Apuntes

de

Probabilidad

Basados en Notas de
Enrique Cabaña del año 1968

2024

Índice

12. Teorema del Límite Central

43

Prefacio

Estos son apuntes para la Unidad Curricular "Probabilidad y Estadística" del segundo semestre del año 2024 de la Facultad de Ingeniería de la Universidad de la República. Están están basadas en gran medida en las notas escritas por Enrique Cabaña para la correspondiente UC del año 1968, accesibles en su página web[1] o directamente aquí [2] pero con modificación en el orden para adaptarlas al actual curso y con la introducción de algunos temas, como Bayes, tablas con algunas distribuciones básicas y las cotas de Barry-Essen.

Eduardo A. Canale.
Montevideo 1ero de agosto de 2024.

1. Algunos problemas que podremos resolver al final de la primera parte de este curso

Problema 1: Tres amigos compiten entre sí en algún juego que se juegue entre dos, como el ping-pong. El que gana juega con el que quedó afuera. Ganará el campeonato cualquiera que gane dos partidos seguidos. ¿Qué conviene más, empezar jugando el primer partido o dejar que los otros dos jueguen dicho partido?

Problema 2: Ninguna moneda real es perfecta, ni ninguna forma de tirarla, por lo tanto la probabilidad de que salga alguno

[1] https://sites.google.com/site/emcabana/
[2] https://drive.google.com/file/d/1Cldc7lLVOVjUJx86hgELhmKcKp7nXMeq/view

de los dos lados será mayor que la probabilidad del otro lado. ¿Hay alguna forma de realizar un experimento con una moneda real (imperfecta) de modo que el experimento tenga solo dos posibles resultados y ambos con la misma probabilidad?

Problema 3: Me ofrecen ganar un premio si gano dos partidos (digamos de tenis) seguido de tres que debo jugar. Mis contrincantes son un amigo que juega más o menos como yo, llamémosle A y otro que juega bastante mejor que yo, llamémosle B. Me ofrecen dos posibilidades, jugar primero con A, luego con B y luego de vuelta con A, o primero con B, luego con A, y devuelta con B. ¿cuál de las dos opciones me conviene más?

2. Espacios de probabilidad

Los resultados de un experimento aleatorios los representaremos con un conjunto llamado **espacio muestral**, que normalmente denotaremos con la letra griega Ω (omega mayúscula).

Ejemplo 2: Si el experimento es tirar un dado, entonces lo resultados posibles son $1, 2, 3, 4, 5, 6$ de modo que el espacio muestra podría ser $\Omega = \{1, 2, 3, 4, 5, 6\}$. Si el mismo dado lo tiramos dos veces, el resultado lo podemos representar como los pares (x, y) con x e y algún número del 1 al 6, o sea el espacio muestral podría ser $\Omega = \{(x, y) : x, y \in \{1, 2, 3, 4, 5, 6\}\}$, aunque también podría ser $\Omega = \{1, 2, \ldots, 36\}$.

Nos interesará calcular la probabilidad de lo que llamaremos

sucesos. Por ejemplo, en el caso de un dado, un suceso podría ser "que salga par" o en el caso de dos dados, "que la suma de los dos dados sea cinco". En lugar de expresar los sucesos con palabras o mejor dicho, proposiciones, los representaremos como subconjuntos del espacio muestral. O sea, en lugar de decir $Q(x)$ con $Q(x) =$ "x es par", diremos que $x \in S$ con S el subconjunto de los números pares, o sea $S = \{\omega \in \Omega : Q(\omega)\} = \{0, \pm 2, \pm 4, \ldots\}$.

Ejemplo 3: Si el experimento es tirar un dado y el primer espacio muestral es el del ejemplo anterior, el suceso de que sea par será el subconjunto $S = \{2, 4, 6\}$ mientras el suceso de que la suma de dos dados sea cinco, en el segundo espacio muestral del ejemplo anterior, será el subconjunto $S = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$.

Al suceso vacío se lo llama **suceso imposible**, al suceso igual al espacio muestral se lo llama **suceso seguro**, y dos sucesos disjuntos se los los llama **sucesos incompatibles**.

Desgraciadamente, por razones técnicas, cuando el espacio muestral es infinito no numerable, ha sido imposible construir una teoría probabilística que considere todos los subconjuntos posibles, o sea, que en general, siempre quedarán subconjuntos raros, aunque fáciles de construir usando el axioma de elección, a los cuales no podremos calcularles su probabilidad, salvo casos triviales. Sin embargo, la teoría se ha logrado desarrollar sobre familias de subconjuntos lo suficientemente generales como para atacar todos los problemas prácticos que se han presentado hasta el momento. Dichas familias se llaman **σ-álgebras** y son aquellas familias no vacías cerrada bajo complementaciones y bajo uniones numerables. O sea, que si $\mathcal{A}$ es una de esas familias de partes[3] del conjunto Ω, ella debe no ser vacía y satisfacer las siguientes condiciones:

(a_1) Si $A \in \mathcal{A}$, entonces su complemento $A^c = \Omega \setminus A \in A$.

(a_2) Si $A_1, \ldots, A_n, \ldots \in \mathcal{A}$, entonces $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{A}$.

Ejemplo 4: Algunas σ-álgebras triviales son:
1) El conjunto $\mathcal{A} = 2^{\Omega}$ de todos los subconjuntos de Ω.
2) $\mathcal{A} = \{\varnothing, \Omega\}$ para cualquier Ω .
3) $\mathcal{A} = \{\varnothing, \{2024\}, \{2024\}^c, \mathbb{R}\}$.

La pareja $(\Omega, \mathcal{A})$ en la que Ω es un conjunto no vacío y $\mathcal{A}$ es una σ−álgebra de partes de Ω se llama **espacio probabilizable**, y los elementos de $\mathcal{A}$ se llaman **sucesos.**

De las propiedades anteriores se pueden deducir las siguientes cuya demostración se deja a cargo del lector

Proposición 2.1. *Toda σ−álgebra sobre Ω verifica las siguientes propiedades*

(a_0') *Contiene a Ω y al suceso vacío $\varnothing$.*

(a_2') *Si $A_1, \ldots, A_n \in \mathcal{A}$, entonces $\bigcup\limits_{i=1}^{n} A_i \in \mathcal{A}$.*

(a_3) *Si $A, B \in \mathcal{A}$, entonces $A \cap B \in \mathcal{A}$.*

(a_3') *Si $A_1, \ldots, A_n \in \mathcal{A}$, entonces $\bigcap\limits_{i=1}^{n} A_i \in \mathcal{A}$.*

□

[3] subconjuntos

Dado un espacio probabilizable $(\Omega, \mathcal{A})$, consideremos ahora una función $\mathsf{P} : \mathcal{A} \to \mathbb{R}^+$, tal que:

(p_1) $\mathsf{P}(\Omega) = 1$.

(p_2) si $A_1, A_2, \ldots$ son sucesos dos a dos incompatibles[4] entonces

$$\mathsf{P}\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathsf{P}(A_n).$$

Una función que satisface las propiedades (p_1) y (p_2), es decir que es una función de conjunto numerablemente aditiva, y normalizada en el sentido de que la imagen del conjunto Ω es 1, se llama **probabilidad**.

La terna $(\Omega, \mathcal{A}, \mathsf{P})$ en la que $(\Omega, \mathcal{A})$ es un espacio probabilizable y P es una probabilidad sobre $(\Omega, \mathcal{A})$ se llama **espacio de probabilidad**.

Ejemplo 5:

1) $\Omega = \{Cara, Cruz\}$, $\mathcal{A} = 2^\Omega = \{\varnothing, \{Cara\}, \{Cruz\}, \Omega\}$, $\mathsf{P}(\varnothing) := 0$, $\mathsf{P}(\{Cara\}) := \mathsf{P}(\{Cruz\}) := 1/2$, $\mathsf{P}(\Omega) := 1$.

2) $\Omega = \{0, 1\}, \mathcal{A} = 2^\Omega, \mathsf{P}(\varnothing) := 0, \mathsf{P}(\{0\}) := 1/10; \mathsf{P}(\{1\}) := 9/10; \mathsf{P}(\Omega) := 1$.

3) $\Omega = [0,1] \times [0,1]$, $\mathcal{A} = \{S$ con área medible$\}$, $\mathsf{P}(S) :=$ área(S).

A veces para simplificar la notación escribiremos $\mathsf{P}(a)$ en lugar de $\mathsf{P}(\{a\})$.

