APUNTES PARA EL CURSO DE PROBABILIDAD

Agosto – Diciembre 2012

DAVID RUELAS RODRÍGUEZ

Estos apuntes fueron revisados y ajustados para cubrir por completo el temario del curso de **Probabilidad** del Departamento de Estadística del Instituto Tecnológico Autónomo de México (ITAM) para el semestre Agosto – Diciembre 2012.

El documento no es un libro de texto, sino una síntesis de las principales definiciones, teorema, conceptos y métodos de la Probabilidad. Sin embargo, esta compilación puede tenerse como referencia en el estudio de otras materias que dependen fundamentalmente de la Probabilidad; por ejemplo, Estadística, Econometría, Procesos Estocásticos y Finanzas.

Cada sección fue escrita con la finalidad de revisar rápidamente la teoría y dedicar la mayor parte del tiempo de clase a: (i) la demostración y análisis de los principales resultados; y (ii) a la resolución de ejercicios que ejemplifican la aplicación de la Probabilidad. Los apuntes de la Sección 1 sobre Fundamentos de Probabilidad son más largos que los demás pues incluyen detalles sobre la Teoría de Conjuntos y teoremas sobre sumas y series que son útiles en el cálculo de probabilidades y el estudio de las variables aleatorias.

Al inicio del documento se incorporó un índice para facilitar la búsqueda de algún tema en particular. Al final del documento se presenta como anexo el *Material de Apoyo para el Curso de Probabilidad* repartido durante las clases de este semestre, incluyendo las preguntas de los 148 ejercicios revisados en clase.

Esta es la cuarta versión de estos apuntes completos. Esta versión es prácticamente igual a la del semestre Agosto – Diciembre 2011, pero se corrigieron algunas referencias a las tareas correspondientes al semestre Agosto – Diciembre 2012. Sé que todavía pueden existir errores y que algunos textos pueden ser mejorados, así que cualquier comentario es bienvenido al correo electrónico davidruelas@hotmail.com.

Agradezco la participación de mis alumnos de este semestre que con sus preguntas y comentarios contribuyen a la mejora de estos apuntes.

David Ruelas Rodríguez

Diciembre 2012

Índice

1. Fundamentos de Probabilidad

1.1. Fenómenos aleatorios e incertidumbre

Figura 1.1

FENÓMENOS ALEATORIOS

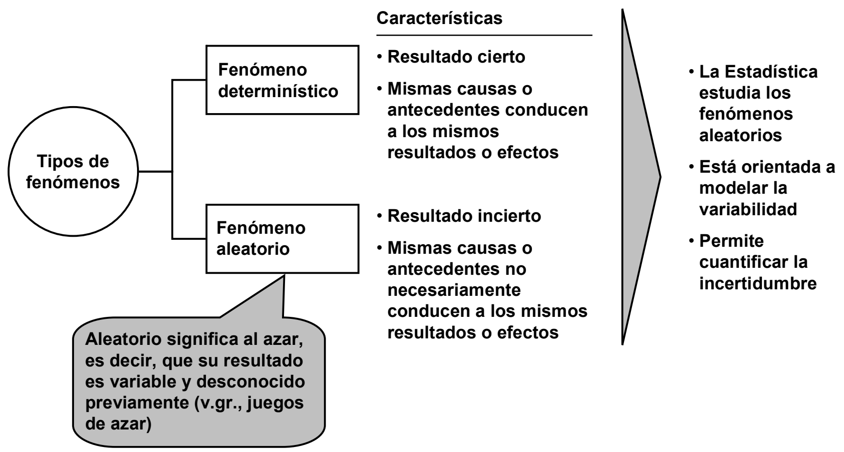


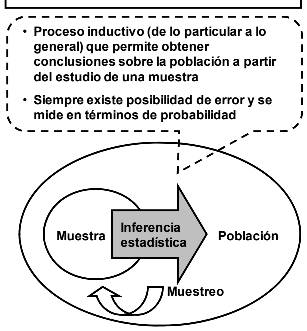
Figura 1.2

PROBABILIDAD, INFERENCIA ESTADÍSTICA Y ECONOMETRÍA

Probabilidad

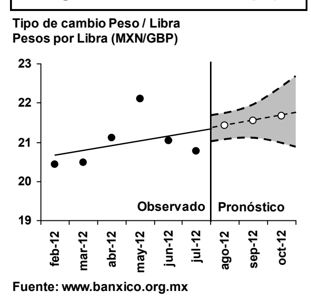
- Si compro un boleto de "Melate", ¿qué tan factible es que gane el premio principal?, ¿es más fácil ganar el "Melate" o la "Lotería"?
- Si entro en una rifa, ¿es importante saber cuántos boletos están a la venta?
- Si hoy llueve, ¿qué tan seguro es que mañana también llueva?
- Si el Índice de Precios y Cotizaciones (IPC) cierra hoy a la alza, ¿es fácil que también mañana lo haga?
- ¿Cuánta mercancía debo tener en el almacén para evitar demanda insatisfecha en mi empresa?

Inferencia Estadística (Estadística Matemática, Muestreo)



- ¿De qué tamaño debe ser una muestra para determinar el porcentaje de la población que está a favor del Partido A?
- ¿Cómo se estima el ingreso promedio de la población?

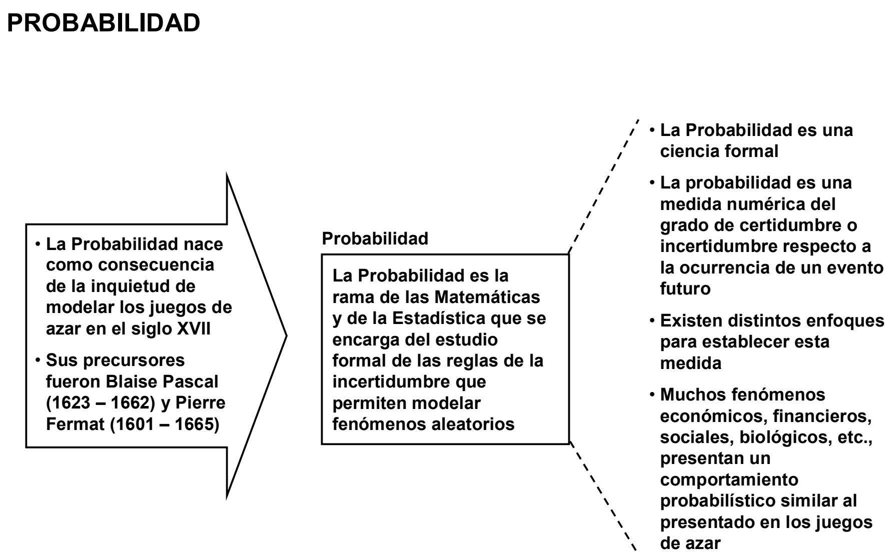
Econometría (Regresión, Pronóstico de Negocios, Series de Tiempo)



- ¿Cuál es el pronóstico para el tipo de cambio Peso / Libra? ¿Qué intervalos garantizan cierto nivel de confianza?
- Con base en un conjunto de precios y cantidades observadas, ¿cuál es la función de demanda estimada?

La Estadística es la rama de las Matemáticas que permite modelar la variabilidad y cuantificar la incertidumbre inherente a los fenómenos aleatorios

Figura 1.3



1.2. Espacios muestrales y eventos

1.2.1. Nociones sobre Teoría de Conjuntos

DEFINICIONES Y REPRESENTACIÓN DE CONJUNTOS

Def. Conjunto. Es una colección o agrupación de elementos.

Notación de conjuntos. Usualmente los conjuntos se denotan mediante letras mayúsculas y sus elementos mediante letras minúsculas. Por ejemplo: $A = \{a_1, a_2, a_3\}$.

Pertenencia. Un elemento x puede:

- **Pertenecer** a un conjunto, denotado por $x \in A$; o
- **No pertenecer** a un conjunto, denotado por $x \notin A$.

Def. Conjunto Universo (Ω). Es el conjunto que contiene a todos los elementos.

Def. Conjunto Vacío (∅). Es el conjunto que no contiene elementos.

Def. Subconjunto. Es un conjunto cuyos elementos corresponden a los de otro conjunto. $B \subset A$ denota que B es un subconjunto de A. Formalmente la **relación de contención** se establece de la siguiente manera:

$$B \subset A \Leftrightarrow x \in B \Rightarrow x \in A$$

Por ejemplo, si C es el conjunto de las letras del alfabeto y D es el conjunto de las vocales, entonces $D \subset C$, es decir, todos los elementos de D están contenidos en C.

Def. Cardinalidad. Es el número de elementos que integran un conjunto o la medida de un conjunto. Si A es un conjunto, $\#A$ o $|A|$ denotan su cardinalidad. Por ejemplo, si C es el conjunto de las letras del alfabeto (en español y sin letras dobles) y D es el conjunto de las vocales, entonces $\#C = 27$ y $\#D = 5$.

Proposición

i) Si $B \subset A$ entonces $\#B \leq \#A$

ii) $\#\varnothing = 0$

Representación de conjuntos. Los conjuntos pueden expresarse **analíticamente**:

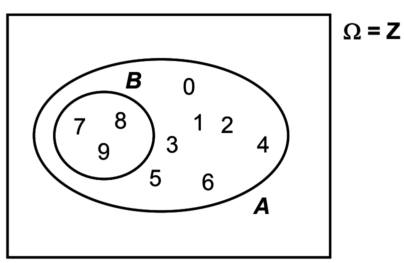
- **Por extensión**, si se escriben cada uno de los elementos del conjunto.
- **Por comprensión**, si se establece una regla que permite identificar a todos los elementos del conjunto.

La representación **gráfica** de los conjuntos puede hacerse mediante **Diagramas de Venn** o Diagramas de Venn-Euler. Por ejemplo, si Ω es el conjunto de los números enteros ($\Omega = \mathbf{Z}$), A el conjunto de los dígitos y B el conjunto de los dígitos mayores a 6:

Tabla 1.1

Conjunto	Descripción	Representación por extensión	Representación por comprensión
A	Dígitos	$A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$	$A = \{x \in \Omega \mid 0 \leq x \leq 9\}$
B	Dígitos mayores a 6	$B = \{7, 8, 9\}$	$B = \{x \in A \mid x > 6\}$

Figura 1.4



Observe cómo $\#A = 10$ y $\#B = 3$. Además, como $B \subset A$, entonces $\#B \leq \#A$.

CONJUNTOS NUMÉRICOS

Def. Números Naturales (N).

N = {1, 2, 3,...}

- Los números naturales se utilizan para contar o enumerar.
- La cardinalidad de **N** es **infinito numerable**.

Def. Números Enteros (Z).

Z = {..., –3, –2, –1, 0, 1, 2, 3,...}

- Los números enteros incluyen a los números naturales (**N** ⊂ **Z**), a sus negativos y al cero.
- La cardinalidad de **Z** también es **infinito numerable** (#**N** = #**Z**) ya que es posible establecer una relación uno a uno entre los elementos de **Z** y los de **N**, como se muestra a continuación:

Figura 1.5

$$\begin{array}{cccccccccc} \mathbf{Z} = \{ & 0, & 1, & -1, & 2, & -2, & 3, & -3, & \ldots \} \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ \mathbf{N} = \{ & 1, & 2, & 3, & 4, & 5, & 6, & 7, & \ldots \} \end{array}$$

Def. Números Racionales (Q).

$$\mathbf{Q} = \left\{ \frac{m}{n} : m, n \in \mathbf{Z}, n \neq 0 \right\}$$

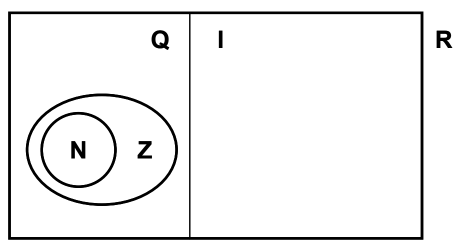
- Los números racionales son todas las fracciones con expansión decimal finita o infinita periódica. Por ejemplo, $\frac{20}{11} = 1.818181... = 1.\overline{81} \in \mathbf{Q}$.
- Los racionales incluyen a los enteros y a los naturales (**N** ⊂ **Z** ⊂ **Q**).
- Sorprendentemente, la cardinalidad de **Q** también es **infinito numerable**. El matemático ruso Georg Cantor (1845 – 1918) ideó la forma de establecer una relación uno a uno entre los elementos de **Q** y los de **N**, como se muestra en la siguiente figura. Los números con un recuadro punteado no hay que enumerarlos pues están repetidos. Para incluir al cero y los negativos simplemente hay que iniciar la enumeración por el cero y luego contar doble cada número (el positivo y el negativo).

Figura 1.6

n \ m	1	2	3	4	5	. . .
1	$\frac{1}{1}$	$\frac{2}{1}$	$\frac{3}{1}$	$\frac{4}{1}$	$\frac{5}{1}$	. . .
2	$\frac{1}{2}$	$\boxed{\frac{2}{2}}$	$\frac{3}{2}$	$\boxed{\frac{4}{2}}$	$\frac{5}{2}$	. . .
3	$\frac{1}{3}$	$\frac{2}{3}$	$\boxed{\frac{3}{3}}$	$\frac{4}{3}$	$\frac{5}{3}$	. . .
4	$\frac{1}{4}$	$\boxed{\frac{2}{4}}$	$\frac{3}{4}$	$\frac{4}{4}$	$\frac{5}{4}$	. . .
5	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{5}{5}$	. . .
⋮	⋮	⋮	⋮	⋮	⋮	

Si a los números racionales se agregan los **Números Irracionales (I)**, es decir, aquellos cuya expansión decimal es infinita y no periódica (v.gr., π, e, $\sqrt{2}$), se obtienen los **Números Reales (R)**.

Figura 1.7



No es posible establecer una relación uno a uno entre los elementos de **N** y los de **I** o los de **R** (*Teorema de Cantor*), de modo que la cardinalidad de **I** y de **R** es **infinito no numerable**, es decir, un infinito mucho mayor que el infinito numerable (#**R** >> #**N**).

Desde el punto de vista matemático, los Números Reales forman un campo ordenado completo con características técnicas muy sofisticadas[1]. Desde el punto de vista probabilístico basta considerar de manera inicial que sobre **R** es posible:

- Hablar de límites y **continuidad** (en el sentido del Cálculo Diferencial e Integral); y
- Definir **medidas** (v.gr., tiempo, longitud, peso, probabilidad).

---

[1] Por ser un **campo**, **R** es un conjunto con dos operaciones (adición y multiplicación), que satisfacen axiomas de asociatividad, conmutatividad, elemento neutro, elemento inverso y distributividad de la multiplicación sobre la adición. Como **R** es **campo ordenado**, es posible establecer una relación de orden (<) y cuenta con un valor absoluto (| · |). Adicionalmente, como **R** es un **campo completo**, satisface el postulado Arquimedeano (todos sus elementos están acotados inferiormente por el recíproco de un número natural) y el axioma del supremo (cada subconjunto cuenta con una mínima cota superior).

OPERACIONES CON CONJUNTOS

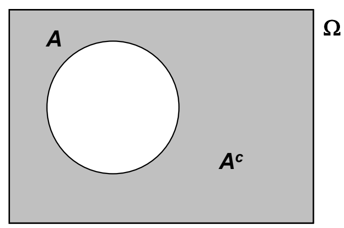
Def. Igualdad de conjuntos. Dos conjuntos son iguales si contienen los mismos elementos. Formalmente:

$A = B \Leftrightarrow A \subset B$ y $B \subset A$

Def. Conjunto Complemento. El conjunto complemento de A, denotado por A^c, es el conjunto de los elementos de Ω que no están en A. Formalmente:

$x \in A^c \Leftrightarrow x \notin A$

Figura 1.8



Teorema

i) $\left(A^c\right)^c = A$

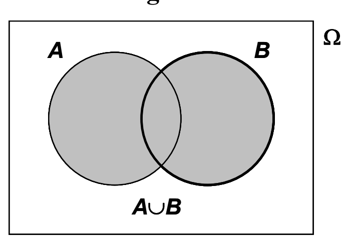
ii) $\Omega^c = \varnothing$

Es importante señalar que los **Diagramas de Venn** son un buen punto de partida para demostrar resultados de la Teoría de Conjuntos, sin embargo, **no constituyen demostraciones formales**.

