Bioestadística (Clase 1):

Probabilidad y Conteo

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Experimentos aleatorios

- Nos interesa un experimento producido bajo una determinada cantidad de condiciones
- El resultado de este experimento podría tener un único resultado (ejemplo: enfriar agua a cero grados a presión atmosférica normal tiene como resultado el congelamiento del agua)
- En algunos casos, los resultados posibles forman una lista, y a priori no hay forma de saber cuál será el resultado del experimento (ejemplo: tirar un dado)
- En el segundo caso estamos en presencia de un experimento aleatorio

Probabilidad

- Listamos el conjunto de resultados posibles en un conjunto Ω[1], que llamamos espacio muestral. Ejemplo:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- El número de resultados posibles en este caso en $N = 6$
- Cuando todos los resultados de un experimento aleatorio son equivalentes, simétricos, le asignamos a cada uno una probalidad de $1/N$, en el ejemplo cada resultado simple tiene probabilidad 1/6.
- Tenemos además sucesos compuestos por varios resultados.

[1] Legra griega mayúscula: Omega

- Por ejemplo el suceso A consiste en ejemplo en obtener un resultado par. Entonces

$$A = \{2, 4, 6\}$$

- En ese caso para calcular la probabilidad tenemos que contar los elementos del suceso A y los del espacio Ω, y definimos la probabilidad del suceso A mediante

$$P(A) = \frac{\text{cantidad de elementos de } A}{\text{cantidad de elementos de } \Omega}$$

- Esta definición se parafrasea como casos favorables sobre casos posibles y fue dada por P. S. de Laplace en 1812.
- En nuestro ejemplo A tiene 3 elementos, y Ω tiene 6, entonces

$$P(A) = \frac{3}{6} = \frac{1}{2}.$$

Ejemplo: los dos dados

- Suponemos ahora que tiramos un dado **negro** y un dado rojo:

- ¿Cuántos resultados posibles tenemos?
- ¿Cuál es el espacio muestral Ω?

Espacio muestral del experimento de tirar dos dados:

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Cuadro: Todos los posibles resultados de tirar 2 dados donde uno tiene los números en rojo y otro en negro.

- Para calcular[2] la cantidad de resultados de Ω multiplicamos

$$6 \times 6 = 36$$

[2] Se aplica aquí la regla del producto

En este espacio nos interesa calcular probabilidades de algunos sucesos:

- Probabilidad de que haya algún resultado 1:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Cuadro: En azul se indican los casos favorable: ¿cuántos son?

$$P(\text{algún } 1) = \frac{11}{36}$$

¿Apostarías a que sale algún uno?

- Probabilidad de la suma sea 6:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Cuadro: En azul se indican los casos de suma 6: ¿cuántos son?

$$P(\text{suma } 6) = \frac{5}{36}$$

¿Cuál es la suma mas probable?

Conteo: regla del producto

- Como vimos, para calcular probabilidades tenemos que contar los resultados de los sucesos que nos interesan.
- Para eso se desarrollan algunas técnicas que nos ayudan en situaciones complejas
- La regla del producto se aplica cuando tenemos que los resultados dependen de dos características, por ejemplo contar los resultados de tirar dos dados: 6×6
- Si queremos contar los números pares de dos cifras, con números del 0 al 6: un números de esos es

$$C_1 \; C_2$$

donde $C_1 = 1, 2, 3, 4, 5, 6$ y $C_2 = 0, 2, 4, 6$. Tenemos entonces $6 \times 4 = 24$ casos.

Problema: Calcular la probabilidad del suceso A consistente en obtener un número par de dos cifras con números del 0 al 6 al elegirlo al azar entre todos los números de dos cifras[3].

- Calculamos el numerador (casos favorables) que era 24
- Cuantos números hay en Ω?
- Tenemos ahora

  $$C_1 \, C_2$$

  donde $C_1 = 1, \ldots, 9$ y $C_2 = 0, \ldots, 9$. Tenemos entonces $9 \times 10 = 90$ casos posibles.

  $$P(A) = \frac{24}{90} = \frac{4}{15}$$

- ¿Cuántas veces aplicamos la regla del producto?

[3] Corregido del original

Problema de ADN:

¿Cuantas tiras de letras distintas obtenemos poniendo las letras A, T, C, G en N lugares?

- La primer letra puede ser cualquiera de cuatro
- La segunda también: $4 \times 4 = 16$
- La tercera daría $4 \times 4 \times 4 = 4^3 = 64$
- Si fuesen N lugares tenemos

$$\underbrace{4 \times 4 \times \cdots \times 4}_{N} = 4^N.$$

- ¿Cuánto vale N en un problema de biología humana?
- Los cromosomas humanos tienen entre N=50 000 000 a N=300 000 000 pares de bases

Conteo: permutaciones

Este es nuestro problema:

- ¿De cuantas formas distintas pueden sentarse 20 estudiantes en 20 sillas?
- Como parece un poco difícil, empezamos con dos estudiantes, digamos A (Alicia) y B (Beto):

- Son dos formas:

$$A\,B, \qquad B\,A.$$

Llega Carlos. Tenemos para cada una de las dos formas anteriores, las siguientes tres posibilidades:

$$C\ A\ B, \qquad A\ C\ B, \qquad A\ B\ C.$$

$$C\ B\ A, \qquad B\ C\ A, \qquad B\ A\ C.$$

Entonces, tenemos tres formas (izquierda, centro y derecha) para cada una de las dos primeras posibilidades:

$$3 \times 2 = 6.$$

- ¿Qué regla estamos aplicando?
- ¿Qué ocurre cuando llega Daniela?

- Tenemos seis formas de sentarse tres estudiantes, supongamos que una de ella es $X\ X\ X$ (no importa cual)
- ¿En cuantos lugares distintos se puede sentar Daniela?

$$D\ X\ X\ X \qquad X\ D\ X\ X \qquad X\ X\ D\ X \qquad X\ X\ X\ D$$

- Aplicando la regla del producto, tenemos

$$4 \times 6 = 24$$

  formas distintas de que se sienten cuatro estudiantes.
- Además, la cuenta total que hicimos es

$$4 \times 3 \times 2 = 24.$$

- Al llegar Emilio, tenemos

$$5 \times 24 = 120$$

formas de sentarse.

- Y para nuestros **veinte** estudiantes tenemos

$$20 \times 19 \times 18 \times \cdots \times 3 \times 2 = 479001600 = 4{,}8 \times 10^7$$

- Unos quinientos millones: 500 000 000.

Definición. Llamamos permutaciones de n elementos a la cantidad de formas de ordenar n elementos en una lista.

- Las permutaciones de n se designan mediante el símbolo $n!$ (se lee n factorial)
- Según calculamos
$$2! = 2, \qquad 3! = 6, \qquad 4! = 24,$$
- Mas en general
$$n! = n \times (n-1) \times \cdots \times 3 \times 2.$$
- Además se cumple
$$n! = n \times (n-1)!$$

Si uno pregunta la solución de un problema,

el conocimiento NO permanece.

Es como si uno lo hubiera pedido prestado.

En cambio, si lo piensa uno,

es como haberlo adquirido para siempre.

ADRIAN PAENZA

La clase de hoy la dedicamos a:

Adrián Paenza y Emanuel (Manu) Ginóbili.

Bioestadística (Clase 2):

Combinaciones

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Permutaciones (repaso)

Definición. Llamamos permutaciones de n elementos a la cantidad de formas de ordenar n elementos en una lista.

- Las permutaciones de n se designan mediante el símbolo $n!$ (se lee n factorial)
- Definimos $0! = 1$ (para darle coherencia a las notaciones)
- Calculamos la clase pasada que:

$$n! = n \times (n-1) \times \cdots \times 3 \times 2.$$

- Además se cumple

$$n! = n \times (n-1)!$$

Aplicación

- ¿De cuantas formas distintas pueden sentarse 2 estudiantes en 2 sillas, digamos A (Alicia) y B (Beto)?:

- Son dos formas:

$$A\,B, \qquad B\,A.$$

Si son por ejemplo Alicia, Beto, Carlos, Daniela y Emilio (**5** estudiantes) tenemos

$$5! = 5 \times 4 \times 3 \times 2 = 120$$

formas distintas de sentarse en 5 sillas.

Combinaciones

- El 23 de agosto de 2004, Ruben Magnano, entrenador de la selección argentina de Basketball no podía dormir.
- Al día siguiente, debía enfrentar en la final por la medalla de oro en Atenas, al equipo de Estados Unidos.
- No se decidía por la formación inicial: tenía que elegir 5 jugadores entre 12.
- ¿Tendría que integrar al inicio a Manu Ginóbili, o hacerlo jugar de 6 como en San Antonio Spurs?
- ¿Cuántos equipos distintos podría formar?

- Como sabía permutaciones, imaginó ordenados sus 12 jugadores en fila. Sabía que eso podía hacerse de 12! formas distintas.

- Luego imaginó que los 5 primeros jugadores de la fila serían el equipo titular, y los últimos 7 los suplentes

- Entonces pensó que habían 12! equipos diferentes.
- Como no conciliaba el sueño, se dió cuenta que muchas de esas permutaciones producían el mismo equipo, los equipos posibles diferentes eran menos. ¿pero cuántos menos?
- Y se dió cuenta que el mismo equipo inicial aparecía 5! veces . . . tenía que dividir 12! entre 5!

- Pero lo mismo pasaba con los suplentes, que eran 7. Tenía que dividir también por 7!
- Llegó a la conclusión de que tenía

$$E = \frac{12!}{5!7!} = 792$$

equipos posibles.

- De otra forma:
  - Sea E la cantidad de equipos posibles
  - Si multiplicamos E por $5!$ obtenemos todos los ordenes posibles de los titulares.
  - Si multiplicamos el resultado por $7!$ obtenemos además todos los posibles ordenes de los suplentes.
  - El resultado es $E \times 5! \times 7!$ produce todas las $12!$ ordenaciones posibles de todos los jugadores (digamos que todas las fotos posibles)

  Despejando
  $$E \times 5! \times 7! = 12!$$
  obtenemos la fórmula.

Combinaciones

Definición. Llamamos combinaciones de n elementos tomados de a k a la cantidad de subconjuntos posibles de k elementos de un conjunto de n elementos. Ese número lo escribimos como $\binom{n}{k}$.

Observaciones

- Un argumento general como al anterior nos permite afirmar que

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

- Se verifica además

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

(es lo mismo elegir los titulares que elegir los suplentes)

Paseo al azar y triangulo de Pascal

- Consideremos una partícula que se desplaza, una unidad para arriba o para abajo en cada instante de tiempo:
- Llamamos paseo al azar a cada posible trayectoria.

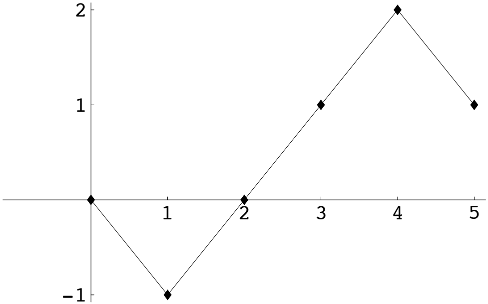
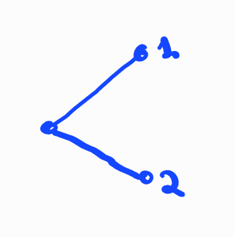


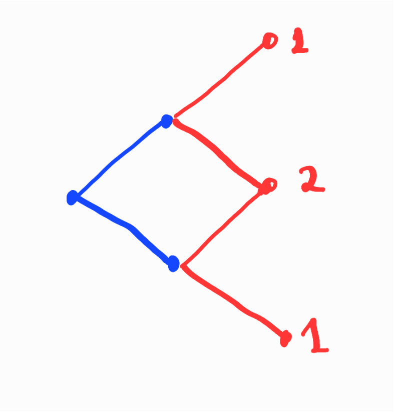
Figura: Una trayectoria del paseo al azar

Contemos la cantidad de paseos posibles en n pasos

- Si tenemos $n = 1$ paso son 2 caminos posibilidades.

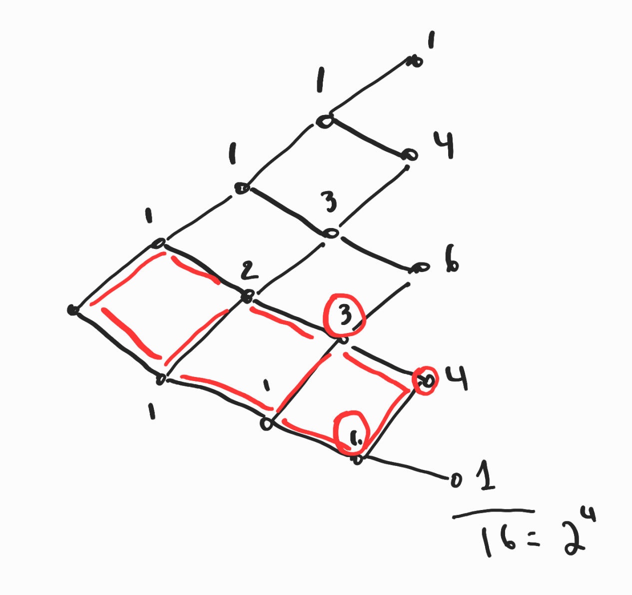


- Si tenemos $n = 2$ (dos pasos)...



- Tenemos 3 destinos finales, al primero llega un camino, al segundo 2, al tercero 1, en total 4 caminos.

- Si tenemos cuatro pasos:



- En síntesos, en en n pasos la particula puede recorrer 2^n caminos diferentes
- La idea es que en cada paso la partícula sube o baja, y se aplica la regla del producto.
- Supongamos que queremos numerar cada uno de estos caminos:
    - Cada camino tienen un número de n dígitos
    - Si en el paso sube ponemos 1, si en el paso baja, ponemos un 0.

- Por ejemplo si $n = 4$
  - el número 1111 corresponde al camino que sube las 4 veces
  - el número 0000 corresponde al camino que baja las 4 veces
  - el número 0101 corresponde al que baja-sube-baja-sube
- Es claro que la cantidad de tiras posibles de 0 y 1 corresponde a la cantidad de caminos posibles.
- Tenemos entonces 2^n tiras de ceros y unos.

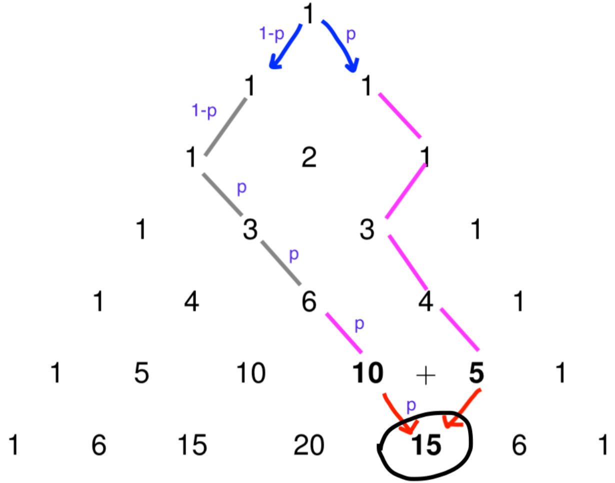
Nos interesa calcular la cantidad de caminos diferentes que conducen a una altura dada en una cantidad de pasos n

- Si un camino tiene la misma cantidad de ceros y unos, llegará a la misma altura.
- Nuestro problema es contar de cuantas formas distintas se pueden disponer k unos en n lugares:

$$\underbrace{\underline{0}\ \underline{0}\ \underline{1}\ \underline{0}\ \underline{1}\ \underline{0}\ \underline{1}}_{7 \text{ lugares}}$$

$$C^{7}_{3} = \frac{7!}{3!\,4!} = 35$$

Si representamos estas cantidades en una tabla obtenemos el triángulo de Pascal:



- La regla de formación de este triángulo es que cada línea se obtiene de la superior sumando los dos números que se encuentran arriba a la derecha y a la izquierda
- Por ejemplo $15 = 5 + 10$.
- Estos números cuentan la cantidad de caminos posibles por los que la partícula llega a las diferentes alturas en n pasos, que son $\binom{n}{k}$
- Como la cantidad total de caminos que puede seguir la partícula es 2^n, y la cantidad de subidas posibles es $0, 1, 2, \dots, n$, obtenemos

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{k} + \cdots + \binom{n}{n} = 2^n$$

- Por ejemplo

$$1 + 6 + 15 + 20 + 15 + 6 + 1 = 2^6 = 64.$$

- Además vemos nuevamente que

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

porque los caminos que suben k veces son la misma cantidad que los que bajan k veces

- Obtenemos además la regla de formación:

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

llamada fórmula de Stieffel

Bioestadística (Clase 3):

Propiedades de la probabilidad, probabilidad condicional

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Experimentos aleatorios

- Nos interesa un experimento producido bajo una determinada cantidad de condiciones
- Cuando los resultados posibles forman una lista, y a priori no hay forma de saber cuál será el resultado del experimento (ejemplo: tirar un dado) estamos en presencia de un experimento aleatorio.
- Ejemplo: tirar un dado.

Probabilidad

- Listamos el conjunto de resultados posibles en un conjunto Ω, que llamamos espacio muestral. Ejemplo:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- El número de resultados posibles en este caso en $N = 6$
- Cuando todos los resultados de un experimento aleatorio son equivalentes, simétricos, le asignamos a cada uno una probalidad de $1/N$, en el ejemplo cada resultado simple tiene probabilidad 1/6.
- Tenemos además sucesos compuestos por varios resultados.

- Por ejemplo el suceso A consiste en ejemplo en obtener un resultado par. Entonces

$$A = \{2, 4, 6\}$$

- En ese caso para calcular la probabilidad tenemos que contar los elementos del suceso A y los del espacio Ω, y definimos la probabilidad del suceso A mediante

$$\boxed{P(A) = \frac{\text{cantidad de elementos de } A}{\text{cantidad de elementos de } \Omega} = \frac{|A|}{|\Omega|}}$$

- Esta definición se parafrasea como casos favorables sobre casos posibles y fue dada por P. S. de Laplace en 1812.

- En nuestro ejemplo A tiene 3 elementos, y Ω tiene 6, entonces

$$P(A) = \frac{3}{6} = \frac{1}{2}.$$

Observaciones

- La probabilidad es no negativa y menor o igual que 1. Para cualquier suceso A tenemos

$$0 \leq P(A) \leq 1.$$

- Como el conjunto vacío $\emptyset$ es un subconjunto de Ω sin elementos, tenemos que su probabilidad es nula, es decir $P(\emptyset) = 0$. Es por ésto que lo llamamos suceso imposible.
- Por otra parte obtenemos que $P(\Omega) = 1$, por lo que llamamos suceso seguro al espacio de sucesos elementales Ω.