[4] $i \neq j \Rightarrow A_i \cap A_j = \varnothing$

Proposición 2.2. *Sea* $\Omega = \{w_1, w_2, \ldots, w_n, \ldots\}$ *un conjunto numerable,* $\mathcal{A} = 2^{\Omega}$ *la familia de todos lo subconjuntos de* Ω *y* $p_1, p_2, \ldots, p_n, \ldots$ *una sucesión de números no negativos tales que*

$$\sum_{n=1}^{\infty} p_n = 1,$$

entonces la función P *definida por*

$$\mathsf{P}\,(\,S\,) = \sum_{w_n \,\in\, S} p_n$$

es una probabilidad sobre $(\Omega, \mathcal{A})$. *En particular si* Ω *es finito con* $|\Omega| = n$ *elementos y todas las probabilidades* p_i *son iguales, o sea, iguales a* $1/n$*, entonces la suma anterior queda igual a*

$$\mathsf{P}\,(\,S\,) = \sum_{w \,\in\, S} \frac{1}{n} = \frac{|S|}{|\Omega|} = \frac{|S|}{n}.$$

En este último caso el espacio $(\Omega, \mathcal{A}, \mathsf{P})$ *es llama* **equiprobable**. $\square$

En el espacio de equiprobabilidades anterior, a los elementos de S se los llama **casos favorables** al suceso S y a los elementos de Ω **casos posibles**. Por ejemplo si queremos saber cual es la probabilidad de que al tirar un dado salga un número par, suponiendo que el dado no está "cargado", todos las caras tienen la misma probabilidad de salir, de modo que estamos en el caso anterior. Los casos favorables son si sale 2, 4 o 6, o sea que $S = \{2, 4, 6\}$ y los casos posibles son $\Omega = \{1, 2, 3, 4, 5, 6\}$, de modo que $\mathsf{P}\,(\,S\,) = |S|/6 = 3/6 = 1/2$. En estos casos de equiprobabilidad, el cálculo de probabilidades se reduce a problemas de conteo. Para ésto es útil conocer algunas reglas de conteo:

2.1. Muestreo y Distribución de bjetos

Dos modelos usuales en temas de conteo son el "muestreo" o también "selección" de objetos y la "distribución" de objetos en cajas. Por ejemplo, tenemos una urna con cierta cantidad de objetos, alguno de los cuales pueden ser indistinguibles, y seleccionamos (muestreamos) cierta cantidad. Podemos ir tomando los objetos en orden o de una sola vez. Si los tomamos de a uno, podemos volverlos a colocar en la urna o no. En esos casos diremos que el muestreo es con o sin reposición respectivamente. Al seleccionar los objetos, podemos considerar que el orden en que sale es importante, como cuando se hace un sorteo y se van eligiendo lo dígitos de un número muy grande, cada dígito se vuelve a colocar en la urna, o cuando se reparten cartas, que no importa el orden en que nos dan las cartas. En el caso de distribuir objetos en cajas, se pueden considerar los casos en que los objetos y/o las cajas son distinguibles o no. También se suele introducir cotas a la cantidad de objetos por caja.

Una técnica muy usual para contar es la: **regla del producto**: *Si un muestro se realiza en etapas y para cada etapa se tiene la misma cantidad de posibilidades, independientemente del resultado de las etapas anteriores, entonces la cantidad de posibilidades finales es el producto de las posibilidades de cada etapa.*

Ejemplo 6:

1) Para ver la cantidad de números en base b de n dígitos,

podemos idear una forma de construir cada uno de dichos números en n etapas, de modo que en la etapa i-ésima elegimos el i-ésimo dígito. Como siempre tenemos b formas de elegir un dígito, la cantidad total de posibilidades es b^n.

2) Contemos la cantidad de números en base b de n dígitos distintos. Nuevamente hacemos como antes, pero esta vez, en la etapa i-ésima tendremos $b-(i-1)$ posibles dígitos para elegir, sin importar cuales hayamos elegidos, pues los que elegimos deben ser todos distintos. Entonces la cantidad total es $b(b-1)(b-2)...(b-(i-1))$ siempre que $i \leq b$. Obviamente si $n > b$ nunca podremos elegir n dígitos distintos de entre b posibles, por lo que la cantidad será cero.

3) Contemos la cantidad de números en base b de n dígitos con dígitos adyacentes distintos. Nuevamente hacemos como antes, pero esta vez, en la etapa i-ésima tendremos $b-1$ posibles dígitos para elegir si $i > 1$, y b posibilidades en la etapa 1. Entonces la cantidad total es $b(b-1)^{n-1}$.

2.1.1. Arreglos y combinaciones

La cantidad de formas de seleccionar en forma ordenada n objetos con reposición de entre m objetos distinguibles es m^n. Dicha cantidad es llamada **arreglos con repeticion de** m **en** n y se denota a veces como AR_n^m.

La cantidad de formas de seleccionar en forma ordenada n objetos sin reposición de entre m objetos distinguibles es $m(m-1)\ldots(m-n+1)$. Dicha cantidad es llamada **arreglos (sin repetición) de** m **en** n y denotada a veces como A_n^m, a veces como $P_{m,n}$ y a veces como $(m)_n$. Si $m = n$ la cantidad es $m! = m(m-1)\dots 1$ y termina siendo la cantidad de formas de ordenar o **permutar** m objetos distinguibles.

La cantidad de formas de seleccionar sin importar el orden n objetos sin reposición de entre m objetos distinguibles es $A^m_n/n!$. Dicha cantidad es llamada **combinaciones (sin repetición) de** m **en** n y se denota a veces como C^m_n, a veces como $C_{m,n}$ y a veces $\binom{m}{n}$.

La cantidad de formas de seleccionar, sin importar el orden n objetos con reposición de entre m objetos distinguibles es C^{m+n-1}_n, denotada **combinaciones con repetición de** m **en** n y simbolizado a veces como CR^m_n.

2.2. Propiedades de la Probabilidad

A partir de los axiomas (p_1) y (p_2) de una función de probabilidad podemos deducir las siguientes propiedades, cuya demostración la dejamos para el lector.

Proposición 2.3. *Toda función* P *de probabilidades verifica:*

(p'_1) $\mathsf{P}(\varnothing) = 0$.

(p'_2) *Si* $A_1, \dots, A_n$ *son sucesos dos a dos incompatibles, entonces*

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{P}(A_i) .$$

(p_3) $\mathsf{P}(A^c) = 1 - \mathsf{P}(A)$.

(p_4) *Si* $A \subset B$, *entonces* $\mathsf{P}(A) \le \mathsf{P}(B)$.

(p_5) **Regla de Adición**:

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

(p_5') **Principio de Inclusión-Exclusión**:

$\mathsf{P}(A_1 \cup \cdots \cup A_n) = \mathsf{P}(A_1) + \cdots + \mathsf{P}(A_n) - \big(\mathsf{P}(A_1 \cap A_2) + \mathsf{P}(A_1 \cap A_3) + \cdots + \mathsf{P}(A_{n-1} \cap A_n)\big) + \cdots + (-1)^{n-1}\mathsf{P}(A_{n-1} \cap \cdots \cap A_n)$.

□

Teorema 2.4 (Continuidad de la Probabilidad).

(i) *Si* (A_n) *es una sucesión creciente de sucesos*[5] *y* $\bigcup_{n=1}^{\infty} A_n = A$, *entonces* $\mathsf{P}(A) = \text{lím}_n \mathsf{P}(A_n)$.

(ii) *Si* (B_n) *es una sucesión decreciente de sucesos y* $\bigcap_{n=1}^{\infty} B_n = B$, *entonces* $\mathsf{P}(B) = \text{lím}_n \mathsf{P}(B_n)$.

Dem: Está claro que por (p_4), las sucesiones $\mathsf{P}(A_n)$ y $\mathsf{P}(B_n)$ son monótonas. Además por ser probabilidades son acotadas, por lo que ambas sucesiones son convergentes. Lo que se desea probar es que los correspondientes límites son $\mathsf{P}(A)$ y $\mathsf{P}(B)$.

Vamos a probar primeramente que (i) implica (ii) (el argumento puede hacerse a la inversa, lo que prueba que en realidad (i) y (ii) son equivalentes).

Si $B_n \supset B_{n+1}$ entonces $B_n^c \subset B_{n+1}^c$. Además $\bigcup_{n=1}^{\infty} B_n^c = \left(\bigcap_{n=1}^{\infty} B_n\right)^c = B^c$, de modo que (i) implica $\mathsf{P}(B^c) = \text{lím}_n \mathsf{P}(B_n^c) = \text{lím}\, 1 - \mathsf{P}(B_n) = 1 - \text{lím}\, \mathsf{P}(B_n) \Rightarrow \text{lím}\, \mathsf{P}(B_n) = 1 - \mathsf{P}(B^c) = \mathsf{P}(B)$.

