general X e Y no actúan de forma coordinada, sino al contrario. Es razonable, cuantas más bolitas rojas obtenga es menos probable obtener más bolitas blancas.

Veamos algunas propiedades adicionales que nos ofrece la covarianza:

Theorem 9.4. *Sea (X, Y) un vector aleatorio. Luego:*

1. $Cov(aX, Y) = aCov(X, Y)\ \forall a \in \mathbb{R}$.
2. $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$.
3. $Cov(X, Y) = Cov(Y, X)$.
4. $Cov(X, X) = Var(X)$.
5. $V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$.

Demostración. Son todas estas cuentas inmediatas de la aplicación de $Cov(X, Y) = E(XY) - E(X)E(Y)$. Es un buen ejercicio para el lector desarrollarlas. □

Seguimos introduciendo más conceptos que nos serán útiles. Teníamos de capítulos anteriores definida la idea de independencia entre eventos, es posible generalizar esto a independencia entre variables aleatorias. Por ejemplo, podemos preguntarnos si por ejemplo son independientes el peso y la altura de una persona. O cualquier otro par de factores que se recabe sobre un individuo, por ejemplo aspectos socio-económicos y preferencias de opinión electoral.

Recordemos que dos eventos A y B son independientes si $P(A \cap B) = P(A)P(B)$. Luego, lo natural sería decir que X e Y serán **variables aleatorias independientes** si

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)\ \forall A, B \subset \mathbb{R}.$$

En el caso discreto, no es difícil probar que esto se reduce a que

$$p_{XY}(x, y) = p_X(x)p_Y(y)\ \forall x, y \in \mathbb{R}.$$

También, a su vez, esto implicaría directamente que $p_{X|Y=y}(x) = p_X(x)$.

Un caso muy particular e interesante, que permite descartar independencia entre dos variables, se da con el cero. Notemos que si $p_{XY}(x, y) = 0$ (una celda de la tabla en donde hay un cero), luego si son independientes esto implica que o bien $p_X(x)$ es cero o bien $p_Y(y)$ lo es, pudiendo ambos ser cero. Acá estamos mirando los márgenes de la tabla.

Así pues, si tenemos dos márgenes que no son nulos que confluyen en una celda que es nula, las variables X e Y NO podrán ser independientes.

En nuestro caso de ejemplo, eso pasa con $p_{XY}(2, 2)$ que vale cero, y sin embargo ni $p_X(2)$ ni $p_Y(2)$ valen cero. Luego, en ese ejemplo X e Y no podrán ser independientes.

Una observación cuidadosa de la tabla nos dará el siguiente corolario:

Proposition 9.5. *Si las variables aleatorias discretas X e Y son independientes, luego $R_{XY} = R_X \times R_Y$. Es decir, el soporte del vector aleatorio (X, Y) deberá ser rectangular.*

Es importante notar que NO vale la vuelta, el soporte del vector aleatorio podría ser rectangular y, pese a eso, las dos variables ser dependientes. No es difícil armar contraejemplos diversos.

Finalmente, tenemos otro criterio sumamente interesante para descartar independencia:

Theorem 9.6. *Si las variables aleatorias discretas X e Y son independientes, luego $Cov(X, Y) = 0$.*

Nuevamente, no es cierto que valga la vuelta. Probemos este resultado.

Demostración. Supongamos X e Y independientes. Luego, queremos calcular $Cov(X, Y) = E(XY) - EX(X)E(Y)$. Veamos lo que pasa con $E(XY)$:

$$E(XY) = \sum_{x,y \in R_X \times R_Y} p_{XY}(x, y) = \sum_{x \in R_X} \sum_{y \in R_Y} p_X Y(x, y).$$

Pero como X e Y son independientes, luego $p_X Y(x, y) = p_X(x) p_Y(y)$. Entonces, nos queda

$$E(XY) = \sum_{x \in R_X} \sum_{y \in R_Y} p_X(x) p_Y(y) = \left(\sum_{x \in R_X} x \cdot p_X(x) \right) \left(\sum_{y \in R_Y} y \cdot p_Y(y) \right) = E(X)E(Y).$$

Así pues, bajo independencia tendremos que $E(XY) = E(X)E(Y)$, de donde se deduce inmediatamente que $Cov(X, Y) = E(XY) - E(X)E(Y) = 0$.

□

Teníamos además, de las propiedades de covarianza, que $V(X + Y) = V(X) + V(Y) - 2Cov(X, Y)$. Notemos que si X e Y son independientes, luego $Cov(X, Y) = 0$, con lo cual obtendremos la aditividad de la varianza: $V(X + Y) = V(X) + V(Y)$.

En nuestro caso, ya teníamos que $Cov(X, Y) = -0{,}20$ que no es cero, es otra forma (quizás más laboriosa) de probar que X e Y no son independientes.

Concluimos la exposición de esta sección con una última definición. Recordemos que teníamos la covarianza, como una medida de movimiento coordinado entre dos variables aleatorias. Sabemos que si es positivo, habla de una suerte de coordinación ”virtuosa” entre ambas variables. Si es negativa, la coordinación es ”viciosa” u opuesta.

Cuando es 0, diremos que X e Y están no correlacionadas, lo cual no quiere decir que sean independientes. Una de las dificultades que tiene la covarianza es arrastrar la unidad de las variables aleatorias. A fin de adimensionalizarla, se propone la siguiente forma de estandarizarla, conocida como la **correlación lineal**:

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{V(X)} \sqrt{V(Y)}}.$$

Daremos el siguiente listado de propiedades de la correlación, sin demostrarlas:

Theorem 9.7. *Sean X e Y variables aleatorias. Luego:*

1. $-1 \leq \rho_{XY} \leq 1$.
2. *X e Y independientes implica que $\rho_{XY} = 0$.*
3. *$\rho_{XY} = 1$ si y sólo si existen $a > 0$ y $b \in \mathbb{R}$ tal que $Y = aX + b$ (existe una correlación lineal perfecta y positiva entre las dos variables).*
4. *$\rho_{XY} = -1$ si y sólo si existen $a < 0$ y $b \in \mathbb{R}$ tal que $Y = aX + b$ (existe una correlación lineal perfecta y negativa entre las dos variables).*

Volviendo a nuestro ejemplo ya canónico, obtengamos el coeficiente de correlación. Ya sabíamos que $Cov(X,Y) = -0{,}20$, con lo cual la correlación deberá ser necesariamente negativa. Obtengamos las varianzas de cada variable.

$$V(X) = E(X^2) - E^2(X) = 0^2 \cdot 84/220 + 1^2 \cdot 108/220 + 2^2 \cdot 27/220 + 3^2 \cdot 1/220 - (0{,}75)^2 = 0{,}46$$

$$V(Y) = E(Y^2) - E^2(Y) = 0^2 \cdot 56/220 + 1^2 \cdot 112/220 + 2^2 \cdot 48/220 + 3^2 \cdot 4/220 - 1^2 = 0{,}55$$

Luego,

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} \frac{-0{,}20}{\sqrt{0{,}46}\sqrt{0{,}55}} = -0{,}40.$$

9.3. Caso Continuo

Procederemos a realizar una traducción al caso continuo de todos los conceptos vistos en el caso discreto, siguiendo a partir de un ejemplo.

Supongamos que tenemos (X,Y) un vector aleatorio continuo. Tendremos entonces ahora una densidad continua $f_{XY}(x,y)$, con estas propiedades:

Theorem 9.8. *1.* $f_{XY}(x,y) \geq 0$.

2. $\int_{\mathbb{R}^2} f_{XY}(x,y)dxdy = 1$.

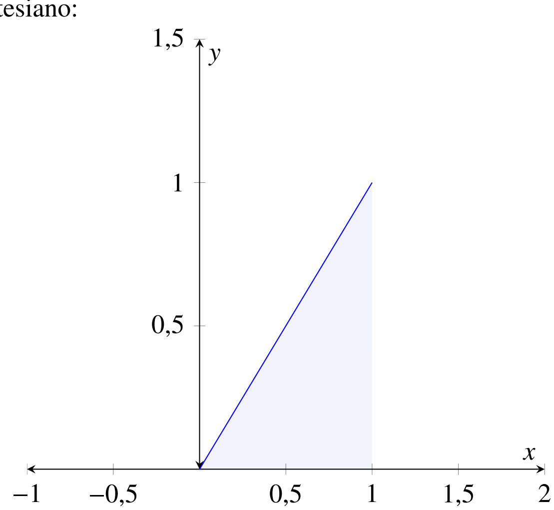
3. (Marginalización en x) $f_X(x) = \int_{\mathbb{R}} f_{XY}(x,y)dy$.

4. (Marginalización en y) $f_Y(y) = \int_{\mathbb{R}} f_{XY}(x,y)dx$.

Veamos un ejemplo sencillo en primer lugar. Tenemos la siguiente densidad conjunta:

$$f_{XY}(x,y) = 2 \cdot I_{(0,1)}(x) I_{(0,x}(y).$$

Nótese la doble indicadora, sobre ambas variables, denotando el **soporte**, región donde la densidad es positiva. Son valores de x entre 0 y 1, y valores de y entre 0 y la recta $y = x$. Grafiquemos esta región en el plano cartesiano:



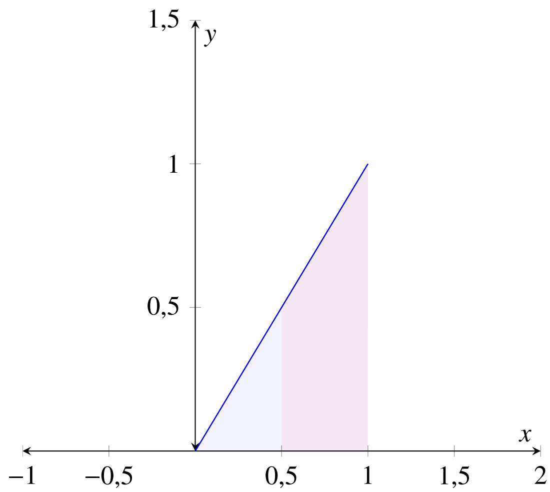
Es altamente recomendable hacer este gráfico para entender la situación de trabajo, en especial en escenarios potencialmente más complejos.

Verifiquemos primero que es efectivamente una densidad, viendo que integra 1 (es claro que es no negativo).

$$\int_{\mathbb{R}^2} f_{XY}(x,y)dxdy = \int_{\mathbb{R}^2} 2 \cdot I_{(0,1)}(x)I_{(0,x)}(y)dxdy = \int_0^1 \int_0^x 2dydx =$$

$$\int_0^1 2xdx = 1.$$

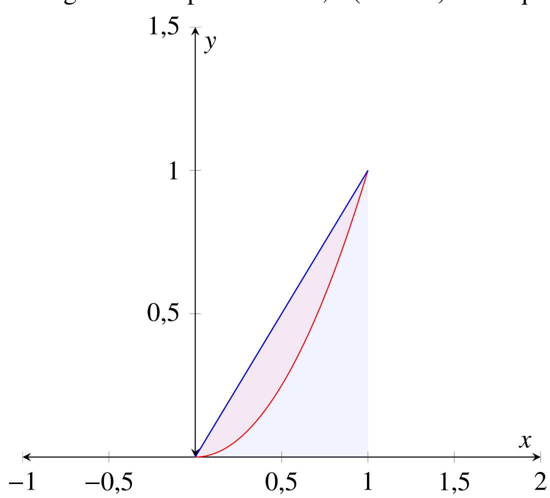
Como primer ejercicio, mirando el gráfico calculemos $P(X > 0{,}5)$. Deberemos recurrir a una integral doble de la siguiente región en rojo:



Bien, esto se puede obtener como

$$P(X > 0{,}5) = \int_{0{,}5}^1 \int_0^x 2dydx = \int_{0{,}5}^1 2xdx = 1^2 - (0{,}5)^2 = 0{,}75.$$

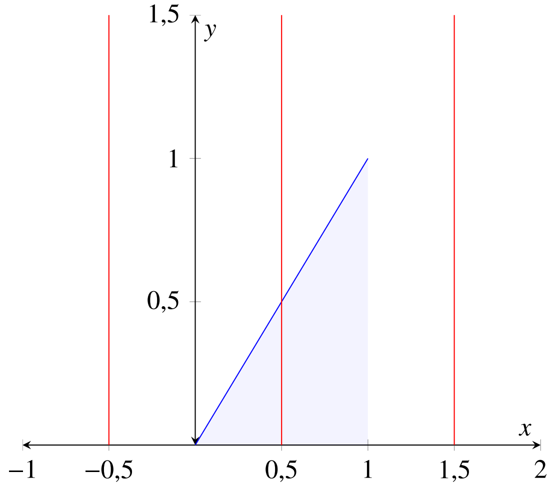
Obtengamos otra probabilidad, $P(Y > X^2)$. Grafiquemos la región de interés:



Calculamos para obtener:

$$P(Y > X^2) = \int_0^1 \int_{x^2}^{x} 2\, dydx = \int_0^1 2x - 2x^2 dx = 1 - 2/3 = 1/3.$$

Procedamos ahora a calcular las densidades marginales. Para estas cosas siempre conviene tener presente el gráfico del soporte. Por ejemplo, si estamos considerando calcular f_X tendremos que integrar a f_{XY} con respecto a y. Veamos los tres escenarios que se nos presentan:



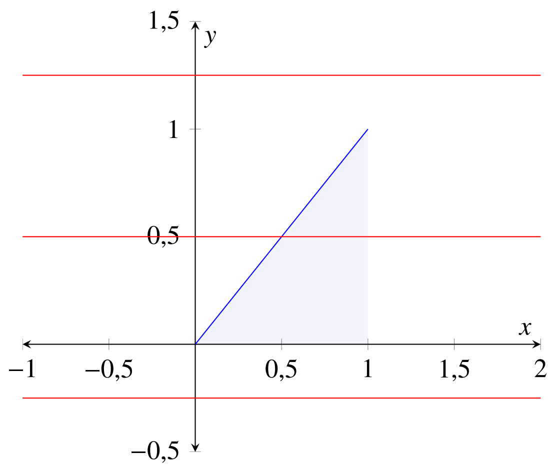
Mirando fijamente las tres rectas, vemos que si x es menor que cero, entonces $f_X(x)$ se obtiene marginalizando sobre una región fuera del soporte del vector aleatorio, con lo cual la densidad será cero. Lo mismo pasa si x es mayor que 1. En $0 < x < 1$, tendremos que

$$f_X(x) = \int_0^x 2\, dxdy = 2x.$$

Así pues, tenemos que en general

$$f_X(x) = 2x \cdot I_{(0,1)}(x).$$

Veamos ahora la marginal en y. Tenemos ahora:



Nuevamente, como antes, tenemos que marginalizamos por fuera del soporte en el caso en que y sea menor que 0 o mayor que 1. En otro caso, entre 0 y 1, tendremos que

$$f_Y(y) = \int_y^1 2 \cdot dxdy = 2 - 2y.$$

Luego, en general, tenemos que

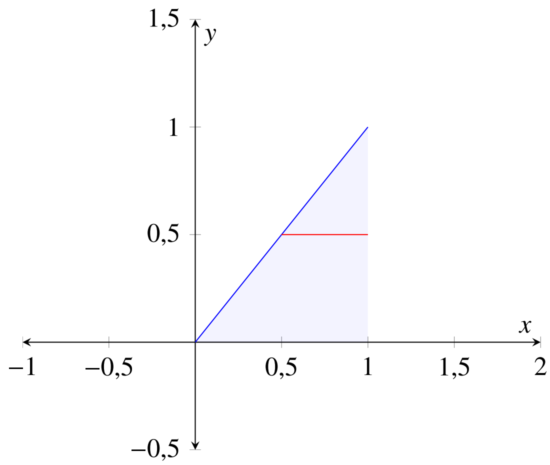
$$f_Y(y) = (2 - 2y) \cdot I_{(0,1)}(y).$$

Así como en el caso discreto teníamos definida la noción de probabilidad puntual condicional $p_{X|Y=y}$, idénticamente tendremos pues la **densidad condicional**:

$$f_{X|Y=y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

definido siempre y cuando $f_Y(y) > 0$.

Por ejemplo, obtengamos $f_{X|Y=0,5}$. Es decir, es una densidad condicional que opera sobre



Así pues,

$$f_{X|Y=0,5}(x) = \frac{f_{XY}(x,0,5)}{f_Y(0,5)} = \frac{2}{2 - 2 \cdot 0,5} I_{(0,5,1)}(x) = 2 \cdot I_{(0,5,1)}(x)$$

Extrapolemos ahora las definiciones de esperanza. Si $h : \mathbb{R}^2 \to \mathbb{R}$, luego

$$E(h(X,Y)) = \int_{\mathbb{R}^2} h(x,y) f_{XY}(x,y) dxdy.$$

Análogamente, tenemos que

$$Cov(X,Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y),$$

sin cambio alguno con respecto al caso discreto. Obtengamos por ejemplo la covarianza para esta distribución. Teniendo expresiones para f_X y para f_Y, es fácil calcular, como hacíamos en el caso univariado, los valores esperados de ambas variables:

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx = \int_0^1 x \cdot 2x dx = \int_0^1 2x^2 dx = 2/3.$$

$$E(Y) = \int_{-\infty}^{+\infty} y \cdot f_Y(y) dy = \int_0^1 y \cdot (2 - 2y) dy = \int_0^1 2y - 2y^2 dy = 1 - 2/3 = 1/3.$$

Ahora, calculemos $E(XY)$:

$$E(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy \cdot f_{XY}(x,y) dxdy = \int_0^1 \int_0^x xy \cdot 2dxdy = \int_0^1 x^3 dx = 1/4 = 0{,}25.$$

Luego,

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 0{,}5 - 2/3 \cdot 1/3 = 0{,}25 - 2/9 = 0{,}028.$$

De la misma forma, podemos calcular la correlación ρ_{XY}. Ya tenemos el numerador, que es la covarianza, nos falta $V(X)$ y $V(Y)$. Para eso, necesitaremos respectivamente $E(X^2)$ y $E(Y^2)$, que valen:

$$E(X^2) = \int_0^1 x^2 2x dx = 2 \int_0^1 x^3 dx = 0{,}5.$$

$$E(Y^2) = \int_0^1 y^2(2-2y)dy = 2 \int_0^1 y^2 - y^3 dy = 2(1/3 - 1/4) = 0{,}17.$$

Entonces,

$$V(X) = E(X^2) - E^2(X) = 0{,}5 - (2/3)^2 = 0{,}5 - 4/9 = 0{,}06$$

y

$$V(Y) = E(Y^2) - E^2(Y) = 0{,}17 - (1/3)^2 = 0{,}17 - 1/9 = 0{,}06$$

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{0{,}028}{\sqrt{0{,}06}\sqrt{0{,}06}} = \frac{0{,}028}{0{,}06} = 0{,}46.$$

Hay una cierta correlación positiva, indicando una tendencia a un comportamiento monótono creciente entre las variables X e Y. Es decir, a mayor valor de X es probable que tengamos un mayor valor de Y.

