FACULTAD DE MATEMÁTICAS
PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE

Pontificada Universidad Católica

EYP1027

Modelos Probabilísticos

Autor:
Sebastián Lepe V.

2 de julio de 2025

Índice

1. Teoría de Conjuntos

1.1. Conjuntos, Contables y No Contables

No entraremos tan profundamente en la teoría de conjuntos. Pero trabajaremos con conceptos que requieren una definición, formas y unos restuldados importante para la teoría de probabilidad.

Un conjunto es un objeto matemático que contiene elementos cualesquiera. Puede ser de animales, números, autos, etc. Por lo general usamos las letras mayúsculas para denotar conjuntos, por ejemplo,

$$A = \{\text{lunes, martes, miercoles,...}\}$$

Definición 1.1. (Contable) *Un conjunto A se dice contable o discreto si es finito o existe un mapeo biyectivo con los naturales. En caso contrario diremos que A es no contable.*

Ejemplo 1.1. Los conjuntos $A_1 = \{0,1\}, A_2 = \{1,2,3,4,5,6\}, A_3 = \{1,3,5,\dots\}$. Son contables. Los dos primeros porque son finitos y el último podemos hacer el siguiente mapeo

$$\begin{aligned} i : A_3 &\to \mathbb{N} \\ x \mapsto i(x) &:= \frac{x+1}{2} \end{aligned}$$

Podemos ver que i es una biyección.

Ejemplo 1.2. Los conjuntos $A_1 = (0,1), A_2 = [0,\infty), A_3 = \{(x,y) : x^2 + y^2 \leq 1\}$ no son contables. Verificarlo resulta más complicado y no lo haremos, ya que no es nuestro enfoque.

1.2. Definiciones y Propiedades

Definción 1.2. *Sean A, B conjuntos. Decimos que A es subconjunto de B o que A está contenido en B si para todo $x \in A$, se tiene que $x \in B$. Si A es subconjunto de B, entonces lo denotamos por $A \subseteq B$.*

Nota 1.1. Puede pasar que $A \subseteq B$, pero $A \neq B$, en ese caso decimos que $A \subset B$ y que A es un subconjunto propio de B.

Nota 1.2. Algo importante a recordad, es que $A = B$ si y sólo si $A \subseteq B$ y $B \subseteq A$. También que para todo conjunto A, se tiene que $\emptyset \subseteq A$ y $A \subseteq A$. ($\emptyset$ es el conjunto vacío)

Definción 1.3. *Sean A, B conjuntos. Se definen la unión por:*

$$A \cup B := \{x : x \in A \ \ o \ \ x \in B\}$$

Se define la intersección por:

$$A \cap B := \{x : x \in A \ \ y \ \ x \in B\}$$

Se define el complemento de A por:

$$A^C := \{x : x \notin A\}$$

Se define la diferencia de A sobre B por:

$$A - B = A \setminus B := \{x : x \in A, x \notin B\}$$

(cuando ponemos una coma, nos referimos al conector "y").

Observación 1.1. Se puede ver que

$$A \setminus B = A \cap B^C$$

Definición 1.4. *Sean A, B conjuntos. Decimos que son disjuntos entre si, si:*

$$A \cap B = \emptyset$$

Sea $\{a_i\}_{i\in\mathbb{N}}$ una secuencia, sucesión, colección de conjuntos. Se dice que es mutuamente disjuntas (dos a dos) si

$$A_i \cap A_j = \emptyset, \text{ para todo } i \neq j$$

El siguiente teorema solo será para recordad las propiedades, no se demostrará.

Teorema 1.1. *Sean A, B, C conjuntos. Entonces se cumple las siguientes propiedades:*

1. ***Conmutatividad.***

$$A \cup B = B \cup A, \ \ A \cap B = B \cap A$$

2. ***Asociatividad.***

$$A \cup (B \cup C) = (A \cup B) \cup C, \ \ A \cap (B \cap C) = (A \cap B) \cap C$$

3. ***Leyes distributivas.***

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \ \ A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

4. ***Leyes de Morga.***

$$(A \cup B)^C = A^C \cap B^C, \ \ (A \cap B)^C = A^C \cup B^C$$

Definición 1.5. *Sea $\{A_i\}_{i\in\mathbb{N}}$ una colección de conjuntos definidos en Ω, entonces se definen*

$$\bigcup_{i=1}^{\infty} A_i = \{x : x \in A_i \text{ para algún } i\}$$
$$\bigcap_{i=1}^{\infty} A_i = \{x : x \in A_i \text{ para todo } i\}$$

Las secuencias infinitas contables, nos permiten construir conjuntos más elaborados. Por ejemplo, si $\Omega = (0,1]$ y sea $A_i := \left[\frac{1}{i}, 1\right]$ para todo $i \in \mathbb{N}$. Entonces

$$\bigcup_{i=1}^{\infty} A_i = \{1\} \cup \left[\frac{1}{2}, 1\right] \cup \left[\frac{1}{3}, 1\right] \cup \cdots = (0,1]$$

$$\bigcap_{i=1}^{\infty} A_i = \{1\} \cap \left[\frac{1}{2}, 1\right] \cap \left[\frac{1}{3}, 1\right] \cap \cdots = \{1\}$$

Nota 1.3. Se puede definir colecciones sobre índices no contables, e intersectarlos o unirlo. Por ejemplo si Γ es un conjunto no contable, entonces

$$\bigcup_{a \in \Gamma} A_a$$

es la unión de todos los elementos de $a \in \Gamma$. Otro ejemplo, si $\Gamma := \{x \in \mathbb{R} : x > 0\}$ y $A_a := (0, a]$, entonces

$$\bigcup_{a \in \Gamma} A_a = (0, \infty)$$

Observación 1.2. Las unionces e intersecciones sobre una colección infinita contable o no, no necesariamente genera conjuntos infinitos, vimos que pueden generar conjuntos singulares, entre otros.

Definición 1.6. *Sea $\{A_i\}_{i\in\mathbb{N}}$ una secuencia de subconjuntos de Ω. Decimos que está colección es una partición de Ω si*

1. $\bigcup_{i\geq 1} A_i = \Omega$.
2. $A_i \cap A_j = \emptyset$ *para todo* $i \neq j$.

Nota 1.4. La partición no necesariamente es sobre un índice infinito. Perfectamente una cantidad finita de subconjuntos de Ω particionar a Ω.

Ejemplo 1.3. Sea

$$A_i := [i, i+1)$$

para todo $i = 0, 1, \ldots$. Podemos probar que la colección $\{A_i\}$ es una partición de $[0, \infty)$ ya que $A_i \cap A_j = \emptyset$ para todo $i \neq j$ y

$$\bigcup_{i=0}^{\infty} A_i = [0, \infty)$$

Definición 1.7. *Sea $\{A_n\}_{n\in\mathbb{N}}$ una secuencia. Decimos que es monótona si:*

1. $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$, *es decir* $\{A_n\}$ *es creciente. En este caso denotamos por* $A_n \uparrow$

2. $A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$, *es decir* $\{A_n\}$ *es decreciente. En este caso denotamos por* $A_n \downarrow$

Lema 1.1. *Sea* $\{A_i\}_{i=1}^{\infty}$ *una colección de conjuntos. Sea la colección* $\{B_i\}_{i=1}^{\infty}$ *definida por*

$$\begin{aligned} B_1 &:= A_1 \\ B_2 &:= A_1^C A_2 \\ &\vdots \\ B_k &:= A_1^C A_2^C \dots A_{k-1}^C A_k \end{aligned}$$

Entonces

1. *La colección* $\{B_i\}_{i=1}^{\infty}$ *es disjunta entre todos sus elementos, es decir,* $B_i \cap B_j = \emptyset$ *para todo* $i \neq j$.
2. $$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$$

Dem.

1. Sean i, j naturales fijo, supongamos sin pérdida de generalidad que $i < j$, entonces
$$B_i \cap B_j = A_1^C A_2^C \dots A_i A_1^C A_2^C \dots A_j$$
por la conmutatividad podemos expresarlo de la siguiente forma
$$B_i \cap B_j = A A_i A_i^C = A\emptyset = \emptyset$$
donde A es el conjunto $A_1^C A_1^C A_2^C A_2^C \dots$.
2. Probaremos por doble inclusión. Sea $x \in \bigcup B_i$, entonces existe un $l \in \mathbb{N}$ talque $x \in B_l$, luego por definición de B_l se tiene que $x \in A_1^C, x \in A_2^C, \dots, x \in A_l$, de está última pertenencia deducimos que $x \in \bigcup A_i$, por lo que
$$\bigcup_{i=1}^{\infty} B_i \subseteq \bigcup_{i=1}^{\infty} A_i$$
Sea $x \in \bigcup A_i$, y sea el conjunto
$$I := \{i \in \mathbb{N} : x \in A_i\} \subseteq \mathbb{N}$$
El conjunto I no es vacío por definición de unión y por lo tanto, existe un valor mínimo, sea $n := \min\{i \in \mathbb{N} : x \in A_i\}$. Ahora con respecto a la colección B_i, podemos ver que $x \notin B_j$ para todo $j > n$. Notemos también que $x \notin A_1, \dots, A_{n-1}$ en virtud de que n es

el mínimo de I. Luego x no puede estar en $B_1, \ldots, B_{n-1}$, dejando como último candidato B_n, y en efecto lo está. Dado que $x \notin A_1, A_2, \ldots, A_{n-1}$ podemos entonces decir que

$$x \in \bigcap_{i=1}^{n-1} A_i^C =: A$$

si al conjunto A lo intersectamos con A_n, obtenenemos el conjunto B_n y si $x \in A_n$, entonces $x \in A \cap A_n = B_n$. Probando así que

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$$

Probando el lema. ■

Definición 1.7. *Sea* $\{A_i\}_{i=1}^{\infty}$ *una colección infinita de conjuntos. Se define el límite superior de la colección por:*

$$\limsup_{n\to\infty} A_n := \bigcap_{k=1}^{\infty} \bigcup_{i=k}^{\infty} A_i$$

De forma análoga se define el límite inferior de la colección por:

$$\liminf_{n\to\infty} A_n = \bigcup_{k=1}^{\infty} \bigcap_{i=k}^{\infty} A_i$$

Proposición 1.2. *Sea* $\{A_i\}_{i=1}^{\infty}$ *una colección de conjuntos infinitos. Entonces*

$$\liminf A_n \subseteq \limsup A_n$$

Dem. Por convenio definiremos

$$C_n := \bigcap_{i=n}^{\infty} A_i$$

de forma que $\liminf A_n = \bigcup C_k$. Sea $x \in \liminf A_n$, entonces existe un valor mínimo $l \in \mathbb{N}$ (Idea de la demostración anterior) talque se tiene que $x \in C_l$, de aquí se deduce que $x \in A_j$ para todo $j \geq l$, lo importante es que para todo $k \in \mathbb{N}$ siempre ocurrirá que $x \in \bigcup_{i=k}^{\infty} A_i$ ya que i en algún momento tomará el valor de l o mayor a l. Luego al aplicar la intersección de 1 al infinito a estas uniones, veremos que x pertenece a ese conjunto. Probando así que

$$\liminf A_n \subseteq \limsup A_n$$

como queriamos. ■

Definición 1.7. *Sea* $\{A_n\}_{n\geq 1}$ *una sucesión de conjuntos. Se dice que es convergente si*

$$\limsup A_n \subseteq \liminf A_n$$

Observación 1.3. Notemos que si una colección de sucesos es convergente, entonces el límite superior es el mismo que el límite inferior.

Lema 1.2. *Toda secuencia monótona, es convergente.*

Dem. Sea $\{A_n\}$ una sucesión de eventos. Sin pérdida de generalidad, supongamos que la secuencia es creciente. Debemos probar que el límite superior de A_n está contenido en el límite inferior de A_n. En virtud de que A_n es creciente, podemos afirmar que para todo $k \in \mathbb{N}$ se tiene que

$$\bigcap_{i=k}^{\infty} A_i = A_k$$

Por lo que el límite inferior toma como valor

$$\liminf A_n = \bigcup_{k=1}^{\infty} A_k$$

Sea $x \in \limsup A_n$, entonces x está contenido para todo $k \in \mathbb{N}$ en el conjunto

$$\bigcup_{i=k}^{\infty} A_i$$

Luego para algún $l \in \mathbb{N}$, se tiene que $x \in A_i$, y esto implica que

$$x \in \bigcup_{k=1}^{\infty} A_k = \liminf A_n$$

De forma que la secuencia converge. ■

Nota 1.2. Si una secuencia converge, en vez de denotar por límite superior e inferior, denotaremos por

$$\lim_{n\to\infty} A_n$$

Teorema 1.1. *Sea* $\{A_n\}_{n\geq 1}$ *una secuencia infinita. Si*

1. *Es creciente, entonces*

$$\lim_{n\to\infty} A_n = \bigcup_{i=1}^{\infty} A_i$$

2. *Es decreciente, entonces*

$$\lim_{n\to\infty} A_n = \bigcap_{i=1}^{\infty} A_i$$

Dem. La demostración consiste en expandir la idea de la demostración del lema 1.2, hemos probado que si la secuencia es creciente, entonces

$$\lim A_n = \bigcup_{k=1}^{\infty} A_k$$

Si la secuencia es decreciente, entonces debemos notar que para todo $k \in \mathbb{N}$ se tiene que

$$\bigcup_{i=k}^{\infty} A_i = A_k$$

Luego

$$\bigcap_{k=1}^{\infty} A_k = \limsup A_n = \lim A_n$$

Probando el teorema. ■

1.3. Conjuntos σ-álgebra

Estudiaremos un tipo especial de conjuntos que son muy importante para la teoría de probabilidad ya que sus propiedades se relacionan bien con las probabilidades.

Definición 1.8. *Sea $\mathcal{A}$ una colección de subconjuntos de Ω. Diremos que $\mathcal{A}$ es un σ-álgebra sobre Ω si satisface las siguientes condiciones:*

1. $\Omega \in \mathcal{A}$
2. *Si $A \in \mathcal{A}$, entonces $A^C \in \mathcal{A}$*
3. *Si $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{A}$ es una colección de conjuntos, entonces*

$$\bigcup_{i\geq 1} A_i \in \mathcal{A}$$

Nota 1.3. Cuando hablamos unσ-álgebra sobre Ω, consideramos a Ω como nuestro universo. Por lo que todo complemento es subconjuntos de Ω.

Proposición 1.1. *Sea $\mathcal{A}$ un σ-álgebra sobre Ω. Entonces se verifican las siguientes propiedades:*

1. $\emptyset \in \mathcal{A}$
2. *Sea $\{A_i\}_{i=1}^{n} \subseteq \mathcal{A}$ una colección de conjuntos. Entonces*

$$\bigcup_{i=1}^{n} A_i, \bigcap_{i=1}^{n} A_i \in \mathcal{A}$$

3. *Si $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{A}$ es una colección de conjuntos. Entonces*

$$\bigcap_{i\geq 1} A_i \in \mathcal{A}$$

Dem.

1. Si $\Omega \in \mathcal{A}$, entonces también lo está el complemento, este caso $\emptyset = \Omega^C \in \mathcal{A}$

2. Consideremos la siguiente colección $\{B_i\}_{i\in\mathbb{N}}$ donde $B_i := A_i \in \mathcal{A}$ si $i \leq n$ y $B_i = \emptyset$ si $i > n$. Claramente tenemos una colección infinita. Luego

$$\bigcup_{i=1}^{n} A_i = \bigcup_{i\geq 1} B_i \in \mathcal{A}$$

   Para probar la intersección, basta ver que dado $\{A_i\}_{i=1}^n$, existe la colección de los complementos $\{A_i^C\}_{i=1}^n$ que son medibles, luego uniendo los complementos llegamos a un conjunto y, para concluir el resultado basta aplicar el complemento, concluyendo que

$$\left(\bigcup_{i=1}^{n} A_i^C\right)^C = \bigcap_{i=1}^{n} A_i \in \mathcal{A}$$

3. Recordemos la ideal del segundo punto, esta vez consideremos la colección complemento $\{A_i^C\}_{i\in\mathbb{N}}$, luego aplicando el segundo axioma de σ-álgebra a la unión de la colección complemento se llega al restuldado.

Probando la proposición. ■

Algo importante sobre los σ-álgebra, es que son conjuntos "medibles", concepto que aun no definimos, pero nos referiremos a que todo lo que tomemos en σ-álgebra sea funcional al momento de definir una función. Por lo que siempre que tomemos un elemento de σ-álgebra, este de alguna forma, es medible.

Definición 1.9. *Sea $\mathcal{A}$ un σ-álgebra sobre Ω. Al par $(\Omega, \mathcal{A})$ se le denomina espacio medible o espacio de suceso. Y si $A \in \mathcal{A}$, entonces se dice que A es medible.*

Ejemplo 1.4.

- Sea $\mathcal{A} = \{\emptyset, \Omega\}$, podemos ver que es un σ-álgebra. A este conjunto le llamamos σ-álgebra trivial

- Sea $A \subseteq \Omega$, luego el conjunto $\mathcal{A} = \{\emptyset, A, A^C, \Omega\}$ es un σ-álgebra.

- El conjunto $\mathcal{A} = \mathcal{P}(\Omega) := \{X : X \subseteq \Omega\}$ es un σ-álgebra. Probemos que lo es

  1. Por definición $\Omega \in P(\Omega)$

  2. Por definición Ω es el universo, por lo que el complemento siempre será subconjunto de Ω.

  3. Sea $\{A_i\}_{i\in\mathbb{N}} \subseteq P(\omega)$ una colección de conjuntos, por propiedades de conjuntos se tiene que

$$\bigcup_{i\geq 1} A_i \subseteq \Omega$$

Viendo que $P(\Omega)$ es un σ-álgebra. Otra forma de nombrar este conjunto es por 2^{Ω} σ-álgebra.

4. Sea $C = \{A_1, A_2\}$, donde $A_1, A_2 \subseteq \Omega$. Para encontrar un σ-álgebra que tiene el elemento C, debemos considerar que este σ-álgebra contiene a Ω, C y su complemento, y la propiedad de que toda unión infita contable, este en el σ-álgebra. Un candidato podría ser

$$\mathcal{A} = \{\emptyset, A_1, A_1^C, A_2, A_2^C, A_1 \cup A_2, A_1^C \cup A_2^C, A_1 \cup A_2^C, A_1^C \cup A_2, \Omega\}$$

Se puede ver que $\mathcal{A}$ es un σ-álgebra.

Ejemplo 1.5. Digamos que $\Omega = \mathbb{R}$, y sea $\mathcal{A}$ el conjunto de todos los intervalos de la forma

$$[a, b], \ \ (a, b], \ \ (a, b) \ \ [a, b)$$

donde $a < b$. Además $\mathcal{A}$ contiene las uniones e intersecciones de todos los subconjuntos de esa formar por las propiedades de σ-álgebra. Este conjunto es un σ-álgebra y se le conoce como σ-álgebra de Borel, denotado por $\mathcal{B}$. Es facil comprobar los axiomas de σ-álgebra. Como particularidad de este conjunto de Borel, podemos extender a los $\mathbb{R}^n$, generando los Borelianos y se denota por $\mathcal{B}_n$.

Teorema 1.2. *Sea Ω un conjunto no vacío. Y sean $\{\mathcal{A}_i\}_{i\in\mathbb{N}}$ una colección de σ-álgebras sobre Ω. Entonces*

$$\bigcap_{i\geq 1} \mathcal{A}_i$$

es un σ-álgebra sobre Ω.

Dem.

1. Notemos que $\Omega \in \mathcal{A}_i$ para todo $i \in \mathbb{N}$, luego

$$\Omega \in \bigcap_{i\geq 1} \mathcal{A}_{\rangle}$$

2. Sea

$$A \in \bigcap_{i\geq 1} \mathcal{A}_i$$

un evento, esto implica que $A \in \mathcal{A}_i$ para todo $i \in \mathbb{N}$, y como son σ-álgebras, se tiene que $A^C \in \mathcal{A}_i$, por tanto

$$A^C \in \bigcap_{i\geq 1} \mathcal{A}_i$$

3. Sea $\{A_i\}_{i\in\mathbb{N}}$ una colección de eventos tales que

$$A_k \in \bigcap_{i\geq 1} \mathcal{A}_i$$

para todo $k \in \mathbb{N}$. Luego

$$\bigcup_{j\geq 1} A_j \in \mathcal{A}_i$$

para todo $i \in \mathbb{N}$. Por tanto

$$\bigcup_{j\geq 1} A_j \in \bigcap_{i\geq 1} \mathcal{A}_i$$

Probando que la intersección infinita contable de σ-álgebras, es un σ-álgebra. ■

Definición 1.10. (σ-álgebra generado por una colección) *Sea Ω un conjunto no vacío. Sea $\mathcal{X}$ una colección de subconjuntos de Ω, y sea*

$$\mathcal{M} := \{\mathcal{A} : \mathcal{A} \text{ es un } \sigma\text{-álgebra sobre } \Omega \text{ y que contiene a } \mathcal{X}\}$$

Entonces

$$\sigma(\mathcal{X}) := \bigcap_{\mathcal{A}\in\mathcal{M}} \mathcal{A}$$

es la σ-álgebra más pequeña sobre Ω que contiene a $\mathcal{X}$. A $\sigma(\mathcal{X})$ se le llama σ-álgevra generado por $\mathcal{X}$.

Es facil ver que es el más pequeño, porque estamos intersectando todo los σ-álgebra que contiene a $\mathcal{X}$, y la intersección es casi sinónimo de reducir, ya que vamos tomando elementos en común de cada conjunto.

2. Modelos Probabilísticos

2.1. Experimentos Aleatorios y Espacio Muestral

Luego de revisar la teoría de conjuntos, estamos listo para recrear y estudiar los modelos probabilísticos. Sentaremos la base de la probabilidad empezando por definiciones importantes.

Un modelo o espacio de probabilidad, es una representación matemática de una situación con resultado con cierta incertidumbre, en ese caso, a fenónemos aleatorios.

Definición 2.1. *Llamaremos experimento aleatorio a un proceso bien estructurado que entrega resultados que no puede ser predecidos en su totalidad.*

Notación. En algunas ocasiones escribiremos E.A para referirnos a un experimento aleatorio.

Nota 2.1. Un E.A se le puede considerar varios tipos de resultado, todo depende del tipo de resultado que queremos.

Ejemplo 2.1.

- Lanzar una moneda y ver que sale, se considera un experimento aletaorio, ya que no podemos predecir si la moneda saldrá cara o saldrá sello.
- Tomar el tiempo de duración de una ampolleta.
- Contar la cantidad de persona que toman el metro en un día.
- El tiempo de sueño de una persona.
- La nota que tendrá un alumno en un examen.
- Cantidad de Kilómetros recorrido en un lapso de tiempo fijo.

Definición 2.2. *Dado un E.A, el conjunto de todos los resultados posibles del E.A se le llama espacio muestral (E.M) y se denota por el símbolo Ω.*

Es importante interpretar bien el experimento aleatorio para determinar el espacio muestral.

Ejemplo 2.2.

- Lanzar una moneda tiene como espacio muestral el conjunto $\Omega = \{C, S\}$ donde C es cara y S es sello.
- Si hay n artefactos en una frábica, y si la cantidad de cuantos fallan es nuestro experimento aleatorio, entonces el espacio muestral es $\Omega = \{0, 1 \ldots, n\}$.
- El tiempo de duración de una ampolleta tiene como espacio muestral $\Omega = \mathbb{R}_{\geq 0}$.

Definición 2.3. *Decimos que ω es un resultado posible del E.A si y sólo si $\omega \in \Omega$ donde Ω está dado por el E.A.*

Ejemplo 2.3. Del ejemplo anterior tenemos que

- C es un posible resultado del E.A lanzar una moneda y ver que cae, ya que $C \in \Omega$. Lo mismo se puede decir de S.
- 6 es un resultado posible del E.A ya que $6 \in \Omega$, ene general, todo elemento de $\mathbb{N} \cup \{0\}$ es un resultado posible del E.A.
- π es un resultado posible del E.A ya que $\pi \in \mathbb{R}_{\geq 0}$.

Nota 2.2. Un resultado $\omega \in \Omega$ también se le conoce como punto muestral o suceso elemental.

Nota 2.3. El espacio muestra Ω de un experimento aleatorio debe ser bien definido y puede ser tan abstracto como se quiera. Pero lo más importante es que nunca sea vacío y aunque no se especifique, se debe asumir por convenio que no lo es.

Definición 2.4. *Sea un E.A que genera un E.M Ω. Diremos que Ω es un:*

1. *Espacio Muestral Discreto (E.M.D) si Ω es discreto.*
2. *Espacio Muestral Continuo (E.M.C) si Ω es no contable.*

Aunque es redundante hablar de conjuntos discretos y continuos, es importante ya que la probabilidad se construye a partir de las caracteristica del conjunto Ω.

Ejemplo 2.4. Del ejemplo 2.2 tenemos que

- Lanzar una mondea y ver que sale, genera un espacio muestral discreto.
- Ver de n artefactos, cuantos sales defectuoso, genera un espacio muestral discreto.
- El tiempo de duración de una ampotella, tiene como espacio muestral Ω.

2.2. Eventos

Definición 2.5. *Sea E.A con espacio muestral Ω, decimos que A es un evento de Ω si y sólo si $A \subseteq \Omega$. Además, decimos que el evento A ocurre si y sólo existe un $\omega \in \Omega$ talque $\omega \in A$.*

Dicho de otra una forma, un evento de Ω ocurre si ocurre un resultado ω y está en el evento.

Nota 2.4. Puede ocurrir que el evento A sea todo Ω o sea vacío. Si $A = \Omega$ entonces decimos que A es un evento seguro y si $A = \emptyset$, decimos que A es un evento imposible. Si $A = \{\omega\}$ con $\omega \in \Omega$, entonce a A se le llama evento simple o elemental. De forma paradójica, tanto $A = \emptyset$ y $A = \Omega$, siempre van a ocurrir si ocurre $\omega \in \Omega$.

Nota 2.5. Si un evento $A \subseteq \Omega$ ocurre, entonces por definición $A \neq \emptyset$.

Podemos generar nuevo eventos y saber cuando estos nuevos eventos ocurren.

Proposición 2.1. *Sean $A, B \subseteq \Omega$ un eventos que ocurren de un E.A. Entonces los siguientes conjuntos son eventos que además ocurren:*

1. A^C
2. $A \cap B$
3. $A \cup B$
4. $A \triangle B := (A^C \cap B) \cup (A \cap B^C)$

Si ocurre A y no B, entonces el evento

$$A \setminus B$$

ocurre.

Dem. Es evidente que los cuatros conjuntos son eventos por definición mismo de evento. Nos queda probar que si A, B son eventos ocurrentes, entonces los conjuntos presentados también lo son.

1. Si $A = \Omega, \emptyset$, entonces $A^C = \emptyset, \Omega$ y por tanto ocurre, si $A \neq \Omega$, entonces $A^C \subseteq \Omega$, es decir, es un evento. Y si ocurre que A, entonces existe un $\omega \notin A$ talque $\omega \in A^C$, luego ocurre A^C.
2. Si $A \cap B = \Omega, \emptyset$ estamos listos. Si $A \cap B \neq \Omega, \emptyset$, entonces existe un elemento $\omega \in A \cap B$, luego por definición $\omega \in \Omega$, por lo que es un evento que ocurre.
3. Si $A \cup B = \Omega, \emptyset$ estamos listos. Si A y B son eventos, entonces necesariamente hay un elemento $\omega \in \Omega$ talque $\omega \in A \cup B$, de forma que es un evento que ocurre.

4. Por definición, la diferencia simétrica
$$A \triangle B := (A^C \cap B) \cup (A \cap B^C)$$
luego tenemos uniones, intersecciones y complementos de eventos que ocurren. Por tanto la diferencia simétrica de A y B es un evento que ocurre.

Supongamos ahora que ocurre el evento A y no el evento B, entonces existe un $\omega \in \Omega$ talque $\omega \in A$ y $\omega \notin B$, esto implica que $\omega \in B^C$, entonces $\omega \in A \cap B^C = A \setminus B$. Por tanto $A \setminus B$ es un evento que ocurre. Probando la proposición. ■

Nota 2.6. Al evento A^C se le llama, evento complementario.

Observación 2.1. Si $A, B \subseteq \Omega$ tales que $A \subseteq B$, entonces si ocurre el evento A, necesariamente ocurre el evento B.

Definición 2.6. *Sean $A, B \subseteq \Omega$ eventos. Si $A \cap B = \emptyset$ decimos que son eventos incompatibles o eventos mutuamente excluyentes.*

Ejemplo 2.5. Consideremos el lanzamiento de una modena y sean los eventos $A =$"resultado cara" y $B =$"resultado sello". Se puede ver que $A \cap B = \emptyset$, por lo que son incompatibles.

Notación. Sean A, B eventos de un espacio muestral. Usaremos las siguientes notaciones:

1. $A \cap B = AB$
2. Si $AB = \emptyset$, entonces $A \cup B = A + B$
3. Si $A \subset B$ (subconjunto propio), entonces $A^C B = B - A$

2.3. Modelos de Probabilidad

Construiremos el modelo de probabilidad en base a la frecuencia relativa, estudiaremos su comportamiento y como este construye los axiomas de Kolmogorv sobre la medidad de probabilidad.

Definición 2.7. *Sea A un evento, el número*
$$fr(A) := \frac{n(A)}{n}$$
se llama frecuencia relativa de A, donde $n(A)$ indica la cantidad de veces que ocurrió el evento A en n repeticiones del experimento.

El valor que toma la frecuencia relativa de A es relativo, depende más de seguir datos más que una fórmula general, esto signifca que el valor de $fr(A)$ no es constante y depende n. Pero se ha observado que cuando $n \to \infty$ (n crece sin límite) la función frecuencia se empieza a estabilizar a un valor fijo en el conjunto $[0, 1]$. Por ejemplo, sea el experimento tirar el dado, y sea el evento $A =$"obtener el número 3". Podemos crear una tabla que dicte los valores de la frecuencia relativa.

Este valor donde estabiliza $fr(A)$ no es coincidencia y a este valor "fijo" se le asocía al valor $P(A)$ que es un indicador de posibilidad de que ocurra el evento A. Pero más importante que eso, vemos que

n	Frecuencia	Frecuencia relativa
100	14	0.14
200	29	0.145
300	51	0.17
400	65	0.1625
500	83	0.166

Cuadro 1: Tabla de $fr(A)$

1. $fr(A) \geq 0$ ya que $n(A) \geq 0$, por lo que la posibilidad $P(A) \geq 0$.
2. Si $n(\Omega) = n$ entonces $fr(\Omega) = 1$, dado que Ω es un evento que siempre ocurre, por lo tanto $P(\Omega) = 1$
3. Si A y B son eventos incompatibles o mutuamente excluyente, entonces $n(A + B) = n(A) + n(B)$, luego $fr(A + B) = fr(A) + fr(B)$, por lo que $P(A + B) = P(A) + P(B)$ (recordad que si $A \cap B = \emptyset$ entonces $A \cup B = A + B$).

Y esta son las condiciones que define una medida de probabilidad.

Definición 2.8 (Medida de probabilidad) *Sea $(\Omega, \mathcal{A})$ un espacio medible. Decimos que una función P es una medida de probabilidad sobre $(\Omega, \mathcal{A})$ dada por:*

$$\begin{aligned} P : \mathcal{A} &\to [0, 1] \\ A &\mapsto P(A) \end{aligned}$$

si satisface los siguientes axiomas:

1. ***No negativa.*** *Para todo $A \in \mathcal{A}$, $P(A) \geq 0$*
2. ***Normalizado.*** $P(\Omega) = 1$
3. ***Aditividad contable.*** *Sea $\{A_i\}_{i\in\mathbb{N}}$ una colección de eventos medibles disjuntos (contenido en $\mathcal{A}$). Entonces:*

$$P\left(\bigcup_{i\geq 1} A_i\right) = \sum_{i\geq 1} P(A_i)$$

La terna $(\Omega, \mathcal{A}, P)$ se llama modelo o espacio de probabilidad, y el valor númerico $P(A)$ se denomina probabilidad del evento $A \in \mathcal{A}$.

Recordemos que dado $(\Omega, \mathcal{A})$ un espacio medible, A es medible si y sólo si $A \in \mathcal{A}$, esto nos dice que todo elemento del σ-álgebra son medibles por convenio. Por otro lado nos afirma que existen eventos que no pueden ser medidos, los cuales claramente ignoraremos.

Ejemplo 2.6. Sea $\Omega = \{1, 2, 3\}$ y sea $\mathcal{A} = \{\emptyset, \{1\}, \{2, 3\}, \omega\}$. Entonces, la función $P(A)$, $A \in \mathcal{A}$, definida por

$$P(A) := \begin{cases} 1 & \text{si } 3 \in A \\ 0 & \text{si } 3 \notin A \end{cases}$$

es una medida de probabilidad sobre $(\Omega, \mathcal{A})$. Probemos que se cumple, para ello debemos probar que $(\Omega, \mathcal{A})$ es un espacio medible y que P satisface los axiomas de medida de probabilidad:

- $(\Omega, \mathcal{A})$ **espacio medible.** Notemos que $\mathcal{A}$ contiene a Ω, todo complemento está en $\mathcal{A}$ y la unión siempre está en $\mathcal{A}$. Siendo un espacio medible.
- **P una medida de probabilidad.** Notemos que $P(A) \geq 0$ para todo $A \in \mathcal{A}$, $P(\Omega) = 1$ ya que $3 \in \Omega$ y dada una colección $\{A_i\}_{i\in\mathbb{N}}$ de conjuntos de $\mathcal{A}$, disjuntos. Supongamos que en la colección no esté $\{2,3\}, \Omega$, luego

$$P\left(\bigcup_{i\geq 1} A_i\right) = 0 = \sum_{i\geq 1} 0 = \sum_{i\geq 1} P(A_i)$$

Si en la colección está el conjunto $\{2,3\}$ o Ω, entonces

$$P\left(\bigcup_{i\geq 1} A_i\right) = 1 = 1 + 0 = \sum_{i_1\geq 1, 3\in A_{i_1}} P(A_{i_1}) + \sum_{i_2\geq 1, 3\notin A_{i_2}} P(A_{i_2}) = \sum_{i\geq 1} P(A_i)$$

Siendo una medida de probabilidad.

Observación 2.2. En el ejemplo 2.6, al probar que P estaba bien definido hicimos algo de trampa. Al momento de considerar una colección de conjunto pertenecientes a $\mathcal{A}$, ocurren algunas cosas, si la colección no tiene ni a $\{2,3\}$ y a Ω, entonces es de la forma $\{\emptyset, \{1\}\}$, si contiene a $\{2,3\}$ o a Ω ocurre dos situaciones, si contiene a $\{2,3\}$, entonces Ω no puede estar en la colección ya que tomamos conjuntos disjuntos, siendo de la forma $\{\emptyset, \{1\}, \{2,3\}\}$, y su está Ω, entonces no tiene que esta ni $\{1\}$, ni $\{2,3\}$, siendo de la forma $\{\emptyset, \Omega\}$. Si se verfica se podrá ver que se cumple la tercera propiedad.

Ejemplo 2.5. Sean $\Omega = \mathbb{N} \cup \{0\}$ bien definido, $\mathcal{A} := \mathcal{P}(\Omega)$ y P una función definida como:

$$P(\{i\}) = (1-q)q^i, \text{ donde } i = 0, 1, \ldots, \; 0 < q < 1$$
$$P(A) = \sum_{i:i\in A} P(\{i\}) = (1-q)\sum_{i:i\in A} q^i$$

Es un espacio de probabilidad. Verifiquemos este hecho.

1. $P(A) \geq 0$ por definición.
2. $$P(\Omega) = (1-q)\sum_{i:i\in\mathbb{N}_0} q^i = (1-q)\frac{1}{(1-q)} = 1$$
3. Sea $\{A_i\}$ una colección disjunta de $\mathcal{P}(\Omega)$. Entonces

$$\begin{aligned} P\left(\bigcup_{i\geq 1} A_i =: B\right) &= (1-q)\sum_{i:i\in B} q^i \\ &= (1-q)\left(\sum_{i_1:i_1\in A_1} q^{i_1} + \sum_{i_2:i_2\in A_2} q^{i_2} + \ldots\right) = \sum_{i\geq 1} P(A_i) \end{aligned}$$

Verificando que P es una medida de probabilidad sobre $(\Omega, \mathcal{A})$.

Ejemplo 2.6. Sea $\Omega = \mathbb{N}, \mathcal{A} = \mathcal{P}(\Omega)$. Considere la función definida por

$$P(\{i\}) = \frac{2}{3}\left(\frac{1}{2}\right)^{i-1}, \ i = 1, , 2, \ldots$$

Entonces P no es una medida de probabilidad sobre $(\Omega, \mathcal{A})$, porque si lo fuera, entonces, para la colección $\{\{i\}\}_{i\in\mathbb{N}}$ que es una partición de Ω, se tendría que

$$1 = P(\Omega) = P\left(\bigcup_{i\geq 1}\{i\}\right) = \sum_{i\geq 1} P(\{i\}) = \frac{2}{3}\sum_{i\geq 1}\left(\frac{1}{2}\right)^{i-1} = \frac{2}{3}\cdot 2 = \frac{4}{3}$$

Ejemplo 2.7. Sea $\Omega = (0, \infty)$ y $\mathcal{A} = \mathcal{B}$. Se define la medida de probabilidad P como

$$P(B) := \int_B e^{-x}dx$$

Entonces P es una medida de probabilidad.

1. Si B tiene límite de intervalo $a \leq b$, entonces

$$\int_a^b e^{-x}dx = -e^{-b} + e^{-a} \geq 0$$

2.

$$P(\Omega) = \int_\Omega e^{-x}dx = \lim_{b\to\infty}\int_0^b e^{-x} = 0 + 1 = 1$$

3. Sea $\{B_i\}_{i\in\mathbb{N}}$ una colección disjuntas. Entonces

$$P\left(\bigcup_{i\geq 1} B_i\right) = \int_{\bigcup_{i\geq 1} B_i} e^{-x}dx = \sum_{i\geq 1}\int_{B_i} e^{-x}dx = \sum_{i\geq 1} P(B_i)$$

Teorema 2.1. *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Entonces:*

1. *$P(\emptyset) = 0$.*
2. *Si $A, B \in \mathcal{A}$ y $A \cap B = \emptyset$, entonces $P(A \cup B) = P(A) + P(B)$.*
3. *Para todo $A \in \mathcal{A}$ se tiene que $P(A^C) = 1 - P(A)$.*
4. *Si $A \subseteq B$, entonces $P(A) \leq P(B)$ y $P(B \setminus A) = P(B) - P(A)$. Además $P(A) \leq 1$ para todo $A \in \mathcal{A}$.*
5. *Para todo $A, B \in \mathcal{A}$ se tiene que $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.*

6. *Sea $\{A_n\}_{i\in\mathbb{N}}$ una colección creciente de eventos de $\mathcal{A}$. Entonces*

$$P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n), \quad \textit{donde} \ \lim_{n\to\infty} A_n := \bigcup_{n\geq 1} A_n$$

7. *Sea $\{A_n\}_{i\in\mathbb{N}}$ una colección decreciente de eventos de $\mathcal{A}$. Entonces*

$$P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n), \quad \textit{donde} \ \lim_{n\to\infty} A_n := \bigcap_{n\geq 1} A_n$$

8. *Sea $\{B_i\}$ una partición de Ω, entonces*

$$P(A) = \sum_{i\geq 1} P(A\cap B_i)$$

9. *Sea $\{A_i\}_{i\in\mathbb{N}}$ una colección de eventos. Entonces*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i\geq 1} P(A_i)$$

(Desigualdad de Boole)

Dem. Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad.

1. Si $P(\Omega) = 1$, sea la colección de eventos $\{A_i\}_{i\geq 1}$ donde $A_i = \emptyset$ para todo $i \geq 2$, luego

$$1 = P\left(\Omega\right) = P(\Omega) + \sum_{i\geq 2} P(A_i) = 1 + \sum_{i\geq} P(\emptyset)$$

Podemos ver que

$$\sum_{i\geq 2} P(\emptyset) = 0$$

Necesariamente $P(\emptyset) = 0$.

2. Sean $A, B, C_1, C_2, \dots \in \mathcal{A}$ una colección de eventos donde $C_i = \emptyset$ y $A\cap B = \emptyset$, entonces

$$P(A\cup B) = P\left(A\cup B\cup \bigcup_{i\geq} C_i\right) = P(A) + P(B) + \sum_{i\geq 1} P(C_i) = P(A) + P(B)$$

3. Sea $A\in\mathcal{A}$, si $A\cap A^C = \emptyset$ entonces

$$1 = P(A\cup A^C) = P(A) + P(A^C)$$

Luego $P(A^C) = 1 - P(A)$

4. Sea $A \subseteq B$, si $A = B$ entonces $P(A) = P(B)$, si $A \subset B$ entonces existe un conjunto $C := B \cap A^C \neq \emptyset$, notemos que $A \cap C = \emptyset$, luego $P(A) + P(C) = P(A \cup C) = P(B)$, entonces si $0 \leq P(C)$, entonces $P(A) \leq P(A) + P(C) = P(B)$. Además concluimos que $P(C) = P(B \setminus A) = P(B) - P(A)$. Si $B = \Omega$, entonces $P(A) \leq P(\Omega) = 1$.

5. Sean $A, B \in \mathcal{A}$. Podemos ver que $A \cap B \subseteq B$, que $A \cap (B \setminus (A \cap B)) = \emptyset$ y que $A \cup (B \setminus (A \cap B)) = A \cup B$, de esta forma

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

6. Sea $\{A_n\}_{n\in\mathbb{N}}$ una colección creciente de eventos de $\mathcal{A}$. Sea la colección $\{C_n\}_{n\in\mathbb{N}}$, definida por $C_1 := A_1, C_n := A_n \setminus A_n - 1$, podemos ver $\{C_n\}$ son disjuntos ya que dados $j < i$ se tiene que

$$A_j \cap A_{j-1}^C \cap A_i \cap A_{i-1}^C = A_j \cap A_{i-1}^C = \emptyset$$

porque si $j < i$, entonces $j \leq i_1$. Y además

$$\bigcup_{n\geq 1} C_n = \bigcup_{n\geq 1} A_n$$

Para ver esto, dado $x \in \bigcup C_n$, entonces existe un $i \in \mathbb{N}$ talque $x \in C_i = A_i A_{i-1}^C$, por definición $x \in A_i \subseteq \bigcup A_i$. Por otro lado, si $x \in \bigcup A_i$, existe un $n \in \mathbb{N}$ minimal talque $x \in A_n$ (con decir minimal, nos referimos a que n es el menor natural talque $x \in A_n$). Luego $x \notin C_j$ con $j < n$ ya que entonces n no sería minimal y si $j > n$ tampoco puede estar, ya que $x \in C_j$ para todo $j > n$ por definición quita todo los elementos que están en A_n. Probemos que $x \in C_n$, notemos que $x \notin A_{n-1}$ ya que n es minimal, luego $x \in A_{n-1}^C \cap A_n = C_n$. Por tanto $x \in \bigcup C_i$. Probando la igualdad. Volviendo a la demostración, se tiene observa que

$$\begin{aligned} P\left(\bigcup_{n\geq 1} A_n\right) &= P\left(\bigcup_{n\geq 1} C_n\right) \\ &= \sum_{n\geq 1} P(C_n) = \sum_{n\geq 2} P(A_n) - P(A_{n-1}) + P(A_1) = \lim_{n\to\infty} P(A_n) \end{aligned}$$

7. Vamos a usar el punto anterior para demostrar el límite. Sea $\{A_n\}_{n\in\mathbb{N}}$ una secuencia decreciente, entonces la secuencia $\{B_n\}_{n\in\mathbb{N}}$ definida por $B_n := A_n^C$, es creciente. Luego

$$\lim_{n\to\infty} B_n = \bigcup_{n=1}^{\infty} B_n$$

Luego

$$P\left(\lim_{n\to\infty} B_n\right) = \lim_{n\to\infty} P(B_n)$$

Luego tomando el complemento, tenemos que

$$P\left(\bigcup_{n=1}^{\infty} A_n^C\right) = \lim_{n\to\infty}(1 - P(A_n)) \Leftrightarrow 1 - P\left(\bigcap_{n=1}^{\infty} A_n\right) = 1 - \lim_{n\to\infty} P(A_n)$$

Concluyendo el resultado.

$$\begin{aligned} P(A) = P(A\cap\Omega) &= P\left(A\cap\bigcup_{i\geq 1} B_i\right) \\ &= P\left(\bigcup_{i\geq 1}(A\cap B_i)\right) = \sum_{i\geq 1} P(A\cap B_i) \end{aligned}$$

(notar que $(A\cap B_i)\cap(A)\cap B_j$ son disjuntos para todo $i\neq j$.)

8. Para este problema usaremos el lema 1.1. Sea $\{A_i\}_{i\in\mathbb{N}}$ una colección de eventos medibles, luego la colección $\{B_i\}$ definida por $B_1 := A_1, B_n := A_1^C A_2^C \ldots A_{n-1}^C A_n$ es talque los $B_i \cap B_j$ son disjuntos y

$$\bigcup_{i\geq 1} B_i = \bigcup_{i\geq 1} A_i$$

Luego

$$P\left(\bigcup_{i\geq 1} A_i\right) = P\left(\bigcup_{i\geq 1} B_i\right) = \sum_{i\geq 1} P(B_i) \leq \sum_{i\geq 1} P(A_i)$$

ya que para todo $i\in\mathbb{N}$ se tiene que $B_i\subseteq A_i$, luego $P(B_i)\leq P(A_i)$.

Probando el teorema. ■

Nota 2.7. En general, Si $A, B\in\mathcal{A}$, entonces

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

Ya que $B = B\setminus A + A\cap B$, luego por propiedades de medidad de probabilidad, se concluye el resultado.

Nota 2.8. El límite $\lim P(A_n)$ está bien definido para una colección $\{A_n\}$ monótona, ya que se puede ver que $\{P(A_n)\}$ es una sucesión monótona y acotada y por tanto, converge a un valor $L\in[0,1]$. Lo mismo se concluye con la sucesión

$$\left\{P\left(\bigcup_{i=1}^{n} A_i\right)\right\}_{i\in\mathbb{N}}$$

Nota 2.9. Se puede ver que $A\cup B = A + A^C B$ y $B = AB + A^C B$.

Teorema 2.2. *Sea* $\Omega = \{\omega_1, \dots, \omega_{N(\Omega)}\}$ *un espacio finitos. Sea* $\mathcal{A}$ *cualquier* σ*-álgebra sobre* Ω*. Sean* $p_1, \dots, p_{N(\Omega)}$ *números no negativos que suman* 1*. Para cualquier* $A \in \mathcal{A}$*, se define* $P(A)$ *por:*

$$P(A) := \sum_{\{i:w_i \in A\}} p_i$$

Entonces P *es una medida de probabilidad sobre* $\mathcal{A}$*. También se cumple si* $\Omega = \{\omega_1, \dots\}$ *es un conjunto conbtable y* $p_1, \dots$ *son números no negativos tales que* $\sum_{i \geq 1} p_i = 1$*. En tales casos, decimos que* $(\Omega, \mathcal{A}, P)$ *es un espacio de probabilidad discreto.*

Dem. Probaremos el caso finito ya que el infinito es una extensión de la idea. Sea $\Omega := \{\omega_1, \dots, \omega_{N(\Omega)}\}$ y sea un $\mathcal{A}$ un σ-álgebra. Debemos probar que la función P es no negativa, $P(\Omega) = 1$ y que dada una colección disjuntas de elementos de $\mathcal{A}$, la probabilidad de la unión es la suma de las probabilidad por separada.

1. Sea $A \in \mathcal{A}$, por definición

$$P(A) = \sum_{\{i:w_i \in A\}} p_i \geq 0$$

2. Notemos que

$$P(\Omega) = \sum_{i=1}^{N(\Omega)} p_i = 1$$

3. Sea $\{A_n\}_{n \in \mathbb{N}}$ una colección disjunta de elementos de $\mathcal{A}$. Digamos que $\bigcup A_n = B$, luego

$$\begin{aligned} P\left(\bigcup_{n \geq 1} A_n\right) &= \sum_{\{i:\omega_i \in B\}} p_i \\ &= \sum_{\{i_1:\omega_{i_1} \in A_1\}} p_{i_1} + \cdots + \sum_{\{i_j:\omega_{i_j} \in A_j\}} p_{i_j} \\ &= \sum_{n=1}^{j} P(A_n) \end{aligned}$$

(Podemos tomar j conjuntos A_j, ya que Ω es finito, si tomamos una colección infinita, entonces tanto $\emptyset, \Omega$ estaría en la colección.)

Probando que P es una medida de probabilidad. Si Ω fuera contable infinito, la única diferencia es que con respecto a ω_{i_j} podemos ver que j está indesada por $\mathbb{N}$.

Por tanto $(\Omega, \mathcal{A}, P)$ es un espacio medible discreto. ■

Observación 2.3. Del teorema 2.2, podemos desprender que $(\Omega, \mathcal{A}, P)$ es discreto si y sólo si Ω es discreto, es decir, cuando Ω es finito o es infinito contable.

Ejemplo 2.8. Sea $\Omega = \{\omega_1, \dots, \omega_{N(\Omega)}\}$ un espacio muestral equiprobable, es decir:

$$p_i := P(\{\omega\}) = \frac{1}{N(\Omega)}, \text{ para todo } i = 1, \dots, N(\Omega)$$

En virtud de que estamos suponiendo que es una medida de probabilidad, tenemos que para cualquier $A \in \mathcal{A}$ se tiene que

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{\text{Casos Favorables}}{\text{Casos Totales}}$$

En tal caso decimos que $(\Omega, \mathcal{A}, P)$ es un espacio de probabilidad discreto equiprobable.

Observación 2.4. El espacio de probabilidad discreto equiprobabe es uno de los más fáciles de estudiar, ya que estamos aplicando una probabilidad intuitiva. Algunos ejemplos conocidos son lanzar una moneda, lanzar dados o n dados, la probabilidad asociada al conteo etc. Siempre tomando que son objetos justos.

Ejemplo 2.9. Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad con $\Omega = \{1, 2, 3, 4\}$, $\mathcal{A} = \{\emptyset, \{1\}, \{4\}, \{2, 3\}, \{1, 4\},$ y $P : \mathcal{A} \to [0, 1]$ dada por

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

Podemos ver que es un espacio discreto. Aunque no sabemos si es equiprobable.

Ejemplo 2.10. Sea $(\Omega, \mathcal{P}(\Omega), P)$ un espacio de probabilidad discreto con $\Omega = \{a, b, c\}$ y P dado por el vector de probabilidad $p = \left(\frac{1}{7}, \frac{4}{7}, \frac{2}{7}\right)$. Es decir

$$P(\{a\}) = \frac{1}{7},\ P(\{b\}) = \frac{4}{7},\ P(\{c\}) = \frac{2}{7}$$

Claramente es discreto por hipótesis, y también por construcción, pero no es equiprobable ya que la probabilidad de que ocurre $\{b\}$ es distinto a que ocurra $\{a\}, \{c\}$.

Definición 2.9 (Modelo de Probabilidad Equiprobable) *Un modelo de probabilidad $(\Omega, \mathcal{A}, P)$ con Ω finito, $\mathcal{A} = \mathcal{P}(\Omega)$ y $P(\{w\}) = \frac{1}{N(\Omega)}$ para todo $\omega \in \Omega$, donde $N(\Omega)$ es la cantidad de elementos de Ω. A este modelo, le llamamos modelo de probabilidad equiprobable o de Laplace. En tal caso, decimos P es la distribución uniforme o clásica sobre Ω.*

Nota 2.10. Notemos que la condiciones Ω ser finita y $\mathcal{A} = \mathcal{P}(\Omega)$, no son innecesaria. Claramente Ω debe ser finita para que la probabilidad esté bien definida. Mientras la otra condición podemos probarlo. Si todo evento de la forma $\{\omega\}$ con $\omega \in \Omega$ tiene su probabilidad, entonces si $A \subseteq \Omega$, existen $\omega_{n_1}, \dots, \omega_{n_j} \in \Omega$, tales que

$$\bigcup_{k=1}^{j} \{\omega_{n_k}\} = A \in \mathcal{A}$$

por propeidades de σ-álgebra, por tanto

$$\mathcal{P}(\Omega) \subseteq \mathcal{A}$$

y luego por definición se obtiene la igualdad.

En virtud de las propiedades de modelos de probabilidad, podemos determinar la probabilidad de un evento, en una medida equiprobable. Sea A un evento y digamos que tiene $N(A)$ elementos. Entonces dado P una distribución clásica sobre Ω, se observa que

$$P(A) = P\left(\bigcup_{\omega\in A}\{\omega\}\right) = \sum_{\omega\in A}\frac{1}{N(\Omega)} = \frac{N(A)}{N(\Omega)}$$

en otras palabras, casos favorables de A sobre los casos totales.

2.4. Técnicas de Conteo

Nos centraremos en espacios de probabilidad discreto equiprobable. El único problema de estos problema es que no se nos entrega el espacio muestra y no sabemos determinar bien los eventos, ya que requiere de técnicas específicas de conteo para ser determinados, cosa que estudiaremos en toda la sección.

Definición 2.10. *Sean A, B conjuntos no vacío. Se define el producto cartesiano de A luego B por*

$$A \times B := \{(x, y) : x \in A, y \in B\}$$

Podemos realizar infinita multiplicaciones por producto cartesiano.

Teorema 2.3. (Principio Multiplicativo) *Supongamos que un experimento aleatorio se divide en dos etapas, E_1 y E_2. Si E_1 tiene N_1 resultados posibles con $\Omega_1 = \{x_1, \ldots, x_{N_1}\}$ e independiemente de cual ocurra, E_2 tiene N_2 resultados posibles con $\Omega_2 = \{y_1, \ldots, y_{N_2}\}$. Entonces el espacio muestral Ω al experimento, se determina con $\Omega_1 \times \Omega_2$ con $N(\Omega) = N_1N_2$ resultados posibles.*

Dem. Si Ω_1 y Ω_2 son independiente de que ocurran, es decir, si ocurre x_i, entonces puede ocurrir y_j con $j = 1, \ldots, N_2$. Es claro que $\Omega := \{(x, y) : x \in \Omega_1, y \in \Omega_2\}$ tiene $N_1 \cdot N_2$ resultados posibles. ■

Nota 2.11. Siempre es importante diferenciar independencia de las etapas , puede existir un factor tiempo o un factor de que ocurre al mismo tiempo pero al ser independiente, los podemos trabajar de forma separada.

Ejemplo 2.12. Si lanzamos dos dados y observamos los números que salen. Se genera el espacio muestral $\Omega = \{(i, j) : i, j \in \{1, \ldots, 6\}\}$ con 6^2 casos posibles. También podemos representarlo por una tabla

Teorema 2.4. (Extensión del Principio Multiplicativo) *Sea un experimento aleatorio con $E_1, E_2, \ldots, E_n$ etapas donde E_i tiene N_i posibles resultados para todo $i = 1, \ldots, n$ y además son independientes entre si, es decir, los resultados de E_1 puede ser seguido por cualquier resultado de E_2 y así sucesivamente. Luego el espacio muestral Ω tiene $N(\Omega) = N_1 \cdot N_2 \ldots N_n$ resultados posibles.*

	1	2	...	6
1	(1,1)	(1,2)	...	(1,6)
2	(2,1)	(2,2)	...	(2,6)
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
6	(6,1)	(6,2)	...	(6,6)

Cuadro 2: Tabla de los resultados posibles

Dem. La demostración no es complicada, de hecho una recursión. Es claro que es solo contar por casos. Si $x_{i,j} \in \Omega_i$ donde $j = 1, \ldots, N_i$, y se fijan $x_{1,j}, \ldots, x_{n-1,j}$ existen N_n formas de tomar el último espacio muestral, si se fija hasta los $n-2$, habrán $N_{n-1}N_n$ formas de tomar los últimos dos espacios muestrales y así recusivamente. ■

Ejemplo 2.13.

- Lanzar n veces una moneda genera el espacio muestral
$$\Omega = \{C, S\}^n$$
donde $N(\Omega) = 2^n$. Es claro que cada lanzamiento es una etapa y que son independientes entre si.
- Para n lanzamiento de un dado, se tiene que
$$\Omega = \{1, 2, 3, 4, 5, 6\}^n$$
y $\mathbb{N}(\Omega) = 6^n$.
- Lanza primero una moneda y luego un dado, entonces
$$\Omega = \{C, S\} \times \{1, 2, 3, 4, 5, 6\}$$
Y $N(\Omega) = 12$.

Definición 2.11. *Un Muestreo Aleatorio es la selección al azar, una muestra dada una población. Esta selección al azar, garantizar que cada sujeto tiene la misma probabilidad de quedar o no en la muestra, esto implica que toda muestra del mismo tamaño, tiene la misma probabilidad de ser extradio. Por tanto, esto es un modelo de probabilidad equiprobable.*

Puede ser complejo entender que es equiprobable, pero tengamos en cuenta que solo estamos sacando elementos de un conjunto sin ninguna condición, es similar al sacar una pelota de una urna, o tomar n sin mirar.

Definición 2.12. *Se define el experimento por extraer una muestra aleatoria de n elementos de N elementos.*

El espacio muestral generado Ω es el conjunto finito de todas las muestras posibles de n elementos. Cada $\omega \in \Omega$ es una n-tupla de la forma $\omega = (s_1, \ldots, s_n)$ donde $s_i \in \{1, \ldots, N\}$. Si $N(\Omega)$ es la cantidad de elementos de Ω, entonces existe $N(\Omega)$ muestras con tamaño n extraidas de una población de tamaño N. Si este muestreo es aleatorio, estaremos generando una medida de probabilidad definida sobre $\mathcal{P}(\Omega)$, la cual es equiprobable ya que estamos con σ-álgebra la potencia de Ω y con probabilidad

$$P(\{w\}) = \frac{1}{N(\Omega)}$$

para todo $\omega \in \Omega$, luego si A es un evento, entonces

$$P(A) = \frac{N(A)}{N(\Omega)}$$

para todo $A \in \mathcal{P}(\Omega)$.

Pero nos falto aclarar algo, el como se selecciona $N(\Omega)$. Esto tiene que ver de las condiciones existentes en el experimento, pero siempre se estudiara 4 situaciones de escoge las muestras que componen tantos a los eventos como a Ω.

Muestra	Con Devolución	Sin Devolución
Ordenado	CO	SO
No ordenado	CN	SN

Cuadro 3: Table de formas de sacar la muestra

Antes de entrega un resultado entendamos que significa cada cosa. Sea P una población de N sujetos, y sea Ω las muestras con tamaño $n \leq N$ sujetos. Podemos decir al sacar n sujetos, no es lo mismo tomar $(a_1, a_2 \ldots, a_n)$ que tomar $(a_2, a_1, \ldots, a_n)$ si es que existe un orden importante, si el orden no importara, entonces

$$(a_1, a_2, \ldots, a_n) = (a_{j_1}, \ldots, a_{j_n})$$

donde $j_i = 1, \ldots, n$ entrega un orden cualesquiera y $j_i \neq j_s$ si $i \neq s$.

Mientras que la devolución entrega la posibilidad de tomar el mismo sujeto ya tomado con anterioridad aumentando la cantidad posibles. Y si no tiene devolución, el sujeto se queda en la muestra.

Podemos concluir que una muestra sea con o sin orden y con o sin devolución, hace variar su formas posibles, en particular

$$\mathcal{C}(SN) \leq \mathcal{C}(CN), \mathcal{C}(SO) \leq \mathcal{C}(CO)$$

donde $\mathcal{C}$ es la cantidad que genera un tipo de muestra extraida.

Nota 2.10. No damos una comparación directa de $\mathcal{C}(CN)$ y $\mathcal{C}(SO)$ ya que aun no sabemos como actuan entre si, por otro lado sabemos como funciona el resto. Todo siguiendo una idea intuitiva.

Teorema 2.5. *Sea N la cantindad de una problación. Se quiere extraer una muestra de n sujetos de un total de N. Entonces:*

1. *El número total de muestras ordenadas sin devolución es:*
$$(N)_n := \frac{N!}{(N-n)!} = N(N-1)(\dots)(N-n+1)$$
2. *El número total de muestras ordenadas con devolución es:*
$$N^n$$
3. *El número total de muestras no ordenadas sin devolución es:*
$$\binom{N}{n} := \frac{N!}{n!(N-n)!}$$
4. *El número total de muestras no ordenadas con devolución es:*
$$\binom{N+n-1}{n} = \frac{N+n-1}{n!(N-1)!}$$

Con esto construimos una tabla mejorada

Muestra	Con Devolución	Sin Devolución
Ordenado	N^n	$(N)_n$
No ordenado	$\binom{N+n-1}{n}$	$\binom{N}{n}$

Cuadro 4: Table de formas de sacar la muestra

Dem.