[5] $A_n \subset A_{n+1} \quad \forall n$

Para probar (i), definimos $C_n = A_n \setminus A_{n-1}$ para $n = 2, 3, \ldots$ y $C_1 = A_1$, de donde $A_n = \bigcup_{n=1}^{\infty} C_n$. Los conjuntos C_n que intervienen en esta unión son disjuntos, y entonces $\mathsf{P}(A) = \sum_{n=1}^{\infty} \mathsf{P}(C_n)$. Por otra parte $A_n = \bigcup_{i=1}^{n} C_i$, de donde $\mathsf{P}(A_n) = \sum_{i=1}^{n} \mathsf{P}(C_i)$ y entonces

$$\lim_{n\to\infty} \mathsf{P}(A_n) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathsf{P}(C_i) = \sum_{n=1}^{\infty} \mathsf{P}(C_n) = \mathsf{P}(A).$$

$\square$

3. *Probabilidades condicionales*

Dado un espacio de probabilidad $(\Omega, \mathcal{A})$, y un suceso A tal que $\mathsf{P}(A) > 0$, se define la función **probabilidad condicional dado** A, como la función

$$\mathsf{P}(\cdot \mid A) : \mathcal{A} \to \mathbb{R}^+$$

tal que

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

Ejemplo 7: Sea $(\Omega, 2^\Omega, \mathsf{P})$ el espacio de equiprobabilidad probabilidad sobre el conjunto Ω de todos los uruguayos. Consideremos el suceso A de aquellos que ganan más de un millón de dólares por año, y el suceso B de aquellos que son jugadores de fútbol de la Liga Española. La probabilidad de A es cercana a $0{,}3 \times 10^{-6}$, lo mismo para la probabilidad de B, y por lo tanto la probabilidad de $A \cap B$ será aún menor (será menor que ambas). Sin embargo la probabilidad de $A|B$ puede relativamente grande, sobre todo si se la compara con la probabilidad de $A|C$ siendo C el suceso de ser un profesor de la FING.

Ejercicio 2: Probar qué $\left(\Omega, \mathcal{A}, \mathsf{P}(\cdot \mid A)\right)$ es un nuevo espacio de probabilidad.

Ejercicio 3: Probar las siguientes propiedades: (se supondrá que las probabilidades condicionales que aparecen están definidas, es decir, que los sucesos "condicionantes" tienen probabilidad positiva)

(p_6) $\mathsf{P}(A_1 \cap A_2) = \mathsf{P}(A_1)\,\mathsf{P}(A_2 \mid A_1)$.

(p_6') $\mathsf{P}(A_1 \cap A_2 \cap \dots) = \mathsf{P}(A_1)\,\mathsf{P}(A_2 \mid A_1)\,\mathsf{P}(A_3 \mid A_2 \cap A_1) \dots$
(Regla de multiplicación de probabilidades).

(p_7) $\mathsf{P}(A \mid B) = \mathsf{P}(B \mid A)\,\mathsf{P}(A) / \mathsf{P}(B)$.

3.1. Fórmula de Bayes

Proposición 3.1 (Fórmula de la Probabilidad Total). *Si $A_1, A_2, \dots$ es una partición de $\mathcal{A}$ en conjuntos con probabilidad positiva entonces*

$$\mathsf{P}(B) = \mathsf{P}(B \mid A_1)\,\mathsf{P}(A_1) + \mathsf{P}(B \mid A_2)\,\mathsf{P}(A_2) + \cdots. \quad (1)$$

$\square$

Ejemplo 8: Queremos estimar a cuantos uruguayos son hinchas de cierto cuadro de fútbol. Para ello encuestamos a 99 y encontramos que 30 de ellos son hinchas de dicho cuadro. Entonces estimamos que la cantidad de uruguayos hinchas del cuadro serán $3\,444\,263 \times 30/99 \approx 1\,043\,716$. Sin embargo, quien hizo la encuesta se le ocurrió preguntarles si tenían menos de 15 años o más de 65 año o ninguna de las anteriores. En esos tres grupos habían 15 hinchas, en el segundo habían 10 y en el tercero solo 5, mientras que los tres grupos tenían 33 personas cada uno. Si del último censo sabemos que la distribución etaria del Uruguay es de $19{,}02\,\%$ menores de 15 años, $65{,}55\,\%$ el entre 15 y 64 años el y $15{,}43\,\%$ mayores de 65 años, entonces usando la fórmula de la probabilidad total estimaremos la proporción de hinchas (que será la probabilidad de elegir uno si elegimos a cualquier uruguayo con la misma probabilidad) de

$$p = 0{,}1902 \times (15/33) + 0{,}6555 \times (5/33) + 0{,}1 \times (10/33) = 0{,}2325$$

De modo que estimamos la cantidad de hinchas total como $3\,444\,263 \times 0{,}2325 = 800\,791$ bastante menor que la primer estimación. Decidir cual de las dos estimaciones es mejor o cual es el error que podemos estar cometiendo es un tema de la matemática estadística.

De ésta fórmula y de la propiedad (p_7) obtenemos la:

Proposición 3.2 (Fórmula de Bayes). *En las condicione de* *la proposición anterior se cumple que*

$$\mathsf{P}(A_1 \mid B) = \frac{\mathsf{P}(B \mid A_1)\,\mathsf{P}(A_1)}{\mathsf{P}(B \mid A_1)\,\mathsf{P}(A_1) + \mathsf{P}(B \mid A_2)\,\mathsf{P}(A_2) + \cdots}. \tag{2}$$

□

Ejemplo 9: Supongamos ahora que cierta página web preguntamos a quienes entran si son hinchas del cuadro del ejemplo anterior ¿cuál será la probabilidad de que un usuario que dijo ser hincha del cuado tenga menos de 15 años? Para ello usamos Bayes: B Sería el suceso de ser hincha del cuadro, A_1 el de tener menos de 15 años. De la encuesta estimamos que $\mathsf{P}(B \mid A_1) = 15/33$, mientras que del censo sabemos que $\mathsf{P}(A_1) = 0{,}1902$, de modo que

$$\mathsf{P}(A_1 \mid B) = \frac{(15/33) \times 0{,}1902}{0{,}2325} = 0{,}37.$$

O sea que habrá una probabilidad del 37 % de que sea menor de 15 años.

3.2. *Independencia*

Cuando $\mathsf{P}(B \mid A) = \mathsf{P}(B)$ se dice que el suceso B es independiente del suceso A. Usando la definición de $\mathsf{P}(B \mid A)$ resulta que esto implica $\mathsf{P}(B \cap A) = \mathsf{P}(B)\ \mathsf{P}(A)$, lo cual implica que si $\mathsf{P}(B) \neq 0$, también es A independiente de B. Teniendo en cuenta esto diremos que A, B son **sucesos independientes**, cuando $\mathsf{P}(A \cap B) = \mathsf{P}(A)\ \mathsf{P}(B)$.

Más en general, se dice que el conjunto $\mathcal{C} \subset \mathcal{A}$ es **independiente** (o que los sucesos de $\mathcal{C}$ son independientes) cuando para todo subconjunto finito $\{A_1, \ldots, A_k\} \subset \mathcal{C}$ se cumple

$$\mathsf{P}\left(A_1 \cap \cdots \cap A_k \right) = \mathsf{P}\left(A_1 \right) \times \cdots \times \mathsf{P}\left(A_k \right) .$$

4. *Variables aleatorias*

Dado un espacio probabilizable $(\Omega, \mathcal{A})$ se llama **variable aleatoria (real)** a cualquier función $X : \Omega \to \mathbb{R}$, tal que la preimagen de todo intervalo es un suceso.

En lo sucesivo los sucesos de la forma $\{\omega \in \Omega : Q(X(\omega))\}$, siendo $Q(x)$ una proposición, los escribiremos como $\{Q(X)\}$. Por ejemplo si $Q(x) =$ "$x < 3$", entonces $\{X < 3\} = \{\omega \in \Omega : X(\omega) < 3\}$. Y en lugar de escribir $\mathsf{P}\left(\{Q(X)\} \right)$, escribiremos $\mathsf{P}\left(Q(X) \right)$ o $\mathsf{P}\{Q(X)\}$. En el ejemplo sería $\mathsf{P}\left(X < 3 \right)$ en lugar de $\mathsf{P}\left(\{X < 3\} \right)$.