Def. Unión de Conjuntos ($\cup$). La unión de los conjuntos A y B, denotada por $A\cup B$, es el conjunto formado por todos los elementos de A más todos los elementos de B. Formalmente:

$x \in A\cup B \Leftrightarrow x \in A$ ó $x \in B$

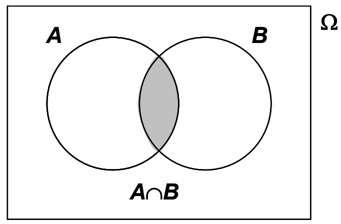
Figura 1.9



Def. Intersección de Conjuntos ($\cap$). La intersección de los conjuntos A y B, denotada por $A \cap B$ o AB, es el conjunto formado por todos los elementos comunes de A y B. Formalmente:

$$x \in A \cap B \Leftrightarrow x \in A \text{ y } x \in B$$

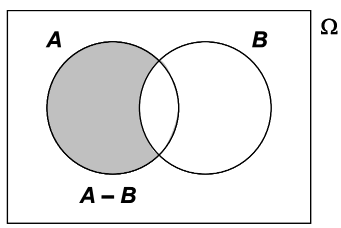
Figura 1.10



Def. Diferencia de 2 conjuntos. La diferencia de los conjuntos A menos B, denotada por $A - B$, es el conjunto formado por los elementos de A que no están en B. Formalmente:

$$x \in A - B \Leftrightarrow x \in A \text{ y } x \notin B$$

Figura 1.11



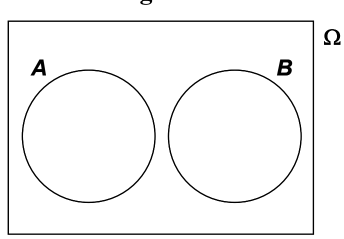
Teorema.

$$A - B = A \cap B^c$$

Def. Conjuntos Mutuamente Excluyentes (m.e.). Se dice que A y B son mutuamente excluyentes si no tienen elementos comunes. Formalmente:

$$A \text{ y } B \text{ son mutuamente excluyentes} \Leftrightarrow A \cap B = \varnothing$$

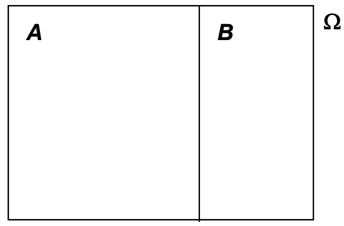
Figura 1.12



Def. Conjuntos Exhaustivos. Se dice que A y B son conjuntos exhaustivos si sus elementos abarcan por completo el conjunto universo. Formalmente:

A y B son exhaustivos $\Leftrightarrow A \cup B = \Omega$

Figura 1.13



Teorema.

i) A y A^c son mutuamente excluyentes, i.e., $A \cap A^c = \varnothing$

ii) A y A^c son exhaustivos, i.e., $A \cup A^c = \Omega$

Ejercicios $\boxed{\text{E1}}$ y $\boxed{\text{E2}}$.

PROPIEDADES DE LAS OPERACIONES CON CONJUNTOS

Propiedades de conmutatividad, asociatividad y distributividad de $\cup$ y $\cap$.

i) $A \cup B = B \cup A$ (conmutatividad de la unión)

ii) $A \cap B = B \cap A$ (conmutatividad de la intersección)

iii) $A \cup (B \cup C) = (A \cup B) \cup C$ (asociatividad de la unión)

iv) $A \cap (B \cap C) = (A \cap B) \cap C$ (asociatividad de la intersección)

v) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (distributividad de la unión sobre la intersección)

vi) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (distributividad de la intersección sobre la unión)

Propiedades de la unión e intersección con Ω y $\varnothing$.

i) $A \cup \Omega = \Omega$

ii) $A \cap \Omega = A$

iii) $A \cup \varnothing = A$

iv) $A \cap \varnothing = \varnothing$

Propiedades de la unión e intersección con subconjuntos.

i) $B \subset A \Rightarrow A \cup B = A$

ii) $B \subset A \Rightarrow A \cap B = B$

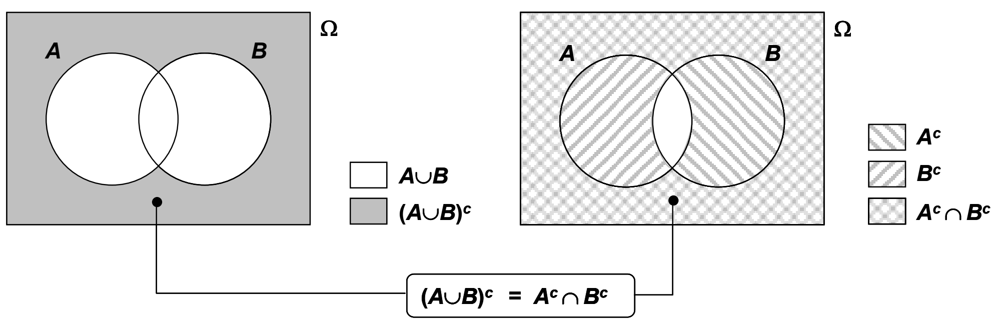
Leyes de De Morgan. Son propiedades de la unión e intersección de complementos, propuestas por el matemático inglés August De Morgan (1806 – 1871).

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

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

Gráficamente, la primera Ley de De Morgan se puede verificar como se muestra a continuación. Note que estos gráficos no constituyen una demostración formal.

Figura 1.14



La unión e intersección de conjuntos se puede generalizar. Si $A_1, A_2, \ldots, A_n$, $n \in \mathbf{N}$, son n conjuntos (pudiendo incluso ser $n = \infty$), entonces:

- $A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i$ denota la unión de todos los conjuntos; y
- $A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} A_i$ denota la intersección de todos los conjuntos.

Leyes de De Morgan (para *n* conjuntos).

i) $\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c$

ii) $\left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c$

Ejercicios $\boxed{\text{E3}}$, $\boxed{\text{E4}}$, $\boxed{\text{E5}}$ y $\boxed{\text{E6}}$.

1.2.2. Espacios muestrales y eventos.

Def. Experimento Aleatorio (EA). Es un proceso mediante el cual se obtiene una observación o un dato, y cuyo resultado no puede predecirse antes de su realización y, por lo tanto, está sujeto al azar.

Por ejemplo, son experimentos aleatorios:

- El lanzamiento de un volado.
- El giro de una ruleta.
- El desarrollo de una partida de Bingo, Yak o Keno.
- El número de palabras contenidas en 5 páginas elegidas al azar en un diccionario.

Def. Espacio Muestral (Ω). Es el conjunto integrado por todos los posibles resultados de un experimento aleatorio o de un fenómeno aleatorio. A sus **elementos** se les denota por ω_1, ω_2,…, y se denominan **puntos del espacio muestral**.

Def. Evento. Es un subconjunto del espacio muestral. Se dice que un evento **ocurre** si al realizar el experimento aleatorio se observa cualquiera de sus elementos.

Por ejemplo, si se considera el experimento aleatorio en el que se lanza un dado honesto con caras numeradas del 1 al 6, entonces el espacio muestral es

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

Si además se definen

A = Evento en que el resultado es par; y
B = Evento en que el resultado es 6,

entonces

$$A = \{2, 4, 6\} \subset \Omega$$
$$B = \{6\} \subset \Omega.$$

Def. Espacio de Eventos ($\mathscr{A}$). Es la clase de todos los eventos asociados a un experimento aleatorio.

Matemáticamente, el espacio de eventos es un **sigma álgebra** (σ-algebra) o *Campo de Borel*, y cumple con las siguientes propiedades.

Propiedades del espacio de eventos.

i) $\varnothing \in \mathscr{A}$, es decir, siempre incluye al **vacío** (*evento imposible*);

ii) $A \in \mathscr{A} \Rightarrow A^c \in \mathscr{A}$, es decir, los **complementos** están incluidos; y

iii) $A_1, A_2, \ldots \in \mathscr{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathscr{A}$, es decir, las **uniones numerables** están incluidas.

De *i*) y *ii*) se infiere que el propio **espacio muestral** Ω esta incluido en el espacio de eventos $\mathscr{A}$ (*evento seguro*), y aplicando las Leyes de De Morgan con las propiedades *ii*) y *iii*) se puede concluir que también las **intersecciones numerables** están incluidas en $\mathscr{A}$.

Por ejemplo, si una urna contiene 3 bolas numeradas del 1 al 3 y se considera el experimento aleatorio en que se elige una bola al azar y se registra su numero, entonces

$$\Omega = \{1, 2, 3\},$$

y $\mathscr{A}$ incluye:

$\varnothing$	$\{1\}$	$\{1, 2\}$	$\{1, 2, 3\} = \Omega$
	$\{2\}$	$\{1, 3\}$	
	$\{3\}$	$\{2, 3\}$	

$\mathscr{A}$ es el **conjunto potencia** de Ω, es decir, el conjunto que incluye a todos los posibles subconjuntos de Ω.

De acuerdo con la cardinalidad de Ω, los espacios muestrales se clasifican en discretos o continuos:

- **Espacio muestral discreto**: Ω finito o infinito numerable; y
- **Espacio muestral continuo**: Ω infinito no numerable.

Ejercicios $\boxed{\text{E7}}$, $\boxed{\text{E8}}$ y $\boxed{\text{E9}}$.

1.3. Enfoques de la probabilidad

Concepto de probabilidad.

La probabilidad es una **medida** numérica del grado de certidumbre o incertidumbre respecto a la ocurrencia de un evento definido dentro de un espacio muestral.

Existen distintos **enfoques** para establecer esta medida:

- Probabilidad Clásica.
- Probabilidad Frecuentista.
- Probabilidad Subjetiva.

1.3.1. Probabilidad Clásica

Figura 1.15

ENFOQUE CLÁSICO DE LA PROBABILIDAD

Características

- No requiere del desarrollo de ningún experimento aleatorio
- Se define con base en el razonamiento lógico, antes de efectuar el experimento
- Se utiliza para calcular probabilidades asociadas a espacios equiprobables, es decir, aquellos en los que cada punto del espacio muestral tiene la misma posibilidad de ser observado

Definición de probabilidad

Bajo el enfoque clásico, la probabilidad del evento A se define como:

$$P(A) = \frac{\text{Casos favorables}}{\text{Casos totales}} = \frac{\#A}{\#\Omega}$$

$\#A$ y $\#\Omega$ pueden ser incluso medidas, por ejemplo, longitud en R, área en R^2, etc.

Observación:

- Como $A \subset \Omega$, la cardinalidad del evento A es un entero positivo menor o igual a la cardinalidad del espacio muestral: $0 \leq \#A \leq \#\Omega$
- Dividiendo entre $\#\Omega > 0$: $\frac{0}{\#\Omega} \leq \frac{\#A}{\#\Omega} \leq \frac{\#\Omega}{\#\Omega}$
- Y por definición de probabilidad bajo el enfoque clásico: $\boxed{0 \leq P(A) \leq 1}$

Ejercicios $\boxed{\text{E10}}$ y $\boxed{\text{E11}}$.

1.3.2. Probabilidad Frecuentista

Figura 1.16

ENFOQUE FRECUENTISTA DE LA PROBABILIDAD

Características

- Bajo el enfoque frecuentista la probabilidad es la frecuencia relativa límite de un evento
- Es adecuada para modelar fenómenos en los que existe regularidad estadística, es decir, que los resultados tienen la misma posibilidad de ser observados al repetir sucesivamente un experimento bajo las mismas condiciones
- El cálculo de probabilidades bajo este enfoque requiere que el experimento aleatorio pueda repetirse en condiciones similares

Definición de probabilidad

Si un experimento se repite n veces y el evento A ocurre $n(A)$ veces, entonces la probabilidad del evento A bajo el enfoque frecuentista se define como:

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

Observación:

- $\frac{n(A)}{n}$ es la frecuencia relativa de la n-ésima repetición
- $n(A)$ cuenta el número de veces que ocurre A al realizar n repeticiones, entonces: $0 \leq n(A) \leq n$
- Dividiendo entre n y tomando el límite cuando $n \to \infty$: $\lim_{n \to \infty} \frac{0}{n} \leq \lim_{n \to \infty} \frac{n(A)}{n} \leq \lim_{n \to \infty} \frac{n}{n}$
- Y por definición de probabilidad bajo el enfoque frecuentista: $\boxed{0 \leq P(A) \leq 1}$

Ejercicio $\boxed{\text{E12}}$.

1.3.3. Probabilidad Subjetiva.

Figura 1.17

ENFOQUE SUBJETIVO DE LA PROBABILIDAD

Características de la probabilidad bajo el enfoque subjetivo

- **La probabilidad del evento *A* se basa en el grado de credibilidad sobre la ocurrencia de dicho evento y de forma que $0 \leq P(A) \leq 1$**
- **Refleja el sentimiento de una persona respecto a la confianza sobre la veracidad de una proposición (subjetiva)**
- **Se asigna a eventos que ocurren una sola vez (v.gr., éxito en la perforación de un pozo petrolero, sismos de gran intensidad, huracanes)**
- **Se relaciona con la Estadística Bayesiana**
- **Por ejemplo: Un ingeniero geofísico considera que la probabilidad de que durante el próximo año ocurra un sismo de más de 8 grados Richter es 0.0001**

1.4. Desarrollo axiomático de la probabilidad.

1.4.1. Técnicas de conteo y cálculo de probabilidades.

La **Combinatoria** o **Cálculo Combinatorio** es el conjunto de técnicas matemáticas que permiten contar de manera eficiente.

Def. Muestreo: Es el procedimiento para elegir los elementos de una muestra a partir de una población.

Hay dos **tipos de muestreo**:

- **Sin reemplazo** o **sin repetición**: Se elige el elemento, se observan sus características y se aparta de la población. Cada elemento puede ser elegido a lo más una vez.
- **Con reemplazo** o **con repetición**: Se elige el elemento, se observan sus características y se incorpora nuevamente a la población. Cada elemento puede ser elegido más de una vez.

TÉCNICAS DE CONTEO

Diagramas de árbol: Gráfico que muestra todos los posibles resultados en un experimento aleatorio y que facilita el conteo de los elementos del espacio muestral y de los eventos.

Ejercicio $\boxed{\text{E13}}$.

Regla de la Multiplicación (Conteo)

Si un experimento puede descomponerse en r partes y la i-ésima parte se puede hacer de n_i formas, $i = 1, 2,\ldots, r$, entonces el experimento se puede realizar de $n_1 n_2 \cdots n_r$ formas.

Ejercicio $\boxed{\text{E14}}$.

Regla de la Suma (Conteo)

En ocasiones es necesario descomponer el experimento en partes que no pueden ocurrir simultáneamente (m.e.); en este caso sus posibilidades deben sumarse.

Ejercicios $\boxed{\text{E15}}$ y $\boxed{\text{E16}}$.

Def. Factoriales

Si $n \in \mathbf{Z}^+$, entonces $\underbrace{n! = n(n-1)!}_{\text{Relación de recurrencia}}$, donde $\underbrace{0! = 1}_{\text{Condición inicial}}$

Como el factorial de n depende del factorial de $n - 1$, se dice que los factoriales se definen a través de un *relación de recurrencia* o *relación recursiva*. Este tipo de relaciones requieren de una condición inicial (0! = 1 en el caso de los factoriales).

Observación: $n = 4 \;\Rightarrow\; 4!$

$$\begin{aligned} 4! &= 4 \cdot 3! \\ &= 4 \cdot 3 \cdot 2! \\ &= 4 \cdot 3 \cdot 2 \cdot 1! \\ &= 4 \cdot 3 \cdot 2 \cdot 1 \cdot 0! \text{ , y como } 0! = 1 \\ &= 1 \cdot 2 \cdot 3 \cdot 4 \end{aligned}$$

Entonces, alternativamente los factoriales se pueden definir de la siguiente manera:

Si $n \in \mathbf{Z}^+$, entonces $n! = 1 \cdot 2 \cdot 3 \cdots (n-1)(n)$

Comúnmente los factoriales se pueden obtener fácilmente en las calculadoras a través del botón **x!** .