1. Sea una población N sujetos y queremos extraer n sujetos para una muestra ordenada sin devolución. Entonces tenemos n etapas, donde la primera tiene un elemento más que la segunda y así sucesivamente, luego la primera etapa tiene N formas, la segunda tiene $N-1$ formas y así hasta la etapa n que tiene $N-n+1$ formas de escoger. Luego por el principio multiplicativo, tenemos
$$N(N-1)(\dots)(N-n+1) = \frac{N!}{(N-n)!} = (N)_n$$
2. Las etapas no varian de cantidad por lo tanto, por el principio multiplicativo la cantidad de meustras son:
$$N \dots N = N^n$$
3. Sea M el numero de muestras no ordenadas sin devolución. Podemos ver que M multiplicado por el número de formas de ordenar a los n sujertos de la muestra, es igual a la cantindad de muestras ordenadas sin devoluciones, es decir
$$M \cdot (n!) = (N)_n$$
Por tanto
$$M = \frac{(N)_n}{n!} = \binom{N}{n}$$

4. **Por demostrar**

Con las técnicas de conteo, podemos realizar una analogía. Sea una población de tamaño N de pelotitas enumeradas del 1 al N. Queremos extraer n pelotas.

1. La extración (no) ordenada de n sujetos, equivale a distribuir n bolitas (no) diferenciable.
2. La extración con (sin) devolución de n sujetos, equivale a distribuir n bolitas sin (con) exclusión.

Ejemplo 2.13.

Sea la población $\{a, b, c, d\}$, queremos escoge dos objetos al azar. Entonces

1. Hay $(4)_2 = \frac{4!}{2!} = 12$ formas de escoger meustras con orden y sin devolución.
2. Hay $4^2 = 16$ formas de escoger muestras con orden y con devolución.
3. Hay $\binom{4}{2} = 6$ formas de escoger muestras sin orden y sin devolución.
4. Hay $\binom{4+2-1}{2} = 10$ formas de escoger muestras sin orden y con devolución.

Ejemplo 2.14. Una urna contiene 5 bolitas enumeradas del 1 al 5. Se sacan sucesivamente al azar 5 bolitas sin reposición. Se desea calcular la probabilidad de que al juntas los números de cada bolita según el orden de extracción, resulta el número 21345.

Sol. Tenemos una población y muestras, por que se genera una probabilidad de dsitribución clásica. Por lo que debemos encontrar el valor de

$$\frac{\text{Casos Favorables}}{\text{Casos Totales}}$$

Dado que el muestreo es con orden y sin reposición, tenemos por casos totales $5! = 120$. Ahora para los casos favorables, notemos que queremos en orden 21345, pero esto es una sola posibilidad, siendo 1 el caso favorable. Luego

$$P(A) = \frac{1}{120}$$

es la probabilidad de extraer en orden 21345.

Ejemplo 2.15. En cierta lotería, se eligen seis números del 1 al 49. La probabilidad de que los números elegidos sean $1, 2, 3, 4, 5$ y 6 es igual a

$$\frac{1}{\binom{49}{6}}$$

Ya que los estamos tomando muestras sin orden y sin reposición, luego los casos totales son $\binom{49}{6}$, y existe un solo caso donde se escogen $1, 2, 3, 4, 5, 6$. Y no necesariamente deben seguir una secuencia, incluso la probabilidad de que salgan los números $2, 41, 7, 39, 34$ y 37 es la misma.

Ejemplo 2.16. Suponga que los 365 días del año tienen la misma probabilidad de ser el dia de cumpleaños de una persona. Determine la probabilidad p de que en un grupo de n personas, ningulo de ellos tengan el mismo cumpleaños.

Sol. Tenemos 365 fechas posibles de cumpleaños a la cual podemos asociarlo a n persona, notemos que las personas son distintas (con orden) y que pueden repetir fechas (con devolución). Luego los casos totales son $N(\Omega) = 365^n$, queremos ahora la cantindad de muestras donde las personas no tenga la misma fecha de nacimiento, esto es tomar con orden y sin devolución. Luego la probabilidad p es:

$$p = \frac{(365)_n}{365^n} = \frac{365!}{n!365^n} = \left(1 - \frac{1}{365}\right) \dots \left(1 - \frac{n-1}{365}\right)$$

Con esta probabilidad, podemos concluir que la probabilidad de que al menos dos personan tenga la misma fecha de cumpleaños es

$$1 - p = 1 - \frac{(365)_n}{365^n}$$

Una curiosidad es que esta probabilidad p cuando tomamos $n = 23$, la probabilidad es casi igual a $\frac{1}{2}$, es decir, en un grupo de 23 personas, la probabilidad de al menos dos tengan el mismo cumpleaños es $\frac{1}{2}$, cosa que parece alucinante, si tomamos $n = 50$, un poco más que el doble de 23, la probabilidad cambia abruscamente a $q = 0,970$, osea es muy probable que al menos dos personas tenga la misma fecha de cumpleaños. Y solo apenas con 50 personas.

Ejemplo 2.17. (Modelo de Urnas) Sea una población de N bolitas, donde $R \leq N$ es la cantidad de bolas rojas y $N - R$ es la cantidad de bolas blancas. Se extraen n bolitas aleatoriamente de la urna. Se desea calcular la probabilidad de que exactamente $k \leq n$ de las bolas, sean de color rojo.

Vamos a suponer que las bolitas están enumeradas del 1 al N de tal forma que las rojas están enumeradas del 1 al R. Hay dos casos importante, sin reemplazo y con remplazo. Sea A_k sacar exactamente $k \leq n$ bolitas rojas.

- **Muestreo ordenado sin reemplazo.** En este caso el espacio muestral es

  $$\Omega = \{(s_1, \dots, s_n) : s_j \in \{1, \dots, N\}, s_i \neq s_j, \forall i \neq j\}$$

  Donde $N(\Omega) = \binom{N}{n}$. A_k consiste en tomar los elementos de Ω, donde k de ellos son extactamente menores o iguales a R. Por tanto

  $$N(A_k) = \binom{R}{k}\binom{N-R}{n-k}$$

  Obteniendose el Modelo Hipergeométrico:

  $$P(A_k) = \frac{N(A_k)}{N(\Omega} \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}}$$

  Es importante ver que en la muestra sin reemplazo, A_k es independiente del orden, ya que solo nos interesa que hayan k rojas.

Un truco para determinar $N(A_k)$ es pensar que de n bolitas, queremos que k sean rojas, pero en la población hay R rojas, por lo que si fijamos $n-k$ bolitas blancas, tenemos que hay

$$\binom{R}{k}$$

formas de tomar esas pelotitas rojas. Ahora, dado que tomamos fijo $n-k$ blancas, podemos contar la cantindad de tomar esas blancas e ir sumando a las rojas, de forma que

$$\binom{R}{k}\binom{N-R}{n-k}$$

La cantidad que buscamos.

- **Muestreo ordenado con reemplazo.** En este caso hay devolución. Luego el espacio muestral, viene dado por

$$\Omega = \{(s_1, \ldots, s_n) : s_j \in \{1, \ldots, N\}, j = 1, \ldots, N\}$$

El evento A_k, consiste en las n tuplas de Ω con exactamente k componentes inferiores o iguales a R. Entonces

$$N(\Omega) = N^n, \text{ y } N(A_k) = \binom{n}{k} R^k (N-R)^{n-k}$$

y en consecuencia, se obtiene el Modelo Binomial.

$$P(A_k) = \frac{N(A_k)}{N(\Omega)} = \binom{n}{k} p^k q^{n-k}$$

donde $p = \frac{R}{N}$ y $q = 1 - p$.

En este caso estamos tomando el orden y para llegar a que

$$N(A_k) = \binom{n}{k} R^k (N-R)^{n-k}$$

Para ello notemos que la n-tupla

$$(a_1, \ldots, a_n)$$

Se pueden tomar R rojos k veces al ser con reemplazo, lo mismo con las blancas, obteniendo $R^K(N-R)^{n-k}$, pero nos falta las combinaciones en que podemos poner cada bola, y esto es tomar n bolas y ponerlas en k espacios sin orden y sin devolución, de forma que aparece la combiantoria $\binom{n}{k}$.

2.5. Probabilidad Geométrica

La probabilidad geométrica se usa en el caso continuo y dada en una distribución uniforme donde se define el espacio muestral.

Definición 2.13. *Sea $(\Omega, \mathcal{A})$ un espacio medible, y supongamos que m es una función donde representa una longitud, área o volumen. Entonces se define la probabilidad geométrica de un evento A de la siguiente forma:*

$$P(A) = \frac{m(A)}{m(\Omega)}$$

En otras palabras comparamos tamaños específicos sobre totales.

Ejemplo 2.18. Encuentra la probabilidad de que un punto elegio al azar, se encuentre en un segmento $\overline{AC}$ de largo 4 de una linea $\overline{AB}$ de largo 15.

Claramente estamos presente sobre un modelo geométrico, donde m es la longitud de un segmento. Entonces la probabilidad del evento A es

$$P(A) = \frac{\text{Longitud de } \overline{AC}}{\text{Longitud } \overline{AB}} = \frac{4}{15} \tag{1}$$

Ejemplo 2.19. María y Carlos acordaron encontrarse en el centro de la ciudad entre las 12 pm y las 1 pm. Ambos pueden llegar en cualquier momento entre ese tiempo. Si los horarios de llegadas son independientes, encuentre:

1. La probabilidad de que Carlos y María se encuentren si ambos esperan por el otro 10 minutos como máximo.
2. La probabilidad de que Carlos y Maria se encuentren si Maria espera 5 minutos pero Carlos espera 20.

Sean X e Y eventos definidos como

$$X = \{\text{Hora de llegada de María}\}$$
$$Y = \{\text{Hora de llegada de Carlos}\}$$

En este caso, el espacio, muestral viene dado por

$$\Omega = \{(x, y) : 0 \leq x \leq 60, 0 \leq y \leq 60\}$$

Para calcula la primera probabilidad debemos pensar que si María llega a x hora, entonces y debe llegar antes o igual o despues generando una diferencia menor o igual a 10 minutos. Es decir

$$T = \{(x, y) : |x - y| \leq 10\}|$$

es nuestro evento. Para determinar la probabilidad debemos graficar en $\mathbb{R}^2$ y luego dividirlo en 60^2.

$$P(T) = \frac{11}{36}$$

Para la segunda probabilidad tenemos que

$$T = \{(x, y) :$$

Termina

2.6. Probabilidad Condicional

Hasta ahora hemos estudiados probabilidades incondicionales. Es decir, de formar separadas sin ninguna restricción. Pero en algunas situaciones es necesario cambia el espacio muestral de forma parcial para determinar probabilidades. Para estudiar entender mejor la probabilidad condicional, empezaremos con un ejemplo simple.

Ejemplo 2.20. Sea $\Omega = \{(i, j) : i, j = 1, \dots, 6\}$, sea $\mathcal{A} = \mathcal{P}(\Omega)$ y que además se sigue un modelo equiprobable. Si

$$A = \{(3, j) : j = 1, \dots, 6\}$$
$$B = \{(i, j) : i + j > 6\}$$

Entonces podemos determinar las distitnas probabilidad, por ejemplo

$$P(A) = \frac{6}{36} = \frac{1}{6}$$
$$P(B) = \frac{21}{36} = \frac{7}{12}$$
$$P(AB) = \frac{3}{36} = \frac{1}{12}$$
$$P(A^C B) = P(B) - P(AB) = \frac{1}{2}$$

¿Qué pasa si sabemos que el evento A ocurre en cualquier circustancia?, esto significa se toma los valores $(3, 1), \dots, (3, 6)$ como son elementos de A, ahora la probabilidad de B, es que ocurrido A, se tiene que la suma de las coordenadas son mayores a 6 y solo existen 3 casos posible, entonces tal probabilidad debe ser

$$\frac{P(AB)}{P(A)} = \frac{3}{6}$$

ya que estamos restrigiendo a A como nuevo espacio muestral, y queremos ver los casos que ocurren B, pero al ocurrir A por hipótesis, necesariamente estamos estudiando la intersección de A, B. En otras palabras, estamos estudiando la probabilidad de B dado que ocurrió A.

Definición 2.14. *Sea* $(\Omega, \mathcal{A}, P)$ *un espacio de probabilidad. Si* $B \in \mathcal{A}$ *es talque* $P(B) > 0$*, entonces se define la probabilidad condicional de* A *dado* B *por:*

$$P(A/B) := \frac{P(AB)}{P(B)}$$

para todo $A \in \mathcal{A}$*.*

Lo que decimos, es que la probabilida de A se reduce en el caso de que ocurra B, es decir, si $A = AB + AB^C$, estamos considerando que ocurrea AB. Si $P(B) = 0$ puede ocurrir dos casos, una donde $P(A/B) = 0$ y la otra $P(A/B) = P(A)$.

Proposición 2.2. *Sea* $(\Omega), \mathcal{A}, P$ *un espacio de probabilidad. Sea* $B \in \mathcal{A}$ *fijo ,talque* $P(B) \neq 0$*, entonces la función*

$$f(A) := P(A/B) = \frac{P(AB)}{P(B)}$$

para todo $A \in \mathcal{A}$*, es una mdidad de probabilidad*

Dem. Debemos probar las tres condiciones.

1. Sea $A \in \mathcal{A}$, luego $P(A/B) = \frac{P(AB)}{P(B)} \geq 0$.
2. Si $A = \Omega$, entonces $P(\Omega/B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1$.
3. Sea $\{A_i\}_{i\in\mathbb{N}}$ una colección de eventos disjuntos. Entonces

$$P\left(\bigcup_{i\geq 1} A_i/B\right) = \frac{P\left(\bigcup_{i\geq 1}(A_iB)\right)}{P(B)} = \frac{\sum_{i\geq 1} P(A_iB)}{P(B)} = \sum_{i\geq 1} P(A_i/B)$$

Probando que f es una medida de probabilidad. ■

Teorema 2.6. *Sea* $(\Omega, \mathcal{A}, P)$ *un espacio de probabilidad, y sea* $A \in \mathcal{A}$ *con* $P(A) > 0$*. Entonces:*

1. *Si* $A \cap B = \emptyset$ *entonces* $P(B/A) = 0$
2. $P(B \cap C/A) = P(B/A \cap C)P(C/A)$ *si* $P(A \cap C) > 0$
3. *Si* $A_1, \dots, A_n \in \mathcal{A}$ *con* $P(A_1 \cap \dots \cap A_{n-1}) > 0$*, entonces se tiene la siguiente regla multiplicativa:*

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2)\dots$$
$$P(A_n/A_1 \cap \dots \cap A_{n-1})$$

Dem.

1. Si $A \cap B = \emptyset$ entonces

$$P(B/A) = \frac{P(AB)}{P(A)} = 0$$

2. Por definición tenemos que

$$P(B \cap C/A) = \frac{P(ABC)}{P(A)} = \frac{P(AC)}{P(AC)}\frac{P(ABC)}{P(A)} = P(B/A \cap C)P(C/A)$$

3. Supongamos que $P(A_1A_2 \ldots A_{n-1}) > 0$ y esto es una condicción justa y necesaria ya que implica que $P(A_1A_2 \ldots A_i) > 0$ para todo $1 \leq i \leq n-1$, ya que si para algún i es cero, entonces

$$A := A_1A_2 \ldots A_{n-1} \subseteq A_1A_2 \ldots A_i =: B_i$$

Luego $P(A) \leq P(B_i) = 0$, es decir $P(A) = 0$, llegando a una contradicción. Vamos a probar por inducción, si

$$P(A_1A_2) = P(A_2/A_1)P(A_1)$$

donde $P(A_1) > 0$. Supongamos que $P(A_1 \ldots A_{n-1}) > 0$ talque

$$P(A_1 \ldots A_n) = P(A_1) \ldots P(A_n/A_1 \ldots A_{n-1})$$

Luego si $P(A_1 \ldots A_n) > 0$ entonces

$$\begin{aligned} P(A_1 \ldots A_nA_{n+1}) &= P(A_1 \ldots A_n) = P(A_1 \ldots A_n)P(A_{n+1}/A_1 \ldots A_n) \\ &= P(A_1)P(A_2/A_1) \ldots P(A_n/A_1 \ldots A_{n-1})P(A_{n+1}/A_1 \ldots A_n) \end{aligned}$$

Probando por inducción.

Probando el teorema. ■

Ejemplo 2.21. Se reparte cuatro cartas desde la parte superior de una baraja bien barajada. ¿Cuál es la probabilidad de que las cuatro cartas sean ases?

Primero pensemos en los casos totales, notemos que el orden no nos interesa y es sin reposición. Luego

$$N(\Omega) = \binom{52}{4}$$

Ahora ¿qué es lo queremos encontrar y qué tenemos? Claramente queremos los cuatro ases, y esto ocurre solo una vez, entonces la probabilidad es:

$$\frac{1}{\binom{52}{4}}$$

Pero también podemos determinarlo usando regla multiplicativo. Sea

$$A_i = \{\text{la i-ésima carta es un as}\}$$

Luego

$$P(A_1A_2A_3A_4) = P(A_1)P(A_2/A_1)P(A_3/A_1A_2)P(A_4/A_1A_2A_3) = \frac{4}{52}\frac{3}{51}\frac{2}{50}\frac{1}{49} = \frac{1}{\binom{52}{4}}$$

Ejemplo 2.22. Calculemos ahora

$$P(\text{4 ases en 4 cartas/i ases en i cartes})$$

donde $i = 1, 2, 3$. Por definición de probabilidad condicional tenemos que

$$\frac{P(\text{4 ases en 4 cartas y i ases en i cartas})}{P(\text{i ases en i cartas})} = \frac{P(\text{4 ases en 4 cartas})}{P(\text{i ases en i cartas})}$$

El numerador ya fue calculado, por lo que debemos calcular el denominador. Sabemos se reparte 4 cartas, de los cuales queremos que i sean ases, eso implica que los que son ases, debemos repartirlo en i y el resto no, es decir

$$P(\text{i ases en i cartas}) = \frac{\binom{4}{i}}{\binom{52}{i}}$$

Luego, nuestra probabilidad es

$$p = \frac{1}{\binom{52-i}{4-i}}$$

La razón de la probabilidad de sacar i ases es i cartas es lo que mostramos, es que estamos condicionando indiréctamente, queremos de i cartas i sean ases, sabiendo que existen 4 ases, entonces nuestro casos totales son la combinatoria de 52 sobre i, y hay $\binom{4}{i}$ de repartir esos i ases.

Ejemplo 2.23. (Urna de Polya) Una caja contiene inicialmente r bolitas rojas y b colitas blancas. En cada prueba se extrae una bolita al azar, se observa el color y se devuelde a la caja junto a otras c bolitas de su mismo color. Se desea calcular la probabilida de que aparezca una bolita roja en cada de las tres primeras pruebas.

Sea A_k : extraer una bolita roja a la prueba k. Entonces

$$\begin{aligned} P(A_1A_2A_3) &= P(A_1)p(A_2/A_1)P(A_3/A_1A_2) \\ &= \frac{r}{r+b}\frac{r+c}{r+b+c}\frac{r+2c}{r+b+2c} \end{aligned}$$

Cuando se resuelven problema de este estilo es importante que leer con cuidado, en el problema se especifica que al ver el color de una bola, a la caja le agregamos c bolitas, además de devolver la que ya teniamos.

Ejemplo 2.24. Tres adolescente quieren entrar a ver una película con clasificación XXX. En la taquilla se les pide que presenten sus identificaciones; después de que el empleado las revisa y les niega la entrada, devuelve las identificiaciones al azar. Encuentre la probabilidad de que ninguno de los adolescente obtenga su propia identificación. Sean $A = \{$ Ninguno de los adolescentes tiene su propia ID$\}$, $B_i = \{$ El i-ésimo adolescente obtiene su propio ID$\}$. Entonces

$$\begin{aligned}
P(A) &= P(B_1^C B_2^C B_3^C) \\
&= 1 - P(B_1 \cup B_2 \cup B_3) \\
&= 1 - P(B_1) - P(B_2) - P(B_3) \\
&+ P(B_1B_2) + P(B_2B_3) + P(B_1B_3) - P(B_1B_2B_3)
\end{aligned}$$

Si $P(B_i) = \frac{1}{3}$ para todo $i = 1, 2, 3$

$$P(B_iB_j) = P(B_i)P(B_j/B_i) = \frac{1}{3} \cdot \frac{1}{2}$$

para $i \neq j$ y

$$P(B_1B_2B_3) = P(B_1)P(B_2/B_1)P(B_3/B_1B_2) = \frac{1}{6}.$$

Por tanto

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

Teorema 2.7. (Teorema de Probabilidad Total) *Sea $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ una partición contable de Ω (también puede ser finita). Tales que $P(A_n) > 0$ para todo $n \in \mathbb{N}$. Entonces para cualquier $B \in \mathcal{A}$, se tiene que:*

$$P(B) = \sum_{n\geq 1} P(B/A_i)P(A_i)$$

Dem. Sea $B \in \mathcal{A}$, por el teorema 2.1 tenemos que

$$P(B) = \sum_{n\geq 1} P(B \cap A_n) = \sum_{n\geq 1} P(B/A_n)P(A_n)$$

Concluyendo el teorema. ■

Este teorema nos dice que la probabilidad de un eventos, se puede determinar por la suma de que ocurra B dado A_i multiplicado por la ocurrencia de A_i. Esto en forma de dibujo tiene mucho sentido.

Corolario 2.1. (Reglas de Bayes) *Sea $\{A_n\}_{n\in\mathbb{N}}$ una partición contable de Ω (también puede ser finita) talque $p(A_n) > 0$ para todo n. Entonces para cualquier $B \in \mathcal{A}$ con $P(B) > 0$ se tiene que*

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_{n\geq 1} P(B/A_n)P(A_n)}$$

para todo i.

Dem. Sea $\{A_n\}$ una partición de Ω de eventos medibles. Sea B un evento medible talque $P(B) > 0$, luego

$$P(A_i/B) = \frac{P(A_iB)}{P(B)} = \frac{P(A_i)P(B/A_i)}{\sum_{ng\geq 1} P(B/A_n)P(A_n)}$$

en virtud del teorema 2.7. ■

Esto es indirectamente tomar casos favorables sobre totales, donde nuestro nuevo total es B, y como vimos

$$P(B) = \sum_{n=1}^{\infty} P(A_n \cap B) = \sum_{n=1}^{\infty} P(B/A_n)P(A_n)$$

Ejemplo 2.25. Suponga que sólo una cuarta parte de la población mundial está vacunada contrar una determinada enfermedad contagiosa. Durante el curso de una epidemia debido a dicha enfermedad, se observa que de cada 5 personas enfermas, sólo una fue vacunada. También se sabe que cada 12 personas vacunadas, sólo uno está enfermo. Deseamos la probabilidad de que una persona no vacunada esté enferma.

Sean los eventos $V = \{\text{Está vacunado}\}$ y $E = \{\text{Está enferma}\}$. Tenemos que $P(V/E) = \frac{1}{5}, P(E/V) = \frac{1}{12}, P(V) = \frac{1}{4}$, luego

$$x := P(E/V^C) = \frac{P(V^C/E)P(E)}{P(V^C)} = \frac{\frac{4}{5}(\frac{1}{12} \cdot \frac{1}{4} + \frac{3}{4}x)}{\frac{3}{4}}$$

Luego debemos despejar x

$$\frac{3x}{4} = \frac{1}{5}\left(\frac{1}{12} + 3x\right) \Leftrightarrow 15x = \frac{1}{3} + 12x \Leftrightarrow x = \frac{1}{9}$$

Encontrando la probabilidad.

Notemos que la probabilibad de que esté enfermo dado que no se vacuno es de forma indirecta, ya que aparece implícita en la fórmula aplicada.

Ejemplo 2.26. Se sabe que la población de una determinada ciudad consta de 45 % mujeres y 55 % hombres. Suponga que 70 % de los hombres y 10 % de las mujeres fuman. Encuentre la probabilidad de que un fumador sea hombre. Sea F el evento de que una persona sea fumadora, H el evento de que una persona sea hombre y M el evento de que una persona sea mujer. Es claro que

$$P(H) = \frac{55}{100}$$
$$P(M) = \frac{45}{100}$$
$$P(F/H) = \frac{70}{100}$$
$$P(F/M) = \frac{10}{100}$$

Queremos determinar $P(F/H)$, entonces

$$P(H/F) = \frac{P(F \cap H)}{P(F)} = \frac{P(F/H)P(H)}{P(F)}$$

si $P(F) = P(F/M) + P(F/H) = \frac{80}{100}$, luego

$$P(H/F) = \frac{55 \cdot 70}{80 \cdot 100} = 0{,}895$$

Ejemplo 2.27. Dunlop Company produce neumáticos que pasan por una máquina de prueba automática. Se observa que 5 % de los neumáticos que ingresan a la máquina de prueba están defectuosos. Sin embargo, la máquina de prueba automática no es completamente confiable. Si un neumático es defectuoso, hay una probabilidad de 0,04 de que no sea rechazado. Si un neumático no esta defectuoso, hay una probabilidad 0,06 de que sea rechazado. ¿Cuál es la probabilidad de que los neumáticos rechazados no esten defectuosos ? Además, ¿qué fracción de los no rechazados son defectuosos ?

Sean D ser defectuoso, R ser rechazado, A ser aceptado. Entonces

$$\begin{aligned} P(D) &= \frac{5}{100} \\ P(A/D) &= \frac{4}{100} \\ P(R/D^C) &= \frac{6}{100} \end{aligned}$$

Y queremos determinar $P(D^C/R)$. Entonces

$$P(D^C/R) = \frac{P(R/D^C)P(D^C)}{P(R)} = \frac{P(R/D^C)P(D^C)}{P(R/D)P(D) + P(R/D^C)P(D^C)}$$

Luego

$$P(D^C/R) = \frac{\frac{6}{100}\frac{95}{100}}{\frac{96}{100}\frac{5}{100} + \frac{6}{100}\frac{95}{100}} = 0,542$$

Lo otro que nos piden es $P(D/A)$, de forma análoga se tiene que

$$P(D/A) = \frac{P(A/D)P(D)}{P(A)} = 0,002$$

Defınción 2.15. *Sea $\{A_n\}_{n\in\mathbb{N}}$ una partición finita o contable de Ω, talque $P(A_n) \geq 0$ para todo n. Si $B \in \mathcal{A}$ con $P(B) > 0$. A la secuencia $\{P(A_n)\}_n$ se le llama distribución a priori, es decir, antes de que B ocurra y a $\{P(A_n/B)\}_n$ se le llama distribución posterior, es decir, despues que B ocurrió.*

Ejemplo 2.28. En una ciudad, se realizan pruebas para detectar una determinada enfermedad. Suponga que 1 % de las personas sanas están registradas como enfermas, 0,1 % de la población está realmente enferma y 90 % de los enfermos se reportan como tales. Deseamos calcular la probabilidad de que una persona elegida al azar que se reporta como enferma esté realmente enferma.

Sean los eventos E estar realmente enferma, R resgistrado como enfermo. Entonces

$$\begin{aligned} P(R/E^C) &= \frac{1}{100} \\ P(E) &= \frac{1}{1000} \\ P(R/E) &= \frac{90}{100} \end{aligned}$$

Queremos determinar

$$P(E/R) = \frac{P(R/E)P(E)}{P(R)}$$

si $P(R) = P(R/E)P(E) + P(R/E^C)P(E^C) = \frac{90}{100}\frac{1}{1000} + \frac{1}{100}\frac{999}{1000}$, luego

$$P(E/R) = \frac{\frac{90}{100}\frac{1}{1000}}{\frac{90}{100}\frac{1}{1000} + \frac{1}{100}\frac{999}{1000}} = \frac{90}{1089}$$

Luego tenemos que la distribución a priori es

$$(P(E), P(E^C)) = (0,001; 0,999)$$

Y la posteriori

$$(P(E/R), P(E^C/R)) = (0,083; 0,917)$$

2.7. Independencia

A veces la ocurrencia de un evento A no afecta en nada a la probabilidad de otro evento B, esto se traduce como

$$P(B/A) = P(A)$$

en este caso decimos que B es independiente del evento A. Aun así pueden ocurrir otros inconvenientes como que $P(A) = 0$. Para evitar este tipo de problema definimos de manera formal la independencia de dos eventos.

Definición 2.16. *Sean A, B eventos. Se dicen que son independientes si y sólo si*

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

En caso contrario, decimos que los eventos son dependiente.

Ejemplo 2.29. Supongamos que se tiran dos veces un dado, consideremos los eventos

$$A = \{\text{La suma de los resultados obtenidos es un par}\}$$
$$B = \{\text{El resultado del segundo lanzamiento es par}\}$$

En este caso

$$P(A) = P(B) = \frac{1}{2} \text{ y } P(A \cap B) = \frac{1}{4} = P(A)P(B)$$

Por tantos son eventos independientes.

Ejemplo 2.30. Supongamos que en una baraja se extraen al azar cartas. Sean los eventos

$$A = \{\text{Sale Rey}\}$$
$$B = \{\text{Sale Corazón}\}$$

Entonces $A \cap B = \{\text{Sale Rey de Corazón}\}$. Luego

$$P(A) = \frac{1}{13} \text{ y } P(B) = \frac{1}{4}$$

Y además $P(A \cap B) = \frac{1}{52} = \frac{1}{13}\frac{1}{4}$. Por tanto son eventos independientes.

Teorema 2.8. *Sean A, B eventos independientes bajo una medida de probabilidad P. Entonces los siguientes pares de eventos son independientes:*

1. A, B^C

2. A^C, B

3. A^C, B^C

Dem.

1. Notemos que

$$P(A \cap B^C) = P(A) - P(A \cap B) = P(A)(1 - P(B)) = P(A)P(B^C)$$

2. Se aplica el mismo método usado en el punto 1.

3. Tenemos que

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

Probando el teorema. ■

A veces es necesario estudiar la independencia de una mayor colección de eventos, de forma que es importante definir la independencia de una familia de eventos.

Definición 2.17. *Una colección de eventos $A_1, \ldots, A_n$ es mutuamente independiente si para cualquier subcolección $A_{i_1}, \ldots, A_{i_k}$ se tiene que*

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j})$$

Observación 2.3. Claramente que una familia sea mutuamente indpendiente, implica que son independientes entre sí.

Nota 2.11. Si $I = \{1, \ldots, n\}$, entonces se verifica que $2^n - (n+1)$ casos para que $A_1, \ldots, A_n$ sean eventos mutuamente independientes.

Para ver esto es solo aplicar inducción. Si $n = 2$ entonces es claro que debemos estudiar el caso $P(A_1A_2)$, donde $2^2 - 2 - 1 = 1$. Si suponemos que se cumple para $n = k$, entonces dado $A_1, \ldots, A_{k+1}$, basta estudiar A_i con A_{k+1}, luego estudiar A_iA_j con A_{k+1} y así con todo los casos posibles. Entonces tendriamos que estudiar en total

$$\begin{aligned}
&2^k - k - 1 + \left(\binom{k}{1} + \cdots + \binom{k}{k}\right) \\
&= 2^k - k - 1 + (2^k - 1) = 2^{k+1} - k - 2 = 2^{k+1} - (k+1) - 1
\end{aligned}$$

Probando por inducción.

De la observación anterior podemos definir la independencia a pares.

Definición 2.18. *Sea $\{A_i\}_{i\in I}$ una familia de eventos. Decimos que son independientes a apres si:*

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

para todo $i \neq j$.

Observación 2.4. El ser independiente a pares, no implica la independencia mutua.

Ejemplo 2.31. Sea Ω el espacio muestral dada por

$$\Omega := \begin{bmatrix} aaa & bbb & ccc \\ abc & bca & bca \\ acb & bac & cab \end{bmatrix}$$

Se define sobre una distribución clásica. Sea el evento

$$A_i := \{\text{el } i\text{-ésimo lugar en la terna está ocupado por } a$$

donde $i = 1, 2, 3$. Notemos que

$$A_1 \cap A_2 = A_1 \cap A_3 = A_2 \cap A_3 = A_1 \cap A_2 \cap A_3 = \{aaa\}$$

Es facil ver que

$$P(A_i \cap A_j) = \frac{1}{9} = P(A_i)P(A_j)$$

Por tanto A_1, A_2, A_3 son independientes a pares. Por otro lado, no son independientes entre sí, ya que

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{9} \neq \frac{1}{27} = P(A_1)P(A_2)P(A_3)$$

3. Variables Aleatorias

3.1. introducción y Definiciones

En la sección anterior sentamos las bases de la probabilidad como un espacio de medida, también revisamos algunos tipos de probabilidad incluyendo la probabilidad clásica, donde podemos aplicar técnicas de conteo, también estudiamos los compotamientos de los eventos, la condicionalidad, ser independiente, disjunto, etc.

Pero en esta sección nos centraremos en el estudio número de los resultados, es decir, dado Ω, definiremos una variable aleatoria como una función que toma un resultado y lo asocia a un número real $\mathbb{R}$ cualquiera. Por ejemplo, sea $\Omega = \{CC, SS, CS, SC\}$ el espacio formado por lanzar dos monedas y estudiar lo que sale, podemos considerar una variable aleatoria X que cuenta la cantindad de caras, de esta forma

$$\begin{aligned} SS &\mapsto X(SS) = 0 \\ CS, SC &\mapsto X(CS) = X(SC) = 1 \\ CC &\mapsto X(CC) = 2 \end{aligned}$$

Defినción 3.1. *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Consideremos la función*

$$\begin{aligned} X : \Omega &\to \mathbb{R} \\ \omega &\mapsto X(\omega) \end{aligned}$$

Decimos que X es una variable aleatoria si para todo $x \in \mathbb{R}$, el conjunto

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A}$$

Nota 3.1. En muchas ocasiones el recorrido de X se denota por $\mathcal{X}$, es decir

$$X : \Omega \to X(\Omega) = \mathcal{X} \subseteq \mathbb{R}$$

Para comprobrar que X es una variable aleatoria, debemos ver que para todo x, el conjunto $\{\omega : X \leq x\}$ es un elemento del σ-álgebra. Esto implica que tiene complemento, y que se puede unir de manera infinita.

Las variables aleatorias sirven para el análisis estadístico. Hay dos clases principales, las variables aleatorias discretas y las continuas, aunque también existen las mixtas.

Observación 3.1. Supongamos que X es una variable aleatorio, sea $x \in \mathbb{R}$ fijo, entonces

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

Es decir, podemos reinterpretar el evento como la pre-imagen de X.

Ejemplo 3.1. Consideremos el experimento lanzar dos veces una moneda justas, tenemos que

$$\Omega = \{(S, S), (C, S), (S, C), (C, C)\}$$

Sea X el número de caras, entonces

$$X(\omega) := \begin{cases} 0 & \omega = (S,S) \\ 1 & \omega = (C,S),(S,C) \\ 2 & \omega = (C,C) \end{cases}$$

Tomando $\mathcal{A} = \mathcal{P}(\Omega)$ vemos que es una variable aleatoria , donde $\mathcal{X} = \{0,1,2\}$. Para ver porque es una variable aleatoria basta probar por definición, notemos lo siguiente

$$X^{-1}((-\infty,x]) := \begin{cases} \emptyset, & x < 0 \\ \{(S,S)\}, & 0 \leq x < 1 \\ \{(C,S),(S,C),(S,S)\}, & 1 \leq x < 2 \\ \Omega, & 2 \leq x < 3 \end{cases}$$

Luego al tomar cualquier $x \in \mathbb{R}$, se ve que es un elemento de $\mathcal{A}$.

Ejemplo 3.2. Sea el experimento observar la altura en centímetros de una persona al azar. Sea X la altura de la persona, entonces

$$\mathcal{X} = (0,\infty)$$

Ejemplo 3.3. Sea el experimento lanzar dos veces un dado, entonces

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

Sea $\mathcal{A} = \mathcal{P}(\Omega)$, sea X la variable aleatoria dada por $X(\omega) = i + j$. Luego

$$\mathcal{X} = \{2,\dots,12\}$$

Claramente es una variable aleatoria ya que para todo $x \in \mathbb{R}$

$$\{\omega \in \Omega : X(\omega) \leq x\} \subseteq \Omega \Leftrightarrow \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A}$$

Ejemplo 3.4. Sea $(\Omega,\mathcal{A},P)$ un espacio de probabilidad con $\Omega = \{a,b,c\}$ y $\mathcal{A} = \{\emptyset,\{a\},\{b,c\},\Omega\}$. Sea $X : \Omega \to \mathbb{R}$ dada por

$$X(\omega) = \begin{cases} 0, & \omega = a \\ 1, & \omega = b,c \end{cases}$$

Entonces X es una variable aleatoria ya que

$$\{\omega \in \Omega : X(\omega) \leq x\} = \begin{cases} \emptyset, & x < 0 \\ \{a\}, & 0 \leq x < 0 \\ \Omega, & x \geq 1 \end{cases}$$

Por otro lado, si consideramos la función $Y : \Omega \to \mathbb{R}$ dada por

$$Y(\omega) = \begin{cases} 0, & \omega = b \\ 1, & \omega = a, c \end{cases}$$

no es variable aleatoria ya que $\{\omega \in \Omega : X(\omega) \leq x < 1\} = \{b\} \notin \mathcal{A}$. Si consideramos ahora $\mathcal{A} = \{\emptyset, \{b\}, \{a, c\}, \Omega\}$, en este caso Y si sería una variable aleatoria. Esto significa como ya dijimos, que la variable depende de donde se miden los eventos.

Convenio. De forma general se tomara por convenio $\mathcal{A} = \mathcal{P}(\Omega)$, en caso de no especificarse.

Notación. Por convención, se usan las letras mayúsculas: $X, Y, Z, \ldots$ para representar variables aleatorias. Y usamos letras minúsculas para los posibles resultados: $x, y, z \ldots$.

Nota 3.2.

1. Si para todo x, $\{\omega : X(\omega) \leq x\} \in \mathcal{A}$, esto significa que es un evento medible y por tanto se le puede asociar a una probabilidad, en tal caso denotaremos por

   $$P(X \leq x)$$

   donde

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

   En general podemos determinar $P(X \in B)$, donde

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

   $B \subseteq \mathbb{R}$. (Borel-medible).

2. Cuando estamos en un espacio de porbabilidad discreto, y si $\mathcal{A} = \mathcal{P}(\Omega)$, entonces para todo $x \in \mathbb{R}$ se tiene que

   $$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A}$$

   Esto es claro ya que por definición este conjunto es subconjunto de Ω.

3. En algunos experimentos cada resultado puede llegar a ser un número, es decir $\Omega \subseteq \mathbb{R}$, de forma que la función identidad $X(\omega) = \omega$ es una variable aleatoria.

Defínción 3.2. *La distribución de probabilidad de una variable aleatoria X definida sobre un modelos de probabilidad $(\Omega, \mathcal{A}, P)$ se define como:*

$$P_X(B) := P(\{X \in B\}) = P(X \in B)$$

para todo $B \in \mathcal{B}$.

Estamos caracterizando la probabilidad bajo una variable aleatoria.

Ejemplo 3.5. Sea el experimento lanzar una moneda tres veces. Sea X el número de caras obtenidas en tres lanzamientos. Entonces

$$\begin{aligned} P_X(\{x\}) &= P(X = x) \\ &= P(X^{-1}(x)) = \frac{i}{8} \end{aligned}$$

si $x = 1, 2$ entonces $i = 3$, si $x = 0, 3$ entonces $i = 1$. También podemos ver que

$$\begin{aligned} P_X(\{0,1\}) &= P_X(X^{-1}(\{0,1\})) \\ &= \frac{4}{8} \end{aligned}$$

Y tomando un intervalo continuo

$$\begin{aligned} P_X((-\infty, 0]) &= P_X(X^{-1}(-\infty, 0]) \\ &= \frac{1}{8} \end{aligned}$$

Nota 3.3. En general, si Ω es discreto, entonces $\mathcal{X}$ es también discreto. Se puede ver por definición

$$X : \Omega \to \mathcal{X} \subseteq \mathbb{R}$$

Teorema 3.1. *Sea X una variable aleatoria definida sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. La distribución de probabilidad de la variable X, definida por*

$$P_X(B) = P(\{X \in B\})$$

para todo $B \in \mathcal{B}$ es una medida de probabilidad sobre $(\mathbb{R}, \mathcal{B})$. Luego $(\mathbb{R}, \mathcal{B}, P_X)$ es un espacio de probabilidad.

Dem. Notemos que P_X está bien definido para todo $B \in \mathcal{B}$ ya que X al ser variable aleatoria podemos ver que

$$\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}$$

Verifiquemos los tres axioma.

1. Sea $B \in \mathcal{B}$, entonces por defineión

$$P_X(B) = P(\{X \in B\}) \geq 0$$

2. Si $B = \mathbb{R}$, entonces

$$P_X(\mathbb{R}) = P(\{\omega \in \Omega : X(\omega) \in \mathbb{R}\}) = P(\Omega) = 1$$

3. Sean $\{B_n\} \subseteq \mathcal{B}$ una colección de conjuntos disjuntos. Entonces

$$\begin{aligned} P_X\left(\bigcup_{n\geq 1} B_n\right) &= P\left(X \in \bigcup_{n\geq 1} B_n\right) \\ &= P\left(\bigcup_{n\geq 1}\{X \in B_n\}\right) \\ &= \sum_{n\geq 1} P(\{X \in B_n\}) \\ &= \sum_{n\geq 1} P_X(B_n) \end{aligned}$$

Probando que P_X es una medida de probabilidad. ■

Con respecto a la demostración, en el tercer punto, en particular, se puede ver que si B_i, B_j son disjuntos, entonces los conjuntos

$$\{\omega : X(\omega) \in B_i\}, \{\omega : X(\omega) \in B_j\}$$

son disjuntos, si existe un ω que está en ambos conjuntos, entonces $X(\omega) \in B_i \cap B_j$, por tanto tal ω no existe. Luego se puede jugar con las uniones y deducir el resultado.

Sea X una variable aleatoria bajo P_X, podemos restringir X en intervalos.

$$\begin{aligned} X \in (a,b) &= a < X < b \\ X \in [a,b) &= a \leq X < b \\ X \in [a,b] &= a \leq X \leq b \\ X \in (a,b] &= a < X \leq b \end{aligned}$$

donde $a < b$.

Basta conocer P_X para una clase de intervalos, (solo basta conocer $a < X < b$o similares y el resto se deduce) Esto se debe a el σ-álgebra $\mathbf{B}$. Es decir, el evento de la forma $\{a < X \leq b\}$ donde $a < b$, determina completamente y de forma única la probabilidad de la forma $\{a < X < b\}$, $\{a \leq X \leq b\}$ y $\{a \leq X < b\}$. Para esto basta probar que:

$$\begin{aligned} \{a < X < b\} &= \bigcup_{n=1}^{\infty}\left\{a < X \leq b - \frac{1}{n}\right\} \\ \{a \leq X \leq b\} &= \bigcap_{n=1}^{\infty}\left\{a - \frac{1}{n} < X \leq b - \frac{1}{n}\right\} \\ \{a \leq X < b\} &= \bigcap_{n=1}^{\infty}\bigcup_{m=1}^{\infty}\left\{a - \frac{1}{n} < X \leq b - \frac{1}{m}\right\} \end{aligned}$$

Probar.

En general, cada conjunto $B \in \mathcal{B}$ se puede generar a partir de uniones e/o intersecciones de secuencias (contable) de intervalos de la forma $(a, b]$.

Otra cosa que podemos ver es que

$$\{a < X \leq b\} = \{X \leq b\} \setminus \{X \leq a\}$$

por propiedades de conjuntos, luego

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a)$$

Es decir, la probabilidad $\{a < X \leq b\}$ se puede determinar de eventos de la forma

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

para todo $x \in \mathbb{R}$.

Defínción 3.3. (Función de Distribución Acumulada) *La función de distribución acumulada o fda de una variable aleataria, denotada por $F_X(x)$, se define por*

$$F_X(x) := P_X((-\infty, x]) = P(X \leq x)$$

para todo $x \in \mathbb{R}$.

Ejemplo 3.6. Considere el experimento lanzar tres monedas justas y sea X el número de caras observadas. Podemos ver que la fda de X es

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

Podemos ver que el comportamiento de F es ser escalonada. Notemos también que podemos determinar la acumulada si $x = 2,3531\ldots < 3$ por la definición misma. También se puede ver que F_X tiene los saltos en $x \in \mathcal{X}$ y el tamaño del salto en x es $P_X(X = x)$.

Teorema 3.2. *Sea X una variable aleatoria definida sobre $(\Omega, \mathcal{A}, P)$. Entonces la fda F_X satisface las siguientes propiedades:*

1. *Si $x \leq y$, entonces $F_X(x) \leq F_X(y)$.*
2. *Cuando $x \to \infty$, entonces $F_X(x) \to 1$ y cuando $x \to -\infty$, entonces $F_X(x) \to 0$.*
3. *Cuando $h \to 0^+$ entonces $F_X(x + h) \to F(x)$ (continuidad por la derecha.)*

Dem.

1. Sea $x \leq y$, tenemos por definición que

$$\{\omega : X(\omega) \leq x\} \subseteq \{\omega : X(\omega) \leq y\}$$

Como son eventos medibles se tiene que

$$P_X(X \leq x) \leq P_X(X \leq y) \Leftrightarrow F_X(x) \leq F_X(y)$$

2. Sea $A_n = \{\omega : X(\omega) \leq n\}$, entonces $\{A_n\}$ es una sucesión creciente de eventos, luego converge, es más

$$\Omega = \bigcup_{n=1}^{\infty}$$

Luego

$$\begin{aligned}
\lim_{x \to \infty} F_X(x) &= \lim_{n \to \infty} F_X(n) \\
&= \lim_{n \to \infty} P_X(A_n) \\
&= P_X\left(\lim_{\infty} A_n\right) \\
&= P_X(\Omega) = 1
\end{aligned}$$

Para probar el otro límite basta considerar $B_n := \{\omega : X(\omega) \leq -n\}$, que forma una sucesión decreciente. Y

$$\bigcap_{n=1}^{\infty} B_n = \emptyset$$

3. Sea $A_h = \{\omega : x < X(\omega) \leq x + h\}$. Luego $A_h \to \emptyset$ cuando $h \to 0^+$, ya que

$$P_X(A_h) = F_X(x + h) - F_X(x)$$

Concluyendo el resultado, luego $F_X(x)$ es continua por la derecha.

Probando el teorema. ■

Nota 3.3. Se puede demostrar que si una función $F : \mathbb{R} \to \mathbb{R}$ que satisface las tres condiciones del teorema 3.2, entonces F es un fda de algún variable aleatoria.

Corolario 3.1. *Sea X una variable aleatoria definida sobre $(\Omega, \mathcal{A}, P)$, sea F_X su fda y sean $a, b \in \mathbb{R}$ tales que $a < b$, entonces*

1. $F_X(x^-) := \lim_{h \to 0^+} F_X(x - h) = P_x(X < x)$
2. $P_X(a \leq X \leq b) = F_x(b) - F_X(a^-)$
3. $P_X(a \leq X < b) = F_X(b^-) - F_X(a^-)$
4. $P_X(X = a) = F_X(a) - F_X(a^-)$

Dem.

1. Notemos que

$$\begin{aligned}
\lim_{h\to 0^+} F_x(x-h) &= \lim_{h\to 0^+} P(\{\omega : X(\omega) \leq x-h\}) \\
&= \lim_{n\to\infty} P\left(\left\{\omega : X(\omega) \leq x - \frac{1}{n}\right\}\right) \\
&= P\left(\bigcup_{n\geq 1}\left\{\omega : X(\omega) \leq x - \frac{1}{n}\right\}\right) \\
&= P_X(X < x)
\end{aligned}$$

2. Por definición tenemos que

$$\begin{aligned}
F_X(b) - F_X(a^-) &= P_X(X \leq b) - P_X(X < a) \\
&= P(\{\omega : X(\omega) \leq b\} \setminus \{\omega : X(\omega) < a\})
\end{aligned}$$

Terminar

Ejemplo 3.7. Sea el experimento lanzar una moneda hasta que aparezca una cara. Supongamos que p es la probabilidad de una cara en cualquier lanzamiento, sea X la variable aleatoria que cuenta los lanzamientos necesarios para obtener una cara. Entonces para cualquier $x = 1, 2, \ldots,$ se tiene que $P(X = x) = (1-p)^{x-1}p$.

Para ver esto notemos que $\Omega = \{1, \ldots\}$ donde $i \in \Omega$ significa que en el lanzamiento i, sale una cara, si $\mathcal{A} = \mathcal{P}(\Omega)$ y tomamos P la distribución clásica, tenemos que

$$\begin{aligned}
P(X = x) &= P(\{\omega : X(\omega) = x\}) \\
&= P(\{x\}) \\
&= P(\underbrace{\text{salga cara en el } x \text{ lanzamiento}}_{A}, y \underbrace{\text{que no haya salido cara}}_{B})\}
\end{aligned}$$

Es claro que son eventos independiente, luego

$$P(X = x) = P(A \cap B) = P(A)P(B) = p(1-p)^{x-1}$$

Sea $x \in \mathbb{R}$, entonces

$$\begin{aligned}
F_X(x) := P_X(X \leq x) &= \sum_{i=1}^{x} P_X(X = i) \\
&= \sum_{i=1}^{x} (1-p)^{i-1} p \\
&= 1 - (1-p)^x
\end{aligned}$$

para todo $x = 1, \ldots,$. Claramente si $p \in (0, 1)$ se ve que F_X es una fda. Y en efecto

1. Si $x \leq y$ naturales, entonces
$$F_X(x) = \sum_{i=1}^{x}(1-p)^{i-1}p \leq \sum_{i=1}^{y}(1-p)^{i-1}p = F_X(y)$$
Por tanto es monótona creciente.

2. Si $x \to \infty$, entonces por la serie geométrica se tiene que
$$p\sum_{n=1}^{\infty}(1-p)^n = \frac{p}{p} = 1$$
Si $x \to -\infty$, entonces para $x < 0$ se tiene que
$$F_X(x) = 0$$
luego es claro que
$$\lim_{x\to-\infty} F_X(x) = 0$$

3. Probemos la continuidad por la derecha. Notemos que
$$F_X(x+h) = \sum_{i=1}^{x+h}(1-p)^{i-1}p$$
cuando $0 < h < 1$, entonces
$$F_X(x+h) = \sum_{i=1}^{x+h}(1-p)^{x+h}p = \sum_{i=1}^{x}(1-p)^{i-1}p = F_X(x)$$
ya que solo consideramos los x naturales. Luego es continua por la derecha.

Por tanto F_X es un fda.

Ejemplo 3.8. Revisemos una fda continua. Digamos que se define fda por
$$F_X(x) := \frac{1}{1+e^{-x}}$$
claramente cumple las tres condiciones, ya que

1. Si $x \leq y$, entonces
$$F_X(x) = \frac{1}{1+e^{-x}} \leq \frac{1}{1+e^{-y}} = F_X(y)$$

2. Si $x \to \infty$, $F_X(x) \to 1$, y si $x \to -\infty$ entonces $F_X(x) \to 0$.

3. Y que por la continuidad se puede ver que

$$\lim_{h\to 0^+} F_X(x) = \frac{1}{1+e^{\left(\lim_{h\to 0^+} -x-h\right)}} = \frac{1}{1+e^{-x}} = F_X(x)$$

Por tanto F_X es un fda. Asociado a un X talque $\mathcal{X} = \mathbb{R} \setminus \{0\}$.

Con el ejemplo 3.8 y 3.7 pudimos ver que existen dos tipos de funciones acumladas, que depende de como es X, si es discreto, claramente tiene un comportamiento escalera, y si es continua, probablemente sea continua.

Defnición 3.4. *Sea X una variable aleatoria. Decimos que X es continua Si $F_X(x)$ es continua y decimos que es discreta si $F_X(x)$ es una función escalera de x.*

La importancia de F_X, es que determina la distribución de la variable aleatoria X, aunque no entrega un comportamiento claro, puede pasar que dos variables aleatorios X, Y tengan igual fda, pero tiene distintas probabilidad para cada evento. Para arreglar esto claramente cada probabilidad debe ser la misma.

Defnición 3.5. *Sean X, Y variables aleatorias. Decimos que tienen la misma distribución si para todo $B \in \mathcal{B}$, se tiene que*

$$P(X \in B) = P(Y \in B)$$

Nota 3.4. Que dos variables tengan la misma distribución, no implica que sean iguales. Puede ser definida de distintas formar y tener la misma distribución.

Ejemplo 3.9. Consideremos el experimento lanzar una moneda justa tres veces. Defina las variables aleatorias X e Y como, X = número de caras observadas e Y = número de sellos observados. La distribución de X está dada en el Ejemplo 1.5 y se puede verificar que la distribución de Y es exactamente la misma. Es decir, para cada $x = 0, 1, 2, 3$, tenemos $P_X(X = x) = P_Y(Y = x)$. Así X e Y son identicamente distribuídas. Sin embargo, para ningún punto en el espacio muestral tenemos que $X(\omega) = Y(\omega)$

Teorema 3.3. *La siguientes dos afirmaciones son equivalentes:*

1. *Las variables aleatorias X e Y son identicamente distribuídas.*
2. *$F_X(x) = F_Y(x)$ para cada x.*

Dem. Supongamos que X e Y son indenticamente distribuídas. Luego por definición tenemos que

$$\begin{aligned} F_X(x) &= P_X(X \leq x) \\ &= P(X \in (-\infty, x]) \\ &= P(Y \in (-\infty, x]) \\ &= P_Y(Y \leq x) = F_Y(x). \end{aligned}$$

Falta la otra dirección.

Dada una variable aleatoria X y su fda F_X, existe otra función llamada funci´pn densidad de probabilidad (fdp) o función de masa de probabilidad (fmp). Si es fdp nos referimos a un caso continuo, mientras que fmp se refiere a un caso discreto.

Defínción 3.6. *La fmp de una variable aleatoria X discreto, está dada por:*

$$f_X(x) := P_X(X = x)$$

para todo $x \in \mathbb{R}$.

Ejemplo 3.10. La distribución geométrica tiene fmp por

$$f_X(x) = P_X(X = x) = \begin{cases} (1-p)^{x-1}p, & \text{para x=1,2,}\dots, \\ 0, & \text{en otro caso (e.o.c)} \end{cases}$$

Con la fmp podemos calcular probabilidad fijas y también si tneemos un intervalo $[a, b]$, entonces, en este caso

$$P_X(a \leq X \leq b) = \sum_{x\in[a,b]} f_X(x) = \sum_{i\in[a,b]\cap\mathbb{N}} (1-p)^{x-1}p$$

En el caso continuo tenemos lo siguiente

$$P_X(X \leq x) = F_x(x) = \int_{-\infty}^{x} f_X(t)dt$$

donde f_X es continua y luego F_X es derivable, en particular

$$\frac{d}{dx}F_X(x) = f_X(x)$$

Defínción 3.7. *La función de densidad o fdp, de una variable aleatoria continua X, es la función que satisface*

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

para todo $x \in \mathbb{R}$.

A partir del teorema fundamental del cálculo podemos ver que

$$\frac{d}{dx}F_X(x) = f_X(x)$$

Claramente que el caso continuo ocurre que $P_X(X = x) = 0$, en efecto, si

$$\{X = x\} \subseteq \{x - \varepsilon < X \leq x\}$$

para cualquier $\varepsilon > 0$, tenemos que

$$P_X(X = x) \leq P_X(x - \varepsilon < X \leq x) = F_X(x) - F_x(x - \varepsilon)$$

luego si $\varepsilon \to 0^+$, entonces se llega al resultado. Por la continuidad derecha de F_X. Otra cosa importante, es que en una variable aleatoria continua tenemos que la fda no varía su valor cuando está en intervalos cerrados, abierto, etc. Al ser una integral.

Ejemplo 3.11. Sea $F_X(x) = \frac{1}{1+e^{-x}}$ que es una fda y como tiene comportamiento continuo, es claro que X debe ser continuo. Luego la función densidad está dada por:

$$f_X(x) = \frac{d}{dx}F_X(x) = \frac{e^{-x}}{(1+e^{-x})^2}$$

Teorema 3.4. *una función $f_X(x)$ es un fdp o fmp de una variable aleatoria X si y sólo si*

1. $f_X(x) \geq 0$ *para todo* $x \in \mathbb{R}$
2. $\sum_{x\in\mathbb{R}} f_X(x) = 1$ *(fmp) o* $\int_{-\infty}^{\infty} f_X(t)dt = 1$ *(fdp)*

Dem. Probaremos el caso continuo y discreto.

- **Discreto.** Sea X una variable aleatoria discreta, sea $f_X(x)$ la fmp. Notemos que por definición

$$f_X(x) = P_X(X = x) = P(\{\omega : X(\omega) = x\}) \geq 0$$

Y que si F_X es el fda asociado a X y a $f_X(x)$, entonces

$$\begin{aligned} 1 = \lim_{x\to\infty} F_X(x) &= \lim_{x\to\infty} P_X(X \leq x) \\ &= \lim_{x\to\infty} \sum_{y\leq x} P_X(X = y) \\ &= \sum_{y<\infty} P_X(X = y) = \sum_{x\in\mathbb{R}} f_X(x) \end{aligned}$$

Notemos que podemos separar $P(X \leq x)$ en una suma ya que estamos en un caso discreto, luego $X^{-1}(-\infty, x]) \subseteq \Omega$ es discreto.

Supongamos ahora que $f_X(x)$ es una función talque $f_X(x) \geq 0$ y $\sum f_X(x) = 1$. Debemos probar que de alguna forma es un fmp, es decir, $f_X(x) = P_X(X = x)$. Sea

$$F_X(x) := \sum_{z\leq x} f_X(z)$$

Probemos que es un fda. Sean $x, y \in \mathbb{R}$ tales que $x \leq y$, luego si $f_X(x) \geq 0$ para todo $x \in \mathbb{R}$, entonces

$$F_X(x) = \sum_{z\leq x} f_X(z) \leq \sum_{z\leq y} f_X(z) = F_X(y)$$

Si $x \to \infty$, es claro que

$$\lim_{x\to\infty} \sum_{z\leq x} f_X(x) = \sum_{x\in\mathbb{R}} f_X(x) = 1$$

y si $x \to -\infty$....

Nos falta ver la continuidad por la derecha. Notemos que

$$F_X(x+h) - F_X(x) = \sum_{x<z\leq x+h} f_X(z)$$

Si $h \to 0$, entonces

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

ya que no existe un z talque $x < z \leq x$. Luego

$$\lim_{h\to 0} F_X(x+h) = F_X(x)$$

De forma que F_X es un fda de X y luego

$$P_X(X=x) = F_X(x) - F_X(x^-) = \sum_{z\leq x} f_X(x) - \sum_{z<x} f_X(x) = f_X(x)$$

Luego $f_X(x)$ es un fmp.

- **Continuo.** Es la misma idea que el discreto. Si X es una variable aleatoria continua y $f_X(x)$ es el fdp, sea F_X el fda, entocne si F_X es creciente, necesariamente $f_X \geq 0$ ya que si $f_X < 0$, por definición de derivada, F_X sería decreciente estríctamente. Una contradicción.

  Además

  $$\begin{aligned} 1 = \lim_{x\to\infty} F_X(x) &= \lim_{x\to\infty} \int_{-\infty}^{x} f_X(t)dt \\ &= \int_{-\infty}^{\infty} f_X(t)dt \end{aligned}$$

  Supongamos ahora que $f_X(x)$ es una función talque $f_X(x) \geq 0$ para todo $x \in \mathbb{R}$ y

  $$\int_{-\infty}^{\infty} f_X(t)dt = 1$$

  Sea

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

  ($f_X(x)$ está bien definida para ser integrable). Probemos que es fda de X y luego por definición f_X es el fdp. Sean $x, y \in \mathbb{R}$ tales que $x \leq y$, luego

  $$\int_{-\infty}^{y} f_X(t)dt - \int_{-\infty}^{x} f_X(t)dt = \int_{x}^{y} f_X(t)dt \geq 0$$

  Luego F_X es creciente.

  Si $x \to \infty$, claramente converge a 1. Si $x \to -\infty$**Terminar**

  Para ver la continuidad por la derecha, basta ver que al ser definida como una integral definida, está es continua. Por lo tanto, también es continua por la derecha, luego F_X es un fda y por lo tanto f_X es fdp.