Se llama **función de distribución** de X a la función $F_X : \mathbb{R} \to \mathbb{R}$ definida por

$$F_X(x) = \mathsf{P}\left(X \leq x \right) = \mathsf{P}_X(\,(-\infty, x]\,)$$

Obsérvese que la probabilidad de que la variable X pertenezca al intervalo $(a, b]$ es:

$$\mathsf{P}_X\left((a, b] \right) = \mathsf{P}\left(X \in (a, b] \right) = F_X\left(b \right) - F_X\left(a \right) .$$

Teorema 4.1. *La función de distribución de probabilidad de una variable aleatoria real X satisface las siguientes propiedades:*

(d_1) F_X *es no decreciente.*

(d_2) F_X *es continua por la derecha.*

(d_3) $F_X(-\infty) = 0$*;* $F_X(\infty) = 1$.

Dem: Si $x < y, F_X = P(X \leq x) \leq P(X \leq x) + P(x < X \leq y) = P(X \leq y) = F_X(y)$ y eso prueba (d_1).

Para probar (d_2) aplicamos el Teorema 1.1 a la sucesión decreciente de conjuntos de Borel $((-\infty, x_n])$ donde $x_n \searrow x$. Resulta $\lim F_X(x_n) = \lim P_X((-\infty, x_n])) = P_X(\cap_{n=1}^{\infty}(-\infty, x_n]) = P_X((-\infty, x]) = F_X(x)$.

La demostración de (d_3) puede hacerse en forma similar y queda como ejercicio. $\square$

El teorema anterior tiene el siguiente recíproco:

Teorema 4.2. *Si* $F : \mathbb{R} \to \mathbb{R}$ *satisface* (d_1), (d_2) *y* (d_3)*, entonces existe una variable aleatoria cuya función de distribución es* F.

$\square$

De modo tal que *una función* F *es la función de distribución de alguna variable aleatoria si y solo si verifica las propiedades* (d_1)*,* (d_2) *y* (d_3).

Ejercicio 4 : Se considera la función de distribución F de la variable X, definida por: $F(x) = \begin{cases} 1 & 1 \leq x, \\ 0 & x < -1, \\ (x+2)/4 & -1 \leq x < 1. \end{cases}$

Calcular:

1. $P_X(\,(-1/2, 1/2]\,)$.
2. $P_X(\,\{0\}\,) = P(\,X = 0\,)$.
3. $P_X(\,\{1\}\,)$.
4. $P_X(\,(2, 3]\,)$.

4.1. Variables Aleatorias Discretas

Una variable aleatoria se dice **discreta** cuando su recorrido es un **conjunto numerable**.[6]

Si X es una variable aleatoria discreta con recorrido $R_X = \{x_i\}$ $i = 1, 2, \ldots$, las probabilidades $p_i = P(\,X = x_i\,)$ $(i = 1, 2, \ldots)$ bastan para determinar la función de distribución

$$F(x) = \sum_{x_i \leq x} p_i$$

y por lo tanto la distribución en virtud del Teorema 4.2. A las probabilidades p_i las podemos ver como la evaluación de una función $p_X : \mathbb{R} \to \mathbb{R}$ en el recorrido de X. A dicha función p_X que vale p_i en x_i y 0 en otros lugares se la conoce como **función de probabilidad** de la variable X.

Es fácil ver que *una función* $p : \mathbb{R} \to \mathbb{R}$ *es la función de probabilidad de una variable discreta si y solo sí:*

1. *es no negativa,*

[6] Se puede ser más fino y pedir que exista un conjunto numerable tal que la probabilidad P_X de dicho conjunto sea uno.

2. *es positiva en un conjunto numerable y*
3. *la suma de sus valores positivos vale 1.*

Ejercicio 5: A continuación se indican funciones de probabilidad p_X de variables discretas. Hallar en cada caso la función de distribución y dibujar su gráfico.

1. $p_X(0) = 1$.
2. $p_X(x) = x/15, \quad x = 1, 2, 3, 4, 5$.
3. $p_X(-1) = p_X(0) = p_X(1) = 1/3$.

Ejercicio 6: La variable aleatoria X puede tomar los valores $1, 2, \ldots, n$ con probabilidades $\mathsf{P}\left(X = r \right)$ proporcionales a r $(r = 1, 2, \ldots, n)$. Hallar F_X.

Observaciones: De las definiciones dadas en el Cuadro 1 se puede ver que:

- $\text{Ber}(p) = \text{Bin}(1, p)$.
- "$\text{Hiper}(m, k, n) \to \text{Bin}(n, p)$" cuando $k/(m + k) \to p$.
- "$\text{Bin}(n, p) \to \mathcal{P}(\lambda)$" cuando cuando $n \to +\infty$ y $np \to \lambda$.

Nombre	Notación	Recorrido	$p_X(x_i)$
Bernoulli	$\text{Ber}(p)$	$\{0,1\}$.	$p^i q^{1-i}$
Binomial	$\text{Bin}(n,p)$	$\{0,1,\ldots,n\}$.	$\binom{n}{i} p^i q^{n-i}$
Geométrica v1	$\text{Geo}(p)$	$\{0,1,2,\ldots\}$	qp^i
Geométrica v2	$\text{Geo}(p)$	$\{1,2,\ldots\}$	qp^{i-1}
Binomial Negativa	$\text{BN}(n,p)$	$\{0,1,2,\ldots\}$	$\binom{i+n-1}{i} p^n q^i$
Hipergeométrica	$\text{Hiper}(m,k,n)$	$\{0,1,2,\ldots,k\}$	$\frac{\binom{k}{i}\binom{m}{n-i}}{\binom{m+k}{n}}$
Poisson	$\mathcal{P}(\lambda)$	$\{0,1,2,\ldots\}$	$e^{-\lambda}\frac{\lambda^i}{i!}$

Cuadro 1: Tabla de distribuciones discretas básicas.En todos los casos $p \in [0,1]$, $q = 1 - p$.

4.2. Variables Aleatorias Continuas

Una variable aleatoria real cuya función distribución F_X se dice ella misma **absolutamente continua** (A.C.), cuando existe $f_X : \mathbb{R} \to \mathbb{R}^+$ tal que

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

Nótese que la función f_X puede ser mayor que 1.

Ejemplo 10 : Si $F_X(x) = x$ si $x \in [0,1]$ y 0 sino, entonces $f_X(x) = 1$ si $x \in [0,1]$ y 0 sino. Verificar que $F_X(x) = \int_{-\infty}^{x} f_X$.

La función f_X se llama **densidad** de la distribución.

Si X es una v.a. A.C. con distribución F y densidad f, entonces

$$P_X\left((a,b] \right) = P\left(X \le b \right) - P\left(X \le a \right) = F(b) - F(a) = \int_a^b f(x)\, dx.$$

Ejercicio 7 : Probar que, *si X tiene distribución A.C., entonces* $P\left(X = a \right) = 0$ *para todo a, y por tanto*
$P\left(a < X \le b \right) = P\left(a \le X < b \right) = P\left(a < X < b \right) = P\left(a \le X \le b \right)$.

La función F_X queda determinada por la función f_X (puesto que $F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$).

Obviamente *toda densidad de probabilidad f debe ser no negativa y satisfacer $\int_{-\infty}^{+\infty} f(x)\, dx = 1$*. Recíprocamente, *si satisface estas condiciones, es claro que $F(x) = \int_{-\infty}^{x} f(t)\, dt$ satisface*

(d_1), (d_2) *y* (d_3) *y por consiguiente* f *es una densidad de probabilidad.*

Ejercicio 8 : A continuación se indican funciones de densidad f_X de v.a. A.C. Hallar en cada caso la función de distribución y dibujar su gráfico.

1. $f_X(x) = \begin{cases} 0 & x \leq 1, \\ 1/x^2, & x > 1. \end{cases}.$

2. $f_X(x) = \begin{cases} 3(1-x^2), & 0 < x < 1; \\ 0, & x \leq 0, x \geq 1 \end{cases}.$

3. $f_X(x) = \begin{cases} 1/3, & 0 < x < 1, 2 < x < 4 \\ 0 & \text{sino.} \end{cases}$

Ejercicio 9: X tiene densidad constante en $(0, 1)$ y densidad cero fuera de ese intervalo. Hallar la distribución de la $n-$ésima cifra de la expresión decimal de X.

Ejercicio 10: X tiene distribución continua F_X y se define una nueva variable aleatoria $Y = g(X)$ por medio de la función derivable y estrictamente creciente g.

1. Calcular F_Y y f_Y.

2. Ídem en el caso en que g sea estrictamente decreciente.