EXCEL. *Factoriales*. La función =FACT(n) calcula n!

Ejercicios E17 y E18 .

Def. Permutaciones (ordenaciones sin repetición)

Si $n, r \in \mathbf{Z}^+$, y $r \le n$, entonces $n\,P\,r \equiv P(n,r) = \dfrac{n!}{(n-r)!}$

Las **permutaciones de *n* en *r***, $P(n, r)$, indican el **número de arreglos ordenados** de tamaño r que se puden formar a partir de n objetos distintos elegidos **sin remplazo** (sin que los objetos se puedan repetir en un mismo arreglo).

Por ejemplo, si se quiere determinar el número de formas posibles de elegir el cuadro de honor (primero, segundo y tercer lugares) en un salón de 20 alumnos, ese número es:

Cuadro de honor:	1º	2º	3º		
Posibilidades:	(20)	(19)	(18)	=	6,840

O bien: $$P(20,3) = \frac{20!}{(20-3)!} = \frac{20!}{17!} = \frac{20 \cdot 19 \cdot 18 \cdot 17!}{17!} = (20)(19)(18) = 6{,}840\,.$$

Comúnmente las permutaciones se pueden obtener fácilmente en las calculadoras a través de la tecla **nPr** .

EXCEL. *Permutaciones*. La función =PERMUTACIONES(n, r) calcula $P(n, r)$.

Note cómo las permutaciones de n en n es simplemente $n!$:

$P(n,n) = \dfrac{n!}{(n-n)!} = \dfrac{n!}{0!} = n!$

Ejercicio E19 .

Def. Combinaciones

Si n, $r \in \mathbf{Z}^+$, y $r \leq n$, entonces $n\,C\,r \equiv C(n,r) \equiv \binom{n}{r} = \dfrac{n!}{r!(n-r)!}$

Las **combinaciones de *n* en *r***, $\binom{n}{r}$, es el **número de subconjuntos (sin orden y sin repetición)** de tamaño r que se pueden formar a partir de un conjunto de n objetos distintos.

Por ejemplo, si Ana (A), Beatriz (B), Carlos (C) y Daniel (D) integran un equipo de su clase de Probabilidad y dos de ellos deben exponer los resultados de un trabajo, ¿de cuántas formas se puede elegir a los integrantes de la pareja que expondrá?

Este problema se traduce en contar cuántos subconjuntos de tamaño 2 se pueden generar a partir de un conjunto de tamaño 4. Este número es $\binom{4}{2} = \dfrac{4!}{2!\cdot(4-2)!} = \dfrac{4\cdot 3\cdot 2!}{2\cdot 1\cdot 2!} = 6$.

Si el equipo de tamaño 4 es: {A, B, C, D}

Las 6 posibles parejas son: {A, B}, {A, C}, {A, D},
{B, C}, {B, D}, y
{C, D}

Si en este ejemplo sólo uno de los integrantes del equipo debiera exponer los resultados del trabajo es evidente que existen sólo $\binom{4}{1} = \dfrac{4!}{1!\cdot(4-1)!} = \dfrac{4\cdot 3!}{3!} = 4$ posibilidades: {A}, {B}, {C} o {D}.

Comúnmente las combinaciones se pueden obtener fácilmente en las calculadoras a través de la tecla **_nCr_** .

EXCEL. *Combinaciones*. La función =COMBINAT(n, r) calcula $\binom{n}{r}$.

Proposición

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

ii) $\binom{n}{0} = \binom{n}{n} = 1$

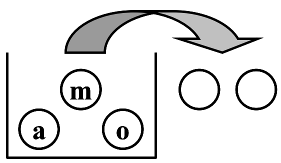
Observación: $\binom{n}{0} = \dfrac{n!}{0!(n-0)!} = \dfrac{n!}{n!} = 1$ es decir, sólo hay un subconjunto de tamaño 0 (el vacío, $\varnothing$).

$\binom{n}{n} = \dfrac{n!}{n!(n-n)!} = \dfrac{n!}{n!\cdot 0!} = 1$ es decir, sólo hay un subconjunto de tamaño n (el universo, Ω).

Diferencia entre las Combinaciones y las Permutaciones.

Para ilustrar la diferencia entre las combinaciones y la permutaciones, suponga que una urna contiene 3 pelotas con las letras “a”, “m” y “o”, y considere el experimento aleatorio en que extraen de la urna 2 pelotas al azar y sin reemplazo. Las permutaciones y las combinaciones asociadas a este experimento aleatorio se muestran a continuación.

Figura 1.18



Extracción de 2 pelotas
sin reemplazo

Permutaciones (con orden):

$$\left.\begin{array}{c} (\,a\,,\,m\,) \\ (\,a\,,\,o\,) \\ (\,m\,,\,a\,) \\ (\,m\,,\,o\,) \\ (\,o\,,\,a\,) \\ (\,o\,,\,m\,) \end{array}\right\} P(3,2) = \frac{3!}{(3-2)!} = \frac{3!}{1!} = \frac{3\cdot 2\cdot 1!}{1!} = 6 \text{ permutaciones}$$

Combinaciones (sin orden):

$$\left.\begin{array}{c} \{\,a\,,\,m\,\} \\ \{\,a\,,\,o\,\} \\ \{\,m\,,\,o\,\} \end{array}\right\} \binom{3}{2} = \frac{3!}{2!(3-2)!} = \frac{3!}{2!\cdot 1!} = \frac{3\cdot 2!}{2!} = 3 \text{ combinaciones}$$

Note cómo si $r \geq 1$, entonces:

$$r! \geq 1 \quad \Rightarrow \quad \frac{1}{r!} \leq 1 \quad \Rightarrow \quad \binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{1}{r!}\left[\frac{n!}{(n-r)!}\right] = \frac{1}{r!}P(n,r) \leq P(n,r)$$

Es decir, en general, las permutaciones son mayores a las combinaciones, y la relación entre ambos conceptos es la siguiente:

$$\boxed{\binom{n}{r} = \frac{1}{r!}P(n,r) \leq P(n,r)}$$

Ejercicios $\boxed{\text{E20}}$ y $\boxed{\text{E21}}$.

Sumas y Productos

Notación: Si $x_1, x_2, \ldots, x_n$ representan n números reales, entonces:

- $x_1 + x_2 + \cdots + x_n = \sum_{i=1}^{n} x_i$ denota su **suma**; y
- $x_1 x_2 \cdots x_n = \prod_{i=1}^{n} x_i$ denota su **producto**.

Los operadores $\sum$ y $\prod$ se denominan **sumatoria** y **productoria**, respectivamente.

Los valores por debajo y por arriba de la sumatoria determinan el inicio y el fin de la suma y se denominan **índice** o **contador**. $\sum_i x_i = \sum x_i$ significa que la suma se debe realizar sobre todos los posibles valores del índice i.

Ejercicio $\boxed{\text{E22}}$. *Solución*: a) 20; b) 325; c) 791.028; d) 1,847.1; e) 7,052.5; f) 39.0625

Importante

Si $x_1, x_2, \ldots, x_n$ y $y_1, y_2, \ldots, y_n$ son conjuntos de números reales, en general, se puede afirmar que:

i) $\sum_i x_i y_i \neq \sum_i x_i \sum_i y_i$

ii) $\sum_i x_i^2 \neq \left(\sum_i x_i \right)^2$

Propiedades de la suma

Si $x_1, x_2, \ldots, x_n$ y $y_1, y_2, \ldots, y_n$ son conjuntos de números reales y c una constante, entonces:

i) $\sum_{i=1}^{n} c = nc$

ii) $\sum_{i=1}^{n} c x_i = c \sum_{i=1}^{n} x_i$

iii) $\sum_{i=1}^{n} (x_i \pm y_i) = \sum_{i=1}^{n} x_i \pm \sum_{i=1}^{n} y_i$

Ejercicios $\boxed{\text{E23}}$ y $\boxed{\text{E24}}$.

SUMAS ÚTILES EN PROBABILIDAD

Sumas de naturales

i) $\sum_{i=1}^{n} i = 1 + 2 + \cdots + n = \frac{n(n+1)}{2}$ (Suma Gaussiana)

ii) $\sum_{i=1}^{n} i^2 = 1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$ (Suma de cuadrados)

Estas fórmulas se pueden demostrar por **inducción matemática** sobre *n*, que es el número de sumandos de cada suma. La Suma Gaussiana recibe su nombre en honor al matemático alemán Kart Friedrich Gauss (1777 – 1855).

Series Geométricas

i) $\sum_{k=0}^{\infty} r^k = 1 + r + r^2 + \cdots = \frac{1}{1-r}$ si $|r| < 1$ (Serie geométrica)

ii) $\sum_{k=0}^{n} x^k = 1 + x + x^2 + \cdots + x^n = \frac{1 - x^{n+1}}{1-x}$ si $x \neq 1$ (Serie geométrica truncada)

La serie geométrica truncada se puede demostrar fácilmente multiplicando $1 - x$ por $1 + x + x^2 + \cdots + x^n$. La serie geométrica es la expansión de la Serie de Taylor de la función $f(x) = \frac{1}{1-x}$ para $|x| < 1$, evaluada en *r* y con centro en cero.

Serie de Taylor

Si *f* es una función de clase C^{∞} (es decir, que tiene infinitas derivadas continuas) y *a* una constante, entonces $f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)(x-a)^k}{k!}$ es su Serie de Taylor con centro en *a*.

Si $a = 0$, la Serie de Taylor se denomina **Serie de Maclaurin**.

Corolario. Serie de Maclaurin de e^x. (Serie de Taylor de $f(x) = e^x$ con $a = 0$)

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

Note cómo $e = e^1 = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots \approx 2.718281...$

Ejercicio $\boxed{\text{E25}}$.

Teorema Binomial o Binomio de Newton

Si $x, y \in \mathbf{R}$ entonces para $n = 1, 2, \ldots$

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

La demostración de este teorema se puede realizar mediante argumentos combinatorios. El Binomio de Newton permite calcular **productos notables**:

Binomio al cuadrado (n = 2). Cuadrado del primero más el doble del primero por el segundo, más cuadrado del segundo.

$$\begin{aligned}(x+y)^2 &= \sum_{k=0}^{2} \binom{2}{k} x^k y^{2-k} = \binom{2}{0} x^0 y^{2-0} + \binom{2}{1} x^1 y^{2-1} + \binom{2}{2} x^2 y^{2-2} = y^2 + 2xy + x^2 \\ &= x^2 + 2xy + y^2\end{aligned}$$

Binomio al cubo (n = 3). Cubo del primero, más triplo del primero al cuadrado por el segundo, más triplo del primero por el segundo al cuadrado, más cubo del último.

$$\begin{aligned}(x+y)^3 &= \sum_{k=0}^{3} \binom{3}{k} x^k y^{3-k} = \binom{3}{0} x^0 y^{3-0} + \binom{3}{1} x^1 y^{3-1} + \binom{3}{2} x^2 y^{3-2} + \binom{3}{3} x^3 y^{3-3} \\ &= y^3 + 3xy^2 + 3x^2 y + x^3 = x^3 + 3x^2 y + 3xy^2 + y^3\end{aligned}$$

Una forma fácil de calcular los coeficientes binomiales que aparecen en el Binomio de Newton es a través del **Triángulo de Pascal**, como se muestra a continuación.

Figura 1.19

	Triángulo de Pascal	Coeficientes Binomiales	Suma de Coeficientes Binomiales
$n = 0$	1	$\binom{0}{0}$	$1 = 2^0$
$n = 1$	1 1	$\binom{1}{0}\ \binom{1}{1}$	$2 = 2^1$
$n = 2$	1 2 1	$\binom{2}{0}\ \binom{2}{1}\ \binom{2}{2}$	$4 = 2^2$
$n = 3$	1 3 3 1	$\binom{3}{0}\ \binom{3}{1}\ \binom{3}{2}\ \binom{3}{3}$	$8 = 2^3$
$n = 4$	1 4 6 4 1	$\binom{4}{0}\ \binom{4}{1}\ \binom{4}{2}\ \binom{4}{3}\ \binom{4}{4}$	$16 = 2^4$
$n = 5$	1 5 10 10 5 1	$\binom{5}{0}\ \binom{5}{1}\ \binom{5}{2}\ \binom{5}{3}\ \binom{5}{4}\ \binom{5}{5}$	$32 = 2^5$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Corolario

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

Demostración. El Teorema Binomial, $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$, se establece para todo $x, y \in \mathbf{R}$. Tomando en particular $x = y = 1$ se obtiene $(1+1)^n = \sum_{k=0}^{n} \binom{n}{k} (1)^k (1)^{n-k}$, y por lo tanto $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

Ejercicio $\boxed{\text{E26}}$.

El corolario anterior permite demostrar el siguiente teorema, que establece que si el espacio muestral Ω es finito y tiene n elementos, entonces el espacio de eventos $\mathscr{A}$ tiene 2^n elementos.

Teorema

$$\#\Omega = n \geq 1 \Rightarrow \#\mathscr{A} = 2^n$$

La generalización del Coeficiente Binomial se denomina **Coeficiente Multinomial** y permite contar el número de posibles divisiones de n objetos en r distintos grupos de tamaños $n_1, n_2, \ldots, n_r$.

Def. Coeficiente Multinomial

Si $n_1 + n_2 + \cdots + n_r = n$, $n_i \geq 0$, $r \geq 1$, entonces $\binom{n}{n_1, n_2, \ldots, n_r} = \dfrac{n!}{n_1! \cdot n_2! \cdots n_r!}$

El Coeficiente Multinomial se define de esta manera ya que para dividir los n objetos en los r grupos distintos se puede aplicar la regla de la multiplicación (de conteo) procediendo de la siguiente manera:

- Tomar el número de formas de elegir los n_1 elementos del grupo 1 a partir de los n elementos totales, $\binom{n}{n_1}$.
- Multiplicar el resultado anterior por el número de formas de elegir a los n_2 elementos del grupo 2 a partir de los $n - n_1$ elementos restantes, $\binom{n - n_1}{n_2}$.
- Continuar multiplicando por factores de la forma $\binom{n - n_1 - \cdots - n_{i-1}}{n_i}$ hasta $i = r$, como se muestra a continuación.

$$
\begin{aligned}
\binom{n}{n_1, n_2, \ldots, n_r} &= \binom{n}{n_1}\binom{n - n_1}{n_2}\binom{n - n_1 - n_2}{n_3}\cdots\binom{n - n_1 - \cdots - n_{r-1}}{n_r} \\
&= \frac{n!}{n_1! \cdot (n - n_1)!} \cdot \frac{(n - n_1)!}{n_2! \cdot (n - n_1 - n_2)!} \cdot \frac{(n - n_1 - n_2)!}{n_3! \cdot (n - n_1 - n_2 - n_3)!} \cdots \frac{(n - n_1 - \cdots - n_{r-1})!}{n_r! \cdot (n - \underbrace{n_1 - n_2 - \cdots - n_r}_{})!} \\
&\qquad \text{donde } n_1 + n_2 + \cdots + n_r = n \\
&= \frac{n!}{n_1! \cdot n_2! \cdot n_3! \cdots n_r! \cdot 0!} = \frac{n!}{n_1! \cdot n_2! \cdots n_r!}
\end{aligned}
$$

Ejercicio $\boxed{\text{E27}}$.

1.4.2. Axiomas de Probabilidad.

Axiomas de Probabilidad (función de probabilidad)

$P : \mathscr{A} \to [0, 1]$ se denomina función de probabilidad y satisface los siguientes axiomas:

i) $P(A) \geq 0$ para todo $A \in \mathscr{A}$

ii) $P(\Omega) = 1$

iii) Si $A_1, A_2, \ldots,$ es una sucesión de eventos m.e. en $\mathscr{A}$ ($A_i \cap A_j = \varnothing$ para toda $i \neq j$), entonces $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

El axioma *iii*) se denomina **aditividad infinita** y se puede demostrar que también es válido para un número finito de eventos.