■

Ejemplo 3.12. Sea un fdp de X, de la forma

$$f_X(x) = \begin{cases} \frac{x}{8}, & 0 < x < 4 \\ 0, & \text{e.o.c} \end{cases}$$

como estamos en un fdp, X debe ser continua y por tanto la fda se determina por

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 0, & x \leq 0 \\ \frac{x^2}{16} & 0 < x < 4 \\ 1, & 4 \leq x \end{cases}$$

Luego

$$P_X(1 \leq X \leq 2) = F_X(2) - F_X(1^-) = \frac{4}{16} - \frac{1}{16} = \frac{3}{16}$$

Y

$$P_X(X > 2) = \left(\lim_{x\to\infty} F_X(x)\right) - F_X(2) = 1 - \frac{1}{4} = \frac{3}{4}$$

Ejemplo 3.13. Una urna contiene dos bolas blancas y tres rojas. Se saca una muestra de tres bolas simultáneamente. Encontrar la fmp y fda de X donde X es el número de bolas rojas en la muestra.

Claramente el número de bolas rojas es 1 o 2 o 3. Entonces el fmp es

$$f_X(x) = P_X(X = x) = \begin{cases} \frac{\binom{3}{x}\binom{2}{3-x}}{\binom{5}{3}}, & x = 1, 2, 3 \\ 0, & eoc \end{cases}$$

Para obtener esto basta ver que estamos en una espacios discreto con muestras sin orden y sin reposición. Luego

$$F_X(x) = \begin{cases} 0, & x < 1 \\ 3/10, & 1 \leq x < 2 \\ 9/10, & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

Ejemplo 3.14. De un lote que contiene 25 artículos, 5 de los cuales son defectuosos, se eligen 4 al azar. Sea X el número de artículos defectuoso encontrados. Si los artículos se escogen con reemplazo entonces

$$f_X(x) = P_X(X = x) = \begin{cases} \binom{4}{x}\left(\frac{1}{5}\right)^x\left(\frac{4}{5}\right)^{4-x}, & x = 0, 1, 2, 3, 4 \\ 0, & eoc \end{cases}$$

Estamos ante un muestro ordenado con reemplazo. Luego el fda está dada por:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 0{,}4096, & 0 \leq x < 1 \\ 0{,}8192, & 1 \leq x < 2 \\ 0{,}9728, & 2 \leq x < 3 \\ 0{,}9984, & 3 \leq x < 4 \\ 1, & x \geq 4 \end{cases}$$

Por otro lado, si sería con reemplazo, tendriamos que el fmp es:

$$f_x(x) = P_X(X = x) = \begin{cases} \frac{\binom{5}{x}\binom{20}{4-x}}{\binom{25}{5}}, & x = 0, 1, 2, 3, 4 \\ 0, & eoc \end{cases}$$

Ejemplo 3.15. Supongamo que una variable aleatoria X tiene fdp dada por:

$$f_X(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & eoc \end{cases}$$

Calculemos $P_X(B|A)$ donde $A = (0, 1/2], B = [1/3, 1/2]$, por definición es la probabilidad de que ocurra A dado que ocurrio B, es decir

$$P_X(B|A) = \frac{P_X([1/3, 1/2])}{P_X((0, 1/2])}$$

Donde

$$\begin{aligned} P_X([1/3, 1/2]) = F_X(1/2) - F_X(1/3) &= \int_{-\infty}^{1/2} 2x dx - \int_{-\infty}^{1/3} 2x dx \\ &= x^2 \Big|_{1/3}^{1/2} = \frac{1}{4} - \frac{1}{9} = \frac{5}{36} \end{aligned}$$

Por otro lado

$$P_X((0, 1/2]) = \frac{1}{4}$$

Luego

$$P_X(B|A) = \frac{5}{9}$$

3.2. Funciones de una Variable Aleatoria

3.2.1. Introducción

En situaciones cotidianas nos interesa estudiar situaciones que dependen de otra. Por ejempo, si X es la tasa a la que se atiende a los clientes en una cola, entonces $1/X$ es el tiempo de espera promedio.

Para ello queremos construir una variable aleatoria Y que depende de X y que esten asociada mendiante una función g bien definida. Recordemos que una variable aleatoria es una función

$$Y : \Omega \to \mathbb{R}$$

donde

$$\{\omega \in \Omega : Y(\omega) \leq y\} \in \mathcal{A}$$

para todo $y \in \mathbb{R}$, entonces de alguna forma si queremos pasar de X , luego a Y tomando g, necesariamente la función g debe ser de la forma

$$g : \mathcal{X} \to \mathcal{Y}$$

donde $\mathcal{X}, \mathcal{Y}$ son los recorridos. Ambos subconjuntos de los reales. De forma que cambia todo el estudio de probabilidad a un cambio mediante una función g. Cuando esto ocurre denotamos $Y = g(X)$. Entonces, tenemos la siguiente relación

$$\Omega \xrightarrow{X} \mathcal{X} \xrightarrow{g} g(\mathcal{X}) =: \mathcal{Y}$$

Luego basta considerar

$$g(X) = Y : \Omega \longrightarrow \mathcal{Y} \subseteq \mathbb{R}$$

Construyendo una nueva variable aleatoria.

Por ejemplo, si $\mathcal{X} = \{-2, -1, 0, 1, 2\}$ y me interesa estudiar la probabilidad con valores positivos, podemos tomar $Y = X^2$ de forma que

$$\mathcal{Y} = \{0, 1, 4\}$$

donde claramente $g(x) := x^2$.

3.2.2. Y como variable aleatoria

Antes de continuar con algunos ejemplos es importante entender como se ven las cosas al aplicar g a una variable X, como si de verdad Y es una variable aleatoria o como calcular las probabilidad.

- **Variable aleatoria.** Sea $Y = g(X)$, sabemos que por definición

$$Y : \Omega \to \mathbb{R}$$

nos falta ver que podemos formar eventos de $\mathcal{A}$. Sea $y \in \mathbb{R}$ fijo, entonces

$$\{\omega \in \Omega : Y(\omega) \leq y\} = \{\omega \in \Omega : g(X(\omega)) \leq y\}$$

todo se reduce a estudiar las preimagenes. En particular, notemos que

$$\{\omega : g(X)(\omega) \leq y\} = \bigcup_{x \in \mathcal{X}: g(x) \leq y} X^{-1}(\{x\})$$

Ya que queremos todos los ω que forman $X(\omega) = x$, tales que $g(x) = y$. Probemos la igualdad, si $\omega \in \Omega$ es talque $g(X)(\omega) \leq y$, entonces $X(\omega) = x$ talque $g(x) =\leq y$, para algún $x \in \mathcal{X}$, luego $\omega \in X^{-1}(\{x\})$. Y si $\omega \in X^{-1}(\{x\})$ para algún $x \in \mathcal{X}$ talque $g(x) \leq y$, es claro que $X(\omega) = x$ y entonces $g(X(\omega)) \leq y$. Probando la igualdad

De esto podemos concluir que

$$\{\omega \in \Omega : Y(\omega) \leq y\}$$

no es más que una unión de $X^{-1}(\{x\}) = \{\omega : X(\omega) = x\} \in \mathcal{A}$, por lo tanto

$$\{\omega \in \Omega : Y(\omega) \leq y\} \in \mathcal{A}$$

Por otro lado, notemos que es una unión disjunta, ya que si $\omega \in X^{-1}(x), X^{-1}(z)$, donde $z \neq x$ y donde $g(x), g(z) \leq y$, entonces $X(\omega) = x = z$. Una contradicción. Por lo tanto, sobre una medida de probabilidad P, se tiene que

$$\begin{aligned}
P_Y(Y \leq y) &= P_Y(g(X) \leq y) \\
&= P\left(\bigcup_{x \in \mathcal{X}: g(x) \leq y} X^{-1}(x)\right) \\
&= \sum_{x \in \mathcal{X}: g(x) \leq y} P(X^{-1}(x)) \\
&= \sum_{x \in \mathcal{X}: g(x) \leq y} P_X(X = x)
\end{aligned}$$

Tomando $\{x \in \mathcal{X} : g(x \leq y\}) = g^{-1}((-\infty, y])$, llegamos a una forma muy conveniente

$$P_Y(Y \leq y) = \sum_{x \in g^{-1}((-\infty, y])} P_X(X = x)$$

Esto sirve tanto para Y continua como discreta, aunque es más efectivo estudiarlo para discreta ya que en Y continua, es casi imposible estudiar la suma, por lo que se asocia directamente a integrales.

En resumidad, Y es una variable aleatoria bien definida por g, podemos definir sin problemas:

$$F_Y(y) := P_Y(Y \leq y), \;\; f_Y(y) = P_Y(Y = y)$$

como una medida y por tanto, se puede definir el espacio de probabilidad $(\mathbb{R}, \mathcal{B}, P_Y)$

- **La variable aleatoria Y como Discreta.** Sea X una variable aleatoria discreta, luego $\mathcal{X}$ es discreto. Sea $Y = g(X)$ una variable discreta a paritr de g, entonces claramente Y es discreta. Es decir, el conjunto

$$\mathcal{Y} = \{y : \exists x \in X, g(x) = y\}$$

es discreto. Como X es discreto podemos formar una función $f_X(x)$ fmp. ¿Cómo sera el fmp de Y?, para ello realizaremos el mismo truco de caracterización del conjunto de los ω que al ser evaluado en Y, dan y. Entonces, por definición, dado $y \in \mathbb{R}$ se tiene que

$$\begin{aligned} f_Y(y) &= P_Y(Y = y) = P(\{\omega : Y(\omega) = y\}) \\ &= P\left(\bigcup_{x\in\mathcal{X}:g(x)=y} X^{-1}(\{x\})\right) \end{aligned}$$

Notemos que los $x \in \mathcal{X} : g(x) = y$ son los x que son preimagenes de $\{y\}$, es decir, $x \in g^{-1}(\{y\})$ y que el conjunto $X^{-1}(\{x\})$ es disjunto para todo $x \in g^{-1}(\{y\})$, por tanto podemos separarlo en una suma y por lo tanto

$$\begin{aligned} f_Y(y) = P\left(\bigcup_{x\in\mathcal{X}:g(x)=y} X^{-1}(\{x\})\right) &= \sum_{x\in g^{-1}(\{y\})} P(X^{-1}(\{x\})) \\ = \sum_{x\in g^{-1}(\{y\})} P_X(X = x) &= \sum_{x\in g^{-1}(\{y\})} f_X(x) \end{aligned}$$

Caracterízando la función fmp de Y. Es decir, el fmp de Y se determina por

$$f_Y(y) = \sum_{x\in g^{-1}(y)} f_X(x)$$

para todo $y \in \mathcal{Y}$. (si tomamos $y \notin \mathcal{Y}$, entonces es claro que no puede ser determinado por g y por tanto $f_Y(y) = 0$.) Por convenio denotamos $g^{-1}(\{y\}) = g^{-1}(y)$ (lo mismo para $X^{-1}(x)$)

De lo anterior se puede deducir el fda de Y.

$$F_Y(y) = \sum_{y}\left(\sum_{x\in g^{-1}(y)} f_X(x)\right)$$

- **La variable aleatoria Y como continua.** Cuando X es una variable aleatoria continua, no necesariamente Y es continua, ya que podemos usar $g : \mathcal{X} \to \{0\}$, por ello supondremos que g es una función bien definida, talque Y es una variable aleatoria continua. El fda de

Y se determina de la siguiente forma

$$\begin{aligned}
F_Y(y) &= P_Y(Y \leq y) \\
&= P_X(g(X) \leq y) \\
&= P_X(\{x \in \mathcal{X} : g(x) \leq y\}) \\
&= \int_{x \in \mathcal{X}: g(x) \leq y} f_X(x)dx
\end{aligned}$$

Viendo que el fda de Y es bastante complejo de determinar. Ya que consiste en determinar la integral en un área variada.

Solo nos falta determinar el fdp de Y, pero como habrá intuido, es bastante más complejo de determinar. Por lo que por el momento no lo estudiaremos, más adelante se tomarán condiciones y se probaran, que se puede determinar el fdp de Y bajo esas condiciones.

Ejemplo 3.16. Sea X una variable aleatoria discreta con fmp dada por al tabla Sea $Y = X^2$

x	-1	0	1	2	3
$P_X(x)$	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{1}{7}$

una nueva variable aleatoria. Claramente $y = g(x) = x^2$, si $\mathcal{X} = \{-1, 0, 1, 2, 3\}$, entonces $\mathcal{Y} = \{0, 1, 4, 9\}$, luego la distribución de probabilidad de Y está dada por la tabla Así por

y	0	1	4	9
$P_Y(y)$	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{1}{7}$

ejemplo

$$P_Y(Y = 1) = P_Y(X^2 = 1) = P_X(X = 1) + P_X(X = -1) = \frac{2}{7}$$

3.2.3. Función Acumulada de Y continua

Ejemplo 3.17. Sea X una variable aleatoria con distribución exponencial, donde fdp está dada por:

$$f_X(x) = \exp(-x)$$

para todo $0 < x < \infty$. Sea $Y := 2X + 1$, entonces el fda de Y está dada por

$$\begin{aligned}
F_Y(y) &= P_Y(Y \leq y) \\
&= P_X(2X + 1 \leq y) \\
&= P(\{\omega : 2X(\omega) + 1 \leq y\}) \\
&= P(\{\omega : X(\omega) \leq (y - 1)/2\}) \\
&= P_X(X \leq (y - 1)/2) \\
&= F_X(x)
\end{aligned}$$

donde $x = g^{-1}(y) = (y-1)/2$, entonces

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{2}\exp(-(y-1)/2)$$

Para determinar la derivada basta usar regla de la cadena, en particular

$$\frac{d}{dy}F_Y(y) = (G\left(\frac{y-1}{2}\right))' = G'\left(\frac{y-1}{2}\right)\frac{1}{2}$$

Donde

$$G(x) := \int_{-\infty}^{x} e^{-t}dt$$

Esto lo pudimos hacer ya que la función $g(x) = 2x+1$ es muy buena, es biyectiva, creciente, continua, entre otros. Veremos que el fda depende mucho de la función g, ya que puede pasar que g biyectiva, que si $g(X) \leq y$, entonces $X \geq g^{-1}(y)$, o que g no sea biyectiva y aun así pueda encontrarse el fda de Y.

Nota 3.5. Cuando hacemos transformaciones de X a $Y := g(X)$, lo más conveniente es usar:

$$\mathcal{X}' = \{x \in \mathcal{X} : f_X(x) > 0\}$$
$$\mathcal{Y}' = \{y \in \mathcal{Y} : y = g(x), \text{ para algún } x \in \mathcal{X}'\}$$

hacemos esto ya que los casos cuando $f_X(x) = 0$ claramente se terminan ignorando al hacer los cálculos.

Teorema 3.5. *Sea X una variable aleatoria continua con fdp $f_X(x)$. Sea $Y := g(X)$ donde $g : \mathcal{X}' \to \mathcal{Y}'$ es una función monótona, con $\mathcal{X}', \mathcal{Y}'$ como hemos definido. Supongamos que $f_X(x)$ es continua sobre $\mathcal{X}'$ y que $g^{-1}(y)$ tiene derivada continua sobre $\mathcal{Y}'$. Entonces la fdp de Y está dada por:*

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|, & \textit{para todo } y \in \mathcal{Y}' \\ 0, & \textit{eoc} \end{cases}$$

Dem. Notemos algunas cosas: g es una función biyectiva ya que $im(g) = \mathcal{Y}'$ por definición, y es monótona. Luego g^{-1} es monótona. $g^{-1}(y)$ tiene derivada continua en $\mathcal{Y}'$ por lo que podemos derivar sin problema. Y por último, la continuidad de $f_X(x)$ en $\mathcal{X}'$ donde $\mathcal{X}'$ continua, implica integrabilidad.

Supongamos que g es creciente. Luego para $y \in \mathcal{Y}'$

$$\begin{aligned} F_Y(y) &= P_Y(Y \leq y) \\ &= P_Y(g(X) \leq y) \\ &= P(\{\omega \in \Omega : g(X(\omega)) \leq y\}) \\ &= P(\{\omega \in \Omega : X(\omega) \leq g^{-1}(y)\} \\ &= P_X(X \leq g^{-1}(y)) \\ &= \int_{-\infty}^{g^{-1}(y)} f_X(t)dt \end{aligned}$$

Sea

$$G(x) := \int_{-\infty}^{x} f_X(t)dt$$

Luego

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}G(g^{-1}(y)) = G'(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|$$

como la derivada de g es positiva. Luego

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|, & \text{para todo } y \in \mathcal{Y}' \\ 0, & eoc. \end{cases}$$

Si g fuera decreciente, entonces

$$\begin{aligned} F_Y(y) &= P_Y(Y \leq y) \\ &= P_Y(g(X) \leq y) \\ &= P(\{\omega \in \Omega : g(X(\omega)) \leq y\}) \\ &= P(\{\omega \in \Omega : X(\omega) \geq g^{-1}(y)\} \\ &= P_X(X \geq g^{-1}(y)) \\ &= 1 - P_X(X \leq g^{-1}(y)) \\ &= 1 - \int_{-\infty}^{g^{-1}(y)} f_X(t)dt \end{aligned}$$

Luego derivando se llega a que

$$f_Y(y) = \frac{d}{dy}F_Y(y) = -\frac{d}{dy}G(g^{-1}(y)) = G'(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|$$

Concluyendo el teorema. ■

Podemos escribir el teorema de una forma más conveniente, recordemos que si $g(x) = y$ talque $g'(x) \neq 0$, entonces

$$\frac{d}{dy}g^{-1}(y) = \frac{1}{g'(x)}$$

Luego tomando las mismas condiciones del teorema 3.5, se tiene que

$$f_Y(y) = \begin{cases} \left|\frac{f_X(g^{-1}(y))}{g'(x)}\right|, & g'(x) \neq 0, \text{para todo } g(x) = y \in \mathcal{Y}' \\ 0, & eoc. \end{cases}$$

Ejemplo 3.18. Sea X una varaible aleatoria contina con fdp

$$f_X(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & eoc \end{cases}$$

Sea $Y := \exp(-X)$. Queremos determinar $f_Y(y)$. Antes de ello notemos que

$$g(x) = e^{-x} : (0,1) \to \left(\frac{1}{e}, 1\right)$$

que es una función, decreciente y monótona, ademas de biyectiva. Y $g^{-1}(x) = -\log(x)$, que es continua y decreciente. Luego se puede aplicar el teorema, además $g'(x) = -e^{-x}$ que no se anula en $(0,1)$, por lo que podemos derivar sin problemas, entonces

$$f_Y(y) = \begin{cases} \left|\frac{2(-log(y))}{-y}\right| = -2\frac{\log(y)}{y}, & y \in \left(\frac{1}{e}, 1\right) \\ 0, & eoc. \end{cases}$$

donde $g'(x) = -e^{-x} = -y$ y si $y \in \left(\frac{1}{e}, 1\right)$, entonces $x \in (0,1)$, luego $g^{-1}(y) \in (0,1)$ y se aplica la definición de $f_X(x)$.

Ejemplo 3.19. Sea X una variable aleatoria continua con f_X continua, en todo $\mathbb{R} \setminus \{0\} = \mathcal{X}'$. Sea $Y = X^2$, entonces $\mathcal{Y}' = (0, \infty)$. Notemos que g no es monótona, por lo que debemos aplicar otro truco para determinar el fda y el fdp. Por definición dado $y \in \mathcal{Y}'$ se tiene que

$$\begin{aligned} F_Y(y) &= P_Y(Y \leq y) \\ &= P(\{\omega : X^2(\omega) \leq y\}) \\ &= P(\{\omega : -\sqrt{y} \leq X(\omega) \leq \sqrt{y}\}) \\ &= P_X(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}^-) \\ &= \int_{-\infty}^{\sqrt{y}} f_X(x)dx - \int_{-\infty}^{-\sqrt{y}} f_X(x)dx \end{aligned}$$

Luego, podemos derivar sobre $\sqrt{y}$ ya que $y \in (0, \infty)$, luego

$$f_Y(y) = \frac{d}{dy}F_y(y) = \begin{cases} \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y})), & y \in (0,\infty) \\ 0, & eoc. \end{cases}$$

Hemos encontrado el fdp y el fda de una variable aleatoria Y generada por una función g que no era biyectiva ni monótona. Por lo que se puede determinar los f_Y por definición y aplicando algunos trucos. Lo que si se puede ver es que g se puede construir como por dos funciones que si son biyectica, monótonas que generan inversas monótonas continuas. Es decir, podemos separar una función g, de forma que es compuesta por g_i monótonas biyectivas, con derivadas continuas, y sumarlas entre si.

Corolario 3.2. *Suponga que X tiene un fdp $f_X(x)$, sea $Y := g(X)$ y sea el conjunto $\mathcal{X}'$ como se definió en el teorema anterior. Sea $\{A_i\}_{i=0}^k$ una partición de $\mathcal{X}'$ talque $P(X \in A_0) = 0$ y $f_X(x)$ continua en todo A_i. sean $g_1(x), \dots, g_k(x)$ funciones definidas sobre $A_1, \dots, A_k$ respectívamente, tales que*

1. *$g(x) = g_i(x)$ para cada A_i.*
2. *$g_i(x)$ es monótona sobre A_i.*
3. *El conjunto $\mathcal{Y} = \{y : y = g_i(x)$ para algún $x \in A_i\}$ es el mismo para cada $i = 1, \dots, k$ y*
4. *$g_i^{-1}(y)$ tiene derivada continua sobre $\mathcal{Y}$, para cada $i = 1, \dots, k$*

Entonces

$$f_Y(y) = \begin{cases} \sum_{i=1}^k f_X(g^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right|, & y \in \mathcal{Y} \\ 0, & eoc \end{cases}$$

Ejemplos 3.20. Sea X una varaible aleatoria, con fdp dada por

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

donde $x \in \mathbb{R}$. Notemos que f_X es continua en todo $\mathbb{R}$. Sea $Y := X^2$ la variable aleatoria dada por $g(x) := x^2$, que es mónotona en $(-\infty, 0)$ y en $(0, \infty)$. Sea la partición A_0, A_1, A_2 dada por

$$\begin{aligned} &A_0 = \{0\} \\ &A_1 = (-\infty, 0), \;\; g_1(x) = x^2, \;\; g_1^{-1}(y) = -\sqrt{y} \\ &A_2 = (0, \infty), \quad\; g_2(x) = x^2, \;\; g_2^{-1}(y) = \sqrt{y} \end{aligned}$$

Luego por el corolario 3.1 tenemos que

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}} e^{-(-\sqrt{y})^2/2} \left| -\frac{1}{2\sqrt{y}} \right| + \frac{1}{\sqrt{2\pi}} e^{-(\sqrt{y})^2/2} \left| \frac{1}{2\sqrt{y}} \right| \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-y/2} \end{aligned}$$

para todo $y \in (0, \infty)$.

En virtud de que podemos formar variables aleatorias sobre funciones g, podemos tomar cualquier variables aleatorio X continuo con $F_X(x)$ continuo, se puede formar una variable aleatorio Y con uan distribución uniforme.

Teorema 3.6. (Transformación integral de probabilidad) *Suponga que la fda de X, $F_X(x)$ es continua y defina la variable aleatoria como $Y = F_X(X)$. Entonces Y tiene una distribución uniforme sobre $(0,1)$, es decir, $P_Y(Y \leq y) = y$ para todo $0 < y < 1$.*

Dem. Sea $F_X : \mathbb{R} = \mathcal{X}' \to (0,1) = \mathcal{Y}'$ la acumulada, suponiendo que no alcanza ni 0 ni a 1. Sabemos que es creciente, es continua , acotada y que

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

Por lo que es sobreyectiva. Entonces existe la inversa $F_X^{-1} : (0,1) \to \mathbb{R}$ continua y creciente. Sea $y \in (0,1)$, entonces

$$\begin{aligned} P_Y(Y \leq y) &= P_Y(F_X(X) \leq y) \\ &= P_X(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) = y \end{aligned}$$

Aplicando el crecimiento de la inversa y definición. Observando que la nueva variable tiene un comportamiento unforme en $(0,1)$. $\blacksquare$

Por tanto, podemos hacer que X pase a una variable Y con distribución uniforme. Además se puede concluir que la fdp es

$$f_Y(y) = \begin{cases} 1, & 0 < y < 1 \\ 0, & eoc. \end{cases}$$

3.3. Momentos centrados y No centrados

3.3.1. Esperanza

Definición 3.8. *Sea X una variable aleatoria definida sobre el espacio de probabilidad $(\Omega, \mathcal{A}, P)$. El valor esperado o medida de variable aleatoria $g(X)$, se denota por $Eg(X)$ y se determina por*

$$Eg(X) := \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x) dx, & X \text{ continua} \\ \sum_{x\in\mathcal{X}} g(x) f_X(x) = \sum_{x\in\mathcal{X}} g(x) P_X(X = x), & X \text{ discreta} \end{cases}$$

Siempre y cuando existan tales sumas e integrales. Si $E|g(X)| = \infty$, no existe $Eg(X)$. En particula, si $g(X) = X$, entonces

$$E(X) := \begin{cases} \int_{-\infty}^{\infty} x f_X(x) dx, & X \text{ continua} \\ \sum_{x\in\mathcal{X}} x f_X(x) = \sum_{x\in\mathcal{X}} x P_X(X = x), & X \text{ discreta} \end{cases}$$

Ejemplo 3.22. Sea X una varaible aleatoria continua con fdp dada por

$$f_X(x) := \begin{cases} 2x, & 0 < x < 1 \\ 0, & eoc \end{cases}$$

Entonces

$$\begin{aligned} E(X^3) &= \int_{-\infty}^{\infty} x3 f_X(x)dx \\ &= \int_0^1 2x^4 dx \\ &= \frac{2}{5} \end{aligned}$$

Ejemplo 3.23. Sea X una variable aleatoria discreta con fmp dada por:

$$p_x(x) := \begin{cases} e^{-3}\frac{3^x}{x!}, & x = 0, 1, 2 \dots \\ 0, & eoc \end{cases}$$

Luego

$$\begin{aligned} E(X) &= \sum_{x=1}^{\infty} x e^{-3}\frac{3^x}{x!} = e^{-3}\sum -x = 1^{\infty}\frac{3^x}{(x-1)!} \\ &= e^{-3}\sum_{j=0}^{\infty}\frac{3^{j+1}}{j!} = 3e^{-3}\sum_{j=0}^{\infty}\frac{3^j}{j!} \\ &= 3e^{-3}e^3 = 3 \end{aligned}$$

Teorema 3.7. *Sea X una variable aleatoria y a, b y c constantes. Entonces para cualquier par de funciones g_1, g_2 cuya esperanza exista, se tiene que:*

1. $E(ag_1(X) + bg_2(X) + c) = aEg_1(X) + bEg_2(X) + c$
2. *Si $g_1(x) \geq 0$ para todo x, entonces $Eg_1(X) \geq 0$.*
3. *Si $g_1(x) \geq g_2(x)$ para todo x, entonces $Eg_1(X) \geq Eg_2(X)$.*
4. *Si $a \leq g_1(x) \leq b$ para todo x, entonces $a \leq Eg_1(X) \leq b$.*

Dem. Probaremos el caso discreto.

1. Sea $g(x) := ag_1(x) + bg_2(x) + c$, luego $Y = g(X)$ está bien definido y entonces

$$\begin{aligned} E(ag_1((X) + bg_2(X) + c) = \sum_{x\in\mathcal{X}}(ag_1(x) + bg_2(x) + c)f_X(x) = \\ a\sum_{x\in\mathcal{X}} g_1(x)f_X(x) + b\sum_{x\in\mathcal{X}} g_2(x)f_X(x) + c\sum_{x\in\mathcal{X}} f_X(x) \\ = aEg_1(X) + bEg_2(X) + c \end{aligned}$$

ya que $\sum_{x\in\mathcal{X}} f_X(x) = 1$.

2. Sea $g_1(x) \geq 0$, entonces

$$Eg_1(X) = \sum_{x \in \mathcal{X}} g_1(x) f_X(x) \geq 0$$

ya que tenemos una suma contable de productos no negativos.

3. Si $g_1(x) - g_2(x) \geq 0$, para todo x, entonces

$$E(g_1(X) - g_2(X)) = Eg_1(X) - Eg_2(X) \geq 0$$

4. si $a \leq g_1(x) \leq b$, entonces podeos pensar a, b como funciones constantes, luego

$$Ea(X) = a \leq Eg_1(X) \leq b = Eb(X)$$

Probemos el caso continuo.

1. Sea $g(x) := ag_1(x) + bg_2(x) + c$, luego $Y = g(X)$ y entonces

$$\begin{aligned} E(ag_1(X) + bg_2(X) + c) &= \int_{-\infty}^{\infty} (ag_1(x) + bg_2(x) + c) f_X(x) dx \\ &= a \int_{-\infty}^{\infty} g_1(x) f_X(x) dx + b \int_{-\infty}^{\infty} g_2(x) f_X(x) dx + c \int_{-\infty}^{\infty} f_X(x) dx \\ &= aEg_1(X) + bEg_2(X) + c \end{aligned}$$

ya que $\int_{-\infty}^{\infty} f_X(x) dx = 1$.

2. Si $g_1(x) \geq 0$, entonces para todo $x \in \mathbb{R}$ $g_1(x) f_X(x)$, luego

$$Eg_1(X) = \int_{-\infty}^{\infty} g_1(x) f_X(x) dx$$

3. Si $g_1(x) - g_2(x) \geq 0$, entonces

$$\int_{-\infty}^{\infty} (g_1(x) - g_2(x)) f_X(x) dx \geq 0$$

Luego es solo separar las integrales.

4. Notemos que a, b son funciones constantes, luego por 3. tenemos que

$$Ea(X) = a \leq Eg_1(X) \leq b = Eb(X)$$

Probando el teorema. ■

Esto implica que la esperanza tiene un comportamiento lineal, es creciente sobre las funciones y mantiene la no negatividad.

3.3.2. Varianza

Definición 3.9. *La varianza de una varaible aleatoria X es el segundo momento central (más adelante se define) y se define por:*

$$Var(X) = E((X - E(X))^2)$$

La raíz positiva de $Var(X)$ se le llama desviación estándar de X.

Corolario 3.1. *Sea X una variable aleatoria con varianza bien definida, entonces*

$$Var(X) = E(X^2) - E^2(X)$$

Dem. Basta usar la definición de varianza. Supongamos que X es continua, luego

$$\begin{aligned}
\mathrm{Var}(X) &= \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)dx \\
&= \int_{-\infty}^{\infty} x^2 f_X(x)dx - 2\mu_X \int_{-\infty}^{\infty} x f_X(x)dx + \mu_X^2 \int_{-\infty}^{\infty} f_X(x)dx \\
&= E(X^2) - 2\mu_X\mu_X + \mu_X^2 \\
&= E(X^2) - \mu_X^2 = E(X^2) - E^2(X)
\end{aligned}$$

Teorema 3.8. *Sea X una variable aleatoria cuyo valor esperado existe y $a, b \in \mathbb{R}$ son constantes. Entonces*

1. $Var(X) \geq 0$.
2. $Var(a) = 0$.
3. $Var(aX) = a^2 Var(X)$-
4. $Var(aX + b) = a^2 Var(X)$.

Dem. Usaremos la definición.

1. Por definición de Var, necesariamente es no negativa.
2. Por definición

$$\mathrm{Var}(a) = E((a - E(a))^2) = E((a - a)^2) = E(0) = 0$$

3. Por el corolario 3.1, se tiene que

$$\begin{aligned}
\mathrm{Var}(aX) &= E(a^2X^2) - E^2(aX) \\
&= a^2E(X^2) - a^1E^2(X) \\
&= a^2(E(X^2) - E^2(X)) \\
&= a^2\mathrm{var}(X)
\end{aligned}$$

4. Por el corolario 3.1 se tiene que

$$\begin{aligned}\text{Var}(aX+b) &= E(a^2X^2+2abX+b^2)-E^2(aX+b)\\ &= a^2E(X^2)+2abE(X)+b^2-(a^2E^2(X)+2abE(X)+b^2)\\ &= a^2(E(X^2)-E^2(X))\\ &= a^2\text{Var}(X)\end{aligned}$$

Probando el teorema. ■

Por lo tanto la varianza tiene un comportamiento distinto de la esperanza o media. Es siempre positiva, cosa que la esperanza no siempre es, se anula en una constante, las constante en el argumento pasan afuera al cuadrado y dada una función afín, solo sobrevive el coeficiente no libre al cuadrado.

Notación. Sea X una variable aleatoria con esperanza bien definida. Denotamos

$$\mu_x := E(X)$$

y la varianza la denotamos por

$$\sigma_x^2 := E((X-\mu_x)^2)$$

Imporante ver que la varianza se considera la función $g(x) := (x-\mu_x)^2$, de forma que $\sigma_x^2 = E(g(X))$. Estos son tipos especiales de momentos no centrado y centrados de una variable aleatoria X.

Observación 3.1. Si g es una función muy complicada, entonces $E(g(X))$ o $\text{Var}(g(X))$, pueden llegar a ser sumas o integrables muy complicadas. Por lo que es importante usar aproximaciones.

Teorema 3.9. *Sea X una variable aleatoria con experanza y varianza definida, donde $Var(X) = \sigma^2, E(X) = \mu$. Si g es una función a lo menos diferenciable dos veces en el punto $x = \mu$. Entonces*

$$\begin{aligned}E(Y) &\simeq g(\mu)+\frac{g''(\mu)}{2}\sigma^2\\ Var(Y) &\simeq [g'(\mu)]^2\sigma^2\end{aligned}$$

Dem.

3.3.3. Momentos centrados y No centrados en general

Definición 3.10. *Sea X una variable aleatoria en $(\Omega,\mathcal{A},P)$ y asumamos que existen las sumatorias o integrales requeridas para cada entero positivo k:*

1. *El k-ésimo momento no centrado de X, se define como:*

$$E(X^k) = \begin{cases}\sum_{x\in\mathcal{X}} x^k f_X(x), & X \text{ discreta}\\ \int_{-\infty}^{\infty} x^k f_X(x)dx, & X \text{ continua}\end{cases}$$

Si $k=1$, entonces $E(X)=\mu_X$, la media de X.

2. El k-ésimo momento centrado de X, se define como:

$$E((X-\mu_k)^k) = \begin{cases} \sum_{x\in\mathcal{X}}(x-\mu_X)^k f_X(x), & X \text{ discreta} \\ \int_{-\infty}^{\infty}(x-\mu_X)^k f_X(x)dx, & X \text{ continua} \end{cases}$$

Si $k=2$, entonces $E((X-\mu_X)^2)=\sigma_X^2$, la varianza de X.

Podemos interpretar la media de X, μ_X y la varianza σ_X^2 como medidas de localización y de dispersión

Definición 3.11. *Sea X una variable aleatoria de forma que están bien definidos los momentos centrados y no centrados. Definimos*

1. El índice de la asimetría de la dsitribución X por:

$$\gamma_X := E\left\{\left(\frac{X-\mu_X}{\sigma_X}\right)^3\right\}$$

2. El índice de curtosis de la distribución de X por:

$$\kappa_X := E\left\{\left(\frac{X-\mu_X}{\sigma_X}\right)^4\right\}$$

Estos índices no son coincidencia, son los momentos tercero y cuarto respectívamente de la variable aleatoria estandarizada $Z := \frac{X-\mu_X}{\sigma_X}$, donde $E(Z)=0$ y $\mathrm{Var}(Z)=1$.

De forma simila se definen $E(|X|^k)$ y $E(|X-\mu_X|^k)$, llamados momento absolutos no centrado y centrado de orden k respectívamente.

Teorema 3.10. *Sea X una variable aleatoria con k-ésimo momento centrado y no centrado bien definido para la sumas e integrales (no necesariamente son valores finitos). Entonces*

1. $|E(X^k)| \leq E(|X|^k)$

2. Si $E(X^k)$ es finito, entonces $E((X-\mu_X)^k)$ es finito.

3. Si $E(X^k)$ es finito, entonces $E(X^m)$ es finito para todo $0<m<k$.

*4. **Desigualdad de Markov.** Para todo $\lambda>0$ real y $k>0$ entero, se tiene que*

$$P(|X|\geq\lambda) \leq \frac{E(|X|^k)}{\lambda^k}$$

En particular, si la varianza de X es finito, entonces

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

para todo $\varepsilon>0$.

5. *Si X es una variable aleatoria arbitraria, con k-ésimo momento finito, entonces, para $k = 1, 2, \ldots$ se tiene que*

$$E(|X|^k) = k\left\{\int_0^{\infty} x^{k-1}(1 - F_X(x))dx - \int_{-\infty}^{0} x^{k-1}F_X(x)dx\right\}$$

Dem.

1. Supongamos que X es continua. Por definición

$$E(|X|^k) = \int_{-\infty}^{\infty} |x|^k f_X(x)$$

Si

$$\left|\int_{-\infty}^{\infty} x^k f_X(x)dx\right| \leq \int_{-\infty}^{\infty} |x^k f_X(x)dx| = \int_{-\infty}^{\infty} |x|^k f_X(x)dx$$

(Dado que $f_X(x) \geq 0$). Luego

$$|E(X^k)| \leq E(|X|^k)$$

2. Supongamos que X es continua y $E(X^k)$ es un valor finito, digamos que $E(X^k) = L$, entonces

$$\int_{-\infty}^{\infty} (x - L)^k f_X(x)dx$$

...

3. Supongamos que $E(X^k) = L < \infty$, sea $1 < m < k$, si

$$\int_{-\infty}^{\infty} x^k f_X(x)dx = L < \infty$$

Y

$$\left|\int_{-\infty}^{\infty} x^m f_X(x)dx\right| \leq \int_{-\infty}^{\infty} |x|^m f_X(x)dx \leq \int_{-\infty}^{\infty} |x|^k f_X(x)dx$$

...

Ejemplo 3.24. Bajo ciertas condiciones la tensión superficial de un líquido (dina/cm) está dada por $S = 2(1 - 0{,}005T)^{1,2}$, donde T es la temperatura del líquido (grados centigrados). Suponga que T es una variable aleatoria continua con la siguiente fdp:

$$f_T(t) = \begin{cases} 3000t^{-4}, & t \geq 10 \\ 0, & eoc. \end{cases}$$

Luego

$$\begin{aligned} E(T) &= \int_{10}^{\infty} 3000t^{-3}dt = 15 \\ \mathrm{Var}(T) &= E(T^2) - 15^2 = 75 \end{aligned}$$

Para calcular $E(S)$ y $\mathrm{Var}(S)$ hay que calcular las integrales:

$$\int_0^{\infty} (1 - 0{,}005t)^{1{,}2} t^{-4} dt$$
$$\int_0^{\infty} (1 - 0{,}005t)^{2{,}4} t^{-4} dt$$

En este caso lo que haremos es usar una aproximación. Por el teorema 3.9 debemos estudiar $g'(15)$ y $g''(15)$, donde $g(t) := 2(1 - 0{,}005t)^{1{,}2}$, entonces

$$\begin{aligned} g'(t) &= -0{,}012(1 - 0{,}005t)^{0{,}2} \\ g''(t) &= 0{,}000012(1 - 0{,}005t)^{-0{,}8} \end{aligned}$$

Por lo tanto

$$g(15) = 1{,}82, g'(15) = 0{,}01, g''(15) \simeq 0^+$$

Luego

$$\begin{aligned} E(S) &\simeq 1{,}82 \\ \mathrm{Var}(S) &\simeq 0{,}87 \end{aligned}$$

Ejemeplo 3.25. Supongamos que X es una variable aleatoria continua con fdp

$$f_X(x) = \begin{cases} 1 + x, & -1 \leq x \leq 0 \\ 1 - x, & 0 \leq x \leq 1 \\ 0, & \textit{eoc.} \end{cases}$$

Sea $Y = |X|$, notemos que $|X| \leq 1$, luego $|X|^k \leq 1$ para todo $k \in \mathbb{N}$, por la monotonía de la esperanza, $E(|X|^k) \leq E(1) = 1$, es decir,

$$|E(X^k)| \leq E(|X|^k) \leq 1$$

por lo que todo momento no centrado de orden $k \in \mathbb{N}$, puede ser finito. (Puede ocurrir que la integral esté acotada entre -1 y 1 y aun así diverger.) Si además f_X es simétrico por el eje y, ($f_X(x) = f_X(-x)$), entonces sin aplicar mucho cálculo, se tiene que

$$\int_{-1}^{1} x f_X(x) dx = -\int_{-1}^{1} x f_X(x) dx$$

Luego $E(X) = 0$. Por lo tanto, Los momentos centrados y no centrado conciden, ya que para todo k

$$E((X-\mu)^k) = E(X^k)$$

Además si k es impar, se tiene que $x^k f_X(x)$ es impar, y luego

$$\int_{-1} 1x^k f_X(x)dx = -\int_{-1} x^k f_X(x)dx$$

Y así $E(X^k) = 0$ para todo k impar. Estudiemos el caso par. Sea $K = 2n$ par, luego

$$\begin{aligned}
\int_{-1}^{1} x^{2n} f_X(x) &= \int_{-1}^{0} x^{2n}(1+x)dx + \int_{0}^{1} x^{2n}(1-x)dx \\
&= \int_{-1}^{0} x^{2n}dx + \int_{-1}^{0} x^{2n+1}dx + \int_{0}^{1} x^{2n}dx - \int_{0}^{1} x^{2n+1}dx \\
&= 2\left(\int_{0}^{1} x^{2n} - \int_{0}^{1} x^{2n+1}\right) \\
&= 2\cdot\left(1\frac{1}{2n+1} - \frac{1}{2n+2}\right) \\
&= 2\left(\frac{1}{k+1} - \frac{1}{k+2}\right) \\
&= \frac{2}{(k+1)(k+2)}
\end{aligned}$$

Claramente finito para todo k. Por tanto, todo momento centrado y no centrado, es finito. Además

$$\mathrm{Var}(X) = E(X^2) = \frac{1}{6}$$

Luego $\sigma_X = \frac{1}{\sqrt{6}}$. También se deduce que

$$\begin{aligned}
\gamma_X &= E\left(\frac{X^3}{\sigma_X^3}\right) = 0 \\
\kappa_X &= E\left(\frac{X^4}{\sigma_X^4}\right) = \frac{36}{15}
\end{aligned}$$

Ejemplo 3.26. Sea X la suma de valores en las caras superioes de dos dados arrojados al azar. Es decir $\mathcal{X} = \{2, 3, \ldots, 12\}$, con fmp dada por

$$f_X(x) = P(X = x) = \begin{cases} \frac{1}{36} \times \mathrm{mín}(x-1, 13-x), & x \in \mathcal{X} \\ 0, & eoc. \end{cases}$$

Luego

1. $\mu_X = \frac{1}{36}\sum_{x=2}^{12} x \times \text{mín}(x-1, 13-x) = 7$.
2. $\sigma_X^2 = \frac{1}{36}\sum_{x=2}^{12}(x-7)^2 \times \text{mín}(x-1, 13-x) = 5{,}833$.
3. $E((X-7)^3) = \sum_{x=2}^{12}(x-7)^3 \text{mín}(x-1, 13-x) = 0$. Luego $\gamma_X = 0$ y la distribución es simétrica.
4. $E((X-7)^4) = \frac{1}{36}\sum_{x=2}^{12}(x-7)^4 \times \text{mín}(x-1, 13-x) = 80{,}5$. Luego $\kappa_X = 2{,}366$.

3.4. Función generadora de momento

3.4.1. Definición y Propiedades

La distribución de probabilidad de una variable aleatoria X, también puede ser caracterizada mediante la denomianda función generadora de moment (fgm). La existencia de la fgm determina la existencia de los momentos de la variable aleatoria X, y como su nombre lo indica, genera los momentos de X. La utilidad de los fgm es caracterizas una distribución. Es decir, si existe, es única y viceversa.

Defínción 3.12. (Función generadora de momentos) *Sea X una variable aleatoria con fda F_X. La función generadora de momento (fgm) de X (o F_X), denotada por $M_X(t)$, se define como:*

$$M_X(t) = E(e^{tX})$$

siempre que exista la esperanza para t en agluna vencidad de 0. Es decir existe $h > 0$ talque, para todo $t \in (-h, h)$, $E(e^{tX})$ exista. Si la esperanza no existe en una vecindad de 0, decimos que la fgm de X no existe.

Nota 3.1. La función $C_X(t) = \log M_X(t)$ se le llama función generadora de cumulantes.

De forma explícita, se puede determinar la fgm de X como

$$M_X(t) = \begin{cases} \sum_{x\in\mathcal{X}} e^{tx} f_X(x), & X \text{ es discreta} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)dx, & X \text{ es continua} \end{cases}$$

Veamos cómo los fgm generan los momentos de X. Notemos que si $|t| < h$ para algún $h > 0$ bien definido, entonces

$$e^{tx} = \sum_{k=0}^{\infty} \frac{t^k}{k!} x^k$$

Luego

$$M_X(t) = E(e^{tX}) = E\left(\sum_{k=0}^{\infty} \frac{t^k}{k!} X^k\right) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k)$$

por la linealdiad de la esperanza. Derivando k veces a t y evaluando $t = 0$, obtenemos el siguiente teorema.

Teorema 3.11. *Si X tiene un fgm $M_X(t)$ para una vecindad de t, supognamos que M_X es infinitamente diferenciable en $t = 0$ entonces*

$$E(X^k) = M_X^{(k)}(0)$$

donde

$$M_X^{(k)}(0) = \frac{d^k}{dt^k} M_X(t)\bigg|_{t=0}$$

Es decir, el k-ésimo momento de X es igual a la k-ésima derivada de $M_X(t)$ evaluada en $t = 0$.

No necesariamente debe ser infinitamente diferenciable, pero esto implica que solo se puede deteminar hasta la derivada de mayor grado.

Es importante notar que si existe $M_X(t)$, entonces necesariamente los momentos no centrados de X deben estar bien definidos, ya que si no $M_X(t)$ no lo estaría. Además puede pasar que $M_X(t)$ sea diferenciable n-veces, pero no exista la derivada n-ésima para $t = 0$. Mientras exista la esperanza, es decir, la espernza no depende exclusivamente de la función generadora, pero si la puede llegar a determinar.

Dem. Vamos a probar de forma alternativa el teorema. Notemos que

$$\begin{aligned}
\frac{d}{dt} M_X(t) &= \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f_X(x) dx \\
&= \int_{-\infty}^{\infty} \left(\frac{d}{dt} e^{tx} \right) f_X(x) dx \\
&= \int_{-\infty}^{\infty} x e^{tX} f_X(x) dx \\
&= E(X e^{tx})
\end{aligned}$$

Así si $t = 0$, entonces

$$\frac{d}{dt} M_X(t)\bigg|_{t=0} = E(X)$$

Procediendo de forma análoga podemos probar que

$$\frac{d^k}{dt^k} M_X(t)\bigg|_{t=0} = E(X^k)$$

Probando el teorema. ■

Ejemplo 3.27. Sea X una variable aleatora con fdp dada por:

$$f_X(x) = \begin{cases} 2e^{-2x}, & x \geq 0 \\ 0, & eoc. \end{cases}$$

Luego

$$\begin{aligned} M_X(t) &= \int_0^\infty e^{tx} 2e^{-2x} dx \\ &= \left(1 - \frac{t}{2}\right)^{-1} \end{aligned}$$

con $t < 2$. Entonces el fgm está bien definido y por tanto podemos determinar todos los momentos.

$$\begin{aligned} E(X) = M_X'(0) &= \frac{(2-t)+2}{(2-t)^2} = \left.\frac{4-t}{(2-t)^2}\right|_{t=0} = 1 \\ E(X^2) = M_X''(0) &= \left.\frac{-1+2(4-t)(2-t)}{(2-t)^4}\right|_{t=0} = \frac{15}{16} \end{aligned}$$

Y entonces $\text{var}(X) = E(X^2) - E^2(X) = \frac{15}{16} - 1 = -\frac{1}{16}$.

Ejemplo 3.28. Decimos que una variable aleatoria X tiene una distribución normal estandad, denotada por $X \sim N(0,1)$ si la fdp de X está dada por

$$f_X(x) := \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

donde $x \in \mathbb{R}$. Notemos que $\phi(-x) = \phi(x)$ para todo x, por lo que la distribución de X es simétrica con respecto al 0, luego los momentos impares existen y son nulos, además los momentos no centrados y los centrados coinciden en su valor. En este caso el momento k-ésimo no centado se determina por la integral

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

Para ello debemos integrar por parte de forma recursiva. En cambio, si notamos que

$$e^{tx}\phi(x) = e^{t^2/2}\phi(x-t)$$

el calculo se vuelve algo más simple. Entonces

$$\begin{aligned} M_X(t) &= \int_{-\infty}^\infty e^{t^2/2}\phi(x-t) dx \\ &= e^{t^2/2} \end{aligned}$$

Luego el fmg está bien definido para todo $t \in \mathbb{R}$, luego al derivar obtenemos que

$$M_X'(t) = te^{t^2/2}$$

Y entonce $E(X) = 0$, volviendo a derivar

$$M_X''(t) = e^{t^2/2} + t^2 e^{t^2/2}$$

Luego $E(X^2) = 1$, además por la caracterización de varianza

$$\text{Var}(X) = E(X^2) - E^2(X) = 1 - 0 = 1$$

Derivadno por tercera vez

$$M_X^{(3)}(t) = te^{t^2/2} + 2te^{t^2/2} + t^3 e^{t^2/2}$$

Luego $E(X^3) = 0$ y entonces $\gamma_X = 0$. Derivando por última vez

$$M_X^{(4)}(t) = e^{t^2/2} + t^2 e^{t^2/2} + 2e^{t^2/2} + 2t^2 e^{t^2/2} + 3t^2 e^{t^2/2} + t^4 e^{t^2/2}$$

Luego $\kappa_X = 3$. Por tanto, en una distribución normal estanda, se tiene que

$$\begin{aligned} \mu_X &= 0 \\ \sigma_X^2 &= 1 \\ \gamma_X &= 0 \\ \kappa_X &= 3 \end{aligned}$$

Ejemplo 3.29. Sea X una variable discreto con valores $\mathcal{X} = \{x_1, \ldots, x_m\}$. Entonces

$$M_X(t) = \sum_{i=1}^{m} e^{tx_i} P(X = x_i)$$

Si $\mathcal{X} = \{-1, 1\}$ y $P(X = -1) = P(X = 1) = \frac{1}{2}$, entonces

$$M_X(t) = \frac{1}{2}(e^{-t} + e^t)$$

Luego está definido para todo $t \in \mathbb{R}$. Entonces la media de X está dada por

$$E(X) = \frac{d}{dt} M_X(t)\bigg|_{t=0} = 0$$

Sean X, Y dos variables aleatorias, puede ocurrir que ambas tengan distitnas fdp pero iguales momentos, todo a menos que tengan soporte acotado. Por otro lado, cuando existe el fmg de una variable aleatoria, esta caracteriza a la variable. Todo resumido en un teorema.

Teorema 3.12. *Sean $F_X(x)$ y $F_Y(y)$ dos fda cuyos momentos exsiten.*

1. *Si X e Y tienen soporte acotado, entonces $F_X(u) = F_Y(u)$ para todo u si y sólo si $E(X^r) = E(Y^r)$ para todo $r = 0, 1, 2, \ldots$*
2. *Si el fgm existe y $M_X(t) = M_Y(t)$ para todo t en alguna vecindad de 0, entonces $F_x(u) = F_Y(y)$ para todo u.*

Dem. Revisar Casellea , Berger 2002.

Teorema 3.13. *Sean a, b constantes, la fgm de la variable aleatoria $aX + b$ está dada por:*

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

Dem. Sea X bien definida y a, b tales que $Y := aX + b$ está bien definida. Entonces por definición

$$\begin{aligned} M_{aX+b}(t) &= E(e^{(aX+b)t}) \\ &= e^{bt} E(e^{at} X) \\ &= e^{bt} M_X(at) \end{aligned}$$

Probando el teorema. ■

Ejemplo 3.30. Sabemos que $X \sim N(0, 1)$, si y sólo si $M_X(t) = e^{t^2/2}$ con $t \in \mathbb{R}$. Entonces por el teorema 3.13, la fgm de $Y = \sigma X + \mu$, donde $\sigma > 0$ y $\mu \in \mathbb{R}$, está dada por:

$$M_Y(t) = e^{t\mu + t^2\sigma^2/2}$$

Tal fgm carateriza la denominada distribución normal con media μ y varianza σ^2, denotada por $Y \sim N(\mu, \sigma 2)$. Si $E(X) = 0$, $\mathrm{Var}(X) = 1$, entonces

$$\begin{aligned} E(Y) &= \sigma E(X) + \mu = \mu \\ \mathrm{Var}(Y) &= \sigma^2 \mathrm{Var}(X) = \sigma^2 \end{aligned}$$

Además, es posible probar que la fdp de $Y = \sigma X + \mu$ está dada por:

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2}$$

con $y \in \mathbb{R}$.

En general, si X es una variable aleatoria de un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ se puede definir la función generadora definida por:

$$G_X(t) := E(t^X)$$

Siempre que la esperanza exista. Se puede notar que:

1. Si $M_X(t)$ existe en una vecindad de 0, entonces, $G_X(t) = M_X(\log t)$ está bien definido con $t > 0$ ya que

   $$M_X(\log t) = E(e^{(\log t)x}) = E(t^X) = G_X(t)$$

   $M_X(t)$ está bien definido en una vecindad, se puede tomar $t > 0$ talque $\log(t)$ está definida en la vecindad de $M_X(t)$, y luego se peude ver que $G_X(t)$ está bein definido.

2. Si X es una variable aleatoria discreta con recorrido $\mathcal{X} = \{0, 1, 2, \dots\}$, entonces

$$G_X(t) = \sum_{x=0}^{\infty} t^x P(X = x)$$

y se llama función generadora de probabilidad (fgp). En tal caso, $G_X(t)$ converge al menos para $|t| \leq 1$ ya que entonces $|t|^x \leq 1$ con $x = 0, 1, \dots$ Para ver esto se necesita probar que la suma parcial

$$\sum_{x=0}^{n} t^x P(X = x)$$

es Cauchy, y en efecto, si $n \leq m$, entonces

$$\left| \sum_{x=n}^{m} t^x P(X = x) \right| \leq \sum_{x=n}^{m} |t^x| P(X = x)$$

Si $0 \leq |t| < 1$, se puede tomar n suficientemente grande, talque $|t|^n < \varepsilon$, luego

$$\left| \sum_{x=n}^{m} t^x P(X = x) \right| \leq \varepsilon \sum_{x=n}^{m} P(X = x) < \varepsilon$$

Luego congerge. Si $t = 1$, es evidente que converge ya que $G_X(1) = E(1) = 1$.

3. Si X es continua y $E(|X|^k) < \infty$, entonces $E(X(X-1)\dots(X-k+1)) = G_X^{(k)}(1)$ donde $G_X^{(k)}(1) = \frac{d^k}{dt^k} G_X(t)|_{t=1}$. Por el punto 1. sabemos que

$$G_X(t) = M_X(\log(t))$$

Luego

$$\frac{d}{dt} G_X(t) = M_X'(\log(t)) \frac{1}{t}$$

Si $t = 1$, entonces

$$G_X'(1) = M_X'(0) = E(X)$$

Volviendo a derivar

$$\frac{d^2}{dt^2} G_X(t) = M_X''(\log(t)) \frac{1}{t^2} - M_X'(\log(t)) \frac{1}{t^2}$$

Si $t = 1$, entonces

$$G_X''(1) = E(X^2) - E(X) = E(X(X-1))$$

Luego por inducción se ...

3.4.2. Función Generatriz de Probabilidad

La importancia de la fgp, es que se utiliza en muchas áreas de probabilidad aplicada, en particular, en procesos estocásticos. Una definición más formal del fgp está dada a continuación

Definición 3.13. (Función Generatriz de Probabilidad) *Sea X una variable aleatoria discreta, no negativa con valores enteros y sea $p_k := P(X = k)$, con $k = 0, 1, \ldots$. Talque $\sum_{k=0}^{\infty} p_k = 1$. La fgp de X se define como:*

$$G_X(t) = E(t^X) = \sum_{k=0}^{\infty} p^k t^k$$

con $|t| < 1$

Ejemeplo 3.31. Sea X una variable aleatoria con fmp dada por:

$$f_X(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, \ldots, n \\ 0, & eoc. \end{cases}$$

donde $n \in \{1, 2, \ldots\}$ y $p \in (0, 1)$. Entonces

$$\begin{aligned} G_X(t) &= \sum_{x=0}^{n} t^x \binom{n}{x} p^x (1-p)^{n-x} \\ &= (1-p)^n \sum_{x=0}^{n} \binom{n}{x} \left(\frac{pt}{1-p}\right)^x \\ &= (1-p)^n \left(\frac{pt}{1-p} + 1\right)^n = (pt + 1 - p)^n \end{aligned}$$

con $|t| > 0$. Si $t = 0$, entonces $G_X(0) = E(0) = 0$. Es decir, la función generatriz de probabilidad de X, existe para todo t. Por lo tanto $M_X(s)$ está bien definido para todo $s \in \mathbb{R}$, ya que si $G_X(t)$ se define para todo $t \in \mathbb{R}$, se puede considerar $s := \log(t)$ donde $t > 0$, luego $s \in \mathbb{R}$ y por la identidad

$$G_X(t) = M_X(\log(t))$$

Se ve que está bien definido.

Las funciones generadoras no siempre existen, pero podemos definir un alterno con recorrido en los complejos.

Definición 3.14. (Función característica) *La función característica (fc) de una variable aleatoria X se define como:*

$$\varphi_X(t) = E(e^{itX}) = E(\cos tX) + iE(\sin tX)$$

donde $i = \sqrt{-1}$, es decir, $\varphi_X : \mathbb{R} \to \mathbb{C}$

Nota 3.2. Como $|E(e^{itX})| \leq E(|e^{itX}|) = E(\sqrt{\cos^2 tX + \sin^2 tX}) = E(1) = 1$, entonces $\varphi_X(t)$ existe para todo t y para toda variable aleatoria X.

Proposición 3.1. *Sea X una variable aleatoria, entonces con respecto a la función característica se cumple:*

1. $|\varphi_X(t)| \leq \varphi_X(0) = 1$ *para todo* t.
2. $\overline{\varphi_X(t)} = E(e^{-itX}) = \varphi_X(-t)$.
3. $\varphi_X(t) = \varphi_Y(t)$ *para todo* t *si y sólo si* $X \overset{d}{=} Y$ *si y sólo si* $F_X(z) = F_Y(z)$ *para todo* z.
4. $\varphi_X(t)$ *es real si y sólo si* $X \overset{d}{=} -X$ *(simetría con respecto de* 0*).*
5. $\varphi_{aX+b}(t) = e^{ibt} = \varphi_X(at)$
6. *Si* $E(|X|^k) < \infty$, *entonces* $\varphi_X^{(k)}(0) = i^k E(X^k)$, $k = 1, 2, \ldots$.
7. *Si* $M_X(t)$ *existe en una vecindad de* 0, *entonces* $\varphi_X(t) = M_X(it)$ *o* $M_X(t) = \varphi_X(-it)$.

Ejemplo 3.32. Sea X una variable aleatoria con:

$$P_X(X = -1) = P_X(X = 1) = 1/2$$

entonces

$$\begin{aligned}
\varphi_X(t) &= \frac{1}{2}(\cos(t) + \cos(-t)) + \frac{i}{2}(\text{sen}(t) + \text{sen}(-t)) \\
&= \cos(t) \Rightarrow X \overset{d}{=} -X
\end{aligned}$$

Sea X una variable aleatoria con fdp $f_X(x) = \frac{1}{\pi(1+x^2)}$ con $x \in \mathbb{R}$. Entonces

$$\varphi_X(t) = \int_{-\infty}^{\infty} \frac{e^{itx}}{\pi(1+x^2)} dx = e^{-|t|}$$

En este caso, el fgm de X no existe.

Teorema 3.14. (Regla de Leibnitz) *Sean $f(x, \theta), a(\theta), b(\theta)$ funciones diferenciables, respecto a θ, entonces*

$$\begin{aligned}
\frac{d}{d\theta} \int_{a(\theta)}^{b(\theta)} f(x, \theta) dx = f(b(\theta), \theta) \frac{d}{d\theta} b(\theta) - f(a(\theta), \theta) \frac{d}{d\theta} a(\theta) \\
+ \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta} f(x, \theta) dx
\end{aligned}$$

En particular, si $a(\theta) = a, b(\theta) = b$ (abuso de notación) son constantes, entonces

$$\frac{d}{d\theta} \int_{a}^{b(\theta)} f(x, \theta) dx = \int_{a}^{b} \frac{\partial}{\partial\theta} f(x, \theta) dx$$

4. Distribuciones Comunes

Empezaremos con el estudio más aplicativo y detallado de distribuciones comunes, veremos las medias, esperanza, entre otros. Estudiando situaciones cotiandas y dandole sentidos a estas.

Notación. Para simplificar la escritura usaremos los siguientes abusos de notación: $F_X = F, f_X = f, M_X = M, \mu_X = \mu, \sigma_X^2 = \sigma^2$, etc.