La definición de independencia en el caso continuo es más delicado si se quiere que en el caso discreto. Lo que haremos es dar una breve exposición sobre ese tema de forma simplificada, aunque cabe aclarar que hay aspectos formales que serán soslayados en aras de no dificultar y hacer empalagoso la explicación, apuntada a transmitir las ideas fundamentales.

Diremos que X e Y variables aleatorias continuas serán independientes si $f_{XY}(x,y) = f_X(x)f_Y(y)$ para todo par x, y que sean puntos de continuidad de las tres funciones. Esta definición no es perfecta para los fines más formales, pero sirve para los casos que analizaremos.

Un resultado bastante útil que permite descartar independencia es el siguiente, que daremos sin demostración:

Theorem 9.9. *Si X e Y son variables aleatorias continuas independientes, entonces el soporte del vector aleatoria necesariamente deberá ser rectangular.*

La idea de esto es que si no es rectangular, podré encontrar puntos (x,y) en donde $f_{XY}(x,y)$ es igual a cero y sin embargo ni $f_X(x)$ ni $f_Y(y)$ lo son, lo cual invalidaría la definición de independencia.

En nuestro ejemplo que venimos desarrollando, tendremos claramente que X e Y no son independientes pues el soporte es un triángulo, no es por lo tanto rectangular y, por ende, se descarta la independencia.

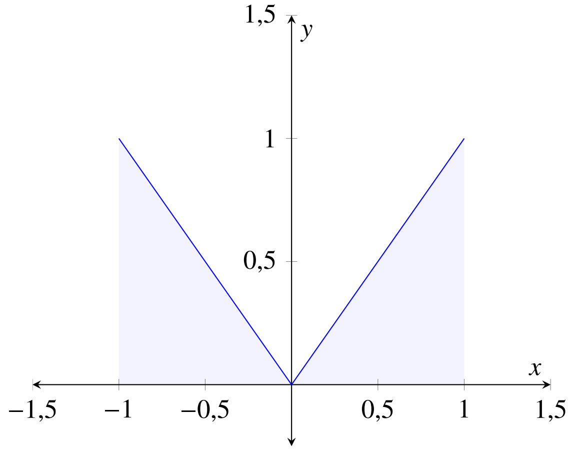
9.3.1. Otro ejemplo

Consideremos un vector aleatorio continuo (X, Y) con densidad:

$$f_{XY}(x, y) = cx^2 y \cdot I_{(-1,1)}(x) I_{0,|x|}(y)$$

siendo $c > 0$ desconocida.

Primero, vamos a graficar el soporte de esta densidad. Notemos que mientras x se mueve entre -1 y 1, y deberá moverse entre 0 y el módulo de x. No es difícil entonces identificar el siguiente soporte:



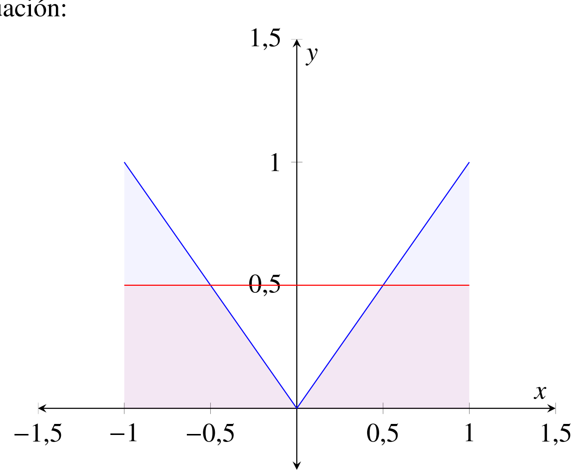
Veamos ahora la constante c. Sabemos que

$$1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y)\, dxdy = \int_{-1}^{0} \int_{0}^{-x} cx^2 y\, dydx + \int_{0}^{1} \int_{0}^{x} cx^2 y\, dydx$$

$$= c \int_{-1}^{0} \frac{x^4}{2} dx + c \int_{0}^{1} \frac{x^4}{2} dx = c(2/5 + 2/5) = c4/5 = 1$$

de donde se deduce que $c = 5/4$.

Veamos ahora como ejemplo el cómputo de $P(Y < 1/2)$. Grafiquemos primero la región para entender la situación:



Podemos calcular esto como sigue:

$$P(Y < 0{,}5) = \int_{0}^{0{,}5} dy \left(\int_{-1}^{-y} cx^2 y\, dx + \int_{y}^{1} cx^2 y\, dx \right)$$

$$= \int_0^{0,5} dy\left(\frac{-cy^4}{3} + \frac{cy}{3} + \frac{cy}{3} - \frac{cy^4}{3}\right)$$

$$= (2/3)c\int_0^{0,5}(y - y^4)dy = (2/3)c(0{,}125 - 0{,}00625) == (2/3)(5/4)0{,}11875 = 0{,}099.$$

Calculemos las distribuciones marginales. Vamos a empezar con f_X. Es claro que si $|X| > 1$ luego la densidad marginal será cero pues integraremos en regiones que están fuera del soporte. Para $|x| \leq 1$ tendremos:

$$f_X(x) = \int_0^{|x|} cx^2 y\, dy = cx^4.$$

Luego,

$$f_X(x) = cx^4 I_{(-1,1)}(x).$$

Veamos ahora f_Y, notemos que si $y < 0$ o bien $y > 1$ la densidad marginal será cero pues caeremos fuera del soporte. Entonces, si $0 < y < 1$ tendremos:

$$f_Y(y) = \int_{-1}^{-y} cx^2 y\, dx + \int_y^1 cx^2 y\, dx = \frac{2cy}{3} - 2\frac{cy^3}{3}.$$

Luego,

$$f_Y(y) = \left(\frac{2cy}{3} - 2\frac{cy^3}{3}\right) I_{(0,1)}(y).$$

Finalmente, veamos si tenemos una situación de independencia. Varias maneras tenemos de calcular esto, por ejemplo obteniendo la covarianza y viendo si es distinta de cero (ojo que si llega a ser cero eso no implica independencia!). Lo más sencillo es notar que el soporte no es rectangular, condición necesaria (pero nos suficiente!) para la independencia.

Por ende, X e Y no son variables independientes.

9.4. Distribución Multinomial

Así como teníamos, en el caso de familias de variables aleatorias discretas, el caso de la binomial, generalizaremos el mismo para la situación en donde en vez de tener dos clases (”Éxito” o ”Fracaso”), ahora tenemos $K > 1$ clases. La clase $1 \leq k \leq K$ tiene probabilidad $0 < p_k < 1$ de ocurrir en cada ensayo. Así pues, definimos el vector aleatorio $(X_1, \ldots, X_K)$ de forma tal que X_k =”cantidad de veces que salió la clase k en los n ensayos”.

Por ejemplo, si tenemos una urna con 10 bolitas rojas, 8 azules y 5 verdes, podemos considerar un experimento que sea extraer 12 bolitas con reposición y luego contar las bolitas obtenidas de cada color. Es decir, tendremos un vector aleatorio (X_R, X_A, X_V) de forma tal que conjuntamente tendrá una distribución **multinomial** de n ensayos con $p_R = 10/23$, $p_A = 8/23$ y $p_V = 5/23$. Lo notaremos como

$$(X_R, X_A, X_V) \sim Mu(n, p_R, p_A, p_V).$$

En general, tendremos que $(X_1, \ldots, X_K \sim Mu(n, p_1, \ldots, p_K)$ con $0 \leq p_k \leq 1$ y $\sum_{k=1}^K p_k = 1$. Vamos a efectuar un esbozo de cálculo de la probabilidad puntual $p_{X_1,\ldots,X_K}(x_1, \ldots, x_K)$. En primer lugar, cada x_i deberá ser entero no negativo y de forma tal que $\sum_{k=1}^K x_k = n$, o de lo contrario la probabilidad será cero. La suma de las marginales X_k debe dar el total de ensayos que se ejecutaron.

Por otro lado, tenemos que si tenemos una realización dada $(x_1, \ldots, x_k)$, entonces podemos pensar en que tenemos n celdillas, de forma tal que en cada celda puede haber una entre K etiquetas, y la suma de las etiquetas de tipo k son x_k.

Luego, primero elegimos las celdillas que corresponden a la etiqueta 1, que serán $\binom{n}{x_1}$. Luego, elegimos sobre las restantes $n - x_1$, x_2 celdillas para la etiqueta x_2, o sea que multiplicamos por $\binom{n-x_1}{x_2}$. Así seguimos hasta el último, que será $\binom{n-\sum_{k=1}^{K-1} x_i}{k_k} = \binom{x_K}{x_K} = 1$. Queda como ejercicio para el lector verificar que

$$\binom{n}{x_1}\binom{n-x_1}{x_2}\binom{n-x_1-x_2}{x_3}\dots\binom{n-x_1-x_2-\dots-x_{K-1}}{x_k} = \frac{n!}{x_1!x_2!\dots X_K!}.$$

Por último, teniendo una tira de n celdillas fijas con x_1 celdas con primer etiqueta con probabilidad p_1, x_2 con etiquetas 2 con probabilidad p_2 y así, por independencia tendremos que su probabilidad será $p_1^{x_1}\dots p_K^{x_K}$.

Así pues, finalmente, la probabilidad puntual buscada será:

$$p_{(X_1,\dots,X_K)}(x_1,\dots,x_K) = \frac{n!}{x_1!\dots x_K!}p_1^{x_1}\dots p_K^{x_K}.$$

En nuestro ejemplo de la urna, con las 12 extracciones podríamos calcular la probabilidad de $p_{X_R,X_A,X_V}(3,5,4) = \frac{12!}{3!5!4!}(10/23)^3(8/23)^5(5/23)^4$.

Un resultado interesante, que sale de la descripción conceptual del experimento, es que si tenemos $(X_1,\dots,X_K) \sim Mu(n,p_1,\dots,p_K)$, luego si tomamos una de las marginales X_k, esto representa la probabilidad de obtener X_k la etiqueta k. Si pensamos en un nuevo escenario en donde consideramos "Éxito" a obtener la etiqueta k y "Fracaso" a no obtenerla, luego en definitiva nos quedará que X_k será una binomial $Bi(n,p_k)$.

Por ejemplo, en el caso de las bolillas tendremos que X_R será una binomial de parámetro n y $p = 10/23$.

9.5. Sumas de Variables Aleatorias: Convoluciones

En esta sección haremos un breve camino para analizar como analizar sumas de variables aleatorias independientes. Es decir, tenemos X e Y variables aleatorias independientes, queremos calcular la distribución de $U = X+Y$.

Vamos a empezar con el caso discreto, supongamos entonces que X e Y son dos variables aleatorias discretas e independientes, con probabilidad puntual p_X y p_Y respectivamente. Queremos obtener una expresión para p_U. Es decir:

$$p_U(u) = P(U=u) = P(X+Y=u).$$

El "truco", para analizar esta expresión, es usar probabilidad total sobre alguna de las dos variables, por ejemplo la Y. Consideremos todos sus posibles valores, dados por R_Y, y tendremos así que

$$P(X+Y=u) = \sum_{y\in R_Y} P(X+Y=u\cap Y=y) = \sum_{y\in R_Y} P(X+y=u\cap Y=y) = \sum_{y\in R_Y} P(X=u-y\cap Y=y) = \sum_{y\in R_Y} P(X=u-y)P(Y=y)$$

donde usamos aquí que X e Y. Luego, esto nos queda

$$p_U(u) = \sum_{y\in R_Y} p_X(u-y)p_Y(y),$$

que es lo que se conoce como una **suma convolucional** entre p_X y p_Y.

Vamos a aplicar esto a un ejemplo concreto, supongamos que tenemos $X \sim \mathcal{P}(\lambda_1)$ e $Y \sim \mathcal{P}(\lambda_2)$. Es decir, dos distribuciones de Poisson. Queremos obtener la distribución de $U = X+Y$. Recordemos que $R_X = R_Y = \mathbb{N}_0$, luego

$$p_U(u) = \sum_{y=0}^{+\infty} p_X(u-y)p_Y(y).$$

Pero si $u-y<0$ luego $p_X(u-y)$ daría cero, entonces en realidad deberíamos tener:

$$p_U(u) = \sum_{y=0}^{u} p_X(u-y)p_Y(y).$$

Es decir,

$$p_U(u) = \sum_{y=0}^{u} \frac{e^{-\lambda_1}\lambda_1^{(u-y)}}{(u-y)!}\frac{e^{-\lambda_2}\lambda_2^{y}}{y!} = e^{-(\lambda_1+\lambda_2)}\sum_{y=0}^{u}\frac{\lambda_1^{(u-y)}\lambda_2^{y}}{(u-y)!y!} = \frac{e^{-(\lambda_1+\lambda_2)}}{u!}\sum_{y=0}^{u}\binom{u}{y}\lambda_1^{(u-y)}\lambda_2^{y}$$

donde el último paso se obtuvo multiplicando y dividiendo por $u!$.

El teorema de la expansión binomial afirma que $(a+b)^n = \sum_{i=0}^{n}\binom{n}{i}a^i b^{n-i}$. Aplicándolo, tendremos que

$$p_U(u) = \frac{e^{-(\lambda_1+\lambda_2)}}{u!}(\lambda_1+\lambda_2)^u$$

que es la expresión de $\mathcal{P})(\lambda_1+\lambda_2)$.

Por ejemplo, si en una determinada ciudad en un día dado llegan aviones siguiendo una variable de Poisson con valor esperado de 5 aviones por día, y los barcos llegan siguiendo una variable de Poisson con valor esperado de 3 barcos por día, luego el total de vehículos que llegan al día será según una variable de Poisson de valor esperado (λ) de 5+3 = 8 vehículos por día.

Supongamos ahora que X e Y son dos variables aleatorias continuas independientes. Veamos la distribución de $U = X + Y$. El truco será recurrir a F_U, que es una probabilidad a diferencia de f_U.

Entonces,

$$F_U(u) = P(U \leq u) = P(X+Y \leq u) = P(Y \leq u - X).$$

Debemos trazar en el plano la recta $Y = y - X$ e integrar en el semiplano inferior. Es decir,

$$\begin{aligned} F_U(u) &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{u-x} f_{XY}(x,y)dy = \int_{-\infty}^{\infty} dx \int_{-\infty}^{u-x} f_X(x)f_Y(y)dy \\ &= \int_{-\infty}^{\infty} f_X(x)F_Y(u-x)dx \end{aligned}$$

donde hemos usado que $F_Y(-\infty) \to 0$. Entonces, para obtener la densidad debemos derivar esto con respecto a u, que al integrar dentro del signo integral nos llega a

$$f_U(u) = F'_U(u) = \int_{-\infty}^{\infty} f_X(x)f_Y(u-x)dx.$$

Este resultado es interesante, y recibe el nombre de **convolución** y tiene una notación especial. En efecto, notamos como $f_U = f_X * f_Y$ siendo $*$ el operador de convolución.

Hagamos un ejemplo sencillo, supongamos que $X \sim U(0,1)$ e $Y \sim U(0,1)$ independientes. Veamos la distribución de $U = X + Y$. Según lo que hemos visto, luego $f_U = f_X * f_Y$. Es decir,

$$f_U(u) = \int_{-\infty}^{\infty} f_X(x)f_Y(u-x)dx$$

Ahora bien, solamente tiene sentido integrad en x cuando este se encuentra entre 0 y 1 (o de lo contrario $f_X(x)$ será 0) o bien cuando $u - x$ esté entre 0 y 1 (o de lo contrario $f_Y(u-x)$ será 0).

Es fácil verificar que U puede tomar valores entre 0 y 2, puesto que tanto X como U lo hacen entre 0 y 1 y además $U = X + Y$. Entonces, si u es menor que 0 o mayor que 2, luego $f_U(u) = 0$.

Tenemos las condiciones:

$$0 \leq x \leq 1$$

$$0 \leq u - x \leq 1$$

Es decir, las condiciones son

$$0 \leq x \leq 1$$

$$x \leq u$$

$$x \geq u - 1$$

Juntando todas estas condiciones, tenemos que

$$max(0, u - 1) \leq x \leq min(1, u)$$

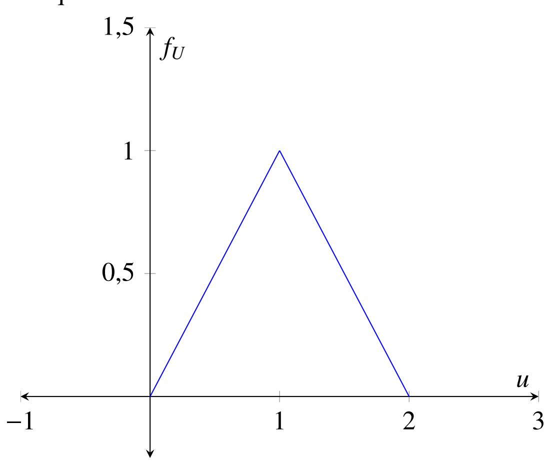
Suponiendo que se cumplan ambas cuestiones, luego $f_X(x) = 1$ y $f_Y(y) = 1$, por ser ambas $U(0, 1)$. Luego,

$$f_U(u) = \int_{max(0,u-1)}^{min(1,u)} 1\ dx = min(1, u) - max(0, u - 1).$$

Partiendo en casos, si $0 < u < 1$, luego $min(1, u) = u$ y $max(0, u - 1) = 0$, con lo cual $f_U(u) = u - 0 = u$. Caso $1 \leq 1 < 2$, entonces $min(1, u) = 1$ y $max(0, u - 1) = u - 1$, con lo cual $f_U(u) = 2 - u$. Así pues, tenemos que

$$f_U(u) = \begin{cases} 0 & u \leq 0 \\ u & 0 < u \leq 1 \\ 2 - u & 1 < u \leq 2 \\ 0 & u > 2 \end{cases}$$

Grafiquemos esta densidad:



Notemos que la suma de dos uniformes ya no es más uniforme, pasa a ser una distribución más concentrada en el valor medio, que es 1. Los números cercanos a 1 se vuelven más probables que los extremos para la variable aleatoria $U = X + Y$.