Teorema. Aditividad finita

Si $A_1, A_2, \ldots, A_n$, $n \geq 1$, son eventos m.e. en $\mathscr{A}$, entonces $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$

Def. Espacio de Probabilidad

Es la tripleta $(\Omega, \mathscr{A}, P)$

El espacio de probabilidad es un concepto sintético que asume la existencia de un espacio muestral Ω, un espacio de eventos $\mathscr{A}$, y una función de probabilidad P.

A partir de los axiomas de probabilidad es posible demostrar las siguientes propiedades que satisface cualquier función de probabilidad P. Estas propiedades son de gran utilidad en el cálculo de probabilidades.

Teorema. Propiedades de la probabilidad

Si $(\Omega, \mathscr{A}, P)$ es un espacio de probabilidad con A y B eventos en $\mathscr{A}$, entonces:

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

ii) $P(\varnothing) = 0$

iii) Si $B \subset A$ entonces $P(A - B) = P(A) - P(B)$

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

Ejercicio $\boxed{\text{E28}}$.

La propiedad *i*) es de gran utilidad en el cálculo de probabilidad, pues en muchos casos es más fácil calcular la probabilidad del evento complemento (v.gr., el problema del cumpleaños).

En relación a la propiedad *ii*), es **importante** resaltar lo siguiente:

$P(A) = 0$ no implica $A = \varnothing$

Contraejemplo: Considere el experimento aleatorio en que se elige al azar un número real. En este caso $\Omega = \mathbf{R}$ de modo que $\#\Omega = \#\mathbf{R}$ es infinito no numerable. Sean $x^* \in \mathbf{R}$ un número fijo y A el evento en que el número elegido es x^*, de modo que $\#A = 1$. Entonces, bajo el enfoque clásico de la probabilidad, $P(A) = \dfrac{\# A}{\#\Omega} = 0$ (constante entre infinito), sin embargo $A = \{x^*\} \neq \varnothing$.

En relación a la propiedad *iii*) es de gran utilidad el siguiente corolario, que permite calcular la **probabilidad de la diferencia** de dos eventos arbitrarios.

Corolario

$P(A - B) = P(A) - P(A \cap B)$ para todo A y B eventos en $\mathscr{A}$

Demostración. Observe cómo para todo A y B eventos en $\mathscr{A}$, $A - B = A - (A \cap B)$ y además $A \cap B \subset A$, entonces considerando la propiedad *iii*) del teorema anterior, tenemos que $P(A - B) = P(A - (A \cap B)) = P(A) - P(A \cap B)$.

Por último, en relación a la propiedad *iv*), la fórmula para el cálculo de la **probabilidad de la unión** de conjuntos $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ se puede generalizar a través de la **fórmula de inclusión-exlusión**.

Teorema. Fórmula de inclusión-exclusión.

Si $A_1, A_2, \ldots, A_n$ son eventos en $\mathscr{A}$, entonces:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \underbrace{\sum_{i<j}\sum P(A_i A_j)}_{\text{Intersecciones binarias}} + \underbrace{\sum_{i<j<k}\sum\sum P(A_i A_j A_k)}_{\text{Intersecciones de tres eventos}} \cdots + \underbrace{(-1)^{n+1} P(A_1 A_2 \cdots A_n)}_{\text{Intersección de todos los eventos}}$$

La demostración de este resultado se puede realizar:

- Por inducción sobre el número de eventos n; o
- Mediante argumentos combinatorios, mostrando cómo cada lado de la igualdad incluye los mismos puntos del espacio muestral.

En particular, considerando la fórmula de inclusión-exclusión para n = 3 se obtiene la siguiente expresión:

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1A_2) - P(A_1A_3) - P(A_2A_3) + P(A_1A_2A_3)$$

Este resultado se puede demostrar a partir de la propiedad *iv*) y mediante la asociatividad de conjuntos.

Ejercicios E29 , E30 y E31 .

1.5. Probabilidad condicional e independencia

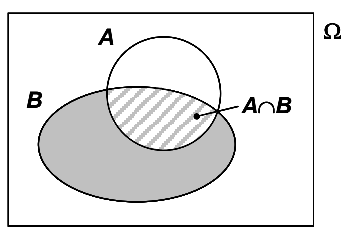
1.5.1. Probabilidad Condicional

Def. Probabilidad Condicional

Si A, $B \in \mathscr{A}$, la probabilidad condicional de A dado B, denotada por $P(A|B)$, se define por $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ si $P(B) > 0$; y no se define si $P(B) = 0$.

Si $P(B) > 0$ (o bien $P(B) \neq 0$), la probabilidad condicional $P(A|B)$ considera que el evento B ya ocurrió, de modo que ahora $P(B)$ representa el 100% (como si se redefiniera el espacio muestral) y el cálculo de la probabilidad de A queda restringido a $P(A \cap B)$ como se muestra en la siguiente figura.

Figura 1.20



La ocurrencia del evento B puede afectar las circunstancias del fenómeno aleatorio y, posiblemente, afectar la probabilidad del evento A. El evento A se denomina **evento de interés** y al evento B se le llama **evento condicionante**.

Def. Eventos Favorables y Eventos Desfavorables

i) Si $P(A|B) > P(A) \Rightarrow B$ es favorable a A.

ii) Si $P(A|B) < P(A) \Rightarrow B$ es desfavorable a A.

Es decir, un evento condicionante es favorable (o desfavorable) al evento de interés si aumenta (o disminuye) su probabilidad de ocurrencia. Si B no es favorable ni desfavorable a A entonces A y B son *eventos independientes*. Por su importancia, este último concepto se estudiará con detalle más adelante.

A partir de las definiciones anteriores se puede demostrar que si el evento B es favorable al evento A, entonces también A es favorable a B. Y en forma análoga, si B es desfavorable a A, entonces también A es desfavorable a B.

Proposición

i) $P(A|B) > P(A) \Rightarrow P(B|A) > P(B)$

ii) $P(A|B) < P(A) \Rightarrow P(B|A) < P(B)$

Ejercicios $\boxed{\text{E32}}$ y $\boxed{\text{E33}}$.

Teorema. Axiomas de Probabilidad de la Probabilidad Condicional

Si $(\Omega, \mathscr{A}, P)$ es un espacio de probabilidad con $P(B) > 0$ para $B \in \mathscr{A}$ entonces $P(\cdot\,|B)$ satisface los mismos axiomas que $P(\cdot)$, es decir:

i) $P(A|B) \geq 0$ para todo $A \in \mathscr{A}$

ii) $P(\Omega|B) = 1$

iii) Si $A_1, A_2, \ldots$, es una sucesión de eventos m.e. en $\mathscr{A}$ ($A_i \cap A_j = \varnothing$ para toda $i \neq j$), entonces $P\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) = \sum_{i=1}^{\infty} P(A_i|B)$

Como la Probabilidad Condicional satisface los Axiomas de Probabilidad, la Probabilidad Condicional posee las mismas propiedades que la Probabilidad No Condicional o Probabilidad Incondicional.

Proposición. Propiedades de la Probabilidad Condicional

Si $(\Omega, \mathscr{A}, P)$ es un espacio de probabilidad con A, B y C eventos en $\mathscr{A}$, y $P(C) > 0$, entonces:

i) $P(A^c|C) = 1 - P(A|C)$

ii) $P(\varnothing|C) = 0$

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

PROBABILIDAD CONJUNTA Y PROBABILIDAD MARGINAL

Def. Distribución de frecuencias absolutas. Es una tabla que resume los conteos de la ocurrencia de uno o varios eventos.

Con base en el *enfoque frecuentista* de la Probabilidad, es posible calcular probabilidades a partir de la distribución de frecuencias absolutas dividiendo entre el número total de observaciones. A estas **frecuencias relativas** también se les denomina **probabilidades empíricas**.

Al considerar **distribuciones de frecuencias relativas** (probabilidades empíricas) que involucran más de un evento se denomina

- **Probabilidades conjuntas**, a las probabilidades de las intersecciones; y
- **Probabilidades marginales** (ubicadas en los márgenes), a las probabilidades que involucran sólo a un evento, libre de los demás eventos.

Ejercicio $\boxed{\text{E34}}$.

REGLA DE LA MULTIPLICACIÓN

Proposición. Regla de la Multiplicación (Probabilidad)

Si $P(B) > 0$, $P(A \cap B) = P(B)\, P(A|B)$

La Regla de la Multiplicación es útil para calcular **probabilidades de intersecciones**.

Demostración. Por definición de probabilidad condicional, $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ ya que $P(B) > 0$. Entonces despejando $P(A \cap B)$ se obtiene que $P(A \cap B) = P(B)\, P(A|B)$.

La Regla de la Multiplicación se puede **generalizar** a n eventos como se muestra en el siguiente Teorema.

Teorema. Regla de la Multiplicación (caso generalizado)

Si $A_1, A_2, \ldots, A_n \in \mathscr{A}$, $n \geq 2$, y $P(A_1 A_2 \cdots A_j) > 0$ para $j = 1, 2, \ldots, n-1$, entonces:

$$P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2|A_1) P(A_3|A_1 A_2) \cdots P(A_n|A_1 A_2 \cdots A_{n-1})$$

Diagramas de árbol con probabilidades y Regla de la Multiplicación.

En ocasiones es útil para el cálculo de probabilidades considerar **diagramas de árbol** agregando probabilidades en cada una de sus ramas:

- Las probabilidades de las **primeras ramas** son probabilidades incondicionales.
- Las probabilidades de la **ramas internas** son condicionales ya que consideran lo ocurrido previamente.

El **caso generalizado de la Regla de la Multiplicación** muestra cómo para calcular la probabilidad de toda una rama de un diagrama de árbol (probabilidad de una intersección) basta **multiplicar** todas las probabilidades involucradas en esa rama.

Ejercicios E35 y E36 .

1.5.2. Independencia estadística

Def. Eventos Independientes (ind.)

Sean A, $B \in \mathscr{A}$. Se dice que A y B son eventos independientes si cualquiera de las siguientes condiciones se cumple:

i) $P(A|B) = P(A)$, si $P(B) > 0$

ii) $P(B|A) = P(B)$, si $P(A) > 0$

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

De acuerdo con la definición anterior, A y B son eventos independientes **si la ocurrencia de uno no influye en la probabilidad del otro**. Esto es evidente en las condiciones *i*) y *ii*). La condición *iii*) es equivalente a las dos anteriores. A continuación se demuestra cómo *i*) implica *iii*).

Demostración. De la Regla de la Multiplicación sabemos que $P(A \cap B) = P(B)\, P(A|B)$ pero como A y B son independientes $P(A|B) = P(A)$, y entonces $P(A \cap B) = P(A)\, P(B)$.

Intuitivamente es posible hacer hipótesis acerca de la independencia de eventos, sin embargo, para **demostrar la independencia de eventos** es necesario verificar que se cumpla cualquiera de las condiciones de la definición. Frecuentemente la independencia se verifica mediante la definición *iii*) pues establece que “la probabilidad de la intersección es igual al producto de las probabilidades” o que “la probabilidad conjunta es igual al producto de las probabilidades marginales”.

Se dice que A y B son **eventos dependientes** si no son independientes.

Ejercicios E37 y E38 .

Importante. Eventos independientes y eventos mutuamente excluyentes son conceptos muy distintos; sin embargo, se pueden relacionar a través de la siguiente proposición.

Proposición

Si A y B son eventos mutuamente excluyentes tales que $P(A) > 0$ y $P(B) > 0$ entonces no pueden ser independientes.

Note cómo el simple hecho de saber que A y B son m.e. hace que sepamos que si ocurre B entonces no puede ocurrir A y viceversa; entonces, A y B no son independientes.

Ejercicio $\boxed{\text{E39}}$.

Teorema

Si A y B son eventos independientes, entonces también lo son:

i) A y B^c

ii) A^c y B

iii) A^c y B^c

Def. Eventos Completamente Independientes

Se dice que $A_1, A_2, \ldots, A_n \in \mathscr{A}$., son eventos completamente independientes sí y sólo sí:

$P(A_i \cap A_j) = P(A_i)P(A_j)$ para toda $i \neq j$ (independencia dos a dos)

$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k)$ para toda $i \neq j$, $i \neq k$ y $j \neq k$ (ind. tres a tres)

$$\vdots$$

$$P\left(\bigcap_{i=i}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

Importante. Independencia dos a dos no implica independencia completa.

Ejercicios $\boxed{\text{E40}}$ y $\boxed{\text{E41}}$.

1.5.3. Regla de Probabilidades Totales y Teorema de Bayes

Def. Partición

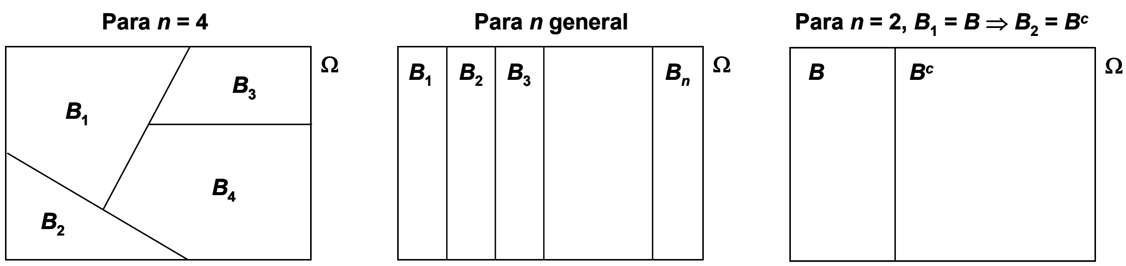
Se dice que $B_1, B_2, \ldots, B_n$ forman una partición de Ω si:

i) $B_i \cap B_j = \varnothing$ para toda $i \neq j$ (eventos m.e.); y

ii) $\bigcup_{i=1}^{n} B_i = \Omega$ (eventos exhaustivos)

Gráficamente una partición se puede representar mediante Diagramas de Venn como se muestra a continuación.

Figura 1.21



Proposición

B y B^c forman una partición de Ω

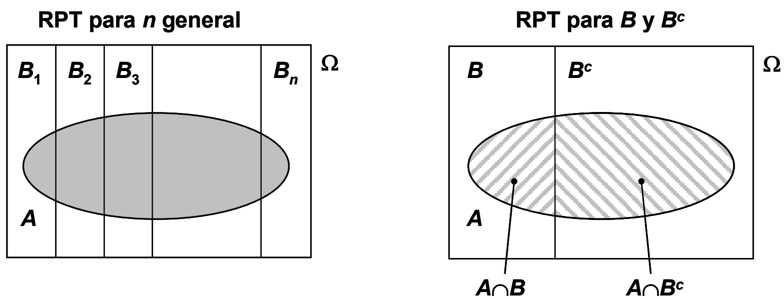
Proposición. (ver Figura 1.22)

Si $B_1, B_2, \ldots, B_n \in \mathscr{A}$ forman una partición de Ω y A es un evento inmerso en la partición, entonces $B_1 \cap A, B_2 \cap A, \ldots, B_n \cap A$ son eventos m.e.

Teorema. Regla de Probabilidades Totales (RPT)

Si $B_1, B_2, \ldots, B_n \in \mathscr{A}$ forman una partición de Ω y $P(B_i) > 0$, $i = 1, 2, \ldots, n$; entonces para $A \in \mathscr{A}$, $P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$

Figura 1.22



La RPT permite calcular la probabilidad de un evento A inmerso en una partición B_1, $B_2, \ldots, B_n$, condicionando sobre cada elemento de la partición.

Corolario. RPT para B y B^c.

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

Demostración. Como B y B^c forman una partición de Ω, $A = (A \cap B) \cup (A \cap B^c)$, donde $A \cap B$ y $A \cap B^c$ son m.e., entonces por aditividad finita $P(A) = P(A \cap B) + P(A \cap B^c)$. Finalmente, aplicando la Regla de la Multiplicación para la calcular probabilidad de la intersección en cada sumando se obtiene $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$.

La demostración de la RPT para el caso generalizado se puede realizar siguiendo este mismo razonamiento.