4.1. Distribuciones Discretas

4.1.1. Distribución Uniforme Discreta

Definición 4.1. *Una variable aleatoria X tiene una distribución uniforme discreta sobre $\mathcal{X} = \{x_1, \ldots, x_N\}$, donde N es un entero positivo, si su fmp está dada por:*

$$p(x) := \begin{cases} \frac{1}{N}, & x \in \mathcal{X} \\ 0, & eoc. \end{cases}$$

Notación. $X \sim UD(\{x_1, \ldots, x_N\})$

Teorema 4.1. *Si $X \sim UD(\{x_1, x \ldots, x_N\})$, entonces*

1. $E(X^r) = \frac{1}{N}\sum_{i=1}^N x_i^r$, $k = 1, 2, \ldots$
2. $M_X(t) = \frac{1}{N}\sum_{i=1}^N e^{tx_i}$

Dem. Es solo aplicar definición.

1. $E(X^r) = \sum_{i=1}^N x_i^r \frac{1}{N} = \frac{1}{N}\sum_{i=1}^N x_i^r$
2. $M_X(t) = E(e^{tX}) = \frac{1}{N}\sum_{i=1}^N e^{tx_i}$

$\blacksquare$

Corolario 4.1. *Si $\mathcal{X} = \{1, \ldots, N\}$, entonces*

1. $E(X) = \frac{N+1}{2}$
2. *Var*$(X) = \frac{N^2-1}{12}$
3. $M_X(t) = \frac{1}{N}\sum_{x=1}^N e^{xt}$.

Dem.

1. Por el teorema 4.1 sabemos que si $r = 1$, entonces

$$E(X) = \frac{1}{N}\sum_{i=1}^N i = \frac{1}{N}\frac{N(N+1)}{2} = \frac{N+1}{2}$$

2. Si $r = 2$, entonces

$$E(X^2) = \frac{1}{N}\frac{N(N+1)(2N+1)}{6} = \frac{2N^2+3N+1}{6}$$

Y si

$$\operatorname{Var}(X) = E(X^2) - E^2(X)$$

entonces

$$\begin{aligned}\operatorname{Var}(X) &= \frac{2N^2+3N+1}{6} - \frac{N^2+2N+1}{4}\\ &= \frac{N^2-1}{12}\end{aligned}$$

3. Solo basta usar la definición. Entonces

$$M_X(t) = \frac{1}{N}\sum_{x=1}^{N} e^{xt}$$

■

Ejemplo 4.1. Sea X el número obtenido al lanzar un dado equilibrado. El espacio muestral es $\mathcal{X} = \{1,2,3,4,5,6\}$ y cada resultado tiene probabilidad $p(x) = \frac{1}{6}$ por las condiciones impuestas. Por lo tanto, es evidente que $X \sim UD$, luego

$$\begin{aligned}E(X) &= \frac{7}{2} = 3{,}5\\ \operatorname{Var}(X) &= \frac{35}{12}\end{aligned}$$

Determinemos el fgm de X. Por definición

$$M_X(t) = \frac{1}{6}\sum_{x=1}^{6} e^{xt} = \frac{e^t+e^{2t}+e^{3t}+e^{4t}+e^{5t}+e^{6t}}{6} = \frac{e^t(e^{6t}-1)}{6(e^t-1)}$$

4.1.2. Distribución Binomial

La distribución binomial, es una de las distribuciones discretas más útiles. Se basa en la idea de un ensayo de Bernoulli. El ensayo de Bernoulli es un experimente con sólo dos resultados posibles. $\Omega = \{E, F\}$, con E éxito, F fracaso, donde $P(E) = p$ donde $0 < p < 1$.

Definición 4.2. *Una variable aleatoria X tiene una distribución binomial con parámetros n y p, si su fmp está dada por:*

$$p(x) := \begin{cases}\binom{n}{x}p^x(1-p)^{n-x}, & x = 0,1,\ldots,n\\ 0, & eoc.\end{cases}$$

donde n es un entero positivo y $0 < p < 1$. Si $n = 1$, la distribución binomial es denominada una distribución Bernoulli con parámetro p.

Notación. $X \sim Bin(n, p)$ y si $n = 1$, entonces $X \sim Ber(p)$.

Teorema 4.2. *Sea $X \sim Bin(n, p)$, entonces*

1. $E(X) = np$.
2. *Var*$(X) = npq$.
3. $M_X(t) = (pe^t + q)^n$, *donde* $q = 1 - p$.

Dem. Determinemos $M_X(t)$

$$\begin{aligned} M(t) &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\ &= (pe^t + q)^n \end{aligned}$$

Luego está bien definido para todo $t \in \mathbb{R}$. Ahora para determinar la media y la varianza, basta derivar fgm.

$$M'(t) = n(pe^t + q)^{n-1}(pe^t)$$

si $t = 0$, entonces

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

Volviendo a derivar

$$M''(t) = n(n-1)(pe^t + q)^{n-2}(pe^t)^2 + n(pe^t + q)^{n-1}(pe^t)$$

luego si $t = 0$, entonces

$$E(X^2) = n(n-1)p^2 + np$$

Luego

$$\text{Var}(X) = n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p) = npq$$

■

Ejemplo 4.2. Una prueba de selección multiple consta de 10 preguntas y cada pregunta tiene 4 respuesta de las cuales sólo una es la correcta. El rendimiento de un postulante se considera satisfactorio si obtiene por lo menos 6 respuestas corretas. Un postulante no preparado elige las respuesta al azar. Calculemos la probabilidad de que el rendimiento del postulante sea satisfactorio.

Notemos que tenemos 4 alternativas, a),b),c),d) (se toman estos por comodidad), luego tenemos un muestreo ordenado con reemplazo, ya que cada 4 alternativa está asociada a una pregunta, luego los casos totales son 4^{10} y los favorables son $\binom{10}{x}(1)^x(3)^{10-x}$ para tener x buenas de 10. Luego

$$P(\text{Tener } x \text{ buenas}) = \binom{10}{x}(0{,}25)^x(0{,}75)^{10-x}$$

Si consideramos la variable aleatoria X como el número de correctas, observamos que $X \sim Bin(10, 0{,}25)$, luego queremos determinar $P(X \geq 6)$, es decir

$$\sum_{6}^{10}\binom{10}{x}(0{,}25)^x(0{,}75)^{10-x} = 0{,}20$$

es la probabilidad de tener un rendimiento satisfactorio. Además

$$E(X) = 10(0{,}25) = 2{,}5$$
$$\mathrm{Var}(X) = (2{,}5)(0{,}75) = 1{,}875$$

4.1.3. Distribución Hipergeométrica

La distribución hipergeométrica tiene muchas aplicaciones en el muestreo de pobablciones finitas.

Definición 4.3. *Supongamos que elegimos al azar y sin reemplazo n objetos de una pobalción de N, de los cuales K poseen una característica de interés, en este caso digamos A. Luego $N(\Omega) = \binom{N}{n}$. Sea X el número de objetos con la características A, de los n elegidos. Entonces el fmp de X está dada por:*

$$p(x) := \begin{cases} \dfrac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, & x = 0, 1, 2, \dots, \text{mín}\{n, K\} \\ 0, & eoc. \end{cases}$$

En este caso se dice que X tiene una distribución Hipergeométrica.

Notación. $X \sim Hip(n, K, N)$.

Ejemplo 4.3. Una compañia recibe una partida de 20 items. Debio a que la inspección de cada items es costosa, se chequea al azar 6 artículos. La partida se acepta si no más uno de los items inspeccionados es defectuoso. Calculemos la probabilidad que en una partida con 5 item defectuosos, sea aceptada. Sea X el número de artículos defectuoso en los 6 items tomados, notemos estamos presenta en un muestro sin orden y sin repetición, luego

$$p(x) = \begin{cases} \dfrac{\binom{5}{x}\binom{15}{6-x}}{\binom{20}{6}}, & x = 0, 1, 2, 3, 4, 5. \\ 0, & eoc. \end{cases}$$

es decir, $X \sim Hip(6, 5, 20)$ y se pide $P(X \leq 1) = p(0) + p(1) = 0{,}516$.

Ejemplo 4.4. Se embarcan motores elétricos pequeños en lotes de 50. Antes de que tal cargamento sea aceptado, un inspector elgie 5 motoroes y los inspecciona. Si ninguno de los motores probados es defectuoso, el lote se acepta. Si se encunetra que uno o más son defectuoso, se inspecciona el cargamente completo. Supongamos que, en realidad hay tres motores defectuoso en el lote. Determinemos la probabilibdad de que sea necesaria la inspección de 100 %. Podemos tomar X el número de motores defectuoso, dode claramente $X \sim Hip(5, 3, 50)$, luego queremos determinar $P(X \geq 1)$, es decir

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{3}{0}\binom{47}{5}}{\binom{50}{5}} = 0{,}28$$

Teorema 4.3. *Sean $X \sim Hip(n, K, N), p = \frac{K}{N}$ y $q = 1 - p$. Entonces*

1. $E(X) = np$
2. $Var(X) = npq\frac{N-n}{N-1}$.
3. $p(x) \simeq \binom{n}{x}p^x(1-p)^{n-x}$, *para N grande.*

Dem. Por terminar

Ejemplo 4.5. En una ciudad con 2 millones de habitantes, 60 % pertenecen al partido político A. Cien personas son elegidas al azar de los habitantes. La distribución del número de personas, entre las 100 elegidas que pertenecen al partido A, es una variable aleatoria X con distribución $Hip(100, 120000, 2000000)$. Aplicando la aproximación podemos ver que

$$P(X = 40) \simeq \binom{100}{40}(0{,}6)^{40}(0{,}4)^{60} = 2{,}4425 \times 10^{-5}$$

4.1.4. Distribución Poisson

Definición 4.4. *Una variable aleatoria X tiene distribución de Poisson con parámetro $\lambda > 0$ si su fmp está dada por:*

$$p(x) := \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, & eoc. \end{cases}$$

Notación $X \sim P(\lambda)$.

Teorema 4.4. *Sea $X \sim P(\lambda)$, entonces*

1. $E(X) = \lambda$.
2. $Var(X) = \lambda$.
3. $M(t) = \exp(\lambda(e^t - 1))$

Dem. Calculemos el fgm. Por definición

$$\begin{aligned} M(t) &= \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\ &= e^{-\lambda} e^{\lambda e^t} \\ &= e^{\lambda(e^t - 1)} \end{aligned}$$

La primera y segunda derivada de fgm es

$$\begin{aligned} M'(t) &= e^{\lambda(e^t-1)} \lambda(e^t) = \lambda e^{\lambda(e^t-1)+t} \\ M''(t) &= \lambda e^{\lambda(e^t-1)+t}(\lambda e^t + 1) \end{aligned}$$

Luego si $t = 0$, entonces

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

Ejemplo 4.6. El número de pacientes que acuden diaramente a la sala de emergencia (ER) de cierto hosítal tiene una distribución de Poisson con una media de 10 personas/día. Determinemos la probabilidad que durante un día normal, eñ número de pacientes admitidos en la sala de emergencia del hospital sea menor o igual a 3. Consideremos X la cantidad de pacientes que acuden diaramente a la ER, del enunciado del problema, sabemos que $X \sim P(10)$ y por lo tanto

$$\begin{aligned} P(X \leq 3) &= e^{-10} + 10e^{-10} + 50e^{-10} + \frac{1000}{6} e^{-10} \\ &= 1{,}0336 \times 10^{-2} \end{aligned}$$

Teorema 4.5. *Sea X una variable aleatoria con distribución binomial con parámetros n y p. Es decir:*

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \;\; x = 0, 1, \ldots, n$$

Supongamos que cuando $n \to \infty$, entonces $np = \lambda$. Entonces bajo estas condiciones se tiene que:

$$\lim_{n \to \infty} P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \;\; x = 0, 1, \ldots$$

la distribución de Poisson con parámetro λ.

Dem. Sea $X \sim Bin(n,p)$, si

$$\begin{aligned} P(X=x) &= \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= \frac{n(n-1)(\dots)(n-x+1)}{x!} p^x \sum_{i=0}^{n-x} \binom{n-x}{i} (-p)^i \end{aligned}$$

notemos que

$$\lim_{n\to\infty} np = \lim_{n\to\infty} (n-1)p = \cdots = \lim_{n\to\infty} (n-x+1)p = \lambda$$

Luego

$$\begin{aligned} P(X=x) &= \lim_{n\to\infty} \frac{n(n-1)(\dots)(n-x+1)}{x!} p^x (1-p)^{n-x} \\ &= \frac{\lambda^x}{x!} \lim_{n\to\infty} \sum_{i=0}^{n-x} \binom{n-x}{i} (-p)^i \end{aligned}$$

Nos falta determinar el otro límite, en particular

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

Luego si $n \to \infty$, llegamos a

$$\lim_{n\to\infty} \sum_{i=0}^{n-x} \binom{n-x}{i} (-p)^i = \sum_{i=0}^{\infty} \frac{(-\lambda)^i}{i!} = e^{-\lambda}$$

Por lo tanto, si $n \to \infty$, entonces

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

con $x = 0, 1, \dots$. Por lo tanto $X \sim Bin(n,p)$ se convierte Poisson cuando $n \to \infty$ si $np \to \lambda \in \mathbb{R}^+$.

Digamos que $np \overset{n\to\infty}{\longrightarrow} 1 = \lambda$

x	$P(\lambda)$	$Bin(100, 0{,}01)$	$Bin(10, 0{,}1)$
0	0,3678	0,3660	0,3487
1	0,3679	0,3697	0,3874
2	0,01839	0,1848	0,1937
3	0,0613	0,0610	0,0574
4	0,0153	0,0149	0,0112

Ejemplo 4.7. Hay 135 estudiantes dentro de una sala de conferencias. La probabilidad de que uno de los estudiantes celebre su cumpleaños hoy es $\frac{1}{365}$.

Entonces tenemos un muestreo ordenado y con reemplazo, donde la persona i puede o no estar de cumpleaño ese día. Por lo que es claro que $X \sim Bin(135, 1/365)$, es decir, dado $x = 0, \ldots, 135$, entonces

$$p(x) = \binom{135}{x} p^x (1-p)^{135-x}$$

Calculemos la probabilidad de que dos o más estudiantes en la sala de conferencia estén celbrando sus cumpleaños hoy. Sea X el número de estudiantes que celebran su cumpleaños. Usando la aproximación de Poisson (que es mala ya que n no es tán grande), se tiene que $np = \lambda$, luego

$$\overline{p}(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$\begin{aligned} P(x \geq 2) &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - e^{-\frac{27}{73}} - \frac{27}{73} e^{-\frac{27}{73}} \\ &= 5{,}3659 \times 10^{-2} \end{aligned}$$

4.1.5. Distribución Binomial negativa y Geométrica

Recordemos que la distribución binomial cuenta la cantindad de éxito fijo de ensayos de Bernoulli. Si contamos el número de ensayos necesarios para obtener r éxito se genera la distribución negativa ($r \geq 1$) y geométrica ($r = 1$).

Se usa la distribución binomial para llegar a esta distribución, en efecto, el evento $\{X = x\}$ puede ocurrir sólo si hay extactamente $r - 1$ éxitos en los primeros $x - 1$ ensayos y un éxito en el ensayo x-ésimo.

Entonces, tenemos x ensayos, de los cuales en $x - 1$ primeros ensayos deben haber $r - 1$ éxitos, que como sabemos, se determina por la distribución binomial, es decir, la probabilidad de que hayan $r - 1$ éxito en los $x - 1$ ensayos, es

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

Ahora nos falta el último ensayo, con probabilidad p, por lo tanto

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

Notemos que $x \geq r$ ya que para que hayan r éxitos, se necesitan al menos r ensayos.

Definición 4.5. *Se dice que una variable aleatoria X tiene distribución binomial negativa (o de Pascal) con parámetros r, p si su fmp está dada por:*

$$p(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r}, & x = r, r+1, \ldots \\ 0, & eoc. \end{cases}$$

Si $r = 1$, se dice que la variable aleaotoria tiene distribución geométrica con parámetro p, y su fmp está dada por:

$$p(x) = \begin{cases} p(1-p)^{x-1}, & x = 1, 2, \ldots \\ 0, & eoc. \end{cases}$$

Notación. $X \sim BN(r, p)$ para la distribución binomial negativa y $X \sim Geo(p)$ para la distribución geometríca.

Ejemplo 4.8. En un departamente de control de calidad, se inspeccionan las unidades que provienen de una línea de ensamblaje. Si la proporición de unidades defectuosas es 0,03. Determinemos la probabilidad de que la vigésima unidad inspeccionada sea la tercera defectuosa. Sea X el número de unidad a inspeccionar para obtener tres defectuosas. Luego tenemos que $X \sin BN(3, 00{,}3)$, por lo tanto

$$\begin{aligned} P(X = 20) &= \binom{19}{2} (0{,}00)^3 (1 - 0{,}03)^{17} \\ &= 2{,}7509 \times 10^{-3} \end{aligned}$$

Teorema 4.6. *Sea $X \sim BN(r, p)$, entonces*

1. $E(X) = r/p$.
2. $Var(X) = r(1-p)/p^2$.
3. $M(t) = \left(\frac{pe^t}{1-(1-p)e^t} \right)^r$

Dem. El k-ésimo momento de X alrdededor del origen está dado por:

$$\begin{aligned} E(X^k) &= \sum_{j=r}^{\infty} j^k \binom{j-1}{r-1} p^r (1-p)^{j-r} \\ &= \frac{r}{p} \sum_{j=r}^{\infty} j^{k-1} \binom{j}{r} p^{r+1} (1-p)^{j-r} = \frac{r}{p} E((Y-1)^{k-1}) \end{aligned}$$

donde $Y \sim BN(r+1, p)$. Por lo tanto

$$E(X) = \frac{r}{p} \text{ y } E(X^2) = \frac{r}{p} E(Y-1) = \frac{r}{p} \cdot \frac{r+1-p}{p}$$

Luego

$$\begin{aligned}\mathrm{Var}(X) &= \frac{r^2 + r - rp}{p^2} - \frac{r^2}{p^2} \\ &= \frac{r(1-p)}{p^2}\end{aligned}$$

Por otro lado

$$M(t) = \sum_{j=r}^{\infty} e^{tj} \binom{j-1}{r-1} p^r (1-p)^{j-r} = (pe^t)^r (1-(1-p)e^t)^{-r}$$

Corolario 4.2. *Si $X \sim Geo(p)$, entonces*

1. $E(X) = 1/p$.
2. $Var(X) = (1-p)/p^2$.
3. $M(t) = \left(\frac{pe^t}{1-(1-p)e^t}\right)$.

Basta considerar la variable aleatoria $Y = X - r$, entonces

$$\begin{aligned}E(Y) &= \frac{r}{p} - r = \frac{r(1-p)}{p} \\ \mathrm{Var}(Y) &= \mathrm{Var}(X) = \frac{r(1-p)}{p^2} \\ M_Y(t) &= \left(\frac{p}{1-(1-p)e^t}\right)^r\end{aligned}$$

La distribución geométrica tiene una propiedad conocidad como "sin memoria". Consideremos $s > t$, entonces

$$P(X > s | X > t) = P(X > s - t)$$

Para ver esto, basta usar definición, entonces si $s > t$ entonces

$$\begin{aligned}P(X > s | X > t) &= \frac{P(X > s \text{ y } X > t)}{P(X > t)} \\ &= \frac{P(X > s)}{P(X > t)} \\ &= \frac{(1-p)^s + (1-p)^{s+1} + \dots}{(1-p)^t + (1-p)^{t+1} + \dots} \\ &= (1-p)^{s-t} = P(X > s - t)\end{aligned}$$

Ejemeplo 4.9. (Tiempos de falla) La distribución geométrica a veces se usa para modelar tiempo hasta la falla de ciertos componenetes. Por ejemplo, la probabilidad de que una bombilla falle en una día, es 0,0001, entonces la probabilidad de que dure almenos 30 días es

$$P(X > 30) = \sum_{x=31}^{\infty} 0{,}001(1 - 0{,}01)^{x-1} = (0{,}999)^{30} = 0{,}970$$

La propiedad sin memoria de la distribución geométrica, describe otra propiedad especial de "falta de nvejicimiento". Este indica que la distribución geométrioca no es aplicable al modelado de tiempos de falla, para las cuales se espera que la probabilidad de falla aumente con el tiempo.

4.2. Distribuciones Continuas

Pasamos a las distribuciones continuas más comunes. Recordemos que una función integrable no negativa se puede transformar en una pdf.

4.2.1. Distribución Uniforme.

Definiciín 4.6. *La distribución uniforme continua, se define extendiendo la masa uniformemente en un intervalo* (a, b). *Su fdp está dada por:*

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b) \\ 0, & eoc. \end{cases}$$

Notación. $X \sim U(a, b)$ o $X \sim U([a, b])$

También podemos considerar los bordes de (a, b).

Teorema 4.7. *Si* $X \sim U(a, b)$, *entonces*

1. $E(X) = \frac{a+b}{2}$.
2. $Var(X) = \frac{(b-a)^2}{12}$.
3. $M(t) = \frac{e^{bt}-e^{at}}{t(b-a)}$, $t \neq 0$.

Dem. Determinemos el fgm de X, sea $t \neq 0$, entonces

$$\begin{aligned} M(t) &= \int_b^a e^{tx} \frac{1}{b-a} dx = \frac{1}{b-a} \frac{e^{tb} - e^{ta}}{t} = \\ &= \frac{e^{bt} - e^{at}}{t(b-a)} \end{aligned}$$

Notemos que no $M'(t)$ no es evaluable en $t = 0$, por que no está bien definido en una vecindad de 0, aun así es determinable en un entorno $V \setminus \{0\}$ de 0. Para calcula la media y la varianza, hay que hacerlo directamente.

$$E(X) = \int_b^a \frac{x}{b-a} dx = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

Y

$$E(X^2) = \int_a^b \frac{x^2}{b-a}dx = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}$$

Luego

$$\begin{aligned}\mathrm{Var}(X) = E(X^2) - E^2(X) &= \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} \\ &= \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}\end{aligned}$$

Ejemeplo 4.10. Sea $X \sim U(-3,2)$. Calculemos $P(X \geq 0)$ y de $P(-5 \leq X \leq \frac{1}{2})$. En este caso la fdp de X está dada por:

$$f(x) = \begin{cases} \frac{1}{5}, & -3 < x < 2 \\ 0, & eoc. \end{cases}$$

Por lo tanto

$$P(X \geq 0) = \int_0^2 \frac{1}{5}dx = \frac{2}{5}$$

y

$$P(-5 \leq X \leq \frac{1}{2}) = \int_{-3}^{\frac{1}{2}} \frac{1}{5}dx = \frac{1}{5}\left(\frac{1}{2}+3\right) = \frac{7}{10}$$

4.2.2. Distribución Normal/ Normal Estándar

La distribución normal es quizás una de las más importantes, debido a que tiene un rol central tanto en la teoría de probabilidad como en la teoría estadística. También es conocido como distribución gaussiana en honor a Gauss. La importancia de la distribución normal se debe al famoso teorema del límite central-

Definición 4.7. *Se dice que una variable aleatoria X tiene distribución normal con parámetros μ (número real) y σ (real positivo), si su fdp está dada por:*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \ x \in \mathbb{R}$$

Notación. $X \sim N(\mu, \sigma^2)$

Insertar figura

Definición 4.8 (Distribución Normal Estándar) *Si $Z \sim N(0,1)$ se dice que Z tiene una distribución normal estándar. Su fdp está dada por:*

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right), \ z \in \mathbb{R}$$

Su fda, se denota por Φ y está dada por:

$$\Phi(z) := \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx, \ z \in \mathbb{R}$$

En general lo valores se usan en $-4 < z - 4$. Además

$$\Phi(0) = \frac{1}{2} \text{ y } \Phi(z) + \Phi(-z) = 1, \text{ para todo } z \in \mathbb{R}$$

Teorema 4.8. *Sea $X \sim N(\mu, \sigma^2)$. Entonces*

1. $E(X) = \mu$.
2. $Var(X) = \sigma^2$.
3. $M(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$

Calculemos el fgm de X

$$\begin{aligned} M(t) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(tx - \frac{1}{2\sigma^2}(x-\mu)^2\right) dx \\ &= \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x - (\mu + \sigma^2 t)^2)\right) dx \\ &= \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right), \ t \in \mathbb{R} \end{aligned}$$

Teorema 4.8. *Sea $X \sim N(\mu, \sigma^2)$. Entonces*

1. $Y = a + bX \sim N(a + b\mu, b^2\sigma^2)$ $Z = (X - \mu)/\sigma \sim N(0, 1)$

Dem. Basta notar que

$$\begin{aligned} M_Y(t) &= \exp(at)M_X(bt) = \exp(at)\exp(\mu bt + \frac{1}{2}\sigma^2 b^2 t^2) \\ &= \exp\left((a + b\mu)t + \frac{1}{2}(\sigma b)^2 t^2\right) \end{aligned}$$

que corresponde a la fgm de una distribución normal con media $a + b\mu$ y varianza $\sigma^2 b^2$. Luego $Y \sim N(a + b\mu, b^2\sigma^2)$.

Ejemeplo 4.11. Sea $X \sim N(1, 4)$. Calculemos $P(0 \leq X < 1)$ y $P(X^2 > 4)$, entonces

$$\begin{aligned} P(0 \leq X < 1) &= P\left(-\frac{1}{2} \leq \frac{X-1}{2} < 0\right) \\ &= \Phi(0) - \Phi\left(-\frac{1}{2}\right) \\ &= 0{,}5 - 0{,}30854 \\ &= 0{,}19146 \end{aligned}$$

Similarmente

$$\begin{aligned}
P(X^2 > 4) &= 1 - P(|X| \leq 2) \\
&= 1 - P(-2 \leq X \leq 2) \\
&= 1 - P\left(-\frac{3}{2} \leq \frac{X-1}{2} \leq \frac{1}{2}\right) \\
&= 1 - \left(\Phi\left(\frac{1}{2}\right) - \Phi\left(-\frac{3}{2}\right)\right) \\
&= 1 - (0{,}69146 - 0{,}06681) \\
&= 0{,}37535
\end{aligned}$$

Ejemplo 4.12. Sean X_1, X_2 las duraciones de dos diapositivos electrónicos. Asuma que $X_1 \sim N(40, 36)$ y $X_2 \sim N(45, 9)$. Si el dispositivo se va a utilizar durante 45 horas, ¿cuál dispositivo sería el preferido?. Y si se va utilizar durante 42 horas, ¿cuál debería preferise ?. Para determinar cada dispositivos debemos determinar la mayor probabilidad de vida útil de más de 45 horas:

$$\begin{aligned}
P(X_1 > 45) &= P\left(Z > \frac{45-40}{6}\right) = 1 - \Phi\left(\frac{5}{6}\right) \\
&= 1 - 0{,}7995 \\
&= 0{,}2005
\end{aligned}$$

y

$$\begin{aligned}
P(X_2 > 45) &= P\left(Z > \frac{45-45}{3}\right) = 1 - \Phi(0) \\
&= 1 - 0{,}5 = 0{,}5
\end{aligned}$$

Por lo tanto, se debe preferir el despositivo X_2. Ahora hay que encontrar qué dispositivo tiene mayor probabilidad de vida útil de más de 42 horas. Los cálculo son similares:

$$\begin{aligned}
P(X_1 > 42) &= 1 - \Phi\left(\frac{1}{3}\right) \\
&= 0{,}3707 \\
P(X_2 > 42) &= 1 - \Phi(-1) \\
&= 0{,}8413
\end{aligned}$$

En este caso también se prefiere el dispositivo X_2.

4.2.3. Distribución Gamma

La distribución gamma se usa de manera extensa en una variedad de áreas, como por ejemplo: descrbir los intervalos de tiempo entre dos fallas consecutivas en el motor de un avión, o los intervalos de tiempo entre las llegadas de clientes a una cola en el cajero de un supermercado.

Definición 4.9. *Se dice que la variable aleatoria X tiene distribución gamma con los parámetros $\alpha > 0$ y $\lambda > 0$ si su fdp está dada por:*

$$f(x) = \begin{cases} \frac{\lambda}{\Gamma(\alpha)}(\lambda x)^{\alpha-1}\exp(-\lambda x), & x > 0 \\ 0, & eoc. \end{cases}$$

Recordemos que la función gamma se define por:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}\exp(-t)dt$$

donde $\Gamma(\alpha+1) = \alpha\Gamma(\alpha), \Gamma(1/2) = \sqrt{\pi}$ y $\Gamma(n) = (n-1)!$ si $n = 1, 2, \ldots$.

Notación. $X \sim Gamma(\alpha, \lambda)$

Insertar Gráfico

Teorema 4.10. *Si $X \sim Gamma(\alpha, \lambda)$, entonces*

1. $E(X) = \alpha/\lambda$.
2. $Var(X) = \alpha/\lambda^2$.
3. $M(t) = (1 - t/\lambda)^{-\alpha}$, *si* $t < \lambda$.

Dem. Calculemos el fgm de X

$$\begin{aligned} M(t) &= \int_0^\infty e^{tx}\frac{\lambda}{\Gamma(r)}(\lambda x)^{\alpha-1}e^{-\lambda x}dx \\ &= \left(\frac{\lambda}{\lambda - t}\right)^\alpha \int_0^\infty \frac{(\lambda-t)^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-(\lambda-t)x}dx \\ &= \left(1 - \frac{t}{\lambda}\right)^{-\alpha}, \ \text{si}\lambda - t > 0 \end{aligned}$$

Luego

$$\begin{aligned} E(X) &= \frac{d}{dt}M(t)\bigg|_{t=0} = \frac{\alpha}{\lambda} \\ E(X^2) &= \frac{d^2}{dt^2}M(t)\bigg|_{t=0} = \frac{\alpha^2+\alpha}{\lambda^2} \end{aligned}$$

de donde sigue que $\text{Var}(X) = \alpha/\lambda^2$.

Casos Particulares

1. Si $\alpha = 1$, entonces $X \sim Exp(\lambda)$, a la cual llamamos distribución exponencial. En tal caso

$$E(X) = \frac{1}{\lambda}, \ \text{Var}(X) = \frac{1}{\lambda^2}, \ M(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}, \ t < \lambda$$

2. Si $\alpha = \frac{\nu}{2}$ y $\lambda = \frac{1}{2}$, entonces $X \sim \chi^2_\nu$, a la cual llamamos distribución chi-cuadrado con ν grados de libertad. En este caso

$$E(X) = \nu,\ \ \mathrm{Var}(X) = 2\nu,\ \ M(t) = (1-2t)^{-\nu/2},\ t < 1/2$$

Si $\alpha > 1$, entonces $X \sim Erlag(\alpha, \lambda)$, distribución de Erlang.

5. Vectores Aleatorios

5.1. Introducción y Definición

Al realizar un experimento aleatorio, podemos considerar dos o más características, por ejemplo, sacar una pieza manufacurada de acero y ver la dureza (X) y la tensión (Y), o bien, en una persona, observar la altura (X) y el peso (Y). En ambos casos se puede asociar $(x, y) = (X(\omega), Y(\omega))$.

Con esto extendemos el concepto de variable aleatoria unidimensional a multidimensional. Aun así, todo se construye en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$.

Definición 5.1. *Un vector aleatorio n-dimensional es una función $(X_1, \ldots, X_n)$ desde el espacio muestral Ω en $\mathbb{R}^n$, el espacio Euclidiano n-dimensional:*

$$\begin{aligned}(X_1, \ldots, X_n) : \Omega &\longrightarrow \mathbb{R}^n \\ \omega &\longmapsto (X_1(\omega), \ldots X_n(\omega))\end{aligned}$$

Es decir, $(X_1, \ldots, X_n)$ es un vector aleatorio en $(\Omega, \mathcal{A}, P)$ si y sólo si $X_1, \ldots, X_n$ son variables aleatorias en $(\Omega, \mathcal{A}, P)$

Ejemplo 5.1. Para n lanzamientos consecutivos de una moneda, sea

$$X_i := \begin{cases} 1, & \text{si el } i\text{-ésimo lanzamiento dio cara.} \\ 0, & eoc. \end{cases}$$

Para $i = 1, \ldots, n$. Entones $(X_1, \ldots, X_n)$ es un vector aleatorio en $(\Omega, \mathcal{A}, P)$, donde

$$\Omega = \{c, s\}^n, \ \ \mathcal{A} = 2^\Omega, \ \ P(\{c\}) = p$$

con $0 < p < 1$. Cosa que hay que tener en cuenta, es que Ω puede ser perféctamente $\{c, s\}$ y estar bien definido el vector aleatorio.

En este ejemplo, para cada $\omega = (\omega_1, \ldots, \omega_n) \in \Omega = \{c, s\}^n$, se tiene que

$$X_i(\omega) = \begin{cases} 1, & \omega_i = c \\ 0, & \omega_i = s \end{cases}$$

$i = 1, \ldots, n$. Luego, $(X_1(\omega), \ldots, X_n(\omega)) \in \mathcal{X} = \{0, 1\}^n \subseteq \mathbb{R}^n$. Por ejemplo, si $n = 2$ se construye la siguiente tabla

$\omega = (\omega_1, \omega_2) \in \{c, s\}^2$	$(X_1(\omega), X_2(\omega)) \in \{0, 1\}^2$
(s, s)	$(0, 0)$
(s, c)	$(0, 1)$
(c, s)	$(1, 0)$
(c, c)	$(1, 1)$

5.2. Distribución Conjuntas

5.2.1. Caso general

Sea $(X_1, \dots, X_n)$ un vector aleatorio en $(\Omega, \mathcal{A}, P)$. Entonces, para todo $B \in \mathcal{B}^n = \mathcal{B}(\mathbb{R}^n)$, se tiene que

$$\{(X_1, \dots, X_n) \in B\} = \{\omega \in \Omega : (X_1(\omega), \dots, X_n(\omega)) \in B\} \in \mathcal{A}$$

siempre y cuando $\mathcal{A}$ está bien definido. Es decir, es un evento y por lo tanto

$$P_{X_1,\dots,X_n}(B) = P\{(X_1, \dots, X_n) \in B\}, \;\; B \in \mathcal{B}^n$$

define la distribución de probabilidad del vector aleatorio $(X_1, \dots, X_n)$ o distribución de probabilidad conjunta de las variables aleatorias $X_1, \dots, X_n$.

De esta forma, si $\mathcal{X} \subseteq \mathbb{R}^n$ es el recorrido conjunto de $(X_1, \dots, X_n)$, entonces

$$(\mathcal{X}, \mathcal{B}^n, P_{X_1,\dots,X_n})$$

define un modelo o espacio de probabilidad multivariado. Esto es claro ya que por definición, la función $P_{X_1,\dots,X_n}$ está asociado a P por lo que no es negativa, si evaluamos en $\mathcal{X}$, por definición se obtiene

$$P_{X_1,\dots,X_n}(\mathcal{X}) = P(\Omega) = 1$$

Y por último, dada una colección disjuntas de intervalos, claramente obtendremos conjuntos disjuntos evaluado en P, y entonces, obtenemos

$$P_{X_1,\dots,X_n}\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P_{X_1,\dots,X_n}(B_i)$$

donde $\{B_i\}_{i \in \mathbb{N}}$ es la colección disjunta.

En particular, para la región $B = (-\infty, x_1] \times \cdots \times (-\infty, x_n]$ se tiene

$$\begin{aligned}
\{(X_1, \dots, X_n) \in B\} &= \{\omega \in \Omega : (X_1(\omega), \dots, X_n(\omega)) \in B\} \\
&= \{\omega \in \Omega : X_1(\omega) \le x_1, \dots, X_n(\omega) \le x_n\} \\
&= \{X_1 \le x_1, \dots, X_n \le x_n\} \in \mathcal{A}, \; \forall (x_1, \dots, x_n) \in \mathbb{R}^n
\end{aligned}$$

Entonces, se puede definir la función acumulada para n-dimensiones. En particular, la función definida por:

$$\begin{aligned} P_{X_1,\ldots,X_n}(B) &= P(X_1 \leq x_1, \ldots, X_n \leq x_n) \\ &= F_{X_1,\ldots,X_n}(x_1, \ldots, x_n), \ (x_1, \ldots, x_n) \in \mathbb{R}^n \end{aligned}$$

se llama función de distribución (acumulada) (fda) del vector aleatorio $(X_1, \ldots, X_n)$ o distribución (acumulada) conjunta de las variables aleatorias $X_1, \ldots, X_n$.

Definición 5.2. (Distribución de probabilidad conjunta) *La distribución de probabilidad conjunta de $X_1, \ldots, X_n$ es la colección de probabilidades:*

$$P\{(X_1, \ldots, X_n) \in B\}$$

Para todos los subconjuntos B de $\mathbb{R}^n$.

Determinar el valor de una probabilidad conjunta depende que lo que se pide, y de como es la distribución de la variables aleatorias como hemos visto.

Definición 5.3. (Función de distribución conjunta) *La función de distribución (acumulada) conjunta de $X_1, \ldots, X_n$, se define como:*

$$F : \mathbb{R}^n \to \mathbb{R}$$

Dada por:

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) := P_{X_1,\ldots,X_n}(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

En una dimensión sabemos que la función distribución acumulada es tomar toda la parte izquierda de un valor x_0, y asociarle a un valor real, en varias dimensiones es tomar n-cubos y asociarlo a un valor real. Notemos que

$$F_{X_1,\ldots,X_n} : \mathbb{R}^n \longrightarrow [0, 1]$$

Ya que

$$P_{X_1,\ldots,X_n}(\Omega) = P(\{\omega : (X_1(\omega), \ldots, X_n(\omega)) \in \Omega\} = P(\Omega) = 1$$

Luego se toma el complemento de Ω y luego la acumulada alcanza tanto el 0 como el 1.

5.2.2. Caso bidimensional

Estudiemos el caso bivariable. Sea $n = 2$, con $X_1 = X$ y $X_2 = Y$ variables aleatorias, entonces la función:

$$\begin{aligned} F_{X,Y}(x, y) &= P(X \leq x, Y \leq y) \\ &= P_{X,Y}\{(-\infty, x] \times (-\infty, y]\}, \ (x, y) \in \mathbb{R}^2 \end{aligned}$$

es la función de distribución (acumulada) conjunta de las variables aleatorias X e Y o simplemente la función de distribución (acumulada) del vector aleatoria (X,Y). $F_{X,Y}(x,y)$ es la probabilidad de que (X,Y) tome algún valor en la región $(-\infty,x]\times(-\infty,y]$.

Como en el caso unidimensional, la función distribución conjunta o acumulada, también tiene propiedades importante de como se comportan $F_{X,Y}$.

Teorema 5.1. *Sea $F_{X,Y}(x,y)$ la función de distribución conjunta de X e Y. Entonces:*

(a) $\lim_{x,y\to\infty} F_{X,Y}=1$. *Necesariamente x,y convergen al infinito.*

(b) $\lim_{x\to-\infty} F_{X,Y}(x,y)=\lim_{y\to-\infty} F_{X,Y}(x,y)=0$, *para cada valor valor fijo del otro argumento.*

(c) $F_{X,Y}(x,y)$ *es no decreciente en cada uno de sus argumentos. Es decir*

$$F_{X,Y}(x,y)\le F_{X,Y}(x+h,y)$$
$$F_{X,Y}(x,y)\le F_{X,Y}(x,y+h)$$

para todo $h\ge 0$.

(d) $F_{X,Y}(x,y)$ *es continua por la derecha en cada uno de sus argumentos. Es decir*

$$\lim_{h\to0^+} F_{X,Y}(x+h,y)=\lim_{h\to0^+} F_{X,Y}(x,y+h)=F_{X,Y}(x,y)$$

(e) Si $a_1<b_1$ y $a_2<b_2$, entonces

$$\underbrace{F_{X,Y}(b_1,b_2)-F_{X,Y}(a_1,b_2)-F_{X,Y}(b_1,a_2)+F_{X,Y}(a_1,a_2)}_{P_{X,Y}\{(a_1,b_1]\times(a_2,b_2]\}=P(a_1<X\le b_1,a_2<Y\le b_2)}\ge 0$$

Observación 5.1. Al igual que en una variable, si una función bivariada $F:\mathbb{R}^2\to\mathbb{R}$ que satisface las propiedades del (a) al (e), entonces está asociada a un vector aleatorio (X,Y) como una función distribución conjunta. Es más, esto también se cumple para el caso n-dimensional con leves diferencias, en el caso (e) es solo seguir la tendencia con n variabesl, la cual no mostraremos.

Observación 5.2. La condición (e) asigna probabilidades no negativas a todos los rectángulos de $\mathbb{R}^n$ de la forma

$$(a_1,b_1]\times\cdots\times(a_n,b_n]$$

Por lo que en el caso n-dimensiona, es tomar un cualquier n-cubo y ver que tiene valor mayor o igual que 0.

Ejemplo 5.2. Sea

$$F_{X,Y}(x,y)=\begin{cases}(1-e^{-x})(1-e^{-y}), & x,y\ge 0\\ 0, & eoc.\end{cases}$$

Probemos que es una función distribución conjunta de X,Y.

(a) Notemos que $F_{X,Y}$ es continua ya que e^{-x} es continua, luego si $x, y \to \infty$, entonces

$$\lim_{x,y\to\infty} F_{X,Y}(x,y) = 1 \cdot 1 = 1$$

(b) Sea y fijo, entonces

$$\lim_{x\to-\infty} F_{X,Y}(x,y) = (1-e^0)(1-e^{-y}) = 0$$

Lo mismo si se fija x. Si $y < 0$ de forma directa $F_{X,Y}(x,y) = 0$, lo mismo si $x < 0$.

(c) Sea $h > 0$, entonces

$$x \leq x+h \Leftrightarrow e^{-x} \geq e^{-x-h} \Leftrightarrow 1-e^{-x} \leq 1-e^{-x-h}$$

Si $1-e^{-y}$ es no negativo para todo $y \geq 0$, entonces

$$F_{X,Y}(x,y) \leq F_{X,Y}(x+h,y)$$

Lo mismo se concluye para el segundo argumento.

(d) Como es continua, es claramente continua por la derecha en cada argumento.

(e) Sean $b_2 > a_2, b_1 > a_1$, entonces

$$\begin{aligned} &F_{X,Y}(b_1,b_2) - F_{X,Y}(a_1,b_2) - F_{X,Y}(b_1,a_2) + F_{X,Y}(a_1,a_2) \\ &= (e^{-b_1} - e^{-a_1})(e^{-b_2} - e^{-a_2}) \geq 0 \end{aligned}$$

Cumpliendo todas las propiedades de función distribución conjunta, es decir, F una función distribución conjunta. Ahora determinemos algunas probabilidades

$$\begin{aligned} P(0 < X \leq 1, 0 < Y \leq 1) &= F_{X,Y}(1,1) - F_{X,Y}(1,0) - F_{X,Y}(0,1) + F_{X,Y}(0,0) \\ &= (1-e^{-1})(1-e^{-1}) = (1-e^{-1})^2 \geq 0 \end{aligned}$$

Ejemplo 5.3. Sea

$$F(x,y) = \begin{cases} 1, & x \geq 0, y \geq 0, x+y \geq 1 \\ 0, & eoc. \end{cases}$$

Veamos si es una función distribución conjunta.

(a) Si x, y son suficientemente grande, entonces $F = 1$, por lo que

$$\lim_{x,y\to\infty} F(x,y) = 1$$

(b) Sea y fijo, entonces basta tomar $x < 0$, de forma que $F = 0$, lo mismo con x fijo.

(c) Es creciente por cada argumento.

(d) Es continua por la derecha.

(e) Notemos que

$$\begin{aligned} 0 \leq P(0 < X \leq 1, 0 < Y \leq 1) &= F(1,1) - F(1,0) - F(0,1) + F(0,0) \\ &= -1 \end{aligned}$$

Una contradicción

Por lo tanto F no puede ser una distribución conjunta bivariada, ya que no cumple la última propiedad.

5.3. Vectores Aleatorios Continuos y Discretos

Al igual que el caso univariado, hay varios tipos de vectores aleatorios, de acuerdo con la naturaleza de su coordenadas. A continuación, se consideran solo dos tipos, los vectores aleatorios discretos y continuos.

5.3.1. Caso Discreto

Definición 5.4. (Vector aleatorio discreto) *Un vector aleatorio $(X_1, \ldots, X_n)$ se dice discreto si su recorrido, $\mathcal{X}$, es un subconjunto contable (finito o infinito) de $\mathbb{R}^n$, es decir, $(X_1, \ldots, X_n)$ es discreto si y sólo las coordenadas $X_1, \ldots, X_n$ son variables aleatorias discretas.*

En tal caso, la distribución de probabilidad conjunta de $X_1, \ldots, X_n$ queda completamente determinada por la función de masa de probabilidad (fmp) conjunta definida por:

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n),\ (x_1, \ldots, x_n) \in \mathbb{R}^n$$

De este modo se tiene que

(a) $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \geq 0$ para todo $(x_1, \ldots, x_n)$.

(b) $\sum \cdots \sum_{\{(x_1,\ldots,x_n) \in \mathbb{R}^n\}} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = 1$.

(c) $P_{X_1,\ldots,X_n}(B) = \sum \cdots \sum_{\{(x_1,\ldots,x_n) \in B\}} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ para todo subconjunto B de $\mathbb{R}^n$.

Siendo las propiedades de la función de masa de probabilidad.

Ejemplo 5.4. Al lanza una moneda justa tres veces consecutivas, defina las siguientes variables aleatorias

$$\begin{aligned} X &= \text{Número de caras obtenidad en los primeros dos lanzamientos} \\ Y &= \text{Número de caras obtenidad en los primeros dos lanzamientos} \end{aligned}$$

Claramente (X, Y) es un vector aleatorio discreto como X, Y lo son. Con recorrido

$$\mathcal{X} = \{(0,0), (0,1), (1,0), (1,1), (1,2), (2,1), (2,2)\}$$

Entonces el fmp de la conjuntes cumple

$$\begin{aligned}
f_{X,Y}(0,0) &= P(X=0,Y=0) = P(S,S,S) = \frac{1}{8}\\
f_{X,Y}(0,1) &= P(X=0,Y=1) = P(S,S,C) = \frac{1}{8}\\
f_{X,Y}(1,0) &= P(X=1,Y=0) = P(C,S,S) = \frac{1}{8}\\
f_{X,Y}(1,1) &= P(X=1,Y=1) = P((C,S,C),(S,C,S)) = \frac{2}{8}\\
f_{X,Y}(1,2) &= P(X=1,Y=2) = P(S,C,C) = \frac{1}{8}\\
f_{X,Y}(2,1) &= P(X=2,Y=1) = P(C,C,S) = \frac{1}{8}\\
f_{X,Y}(2,2) &= P(X=2,Y=2) = P(C,C,C) = \frac{1}{8}
\end{aligned}$$

Y $f_{X,Y} = 0$ para todo $(x,y) \notin \mathcal{X}$.

El fmp de (X,Y) se puede resumir en una tabla, como el de la siguiente Esta tabla se le conoce

$X \setminus Y$	0	1	2
0	$\frac{1}{8}$	$\frac{1}{8}$	0
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
2	0	$\frac{1}{8}$	$\frac{1}{8}$

como tabla de contingencia.

Para determinar la función distribución acumada, basta considerar

$$B = \{(y_1,\ldots,y_n) \in \mathbb{R}^n : y_1 \leq x_1,\ldots,y_n \leq x_n\}$$

Luego en el caso discreto:

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \sum_{y_1=k_1}^{x_1} \cdots \sum_{y_n=k_n}^{x_n} f_{X_1,\ldots,X_n}(y_1,\ldots,y_n)$$

Si lo pensamos en una tabla es más fácil de determinar. Por ejemplo, de forma directa, del ejemplo 5.4, para determinar la acumulada de $(1,1)$ basta sumar

$$F_{X,Y}(1,1) = P(X \leq 1, Y \leq 1) = \sum_{x=0}^{1}\sum_{y=0}^{1} f_X, Y(x,y) = \frac{5}{8}$$

Pero para la tabla, basta sumar el cuadrado que se genera al tomar $x=1, y=1$. Que es $\frac{5}{8}$.

5.3.2. Caso Continuo

Definición 5.5. (Vector aleatorio continuo) *Un vector aleatorio $(X_1, \ldots, X_n)$ se dice (absolutamente) continuo si existe una función no negativa*

$$f_{X_1,\ldots,X_n} : \mathbb{R}^n \to \mathbb{R}$$

talque

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1,\ldots,X_n}(u_1, \ldots, u_n) du_n \ldots du_1$$

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

En tal caso

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \frac{\partial^n F_{x_1,\ldots,x_n}(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}$$

y se llama función densidad de probabilidad (fdp) conjunta de $X_1, \ldots, X_n$. La fdp satisface

(a) $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \geq 0$ para todo $(x_1, \ldots, x_n)$.

(b) $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n = 1$.

(c) $P_{X_1,\ldots,X_n}(B) = \underset{B}{\int \cdots \int} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_n \ldots dx_1$. Para cada subconjunto B de $\mathbb{R}^n$.

Con respecto a la definición de una función distribución conjunta sobre un vector continuo, consideramos un caso muy conveniente, donde se puede integrar de forma independientes, todo porque en general se toma $f_{X_1,\ldots,X_n}$ continua. La forma general es que si $B = \{(y_1, \ldots, y_n) \in \mathbb{R}^n : y_1 \leq x_1, \ldots, y_n \leq x_n\}$, entonces

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \underset{B}{\int \cdots \int} f_{X_1,\ldots,X_n}(u_1, \ldots, u_n) du_1 \ldots du_n$$

Nota 5.1.

- Si $(X_1, \ldots, X_n)$ es un vector (absolutmente) continuo, entonces $X_1, \ldots, X_n$ son variables aleatorias (absolutamente).
- Sin embargo, el que $X_1, \ldots, X_n$ sean variables aleatorias (absolutamente) continuas no implica que el vector aleatorio $(X_1, \ldots, X_n)$ sea (absolutamente) continuo.

Por ejemplo, en el caso bivariable, si $n = 2$ con $X_1 = X, X_2 = Y$, se tiene que

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) dv du$$

y luego, por el teorema fundamental del cálculo,

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

en los puntos de continuidad de $f_{X,Y}(x,y)$.

Ejemplo 5.5. Sean X, Y variables aleatorias con fdp conjunta dada por:

$$f_{X,Y}(x,y) = \begin{cases} 6xy^2, & 0 < x < 1,\ \ 0 < y < 1 \\ 0, & eoc. \end{cases}$$

Notemos que al se la fdp continuo del vector aleatorio, está asociado a una función distribución $F_{X,Y}$ dada por

$$\begin{aligned} F(x,y) := F_{X,Y}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} 6uv^2 dvdu \\ &= \int_{0}^{x}\int_{0}^{y} 6uv^2 dvdu \\ &= \int_{0}^{x} 6u\frac{y^3}{3}du \\ &= \frac{y^3}{3}\frac{6x^2}{2} = x^2y^3 \end{aligned}$$

para todo $0 < x < 1, 0 < y < 1$. Entonces

$$\begin{aligned} P\left(\frac{1}{2} < X \le \frac{3}{4}, 0 \le Y \le \frac{1}{3}\right) &= F\left(\frac{3}{4},\frac{1}{3}\right) - F\left(\frac{1}{2},\frac{1}{3}\right) - F\left(\frac{3}{4},0\right) + F\left(\frac{1}{2},0\right) \\ &= \frac{9}{16}\frac{1}{27} - \frac{1}{4}\frac{1}{27} = \frac{1}{48} - \frac{1}{108} = 1{,}1574 \times 10^{-2} \end{aligned}$$

Tambien se puede determinar de forma directa estudiando

$$\int_{0}^{\frac{1}{3}}\int_{\frac{1}{2}}^{\frac{3}{4}} 6xy^2 dxdy$$

Supongamos que nos interesa solo encontrar la probabilidad de X entre $\frac{1}{2}, \frac{3}{4}$, entonces decimos que Y recorre de 0 a 1, es decir

$$P\left(\frac{1}{2} < X \le \frac{3}{4}\right) = \int_{\frac{1}{2}}^{\frac{3}{4}}\int_{0}^{1} 6xy^2 dydx = x^2\Big|_{\frac{1}{2}}^{\frac{3}{4}} = \frac{9}{16} - \frac{1}{4} = \frac{5}{16} = 0{,}3125$$

Por último, que los valores de X e Y dependan de entre sí. Queremos determinar la probabilidad $P(X + Y \ge 1)$, para ello debemos intregar sobre la región $B = \{(x,y) : x + y \ge 1, 0 < x < 1, 0 < y < 1\}$. Notemos que

$$\begin{aligned} B &= \{(x,y) : x \ge 1 - y, 0 < x < 1, 0 < y < 1\} \\ &= \{(x,y) : 1 - y \le x < 1, 0 < y < 1\} \end{aligned}$$

Luego

$$\begin{aligned} P(X + Y \ge 1) &= \iint_B 6xy^2 dxdy = \int_{0}^{1}\int_{1-y}^{1} 6xy^2 dxdy \\ &= \int_{0}^{1} 6y^2\left(\frac{1}{2} - \frac{(1-y)^2}{2}\right)dy = \frac{9}{10} \end{aligned}$$

Nota 5.2. Si (X, Y) es un vector aleatorio continuo, entonces la probabilidad de que ocurra el evento

$$\{a \leq X \leq b, c \leq Y \leq d\}$$

no cambia si no se considera los extremos, es decir

$$P(a \leq X \leq b, c \leq Y \leq d) = P(a < X \leq b, c \leq Y \leq d) = \dots$$

Y así con el resto de casos.

5.4. Distribuciones Marginales e Independencia

5.4.1. Fda, Fmp, Fdp Marginales

Sea la distribución conjunta $X_1, \dots, X_n$, de aquí se puede obtener cualquier subvector de $(X_1, \dots, X_n)$ y en particular las distribuciones marginales de cada una de sus coordenadas.

Por ejemplo, el subvector $(X_1, \dots, X_k)$ que es tomar las primeras k componentes del vector aleatorio. El fda se puede determinar por:

$$F_{X_1,\dots,X_k}(x_1, \dots, x_k) = F_{X_1,\dots,X_k,X_{k+1},\dots,X_n}(x_1, \dots, x_k, \infty, \dots, \infty)$$

para todo $(x_1, \dots, x_k) \in \mathbb{R}^k$. Donde

$$F_{X_1,\dots,X_k,X_{k+1},\dots,X_n}(x_1, \dots, x_k, \infty, \dots, \infty) = \lim_{x_{k+1},\dots,x_n \to \infty} F_{X_1,\dots,X_n}(x_1, \dots, x_n)$$

De la misma forma se obtiene cualquier fda de cualquier subvector del vector aleatorio principal.

Definición 5.6. (Marginal) *Sea el vector aleatorio $(X_1, \dots, X_n)$. Se define la marginal X_i, como el subvector (X_i) con $i = 1, \dots, n$ donde su fda está dado por:*

$$F_{X_i}(x_i) = F_{X_1,\dots,X_n}(\infty, \dots, \infty, x_i, \infty, \dots, \infty)$$

para todo $x_i \in \mathbb{R}$.

De igual forma se puede definir los fmp o fdp conjunta de un subvector $(X_1, \dots, X_k)$, la cual se determina por:

$$\begin{aligned} &f_{X_1,\dots,X_k}(x_1, \dots, x_k) \\ &\quad = \begin{cases} \sum_{x_{k+1}\in\mathbb{R}} \cdots \sum_{x_n\in\mathbb{R}} f_{X_1,\dots,X_n}(x_1, \dots, x_k, x_{k+1}, \dots, x_n), & \text{c.d.} \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\dots,X_n}(x_1, \dots, x_n) dx_{k+1} \dots dx_n, & \text{c.c.} \end{cases} \end{aligned}$$

Donde c.d es el caso discreto y c.c es el caso continuo. Dicho de otra forma, para obtener el fmp o fdp de un subvector, debemos rellenar los espacios que no nos sirve.

De igual forma se obtiene el fmp o fdp para cualquier subvector, entonces la marginal fmp o fdp de X_1 se determina por:

$$f_{X_1}(x_1) = \begin{cases} \sum_{x_2\in\mathbb{R}} \cdots \sum_{x_n\in\mathbb{R}} f_{X_1,\dots,X_n}(x_1, \dots, x_k, x_{k+1}, \dots, x_n), & \text{c.d.} \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\dots,X_n}(x_1, \dots, x_n) dx_2 \dots dx_n, & \text{c.c.} \end{cases}$$

De forma análoga se obtiene las fmp, fdp marginales de $X_2, \ldots, X_n$. La fda marginal de X_1 sería

$$F_{X_1}(x_1)) = \begin{cases} \sum_{i \leq x_1} \left(\sum_{x_2 \in \mathbb{R}} \cdots \sum_{x_n \in \mathbb{R}} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \right), & \text{c.d.} \\ \int_{-\infty}^{x_1} \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(u, \ldots, x_n) dx_n \ldots dx_2 \right) du, & \text{c.c.} \end{cases}$$

5.4.2. Independencia

La independencia es una propiedad importante de la fda conjunta que está relacionada con los fda's marginales.

Definición 5.7. (Independencia fda) *Sea $(X_1, \ldots, X_n)$ un vector aleatorio (arbitrario) con fda conjunta $F_{X_1,\ldots,X_n}$ y fda's marginales $F_{X_1,\ldots,X_n}$ respéctivamente. Entonces las variables aleatorias $X_1, \ldots, X_n$ se dicen (mutuamente) independiente si y sólo si:*

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

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

La independencia por fda's es la definición general de independencia. Si el vector aleatorio es discreto o continuo, entonces podemos definir de forma equivalente con los fmp's o fdp's conjunta y marginales.

Definición 5.8. (Independencia fdp, fmp) *Si $(X_1, \ldots, X_n)$ es un vector aleatorio (discreto o continuo) con fmp conjunta o fdp conjunta $f_{X_1,\ldots,X_n}$, y fmp's marginales o fdp's marginales $f_{X_1}, \ldots, f_{X_n}$, respéctivamente. Entonces, las variables aleatorias $X_1, \ldots, X_n$ se dicen (mutuamente) independientes si y sólo si*

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

para todo $(x_1, \ldots, x_n) \in \mathbb{R}^n$

Nota 5.3.

- Sean $\mathcal{X}, \mathcal{X}_1, \ldots, \mathcal{X}_n$ los recorridos del vector y de las marginales $X_1, \ldots, X_n$ respéctivamente. Entonces una condición necesaria (pero no suficiente), es que para que $X_1, \ldots, X_n$ sean (mutuamente) independientes, necesariamente:

$$\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$$

- Si $X_1, \ldots, X_n$ son variables (mutuamente) independientes, entonces todos los subvectores son independiente. Esto es evidente, si $(X_1, \ldots, X_k)$ es un subvector, entonces es independiente por definición por parte de la fda, ya que

$$\begin{aligned}
F_{X_1,\ldots,X_k} &= \lim_{x_{k+1},\ldots,x_n\to\infty} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\\
&= \lim_{x_{k+1},\ldots,x_n\to\infty} F_{X_1}(x_1)\ldots F_{X_n}(x_n)\\
&= F_{X_1}(x_1)\ldots F_{X_k}(x_k)F_{X_{k+1}}(\infty)\ldots F_{X_n}(\infty)\\
&= \prod_{i=1}^{k} F_{X_i}(x_i)
\end{aligned}$$

Notar que $F_{X_i}(\infty) = \text{lím}_{x_1,\ldots,x_n\to\infty} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = 1$.

- Si $X_1,\ldots,X_n$ son variables (absolutamente) continuas (mutuamente) independientes, entonces el vector aleatorio $(X_1,\ldots,X_n)$ es (absolutamente) continuo. Esto caracteriza un vector aleatorio, ahora podemos decir que si $X_1,\ldots,X_n$ son independientes, entonces

$$(X_1,\ldots,X_n) \text{ es v.a continuo si y sólo si } X_1,\ldots,X_n \text{ son v.a's continuas.}$$

- Si $X_1,\ldots,X_n$ son variables independientes e idénticamente distribuidas (iid), es decir, $F_{X_i}(x) = F_{X_j}(x)$ para todo $x\in\mathbb{R}$, para todo $i=1,\ldots,n$ (fmp/fdp f). Escribimos

$$X_1,\ldots,X_n \overset{\text{iid}}{\sim} F \text{ o } X_1,\ldots,X_n \overset{\text{iid}}{\sim} f$$

También es importante recordad de que si X_i tiene fdp o fmp o fgm igual a una distribución conocidad, entonces por unicidad se tiene que X_i se distribuye en la distribución conocida.