3. Ídem en el caso $Y = aX + b$.

4. Ídem para $Y = X^2$ y para $Y = e^X$.

Nombre	Parámetros	Notación	Soporte	$f_X(x)$
Uniforme	$a < b$	$\mathcal{U}(a, b)$	$[a, b]$.	$1/(b-a)$
Exponencial	$\lambda > 0$	$\text{Exp}(\lambda)$	$\mathbb{R}^+$	$\lambda e^{-\lambda x}, \quad x \geq 0$
Normal	μ, σ^2	$N(\mu, \sigma^2)$	$\mathbb{R}$	$\dfrac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$

Cuadro 2: Tabla de distribuciones A.C. básicas.

Es fácil ver que si $X \sim \mathcal{U}(0, 1)$ entonces $(b-a)X+a \sim \mathcal{U}(0, 1)$.[7]

También es fácil calcular la función de distribución de una v.a. exponencial:

$$X \sim \text{Exp}(\lambda) \Rightarrow F_X(x) = 1 - e^{-\lambda x} \quad \forall x \geq 0.$$

Con un poco más de esfuerzo se puede ver que si $Z \sim \mathcal{N}(0, 1)$ entonces $Z/\sigma + \mu \sim \mathcal{N}(\mu, \sigma^2)$, pero su función de distribución no es expresable en términos sencillos. A la función de distribución F_Z de una $Z \sim \mathcal{N}(0, 1)$ se la denota $\Phi(x)$.

[7]Nótese como observamos antes que la función de densidad, al contrario de las funciones de probabilidad de una v.a. discreta, no tienen porque ser menores que 1. En el caso de la v.a. uniforma, si $b - a < 1$, entonces su densidad será mayor que 1 entre a y b.

5. Distribución Conjunta

En un experimento aleatorio se pueden hacer mediciones de diferentes parámetros, como ser la altura y el peso de una persona. Eso da lugar a más de una variable aleatoria. Dichas variables si bien son aleatorias pueden estar relacionadas, como por ejemplo el peso de una persona alta suele ser mayor al de una baja. Para estudiar esta correlación entre variables aleatorias, podemos considerar, en el caso de dos variables X y Y, el vector (X, Y) y la probabilidad de que dicho vector pertenezca a un subconjunto de $\mathbb{R}^2$. Para ello definiremos la **distribución conjunta** de X e Y como la función $F_{X,Y} : \mathbb{R}^2 \to [0, 1]$ dada por

$$F_{X,Y}(x, y) = P(X \le x, Y \le y) .$$

Proposición 5.1 (Propiedades de la función de distribución conjunta)**.** *Para cualesquiera dos variables aleatorias X y Y se cumple:*

1. $P((X, Y) \in (a, b] \times (c, d]) = F_{X,Y}(b, d) - F_{X,Y}(a, d) - F_{X,Y}(b, c) + F_{X,Y}(a, c)$.
2. $F_{X,Y}(x, y)$ *es monótona creciente en x y en y.*
3. $\lim_{y\to+\infty} F_{X,Y}(x, y) = F_X(x)$ *(**distribución marginal**).*
4. $\lim_{x\to+\infty} F_{X,Y}(x, y) = F_Y(y)$ *(distribución marginal).*
5. $\lim_{x,y\to+\infty} F_{X,Y}(x, y) = 1$.
6. $\lim_{x\to-\infty} F_{X,Y}(x, y) = \lim_{y\to-\infty} F_{X,Y}(x, y) = 0$.
7. *Continua por derecha en cada variable.*

Demostración. 1. llamemos I_x al intervalo $(-\infty, x]$ entonces
$\mathsf{P}(\,(X,Y) \in (a,b] \times (c,d]\,)$
$= \mathsf{P}\left((X,Y) \in I_b \times I_d \setminus \left(\overbrace{I_a \times I_d}^{A} \cup \overbrace{I_b \times I_c}^{B}\right)\right.$
$= \mathsf{P}(\,(X,Y) \in I_b \times I_d)\,) - \mathsf{P}(\,(X,Y) \in A \cup B\,)$,
pero
$\mathsf{P}(\,(X,Y) \in A \cup B\,) = \mathsf{P}(\,(X,Y) \in A\,) + \mathsf{P}(\,(X,Y) \in B\,) - \mathsf{P}(\,(X,Y) \in A \cap B\,)$ ✓

2. Si $x < x'$ entonces $X \leq x, Y \leq y \Rightarrow X \leq x', Y \leq y$ de donde el primer suceso está incluido en el segundo y de allí la desigualdad.

3. Por 2., $F_{X,Y}$ es creciente en y de donde
$\text{lím}_{y \to +\infty} F_{X,Y}(x,y) = \text{lím}_{n \to +\infty} F_{X,Y}(x,n)$
$= \text{lím}_{n \to +\infty} \mathsf{P}(\,(X \leq x, Y \leq n\,) = \mathsf{P}(\,X \leq x\,) = F_X(x)$.

4. Igual que 3.

5. Por 3., $\text{lím}_{x,y \to +\infty} F_{X,Y}(x,y) = \text{lím}_{x \to +\infty} F_X(x) = 1$.

6. Por 2.
$$\lim_{x \to -\infty} F_{X,Y}(x,y) \leq \lim_{x \to -\infty} \lim_{y \to +\infty} F_{(X,Y)}(x,y) = \lim_{x \to -\infty} F_X(x) = 0.$$

7.
$$F_{X,Y}(a^+, b) = \lim_{x \to a^+} \mathsf{P}(\,X \leq x, Y \leq b\,) = \lim_{x \to a^+} \frac{\mathsf{P}(\,X \leq x | Y \leq b\,)}{\mathsf{P}(\,Y \leq b\,)}$$
$$= \lim_{x \to a^+} \frac{F_{X|_{Y \leq b}}(x)}{F_Y(b)} = \frac{\mathsf{P}_{X|Y \leq b}(\,a\,)}{F_Y(b)} = \frac{\mathsf{P}(\,X \leq a | Y \leq b\,)}{\mathsf{P}(\,Y \leq b\,)} = F_{X,Y}(a,b).$$

Donde $X|Y \leq b$ es la variable aleatoria X restringida a $\Omega' = \Omega \cap \{\omega : Y(\omega) \leq b\}$. Las igualdades anteriores son válidas si $\mathsf{P}(\,Y \leq b\,) > 0$ en caso de que sea 0, la igualdad es trivial. $\square$

Proposición 5.2. *Una función* $F : \mathbb{R}^2 \to [0, 1]$ *es la función de distribución conjunta de dos variables aleatorias sii cumple las propiedades 1. para cualesquiera* a, b, c, d *tales que* $a < b, c < d$ *y también las propiedades 2., 5. y 6..*

5.1. Vectores aleatorios discretos

Se dice que el vector aleatorio (X, Y) es **discreto**,cuando su recorrido $R_{X,Y}$ es numerable.

Proposición 5.3. *Un vector aleatorio es discreto sii lo son sus componentes.* □

A la probabilidad $p_{X,Y}(x_i, y_i)$ de que $X = x_i$ y $Y = y_i$ se le llama **función de probabilidad conjunta** de las v.a. X y Y. Claramente tendremos que

$$P\{(X, Y) \in S\} = \sum_{(x_i, y_i) \in S} p_i.$$

En particular

$$F_{X,Y}(x, y) = \sum_{x_i \leq x, y_{\leq} y} p_{X,Y}(x_i, y_i).$$

5.2. Vectores aleatorios absolutamente continuos

Se dice que la pareja de variables aleatorias X, Y tiene **distribución conjunta absolutamente continua**, y escribiremos **A.C.**, cuando existe una función no negativa $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ tal que la probabilidad de que la pareja (X, Y) pertenezca a un conjunto S del plano se puede calcular en la forma.

$$\mathsf{P}\left((X, Y) \in S \right) = \iint_S f(x, y)\, dx\, dy.$$

La función $f_{X,Y}$ se llama **densidad de probabilidad conjunta de las variables**. [8] A partir de esta igualdad podemos calcular la función de distribución de un vector aleatorio como

$$F_{X,Y}(s, t) = \iint_{(-\infty,s]\times(-\infty,t]} f(x, y)\, dx\, dy = \int_{-\infty}^{s} dx \int_{-\infty}^{t} f(x, y)\, dx\, dy.$$

Viceversa a partir de la función de distribución conjunta $F_{X,Y}(s, t)$ podemos obtener la densidad derivando parcialmente:

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).$$

6. Independencia de Variables Aleatorias

Las v.a. $X_1, \ldots, X_n, \ldots$ se dicen **independientes**, cuando para cualesquiera intervalos, $B_1, B_2, \ldots$ las preimagenes $X_1^{-1}(B_1)$, $X_2^{-1}(B_2), \ldots$, son sucesos independientes.