Propiedades de la probabilidad

- La unión de los sucesos A y B son los puntos que componen A o B.
- La intersección de los sucesos A y B son los puntos que componen A y B.
- Dados A y B, designamos mediante $A \cup B$ su unión, y $A \cap B$ a su intersección.
- Cuando los conjuntos no contienen puntos comunes, decimos que son sucesos disjuntos. Escribimos en este caso
$$A \cap B = \emptyset.$$

Ejemplo. Tiremos un dado, y designemos

- A el suceso de obtener una cantidad par de puntos;
- B una cantidad múltipo de tres.
- El suceso $A = \{2, 4, 6\}$, mientras que $B = \{3, 6\}$. Por eso

$$A \cup B = \{2, 3, 4, 6\}.$$

- El resultado 6 aparece en ambos sucesos, no son disjuntos:
- La intersección de A y B es el suceso 6, es decir

$$A \cap B = \{6\}.$$

Monotonía

- Si un suceso B tiene todos los puntos de otro A y eventualmente más, decimos que B contiene a A y escribimos

$$A \subset B.$$

- Como tiene más puntos (o la misma cantidad), tenemos $|A| \leq |B|$.
- De aquí tenemos

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

Probabilidad de la unión de sucesos

Propiedad. Si la unión de dos sucesos A y B es disjunta, su probabilidad es la suma de las respectivas probabilidades. Es decir, si $A \cap B = \emptyset$, tenemos

$$\boxed{P(A \cup B) = P(A) + P(B).}$$

Demostración. Si A y B no tienen puntos en común, se tiene

$$|A \cup B| = |A| + |B|.$$

Por eso

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

$\square$

¿Qué pasa con tres conjuntos?

- Supongamos que A, B y C no tienen puntos en común.
- Consideramos $D = B \cup C$ que no tiene puntos en común con A. Eso nos da

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

- Además

$$P(D) = P(B) + P(C)$$

- Si sustituimos la última igualdad, tenemos

$$\boxed{P(A \cup B \cup C) = P(A) + P(B) + P(C).}$$

Partición de un conjunto

- Supongamos que Ω se compone de n conjuntos que no tienen puntos en común, llamados $A_1, \ldots A_n$.
- Estamos diciendo que
$$\Omega = A_1 \cup A_2 \cup \cdots \cup A_n.$$
- y que
$$A_i \cap A_j = \emptyset \quad \text{siempre que } i \neq j$$
- Como $P(\Omega) = 1$, un razonamiento análogo nos da
$$P(A_1) + P(A_2) + \cdots + P(A_n) = 1.$$

Probabilidad del complemento de un suceso

- El complemento de un suceso A contiene todos los puntos que no están en A,
- Lo designamos A^c.
- Tenemos
$$\Omega = A \cup A^c.$$
- Además, son disjuntos, no tienen puntos en común. Entonces, por la propiedad anterior
$$1 = P(\Omega) = P(A) + P(A^c)$$
- Despejando
$$\boxed{P(A^c) = 1 - P(A)}$$
- A y A^c son una partición de Ω.

Probabilidad condicional

Definición. Dados dos sucesos A y B, con $P(B) > 0$ definimos la probabilidad condicional de A dado B, mediante

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

Nota: Precisamos que $P(B) > 0$ porque aparece en el denominador

Es útil escribir la fórmula anterior como:

$$\boxed{P(A \cap B) = P(A \mid B) \times P(B).}$$

que llamamos la fórmula dos.

Comentarios

- La probabilidad condicional se interpreta como la probabilidad de que ocurra A cuando sabemos que ocurre B.
- Es como pensar que el espacio de probabilidad es B (porque no ocurren puntos fuera de B)
- Por eso contamos los puntos de A que están en B, y dividimos por los de B.

Ejemplo

Supongamos que tiramos el dado negro y el rojo, y consideramos los sucesos:

- $A = \{$el producto es 10$\}$.
- $B = \{$la suma de los dados es 7$\}$.

Pregunta: Calcular e interpretar $P(A \mid B)$.

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Cuadro: En verde las sumas 7, en violeta los productos 10

Solución Tenemos

- $P(A) = \frac{2}{36} = \frac{1}{18}$
- $P(B) = \frac{6}{36} = \frac{1}{6}$.
- $P(A \cap B) = P(A) = \frac{1}{18}$

$$P(A \mid B) = \frac{1/18}{1/6} = \frac{6}{18} = \frac{1}{3}.$$

- Ocurre que sabemos que ocurre B (la suma es 7), entonces tenemos 2 casos favorables en 6 posibles.

Fórmula de la probabilidad total

Ejemplo: Supongamos que se hacen test de covid. El espacio de resultados posibles Ω se divide en cuatro categorías:

- Los verdaderos positivos: el test da positivo y el paciente está enfermo.
- Los falsos positivos: el test da positivo pero el paciente no está enfermo.
- Los verdaderos negativos: el test da negativo y el paciente está sano.
- Los falsos negativos: el test da negativo pero el paciente está enfermo

Alternativamente Ω se divide en S (estar sano) y $S^c = E$ (estar enfermo), y también test positivo T^+ y test negativo T^-.

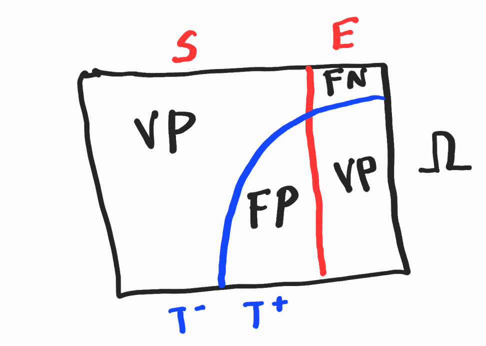
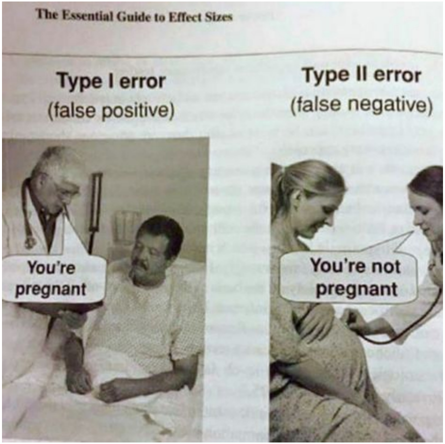


Figura: El espacio se divide en cuatro

Dos situaciones ¿posibles?



Tenemos que los positivos se dividen en 2:

$$T^+ = VP \cup FP = (E \cap T^+) \cup (S \cap T^+)$$

Como estas dos clases son disjuntas, tenemos

$$P(T^+) = P(E \cap T^+) + P(S \cap T^+)$$

Aplicando ahora la "fórmula dos"[4], tenemos

$$\boxed{P(T^+) = P(T^+ \mid E)P(E) + P(T^+ \mid S)P(S)}$$

Esta expresión se llama fórmula de la probabilidad total.
Veamos su formulación general.

[4] $P(A \cap B) = P(A \mid B) \times P(B)$

Fórmula de la probabilidad total

- Supongamos que el espacio Ω se particiona en los conjuntos $A_1, \ldots, A_n$.
- Supongamos que tenemos un suceso de interés B.
- La fórmula de la probabilidad total nos permite calcular B mediante

$$\boxed{P(B) = P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_n)P(A_n)}$$

que se llama fórmula de la probabilidad total.

Ejemplo

- El próximo jueves 2 Uruguay enfrenta a Perú por las eliminatorias sudamericanas.
- Supongamos que Uruguay tirará un penal, nos interesa saber cual es la probabilidad de que se convierta en gol, sea
$$B = \{\text{se convierte el gol}\}$$
- Habitualmente los penales los tira Suarez (que tiene una probabilidad de convertir de 0.95), Cavani (que tiene una probabilidad de convertir de 0.90) o un tercer jugador, que tiene una probabilidad de convertir de 0.60.

- Como no sabemos quien va a tirar, les asignamos una probabilidad de 1/3 a cada evento (tira S, tira C, tira T). Es decir
$$\Omega = S \cup C \cup T.$$

- Podemos calcular la probabilidad de B mediante

$$\begin{aligned} P(B) &= P(B \mid S)P(S) + P(B \mid C)P(C) + P(B \mid T)P(T) \\ &= 0{,}95 \times \frac{1}{3} + 0{,}90 \times \frac{1}{3} + 0{,}60 \times \frac{1}{3} = \mathbf{0.82} \end{aligned}$$

El homenaje de hoy:

- Thomas Bayes 1702-1761, matemático y presbítero británico que descubrió la una fórmula para la probabilidad inversa.
- Este descubrimiento, que fundamenta una amplia rama de la estadística, fue publicado luego de su muerte por Richard Price

Bioestadística (Clase 4):

Fórmula de probabilidad total, fórmula de Bayes, Independencia

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Probabilidad condicional

Definición. Dados dos sucesos A y B, con $P(A) > 0$ definimos la probabilidad condicional de B dado A, mediante

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

Nota: Precisamos que $P(A) > 0$ porque aparece en el denominador

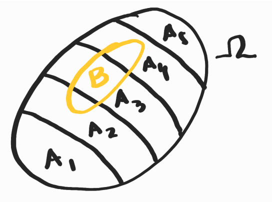
Es útil escribir la fórmula anterior como:

$$\boxed{P(B \cap A) = P(B \mid A) \times P(A).}$$

que llamamos la fórmula dos.

Fórmula de la probabilidad total

- Supongamos que el espacio Ω se particiona en los conjuntos $A_1, \ldots, A_n$ y tenemos un suceso de interés B:



- La fórmula de la probabilidad total nos permite calcular B mediante

$$\boxed{P(B) = P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_n)P(A_n)}$$

que se llama fórmula de la probabilidad total.

Demostración

- Como los $A_1, \ldots, A_n$ parten el espacio, también parten el conjunto B, es decir

$$B = (B \cap A_1) \cup \cdots \cup (B \cap A_n)$$

- Como estos conjuntos son disjuntos (por serlo los A_i) aplicamos la suma de probabilidades:

$$P(B) = P(B \cap A_1) + \cdots + P(B \cap A_n)$$

- Finalmente a cada sumando le aplicamos la regla dos[5]

$$P(B) = P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_n)P(A_n).$$

que es la fórmula buscada. $\square$

---

[5] $P(B \cap A) = P(B \mid A)P(A)$

Revisitamos el problema del penal[6].

- Penal para Uruguay. Tiran Suárez, Cavani, o un Tercer jugador. Sea

$$G = \{\text{se convierte el gol}\}$$

- Vamos a `https://www.transfermarkt.es/` (2008-2021):
  - Suárez convirtió 48 penales y erró 9 de los 57 pateados
  - Cavani convirtió 57 y erró 13 de los 70 pateados en el mismo período.
  - Otros jugadores convirtieron 20 y erraron 12 de los 32 pateados.

[6] Recalculando

- Estimamos probabilidades a partir de estas frecuencias:
- Para Suárez:
$$P(G \mid S) = \frac{48}{57} = 0{,}84,$$
- Para Cavani:
$$P(G \mid C) = \frac{57}{70} = 0{,}81,$$
- Para T (un tercer jugador)
$$P(G \mid T) = \frac{20}{32} = 0{,}63.$$

- Patearon penales por Uruguay entre 2009 y 2021: **S**uárez 18 veces, **C**avani 10 veces, un **T**ercer jugador 32 veces (total 60):

$$\Omega = S \cup C \cup T.$$

- Estimamos probabilidades para el tirador:

$$P(S) = \frac{18}{60}, \quad P(C) = \frac{10}{60}, \quad P(T) = \frac{32}{60}.$$

- Tenemos todos los ingredientes:

$$\begin{aligned} P(G) &= P(G \mid S)P(S) + P(G \mid C)P(C) + P(G \mid T)P(T) \\ &= \frac{48}{57} \times \frac{18}{60} + \frac{57}{70} \times \frac{10}{60} + \frac{20}{32} \times \frac{32}{60} = \mathbf{0.72} \end{aligned}$$

Fórmula de Bayes

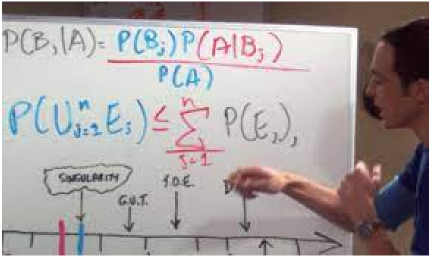
- Suponemos conocer $P(B \mid A)$ y queremos calcular $P(A \mid B)$.
- En el ejemplo anterior, sería: quien pateó el penal, dado que fue gol.
- Bayes hizo esta cuenta:

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

- Obtuvo su famosa fórmula

$$\boxed{P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}}$$

¿Lo conocen?



El físico Sheldon Cooper aplica la fórmula de Bayes, en el California Institute of Technology (Caltech)

- Está fórmula da origen a la estadística bayesiana
- $P(A)$ se llama probabilidad a priori
- $P(B \mid A)$ es la verosimilitud
- $P(A \mid B)$ se llama probabilidad a posteriori
- El cociente $\frac{P(B|A)}{P(B)}$ es el impacto de B en la probabilidad de A:

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

Ejemplo

- Supongamos que se tiró un penal y fue gol. ¿Cual es la probabilidad de que haya pateado Suárez?
- Es decir, queremos calcular $P(S \mid G)$.
- Aplicando la fórmula de Bayes:

$$P(S \mid G) = \frac{P(G \mid S)P(S)}{P(G)} = \frac{0{,}84}{0{,}72} \times 0{,}30 = 0{,}35$$

Fórmula general de Bayes

- La fórmula obtenida es

$$\boxed{P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}}$$

- Cuando no conocemos $P(B)$, si tenemos una partición de Ω con $A = A_1, \ldots A_n$, utilizamos en el denominador la fórmula de la probabilidad total. Obtenemos

$$\boxed{P(A_1 \mid B) = \frac{P(B \mid A_1)P(A_1)}{P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_n)P(A_n)}}$$

Ejemplo[7]

Tenemos una fábrica que tiene una alarma que suena en caso de accidente ... Pero:

- La probabilidad de que haya un accidente es $0,1$.
- La probabilidad de que suene la alarma si hay un accidente es $0,97$
- La probabilidad de que suene si no hay accidente es $0,02$ (falsa alarma)
- Escuchamos sonar la alarma
- ¿cuál es la probabilidad de que no haya habido un accidente?

[7] Tomado de las notas del Curso de A. Cholaquidis

Identifiquemos los sucesos:

- I: hubo un acciente
- I^c: no hubo accidente
- A: sonó la alarma

Sabemos

- $P(I) = 0,1$ (hubo accidente)
- $P(A \mid I) = 0{,}97$ (sonó la alarma dado que hubo accidente)
- $P(A \mid I^c) = 0{,}02$ (sonó la alarma dado que no hubo accidente)
- Queremos calcular

$$P(I^c \mid A) = \frac{P(A \mid I^c)}{P(A)} P(I^c).$$

- Primero calculamos el denominador $P(A)$ por probabilidad total:

$$\begin{aligned} P(A) &= P(A \mid I)P(I) + P(A \mid I^c)P(I^c) \\ &= 0{,}97 \times 0{,}1 + 0{,}02 \times 0{,}9 = 0{,}115. \end{aligned}$$

- Ahora

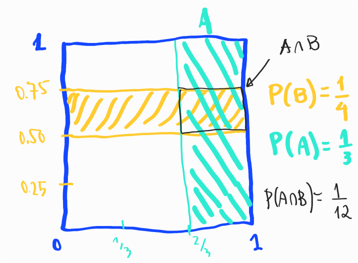
$$P(I^c \mid A) = \frac{P(A \mid I^c)}{P(A)} P(I^c) = \frac{0{,}02}{0{,}115} \times 0{,}9 = 0{,}157.$$

Sucesos independientes

- Decimos que dos sucesos A y B son independientes cuando se verifica

$$\boxed{P(A \cap B) = P(A) \times P(B).}$$

- La independencia es un concepto matemático.
- Si las probabilidades fuesen áreas, estos sucesos serían independientes:



Ejemplo

Supongamos que tiramos dos dados (el negro y el rojo)

- Veamos que el resultado del dado negro es independiente del dado rojo:
- Sea A el dado negro es par
- Sea B el dado rojo es menor que 3.

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	**(2,2)**	**(2,3)**	**(2,4)**	**(2,5)**	**(2,6)**
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	**(4,2)**	**(4,3)**	**(4,4)**	**(4,5)**	**(4,6)**
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	**(6,2)**	**(6,3)**	**(6,4)**	**(6,5)**	**(6,6)**

Cuadro: En negrita si el dado negro (el primero) es par. A la izquierda si el segundo dado es 1 o 2

- Tenemos $P(A) = \frac{18}{36} = \frac{1}{2}$.
- Además $P(B) = \frac{12}{36} = \frac{1}{3}$.
- Por último $P(A \cap B) = \frac{6}{36} = \frac{1}{6}$
- Como
$$P(A \cap B) = \frac{1}{6} = \frac{1}{2} \times \frac{1}{3} = P(A)P(B)$$
tenemos independencia de los sucesos A y B.

- Más en general, si N, R son los resultados respectivos de los dados del primer dado,
- El primer suceso es $N = k$ (el dado negro toma el valor k)
- El segundo suceso es $R = j$ (el dado rojo toma el valor j)
- Tenemos

$$P(N = k, R = j) = \frac{1}{36} = \frac{1}{6} \times \frac{1}{6} = P(N = k)P(R = j)$$

- Decimos que los resultados N y R son independientes.

Independencia y probabilidad condicional

- Supongamos que A y B son independientes, y que $P(A) > 0$. En ese caso tenemos

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

- La ocurrencia de A no modifica la probabilidad de B cuando son independientes.

Observaciones

- El suceso Ω es independiente de cualquier suceso A:

$$P(A \cap \Omega) = P(A) = P(A) \times 1 = P(A) \times P(\Omega)$$

- El suceso $\emptyset$ también es independiente de cualquier suceso A:

$$P(A \cap \emptyset) = P(\emptyset) = 0 = 0 \times P(A) = P(\emptyset) \times P(A)$$

- Si dos sucesos son disjuntos, y tienen probabilidad positiva, no pueden ser independientes:

$$0 = P(A \cap B) \neq P(A)P(B) > 0.$$

Mas observaciones

Si A, B son independientes, también lo son A, B^c, A^c, B y A^c, B^c. Por ejemplo

$$A = (A \cap B) \cup (A \cap B^c)$$

es una unión disjunta, y resulta

$$P(A) = P(A \cap B) + P(A \cap B^c) = P(A)P(B) + P(A \cap B^c)$$

Despejando

$$P(A \cap B^c) = P(A) - P(A)P(B) = P(A)[1 - P(B)] = P(A)P(B^c)$$

nos da la independencia. Los otros resultados son análogos.

Hoy demostrarmos que: La mala suerte no es verdad

Bioestadística (Clase 5):

Variables aleatorias discretas

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Sucesos independientes

- Decimos que dos sucesos A y B son independientes cuando se verifica

$$\boxed{P(A \cap B) = P(A) \times P(B).}$$

- La independencia es un concepto matemático.

Ejemplo

Supongamos que tiramos dos dados (el negro y el rojo)

- Veamos que el resultado del dado negro es independiente del dado rojo si N, R son los resultados respectivos.
- El primer suceso es $N = k$ (el dado negro toma el valor k)
- El segundo suceso es $R = j$ (el dado rojo toma el valor j)
- Tenemos

$$P(N = k, R = j) = \frac{1}{36} = \frac{1}{6} \times \frac{1}{6} = P(N = k)P(R = j)$$

- Decimos que los resultados N y R son independientes.

Ejemplo de sucesos dependientes

Consideremos los sucesos:

- $A = \{\text{la suma es mayor estricta que 7}\} = \{N + R > 7\}$
- $B = \{\text{el dado negro es mayor o igual a 4}\} = \{N \geq 4\}$
- Veamos si estos sucesos A y B son independientes

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Cuadro: En azul la suma mayor que 7, a la derecha de la línea el dado negro mayor o igual a 4

$$P(A) = \frac{15}{36}, \quad P(B) = \frac{18}{36} = \frac{1}{2}, \quad P(A \cap B) = \frac{12}{36} = \frac{1}{3}$$

$$P(A)P(B) = \frac{15}{72} \neq P(A \cap B) = \frac{1}{3} = \frac{24}{72}.$$

Los sucesos A y B no son idependientes.