5.4.3. Caso Bivariado

Para reforzar los conceptos previos, estudiaremos el caso de dos dimensión. Sea $X_1 = X, X_2 = Y$ y sea (X,Y) un vector aleatorio con fmp conjunta o fdp conjunta $f_{X,Y}(x,y)$, entonces

(a) La distribución de probabilida de (X,Y) es

$$\begin{aligned}
P_{X,Y}(B) &= P\{(X,Y)\in B\}\\
&= \begin{cases} \sum_{(x,y)\in B} f_{X,Y}(x,y), & \text{Caso Discreto.}\\ \\ \int_{(x,y)\in B} f_{X,Y}(x,y)dydx, & \text{Caso Continuo}\end{cases}
\end{aligned}$$

para todo $B\subseteq\mathbb{R}^2$.

(b) La fda de (X,Y) es

$$\begin{aligned}
F_{X,Y}(x,y) &= P(X\le x, Y\le y)\\
&= \begin{cases} \sum_{u\le x}\sum_{v\le y} f_{X,Y}(u,v), & \text{Caso Discreto.}\\ \\ \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)dvdy, & \text{Caso Continuo}\end{cases}
\end{aligned}$$

para todo $(x, y) \in \mathbb{R}^2$. Además, las fda's marginales de X e Y son respectívamente:

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x, y) := F_{X,Y}(x, \infty), \text{ para todo } x \in \mathbb{R}$$
$$F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x, y) := F_{X,Y}(\infty, y), \text{ para todo } y \in \mathbb{R}$$

(c) Si (X, Y) es discreto, las fmp marginales de X e Y son respectívamente:

$$\begin{aligned} f_X(x) = P(X = x) &= P(X = x, Y \in \mathbb{R}) \\ &= P(\{X = x\} \cap (\cup_{y\in\mathbb{R}}\{Y = y\}) \\ &= P(\cup_{y\in\mathbb{R}}\{X = x, Y = y\}) \\ &= \sum_{y\in\mathbb{R}} P(X = x, Y = y) \\ &= \sum_{y\in\mathbb{R}} f_{X,Y}(x, y) \end{aligned}$$

para todo $x \in \mathbb{R}$. Y

$$\begin{aligned} f_Y(y) = P(Y = y) &= P(X \in \mathbb{R}, Y = y) \\ &= P((\cup_{x\in\mathbb{R}}\{X = x\}) \cap \{Y = y\}) \\ &= P(\cup_{x\in\mathbb{R}}\{X = x, Y = y\}) \\ &= \sum_{x\in\mathbb{R}} P(X = x, Y = y) \\ &= \sum_{x\in\mathbb{R}} f_{X,Y}(x, y) \end{aligned}$$

Para todo $y \in \mathbb{R}$.

Teorema 5.2. *Sea (X, Y) un vector aleatorio bivariado discreto con fmp conjunta $f_{X,Y}(x, y)$. Entonces las fmp marginal de X e Y, $f_X(x) = P(X = x)$ y $f_Y(y) = P(Y = y)$, están dadas por:*

$$f_X(x) = \sum_{y\in\mathbb{R}} f_{X,Y}(x, y) \;\text{ y }\; f_Y(y) = \sum_{x\in\mathbb{R}} f_{X,Y}(x, y)$$

Dem. Basta con aplicar la observacion anterior. ■

(d) Si (X, Y) es un vector aleatorio continuo, entonces X e Y también son variables aleatorias continuas. Por lo que las fdp's marginales de X e Y se pueden definir. En particular, las fdp's marginales de X e Y están dados por:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy, \;\; x \in \mathbb{R}$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx, \;\; y \in \mathbb{R}$$

Como X e Y son variables aleatorias continuas, las fdp's marginales de X e Y, se pueden obtene de la derivada de las fda's marginales, es decir

$$f_X(x) = \frac{dF_X(x)}{dx} \text{ y } f_Y(y) = \frac{dF_y(y)}{dy}$$

(e) Finalmente, X e Y son variables aleatorias independientes si y sólo si

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

para todo $(x,y) \in \mathbb{R}^2$. Equivalentemente, si las variables aleatorias son discretas o continuas, ellas son independientes si y sólo si

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

para todo $(x,y) \in \mathbb{R}^2$.

Ejemplo 5.6. (Caso Discreto:) Sean A, B dos eventos de Ω. Sean las variables aleatorias

$$X = I_A := \begin{cases} 1, & \text{si ocurre } A \\ 0, & \text{si no} \end{cases} \quad \text{e} \quad Y = I_B := \begin{cases} 1, & \text{si ocurre } B \\ 0, & \text{si no} \end{cases}$$

Notemos que $\Omega \overset{(X,Y)}{\longrightarrow} \mathcal{X} = \{(0,0);(0,1);(1,0);(1,1)\}$. En particular $\mathcal{X}_1 = \mathcal{X}_2 = \{0,1\}$, luego

$$\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$$

Por lo que X, Y pueden ser independientes. Si

$$\begin{aligned} AB &\overset{(X,Y)}{\longrightarrow} (1,1) \\ A^cB &\overset{(X,Y)}{\longrightarrow} (0,1) \\ AB^c &\overset{(X,Y)}{\longrightarrow} (1,0) \\ A^cB^c &\overset{(X,Y)}{\longrightarrow} (0,0) \end{aligned}$$

Entonces la fmp conjunta de X e Y es

$$\begin{aligned} f_{X,Y}(x,y) &= P(X=x, Y=y) \\ &= \begin{cases} P(A^cB^c), & x=0, y=0 \\ P(A^cB), & x=0, y=1 \\ P(AB^c), & x=1, y=0 \\ P(AB), & x=1, y=1 \\ 0, & eoc. \end{cases} \end{aligned}$$

La fmp conjunta $f_{X,Y}(x,y) = P(X=x, Y=y)$ anterior se puede reescribir como

$x \setminus y$	0	1	$P(X=x)$
0	$P(A^cB^c)$	$P(A^cB)$	$P(A^c)=1-P(A)$
1	$P(AB^c)$	$P(AB)$	$P(A)$
$P(Y=y)$	$P(B^c)=1-P(B)$	$P(B)$	1

Donde las marginales se obtiene de fijar un argumento y sumar todos los casos posibles del otro argumento. Es decir

$$\begin{aligned} f_X(x) &= P(X=x) \\ &= \sum_{y\in\{0,1\}} P(X=x, Y=y) = \begin{cases} 1-P(A), & x=0 \\ P(A), & x=1 \\ 0, & eooc \end{cases} \\ f_Y(y) &= P(Y=y) \\ &= \sum_{x\in\{0,1\}} P(X=x, Y=y) = \begin{cases} 1-P(B), & y=0 \\ P(B), & y=1 \\ 0, & eooc \end{cases} \end{aligned}$$

En particular, si A y B son independientes, entonces

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

En tal caso se tiene que

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

para todo (x,y). Esto significa que X e Y son independientes.

Es decir, a partir de dos conjuntos A, B independientes donde X es asociado a A e Y es asociado a B, se puede determinar que las variables aleatorias X e Y son independiente

Aplicación.

1. Lanzar una moneda al aire dos veces consecutivas. Sean

$$\begin{aligned} A &= \text{sale cara en el primer lanzamiento.} \\ B &= \text{sale cara en el segundo lanzamiento.} \end{aligned}$$

   Entonces A y B son eventos independientes ya que

$$P(AB) = \frac{1}{4} = \frac{1}{2}\cdot\frac{1}{2} = P(A)P(B)$$

Luego $AB^c, A^cB.A^cB^c$ son independientes. En este caso

$$\begin{aligned} P(X=x,Y=y) &= P(X=x)P(Y=y) \\ &= \begin{cases} p^{x+y}(1-p)^{2-(x+y)} & x,y=0,1 \\ 0, & eoc. \end{cases} \end{aligned}$$

donde p es la probabilidad de obtener cara al lanzar la moneda. Aquí, se tiene que $X, Y \overset{\text{iid}}{\sim} Ber(p)$, es decir, con fmp $f(z) = p^z(1-p)^{1-z}I_{0,1}(z)$.

2. Sacar una carta al azar de una baraja de 52 cartas. Sean

$$\begin{aligned} A &= \text{sale rey} \\ B &= \text{sale corazón} \end{aligned}$$

Entonces A y B son eventos independientes. En este caso

$$\begin{aligned} P(X=x,Y=y) &= P(X=x)P(Y=y) \\ &= \begin{cases} \left(\frac{1}{13}\right)^x \left(1-\frac{1}{13}\right)^{1-x} \left(\frac{1}{4}\right)^y \left(1-\frac{1}{4}\right)^{1-y} & x,y=0,1 \\ 0, & eoc. \end{cases} \end{aligned}$$

Entonces X e Y son variables aleatorias independientes, con $X \sim Ber(1/13)$ e $Y \sim Ber(1/4)$.

3. Lanza un dado honesto dos veces consecutivas. Sean

$$\begin{aligned} A &= \text{la suma es impar} \\ B &= \text{la suma es par} \end{aligned}$$

entonces $P(A) = 1/2, P(B) = 1/2$. Entonces A y B no son eventos independientes. ($AB = \emptyset, A + B = \emptyset$, luego $B = A^c$), y entonces $X + Y = 1$ y $X \overset{d}{=} Y \sim Ber(1/2)$, pero X e Y no son independientes.

En este caso

$$P(X=x,Y=y) = \begin{cases} 1/2, & x=0,1, x+y=1 \\ 0, & eoc. \end{cases}$$

Ejemplo 5.7. (Caso Continuo) Sean X e Y variables aleatorias continuas con fda conjunta dada por

$$F_{X,Y}(x,y) = \begin{cases} (1-e^{-x})(1-e^{-y}), & x,y>0 \\ 0, & eoc. \end{cases}$$

Sigue que

$$F_X(x) = F_{X,Y}(x,\infty) = \begin{cases} 1-e^{-x}, & x>0 \\ 0, & eoc. \end{cases}$$

$$F_Y(y) = F_{X,Y}(\infty,y) = \begin{cases} 1-e^{-y}, & y>0 \\ 0, & eoc. \end{cases}$$

Y entonces

$$F_{X,Y}(x,y) = F_X(x)F_y(y)$$

para todo (x,y). De forma alternativa, como

$$f_{X,Y} = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} = \begin{cases} e^{-x}e^{-y}, & x,y>0 \\ 0, & eoc. \end{cases}$$

Y

$$f_X(x) = \frac{\partial F_X(x)}{\partial x} = \begin{cases} e^{-x}, \ x>0 \\ 0, & eoc. \end{cases}$$

$$f_Y(y) = \frac{\partial F_Y(y)}{\partial y} = \begin{cases} e^{-y}, \ y>0 \\ 0, & eoc. \end{cases}$$

En este caso, se tiene $X,Y \overset{\text{iid}}{\sim} \exp(1)$. Note, además, que $\mathcal{X} = \mathbb{R}_+^2 = \mathbb{R}_+ \times \mathbb{R}_+ = \mathcal{X}_1 \times \mathcal{X}_2$.

Ejemplo 5.8. Sea

$$f(x,y) = \begin{cases} c(|x|+|y|), & |x|+|y| \leq 1 \\ 0, & eoc. \end{cases}$$

(a) Notemos que es una piramide sin interior con punta en $(0,0,0)$.

(b) Determinemos la constante c, a priori para que $f(x,y)$ sea fdp bivariada, necesariamente

$$\iint_B f(x,y)dxdy = 1$$

Donde $B = \{(x,y) : |x|+|y| \leq 1\}$. Pero por la parte (a), sabemos que es una piramide, por lo que basta estudiar un cuarto de la región, ya que el resto de la región general el mismo valor, sea $A = \{(x,y) : x+y \leq 1, 0 \leq x \leq 1, 0 \leq y \leq 1\}$, donde

$$\iint_B c(x+y)dxdy = \frac{1}{4}$$

Como tomamos solo los x, y positivo. Entonces

$$\begin{aligned}\frac{1}{4} &= \int_0^1 \int_0^{1-x} c(x+y)dydx \\ &= \frac{c}{2}\int_0^1 (1-x^2)dx \\ &= \frac{c}{3}\end{aligned}$$

Por lo tanto $c = \frac{3}{4}$.

(c) Ahora, es claro que $P(A) = \frac{1}{4}$ como acabos de concluir.

(d) Determinemos las fdp marginales. Con respecto a la variables aleatoria X, en particular

$$f_X(x) = \begin{cases} \frac{3}{4}(1-x^2), & x \geq 0 \\ \frac{3}{4}(3x^2 - 4x + 1), & x < 0 \end{cases}$$

Es continua cuando $x > 0, x > 0$ y además

$$\lim_{x \to 0} f(x) = \frac{3}{4}$$

Siendo continua, y por tanto X es una variable aleatoria continua. Lo mismo se concluye de Y.

5.5. Funciones de un Vector Aleatorio I

Al igual que el caso unidimensiona, podemos definir funciones para generar nuevas variables aleatorias, la cuestión que en mayores dimensiones las funciones varían. Empezaremos estudiando funciones de varias variables a una variable, es decir, funciones de la forma

$$\begin{aligned} F : \mathbb{R}^n &\to \mathbb{R} \\ (x_1, \ldots, x_n) &\mapsto F(x_1, \ldots, x_n) = y \end{aligned}$$

Sea $(X_1, \ldots, X_n)$ un vector aleatorio en $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad bien definido, sea $g : \mathbb{R}^n \to \mathbb{R}$ una funcion de varias variables a una variable, entonces las transformación $Y = g(X_1, \ldots, X_n)$ define una variable aleatoria en $(\Omega, \mathcal{A}, P)$, ya que para cada $\omega \in \Omega$ se tiene que $Y(\omega) = g(X_1(\omega), \ldots, X_n(\omega)) \in \mathbb{R}$. Con respecto a Y, se genera una variable aleatoria de la forma

$$Y : \Omega \to \mathbb{R}$$

Es importante entender el comportamiento de Y, pero podemos prenscindir de ella por el momento, para el cálculo de su esperanza (si es que éxiste).

Teorema 5.3. *Sea* $(X_1, \ldots, X_n)$ *un vector aleatorio, sea* $g : \mathbb{R}^n \to \mathbb{R}$ *una función real valorada, entonces*

$$\begin{aligned} &E(g(X_1, \ldots, g_n)) \\ &= \begin{cases} \sum \cdots \sum_{(x_1,\ldots,x_n)\in\mathbb{R}^n} g(x_1, \ldots, x_n) f_{X_1,\ldots,X_j}(x_1 \ldots, x_n), & c.d. \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n, & c.c. \end{cases} \end{aligned}$$

5.5.1. Esperanza de Funciones Lineales

Si $g(x_1, \ldots, x_n) = \sum_{i=1}^n a_i x_i + b$, donde $a_1, \ldots, a_n, b$ son constantes reales, entonces

$$E\left(\sum_{i=1}^n a_i X_i + b\right) = \sum_{i=1}^n a_i E(X_i) + b$$

provisto que $E(|X_i|) < \infty$ para todo $i = 1, \ldots, n$. Esto se ve por definición, supongamos que el vector aleatorio es continuo, entonces

$$\begin{aligned} E\left(\sum_{i=1}^n a_i X_i + b\right) &= \int_{\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\sum_{i=1}^n a_i u_i + b\right) f_{X_1,\ldots,X_n}(u_1, \ldots, u_n) du_n \ldots du_1 \\ &= \sum_{i=1}^n a_i \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u_i f_{X_1,\ldots,X_n}(u_1, \ldots, u_n) du_n \ldots du_1 + b \\ &= \sum_{i=1}^n a_i E(X_i) + b \end{aligned}$$

Notemos que $E(X_i)$ es la esperanza de la marginal X_i que además, se puede interpretar como la esperanza habitual, es decir

$$E(X_i) = \int_{-\infty}^{\infty} u_i f_{X_i}(u_i) du_i$$

Y esto es en virtud de estamos tomando la función continua $g(x_1, \ldots, x_n) = x_i$, por lo que

$$\begin{aligned} E(X_i) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u_i f_{X_1,\ldots,X_n}(u_1, \ldots, u_n) du_1 \ldots du_n \\ &= \int_{-\infty}^{\infty} u_i \underbrace{\left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(u_1, \ldots, u_n) du_1 \ldots du_n\right)}_{f_{X_i}(u_i)} du_i \\ &= \int_{-\infty}^{\infty} u_i f_{X_i}(u_i) du_i \end{aligned}$$

Por lo tanto, podemos determinar de forma más comoda estos tipos de funciones. Por ejemplo, se se toma la función

$$F : \mathbb{R}^n \to \mathbb{R}$$

$$(x_1, \ldots, x_n) \mapsto \overline{x} = \sum_{i=1}^n x_i$$

Entonces, si $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ es la media muestral de $X_1, \ldots, X_n$. Entonces

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

En particular, si todas las X_i's tienen la misma media, digamos que $E(X_i) = \mu$, para $i = 1, \ldots, n$, entonces

$$E(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu \quad (\star)$$

para todo $n = 1, 2, \ldots$. Esto ocurre, por ejemplo, cuando $X_1, \ldots, X_n$ corresponde a una muestra aleatoria (m.a) de una distribución F (probalción) con media μ, lo cual equivale a decir que $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$, donde F una distribución con media μ.

El resultado $(\star)$ indica que la media muestral $\overline{X}$ es un predictor insesgado de la media poblacional μ.

Ejemplo 5.9. Sean $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \exp(\lambda)$. Sabemos que a media de la distribución exponencial es

$$\mu = \int_0^\infty x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}\int_0^\infty t e^{-t} dt = \frac{\Gamma(2)}{\lambda} = \frac{1}{\lambda}$$

donde $t = x\lambda$. Por lo tanto

$$E(\overline{X}) = \frac{1}{\lambda}$$

Ejemplo 5.10. Sean X_1 y X_2 variables aleatorias independientes, donde $X_1 \sim U(0,1)$ con fdp $f_{X_1}(x) = I_{(0,1)}(x)$ (toma valor 1 en el intervalo $(0,1)$) y $X_2 \sim N(0,1)$ con fdp $f_{X_2}(x) = \phi(x)$ donde

$$\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$$

para cada $x \in \mathbb{R}$. Entonces

$$E(X_1) = \int_0^1 x dx = \frac{1}{2} \text{ y } E(X_2) = \int_{-\infty}^{\infty} x\phi(x) dx = 0$$

Por lo tanto, si

$$\begin{aligned} E(2X_1 + x_2) &= 2E(X_1) + E(X_2) = 1 \\ E(2X_1 - X_2 - 1) &= 2E(X_1) - E(X_2) - 1 = 0 \end{aligned}$$

Ejemplo 5.11. Sean X_1, X_2, X_3 variables aleatorias con fdp conjunta dada por:

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = \begin{cases} \frac{3}{4}(x_1^2 + x_2^2 + x_3^2), & 0 < x_1, x_2, x_3 < 1 \\ 0, & eoc. \end{cases}$$

Calculemo la esperanza de $X = (X_1 + X_2 + X_3)/3$. Calculemos primero usando medias marginales

$$\begin{aligned}
E(X_1) &= \int_0^1 \int_0^1 \int_0^1 x_1 \frac{3}{4}(x_1^2 + x_2^2 + x_3^2) dx_1 dx_2 dx_3 \\
&= \frac{3}{4} \int_0^1 \int_0^1 \frac{1}{4} + \frac{1}{2}x_2^2 + \frac{1}{2}x_3^2 dx_2 dx_3 \\
&= \frac{3}{4} \int_0^1 \frac{1}{4} + \frac{1}{6} + \frac{1}{2}x_3^2 dx_3 \\
&= \frac{3}{4}\left(\frac{1}{4} + \frac{1}{6} + \frac{1}{6}\right) \\
&= \frac{7}{16}
\end{aligned}$$

Notemos que $E(X_1) = E(X_2) = E(X_3)$, por lo que

$$E((X_1 + X_2 + X_3)/3) = \frac{1}{3} \cdot 3 \cdot \frac{7}{16} = \frac{7}{16}$$

Veamos si las variables son idependientes. Notemos que

$$\begin{aligned}
f_{X_1}(x_1) &= \int_0^1 \int_0^1 \frac{3}{4}(x_1^2 + x_2^2 + x_3^2) dx_2 dx_3 \\
&= \frac{3}{4} \int_0^1 x_1^2 + \frac{1}{3} + x_3^2 dx_3 \\
&= \frac{3}{4}\left(x_1^2 + \frac{1}{3} + \frac{1}{3}\right) \\
&= \frac{3}{4}x_1^2 + \frac{1}{2}
\end{aligned}$$

De forma análoga se ve que

$$f_{X_2}(x_2) = \frac{3}{4}x_2^2 + \frac{1}{2}, \quad f_{X_3}(x_3) = \frac{3}{4}x_3^2 + \frac{1}{2}$$

Y entonces

$$f_{X_1}(x_1) f_{X_2}(x_2) f_{X_3}(x_3) \neq f_{X_1,X_2,X_3}(x_1, x_2, x_3)$$

Es decir, no son independientes.

5.5.2. Función Generadora de Momentos Multivariada (fgm)

Sea la función definida por

$$\begin{aligned}
g : \mathbb{R}^n &\to \mathbb{R} \\
(x_1, \ldots, x_n) &\mapsto \prod_{i=1} e^{t_i x_i} = e^{\sum_{i=1}^n t_i x_i}
\end{aligned}$$

donde $t_1, \dots, t_n$ fijos. Entonces

$$E(g(X_1, \dots, X_n))$$

define la fgm conjunta de $X_1, \dots, X_n$, para $t_1, \dots, t_n$ fijos en caso de que exista.

Definición 5.9. *La fgm de conjunta de $X_1, \dots, X_n$ se define como*

$$\begin{aligned} M_{X_1,\dots,X_n}(t_1, \dots, t_n) &= E\left(e^{\sum_{i=1}^n t_i X_i}\right) \\ &= \begin{cases} \sum \cdots \sum_{(x_1,\dots,x_n)\in\mathbb{R}^n} e^{\sum_{i=1}^n t_i x_i} f_{X_1,\dots,x_n}(x_1, \dots, x_n), & c.d. \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{\sum_{i=1}^n t_i x_i} f_{X_1,\dots,X_n}(x_1, \dots, x_n) dx_1 \dots x_n, & c.c. \end{cases} \end{aligned}$$

Siempre y cuando exista para todo $(t_1, \dots, t_n) \in \mathbb{R}^n$ talque $|t_k| < h_k$ para algún $(h_1, \dots, h_k) \in \mathbb{R}^n$, con $h_k > 0$ para todo $k = 1, \dots, n$.

Al igual que en una variable, la fgm multivariada entrega momentos, veamos algunas propiedades de esta.

Proposición 5.1. (Propiedades de fgm multivariada)

(a) *Se puede definir sobre un subvector, si $(X_1, \dots, X_k)$ es un subvector de $(X_1, \dots, X_n)$, entonces:*

$$M_{X_1,\dots,X_k}(t_1, \dots, t_k) = M_{X_1,\dots,X_k,X_{k+1},\dots,X_n}(t_1, \dots, t_k, 0, \dots, 0)$$

para todo $k = 1, \dots, n-1$.

(b)

$$\left.\frac{\partial^{k_1+\cdots+k_n} M_{X_1,\dots,X_n}(t_1, \dots, t_n)}{\partial t_1^{k_1} \dots \partial t_n^{k_n}}\right|_{t_1=\cdots=t_n=0} = E(X_1^{k_1} \times \cdots \times X_n^{k_n})$$

(c) *$X_1, \dots, X_n$ son variables aleatorias independientes, si y sólo si*

$$M_{X_1,\dots,X_n}(t_1, \dots, t_n) = \prod_{i=1}^n M_{X_i}(t_i)$$

para todo $(t_1, \dots, t_n)$ donde las fgm's existen.

(d) *Si $Y = \sum_{i=1}^n a_i X_i + b$, entonces*

$$M_Y(t) = e^{bt} M_{X_1,\dots,X_n}(a_1 t, \dots, a_n t)$$

Es importante ver que la fgm tiene un comportamiento similar a la esperanza, con respecto a subvectores, solo que evaluamos en 0 en las variables que no nos interesa. Para determinar la esperanza a partir de la fgm es similar en una dimensión, solo que hay que derivar donde es necesario, asi para determinar $E(X_1 \dots X_n)$, basta estudiar

$$\left.\frac{\partial^n M_{X_1,\dots,X_n}(t_1, \dots, t_n)}{\partial t_1 \dots \partial t_n}\right|_{t_1=\cdots=t_n=0}$$

Con respecto a la independencia, el valor del término

$$M_{X_i}(t_i) = \int_{-\infty}^{\infty} e^{t_i x_i} f_{X_i}(x_i) dx_i$$

donde f_{X_i} es la marginal fdp de la conjunta. Siempre pensando en un vector continuo.

Otra caracterización intersante, es que podemos ver cuando $X_1, \ldots, X_n$ son variables aleatorias independientes, usando la fgm, siempre y cuando se cumpla para todo t que la satisfaga.

Si $X_1, \ldots, X_n$ son variables aleatorias independientes, entonces

$$\begin{aligned} M_{\sum_{i=1}^n a_i X_i + b}(t) &= e^{bt} M_{X_1,\ldots,X_n}(a_1 t, \ldots, a_n t) \\ &= e^{bt} \prod_{i=1}^{n} M_{X_i}(a_i t) \end{aligned}$$

En particular, si $X_1, \ldots, X_n \overset{\text{iid}}{\sim} M(t)$ e $Y = \sum_{i=1}^n X_i$, entonces

$$M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^{n} M(t) = (M(t))^n$$

Ejemplo 5.12. Sean $X_1, \ldots, X_n \overset{\text{iid}}{\sim} Ber(p)$. Probemos que $Y := \sum_{i=1}^n X_i \sim Bin(n,p)$. Si $X_1, \ldots, X_n \overset{\text{iid}}{\sim} Ber(p)$, entonces

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} M(t) = (1 - p + pe^t)$$

Luego, la fgm de $Y = \sum_{i=1}^n X_i$ es,

$$\begin{aligned} M_Y(t) &= M_{\sum_{i=1}^n X_i(t)} \\ &= (M(t))^n \\ &= (1 - p + pe^t)^n \end{aligned}$$

Entonces $M_Y(t) =$ fgm de la distribución $Bin(n,p)$, como Y tiene fgm $Bin(n,p)$ entonces necesariamente

$$Y \sim Bin(n,p)$$

Por el teorema 3.12.

Ejemplo 5.13. Si $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \exp(\lambda)$, entonces $M(t) = (1 - t/\lambda)^{-1}$ donde $t < \lambda$, luego si $Y = \sum_{i=1}^n X_i$

$$M_Y(t) = (1 - t/\lambda)^{-n}$$

Pero esta es la fgm de la disgtribución $Gamma(n, \lambda)$, por lo tanto por unicidad de fgm

$$Y \sim Gamma(n, \lambda)$$

Ejemplo 5.14. Sean X_1, X_2 variables aleatorias independientes, con $X_i \sim N(\mu_i, \sigma^2)$ $i = 1, 2$. Encontremos la distribución $Y = X_1 - X_2$.

Si $X_i \sim N(\mu_i, \sigma^2)$, entonces $M_{X_i}(t) = e^{\mu_i t + \frac{1}{2}\sigma^2 t^2}$ para $i = 1, 2$. Además, como X_1 y X_2 son variables aleatorias independientes, entonces la fgm de $Y = X_1 - X_2$ es

$$\begin{aligned} M_{X_1 - X_2}(t) &= M_{X_1, X_2}(t, -t) \\ &= M_{X_1}(t) M_{X_2}(-t) \\ &= e^{\mu_1 t + \frac{1}{2}\sigma^2 t^2} e^{-\mu_2 t + \frac{1}{2}\sigma^2 t^2} \\ &= e^{(\mu_1 - \mu_2)t + \frac{1}{2}(2\sigma^2)t^2} \end{aligned}$$

para todo $t \in \mathbb{R}$. Entonces a partir de la fgm, se ve que la media es $\mu_1 - \mu_2$ y la varianza $2\sigma^2$, entonces

$$M_Y(t) = M_Z(t)$$

para todo $t \in \mathbb{R}$, donde $Z \sim N(\mu_1 - \mu_2, 2\sigma^2)$. Por la unicidad de fgm se concluye que

$$X_1 - X_2 \sim N(\mu_1 - \mu_2, 2\sigma^2)$$

Esto también caracteriza la distribución de $Y = \sum_{i=1}^n X_i$, donde $X_i \sim N(\mu_i, \sigma^2)$, por lo que es fácil ver que

$$M_Y(t) = e^{\sum_{i=1}^n \mu_i + \frac{1}{2}(n\sigma^2)t^2}$$

Por lo tanto

$$Y \sim N\left(\sum_{i=1}^n \mu_i, n\sigma^2\right)$$

Si se toma $X_i \sim N(\mu_i, \sigma_i^2)$ entonces

$$Y := \sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

5.5.3. Covarianza y Correlación

Estudiaremos un nuevo tipo de " varianza "la cual le diremos covarianza, donde se puede sacar la varianza habitual.

Definición 5.10. (Covarianza) *La covarianza entre dos variables aleatorias X e Y se define como*

$$\begin{aligned} Cov(X,Y) &= E((X-\mu_X)(Y-\mu_Y)) \\ &= \begin{cases} \sum_{x\in\mathbb{R}}\sum_{y\in\mathbb{R}}(x-\mu_X)(y-\mu_Y)f_{X,Y}(x,y), & c.d. \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f_{X,Y}(x,y)dydx, & c.c. \end{cases} \end{aligned}$$

Notación. Una notación alternativa es

$$\sigma_{XY} := \mathrm{Cov}(X,Y)$$

Proposición 5.2. *Sean X e Y variables aleatorias definidas sobre un mismo espacio de probabilidad, entonces*

(a) ***Forma alternativa.*** $Cov(X,Y) = E(X,Y) - E(X)E(Y)$.

(b) ***Operador positivo definido.*** $Cov(X,X) = Var(X) \geq 0$.

(c) ***Simetría.*** $Cov(X,Y) = Cov(Y,X)$.

(d) $Cov(X,c) = Cov(c,X) = 0$.

(e) Si X e Y son variables aleatorias independientes, entonces

$$Cov(X,Y) = 0$$

(f) $Cov(aX+b,Y) = a\,Cov(X,Y)$ *y* $Cov(aX+b,X) = a\,Var(X)$.

(g) ***Desigualdad de Cauchy-Schwarz.*** $|Cov(X,Y)| \leq \sqrt{Var(X)\,Var(Y)}$.

Dem. Sean X,Y variables aleatorias bien definida sobre un espacio de probabilidad, con covarianza bien definida. Entonces

(a) Por definición

$$\begin{aligned} \mathrm{Cov}(X,Y) &= E((X-\mu_X)(Y-\mu_Y)) \\ &= E(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y) \\ &= E(XY) - 2\mu_X\mu_Y + \mu_X\mu_Y \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

(b) Por (a)

$$\mathrm{Cov}(X,X) = E(X^2) - E^2(X) = \mathrm{Var}(X) \geq 0$$

(c) Es claro que simétrica.

(d) Sea c constante, entonces

$$\mathrm{Cov}(X,c) = cE(X) - E(X)E(c) = 0$$

(e) Sean X, Y independientes, continua, (X, Y) es un vector continua y entonces

$$\begin{aligned}\text{Cov}(X,Y) &= E(XY) - E(X)E(Y) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{X,Y}(x,y)dxdy \\ &= \int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dxdy - E(X)E(Y) \\ &= \left(\int_{-\infty}^{\infty} xf_X(x)dx\right)\left(\int_{-\infty}^{\infty} yf_Y(y)dy\right) - E(X)E(Y) \\ &= E(X)E(Y) - E(X)E(Y) = 0\end{aligned}$$

(También se puede probar de forma general usando la fda y límites)

(f) Sean a, b constantes, luego

$$\begin{aligned}\text{Cov}(aX+b,Y) &= E((aX+b)Y) - E(aX+b)E(Y) \\ &= E(aXY+bY) - aE(X)E(Y) - bE(Y) \\ &= aE(XY) + bE(Y) - aE(X)E(Y) - bE(Y) \\ &= a\text{Cov}(X,Y)\end{aligned}$$

Si $X = Y$, entonces

$$\text{Cov}(aX+b,X) = a\text{Cov}(X,X) = a\text{Var}(X)$$

(g) ...

Probando la proposición. ■

Nota 5.5. Si X e Y son independientes, entonces $g(X), h(Y)$ también son independientes para cualquier función g y h.

Definición 5.11. *La correlación (coeficiente de correlación) entre dos variables aleatorias X e Y, se define como*

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

Observación 5.1. De la propiedad (g), $|\rho_{XY}| \leq 1$, donde es igual, si y sólo s $Y = aX + b$ con $a \neq 0$, de abuerdo a la propiedad (f).

Teorema 5.4. *Suponga que X es una variable aleatoria talque $0 < \sigma_X^2 < \infty$. Si existen constantes $a \neq 0$ y b talque $Y = aX + b$, entonces, $\rho_{XY} = 1$ si $a > 0$ (asociación lineal +) y $\rho_{XY} = -1$ si $a < 0$ (asociación linea −).*

Dem. Basta usar la proposición 5.2, como

$$\text{Var}(aX+b) = a^2\text{Var}(X) \text{ y } \text{Cov}(X,aX+b) = a\text{Var}(X)$$

por la definición de correlación se tiene

$$\rho(X, aX+b) = \frac{\mathrm{Cov}(X, aX+b)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(aX+b)}} = \frac{a}{\sqrt{a^2}} = \begin{cases} 1, & a > 0 \\ -1, & a < 0 \end{cases}$$

Si $a = 0$ entonces tendriamos que

$$\rho(X, b) = \frac{\mathrm{Cov}(X, b)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(aX+b)}} = 0$$

■

Ejemplo 5.15. Sean $X = 3Z+1$ e $Y = -\frac{1}{3}Z-1$, donde $Z \sim N(0,1)$. SI $E(Z) = 0, \mathrm{Var}(Z) = 1$, entonces

$$\begin{aligned} E(X) &= 3E(Z) + 1 = 1, \quad E(Y) = -3E(Z) - 1 = -1 \\ \mathrm{Var}(X) &= 3^2\mathrm{Var}(Z) = 9, \quad \mathrm{Var}(Y) = \left(-\frac{1}{3}\right)^2 \mathrm{Var}(Z) = \frac{1}{9} \\ \mathrm{Cov}(X,Y) &= \mathrm{Cov}(3Z+1, Y) \\ &= 3\mathrm{Cov}(Z, Y) \\ &= 3\left(-\frac{1}{3}\right)\mathrm{Var}(Z) \\ &= -\mathrm{Var}(Z) = -1 \end{aligned}$$

Por lo tanto

$$\rho_{XY} = -1/\sqrt{9 \cdot \frac{1}{9}} = -1$$

Ejemplo 5.16. Tomando X, Y del ejemplo anterior. Si $Z \sim U(-1,1)$, entonces

$$E(Z) = \frac{1+(-1)}{2} = 0 \;\text{ y }\; \mathrm{Var}(Z) = \frac{(1-(-1))^2}{12} = \frac{1}{3}$$

Luego

$$\begin{aligned} E(X) &= 1, \quad E(Y) = -1 \\ \mathrm{Var}(X) &= 9\mathrm{Var}(Z) = 3, \quad \mathrm{Var}(Y) = \frac{1}{9}\mathrm{Var}(Z) = \frac{1}{27} \end{aligned}$$

$$\mathrm{Cov}(X,Y) = -\mathrm{Var}(Z) = -\frac{1}{3}$$

Entonces

$$\rho_{XY} = -1$$

5.5.4. Varianza de Funciones Lineales

Sea una función lineal

$$Y = \sum_{i=1}^{n} a_i X_i + b$$

Calculemos su varianza, sean $X_1, \ldots, X_n$ variables aleatorias finitas y $a_1, \ldots, a_n, b \in \mathbb{R}$ constantes, vimos que

$$E\left(\sum_{i=1}^{n} a_i X_i + b\right) = \sum_{i=1}^{n} a_i E(X_i) + b \quad (\star)$$

Ahora, podemos también afirmar que

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i + b\right) = \sum_{i=1}^{n} a_i^2 \mathrm{Var}(X_i) + 2 \sum_{\substack{1 \leq i,j \leq n \\ i<j}} a_i a_j \mathrm{Cov}(X_i, X_j) \quad (\star\star)$$

En particular, si las variables aleatorias $X_1, \ldots, X_n$ son independientes, entonces $\mathrm{Cov}(X_i, X_j) = 0$ para todo $i \neq j$ y de $(\star\star)$ se tiene que

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i + b\right) = \sum_{i=1}^{n} a_i^2 \mathrm{Var}(X_i)$$

Por ejemplo, si $X_1, \ldots, X_n$ son variables aleatorias iid con media μ y varianza σ^2, entonces la media y la varianza de la media muestral $X = \sum_{i=1}^{n} \frac{X_i}{n}$ están dadas por

$$E(X) = \mu \text{ y } \mathrm{Var}(X) = \frac{\sigma^2}{n}$$

Además, la desigualdad de Chebyshev implica que para todo $\varepsilon > 0$

$$0 \leq P(|X - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \longrightarrow, \text{ cuando } n \longrightarrow \infty$$

Es decir, la probabilidad de que X difiera de μ es despreciable para una cantidad n suficientemente grande.

Ejemplo 5.17.

- Sean $X, Y \overset{\text{iid}}{\sim} N(0,1)$, entonces $E(X) = E(Y) = 0, \mathrm{Var}(X) = 1\mathrm{Var}(Y) = 0$ y $\mathrm{Cov}(X,Y) = 0$. Luego

  $$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y) = 1 + 1 + 2 \cdot 0 = 2$$

  Análogamente

  $$\mathrm{Var}(X-Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X,Y) = 1 + 1 - 2 \cdot 0 = 2$$

Además, como $E(X+Y) = E(X-Y)$, luego por propiedades de covarianza, se tiene que

$$\begin{aligned}\text{Cov}(X+Y, X-Y) &= E((X+Y)(X-Y)) - 0\\ &= E(X^2 - Y^2) = E(X^2) - E(Y^2)\\ &= 1 - 1 = 0\end{aligned}$$

Entonces $\rho_{X+Y,X-Y} = 0$.

- Si $X, Y \stackrel{\text{iid}}{\sim} \exp(1)$, entonces $X+Y, X-Y$ tambien tiene correlación nula. En efecto, en este caso

$$E(X) = E(Y) = \text{Var}(X) = \text{Var}(Y) = 1 \text{ y } \text{Cov}(X, Y) = 0$$

Luego

$$E(X+Y) = 2,\ \ E(X-Y) = 0,\ \ \text{Var}(X+Y) = \text{Var}(X-Y) = 2$$

y

$$\begin{aligned}\text{Cov}(X+Y, X-Y) &= E((X+Y)(X-Y)) - 2\cdot 0 = E(X^2) - E(Y^2)\\ &= \text{Var}(X) + E(X)^2 - (\text{Var}(Y) + E(Y)^2)\\ &= 2 - 2 = 0\end{aligned}$$

Entonces $\rho_{X+Y,X-Y} = 0$.

- Si $X \sim N(0,1)$ e $Y \sim N(1,1)$ son variables aleatorias independientes, entonces

$$\begin{aligned}E(X+Y) &= 1,\ \ E(X-Y) = -1\\ \text{Var}(X+Y) &= \text{Var}(X-Y) = 2\\ \text{Cov}(X+Y, X-Y) &= E((X+Y)(X-Y)) - E(X+Y)E(X-Y)\\ &= 1 - (1+1) - 1\cdot(-1) = 0\end{aligned}$$

Luego $\rho_{X+Y,X-Y} = 0$

Ejemplo 5.18.

- Sean $X, Y \stackrel{\text{iid}}{\sim} N(0,1)$. Veamos si $X+Y, X-Y$ son variables aleatorias independientes con distribución $N(0,2)$. Recordemos que la función generadora de momentos sobre $N(0,1)$ es $M(t) = e^{t^2/2}$, entonces

$$M_{X+Y}(t) = M_{X,Y}(t,t) = M_X(t)M_y(t) = e^{t^2}$$

Luego por unicidad de fgm se tiene que $X+Y \sim N(0,2)$, por otro lado

$$M_{X-Y}(t) = M_X(t)M_Y(-t) = e^{t^2}$$

De forma que $X-Y \sim N(0,2)$. Veamos si son independientes, digamos que $Z_1 = X+Y, Z_2 = X-Y$, entonces por definición

$$f_{Z_1,Z_2}(z_1, z_2) = P(X+Y = z_1, X-Y = z_2)$$

Notemos que debemos resolver un sistema de ecuaciones de la forma

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

Con matriz invertible y luego, con una única solución, en donde $a = (z_1 + z_2)/2, b = (z_1 - z_2)/2$, luego

$$\begin{aligned} f_{Z_1,Z_2}(z_1, z_2) &= P(X = (z_1 + z_2)/2, Y = (z_1 - z_2)/2) \\ &= P(X = (z_1 + z_2)/2)P(Y = (z_1 - z_2)/2) \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}\frac{(z_1^2+z_2^2)}{2}} \end{aligned}$$

Por otro lado

$$f_{Z_1}(z_1) = \frac{1}{2\sqrt{\pi}} e^{-\frac{1}{2}\frac{z_1^2}{2}}, \quad f_{Z_2}(z_2) = \frac{1}{2\sqrt{\pi}} e^{-\frac{1}{2}\frac{z_2^2}{2}}$$

Claramente $f_{Z_1,Z_2}(z_1, z_2) = f_{Z_1}(z_1)f_{Z_2}(z_2)$. Por lo tanto son independientes.

- Sean $X, Y \overset{\text{iid}}{\sim} \exp(1)$. Veamos si las variables $X + Y, X - Y$ son o no independientes. Notemos que

$$\begin{aligned} M_{X+Y,X-Y}(t_1, t_2) &= E(e^{t_1(X+Y)+t_2(X-Y)}) \\ &= E(e^{X(t_1+t_2)+Y(t_1-t_2)}) \\ &= M_{X,Y}(t_1 + t_2, t_1 - t_2) \\ &= M_X(t_1 + t_2)M_Y(t_1 - t_2) \\ &= (1 - t_1 - t_2)^{-1}(1 - t_1 + t_2) \\ &= ((1 - t_1)^2 - t_2^2) \end{aligned}$$

Mientras que

$$\begin{aligned} M_{X+Y}(t) &= (1 - t)^{-2} \\ M_{X-Y}(t) &= (1 - t^2)^{-1} \end{aligned}$$

Pero

$$M_{X+Y,X_Y}(t, s) \neq M_{X,Y}(t), M_{X,Y}(s)$$

- Sean $X \sim N(0, 1)$ e $Y \sim N(1, 1)$ variables aleatorias independientes. Obtengamos las fgm's conjuntas y marginales de $X+Y, X-Y$. Por definición, si $Z_1 := X+Y, Z_2 := X-Y$, entonces

$$\begin{aligned} M_{Z_1,Z_2}(t_1, t_2) &= M_X(t_1 + t_2)M_Y(t_1 - t_2) \\ &= e^{\frac{1}{2}(t_1+t_2)^2} e^{t_1-t_2+\frac{1}{2}(t_1-t_2)^2} \\ &= e^{t_1^2+t_2^2+t_1-t_2} \end{aligned}$$

Encontrando la fgm conjunta, encontremos las marginales.

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)M_Y(t) \\ &= e^{\frac{1}{2}t^2}e^{t+\frac{1}{2}t^2} \\ &= e^{t^2+t} \\ M_{X-Y}(t) &= M_X(t)M_Y(-t) \\ &= e^{\frac{1}{2}t^2}e^{-t+\frac{1}{2}t^2} \\ &= e^{t^2-t} \end{aligned}$$

Luego claramente

$$M_{Z_1,Z_2}(t_1,t_2) = M_{Z_1}(t_1)M_{Z_2}(t_2)$$

Siendo Z_1, Z_2 independientes.

5.6. Esperanza y Varianza de Vectores Aleatorios

5.6.1. Definición y Ejemplos

Hasta ahora hemos trabajado con la esperanza de una variable aleatoria generada a partir de un vector aleatorio, ahora definiremos la esperanza de un vector aleatorio.

Notación. A partir de ahora usaremos vectores, por lo que necesitamos una notación. Cuando hablemos de vectores aleatorios, usaremos $\mathbf{X}$, cuando hablamos de la esperanza de un vector, en vez de escribir μ, escribimos $\boldsymbol{\mu}$ y así sucesivamente. Es decir, cuando está en un tono marcado, nos referimos a un vector.

Otra cosa a mencionar, antes hemos trabajado los vectores aleatorios con la notación $(X_1, \ldots, X_n)$, pero ahora le daremos un sentido como matriz, usaremos la expresión

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = (X_1, \ldots, X_n)^T$$

para expresar un vector aleatorio.

Definición 5.12. (Esperanza de un vector aleatorio) *Sea $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ un vector aleatorio de dimensión n. La esperanza de $\boldsymbol{X}$ denotado como $E(\boldsymbol{X})$, se define por:*

$$E(\boldsymbol{X}) = (E(X_1), E(X_2), \ldots, E(X_n))^T$$

Es decir, si $\boldsymbol{\mu} := E(\boldsymbol{X})$ y $\mu_i = E(X_i)$ para $i = 1, \ldots, n$, entonces $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$ y también se llama vector esperado.

Definición 5.13. (Matriz de varianza-covarianza) *Sea $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ un vector aleatorio de dimensión n. La matriz de varianza-covarianza de $\boldsymbol{X}$ se define como:*

$$\begin{aligned} Var(\boldsymbol{X}) &= E([\boldsymbol{X} - E(\boldsymbol{X})][\boldsymbol{X} - E(\boldsymbol{X})]^T) \\ &= \begin{pmatrix} Cov(X_1, X_1) & Cov(X_1, X_2) & \ldots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Cov(X_2, X_2) & \ldots & Cov(X_2, X_n) \\ \vdots & \vdots & \ldots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \ldots & Cov(X_n, X_n) \end{pmatrix} \end{aligned}$$

Es decir, si $\Sigma := Var(\boldsymbol{X})$ y $\sigma_{ij} = Cov(X_i, X_j) = E((X_i - \mu_i)(X_j - \mu_j))$ donde $i, j = 1, \ldots, n$, entonces $\Sigma = \{(\sigma_{ij})\}_{i,j=1,\ldots,n}$

Recordemos que $\text{Cov}(X_i, X_i) = \text{Var}(X_i)$. Por lo que también se puede interpreta como:

$$\text{Var}(\mathbf{X}) = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \ldots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \ldots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ldots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \ldots & \text{Var}(X_n) \end{pmatrix}$$

Notar que la definición tiene sentido, ya que la matriz $(\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))^T$ genera la matriz

$$\begin{pmatrix} (X_1 - E(X_1))(X_1 - E(X_1)) & \ldots & (X_1 - E(X_1))(X_n - E(X_n)) \\ \vdots & \ldots & \vdots \\ (X_n - E(X_n))(X_1 - E(X_1)) & \ldots & (X_n - E(X_n))(X_n - E(X_n)) \end{pmatrix}$$

Luego es solo distribuir la esperanza, formando varianzas y covarianzas.

Definición 5.14. (Matriz de covarianza) *Sean $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ e $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^T$ dos vectores aleatorios definidos en un mismo espacio de probabilidad. La matriz de covarianza entre $\boldsymbol{X}$ e $\boldsymbol{Y}$ se define como:*

$$\begin{aligned} Cov(\boldsymbol{X}, \boldsymbol{Y}) &= E([\boldsymbol{X} - E(\boldsymbol{X}][\boldsymbol{Y} - E(\boldsymbol{Y})]^T) \\ &= \begin{pmatrix} Cov(X_1, Y_1) & Cov(X_1, Y_2) & \ldots & Cov(X_1, Y_m) \\ Cov(X_2, Y_1) & Cov(X_2, Y_2) & \ldots & Cov(X_2, Y_m) \\ \vdots & \vdots & \ldots & \vdots \\ Cov(X_n, Y_1) & Cov(X_n, Y_2) & \ldots & Cov(X_n, Y_m) \end{pmatrix} \end{aligned}$$

Ejemplo 5.19. Sean X e Y variables aleatorias con fmp conjunta dada por:

$x \setminus y$	0	1	$P(X = x)$
-1	1/7	1/7	2/7
0	2/7	1/7	3/7
1	1/7	1/7	2/7
$P(Y = y)$	4/7	3/7	1

Es claro que

$$\begin{aligned}
E(X) &= -1 \cdot \frac{2}{7} + 1 \cdot \frac{2}{7} = 0 \\
\text{Var}(X) &= E(X^2) = \frac{4}{7} \\
E(Y) &= 1 \cdot \frac{3}{7} = \frac{3}{7} \\
\text{Var}(Y) &= \frac{12}{49} \\
\text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\
&= 0
\end{aligned}$$

Por lo que el vector esperador y la matriz varianza-covarianza de $\mathbf{X}$ es

$$\mu = \begin{pmatrix} 0 \\ 3/7 \end{pmatrix}, \;\; \Sigma = \begin{pmatrix} 4/7 & 0 \\ 0 & 12/49 \end{pmatrix}$$

Ejemplo 5.20. Sea (X, Y) un vector fijo con fdp conjunta dada por

$$f_{X,Y} = \begin{cases} \frac{1}{x}, & 0 < x \leq 1, 0 < y \leq x \\ 0, & eoc. \end{cases}$$

En este caso tenemos

$$\begin{aligned}
E(X) &= \int_0^1 \int_y^1 x \frac{1}{x} dx dy = \frac{1}{2} \\
E(Y) &= \int_0^1 \int_0^x y \frac{1}{x} dy dx = \frac{1}{4} \\
E(X^2) &= \int_0 1 \int_0^x x^2 \frac{1}{x} dy dx = \frac{1}{3}
\end{aligned}$$

Por lo tanto $\text{Var}(X) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$. Por último

$$\begin{aligned}
E(Y^2) &= \int_0^1 \int_0^x y^2 \frac{1}{x} dy dx = \frac{1}{9} \\
E(XY) &= \int_0^1 \int_0^x xy \frac{1}{x} dy dx = \frac{1}{6}
\end{aligned}$$

Y entonces $\text{Var}(Y) = \frac{1}{9} - \frac{1}{16} = \frac{7}{144}$,$\text{Cov}(X,Y) = \frac{1}{6} - \frac{1}{8} = \frac{1}{24}$. Por lo tanto, el valor esperado $\mu = E(\mathbf{X})$ es

$$E(\mathbf{X}) = \mu = \begin{pmatrix} 1/2 \\ 1/4 \end{pmatrix}$$

y la matriz varianza-covarianza $\Sigma = \text{Var}(\mathbf{X})$ es

$$\Sigma = \begin{pmatrix} 1/12 & 1/24 \\ 1/24 & 7/144 \end{pmatrix}$$

5.6.2. Matriz de Correlación

Definición 5.15. (Matriz de Correlación) *Sea $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ un vector aleatorio n dimensional. La matriz de correlación de $\boldsymbol{X}$, denotara como $\boldsymbol{R}$, se define como*

$$\boldsymbol{R} := \begin{pmatrix} \rho_{X_1,X_1} & \rho_{X_1,X_2} & \cdots & \rho_{X_1,X_n} \\ \rho_{X_2,X_1} & \rho_{X_2,X_2} & \cdots & \rho_{X_2,X_n} \\ \vdots & \vdots & \ldots & \vdots \\ \rho_{X_n,X_1} & \rho_{X_n,X_2} & \cdots & \rho_{X_n,X_n} \end{pmatrix}$$

De forma similar se define la matriz de correlación para dos vectores aleatorios, $\boldsymbol{R}_{XY} = \{(\rho_{X_i,Y_j})\}_{n\times m}$, donde $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ e $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^T$, de modo que la matriz $\boldsymbol{R}$ se tiene para cuando $\boldsymbol{X} = \boldsymbol{Y}$.

Ejemplo 5.21. Sea $\mathbf{X} = (X, Y)^T$ un vector aleatorio bi-dimensional con fdp conjunta dada por:

$$f(x,y) = \begin{cases} 2, & 0 < x < y, 0 < y < 1 \\ 0, & eoc. \end{cases}$$

Tenemos entonces

$$\begin{aligned} E(X) &= \int_0^1 \int_0^y 2x dx dy = \frac{1}{3} \\ E(Y) &= \int_0^1 \int_x^1 2y dy dx = \frac{2}{3} \\ \text{Var}(X) &= \int_0^1 \int_0^y 2\left(x - \frac{1}{3}\right)^2 dx dy = \frac{1}{18} \\ \text{Var}(Y) &= \int_0^1 \int_x^1 2\left(y - \frac{2}{3}\right)^2 dy dx = \frac{1}{18} \\ \text{Cov}(X,Y) &= \int_0^1 \int_0^y 2\left(x - \frac{1}{3}\right)\left(y - \frac{2}{3}\right) dx dy = \frac{1}{36} \end{aligned}$$

Por lo tanto, la matriz de correlación está dada por:

$$\mathbf{R} = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$$

Proposición 5.3. *Sean* $\boldsymbol{X} = (X_1, \dots, X_n)^T$, $\boldsymbol{Y} = (Y_1, \dots, Y_m)^T$, $\boldsymbol{A} \in M_{a\times n}, \boldsymbol{B} \in M_{b\times m}, \boldsymbol{C} \in M_{a\times 1}, \boldsymbol{D} \in M_{b\times 1}$, *entonces*

(a) ***Linealidad.***

$$E(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}) = \boldsymbol{A}E(\boldsymbol{X}) + \boldsymbol{b}$$

y

$$Var(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}) = \boldsymbol{A}\,Var(\boldsymbol{X})\boldsymbol{A}^T$$

(b) ***Simetría.*** $\Sigma = Var(\boldsymbol{X})$ *es una matriz simétrica,* $\Sigma = \{(\sigma_{ij})\} = \{(\sigma_{ji})\} = \Sigma^T$.

(c) ***Semidefinida positiva.*** $\Sigma = Var(\boldsymbol{X})$ *es una matriz semidefinida positiva, es decir*

$$\boldsymbol{a}^T\Sigma\boldsymbol{a} \geq 0$$

para todo $\boldsymbol{a} \in \mathbb{R}^n$.

(d) ***Independencia y definida positiva.*** $\Sigma = Var(\boldsymbol{X})$ *es una matriz definida positiva, es decir*

$$\boldsymbol{a}^T\Sigma\boldsymbol{a} \geq 0$$

para todo $a \neq 0$, *si y solo si* $X_1, \dots, X_n$ *son linealmente independiente.*

(e) ***Forma alternativa.*** $Cov(\boldsymbol{X}, \boldsymbol{Y}) = E(\boldsymbol{X}\boldsymbol{Y}^T) - E(\boldsymbol{X})E(\boldsymbol{Y})^T, Cov(\boldsymbol{X}, \boldsymbol{X}) = Var(\boldsymbol{X})$ *y* $Cov(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{C}, \boldsymbol{B}\boldsymbol{Y} + \boldsymbol{D}) = \boldsymbol{A}\,Cov(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{B}^T$.

Note que $\text{Cov}(\mathbf{X}, \mathbf{Y}) = \text{Cov}(\mathbf{Y}, \mathbf{X})^T$, de modo que $\text{Var}(\mathbf{X} \pm \mathbf{Y}) = \text{Var}(\mathbf{X} \pm \mathbf{Y}) = \text{Var}(\mathbf{X}) + \text{Var}(\mathbf{Y}) \pm \text{Cov}(\mathbf{X}, \mathbf{Y}) \pm \text{Cov}(\mathbf{Y}, \mathbf{X})$

6. Distribuciones Multivariadas Especiales

6.1. Distribución Multinomial

Consideremos un experimento aleatorio con k resultados posibles, $R_1, \dots, R_k$, exhaustivos y muamente excluyentes, es decir

$$\bigcup_{i=1}^{k} R_i = \Omega, \quad R_i \cap R_j = \emptyset \quad i \neq j$$

Sea $p_i = P(R_i)$ para $i = 1, \dots, k$, por lo que

$$p_i \geq 0 \ \text{ y } \ \sum_{i=1}^{k} p_i = 1 \quad (\star\star)$$

Para n repeticiones independientes de este experimento definamos las variables aleatorias

$$X_i = \text{número de veces que ocrre } R_i$$

Entonces $X_i(\omega) \in \mathcal{X}_i = \{0, 1, \dots, n\}$ y $\sum_{i=1}^{k} X_i(\omega) = n$ para todo $\omega \in \Omega$. Luego en n repeticiones independientes, la probabilidad del evento

$$\begin{aligned}\{X_1 = x_1, \dots, X_k = x_k\} :&= \bigcap_{i=1}^{k} \{X_i = x_i\} \\ &= \bigcap_{i=1}^{k} \{R_i \text{ ocurre exactamente } x_i \text{ veces}\}\end{aligned}$$

está dada por

$$P(X_1 = x_1, \dots, X_k = x_k) = \begin{cases} \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}, & (x_1, \dots, x_k) \in \mathcal{X} \\ 0, & eoc. \end{cases} \quad (\star)$$

Donde $\mathcal{X} = \{(x_1, \dots, x_k) : x_i = 1, \dots, n, \sum_{i=1}^{k} x_i = n\}$ es el recorrido del vector.

Lo que acabamos de realizar es construir un nuevo tipo de distribución sobre vectores aleatorios.

Definición 6.1. *Se dice que un vector aleatorio discreto $(X_1, \dots, X_k)$ tiene distribución multinomial con parámetros n y $(p_1, \dots, p_k)$, si las variables aleatorias $X_1, \dots, X_k$ tiene fmp conjunta dada $(\star)$, en tal caso se denota*

$$(X_1, \dots, X_k) \sim Mult(n, p_1, \dots, p_k)$$

donde $n \in \{1, 2, \dots\}$ y $p_1, \dots, p_k$ satisface $(\star\star)$.

Nota 6.1. Ya que $\sum_{i=1}^{n} X_i = n$ ($X_1, \dots, X_k$ son linealmente dependientes), basta con especificar la distribución conjunta de $X_1, \dots, X_{k-1}$.

Ejemplo 6.1. En 12 lanzamientos de un dado honesto, sea X_i = número de veces que aparece el i-ésimo resultado $i = 1, \dots, 6$, entonces

$$(X_1, \dots, X_6) \sim Mult(n = 12, 1/6, \dots, 1/6)$$

Luego

$$P(X_1 = 2, X_2 = 3, X_3 = 2, X_4 = 2, X_5 = 2, X_6 = 3) = \frac{12!}{(2!)^3(3!)^3}(1/6)^{12}$$

Propiedades. Sea $(X_1, \dots, X_n)$ un vector aleatorio.

(a) Si $(X_1, \dots, X_n) \sim Mult(n, p_1, \dots, p_k)$, entonces la fmg conjunta de $X_1, \dots, X_k$ es

$$\begin{aligned} M_{X_1,\dots,X_k}(t_1, \dots, t_k) &= E(e^{\sum_{i=1}^k t_i X_i}) \\ &= \sum_{(x_1,\dots,x_k)\in\mathcal{X}} e^{\sum_{i=1}^n t_i x_i} \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} \\ &= \sum_{(x_1,\dots,x_k)\in\mathcal{X}} \frac{n!}{x_1! \dots x_k!} (p_1 e^{t_1})^{x_1} \dots (p_k e^{t_k})^{x_k} \\ &= (p_1 e^{t_1} + \dots + p_k e^{t_k})^n \end{aligned}$$

para todo $(t_1, \dots, t_k) \in \mathbb{R}^k$.

(b) Para todo $m = 1, \dots, k$

$$(X_1, \dots, X_m, n - \sum_{i=1}^m X_i) \sim Mult(n, p_1, \dots, p_m, 1 - \sum_{i=1}^m p_i)$$

En particular

$$(X_i, n - \sum_{l\neq i} X_i) \sim Bin(n, p_i, 1 - \sum_{l\neq i} p_l)$$

Luego $E(X_i) = np_i$ y $\text{Var}(X_i) = np_i(1 - p_i)$, con $i = 1, \dots, k$.

$$(X_i, X_j, n - \sum_{l\neq i,j} X_l) \sim Trin(n, p_i, p_j, 1 - \sum_{l\neq i,j} p_l)$$

Entonces $\text{Cov}(X_i, X_j) = -np_ip_j$ para todo $i \neq j$.