Teorema 6.1. *Si la pareja de v. a.* X, Y *tienen distribución conjunta A. C. y cada una de las v.a. tiene distribución A.C. y, además, las variables* X, Y *son independientes, la densidad conjunta* $f_{X,Y}$ *es el producto de las densidades* f_X *y* f_Y *de cada una de las variables; más precisamente.*

$$f_{X,Y}(x, y) = f_X(x).f_Y(y).$$

[8] La elección de los conjuntos S el plano deberá limitarse a la de aquellos para los cuál es tenga sentido considerar la integral doble.

Dem: Por ser independientes las variables se cumple

$$\begin{aligned} P(a < X \le x, c < Y \le y) &= \int_a^x du \int_c^y f(u,v)\,dv \\ &= \int_a^x f_X(u)\,du \int_c^y f_Y(v)\,dv, \end{aligned}$$

y derivando sucesivamente respecto de x, y, obtenemos el resultado indicado arriba. □

Teorema 6.2. *Si la pareja de v.a.* X, Y *tiene distribución discreta, cada una de las variables tiene distribución discreta, además, los variables* X, Y *son independientes, entonces la función de probabilidad* $p_{X,Y}(x_i, y_j) = P(X = x_i, Y = y_j)$ *es producto de las funciones de probabilidades* $P(X = x_i)$ *y* $P(Y = y_j)$. □

Teorema 6.3. *Dos v.a. son independientes su función de distribución conjunta es el producto de las funciones de distribución. Más concretamente sii* $F_{X,Y}(x, y) \equiv F_X(x)F_Y(y)$. □

6.1. Vectores Aleatorios de mayor dimensión

Dadas n variables aleatorias $X_1, \ldots, X_n$ definimos $F_{X_1,\ldots,X_n} : \mathbb{R}^n \to [0,1]$ como

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 \le x_1, \ldots, X_n \le x_n).$$

Proposición 6.4. *Las v. a.* $X_1, \ldots, X_n$ *son independientes sii para cualesquiera* $x_1, \ldots, x_n$ *se cumple que*

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \times F_{X_2}(x_2) \times \ldots \times F_{X_n}(x_n).$$

□

7. Esperanza, Media o Valor Esperado

Cuando integramos una función consideramos el área bajo la curva. ¿Qué sucede si dilatamos las abscisas de modo que la longitud total de dicho eje sea 1 y algunso puntos puedan tener longitud no nula? La respuesta es una nueva forma de integrar llamada **integral de Riemann-Stieljes** y tiene estrecha relación con ciertos parámetros de las variables aleatorias. Dicha relación viene dada porque la dilatación quedará definida por la función de distribución de alguna variable aleatoria. Más concretamente, la integral de Riemann-Stieljes en el intervalo $[a, b]$ de la función g con respecto a la función de distribución F de la variable aleatoria X se define de la siguiente forma: a cada partición $a = x_0 \leq x_1 \leq \cdots \leq x_n = b$ del intervalo $(a, b]$ en cada uno de cuyos intervalos $dx_i = (x_{i-1}, x_i]$ se ha elegido un punto ξ_i, se asocia, la suma

$$\sum_{i=1}^{n} g(\xi_i)\ P\left(X \in\ dx_i \right) = \sum_{i=1}^{n} g(\xi_i) \left(F(x_i) - F(x_{i-1}) \right). \qquad (3)$$

Si a cualquier sucesión de particiones cuya **norma**[9] tiende a cero corresponde, cualquiera sea la elección de los ξ_i, una sucesión de sumas (3) con un mismo límite, se dice que existe la integral R.-S. de g con respecto a F en $[a, b]$ (o que g es R.-S. integrable con respecto a F en $[a, b]$), y que dicha integral es igual a ese límite. La integral se denota

$$\int_a^b g(x)\ dF(x).$$

[9]o sea el máximo de los $|\ dx_i|$

Lema 7.1. *Si* g *es continúa entonces siempre existe la integral de* g *respecto a* F. $\square$

Si existe el límite que se indica definimos:

$$\int_{-\infty}^{+\infty} g(x)\, dF(x) = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b g(x)\, dF(x).$$

Si $\int_{-\infty}^{+\infty} |x|\, dF_X(x) < \infty$, llamamos **valor esperado** o **media** de X a

$$\mathsf{E}\,(X) = \int_{-\infty}^{+\infty} x\, dF_X(x). \tag{4}$$

La convergencia absoluta de la integral que aparece en (4) (de la cual se puede deducir como es sabido la convergencia simple) se exige por razones técnicas que nos vamos a indicar en este curso.

Proposición 7.2. *Si existe* $\int_{-\infty}^{+\infty} |g(x)|\, dF_X(x)$*,* g *es continua entonces*

$$\int_{-\infty}^{+\infty} g(x)\, dF_X(x) = \begin{cases} \displaystyle\sum_{x_i \in R_X} g(x_i) p_X(x_i) & \textit{si } X \textit{ es dis. con f. de p. } p_X \\ \\ \displaystyle\int_{-\infty}^{+\infty} g(x) f_X(x)\, dx & \textit{si } X \textit{ es A.C. con densidad } f_X \end{cases}$$

Corolario 7.3. *Si existe* $\int_{-\infty}^{+\infty} |x|\, dF_X(x)$*, entonces*

$$\mathsf{E}\,(X) = \begin{cases} \displaystyle\sum_{x_i \in R_X} x_i p_X(x_i) & \textit{si } X \textit{ dis. con f. de p. } p_X, \\ \\ \displaystyle\int_{-\infty}^{+\infty} x f_X(x)\, dx & \textit{si } X \textit{ es A.C. con densidad } f_X. \end{cases}$$

Puede probarse que el valor esperado, tiene las propiedades que se resumen en el siguiente enunciado;

Proposición 7.4. *Si* X *y* Y *son v.a. y* $\lambda, \mu, C \in \mathbb{R}$*, entonces*

(i) *Si* $X \geq 0$*, entonces* $\mathsf{E}\left(X\right) \geq 0$ *(monotonía).*

(ii) $\mathsf{E}\left(\lambda X + \mu Y\right) = \lambda \mathsf{E}\left(X\right) + \mu \mathsf{E}\left(Y\right)$ *(linealidad),*

(iii) *Si* X *es constante igual a* C*, entonces* $\mathsf{E}\left(X\right) = C$.

Corolario 7.5. *Si* $X_1, \ldots, X_n$ *son variables aleatorias con la misma media* μ*, entonces su promedio* $\overline{X}_n = (X_1 + \cdots + X_n)/n$ *tiene media* μ. □

Ejemplo 11: Si $X \sim \mathrm{Bin}(n, p)$ el valor esperado de X es, por la Proposición 7.2 igual a

$$\mathsf{E}\left(X\right) = \sum_{n=1}^{\infty} i \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^{n} \frac{n!}{(i-1)!(n-i)!} p^i (1-p)^{n-i}$$

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

$$= np.$$

Nótese que la última sumatoria es la suma de las probabilidades puntuales de una $\mathrm{Bin}(n-1, p)$.

Ejemplo 12:

Distribución	Esperanza
$\mathrm{Ber}(p)$	p
$\mathrm{Bin}(n, p)$	np
$\mathrm{Geo}(p)$ v1	q/p
$\mathrm{Geo}(p)$ v2	$1/p$
$\mathrm{BN}(n, p)$	nq/p
$\mathrm{Hiper}(m, k, n)$	$n\dfrac{k}{m+k}$
$\mathcal{P}(\lambda)$	λ
$\mathcal{U}(a, b)$	$(a+b)/2$
$\mathrm{Exp}(\lambda)$	$1/\lambda$
$N(\mu, \sigma^2)$	μ

Cuadro 3: Medias de distribuciones básicas, donde $q = 1 - p$.

Ejemplo 13: Si X tiene distribución A.C. con densidad cero fuera del intervalo $(0, 1)$ y $f(x) = 2x$ en dicho intervalo, entonces:

$$\mathsf{E}\,(X) = \int_{-\infty}^{+\infty} xf(X)\;dx = \int_0^1 2x^2\;dx = 2/3.$$

El significado del valor esperado de una variable aleatoria puede aclararse considerando la siguiente *analogía mecánica*. Si se considera en la recta una distribución de masa continua con densidad ρ, la abscisa del baricentro vale.

$$\frac{\int x\rho(x)\;\;dx}{\int \rho(x)\;\;dx}.$$

Si a una variable aleatoria continua con densidad f asociamos entonces una distribución de masa con densidad $\rho = f$, el valor esperado de X coincide con la abscisa del baricentro puesto que en este caso $\int \rho(x)\;dx = 1$.