9.6. Distribuciones Jerárquicas

Esto es un intento de formalizar algo que probablemente ya nos hemos topado antes en ejercicios y ejemplos. La idea de esto es considerar una variable aleatoria cuya distribución tiene parámetros que dependen, a su vez, de otra distribución.

Vamos a ilustrar con un ejemplo. Supongamos que tenemos una tienda en donde la cantidad de clientes que llegan en una hora sigue un proceso de Poisson de tasa λ clientes por hora. Ahora bien, cada cliente que llega tiene una probabilidad $0 < p < 1$ de comprar o no un producto. Llamemos X a la variable aleatoria que modela la cantidad de ventas realizadas en una hora, e Y la cantidad de clientes que llegan a la tienda.

Entonces, tendremos que

$$X|Y = y \sim Bi(y, p)$$
$$Y \sim \mathcal{P}(\lambda).$$

Nos interesa obtener la distribución de X en este caso, la cantidad de ventas concretadas. No siempre resulta fácil obtener una buena expresión para esto, pero en este caso resultará en efecto posible, apelando a cuentas a las cuales ya estamos acostumbrados a recurrir:

$$p_X(x) = \sum_{y=x}^{\infty} p_{XY}(x, y) = \sum_{y=x}^{\infty} p_{X|Y=y}(x) p_Y(y)$$

Es decir, hacemos probabilidad total sobre todos los posibles valores de y, pero considerando solamente aquellos en donde $p_{XY}(x, y)$ no sea cero, que serían para los y mayores o iguales que x, no tiene sentido que vengan menos clientes que las ventas realizadas.

Bien, $p_Y(y) = \frac{e^{-\lambda}}{\lambda}^u y!$, y $p_{X|Y=y}(x) = \binom{y}{x} p^x (1-p)^{y-x}$. Luego pues, nos queda que

$$\begin{aligned}
p_X(x) &= \sum_{y=x}^{\infty} \binom{y}{x} p^x (1-p)^{y-x} \frac{e^{-\lambda}\lambda^y}{y!} \\
&= \sum_{y=x}^{\infty} \frac{y!}{x!(y-x)!} p^x (1-p)^{n-x} \frac{e^{-\lambda}\lambda^{x+(y-x)}}{y!} \\
&= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{y=x}^{\infty} \frac{((1-p)\lambda)^{y-x}}{(y-x)!}.
\end{aligned}$$

Efectuamos la sustitución $y = y - x$, con lo cual nos queda:

$$p_X(x) = \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{t=0}^{\infty} \frac{((1-p)\lambda)^t}{t!}$$

Estamos sumando sobre una probabilidad puntual de una Poisson de parámetro $(1-p)\lambda$, con lo cual toda esa suma debería dar 1. Entonces, finalmente nos queda:

$$p_X(x) = \frac{(\lambda p)^x e^{-\lambda}}{x!},$$

que es exactamente la densidad de una Poisson de parámetro $p\lambda$. Es decir, la cantidad de ventas concretadas en la tienda tendrá un proceso de Poisson de parámetro $p\lambda$ ventas por hora.

9.7. Estadísticos de Orden

Se entiende como **muestra aleatoria** a una colección de variables $X_1, \ldots, X_n$ independientes y todas ellas con la misma distribución. Es decir, todas tienen la misma densidad (si fuesen por ejemplo continuas) o probabilidad puntual (en el caso discreto) y son además independientes. Supongamos que las quisiéramos tener ordenadas. Es decir, permutamos a la muestra de forma que queden ordenadas de manera ascendente

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}.$$

Cuando indexamos a las variables aleatorias con un (i) estamos indicando que son estadísticos de orden. Por ejemplo, si tenemos $X_1 = 4, X_2 = 1, X_3 = 9, X_4 = 3$, luego $X_{(1)} = 1$, $X_{(2)} = 3$, $X_{(3)} = 4$ y $X_{(4)} = 9$.

Preguntas que nos podemos hacer: ¿qué distribución tienen los estadísticos de orden? Trataremos de responder esas preguntas. Lo haremos en un caso simplificado, suponiendo que las variables tienen distribución continua. Eso impide que hayan empates en las variables aleatorias. En efecto, por la continuidad se puede probar que la probabilidad de que exista un par $i \neq j$ tal que $P(X_i = X_j)$ vale 0.

9.7.1. Distribución del máximo

Tomemos $U = \max\{X_1, \ldots, X_n\} = X_{(n)}$ el máximo, veamos su distribución.
Para eso, trabajaremos primero con F_U, la acumulada de U. Obtengamos una expresión para esto:

$$F_U(u) = P(U \leq u) = P(\max_i X_i \leq u).$$

Ahora bien, si el máximo de n numeros es menor o igual que u, es porque todos ellos son menores o iguales que u. Con lo cual

$$F_U(u) = P(X_1 \leq u \cap \ldots \cap X_n \leq u) = P(X_1 \leq u) \ldots P(X_n \leq u) = F_X(u) \ldots F_X(u) = F_X^n(u).$$

Tenemos así una expresión para la acumulada del máximo. Veamos su densidad:

$$f_U(u) = F_U(u)' = (F_X^n(u))' = nF_X^{n-1}(u) f_X(u).$$

Veamos un ejemplo. Supongamos una muestra aleatoria de n variables tales que todas ellas tienen distribución $U(0, 1)$. Tenemos que $f_X(x) = I_{(0,1)}(x)$ y que

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

Luego, la acumulada del máximo nos queda

$$F_U(u) = \begin{cases} 0 & u < 0 \\ u^n & 0 \leq u < 1 \\ 1 & u \geq 1 \end{cases}$$

y la densidad

$$f_U(u) = nu^{n-1} I_{(0,1)}(u).$$

Su valor esperado será

$$E(U) = \int_0^1 unu^{n-1} du = n \int_0^1 u^n du = \frac{n}{n+1}.$$

Es interesante el hecho de que a medida que el n se va volviendo más grande, el valor esperado del máximo va tendiendo, por debajo, a 1, lo cual es razonable. Cuantos más elementos tenga, más cerca tenderemos a estar en valor máximo al borde superior del soporte.

9.7.2. Distribución del mínimo

Definamos ahora $V = \min\{X_1, \ldots, X_n\} = X_{(1)}$. Veamos su distribución acumulada primero.

$$F_V(v) = P(V \leq v).$$

El mínimo no se lleva muy bien con una relación de menor o igual, por lo cual apelamos al truco ya conocido de complementar para tener así

$$F_V(v) = 1 - P(V \geq v) = 1 - P(\min\{X_1, \ldots, X_n\} \geq v) = 1 - P(X_1 \geq v, \ldots, X_n \geq v).$$

Por independencia, tendremos

$$F_V(v) = 1 - P(X_1 \geq v) \dots P(X_n \geq v) = 1 - (1 - F_X(v))^n.$$

La densidad será entonces

$$f_V(v) = n(1 - F_X(v))^{n-1} f_X(v).$$

9.7.3. Aplicación

Supongamos que tenemos n componentes en un circuito en serie, y para que funcione la totalidad del circuito deben estar todas las componentes. Sabemos que cada componente tiene una duración en meses dada por una exponencial de parámetro $\lambda > 0$, y que son independientes entre sí.

Obtengamos la distribución de la duración del circuito en sí. Sea V la duración del circuito, tenemos que efectivamente la duración de V estará dada por la duración mínima de los componentes. En la medida que todas las componentes estén andando supongamos por al menos una hora, luego todo el circuito lo hará por al menos una hora. Ahora bien, si una componente se quema, el circuito dejará de andar.

Entonces pues, la densidad de la duración del circuito se corresponde con la del mínimo, luego su densidad será (para $v > 0$, pues sino será cero)

$$f_V(v) = n(1 - F_X(v))^{n-1} f_X(v) = n(1 - (1 - e^{\lambda v}))^{n-1} \lambda e^{-\lambda v} = n\lambda e^{-n\lambda v}$$

que se corresponde exactamente con una distribución de una exponencial de parámetro $n\lambda$.

9.8. Ejercitación

1. Se arrojan dos dados de seis caras equilibradas.
   a) Sea X el valor más grande obtenido en el dado e Y es la suma de ambos dados. Hallar p_X, p_Y y p_{XY}.
   b) Calcular $Cov(X, Y)$ y ρ_{XY}. Son X e Y independientes?
2. Una urna contiene 5 transistores, de los cuales 2 son defectuosos. Se extrae sin reposición de la urna y se va probando cada transistor hasta tener identificados los dos transistores defectuosos. Sea N_1 la cantidad de inspecciones hasta encontrar el primer transistor defectuoso y sea N_2 la cantidad de inspecciones hasta hallar el segundo defectuoso. Encontrar p_{N_1,N_2} y hallar el valor esperado del total de transistores que se deberán revisar $(N_1 + N_2)$.
3. Sea (X, Y) un vector aleatorio continuo con la siguiente densidad conjunta:

   $$f_{XY}(x, y) = c(x + y)I_{(0,3/4)}(x)I_{(x^2,x)}(y)$$

   con $c > 0$ desconocido.
   a) Graficar el soporte.
   b) Determinar el valor de c.
   c) Calcular $P(X < 1/2)$ y $P(Y < 1/2)$.
   d) Obtener las expresiones de f_X y f_Y.
   e) Son independientes X e Y?
4. Se tiene una diana consistente en una circunferencia de radio R.
   a) Se arroja un dardo que puede caer en cualquier lado de la diana, obtenga la probabilidad de que esté a menos de R/2 del centro de la diana.

b) Sea D la distancia al origen. Es decir, $D = \sqrt{X^2 + Y^2}$. Obtenga $F_D(d) = P(D \leq d) = P(\sqrt{X^2 + Y^2} \leq d)$ y f_D.

5. Sea (X, Y) un vector aleatorio continuo con densidad

$$f_{(X,Y)} = (x/5 + cy)I_{(0,1)}(x)I_{(1,5)}(y).$$

   a) Halle el valor de c.

   b) Son X e Y independientes?

   c) Obtener $P(X + Y > 3)$.

6. Demuestre que $f_{XY}(x, y) = 1/xI_{(0,1}(x)I_{(0,x)}(y)$ es una densidad. Grafique su soporte. Halle $E(X)$ y $E(Y)$.

7. Se tiene una vara de longitud $R > 0$ conocido. Se eligen dos puntos totalmente al azar en esta vara, obtenga el valor esperado de distancia entre ambos puntos.

8. Suponga que X_1 y X_2 son ambas variables aleatorias independientes exponenciales, con parámetro λ_1 y λ_2 respectivamente. Calcule $P(X_1 < X_2)$.

9. Augusto y César deben juntarse, ambos llegan entre las 2PM y las 3PM de forma aleatoria siguiendo reglas determinadas. La densidad de probabilidad de la llegada de Augusto sigue una distribución aleatoria continua con densidad

$$f_X(x) = 3x^2 I_{(0,1)}(x)$$

   mientras que la densidad de probabilidad de la hora de llegada de César es

$$f_Y(y) = 2yI_{(0,1)}(y).$$

   La llegada de ambos es independiente.

   a) Obtenga el valor esperado que deberán esperar entre sí. Es decir, obtenga el valor esperado de $|X - Y|$.

   b) Suponga ahora que cada uno está dispuesto a esperar hasta 15 minutos al otro. Obtenga la probabilidad de que efectivamente se encuentren.

10. Suponga que $Y = aX + b$, demuestre que si $a > 0$ luego $\rho_{XY} = 1$. Demuestre que si $a < 0$ entonces $\rho_{XY} = -1$, y si $a = 0$ entonces $\rho_{XY} = 0$.

10. Muestras aleatorias y teoremas límite

10.1. Introducción

En este capítulo trabajaremos con lo que se conoce como una **muestra aleatoria**. La idea es partir de X una variable aleatoria, que podría representar por ejemplo la duración de una determinada componente electrónica. Ahora bien, supongamos que tenemos $X_1, \ldots, X_n$ una colección de n variables aleatorias, independientes entre sí y todas con la misma distribución. Por ejemplo, podríamos tener n componentes electrónicas, construidas de forma independiente. Diremos que $X_1, \ldots, X_n$ conforman una muestra aleatoria.

Otro ejemplo, supongamos que tenemos una mesa que tiene una determinada longitud μ. Queremos medir esa mesa con un instrumento que supondremos impreciso pero insesgado. El hecho de que sea impreciso significa que tiene una variabilidad, que no siempre está midiendo lo mismo. Es decir, si X representa una medición dada, tendremos que X es una variable aleatoria con una determinada varianza $V(X) = \sigma^2$. El hecho de que las mediciones sean insesgadas significa que el instrumento está calibrado de forma tal de que en promedio mide lo que se espera, sin subestimar o sobreestimar la magnitud en cuestión. O sea, tenemos que $E(X) = \mu$ y $V(X) = \sigma^2$. Esto no quiere decir necesariamente que las mediciones tengan distribución normal, de hecho, la distribución exacta de X podría ser desconocida.

Intuitivamente, es un hecho conocido que para mejorar la precisión instrumental lo ideal es medir más de una vez y tomar un promedio. Es decir, generar $X_1, \ldots, X_n$ mediciones y considerar después $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, lo que llamaremos la **media muestral** o promedio, para diferenciarlo de la **media poblacional** que es $E(X) = \mu$.

Por qué el promedio resulta mejor? Es lo que veremos en este capítulo.

10.2. Estadísticos de una muestra

Supongamos que trabajamos con una muestra aleatoria $X_1, \ldots, X_n$, con $E(X_i) = \mu$ y $V(X_i) = \sigma^2$. Un **estadístico** se define como una función $h(X_1, \ldots, X_n)$ de la muestra. Vamos a considerar particularmente dos de ellos. El primero es la **suma** o **total**:

$$S_n = \sum_{i=1}^{n} X_i. \tag{2}$$

El segundo es el **promedio** o **media muestral**:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{3}$$

Algunas observaciones interesantes, vamos a calcular la esperanza y la varianza de estos estadísticos.

$$E(S_n) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \mu = n\mu.$$

$$V(S_n) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} \sigma^2 = n\sigma^2.$$

En el cálculo de la varianza se usa fuerte que las variables son independientes. Vamos a proceder similarmente con las cuentas para el promedio. Usaremos que $V(aX) = a^2 V(X)$. Así pues:

$$E(\overline{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

y

$$V(\overline{X}_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}.$$

Son muy interesantes los resultados obtenidos para el caso del promedio, la esperanza de $\overline{X}_n$ coincide con la de X. O sea, que si el instrumento de medición, volviendo con ese ejemplo, resulta insesgado, el promedio de n mediciones tiene el mismo valor esperado que la longitud de la mesa. Pero ahora bien, y quizás esto sea más interesante, la varianza del promedio es σ^2/n. Es decir, que la varianza del promedio está tendiendo a cero. Una variable que tiene varianza cero es constante, con lo cual estamos teniendo que el promedio en algún sentido tiende a una constante que será el del valor esperado, μ, la verdadera longitud de la mesa.

Vamos a formalizar mejor estos resultados con algunas observaciones teóricas.

10.3. Desigualdad de Tchebyshev

Supongamos que tenemos una variable aleatoria X, por sí sola, de la cual sabemos bien poco. Por ejemplo, sabemos que $E(X) = 5$ y que $V(X) = 4$. No sabemos más nada, no tenemos noción alguna de su distribución. Tenemos el siguiente problema, nos piden calcular la probabilidad de que X esté entre 4 y 6. O sea, que X diste no más de 1 con respecto a su valor esperado. Formalmente, sería, $P(4 < X < 6) = P(|X - 5| < 1)$. Es posible resolver esto? Claramente, la respuesta debería ser **no**. Es imposible obtener la probabilidad **exacta** que nos están pidiendo pues no tenemos la distribución de X. Ni siquiera sabemos si es continua o discreta!

Sin embargo, si somos menos ambiciosos, quizás algo podamos hacer... Por ejemplo, tratar de ver si esta probabilidad se puede **acotar**. Eso ciertamente es pretender menos, y en efecto quizás algo se puede hacer. En ese sentido, viene al rescate un resultado que se conoce como la **desigualdad de Tchebyshev**.

Theorem 10.1. *Sea X una variable aleatoria con esperanza μ y varianza σ^2. Luego, vale que para todo $\epsilon > 0$,*

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

Vamos a dar una demostración de este resultado solamente para el caso continuo, probarlo para el caso general requiere usar algunos argumentos técnicos de teoría de la medida que están fuera del alcance de este curso. Posteriormente, veremos cómo utilizarlo e interpretarlo.