(c) Suponga que se reagrupan los resultados $R_1, \dots, R_k$ en m grupos exhaustivos y excluyentes $S_1, \dots, S_m$, es decir

$$\bigcup_{j=1}^m S_k j = \Omega \quad \text{y} \quad S_i \cap S_j = \emptyset$$

para todo $i \neq j$, de modo que

$$P(S_j) = \pi_j \geq 0, \;\; j = 1, \dots, m \;\text{ y }\; \sum_{j=1}^{m} \pi_j = 1$$

Sea Y_i el número de veces que ocurre S_i en las n repeticiones independientes, $j = 1, \dots, m$. Entonces

$$(Y_1, \dots, Y_m) \sim Mult(n, \pi_1, \dots, \pi_m)$$

6.2. Distribución Normal Multivariada

Sea $\mathbf{Z} = (Z_1, \dots, Z_m)^T$, donde $Z_1, \dots, Z_m$ son variables aleatorias iid a $N(0, 1)$. Entonces

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) &= f_{Z_1,\dots,Z_m}(z_1, \dots, z_m) \\ &= \prod_{i=1}^{m} f_{Z_i}(z_i) \\ &= \prod_{i=1}^{m} (2\pi)^{-1/2} e^{-\frac{1}{2} z_i^2} \\ &= (2\pi)^{-m/2} e^{-\frac{1}{2}\sum_{i=1}^{m} z_i^2} \\ &= (2\pi)^{-m/2} e^{-\frac{1}{2}\mathbf{z}^T\mathbf{z}} \end{aligned}$$

para todo $\mathbf{z} = (z_1, \dots, z_m)^T \in \mathbb{R}^m$. Además

$$\begin{aligned} M_{\mathbf{Z}}(\mathbf{t}) &= M_{Z_1,\dots,Z_m}(t_1, \dots, t_m) \\ &= \prod_{i=1}^{m} M_{Z_i}(t_i) \\ &= \prod_{i=1}^{m} e^{\frac{1}{2} t_i^2} \\ &= e^{\frac{1}{2}\sum_{i=1}^{m} t_i^2} \\ &= e^{\frac{1}{2}\mathbf{t}^T\mathbf{t}} \end{aligned}$$

para todo $\mathbf{t} = (t_1, \dots, t_m)^T \in \mathbb{R}^m$.

Notemo también que

$$\begin{aligned} E(Z_i) &= 0, \;\; i = 1, \dots, m, \;\text{ y} \\ \text{Cov}(Z_i, Z_j) &= \begin{cases} 1, & i = j, \;\; i, j = 1, \dots, m \\ 0, & i \neq j, \;\; i, j = 1, \dots, m \end{cases} \end{aligned}$$

Por lo que

$$\begin{aligned} E(\mathbf{Z}) &= (E(Z_1), \dots, E(Z_m))^T = 0 \in \mathbb{R}^m \\ \text{Var}(\mathbf{Z}) &= \{(\text{Cov}(Z_i, Z_j))\} = \mathbf{I}_m \end{aligned}$$

Sean $X_i = \sum_{j=1}^{n} a_{ij} Z_j + \mu_i$ para $i = 1, \ldots, n$, donde a_{ij} y μ_i, $i, j = 1, \ldots, n$, son constantes reales. Sea

$$\mathbf{X} = (X_1, \ldots, X_n)^T$$

Una matriz $\mathbf{A} = \{(a_{ij})\}$ y $\mu = (\mu_1, \ldots, \mu_n)^T$. En forma de matrices se tiene

$$\mathbf{X} = \mathbf{AZ} + \mu$$

donde claramente $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mu \in \mathbb{R}^{n \times 1}$. Entonces

$$\begin{aligned} E(\mathbf{X}) &= E(\mathbf{AZ} + \mu) = \mathbf{A}E(\mathbf{Z}) + \mu = \mu \\ \mathrm{Var}(\mathbf{X}) &= \mathrm{Var}(\mathbf{AZ} + \mu) = \mathbf{A}\mathrm{Var}(\mathbf{Z})\mathbf{A}^T = \mathbf{AA}^T := \Sigma \end{aligned}$$

Además

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{t}) &= E(e^{\mathbf{t}^T \mathbf{X}}) \\ &= E(e^{\mathbf{t}^T (\mathbf{AZ} + \mu)}) \\ &= E(e^{\mathbf{t}^T \mathbf{AZ} + \mathbf{t}^T \mu}) \\ &= e^{\mathbf{t}^T \mu} E(e^{(\mathbf{A}^T \mathbf{t})^T \mathbf{Z}}) \\ &= e^{\mathbf{t}^T \mu} M_{\mathbf{Z}}(\mathbf{A}^T \mathbf{t}) \\ &= e^{\mathbf{t}^T \mu} e^{\frac{1}{2} (\mathbf{A}^t \mathbf{t})^T (\mathbf{A}^T \mathbf{t})} \\ &= e^{\mathbf{t}^T \mu + \frac{1}{2} \mathbf{t}^T \mathbf{AA}^T \mathbf{t}} \\ &= e^{\mathbf{t}^T \mu + \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t}} \end{aligned}$$

para todo $\mathbf{t} \in \mathbb{R}^n$.

Definición 6.2. *Se dice que un vector aleatorio $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ tiene distribución normal n-variada con vector de medias $\mu = (\mu_1, \ldots, \mu_n)^T$ y matriz varianza-covarianza $\Sigma = \{(\sigma_{ij})\}_{i,j=1,\ldots,n}$ lo cual se escribre como $\boldsymbol{X} \sim N_n(\mu, \Sigma)$ si y sólo si*

$$\boldsymbol{X} = \boldsymbol{AZ} + \mu$$

donde $\boldsymbol{AA}^T = \Sigma$,y $\boldsymbol{Z} = (Z_1, \ldots, Z_m)^T \sim N_m(0, \boldsymbol{I}_m)$, es decir, $Z_1, \ldots, Z_m \overset{iid}{\sim} N(0, 1)$.

Otra forma de definirlo es que

$$\mathbf{X} \sim N_n(\mu, \Sigma) \Leftrightarrow M_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t}^T \mu + \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t}}$$

para todo $t \in \mathbb{R}^n$. Equivalentemente, si $\Sigma > 0$ (matriz definida positiva), entonces

$$\mathbf{X} \sim N_n(\mu, \Sigma) \Leftrightarrow f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-n/2} |\Sigma|^{-1/2} e^{-\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)}$$

para todo $\mathbf{x} \in \mathbb{R}^n$.

Propiedades. Sea $\mathbf{X} \sim N_n(\mu, \Sigma)$. Entonces:

(a) $\mathbf{BX} + \mathbf{b} \sim N_m(\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{B\Sigma B}^T)$ para cualquier matriz $\mathbf{B} \in \mathbb{R}^{m\times n}$ y vector $\mathbf{b} \in \mathbb{R}^{m\times 1}$. En efecto

$$\begin{aligned} M_{\mathbf{BX}+\mathbf{b}}(\mathbf{t}) &= E(e^{\mathbf{t}^T(\mathbf{BX}+\mathbf{b})}) \\ &= E(e^{\mathbf{t}^T\mathbf{BX}+\mathbf{t}^T\mathbf{b}}) \\ &= e^{\mathbf{t}^T\mathbf{b}}E(e^{\mathbf{t}^T\mathbf{BX}}) \\ &= e^{\mathbf{t}^T\mathbf{b}}E(e^{(\mathbf{B}^T\mathbf{t})^T\mathbf{X}}) \\ &= e^{\mathbf{t}^T\mathbf{b}}M_{\mathbf{X}}(\mathbf{B}^T\mathbf{t}) \\ &= e^{\mathbf{t}^T\mathbf{b}}e^{\mathbf{t}^T\mathbf{B}\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}^T\mathbf{B\Sigma B}^T\mathbf{t}} \\ &= e^{\mathbf{t}^T\mathbf{B}\boldsymbol{\mu}+\mathbf{b}+\frac{1}{2}\mathbf{t}^T\mathbf{B\Sigma B}^T\mathbf{t}} \end{aligned}$$

Por lo tanto

$$\mathbf{BX} + \mathbf{b} \sim N_m(\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{B\Sigma B}^T)$$

En particular, si $\Sigma > 0$, entonces $\Sigma^{-1/2}(\mathbf{X} - \mu) \sim N_n(\mathbf{0}, \mathbf{I}_n)$, donde $\mathbf{B} = \Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$ y $\Sigma^{1/2}$ es la única raíz cuadra simétrica de Σ.

(b) Sea la partición

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \longrightarrow \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

donde los $\mathbf{X}_j$'s y μ_j's son vectores $n_j \times 1$, los Σ_{ij}'s son matrices $n_i \times n_j$ y $n_1 + n_2 = n$. Entonces:

a) $\mathbf{X}_j \sim N_{n_j}(\mu_j, \Sigma_{jj})$, $j = 1, 2$, donde $n_1 + n_2 = n$.

Dem....

b) $\mathbf{X}_1$ y $\mathbf{X}_2$ si y sólo si $\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \Sigma_{12} = \Sigma_{21}^T = \mathbf{0}$.

Dem fgm

(c) Si $\Sigma > 0$, entonces $(\mathbf{X} - \mu)^T\Sigma^{-1}(\mathbf{X} - \mu) \sim \chi^2_n$.

6.2.1. Caso Bivariada

Ejemplo 6.2. Sean $X_1 = X, X_2 = Y$, se tiene una distribución normal bivariada dada por:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}\right)$$

donde claramente $\mu_X = E(X), \mu_Y = E(Y), \sigma_X^2 = \text{Var}(X), \sigma_Y^2 = \text{Var}(Y)$ y $\sigma_{XY} = \text{Cov}(X, Y)$. Notemos con respecto a la correlación, que

$$\sigma_{XY} = \rho_{XY}\sigma_X\sigma_Y$$

donde ρ_{XY} es el coeficiente de correlación entre X e Y, σ_X y σ_Y son las desviaciones estándar de X e Y respectívamente. Si $-1 < \rho_{XY} < 1$, entonces $\Sigma > 0$ y por lo tanto, es invertible. En tal caso, (X, Y) tiene fdp normal bivariada dada por:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} \exp\left\{-\frac{1}{2(1-\rho_{XY}^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 \right.\right.$$
$$\left.\left. -2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$$

para $-\infty < x < \infty$ y $-\infty < y < \infty$.

Aunque la fdp anterior parece bastante complicada, la distribución normal bivariada es una de las más utilizadas. Algunas de sus muchas propiedades incluyen:

1. La distribución margina de X es $N(\mu_X, \sigma_X^2)$
2. La distribución marginal de Y es $N(\mu_Y, \sigma_Y^2)$
3. $\rho_{XY} = 0$ si y sólo si X e Y son independientes.
4. Para constantes cualquiera a, b, la distribución de $aX + bY$ es

$$N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho_{XY}\sigma_X\sigma_Y)$$

Insertar figua

Ejemplo 6.3. Suponga que se selecciona al azar una pareja formada por un hombre y una mujer de una determinada población. Sea X la altura de la mujer e Y la altura del hombre, ambas medidas en pulgadas. Se sabe que la distribución conjunta de X e Y es normal bivariada con medias $\mu_X = 66{,}8$ y $\mu_Y = 70$, desviaciones estándar $\sigma_X = \sigma_Y = 2$ y correlación $\rho_{XY} = 0{,}68$. Calculemos la probabilidad de que la mujer sea más alta que el hombre, es decir, $P(X - Y > 0)$.

Tenemos (X, Y) un vector aleatorio con distribución normal bivariada, se tiene la distribución de $X - Y$ también es normal, con una media

$$E(X - Y) = 66{,}8 - 70 = -3{,}2$$

y varianza

$$\begin{aligned}\text{Var}(X - Y) &= \sigma_X^2 + \sigma_Y^2 - 2\text{Cov}(X, Y) \\ &= 4 + 4 - 2(0{,}68)(2)(2) = 2{,}56\end{aligned}$$

Luego, como la desviación estándar de $X - Y$ es $\sigma_{X-Y} = 1{,}6$, entonces la estandarización de $X - Y$ es $Z = (X - Y + 3{,}2)/1{,}6 \sim N(0, 1)$. Luego

$$\begin{aligned}P(X - Y > 0) &= P(Z > 2) = 1 - \Phi(2) \\ &= 0{,}0227\end{aligned}$$

6.3. Funciones de Vectores Aleatorios II

Sea un vector aleatorio $(X_1,\ldots,X_n)$ en $(\Omega,\mathcal{A},P)$ y n funciones real valoradas $g_i:\mathbb{R}^n\to\mathbb{R}$, $i=1,\ldots,n$, considere la transformación de $\mathbb{R}^n$ para $\mathbb{R}^n$

$$(X_1,\ldots,X_n)\longrightarrow(Y_1,\ldots,Y_n)$$

donde

$$\begin{aligned}Y_1&=g_1(X_1,\ldots,X_n)\\&\vdots\\Y_n&=g_n(X_1,\ldots,X_n)\end{aligned}$$

Entonces $(Y_1,\ldots,Y_n)$ es un vector aleatorio en $(\Omega,\mathcal{A},P)$, ya cada una de sus coordenadas $Y_1,\ldots,Y_n$ es una variable aleatoria en $(\Omega,\mathcal{A},P)$. Si conocemos la distribución de $(X_1,\ldots,X_n)$, deseamos determinar la distribución del nuevo vector aleatorio $(Y_1,\ldots,Y_n)$. Notemos que $\mathbf{Y}=g(\mathbf{X})$ con $\mathbf{g}=(g_1,\ldots,g_n):\mathbb{R}^n\to\mathbb{R}^n$ talque dado $\mathbf{x}=(x_1,\ldots,x_n)$ se tiene

$$\mathbf{g}(\ \mathbf{x})=(g_1(\mathbf{x}),\ldots,g_n(\mathbf{x}))$$

Sea

$$\mathbf{h}(B)=\mathbf{g}^{-1}(B):=\{\mathbf{x}:\mathbf{g}(\mathbf{x})\in B\}$$

con $B\subset\mathbb{R}^n$. Entonces, en cualquier situación, la distribución de probabilidad de $\mathbf{Y}$ puede determinarse como

$$\begin{aligned}P_{\mathbf{Y}}(B)&=P(\mathbf{Y}\in B)\\&=P(\mathbf{g}(\mathbf{X})\in B)\\&=P(\mathbf{X}\in\{\mathbf{x}:\mathbf{g}(\mathbf{x})\in B\})\\&=P_{\mathbf{X}}(\mathbf{g}^{-1}(B))\end{aligned}$$

con $B\subseteq\mathbb{R}^n$. Por lo tanto, para determinar la probabilidad de $P_{\mathbf{Y}}(B)$, basta con determinar la probabilidad de $P_{\mathbf{X}}(\mathbf{g}^{-1}(B))$

6.3.1. Caso Discreto

Si $(X_1,\ldots,X_n)$ e $(Y_1,\ldots,Y_n)$ son discretos, entonces la fmp conjunta de $Y_1,\ldots,Y_n$ está determinada por:

$$\begin{aligned}f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n)&=P(Y_1=y_1,\ldots,Y_n=y_n)\\&=P(g_1(X_1,\ldots,X_n)=y_1,\ldots,g_n(X_1,\ldots,X_n)=y_n)\\&=\sum_{\{(x_1,\ldots,x_n):g(x_1,\ldots,x_n)=y_1,\ldots,g_n(x_1,\ldots,x_n)=y_n\}}P(X_1=x_1,\ldots,X_n=x_n)\\&=\sum_{\{(x_1,\ldots,x_n):g(x_1,\ldots,x_n)=y_1,\ldots,g_n(x_1,\ldots,x_n)=y_n\}}f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\end{aligned}$$

Nota 6.1. Por convenio se determina primero el recorrido de $(Y_1, \ldots, Y_n)$

$$\begin{aligned}\mathcal{Y} &= \{(y_1, \ldots, y_n) : y_i = g_i(x_1, \ldots, x_n), i = 1, \ldots, n, \text{ para algún } (x_1, \ldots, x_n) \in \mathcal{X}\} \\ &= \{(y_1, \ldots, y_n) : f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) > 0\}\end{aligned}$$

donde $\mathcal{X} = \{(x_1, \ldots, x_n) : f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) > 0\}$ es el recorrido dd $(X_1, \ldots, X_n)$.

Ejemplo 6.4. Sean $Y_1 = X_1 + X_2$ e $Y_2 = X_1 - X_2$, entonces

$$\begin{aligned}P(Y_1 = y_1, Y_2 = y_2) &= P(X_1 + X_2 = y_1, X_1 - X_2 = y_2) \\ &= \sum_{\{(x_1,x_2)):x_1+x_2=y_1,x_1-x_2=y_2\}} P(X_1 = x_1, X_2 = x_2)\end{aligned}$$

En este caso, $\mathcal{Y} = \{(y_1, y_2) : x_1 + x_2 = y_1, x_1 - x_2 = y_2, \text{ para todo } (x_1, x_2) \in \mathcal{X}\}$.

Por ejemplo, si

x_1	x_2	$P(X_1 = x_1, X_2 = x_2)$
0	2	1/4
3	4	1/8
1	6	1/8
2	8	1/2

entonces

$y_1 = x_1 + x_2$	$y_2 = x_1 - x_2$	$P(Y_1 = y_1, Y_2 = y_2)$
2	-2	1/4
7	-1	1/8
7	-5	1/8
10	-6	1/2

En particular, dados y_1, y_2 tales que $(y_1, y_2) \in \mathcal{Y}$, existe un único $(x_1, x_2) \in \mathcal{X}$ talque $\mathbf{g}(x_1, x_2) = (y_1, y_2)$. ($\mathbf{g}$ es una función biyectiva) Esto se ve estudiando el sistema

$$\begin{aligned} x_1 + x_2 &= y_1 \\ x_1 - x_2 &= y_2 \end{aligned}$$

que tiene una única solución.

Ejemplo 6.5. Sean X_1 y X_2 variables aleatorias con fmp conjunta dada por

$x_1 \setminus x_2$	0	1
-1	1/7	1/7
0	2/7	1/7
1	1/7	1/7

Sean $g_1(x_1, x_2) = x_1 + x_2$ y $g_2(x_1, x_2) = x_1 x_2$. Claramente, las variables aleatorias

$$Y_1 = g_1(X_1, X_2) = X_1 + X_2 \quad \text{e} \quad Y_2 = g_2(X_1, X_2) = X_1 X_2$$

toman valores en $\mathcal{Y}_1 = \{-1, 0, 1, 2\}$ e $\mathcal{Y}_2 = \{-1, 0, 1\}$, respectívamente.

Entonces, la fmp conjunta de Y_1, Y_2 queda determinada por:

$y_1 \setminus y_2$	-1	0	1
-1	0	1/7	0
0	1/7	2/7	0
1	0	2/7	0
2	0	0	1/7

Notemos tres cosas de $\mathbf{g}(\mathbf{X}) = \mathbf{Y}$

1. La probabilidad de que $P(Y_1 = -1, Y_2 = -1)$ es 0, entonces o bien es la suma de probabilidad nulas de X_1, X_2 o no existe un par (x_1, x_2) talque $x_1 + x_2 = -1$ y $x_1 x_2 = -1$. Por la primera tabla $x_1 \setminus x_2$, se tiene que

$$P(Y_1 = -1, Y_2 = -1) = P(\{\emptyset\}) = 0$$

2. $P(Y_1 = -1, Y_2 = 0) = P(X_1 = -1, X_2 = 0) = 1/7$. Una correspondencia única ya que el único $(-1, 0) \in \mathcal{X}$ es el único par que $x_1 + x_2 = -1, x_1 x - 2 = 0$.
3. $P(Y_1 = 0, Y_2 = 0) = P(X_1 = 0, X_2 = 1) = P(X_1 = 0, X_2 = 0) = 2/7$. Es decir, una correspondencia no única, ya que existen dos pares (x_1, x_2) que satisface la condición.

Vemos que $\mathbf{g}$ es inyectivo, sobreyectivo o biyectivo puede influir en la probabilidad y en la forma de determinarlo. Cosa que más adelante explicaremos y estudiaremos.

Formalizemos el caso discreto para $n = 2$. Si (X_1, X_2) es un vector aleatorio discreto, entonces su recorrido

$$\mathcal{X} = \{(x_1, x_2) : f_{X_1,X_2}(x_1, x_2) > 0\}$$

es un subconjunto contable de valores en $\mathbb{R}^2$. Luego, el recorrido de $(Y_1, Y_2) = (g_1(X_1, X_2), g_2(X_1, X_2))$ es

$$\mathcal{Y} = \{(y_1, y_2) : y_1 = g_1(x_1, x_2), y_2 = g_2(x_1, x_2), \text{para algún}(x_1, x_2) \in \mathcal{X}\}$$

También es un subconjunto contable de valores en $\mathbb{R}^2$, es decir, si (X_1, X_2) es discreto, entonces (Y_1, Y_2) también es discreto y su fmp se determina como

$$f_{Y_1,Y_2}(y_1, y_2) = \sum_{\{(x_1,x_2)\in\mathcal{X}:g_1(x_1,x_2)=y_1,g_1(x_1,x_2)=y_2\}} f_{X_1,X_2}(x_1, x_2)$$

Ejemplo 6.6. Sean X_1 y X_2 variables aleatorias discretas independientes. Se desea encontrar la fpm de $X_1 + X_2$.

- **Paso 1.** Primero, transformamos

$$(X_1, X_2) \xrightarrow{\mathbf{g}} (Y_1, Y_2)$$

donde $Y_1 = X_1 + X_2$ (variables de interés) e $Y_2 = X_2$ (variables auxiliar arbitraria), entonces

$$\begin{aligned} P(Y_1 = y_1, Y_2 = y_2) &= P(X_1 + X_2 = y_1, X_2 = y_2) \\ &= P(X_1 = y_1 - y_2, X_2 = y_2) \\ &= P(X_1 = y_1 - y_2)P(X_2 = y_2) \end{aligned}$$

por la independencia.

- **Paso 2.** Vamos a marginar la variable auxiliar Y_2, entonces

$$P(Y_1 = y_1) = \sum_{\{y_2:(y_1-y_2,y_2)\in\mathcal{X}\}} P(X_1 = y_1 - y_2)P(X_2 = y_2)$$

Aplicación. Si $X_1 \sim P(\lambda_1)$ y $X_2 \sim P(\lambda_2)$ son variables aleatorias independientes, entonces $Y = X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$. En efecto, de la parte anterior

$$\begin{aligned} P(Y = y) &= \sum_{\{k\geq 0, y-k\geq 0\}} P(X_1 = y - k)P(X_2 = k) \\ &= \sum_{k=0}^{y} \frac{\lambda_1^{y-k}e^{-\lambda_1}}{(y-k)!}\frac{\lambda_2^{k}e^{-\lambda_2}}{k!} \; (x_1, x_2 \geq 0, \Rightarrow 0 \leq k \leq y, y \geq 0) \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{y!}\sum_{k=0}^{y}\binom{y}{k}\lambda_1^{y-k}\lambda_2^{k}, \;\; y = 0, 1, 2, \ldots \\ &= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^y}{y!}, \;\; y = 0, 1, 2, \ldots \end{aligned}$$

la cual es la fmp de $Y = X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$.

Extensión. Si $X_i \sim P(\lambda_i)$, $i = 1, 2, \ldots, n$ son variableas aleatorias independientes, entonces $Y = \sum_{i=1}^{n} X_i \sim P(\sum_{i=1}^{n} \lambda_i)$.

Probemos que

(a) Si $X_1 \sim Bin(n_1, p)$ y $X_2 \sim Bin(n_2, p)$ son variables aleatorias independientes, entonces $Y = X_1 + X_2 \sim Bin(n_1 + n_2, p)$. **labla** Notemos que

$$\sum_{k=0}^{y} \frac{\binom{y}{k}\binom{n_1+n_2-y}{n_2-k}}{\binom{n_1+n_2}{n_2}}$$

Es la suma de la hipergeométrica, tomanto $N = n_1 + n_2, K = y, n = n_2$, entonces por definición de hipergeométrica, la suma debe ser 1, luego

$$f_Y(y) = \binom{n_1+n_2}{y}p^y(1-p)^{n_1-n_2}$$

Por lo tanto $Y \sim Bin(n_1 + n_2, p)$

(b) Si $X_1 \sim Bin(n_1, p)$ y $X_2 \sim Bin(n_2, p)$ son variables aleatorias independientes, entonces $Y = X_1 + X_2 \in BN(r_1 + r_2, p)$.

(c) Que $Y = \sum_{i=1}^{n} X_i \sim Bin(n, p)$, donde $X_1, \ldots, X_n \overset{\text{iid}}{\sim} Ber(p)$. Para ello usaremos solo la independencia. Notemos que $\mathcal{X} = \{0,1\}^n$, luego

$$\begin{aligned} f_Y(y) = P(Y = y) &= \sum_{\{(x_1,\ldots,x_n)=\mathbf{x}:g(\mathbf{x})=y\}} P(X_1 = x_1, \ldots, X_n = x_n) \\ &= \sum_{\{(x_1,\ldots,x_n)=\mathbf{x}:g(\mathbf{x})=y\}} p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_n} \\ &= \sum_{\{(x_1,\ldots,x_n)=\mathbf{x}:g(\mathbf{x})=y\}} p^y(1-p)^{n-y} \end{aligned}$$

Todo depende de los sumandos tales que $\sum_{i=1}^{n} x_i = y$, que al final es solo coneto, queremos orden y's unos sin orden y sin repetición, es claro que hay $\binom{n}{y}$ formas de hacerlo, luego

$$f_Y(y) = \binom{n}{y} p^y (1-p)^{n-y}$$

Es decir, $Y \sim Bin(n, p)$.

(d) Pruebe que $Y = \sum_{i=1}^{n} X_i \sim Bin(n, p)$, donde $X_1, \ldots, X_n \overset{\text{iid}}{\sim} Geo(p)$.

6.3.2. Caso Continuo

Sean $\mathbf{X}$ e $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ vectores aleatorios continuos. Al igual que el caso univaraido, se distinguen dos casos dependiendo de la transformación

$$\mathbf{g} : \mathcal{X} \subseteq \mathbb{R}^n \longrightarrow \mathcal{Y} \subseteq \mathbb{R}^n$$

si es inyectiva o no. Si $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n : f_{\mathbf{X}}(\mathbf{x}) > 0\}$ e $\mathcal{Y} = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y} = \mathbf{g}(\mathbf{x})$ para algún $\mathbf{x} \in X\}$.

Caso Inyectivo. Sea $y_i = g_i(x_1, \ldots, x_n)$ para $i = 1, \ldots, n$. Supongamos que para cada $(x_1, \ldots, x_n) \in \mathcal{X}$, existe un único $(y_1, \ldots, y_n) \in \mathcal{Y}$ talque

$$x_i = h_i(y_1, \ldots, y_n), \;\; i = 1, \ldots, n$$

donde $(h_1, \ldots, h_n)$ es la transformación inversa de $(g_1, \ldots, g_n)$. Se define la matriz Jacobiana de la transformación como

$$J := \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial x_n}{y_1} & \cdots & \frac{\partial x_n}{y_n} \end{pmatrix}$$

el cual es una función de $(y_1, \ldots, y_n)$, ya que

$$\frac{\partial x_i}{\partial y_j} = \frac{\partial h_i(y_1, \ldots, y_n)}{\partial y_j}, \;\; i, j = 1, \ldots, n$$

Si dichas derivafas existen y son todas continuas con $J \neq 0$ para todo $(y_1, \ldots, y_n) \in \mathcal{Y}$, entonces $Y_1, \ldots, Y_n$ tiene fdp conjunta dada por:

$$
\begin{aligned}
&f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n)\\
&= \begin{cases} |J| f_{Y_1,\ldots,Y_n}(h_1(y_1, \ldots, y_n), \ldots, h_n(y_1, \ldots, y_n)), & (y_1, \ldots, y_n) \in \mathcal{Y} \\ \\ 0, & eoc. \end{cases}
\end{aligned}
$$

Idea de la demostración. Sea $\mathbf{h}(B) = \mathbf{g}^{-1}(B) = \{\mathbf{x} : \mathbf{g}(x) \in B\}$, con $B \subset \mathbb{R}^n$. Entonces

$$
\begin{aligned}
P_{\mathbf{Y}}(B) &= P(\mathbf{Y} \in B)\\
&= P(\mathbf{g}(\mathbf{X}) \in B)\\
&= P(\mathbf{X} \in \mathbf{g}^{-1}(B))\\
&= P(\mathbf{X} \in \mathbf{h}(B))\\
&= \int_{\mathbf{h}(b)} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \;\; (\mathbf{y} = \mathbf{g}(\mathbf{x}), \Rightarrow \mathbf{x} = \mathbf{g}^{-1}(\mathbf{y}) = \mathbf{h}(\mathbf{y}))\\
&\int_{B} f_{\mathbf{X}}(\mathbf{h}(\mathbf{y}))|J| d\mathbf{y}, \;\; (\mathbf{x} \in \mathbf{h}(B) \Rightarrow \mathbf{y} = \mathbf{g}(\mathbf{x}) \in B)
\end{aligned}
$$

de acuerdo con la fórmula para el cambio de variables en integrables múltiples. Este resultado se obtiene a partir del siguiente lema:

Lema 6.1. *Sea $\boldsymbol{Y}$ un vector aleatorio continuo de dimensión n. Si para cada $B \subset \mathbb{R}^n$,*

$$P_{\boldsymbol{Y}}(B) = P(\boldsymbol{Y} \in B) = \int_B f(\boldsymbol{y}) d\boldsymbol{y}$$

entonces $f(\boldsymbol{y}) = f_{\boldsymbol{Y}}(\boldsymbol{y})$, es decir, la fdp de $\boldsymbol{Y}$.

Formalizemos para $n = 2$

Teorema 6.1. *Sea (X_1, X_2) un vector aleatorio continuo con fdp conjunta $f_{(X_1,X_2)}$. Sean $\mathcal{X} = \{(x_1, x_2) : f_{(X_1,X_2)}(x_1, x_2) > 0\} \subseteq \mathbb{R}^2$. Suponga que*

(a) $y_1 = g_1(x_1, x_2)$ e $y_2 = g_2(x_1, x_2)$ define una transformación inyectiva de $\mathcal{X}$ en $\mathcal{Y} = \{(y_1, y_2) : y_1 = g_1(x_1, x_2), y_2 = g_2(x_1, x_2)$ para algún $(x_1, x_2) \in \mathcal{Y}\} \subseteq \mathbb{R}^2$.

(b) Las derivadas parciales de la transformación inversa $x_1 = h_1(y_1, y_2)$ y $x_2 = h_2(y_1, y_2)$ son continua sobre $\mathcal{Y}$.

(c) El Jacobiano de la transformación

$$J = \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} \neq 0$$

para $(y_1, y_2) \in \mathcal{Y}$. Entonces la fdp conjunta de $Y_1 = g_1(X_1, X_2)$ y $Y_2 = g_2(X_1, X_2)$, está dada por:

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} |J| f_{X_1,X_2}(h_1(y_1, y_2), h_2(y_1, y_2)), & (y_1, y_2) \in \mathcal{Y} \\ 0, & eoc. \end{cases}$$

Por ejemplo, para $n = 2$, la transformación

$$\begin{aligned} y_1 = x_1 + x_2 &\longrightarrow g_1(x_1, x_2) = x_1 + x_2 \\ y_2 = x_1 - x_2 &\longrightarrow g_2(x_1, x_2) = x_1 - x2 \end{aligned}$$

es inyectiva en $\mathbb{R}^2$. La transformación inversa es

$$\begin{aligned} x_1 = (y_1 + y - 2)/2 &\longrightarrow h_1(x_1, x_2) = (y_1 + y_2)/2 \\ x_2 = (y_1 - y_2)/2 &\longrightarrow h_2(x_1, x_2) = (y_1 - y_2)/2 \end{aligned}$$

y el jacobiano de la transformación es:

$$J = \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} = \det \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} = -\frac{1}{2}$$

Entonces $|J| = \frac{1}{2}$. Así

$$f_{Y_1,Y_2}(y_1, y_)= \begin{cases} \frac{1}{2} f_{X_1,X_2}((y_1 + y_2)/2, (y_1 - y_2)/2), & (y_1, y_2) \in \mathcal{Y} \\ 0, & eoc. \end{cases}$$

Aplaciones.

(a) Sean $X_1, X_2 \overset{\text{iid}}{\sim} \exp(1)$, es decir

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) = \begin{cases} e^{-x_1 - x_2}, & x_1 > 0, x_2 > 0 \\ 0, & eoc. \end{cases}$$

Queremos determinar la fdp de $(Y_1, Y_2) = (X_1 + X_2, X_1 - X_2)$. Notemos que $x_1, x_2 > 0$ entonces $y_1 = x_1 + x_2 > 0$. Luego

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= \begin{cases} \frac{1}{2} e^{-\frac{y_1+y_2}{2} - \frac{y_1-y_2}{2}}, & \frac{y_1+y_2}{2} > 0, \frac{y_1-y_2}{2} > 0 \\ 0, & eoc. \end{cases} \\ &= \begin{cases} \frac{1}{2} e^{-y_1}, & y_2 > y_1, y_2 < -y_1, y_1 > 0 \\ 0, & eoc. \end{cases} \\ &= \begin{cases} \frac{1}{2} e^{-y_1}, & |y_2| < y_1, y_1 > 0 \\ 0, & eoc. \end{cases} \end{aligned}$$

(b) Sean $X_1, X_2 \overset{\text{iid}}{\sim} U(0, 1)$, es decir

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) = \begin{cases} 1, & 0 < x_1 < 1, 0 < x_2 < 1 \\ 0, & eoc. \end{cases}$$

Entonces, la fdp de $(Y_1, Y_2) = (X_1 + X_2, X_1 - X_2)$ es

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \frac{1}{2}, & 0 < \frac{y_1+y_2}{2} < 1, 0 < \frac{y_1-y_2}{2} < 1 \\ 0, & eoc. \end{cases}$$
$$= \begin{cases} \frac{1}{2}, & -y_1 < y_2 < 2 - y_1, 0 < y_1 < 1, -(2 - y_1) < y_2 < y_1, 1 < y_1 < 2 \\ 0, & eoc. \end{cases}$$

(**Tarea.** Bosquejar en (a),(b) los recorrido y determinar las distribuciones marginales Y_1 e Y_2)

(c) Sean $X_1, X_2 \overset{\text{iid}}{\sim} N(0, 1)$, es decir

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2) = (2\pi)^{-1}e^{-x_1^2/2}e^{-x_2^2/2},\ \ (x_1, x_2) \in \mathbb{R}^2$$

Entonces

$$Y_1 = X_1 + X_2 \sim N(0, 2)$$
$$Y_2 = X_1 - X_2 \sim N(0, 2)$$

(Bosqueje los recorridos ene ste caso. Obtener la fdp conjunta de Y_1 e Y_2 son Y_1 e Y_2 variables aleatorias independientes?)

Ejemplo 6.6. Sean X_1 e X_2 variables aleatorias continuas independientes. Se desea la fdp de $X_1 + X_2$.

- **Paso 1.** Sean $Y_1 = X_1 + X_2$ (variable de interés) e $Y_2 = X_2$ (variables auxiliar arbitraria). Entonces, $y_1 = g_1(x_1, x_2) = x_1 + x_2$ e $y_2 = g_1(x_1, x_2) = x_2$, con inversos $x_1 = h_1(y_1, y_2) = y_1 - y_2$ y $x_2 = h - 2(y_1, y_2) = y_2$. Luego

$$J = \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} = \det \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = 1$$

y

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(y_1 - y_2, y_2) = f_{X_1}(y_1 - y_2)f_{X_2}(y_2)$$

- **Paso 2.** $f_{Y_1}(y_1) = \int_{\{y_2:(y_1-y_2,y_2)\in\mathcal{X}\}} f_{X_1}(y_1 - y_2)f_{X_2}(y_2)dy_2$. Por ejemplo, si $Y = X_1 + X_2$, donde $X_1, X_2 \overset{\text{iid}}{\sim} \exp(\lambda)$, entonces

$$f_Y(y) = \begin{cases} \int_0^y \lambda^2 e^{-\lambda(y-z+z)}dz, & y > 0 \\ 0, & eoc. \end{cases}$$
$$= \begin{cases} \lambda^2 y e^{-\lambda y}, & y > 0 \\ 0, & eoc. \end{cases}$$

Es decir, $Y = X_1 + X_2 \sim Gamma(2, \lambda)$.

Extensión. Si $X_i \sim Gamma(\alpha_i, \lambda)$, $i = 1, \ldots, n$ son variables aleatorias independientes, entonces $Y = \sum_{i=1}^n X_i \sim Gamma(\sum_{i=1}^n \alpha_i, \lambda)$.

Caso No Inyectivo. En muchos casos donde $\mathbf{g} : \mathcal{X} \longrightarrow \mathcal{Y}$ no es inyectivo, también podemos determinar la distribución $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ mediante el método del jacobiano descrito anteriormente.

Para esto, es suficiente que $\mathbf{g} = (g_1, \ldots, g_n)$ sea inyectivo cuando se restringe a cada una de k regiones (abiertas) disjuntas cuya unión contenga el valor de $\mathbf{X}$ con probabilidad 1. El siguiente procedimiento es para el caso bivariado. La extensión para más casos es similar.

Si la transformación no es inyectiva, sea $\{B_0, B_1, \ldots, B_k\}$ una partición de $\mathcal{X} = \{(x_1, x_2) : f_{X_1,X_2}(x_1, x_2) > 0\}$. Supongamos que

(a) El conjunto B_0 (posiblemente vacío) satisface $P((X_1, X_2) \in B_0) = 0$.

(b) Para cada $i = 1, \ldots, k$, la transformación

$$(y_1, y_2) = (g_1(x_1, x_2), g_2(x_1, x_2))$$

es inyectivo desde B_i sobre $\mathcal{Y}$.

Entonces para cada $i = 1, \ldots, k$ existe la transformación inversa desdes $\mathcal{Y}$ a B_i, sea la i-ésima transformación inversa por

$$(x_{1i}, x_{2i}) = (h_{1i}(y_1, y_2), h_{2i}(y_1, y_2)), \quad i = 1, \ldots, k$$

Sea J_i es Jacobiano calculado desde la i-ésima inversa, es decir

$$J_i = \det \begin{pmatrix} \frac{\partial x_{1i}}{\partial y_1} & \frac{x_{1i}}{\partial y_2} \\ \frac{\partial x_{2i}}{\partial y_1} & \frac{\partial x_{2i}}{\partial y_2} \end{pmatrix}, \ i = 1, \ldots, k$$

Supongamos que los Jacobianos no son cero en $\mathcal{Y}$, entonces bajo las condiciones anteriores, la fdp de (Y_1, Y_2) está dada por:

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \sum_{i=1}^k |J_i| f_{X_1,X_2}(h_{1i}(y_1, y_2), h_{2i}(y_1, y_2)), & (y_1, y_2) \in \mathcal{Y} \\ 0, & (y_1, y_2) \notin \mathcal{Y} \end{cases}$$

Ejemplo 6.7. Sean $X_1, X_2 \overset{\text{iid}}{\sim} N(0, 1)$. Considere la transformación

$$Y - 1 = \frac{X_1}{X_2} \ \text{ y } \ Y_2 = |X_2|$$

Esta es una transformación desde $\mathcal{X} = \mathbb{R}^2$ a $\mathcal{Y} = \{(y_1, y_2) : y_2 > 0\} \subseteq \mathbb{R}^2$ que no es inyectiva, ya que los puntos (x_1, x_2) y $(-x_1, -x_2)$ de $\mathcal{X}$ son ambos mapeado en el mismo punto (y_1, y_2) de $\mathcal{Y}$.

Pero si nos restringimos a los valores positivos o negativos de x_2, entonces la transformación restringida es inyectiva en cada caso.

Sean $B_0 = \{(x_1, x_2) : x_2 = 0\}, B_1 = \{(x_1, x_2) : x_2 > 0\}$ y $B_2 = \{(x_1, x_2) : x_2 < 0\}$. Entonces , B_0, B_1 y B_2 forman una partición de $\mathcal{X} = \mathbb{R}^2$ talque

(a) $P((X_1, X_2) \in B_0) = P(X_2 = 0) = 0$.

(b) $\mathcal{Y} = \{(y_1, y_2) : y_2 > 0\}$ es la imagen tanto de B_1 como de B_2 bajo la transformación $(g_1(x_1, x_2), g_2(x_1, x_2)) = (x_1/x_2, |x_2|)$.

Además, las transformaciones inversas de $\mathcal{Y}$ a B_1 e $\mathcal{Y}$ a B_2 están dadas por

$$
\begin{aligned}
\mathbf{h}_1 : \mathcal{Y} \to B_1, \Rightarrow x_{11} = h_{11}(y_1, y-2) = y-1y_2,\; x_{21} = h_{21}(y_1, y_2) = y_2 \\
\mathbf{h}_2 : \mathcal{Y} \to B, \Rightarrow x_{12} = h_{12}(y_1, y_2) = -y_1y_2,\; x_{22} = h_{22}(y_1, y_2) = -y_2
\end{aligned}
$$

Notemos que la primera transformación inversa produce valores positivos de x_2 mientras que el segundo inverso da valores negativos de x_2. También notemos que los jacobianos de las dos transformaciones inversas son $|J_1| = |J_2| = y_2$ para todo $(y_1, y_2) \in \mathcal{Y}$. Luego, usando que $f_{X_1,X_2}(x_1, x_2) = \phi(x_1)\phi(x_2) = (2\pi)^{-1}e^{-x_1^2/2}e^{-x_2^2/2}$ para $< -infty < x_1 < \infty$ y $-\infty < x_2 < \infty$, se obtiene que

$$
\begin{aligned}
&f_{Y_1,Y_2}(y_1, y_2) \\
&= \begin{cases} \frac{|y_2|}{2\pi}e^{-(y_1y_2)^2/2}e^{-y_2^2/2} + \frac{|y_2|}{2\pi}e^{-(-y_1y_2)^2/2}e^{-(-y_2)^2/2}, & y_1 \in \mathbb{R}, y_2 \in (0, \infty) \\ \\ 0, & eoc. \end{cases} \\
&= \begin{cases} \frac{y_2}{\pi}e^{-(1+y_1^2)y_2^2/2}, & y_1 \in \mathbb{R}, y_2 \in (0, \infty) \\ 0, & eoc. \end{cases}
\end{aligned}
$$

La fdp margina de Y_1 puede determinarse como

$$
\begin{aligned}
f_{Y_1}(y_1) &= \int_0^\infty \frac{y_2}{\pi}e^{-(1+y_1^2)y_2^2/2}dy_2 \\
&= \frac{1}{2\pi}\int_0^\infty e^{-(1+y_1^2)z/2}dz \\
&= \frac{1}{\pi(1+y_1^2)}
\end{aligned}
$$

para todo $y_1 \in \mathbb{R}$. De aquí se concluye que la razón de dos variables normales estándar independientes, es una variable aleatoria con distribución de Cauchy.

Además, es fácil ve rque $f_{Y_2}(y_2) = 2\phi(y_2)$ para $y_2 > 0$, es decir, si $X_2 \sim N(0, 1)$, entonces $Y_2 = |X_2| \sim HN(0, 1)$ (media normal/half normal).

Ejemplo 6.8. Asumiendo nuevamente que $X_1, X_2 \overset{\text{iid}}{\sim} N(0, 1)$, pruebe que

$$
\begin{aligned}
Y_1 &= \frac{X_1 + X_2}{2} \sim N(0, 1/2) \\
Y_2 &= \frac{(X_1 - X_2)^2}{2} \sim \chi_1^2 := Gamma(1/2, 1/2)
\end{aligned}
$$

y que Y_1 e Y_2 son variables aleatorias independientes.

Extensión. Sean $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$. Entonces, la media muestra

$$X = \frac{1}{n} \sum_{i=1}^{n} X_i$$

y la varianza muestra

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - X)^2$$

son variables aleatorias independientes, con

$$X \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ Y } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

para cada $n \geq 2$.

Ejemplo 6.9. Suponga que $X_1 \sim N(0,1)$ y $X_2 \sim \chi^2_\eta$ con $(\eta > 0)$ son independientes. Entonces, la distribución de la variables aleatoria

$$Y = \frac{X_1}{\sqrt{X_2/\eta}}$$

se llama t (de student) con η grados de libertad, y se denota como $Y \sim t_\eta := t(0, 1, \eta)$

Tarea. probar que

$$f_Y(y) = \frac{\Gamma[(\eta+1)/2]}{\Gamma(\eta/2)\sqrt{\pi\eta}} \left(1 + \frac{y^2}{\eta}\right)^{-(\eta+1)/2}, \quad -\infty < y < \infty$$

Nota 6.. $P(Y \leq y) = \int_{-\infty}^{y} f_Y(z)dz$ se encuentra tabulada para varios valores de y y η.

Aplicación. Sean $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$. Vimos que $X = \frac{1}{n}\sum_{i=1}^{n} X_i$ y $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - X)^2$ son variables aleatorias independientes, con

$$X \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ y } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Es decir

$$\frac{\sqrt{n}(X-\mu)}{\sigma} \sim N(0,1) \text{ y } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

y son independientes. Luego

$$\frac{\sqrt{n}(X-\mu)/\sigma}{\sqrt{(n-1)S^2/(n-1)\sigma 2}} = \frac{\sqrt{n}(X-\mu)}{S} \sin t_{n-1}$$

Ejemplo 6.9. Sean $X_1 \sim X_r^2$ y $X_2 \sim X_s^2$ con $(r, s > 0)$ variables aleatortias independientes. La distribución de la variable aleatoria

$$Y = \frac{X_1/r}{X_2/s}$$

se llama F (de Fishcer) con r y s grados de libertad, y se denota como $Y \sim F_{r,s}$. Note que $Y^{-1} \sim F_{s,r}$.

Tarea. Probar que

$$f_Y(y) = \begin{cases} \frac{\Gamma[(r+s)/2]}{\Gamma(r/2)\Gamma(s/2)} \left(\frac{r}{s}\right)^{r/2} y^{(r/2)-1} \left(1 + \frac{r}{s}y\right)^{-(r+s)/2}, & y > 0 \\ 0, & eoc. \end{cases}$$

Nota 6.. $P(Y \leq y) = \int_{-\infty}^{y} f_Y(z)dz$ se encuentra tabulada para varios valores de y, r y s.

Aplicación. Suponga que

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu_X, \sigma^2)$$

e

$$Y_1, \ldots, Y_m \overset{\text{iid}}{\sim} N(\mu_Y, \sigma^2)$$

donde los X_i's son independientes de los Y_i's. Entonces

$$(n-1)S_X^2/\sigma^2 = \sum_{i=1}^{n}(X_1 - X)^2/\sigma^2 \sim \chi_{n-1}^2$$

y

$$(m-1)S_Y^2/\sigma^2 = \sum_{i=1}^{m}(Y_i - Y)^2/\sigma^2 \sim \chi_{m-1}^2$$

y son variables independientes. Luego

$$\frac{S_X^2}{S_Y^2} = \frac{\sum_{i=1}^{n}(X_i - X)^2/(n-1)}{\sum_{i=1}^{m}(Y_i - Y)^2/(m-1)} \sim F_{n-1,m-1}$$

Ejemplo 6.10. Sean $X_1 \sim Gamma(\alpha_1, \lambda)$ y $X_2 \sim Gamma(\alpha_2, \lambda)$ con $(\alpha_1, \alpha_2, \lambda > 0)$ variables aleatorias independientes. la distribución de la variable aleatoria

$$Y = \frac{X_1}{X_1 + X_2}$$

se llama beta con parámetros α_1, α_2, y se denota como $Y \sim Beta(\alpha_1, \alpha_2)$.

Tarea. Pruebe que

$$f_Y(y) = \begin{cases} \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha)\Gamma(\alpha_2)} y^{\alpha-1-1}(1-y)^{\alpha_2-1}, & 0 < y < 1 \\ 0, & eoc. \end{cases}$$

Tarea. Pruebe que $Y_1 = \frac{X_1}{X_1+X_2}$ e $Y_2 = X_1 + X_2$ son independientes.

Extensión. Si $X_1, \ldots, X_n \overset{\text{iid}}{\sim} Gamma(\alpha_i, \lambda)$, entonces

$$\frac{X_1 + \cdots + X_m}{X_1 + \cdots + X_m + X_{m+1} + \cdots + X_n} \sim Beta(\alpha_1 + \cdots + \alpha_m, \alpha_{m+1} + \cdots + \alpha_n)$$

donde hay que recordar que $\sum_{i=1}^k X_i \sim Gamma\left(\sum_{i=1}^k \alpha_i, \lambda\right)$ para cada $k = 1, \ldots, n$.

Nota 6.. La función beta se define como

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},\ \ \alpha, \beta > 0$$

Nota 6.. Si $X_1, \ldots, X_n$ son variables aleatorias independientes (arbitrarias), entonces funciones de subvectores disjuntos de los X_i's también son independientes. Por ejemplo $Y_1 = X_1 + X_2 + X_3, Y_2 = (X_5 - X_4)^2$, $Y_3 = \text{máx}\{X_6, X_7\}$ e $Y_4 = e^{-Ys}$, son variables aleatorias independientes. Para ilustrar el procedimiento, se probará el siguiente caso especial mediante la fda conjunta.

Teorema 6.2. *Sean $X_1, \ldots, X_n$ variables aleatorias independientes. Para cada $i = 1, \ldots, n$, suponga que $g_i(x_i)$ es una función solo de x_i. Entonces, las transformaciones $Y_1 = g_1(X_1), \ldots, Y_n = g_n(X_n)$ también son variables aleatorias independientes.*

Dem. La fda conjunta de $Y_1 = g_1(X), \ldots, Y_n = g_n(X_n)$ es

$$\begin{aligned}
F_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) &= P(Y_1 \leq y_1, \ldots, Y_n \leq y_n) \\
&= P(\{g_1(X_1) \leq y_1, \ldots, g_n(X_n) \leq y_n\}) \\
&= P(\{g_1(X_1) \in (-\infty, y_1], \ldots, g_n(X_n) \in (-\infty, y_n]\}) \\
&= P(\{X_1 \in g_1^{-1}((-\infty, y_1]), \ldots, X_n \in g_n^{-1}((-\infty, y_n])\} \\
&= P(\{X_1 \in g_1^{-1}((-\infty, y_1])\}) \times \cdots \times P(\{X_n \in g_n^{-1}((-\infty, y_n])\}) \\
&= P(\{g_1(X_1) \in (-\infty, y_1]\}) \times \cdots \times P(\{g_n(X_n) \in (-\infty, y_n]\}) \\
&= P(\{g_1(X_1) \leq y_1\}) \times \cdots \times P(\{g_n(X_n) \leq y_n\}) \\
&= P(\{Y_1 \leq y_1\}) \times \cdots \times P(\{Y_n \leq y_n\}) \\
&= F_{Y_1}(y_1) \times \cdots \times F_{Y_n}(y_n)
\end{aligned}$$

■

Ejemplo 6.11. Sean $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(0, 1)$. Entonces:

(a) $X_1^2, \ldots, X_n^2 \overset{\text{iid}}{\sim} \chi_1^2$, lo que permite probar también **Tarea.** que $\sum_{i=1}^n X_i^2 \sim \chi_n^2$.

(b) $I_{(0,\infty)}(X_1), \ldots, I_{(0,\infty)}(X_n) \overset{\text{iid}}{\sim} Ber(1/2)$, donde

$$I_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \in A^c \end{cases}$$

(c) $\Phi(X_1), \ldots, \Phi(X_n) \overset{\text{iid}}{\sim} U(0,1)$, donde $\Phi(z), z \in \mathbb{R}$, es la fda de la distribución $N(0,1)$.

6.4. Estadísticos de orden

6.4.1. Definición

Sean $X_1, \ldots, X_n$ variables aleatorias, todas bien definidas sobre el mismo espacio de probabilidad. Se define

$$X_{(k)} = k - \text{ésimo menor de } X_1, \ldots, X_n \Rightarrow X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

Definición 6.3. *El vector aleatoria $(X_{(1)}, \ldots, X_{(n)})$ se llama estadísticos de orden, donde $X_{(k)}$ es el k-ésimo orden, y los ordenes extremos son*

$$X_{(1)} = \text{mín}\{x_1, \ldots, X_n\} \;\; y \;\; X_{(n)} = \text{máx}\{X_1, \ldots, X_n\}$$

Nota 6.. Recordemos que hay $n!$ maneras de reodenar n valores $x_1, \ldots, x_n$.

Entonces, lo que hacemos es reodenar del menor al mayor (si es posible) y el nuevo vector generado se le llama estadístico de orden.

6.4.2. Caso iid F

Estudiaremos el caso donde $X_1, \ldots, X_n$ son variables aleatorias iid con fda F, y con fdp $\int f = F'$ en el caso continuo.

Teorema 6.3. *Sean $X_{(1)} = \text{mín}\{X_1, \ldots, X_n\}$ y $X_{(n)} = \text{máx}\{X_1, \ldots, X_n\}$ donde $X_1, \ldots, X_n$ son variables aleatorias iid con fda F. Entonces*

$$F_{X_{(1)}}(x) = 1 - (1 - F(x))^n \;\; y \;\; F_{X_{(n)}}(x) = (F(x))^n$$

Si ademas, la fda F es continua, entonces

$$\begin{aligned} f_{X_{(1)}}(x) &= \frac{dF_{X_{(1)}}(x)}{dx} = n(1 - F(x))^{n-1} f(x) \\ f_{X_{(n)}}(x) &= \frac{dF_{X_{(n)}}(x)}{dx} = n(F(x))^{n-1} f(x) \end{aligned}$$

Idea de la demostración.

$$
\begin{aligned}
F_{X_(1)}(x) = P(X_{(1)} \leq x) \\
&= P(\text{mín}\{X_1, \dots, X_n\} \leq x) \\
&= 1 - P(\text{mín}\{X_1, \dots, X_n\} > x) \\
&= 1 - P(X_1 > x, \dots, X_n > x) \\
&= 1 - P(\bigcap_{i=1}^{n} \{X_i > x\}) \\
&= 1 - \prod_{i=1}^{n} P(X_i > x), \text{ independencia} \\
&= 1 - \prod_{i=1}^{n} (1 - P(X_i \leq x)) \\
&= 1 - \prod_{i=1}^{n} (1 - F(x)), \ (X_i \sim F, \text{ para todo } i) \\
&= 1 - (1 - F(x))^n
\end{aligned}
$$

Similarmente para el máximo

$$
\begin{aligned}
F_{X_(n)}(x) = P(X_{(n)} \leq x) \\
&= P(\text{máx}\{X_1, \dots, X_n\} \leq x) \\
&= P(\bigcap_{i=1}^{n} \{X_i \leq x\}) \\
&= \prod_{i=1}^{n} P(X_i \leq x), \text{ independencia} \\
&= \prod_{i=1}^{n} F(x) \\
&= (F(x))^n
\end{aligned}
$$

Ejemplo 6.12. Sean $X_1, \dots, X_n \overset{\text{iid}}{\sim} \exp(1)$, es decir

$$
F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x}, & x \geq 0 \end{cases}
$$

$$
f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & eoc. \end{cases}
$$

Entonces

$$f_{X_{(1)}}(x) = \begin{cases} ne^{-nx}, & x \geq 0 \\ 0, & eoc. \end{cases}$$

$$f_{X_{(n)}}(x) = \begin{cases} n(1-e^{-x})^{n-1}e^{-x}, & x \geq 0 \\ 0, & eoc. \end{cases}$$

En este caso, es claro que $X_{(1)} \sim \exp(n)$, de modo que $E(X_{(1)}) = 1/n$ y $\mathrm{Var}(X_{(1)}) = 1/n^2$.

Tarea. 1) Detetminar $E(X_{(n)})$ y $\mathrm{Var}(X_{(n)})$. 2) Determinar el rango esperado y el punto medio esperado en este caso. ¿Qué se neceista para estudiar la distribuciones conjuntas y marginales de R_n, M_n?.

Teorema 6.4. *Sean $X_1, \ldots, X_n \overset{iid}{\sim} F$. Entonces, para $k = 1, \ldots, n$, se tiene que*

$$\begin{aligned} F_{X_{(k)}}(x) &= \sum_{m=1}^{m} \binom{n}{m} (F(x))^m (1-F(x))^{n-m}, \quad (\star) \\ &= \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n+1-k)} \int_0^{F(x)} y^{k-1}(1-y)^{n-k} dy, \quad (\star\star) \end{aligned}$$

ya que

$$\sum_{m=1}^{k} \binom{n}{m} z^m (1-z)^{n-m} = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n+1-k)} \int_0^{z} y^{k-1}(1-y)^{n-k} dy$$

para todo $0 \leq z \leq 1$ y $k = 1, \ldots, n$. De $(\star\star)$ se tiene $F_{X_{(k)}}(x) = F_Z(x)$, donde $Z \sim Beta(k, n+1-k)$. En particular, si $X_1, \ldots, X_n \overset{iid}{\sim} U(0,1)$, entonces el k-ésimo menor se distribuye como

$$X_{(k)} \sim Beta(k, n+1-k), \quad k = 1, \ldots, n$$

Dem. Se define....

Teorema 6.5. *Si$X_1, \ldots, X_n \overset{iid}{\sim} F$, entonces la distribución conjunta de $X_{(1)}$ y $X_{(n)}$ está dada por:*

$$\begin{aligned} F_{X_{(1)}X_{(n)}}(x,y) &= F_{X_{(n)}}(y) - P(X_{(1)} > x, X_{(n)} \leq y) \\ &= \begin{cases} (F(y))^n - (F(y)-F(x))^n, & x < y \\ (F(y))^n, & x \geq y \end{cases} \end{aligned}$$

En particular si F es continua, entonces $X_{(1)}$ y $X_{(n)}$ tienen fdp conjunta dada por:

$$\begin{aligned} &f_{X_{(1)},X_{(n)}}(x,y) \\ &\quad = \begin{cases} n(n-1)(F(y)-F(x))^{n-2}f(x)f(y), & x < y \\ 0, & eoc. \end{cases} \end{aligned}$$

Dem. Terminar...

A partir de estos resultados, también se pueden obtener las distribucipnes conjunta y marginales de las variables aleatorias

$$R_n = X_{(n)} - X_{(1)} \ \text{ y } \ M_n = \frac{X_{(n)} + X_{(1)}}{2}$$

Por ejemplo, para el caso continuo se tiene que

$$\begin{aligned} & f_{R_n}(x) \\ & \quad = \begin{cases} n(n-1)\int_{-\infty}^{\infty}(F(x+y) - F(x))^{n-2} f(x+y) f(y) dy, & x > 0 \\ 0, & eoc. \end{cases} \end{aligned}$$

Ejemplo 6.13. Si $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \exp(1)$, entonces

$$\begin{aligned} & f_{X_{(1)}, X_{(n)}}(x, y) \\ & \quad = \begin{cases} n(n-1)(e^{-x} - e^{-y})^{n-2} e^{-x} e^{-y}, & 0 < x < y < \infty \\ 0, & eoc. \end{cases} \end{aligned}$$

y

$$\begin{aligned} f_{R_n}(r) &= \begin{cases} n(n-1)\int_0^{\infty}(e^{-x} - e^{-x-y})^{n-2} e^{-x-y} e^{-x} dy, & 0 < x < \infty \\ 0, & eoc. \end{cases} \\ &= \begin{cases} ne^{-nx}, & 0 < x < \infty \\ 0m & eoc. \end{cases} \end{aligned}$$

Entonces $R_n = X_{(n)} - X_{(1)} \sim \exp(n)$.

Tarea.

(a) Obtenga la distribución del rango muestral $R_n = X_{(n)} - X_{(1)}$ cuando $X_1, \ldots, X_n \overset{\text{iid}}{\sim} U(0,1)$.

(b) Sean $Y_1 = X_{(1)}$, determine la distribución conjunta de $Y_1, \ldots, Y_n$, y revisar si estas variables son independientes.

(c) Para $n = 2$, con $X_1, X_2 \overset{\text{iid}}{\sim} Geom(p)$, pruebe que el rango y el mínimo son independientes. Compare con la misma situación en el caso $\exp(1)$.

(d) Sean $X_1, \ldots, X_n$ variables aleatorias continuas iid con fdp f. Pruebe que

$$f_{X_{(1)}, \ldots, X_{(n)}}(y_1, \ldots, y_n) = \begin{cases} n! \prod_{i=1}^{n} f(y_i), & y_1 < y_2 < \ldots < y_n \\ 0, & eoc. \end{cases}$$

6.5. Distribución Condicional

6.5.1. Definición y Ejemplos

Sea (X, Y) un vector aleatorio en $(\Omega, \mathcal{A}, P)$, queremos determinar la distribución de probabilidad condicional dde Y cuando se sabe que $X = x$ para algún x en el recorrido de X, es decir, se quiere calcular

$$P(Y \in B | X = x)$$

para cualquier subconjunto B de números reales.

Ejemplo 6.14. Suponga que lanza un dado justo dos veces. Si X_1, X_2 son los puntajes del primer y segundo lanzamiento, respectívamente, defina $X = X_1 + X_2$ e $Y = \text{mín}\{X_1, X_2\}$. Note que X e Y son discretas. Para calcula $P(Y = 2|X = 7)$ considere los eventos $A = \{X = /\}, B = \{Y = 2\}$. Es claro que $P(A) = 6/36$ y $P(A \cup B) = 2/36$. Entonces

$$\begin{aligned} P(Y = 2|X = 7) = P(B|A) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{P(X?7, Y = 2)}{P(X = 7)} \\ &= \frac{2/36}{6/36} = \frac{1}{3} \end{aligned}$$

De forma general, si (X, Y) es un vector aleatorio discreto, entonces

$$P(Y = y|X = x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

donde $f_X(x) = P(X = x) > 0$. Cuando $f_X(x) = 0$, esta probabilidad se define de forma arbitraria, digamos $P(Y = y|X = x) = 0$. Definimos

$$f_{Y|X=x}(y) = P(Y = y|X = x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)}, & f_X(x) > 0 \\ 0, & eoc. \end{cases}$$

Entonces se cumple la siguientes propiedades

(a) $0 \leq f_{Y|X=x} f(y) \leq 1$ para todo $(x, y) \in \mathbb{R}^2$.