Análogamente, el baricentro de una distribución de masas m_i concentradas en los puntos x_i está dado por $\sum m_i x_i / \sum m_i$ y si a una variable aleatoria discreta con probabilidades p_i concentradas en los puntos x_i, asociamos la distribución anterior para lo cual tomamos $m_i = p_i$, el baricentro coincide con el valor esperado.

Así como en mecánica el estudio de los momentos de una distribución de masas, en particular los momentos de segundo orden o de inercia, arroja resultados interesantes, lo mismo ocurre en probabilidad, por lo cual consideraremos las definiciones de la próxima sección.

8. Variancia o Varianza

Cuando $\int_{-\infty}^{+\infty} |x|^n \, dF_X(x) < \infty$, diremos que $\int_{-\infty}^{+\infty} x^n \, dF_X(x)$ es el **momento de orden** n **de** X o de la distribución de X. Cuando $\int_{-\infty}^{+\infty} |x - \mathsf{E}(X)|^n \, dF_X(x) < \infty$ diremos que $\int_{-\infty}^{+\infty} (x - \mathsf{E}(X))^n \, dF_X(x)$ es el **momento central de orden** n **de** X**.**

Al momento central de orden dos se llama **variancia**, y se lo denota $\mathsf{Var}(X)$, es decir

$$\mathsf{Var}(X) = \mathsf{E}\left((X - \mathsf{E}(X))^2\right).$$

Usando propiedades de la esperanza podemos demostrar las siguientes propiedades de la varianza

Proposición 8.1.

1. $\mathsf{Var}(X) = \mathsf{E}\left(X^2\right) - \left(\mathsf{E}(X)\right)^2$.
2. $\mathsf{Var}(aX + b) = a^2 \mathsf{Var}(X)$.
3. $\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y) + 2\mathsf{E}\big((X - \mathsf{E}X)(Y - \mathsf{E}Y)\big) = \mathsf{Var}(X) + \mathsf{Var}(Y) + 2\big(\mathsf{E}(XY) - \mathsf{E}X\mathsf{E}Y\big)$
4. *Si* $\mu = \mathsf{E}X$, $\sigma^2 = \mathsf{Var}X$ *y* $Y = (X - \mu)/\sigma$ *entonces* $\mathsf{E}Y = 0$ *y* $\mathsf{Var}\,Y = 1$.

□

El siguiente teorema permite calcular los momentos de orden mayor a 1, y por la propiedad 1. de la proposición anterior, la varianza de las distribuciones básicas.

Teorema 8.2. *Si* g *es una función continua y* $\int_{-\infty}^{+\infty} |g(x)| \, dF_X(x) < \infty$*, entonces*

$$\mathsf{E}\left(g(X)\right) = \int_{-\infty}^{+\infty} g(x) \, dF_X(x).$$

□

Proposición 8.3. *Si dos variables* X, Y *tienen momentos absolutos de segundo orden* $\mathsf{E}(|X|^2)$, $\mathsf{E}(|Y|^2)$ *y momentos mixtos* $\mathsf{E}(|XY|)$*, entonces valen las siguientes desigualdades*

$$(\mathsf{E}X)^2 \leq (\mathsf{E}|X|)^2 \leq \mathsf{E}\left(|X|^2\right),$$

$$(\mathsf{E}(XY))^2 \leq \mathsf{E}\left(X^2\right)\mathsf{E}\left(X^2\right) \qquad \textit{(Cauchy-Schwarz).}$$

□

Distribución	Esperanza	Distribución	Esperanza
Ber(p)	pq	$\mathcal{U}(a,b)$	$(b-a)^2/12$
Bin(n,p)	npq	Exp(λ)	$1/\lambda^2$
Geo(p) v1 y v2	q/p^2	N(μ,σ^2)	σ^2
BN(n,p)	nq/p		
Hiper(m,k,n)	$n\frac{k}{N}\frac{m}{N}\frac{N-n}{N-1}$		
$\mathcal{P}(\lambda)$	λ		

Cuadro 4: Tabla de varianzas de distribuciones básicas, donde $q = 1 - p$ y $N = m + k$.

Ejercicio 11: Sea $X \sim \mathcal{U}(0,1)$. Hallar las distribuciones X^2, e^X y calcular en cada caso el valor esperado y la varianza.

9. Momentos de un vector aleatorio

Podemos calcular la media y la variancia de un vector usando el siguiente resultado homólogo al correspondiente para variables aleatorias:

Teorema 9.1 (Regla para el cálculo de $\mathsf{E}\,(\,g(X,Y)\,)$)**.** *Si* g *es una función continua de dos variables y las siguientes series o integrales, según corresponda, sean absolutamente convergentes, entonces si* (X,Y) *es*

- *discreta con recorrido* $\{(x_i, y_i)\}$ *y función de probabilidad* p_i*, la esperanza de* $g(X,Y)$ *es*

$$\mathsf{E}\,(g(X,Y)) = \sum_i g(x_i, y_i)\ p_i.$$

- *A.C. con densidad* $f_{X,Y}$ *, la esperanza de* $g(X,Y)$ *es*

$$\mathsf{E}\,(g(X,Y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)\ f_{X,Y}(x,y)\ \ dx\,dy.$$

□

9.1. Correlación entre variables aleatorias

Se llama **covariancia** de las variables X, Y a $\mathsf{Cov}(X,Y) = \mathsf{E}\,((X-\mathsf{E}\,(X))(Y-\mathsf{E}\,(Y)))$, y **coeficiente de correlación** de las mismas variables a

$$\rho(X,Y) = \frac{\mathsf{Cov}(X,Y)}{\sqrt{\mathsf{Var}(X)\mathsf{Var}(Y)}}.$$

El coeficiente de correlación pertenece al intervalo $[-1, 1]$. Si el coeficiente es -1 o 1 las variables están linealmente correlacionadas en el sentido que $Y = aX + b$ con a negativo o positivo respectivamente. Si el coeficiente es nulo se dice que las variables se dicen **no correlacionadas**.

Si X, Y son independientes, entonces su covarianza es nula. El recíproco no es cierto, basta con considerar variables cuya esperanza sea nula así como la de su producto, algo sencillo de encontrar, pero que estén correlacionadas. Por ejemplo $X \sim N(0, \sigma^2)$ y $Y = |X|$.

Proposición 9.2. *Si las variables* $X_1, \dots, X_n$ *están dos a dos no correlacionadas, entonces*

$$\mathsf{Var}\,(X_1 + \cdots + X_n) = \mathsf{Var}\,(X_1) + \cdots + \mathsf{Var}\,(X_n)\,.$$

□

Ejercicio 12: Probar que $\mathsf{E}\,((X - t)^2)$ es mínimo para $t = \mathsf{E}\,(X)$. ¿Para que valor de t es mínimo $\mathsf{E}\,(|X - t|)$?

Ejercicio 13: Sea $X \sim N(0, 1)$. Demostrar que los momentos de orden impar son cero y los de orden par están dados por $\mathsf{E}\left(X^{2k}\right) = (2k)!/(2^k k!)$.

Teorema 9.3. *Si* X, Y *son v.a. independientes y existen los valores esperados* $\mathsf{E}\,X\,\mathsf{E}\,Y$, *entonces existe el valor esperado del* *producto* XY*, y vale*

$$E(XY) = E\,X\,E\,Y.$$

$\square$

Demostración. Lo demostraremos en el caso en que las son v.a. son discretas:

$$E(XY) = \sum_{i,j} x_i y_j P(X = x_i, Y = y_j) = \sum_i \sum_j x_i y_j P_X(x_i) P_Y(y_j)$$

$$= \sum_i x_i P_X(x_i) \sum_j y_j P_Y(y_j) = \sum_i x_i P_X(x_i) E\,Y = E\,Y E\,X.$$

$\square$

Ejercicio 14: Demostrar el Teorema 6.3 para v.a. A.C.

Ejercicio 15: Probar que si las v.a. $X_1, \ldots, X_n$ son independientes, entonces están dos a dos no correlacionadas. Por tanto se cumple

$$\mathrm{Var}\left(\sum_i X_i\right) = \sum_i \mathrm{Var}(X_i).$$

Ejercicio 16: Se repiten n veces independientemente un experimento que resulta de en A con probabilidad p. Mostrar que el número total de veces que ocurre A tiene distribución binomial con parámetros n, p.