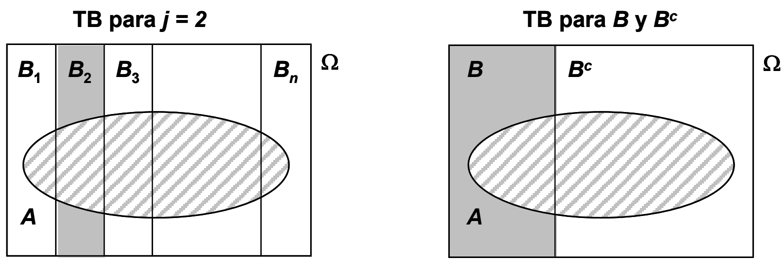
Al considerar **diagramas de árbol** con probabilidades en sus ramas, la RPT es el fundamento probabilístico que permite calcular la probabilidad de un evento de interés **sumando** los productos de las probabilidades de las ramas que conducen a la ocurrencia del evento.

Ejercicios E42 y E43 .

Teorema de Bayes (TB)

Si $B_1, B_2, \ldots, B_n \in \mathscr{A}$ forman una partición de Ω y $P(B_i) > 0$, $i = 1, 2, \ldots, n$; entonces para $A \in \mathscr{A}$ con $P(A) > 0$, $P(B_j|A) = \dfrac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$ para $j = 1, 2, \ldots, n$.

Figura 1.23



A diferencia de la RPT, el Teorema de Bayes permite calcular la probabilidad condicional de uno de los eventos de la partición condicionando en el evento inmerso en la partición.

Note cómo en el Teorema de Bayes:

- El denominador es la RPT; y
- El numerador siempre aparece como un sumando del denominador.

Corolario. Teorema de Bayes para B y B^c.

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} \text{ si } P(A) > 0$$

Demostración. Por definición de probabilidad condicional $P(B|A) = \dfrac{P(B \cap A)}{P(A)}$ para $P(A) > 0$. Aplicando conmutatividad y Regla de la Multiplicación en el numerador se obtiene $P(B \cap A) = P(A \cap B) = P(A|B)\ P(B)$. Por su parte, como B y B^c forman una partición de Ω en el denominador se obtiene $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$ al aplicar la RPT. Finalmente sustituyendo numerador y denominador se obtiene el resultado deseado, es decir, $P(B|A) = \dfrac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$.

Ejercicios $\boxed{\text{E44}}$, $\boxed{\text{E45}}$ y $\boxed{\text{E46}}$.

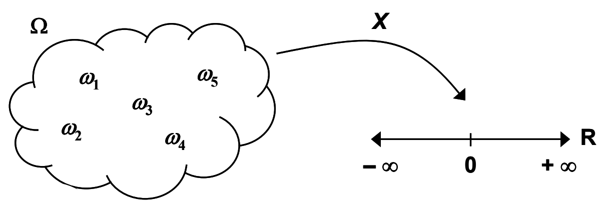
2. Variables Aleatorias

2.1. Definición y propiedades de variables aleatorias discretas y absolutamente continuas.

Def. Variable Aleatoria (v.a.)

Se dice que X es una variable aleatoria si X: $\Omega \to \mathbf{R}$, es decir, una función cuyo dominio es el espacio muestral y su contradominio la recta real.

Figura 2.1



La imagen o rango de la función X se denomina **soporte de la variable aleatoria *X***.

Ejercicio $\boxed{\text{E47}}$.

Clasificación de las variables aleatorias. Con base en su soporte, las variables aleatorias se clasifican en *discretas* y *continuas*.

- **Variables aleatorias discretas.** Se asocian a un espacio muestral discreto, su soporte es finito o infinito numerable y provienen de un proceso de conteo.
- **Variables aleatorias continuas.** Se asocian a un espacio muestral continuo, su soporte es infinito no numerable y provienen de un proceso de medición.

Son ejemplos de variables aleatorias discretas: el número de éxitos al realizar varias repeticiones de un experimento, el número de autobuses que llegan a una central, el número de llamadas telefónicas que recibe una operadora, etc.

Son ejemplos de variables aleatorias continuas: la distancia diaria que recorre un coche elegido al azar, la temperatura promedio de un día, el volumen de lluvia que cae durante una tormenta, la duración de una llamada telefónica, el monto que debe pagar diariamente una aseguradora a sus asegurados, etc.

Ejercicio $\boxed{\text{E48}}$.

2.2. Funciones de masa, densidad y distribución de probabilidad.

2.2.1. Función de masa de probabilidad (f.m.p.)

Def. Función de masa de probabilidad (f.m.p.)

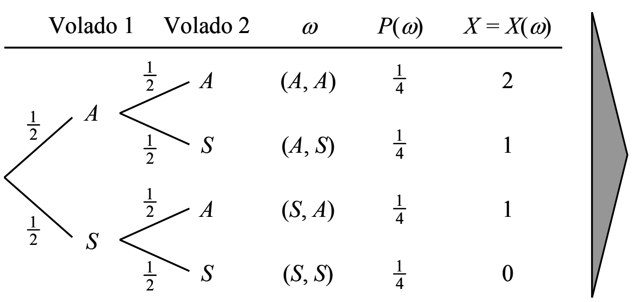
Si X es variable aleatoria discreta con soporte $\{x_1, x_2, \ldots\}$, se dice que la función $f_X(\cdot)$, $f_X : \mathbf{R} \to [0, 1]$, es la función de masa de probabilidad de X si:

i) $f_X(x) = \begin{cases} P[X = x] > 0 & si \;\; x = x_1, x_2, \ldots \\ 0 & en\ otro\ caso\ (e.o.c.) \end{cases}$; y

ii) $\sum_x f_X(x) = 1$

Por ejemplo, si se lanzan 2 volados con una moneda honesta y X es el número de águilas, entonces $X \in \{0, 1, 2\}$, el espacio muestral asociado a este experimento aleatorio se muestra en la figura 2.2.

Figura 2.2



$$f_X(0) = P[X = 0] = P((S, S)) = \frac{1}{4}$$

$$f_X(1) = P[X = 1] = P((A, S) \cup (S, A)) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

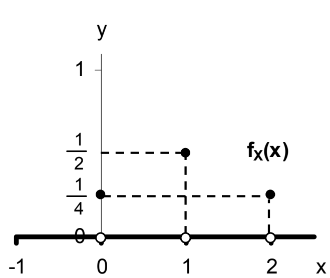
$$f_X(2) = P[X = 2] = P((A, A)) = \frac{1}{4}$$

Consecuentemente, la función de masa de probabilidad de la variable aleatoria X es

$$f_X(x) = \begin{cases} \frac{1}{4} & si \;\; x = 0, 2 \\ \frac{1}{2} & si \;\; x = 1 \\ 0 & e.o.c. \end{cases}$$

y su gráfica se muestra a continuación.

Figura 2.3



Al conjunto de probabilidades asociadas a la variable aleatoria X se le llama **distribución de probabilidades** o **distribución de la variable aleatoria *X***.

Una forma común de representar la *distribución de probabilidades de una variable aleatoria discreta* es a través de la **representación tabular de su f.m.p.**, es decir, una tabla que muestra las parejas $(x, f_X(x))$ para cada valor del soporte de la variable aleatoria X. Note cómo la representación tabular omite el "cero en otro caso".

Importante. Para el *caso discreto* el contradominio de la f.m.p. es [0, 1], ya que $f_X(x)$ es una probabilidad.

Proposición. Cálculo de probabilidades vía f.m.p.

Si X es variable aleatoria discreta y $A \subset \mathbf{R}$, entonces $P[X \in A] = \sum_{x \in A} f_X(x)$

Corolario. En particular:

Si X es variable aleatoria discreta $P[a \leq X \leq b] = \sum_{\{x: a \leq x \leq b\}} f_X(x)$

Ejercicio $\boxed{\text{E49}}$.

Def. Función Indicadora

La función indicadora del conjunto A se define por $I_A(x) = \begin{cases} 1 & si \ \ x \in A \\ 0 & si \ \ x \notin A \end{cases}$

Por ejemplo, si A es el conjunto de los dígitos, entonces $I_A(x) = \begin{cases} 1 & si \ \ x \in \{0, 1, \ldots, 9\} \\ 0 & e.o.c. \end{cases}$, de modo que $I_A(5) = 1$ pero $I_A(10) = 0$, $I_A(1.5) = 0$ y $I_A(-1) = 0$.

Una *notación alternativa* de las funciones indicadoras es $\mathbf{1}(x \in A) = I_A(x)$.

Las funciones indicadoras ayudan a simplificar la notación de las funciones utilizadas en Probabilidad ya que los distintos casos convierten en sumandos y el "cero en otro caso" queda implícito. Por ejemplo, considerando nuevamente el lanzamiento de 2 volados con una moneda honesta, si X es el número de águilas obtenidas, entonces:

$$f_X(x) = \begin{cases} \frac{1}{4} & si \ \ x = 0, 2 \\ \frac{1}{2} & si \ \ x = 1 \\ 0 & e.o.c. \end{cases} \quad \text{se expresaría como} \quad f_X(x) = \frac{1}{4} I_{\{0,2\}}(x) + \frac{1}{2} I_{\{1\}}(x).$$

Propiedades de la Función Indicadora

i) $I_A(x) = 1 - I_{A^c}(x)$

ii) $I_{A_1 \cap A_2 \cap \cdots \cap A_n}(x) = I_{A_1}(x) I_{A_2}(x) \cdots I_{A_n}(x)$

iii) $I_{A_1 \cup A_2 \cup \cdots \cup A_n}(x) = \max\{I_{A_1}(x), I_{A_2}(x), \cdots, I_{A_n}(x)\}$

iv) $I_A^2(x) = I_A(x)$ (idempotencia)

Ejercicios $\boxed{\text{E50}}$, $\boxed{\text{E51}}$ y $\boxed{\text{E52}}$.

Def. Modelos Paramétricos

Un modelo paramétrico es una función de masa de probabilidad que involucra constantes llamadas parámetros y cuyos valores determinan por completo el comportamiento probabilístico de un fenómeno aleatorio (distribución de probabilidad).

Al conjunto de valores posibles de los parámetros se le denomina **espacio paramétrico**.

Por ejemplo, un modelo paramétrico utilizado frecuentemente en Probabilidad es la **Distribución Geométrica**. Se dice que X tiene una Distribución Geométrica si

$$\underset{\substack{\uparrow \\ \text{Parámetro}}}{X \sim Ge(p)} \quad \Rightarrow \quad f_X(x) = p(1-p)^{x-1} I_{\{1,2,\ldots\}}(x), \quad \underbrace{0 < p < 1}_{\substack{\text{Espacio} \\ \text{paramétrico}}}$$

Aplicaciones. Si un experimento aleatorio (i) puede ser éxito con probabilidad p o fracaso con probabilidad $1 - p$, (ii) el resultado de cada experimento es independiente de los demás; y (iii) X es el número de experimentos necesarios hasta obtener el primer éxito, entonces X tiene una Distribución Geométrica con parámetro p. Este modelo paramétrico puede ser utilizado para modelar el número de lanzamientos de un dado necesarios hasta obtener un 6 (en este caso $p = \frac{1}{6}$), el número de hijos que tendrá una pareja que decide tener hijos hasta tener una niña (en este caso el parámetro p es la probabilidad de tener una niña en cierta población), etc.

Ejercicio $\boxed{\text{E53}}$.

2.2.2. Función de densidad de probabilidad (f.d.p.)

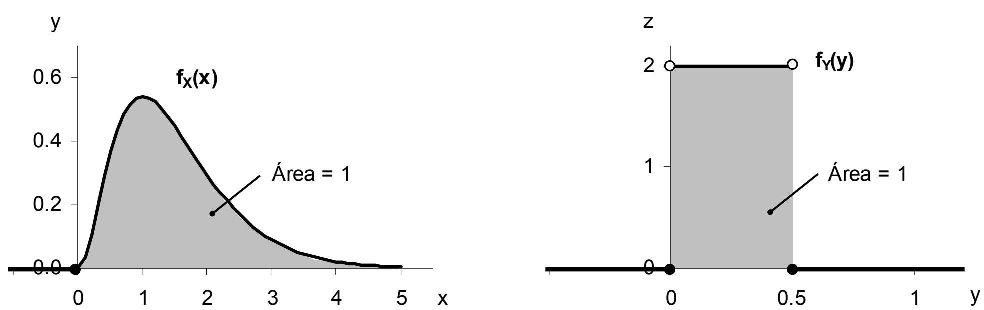
Def. Función de densidad de probabilidad (f.d.p.)

Si X es variable aleatoria continua, se dice que la función $f_X(\cdot)$, $f_X : \mathbf{R} \to [0, \infty)$, es la función de densidad de probabilidad de X si:

i) $f_X(x) \geq 0$ para toda $x \in \mathbf{R}$; y

ii) $\int_{-\infty}^{\infty} f_X(x)dx = 1$

Figura 2.4



La principal diferencia entre las variables aleatorias discretas y continuas es que en el caso discreto se consideran sumas de masas de probabilidad y en el caso continuo se consideran **áreas bajo la curva**, ya que el soporte de las variables aleatorias continuas es denso (propiedad de los números reales). En la figura 2.4 se observa cómo el área bajo la curva de cada f.d.p. es igual a 1 (correspondiente al 100% del espacio muestral).

Importante. Para el *caso continuo* el contradominio de la f.d.p. es $[0, \infty)$ (y no $[0, 1]$ como en el caso discreto) ya que $\boldsymbol{f_X(x)}$ **no es una probabilidad** (de hecho ni siquiera es relevante). Por ejemplo, en el caso de la f.d.p. de la variable aleatoria Y de la figura 2.4, $f_Y(y) > 1$ para $0 < y < 0.5$.

En el caso continuo lo relevante para el cálculo de probabilidades no es el valor que toma la f.d.p. sino el área bajo la curva.

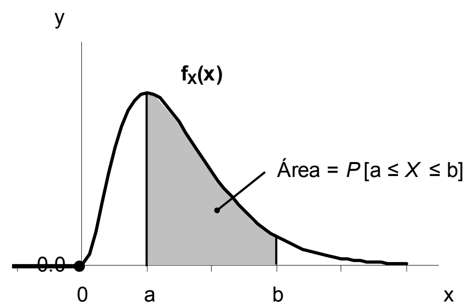
Proposición. Cálculo de probabilidades vía f.d.p.

Si X es variable aleatoria continua y $A \subset \mathbf{R}$, entonces $P[X \in A] = \int_A f_X(x)dx$

Corolario. En particular:

Si X es variable aleatoria continua $P[a \leq X \leq b] = \int_a^b f_X(x)dx$

Figura 2.5



Proposición

Si X es variable aleatoria continua entonces:
i) $P[X = a] = 0$
ii) $P[X \leq a] = P[X < a]$

Demostración. Para el caso *i*) si X es variable aleatoria continua con f.d.p. $f_X(x)$ y $h > 0$ entonces $P[a \leq X \leq a+h] = \int_a^{a+h} f_X(x)dx$, de modo que para h suficientemente pequeña se obtiene que $P[X = a] = \lim_{h \to 0} P[a \leq X \leq a+h] = \lim_{h \to 0} \int_a^{a+h} f_X(x)dx = \int_a^a f_X(x)dx = 0$. Para el caso *ii*) cómo X es variable aleatoria continua $\{X \leq a\} = \{X < a\} \cup \{X = a\}$, en donde $\{X < a\}$ y $\{X = a\}$ son eventos m.e., entonces considerando el resultado demostrado en el caso *i*) se tiene que $P[X \leq a] = P[X < a] + P[X = a] = P[X < a] + 0 = P[X < a]$.

Proposición. Generalización del caso *ii*) de la proposición anterior.

Si X es variable aleatoria continua entonces para $a \leq b$:
$P[a \leq X \leq b] = P[a < X < b] = P[a \leq X < b] = P[a < X \leq b]$

Ejercicios $\boxed{\text{E54}}$, $\boxed{\text{E55}}$ y $\boxed{\text{E56}}$.