Variables aleatorias discretas

- El ejemplo anterior nos sirvió para introducir las variables aleatorias R y N
- Una variable aleatoria es una función que nos lleva del espacio Ω a $\mathbb{R}$
- Como estamos en $\mathbb{R}$ nos permite utilizar herramientas de cálculo
- La variable aleatoria es discreta cuanto toma un conjunto finito $\{0, 1, 2, \ldots, N\}$ o numerable $\{0, 1, 2, \ldots\}$ de valores.
- Nos da flexibilidad para considerar modelos de probabilidad no equiprobables.

Variable aleatoria uniforme

- El primer ejemplo que consideramos es una variable aleatoria X que toma valores $1, \ldots N$.
- Asumimos que todos los valores son equiprobables.
- Tenemos

$$P(X = k) = \frac{1}{N}, \quad \text{para } k = 1, \ldots, N$$

- Este modelo:
  - Si $N = 2$ corresponde a tirar una moneda equilibrada, y asignar 1 y 2 a cara y número.
  - Si $N = 6$ corresponde a tirar un dado equilibrado.

Variable aleatoria de Bernoulli

- Nuestra variable toma dos valores:

$$X = \begin{cases} 0, & \text{que corresponde al fracaso} \\ 1, & \text{que corresponde al éxito} \end{cases}$$

  de un experimento aleatorio con dos resultados posibles

- Los valores **no** son equiprobables. Para un $p \in (0,1)$:

$$P(X = 0) = 1 - p, \qquad P(X = 1) = p.$$

- Este modelo se puede aplicar en muchas situaciones:
  - Si un producto es defectuoso o no,
  - Si salvo un examen o lo pierdo,
  - Si en un texto en un lugar tengo una vocal o una consonante . . . etc.

Variable aleatoria Binomial

- Es de mucho interés la repetición de experimentos independientes de Bernoulli, por ejemplo n veces.
- En ese caso, el resultado sería una tira de ceros y unos, correspondientes a cada éxito o fracaso:
- Una tal tira sería
$$0010100010$$
En este caso $n = 10$ y tenemos 3 éxitos y 7 fracasos.
- La variable aleatoria Binomial Y cuenta la cantidad de éxitos en los n experimentos.
- ¿Qué valores puede tomar Y?

- Si no hay ningún éxito tenemos $Y = 0$, si todos son éxitos tenemos $Y = n$.
- Entonces Y toma alguno de los valores $0, 1 \ldots, n$.
- ¿Y con qué probabilidades toma Y cada uno de estos valores?
- Supongamos $n = 2$. ¿Cuántas tiras con 2 ceros o unos tenemos?

- Tenemos 4:

$$00, \quad 01, \quad 10, \quad 11.$$

- Como los resultados son independientes, las probabilidades para estas tiras son:

$$\begin{aligned}
P(00) &= P(0) \times P(0) = (1-p)^2, \\
P(01) &= P(0) \times P(1) = (1-p)p, \\
P(10) &= P(1) \times P(0) = p(1-p), \\
P(11) &= P(1) \times P(1) = p^2,
\end{aligned}$$

- Obtenemos

$$P(Y=0)=(1-p)^2,\ P(Y=1)=2p(1-p),\ P(Y=2)=p^2.$$

- Además

$$\begin{aligned} P(Y=0)+P(Y=1)+P(Y=2) \\ &= (1-p)^2+2p(1-p)+p^2=(1-p+p)^2=1. \end{aligned}$$

Veamos el caso $n = 3$

- Cuántas tiras tenemos:

$$\_\,\_\,\_$$

- En este caso $n = 3$ y $2^3 = 8$.
- Contamos la mala suerte pura (todos fracasos)

$$P(Y = 0) = (1 - p)^3$$

- Contamos la buena suerte pura (todos éxitos)

$$P(Y = 3) = p^3$$

- ¿Cómo son los casos intermedios?
- Por ejemplo, para la tira 100 tenemos

$$P(100) = P(1)P(0)^2 = p(1-p)^2.$$

- ¿Cuántas tiras tienen un solo éxito?

$$\underline{1}\ \_\ \_$$

- Son $3 = \binom{3}{1}$. Entonces

$$P(Y=1) = 3p(1-p)^2 = \binom{3}{1} p(1-p)^2.$$

- Análogamente, dos éxitos es un fracaso, entonces

$$P(Y = 2) = 3p^2(1 - p) = \binom{3}{1} p^2(1 - p).$$

- Y se verifica

$$P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3) =$$
$$= (1-p)^3 + 3(1-p)^2 p + 3(1-p)p^2 + p^3 = (1-p+p)^3 = 1.$$

- Escrito con sumatoria:

$$\sum_{k=0}^{3} P(Y = k) = \sum_{k=0}^{3} \binom{3}{k} (1 - p)^{3-k} p^k = 1.$$

¿Nos animamos al caso general?

- Tenemos 2^n tiras de ceros y unos
- Si tenemos k éxitos ($0 \leq k \leq n$) los podemos ubicar en n lugares que elegimos de $\binom{n}{k}$ formas diferentes.
- Cada una de esas formas me da un caso de $Y = k$.
- La probabilidad de cada forma es $p^k(1-p)^{n-k}$
- Entonces

$$\boxed{P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n}$$

- La variable aleatoria Y se llama binomial, con parámetros n, p.

- Veamos que el modelo está correctamente definido, es decir, la suma de todos los resultados posibles tiene probabilidad uno.
- Para eso tenemos que sumar:

$$\sum_{k=0}^{n} P(Y=k) = \sum_{k=0}^{n} \binom{n}{k} (1-p)^{n-k} p^k = (p+1-p)^n = 1.$$

Ejemplo

En la generala se tiran 5 dados.

(a) Calcular la probabilidad de obtener dos dados con 6.

(b) Calcular la probabilidad de sacar al menos un 6.

Solución:

- El éxito es sacar un 6 al tirar el dado.
- La probabilidad de éxito es $p = 1/6$.
- Y es una binomial con parámetros $n = 5$ y $p = 1/6$
  Tenemos

$$P(Y = 2) = \binom{5}{2} p^2 (1-p)^3 = 10 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3 = 0{,}16.$$

- Para la parte (b) tenemos que calcular

$$P(Y \geq 1)$$

- Tenemos

$$\begin{aligned} P(Y \geq 1) = P(Y = 1) + P(Y = 2) + P(Y = 3) \\ + P(Y = 4) + P(Y = 5) \end{aligned}$$

- Es más sencillo:

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - (1 - p)^5 = 1 - \left(\frac{5}{6}\right)^5 = 0{,}60.$$

Paseo al azar, triangulo de Pascal y variable aleatoria Binomial

- Consideremos una partícula que se desplaza, eligiendo con probabilidad p una unidad para arriba (éxito) o para abajo en cada instante de tiempo (fracaso):
- Llamamos paseo al azar a cada posible trayectoria.

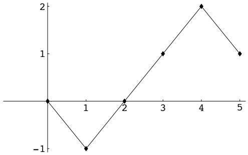


Figura: ¿Qué probabilidad tiene esta trayectoria?

$$P(01110) = p^3(1-p)^2.$$

- ¿Cuál es la probabilidad de que llegue (por cualquier camino) a 1?

$$P(Y = 3) = \binom{5}{3} p^3 (1-p)^2.$$

Nos interesa calcular la probabilidad de llegar a una altura dada en una cantidad de pasos n

- Si un camino tiene la misma cantidad de ceros y unos, llegará a la misma altura.

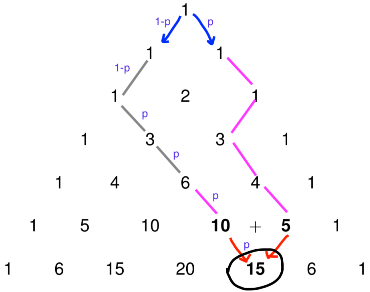
- Nuestro problema es contar de cuantas formas distintas se pueden disponer k unos en n lugares:

$$\underbrace{\underline{0}\ \underline{0}\ \underline{1}\ \underline{0}\ \underline{1}\ \underline{0}\ \underline{1}}_{7 \text{ lugares}}$$

$$C^7_3 = \frac{7!}{3!\,4!} = 35$$

- Y calcular la probabilidad indvidual de esos caminos, que es $p^3(1-p)^4$

Si representamos estas cantidades en una tabla obtenemos el triángulo de Pascal, al que le agregamos probabilidades:



La probabilidad de llegar a ese lugar consiste en tener 4 éxitos y 2 fracasos:

$$P(Y=4)=\binom{6}{4}p^4(1-p)^2=15p^4(1-p)^2.$$

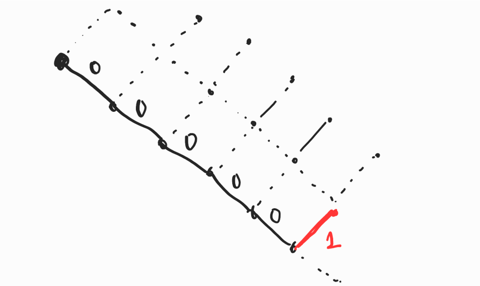
Variable geométrica (la mala suerte no es verdad)

- Se trata de realizar experimentos de Bernoulli hasta el primer éxito.
- El número Y de fracasos necesarios hasta el primer éxito es (por definición) una variable aleatoria geométrica
- Si soy muy afortunado, mi primer experimento es un éxito, me da $Y = 0$
- Si tengo un fracaso y un éxito, tengo $Y = 1$.
- Puede ocurrir que tenga una cantidad de fracasos arbitrariamente grande.
- Entonces Y toma todos los valores naturales:

$$0, 1, \ldots, n, \ldots$$

¿Y cuales son las probabilidades?

- Tenemos: $P(Y = 0) = P(1) = p$.
- Si fracaso una sóla vez: $P(Y = 1) = P(01) = (1 - p)p$.
- Si fracaso dos veces: $P(Y = 2) = P(001) = (1 - p)^2 p$.
- Si fracaso k veces: $P(Y = k) = P(\underbrace{0 \ldots 0}_{k} 1) = (1 - p)^k p$.
- Son las trayectorias de la siguiente forma:



- Veamos que en algún momento tenemos un éxito (aunque p sea muy pequeña):

$$\begin{aligned} P(Y=0)+P(Y=1)+\cdots+P(Y=n)+\cdots \\ = p+(1-p)p+\cdots+(1-p)^n p+\cdots \\ = p\times\left[1+(1-p)+\cdots+(1-p)^n+\cdots\right] \end{aligned}$$

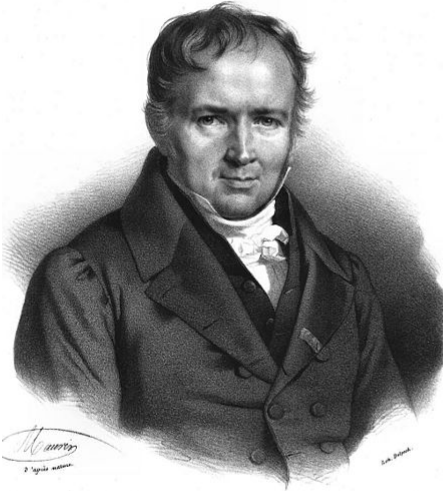
- Y la expresión en azul es una serie geométrica de razón $r=1-p$. Tenemos

$$1+r+r^2+\cdots+r^n+\cdots=\frac{1}{1-r}$$

- Entonces

$$
\begin{aligned}
P(Y=0)+P(Y=1)+\cdots+P(Y=n)+\cdots \\
&= p \times \left[1+(1-p)+\cdots+(1-p)^n+\cdots\right] \\
&= p\frac{1}{1-(1-p)} = 1.
\end{aligned}
$$

- En algún momento tenemos suerte.



Siméon Denis Poisson. Matemático francés, 1781-1840

Bioestadística (Clase 6):

Modelos Multinomial, Hipergeométrico y de Poisson

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Variables aleatorias discretas (repaso)

- Una variable aleatoria es una función que nos lleva del espacio Ω a $\mathbb{R}$
- Como estamos en $\mathbb{R}$ nos permite utilizar herramientas de cálculo
- La variable aleatoria es discreta cuanto toma un conjunto finito $\{0, 1, 2, \ldots, N\}$ o numerable $\{0, 1, 2, \ldots\}$ de valores.
- Nos da flexibilidad para considerar modelos de probabilidad no equiprobables.

Variable aleatoria uniforme

- Es una variable aleatoria X que toma los valores $1, \ldots N$.
- Asumimos que todos los valores son equiprobables.
- Tenemos

$$P(X = k) = \frac{1}{N}, \quad \text{para } k = 1, \ldots, N$$

Variable aleatoria de Bernoulli

- Nuestra variable toma dos valores:

$$X = \begin{cases} 0, & \text{que corresponde al fracaso} \\ 1, & \text{que corresponde al éxito} \end{cases}$$

  de un experimento aleatorio con dos resultados posibles

- Los valores **no** son equiprobables. Para un $p \in (0,1)$:

$$P(X=0) = 1-p, \qquad P(X=1) = p.$$

Variable aleatoria Binomial

- La variable aleatoria binomial cuenta la cantidad de éxitos en una serie de n experimentos de Bernoulli independientes.
- Cada experimento tiene una probabilidad de éxito $p \in (0, 1)$, la variable binomial tiene parámetros n y p.
- Una variable binomial Y toma alguno de los valores $0, 1 \ldots, n$:

$$\boxed{P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n}$$

Variable geométrica

- Una variable aleatoria geométrica Y que cuenta el número de fracasos necesarios hasta el primer éxito es (por definición) de una serie arbitrariamente larga de experimentos de Bernoulli.

- La variable Y toma todos los valores naturales:

$$0, 1, \ldots, n, \ldots$$

- Si fracaso k veces: $P(Y = k) = P(\underbrace{0 \ldots 0}_{k}\, 1) = (1 - p)^k p.$

Distribución Multinomial

- Modela la repetición de n experimentos independientes con más de dos o más resultados posibles.
- En el caso de ser 2, tenemos la distribución Binomial.
- Supongamos que tenemos un experimento con tres resultados posibles, que etiquetamos A,B y C.
- Cada uno de los resultados tiene una probabilidad p, q, r que cumplen $p + q + r = 1$
- Nos preguntamos entonces por la probabilidad de tener x_1 resultados 0, x_2 resultados 1, x_3 resultados 2.
- Como son n experimentos tenemos $x_1 + x_2 + x_3 = n$

- Por ejemplo $n = 6$, $x_1 = 2$, $x_2 = 1$, $x_3 = 3$.
- Los resultados posibles entonces son tiras de la forma

$$A\ A\ B\ C\ C\ C$$

- Como son independientes, la probabilidad de esta tira es

$$P(A\ A\ B\ C\ C\ C) = p^2 q\, r^3.$$

- ¿Cuántas tiras disitintas tienen dos A, una B, y tres C?

- Para contarlas, primero ubico las A en los 6 lugares:

$$\_\ \underline{A}\ \_\ \underline{A}\ \_\ \_$$

- Tenemos $\binom{6}{2}$ formas de hacerlo
- Luego ubicamos la B en los 4 lugares restantes:

$$\_\ \underline{A}\ \underline{\color{red}{B}}\ \underline{A}\ \_\ \_$$

- Tenemos $\binom{4}{1}$ formas de hacerlo por cada una de las anteriores.
- Los restantes lugares se ocupan con las tres C:

$$\underline{\color{blue}{C}}\ \underline{A}\ \underline{\color{red}{B}}\ \underline{A}\ \underline{\color{blue}{C}}\ \underline{\color{blue}{C}}$$

- Si aplicamos la regla del producto, tenemos $\binom{6}{2}\binom{4}{1}$ formas diferentes de ubicar los resultados,
- Calculando

$$\binom{6}{2}\binom{4}{1} = \frac{6!}{2!4!} \times \frac{4!}{1!3!} = \frac{6!}{2!1!3!}$$

- El número que obtuvimos es el coeficiente de la probabilidad (es decir, la cantidad de tiras con dos A, una B y tres C), y se llama coeficiente multinomial, escribiéndose

$$\binom{6}{2,1,3} = \frac{6!}{2!1!3!}$$

- Nuestra conclusión es que

$$\begin{aligned} P(\text{dos A, una B, tres C}) &= P(x_1 = 2, x_2 = 1, x_3 = 3) \\ &= \frac{6!}{2!1!3!} p^2 q^1 r^3. \end{aligned}$$

Fórmula general

- Tenemos k resultados posibles de un experimento que se repite n veces

- Las probabilidades de los resultados de un experimento son $p_1, \ldots, p_k$, que verifican

$$p_1 + \cdots + p_k = 1.$$

- Nos interesa que de los n experimentos, el primer resultado salga x_1 veces, el segundo x_2 veces ... el k-ésimo x_k veces, y se verifica

$$x_1 + \cdots + x_k = n.$$

- La probabilidad es

$$P((x_1, \ldots, x_k)) = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} \cdots p_k^{x_k}.$$

- Escrita de otra forma

$$P((x_1, \ldots, x_k)) = \frac{n!}{x_1! \cdots , x_k!} p_1^{x_1} \cdots p_k^{x_k}.$$

Ejemplo

- Tiramos 5 dados al jugar a la generala.
- ¿Cuál es la probabilidad de sacar un uno, un dos , y tres seis?

Solución

- Tenemos $n = 5$ experimentos con 6 resultados posibles cada uno.
- Cada uno de los 6 resultados tiene probabilidad 1/3.
- Queremos tiras de la forma

$$\underline{1}\ \underline{2}\ \underline{6}\ \underline{6}\ \underline{6}$$

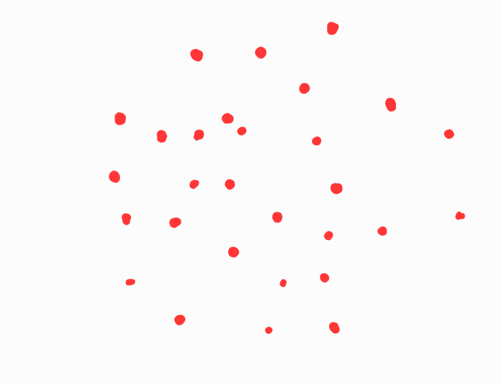
- Para eso precisamos que no salgan ni tres, ni cuatros, ni cincos.
- Es decir, tenemos
$$x_1 = 1,\ x_2 = 1,\ x_3 = 0,\ x_4 = 0,\ x_5 = 0,\ x_6 = 3.$$
- Aplicando la probabilidad multinomial
$$P(x_1 = 1, x_2 = 1, x_3 = 0, x_4 = 0, x_5 = 0, x_6 = 3)$$
$$= \binom{5}{1,1,0,0,0,3} \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^3$$
$$= \frac{5!}{3!} \frac{1}{6^5} = 0{,}0026$$

Distribución de Poisson

Supongamos que estamos observando bacterias en una placa de Petri:

O elefantes en una foto aérea:

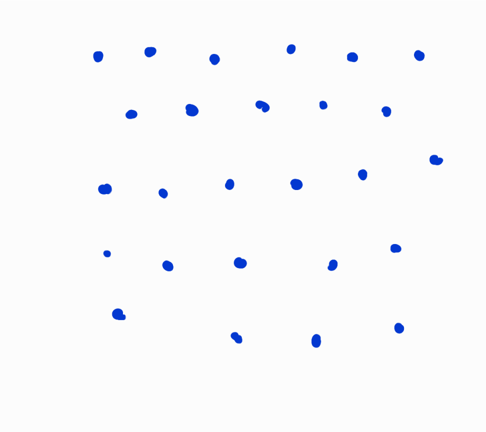
Digamos que tenemos una distribución de puntos en el plano:



Y nos interesa saber si los puntos entre sí

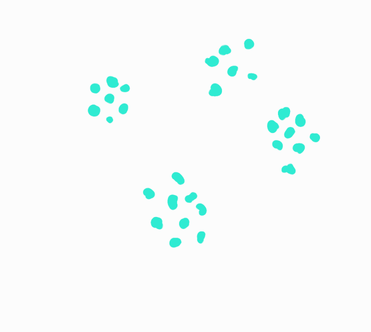
- Se atraen
- Se repelen
- Son indiferentes

Estos puntos se repelen:

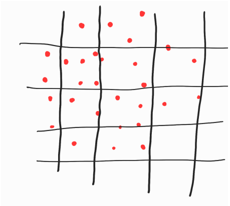
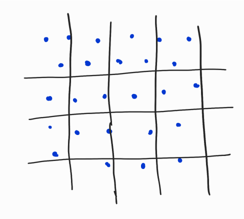
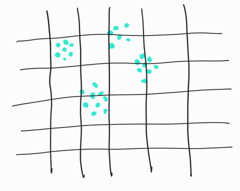


Podrían ser animales que compiten por recursos (comida, agua, etc.)