Recalcular la media y la variancia de la distribución binomial aprovechando esta circunstancia.

Ejercicio 17: En un experimento aleatorio, la probabilidad de que el resultado sea un suceso a A es p. Se repite el experimento independientemente hasta que ocurre por primera vez A. Sea n el número de veces que es necesario repetir el experimento; calcular $\mathsf{E}\,(\,n\,)$ y $\mathsf{Var}(\,n\,)$.

Ejercicio 18: X e Y son independientes con distribución uniforme en $[0,1]$. Hallar la distribución de $Z = \text{máx}\{x,y\}$, $W = \text{mín}\{x,y\}$ y calcular $\mathsf{E}\,(\,Z\,)$ y $\mathsf{E}\,(\,W\,)$.

10. Desigualdad de Chebyshev

Teorema 10.1 (Desigualdad de Chebyshev)**.** *Si X es una variable aleatoria, $g : \mathbb{R} \to \mathbb{R}$ una función creciente, no negativa, y $g(a) > 0$ entonces*

$$\mathsf{P}\,(\,X \geq a\,) \;\leq\; \frac{\mathsf{E}\left(\,g(X)\,\right)}{g(a)}.$$

Dem: $\mathsf{E}\,(g(X)) = \int_{-\infty}^{+\infty} g(x)\; dF(x) \geq \int_{a}^{+\infty} g(x)\; dF(x)$
$\geq g(a) \int_{a}^{+\infty}\; dF(x) = g(a)\,\mathsf{P}\,(\,X \geq a\,)\,.$ □

Ejercicio 10.2. *Justifique las (des)igualdades en la demostración anterior.*

Corolario 10.3. *Si Y es una v.a. con variancia entonces*

$$\mathsf{P}\,(\,|Y - \mathsf{E}\,(Y)| \geq \epsilon\,) \;\leq\; \frac{\mathsf{Var}(\,Y\,)}{\epsilon^2}.$$

Dem: Basta elegir en el Teorema 10.1

$$g(x) = \begin{cases} 0 & x < 0 \\ x^2 & x \geq 0. \end{cases}$$

y $X = |Y - E(Y)|$. $\square$

11. Leyes de los grandes números

Se dice que la sucesión de variables aleatorias (X_n) **converge en probabilidad a** L, o que

$$\operatorname*{plím}_{n\to\infty} X_n = L$$

cuando para cada $\varepsilon > 0$

$$P(|X_n - L| \geq \varepsilon) = P(\{\omega : |X_n(\omega) - L| \geq \varepsilon\}) \underset{n\to+\infty}{\longrightarrow} 0.$$

También se denota $X_n \xrightarrow{P} L$.

Se dice que la sucesión de variables aleatorias (X_n) **converge casi seguramente a** L, o que

$$X_n \xrightarrow{c.s.} L$$

sii

$$P(X_n \to L) = P(\{\omega : X_n(\omega) \to L\}) = 1.$$

Se puede demostrar que *si* X_n *converge en probabilidad entonces converge casi seguramente.*

Se dice que el conjunto de variables aleatorias $X_1, \ldots, X_n$ es una **muestra aleatoria simple** o **M.A.S.** de una distribución de probabilidad, cuando todas las X_i tienen esa distribución de probabilidad y son independientes, por ejemplo si son el resultado de repeticiones independientes de un mismo experimento.

Teorema 11.1 (Ley débil de los grandes números)**.** *Para cada* n*, sean* $X_1, \dots, X_n$ *una M.A.S. de una distribución con momento de segundo orden finito. Entonces, si* μ *es el valor esperado común de todas las variables y* $\overline{X}_n = (X_1 + \cdots + X_n)/n$

$$\operatorname*{plím}_{n \to \infty} \overline{X}_n = \mu.$$

Dem: Sea σ^2 el valor común de las variancias. Entonces $\mathsf{E}\left(\overline{X}_n\right) = \mu$. $\mathsf{Var}(\overline{X}_n) = \sigma^2/n$ y $\mathsf{P}\left(|\overline{X}_n - \mu| \geq \varepsilon\right) \leq \mathsf{Var}(\overline{X}_n)/\varepsilon^2 = \sigma^2/(n\varepsilon^2) \to 0$ por el corolario del Teorema 7.1. □

Ejercicio 19 : (Teorema de Bernoulli) Probar que si las variables $X_1, \dots, X_n, \dots$ tienen distribución binomial con parámetros n, p, entonces

$$\operatorname*{plím}_{n \to \infty} \overline{X}_n = p.$$

Teorema 11.2 (Ley fuerte de los grandes números)**.** *Para cada* n*, sean* $X_1, \dots, X_n$ *variables aleatorias independientes todas con la misma distribución, con momento de cuarto orden finito. Entonces, si* μ *es el valor esperado común de todas las variables, entonces*

$$\overline{X}_n \xrightarrow{\text{c.s.}} \mu.$$

□

12. Teorema del Límite Central

Se llama **función de distribución de una muestra** $X_1, \ldots, X_n$, a la función F_n dada por

$$F_n(x) = \frac{1}{n}[\text{número de variables en } X_1, \ldots, X_n \text{ que son} \leq x].$$

(Si χ es la función característica de $(-\infty, x]$, entonces $F_n(x) = (1/n)\sum_{i=1}^{n}\chi(X_i \leq x)$). Siendo $\chi(X_i \leq x)$ es la función indicatriz de que X_i sea menor o igual a x o sea

$$\chi(X_i \leq x) = \begin{cases} 1 & X_i \leq x, \\ 0 & X_i > 0. \end{cases}$$

Ejercicio 20: Para cada n, $X_1, \ldots, X_n$ es una muestra aleatoria simple de la distribución de probabilidades con función de distribución F. Sea F_n la función de distribución de $X_1, \ldots, X_n$.
a) Hallar la distribución de $F_n(x)$.
b) Probar que para cada x, plím $F_n(x) = F(x)$.

Ejercicio 21: Usar la desigualdad de Chebyshev para determinar cuantas veces debe tirarse una moneda ($P(\text{ CARA }) = P(\text{ CRUZ }) = 1/2$) de modo que la probabilidad de que la proporción de caras en el conjunto de resultados esté entre 0.4 y 0.6, sea por lo menos de 0.9.

Ejercicio 22:

Probar que *el coeficiente de correlación de dos variables aleatorias* X, Y *con medias* $\mathsf{E}(X) = \mathsf{E}(Y) = 0$*, cumple* $\rho(X, Y) = 1$ *sii las variables son linealmente independientes, con probabilidad 1* (Se sugiere probar que hay una combinación lineal $\lambda X + \mu Y$ tal que $\mathsf{P}(|\lambda X + \mu Y| \geq \varepsilon) = 0$ para todo $\varepsilon > 0$).

Agregaremos finalmente el enunciado de otro teorema que no demostraremos. Este teorema justifica en parte la gran importancia que tiene en estadística la llamada "distribución normal", que estudiaremos en Estadística.

Teorema 12.1 (Teorema del Límite Central[10])**.** *Sean,* $X_1, \ldots, X_n$ *para cada* n*, variables independientes todas con la misma distribución, con momento de orden* $(2+a)$ *finito para algún* $a > 0$*. Sea* $\mu = \mathsf{E}(X_i)$*,* $\sigma^2 = \mathsf{Var}(X_i)$ $(i = 1, 2, \ldots, n)$ *y* $\Phi(x)$ *es la F.D. de una v.a.* $N(0, 1)$ *o sea*

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

Entonces para cada x *se cumple*

$$\lim_{n \to +\infty} \mathsf{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq x \right) = \Phi(x)$$

$\square$

Una pregunta natural es cuál será la velocidad de la convergencia del límite anterior o una cota de dicha velocidad. El siguiente resultado en ese sentido fue demostrado por Andrew C. Berry y Carl-Gustav Esseen, que con mejoras dice así:

---

[10] Conocido como "Teorema central del límite" en castellano

Teorema 12.2. *En las condiciones del teorema del límite central, si* $\rho = \mathsf{E}\,|X_1 - \mu|^3$ *y*

$$\left| \mathsf{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq x \right) - \Phi(x) \right| \leq \frac{0{,}4748\rho}{\sigma^3\sqrt{n}}.$$

□

Se sabe, gracias a Essen, que la constante $0{,}4748$ no se puede mejorar más que hasta $0{,}4097$.

Ejercicio 23 : Resolver el Ejercicio 21 usando el teorema anterior.

Índice alfabético