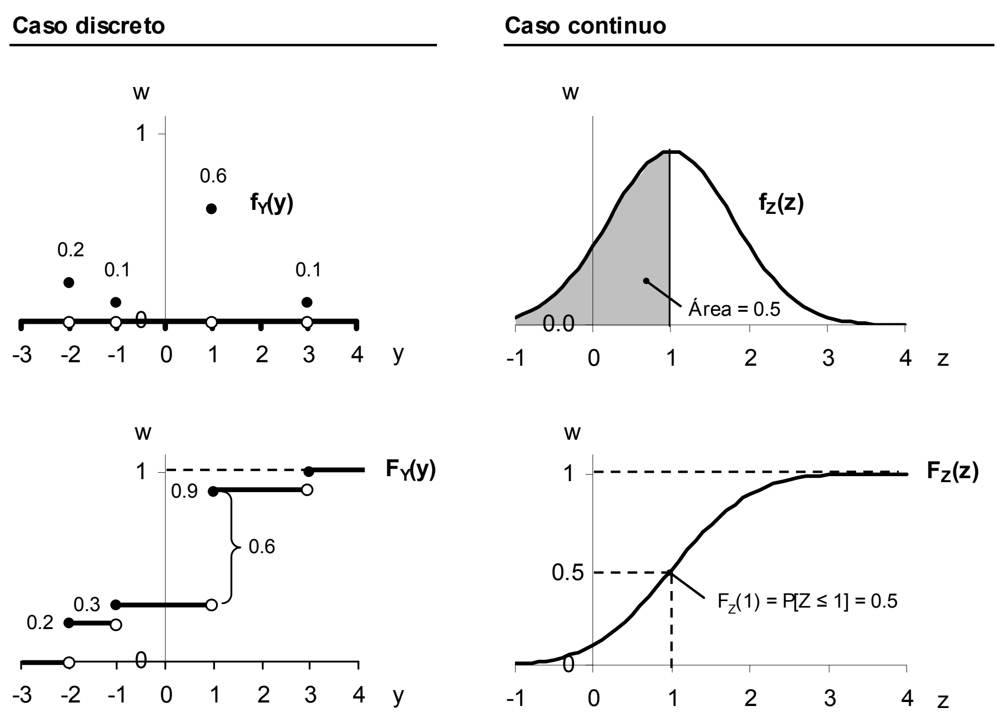
2.2.3. Función de distribución acumulada (f.d.a.)

Def. Función de distribución acumulada (f.d.a.)

Se dice que la función $F_X(\cdot)$, $F_X : \mathbf{R} \to [0, 1]$ es la función de distribución acumulada de la variable aleatoria X si $F_X(x) = P[X \le x]$ para toda $x \in \mathbf{R}$.

La f.d.a. tiene la **misma definición para variables aleatorias discretas o continuas**. La figura 2.6 contrasta la relación que hay entre la f.m.p. y su correspondiente f.d.a. (caso discreto) y la relación que hay entre la f.d.p. y su correspondiente f.d.a. (caso continuo).

Figura 2.6



Proposición

i) Si W es variable aleatoria discreta con soporte $\{w_1, w_2, \ldots\}$ entonces
$$F_W(w) = P[W \le w] = \sum_{\{w_i : w_i \le w\}} f_W(w_i) \text{ para toda } w \in \mathbf{R} \quad \text{(función escalonada)}$$

ii) Si Z es variable aleatoria continua entonces
$$F_Z(z) = P[Z \le z] = \int_{-\infty}^{z} f_Z(t)\,dt \text{ para toda } z \in \mathbf{R} \quad \text{(función continua)}$$

Note cómo el **cálculo de la f.d.a. en el caso continuo requiere un cambio de notación** en la variable del integrando para evitar ambigüedad. Este cambio es válido ya que en cualquier integral definida la variable del integrando es una *variable muda*.

Caracterización de la f.d.a.

Se dice que la función $F_X(\cdot)$, $F_X : \mathbf{R} \to [0, 1]$ es la función de distribución acumulada de la variable aleatoria X si y sólo si:

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

ii) $F_X(\cdot)$ es no decreciente, es decir, $a < b \Rightarrow F_X(a) \le F_X(b)$; y

iii) $F_X(\cdot)$ es continua por la derecha, es decir, $\lim_{h\to 0} F_X(x+h) = F_X(x)$ para $h > 0$.

En la figura 2.6 se puede apreciar cómo tanto $F_Y(y)$ (caso discreto) como $F_Z(z)$ (caso continuo) satisfacen las 3 condiciones que caracterizan a cualquier f.d.a.

Notación. Por practicidad, se considerarán las siguientes simplificaciones en la notación que involucre límites:

- Límites al infinito: $\lim_{x\to\infty} F_X(x) = F_X(\infty)$.
- Límites por la derecha: $\lim_{h\to 0} F_X(a+h) = \lim_{x\to a^+} F_X(x) = F_X(a^+)$ para $h > 0$.
- Límites por la izquierda: $\lim_{h\to 0} F_X(a-h) = \lim_{x\to a^-} F_X(x) = F_X(a^-)$ para $h > 0$.

Ejercicios $\boxed{\text{E57}}$, $\boxed{\text{E58}}$ y $\boxed{\text{E59}}$.

Proposición

i) Si W es variable aleatoria discreta con $a < b$, $P[a \le W \le b] = F_W(b) - F_W(a^-)$

ii) Si Z es variable aleatoria continua con $a < b$, $P[a \le Z \le b] = F_Z(b) - F_Z(a)$

Construcción de la f.m.p. (caso discreto) y de la f.d.p. (caso continuo) vía f.d.a.

i) Si W es variable aleatoria discreta entonces $f_W(w) = F_W(w) - F_W(w^-)$

ii) Si Z es variable aleatoria continua entonces $f_Z(z) = \frac{d}{dz} F_Z(z)$

Ejercicios $\boxed{\text{E60}}$ y $\boxed{\text{E61}}$.

Def. Función de Supervivencia

Se dice que la función $S_X(\cdot)$, $S_X : [0, \infty) \to [0, 1]$ es la función de supervivencia de la variable aleatoria X si $S_X(x) = P[X > x]$ para $x \geq 0$.

Las funciones de supervivencia son útiles para modelar el **tiempo futuro de vida**. Por ejemplo, el tiempo a la muerte de algunas especies (v.gr., seres humanos, animales, bacterias, virus y algunas otras especies biológicas), o la **vida útil** de algunos objetos (v.gr., maquinaria, mobiliario, equipo de cómputo, artículos electrodomésticos). A la rama de la Estadística encargada del estudio de esta clase de fenómenos se le conoce como *Análisis de Supervivencia*.

Proposición

$S_X(x) = 1 - F_X(x)$ para $x \geq 0$

Demostración. Por definición, si $x \geq 0$: $S_X(x) = P[X > x] = 1 - P[X \leq x] = 1 - F_X(x)$.

Ejercicio $\boxed{\text{E62}}$.

2.3. Características numéricas de las variables aleatorias

2.3.1. Esperanza, varianza y momentos de variables aleatorias

Def. Esperanza

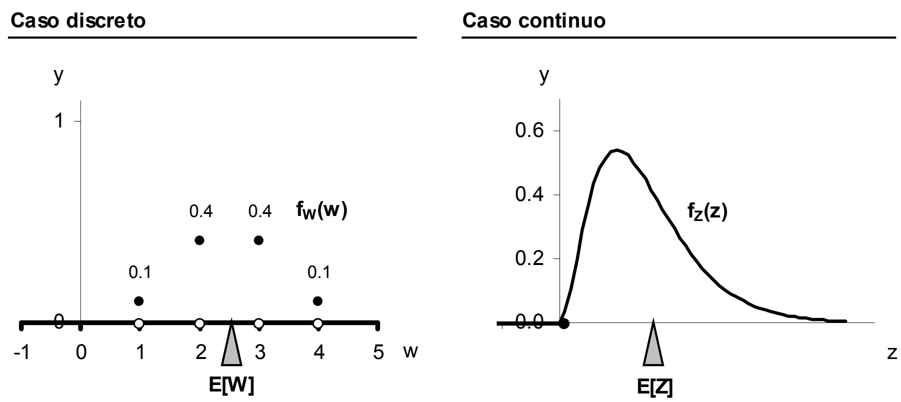
Si X es variable aleatoria, entonces su esperanza o media es:

i) $E[X] = \sum_x x\, f_X(x)$ si X es discreta.

ii) $E[X] = \int_{-\infty}^{\infty} x\, f_X(x)dx$ si X es continua.

La esperanza, también conocida como esperanza matemática o media, es el **promedio ponderado** de los valores que toma X, ponderados por sus respectivas probabilidades. En el caso continuo este promedio ponderado llevado al límite se convierte en una integral.

Figura 2.7



Por ejemplo, la esperanza de la variable aleatoria W de la figura 2.7 es:

$$E[W] = \sum_{w} w f_W(w) = (1)(0.1) + (2)(0.4) + (3)(0.4) + (4)(0.1) = 2.5$$

Geométricamente, $E[X]$ es el **centro de masa** de la función $f_X(x)$, es decir, el punto en el cuál la f.m.p. (caso discreto) o la f.d.p. (caso continuo) se equilibran, como se muestra en la figura 2.7.

Cuando $E[X]$ es de la **forma ∞ o ∞ – ∞,** se dice que **la esperanza no existe**. La esperanza existe únicamente si $E[X] < \infty$, es decir, cuando converge a un valor.

Proposición

Si $E[X]$ existe y $a \leq X \leq b \Rightarrow a \leq E[X] \leq b$

Si X es variable aleatoria discreta, $E[X]$ puede tomar un valor acotado por su soporte pero que no necesariamente pertenece al soporte, como se observa en la figura 2.7.

Ejercicios E63 , E64 , E65 , E66 y E67 .

Proposición. Esperanza de v.a.'s continuas vía f.d.a.

Si X es variable aleatoria continua, $E[X] = \int_0^{\infty} [1 - F_X(x)]dx - \int_{-\infty}^{0} F_X(x)dx$

La demostración de esta proposición se puede realizar resolviendo la doble integral que aparece al sustituir $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ y aplicando el *Teorema de Fubini* (intercambio en el orden de integración sobre la misma región).

El concepto de *esperanza* se puede generalizar a través del *valor esperado*.

Def. Valor Esperado

Si X es variable aleatoria y $g(X)$ una transformación de ésta, g: $\mathbf{R} \to \mathbf{R}$, entonces el valor esperado de $g(X)$ se define por:

i) $E[g(X)] = \sum_x g(x) f_X(x)$ si X es discreta.

ii) $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$ si X es continua.

Son ejemplos de la transformación g: $g(X) = X^2$, $g(X) = \ln(X)$, $g(X) = (X - c)^2$ donde c es una constante, etc. El **valor esperado** es la **generalización** de la esperanza ya que $Y = g(X)$ es una nueva variable aleatoria. La **esperanza** es un **caso particular** del valor esperado, considerando la transformación identidad $g(X) = X$.

Por la forma en que se define, al cálculo del valor esperado también se le conoce como *ley del estadístico inconsciente* (LEI), pues para calcular el valor esperado de la nueva variable aleatoria $Y = g(X)$ basta aplicar la transformación $g(\cdot)$ a X (lado izquierdo) y a sus posibles valores x (lado derecho).

Igual que en el caso de la esperanza, cuando $E[g(X)]$ es de la **forma ∞ o ∞ – ∞,** se dice que **el valor esperado no existe**. El valor esperado existe únicamente si $E[g(X)] < \infty$, es decir, cuando converge a un valor.

Aunque X es la variable aleatoria para la cuál se conoce $f_X(x)$, la definición de valor esperado permite calcular $E[Y] = E[g(X)]$ sin necesidad de conocer $f_Y(y)$. Por ejemplo, si X denota la cantidad producida de algún bien con $f_X(x)$ conocida y $C(X)$ es su función de costos, entonces el **costo esperado** es $E[C(X)]$. Del mismo modo es posible hablar de **utilidad esperada**, **ingreso esperado**, etc.

Cálculo de probabilidades de $Y = g(X)$ vía $f_X(x)$.

Si $Y = g(X)$ y $f_X(x)$ es conocida pero $f_Y(y)$ no, es posible hacer el cálculo de probabilidades asociadas a la variable aleatoria Y a partir de $f_X(x)$ **"despejando adecuadamente"** a X de la ecuación $Y = g(X)$.

Ejercicios $\boxed{\text{E68}}$ y $\boxed{\text{E69}}$.

Def. Momentos y Momentos Centrales

Si X es variable aleatoria y $k \in \mathbf{Z}^+$, entonces:
i) $\mu'_k = E\left[X^k\right]$ es su *k*-ésimo momento.
ii) $\mu_k = E\left[(X - E[X])^k\right]$ es su *k*-ésimo momento central.

El **cálculo de los momentos y de los momentos centrales** se realiza aplicando la *ley del estadístico inconsciente* considerando las transformaciones $g(X) = X^k$ y $g(X) = (X - E[X])^k$, respectivamente.

Note cómo $k = 1 \Rightarrow \mu'_1 = E\left[X^1\right] = E[X]$, es decir, **el primer momento de una variable aleatoria es su esperanza**.

Algunos momentos (centrales y no centrales) tienen **interpretación**, algunos otros sólo sirven de manera auxiliar en el cálculo de otras características numéricas importantes de las distribuciones de probabilidad de las variables aleatorias.

Ejercicio $\boxed{\text{E70}}$.

Proposición. Momentos centrales vía momentos no centrales

$$\mu_k = \sum_{i=0}^{k} (-1)^i \binom{k}{i} (\mu'_1)^i \mu'_{k-i} \quad \text{para } k \in \mathbf{Z}^+, \text{ con } \mu'_0 = 1$$

La proposición establece que cualquier momento central se puede calcular a partir de los momentos no centrales. Considerando la fórmula anterior para k = 2, 3 y 4 se obtienen los resultados del siguiente corolario.

Corolario. Fórmula para el segundo, tercer y cuarto momentos centrales

Si X es variable aleatoria y $E[X] = \mu$, entonces:
i) $\mu_2 = \mu'_2 - (\mu'_1)^2$, es decir, $E\left[(X-\mu)^2\right] = E\left[X^2\right] - (E[X])^2$
ii) $\mu_3 = \mu'_3 - 3\mu'_1\mu'_2 + 2(\mu'_1)^3$, es decir,
$E\left[(X-\mu)^3\right] = E\left[X^3\right] - 3E[X]E\left[X^2\right] + 2(E[X])^3$ (ver formulario)
iii) $\mu_4 = \mu'_4 - 4\mu'_1\mu'_3 + 6(\mu'_1)^2\mu'_2 - 3(\mu'_1)^4$, es decir,
$E\left[(X-\mu)^4\right] = E\left[X^4\right] - 4E[X]E\left[X^3\right] + 6(E[X])^2 E\left[X^2\right] - 3(E[X])^4$ (ver formulario)

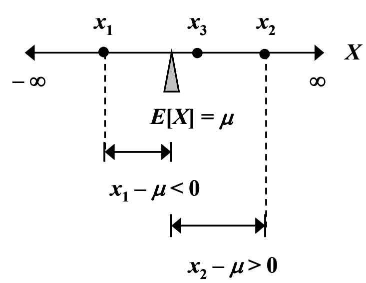
Def. Varianza

Si X es variable aleatoria, entonces su varianza se define por $Var[X] = E[(X - E[X])^2]$

Note cómo **la varianza es el segundo momento central**.

La **varianza** es una **medida de dispersión** de la distribución de la variable aleatoria X respecto a su esperanza como se muestra en la figura 2.8.

Figura 2.8



Note cómo:

- $X - E[X]$ es una distancia aleatoria **con signo** (positiva para algunas realizaciones de X y negativa para algunas otras, como se muestra en la figura 2.8); y
- $g(X) = (X - E[X])^2 \geq 0$ es una distancia aleatoria positiva (o cero), pero con **unidades al cuadrado**.

El cálculo de la varianza de X es simple y sencillamente el cálculo del valor esperado de $g(X) = (X - E[X])^2 \geq 0$, es decir, de una transformación de la variable aleatoria X.