Estos puntos se atraen:



Para decidir que situación tenemos, cuadriculamos el plano:

# punto	# cuadros
0	1
1	1
2	5
3	2
4	1
5	1
6	0
7	0
⋮	⋮

# Pts	# □
0	0
1	7
2	5
3	3
4	0
5	0
6	0
7	0
⋮	⋮

# Pt	# □
0	8
1	0
2	0
3	2
4	1
5	1
6	1
7	0
8	
⋮	

- De cantidades podemos pasar a frecuencias.
- De frecuencias a probabilidades
- La situación en que los puntos son indiferentes, dan las probabilidades de Poisson
- Una variable de Poisson X con intensidad λ toma los valores en los naturales $\{0, 1, \ldots, n, \ldots\}$ con probabilidades

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

- El λ representa la cantidad media de puntos por cuadro.

La distribución de Poisson se aplica en muchas situaciones:

- La cantidad de llamadas telefónicas a una central en un intervalo de tiempo
- La cantidad de estrellas de una cierta magnitud en el cielo
- La cantidad de clientes en un supermercado durante un lapso de tiempo
- La cantidad de tareas que recibe un procesador de computadora
- $\vdots$

Ejercicio

La cantidad de errores tipográficos en un texto sigue una distribución de Poisson, y son aproximadamente una por millón.

- El Quijote tiene 377 032 palabras y 1 687 570 caracteres
- Calcular la probabilidad de que en todo haya alguna errata.

Solución. El promedio de errores es

$$\lambda = 1,687570 \sim 1,7$$

Entonces

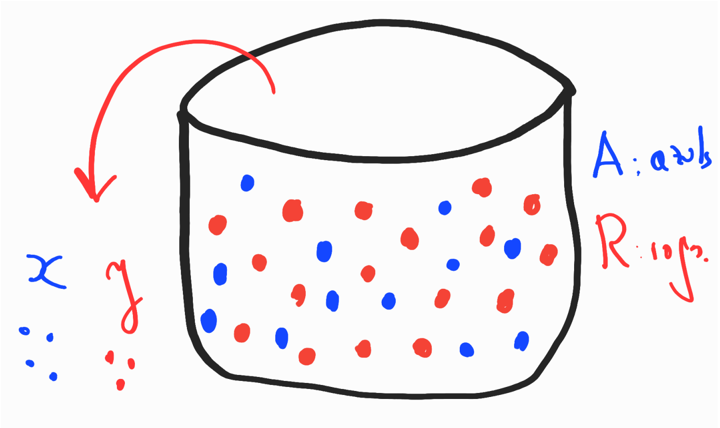
$$P(\text{cero error}) = P(X = 0) = e^{-\lambda} = e^{-1{,}7} = 0{,}18.$$

$$P(\text{alguna errata}) = 1 - P(X = 0) = 0{,}82.$$

Distribución Hipergeométrica

- Extraemos n bolas de una urna que tiene A bolas azules y R bolas rojas.

- ¿Cuál es la probabilidad de extraer exactamente x bolas azules, y bolas rojas?



- Extraemos $x + y$ bolas de un total de $A + R$
- Tenemos
$$\binom{A+R}{x+y}$$
extracciones distintas, cada una es un caso posible.
- Los favorables son cuando hay exactamente x azules, y rojas.
- Como hay un total de A azules, podríamos seleccionar
$$\binom{A}{x}$$
bolas azules, cada una conformaría un caso favorable

- A su vez, para cada configuración azul, preciso extraer y rojas entre las R posibles. Eso me da

$$\binom{R}{y}$$

posibilidades.

- Como cada posibilidad de las azules y las rojas se combina, aplicando la regla del producto, tengo

$$\binom{A}{x}\binom{R}{y}$$

casos favorables.

- La probabilidad buscada es la distribución hipergeométrica:

$$P((x,y)) = \frac{\binom{A}{x}\binom{R}{y}}{\binom{A+R}{x+y}}$$

Ejemplo Un mazo de 52 cartas tiene la mitad rojas y la mitad negras. ¿Cual es la probabilidad, al extraer 10 cartas, de obtener exactamente 5 rojas y 5 negras?

Solución

$$P((5,5)) = \frac{\binom{26}{5}\binom{26}{5}}{\binom{52}{10}} = 0{,}27.$$

Clase 7 de Bioestadística

Variables aleatorias continuas

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Introducción

Nos interesa modelar un experimento cuyo resultado es un número real:

- La medición de una longitud, con un dispositivo sujeto a error,
- El tiempo hasta que se rompe una lamparita,
- La medición de la altura de una persona elegida al azar en una población,
- El peso de un bebé al nacer,
- El tiempo de desintegración radiactiva de un núcleo atómico inestable
- ¿más ejemplos?

Identificamos el conjunto de valores I que puede tomar la variable, que pude ser

- Todo $I = \mathbb{R} = \{x\colon -\infty < x < \infty\}$
- Los reales no-negativos $I = \mathbb{R}^{+} = [0, \infty) = \{x\colon x \geq 0\}$
- Un intervalo $I = [a, b] = \{a \leq x \leq b\}$

Estudiamos entonces variables aleatorias continuas

$$X : \Omega \to I.$$

Suponemos que hay un universo Ω en donde mediante un experimento se elige al azar un $\omega \in \Omega$ y que nosotros observamos

- $X(\omega)$ antes del experimento
- $X(\omega) = x \in I$ después del experimento

Lo que estudiamos a continuación es cómo se distribuyen las probabilidades en I

Variable aleatoria Uniforme

- Supongamos que $I = [0, 1]$ y elegimos un punto al azar X en I.
- Supongamos que queremos indiferencia para la ubicación del resultado en $[0, 1]$
- Para eso suponemos que los sucesos

$$X \in \left[0, \frac{1}{2}\right], \qquad X \in \left[\frac{1}{2}, 1\right]$$

  tienen la misma probabilidad.
- ¿Qué probabilidad tendría que tener cada intervalo?

- Si queremos que además los sucesos

$$X \in \left[0, \frac{1}{4}\right], \quad X \in \left[\frac{1}{4}, \frac{1}{2}\right], \quad X \in \left[\frac{1}{2}, \frac{3}{4}\right], \quad X \in \left[\frac{3}{4}, 1\right].$$

  tienen la misma probabilidad, deberían tener probabilidad $\frac{1}{4}$

- La distribución uniforme es la que asigna para $0 \leq x < y \leq 1$:

$$\boxed{P(X \in [x, y]) = y - x}$$

- Es decir, a cada intervalo le asignamos como probabilidad su longitud.

Ejercicio

- Se elige un número X al azar en $[0, 1]$ con distribución uniforme.
- ¿Cuál es la probabilidad de que la primera cifra decimal sea un 3?
$$x = 0, \underline{3}\ \_\ \cdots$$
- ¿Cuál es la probabilidad de que la segunda cifra decimal sea un 5?
$$x = 0, \_\ \underline{5}\ \_\ \cdots$$

Solución

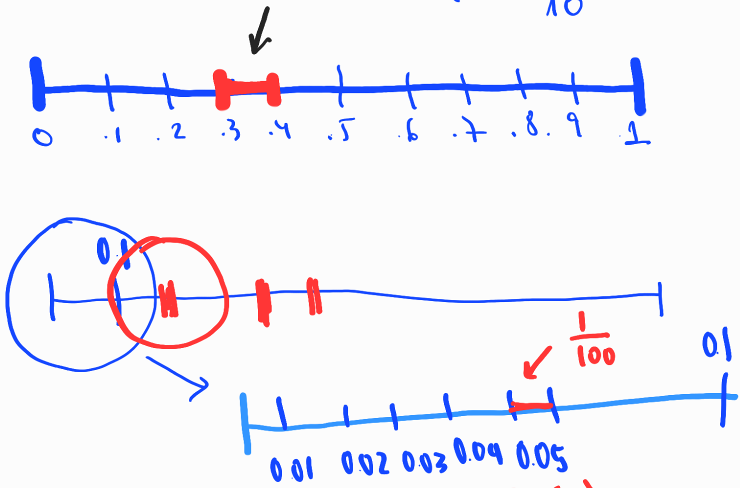
- Sean los sucesos

$$A = \{\text{la primer cifra es un 3}\}$$

$$B = \{\text{la segunda cifra es un 5}\}$$

$$C = A \cap B = \{\text{primer cifra 3 y segunda cifra 5}\}$$

0.3- - - - $P(A) = \frac{1}{10}$



Total: $10 \times \frac{1}{100} = \frac{1}{10} = P(B)$

Observación

- Tenemos
$$P(C) = \frac{1}{100}$$
- Entonces
$$P(A \cap C) = P(A)P(B)$$
Los sucesos A y B son independientes

- Pero ... ¿cómo? ¿independientes?
- ¡Sí! Cumplen la definición
- La primer y segunda cifra de un mismo número aleatorio en $[0, 1]$ son independientes.
- Ocurre que el conocimiento de uno no tiene información sobe el resultado del otro

Variable uniforme en $[a, b]$

- ¿Cómo tenemos que modificar el resultado anterior si queremos un resultado en un intervalo cualquiera $[a, b]$?
- Tomamos la longitud relativa al intervalo total
- Si X es tiene distribución uniforme en $[a, b]$, se verifica

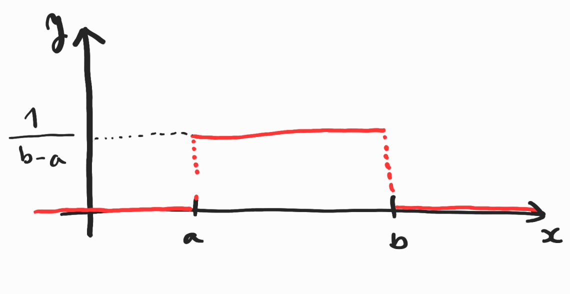
$$P(X \in [x, y]) = \frac{y - x}{b - a} \qquad a \leq x < y \leq b.$$

Densidad de la variable aleatoria uniforme en $[a, b]$

- La densidad de la variable aleatoria X uniforme en $[a, b$ es la función

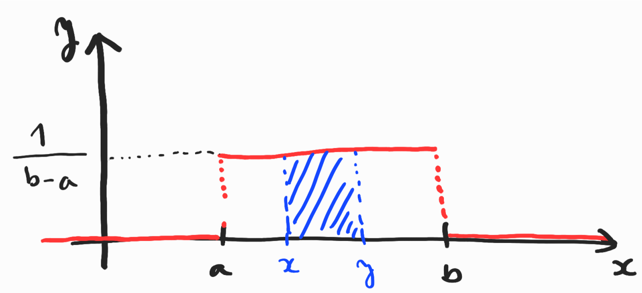
$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{cuando } a \leq x \leq b \\ 0, & \text{en otro caso} \end{cases}$$

se llama densidad de la variable aleatoria X, uniforme en $[a, b]$



- Las probabilidades se calculan como áreas:
- Para calcular áreas usamos integrales:

$$P(X \in [x, y]) = \int_x^y f(t)dt = \int_x^y \frac{1}{b-a}dt = \frac{y-x}{b-a}$$



Distribución de una variable aleatoria uniforme

- Más en general nos interesa la siguiente función:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

llamada función de distribución o también función de distribución de probabilidades e la variable aleatoria X

Calculamos

- Primero observamos que el cálculo es entre a y x (porque la $f(t)$ vale cero fuera de $[a, b]$:

$$F(x) = \int_a^x f(t)dt = \int_a^x \frac{1}{b-a}dt$$

$$= \frac{1}{b-a}\int_a^x 1dt = \frac{x-a}{b-a}$$

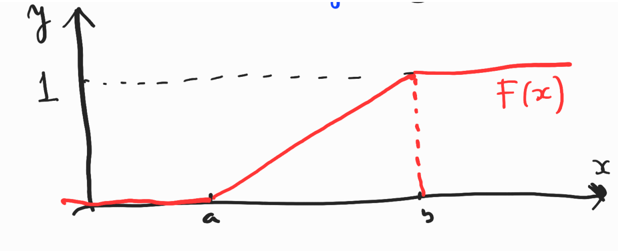
- Eso si $x \leq b$. En el caso en que $x > b$, tenemos

$$F(x) = \int_a^b f(t)dt = \int_a^b \frac{1}{b-a}dt$$

$$= \frac{1}{b-a}\int_a^b 1dt = \frac{b-a}{b-a} = 1$$

Obtenemos

$$F(x) = \begin{cases} 0, & \text{cuando } x < a \\ \frac{x-a}{b-a} & \text{cuando } a \leq x \leq b \\ 1, & \text{cuando } b \leq x \end{cases}$$



- También podemos obtener la densidad f a partir de la distribución F.
- Sabemos que se verifica

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

- Si derivamos con respecto a x obtenemos

$$F'(x) = f(x)$$

- La densidad f y la distribución F tienen la misma información: se puede obtener una a partir de la otra

- Con la función de distribución podemos calcular probabilidades.

- Si A es el suceso de obtener un 3 como primer dígito de X

$$
\begin{aligned}
P(A) = P(X \in [0{,}3, 0{,}4]) &= \int_{0{,}3}^{0{,}4} f(t)dt \\
&= F(0{,}4) - F(0{,}3) = 0{,}1.
\end{aligned}
$$

Simulación de una variable uniforme

Cada variable aleatoria tiene en R cuatro comandos asociados, que en el caso de la uniforme son

- `dunif` me da la densidad uniforme $f(x)$
- `punif` me da la distribución de probabilidades $F(x)$
- `qunif` me da la función inversa F^{-1} de la distribución F
- `runif` me numeros aleatorios (random) con distribución uniforme

Por ejemplo `runif(10,1,2)` me da 10 numeros aleatorios con distribución uniforme en $[1, 2]$:

```
> runif(10,1,2)
 [1] 1.306220 1.863714 1.076289 1.583812
 [5] 1.015811 1.149963 1.841357 1.572209
 [9] 1.837909 1.232794
```

Ahora ... ¿cómo genera el R los números aleatorios?

- En primer lugar genera una variable aleatoria discreta con un N muy grande y divide por N.
- De esa forma obtiene un número en $[0, 1]$ que es casi continuo (tiene muchas cifras pero no infinitas)
- En segundo lugar, se genera una cantidad n (para ver como se distribuyen)
- Eso lo hace mediante un algoritmo
- Los números no son realmente aleatorios: si sabemos el algoritmo podemos “adivinarlos”.
- Se llaman pseudo-aleatorios

- El número $n+1$ se obtiene a partir del X_n:

$$X_{n+1} = (aX_n + c) \bmod m$$

  donde

  - $m > 0$ es el módulo
  - $(x) \bmod m$ es el resto de dividir x por m:
  - $(100) \bmod 7 = 2$: siempre es un número entre 0 y $m = 6$
  - a tal que $0 < a < m$ es el multiplicador
  - c tal que $0 \leq c < m$ es el incremento
  - X_0 es la semilla

Variable aleatoria exponencial

- El segundo modelo continuo que consideramos es una variable aleatoria[8] T con distribución exponencial
- Es una variable que toma valores positivos: $I = [0, \infty)$
- Se puede definir a partir de su distribución:

$$F(t) = P(T \leq t) = 1 - e^{-\lambda t} \quad (t \geq 0)$$

- Derivando tenemos la densidad exponencial:

$$f(t) = F'(t) = \lambda e^{-\lambda t} \quad (t \geq 0)$$

[8] usamos la T porque podría tratarse de un tiempo aleatorio

- El modelo exponencial se utiliza para medir duraciones de componentes que no envejecen
- Por ejemplo, que se rompen por motivos externos a su desgaste
- Se puede usar para modelar el tiempo que dura una lámpara, que se quema por un pico de corriente externo
- Eso es porque cumple la propiedad de pérdida de memoria.
- Para eso usamos la sobrevida de T, que es

$$P(T > t) = e^{-\lambda t}.$$

Pérdida de memoria

- La pérdida de memoria es una propiedad matemática:

$$P(T > t + h \mid T > t) = P(T > h)$$

- Primero verificamos la propiedad:

$$\begin{aligned}
P(T > t + h \mid T > t) &= \frac{P(\{T > t + h\} \cap \{T > t\})}{P(T > t)} \\
&= \frac{P(\{T > t + h\})}{P(T > t)} \\
&= \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda(t+h)+\lambda t} \\
&= e^{-\lambda h} = P(T > h)
\end{aligned}$$

Ahora la interpretamos:

$$P(T > t + h \mid T > t) = P(T > h)$$

- Si T no ocurrió en tiempo t, esperar h más (hasta $t + h$) es equivalente a esperar h a partir de 0
- Por eso no importa cuanto tiempo ocurrió, si la lámpara no se rompió, es como que fuese nueva

Simulación de una variable exponencial

En R los cuatro comandos de una variable exponencial son:

- `dexp` me da la densidad exponencial
- `pexp` me da la distribución de probabilidades expnencial
- `qexp` me da la función inversa F^{-1} de la distribución F
- `rexp` me numeros aleatorios (random) con distribución exponencial
- Por ejemplo:

```
> rexp(3,1)
[1] 1.1611435 2.1357145 0.3589149
```

simula 3 variables exponenciales con parámetro $\lambda = 1$.