Bien, suponiendo X una variable aleatoria continua, tenemos que tiene una densidad f_X que es lo que permite calcular probabilidades. Además, tenemos en general que $\sigma^2 = V(X) = E((X - \mu)^2)$.

$$\sigma^2 = E((X - \mu)^2) = \int_{\mathbb{R}} (x - \mu)^2 f_X(x)dx.$$

Vamos a partir la recta real en dos regiones, la región A es donde x es tal que $|x - \mu| \geq \epsilon$, y la región B es su complemento, es decir, $|x - \mu| < \epsilon$. Bien pues, entonces tendremos que

$$\int_{\mathbb{R}} (x - \mu)^2 f_X(x)dx = \int_{|x-\mu|\geq\epsilon} (x - \mu)^2 f_X(x)dx + \int_{|x-\mu|<\epsilon} (x - \mu)^2 f_X(x)dx.$$

Bien, la segunda integral es algo positivo o en el peor de los casos 0, podemos acotar esta cuenta por algo en donde no esté la segunda integral. Es decir,

$$\int_{\mathbb{R}} (x - \mu)^2 f_X(x)dx \geq \int_{|x-\mu|\geq\epsilon} (x - \mu)^2 f_X(x)dx.$$

Como estamos en una región en donde $|x - \mu| \geq \epsilon$, podemos acotar otra vez, esta vez haciendo que $(x - \mu)^2 \geq \epsilon^2$. Nos queda:

$$\int_{|x-\mu|\geq\epsilon}(x-\mu)^2 f_X(x)dx \geq \int_{|x-\mu|\geq\epsilon}\epsilon^2 f_X(x)dx = \epsilon^2\int_{|x-\mu|\geq\epsilon} f_X(x)dx = \epsilon^2 P(|x-\mu|\geq\epsilon).$$

O sea, que

$$\sigma^2 \geq \epsilon^2 P(|x-\mu|\geq\epsilon),$$

despejamos y sale que

$$P(|x-\mu|\geq\epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

Hemos probado así el resultado, aunque sea para el caso continuo.

Una forma de interpretar este resultado es viendo que $P(|X-\mu|\geq\epsilon)$ calcula la probabilidad de que X se encuentro "lejos" de su valor central que es μ. La probabilidad de que eso pase no puede ser arbitrariamente grande, y se acota por algo que depende de la varianza (σ^2) y un término (ϵ) que demarca la "tolerancia" que estoy dispuesto a permitir como distancia del centro.

Veamos cómo usarlo en nuestro problema.

Teníamos X una variable aleatoria con $E(X) = 5$ y $V(X) = 4$, queremos ahora acotar (en vez de calcular exactamente) la probabilidad de que X caiga entre 4 y 6. Es decir, $P(4 < X < 6) = P(|X-5| < 1)$. La idea es usar Tchebyshev, pero para eso necesitamos que la desigualdad esté "al revés". Complementamos y obtenemos que $P(|X-5| < 1) = 1 - P(|X-5| \geq 1)$. Pues bien, afortunadamente estamos en un caso en donde la expresión involucra un módulo centrado en 5 que es precisamente la esperanza de X. Luego, $P(|X-5| \geq 1) \leq \frac{\sigma^2}{1^2} = 4/1 = 4$.

Luego pues, $P(|X-5| < 1) = 1 - P(|X-5| \geq 1) \geq 1 - 4 = -3$.

O sea, obtuvimos que la probabilidad deseada era mayor o igual a -3 (!!!!). Esta estó "mal"?? En rigor de verdad no, pues una probabilidad en efecto es al menos cero. Lo que está pasando es que, en este caso, Tchebyshev nos está dando una cota demasiado "conservadora", tanto que el resultado final que nos da es en efecto inútil! Eso pasa porque la varianza en este caso es bastante grande en comparación con el ϵ que nos habíamos planteado, bien puede ocurrir que la desigualdad de Tchebyshev sea, si bien siempre correcta, no informativa. Modifiquemos un poco el problema.

Supongamos ahora que tenemos $X_1, \ldots, X_n$ una muestra aleatoria, con las mismas características de antes, es decir, $E(X_i) = 5$ y $V(X_i) = 4$. Supongamos $n = 100$ el tamaño muestral. Pero ahora, queremos obtener una cota para que el promedio (media muestral) diste menos de 1 con respecto a 5. Es decir, buscamos una cota para $P(4 < \overline{X}_n < 6) = P(|\overline{X} - 5| < 1)$.

Pensemos entonces como antes: $P(|\overline{X}-5| < 1) = 1 - P(|\overline{X}-5| \geq 1)$. Usamos Tchebyshev, tenemos que $P(|\overline{X}-5| \geq 1) \leq \frac{4}{1^2 n}$, donde hemos usado que $V(\overline{X}) = \frac{4}{n}$. Luego,

$$P(|\overline{X}-5| < 1) \geq 1 - \frac{4}{n}.$$

Como $n = 100$, tenemos que el lado derecho es $1 - 4/100 = 0{,}96$. Notemos que ahora sí nos está dando algo de información interesante, nos está diciendo que la probabilidad de que el promedio se encuentre entre 4 y 6 es AL MENOS de 0.96. O sea, que es bastante probable.

Otro tipo de pregunta que podría surgir es para qué valor de n podemos asegurar que una determinada probabilidad, por ejemplo, la probabilidad de que el promedio esté entre 4 y 6, sea mayor a un determinado número, por ejemplo, 0.99.

En esencia, es repetir las cuentas de antes, tendremos que:

$$P(|\overline{X}-5| < 1) \geq 1 - \frac{4}{n} \geq 0{,}99.$$

De despejar $1 - \frac{4}{n} \geq 0{,}99$ nos sale que $n \geq 4/0{,}01 = 400$. O sea, con 400 observaciones podré asegurar la calidad pretendida en la probabilidad. Importante: a veces de esta inecuación puede salir un n no entero, en ese caso se debe tomar el primer entero mayor al número en cuestión. El n, como representa cantidad de mediciones, siempre debe ser un número entero.

10.4. Ejercicios extras de desigualdad de Tchebyshev

Ejercicio 10.2. *Una fábrica produce bombones, que se componen en parte de un envoltorio cuyo peso sigue una distribución uniforme $U(8, 12)$ y por otro lado de contenido neto de chocolate, cuyo peso es una distribución $N(\mu = 5, \sigma^2 = 1)$. Un bombón en sí es una parte envoltorio y otra parte chocolate. La producción de todas las partes y de los bombones entre sí de hace de forma independiente. Hallar una cantidad n de bombones que se habrá de producir como para poder asegurar de que la probabilidad de que el peso promedio de los n bombones esté entre 14.9 y 15.1 sea de al menos 0.995.*

Ejercicio 10.3. *Se está tratando de medir una mesa cuya longitud exacta se sabe que es de 5 metros, usando un instrumento cuya precisión está dada por una distribución desconocida pero que cuya media es el valor exacto de lo que se pretende medir (en este caso, 5 metros) y cuya varianza es un valor desconocido σ^2. Se busca hacer una calibración del instrumento regulando el σ^2 de forma tal de asegurar que si efectúo 100 mediciones, entonces la probabilidad de que la medición promedio diste de la longitud real de la mesa en menos de 0.01 metros sea con una probabilidad de al menos 0.99. Obtener el máximo valor de σ^2 que pueda asegurar eso.*

10.5. Convergencia en Probabilidad y Ley de los Grandes Números

Dada una sucesión de variables aleatorias $\{X_n\}_n$, podemos definir una noción de ”límite probabilístico”. Diremos que la sucesión X_n **converge en probabilidad** a la variable aleatoria X, y lo notaremos como $X_n \xrightarrow{p} X$, si se cumple la siguiente condición: para todo $\epsilon > 0$, vale que

$$P(|X_n - X| > \epsilon) \to 0.$$

La idea es que a medida que crece n, el X_n se volverá cada vez más parecido a X medido como que la probabilidad de obtener un evento ”extremo” (es decir, una situación en donde X_n y X disten más de ϵ) deberá tender a cero.

No es la única forma de definir un límite con variables aleatorias, pero es la que usaremos en este curso pues a fines prácticos nos alcanzará. La convergencia en probabilidad respeta las propiedades usuales de álgebra de límites, por ejemplo, si $X_n \xrightarrow{p} X$ e $Y_n \xrightarrow{p} Y$ entonces vale que $X_n + Y_n \xrightarrow{p} X + Y$, que $X_nY_n \xrightarrow{p} XY$ y, si Y no toma valores nulos, entonces $X_n/Y_n \xrightarrow{p} X/Y$.

Por otro lado, si tenemos una función real continua g, vale que $g(X_n) \xrightarrow{p} X$.

Lo más interesante es cuando la convergencia es a un número, cosa que por ejemplo ocurrirá con los promedios muestrales a medida que el n crece. En efecto, si tenemos una muestra aleatoria $X_1, \ldots, X_n$, con n ”grande”, que podrían representar por ejemplo mediciones insesgadas de una mesa de longitud μ, entonces el promedio $\overline{X}_n$ tenderá a distar poco de μ. Más formalmente, tenemos que:

Theorem 10.4. *(Ley de los grandes números) Sea $\{X_n\}_n$ una sucesión de variables aleatorias independientes. Supongamos que $E(X_i) = \mu$ y que $V(X_i) = \sigma^2 < \infty$. Entonces, vale que*

$$\overline{X}_n \xrightarrow{p} \mu.$$

Demostración:

Para ver que $\overline{X}_n \xrightarrow{p} \mu$, veamos que se verifica la definición. Fijemos un $\epsilon > 0$, veamos efectivamente que $P(|\overline{X}_n - \mu| > \epsilon) \to 0$. Pues bien, $P(|\overline{X}_n - \mu| > \epsilon) \leq P(|\overline{X}_n - \mu| \geq \epsilon)$. Notemos que $E(\overline{X}_n) = \mu$ y $V(\overline{X}_n) = \sigma^2/n$. Usando la desigualdad de Tchebyshev, tendremos que

$$P(|\overline{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2 n}$$

y esto tiende a cero a medida que n tiende a infinito, pues tanto σ^2 como ϵ están fijos, con lo cual demostramos el teorema.

Este teorema tiene una consecuencia fundamental, afirma que una media (un promedio poblacional) se puede aproximar por un promedio muestral.

Esto tiene varias consecuencias prácticas. Por un lado, permite terminar de cerrar el argumento de por qué en la ruleta una esperanza negativa (-0.0027 en el caso de la ruleta europea) es malo para el jugador. En efecto, como el promedio muestral de varias jugadas con n grande se aproxima a la esperanza (negativa), entonces eso quiere decir que si n es muy grande, en promedio redundará en una pérdida económica que se aproximará a como si en cada jugada perdiera 2.7 centavos por cada peso apostado.

Otra aplicación de este resultado es que en virtud del mismo es que tiene sentido por ejemplo algunos aspectos de la estadística, como la confección de una encuesta. Supongamos una población muy grande ($n >> 1$), en donde cada individuo tiene una decisión tomada respecto a una determinada política de acción (por ejemplo, en el caso electoral referido a si votaría o no a un determinado candidato ”A”).

Habrá una proporción $0 \leq p \leq 1$ de personas que votarían por ese candidato A. Nuestro objetivo, como encuestadora, es determinar esa verdadera proporción p. Idealmente, para poder saber con precisión dicho valor, deberíamos efectuar lo que se conoce como un **censo**, es decir, una encuesta que cubra la población entera.

Como esto en términos prácticos no está al alcance de una encuestadora, es que la misma efectúa un **muestreo** y obtiene una muestra de individuos, idealmente de forma independiente (lo cual no siempre es fácil), $X_1, \ldots, X_n$. Podemos pensar que cada X_i es un ensayo de Bernoulli, que vale 1 con probabilidad p. Recordemos que en el caso del ensayo de Bernoulli, teníamos que $E(X_i) = p$.

Luego, tenemos que dada una muestra $X_1, \ldots, X_n$, si considero el promedio $\overline{X}_n$, tendremos que por la ley de los grandes números

$$\overline{X}_n \xrightarrow{p} p.$$

O sea, que en teoría con una muestra suficientemente grande podríamos aproximar al verdadero valor p de la proporción, lo que le daría la justificación probabilística a la encuesta sobre una muestra. Por supuesto que no es todo tan sencillo en la práctica, hay muchos considerandos a tomar en cuenta (entre ellos la independencia, la posibilidad de la no respuesta, la respuesta falsa, etc).

10.5.1. Aplicación: integración Monte Carlo

Supongamos que tenemos $g : [0, 1] \to \mathbb{R}$ una función a la cual le queremos calcular la integral. Existen varias opciones, una de ellas es obtener la primitiva y aplicar la regla de Barrow. Esto claramente no siempre es posible puesto que la dificultad de obtener primitivas puede ser alta, o directamente imposible.

Otra alternativa es recurrir a una cuadratura numérica, en esencia son variaciones de la las sumas de Riemann, como el método de los trapecios.

Veremos otro esquema, una método ”Monte Carlo”, para aproximar esta integral. La idea es definir $X_1, \ldots, X_n$ muestra aleatoria de distribución $U(0, 1)$.

Por la Ley de los grandes números, tenemos que

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{p} E(g(X)).$$

Calculemos $E(g(X))$, por la fórmula del estadístico inconsciente, esto será

$$E(h(X)) = \int_0^1 g(x) f_X(x) dx = \int_0^1 g(x) dx$$

donde hemos usado que la densidad de una uniforme es $f_X(x) = 1 \cdot I_{(0,1)}$. Nótese que hemos llegado a la integral que originalmente queríamos obtener.

Tenemos así un método relativamente sencillo para aproximar la integral, se necesita generar una cantidad ”grande” de datos uniformes, evaluar la función que se quiere integrar en cada uno de estos datos y promediar, así tenemos pues la integral deseada.

Ejercicio 10.5. *Suponga que $h : [a, b] \to \mathbb{R}$, ¿cómo aproximaría su integral?*

10.5.2. Aplicación: vuelta a la primer definición de probabilidad

Recordemos que habíamos ofrecido al comienzo un esbozo de definición de probabilidad. Habíamos dicho que dado $A \subset \Omega$ un evento, definíamos a $P(A)$ como el siguiente límite:

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n}$$

en donde en esencia se está repitiendo n veces el experimento aleatoria y se cuenta en n_A la cantidad de experimentos en donde pasó el evento A. Vamos a tratar de justificar esto a través de la ley de los grandes números.

En efecto, tenemos el evento A y definimos la variable aleatoria X_i como 1 si pasa el evento A en el ensayo i-ésimo y 0 si no. Luego, notemos que X_i es un ensayo de Bernoulli, con probabilidad $p = P(A)$.

Ahora bien, $n_A = \sum_{i=1}^{n} X_i$, con lo cual entonces

$$\frac{n_A}{n} = \overline{X}_n \xrightarrow{p} E(X) = p = P(A).$$

Así pues, el promedio muestral de n experimentos que cuentan la cantida de veces que ocurre A tiene en probabilidad a la probabilidad ”verdadera” de A, recuperando en algún sentido la primer definición que habíamos dado de probabilidad.

10.6. Teorema Central del Límite

Algunas propiedades que no hemos visto de la distribución normal, que las daremos sin demostración.

Theorem 10.6. *Si $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$ independientes, luego*

$$X + Y \sim N(\mu_x + \mu_Y, \sigma_x^2 + \sigma_y^2).$$

Es decir, la suma de dos variables normales independientes es una variable normal independiente. Los parámetros salen naturalmente de la aditividad de la esperanza y de la varianza ante independencia.

Análogamente, se puede probar que

Theorem 10.7. *Si $X \sim N(\mu_x, \sigma_x^2)$ y $a \neq 0$, luego*

$$aX \sim N(a\mu_x, a^2\sigma_x^2).$$

O sea, multiplicar por un escalar no nulo preserva la normalidad, con los parámetros ajustados como deben ser, recordando que $E(aX) = aE(X)$ y que $V(aX) = a^2V(X)$.

Supongamos, para simplificar las cosas, que $X_1, \dots, X_n$ es una muestra aleatoria de distribución normal con media μ y varianza σ^2. Recurriendo a los dos teoremas anteriores, tenemos que tanto la suma S_n como el promedio $\overline{X}_n$ serán normales, con los parámetros que determinamos en la primer sección de este capítulo. Es decir,

$$S_n \sim N(n\mu, n\sigma^2)$$

y

$$\overline{X}_n \sim N(\mu, \sigma^2/n).$$

Al estandarizarlo, tendríamos que

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \sim N(0, 1)$$

y que

$$\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1).$$

Todo esto lo hemos podido hacer porque teníamos la normalidad de los datos. Qué pasa si falla esa hipótesis? Es decir, qué pasa si tenemos que los datos NO son necesariamente con distribución normal, pero estamos trabajando con S_n y $\overline{X}_n$. La respuesta, es que en el fondo las cosas cambian pero no tanto. Si el n es "grande", podremos utilizar la expresión anterior de la estandarización y, bajo una aproximación, tener un resultado que es aproximadamente el de una normal estándar.

Es decir, cuando trabajamos con los agregados estadísticos S_n y $\overline{X}_n$, podemos suponer que ambos tienen una distribución **aproximadamente normal**. No precisaremos mejor la idea de este "aproximado", requeriría meternos con un tipo de convergencia adicional. Enunciaremos, sin demostración, una aproximación formal de lo que se conoce como el Teorema Central del Límite:

Theorem 10.8. *(Teorema Central del Límite) Sean $X_1, \ldots, X_n$ una muestra aleatoria tal que $E(X_i) = \mu$ y $V(X_i) = \sigma^2$. Luego, vale que para n grande (asintótico), tenemos que*

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \overset{a}{\sim} N(0,1)$$

y que

$$\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \overset{a}{\sim} N(0,1).$$

Debemos entender $\overset{a}{\sim}$ como un "se distribuye aproximadamente como".

Este resultado sirve para aproximar probabilidades en situaciones en donde la cuenta exacta es impráctica o imposible de obtener, permitiendo un pasaje a la distribución normal.

Vamos a ser una serie de ejercicios de utilización de este teorema.

Ejercicio 10.9. *Sean $X_1, \ldots, X_n$ una muestra aleatoria tal que $E(X_i) = 5$ y $V(X_i) = 4$.*

1. *Supongamos $n = 100$, calcule de forma aproximada la probabilidad de que $P(\overline{X}_n > 5{,}1)$.*
2. *Supongamos $n = 100$, calcule de forma aproximada la probabilidad de que $P(S_n > 508)$.*
3. *Hallar un valor n de forma tal que la probabilidad aproximada de que el promedio $\overline{X}_n$ sea mayor a 4.9 sea al menos 0.992.*
4. *Hallar un valor n de forma tal que la probabilidad aproximada de que la suma S_n sea mayor a 495 sea al menos 0.995.*

La clave en todos estos ejercicios está en el término "probabilidad aproximada". Esa será la pauta que permite la utilización del teorema central del límite.