(b) $\sum\limits_{y \in \mathbb{R}} f_{Y|X=x}(y) = \sum\limits_{y \in \mathbb{R}} \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\sum_{y \in \mathbb{R}} f_{X,Y}(x,y)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1.$

Entonces, a partir de las propiedades, se deduce que la función $f_{Y|X=x}(y)$ es una fmp en $\mathbb{R}$, llamada fmp condicional de Y dado $X = x$. De forma análoga se define $f_{X|Y=y}(x)$ como la fmp condicional de X dado $Y = y$.

Para la forma continua se procede de forma similar.

Nota 6.. Otra notación para las fmp o fdp condiciona de Y dado $X = x$ es $f(y|x)$ o $(f(x|y))$.

Definición 6.4. *Sea (X, Y) un vector aleatorio discreto o continup con fmp (c.d.) o fdp (c.c.) conjunta $f_{X,Y}(x, y)$ y fmp's (c.d.) o fdp's (c.c.) marginales $f_X(x)$ y $f_Y(y)$. La fmp (c.d.) o fdp (c.c.) condicional de Y dado $X = x$ se define como*

$$f_{Y|X=x}(y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)}, & f_X(x) > 0 \\ 0, & eoc. \end{cases}$$

Análogamente, fmp (c.d.) o fdp (c.c.) condicional de X dado $Y = y$ se define como

$$f_{X|Y=y}(x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)}, & f_Y(y) > 0 \\ 0, & eoc. \end{cases}$$

Teorema 6.6. *La distribución de probabilidad condicional de Y dado $X = x$, estpa dada por:*

$$P(Y \in B|X = x) = \begin{cases} \sum_{y \in B} f_{Y|X=x}(y), & c.d. \\ \int_{y \in B} f_{Y|X=x}(y)dy, & c.c. \end{cases}$$

para todo $B \subseteq \mathbb{R}$. En particula, la fda condicional de Y dado $X = x$, está dada por:

$$P_{Y|X=x}(y) = P(Y \leq y|X = x) = \begin{cases} \sum_{z \leq y} f_{Y|X=x}(z), & c.d. \\ \int_{-\infty}^{y} f_{Y|X=x}(z)dz, & c.c. \end{cases}$$

para todo y.

Nota 6.. La definición y resultados anteriores son análogos si $(\mathbf{X}, \mathbf{Y})$ es un vector aleatorio discreto o continuo, con $\mathbf{X} : \Omega \to \mathcal{X} \subseteq \mathbb{R}^n$ e $\mathbf{Y} : \Omega \to \mathcal{Y} \subseteq \mathbb{R}^m$.

Ejemplo 6.15. Se extrae al azar un bolita de una urna con N bolitas numeradas del 1 al N. Luego se lanza una moneda tantas veces como lo indica el número de la bolita seleccionada. Sea X el número de la bolita extraída. Si $X = x$, entonces se lanza la moneda x veces. Si Y es el número caras obtenidas en los x lanzamientos de la moneda, entonces, $Y|X = x \sim Bin(x, p)$, donde p es la probabilidad de obtener cara al lanzar la moneda, es decir

$$\begin{aligned} f_{Y|X=x}(y) &= P(Y = y|X = x) \\ &= \begin{cases} \binom{x}{y} p^y (1-p)^{x-y}, & y = 0, 1, \ldots, x, \text{ para } x = 1, 2, \ldots, N. \\ 0, & eoc. \end{cases} \end{aligned}$$

Ejemplo 6.16. Sean X e Y variables aleatorias continuas con fdp conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} 2, & x > 0, y > 0, x + y < 1 \\ 0, & eoc. \end{cases}$$

Entonces

$$f_X(x) = \begin{cases} \int_0^{1-x} 2dy = 2(1 - x), & 0 < x < 1 \\ 0, & eoc. \end{cases}$$

Luego

$$f_{Y|X=x}(y) = \begin{cases} \frac{1}{1-x}, & 0 < y < 1-x, \text{ 0¡x¡1} \\ 0, & eoc. \end{cases}$$

Notemos que $Y|X = x \sim U(0, 1-x)$ para cada $x \in (0,1)$. Análogamente, se tiene que $X|Y = y \sim U(0, 1-y)$ para cada $y \in (0,1)$.

(a) $f_{X,Y}(x,y) = f_{Y|X=x}(y)f_X(x) = f_{X|Y=y}(x)f_Y(y)$ para todo (x,y).

(b) Si X e Y son variables aleatorias independientes, entonces

$$f_{Y|X=x}(y) = f_Y(y) \text{ y } f_{X|Y=y} = f_X(x)$$

para todo (x,y).

(c) Para cada y fijo, se tiene que $f_{Y|X=x}(y) = g(x)$ es una función (no aleatoria) de x definida sobre el recorrido de x. Por ejemplo, si $Y|X = x \sim Bin(x,p)$, entonces

$$f_{Y|X=x} = \binom{x}{y} p^y (1-p)^{x-y} = g(x)$$

para cada $y = 0, 1, 2, \ldots, x$. Similarmente, si $Y|X = x \sim U(0, 1-x)$, entonces

$$f_{Y|X=x}(y) = \frac{1}{(1-x)} = g(x)$$

para cada $y \in (0, 1-x)$.

6.5.2. Esperanza Condicional

Consideremos la función aleatoria $g(X) = f_{Y|X}(y)$, entonces

$$E(g(X)) = E(f_{Y|X}(y)) = f_Y(y)$$

En efecto, consideremos el caso continuo, entonces

$$\begin{aligned} E(g(X)) &= \int_{-\infty}^{\infty} g(x)f_X(x)dx \\ &= \int_{-\infty}^{\infty} f_{Y|X=x}(y)f_X(x)dx \\ &= \int_{-\infty}^{\infty} f_{Y,X}(x,y)dx \\ &= f_Y(y) \end{aligned}$$

De aquí, también es directo que

$$E(P(Y \in B|X)) = P(Y \in B)$$

para todo $B \subseteq \mathbb{R}$.

Definición 6.5. *La esperanza condicional de Y dado $X = x$, provisto que exista, se define como:*

$$E(Y|X = x) = \begin{cases} \sum_{y\in\mathbb{R}} y f_{Y|X=x}(y), & c.d. \\ \int_{-\infty}^{\infty} y f_{Y|X=x}(y)dy, & c.c. \end{cases}$$

La esperanza condicional de X dado $Y = y$, se define la forma análoga.

Nota 6.. Si Y tiene esperanza finita, entonces la esperanza condicional de Y dado $X = x$, también es finita.

Más generalmente, si $g(Y)$ tiene esperanza finita, entonces la esperanza condicional de $g(Y)$ dado $X = x$, se define como:

$$E(g(Y)|X = x) = \begin{cases} \sum_{y\in\mathbb{R}} g(y) f_{Y|X=x}(y), & \text{c.d.} \\ \int_{-\infty}^{\infty} g(y) f_{Y|X=x}(y)dy, & \text{c.d.} \end{cases}$$

Ejemplo 6.17.

(a) Si $Y|X = x \sim Bin(x, p)$, entonces

$$E(Y|X = x) = \sum_{y=0}^{x} y \binom{x}{y} p^y (1-p)^{x-y} = xp$$

(b) Si $Y|X = x \sim U(0, 1-x)$, entonces

$$E(Y|X = x) = \int_{y=0}^{1-x} y \frac{1}{1-x} dy = \frac{1-x}{2}$$

A partir de los ejemplos anteriores que

$$E(Y|X = x) = h(x), \text{ (función no aleatoria de } x)$$

Sea

$$h(X) = E(Y|X), \text{ (función aleaotria de } X)$$

Teorema 6.7. (Ley de Probabilidad Total para Esperanzas) *Sean X e Y variables aleatorias definidas en el mismo espacio de probabilidad. Si Y tiene esperanza finita, entonces*

$$E(Y) = E(E(Y|X))$$

Dem....

Ejemplo 6.18.

(a) Si $Y|X = x \sim Bin(n,p)$, entonces $E(Y|X = x) = xp$, de modo que $E(Y|X) = Xp$, luego

$$E(Y) = E(E(Y|X)) = E(Xp) = E(X)p$$

Así, si $X \sim P(\lambda)$, entonces $E(X) = \lambda$ y por tanto $E(Y) = \lambda p$.

(b) Si $Y|X = x \sim U(0, 1-x)$, entonces $E(Y|X = x) = (1-x)/2$, de modo que $E(Y|X) = (1-X)/2$, luego

$$E(Y) = E(E(Y|X)) = E((1-X)/2) = (1-E(X))/2$$

Así, si $f_X(x) = 2(1-x)I_{(0,1)}(x)$, entonces

$$E(X) = \int_0^1 2x(1-x)dx = \frac{1}{3}$$

y por tanto $E(Y) = \frac{1}{3}$.

Ejemplo 6.19. (Encuesta de Hogares) Sean X el número de miembros en un hogar seleccionado aleaotriamente en la encuesta, e Y el número de automóviles de propiedad de dicho hogar. Los 250 hogares encuestados tienen la misma probabilidad de ser seleccionados, por lo que $P(X = x, Y = y)$ es igual al número de hogares con x miembros e y autos, dividido por 250, estas probabilidades se presentan en la tabla 1 dada a continuación. Suponga que el hogar seleccionado tiene $X = 4$ miembros. La fmp condicional de Y dado $X = 4$ es $f_{Y|X=4}(y) = f_{X,Y}(4,y)/f_X(4)$, y corresponde a los valores de la columna $x = 4$ de la tabla 1 dividido por $f_X(4) = 0{,}208$, es decir

$$f_{Y|X=4}(0) = 0{,}0385, \quad f_{Y|X=4}(1) = 0{,}5769$$
$$f_{Y|X=4}(2) = 0{,}2885, \quad f_{Y|X=4}(3) = 0{,}0962$$

talble

Cuadro 5: Caption

La media condicional de Y dado $X = 4$ es

$$E(Y|X=4) = 0 \times 0{,}0385 + 1 \times 0{,}5769 + 2 \times 0{,}2885 + 3 \times 0{,}0962 = 1{,}442$$

Similarmente, podemos calcular $E(Y|X = x)$ para los ocho valores de x, estos son

x	1	2	3	4	5	6	7	8
$E(Y\|X=x)$	0,609	1,057	1,317	1,442	1,538	1,533	1,75	2

La variable aleatoria $h(X)$ que toma el valor 0,609 cuando el hogar muestreado tiene un miembro, toma el valor 1,057 cuando el hogar muestreado tiene dos miembros, y así sucesívamente, es $h(X) = E(Y|X)$, es decir, la esperanza condicional de Y dado la variable aleatoria X. Además de la propiedad importante de que $E(E(Y|X)) = E(Y)$, la esperanza condicional posee (condicionalmente) todas las propiedades de la esperanza ordinaria, ya que es la media de la distribución condicional. A continuación se enuncian sólo algunas de estas propiedades

(a) $E(aY + b|X = x) = aE(Y|X = x) + b$.

(b) $E(g(X,Y)|X = x) = E(g(x,Y)|X = x)$, en parituclar $E(XY|X = x) = xE(Y|X)$. Además, $E(g(X)h(Y)) = E(g(X)E(h(Y|X)))$, por ejemplo $E(XY) = E(XE(Y|X))$.

(c) Si X e Y son independientes, entonces la distribución condicional de Y dado $X = x$ coincide con la distribución marginal de Y para todo x, es decir, $P(Y \in B|X = x) = P(Y \in B)$ para todo x y todo B, luego $E(Y|X = x) = E(Y)$, del mismo modo se tiene que $E(X|Y = y) = E(X)$.

6.5.3. Varianza Condicional

Tal como la media condicional, la varianza condiciona es simplemente la varianza de la distribución condicional como se define a continuación, por ende también satisface todas las propiedades de la varianza ordinaria.

Definición 6.6. *La varianza condicional de Y dado $X = x$, se define como*

$$\begin{aligned} Var(Y|X = x) &= E((Y - E(Y|X = x))^2|x) \\ &= \begin{cases} \sum_{y\in\mathbb{R}}(y - E(Y|X = x))^2 f_{Y|X=x}(y), & c.d. \\ \int_{-\infty}^{\infty}(y - E(Y|X = x))^2 f_{Y|X=x}(y)dy, & c.c. \end{cases} \end{aligned}$$

probisto que la esperanza exista.

Tarea. Pruebe que la varianza condicional de Y dado $X = x$, también puede calcularse como $\mathrm{Var}(Y|X = x) = E(Y^2|X = x) - (E(Y|X = x))^2$.

Ejemplo 6.20.

(a) Si $Y|X = x \sim Bin(n,p)$, entonces, $E(Y|X = x) = xp$, de modo que $E(Y|X) = Xp$, luego

$$\mathrm{Var}(Y|X = x) = \sum_{y=0}^{x}(y - xp)2\binom{x}{y}p^y(1-p)^{x-y} = xp(1-p)$$

(b) Si $Y|X = x \sim U(0, 1-x)$, entonces, $E(Y|X = x) = (1-x)/2$, luego

$$\mathrm{Var}(Y|X = x) = \int_{y=0}^{1-x}\left(y - \frac{1-x}{2}\right)^2 \frac{1}{1-x}dy = \frac{(1-x)^2}{12}$$

De los ejemplos anteriores se desprende que $\mathrm{Var}(Y|X = x) = v(x)$ (función no aleatoria), mientas que $\mathrm{Var}(Y|X) = v(x)$ (función aleatoria)

Teorema 6.8. (Ley de Probabilidad Total para Varianzas) *Sean X e Y variables aleatorias definidas en un mismo espacio de probabilidad. Si $E(Y^2)$ es finita, entonces*

$$Var(Y) = Var(E(Y|X)) + E(Var(Y|X))$$

Dem.

Ejemplo 6.21. Si $Y|X = x \sim Bin(n,p)$, entonces ,$E(Y|X = x) = xp$ y $\mathrm{Var}(Y|X = x) = xp(1-p)$, luego

$$\begin{aligned}\mathrm{Var}(Y) &= \mathrm{Var}(Xp) + E(Xp(1-p)) \\ &= p^2\mathrm{Var}(Xp) + p(1-p)E(X)\end{aligned}$$

Si $X \sim P(\lambda)$, entonces $\mathrm{Var}(X) = E(X) = \lambda$, de modo que $\mathrm{Var}(Y) = E(Y) = \lambda p$.

Tarea. Pruebe que si $Y|X = x \sim Bin(n,p)$ y $X \sim P(\lambda)$, entonces $Y \sim P(\lambda p)$.

Ejemplo 6.22. Si $Y|X = x \sim U(0, 1-x)$, entonces, $E(Y|X = x) = (1-x)/2$ y $\mathrm{Var}(Y|X = x) = (1-x)^2/12$, luego

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\frac{1-X}{2}\right) + E\left(\frac{(1-X)^2}{12}\right)$$
$$= \frac{1}{4}\mathrm{Var}(X) + \frac{1}{2}E((1-X)^2)$$

Tarea Terkine el ejmplo para $X \sim f_X(x) = 2(1-x)I_{(0,1)}(x)$.

Ejemplo 6.21. Sea N el número de personas por día que entra a un supermercado. Sean $X_1, \ldots, X_N$ las cantidades gastadas por cada una de las N personas que ingreso al supermercado durante un determinado día. Suponga que N y $X_1, \ldots, X_N$ son variables aleatorias independientes. Encuentre la media y la varianza del ingreso total del supermercado durante un día. Sea $Y = \sum_{i=1}^{N} X_i$ el ingreso diario total del supermercado, donde es N una variable aleatoria con valores en los enteros positivos, mientras que los X_i son variables aleatorias continuas con valroes positivos. Asuma también que el gasto de cada persona que entra al supermercado durante un día tiene la misma distribución.

i) **Esperanza de Y.** Se tiene que $E(Y) = E(E(Y|N))$, donde $E(Y|N = n) = E(\sum_{i=1}^{M} X_i|N = n) = E(\sum_{i=1}^{n} X_i) = nE(X_1)$, luego

$$\begin{aligned}E(Y) &= E(E(Y|N)) \\ &= E(NE(X_1)) = E(N)E(X_1)\end{aligned}$$

ii) **Varianza de Y.** Sabemos que

$$\mathrm{Var}(Y) = E(\mathrm{Var}(Y|N)) + \mathrm{Var}(E(Y|N))$$

Ahora

$$\text{Var}(Y|N = n) = \sum_{i=1}^{n} \text{Var}(X_i) = n\text{Var}(X_1)$$

Luego

$$\begin{aligned}\text{Var}(Y) &= E(\text{Var}(Y|N)) + \text{Var}(E(Y|N)) \\ &= E(N)\text{Var}(X_1) + \text{Var}(NE(X_1)) \\ &= E(N)\text{Var}(X_1) + \text{Var}(N)(E(X_1))^2\end{aligned}$$

Tarea. Concluya el ejmplo asumiendo que $N \sim P(\lambda)$ y $X_1 \sim U(0, \theta)$

6.5.4. Predicción

Considere dos variables aleatorias X e Y con fdp (o fmp) conjunta $f_{X,Y}(x, y)$. Supongamos que después de que se haya observado el valor de X, se debe predecir el valor de Y. En otras palabras, el valor predicho de Y puede depender del valor de X. Sunponga que este valor predicho $h(X)$ debe elegirse de modo que minimice el error cuadrático medio $E((Y - h(X))^2)$

Teormea 6.9. *El predictor $h(X)$ que minimiza $E((Y - h(x))^2)$* es $h(X) = E(Y|X)$.

Tarea. Obtenga $h(x) = E(Y|X = x)$ cuando (X, Y) tiene distribución normal bivariada $NB(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$ con $|\rho_{XY}| < 1$.

7. Muestras Aleatorias y Distribuciones

7.1. Definiciones

Una distribución de probabilidad, también llamada distribución poblacional (o simplemente población), describe el compartamiento probabilístico de una determinada variable aleatoria.

También sabemos que toda distribución de probabilidad o población puede ser descrita o representada por su fda (F) o bien por su fmp (f), en el caso discreto o fdp (f) en el caso continuo.

Normalmente no es posible conocer toda la población, por lo sólo disponemos de una muestra representativa de dicha población.

A partir de la muestra, podemos inferir sobre características poblaciones, llamadas parámetros, que nos interesa conocer.

Para ello, se requiere saber primero la distribución de probabilidad (exacta o asintótica) de las versiones muestrales de tales características, las cuales son denominadas estadísticos.

Definición 7.1. *Se dice que $X_1, \ldots, X_n$ es una muestra aleatoria de tamaño n (ma(n)) de una población con fmp o fdp f, si $X_1, \ldots, X_n$ son variables aleatorias mutuamente independientes e identicamente distribuidas (iid) f.*

Como es obvio, si $X_1, \ldots, X_n$ es una ma(n) es

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$$

Cuando la ma(n) proviene de una familia paramétrica $f(x; \theta)$, entonces la fmp o fdp conjunta de $X_1, \ldots, X_n$ es

$$f(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

donde el parámetro θ (posiblemente vectorial) etiqueta la fmp o fdp.

Ejemplo 7.1. Sea $X_1, \ldots, X_n$ una ma(n) de una población exp(λ), donde $\lambda > 0$ es el parámetro asociado. Por ejemplo, $X_1, \ldots, X_n$ podrián corresponder a los tiempos de vida (medidos en años y fracciones) de n artefactos electrónicos idénticos, los que se hacen funcionar hasta que fallen. La fdp conjunta de la ma(n) es

$$\begin{aligned} f(x_1, \ldots, x_n; \lambda) &= \prod_{i=1}^{n} f(x_i; \lambda) \\ &= \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i},\ x_i > 0,\ i = 1, 2, \ldots, n \end{aligned}$$

Con esta fdp se pueden estudiar diferentes propiedades de la muestra, por ejemplo, para calcular la probabilidad de que todos los artefactos duren más de 2 años

$$P(X_1 > 2, \ldots, X_n > 2) = \prod_{i=1}^{n} P(X_i > 2) = \exp(-2n\lambda)$$

A partir de la fda conjunta de la ma(n) también se puede demostrar que la variable aleatoria $\sum_{i=1}^{n} X_i$ tiene distribución $Gamma(n, \lambda)$, y por tanto que

$$2\lambda \sum_{i=1}^{n} X_i \sim \chi^2_{2n}$$

Este resultado juega un rol importante cuando se desea inferir sobre el verdadero valor del parámetro λ, es decir, aquel valor efectivamente generó la ma(n) observada.

Definición 7.2. *Sea $T(x_1, \ldots, x_n)$ una función de valor real o bien vectorial, y sea $X_1, \ldots, X_n$ una ma(n) de una población. Entonces la variable o vector aleatoria(o) $T = T(X_1, \ldots, X_n)$ se denomina estadístico. La distribución de probabilidad de un estadístico T se llama distribución muestral de T.*

La definición de un estadístico es muy amplia, cualquier función de la ma(n) que no dependa de ningún parámetro desconocido es un estadístico. Por ejmplo, un estadístico puede ser la propia ma(n), el mínimo o el máximo de la muestra, el promedio de la muestra o alguna medida de la variabilidad de las observaciones de la muestra.

7.2. La media muestral y La varianza muestral

Habitualmente, los parámetros poblaciones de mayor interés son la media μ y la varianza σ^2 de la población. Por ende, al disponer de una ma(n) de dicha población, es natural preocuparse de la media $\overline{X}$ y la varianza S^2 de la muestra.

Definición 7.3. *Dada una muestra aleatoria $X_1, \ldots, X_n$ de una población f, dos estadísticos de interés inmediato son:*

(a) La media muestral, definida como:

$$\overline{X} := \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1} nX_i$$

(b) La varianza muestral, definida como:

$$S^2 := \frac{(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

También se define la desviación estándar muestral como, $S = \sqrt{S^2}$.

Teorema 7.1. *Para cualquier secuencia de números reales $x_1, \ldots, x_n$ sea $\overline{x} = \sum_{i=1}^{n} x_i/n$. Entonces* *i)* $\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}(x_i-\overline{x})^2 + n(\overline{x}-\mu)^2$ *para todo* $\mu \in \mathbb{R}$.

ii) $\min_{\mu\in\mathbb{R}} \sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}(x_i-\overline{x})^2$, *es decir, para* $\hat{\mu}=\overline{x}$ *la distancia* $\sum_{i=1}^{n}(x_i-\mu)^2$ *es mínima.*

iii) $\sum_{i=1}^{n}(x_i-\overline{x})^2 = \sum_{i=1}^{n}(x_i-\overline{x})x_i = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$

Tarea. Pruebe que $\sum_{i=1}^{n}(x_i-\overline{x}) = 0$.

Dem.

Lema 7.1. *Sea* $X_1,\ldots,X_n$ *una ma(n) de* f*, y sea* g *una función real valorada. Entonces*

i) $g(X_1),\ldots,g(X_n)$ *son variables aleatorias iid.*

ii) $E(\sum_{i=1}^{n} g(X_i)) = nE(g(x_1))$ *provisto que la esperanza exista.*

iii) $Var(\sum_{i=1}^{n} g(X_i)) = n\,Var(g(X_1))$ *provisto que la varianza exista.*

Dem...

Teorema 7.2. *Sea* $X_1,\ldots,X_n$ *una ma(n) desde una población con media* μ *y varianza* σ^2 *finitas. Entonces:*

i) $E(\overline{X}) = \mu$.

ii) $Var(\overline{X}) = \sigma^2/n$.

iii) $E(S^2) = \sigma^2$.

Conclusión. La media muestral $\overline{X}$ y la varianza muestral S^2 de una ma(n) son estimadores insesgados, respectívamente, de la media población μ la varianza población σ^2.

Dem...

Teorema 7.3. *Sea* $X_1,\ldots,X_n$ *una ma(n) desde una población con fgm* $M(t)$*. Entonces la fgm de la media muestral es:*

$$M_X(t) = (M(t/n))^n$$

Dem...

El resultado anterior propociona una forma muy simple para obtener la distribución muestral de $\overline{X}$ cuando $M_X(t)$ es una fgm conocida, como se ilustra en el siguiente ejemplo.

Ejemplo 7.2. (Distribución de la Media) Sea $X_1,\ldots,X_n$ una ma(n) desde una población $N(\mu,\sigma^2)$. Entonces la fgm de la media muestral $\overline{X}$ es

$$\begin{aligned} M_X(t) &= \exp\left(\mu\frac{1}{n} + \frac{\sigma^2(t/n)^2}{2}\right)^n \\ &= \exp\left(n\left(\mu\frac{1}{n} + \frac{\sigma^2(t/n)^2}{2}\right)\right) \\ &= \exp\left(\mu t + \frac{(\sigma^2/n)t^2}{2}\right) \end{aligned}$$

Así, si la ma(n) proviene de una población $N(\mu, \sigma^2)$, entonces la distribución muestral de $\overline{X}$ es $N(\mu, \sigma^2/n)$.

Ejemplo 7.3. Se tiene una máquina de llenado para vaciar 500gr de cereal en una caja de cartón. Supongamos que la cantidad de cereal que se coloca en cada caja es una variable aleatoria normalmente disribuidad con media 500gr y desviacion estándar igual a 20gr.

Para verificar que el peso promedio de cada caja se mantiene en 500gr se toma una muestra aleatoria de 25 de éstas cajas en forma periódica y se pesa el contenido de cada una de ellas.

El gerente de la planta ha decidido detener el proceso y encontrar la falla cada vez que el valor promedio de la muestra sea mayor de 510gr o menos de 490gr. Obtenga la probabilidad de detener el proceso.

Sea $X_1, \ldots, X_{25}$ una ma(25), las cuales representan la cantidad de cereal contenido en las cajas de una muestra aleatoria dada.

Por hipótesis $X_i \sim N(500, 20^2)$ $i = 1, 2, \ldots, 25$. $\overline{X} \sim N(500, 20^2/25)$. La probabilidad deseada es igual a uno menos la probabilidad de que $\overline{X}$ se encuentre 490 y 510gr.

$$\begin{aligned}
P(\text{Detención del proceso}) &= 1 - P(490 < \overline{X} < 510) \\
&= 1 - P\left(\frac{490 - 500}{4} < Z < \frac{510 - 500}{4}\right) \\
&= 1 - P(-2{,}5 < Z < 2{,}5) \\
&= 0{,}0124
\end{aligned}$$

Teorema 7.4. *Sea $X_1, \ldots, X_n$ una ma(n) desde una distribución $N(\mu, \sigma^2)$, y sea $\overline{X} = \sum_{i=1}^n X_i/n$ y $S^2 = \sum_{i=1}^n (X_i - \overline{X})^2/(n-1)$. Entonces*

i) $\overline{X}$ *y* S^2 *son variables aleatorias independientes.*

ii) $\overline{X}$ *tiene una distribución* $N(\mu, \sigma^2/n)$.

iii) $(n-1)S^2/\sigma^2$ *distribuye chicuadrado con* $n-1$ *grados de libertad.*

Dem...

Es decir, S^2 es una función de $(X_2 - \overline{X}, \ldots, X_n - \overline{X})$ solamente. Entonce, basta mostrar que estas variables son independientes de $\overline{X}$.

Supongamos sin pérdida de generalidad que $\mu = 0$ y $\sigma = 1$, entonces la fdp conjunta de la muestra $X_1, \ldots, X_n$ está dada por:

$$f(x_1, \ldots, x_n) = (2\pi)^{-n/2} e^{-\sum_{i=1}^n x_i^2/2}$$

con $-\infty < x_i < \infty$. Consideremos la transformación inyectiva dada por

$$Y_1 = \overline{X}, Y_2 = X_2 - \overline{X}, \ldots, Y_n = X_n - \overline{X}$$

Esta transformación es lineal en $(X_1, \ldots, X_n)$ con un Jacobiano $1/n$. Luego

$$\begin{aligned} f(y_1, \ldots, y_n) &= n(2\pi)^{-n/2} e^{-(y_1 - \sum_{n=2}^{n} y_i)^2/2} \cdot e^{-\sum_{i=2}^{n}(y_i + y_1)^2/2},\ y_i \in \mathbb{R}, i = 1, 2, \ldots, n \\ &= (n/(2\pi))^{1/2} e^{(-ny_1^2)/2} \cdot n^{1/2}(2\pi)^{-(n-1)/2} e^{-(\sum_{i=2}^{n} y_i^2 + (\sum_{i=2}^{n} y_i)^2)/2}, y_i \in \mathbb{R} \\ i &= 1, 2, \ldots, n \end{aligned}$$

Dado que la fdp conjunta de $Y_1, \ldots, Y_n$ se puede escribir como el producto de las marginales, se deduce que $Y_1 = \overline{X}$ es independientes de $Y_2 = X_2 - \overline{X}, \ldots, Y_n = X_n - \overline{X}$ y por lo tanto, que $\overline{X}$ es independiente de S^2.

7.3. Distribuciones

7.3.1. Distribución t-Student

Sean X e Y variables aleatorias independientes tales que $X \sim N(0, 1)$ e $Y \sim \chi^2_{(k)}$. Considere la transformación

$$g(x, y) = (x/\sqrt{y/k}, y)$$

La transformación inversa está dada por:

$$h(x, y) = (x\sqrt{y/k}, y) \tag{2}$$

Mientras que el Jacobiano es

$$J(x, y) = \sqrt{y/k}$$

Por lo tanto, la fdp conjunta de $Z = X/\sqrt{Y/k}$ y $W = Y$ está dada por:

$$f_{Z,W}(z, w) = \sqrt{w/k} f_{X,Y}(z\sqrt{w/k}, w),\ x \in \mathbb{R},\ \ w > 0$$

donde

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{(1/2)^{k/2}}{\Gamma(k/2)} y^{k/2-1} e^{-y/2},\ x \in \mathbb{R},\ y > 0$$

Integrando $f_{Z,W}(z, w)$ con respecto a w, se obtiene la fdp marginal de Z, es decir

$$f_Z(z) = \frac{\Gamma((k+1)/2)}{\sqrt{k\pi}\Gamma(k/2)} \left(1 + \frac{z^2}{k}\right)^{-(k+1)/2},\ \ z \in \mathbb{R}$$

La cual corresponde a la fdp de una distribución t-Student con k grados de libertad.

Notación. $Z \sim t_k$

7.3.2. Distribución F

Sean X e Y variables aleatorias independientes tales que $X \sim \chi^2_m$ e χ^2_n. Considere la transformación

$$g(x,y) = \left(\frac{x/m}{y/n}, y\right)$$

La función inversa de g está dada por:

$$h(x,y) = \left(\frac{m}{n}xy, y\right)$$

y tiene un Jacobiano igual a $J(x,y) = \frac{m}{n}y$. Luego, la fdp conjunta de $Z = \frac{nX}{mY}$ y $W = Y$ está dada por:

$$f_{Z,W}(z,w) = \frac{(1/2)^{m/2}(1/2)^{n/2}}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}zw\right)^{\frac{m}{2}-1} w^{\frac{n}{2}-1}e^{\left(-\frac{1}{2}w\left(\frac{m}{n}z+1\right)\right)}$$

Integrado con respecto a w, encontramos la fdp de Z, dada por:

$$f_Z(z) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{\frac{m}{2}} z^{\frac{m}{2}-1}\left(1+\frac{m}{n}z\right)^{-(m+n)/2}, \; z > 0$$

La cual corresponde a la fdp de una distrubución F con m gl en el numerador y n gl en el denominador.

Notación. $Z \sim F_{m,n}$

Teorema 7.5. *Sea $X_1, \ldots, X_n$ una ma(n) desde una distribución $N(\mu, \sigma^2)$. Entonces la variable aleatoria $(\overline{X} - \mu)/(S/\sqrt{n})$ tiene una distribución t-Studen con $n-1$ grados de libertad.*

Dem...

Teorema 7.6. *Sea $X_1, \ldots, X_n$ una ma(n) de una población $N(\mu_X, \sigma^2_X)$, sea $Y_1, \ldots, Y_m$ una ma(m) una población independiente $N(\mu_Y, \sigma^2_Y)$. La variable aleaotria $F = (S^2_X/\sigma^2_X)/(S^2_Y/\sigma^2_Y)$ tiene distribución F con $n-1$ y $m-1$ grados de libertad.*

Dem...

Teorema 7.7.

i) Si $X \sim F_{p,q}$ entonces $1/X \sim F_{q,p}$, esto es, el recíproco de una variable aleatoria F también tiene una distribución F.

1. Si $X \sim t_q$, entonces $X^2 \sim F_{1,q}$.

2. Si $X \sim F_{p,q}$, entonces $(p/q)X(1 + (p/q)X) \sim Beta(p/2, q/2)$.

Problam. ¿Qué hacemos si no tenemos normalidad?

8. Convergencia Estocástica y Teoremas Límites

8.1. Definición y Ejemplos

El objetivo de este tópico, es estudiar el comportamiento de un secuencia de variables aleatorias $\{Z_n : n \geq 1\}$ cuando $n \to \infty$. La secuencia Z_n puede estar asociada a un estadístico muestral o alguna función de dicho estadístico que tenga algún interés inferencial, en cuyo caso el subíndice n representa el tamaño muestral. A pesar que la noción de un tamaño de muestra infinito es un concepto teórico, el sentido práctico es obtener aproximaciones útiles en el contexto de muestras finitas, ya que generalmente las expresiones se simplifican en el límite.

Ejemplo 8.1. Sea $Z_n = \sum_{i=1}^{n} X_i$, donde para cada $n \geq 1$, $X_1, \ldots, X_n$ son variables aleatorias iid $Ber(p)$. Entonces, $\{Z_n : n \geq 1\}$ es una secuencia de variables aleatorias tales que, para cada $n \geq 1$, $Z_n \sim Bin(n, p)$. En este caso, se puede demostrar, por ejemplo, que:

(a) Para todo $\varepsilon > 0$

$$P\left(\left|\frac{Z_n}{n} - p\right| \geq \varepsilon\right) \overset{n\to\infty}{\longrightarrow} 0$$

(b) Para todo $z \in \mathbb{R}$

$$P\left(\frac{Z_n - np}{\sqrt{np(1-p)}} \leq x\right) \overset{n\to\infty}{\longrightarrow} \Phi(z)$$

La propiedad (a) implica que para n suficientemente grande, la probabilidad de que $\frac{Z_n}{n} = \overline{X}_n$ (media muestral) difiera de $p = E(X_1)$ (media poblacional) en una cantidad (positiva) arbitrariamente pequeña es despreciable. Miestra que la propiedad (b) implica que para n suficientemente grande, se tiene que, $P\left(\frac{Z_n - np}{\sqrt{np(1-p)}} \leq z\right) \simeq \Phi(z)$, es decir, que la distribución de la variable aleatoria $\frac{Z_n - np}{\sqrt{np(1-p)}} = \frac{\sqrt{n}(\overline{X}_n - p)}{\sqrt{p(1-p)}}$ puede aproximarse por la distribución de una variable aleatoria $Z \sim N(0, 1)$.

Lema 8.1. (Desigualdad de Chebyshev) *Si X es una variable aleatoria con segundo momento finito y c es una constante real arbitraria, entonces para todo $\varepsilon > 0$*

$$P(|X - c| \geq \varepsilon) \leq \frac{E((X-c)^2)}{\varepsilon^2}$$

En particual, si $c = \mu := E(X)$, entonces $E((X-c)^2) = Var(X) := \sigma^2$, luego para todo $\varepsilon > 0$

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Asi, si $Z_n \sim Bin(n, p)$, entonces $E\left(\frac{Z_n}{n}\right) = p$ y $\mathrm{Var}\left(\frac{Z_n}{n}\right) = \frac{p(1-p)}{n}$, para cada $n \geq 1$. Luego, para cada $n \geq 1$ y cada $\varepsilon > 0$, se tiene que

$$0 \leq P\left(\left|\frac{Z_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{p(1-p)}{n\varepsilon^2}$$

Tomando $n \to \infty$, llegamos a que

$$\lim_{n\to\infty} P\left(\left|\frac{Z_n}{n} - p\right| \geq \varepsilon\right) = 0$$

La propiedad (b) es una consecuencia del Teorema del Límite central, el cual será enunciado más tarde.

Sea $\{Z_n : n \geq 1\}$ una secuencia de variables aleatorias definidas en un espacio de probabililidad $(\Omega, \mathcal{A}, P)$, y sea θ una constante en $\mathbb{R}$.

Definición 8.1.*Se dice que Z_n converge en probabilidad para θ si, para todo $\varepsilon > 0$*

$$P(|Z_n - \theta| \geq \varepsilon) \overset{n\to\infty}{\longrightarrow} 0$$

o de forma equivalente, para todo $\varepsilon > 0$

$$P(|Z_n - \theta| < \varepsilon) \overset{n\to\infty}{\longrightarrow} 1$$

Notación. $Z_n \overset{P}{\to} \theta$ o $plimZ_n = \theta$.

Nota 8.1. Si $Z_n = T(X_1, \ldots, X_n)$ es un estadístico muestral talque $Z_n \overset{P}{\to} \theta$, se dice que Z_n es un estimador consistente del parámetro poblacional θ.

Nota 8.2.

(a) De la definición anterior, es claro que $Z_n \overset{P}{\to} \theta$ si y sólo si $Z_n - \theta \overset{P}{\to} 0$.

(b) Si $\{Z_n := \theta_n : n \geq 1\}$, es una secuencia no-aleatoria, y $\theta_n \to \theta$ cuando $n \to \infty$, entonces $\theta_n \overset{P}{\to} \theta$.

(c) La convergencia en probabilidad de Z_n también puede ser para una variable aleaotria Z, definida en el mismo espacio de probabilidad que la secuencia $\{Z_n : n \geq 1\}$, es decir

Definición 8.2. *Se dice que Z_n converge en probabilidad para una variable aleatoria Z, y se denota por $Z_n \overset{P}{\to} Z$ si y sólo si $Z_n - Z \overset{P}{\to} 0$.*

(d) Si $\mathbf{Z}, \mathbf{Z}_1, \mathbf{Z}_2, \ldots$ son vectores aleaotrios p-dimensionales definidos en el mismo espacio de probabilidad, entonces, $\mathbf{Z}_n \overset{P}{\to} \mathbf{Z}$ si y sólo si $\|\mathbf{Z}_n - \mathbf{Z}\| \overset{P}{\to} 0$, donde $\|z\| = \sqrt{z^T z}$ es la norma usual en $\mathbb{R}^p$.

Ejemplo 8.2. Sea $Z_n \sim Bin(n, p)$, $n \geq 1$. Vimos que para todo $\varepsilon > 0$

$$P\left(\left|\frac{Z_n}{n} - p\right| \geq \varepsilon\right) \overset{n\to\infty}{\longrightarrow} 0$$

Por lo tanto, $\frac{Z_n}{n} \overset{P}{\to} p$. Ya que $\frac{Z_n}{n}$ corresponde al promedio muestral $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, donde $X_1, \ldots, X_n$ es una ma(n) de la distribución $Ber(p)$, entonces se concluye que $\overline{X}_n \overset{P}{\to} p$, es decir, $\overline{X}_n$ es un estimador consistente de p.

Ejemplo 8.3. Supongamos que $P(Z_n = \theta) = 1 - \frac{1}{n}$ y $P(Z_n = \theta + n^2) = \frac{1}{n}$ para todo $n \geq 1$. Entonces, $Z_n \overset{P}{\to} \theta$. En efecto, del enunciado

$$P(Z_n = z) = \begin{cases} 1 - \frac{1}{n}, & z = \theta \\ \frac{1}{n}, & z = \theta + n^2 \\ 0, & eoc. \end{cases}$$

Luego para todo $\varepsilon > 0$

$$\begin{aligned} P(|Z_n - \theta| < \varepsilon) &= P(\theta - \varepsilon < Z_n < \theta + \varepsilon) \\ &= P(Z_n = \theta) = 1 - \frac{1}{n} \overset{n\to\infty}{\longrightarrow} 1 \end{aligned}$$

Ejemplo 8.4. Sea $Z_n = \max\{X_1, \ldots, X_n\}$, donde $X_1, \ldots, X_n \overset{\text{iid}}{\sim} U(0, \theta)$ con $(\theta > 0)$. Entonces $Z_n \overset{P}{\to} \theta$, es decir, $\max\{X_1, \ldots, X_n\}$ es un estiamdor consistente del parámetro θ. En efecto, sabemos que la fda marginal de los $X_i's$ es

$$F(X) = \begin{cases} 0, & x < 0 \\ \frac{x}{\theta}, & 0 \leq x < \theta \\ 1, & x \geq \theta \end{cases}$$

Además, vimos que la fda $Z_n = \max\{X_1, \ldots, X_n\}$ es

$$\begin{aligned} F_{Z_n}(z) &= (F(z))^n \\ &= \begin{cases} 0, & z < 0 \\ \left(\frac{z}{\theta}\right)^n, & 0 \leq z < \theta \\ 1, & z \geq 0 \end{cases} \end{aligned}$$

Ahora, para todo $\varepsilon > 0$

$$\begin{aligned} P(|Z_n - \theta| < \varepsilon) &= P(-\varepsilon < Z_n - \theta < \varepsilon) \\ &= P(\theta - \varepsilon < Z_n < \theta + \varepsilon) \\ &= F_{Z_n}(\theta + \varepsilon) - F_{Z_n}(\theta - \varepsilon) \end{aligned}$$

(Como F_{Z_n} es continua), donde $F_{Z_n}(\theta + \varepsilon) = 1$, ya que $\theta + \varepsilon \geq 0$ para todo $\varepsilon > 0$ y todo $\theta > 0$ y

$$F_{Z_n}(\theta - \varepsilon) = \begin{cases} 0, & \theta - \varepsilon < 0 \Leftrightarrow \varepsilon > \theta \\ \left(1 - \frac{\varepsilon}{\theta}\right)^n, & 0 \leq \theta - \varepsilon < \theta \Leftrightarrow \varepsilon \leq \theta \end{cases}$$

Luego para todo $\varepsilon > 0$

$$\begin{aligned} P(|Z_n - \theta| < \varepsilon) &= \begin{cases} 1, & \varepsilon > \theta \\ 1 - \left(1 - \frac{\varepsilon}{\theta}\right)^n, & \varepsilon \leq \theta \end{cases} \\ &\overset{n\to\infty}{\longrightarrow} 1 \end{aligned}$$

Por lo tanto $Z_n \xrightarrow{P} \theta$.

Definición 8.3. *Una secuencia de variables aleatorias* $\{Z_n : n \geq 1\}$ *converge en media cuadrática (mc) para una constante* θ *si*

$$E((Z_n - \theta)^2) \overset{n\to\infty}{\longrightarrow} 0$$

Notación. $Z_n \overset{\text{mc}}{\to} \theta$ o de forma equivalente, $Z_n - \theta \overset{\text{mc}}{\to} 0$

Nota 8.3. Similarmente, se dice que Z_n converge en media cuadrática para una variable aleatoria Z (definida en el mismo espacio de probabilida de la secuencia $\{Z_n : n \geq\}$), y se denota como $Z_n \overset{\text{mc}}{\to} Z$ si y sólo si $Z_n - Z \overset{\text{iid}}{\to} 0$.

Teorema 8.1. *Sea* $\{Z_n : n \geq 1\}$ *una secuencia de variables aleatorias. Si* $Z_n \overset{mc}{\to} \theta$, *entocnes* $Z_n \xrightarrow{P} \theta$. *Es decir, si*

$$\lim_{n\to\infty} E((Z_n - \theta)^2) = 0$$

Entonces

$$\lim_{n\to\infty} P(|Z_n - \theta| \geq \varepsilon)$$

para todo $\varepsilon > 0$.

Dem...

Corolario 8.1. *Supongamos que se cumple:*

(a) $\lim_{n\to\infty} E(Z_n) = \theta$.

(b) $\lim_{n\to\infty} Var(Z_n) = 0$

Entonces $Z_n \xrightarrow{P} \theta$.

Dem...

Nota 8.4. La esperanza $E((Z_n - \theta)^2)$ se llama error cuadrático medio de Z_n con respecto a θ, y la diferencia $E(Z_n) - \theta$ se llama sesgo de Z_n, con respecto a θ.

Ejemplo 8.5. Sea $Z_n \sim Bin(n, p)$ con $n \geq 1$, entonces

(a) $E\left(\frac{Z_n}{n}\right) = p$ para todo $n \geq 1$, entonces $\lim_{n\to\infty} E\left(\frac{Z_n}{n}\right) = p$.

(b) $E\left(\frac{Z_n}{n}\right) = \frac{p(1-p)}{n}$ para todo $n \geq 1$, entonces $\lim_{n\to\infty} \mathrm{Var}\left(\frac{Z_n}{n}\right) = \lim_{n\to\infty} \frac{p(1-p)}{n} = 0$.

Por lo tanto, $\frac{Z_n}{n} \xrightarrow{P} p$.

Ejemplo 8.6. Sea $Z_n = \max\{X_1, \ldots, X_n\}$ donde $X_1, \ldots, X_n \overset{\text{iid}}{\sim} U(0, \theta)$ $(\theta > 0)$. Entonces

$$f_{Z_n}(z) = \begin{cases} \frac{n}{\theta}\left(\frac{z}{\theta}\right)^{n-1}, & 0 < z < \theta. \\ 0, & eoc. \end{cases}$$

Luego

$$\begin{aligned} E(Z_n) &= \int_0^\theta z\frac{n}{\theta}\left(\frac{z}{\theta}\right)^{n-1}dz = \frac{n}{n+1}\theta \\ E(Z_n^2) &= \int_0^\theta z^2\frac{n}{\theta}\left(\frac{z}{\theta}\right)^{n-1}dz = \frac{n}{n+2}\theta^2 \end{aligned}$$

Así, se tiene que

$$\begin{aligned} \lim_{n\to\infty} E(Z_n) &= \lim_{n\to\infty}\frac{n}{n+1}\theta = \theta \\ \lim_{n\to\infty} \mathrm{Var}(Z_n^2) &= \lim_{n\to\infty}\left(\frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2\right) \\ &= \lim_{n\to\infty}\frac{n}{n+2}\theta^2 = \lim_{n\to\infty}\left(\frac{n}{n+1}\theta\right)^2 \\ &= \theta^2-\theta^2 = 0 \end{aligned}$$

Por lo tanto, $Z_n \overset{P}{\to} \theta$.

Ejemplo 8.7. Caso donde el teorema 8.1 falla. Sea $\{Z_n : n \geq 1\}$ una secuencia de variables tales que $P(Z_n = \theta) = 1 - \frac{1}{n}$ y $P(Z_n = \theta + n^2) = \frac{1}{n}$ para todo $n \geq 1$. Probamos que $Z_n \overset{P}{\to} \theta$. Pero en este caso

$$E((Z_n-\theta)^2) = 0\times\left(1-\frac{1}{n}\right) + n^4\times\frac{1}{n} = n^3 \overset{n\to\infty}{\longrightarrow} \infty$$

Sin embargo, que

$$E(|Z_n-\theta|^{1/4}) = 0\times\left(1-\frac{1}{n}\right) + n^{1/2}\times\frac{1}{n} = \frac{1}{\sqrt{n}} \overset{n\to\infty}{\longrightarrow} 0$$

Extendamos el resultado del teorema 8.1.

Teorema 8.2. *Sea $\{Z_n : n \geq 1\}$ una secuencia de variables aleatorias. Si, para algún $r > 0$, se tiene que*

$$\lim_{n\to\infty} E(|Z_n-\theta|^r) = 0, \quad (\star)$$

entonces $Z_n \overset{P}{\to} \theta$.

Dem...

Nota 8.5. Si Z_n verifica $(\star)$, se dice que Z_n converge en media de orden r para θ y se denota por $Z_n \overset{\mathrm{mr}}{\to} \theta$

El siguiente teorema estable que la media muestral $\overline{X}_n$ de una ma(n) es un estimador consistente de la media poblacional μ cuando esta es finita.

Teorema 8.3. (Ley debil de los grandes números (LDGN)) *Sea $X_1, X_2, \ldots$ una secuencia de variables aleatorias iid con $E(|X_1|) < \infty$. Entonces*

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu = E(X_1)$$

Nota 8.6. Similarmente, si $\mathbf{X}_1, \mathbf{X}_2, \ldots$ es una secuencia de vectores aleatorias p-dimensionales iid, con vector de medias $\boldsymbol{\mu} = E(\mathbf{X}_1)$ finito, entonces $\overline{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \xrightarrow{P} \boldsymbol{\mu}$.

Dem...

Ejemplo 8.8.

(a) Si $X_1, X_2, \ldots$ son variables aleatorias iid $Ber(p)$, entonces $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} p$.

(b) Si $X_1, X_2, \ldots$ son variables aleaotrias iid $U(0, \theta)$, entonces $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \theta/2$.

(c) Si $X_1, X_2, \ldots$ son vatiables aleatorias iid $\exp(\lambda)$, entonces $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} 1/\lambda$.

(d) Si $X_1, X_2, \ldots$ son variables aleatorias iid $N(\mu, \sigma^2)$, entonces

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu = E(X_1) \text{ y } \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} \mu^2 + \sigma^2 = E(X_1^2)$$

Y así con más casos.

Teorema 8.4. *Sea $\{Z_n : n \geq 1\}$ una secuencia de variables aleatorias talque $Z_n \xrightarrow{P} \theta$ y sea $g : \mathbb{R} \to \mathbb{R}$ una función continua. Entonces $g(Z_n) \xrightarrow{P} g(\theta)$.*

Dem...

Ejemplo 8.9.

(a) Si $Z_n \sim Bin(n, p)$ probamos que $\frac{Z_n}{n} = \overline{X}_n \xrightarrow{P} p$. Sea $g(x) = x(1-x)$, la cual es una función continua para todo $0 < x < 1$. Entonces $g(\overline{X}_n) = \overline{X}_n(1 - \overline{X}_n) \xrightarrow{P} p(1-p)$.

(b) Si $X_1, X_2, \ldots \stackrel{\text{iid}}{\sim} \exp(\lambda)$, la LDGN implica que $\overline{X}_n \xrightarrow{P} 1/\lambda$. Entonces tomando $g(x) = 1/x$, se tiene que $g(\overline{X}_n) = 1/\overline{X}_n \xrightarrow{P} \lambda$.

(c) Si $X_1, X_2, \ldots \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, entonces $\overline{X}_n \xrightarrow{P} \mu$ y, por lo tanto $\overline{X}_n^2 \xrightarrow{P} \mu^2$

8.1.1. Convergencia Casi Segura

Un modo de convergencia más fuerte que implica la convergencia en la probabilidad es la denominada convergencia casi segura.

Definición 8.4. *Una secuencia de variables $\{Z_n : n \geq 1\}$ converge casi seguramente (es) para θ, que se denota como $Z_n \xrightarrow{cs} \theta$ si para todo $\varepsilon > 0$*

$$P\left(\lim_{n\to\infty} |X_n - \theta| < \varepsilon\right) = 1$$

Equivalentemente, $Z_n \overset{cs}{\to} \theta$ *si*

$$P\left(\left\{\omega \in \Omega : \lim_{n\to\infty} Z_n(\omega) = \theta\right\}\right) = 1$$

Nota 8.7. Similarmente, $Z_n \overset{\text{cs}}{\to} Z$ si y sólo si $Z_n - Z \overset{\text{cs}}{\to} 0$, donde Z es una variable aleatoria definida en mismo espacio de probabilidad de la secuencia $\{Z_n : n \geq 1\}$.

El siguiente teorema es una versión más fuerte del teorem 8.3 (LDGN)

Teoremra 8.5. (Ley fuerte de los grandes números (LFGN)) *Sea* $X_1, X_2, \ldots$ *una secuencia de variables aleatorias iid, con* $E(|X_1|) < \infty$. *Entonces*

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{cs}{\to} \mu = E(X_1)$$

Es decir, $P(\{\omega \in \Omega : \lim_{n\to\infty} \overline{X}_n(\omega) = \mu\}) = 1$

Nota 8.8. Similarmente, si $\mathbf{X}_1, \mathbf{X}_2, \ldots$ es una secuencia de vectores aleatorios p-dimensionales iid, con vector de medias $\boldsymbol{\mu} = E(\mathbf{X}_1)$ finito, entonces $\overline{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \overset{\text{cs}}{\to} \boldsymbol{\mu}$.

Nota 8.9. $\mathbf{Z}_n = (Z_{1n}, \ldots, Z_{pn})^T \overset{\text{P o cs}}{\longrightarrow} \mathbf{Z} = (Z_1, \ldots, Z_p)^T$ si y sólo si $Z_{in} \overset{\text{P o cs}}{\longrightarrow} Z_i$ para todo $i = 1, \ldots, p$.

Teorema 8.6. *Sea* $\{Z_n : n \geq 1\}$ *una secuencia de variables aleatorias. Si*

$$Z_n \overset{cs}{\to} \theta$$

entonces $Z_n \overset{P}{\to} \theta$.

Dem...

8.2. Convergencia en distibución

Sea $\{Z_n : n \geq 1\}$ una secuencia de variables aleatorias definidas en un determinado espacio de probabilidad, y sea Z cualquier variable aleatoria cuya distribución no depende de n.

Definición 8.4. *Se dice que* Z_n *converge en distribución para* Z, *si*

$$F_{Z_n}(z) = P(Z_n \leq z) \overset{n\to\infty}{\longrightarrow} P(Z \leq z) = F_Z(z)$$

para todo z *donde* F_Z *es continua. (Es decir, para todo* z *talque* $P(Z = z) = 0$*)*

Notación. $Z_n \overset{d}{\to} Z$ o $F_{Z_n} \Rightarrow F_Z$.

Nota 8.. Similarmente, si $\mathbf{Z}, \mathbf{Z}_1, \mathbf{Z}_2, \ldots$ son vectores aleatorias de dimensiónn p, entonces $\mathbf{Z}_n \overset{d}{\to} \mathbf{Z}$ si y sólo si

$$\lim_{n\to\infty} F_{\mathbf{Z}_n}(z) = F_{\mathbf{Z}}(z)$$

para todo z que sea punto de continuidad de $F_{\mathbf{Z}}$.

Ejemplo 8.10.

(a) $Z_n \xrightarrow{d} Z \sim N(\mu, \sigma^2)$ si y sólo si

$$\lim_{n\to\infty} F_{Z_n}(z) = \Phi\left(\frac{z-\mu}{\sigma}\right), \text{ para todo } z \in \mathbb{R}.$$

(b) Sea $Z \equiv \theta$ (variable aleatoria degenerada), es decir, $P(Z = \theta) = 1$ y $P(Z = z) = 0$ para todo $z \neq \theta$, de modo que la fda de Z está dada por:

$$F_Z(z) = \begin{cases} 0, & z < \theta \\ 1, & z \geq \theta \end{cases}$$

En este caso, se tiene que

$$\begin{aligned} Z_n \xrightarrow{d} Z \Leftrightarrow P(Z_n \leq z) &\to P(Z \leq z), \text{ para todo } z \neq \theta. \\ \Leftrightarrow P(Z_n \leq z) &\to 0, \text{ para todo } z < \theta. \\ &\to 1, \text{ para todo } z > \theta. \end{aligned}$$

Aplicación. Sea $\{Z_n : n \leq 1\}$ tal que $P(Z_n = \theta) = 1 - \frac{1}{n}$ y $P(Z_n = \theta + n^2) = \frac{1}{n}$. Entonces, $Z_n \xrightarrow{d} Z$, en efecto

$$F_{Z_n}(z) = P(Z_n \leq z) = \begin{cases} 0, & z < \theta \\ 1 - \frac{1}{n}, & \theta \leq z < \theta + n^2 \\ 1, & z \geq \theta + n^2 \end{cases}$$

Entonces

$$\lim_{n\to\infty} F_{Z_n}(z) = \begin{cases} 0, & z < \theta \\ 1, & z > \theta \end{cases}$$

Y entonces

$$\lim_{n\to\infty} F_{Z_n}(z) = F_Z(z), \text{ para todo } z \neq \theta$$

(c) Sea $Z_n = \text{máx}\{X_1, \ldots, X_n\}$, donde $X_1, \ldots, X_n \overset{\text{iid}}{\sim} U(0, \theta)$. Entonces, $T_n = n(\theta - Z_n) \xrightarrow{d} T \sim \exp(1/\theta)$.

Sol. Vimos que

$$F_{Z_n}(z) = \begin{cases} 0, & z < 0 \\ \left(\frac{z}{\theta}\right)^n, & 0 \leq z < \theta \\ 1, & z \geq \theta \end{cases}$$

Luego

$$\begin{aligned} F_{T_n}(t) &= P(T_n \leq t) \\ &= P(n(\theta - Z_n) \leq t) \\ &= P(Z_n \geq \theta - t/n) \\ &= 1 - P(Z_n \leq \theta - t/n) \\ &= 1 - F_{Z_n}(\theta - t/n) \end{aligned}$$

Reemplazando $F_{Z_n}(z)$ evaluada en $z = \theta - \frac{t}{n}$ en esta última expresión, se tiene que

$$\begin{aligned} F_{T_n}(t) &= \begin{cases} 1, & \theta - \frac{t}{n} < 0 \Leftrightarrow t > n\theta \\ 1 - \left(1 - \frac{t}{n\theta}\right)^n, & 0 \leq \theta - \frac{t}{n} < \theta \Leftrightarrow 0 < t \leq n\theta \\ 0, & \theta - \frac{t}{n} \geq \theta \Leftrightarrow t \leq 0. \end{cases} \\ &= \begin{cases} 0, & t \leq 0 \\ 1 - \left(1 - \frac{t}{n}\right)^n, & 0 < t \leq n\theta \end{cases} \end{aligned}$$

Por lo tanto

$$\begin{aligned} \lim_{n\to\infty} F_{T_n}(t) &= \begin{cases} 0, & t \leq 0 \\ 1 - e^{-t/\theta}, & 0 < t < \infty \end{cases} \\ & F_T(t), \text{ donde } T \sim \exp(1, \theta) \end{aligned}$$

Teorema 8.7. *Sea $\{Z_n : n \geq 1\}$ una secuencia de variables aleatroia y Z una variable aleatoria que no depende de n.*

(a) $Z_n \xrightarrow{d} Z$ si y sólo si $\lim_{n\to\infty} M_{Z_n}(t) = M_Z(t)$ para todo t en algún intervalo que contega el 0 y donde las fgm's existen.

(b) $Z_n \xrightarrow{d} Z$ si y sólo $\lim_{n\to\infty} \varphi_{Z_n}(t) = \varphi_Z(t)$ para todo t, donde $\varphi_{Z_n}(t)$ es la f.c. de $Z_n, n \geq 1$ y $\varphi_Z(t)$ es la f.c. de Z.

Nota 8.. Si $\mathbf{Z}, \mathbf{Z}_1, \mathbf{Z}_2, \ldots$ son vectores aleatorias p-dimensionales, entonces $\mathbf{Z}_n \xrightarrow{d} \mathbf{Z}$ si y sólo si $\mathbf{a}^T\mathbf{Z}_n \xrightarrow{d} \mathbf{a}^T\mathbf{Z}$ paratodo $\mathbf{a} \in \mathbb{R}^p$.

Aplicación. Para demostrar que $\mathbf{Z}_n \xrightarrow{d} Z$ es suficiente probar que $M_{Z_n}(t) \to M_Z(t)$ (o $\varphi_{Z_n}(t) \to \varphi_Z(t)$) cuando $n \to \infty$.

Ejemplo 8.11.

(a) $Z_n \xrightarrow{d} Z \sim N(\mu, \sigma^2)$ si y sólo $\lim_{n\to\infty} M_{Z_n}(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ para todo $t \in \mathbb{R}$.

(b) Sea $Z_n \sim Bin(n, p_n)$, donde $np_n = \lambda > 0$ para todo $n = 1, 2, \ldots$. Entonces, $Z_n \xrightarrow{d} Z \sim P(\lambda)$.