Johann Carl Friedrich Gauss (1777-1855, Alemania)

$$\boxed{1 + 2 + \cdots + 50 + 51 + \cdots + 99 + 100 = 101 \times 50 = 5050.}$$

Clase 8 de Bioestadística

Variables aleatorias exponenciales y normales

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Introducción

- Nos interesa modelar un experimento cuyo resultado es un número real.
- Identificamos el conjunto de valores I que puede tomar la variable, que pude ser
  - Todo $I = \mathbb{R} = \{x : -\infty < x < \infty\}$
  - Los reales no-negativos $I = \mathbb{R}^+ = [0, \infty) = \{x : x \geq 0\}$
  - Un intervalo $I = [a, b] = \{a \leq x \leq b\}$

Estudiamos entonces variables aleatorias continuas

$$X\colon \Omega \to I.$$

Suponemos que hay un universo Ω en donde mediante un experimento se elige al azar un $\omega \in \Omega$ y que nosotros observamos

- $X(\omega)$ antes del experimento
- $X(\omega) = x \in I$ después del experimento

Lo que estudiamos a continuación es cómo se distribuyen las probabilidades en I

Variable uniforme en $[a, b]$

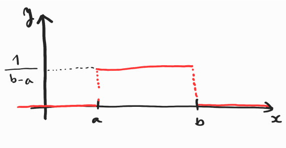
- Si X es tiene distribución uniforme en $[a, b]$, se verifica

$$\boxed{P(X \in [x, y]) = \frac{y - x}{b - a} \qquad a \leq x < y \leq b.}$$

- La densidad de la variable aleatoria X uniforme en $[a, b]$ es la función

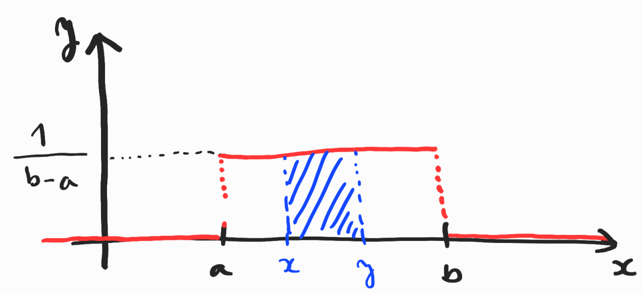
$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{cuando } a \leq x \leq b \\ 0, & \text{en otro caso} \end{cases}$$

se llama densidad de la variable aleatoria X, uniforme en $[a, b]$



- Las probabilidades se calculan como áreas:
- Para calcular áreas usamos integrales:

$$P(X \in [x, y]) = \int_x^y f(t)dt = \int_x^y \frac{1}{b-a}dt = \frac{y-x}{b-a}$$



Distribución de una variable aleatoria uniforme

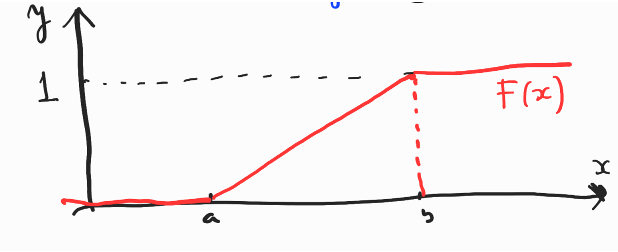
- Más en general nos interesa la siguiente función:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

llamada función de distribución o también función de distribución de probabilidades e la variable aleatoria X

Obtenemos

$$F(x) = \begin{cases} 0, & \text{cuando } x < a \\ \frac{x-a}{b-a} & \text{cuando } a \leq x \leq b \\ 1, & \text{cuando } b \leq x \end{cases}$$



- También podemos obtener la densidad f a partir de la distribución F.
- Sabemos que se verifica

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

- Si derivamos con respecto a x obtenemos

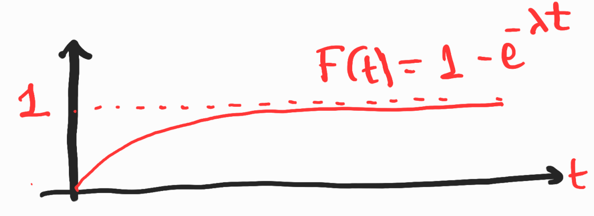
$$F'(x) = f(x)$$

- La densidad f y la distribución F tienen la misma información: se puede obtener una a partir de la otra

Variable aleatoria exponencial

- El segundo modelo continuo que consideramos es una variable aleatoria[9] T con distribución exponencial
- Es una variable que toma valores positivos: $I = [0, \infty)$
- Se puede definir a partir de su distribución:

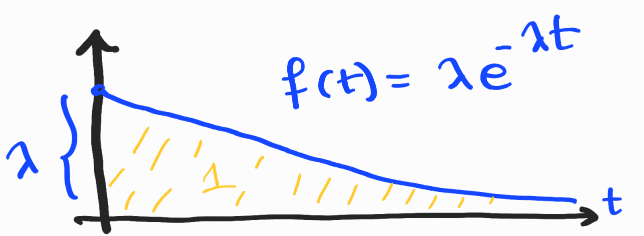
$$F(t) = P(T \leq t) = 1 - e^{-\lambda t} \quad (t \geq 0)$$



[9] usamos la T porque podría tratarse de un tiempo aleatorio

Derivando tenemos la densidad exponencial:

$$f(t) = F'(t) = \lambda e^{-\lambda t} \quad (t \geq 0)$$



- El modelo exponencial se utiliza para medir duraciones de componentes que no envejecen
- Por ejemplo, que se rompen por motivos externos a su desgaste
- Se puede usar para modelar el tiempo que dura una lámpara, que se quema por un pico de corriente externo
- Eso es porque cumple la propiedad de pérdida de memoria.
- Para eso usamos la sobrevida de T, que es

$$P(T > t) = e^{-\lambda t}.$$

Pérdida de memoria

- La pérdida de memoria es una propiedad matemática:

$$\boxed{P(T > t + h \mid T > t) = P(T > h)}$$

- Primero verificamos la propiedad:

$$\begin{aligned}
P(T > t + h \mid T > t) &= \frac{P(\{T > t + h\} \cap \{T > t\})}{P(T > t)} \\
&= \frac{P(\{T > t + h\})}{P(T > t)} \\
&= \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda(t+h)+\lambda t} \\
&= e^{-\lambda h} = P(T > h)
\end{aligned}$$

Ahora la interpretamos:

$$P(T > t + h \mid T > t) = P(T > h)$$

- Si T no ocurrió en tiempo t, esperar h más (hasta $t + h$) es equivalente a esperar h a partir de 0
- Por eso no importa cuanto tiempo ocurrió, si la lámpara no se rompió, es como que fuese nueva

Simulación de una variable exponencial

En R los cuatro comandos de una variable exponencial son:

- `dexp` me da la densidad exponencial
- `pexp` me da la distribución de probabilidades exponencial
- `qexp` me da la función inversa F^{-1} de la distribución F
- `rexp` me numeros aleatorios (random) con distribución exponencial
- Por ejemplo:

```
> rexp(3,1)
[1] 1.1611435 2.1357145 0.3589149
```

simula 3 variables exponenciales con parámetro $\lambda = 1$.

Aplicación: isótopos radiactivos

- El tiempo de desintegración de un isótopo radiactivo se modela mediante una v.a. exponencial de parámetro λ
- La vida media τ se define como $\tau = 1/\lambda$
- La vida media del Uranio 232 es de 70 años.

Problema: Cual es la probabilidad de que *n* átomos de Uranio (independientes) se desintegren todos antes de 100 años.

Solución.

- Sean T_k el tiempo de desintegración del átomo $k = 1, \ldots, n$.
- Cada T_k es una variable exponencial independiente de parámetro $\lambda = 1/70$.
- El tiempo se mide en años.

$$P(T_1 \leq 100) = 1 - e^{-\lambda 100} = 1 - e^{-100/70} = 0{,}76$$

- Como queremos que todos se desintegren antes de 100, queremos calcular

$$P(T_1 \leq 100, \cdots, T_n \leq 100) \\ = P(T_1 \leq 100) \cdots P(T_n \leq 100) = 0{,}76^n.$$

- En $=$ usamos la independencia.
- Si fuesen $n = 10$ átomos, tenemos

$$0{,}76^{10} = 0{,}06.$$

Ejemplo: las aguavivas inmortales

- Las aguavivas Turritopsis nutricula tiene un ciclo de vida en el que se revierte a pólipo después de llegar a su maduración sexual, presentándose como biológicamente inmortal: Turritopsis nutricula
- A pesar de esta remarcable habilidad, la mayoría de medusas Turritopsis suelen caer víctimas de las amenazas habituales de la vida del plancton, incluyendo ser comido por otros animales, o sucumbir a una enfermedad.
- ¿Se podría utilizar un modelo exponencial para modelar el tiempo de vida de estas aguavivas?

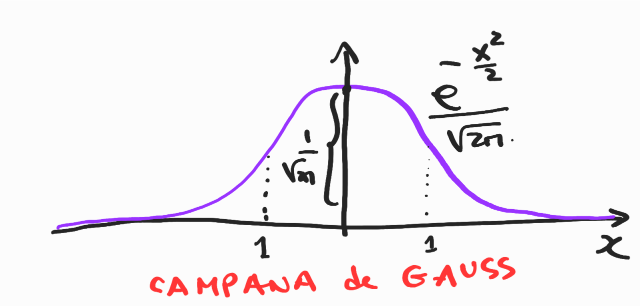
Variable aleatoria normal

Todo el mundo utiliza la variable aleatoria normal, para modelar errores de medición:

- los matemáticos porque piensan que es un hecho experimental,
- y los experimentadores porque suponen que es un teorema matemático.

Definición: Una variable aleatoria X es normal estándar o gaussiana si tiene densidad dada por

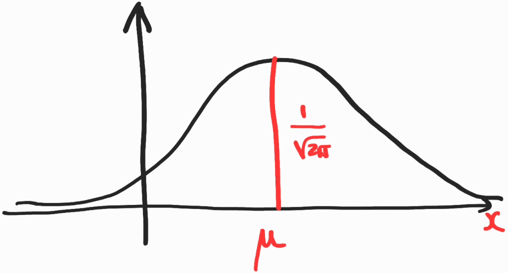
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$



Normal con parámetros μ,1

Nos interesa que el valor central sea un número arbitrario μ. Obtenemos la fórmula

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

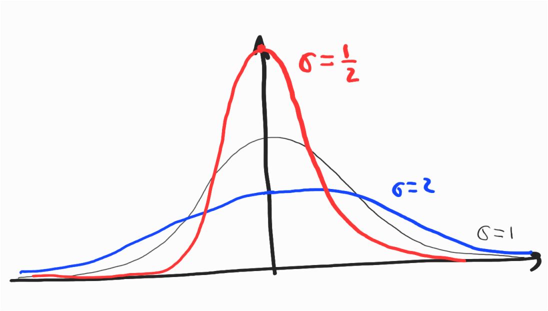


Tenemos $f(\mu) = \frac{1}{\sqrt{2\pi}}$

Normal con parámetros 0,σ

Nos interesa que la variable tenga dispersión arbitraria. Sea para eso $\sigma > 0$, y

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



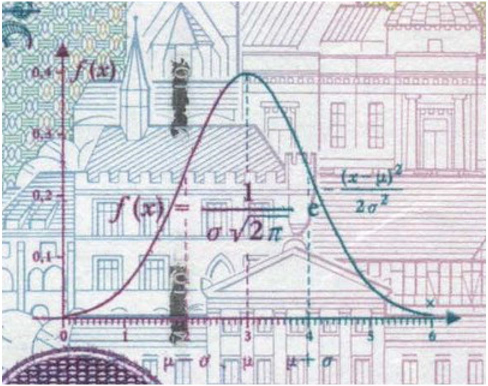
Tenemos $f(0) = \frac{1}{\sigma\sqrt{2\pi}}$

Normal con parámetros μ,σ

Si la posición y la dispersión son arbitrarias, tenemos la densidad

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

Figura: Zehn Deutsche Marks banknote



0,4
$f(x)$
0,3
0,2
0,1
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
0
1
2
3
4
5
6
x
$\mu - \sigma$
μ
$\mu + \sigma$

Áreas y probabilidades

$$\mathbf{P}(a-\sigma \leq X \leq a+\sigma) = 0{,}68$$
$$\mathbf{P}(a-1{,}96\sigma \leq X \leq a+1{,}96\sigma) = 0{,}95$$
$$\mathbf{P}(a-3\sigma \leq X \leq a+3\sigma) = 0{,}997$$

En el gráfico 3.7, el área de la figura delimitada por el gráfico de la función

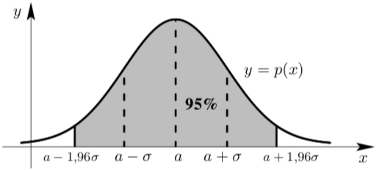


Figura 3.7: Gráfico de la densidad normal $p(x)$ con parámetros (a, σ). El área sombreada es el 95 % del área total

Simulación de una variable normal

En R los cuatro comandos de una variable normal son:

- `dnorm` me da la densidad normal
- `pnorm` me da la distribución de probabilidades normal
- `qnorm` me da la función inversa F^{-1} de la distribución F
- `rnorm` me numeros aleatorios (random) con distribución exponencial
- Por ejemplo:

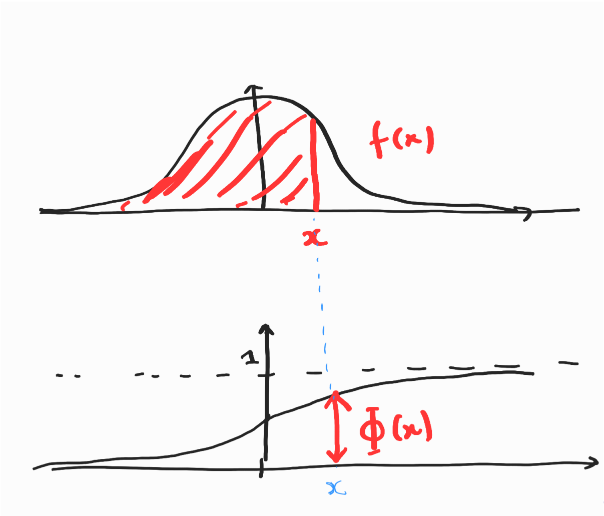
```
> dnorm(0,mean=0,sd=1)
[1] 0.3989423
```

$$\frac{1}{\sqrt{2\pi}} \sim \frac{1}{2{,}5} = 0{,}4.$$

Distribución normal

La distribución normal es la integral de la densidad normal:

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



Clase 9 de Bioestadística

Esperanza de una variable aleatoria

Ernesto Mordecki

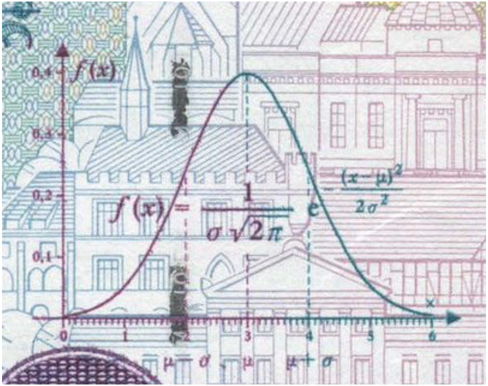
CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Variable aleatoria normal (repaso)

Definición: Una variable aleatoria X es normal o gaussiana con parámetros μ y σ^2 si tiene densidad dada por

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



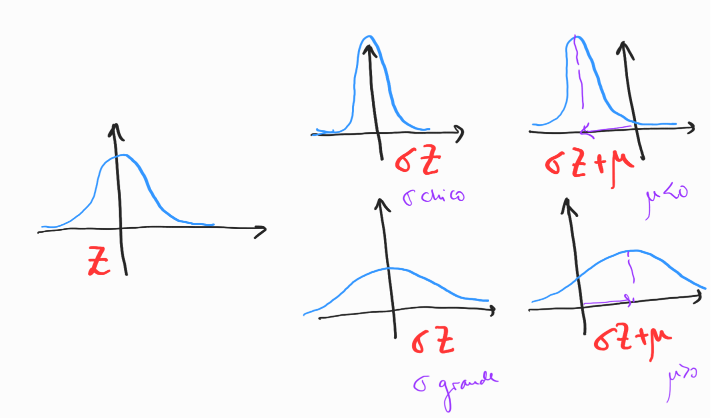
Los parámetros

- El parámetro de posición μ indica el punto de mayor densidad
- El parámetro de escala σ indica la concentración de la probabilidad alrededor de μ:
  - Si σ es pequeño, hay gran concentración,
  - Si σ es grande, hay gran dispersión
- Cuando $\mu = 0$ y $\sigma = 1$ tenemos una variable normal estándar

- La notación es
$$X \sim \mathcal{N}(\mu, \sigma^2)$$
se lee
  - X tiene distribución normal con parámetros μ, σ^2,
  - X es normal μ, σ^2.
- Para las normales estándar usamos la letra Z, es decir
$$Z \sim \mathcal{N}(0, 1)$$

Estandarización

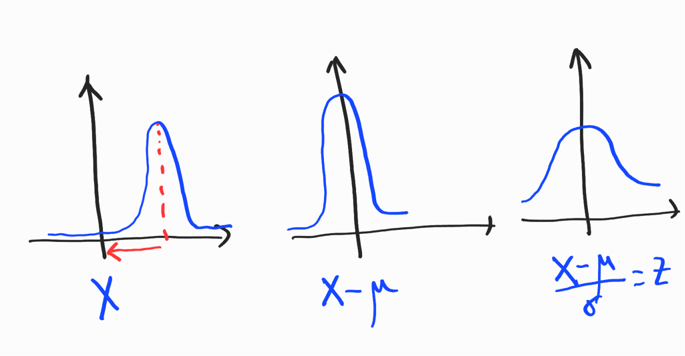
Es el pasaje de una normal estándar $(0, 1)$ a una (μ, σ^2):



Entonces

$$Z \sim \mathcal{N}(0, 1) \qquad \text{implica} \qquad X = \sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2)$$

Ahora pasamos de $X \sim \mathcal{N}(\mu, \sigma^2)$ a $Z \sim \mathcal{N}(0, 1)$



Entonces

$$X \sim \mathcal{N}(\mu, \sigma^2) \qquad \text{implica} \qquad Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Esperanza

- Queremos definir un número que represente el resultado de un experimento aleatorio
- Buscamos entonces un valor central, o medio, que pueda representar el experimento antes que este se realice.
- Si tenemos una variable normal, ese número podría ser μ
- ¿Cómo hacemos con una variable aleatoria genérica X?

Definición Llamamos esperanza de una variable aleatoria X al número $E(X)$ definido como

$$E(X) = \sum_k x_k P(X = x_k)$$

si X es una variable discreta y

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

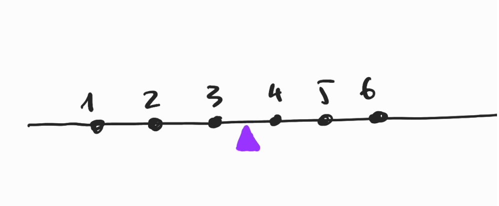
donde f es la densidad de X

También se llama esperanza matemática, o valor esperado.

Esperanza de una variable aleatoria discreta

Supongamos que X es el resultado de tirar un dado. Tenemos

$$\begin{aligned} E(X) &= \sum_k x_k P(X = x_k) \\ &= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} \\ &= \frac{1}{6}(1 + \cdots + 6) = \frac{1}{6}\left(\frac{6 \times 7}{2}\right) = 3{,}5 \end{aligned}$$



Esperanza de una variable uniforme

En el caso de una variable X uniforme discreta, que toma los valores

$$1, 2, \ldots, N$$

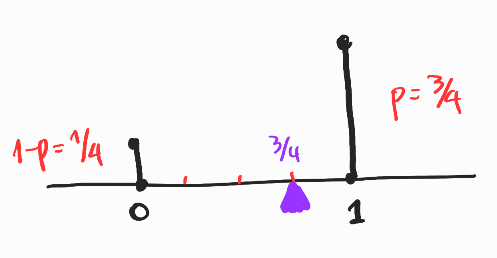
con probabilidad $1/N$ cada uno, tenemos

$$
\begin{aligned}
E(X) &= \sum_{k=1}^{k=N} k \; P(X = k) = \frac{1}{N} \sum_{k=1}^{k=N} k \\
&= \frac{1}{N} (1 + \cdots + N) = \frac{1}{N} \frac{N \times (N+1)}{2} = \frac{N+1}{2}
\end{aligned}
$$

Esperanza de una variable de Bernoulli

Si X tiene distribución de Bernoulli, toma los valores 0 y 1 con probabilidades $1 - p$ y p. Entonces

$$E(X) = 0 \times (1 - p) + 1 \times p = p.$$



Esperanza de una variable aleatoria Binomial

- X cuenta los éxitos en una serie de n experimentos de Bernoulli independientes.