Veamos el primer ítem. Lo que estaría pidiendo es efectivamente $P(\overline{X}_n > 5{,}1)$. La estrategia consiste en estandarizar esto como si $\overline{X}_n$ sea normal, aunque no necesariamente lo sea. Recordemos que $E(\overline{X}_n) = \mu = 5$ y $V(\overline{X}_n) = \sigma^2/n = 5/100$.

Entonces, estandarizamos. Eso es simplemente un proceso algebraico, no hay "trampa" alguna en eso (es decir, aproximación). Luego:

$$P(\overline{X}_n > 5{,}1) = P\left(\frac{\overline{X}_n - 5}{\sqrt{5/100}} > \frac{5{,}1-5}{\sqrt{5/100}}\right).$$

Ahora es cuando aplicamos el Teorema Central del Límite (TCL), una vez que está estandarizada la expresión para el $\overline{X}_n$. Lo que viene a continuación es una aproximación, deja de ser una igualdad, pues está el TCL de por medio. Tenemos que

$$P\left(\frac{\overline{X}_n - \mu}{\sqrt{5/100}} > \frac{5{,}1-5}{\sqrt{4/100}}\right) \approx_{TCL} P\left(Z > \frac{5{,}1-5}{\sqrt{4/100}}\right) = P(Z > 0{,}45)$$

con Z una variable con distribución normal estándar. Seguimos como un clásico problema de distribución normal. Tenemos que

$$P(Z > 0{,}45) = 1 - \Phi(0{,}45) = 1 - 0{,}6700 = 0{,}3300.$$

Así pues, la probabilidad aproximada buscada será 0.33.

Vamos con el segundo ítem, la estrategia será bastante similar, solamente que ahora hay que estandarizar S_n en vez de $\overline{X}_n$. Recordemos que $E(S_n) = n\mu = 500$ y que $V(S_n) = n\sigma^2 = 400$. Entonces pues:

$$P(\overline{X}_n > 508) = P\left(\frac{S_n - 500}{\sqrt{400}} > \frac{508-500}{\sqrt{400}}\right).$$

Nuevamente, aquí aplicamos el TCL y tenemos que

$$P\left(\frac{S_n - 500}{\sqrt{400}} > \frac{508-500}{\sqrt{400}}\right) \approx_{TCL} P\left(Z > \frac{508-500}{\sqrt{400}}\right) = P(Z > 0{,}4) = 1 - Phi(0{,}4) = 1 - 0{,}6554 = 0{,}3446.$$

Es decir, la probabilidad aproximada será 0.3446.

Vamos ahora al tercer ítem. Vamos a traducir en términos matemáticos lo que está planteado coloquialmente. Nos pide un n para que $P(\overline{X}_n > 4{,}9) > 0{,}992$.

La estrategia es como siempre, estandarizamos primero pensando que $\overline{X}_n$ es normal aunque no lo sea. Entonces:

$$P(\overline{X}_n > 4{,}9) = P\left(\frac{\overline{X}_n - 5}{\sqrt{4/n}} > \frac{4{,}9-5}{\sqrt{4/n}}\right).$$

Apelamos al teorema central del límite aquí, quedándonos:

$$P\left(\frac{\overline{X}_n - 5}{\sqrt{4/n}} > \frac{4{,}9-5}{\sqrt{4/n}}\right) \approx_{TCL} P\left(Z > \frac{4{,}9-5}{\sqrt{4/n}}\right) = 1 - \Phi\left(\frac{4{,}9-5}{\sqrt{4/n}}\right) > 0{,}992$$

O sea, queremos que

$$\Phi\left(\frac{4{,}9-5}{\sqrt{4/n}}\right) < 0{,}008,$$

implicando que

$$\frac{4{,}9-5}{\sqrt{4/n}} < \Phi^{-1}(0{,}008).$$

Vamos a obtener, mediante la tabla, el percentil $\Phi^{-1}(0{,}008)$. Habría que buscar del interior de la tabla hacia los márgenes. Recordemos que en este caso ese percentil no lo podemos hallar (corresponde a un z negativo), con lo cual apelamos, como hicimos antes, a la simetrización. Tenemos que $\Phi^{-1}(0{,}008) = -\Phi^{-1}(1 - 0{,}008) = -\Phi^{-1}(0{,}992)$.

Ahora sí, buscamos 0.992 en el interior de la tabla y vemos que se corresponde con 2.41. Luego, $\Phi^{-1}(0{,}008) = -2{,}41$.

Entonces, debemos despejar n de la inecuación:

$$\frac{4{,}9-5}{\sqrt{4/n}} < -2{,}41$$

lo que es relativamente sencillo de hacer, nos queda que

$$\sqrt{n} > 48{,}2.$$

O sea, $n > 2323,24$. Como n debe ser natural, habremos de considerar $n \geq 2324$. A partir de ese valor de n podemos asegurar de que la probabilidad aproximada de que el promedio muestral de n elementos sea mayor a 4.9 será al menos de 0.992.

Veamos el último ítem. Es bastante parecido al anterior, aunque involucra una pequeña complicación adicional que ya apreciaremos. Nos piden n para que $P(S_n > 495)$ sea al menos 0.995, pensando la P como una probabilidad aproximada. Como siempre, empezamos estandarizando:

$$P(S_n > 495) = P\left(\frac{S_n - 5n}{\sqrt{4n}} > \frac{495 - 5n}{\sqrt{4n}}\right) \approx_{TCL} P\left(Z > \frac{495 - 5n}{\sqrt{4n}}\right) = 1 - \Phi\left(\frac{495 - 5n}{\sqrt{4n}}\right) > 0{,}995.$$

De aquí despejamos y obtenemos que $\Phi\left(\frac{495-5n}{\sqrt{4n}}\right) < 0{,}005$. Despejamos para obtener que $\frac{495-5n}{\sqrt{4n}} < \Phi^{-1}(0{,}005) = -\Phi^{-1}(1 - 0{,}005) = -\Phi^{-1}(0{,}995) = -2{,}57$.

Pues bien, tenemos la inecuación

$$\frac{495 - 5n}{\sqrt{4n}} < -2{,}57.$$

Esta ya no es tan sencilla como en el caso del promedio, pues el n aparece dos veces. Vamos a reescribir un poco esta inecuación. Pasando multiplicando el $\sqrt{4n}$ y agrupando todo en un lado de la desigualdad, nos queda la siguiente inecuación:

$$5n - 2{,}57\sqrt{4}\sqrt{n} - 495 > 0.$$

O sea,

$$5n - 5{,}14\sqrt{n} - 495 > 0.$$

Si miramos fijamente esta inecuación, notaremos que en el fondo se está ocultando una expresión cuadrática. En efecto, si hacemos la sustitución $x = \sqrt{n}$, tenemos que nos queda $5x^2 - 5{,}14x - 495 > 0$. O sea, debemos hallar los valores de x que conforman el conjunto de positivdad de la cuadrática $5x^2 - 5{,}14x - 495$. Esta cuadrática tiene concavidad positiva, en el caso de que tenga dos raíces $x_1 < x_2$, el conjunto de positivad serán los elementos a la izquierda de x_1 o bien a la derecha de x_2. Vamos a hallar las raíces, aplicando la fórmula del discriminante cuadrático obtenemos estas raíces $x_1 = -9{,}45$ y $x_2 = 10{,}47$.

Ahora bien, notemos que no tiene sentido mirar los valores de x que estén a la izquierda de $-9{,}45$ puesto que se corresponderían a valores x negativos y no olvidemos que x es $\sqrt{n}$, que necesariamente deberá ser positivo. Luego, tomamos los $\sqrt{n} = x > 10{,}47$. De aquí, nos sale que $n > 109,69$. Nuevamente, como n debe ser un natural entonces tomamos a partir de 110.

Luego, con 110 o más elementos podremos asegurar que la suma será mayor a 495 con probabilidad aproximada de al menos 0.995.

10.7. (Optativo) Aplicaciones del Teorema Central del Límite: tasa de convergencia

Ya sabíamos, por la Ley de los Grandes Números, que $\overline{X}_n \xrightarrow{p} \mu$ bajo una sucesión de variables $\{X_n\}_n$ independientes e idénticamente distribuidas, con $E(X_i) = \mu$.

El TCL da otro tipo de información, nos dice además algo sobre la velocidad de convergencia. En efecto, tenemos que $\overline{X}_n - \mu \xrightarrow{p} 0$. Queremos determinar lo que se conoce como la tasa de convergencia. Formalmente, la idea es examinar el límite de algo de este estilo

$$n^{\alpha}(\overline{X}_n - \mu) \to ??$$

con $\alpha > 0$. Para terminar de entender la idea, pensemos por un segundo en sucesiones de números reales. Supongamos que tenemos $a_n = \frac{n+1}{n}$. Está claro que $a_n \to 1$, es un ejercicio sencillo de Cálculo. Ahora bien, nos preguntamos la velocidad de convergencia en la que esto ocurre. Es rápido? Es lento? Con qué potencia de n va tendiendo esto a 1?

Formalmente, nos preguntamos qué valor de α hace que $n^{\alpha}(a_n - 1)$ tienda a algo que siga siendo cero (α muy pequeño) o bien infinito (α muy grande). Queremos el α "justo" en donde el límite nos de un número real no cero ni infinito. Bien, $a_n - 1 = 1/n$. Está claro que para que $n^{\alpha} 1/n$ tienda a algo que no sea cero ni infinito necesitamos exactamente que α sea 1. La velocidad de convergencia de a_n a 1 es entonces de orden n.

Volviendo a nuestro caso probabilístico, buscamos α tal que $n^{\alpha}(\overline{X}_n - \mu)$ converja a una distribución que no sea la trivialmente cero ni la infinito. Cuál es el valor correcto de α? Pues bien, gracias al TCL tendremos que ese α deberá ser $1/2$. En efecto, tenemos por TCL que

$$\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} = \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \to N(0, 1).$$

Luego, notemos que como aparece $\sqrt{n}$ acompañando a $(\overline{X}_n - \mu)$, con un límite que no es una distribución trivialmente cero o infinito, eso establece su velocidad de convergencia en 1/2.

En cierta forma, es esa la velocidad de convergencia del promedio muestral al promedio poblacional. Eso se traduce en que una muestra mayor (es decir, mayor n) hace que se asemejen más ambos, pero no corre a una velocidad proporcional a n sino a la raíz de n. Entonces, cada vez costará un poco más obtener los mismos resultados de calidad a medida que va creciendo n. Una muestra de tamaño 10000 no es 100 veces "mejor" que una de tamaño 100, sino que en términos prácticas será $\sqrt{100} = 10$ veces mejor.

Esto es de crucial importancia en problemas de estadística, a la hora de realizar estudios sobre estimación con márgenes de error. La precisión de las estimaciones mejorará en función del tamaño muestral, pero no será una mejora proporcional a ella sino a su raíz.

10.8. Ejercitación

1. *a)* Sea X una v.a. continua con distribución desconocida tal que $E(X) = 5$ y $V(X) = 0{,}1$, acote la probabilidad de que X se encuentre entre 4.5 y 5.5.

   b) Sea $X_1, \ldots, X_n$ (con $n = 10$) una muestra aleatoria con $E(X_i) = 5$ y $V(X_i) = 1$. Acotar la probabilidad de que $\overline{X}_n$ (el promedio muestral) esté entre 4.5 y 5.5.

   c) ¿Qué pasa con la cota calculada en b) cuando n tiende a infinito?

   d) Halle un valor n para el cual la probabilidad encontrada en b) sea seguro mayor o igual a 0.999.

2. El número de aviones que aterrizan en un aeropuerto sigue un procesode Poisson de intensidad 10 aviones cada 20 minutos. Determinar una cota para la probabilidad de que el número de aviones que aterrizan en un período de 1 hora esté entre 20 y 40.

3. Sea p la probabilidad de que una persona elegida apoye una determinada moción. Para estimar p se encuesta a n personas elegidas completamente al azar y se suma la cantidad de ellas que apoya la moción, siendo T ese valor.

   a) ¿Qué distribución tiene T?

   b) Se considera el promedio $\overline{X}_n = T/n$, que debe ser una buena forma de estimar p. Hallar una cota superior para $P(|\overline{X}_n - p| > 0{,}1)$ que no dependa de p.

c) ¿A cuánta gente se debería entrevistar para poder asegurar que $P(|\overline{X}_n - p| > 0{,}1)$ sea seguro menor que 0.1?

4. (versión generalizada de la ley de los grandes números) Sea $X_1, \ldots, X_n$ una sucesión de variables aleatorias independientes pero no idénticamente distribuídas. Se sabe que todas ellas comparten la misma esperanza que es μ, pero $V(X_i) = \sigma_i^2$, con $\sigma_i^2 \leq C$ para todo i (es decir, la varianza está acotada por C).

   a) Usando Tchebyshev, demuestre que $P(|\overline{X}_n - \mu| > \epsilon) \leq K/n$ para una constante K que no depende de n pero sí de ϵ.

   b) Concluya que $\overline{X}_n \xrightarrow{p} \mu$.

5. Una empresa láctea produce una cierta variedad de queso en unidades cuyo peso (en kg) es una v.a. con media 2 y varianza 0.04.

   a) Calcular en forma aproximada la probabilidad de que 60 quesos pesen más de 122 kg.

   b) ¿Cuántas unidades serán necesarias para satisfacer un pedido de 5000 kg con probabilidad aproximada mayor o igual que 0.95?

   c) ¿En qué cambiarían los resultados obtenidos en los incisos anteriores si el peso de fuera una variable con distribución $N(\mu = 2, \sigma^2 = 0{,}04)$?

6. Para rellenar una zona baja del río llegan diariamente 3 camiones A, B y C llenos de piedras. Los pesos en toneladas de las cargas de los camiones son v.a. independientes con esperanzas 1.8, 3.8, 4.1 y varianzas 0.1, 0.18 y 0.25 respectivamente.

   a) ¿Cuál es la probabilidad aproximada de que en 360 días se superen las 3500 toneladas?

   b) ¿Cuánto tiempo sería necesario para superar las 3500 toneladas con una probabilidad aproximada de 0.90 por lo menos?

7. Una fábrica produce caramelos de forma independiente cuyo peso neto en gramos sigue una distribución $U(0, 10)$. Se producen n caramelos.

   a) Halle n para que la probabilidad exacta de que el peso promedio de n caramelos caiga entre 4,9 y 5,1 sea al menos 0.989.

   b) Se paga 5\$ cada vez que el caramelo pesa más de 6 gramos, y serán 0\$ si no es así. Halle un n para el cual se pueda asegurar que la probabilidad aproximada de que el total de dinero recaudado por los n caramelos sea mayor a 5000\$ sea mayor o igual a 0.992.

11. Introducción a la Estadística

11.1. ¿Qué es la estadística?

Hasta el momento, el foco de nuestro curso ha estado puesto en teoría de Probabilidades. Hablaremos un poco de Estadística, que sería la otra cara de la moneda.

La probabilidad apunta en algún sentido a responder problemas que tienen que ver con casos particulares sabiendo la generalidad, lo universal. En contrapartida, la estadística busca explicar lo universal a partir de lo particular. En cierto sentido, son disciplinas complementarias.

Vamos a ejemplificar esto. En un problema de probabilidades, podríamos saber por ejemplo que una moneda tiene una probabilidad $p = 1/3$ de que salga cara, siendo p un valor conocido, y nos preguntamos la "particularidad" de que en 15 arrojadas salgan exactamente 8 caras, por ejemplo. Nótese que la universalidad es conocida, sabemos que la probabilidad de que salga cara en la moneda es 1/3, buscamos preguntarnos algo sobre un caso particular, que es la probabilidad de un determinado escenario sabiendo esta universalidad.

Un problema de estadística sería lo opuesto. No sabemos por ejemplo p la probabilidad de que en una determinada moneda salga cara, y buscamos tener una idea de este valor. Una forma de hacerlo es arrojar varias veces la moneda, por ejemplo 100, y contar la cantidad de caras, por ejemplo 43. A partir de eso, con alguna cuenta, estimaríamos el p, es decir, la probabilidad de que salga cara. Parece razonable estimar a p por $\widehat{p} = 43/100$. Este $\widehat{p}$ es lo que se conoce como un **estimador** de p.

No es fácil dar una definición directa de un estimador, el mismo más que matemáticamente se define más bien por su funcionalidad, que es tratar de inferir algo sobre una realidad universal desconocida. El caso de la moneda representa el ejemplo de lo que sería un **estimador puntual**, puesto que buscamos capturar una característica universal a través un valor fijo, no se captura en esto cuestiones de error.

Por ejemplo, en el caso de la moneda, tenemos que efectivamente si tenemos X_i un ensayo de Bernoulli que vale 1 si la moneda sale cara y 0 sino, es fácil deducir que $E(X_i) = p$ y que si tomamos $\widehat{p} = \overline{X}_n$ tendremos que, por la ley de los grandes números, entonces $\overline{X}_n \xrightarrow{p} E(X_i) = p$.

Tenemos en efecto que $\overline{X}_n$ parece un buen estimador de p, pues asintóticamente se asemeja a p. Muchas veces empero este tipo de respuestas puede ser insuficiente, nos podría interesar el error que uno podría estar cometiendo, es decir, una noción de margen de error.

Es en algunos casos importante, por ejemplo si ahora estamos en un proceso en donde buscamos determinar si una determinada legislación podría ser aprobada o no en función de un plebiscito, entonces estamos en un problema similar al de la moneda. En rigor de verdad, si la población es finita, de tamaño N, tenemos K personas que apoyarían la moción y $N - K$ que no la apoyarían (descartamos los indecisos). Obtenemos una muestra de n individuos. El modelo que en principio mejor respondería a esto sería el hipergeométrico, puesto que obtener una muestra es como hacer extracciones sin reposición. Sin embargo, si el $n << N$, se puede, sin cometer un error demasiado grosero, aproximar este modelo por uno binomial, en donde $p = K/N$, y donde buscamos estimar p.

Además, queremos obtener una idea del error que eventualmente podemos estar cometiendo, en especial si el $\widehat{p}$ queda bastante en el borde. De aquí surge la noción de **intervalo de confianza**, que veremos en este capítulo. El intervalo de confianza es un mecanismo que permite justamente dar una idea del error, devolviendo en vez de un único valor para la estimación, como hace el estimador puntual, sino que considerando un rango de posibles valores, con una confianza "alta".