Dem. Como $M_Z(t) = e^{\lambda(e^t-1)}$, $t \in \mathbb{R}$, basta con probar que $\lim_{n\to\infty} M_{Z_n}(t) = e^{\lambda(e^t-1)}$ para todo $t \in \mathbb{R}$. Por otro lado

$$M_{Z_n}(t) = (1 + p_n(e^t - 1))^n = \left(1 + \frac{\lambda(e^t-1)}{n}\right)^n, \ (p_n = \lambda/n)$$

Entonces

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty} \left(1 + \frac{\lambda(e^t-1)}{n}\right)^n = e^{\lambda(e^t-1)}, \text{ para todo } t \in \mathbb{R}$$

(c) Sea $Z_n \sim P(n), n \geq 1$, por ejemplo $Z_n = \sum_{i=1}^n X_i$, donde, para cada $n \geq 1$, $X_1, \ldots, X_n \overset{\text{iid}}{\sim} P(1)$. Entonces

$$T_n = \frac{Z_n - n}{\sqrt{n}} \xrightarrow{d} T \sim N(0,1)$$

cunado $n \to \infty$. Note que $E(Z_n) = \text{Var}(Z_n) = n$, de modo que $E(T_n) = 0$ y $\text{Var}(T_n) = 1$.

Dem. La definición de la fgm de T_n, implica que

$$M_{T_n}(t) = E\left(e^{t\left(\frac{Z_n-n}{\sqrt{n}}\right)}\right) = e^{-t\sqrt{n}M_{Z_n}(t/\sqrt{n})}, \ t \in \mathbb{R}$$

donde $M_{Z_n}(s) = e^{n(e^s-1)}, s \in \mathbb{R}$, ya que $Z_n \sim P(n)$. Luego

$$M_{T_n}(t) = e^{-t\sqrt{n}}e^{ne^{t/\sqrt{n}}-1)} = e^{nh(t/\sqrt{n})}$$

8.3. Algunos modos de convergencia

Sea $\{Z_n : n \geq 1\}$ una secuencia de variables aleatorias definidas en un determinado espacio de probabilidad, y sea Z cualquier variable aleatoria cuya distribución no depende de n.

Definición 8.5. *Se dice que Z_n converge en distribución para Z si,*

$$F_{Z_n}(z) = P(Z_n \leq z) \to P(Z \leq z) = F_Z(z)$$

cuando $n \to \infty$, para todo z donde F_Z es continua.

Notación. $Z_n \xrightarrow{d} Z$ o $F_{Z_n} \Rightarrow F_Z$.

Nota 8.. Similarmente, si $\mathbf{Z}, \mathbf{Z}_1, \mathbf{Z}_2, \ldots$ son vectores aleatorias de dimensión p, entonces $\mathbf{Z}_n \xrightarrow{d} \mathbf{Z}$ si y sólo si $\lim_{n\to\infty} F_{\mathbf{Z}_n}(\mathbf{z}) = F_{\mathbf{Z}}(\mathbf{z})$ para todo $\mathbf{z}$ que sea punto de continuidad de $F_{\mathbf{Z}}$.

Ejemplo 8.12.

(a) $Z_n \xrightarrow{d} Z \sim N(\mu, \sigma^2)$ si y sólo si

$$\lim_{n\to\infty} F_{Z_n}(z) = \Phi\left(\frac{z-\mu}{\sigma}\right)$$

para todo $z \in \mathbb{R}$.

(b) Sea $Z \equiv \theta$ (variable aleatoria degenerada); es decir, $P(Z = \theta) = 1$ y $P(Z = z) = 0$ para todo $z \neq \theta$, de modo que la fda de Z está dada por:

$$F_Z(z) = \begin{cases} 0, & z < \theta \\ 1, & z \geq \theta \end{cases}$$

En este caso, se tiene que

$$\begin{aligned} Z_n \xrightarrow{d} Z &\Leftrightarrow P(Z_n \leq z) \to P(Z \leq z) \text{ para todo } z \neq \theta \\ &\Leftrightarrow P(Z_n \leq z) \to 0 \text{ para todo } z < \theta \\ &\qquad\qquad\quad\;\; \to 1 \text{ para todo } z > \theta \end{aligned}$$

Aplicación. Sea $\{Z_n : n \leq 1\}$ tal que $P(Z_n = \theta) = 1 - \frac{1}{n}$ y $P(Z_n = \theta + n^2) = \frac{1}{n}$. Entonces, $Z_n \xrightarrow{d} Z$, en efecto

$$\begin{aligned} F_{Z_n}(z) = P(Z_n \leq z) &= \begin{cases} 0, & z < \theta \\ 1 - \frac{1}{n}, & \theta \leq z < \theta + n^2 \\ 1, & z \geq \theta + n^2 \end{cases} \\ \Rightarrow \lim_{n\to\infty} F_{Z_n}(z) &= \begin{cases} 0, & z < \theta \\ 1, & z > \theta \end{cases} \\ \Rightarrow \lim_{n\to\infty} F_{Z_n}(z) &= F_Z(z), \text{ para todo } z \neq \theta \end{aligned}$$

(c) Sea $Z_n = \text{máx}\{X_1, \ldots, X_n\}$, donde $X_1, \ldots, X_n \overset{\text{iid}}{\sim} U(0, \theta)$. Entonces, $T_n = n(\theta - Z_n) \xrightarrow{d} T \sim \exp(1/\theta)$.

Dem. Vimos que

$$F_{Z_n}(z) = \begin{cases} 0, & z < 0 \\ \left(\frac{z}{\theta}\right)^n, & 0 \leq z < \theta \\ 1, & z \geq \theta \end{cases}$$

Luego,

$$\begin{aligned} F_{T_n}(t) = P(T_n \leq t) &= P(n(\theta - Z_n) \leq t) \\ &= P(Z_n \geq \theta - t/n) \\ &= 1 - P(Z_n \leq \theta - t/n) \\ &= 1 - F_{Z_n}(\theta - t/n) \end{aligned}$$

Reemplazando $F_{Z_n}(z)$ evaluada en $z = \theta - \frac{t}{n}$ en esta última expresión, se tiene que,

$$\begin{aligned} F_{T_n}(t) &= \begin{cases} 1, & \theta - \frac{t}{n} < 0 \Leftrightarrow t > n\theta \\ 1 - \left(1 - \frac{t}{n\theta}\right)^n, & 0 \leq \theta - \frac{t}{n} < \theta \Leftrightarrow 0 < t \leq n\theta \\ 0, & \theta - \frac{t}{n} \geq \theta \Leftrightarrow t \leq 0 \end{cases} \\ &= \begin{cases} 0, & t \leq 0 \\ 1 - \left(1 - \frac{t}{n\theta}\right)^n, & 0 < t \leq n\theta \end{cases} \end{aligned}$$

Por lo tanto,

$$\lim_{n\to\infty} F_{T_n}(t) = \begin{cases} 0, & t \leq 0 \\ 1 - e^{-t/\theta}, & 0 < t < \infty \end{cases}$$
$$= F_T(t)$$

donde $T \sim \exp(1/\theta)$.

Teorema 8.8. *Sea $\{Z_n : n \geq 1\}$ una secuencia de variables aleatorias y Z una variable que no depende de n.*

(a) $Z_n \xrightarrow{d} Z \Leftrightarrow \lim_{n\to\infty} M_{Z_n}(t) = M_Z(t)$ para todo t en algún intervalo que contega el 0 y donde las fgm's existen.

(b) $Z_n \xrightarrow{d} Z \Leftrightarrow \lim_{n\to\infty} \varphi Z_n(t) = \varphi_Z(t)$ para todo t, donde $\varphi_{Z_n}(t)$ es la f.c. de $Z_n, n \geq 1$, y $\varphi_Z(t)$ es la f.c. de Z.

Nota 8.. Si $\mathbf{Z}, \mathbf{Z}_1, \mathbf{Z}_2, \ldots$ son vectores aleatorias p-dimensionales, entonces, $\mathbf{Z}_n \xrightarrow{d} \mathbf{Z} \Leftrightarrow \mathbf{a}^T\mathbf{Z}_n \xrightarrow{d} \mathbf{a}^T\mathbf{Z}$ para todo $\mathbf{a} \in \mathbb{R}^p$.

Aplicación. Para demostrar que $Z_n \xrightarrow{d} Z$ es suficiente probar que $M_{Z_n}(t) \to M_Z(t)$ (o $\varphi_{Z_n}(t) \to \varphi_Z(t)$), cuando $n \to \infty$.

Ejemplo 8.13.

(a) Para demostrar que $Z_n \xrightarrow{d} Z$ es suficiente probar que $M_{Z_n}(t) \to M_Z(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ para todo $t \in \mathbb{R}$.

(b) Sea $Z_n \sim Bin(n, p_n)$, donde $np_n = \lambda > 0$ para todo $n = 1, 2, \ldots$. Entonces, $Z_n \xrightarrow{d} Z \sim P(\lambda)$.

Dem. Como $M_Z(t) = e^{\lambda(e^t-1)}$, $t \in \mathbb{R}$, basta con probar que $\lim_{n\to\infty} M_{Z_n}(t) = e^{\lambda(e^t-1)}$ para $t \in \mathbb{R}$. Por otro lado,

$$M_{Z_n}(t) = (1 + p_n(e^t - 1))^n = \left(1 + \frac{\lambda(e^t-1)}{n}\right)^n, \ (p_n = \lambda/n)$$
$$\Rightarrow \lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty}\left(1 + \frac{\lambda(e^t-1)}{n}\right)^n = e^{\lambda(e^t-1)}$$

para todo $t \in \mathbb{R}$.

(c) Sea $Z_n \sim P(n), n \geq 1$; por ejemplo, $Z_n = \sum_{i=1}^n X_i$, donde, para cada $n \geq 1$, $X_1, \ldots, X_n \overset{\text{iid}}{\sim} P(1)$. Entonces

$$T_n = \frac{Z_n - n}{\sqrt{n}} \xrightarrow{d} T \sim N(0,1)$$

cuando $n \to \infty$. Note que $E(Z_n) = \text{Var}(Z_n) = n$, de modo que $E(T_n) = 0$ y $\text{Var}(T_n) = 1$.
Dem. La definición de la fgm de T_n implica que,

$$M_{T_n}(t) = E\left(e^{t\left(\frac{Z_n-n}{\sqrt{n}}\right)}\right) = e^{-t\sqrt{n}} M_{Z_n}(t/\sqrt{n})$$

para todo $t \in \mathbb{R}$. Donde $M_{Z_n}(s) = e^{n(e^s-1)}$ con $s \in \mathbb{R}$, ya que $Z_n \sim P(n)$. Luego,

$$M_{T_n}(t) = e^{-t\sqrt{n}} e^{n(e^{t/\sqrt{n}}-1)} = e^{nh(t/\sqrt{n})}$$

donde $h(t/\sqrt{n}) = e^{\frac{t}{\sqrt{n}}} - 1$. Una expasión de Taylor de segundo orden de $e^x|_{x=t/\sqrt{n}}$ alrededor de 0 implica que

$$\begin{aligned} h(t/\sqrt{n}) &= 1 + \frac{t}{\sqrt{n}} + \frac{1}{2}\frac{t^2}{(\sqrt{n})^2} + \frac{1}{6}\frac{t^3}{(\sqrt{n})^3} e^{\xi\frac{t}{\sqrt{n}}} - \frac{t}{\sqrt{n}} - 1 \\ &= \frac{1}{2}\frac{t^2}{n} + \frac{1}{6}\frac{t^3}{n^{3/2}} e^{\xi\frac{t}{n}} \end{aligned}$$

donde $0 < \xi < 1$. Así, multiplicando por n, se tiene que,

$$\begin{aligned} nh(t/\sqrt{n}) &= \frac{1}{2}t^2 + \frac{1}{6}\frac{t^3}{\sqrt{n}} e^{\xi\frac{t}{\sqrt{n}}} \overset{n\to\infty}{\longrightarrow} \frac{1}{2}t^2 \\ \Rightarrow M_{T_n}(t) &= e^{nh(t/\sqrt{n})} \overset{n\to\infty}{\longrightarrow} e^{t^2/2} \end{aligned}$$

donde $M_T(t) = e^{t^2/2}$ esla fgm de $T \sim N(0,1)$.

Tarea. Si $Z_n \sim \chi^2_n, n \geq 1$, demuestre que $(Z_n - n)/\sqrt{2n} \overset{d}{\to} N(0,1)$, cuando $n \to \infty$. Pruebe tambien que $Z_n/n \overset{P}{\to} 1$, cuando $n \to \infty$.

Teorema 8.9. *Si $Z_n \overset{d}{\to} Z$, entonces $g(Z_n) \overset{d}{\to} g(Z)$ para toda función continua g.*

Dem.

Ejemplo 8.14

(a) Si $Z_n = \text{máx}\{X_1, \ldots, X_n\}$, donde $X_1, \ldots, X_n \overset{\text{iid}}{\sim} U(0,\theta)$, vimos que

$$T_n = n(\theta - Z_n) \overset{d}{\to} \exp(1/\theta)$$

cuando $n \to \infty$. Entonces, por el teorema 8.9 se tiene que,

$$g(T_n) = \frac{T_n}{\theta} \overset{d}{\to} \frac{T}{\theta} \sim \exp(1)$$

cuando $n \to \infty$.

(b) Si $Z_n \overset{d}{\to} Z \sim N(\mu, \sigma^2)$, entonces $\frac{Z_n-\mu}{\sigma} \overset{d}{\to} \frac{Z-\mu}{\sigma} \sim N(0,1)$ y, por tanto, $\left(\frac{Z_n-\mu}{\sigma}\right)^2 \overset{d}{\to} \left(\frac{Z_n-\mu}{\sigma}\right)^2 \sim \chi^2_1$, cuando $n \to \infty$.

Teorema 8.10. *Sea $X_1, X_2, \ldots$ una secuencia de variables aleatorias iid con media $E(X_1) = \mu$ y varianza $0 < Var(X_1) = \sigma^2 < \infty$. Entonces*

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \overset{d}{\to} Z \sim N(0,1)$$

cuando $n \to \infty$. Es decir

$$\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z)$$

para todo $z \in \mathbb{R}$, donde $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)dt$ es la fda de $Z \sim N(0,1)$.

Dem. Se define $Y_i = (X_i - \mu)/\sigma$ $i = 1, 2, \ldots$. La secuencia $Y_1, Y_2, \ldots$ son variables aleatorias iid con $E(Y_1) = 0$ y $\text{Var}(Y_1) = E(Y_1^2) = 1$. Sea $M_{Y_1}(t)$ la fgm común de $Y_1, Y_2, \ldots$, la cual se supone que existe para $|t| < \sigma h$, algún $h > 0$. Vimos que si la fgm de una variable aleatoria Y_1 existe en un intervalo que contiene el cero, entonces la expasión Taylor de alrededor de cero produce que,

$$M_{Y_1}(s) = 1 + \sum_{t=1}^{\infty} \underbrace{M_{Y_1}^{(k)}(0)}_{E(Y_1^k)} \frac{s}{k!} = 1 + \frac{s^2}{2} + \underbrace{\sum_{k=3}^{\infty} M_{Y_1}^{(k)}(0)\frac{s}{k!}}_{\text{resto}=\, r(s)},\ (\star)$$

ya que $M_{Y_1}^{(1)}(0) = E(Y_1) = 0$ y $M_{Y_1}^{(2)}(0) = E(Y_1^2) = 1$ y donde $r(s)/s^2 \to 0$ cuando $s \to 0$ de acuerdo con el teorema de Taylor. Por otro lado, en términos de las variables estandarizadas $Y_1, \ldots, Y_n$, tenemos que

$$Z_n = \frac{\sum_{i=1} nX_i - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$$

Luego, la fgm de Z_n está dada por

$$\begin{aligned}
M_{Z_n}(t) &= M_{\sum_{i=1} nY_i/\sqrt{n}}(t) \\
&= M_{\sum_{i=1}^n Y_i}(t/\sqrt{n}) \\
&= \left(1 + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + r\left(\frac{t}{\sqrt{n}}\right)\right)^n, \text{ por } (\star) \\
&= \left(1 + \frac{1}{n}\left(\frac{t^2}{2} + nr\left(\frac{t}{\sqrt{n}}\right)\right)\right)^n
\end{aligned}$$

Finalmente, tomando límite cuando $n \to \infty$, se tiene que

$$\begin{aligned}
\lim_{n\to\infty} M_{Z_n}(t) &= \lim_{n\to\infty} \left(1 + \frac{1}{n}\left(\frac{t^2}{2} + nr\left(\frac{t}{\sqrt{n}}\right)\right)\right)^n \\
&= \exp(t^2/2)
\end{aligned}$$

ya que $nr(t/\sqrt{n}) = t^2 r(t/\sqrt{n})/(t/\sqrt{n})^2 \overset{n\to\infty}{\longrightarrow} 0$.

Nota 8.. La demostración presupone la existencia de la fgm de la secuencia $X_1, X_2, \ldots$ lo cual límita el alcance del TLC. Una demostración similar puede realizarse en base a la función carastarística la cual siempre existe, extendiendo la aplicación del TLC a cualquier secuencias de variables aleatorias iid con segundo momento finito.

Nota 8.. Similarmente, si $\mathbf{X}_1, \mathbf{X}_2, \ldots$ es una secuencia vectores aleatorias P-dimensionales iid con $\boldsymbol{\mu} = E(\mathbf{X}_1)$ y $\boldsymbol{\Sigma} = \text{Var}(\mathbf{X}_1)$ (finita), entonces $\sqrt{n}(\overline{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Sigma})$.

Aplicación. Para n suficientemente grande, se tiene que

$$\frac{n(\overline{X}_n - \mu)}{\sigma} \simeq N(0,1) \text{ o, equivalentemente, } \overline{X}_n \simeq N(\mu, \sigma^2/n)$$

es decir, para n lo suficientemente grande

$$\begin{aligned} P(\overline{X}_n \leq x) &= P\left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \leq \frac{\sqrt{n}(x-\mu)}{\sigma}\right) \\ &\simeq \Phi\left(\frac{\sqrt{n}(x-\mu)}{\sigma}\right) \end{aligned}$$

para todo $x \in \mathbb{R}$.

Ejemplo 8.15.

(a) Sea $X_1, X_2, \ldots$ una secuencia de variables aleatorias iid $Ber(p)$. Note que $E(X_1) = p$ y $\text{Var}(X_1) = p(1-p)$ son ambos finitos. Además, en este caso, se debe notar que, $\sum_{i=1}^n X_i \sim Bin(n,p)$ $n = 1, 2, \ldots,$ es una secuencia de variables aleatorias discretas.

Sin embargo, por el TLC se tiene que

$$\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} = \frac{\sqrt{n}(\overline{X}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{d} N(0,1)$$

cuando $n \to \infty$. Es decir, para n lo suficientemente grande el TLC permite aproximar la distribución $Bin(n,p)$ por una distribución $N(np, np(1-p))$, de modo que si $T_n = \sum_{i=1}^n X_i \sim Bin(n,p)$ y $k_1 < k_2$ son enteros no negativos, entonces

$$P(k_1 \leq S_n \leq k_2) \simeq \Phi\left(\frac{k_2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{np(1-p)}}\right)$$

Nota. Aplicando la denominada correción por continuidad, se tiene que

$$P(k_1 \leq T_n \leq k_2) \simeq \Phi\left(\frac{k_2 + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Por ejemplo, si $n = 36$ y $p = 0{,}5$, un cálculo exacto da

$$P(T_n \leq 21) = \sum_{k=0}^{21} \binom{36}{k} (0{,}5)^{36} = 0{,}8785$$

La aproximación sin la correción por continuidad, da

$$P(T_n \leq 21) \simeq \Phi\left(\frac{21 - np}{\sqrt{np(1-0)}}\right) = \Phi\left(\frac{21-18}{3}\right) = \Phi(1) = 0{,}8413$$

Si se aplica la correción por continuidad, se tiene

$$P(T_n \leq 21) \simeq \Phi\left(\frac{21{,}5 - np}{\sqrt{np(1-p)}}\right) = \Phi\left(\frac{21{,}5 - 18}{3}\right) = \Phi(1{,}17) = 0{,}879$$

la cual es mucho más cercana al valor exacto.

Nota. El resultado sobre la aproximación de la binomial por una normal se conoce como Aproximación (o Teorema) de Demoivre-Laplace.

(b) Si $X_1, X_2, \ldots$ son variables aleatorias iid χ^2_1, entonces

$$\frac{\sum_{i=1}^n X_i - n}{\sqrt{2n}} \xrightarrow{d} N(0,1)$$

cuando $n \to \infty$. Note que $\sum_{i=1}^n X_i \sim \chi^2_n$, con media n y varianza $2n$, $n = 1, 2, \ldots$ y para n suficientemente grande, se tiene que $\sum_{i=1}^n X_i \simeq N(n, 2n)$.

(c) Si $X_1, X_2, \ldots$ son variables aleatorias iid $P(1)$, entonces

$$\frac{\sum_{i=1}^n X_i - n}{\sqrt{n}} \xrightarrow{d} N(0,1)$$

cuando $n \to \infty$. En este caso, $\sum_{i=1}^n X_i \sim P(n)$, con media y varianza n, $n = 1, 2, \ldots$ y para n suficientemente grande, se tiene que, $\sum -i = 1^n X_i \simeq N(n, n)$, y asi entre otros.

Vimos que tanto la convergencia en media cuadrática como la convergencia casi seguramente implica la convergencia en probabilidad. El siguiente resultado establece que la convergencia en probabilidad implica la convergencia en distribución.

Teorema 8.11. *Si $Z_n \xrightarrow{P} Z$, entonces $Z_n \xrightarrow{d} Z$.*

Dem.

Teorema 8.12. *Para Z_n se tiene que $Z_n \xrightarrow{P} \theta$ si y sólo si $Z_n \xrightarrow{d} \theta$.*

Dem. Supongamos que $Z_n \xrightarrow{P} \theta$. Entonces, por definición de probabilidad, se tiene que para todo $\varepsilon > 0$

$$\begin{aligned}
1 = \lim_{n\to\infty} P(|Z_n - \theta| < \varepsilon)& \\
&= \lim_{n\to\infty} P(\theta - \varepsilon < Z_n < \theta + \varepsilon) \\
&= \lim_{n\to\infty} (P(Z_n < \theta + \varepsilon) - P(Z_n < \theta - \varepsilon)) \\
&= \lim_{n\to\infty} P(Z_n < \theta + \varepsilon) - \lim_{n\to\infty} P(Z_n \leq \theta - \varepsilon)
\end{aligned}$$

lo cual ocurre si y sólo si para todo $\varepsilon > 0$

$$\lim_{n\to\infty} P(Z_n < \theta + \varepsilon) = 1 \text{ y } \lim_{n\to\infty} P(Z_n \leq \theta - \varepsilon) = 0$$

Es decir, si y sólo si

$$(\star)\ \lim_{n\to\infty} P(Z_n \leq z) = \begin{cases} 0, & z < \theta \\ 1, & z > \theta \end{cases}$$
$$= F_Z(z)$$

para todo $z \neq \theta$, donde $Z = \theta$ con probabilidad 1. Por lo tanto $Z_n \xrightarrow{d} Z \equiv \theta$. Entonces, por la definición convergencia en distribución, debe tenerse que

$$\lim_{n\to\infty} F_{Z_n}(z) = F_Z(z) = \begin{cases} 0, & z < \theta \\ 1, & z > \theta \end{cases}$$

para todo $z \neq \theta$. Entonces

$$\begin{aligned} \lim_{n\to\infty} P(\theta - \varepsilon < Z_n < \theta + \varepsilon) &= \lim_{n\to\infty} P(Z_n < \theta + \varepsilon) - \lim_{n\to\infty} P(z_n \leq \theta - \varepsilon) \\ &= 1 - 0 \text{ por } (\star) \end{aligned}$$

Por lo tanto, $Z_n \xrightarrow{P} \theta$. ■

Nota 8.. La extensión de los teoremas 8.11, 8.12 al caso p-dimensional es indemiata.

Aplicación. Para demostrar que $Z_n \xrightarrow{P} \theta$, es suficiente probar que $Z_n \xrightarrow{d} \theta$.

Ejemplo 8.16. Si $\varphi_{Z_n}(t) = E(e^{itZ_n}) \overset{n\to\infty}{\longrightarrow} e^{it\theta}$, entonces por la parte (b) del teorema 8.. se tiene que $Z_n \xrightarrow{d} \theta$ y, por lo tanto, el teorema 8.12 implica que $Z_n \xrightarrow{P} \theta$.

Con este resultado, se puede discutir la demostración de la LDGN sin tener que suponer la existencia del segundo momento para la secuencia de variables aleatorias $X_1, X_2 \ldots$ iid con media finita μ. En efecto, la función característica de $\overline{X}_n = \sum_{i=1}^n X_i/n$ es

$$\varphi_{\overline{X}_n}(t) = (\varphi_{X_1}(t/n))^n \overset{(\star\star)}{=} \left(1 + i\mu\frac{t}{n} + o(1/n)\right)^n$$

donde $no(1/n) = o(1) \overset{n\to\infty}{\to} 0$. La igualdad $(\star\star)$ se debe al uso de una expansión de Taylot de primer orden de cero para $\varphi_{X_1}(t/n)$; es decir

$$\varphi_{X_1}(t/n) = \varphi_{X_1}(0) + \varphi'_{X_1}(0)(t/n) + o(1/n)$$

cuando $n \to \infty$, donde $\varphi_{X_1}(0) = 1$ y $\varphi'_{X_1}(0) = i\mu$. Así, el resultado se obtiene al aplicar el siguiente lema:

Lema 8.1.

$$\lim_{n\to\infty} \left(1 + \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^n = e^\lambda$$

Teorema 8.13. (Slutsky) *Sean* $\{X_n : n \geq 1\}$ *e* $\{Y_n : n \geq 1\}$ *secuencias de variables aleatorias tales que* $X_n \xrightarrow{d} X$ *e* $Y_n \xrightarrow{P} b$ *(b constante). Entonces, para cualquier función continua* $g : \mathbb{R}^2 \to \mathbb{R}$*, se tiene que:*

$$g(X_n, Y_n) \xrightarrow{d} g(X, b)$$

Nota 8.. La extensión secuencias multivariadas es inmediata.

Idea de la demostración. Como $X_n \xrightarrow{d} X$ e $Y_n \xrightarrow{P} b$, se tiene que $(X_n, Y_n) \xrightarrow{d} (X, b)$, y por tanto, $g(X_n, Y_n) \xrightarrow{d} g(X, b)$, por el teorema de la aplicación continua.

Corolario 8.1. (Versión clásica de Teorema de Slutsky) *Si* $X_n \xrightarrow{d} X$ *e* $Y_n \xrightarrow{P} b$*, entonces*

(a) $X_n + Y_n \xrightarrow{d} X + b$.

(b) $Y_n X_n \xrightarrow{d} bX$.

(c) $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{b}$ *siempre que* $b \neq 0$.

Aplicación. Supongamos, por ejemplo, que $X \sim N(\mu, \sigma^2)$. Entonces

(a) $X_n + Y_n \xrightarrow{d} X + b \sim N(\mu + b, \sigma^2)$, y por lo tanto

$$P(X_n + Y_n \leq z) \simeq P(X + b \leq z) = \Phi\left(\frac{z - \mu - b}{\sigma}\right)$$

para n suficientemente grande.

(b) $X_n Y_n \xrightarrow{d} Xb \sim N(\mu b, \sigma^2 b^2)$, y por lo tanto

$$P(X_n Y_n \leq z) \simeq P(Xb \leq z) = \Phi\left(\frac{z - \mu b}{\sigma b}\right)$$

para n suficientemente grande.

(c) $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{b} \sim N(\mu/b, \sigma^2/b^2)$ para $b \neq 0$, y por lo tanto

$$P\left(\frac{X_n}{Y_n} \leq z\right) \simeq P\left(\frac{X}{b} \leq z\right) = \Phi\left(\frac{bz - \mu}{\sigma}\right)$$

para n suficientemente grande.

Teorema 8.14. *Sean* $\{X_n : n \geq 1\}$ *e* $\{Y_n : n \geq 1\}$ *secuencias de variables aleatorias tales que* $X_n \xrightarrow{P} a$ *e* $Y_n \xrightarrow{P} b$*, donde* a *y* b *constantes. Entonces, para cualquier función continua* $g : \mathbb{R}^2 \to \mathbb{R}$*, se tiene que:*

$$g(X_n, Y_n) \xrightarrow{P} g(a, b)$$

Ejemplo 8.17.

(a) Sean $X_1, X_2, \ldots$ variables aleatorias iid con media μ y varianza σ^2 (finita). Entonces, por la LDGN se tiene que

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu = E(X_1),\ (\star)$$
$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} \mu^2 + \sigma^2 = E(X_1^2),\ (\star\star)$$

Considerando $(\star)$ y el hecho que la función $g(x) = x^2$ es continua, se tiene que

$$g(\overline{X}_n) = \overline{X}_n^2 \xrightarrow{P} \mu^2,\ (\star\star\star)$$

Luego, usando $(\star\star), (\star\star\star)$ y el teorema 8.14 con $g(x, y) = x - y^2$, se concluye que

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2 \xrightarrow{P} \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

Sea

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,\ n \geq 2$$

Entonces

$$S_n^2 = \underbrace{\frac{n}{n-1}}_{\xrightarrow{P} 1}\underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}_{\xrightarrow{P}\sigma^2} \xrightarrow{P} \sigma^2$$

De aquí, también se tieen que, $S_n = \sqrt{S_n^2} \xrightarrow{P} \sigma$ o bien, $S_n/\sigma \xrightarrow{P} 1$ o $\sigma/S_n \xrightarrow{P} 1$. Por otro lado, por el TLC se sabe que

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$

y como fue visto, $\sigma/S_n \xrightarrow{P} 1$. Luego, aplicando Slutsky (más específicamente, la parte (b) del corolario 8.1), se tiene que

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} = \underbrace{\frac{\sigma}{S_n}}_{\xrightarrow{P} 1}\underbrace{\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}}_{\xrightarrow{d} N(0,1)} \xrightarrow{d} N(0, 1)$$

(b) Sean $X_1, X_2, \ldots$ variables aleatorias iid $N(\mu, \sigma^2)$. Se vio que

$$\begin{cases} \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \sim N(0, 1), & \forall n \geq 1 \\ \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}, & \forall n \geq 2 \end{cases}$$

Y son variables aleatorias independientes. Bajo este contexto, también se vio que

$$T_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \sim t_{n-1}, \text{ (exacta) } \forall n \geq 1$$

donde t_ν denota la distribución t (de Student) con ν grados de libertad. Además, por el ejemplo 8.17, se tiene que

$$T_n \xrightarrow{d} N(0,1)$$

cuando $n \to \infty$ y por tanto que $t_\nu \xrightarrow{d} N(0,1)$, cuando $\nu \to \infty$.

(c) Sean $\{a_n : n \geq 1\}$ una secuencia de números reales y $\{Z_n : n \geq 1\}$ una secuencia de variables aleatorias, tales que

1.) $a_n \xrightarrow{n\to\infty} \infty$ y

2.) $a_n(Z_n - \theta) \xrightarrow{d} X$, para alguna constante θ y variable aleatoria X.

Entonces

$$Z_n \xrightarrow{P} \theta$$

Dem. Basta demostrar que $Z_n - \theta \xrightarrow{P} 0$. Para probar esto, note que

$$Z_n - \theta = \frac{1}{a_n} a_n(Z_n - \theta) = Y_n X_n$$

donde $Y_n = 1/a_n \xrightarrow{n\to\infty} 0$, y por ende que $Y_n \xrightarrow{P} 0$ y $X_n = a_n(Z_n - \theta) \xrightarrow{d} X$. Luego, usando Slutsky, se tiene que $Z_n - \theta = Y_n X_n \xrightarrow{d} 0 \cdot X = 0$, o equivalentemente, $Z_n - \theta \xrightarrow{P} 0$.

Aplicación (c). Sea $X_1, X_2, \ldots$ una secuencia de variables aleatorias iid con media μ y varianza σ^2 (finita). Entonces

$$\begin{aligned}
&\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1) \text{ por el TLC} \\
\Leftrightarrow\ & \sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0,\sigma^2) \\
\Rightarrow\ & \overline{X}_n \xrightarrow{P} \mu \ (a_n = \sqrt{n},\ Z_n = \overline{X}_n,\ \theta = \mu)
\end{aligned}$$

Teorema 8.15. *Sean $\{a_n : n \geq 1\}$ una secuencia de números reales y $\{Z_n : n \geq 1\}$ una secuencia de variables aleatorias, tales que:*

(a) $a_n \xrightarrow{n\to\infty} \infty$ *y*

(b) $a_n(Z_n - \theta) \xrightarrow{d} Z$*, para alguna constante θ y variable aleatoria Z. Si $g : \mathbb{R} \to \mathbb{R}$ es una función continua y diferenciable en θ, entonces*

$$a_n(g(Z_n) - g(\theta)) \xrightarrow{d} g'(\theta) Z$$

cuando $n \to \infty$, donde g' es la derivada de g.

Idea de la demostración. Como fue probado, si $a_n(Z_n - \theta) \xrightarrow{d} Z$, entonces $Z_n - \theta \xrightarrow{P} 0$, o bien, $Z_n \xrightarrow{P} \theta$. Luego, como g es continua, entonces, $g(Z_n) \xrightarrow{P} g(\theta)$, o bien, $g(Z_n) - g(\theta) \xrightarrow{P} 0$. Así, se tiene que, entonces, $g(Z_n) \xrightarrow{P} g(\theta)$, o bien, $g(Z_n) - g(\theta) \xrightarrow{P} 0$. Así, se tiene que

$$\frac{g(Z_n) - g(\theta)}{Z_n - \theta} = \frac{g(\theta + (Z_n - \theta)) - g(\theta)}{Z_n - \theta} \xrightarrow{P} g'(\theta)$$

Por lo tanto, usando Slutsky, se tiene que

$$a_n(g(Z_n) - g(\theta)) = \underbrace{\frac{g(Z_n) - g(\theta)}{Z_n - \theta}}_{\xrightarrow{P} g'(\theta)} \underbrace{a_n(Z_n - \theta)}_{\xrightarrow{d} Z} \xrightarrow{d} g'(\theta)Z$$

Ejemplo 8.18. (Método delta) Suponga que $\sqrt{n}(Z_n - \theta) \xrightarrow{d} Z \sim N(0.\sigma^2)$, cuando $n \to \infty$, y sea g una función continua y diferenciable en θ. Entonces

$$\sqrt{n}(g(Z_n) - g(\theta)) \xrightarrow{d} g'(\theta)Z \sim N(0, (g'(\theta))^2\sigma^2)$$

Nota 8.. Si $g'(\theta) = 0$, entonces $\sqrt{n}(g(Z_n) - g(\theta)) \xrightarrow{d} N(0, 0) \equiv 0$.

Nota 8.. La extensión al caso multivariado en inmediata.

Apliación. Sean $X_1, X_2, \ldots$ variables aleatorias iid con media μ y varianza σ^2 (finita). Entonces, por el TLC, se tiene que

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} Z \sim N(0, \sigma^2)$$

Luego, si g es una función continua y diferenciable en μ, el método delta implica que

$$\sqrt{n}(g(\overline{X}_n) - g(\mu)) \xrightarrow{d} g'(u)Z \sim N(0, (g'(\mu))^2\sigma^2)$$

Ejemplo 8.19.

(a) Si $X_1, X_2, \ldots$ variables aleatorias iid con media μ y varianza σ^2 (finita), entonces $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$, y por tanto

1) $\sqrt{n}(\overline{X}_n^2 - \mu^2) \xrightarrow{d} N(0, 4\mu^2\sigma^2)$ ($g(x) = x^2$ y $g'(x) = 2x$)

2) $\sqrt{n}(e^{\overline{X}_n} - e^{\mu}) \xrightarrow{d} N(0, e^{2\mu}\sigma^2)$ ($g(x) = e^x$ y $g'(x) = e^x$)

3) $\sqrt{n}(\Phi(\overline{X}_n) - \Phi(\mu)) \xrightarrow{d} N(0, \phi(\mu)^2\sigma^2)$ ($g(x) = \Phi(x)$ y $g'(x) = \phi(x)$, la fda y fdp de la distribución $N(0, 1)$).

(b) Si $X_1, X_2, \ldots$ variables aleatorias iid $\exp(\lambda)$ ($\lambda > 0$), entonces $\sqrt{n}(\overline{X}_n - \frac{1}{\lambda}) \xrightarrow{d} N\left(0, \frac{1}{\lambda^2}\right)$, y por tanto

1) $\sqrt{n}\left(\frac{1}{\overline{X}_n} - \lambda\right) \xrightarrow{d} N(', \lambda^2)$ ($g(x) = 1/x$ y $g'(x) = 1/x^2, x > 0$).

2) $\sqrt{n}\left(\log \overline{X}_n - \log \frac{1}{\lambda}\right) = \sqrt{n}\log(\lambda \overline{X}_n) \xrightarrow{d} N(0,1)$ ($g(x) = \log x$ y $g'(x) = 1/x, x > 0$).

Teorema 8.16. *Sean $\{Z_n : n \geq 1\}$ y Z variables aleatorias con valores enteros no negativos (por ejemplo, Poisson, geométrica, binomial, etc.) Entonces, una condición suficiente para que $Z_n \xrightarrow{d} Z$ está dada por*

$$P(Z_n = z) \overset{n\to\infty}{\longrightarrow} P(Z = z)$$

para todo $z = 0, 1, 2, \ldots$.

Ejemplo 8.20.

(a) Sea $Z_n \sim Hip(N, N_1, n)$, por ejemplo, $Z_n =$ número de elementos de tipo 1 en una muestra aleatoria de n elementos extraídos sin devolución de una población con N elementos, de los cuales N_1 son de tipo 1. La fmp de Z_n es

$$P(Z_n = z) = \frac{\binom{N_1}{z}\binom{N-N_1}{n-z}}{\binom{N}{n}}$$

Supongamos que $\frac{N_1}{N} \to p > 0$, cuando $N \to \infty$. Entonces

$$P(Z_N = z) \overset{N\to\infty}{\longrightarrow} P(Z = z) = \binom{n}{z} p^z (1-p)^{n-z}$$

con $z = 0, 1, \ldots, n$. Donde $Z \sim Bin(n, p)$.

(b) Sea $Z_n \sim Bin(n, p_n)$, donde $np_n \overset{n\to\infty}{\longrightarrow} \lambda$. Entonces

$$P(Z_n = z) = \binom{n}{z} p^z (1-p)^{n-z} \overset{n\to\infty}{\longrightarrow} \frac{e^{-\lambda}\lambda^z}{z!} = P(Z = z),\ z = 0, 1, 2, \ldots$$

donde $Z \sim P(\lambda)$.

Tarea. Pruebe los resultados de (a),(b).

8.4. Proceso de Poisson

Sea

$$X_t = \text{número de ocurrencias de un cierto fenómeno (o evento) en un intervalo } (0, t], t \geq 0.$$

Entonces, $\{X_t : t \geq 0\}$ es un proceso estocástico asumiendo valores enteros no negativos, es decir, para cada $t \geq 0$, X_t es una variable aleatoria con valores en $\mathbb{Z}^+ = \{0, 1, 2, \ldots\}$.

Definición 8.6. $\{X_t : t \geq 0\}$ *se llama proceso de Poisson si $X_0 = 0$ y se verifican las siguientes hipótesis:*

H1) Los incrementos son independientes:

$\{X_t = k_1\}$ y $\{X_{t+s} - X_s = k_2\}$ son eventos independientes para todo $k_1, k_2 \in \mathbb{Z}^+$ y para todo $t, s \geq 0$.

H2) Los incrementos son estacionarios:

$$P(X_{t+s} - X_s = k) = P(X_t = k) := p_k(t), \;\; \forall k \in \mathbb{Z}^+ \; y \, \forall t, s \geq 0$$

Es decir, la probabilidad de observar k ocurrencias en un intervalo $(s, s+t]$ sólo depende del tamaño (largo) del intervalo (y no de la posición), o sea, de t y no de s.

H3) No se producen ocurrencias simlultaneas:

$$\begin{aligned}
P(X_t \geq 2 | X_t \geq 1) &= \frac{P(X_t \geq 2)}{P(X_t \geq 1)} \\
&= \frac{1 - P(X_t = 0) - P(X_t = 1)}{1 - P(X_t = 0)} \\
&= 1 - \frac{P(X_t = 1)}{1 - P(X_t = 0)} \xrightarrow{t \to 0} 0 \\
&\Leftrightarrow \frac{P(X_t = 1)}{1 - P(X_t = 0)} = \frac{p_1(t)}{1 - p_0(t)} \xrightarrow{t \to 0} 1
\end{aligned}$$

Se puede demostrar que la solución está dada por:

$$p_k(t) = P(X_t = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \; k = 0, 1, 2 \ldots, \; t \geq 0$$

donde

$$\lambda = \lim_{t \to 0} \frac{1 - p_0(t)}{t} = \text{tasa o intensidad de ocurrrencia por unidad de tiempo}$$

Es decir, se tiene que, $X_t \sim P(\lambda t), t \geq 0$, es un proceso de Poisson con intensidad λ ($\lambda > 0$).

Sea

$$T_1 = \text{tiempo transcurrido hasta que se produce la primer ocurrencia.}$$

Note que

$$\{T_1 > t\} \equiv \{X_t = 0\} = \text{no hay ocurrencias en } (0, t], t \geq 0$$

Equivalentemente

$$\{T_1 \leq t\} \equiv \{X_t = 0\}^c, t \geq 0$$

Luego

$$\begin{aligned}
F_{T_1}(t) &= P(T_1 \leq t) \\
&= P(\{x_t = 0\}^c) \\
&= 1 - P(X_t = 0) \\
&= 1 - \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = 1 - e^{-\lambda t}, \; t \geq 0
\end{aligned}$$

Es decir

$$F_{T_1}(t) = \begin{cases} 0, & t < 0 \\ 1 - e^{-\lambda t}, & t \geq 0 \end{cases}$$

$(\lambda > 0)$ de modo que $T_1 \sim \exp(\lambda)$. Recordadno que

$$P(T_1 > t + s | T_1 > s) = \frac{P(t_1 > t + s)}{P(T_1 > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t},\ \forall t, s \geq 0$$

Se concluye que el proceso de Poisson no tiene memoria.

Tarea. Sea T_k el tiempo transcurrido hasta la k-ésima ocurrencia (de un proceso de Poisson con intensidad λ) $k = 1, \ldots, n$. Pruebe que

(a) $T_1, T_2 - T_1, \ldots, T_n - T_{n-1}$ (tiempos entre dos ocurrencias consecutivas) son variables aleatorias iid $\exp(\lambda)$.

(b) $T_n \sim Gamma(n, \lambda)$.

9. Ejercicios

P1. *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Pruebe que:*

1. *Si $A_1, \ldots, A_n$ son eventos disjuntos. Entonces*

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

2. *Para cualquier colección de eventos $A_1, \ldots, A_n$ en $\mathcal{A}$, se tiene que*

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) + \cdots + (-1)^{n+1} P(A_1A_2\ldots A_n)$$

3. *Para cualquier secuencia contable de eventos $A_1, A_2, \ldots$ en $\mathcal{A}$, se tiene que*

$$P\left(\bigcup_{n\geq 1} A_n\right) \leq \sum_{n\geq 1} P(A_n)$$

Sol. Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad.

1. Sea la colección $\{A_i\}$ infinita de eventos medibles tales que $A_i = \emptyset$ para todo $i > n$ y disjuntos. Entonces por el tercer axioma de medida de probabilidad, se tiene que

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

dado que $\sum_{i=n+1}^{\infty} P(A_i) = 0$.

2. Probaremos por inducción. Si $n = 2$ entonces dados A, B eventos medibles

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

por el teorema 2.1. Supongamos que se cumple para todo $k \leq n$, entonces para $k = n+1$ tenemos que dada una colección $\{A_i\}_{i=1}^{n+1}$ de eventos medibles se tiene que

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(\bigcup_{i=1}^{n} A_i \cup A_{n+1}\right) \\ &= P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1}) - P\left(\bigcup_{i=1}^{n} A_i \cap A_{n+1}\right) \end{aligned}$$

Podemos aplicar la hipótesis inductiva, pero tenemos otro problema, no tenemos una forma conveniente de

$$P\left(\bigcup_{i=1}^{n} A_i \cap A_{n+1}\right)$$

Para ello notemos que

$$\bigcup_{i=1}^{n} A_i \cap A_{n+1} = \bigcup_{i=1}^{n} (A_i \cap A_{n+1})$$

Y a la segunda expresión también se le puede aplicar la hipótesis inductiva. Entonces

$$P\left(\bigcup_{i=1}^{n} A_i \cap A_{n+1}\right) \leq \sum_{i=1}^{n} P(A_i A_{n+1}) - \sum_{i<j} P(A_i A_j A_{n+1}) + \sum_{i<j<k} P(A_i A_j A_k A_{n+1}) + \cdots + (-1)^{n+1} P(A_1 A_2 \ldots A_{n+1})$$

...

3. Sea la secuencia $\{A_i\}_{i\in\mathbb{N}}$ de eventos medibles. Sea $\{B_i\}$ la secuencia definida por $B_1 = A_1, B_n := A_1^C A_2^C \ldots A_n$, por el lema 1.1, los B_i son disjunto y

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$$

luego por el tercer axioma de medida de probabilidad, se tiene que

$$P\left(\bigcup_{n\geq 1} A_n\right) = P\left(\bigcup_{n\geq 1} B_n\right) = \sum_{n\geq 1} P(B_n) = \sum_{n\geq 1} P(A_1^C A_2 \ldots A_n)$$

si $B_n \subseteq A_n$, se tiene que por monotonía de las probadilidades que

$$P\left(\bigcup_{n\geq 1} A_n\right) \leq \sum_{n\geq 1} P(A_n)$$

como queriamos probar.

P2. *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Sean A, B eventos tales que*

$$P(A) = p, \;\; P(B) = q, \;\; P(A \cup B) = r$$

donde $0 < p, q, r < 1$. Entonces

$$\begin{aligned} P(A \cap B) &= p + q - r \\ P(A \setminus B) &= r - q \\ P(A^C \cap B^C) &= 1 - r \\ P(A \cup B^C) &= p - r + 1 \end{aligned}$$

Sol. Usaremos las propiedades medida de probabilidad, si $P(A\cup B) = P(A)+P(B)-P(A\cap B)$, entonces $r = p + q - P(A \cap B)$. Luego $P(A \cap B) = p + q - r$. Para la segunda notemos que $P(A \cup B) = P(B \cup (A \setminus B)) = P(B) + P(A \setminus B)$, luego $P(A \setminus B) = P(r - q)$. La tercera es ver que $P(A^C \cap B^C) = 1 - P(A \cup B) = 1 - r$. Y la última es ver que $P(A \cup B^C) = P(A) + P(B^C) - P(A \cap B^C) = p + (1 - q) - (r - q) = p - r + 1$.

P3. *Muestre los siguientes resultado:*

1. *Eventos con probabilidad* 0 *o* 1 *son independientes de cualquier otro evento.*
2. *Un eventos A es independiente de si mismo si y sólo si $P(A) = 0$ y $P(A) = 1$.*
3. *Si $AB = \emptyset$, entonces los eventos A y B no son independientes, a menos que $P(A) = 0$ o $P(B) = 0$.*
4. *Si $A_1, \ldots, A_n$ son eventos mutuamente independientes, entonces $B_1, \ldots, B_n$ también lo son, donde $B_i := A_i$ o A_i^C.*
5. *Si $A_1, \ldots, A_n$ son eventos mutuamentes independientes. Muestre que*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n}(1 - P(A_i))$$

Sol.

1. Sea A un evento de probabilidad 0, sea B cualquier evento medible, entonces $A \cap B \subseteq A$, entonces

$$P(A \cap B) = 0 = P(A)P(B)$$

Si A es un evento de probabilidad 1, entonces dado B cualquier evento medible se tiene que $A \subseteq A \cup B$, luego

$$\begin{aligned} 1 = P(A \cup B) = P(A) + P(B) - P(A \cup B) &\Leftrightarrow P(B) = P(A \cup B) \\ &\Leftrightarrow P(A)P(B) = P(A \cup B) \end{aligned}$$

2. Sea A un evento talque $P(A \cap A) = P(A)P(A)$, entonces $P(A)(1 - P(A)) = 0$, luego $P(A) = 1$ o $P(A) = 0$.
3. Supongamos que $P(A) \neq 0, P(B) \neq 0$, luego $P(A)P(B) \neq 0$, y si $P(AB) = 0$, entonces A y B no pueden ser independientes. Si $P(A) = 0$ o $P(B) = 0$ claramente son indepdenties.
4. Sean $A_1, \ldots, A_n$ mutuamente independientes
5. Sean $A_1, \ldots, A_n$ mutuamente independientes. Entonces

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - P\left(\bigcap_{i=1}^{n} A_i^C\right) = 1 - \prod_{i=1}^{n} P(A_i^C) = 1 - \prod_{i=1}^{n}(1 - P(A_i))$$

P4. *Considere dos centavos, uno donde $P(C_1) = u$ y otro donde $P(C_2) = w$. Donde se comportan de manera independiente. Defina*

$$\begin{aligned} p_0 &= P(\textit{0 caras}) \\ p_1 &= P(\textit{1 cara}) \\ p_2 &= P(\textit{2 caras}) \end{aligned}$$

¿puede tomares u y w tales que $p_0 = p_1p_2$?

Sol. Armemos el espacio de probabilidad, tomares $\Omega = \{C_1C_2, C_1S_2, S_1C_2, S_1S_2\}$ y $\mathcal{A} := \mathcal{P}(\Omega)$. Es claro que para que no hayan cara, debe salir el caso S_1S_2 y como son independientes entre si, tenemos que

$$P(\{S_1S_2\}) = (1-u)(1-w)$$

Lo mismo se deduce del resto, entonces tenemos que

$$p_0 = p_1 = p_2 \Leftrightarrow (1-u)(1-w) = w(1-u) + u(1-w) = uw$$

En otras palabras debemos resolver el sistema. **revisar Libro pag 38, 1.6**

P5. *Si $P(A) = \frac{1}{3}, P(B^C) = \frac{1}{4}$, pueden ser A, B disjuntos?*

Sol. Notemos que $P(B) = \frac{3}{4}$ y que

$$1 \geq P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{13}{12} - P(A \cap B)$$

luego

$$P(A \cap B) \geq \frac{1}{12}$$

Dicho de otra forma, A, B no pueden ser disjuntos.

P6. *un closet contiene n pares de zapatillas. Si $2r$ zapatillas son tomadas al azar, ¿cuál es la probabilidad de que no esté su par?*(**Page 40, 1.21**)

Sol....

P7. *dos personas lanzan cada una una moneda justa n veces. Encuentre la probabilidad de que saquen el mismo número de caras.*

Sol.

P8. *Sean dos sucesos equiprobables e independientes tal que la probabilidad de su unión es 0.75. Obtener la probabilidad de cada uno de ellos.*

Sol.

Sean A, B los sucesos equiprobables tales que $P(A) = P(B) = k$, si $P(A \cap B) = P(A)P(B)$, entonces

$$\frac{3}{4} = P(A \cup B) = 2k - k^2$$

Por tanto debemos resolver la ecuación.

$$k = \frac{2 - \sqrt{1}}{2} = \frac{1}{2}$$

Encontrando la probabilidad de A, B.

P9. *Se ha realizado una encuesta entre los estudiantes de grado del MIT (Massachusetts Institute of Technology) para conocer sus preferencias tecnológicas. El 35 % de los entrevistados tienen un iPhone y un IPad, el 80 % tienen al menos uno de estos dispositivos y el 60 % no tiene iPad. Se elige un estudiante al azar, calcula la probabilidad de que:*

1. *Disponga de iPhone y no de iPad*
2. *Tenga un iPad pero no un iPhone*
3. *Tenga únicamente uno de los dos dipositivos.*
4. *No disponga de ninguno de los dos dispositivos.*

Sol. Notemos que estamos en un espacio de probabilidad discreto. Con espacio muestral dado por todas las combinaciones de tener o no Iphone y/0 tener o tener un Ipad. Por lo que solo contar. Por convenio supongamos que hay 100 personas, sean los eventos

$$\begin{aligned} A &= \text{ tener un iPhone y un iPad} \\ B &= \text{ Tener un iPhone o una iPad} \\ C &= \text{ no tiener un iPad} \end{aligned}$$

Donde

$$\begin{aligned} P(A) &= \frac{35}{100} \\ P(B) &= \frac{80}{100} \\ P(C) &= \frac{60}{100} \end{aligned}$$

como estamos considerando a 100 personas, podemos ver que 20 no tienen ningún objeto tecnológico, 40 tiene solo un iPhone, 35 solo tienen iPhone y iPad y solo 5 tienen solo un Ipad. Entonces

1. $P(\text{Tener un iPhone y no un Ipad}) = \frac{40}{100}$
2. $P(\text{Tener un iPad y no un iPhone}) = \frac{5}{100}$
3. $P(\text{Tenga únicamente uno de los dos dipositivos}) = \frac{45}{100}$
4. $P(\text{No disponga de ninguno de los dos dispositivos}) = \frac{20}{100}$

P10. *Una encuesta realizada entre inmigrantes proporciona la siguiente información: el 80 % de los jóvenes entre 18 y 25 años no tiene trabajo, el 75 % de los jóvenes en esa edad no están matriculados en ningún centro educativo (no estudian). Si se toma un joven al azar de la muestra, ¿cuál es la probabilidad de que ni estudie ni trabaje? (Otra forma de hacer la misma pregunta, ¿cuál es la proporción de j´ovenes de la encuesta que ni estudian ni trabajan?*

Sol. Sean los eventos

$$A = \text{Trabaja}$$
$$B = \text{Estudia}$$

Donde

$$P(B^C) = \frac{80}{100}$$
$$P(A^C) = \frac{75}{100}$$

Queremos determinar $P(A^C \cap B^C)$, entonces tenemos que

$$P(A^C \cap B^C) = 1 - P(A \cup B) = 1 - P(A) - P(B) + P(A \cap B)$$

el problema es que no podemos determinar $P(A \cap B)$, pero si estimarlo, notemos que

$$P(A^C \cap B^C) = \frac{55}{100} + P(A \cap B)$$

y si $P(A^C \cap B^C) \leq P(B^C)$, entonces

$$\frac{55}{100} \leq P(A^C \cap B^C) \leq \frac{75}{100}$$

P11. *Demostrar que para cualquier par de sucesos, A_1, A_2 se cumple que*

$$P(A_1 \cap A_2) \geq P(A_1) + P(A_2) - 1$$

demuestre en general que

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$$

Sol. Si

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

entonces

$$P(A_1 \cap A_2) = P(A_1) + P(A_2) - P(A_1 \cup A_2) \geq P(A_1) + P(A_2) - 1$$

Probemos la generalización. Notemos que

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P\left(A_1 \cap \bigcap_{i=2}^{n} A_i\right) \geq P(A_1) + P\left(\bigcap_{i=2}^{n} A_i\right) - 1$$

Luego de forma recursiva, llegamos a que

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$$

P12. *La probabilidad de que un componente de una máquina se averíe antes de* 100 *horas es* 0,01*. La máquina tiene* 50 *componentes; calcular la probabilidad de avería de la máquina antes de* 100 *horas en los casos siguientes:*

1. *La máquina se avería cuando lo hace uno o más componentes.*
2. *La máquina se avería cuando fallan dos o más componentes.*
3. *La máquina sólo se avería cuando lo hacen todos los componentes.*

Sol.

1. Sea A el evento donde no se avería ningún componente y sea A_i el evento donde el componente i no se avería, es claro que

$$P(A_i) = \frac{99}{100}$$

ahora notemos que $A = A_1 \cap A_2 \ldots A_5 0$, como son independiente entre si, es claro que

$$P(A) = \left(\frac{99}{100}\right)^{50}$$

Entonces

$$P(A^C) = 1 - \left(\frac{99}{100}\right)^{50}$$

2. Usando el mismo argumento que en el 1, solo nos interesa que uno o ninguno falle. Notemos que el evento B que es que falle un componente es que falle el primero y el resto o falle el segundo y el resto no, y así sucesívamente, luego

$$P(B) = 50\frac{1}{100}\left(\frac{99}{100}\right)^{49}$$

Entonces, la probabilidad pedida es

$$P = 1 - \left(\left(\frac{99}{100}\right)^{50} + 50\frac{1}{100}\left(\frac{99}{100}\right)^{49}\right)$$

3. Ahora queremos que todo los componente se averién, pero es como el primer problema, solo que tomamos $1/100$, entonces

$$P = \left(\frac{1}{100}\right)^{50}$$

P. *Una urna contiene* 3 *bolas azules,* 2 *blancas y* 5 *rojas. Si se extraen* 6 *bolas con reposición, calcular la probabilidad de obtener* 2 *bolas de cada color.*

Sol. Estamos en un muestreo ordenado y con reposición. Para resolver el problema tomemos la siguiente configuración fija

$$(R_1, R_2, A_1, A_2, B_1, B_2)$$

La probabilidad de que esto ocurra es

$$\begin{aligned} P(R_1 \cap R_2 \cap A_1 \cap A_2 \cap B_1 \cap B_2) &= P(R_1)P(R_2)P(A_1)P(A_2)P(B_1)P(B_2) \\ &= \left(\frac{5}{10}\right)^2 \left(\frac{2}{10}\right)^2 \left(\frac{3}{10}\right)^2 \end{aligned}$$

Nos falta contar todas las configuraciones posibles. Claramente son 6! formas de distribuirlos, pero como son dos a dos los colores, debemos quitar casos, por ello la probabilidad pedida es

$$\frac{6!}{2!2!2!} \left(\frac{5}{10}\right)^2 \left(\frac{2}{10}\right)^2 \left(\frac{3}{10}\right)^2$$

P13. *entre estos y Dadciones posibles, clarama la variable aleatoria discreta X, cuya función de probabilidad viene definida por:*

$$P(X = x) = kx$$

para $x = 1, 2, 3, 4, 5$. Determine:

1. *El valor de la constante k.*
2. *$P(X > 2)$*

Sol. Como estamos en un espacio de probabilidad discreto $(\mathbb{R}, \mathcal{B}, P_X)$ entonces

$$1 = P_X(\mathbb{R}) = P(X \in \{1, 2, 3, 4, 5\}) = P(X = 1) + \cdots + P(X = 5)$$

Luego $k = \frac{1}{15}$. Entonces

$$P(X > 2) = P(X = 3) + P(X = 4) + P(X = 5) = \frac{4}{5}$$

P14. *Dada la variable aleatoria X, cuya función de densidad (fdp) es*

$$f_X(x) := \begin{cases} k(1 - x^2), & 0 < x < 1 \\ 0, & e.o.c \end{cases}$$

Entonces

1. *Obtenga k*

2. Calcule la probabnilidad $P(X < 0,3)$

Sol. Claramente X es continua, luego

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

Notemos que $x \to \infty$ se tiene que

$$1 = \lim_{x\to\infty} F_X(x) = \int_0^1 k(1-t^2)dt = k\left(t - \frac{t^3}{3}\bigg|_0^1\right) = \frac{2k}{3}$$

Entonces $k = \frac{3}{2}$.

Encontremos la probabilidad de $P(X < 0,3)$. Como estamos en una distribución continua, claramente

$$P(X < 0,3) = F_X(0,3) = \frac{3}{2}\left(0,3 - \frac{0,09}{2}\right) = 0,405$$

P15. *Un gran almacén guarda cajas que contienen piezas de distinto tipo. La proporción X de piezas de tipo A en una caja se puede considerar una variable aleatoria con función de densidad:*

$$f_X(x) = kx(1-x)$$

con $0 \leq x \leq 1$, 0 si $x \notin [0,1]$. Entonces, calcule la constante k y si se toman 10 cajas al azar, determine la probabilidad de que ninguna de ella contenga una proporción de piezas de tipo A igual o superior al 75 %.

Sol. Notemos que estamos en un fdp, entonces

$$F_X(x) = \int_{-\infty}^{x} kt(1-t)dt$$

Luego

$$1 = \lim_{x\to\infty} F_X(x) = \int_0^1 kt(1-t)dt = k\int_0^1 (t-t^2)dt = k\left(\frac{t^2}{2} - \frac{t^3}{3}\right)\bigg|_{t=1} = \frac{k}{6}$$

Luego claramente $k = 6$.

Nos piden determinar

$$F_X(0,75) = \frac{27}{32} = 0,1828681$$

P16. *Un modelo que habitualmente se utiliza en balística para comprobar la correcta calibración de las armas es:*

$$f_X(x) = \frac{x}{\sigma^2}\exp\left[-\frac{x^2}{2\sigma^2}\right]$$

$x \geq 0, \sigma \leq 0$. *Determine la función distribución. La probabilidad de que* $P(X \leq 5)$

Sol. Esto es un fdp, determines el fda. Entonces

$$F_X(x) = \int_0^x \frac{t}{\sigma^2} e^{-\frac{t^2}{2\sigma^2}} dt = -e^{-\frac{x^2}{2\sigma^2}} + 1 =$$

Determinado el fda, podemo calcular la probabilidad $P(X \leq 5)$, entonces

$$F_X(5) = 1 - e^{-\frac{1}{8}}$$