- Entonces toma los valores $0, 1, \ldots, n$, con probabilidades

$$\boxed{P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n}$$

- Nuestra fórmula nos da

$$\begin{aligned} E(X) &= \sum_{k=0}^{k=n} {\color{blue}k}\ P(X = k) \\ &= \sum_{k={\color{red}0}}^{k=n} {\color{blue}k} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k={\color{red}1}}^{k=n} {\color{blue}k} \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

Calculemos

- Primero vemos que

$$
\begin{aligned}
k\binom{n}{k} &= k\,\frac{n!}{k!(n-k)!} = k\,\frac{n!}{k(k-1)!(n-k)!} \\
&= \frac{n!}{(k-1)!(n-k)!} = \frac{n(n-1)!}{(k-1)!(n-k)!} \\
&= \frac{n(n-1)!}{(k-1)!((n-1)-(k-1))!} \\
&= n\binom{n-1}{k-1}
\end{aligned}
$$

Entonces

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{k=n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{k=n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\
&\overset{\ell=k-1}{=} np \boxed{\sum_{\ell=0}^{\ell=n-1} \binom{n-1}{\ell} p^{\ell} (1-p)^{n-k}} \\
&= np.
\end{aligned}
$$

Cálculo alternativo

- Supongamos que $Y_1, \ldots, Y_n$ toman valores $0, 1$ según el experimento sea fracaso o éxito.
- Tenemos
$$X = Y_1 + \cdots + Y_n$$
para contar los éxitos.
- Si fuese cierto que
$$E(X) = E(Y_1 + \cdots + Y_n) \stackrel{?}{=} E(Y_1) + \cdots + E(Y_n)$$
- Como tenemos $E(Y_k) = p$ siendo n sumandos, obtenemos
$$E(X) = np.$$

Esperanza de una variable geométrica

- Una variable geométrica cuenta los fracasos antes de un éxito.
- Verifica

$$P(X = k) = (1 - p)^k p, \qquad k = 0, 1, \ldots$$

- Sumando una serie[10]

$$E(X) = \sum_{k=0}^{\infty} k P(X = k) = \sum_{k=0}^{\infty} k (1 - p)^k p \stackrel{1}{=} \frac{1 - p}{p}.$$

[10] La suma de esta serie requiere estudio de series de potencias

Esperanza de una variable de Poisson

- La variable de Poisson tiene probabilidades

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

- Asumamos la suma de la serie

$$\boxed{e^{\lambda} = \sum_{k=0}^{k=\infty} \frac{\lambda^k}{k!} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots}$$

- Mutliplicando por $e^{-\lambda}$ obtenemos

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

- Calculemos la esperanza

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

$$= \sum_{k=1}^{k=\infty} e^{-\lambda} \frac{\lambda \lambda^{k-1}}{(k-1)!} \overset{\ell=k-1}{=} \lambda \sum_{\ell=0}^{\ell=\infty} e^{-\lambda} \frac{\lambda^\ell}{\ell!} = \lambda$$

- Como conclusión el parámetro λ es la esperanza de una variable exponencial.

Andrei Nicolaievich Kolmogorov (1903-1987, Rusia)

Figura: Fundador de la probabilidad moderna: propuso un sistema axiomático en 1933

Clase 10 de Bioestadística

Ley fuerte de los grandes números

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Esperanza (repaso)

Definición: Llamamos esperanza matemática de una variable aleatoria X al número $E(X)$ definido como

$$E(X) = \sum_k x_k P(X = x_k)$$

si X es una variable discreta y

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

donde f es la densidad de X

La esperanza (o valor esperado) es un número representativo o central de todos los resultados posibles de una variable aleatoria X

Ejemplos (repaso)

- Si X es el resultado de tirar un dado, $E(X) = 3,5$
- Si X es una variable uniforme discreta en $1, 2, \ldots, N$ entonces $E(X) = \frac{N+1}{2}$
- Si X es una variable de Bernoulli, con parámetro p, entonces $E(X) = p$
- Si X es una variable binomial, con parámetros n, p entonces $E(X) = np$

- Si X es una variable geométrica de parámetro p entonces[11]:

$$\boxed{E(X) = \frac{1-p}{p}}$$

  - Si p es pequeño (tengo poca suerte) $E(X)$ es grande, y tengo que esperar bastante
  - Si $p \sim 1$ entonces $E(X) \sim 0$ (ya el primer experimento es éxito)

- Para una variable de Poisson de parámetro λ, entonces $E(X) = \lambda$

- Por último creemos que si $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces $E(X) = \mu$

[11] Ojo: la clase pasada me equivoqué, esta es la fórmula correcta.

Variables aleatorias independientes

Definición: Dos variables X e Y son independientes cuando generan sucesos independientes. Es decir, cuando todos los sucesos de la forma

$$a < X \leq b, \qquad c < Y \leq d$$

son independientes.

- Por ejemplo si N es el resultado del **dado negro**, e R el del dado rojo, son variables aleatorias independientes.

- Si tiramos un tercer dado azul y llamamos Z al resultado, tenemos (es sencillo de probar) que todos los sucesos de la forma

$$a_1 < X \leq b_1, \qquad a_2 < Y \leq b_2, \qquad a_3 < Z \leq b_3$$

  son independientes, es decir

$$P(a_1 < X \leq b_1, a_2 < Y \leq b_2, a_3 < Z \leq b_3)$$
$$= P(a_1 < X \leq b_1) \times P(a_2 < Y \leq b_2) \times P(a_3 < Z \leq b_3)$$

- Análogamente podemos definir que las v.a. $X_1, \dots, X_n$ son independientes
- Y si tenemos una sucesión infinita $X_1, X_2, \dots$ decimos que son independientes cuando cualquier conjunto finito de ellas es independiente.
- Por ejemplo, si tiramos un dado (hasta aburrirnos) tenemos una sucesión de variables aleatorias independientes, al ir anotando los resultados.

Variables aleatorias idénticamente distribuídas

Definición: Decimos que un conjunto de variables aleatorias $X_1, X_2, \ldots$ son idénticamente distribuídas cuando todas son resultados de experimentos diferentes que tienen los mismos resultados posibles con las mismas probabilidades.

- Por ejemplo N y R (los dados negro y rojo) son tienen la misma distribución.
- Las variables con la misma distribución:
  - No dan los mismos resultados
  - Tienen la misma esperanza

- Un objeto central (abstracto) en estadística es una sucesión
$$X_1, X_2, \ldots, X_n, \ldots$$
de variables aleatorias independientes e idénticamente distribuídas

- Representa los resultados de una serie arbitrariamente larga de experimentos que somos capaces de realizar
  - en forma independiente
  - bajo las mismas condiciones experimentales (idéntica distribución)

- Usamos la notación (X_n) v.a.i.i.d.

- Se lee “variables aleatorias independientes e idénticamente distribuídas”

- ¿Ejemplos?

Ley fuerte de los grandes números[13]

Teorema

- Consideremos una sucesión de v.a.i.i.d:

$$X_1, X_2, \ldots$$

- Supongamos que la esperanza de $E(X_1)$ es finita[12].
- Entonces
$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to E(X_1).$$
cuando n es arbitrariamente grande, tiende a infinito, $(n \to \infty)$

---

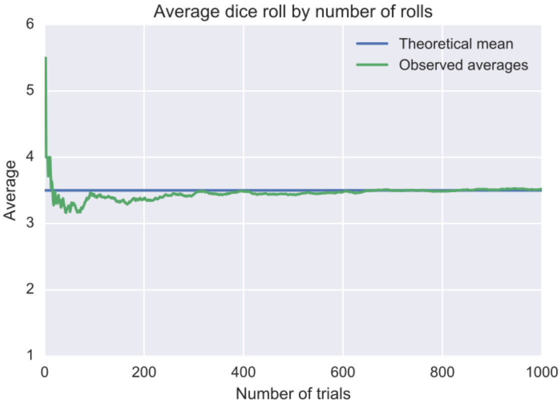
[12] Recordar que las esperanzas son todas iguales por tener la misma distribución

[13] 1930, Kolmogorov

Comentarios

- Se puede abreviar los promedios observados:
$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$
- Si bien los promedios observados son diferentes el límite siempre es el mismo
- Además, ese límite es la esperanza $E(X_1)$:
$$\overline{X}_n \to E(X_1), \qquad (n \to \infty)$$
- Esto tiene una aplicación estadística: para conocer $E(X_1)$ (desconocido) podemos aproximarlo por el promedio de experimentos (que podemos realizar)
- Por último, no todas las variables aleatorias tienen esperanza finita.

Veamos una ilustración:



Clase 11 de Bioestadística

Esperanza de variables aleatorias continuas

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Esperanza (repaso)

Definición: Llamamos esperanza matemática de una variable aleatoria X al número $E(X)$ definido como

$$E(X) = \sum_k x_k P(X = x_k)$$

si X es una variable discreta y

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

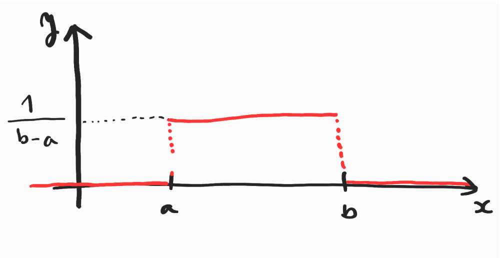
donde f es la densidad de X

Para calcular la esperanza matemática de una variable continua tenemos que conocer la la densidad de la variable aleatoria X.

Ejemplo: variable unforme en $[a, b]$

- Si U es una variable uniforme en $[a, b]$ su densidad vale

$$f(x) = \frac{1}{b-a}, \qquad a \leq x \leq b.$$



- ¿dónde debería estar ubicada la esperanza?
- ¿cuánto debería valer?

Hagamos las cuentas:

Partimos de la definición:

$$E(U) = \int_{-\infty}^{\infty} x\, f(x) dx = \int_a^b x\, f(x) dx$$

$$= \int_a^b x\, \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x\, dx$$

$$= \frac{1}{b-a} \left[\frac{x^2}{2} \right]_{x=a}^{x=b} = \frac{1}{b-a} \left[\frac{b^2 - a^2}{2} \right]$$

$$= \frac{1}{\cancel{b-a}} = \frac{a+b}{2}$$

Aplicación: Ley fuerte de los grandes números (LFGN)

- Sea $U_1, U_2, \ldots$ una sucesión de variables aleatorias independientes e idénticamente distribuídas, uniformes en el intervalo $[1, 3]$

- Nos intersa aproximar el promedio

$$\overline{U}_n = \frac{U_1 + \cdots + U_n}{n}$$

- Aplicamos la LFGN: sabemos que $E(U_1) = 2$.

- Obtenemos

$$\overline{U}_n \sim 2$$

- La calidad de la aproximación dependerá de n (y del azar)

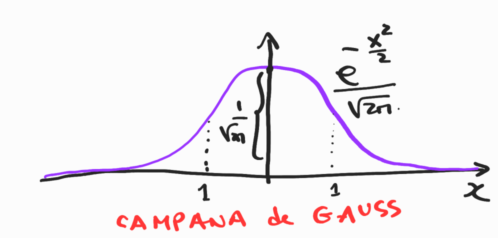
Simulamos en R

```
>
> mean(runif(10,min=1,max=3))
[1] 1.8904
> mean(runif(100,min=1,max=3))
[1] 2.00482
> mean(runif(1000,min=1,max=3))
[1] 1.993221
> mean(runif(10000,min=1,max=3))
[1] 2.002744
> mean(runif(100000,min=1,max=3))
[1] 2.002927
> mean(runif(1e6,min=1,max=3))
[1] 1.999075
> mean(runif(1e7,min=1,max=3))
[1] 2.000061
> |
```

Esperanza de una variable aleatoria normal

- Comenzamos con una variable normal estándar Z, es decir $Z \sim \mathcal{N}(0,1)$ y su densidad vale[14]

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad x \in \mathbb{R}.$$



- ¿dónde debería estar ubicada la esperanza?

- ¿cuánto debería valer?

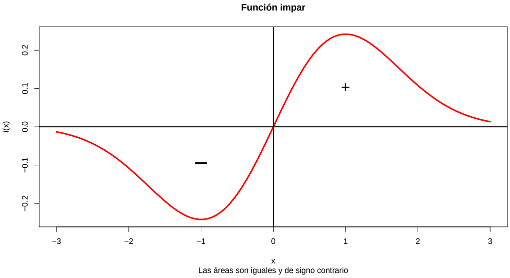
[14] Se utiliza φ para la densidad de $Z \sim \mathcal{N}(0,1)$

Vamos a calcular:

$$E(Z) = \int_{-\infty}^{\infty} x\ \varphi(x)dx = \int_{-\infty}^{\infty} x\ \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x\ e^{-x^2/2} dx = 0,$$

por simetría (el integrando es una función impar):



El código que me produce el gráfico

```
# grafico una función
# lo primero es definirla
i<-function(x) {x*dnorm(x)}
# le indico donde quiero guardar el pdf que produzco
setwd("/home/mordecki/Dropbox/1021_Bioestadistica
      /clases_teoricas/figuras")
# este comando abre un PDF de tamaño 11 x 6
pdf("impar.pdf",width = 11, height = 6)
# el comando "curve" grafica una función
curve(i,-3,3,lwd=3,col="red",main="Función impar",
      sub="Las áreas son iguales y de signo contrario")
# así dibujo los ejes:
abline(h=0,lwd=2)
abline(v=0,lwd=2)
# ahora agrego texto en el dibujo
text(1,0.1,"+",cex=2)
text(-1,-0.1,"-",cex=3)
# por último cierro el PDF y queda guardado en el lugar indicado
dev.off()
```

Esperanza de una normal (μ, σ^2)

Para pasar al caso general, recordamos el resultado de estandarización

$$X \sim \mathcal{N}(\mu, \sigma^2) \quad \text{si y solo si} \quad Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

¿Como se traduce este resultado en términos de densidades?

- Sea $\varphi(x)$ la densidad de Z y $f(x)$ la densidad de X. Tenemos

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)\right]^2$$

- Es decir

$$f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$$

- Donde decía z pusimos $\frac{x-\mu}{\sigma}$. Es un cambio de variable.
- Ahora podemos calcular $E(X)$ para $X \sim \mathcal{N}(\mu, \sigma^2)$

- Según la definición

$$E(X) = \int_{-\infty}^{\infty} x\, f(x) dx = \int_{-\infty}^{\infty} x\, \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right) dx$$

- Ahora cambiamos la variable en la integral de acuerdo a

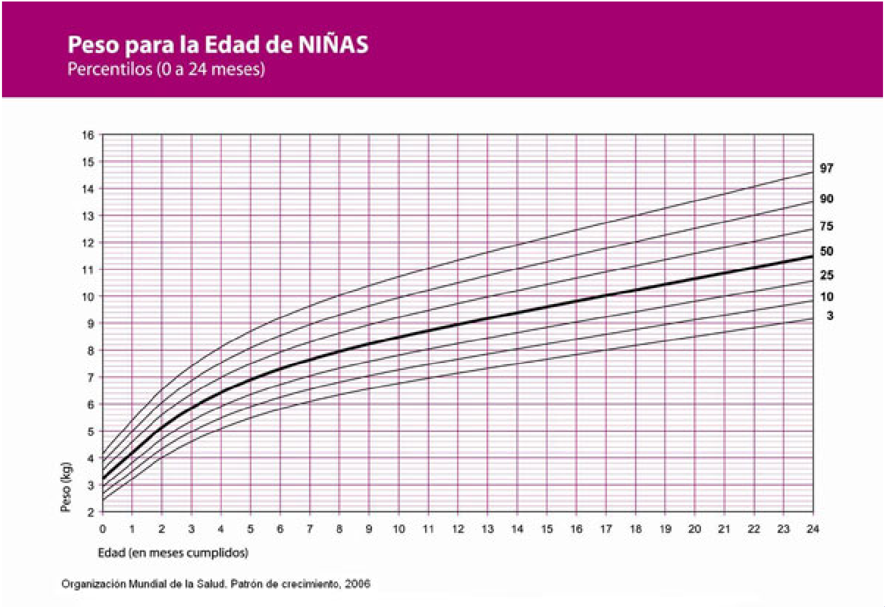
$$z = \frac{x-\mu}{\sigma}, \quad dz = \frac{1}{\sigma} dx, \quad x = \sigma z + \mu.$$

- Los límites de integración no cambian. Entonces

$$E(X) = \int_{-\infty}^{\infty} (\sigma z + \mu)\, \varphi(z) dz$$

$$= \sigma \boxed{\int_{-\infty}^{\infty} z\, \varphi(z) dz} + \mu \boxed{\int_{-\infty}^{\infty} \varphi(z) dz} = \mu.$$

Comentario: pesos de las bebés



Esperanza de una variable exponencial

- En este caso T tiene densidad
$$f(t) = \lambda e^{-\lambda t}$$

- Entonces
$$E(T) = \int_{-\infty}^{\infty} t\ f(t) dt = \lambda \int_{0}^{\infty} t\ e^{-\lambda t} dt$$

- Cambiamos de variable
$$t = \frac{v}{\lambda}, \quad dt = \frac{dv}{\lambda}, \quad v = \lambda t,$$

- Obtenemos
$$E(T) = \lambda \int_{0}^{\infty} \frac{v}{\lambda}\ e^{-v} \frac{dv}{\lambda} = \frac{1}{\lambda} \boxed{\int_{0}^{\infty} v\ e^{-v} dv}$$

La integral en la caja se calcula por partes:

$$\int_a^b f(v)g'(v)dv = f(b)g(b) - f(a)g(a) - \int_a^b f'(v)g(v)dv$$

Tomamos entonces

$$f(v) = v, \quad g'(v) = e^{-v}$$

$$f'(v) = 1, \quad g(v) = -e^{-v}$$

Obtenemos

$$\int_0^\infty v\, e^{-v}dv = \lim_{v\to\infty}(-ve^{-v}) - 0e^{-0} - \int_0^\infty(-e^{-v})dv$$

$$= \int_0^\infty e^{-v}dv = \left[-e^{-v}\right]_{v=0}^{v=\infty} = 1.$$

Conclusión

$$\boxed{E(T) = \frac{1}{\lambda}.}$$

Aplicación: la medusa inmortal

- La medusa inmortal (Turritopsis Nutricula) se considera bioógicamente inmortal.
- Dado que se reproducen, de ser inmortales su cantidad aumentaría exponencialmente.
- Consideramos que mueren entonces por predación o por enfermedades.
- Debido a que no envejecen, podemos suponer que tienen un tiempo de vida exponencial
- Suponiendo que en promedio viven un siglo, ¿cuál es la probabilidad de que un ejemplar viva mas de 500 años?

Solución[15]

- Si T representa el tiempo de vida de un ejemplar, tenemos, medido en años:

$$E(T) = \frac{1}{\lambda} = 100.$$

- Tenemos que calcular

$$P(T > 500) = e^{-\lambda \times 500} = e^{-500/100} = e^{-5} = 0{,}007$$

[15] Ojo: el ejemplo es una especulación:

Clase 12 de Bioestadística

Varianza de una variables aleatoria
Desigualdad de Chebishev

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Esperanza (repaso)

Definición: Llamamos esperanza matemática de una variable aleatoria X al número $E(X)$ definido como

$$E(X) = \sum_k x_k P(X = x_k)$$

si X es una variable discreta y

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

donde f es la densidad de X

Para calcular la esperanza matemática de una variable continua tenemos que conocer la la densidad de la variable aleatoria X.