Entonces, si bien podría ser, en el ejemplo de la consulta, que al entrevistar a $n = 100$ personas obtenemos que 53 apoyan la legislación, distinto es si esto tiene un margen de error que determina un intervalo de confianza entre por ejemplo $(52{,}5, 53{,}5)$ o bien uno en donde tenemos $(49, 57)$. En el primero podemos afirmar con bastante seguridad de que la legislación saldrá aprobada, en el segundo ya no podemos estar tan seguros.

11.2. Principios de Intervalo de Confianza

En general, lo que buscamos construir es un intervalo de confianza que dada una muestra aleatoria con alta probabilidad contenga al parámetro en cuestión. Es decir, buscamos A y B aleatorios dependiente de la muestra de forma tal que si el parámetro que buscamos es θ, entonces el intervalo (A, B) contenga a θ con alta probabilidad.

En particular, lo que se suele hacer es fijar un valor α de forma que $P(A < \theta < B) = 1-\alpha$, donde típicamente se toma para α valores como 0,05 o bien 0,01, por convención. Entonces, el intervalo aleatorio (A, B) contendrá al verdadero parámetro θ con una probabilidad de $1-\alpha$, que podría ser por ejemplo 0.95 o bien 0.99, dependiendo el valor de α escogido.

Si bien tenemos la expresión $1 - \alpha = P(A < \theta < B)$ es importante aclarar que lo aleatorio no es θ, sino A y B, que son funciones de la muestra $X_1, \ldots, X_n$.

Para hacer las cuentas finas con esto, requeriremos apelar al otro resultado importante del capítulo anterior, el Teorema Central del Límite, que nos da una visión más centrada sobre cómo es la convergencia.

11.3. Intervalo de Confianza: caso Normal con σ^2 conocido

Supongamos que tenemos $X_1, \ldots, X_n$ una muestra aleatoria con distribución $N(\mu, \sigma_0^2)$, siendo μ desconocido y σ_0^2 conocido, buscamos estimar μ. Está claro, por la Ley de los Grandes Números, que un posible estimador puntual para μ será $\widehat{\mu} = \overline{X}_n$. Queremos ahora construir un intervalo de confianza para tener una mejor idea de la precisión/error de la estimación.

Recurriremos a un método que se conoce como la ”técnica del pivote”. Un **pivote** se define como un estadístico (una función de la muestra) en donde en su expresión aparecen el parámetro que se busca estimar pero cuya distribución no depende del parámetro. Esto permitirá efectuar un ”despeje” del parámetro, como pasaremos a ver inmediatamente en esta situación.

Partimos de $\overline{X}_n$, que ya sabemos que opera bien como estimador puntual de μ, ya que $E(\overline{X}) = \mu$ y por la Ley de los Grandes Números entonces $\overline{X}_n \xrightarrow{p} \mu$. Bien, tenemos que $\overline{X}_n \sim N(\mu, \sigma_0^2/n)$. Si estandarizamos, tendremos que

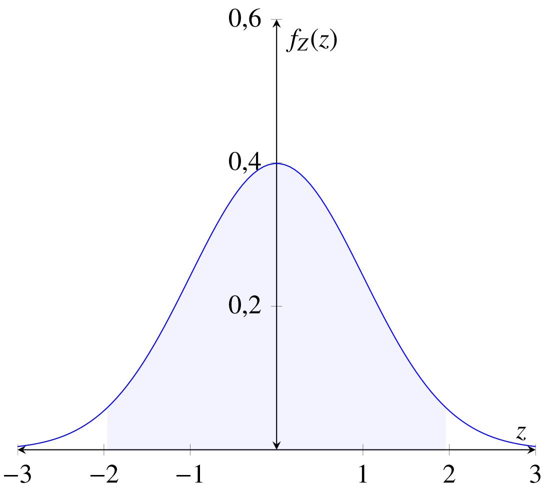
$$Z = \frac{\overline{X}_n - \mu}{\sqrt{\sigma_0^2/n}} \sim N(0, 1).$$

Notemos que Z es efectivamente un buen pivote para nuestro despeje, ya que en su expresión aparece el parámetro que buscamos despejar, μ y su distribución no depende de nada desconocido, siendo una normal estándar.

El procedimiento para deducir el intervalo de confianza es el siguiente. Primero, establecemos el nivel de confianza del intervalo, que sería el $1 - \alpha$. Valores típicos para esto son $\alpha = 0{,}05$ dando un intervalo de nivel 0,95 o bien $\alpha = 0{,}01$, para un intervalo de nivel 0,01. Una vez fijado esto, necesitamos encontrar valores A_1 y A_2 tales que

$$1 - \alpha = P(A_1 \leq Z \leq A_2).$$

O sea, queremos un cubrimiento lo suficientemente bueno en términos de probabilidad del pivote Z, es decir, de probabilidad $1 - \alpha$. Eso quiere decir que queda una probabilidad α fuera de la cobertura. La manera usual de lograr eso, aprovechando la simetría, es eligiendo $A_1 = z_{\alpha/2}$ y $A_2 = z_{1-\alpha/2}$, es decir, percentiles $\alpha/2$ y $1 - \alpha/2$. Por ejemplo, si tomamos $\alpha = 0{,}05$, tendremos que $z_{\alpha/2} = -1{,}96$ y que $z_{1-\alpha/2} = 1{,}96$ (sale de la tabla de la normal estándar). En general, por la simetría, tendremos que $z_{\alpha/2} = -z_{1-\alpha/2}$. Gráficamente:



El área en azul cubre exactamente 0.95, o bien en general $1-\alpha$. Entonces pues, tendremos que

$$1-\alpha = P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = P\left(-z_{1-\alpha/2} < \frac{\overline{X}_n - \mu}{\sqrt{\sigma_0^2/n}} < z_{1-\alpha/2}\right)$$

Despejamos algebraicamente esta inecuación, para que quede μ en el centro, el parámetro de interés. Es relativamente sencillo ver que quedará:

$$1-\alpha = P(\overline{X}_n - z_{1-\alpha/2}\sigma_0/\sqrt{n} < \mu < \overline{X}_n + z_{1-\alpha/2}\sigma_0/\sqrt{n}).$$

O sea que el intervalo de confianza para μ de nivel $1-\alpha$ será:

$$IC_{1-\alpha}(\mu) = (\overline{X}_n - z_{1-\alpha/2}\sigma_0/\sqrt{n}, \overline{X}_n + z_{1-\alpha/2}\sigma_0/\sqrt{n}).$$

O sea, que tenemos una estimación para μ que estará dado por

$$IC_{1-\alpha}(\mu) = \overline{X}_n \pm z_{1-\alpha/2}\sigma_0/\sqrt{n}.$$

Una cosa que interesará en general es la longitud del intervalo, que está asociado a la precisión de la estimación que efectuamos sobre μ. La longitud es lisa y llanamente restar los dos bordes del intervalo, y obtenemos así que

$$LIC_{1-\alpha}(\mu) = 2z_{1-\alpha/2}\sigma_0/\sqrt{n}.$$

Notemos que se tacha $\overline{X}_n$, y en particular cualquier cuestión que sea aleatoria. Esto es algo muy característico de esta situación en particular, tendremos que la longitud del intervalo (la precisión) es posible determinarla de antemano, antes de realizar el muestreo en cuestión. Es importante este punto en muchos casos pues permite hacer un diseño experimental antes de efectuar el experimento en sí, recolectando las muestras y midiendo.

Algunas observaciones:

- A medida que disminuye α, aumenta la confianza y por ende la cobertura, pero esto se paga con valores $z_{1-\alpha}$ que son cada vez más grandes, implicando un aumento en la longitud del intervalo de confianza, o sea, una pérdida de precisión. Luego, más confianza/cobertura implica menos precisión.

- A mayor varianza de los datos muestrales, mayor es la longitud del intervalo y por ende menor precisión. Mayor varianza en los datos se paga con menos precisión.

- Mayor tamaño muestral (n) se traduce en un intervalo más pequeño, luego mayor tamaño de la muestra implica mayor precisión. La tasa de decrecimiento va con la raíz de n.

Hagamos un ejemplo completo de esto para un problema de diseño experimental.

Se desea determinar la longitud μ en centímetros de una mesa efectuando una serie de mediciones con un instrumento con un error de medición Y_i que sigue una distribución normal con $\sigma_0 = 2$ centímetros y que además es insesgado. Es decir, $E(Y_i) = 0$. Para eso, sea realizarán varias mediciones, n, y se tomará como aproximación de la mesa el promedio de las mediciones. Como se quiere tener una idea de margen de error se construirá un intervalo de confianza de nivel $1 - \alpha = 0.95$.

a) Hallar n de forma tal que asegura que a un nivel del 95 % la precisión de la medición, entendida como la longitud del intervalo de confianza, sea de a lo sumo 0.1 centímetros.

Cada medición será una variable aleatoria $X_i = \mu + Y_i$. Notemos que $X_i \sim N(\mu, \sigma_0^2 = 2^2)$.

Tenemos la fórmula antes deducida para la longitud del intervalo de confianza, que es $L = 2z_{1-\alpha/2}\sigma_0/\sqrt{n}$. Aplicado al caso $\alpha = 0{,}95$ tenemos que $z_{1-\alpha/2} = z_{0{,}975} = 1{,}96$ (por tabla) y $\sigma_0 = 2$. Luego,

$$\frac{2 \cdot 1{,}96 \cdot 2}{\sqrt{n}} \leq 0{,}1,$$

despejamos y sale

$$\sqrt{n} \geq 2 \cdot 1{,}96 \cdot 2/0{,}1 = 78{,}4.$$

O sea, $n \geq 6146, 56$. Tomamos así $n = 6147$. Esa será la cantidad de mediciones necesarias para asegurar la precisión deseada.

Habiendo hecho esto, realizamos las 6147 mediciones y obtenemos $\overline{X}_n = 152{,}46$ centímetros como estimación de μ, la medida real de la mesa. En términos de intervalo de confianza,

$$(\overline{X}_n - z_{1-\alpha/2}\sigma_0/\sqrt{n}, \overline{X}_n + z_{1-\alpha/2}\sigma_0/\sqrt{n}) = (152{,}46 - 0{,}05, 152{,}46 + 0{,}05) = (152{,}41, 152{,}51).$$

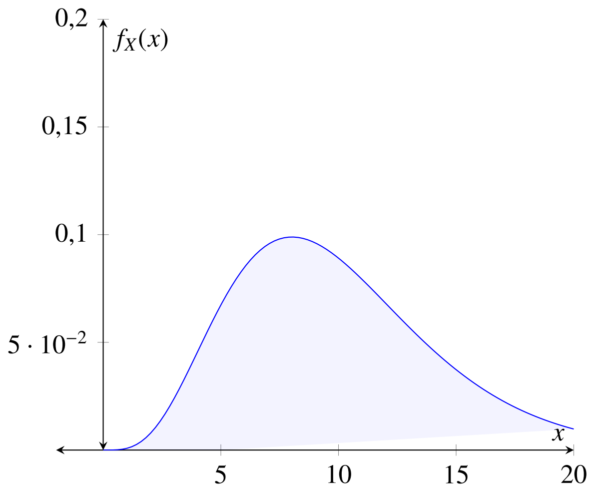
11.4. Distribuciones especiales

A fin de poder elaborar lo que viene a continuación precisaremos trabajar con una serie de distribuciones especiales.

11.4.1. Distribución χ^2 ”Chi-cuadrado”

Consideremos Z una variable aleatoria normal estándar $N(\mu = 0, \sigma^2 = 1)$. En lo que vendrá en las próximas secciones trabajaremos bastante con Z^2 y variaciones de la misma. Se puede calcular la distribución de Z^2, que responde a un tipo especial de familia que no se ha tratado en este escrito, conocido como la distribución Gamma. Z^2 será un tipo especial de distribución Gamma que se denomina χ_1^2, siendo el 1 un parámetro de la distribución que se conoce como los ”grados de libertad”. En general, tendremos que si $Z_1, \ldots, Z_n$ es una muestra aleatoria con distribución normal estándar, entonces a la distribución de $Z_1^2 + \ldots + Z_n^2$ la llamaremos χ_n^2, distribución chi-cuadrado con n grados de libertad.

Mostremos un gráfico de la densidad de una χ_{10}^2:



Esta distribución tiene soporte en los valores reales positivos, su densidad va tendiendo a cero a medida que x tiende a infinito.

Supongamos ahora $X_1, \ldots, X_n$ una muestra aleatoria con distribución $N(\mu, \sigma^2)$, consideremos la siguiente expresión:

$$Q = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 = \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2.$$

A partir de esto, nos surge el siguiente estimados ”natural” para σ^2 en el caso en que μ sea μ_0 conocido:

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_0)^2.$$

En esencia, $\tilde{\sigma}^2$ reemplaza la varianza poblacional por una versión muestral, en donde se usa a la propia muestra para tener una idea de la dispersión de los datos y de la distribución.

Es fácil obtener que

$$\frac{n\tilde{\sigma}^2}{\sigma^2} \sim \chi_n^2$$

lo cual será de utilidad para después.

¿Y qué pasa si μ es desconocido? En ese caso, debemos estimarlo, y lo haremos por su estimador ”natural” que es $\overline{X}_n$. A partir de eso, se construye la **varianza muestral** como sigue

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2.$$

Notar que la idea de este estadístico es aproximar, usando la muestra, a la dispersión de la población. Eso se hace midiendo distancias al cuadrado de cada observación con respecto a su media, pero como esta es desconocida, se reemplaza por un estimador ”natural” de la misma que es el promedio muestral $\overline{X}_n$.

Cabe preguntarse por qué dividimos por $n-1$ en vez de n. No detallaremos las razones en este escrito, pero en esencia obedece más a razones históricas, fundamentadas por el hecho de que al dividir por $n-1$ se obtiene así un estimador que se dice insesgado, es decir, que en promedio no tiende a subestimar o sobreestimar el valor que verdadero que se busca, σ^2.

Se puede demostrar, no lo haremos en el contexto de este curso, que si definimos

$$V = \frac{(n-1)s^2}{\sigma^2}$$

luego $V \sim \chi^2_{n-1}$. Es decir, el hecho de estimar μ por su promedio poblacional $\overline{X}_n$ conlleva la pérdida de un grado de libertad, que implicará como veremos más adelante un ligero perjuicio en la precisión de la estimación.

¿Cómo se opera con esta distribución? En general encontraremos tabulado los percentiles en una tabla. A diferencia de la tabla de la normal estándar, que permite calcular probabilidades y percentiles, la tabla de la χ^2_n solamente permite calcular percentiles, que suele ser el contexto de uso más frecuente. Típicamente, en las tablas tendremos en cada fila los grados de libertad y en las columnas el percentil. Es importante ver si los percentiles son de área acumulada a derecha o a izquierda, prestar atención a los instrucciones de la tabla para eso.

11.4.2. Distribución t de Student

Dado una variable Z con distribución normal estándar y U una variable aleatoria independiente de Z con distribución χ^2_n, se define la distribución **t de Student** con n grados de libertad a

$$V = \frac{Z}{\sqrt{\frac{U}{n}}}.$$

Gráficamente, la distribución t de Student es parecida a la normal, tiene forma acampanada pero su diferencia radica en que tiene colas más pesadas en los extremos. Vale que si n tiende a infinito, luego la distribución t de Student tiende a una normal estándar.

Se puede demostrar apelando a estos criterios que si definimos

$$T = \frac{\overline{X}_n - \mu}{\sqrt{\frac{s^2}{n}}}$$

entonces T tendrá una distribución t de Student con $n-1$ grados de libertad. Ya usaremos prontamente esos resultados.

La operatoria con la tabla de la t de Student es exactamente igual que con la χ^2, accedemos por fila a los grados de libertad y por columna al percentil buscado.

11.5. Intervalo de Confianza: caso Normal con σ^2 desconocido

En muchos casos el σ^2 de la distribución es desconocido, en cuyo casi se nos complica el estadístico de antes ya que depende explícitamente del σ^2.

Recordemos que el pivote utilizado era

$$Z = \frac{\overline{X}_n - \mu}{\sqrt{\sigma_0^2/n}}.$$

No lo podremos usar pues σ_0^2 ya no está más, σ^2 es desconocido. En su lugar, usamos

$$T = \frac{\overline{X}_n - \mu}{\sqrt{\frac{s^2}{n}}}$$

que será un pivote en donde aparece μ (que se busca estimar) y cuya distribución es una t_{n-1} (t de Student con $n-1$ grados de libertad) en donde todo es conocido.

El proceso de despeje sigue exactamente igual que en el caso de σ^2 conocido, trabajando con la distribución t de Student en lugar de la normal estándar. Repitiendo exactamente las mismas cuentas, concluimos que el intervalo de confianza para μ será

$$IC_{1-\alpha}(\mu) = \overline{X}_n \pm t_{n-1,1-\alpha/2}\sigma_0/\sqrt{n}$$

donde $t_{n-1,1-\alpha/2}$ es el percentil $1-\alpha/2$ de la distribución t de Student con $n-1$ grados de libertad.

La longitud del intervalo de confianza será la resta de los dos extremos, obteniendo así

$$LIC_{1-\alpha}(\mu) = 2t_{n-1,1-\alpha/2}s/\sqrt{n}.$$

Un cambio importante es que la longitud, a diferencia con el caso de σ conocido, es aleatorio, no se puede en principio prever de antemano.

Veamos un ejercicio en donde comparamos los dos tipos de intervalos de confianza.