Proposición

$Var[X] \geq 0$

Para resolver el problema de las unidades al cuadrado de la varianza se define la *desviación estándar* como la raíz cuadrada de la varianza.

Def. Desviación Estándar

$$\sigma[X] = \sqrt{Var[X]}$$

La **desviación estándar** es una **medida de dispersión** de la distribución de la variable aleatoria *X* respecto a su esperanza, pero medida en las mismas unidades que la variable.

Ejercicio $\boxed{\text{E71}}$.

El cálculo de la varianza (y de la desviación estándar en consecuencia) se puede simplificar a través del siguiente teorema.

Teorema

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

Ejercicios $\boxed{\text{E72}}$, $\boxed{\text{E73}}$, $\boxed{\text{E74}}$ y $\boxed{\text{E75}}$.

2.3.2. Propiedades del valor esperado y la varianza

Teorema

Si *X* es variable aleatoria; $a, b, c \in \mathbf{R}$; y $f, g : \mathbf{R} \to \mathbf{R}$, entonces:

i) $E[c] = c$

ii) $E[ag(X) + bh(X)] = aE[g(X)] + bE[h(X)]$ (operador lineal)

iii) $Var[c] = 0$

iv) $Var[cX] = c^2 Var[X]$

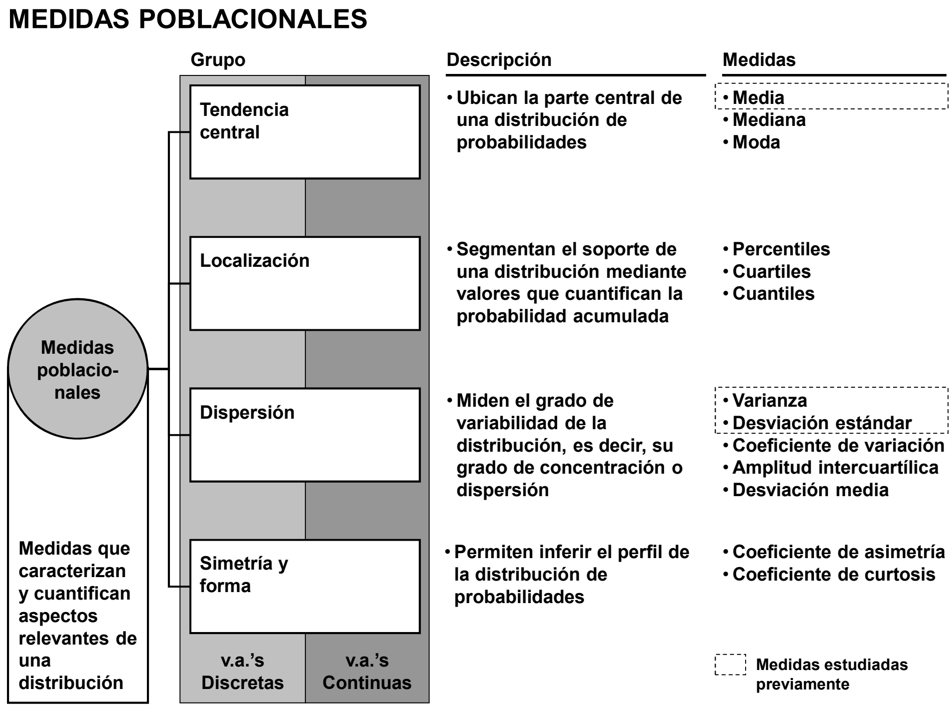
v) $Var[X + c] = Var[X]$

Importante. Cualquier operador que "saca constantes" y que "parte sumas", se denomina operador lineal. En particular, ***E*[·] es un operador lineal** como se observa en la propiedad *ii*) del Teorema anterior.

Ejercicios $\boxed{\text{E76}}$ y $\boxed{\text{E77}}$.

2.3.3. Medidas poblacionales

Figura 2.9

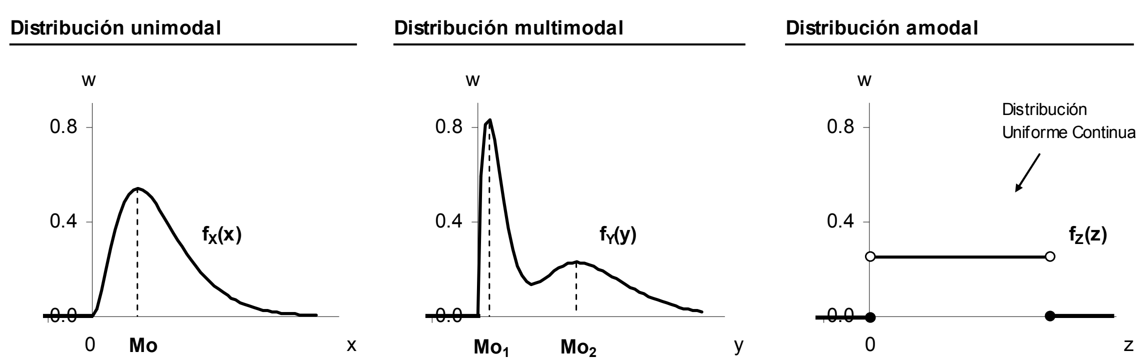


MEDIDAS DE TENDENCIA CENTRAL

Def. Moda

> La moda de la variable aleatoria X, denotada por Mo, es el valor del soporte de X que maximiza a $f_X(x)$, es decir, su f.m.p. (caso discreto) o su f.d.p. (caso continuo).

Figura 2.10



La moda puede no ser única o incluso no existir:

- Si $f_X(x)$ alcanza su máximo en un sólo punto se dice que X tiene una **distribución unimodal**.
- Si $f_X(x)$ alcanza su máximo en más de un punto se dice que X tiene una **distribución multimodal**. En estricto sentido, la moda es el valor para el cuál $f_X(x)$ es un *máximo global*, sin embargo, es común considerar como modas a los valores en los que $f_X(x)$ alcanza un *máximo local* como se muestra en la figura 2.10.
- Si $f_X(x)$ toma el mismo valor en todos los puntos del soporte se dice que X tiene una **distribución amodal**.

Si X es variable *aleatoria discreta*, su **moda** es el **valor más probable**.

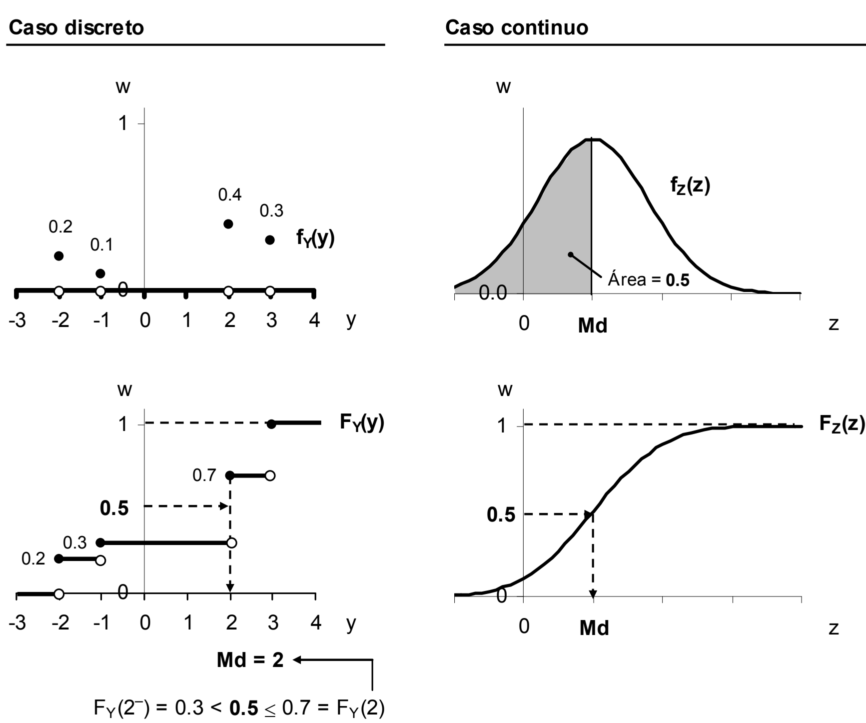
Def. Mediana

Si X es variable aleatoria con f.d.a. $F_X(x)$, entonces su mediana, denotada por Md, es el valor del soporte de X tal que:

i) $F_X\left(Md^-\right) < \frac{1}{2} \le F_X\left(Md\right)$ si X es discreta.

ii) $F_X\left(Md\right) = \frac{1}{2}$ si X es continua.

Figura 2.11



La **mediana** es el valor de X por debajo (y por arriba) del cuál se **acumula el 50% de la distribución** de probabilidades.

En el *caso discreto*, la **mediana** es el **mínimo valor de *x*** tal que $\frac{1}{2} \le F_X(x)$.

Ventajas de la mediana sobre la media:

- La mediana *siempre existe*.
- La mediana es una *mejor medida de tendencia central en distribuciones no simétricas*.

Def. Distribución Simétrica

Si X es variable aleatoria con f.m.p. (caso discreto) o f.d.p. (caso continuo) $f_X(x)$ y $c \in \mathbf{R}$, se dice que X tiene una distribución simétrica respecto a c si $f_X(c-x) = f_X(c+x)$.

Ejercicios $\boxed{\text{E78}}$ y $\boxed{\text{E79}}$.

MEDIDAS DE LOCALIZACIÓN

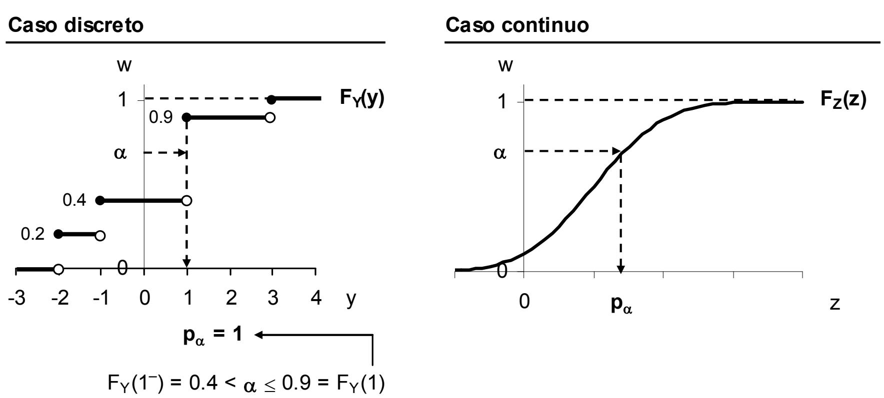
Def. Percentiles

Si X es variable aleatoria con f.d.a. $F_X(x)$, entonces el percentil α, $0 < \alpha < 1$, denotado por p_α, es el valor del soporte de X tal que:

i) $F_X(p_\alpha^-) < \alpha \le F_X(p_\alpha)$ si X es discreta.

ii) $F_X(p_\alpha) = \alpha$ si X es continua.

Figura 2.12



El **percentil α** es el valor de X por debajo del cual se **acumula el 100α% de la distribución** de probabilidades.

En el *caso discreto*, el **percentil α** es el mínimo valor de x tal que $\alpha \leq F_X(x)$.

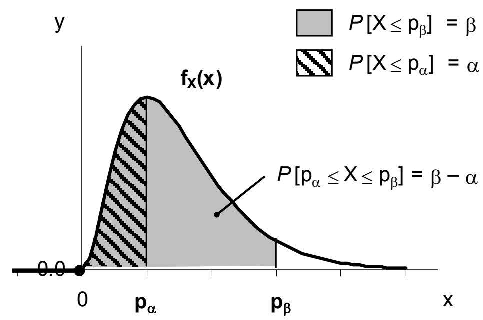
Los percentiles 0.1, 0.2,…, 0.9 se denominan **deciles**, y surgen frecuentemente en aplicaciones de Economía (v.gr., distribución del ingreso).

Proposición

Si X es variable aleatoria continua y $0 < \alpha < \beta < 1$, entonces $P\lfloor p_\alpha \leq X \leq p_\beta \rfloor = \beta - \alpha$

Demostración. $P\lfloor p_\alpha \leq X \leq p_\beta \rfloor = F_X(p_\beta) - F_X(p_\alpha) = \beta - \alpha$.

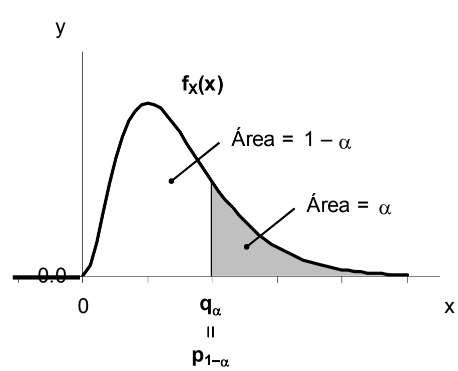
Figura 2.13



El **percentil α** de una distribución involucra un área α, $0 < \alpha < 1$, en la **cola izquierda** de la distribución de la variable aleatoria X.

Def. Cuantil. El valor del soporte de X que deja un área α en la **cola derecha** de la distribución se denomina **cuantil α** y se denota por q_α.

Figura 2.13



Proposición. Cuantiles vía percentiles

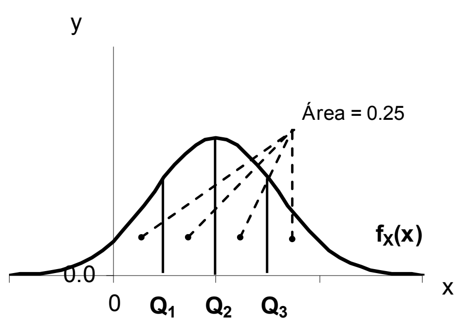
$q_\alpha = p_{1-\alpha}$

Def. Cuartiles

Si X es variable aleatoria entonces:
i) El primer cuartil o cuartil inferior se define por $Q_1 = p_{0.25}$
ii) El segundo cuartil se define por $Q_2 = p_{0.50}$
iii) El tercer cuartil o cuartil superior se define por $Q_3 = p_{0.75}$

Los **cuartiles** son los valores del soporte de X que dividen la distribución de probabilidades en 4 regiones equiprobables (25% en cada región) como se muestra en la figura 2.14.

Figura 2.14



Proposición.

$Md = Q_2 = p_{0.50}$

Ejercicio $\boxed{\text{E80}}$.

MEDIDAS DE DISPERSIÓN

Def. Coeficiente de variación

Si X es variable aleatoria, su esperanza existe y $E[X] \neq 0$, entonces su coeficiente de variación se define por $CV[X] = \dfrac{\sigma[X]}{E[X]}$

Tanto la desviación estándar como la media de la variable aleatoria X se expresan en las mismas unidades y en su cociente quedan canceladas, consecuentemente el **coeficiente de variación** es **adimensional** y en ocasiones se expresa como **porcentaje**. De hecho, es la desviación estándar como porcentaje de la media, es decir, una **medida de dispersión relativa**.

Ejercicio $\boxed{\text{E81}}$.

Def. Amplitud intercuartílica (AI)

$$AI = Q_3 - Q_1$$

La **amplitud intercuartílica** o amplitud intercuartil mide la variabilidad de la distribución de la variable aleatoria X considerando la **distancia entre el primer y el tercer cuartiles**.

Def. Desviación media respecto a la media

Si X es variable aleatoria con $\mu = E[X] < \infty$ su desviación media respecto a la media se define por $DM(\mu) = E[|X-\mu|]$

La transformación $g(X) = |X-\mu|$ es la **variación absoluta** de la variable aleatoria X respecto a su media $\mu = E[X]$. La **desviación media respecto a la media** mide la **variación absoluta promedio** de una variable aleatoria respecto a su media.

El concepto de desviación media se puede aplicar respecto a otras medidas de tendencia central, por ejemplo, moda o mediana:

- **Desviación media respecto a la moda**. $DM(Mo) = E[|X-Mo|]$, siempre y cuando X tenga una distribución unimodal.
- **Desviación media respecto a la mediana**. $DM(Md) = E[|X-Md|]$

La desviación estándar es preferida sobre la desviación media respecto a la media porque, en general, el cálculo $\sigma[X]$ es más simple que el de $DM(\mu)$ y por las propiedades matemáticas de la transformación $g(X) = (X-\mu)^2$ (función cuadrática) frente a las de $g(X) = |X-\mu|$ (función valor absoluto).