Varianza de una variable aleatoria

- La esperanza matemática $E(X)$ de una variable aleatoria resume su posición
- Nos interesa ahora describir cuan lejos están los valores posibles que toma X de $E(X)$.
- Es decir, nos interesa resumir todos los desvíos $X - E(X)$ en un único número
- Para eso vamos a tomamos un promedio de los desvíos: para que no se eliminen los positivos con los negativos[16], tomamos las cantidades $(X - E(X))^2$

[16] El cuadrado "se come" los signos.

- Definimos entonces la varianza de una variable aleatoria X mediante

$$\boxed{Var(X) = E\left[(X - E(X))^2\right]}$$

- Como la varianza tiene unidades al cuadrado, definimos también el desvío estándar de X mediante

$$\boxed{\sigma_X = \sqrt{Var(X)} = \sqrt{E\left[(X - E(X))^2\right]}}$$

- Observación: tanto la varianza como el desvío estándar son siempre cantidades positivas.

Cálculo de la varianza

- Como definimos la varianza de X como la esperanza de $Y = (X - E(X))^2$, podemos usar la fórmula de esperanza.
- En el caso discreto tenemos entonces

$$Var(X) = \sum_k (x_k - E(X))^2 P(X = x_k)$$

si X es una variable discreta y

$$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$$

donde f es la densidad de X

Ejemplo: variable de Bernoulli

- Consideremos una variable X de Bernoulli, que toma los valores 0 y 1, con probabilidades

$$P(X = 0) = 1 - p, \qquad P(X = 0) = p.$$

- Habíamos calculado $E(X) = p$
- Entonces

$$\begin{aligned} Var(X) &= ({\color{blue}0} - p)^2(1-p) + ({\color{blue}1} - p)^2 p \\ &= p^2 - p^3 + (1 - 2p + p^2)p \\ &= p^2 \cancel{-p^3} + p - 2p^2 + \cancel{p^3} = p - p^2 = \boxed{p(1-p)} \end{aligned}$$

Ejemplo: el dado

- El caso del dado es similar pero con 6 resultados. Sabemos que $E(X) = 3{,}5$, y todas las probabilidades son iguales:

$$\begin{aligned} Var(X) &= \frac{1}{6}\left[(1 - 3{,}5)^2 + (2 - 3{,}5)^2 + \cdots + (6 - 3{,}5)^2\right] \\ &= 2{,}92 \end{aligned}$$

Ejemplo: variable Binomial

- En el caso de una variable Binomial B con parámetros n y p, obtenemos
$$Var(B) = np(1-p)$$

- Vale la pena observar que una variable Binomial cuenta los éxitos en n ensayos de Bernoulli:
$$B = X_1 + \cdots + X_n$$
donde X_k es el resultado de cada ensayo, por lo que $Var(X_k) = p(1-p)$.

- El resultado es
$$Var(B) = Var(X_1) + \cdots + Var(X_n) = np(1-p)$$

Otras variables discretas

- El cálculo de la varianza de variables discretas se hace mediante suma de series
- En el caso de una variable geométrica G con parámetro $p \in (0, 1)$ se obtiene

$$Var(G) = \frac{1-p}{p^2}$$

- En el caso de una variable P de Poisson de parámetro $\lambda > 0$, se obtiene

$$Var(P) = \lambda.$$

Vale la pena observar que $E(G) = Var(G)$.

Varianza de variables continuas

- En este caso tenemos que utilizar la fórmula con la integral y la densidad.

$$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$$

- Comencemos con una variable U uniforme en $[0, 1]$. Sabemos que $E(X) = \frac{1}{2}$.

- Tenemos entonces

$$
\begin{aligned}
Var(U) &= \int_{-\infty}^{\infty} \left(x - \frac{1}{2}\right)^2 f(x)dx \\
&= \int_0^1 \left(x - \frac{1}{2}\right)^2 1dx \\
&= \int_0^1 \left(x^2 - x + \frac{1}{4}\right) dx \\
&= \left[\frac{x^3}{3}\right]_0^1 - \left[\frac{x^2}{2}\right]_0^1 + \frac{1}{4} \qquad = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} \\
&= \frac{4}{12} - \frac{3}{12} + \frac{3}{12} = \boxed{\frac{1}{12}}
\end{aligned}
$$

Propiedades de la varianza

Propiedad Sea X una variable aleatoria y dados a, b números reales consideremos la variable aleatoria

$$Y = aX + b$$

Entonces

$$E(Y) = aE(X) + b, \qquad Var(Y) = a^2 \; Var(X)$$

- Observemos que la esperanza de Y se traslada b y se multiplica por a.
- Mientras que la varianza se multiplica por b^2 pero no depende de a.

Demostración

- Vamos a demostrarlo en el caso discreto. El caso continuo es similar. Comenzamos con la esperanza:

$$\begin{aligned} E(aX+b) &= \sum_k (ax_k+b)P(X=x_k) \\ &= a\boxed{\sum_k x_k P(X=x_k)} + b\boxed{\sum_k P(X=x_k)} \\ &= a\boxed{E(X)} + b\boxed{1} = aE(X)+b. \end{aligned}$$

Ahora la varianza:

$$
\begin{aligned}
Var(aX+b) &= E\left[\{aX+b-E(aX+b)\}^2\right] \\
&= E\left[\{aX+\cancel{b}-aE(X)-\cancel{b}\}^2\right] \\
&= E\left[\{aX-aE(X)\}^2\right] \\
&= E\left[\{a(X-E(X))\}^2\right] \\
&= a^2 E\left[\{X-E(X)\}^2\right] = a^2\ Var(X).
\end{aligned}
$$

$\square$

Varianza de una variable uniforme

Para una variable uniforme se puede demostrar que:

$$V = a + (b - a)U \sim U[a, b] \Leftrightarrow U \sim U[0, 1]$$

Entonces

$$Var(V) = (b - a)^2 Var(U) = \frac{(b - a)^2}{12}.$$

Varianza de una variable normal

- Sea $Z \sim \mathcal{N}(0,1)$, es decir, una normal estándar.
- Comenzamos enunciando que

$$Var(Z) = \int_{-\infty}^{\infty} (x - 0)^2 \varphi(x) dx = 1$$

- Entonces si $X \sim \mathcal{N}(\mu, \sigma^2)$, por estandarización se puede escribir

$$X = \sigma Z + \mu$$

- Calculamos

$$Var(X) = Var(\sigma Z + \mu) = \sigma^2 Var(Z) = \sigma^2.$$

- Concluímos que si $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces

$$E(X) = \mu, \qquad Var(X) = \sigma^2.$$

Desigualdad de Chebishev

Teorema. Sea X una variable aleatoria que toma valores no negativos, es decir, $X \geq 0$. Entonces, para cualquier $\varepsilon > 0$ tenemos

$$\boxed{P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}}$$

Demostración. Hacemos la demostración en el caso continuo.

$$\begin{aligned} E(X) &= \int_0^\infty xf(x)dx \geq \int_\varepsilon^\infty x f(x)dx \\ &\geq \int_\varepsilon^\infty \varepsilon f(x)dx = \varepsilon \int_\varepsilon^\infty f(x)dx \\ &= \varepsilon P(X \geq \varepsilon) \end{aligned}$$

Por último, si pasamos ε dividiendo tenemos la desigualdad buscada. $\square$

Desigualdad de Markov

Corolario: Sea Y una variable aleatoria y $Var(Y)$ su varianza. Entonces

$$\boxed{P(|Y - E(Y)| \geq \varepsilon) \leq \frac{Var(Y)}{\varepsilon^2}}$$

Demostración:

- Observemos que

$$\boxed{P(|Y - E(Y)| \geq \varepsilon)} = P(|Y - E(Y)|^2 \geq \varepsilon^2)$$

- La variable $X = (Y - E(Y))^2$ es no negativa, entonces

$$P((Y - E(Y))^2 \geq \varepsilon^2) \boxed{\leq} \frac{E[(Y - E(Y))^2]}{\varepsilon^2} = \boxed{\frac{Var(Y)}{\varepsilon^2}}$$

$\square$

Paul Pierre Lévy (1886-1971, Francia)

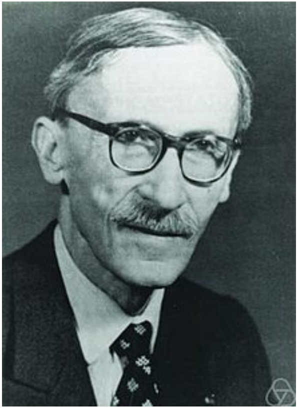


Figura: Demostró el teorema que veremos hoy

Clase 13 de Bioestadística

Teorema Central del Límite

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Esperanza y varianza (repaso)

Llamamos esperanza matemática de una variable aleatoria X al número $E(X)$ definido como

$$E(X) = \sum_k x_k P(X = x_k), \qquad \int_{-\infty}^{\infty} x f(x) dx$$

si X es una variable discreta o continua, y llamamos varianza al número

$$Var(X) = E\left[(X - E(X))^2\right]$$

Sucesión de v.a.i.i.d.

- El objeto central de la clase de hoy es una sucesión de variables aleatorias

$$X_1, X_2, \ldots, X_n, \ldots$$

que suponemos

- Independientes: las probabilidades de sucesos de n variables son el producto de las probabilidades individuales. Por ejemplo

$$P(X_1 \leq a, X_2 \leq b, X_3 \leq c) = P(X_1 \leq a) \times P(X_2 \leq b) \times P(X_3 \leq c)$$

- Idénticamente distribuídas: si bien los resultados son distintos, tienen las mismas probabilidades de ocurrir:

$$P(X_k \leq a) = P(X_1 \leq a) = F(a) \qquad \text{para todo } k = 1, 2, \ldots$$

- Las sucesiones de variables aleatorias independientes e idénticamente distribuídas

  v.a.i.i.d.

  son una idealización teórica de un experimento que se repite una gran cantidad de veces bajo las mismas condiciones.

- El ejemplo que consideramos es tirar un  una gran cantidad de veces

- Podríamos modelar la vacunación masiva de una población y medir la respuesta en anticuerpos al año de aplicar la vacuna

- ¿Ejemplos?

Propiedades

- Como tenemos la misma distribución, tienen la misma esperanza y la misma varianza. Les llamamos

$$\mu = E(X_1), \qquad \sigma^2 = var(X_1).$$

- Ahora vamos a quedarnos con las n primeras variables i.i.d:

$$X_1, X_2, \ldots, X_n$$

- Nos interesa en realidad el promedio

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- Queremos calcular la esperanza y la varianza del promedio en términos de μ y σ^2

Propiedades[17]

Teorema

(a) Sean $Y_1, \ldots, Y_n$ variables aleatorias cualesquiera. Tenemos

$$E(Y_1 + \cdots + Y_n) = E(Y_1) + \cdots + E(Y_n)$$

(b) Sean X e Y variables independientes. Entonces

$$E(X \times Y) = E(X) \times E(Y)$$

[17] Suponemos que las esperanzas y varianzas que aparecen son finitas

Demostraciones

- La propiedad (a) vamos a asumirla verdadera.
- Veamos (b) en el caso que ambas variables sean discretas y finitas.
- X toma valores $x_1, \dots x_n$ con probabilidades $p_1, \dots, p_n$, Y toma valores $y_1, \dots, y_k$ con probabilidades $q_1, \dots, q_k$
- Entonces XY toma los valores $x_i y_\ell$ con probabilidades[18] $p_i q_\ell$
- Tenemos

$$E(XY) = \sum_{i,\ell} x_i y_\ell \times p_i q_\ell = \sum_{i,\ell} (x_i p_i)(y_\ell q_\ell)$$
$$= \sum_i x_i p_i \sum_\ell y_\ell q_\ell = E(X)E(Y)$$

[18] Suponemos que todos los productos dan distinto resultado

Covarianza

- Definimos la covarianza de dos v.a. X e Y mediante

$$cov(X, Y) = E\left[(X - E(X))(Y - E(Y))\right]$$

- Notemos que si $X = Y$ entonces

$$\begin{aligned} cov(X, X) &= E\left[(X - E(X))(X - E(X))\right] \\ &= E\left[(X - E(X))^2\right] = var(X) \end{aligned}$$

- Operemos:

$$
\begin{aligned}
cov(X,Y) &= E\left[(X - E(X))(Y - E(Y))\right] \\
&= E\left[XY - YE(X) - XE(Y) + E(X)E(Y)\right] \\
&= E\left[XY\right] - E\left[YE(X)\right] - E\left[XE(Y)\right] + E\left[E(X)E(Y)\right] \\
&= E\left[XY\right] - E[Y]E(X) - E[X]E(Y) + E(X)E(Y) \\
&= E(XY) - E(X)E(Y)
\end{aligned}
$$

- Conclusión: si X, Y son independientes, entonces

$$\boxed{cov(X,Y) = 0}$$

Prop. Sean X e Y variables independientes. Entonces

$$var(X+Y) = var(X) + var(Y)$$

Dem. Es una cuenta:

$$\begin{aligned}
var(X+Y) &= E\left[(X+Y-E(X+Y))^2\right] \\
&= E\left[(X-E(X)+Y-E(Y))^2\right] \\
&= E\left[(X-E(X))^2+(Y-E(Y))^2\right. \\
&\qquad \left. +2\,(X-E(X))(Y-E(Y))\right] \\
&= var(X)+var(Y)+2cov(X,Y) \\
&= var(X)+var(Y).
\end{aligned}$$

$\square$

Generalización

Prop. Sean $X_1, \ldots, X_n$ variables independientes. Entonces

$$var(X_1 + \cdots + X_n) = var(X_1) + \cdots + var(X_n)$$

Dem. Sea $X = X_1$, $Y = X_2 + \cdots + X_n$. Las variables X e Y son independientes, entonces

$$var(\underbrace{X_1}_{=X} + \underbrace{X_2 + \cdots + X_n}_{=Y}) = var(X_1) + var(X_2 + \cdots + X_n)$$

- Un paso mas igual nos da

$$var(X_1 + X_2 + \cdots + X_n) = var(X_1) + var(X_2) + var(X_3 + \cdots + X_n)$$

- Lo hacemos tantas veces como sea necesario. $\square$

- Estamos en condiciones de calcular la esperanza y la varianza de
$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$
para $X_1, \ldots, X_n$, i.i.d.

- Recordemos que
$$E(aX + b) = aE(X) + b, \qquad var(aX + b) = a^2 var(X).$$

- Entonces
$$E(\overline{X}_n) = E\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n} E\left(X_1 + \cdots + X_n\right)$$
$$= \frac{1}{n}[E(X_1) + \cdots + E(X_n)] = \frac{1}{n}[nE(X_1)] = \mu.$$

- La esperanza del promedio es igual a la esperanza de los sumandos.

La varianza de un promedio

- Tenemos

$$var(\overline{X}_n) = var\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2} var\left(X_1 + \cdots + X_n\right)$$
$$= \frac{1}{n^2}\left[var(X_1) + \cdots + var(X_n)\right] = \frac{1}{n^2}\left[n var(X_1)\right] = \frac{\sigma^2}{n}.$$

- La varianza del promedio es la n-ésima parte de la varianza de cada sumando
- Cuantos más sumandos menos varianza

El problema de la velocidad

- Según la ley fuerte de los grandes números sabemos que el promedio se aproxima a la esperanza
- Es decir
$$\overline{X}_n \to \mu \qquad \text{si } n \text{ tiende a infinito}$$
- Pero . . . , cuánto hay que esperar?
- ¿Cuántos experimentos hay que hacer?
- Cuán pequeño es el error
$$\overline{X}_n - \mu$$

- En primer lugar sabemos que

$$E\left(\overline{X}_n - \mu\right) = E\left(\overline{X}_n\right) - E(\mu) = \mu - \mu = 0.$$

- En cuanto a la varianza

$$var\left(\overline{X}_n - \mu\right) = var\left(\overline{X}_n\right) = \frac{1}{n}\sigma^2.$$

- Multiplico por σ^2 y divido por n, y aplico la propiedad,

$$\frac{\sigma^2}{n} var\left(\overline{X}_n\right) = 1, \qquad var\left(\frac{\sqrt{n}}{\sigma}\left[\overline{X}_n - \mu\right]\right) = 1$$

- Consideremos la variable

$$Z_n = \frac{\sqrt{n}}{\sigma}\left[\overline{X}_n - \mu\right]$$

- Tiene esperanza cero y varianza uno.

Teorema Central del Límite

- Paul Lévy demostró que

$$Z_n \to Z \sim \mathcal{N}(0,1) \qquad \text{cuando } n \text{ tiende a infinito}$$

- Eso quiere decir que

$$P(a \leq Z_n \leq b) \to P(a \leq Z \leq b) = \int_a^b \varphi(x)dx$$

- Podemos escribir entonces, si n es grande

$$Z_n = \frac{\sqrt{n}}{\sigma}\left[\overline{X}_n - \mu\right] \approx Z$$

- Podemos escribir entonces, si n es grande

$$Z_n = \frac{\sqrt{n}}{\sigma}\left[\overline{X}_n - \mu\right] \approx Z$$

- Despejando

$$\left[\overline{X}_n - \mu\right] \approx \frac{\sigma}{\sqrt{n}} Z$$

- Despejando

$$\overline{X}_n \approx \mu + \frac{\sigma}{\sqrt{n}} Z$$

- Es decir, si sustituímos μ (desconocido) por $\overline{X}_n$ cometemos un error de la forma

$$\frac{\sigma}{\sqrt{n}} Z$$

- Cuanto más grande es n menor es el error
- El error decrece como la raíz de n.

Aplicación

- Queremos simular Z
- Sabemos simular uniformes
- Resulta que 12 uniformes aproximan bien a Z

$$Z_{12} = \frac{\sqrt{12}}{\sigma}\left[\overline{X}_{12} - \frac{1}{2}\right] \approx Z$$

- Como la varianza es $\sigma^2 = 1/12$, tenemos $\sigma = \sqrt{1/12}$ se multiplica con $\sqrt{n}$
- En conclusión

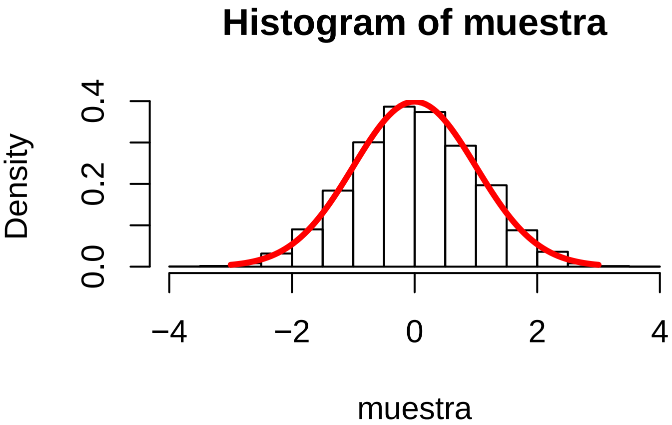
$$Z_{12} = 12\left[\overline{X}_{12} - \frac{1}{2}\right] \approx Z$$

Escribo el siguiente código:

```
# simula una normal Z:
12*mean(runif(12)-0.5)
# preparo un vector para una muestra
muestra<-c()
# tamaño de la muestra
tamano<-10000
# repito la simulacion "tamano" veces
# y guardo el resultado en muestra
for(i in 1:tamano){
  valor<-12*mean(runif(12)-0.5)
  muestra<-c(muestra,valor)
}
# calculo el histograma de mi muestra
hist(muestra,freq = F)
# lo comparo con la densidad normal estándar
curve(dnorm,-3,3,add=T,lwd=3,col="red")
```

Simulo 10000 muestras con 12 uniformes

El gráfico compara el histograma de la muestra con la densidad normal estándar:



La aproximación indica que los valores simulados son aproximadamente normales estándar.