Ejercicio 11.1. *Se tienen $n = 15$ parcelas en donde se cultiva trigo. El rinde en toneladas de cada parcela sigue una distribución $N(mu, \sigma^2)$, todos de forma independiente entre sí. Se busca estimar el rinde medio μ de las parcelas mediante un intervalo de confianza de nivel $1-\alpha = 0{,}95$.*

1. *Suponga conocido σ y que vale 3 toneladas. Se observó un rinde promedio de $\overline{X}_n = 20$ toneladas, estime por intervalo de confianza para μ.*

2. *Igual al ejercicio anterior pero ahora σ es desconocido y se estima por s, en donde se observó $s = 3$ toneladas como desvío estándar muestral. Bajo las mismas condiciones, observamos $\overline{X}_n = 20$ toneladas.*

Rta:

Para el primer ítem ya teníamos la expresión del intervalo de confianza, que es

$$IC_{1-\alpha}(\mu) = \overline{X}_n \pm z_{1-\alpha/2}\sigma_0/\sqrt{n} = 20 \pm 1{,}96 \cdot 3/\sqrt{15} = (18{,}48; 21{,}52).$$

Para el segundo inciso, tenemos que ahora

$$IC_{1-\alpha}(\mu) = \overline{X}_n \pm t_{15-1,1-\alpha/2}s/\sqrt{n}.$$

Buscando en la tabla t de Student, obtenemos que $t_{15-1,1-\alpha/2} = 2{,}14$. Luego, nos queda

$$IC = 20 \pm 2{,}14 \cdot 3/\sqrt{15} = (18{,}34; 21{,}66).$$

El intervalo es más ancho, lo que se traduce en una precisión inferior. Es esperable, en esta situación nos manejamos con más incertidumbre puesto que el desvío estándar fue desconocido y lo tuvimos que estimar.

11.6. Intervalo Normal para σ^2 con μ conocido

Consideremos ahora una muestra aleatoria $X_1, \ldots, X_n$ con distribución $N(\mu_0, \sigma^2)$, buscamos estimar σ^2 sabiendo conocido el $\mu = \mu_0$.

El pivote que usaremos estará basado en el siguiente estadístico:

$$\tilde{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2$$

y tomaremos como pivote

$$U = \frac{n\tilde{\sigma}^2}{\sigma^2} \sim \chi^2_n.$$

Repetiremos la deducción que hicimos en el primer caso para prestar atención a las diferencias.

Necesitamos A_1 y A_2 tales que que

$$1-\alpha = P(A_1 \leq U \leq A2).$$

Elegimos A_1 de forma tal que deje $\alpha/2$ área a su izquierda, es decir, tomamos $A_1 = \chi^2_{n,\alpha/2}$ y A_2 el que deja $\alpha/2$ área a su derecha, es decir, tomamos $A_2 = \chi^2_{n,1-\alpha/2}$.

Así pues,

$$1-\alpha = P\left(\chi^2_{n,\alpha/2} \leq \frac{n\tilde{\sigma}^2}{\sigma^2} \leq \chi^2_{n,1-\alpha/2}\right).$$

Tal como hacemos siempre, despejamos el σ^2 y nos queda que

$$1-\alpha = P\left(\frac{n\tilde{\sigma}^2}{\chi^2_{n,1-\alpha/2}} \leq \sigma^2 \leq \frac{n\tilde{\sigma}^2}{\chi^2_{n,\alpha/2}}\right)$$

O sea que el intervalo de confianza de nivel $1-\alpha$ para σ^2 será

$$IC = \left(\frac{n\tilde{\sigma}^2}{\chi^2_{n,1-\alpha/2}}; \frac{n\tilde{\sigma}^2}{\chi^2_{n,\alpha/2}}\right)$$

11.7. Intervalo Normal para σ^2 con μ desconocido

Bajo la situación de $X_1,\ldots,X_n$ muestra aleatoria con distribución $N(\mu,\sigma^2)$, siendo ambos parámetros desconocidos y buscando estimar σ^2, la cuenta es bastante parecida a la hecha en el inciso anterior. La diferencia es que en vez de $\tilde{\sigma}^2$ usamos s^2, con la consecuente pérdida del grado de libertad.

La deducción es exactamente igual y llegamos a

$$IC = \left(\frac{ns^2}{\chi^2_{n-1,1-\alpha/2}}; \frac{ns^2}{\chi^2_{n-1,\alpha/2}}\right)$$

11.8. Intervalos de nivel asintótico/aproximado

Hasta el momento hemos trabajo con muestras normales, cuando ese no es el caso la cuestión puede complicarse bastante a la hora de definir un pivote que sea útil. En este sentido, si nos conformamos con un poco menos se puede apelar a intervalos de confianza de nivel asintótico.

Dada una muestra aleatoria $X_1,\ldots,X_n$ y un parámetro θ de la distribución de los X_i que buscamos estimar, tenemos un intervalo de confianza de cierto nivel que no diremos cuál es para θ que llamaremos $I_n(\theta)$. Ahora bien, si se cumple que $P(\theta \in I_n(\theta)) \to 1-\alpha$ cuando n tiende a infinito, diremos entonces que los I_n conformarán una sucesión de intervalos de confianza de nivel asintótico $1-\alpha$ para θ.

Es decir, cada intervalo para un n fijo en sí no tiene necesariamente nivel $1-\alpha$, pero para n grande se supondrá que será aproximadamente $1-\alpha$. La herramienta natural para poder construir estos tipos de intervalos será naturalmente el Teorema Central del Límite.

Supongamos $X_1,\ldots,X_n$ muestra aleatoria, con media μ y varianza σ^2 pero no necesariamente con distribución normal. Si consideramos $\widehat{\sigma}^2$ un estimador consistente de la varianza σ^2, entendido eso como una función de la muestra en donde $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$ (por ejemplo, s^2 cumple esta condición), entonces vale que

$$\frac{\overline{X}_n - \mu}{\frac{\widehat{\sigma}}{\sqrt{n}}} \overset{a}{\sim} N(0,1).$$

De aquí podemos deducir un intervalo de confianza de nivel aproximado $1-\alpha$ deduciendo como se hacía para el caso de la normal, obteniendo

$$IC(\mu) = \overline{X}_n \pm z_{1-\alpha/2}\frac{\widehat{\sigma}}{\sqrt{n}}$$

11.8.1. Caso Poisson

Supongamos $X_1, \ldots, X_n$ una muestra aleatoria con distribución de Poisson de parámetro $\lambda > 0$. Buscamos obtener un intervalo de confianza para λ. Con el resultado de la sección anterior, tendríamos que

$$IC(\lambda) = \overline{X}_n \pm z_{1-\alpha/2}\frac{\widehat{\sigma}}{\sqrt{n}}$$

Ahora bien, la pregunta es: ¿qué usamos como $\widehat{\sigma}$? Podríamos usar el $\sqrt{s^2}$ y sería perfectamente válido, aunque usualmente se suele aprovechar el hecho de que $V(X_i) = \lambda$, con lo cual un estimador de σ sería $\sqrt{\overline{X}_n}$ dado que $\overline{X}_n \xrightarrow{p} E(X_i) = \lambda$ por la ley de los grandes números. Luego, $\sqrt{\overline{X}_N} \xrightarrow{p} \sqrt{\lambda} = \sigma$.

Así pues, el intervalo de confianza nos quedará:

$$IC(\lambda) = \overline{X}_n \pm z_{1-\alpha/2}\frac{\sqrt{\overline{X}_n}}{\sqrt{n}}$$

11.8.2. Intervalos para proporciones

Supongamos que tenemos $X_1, \ldots, X_n$ una muestra aleatoria de variables Bernoulli con parámetro p. Es decir, variables que pueden tomar el valor 0 o 1. Podría representar una encuesta electoral, en donde se interrogan a n individuos sobre su preferencia electoral. Podría valer 1 si efectivamente votarían por un candidato A y 0 en caso de que elijan a su oponente.

Quisiéramos armar un intervalo de confianza y ver si podemos concluir si el candidato A será efectivamente elegido a la hora de la elección, que sería sobre el total de la población.

Vale que $\mu = E(X_i) = p$, luego el intervalo de confianza de nivel aproximado $1 - \alpha$ será

$$IC(p) = \overline{X}_n \pm z_{1-\alpha/2}\frac{\widehat{\sigma}}{\sqrt{n}}.$$

¿Qué usamos en lugar de $\widehat{\sigma}$? Pensemos que $\sigma^2 = V(X_i) = p(1-p)$, y esto es estimable por $\overline{X}_n(1-\overline{X}_n)$. Así pues, tendremos que podemos usar como $\widehat{\sigma}$ a $\sqrt{\overline{X}_n(1-\overline{X}_n}$.

Luego, el intervalo de confianza nos queda

$$IC(p) = \overline{X}_n \pm z_{1-\alpha/2}\frac{\sqrt{\overline{X}_n(1-\overline{X}_n}}{\sqrt{n}}.$$

Veamos un ejemplo, supongamos que entrevistamos a 100 personas y obtenemos que 54 de ellos apoyarían al candidato A. ¿Tenemos suficiente evidencia estadística a nivel $1 - 0{,}05 = 0{,}95$ como para marcar una tendencia definitiva de triunfo para A? Veamos efectivamente eso armando el intervalo de confianza de nivel 0.95 aproximado para p, la verdadera proporción de la población que apoyaría a A.

$$IC(p) = \overline{X}_n \pm z_{1-\alpha/2}\frac{\sqrt{\overline{X}_n(1-\overline{X}_n}}{\sqrt{n}} = 0{,}54 \pm 1{,}96\frac{\sqrt{0{,}54(1-0{,}54)}}{\sqrt{100}}$$

Nos queda como IC el siguiente:

$$IC(p) = (0{,}44, 0{,}64).$$

Si bien la tendencia es favorable para el candidato A, todavía es muy temprano para marcar una tendencia ”irreversible”, puesto que tenemos una porción del intervalo debajo de la marca de 0.5.

Supongamos que aumentamos la encuesta, entrevistamos ahora a 10000 personas y obtenemos 5432 personas que votarían por A y el remanente que lo haría por su rival. Rehagamos el ejercicio con estos números:

$$IC(p) = \overline{X}_n \pm z_{1-\alpha/2}\frac{\sqrt{\overline{X}_n(1-\overline{X}_n}}{\sqrt{n}} = 0{,}5432 \pm 1{,}96\frac{\sqrt{0{,}5432(1-0{,}5432)}}{\sqrt{10000}}$$

Nos queda así como IC lo siguiente:

$$IC = (0{,}53; 0{,}55)$$

Con este intervalo de confianza estamos más confiados para afirmar algo más taxativamente de que A resultará el vencedor en la elección, puesto que ahora no tenemos ninguna fracción por debajo del empate marcado en 0.5.

11.9. Ejercitación

1. Se tiene una muestra de 50 botellas de un licor determinado, suponiendo que la graduación alcohólica del licor sigue una distribución $N(\mu, \sigma^2 = 2)$, con μ desconocido. Se busca estimar la graduación alcohólica media, μ, y se arma un intervalo de confianza de nivel exacto 0.95 obteniéndose como intervalo $IC = (8{,}3, 9{,}2)$.

   a) ¿Qué hubiese pasado con un intervalo de confianza de nivel 0.90? ¿Sería más o menos angosto que el original?

   b) ¿Es correcto afirmar de que la probabilidad de que μ caiga entre 8.3 y 9.2 es del 95$?

   c) Si en vez de 50 botellas consideramos 500 botellas, ¿el nuevo intervalo sería más o menos angosto?

2. Las medias de los diámetros de una muestra de 200 almohadones hechos por una determinada máquina durante un período dieron una media muestral de 1.625 centímetros y una desviación estándar muestral de 0.084 centímetros. Suponiendo distribución normal $N(\mu, \sigma^2)$, con ambos parámetros desconocidos, halle intervalos de confianza para la media poblacional μ de nivel $1 - \alpha = 95\,\%$ y $1 - \alpha = 99\,\%$.

3. Se efectúa un ensayo de reacción sobre una muestra de n pacientes. Se sabe, por registris históricos, que la distribución del tiempo de reacción sigue una distribución normal con media desconocida pero con desvío estándar $\sigma = 0{,}05$ segundos. Cuánto debe ser el tamaño muestral n para que la precisión (ancho del intervalo) sea seguro inferior a 0.01 segundos al armar un intervalo de confianza de nivel $1-\alpha = 0{,}97$?

4. Se saben que las califiaciones de un examen de matemática históricamente se distribuyeron de forma normal, con media desconocida y desvío estándar poblacional conocido, de valor 10. Se obtuvo una muestra de 200 calificaciones de un curso, obteníendose un puntaje promedio de 75.

   a) Obtenga un intervalo de confianza de nivel $1 - \alpha = 0{,}95$ para la medida de las calificaciones.

   b) ¿Con qué nivel de confianza se puede decir que la media poblacional es 75 ± 1?

5. Estando en el contexto previo a una elección, se desea diseñar un muestreo (es decir, determinar el tama no muestral n) para determinar la proporción verdadera de personas que aceptarían votar por un candidato dado. La metodología se basará en la construcción de un intervalo de confianza de nivel aproximado $1-\alpha = 0{,}95$ y se desea que la precisión del intervalo sea menor a 0,01. Determine el valor de n necesario.

6. Se realiza un estudio y se encuentra que 156 de 500 varones adultos son fumadores. Obtener un intervalo de confianza asintótico (aproximado) de nivel 99 % para la verdadera proporción de adultos varones que son fumadores.

12. Test de Hipótesis (EN CONSTRUCCIÓN

12.1. Introducción epistemológica

A fin de entender el funcionamiento de esta herramienta, será conveniente hacer un breve repaso de algunas nociones referidas a la metodología de la investigación. Propiamente, repasaremos algunas ideas referidas al desarrollo de la ciencia según el método hipotético deductivo, en una versión si se quiere simplificada.

Como habíamos establecido antes, el principio de la estadística radica en poder dar explicaciones de lo general a partir de lo particular. En tal sentido, es que trabajamos brevemente con estimación puntual y con más detalle con intervalos de confianza.

En el esquema clásico del desarrollo científico, basado en la inducción, los que eran conocidos como ”leyes generales” se deducían, según algunas reglas, a través de diversos casos particulares. El ejemplo paradigmático de esto es el siguiente:

- El oro se dilata con el calor.
- La plata se dilata con el calor.
- El hierro se dilata con el calor.

Conclusión: todos los metales se dilatan con el calor.

Los primeros serían lo que se suele llamar **consecuencias observacionales**, proposiciones que son verificables con la observación de casos particulares, mientras que la conclusión apuntaría a ser una **ley general**, ya que no se remite a un metal o caso particular, sino a todos los metales. Es una apreciación de tipo universal.

Lo importante a destacar es que este tipo de razonamiento inductivo es formalmente, a nivel de planteo deductivo, inválido. Nada quitaría quizás que en un futuro no se descubra un metal que no se dilate en el calor, no se desprende naturalmente la conclusión a partir de los antecedentes.

Bajo el método hipotético deductivo, nosotros trabajaremos con estas leyes generales, que llamaremos **hipótesis**, a partir del cual se deducen los casos particulares, como podrían ser las consecuencias observaciones.

Tendremos en efecto que

$$\text{Hipótesis} \rightarrow \text{Consecuencia observacional.}$$

La implicación (la ”flechita”) permite trasladar la verdad. Es decir, si la hipótesis resulta ser cierta necesariamente las consecuencias observacionales lo deberán ser. Entonces, si planteamos como hipótesis de que todos los metales se dilatan con el calor, necesariamente deberá ser cierta la consecuencia observacional de que el oro se dilata con el calor, por ser el oro un metal y por ser, según suponemos, cierta la hipótesis.

La implicación no traslada sin embargo la falsedad. No es cierto que si fuese falso la hipótesis entonces será falsa la consecuencia observacional. Podría existir un metal que no se dilate con el calor y pese a eso el oro, por una característica propia, dilatarse con el calor, en principio.

Sin embargo, lo que sí es cierto es que si bien la verdad ”desciende” por la flecha y no así la falsedad, la falsedad ”asciende” por la flecha. Eso quiere decir que basta encontrarse una consecuencia observacional que sea falsa para así descartar de lleno la hipótesis.

Por ejemplo, identificando un metal, cualquiera sea, que no se dilate con el calor ya podríamos **refutar**/rechazar la hipótesis.

Nunca podremos afirmar al 100 % que una hipótesis es verdadera, lo más que podemos hacer es lo siguiente:

1. Refutarla: en caso de que una consecuencia observacional sea falsa.
2. Corroborarla: en caso de que una consecuencia observacional resulte verdadera.

Nótese el cuidadoso empleo de los términos, no hablamos de ”verificar” la hipótesis sino de ”corroborarla”. Es decir, no tenemos suficiente evidencia como para rechazar a la hipótesis. Karl Popper, famoso filósofo de la ciencia, diría que la hipótesis ”ha mostrado su temple”.

En el aspecto epistemológico el desarrollo es bastante más elaborado, pues se permite la introducción de hipótesis auxiliares o *ad hoc* a fin de por ejemplo ”salvar” a una hipótesis de ser refutada ante el fallo de una consecuencia observacional, pero no nos meteremos en esos detalles.

¿Cómo se aplicaría esto en la estadística? Básicamente, el principio del **test de hipótesis** es refutacionista, ofrece herramientas para poder llevar a cabo el rechazo/refutación de una hipótesis sobre el cual se tienen dudas. El mecanismo consiste en suponer cierta la hipótesis a rechazar, recolectar datos y demostrar que, bajo el supuesto de esa hipótesis, probabilísticamente hablando lo observado (o algo más extremo que eso) es tan infrecuente o raro que no nos queda más remedio que descartar, razonablemente, la hipótesis.

Sin más, una vez aclarado este punto, vamos a desarrollar un ejemplo bien completo que permita capturar todos los elementos propios de la teoría de **test de hipótesis**.

12.2. Ejemplo introductorio: hipótesis puntuales

Supongamos una fábrica que produce bombones. Dispone para ello de dos máquinas, la primera de ellas, la A, produce bombones con una distribución dada por una normal de media $\mu = 100$ gramos y desvío estándar de $\sigma = 4$ gramos. Ahora bien, tiene otra máquina, la B, para producir bombones de una calidad inferior, es decir, siguiendo una distribución normal de media $\mu = 95$ gramos y el mismo desvío estándar de $\sigma = 4$ gramos.

Supondremos, a modo de simplificar el planteo, que solamente puede estar una máquina prendida a la vez, y que el capataz de la fábrica afirma que trabaja exclusivamente con la máquina A, según lo indicado frente a algún ente de regulación.

Un inspector sospecha de esta afirmación del capataz, y se plantea la **hipótesis** de que se está trabajando con la máquina B en vez de usar la máquina A.