Clase 14 de Bioestadística

Estadística: estimación de la esperanza y la varianza

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Probabilidad y Estadística

- La probabilidad y la estadística son disciplinas hermanas
- Comparten un mismo marco teórico matemático:
  - un espacio de sucesos Ω
  - un modelo de probabilidad P
- Se diferencian en las pregunas que formula cada una:
  - La probabilidad conoce P y calcula a partir de ella
  - La estadística conoce valores $X_1, \ldots, X_n$ y se ocupa de estimar[19] P.
  - La estimación de la probabilidad se designa $\widehat{P}$.

[19] Se trata de dar una buena aproximación del modelo

Ejemplo 1: Modelo normal

Consideramos un ejemplo clásico

- Suponemos el modelo gaussiano o normal donde desconocemos la media μ y la varianza σ^2
- Observamos (conocemos) n valores numéricos de variables aleatorias $X_1, \ldots, X_n$ que corresponden al modelo incóngnita
- ¿Cómo estimamos los parámetros μ y σ^2 a partir de los datos $X_1, \ldots, X_n$?

Ejemplo 2: Estimación de una probabilidad

Consideramos otro ejemplo clásico

- Suponemos el modelo Bernuolli donde desconocemos la probabilidad p
- Observamos (conocemos) n valores numéricos de variables aleatorias $X_1, \ldots, X_n$ que corresponden al modelo incóngnita, que corresponden a una tira de ceros (fracasos) y unos (éxitos)
- ¿Cómo estimamos el parámetro p a partir de esa tira de ceros y unos?

Terminología estadísitica

- El conunto Ω se llama espacio muestral. En probablidad se llama espacio de sucesos
- Los valores $X_1, \ldots, X_n$ se llaman muestra aleatoria simple, se resume m.a.s[20]
- Si el parámetro desconocido se designa θ, la aproximación que hacemos se designa $\widehat{\theta}$.

[20] Es lo que en probabilidad se llama v.a.i.i.d.

Estimación de la esperanza

- La base de la estimación es la ley fuerte de los grandes números (LFGN):
- Sean $X_1, \ldots, X_n$ una m.a.s. de v.a. con esperanza desconocida $\mu = E(X_1)$.
- La LFGN establece que:

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \to E(X_1)$$

- Si tenemos un modelo con una probabilidad P con $\mu = E(X_1)$ desconocido, podemos estimar μ a traves del promedio de los datos

$$\widehat{\mu} = \overline{X}_n$$

- La LFGN nos dice entonces que

$$\widehat{\mu} \to \mu,$$

que se denomina consistencia del estimador.

Estimación de una probabilidad

- Supongamos un modelo de Bernoulli con probabilidad p desconocida.
- Tenemos una m.a.s. $X_1, \ldots, X_n$. Se verifica
$$E(X_1) = p$$
- Entonces se reduce al caso anterior: ¡la probabilidad es la esperanza!

- Entonces, el estimador de la probabilidad es

$$\widehat{p} = \frac{X_1 + \cdots + X_n}{n} = \frac{\text{cantidad de éxitos}}{n}$$

- El estimador también se denomina frecuencia observada de la cantidad de éxitos

- Además, la LFGN nos dice que el estimador es consistente, es decir

$$\widehat{p} \to p.$$

Estimación de la varianza con media conocida

- Tenemos ahora una m.a.s. $X_1, \ldots, X_n$, suponemos que $\mu = E(X_1)$ es conocida, pero $\sigma^2 = var(X_1)$ es desconocida
- Queremos entonces estimar σ^2.
- Recordamos para eso que σ^2 es en realidad una esperanza, de la variable aleatoria

$$Y = (X - E(X))^2 = (X - \mu)^2$$

- Producimos entonces una m.a.s. de Y, mediante

$$Y_i = (X_i - \mu)^2$$

- Estimamos entonces la varianza mediante

$$\widehat{\sigma^2} = \frac{(X_1 - \mu)^2 + \cdots + (X_n - \mu)^2}{n}$$

- Aplicando la LFGN sabemos que

$$\widehat{\sigma^2} = \frac{(X_1 - \mu)^2 + \cdots + (X_n - \mu)^2}{n} \to E(X - \mu)^2 = \sigma^2,$$

es decir, tenemos consistencia.

Estimación de la varianza con media **desconocida**

- La situación más frecuente es cuando no conocemos ni la media μ ni la varianza σ^2 de la variable que queremos estimar.
- Por lo tanto no conocemos μ para ponerlo en la fórmula anterior.
- La idea es entonces primero estimar μ, y luego utilizarlo para estimar la varianza.
- El resultado de este procedimiento es

$$\widehat{\sigma^2} = \frac{(X_1 - \widehat{\mu})^2 + \cdots + (X_n - \widehat{\mu})^2}{n}$$

- En este caso las variables

$$Y_i = (X_i - \widehat{\mu})^2$$

no conforman una muestra aleatoria simple, porque no son independientes.

- Se puede probar sin embargo que

$$\widehat{\sigma^2} = \frac{(X_1 - \widehat{\mu})^2 + \cdots + (X_n - \widehat{\mu})^2}{n} \to \sigma^2,$$

también tenemos consistencia.

Estimación del desvío estándar

- Una vez estimada la varianza, para estimar el desvío estándar se toma la raíz:

$$\widehat{\sigma_X} = \sqrt{\widehat{\sigma^2}} = \sqrt{\frac{(X_1 - \widehat{\mu})^2 + \cdots + (X_n - \widehat{\mu})^2}{n}}$$

$$= \sqrt{\frac{1}{n}\sum_{k=1}^{n}(X_k - \widehat{\mu})^2}$$

Correlación y su estimación

- En el caso en que dos variables no son independientes, decimos que son dependientes
- Una medida de su dependencia es la covarianza, que según vimos se define como

$$cov(X, Y) = E\left[(X - E(X))(Y - E(Y))\right]$$

- Con el objetivo de obtener una cantidad adimensionada, definimos la correlación

$$\rho(X, Y) = \frac{E\left[(X - E(X))(Y - E(Y))\right]}{\sigma_X \sigma_Y}$$

donde σ_X y σ_Y son los desvíos estándar de X e Y respectivamente.

Propiedades

- Se puede demostrar que la correlación es una cantidad que se encuentra entre -1 y 1, es decir

$$-1 \leq \rho(X, Y) \leq 1.$$

- La correlación positiva indica que a valores mayores de una se obtienen valores mayores de la otra,
- La correlación negativa indica lo contrario: valores mayores que la esperanza de una corresponden a valores menores que la esperanza de la otra
- Cuanto más cerca de uno (o de -1) sea el valor de la correlación, mas pronunciado será este fenómeno.

Estimación

- Para estimar la covarianza suponemos conocer una muestra aleatoria simpe del vector aleatorio (X, Y), es decir, v.a.s
$$(X_1, Y_1), \ldots, (X_n, Y_n)$$
- Aquí se supone que las parejas son independientes entre sí, pero no tiene que serlo cada X_k de su compañera Y_k

Estimación con medias y desvíos conocidos

- Supongamos primero que conocemos $\mu_X = E(X_1)$, $\mu_Y = E(Y_1)$ así como las varianzas $\sigma_X^2 = var(X_1)$ y $\sigma_Y = var(Y_1)$
- En ese caso la correación es la esperanza de las variables

$$Z = \frac{(X - \mu_X)(Y - \mu_Y)}{\sigma_X \sigma_Y}$$

- Entonces conseguimos un estimaror consistente aplicando la LFGN:

$$\widehat{\rho}(X, Y) = \frac{1}{n} \sum_{k=1}^{n} \frac{(X_k - \mu_X)(Y_k - \mu_Y)}{\sigma_X \sigma_Y}$$

Estimación: caso general

- En la situación se desconocen las esperanzas y las varianzas.
- La estimación entonces comienza estimando las esperanzas μ_X μ_Y y los desvíos σ_X , σ_Y.
- Primero estimamos las esperanzas:

$$\widehat{\mu_X} = \overline{X}_n, \qquad \widehat{\mu_Y} = \overline{Y}_n$$

- y luego los desvíos estándar

$$\widehat{\sigma_X} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(X_k - \widehat{\mu_X})^2}$$

$$\widehat{\sigma_Y} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(Y_k - \widehat{\mu_Y})^2}$$

- Finalmente construímos el estimador de la correlación:

$$\widehat{\rho}(X, Y) = \frac{1}{n}\sum_{k=1}^{n}\frac{(X_k - \widehat{\mu_X})(Y_k - \widehat{\mu_Y})}{\widehat{\sigma_X}\widehat{\sigma_Y}}$$

- El estimador de la correlación sirve para saber si dos variables son dependientes (¿cómo?)


¡TODO EL DÍA PENSANDO EN NÚMEROS, CIFRAS, CUENTAS, CÁLCULOS MATEMÁTICOS, PORCENTAJES! ¿NO TE DAS CUENTA DE CUÁNTO DAÑA NUESTRA RELACIÓN?
SÍ, UN 63%...
KAPPEL

Clase 15 de Bioestadística

Estadística: distribución empírica y cuantiles

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Probabilidad y Estadística (repaso)

- La probabilidad y la estadística son disciplinas hermanas
- Comparten un mismo marco teórico matemático:
  - un espacio de sucesos Ω
  - un modelo de probabilidad P
- Se diferencian en las preguntas que formula cada una:
  - La probabilidad conoce P y calcula a partir de ella
  - La estadística conoce valores $X_1, \ldots, X_n$ y se ocupa de estimar[21] P.
  - La estimación de la probabilidad se designa $\widehat{P}$.

[21] Se trata de dar una buena aproximación del modelo

Terminología estadísitica

- El conunto Ω se llama espacio muestral. En probablidad se llama espacio de sucesos
- Los valores $X_1, \ldots, X_n$ se llaman muestra aleatoria simple, se resume m.a.s[22]
- Si el parámetro desconocido se designa θ, la aproximación que hacemos se designa $\widehat{\theta}$.

[22] Es lo que en probabilidad se llama v.a.i.i.d.

Estimación de la esperanza

- La base de la estimación es la ley fuerte de los grandes números (LFGN):
- Sean $X_1, \ldots, X_n$ una m.a.s. de v.a. con esperanza desconocida $\mu = E(X_1)$.
- La LFGN establece que:

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \to E(X_1)$$

- Si tenemos un modelo con una probabilidad P con $\mu = E(X_1)$ desconocido, podemos estimar μ a traves del promedio de los datos

$$\widehat{\mu} = \overline{X}_n$$

- La LFGN nos dice entonces que

$$\widehat{\mu} \to \mu,$$

que se denomina consistencia del estimador.

Estimación de una probabilidad (repaso)

- Supongamos un modelo de Bernoulli con probabilidad p desconocida.
- Tenemos una m.a.s. $X_1, \ldots, X_n$. Se verifica

$$E(X_1) = p$$

- Entonces se reduce al caso anterior: ¡la probabilidad es la esperanza!

- Entonces, el estimador de la probabilidad es

$$\widehat{p} = \frac{X_1 + \cdots + X_n}{n} = \frac{\text{cantidad de éxitos}}{n}$$

- El estimador también se denomina frecuencia observada de la cantidad de éxitos

- Además, la LFGN nos dice que el estimador es consistente, es decir

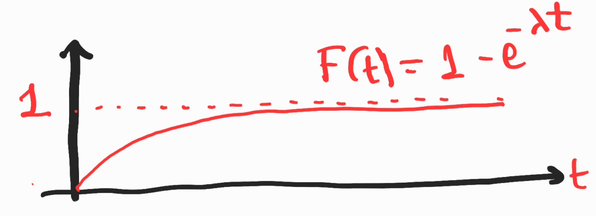
$$\widehat{p} \to p.$$

Estimación de la función de distribución

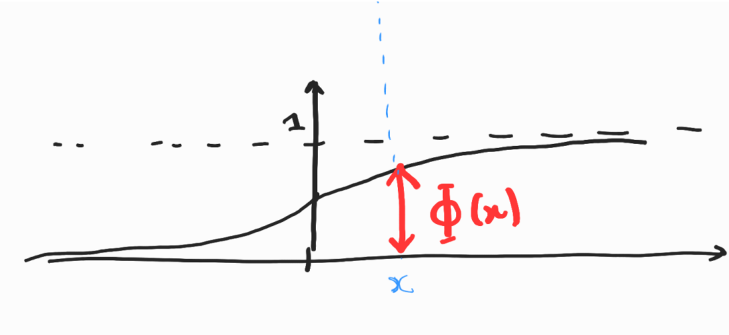
- Dada una variable aleatoria X, definimos su distribución F como la función tal que

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

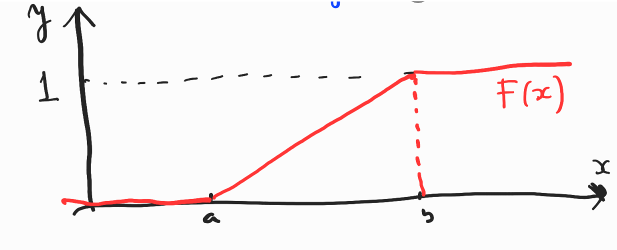
- Distribución exponencial:



- Distribución Normal:



- Distribución uniforme:



Propiedades de la función de distribución

- La distribución $F(x)$ es la probabilidad del suceso

$$\{X \leq x\},$$

por lo tanto

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

- Monotonía: Si $a < b$ tenemos la inclusión de sucesos

$$\{X \leq a\} \leq \{X \leq b\},$$

(si X (el resultado del experimento) es menor o igual que a y a es menor que b entonces el resultado X es menor o igual que b). Eso nos da

$$F(a) \leq F(b)$$

- Si $x = -\infty$ entonces $F(x) = 0$: nunca ocurre $X \leq -\infty$, entonces

$$\{X \leq -\infty\} = \emptyset, \quad \text{luego} \quad F(-\infty) = P(\emptyset) = 0$$

- Si $x = +\infty$ entonces $F(x) = 1$: siempre ocurre $X \leq +\infty$, entonces

$$\{X \leq +\infty\} = \Omega, \quad \text{luego} \quad F(+\infty) = P(\Omega) = 1.$$

Estimación de la función de distribución

Nos planteamos el siguiente problema:

- Tenemos una muestra aleatoria simple

$$X_1, \ldots, X_n$$

  que corresponde a una v.a. con distribución F
- Desconocemos F y queremos estimarla
- F es una función, empecemos por un valor $F(a)$.

- $F(a) = p$ es la probabilidad del suceso

$$\{X \leq a\}$$

- El estimador de una probabilidad p es su frecuencia, es decir, la proporción de casos favorables
- Ejemplo: Supongamos que $a = 0$ y nuestra muestra son los valores

$$0{,}55,\ 0{,}91,\ -0{,}16,\ -0{,}30,\ -0{,}83,\ -0{,}59,\ 0{,}45,\ -0{,}29$$

- ¿Cuáles son nuestros éxitos? Como $a = 0$ el suceso es

$$X \leq 0$$

es decir, son los valores negativos:

$$0{,}55,\ 0{,}91,\ -0{,}16,\ -0{,}30,\ -0{,}83,\ -0{,}59,\ 0{,}45,\ -0{,}29$$

- Entonces nuestra estimación es

$$\widehat{p} = \frac{5}{8}.$$

- Para el estimador de $F(a)$ se usa la notación $F_n(a)$, que se llama distribución empírica de F. Tenemos entonces

$$F_n(0) = \frac{5}{8}$$

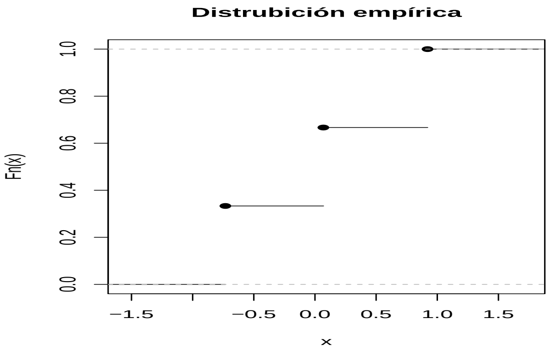
- Hacemos el mismo razonmiento ahora para cualquier a:
- Para estimar $F(a)$ contamos la cantidad de variables de la muestra que son menores o iguales que a y dividimos por el tamaño de la muestra, es decir

$$F_n(a) = \frac{\#\{k \colon X_k \leq a\}}{n}$$

- Miramos ahora la función $F_n(a)$ cuando varía a, si por ejemplo $n = 3$ y tenemos los valores

$$0{,}07,\ 0{,}92,\ -0{,}73$$

si $a < -0{,}73$ tenemos $F_3(a) = 0$.
- Si $-0{,}73 \leq a < 0{,}07$ tenemos $F_3(a) = 1/3$
- Si $0{,}07 \leq a < 0{,}92$ tenemos $F_3(a) = 2/3$
- Si $0{,}98 \leq a$ tenemos $F_3(a) = 1$



Para plotear la distribución empírica en R se usa el comando `ecdf`. Concretamente el código es

```
muestra<-rnorm(3,0,1)
p<-ecdf(muestra)
plot(p,main="Distribución empírica")
```

Consistencia

Ahora, ¿se parece la F_n (estimada) la F (teórica)?

- Para eso tenemos la indicatriz de un conjunto:

$$\mathbf{1}_A = \begin{cases} 1, & \text{si } x \in A \\ 0, & \text{si } x \notin A \end{cases}$$

- Es decir, sumamos

$$\#\{k \colon X_k \leq a\} = \sum_{k=1}^{n} \mathbf{1}_{\{X_k \leq a\}}$$

- Entonces tenemos un promedio, y aplicamos la LFGN:

$$F_n(a) = \widehat{p} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{X_k \le a\}} \to E\left(\mathbf{1}_{\{X_1 \le a\}}\right)$$

- Como

$$E\left(\mathbf{1}_{\{X_1 \le a\}}\right) = 0 \times P(X_1 > a) + 1 \times P(X_1 \le a) = P(X_1 \le a) = F(a)$$

obtenemos la consistencia de la estimación:

$$F_n(a) \to F(a)$$

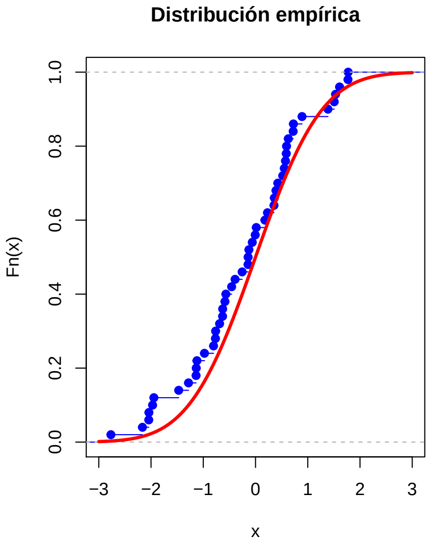


Figura: La distribución empírica (azul) se aproxima a la distribución teórica (roja)