Para eso, obtiene una muestra de 10 bombones (supuestamente fabricados con la máquina A) y obtiene que el peso promedio de estos 10 bombones es de 96.8 gramos.

La pregunta que cabe hacerse es si es razonable o no suponer que se está operando con la máquina A dado que se obtuvo una muestra de 10 bombones con ese peso promedio. Es decir, ¿se puede seguir afirmando que la media verdadera es de 100 gramos o bien quizás podría ser de 95?

Tenemos así dos situaciones (o hipótesis) que son las que siguen:

1. Hipótesis **nula** (H_0): la media es de 100 gramos ($\mu = 100$).
2. Hipótesis **alternativa** (H_1): la media es de 95 gramos ($\mu = 95$).

Ahora bien, el inspector sabe que si la hipótesis que plantea es cierta, lo correcto sería efectuar una acusación con las correspondientes consecuencias que eso implica (clausura, suspensión de actividades, multa, etc). Quiere estar realmente seguro de poder hacer esta afirmación con mucho aval estadístico de su lado. Alcanzará con una observación de 96.5 gramos con 100 bombones o es muy ”débil”?

La realidad es una de dos, o vale H_0 (la hipótesis nula, la que plantea la fábrica) o bien vale H_1 (la hipótesis alternativa, lo que sospecha el inspector).

El inspector puede tomar a su vez dos decisiones, rechazar la hipótesis nula y aceptar la alternativa H_1, o bien no rechazar la hipótesis nula, lo cual no quiere decir aceptarla sino que decide que no juntó suficiente evidencia estadística como para afirmar que H_1 es cierta.

Se presentan cuatro casos:

1. H_0 es verdadera y el inspector no decide rechazarla: en definitiva, se decide **corroborar** (no así aceptar) la hipótesis del fabricante, lo cual sería el paso correcto.
2. H_0 es falsa y el inspector decide rechazarla: en definitiva, el fabricante está produciendo bombones de peor calidad y así lo determinó el inspector, lo cual sería una decisión correcta.

3. H_0 es verdadera y el inspector decide rechazarla: casi el peor de los mundos posibles para el inspector, con todos los problemas que eso le acarreará. Esto se llama **error de tipo I** y es una de las cosas que se querrá controlar en términos de probabilidad. Querremos que las chances de que esto ocurra sea ”baja”.

4. H_0 es falsa y el inspector decide no rechazarla: es decir, deja al fabricante que siga produciendo bombones defectuosos. Esto conforma un **error de tipo II**.

A fin de salvaguardar la reputación del inspector y evitarle problemas, este quiere estar realmente seguro de rechazar con suficiente evidencia estadística la hipótesis nula en caso de sospechar que es falsa. Es decir, quiere mantener bajo control las chances de cometer un error de tipo I, de forma tal de que, como decíamos, la probabilidad de que ocurra sea baja.

Por supuesto que es importante también que el error de tipo II también tenga probabilidad baja, pero lamentablemente no podremos controlar ambos errores de forma simultánea con una muestra dada, pues son en cierta forma contrapuestos.

En el planteo clásico de test de hipótesis, el objetivo en general es controlar el error de tipo I y dejar que pase lo que tenga que pasar con el error de tipo II. Ya veremos más adelante cómo controlar a su vez este error.

El inspector tiene una muestra de 10 bombones y computó su peso promedio $\overline{X}_{10}$. Es claro que cuanto mayor sea este valor, más peso habrá para corroborar H_0, y cuanto menor sea más factible se vuelve el rechazo de la misma.

Entonces, un primer criterio de rechazo de la hipótesis nula podría ser el siguiente:

$$\overline{X} < K.$$

¿Cómo determinamos ese valor de K? Entran en juego varias consideraciones acá, que tienen que ver con el nivel de seguridad que queremos tener respecto a las chances de rechazar incorrectamente la hipótesis nula cuando esta es verdadera. Es decir, queremos controlar el error de tipo I.

Supongamos establecido que el inspector decide admitir hasta un $\alpha = 0{,}05$ de probabilidad de cometer un error de tipo I. A este valor α se lo llama el **nivel del test**. Es decir, está dispuesto a permitir hasta ese nivel de riesgo para procesar incorrectamente al fabricante.

Ahora, la cosa es más sencilla, podemos entonces plantear:

$$0{,}05 = \alpha = P(\text{Error de tipo 1}) = P_{H_0}(\text{rechazar } H_0) = P_{H_0}(\overline{X} < K)$$

donde debemos entender que P_{H_0} es la probabilidad en el caso en que H_0 sea verdadera. También lo podemos escribir como $P_{\mu=100}$, para poner en evidencia que el valor verdadero en este caso de $\mu = \mu_0$ es 100.

En general, cuando trabajamos con distribuciones normales el camino típico consiste en estandarizar, con lo cual trabajaremos en realidad con la versión estandarizada de $\overline{X}$, que será:

$$T = \sqrt{n}\frac{\overline{X} - \mu_0}{\sigma_0} = \sqrt{10}\frac{\overline{X} - 100}{4}$$

que, bajo la hipótesis nula, tendrá distribución $N(0, 1)$. A este T se lo llama **estadístico del test**. Así pues, tenemos que

$$0{,}05 = \alpha = P_{H_0}(\overline{X} < K) = P_{H_0}\left(\sqrt{n}\frac{\overline{X} - \mu_0}{\sigma_0} < \sqrt{n}\frac{K - \mu_0}{\sigma_0}\right) = P_{H_0}\left(\sqrt{n}\frac{\overline{X} - \mu_0}{\sigma_0} < K'\right) = P(Z < K')$$

donde K' es la versión ”estandarizada” de K, que volveremos a llamar K para no sobrecargar los índices. Z es una variable aleatoria con distribución normal estándar bajo el supuesto de H_0, la hipótesis nula que estamos poniendo bajo sospecha.

Entonces, debemos elegir un K de forma tal que

$$0{,}05 = \alpha = P(Z < K)$$

de donde, consultando la tabla de la normal estándar, nos sale que $K = -1{,}65$ aproximadamente.

Entonces, para este test con un nivel de $\alpha = 0{,}05$, la regla de rechazo de la hipótesis nula de forma tal de garantizar un error de tipo I con probabilidad 0.05 será

$$\text{rechazar } H_0 \text{ si } T = \sqrt{n}\frac{\overline{X} - \mu_0}{\sigma_0} < -1{,}65.$$

Evaluemos en nuestro caso el estadístico del test. Tenemos que $\overline{X}$ terminó resultando 96.5, $\mu_0 = 100$, $\sigma_0 = 4$ y $n = 50$. Evaluando todo eso, obtenemos un $T_obs = -2{,}52$. Como es menor que -1.65, puede garantizar, al nivel deseado, de que hay suficiente evidencia estadística como para suponer de que se está utilizando la máquina B en lugar de la A y que el resultado observado es poco probablemente atribuible al azar. Probablemente esté bien justificado, con aval estadístico, iniciar un procesamiento al fabricante, la evidencia respalda al inspector.

Si el nivel es α, en general quedará que el rechazo se dará cuando T_{obs} sea menor que z_α, siendo esto último el percentil α de una distribución normal estándar.

Definimos la **región de rechazo** del test como el conjunto del intervalor real en donde debe caer T_{obs} para que se rechace la hipótesis nula. En nuestro caso, la región de rechazo será $(-\infty, -1{,}65)$.

Por ejemplo, a nivel $\alpha = 0{,}05$ ya vimos que es -1.65, consultando en la tabla. A nivel $\alpha = 0{,}01$, el z_α será -2.33. Todavía podemos rechazar incluso a ese nivel la hipótesis nula. La determinación de qué α, o sea, nivel del test, se debe utilizar es subjetivo y muy dependiente de la disciplina. Valores usuales son $\alpha = 0{,}05$ o bien $\alpha = 0{,}01$, al igual que los que típicamente se usan en intervalos de confianza.

Nos preguntamos entonces cuál es el mínimo nivel para el cual se rechazaría la hipótesis nula. Es lo que veremos a continuación, lo que se pasa a denominar el **p-valor**.

12.2.1. p-valor

Hasta ahora, hemos trabajado exclusivamente con un nivel de test $\alpha = 0{,}05$. A veces, no se encuentra determinado de antemano e interesa saber, dada una observación ya realizada, cuál es el mínimo nivel del test para el cual se rechazaría la hipótesis nula. Este valor se denomina **p-valor** del test.

El p-valor depende directamente del valor observado del estadístico de test, T_{obs}. Tenemos que efectivamente, con el $T_obs = -2{,}52$ rechazaríamos la hipótesis nula a un nivel de $\alpha = 0{,}05$. Cabe la pregunta de si podríamos rechazarlo con ”mayor seguridad”, es decir, con un nivel aún menor. Eso es lo que refleja el p-valor, el mínimo nivel para el cual todavía rechazaríamos la hipótesis nula.

Recordemos que la región de rechazo está determinado por $(-\infty, z_\alpha)$. Si vamos poniéndonos cada vez más estrictos con el test, lo que hacemos es reducir el valor del α. Eso a su vez va reduciendo el valor de z_α hasta que llega un punto en donde coincide con T_{obs}. Más allá de eso no lo podría reducir más puesto que dejaría de rechazar la hipótesis nula.

Es decir, el p-valor se dará en el α tal que $T_obs = z_\alpha$. Es decir, que $P(Z < T_{obs}) = \alpha$. Es decir, el p-valor representaría la probabilidad de obtener un valor más extremo todavía que el que se obtuvo, bajo el supuesto de H_0, la hipótesis nula.

En nuestro caso, el T_{obs} era -2.52. Así, el p-valor será $P(Z < -2{,}52)$, y esto vale aproximadamente 0.0059. Ese será el mínimo valor de α (el nivel del test) para el cual rechazaremos todavía la hipótesis nula. Es decir, si definimos un test de nivel por ejemplo 0.006, rechazaremos la hipótesis nula con el valor observado. Ahora bien, si lo configuramos en 0.0058, o cualquier otro valor menor que 0.0059, no podremos rechazar la hipótesis nula.

12.2.2. Potencia y error de tipo II

La mecánica de test de hipótesis permite mantener controlado el error de tipo I, por ejemplo, a nivel $\alpha = 0{,}05$. Nos preguntamos ahora sobre otro concepto importante de un test de hipótesis, que se llama la **potencia** de un test, definido como $\pi(\mu)$, que es una función de de los posibles valores de μ. En nuestro caso, los posibles valores son $\mu = 100$ (zona de la hipótesis nula) y $\mu = 95$ (zona de la hipótesis alternativa).

Vamos a hacer igualmente la cuenta en genérico, para cualquier μ, pues esto lo usaremos posteriormente cuando trabajemos con otros tipos de hipótesis.

Tenemos que

$$\pi(\mu) = P_\mu(\text{rechaza la hipótesis nula}) = P_\mu(T < z_\alpha) = P_\mu\left(\sqrt{n}\frac{\overline{X} - \mu_0}{\sigma} < z_\alpha\right).$$

Recordemos que como estamos bajo P_μ, eso quiere decir que el verdadero valor de la media de la normal es μ, que no necesariamente es $\mu_0 = 100$, con lo cual no tenemos en principio a T de forma estandarizada. Debemos hacer una ligera manipulación algebraica sobre el mismo a fin de que quede otra vez una normal estándar.

Así pues,

$$P_\mu\left(\sqrt{n}\frac{\overline{X} - \mu_0}{\sigma} < z_\alpha\right) = P_\mu\left(\sqrt{n}\frac{\overline{X} - \mu + \mu - \mu_0}{\sigma} < z_\alpha\right) = P_\mu\left(\sqrt{n}\frac{\overline{X} - \mu}{\sigma} < z_\alpha + \sqrt{n}\frac{\mu_0 - \mu}{\sigma}\right)$$

Ahora sí tenemos, del lado izquierdo de la desigualdad, algo que está correctamente estandarizado, con lo cual pasará a ser una Z con distribución normal estándar. Así pues, nos queda que

$$\pi(\mu) = P\left(Z < z_\alpha + \sqrt{n}\frac{\mu_0 - \mu}{\sigma}\right) = \Phi\left(z_\alpha + \sqrt{n}\frac{\mu_0 - \mu}{\sigma}\right).$$

Algunas observaciones, si evaluamos la potencia en $\mu = \mu_0$, notemos que sale $\pi(\mu_0) = \Phi\left(z_\alpha + \sqrt{n}\frac{\mu_0 - \mu_0}{\sigma}\right) = \Phi(z_\alpha) = \alpha$, lo cual tiene sentido. La potencia representa la probabilidad de rechazar la hipótesis nula, evaluado en un valor μ que corresponde a la hipótesis nula sería en efecto un error de tipo 1 (pues estamos incorrectamente rechazando la hipótesis nula), que ya sabíamos que estaba controlado por α.

Otra observación, es una función decreciente en μ (en este caso). Cuanto mayor es el μ (más cerca de la región de la hipótesis nula), más improbable será el rechazo de H_0.

Notemos también que a medida que $\mu \to -\infty$ (es decir, se aleja de la región nula), entonces la potencia tiende a 1, pues lo que está entre paréntesis de la Φ tiende a $+\infty$. O sea, cuanto más desvíada esté la máquina B, más fácil será detectar su utilización.

Por último, y no menos importante, cuando $n \to \infty$, la potencia también tenderá a 1 si $\mu < \mu_0$. Es decir, el tamaño muestral ayuda a poder detectar más fácilmente desvíos con respecto a la hipótesis nula.

El Error de Tipo II (que se suele notar como β) habíamos dicho que era la probabilidad de no rechazar la hipótesis nula cuando esta es falsa. Es decir, la probabilidad de no detectar que se está usando la máquina de peor calidad cuando en realidad se la está usando. Esto vale $1 - \pi(\mu = 95)$. O sea, el complemento de la probabilidad de rechazar bajo la región de la hipótesis alternativa.

Nos quedará

$$\beta = 1 - \pi(95) = 1 - \Phi\left(z_\alpha + \sqrt{n}\frac{100 - 95}{\sigma}\right).$$

Consideremos con $n = 10$ y $\sigma = 4$ lo que nos devuelve esto. Quedará 1 - *Phi*(2,30) que es 0.01072. O sea, con este tamaño muestral el test parece los bastante potente como para poder reducir a valores bajos (pero nunca cero) el error de tipo II.

Una pregunta que podríamos hacernos es qué tamaño muestral necesitaría para asegurar, a un nivel α dado de por ejemplo, un nivel determinado de error de tipo 2. Por ejemplo, queremos saber el tama no muestral necesario para poder asegurar una probabilidad de error de tipo 2 menor a 0.005.

Hagamos la cuenta. Queremos que $0{,}005 > \beta = 1 - \pi(\mu = 95) = 1 - \Phi\left(z_\alpha + \sqrt{n}\frac{100-95}{\sigma}\right)$. De aquí se despeja que

$$\Phi\left(z_\alpha + \sqrt{n}\frac{100-95}{\sigma}\right) > 0{,}995.$$

O sea,

$$-1{,}65 + \sqrt{n}\frac{100-95}{4} > \Phi^{-1}(0{,}995) = 2{,}58.$$

De aquí despejamos n, y nos queda

$$n \geq \left(\frac{4,23 \cdot 4}{5}\right)^2 = 11{,}45$$

Eso quiere decir que con $n = 12$ podríamos garantizar el requerimiento solicitado.

12.3. Hipótesis compuestas

Esta representa una situación más normal del problema, como en general se lo suele hallar en los textos de estadística.

Supongamos ahora la siguiente situación. No tenemos dos máquinas, sino una sola pero que tiene una forma de "regular" el valor medio del peso de los bombones. Es decir, que tenemos que ahora el μ es totalmente libre. Como antes, el fabricante afirma que el valor de μ con el que produce es de $\mu = 100$, pero ahora lo que está en tela de juicio es que el μ es menor a 100, aunque sin necesariamente especificar un valor específico como antes, que era el $\mu = 95$.

Así pues, las hipótesis en esta situación serían

$$H_0 : \mu = \mu_0 = 100$$

vs

$$H_1 : \mu < \mu_0 = 100.$$

La construcción del test es exactamente igual que antes, se puede deducir que se usará el mismo estadístico de test $T = \sqrt{n}(\overline{X} - \mu_0)/\sigma$, siendo $\mu_0 = 100$ (en este caso) y siendo la región de rechazo, como antes, el percentil α de la normal estándar.

Es decir, se rechaza H_0 en el caso en que $T_{obs} < \Phi^{-1}(\alpha)$.

El p-valor se calcula idénticamente, como $P(Z < T_{obs}) = \Phi(T_{obs})$.

La potencia también, que será $\pi(\mu) = \Phi\left(z_\alpha + \sqrt{n}\frac{\mu_0-\mu}{\sigma}\right)$.

Notemos que no hay un único error de tipo II, como antes era con $\mu = 95$, sino que ahora $\beta(\mu)$ será una función de μ. Como antes, $\beta(\mu) = 1 - \pi(\mu)$, de donde nos sale que

$$\beta(\mu) = 1 - \pi(\mu) = 1 - \Phi\left(z_\alpha + \sqrt{n}\frac{\mu_0 - \mu}{\sigma}\right).$$

También se puede demostrar que no hay cambio alguno si ahora las hipótesis a considerar son

$$H_0 : \mu \geq \mu_0$$

vs

$$H_1 : \mu < \mu_0$$

Es decir, desigualdad también en la hipótesis nula.

Análogamente, podemos pensar una situación de test de hipótesis unilateral a derecha. Es decir,

$$H_0 : \mu \leq \mu_0$$

vs

$$H_1 : \mu > \mu_0.$$

Nuevamente, el estadístico del test será

$$T = \sqrt{n}(\overline{X} - \mu_0)/\sigma$$

con una región de rechazo dada por

$$T_{obs} > \Phi^{-1}(1 - \alpha),$$

un p-valor dado por

$$p - valor = P(Z > T_{obs}) = 1 - \Phi(T_{obs})$$

y una función de Error de tipo II dado por

$$\beta(\mu) = \Phi\left(z_\alpha + \sqrt{n}\frac{\mu - \mu_0}{\sigma}\right).$$

Finalmente, nos queda analizar un último caso, que es un poco más elaborado, que es el que se conoce como **hipótesis bilateral**.