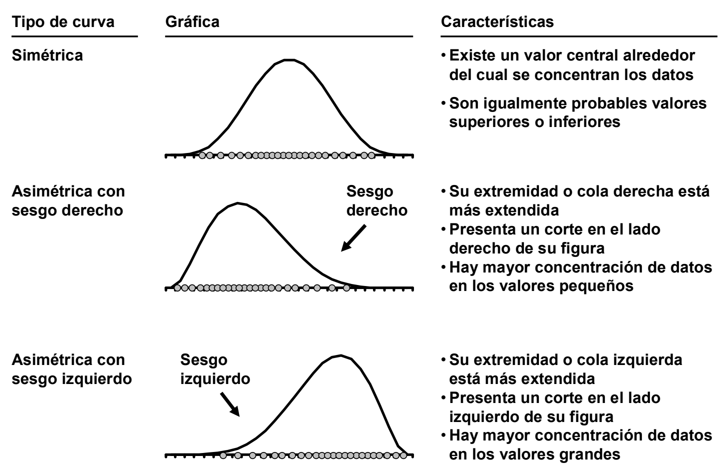
En general, la **desviación estándar sobredimensiona la variabilidad** en comparación con la desviación media respecto a la media.

MEDIDAS DE SIMETRÍA Y FORMA

Las distribuciones de probabilidad (o *curvas poblacionales*) se clasifican de acuerdo con su *grado de simetría* en **simétricas** o en **asimétricas** (con **sesgo derecho o izquierdo**).

Figura 2.15

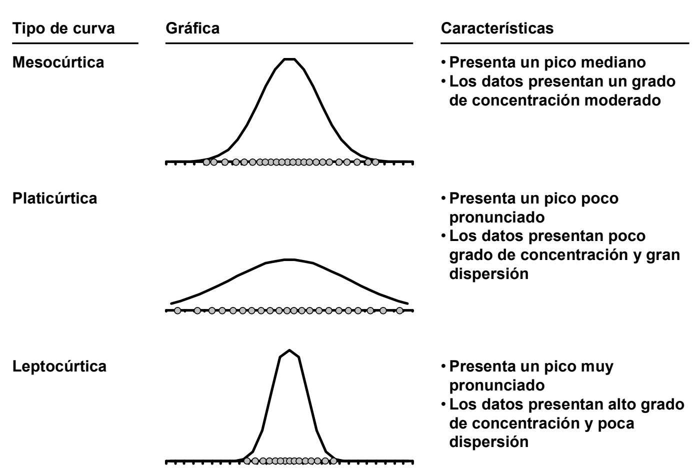
CLASIFICACIÓN DE LAS CURVAS POBLACIONALES DE ACUERDO CON SU GRADO DE SIMETRÍA



De acuerdo con su *picudez* (o *planura*), las distribuciones de probabilidad se clasifican en **mesocúrticas** (curva media), **platicúrticas** (curva plana) o **leptocúrticas** (curva pronunciada).

Figura 2.16

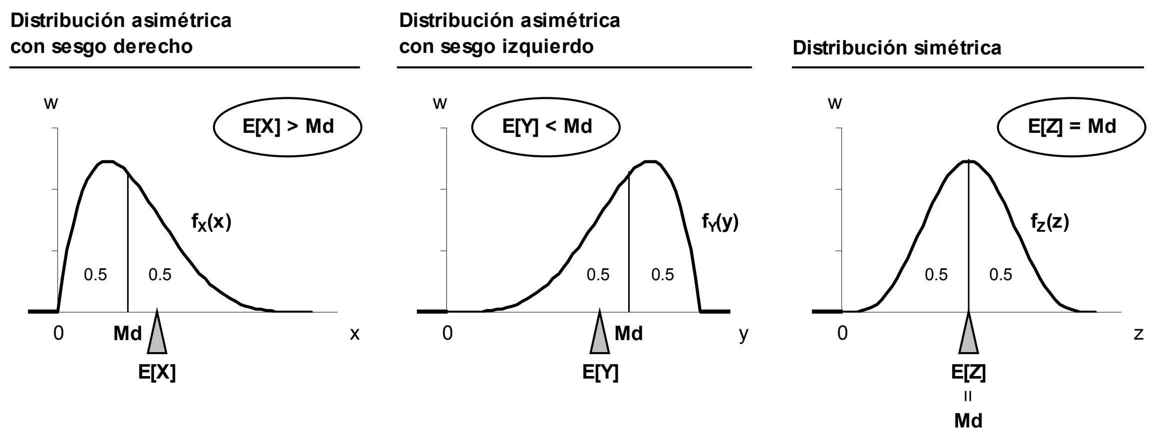
CLASIFICACIÓN DE LAS CURVAS POBLACIONALES DE ACUERDO CON SU "PICUDEZ"



Proposición. Clasificación de simetría vía relación media-mediana.

Si X es tiene distribución unimodal y $E[X] < \infty$, entonces:
i) $E[X] > Md \Rightarrow X$ tiene una distribución asimétrica con sesgo derecho.
ii) $E[X] < Md \Rightarrow X$ tiene una distribución asimétrica con sesgo izquierdo.
iii) $E[X] = Md \Rightarrow X$ tiene una distribución simétrica.

Figura 2.17



Def. Coeficiente de Asimetría y Coeficiente de Curtosis

Si X es variable aleatoria con $\sigma = \sigma[X]$ y $\mu_k = E\left[(X - E[X])^k\right]$, $k \in \mathbf{Z}^+$, entonces:

i) $CA[X] \equiv \alpha_3 = \dfrac{\mu_3}{\sigma^3}$ define su coeficiente de asimetría; y

ii) $CC[X] \equiv \alpha_4 = \dfrac{\mu_4}{\sigma^4}$ define su coeficiente de curtosis.

Importante. El cálculo de los coeficientes de asimetría y curtosis involucra el cálculo de **momentos centrales**. Éstos últimos pueden calcularse **vía momentos no centrales**.

El **coeficiente de asimetría** permite clasificar una distribución de probabilidad de acuerdo con su **grado de simetría**.

Interpretación del Coeficiente de Asimetría

i) $CA[X] > 0 \Rightarrow X$ tiene una distribución asimétrica con sesgo derecho.
ii) $CA[X] < 0 \Rightarrow X$ tiene una distribución asimétrica con sesgo izquierdo.
iii) $CA[X] = 0 \Rightarrow X$ tiene una distribución simétrica.

El **coeficiente de curtosis** permite clasificar una distribución de probabilidad de acuerdo con su **grado de picudez** (planura).

Interpretación del Coeficiente de Curtosis

i) $CC[X] > 3 \Rightarrow X$ tiene una distribución leptocúrtica.
ii) $CC[X] < 3 \Rightarrow X$ tiene una distribución platicúrtica.
iii) $CC[X] = 3 \Rightarrow X$ tiene una distribución mesocúrtica.

Ejercicio $\boxed{\text{E82}}$.

2.4. Función Generadora de Momentos

Def. Función Generadora de Momentos (f.g.m.)

La función generadora de momentos de la variable aleatoria X se define por:
$M_X(t) = E\left[e^{tX}\right]$ para t tal que $-b < t < b$ para alguna $b > 0$.

Se dice que la **función generadora de momentos no existe si**:

- $E\left[e^{tX}\right]$ es de la **forma** ∞ o ∞ – ∞; o
- **No se cumple la condición** $-b < t < b$ para alguna $b > 0$.

La condición $-b < t < b$ (o $|\, t \,| < b$) para alguna $b > 0$ exige que $M_X(t)$ exista en una vecindad del cero (pudiendo no estar definida exactamente en $t = 0$) y garantiza que las derivadas de $M_X(t)$ puedan ser evaluadas en $t = 0$.

Teorema. Cálculo de momentos (no centrales) vía f.g.m.

Si X es variable aleatoria y su función generadora de momentos existe, entonces:
$\left.\dfrac{d^k}{dt^k} M_X(t)\right|_{t=0} = E\left[X^k\right] \equiv \mu_k'$ para $k = 1, 2, \ldots$

Condiciones de regularidad. Si X es una v.a. cuya f.m.p. (caso discreto) o f.d.p. (caso continuo) posee propiedades tales que

$$\frac{\partial}{\partial t} E\left[h(X,t)\right] = E\left[\frac{\partial}{\partial t} h(X,t)\right],$$

es decir, que los operadores esperaza y derivada se pueden intercambiar, entonces se dice que prevalecen condiciones de regularidad en la distribución de X.

Para demostrar fácilmente el Teorema anterior es necesario suponer condiciones de regularidad.

Ejercicios $\boxed{\text{E83}}$, $\boxed{\text{E84}}$, $\boxed{\text{E85}}$ y $\boxed{\text{E86}}$.

Ante la dificultad de la no existencia de la función generadora de momentos es posible definir la *función característica*, que siempre existe pero se define en el campo de los números complejos (**C**).

Def. Función Característica

La función característica de la variable aleatoria X se define por:
$\phi_X(t) = E\left[e^{itX}\right]$, donde $i = \sqrt{-1} \in \mathbf{C}$

La **función característica** es única y define por completo el comportamiento de la distribución de una variable aleatoria. *Si dos variables aleatorias tienen la misma función característica, su distribución de probabilidades es la misma*. Esta misma propiedad se puede establecer con la f.g.m. (si es que existe) y se considerará más adelante para determinar la distribución de una transformación de una variable aleatoria.

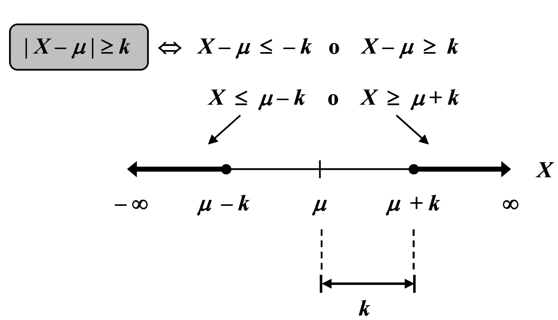
2.5. Desigualdades de Chebyshev y de Jensen

Teorema. Desigualdad de Chebyshev

Si X es una variable aleatoria con $E[X] = \mu$ y $Var[X] = \sigma^2$, ambas finitas, entonces para todo $k > 0$, $P[|X - \mu| \geq k] \leq \dfrac{\sigma^2}{k^2}$

La **Desigualdad de Chebyshev** o Tchebysheff permite establecer una **cota superior** para la **probabilidad** de que la variable aleatoria ***X* se aleje de su media en *k* unidades o más**, como se muestra en la figura 2.18.

Figura 2.18



Note cómo si $\frac{\sigma^2}{k^2} > 1$, la cota resulta poco útil pues $P[X \in A] \leq 1$ para todo $A \subset \mathbf{R}$.

La Desigualdad de Chebyshev es muy útil en la *Teoría Asintótica* de la Estadística, particularmente para demostrar la *Ley Débil de los Grandes Números*[1] (uno de los principales resultados de la Inferencia Estadística). La Desigualdad de Chebyshev se puede demostrar fácilmente mediante la *Desigualdad de Markov*.

Lema. Desigualdad de Markov

Si Z es una variable aleatoria no negativa, entonces para $a > 0$, $P[Z \geq a] \leq \frac{E[Z]}{a}$

Ejercicio $\boxed{\text{E87}}$.

Una versión alternativa de la Desigualdad de Chebyshev permite acotar la probabilidad de que la variable aleatoria X se aleje r desviaciones estándar de su media ($r > 0$). Si en la Desigualdad de Cheyshev $P[|X-\mu| \geq k] \leq \frac{\sigma^2}{k^2}$ se toma $k = r\sigma$, entonces se obtiene $P[|X-\mu| \geq r\sigma] \leq \frac{\sigma^2}{(r\sigma)^2} = \frac{1}{r^2}$.

Considerando la probabilidad del complemento en la Desigualdad de Chebyshev se obtiene una **cota inferior** para la **probabilidad** de que la variable aleatoria ***X* se aleje de su media en menos de *k* unidades**, como se muestra a continuación:

$$P[|X-\mu| \geq k] \leq \frac{\sigma^2}{k^2} \quad \Leftrightarrow \quad 1 - P[|X-\mu| < k] \leq \frac{\sigma^2}{k^2} \quad \Leftrightarrow \quad P[|X-\mu| < k] \geq 1 - \frac{\sigma^2}{k^2}$$

En resumen, la figura 2.19 muestra las cuatro versiones que se pueden tener de la Desigualdad de Chebyshev dependiendo del tipo de cota (superior o inferior) y las unidades en que X se aleja de μ (k unidades o r desviaciones estándar).

---

[1] La **Ley Débil de los Grandes Números** establece que si $X_1, X_2, \ldots, X_n$, forman una muestra aleatoria (variables aleatorias independientes e idénticamente distribuidas) con media común μ (llamada *media poblacional*), entonces la *media muestral* definida por $\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$ converge en probabilidad al valor de la *media poblacional* μ. Es decir, $\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{P} \mu$, o bien, $\lim_{n\to\infty} P[|\overline{X} - \mu| > \varepsilon] = 0$.

Figura 2.19

Distancia entre la v.a. X y su media μ	**Cota superior**	**Cota inferior**
k unidades	$P[\lvert X-\mu\rvert \geq k] \leq \frac{\sigma^2}{k^2}$	$P[\lvert X-\mu\rvert < k] \geq 1-\frac{\sigma^2}{k^2}$
r desviaciones estándar	$P[\lvert X-\mu\rvert \geq r\sigma] \leq \frac{1}{r^2}$	$P[\lvert X-\mu\rvert < r\sigma] \geq 1-\frac{1}{r^2}$

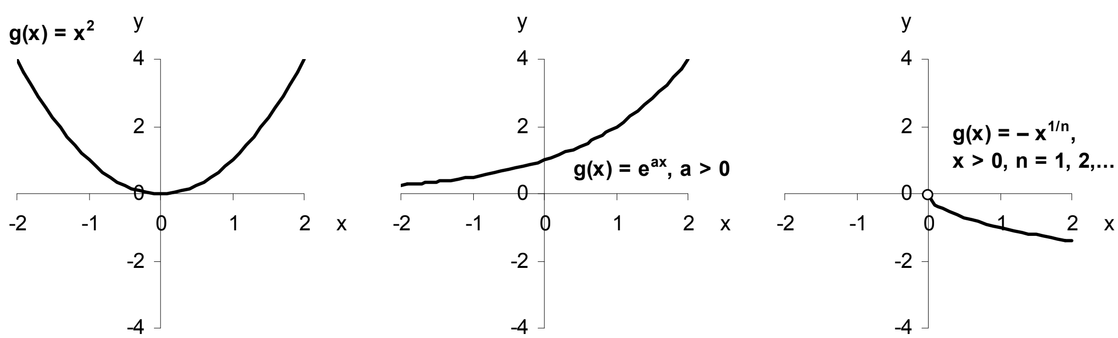
Ejercicio $\boxed{\text{E88}}$.

Def. Funciones convexas y cóncavas

Si $g: D \subset \mathbf{R} \to \mathbf{R}$ es de clase C^2, se dice que:
i) $g(x)$ es función convexa si $g''(x) \geq 0$ para toda $x \in D$
ii) $g(x)$ es función cóncava si $g''(x) \leq 0$ para toda $x \in D$

Son ejemplos de **funciones convexas**: $g(x) = x^2$, $g(x) = e^{ax}$ para $a > 0$ o $g(x) = -\sqrt[n]{x}$ para $x \geq 0$ y $n = 1, 2,\ldots$, como se muestra en la figura 2.20.

Figura 2.20



Teorema. Desigualdad de Jensen

Si $g(x)$ es una función convexa, entonces $E[g(X)] \geq g(E[X])$, considerando que los valores esperados existen.

Aplicaciones inmediatas de este resultado son:

- $E\left[\frac{1}{X}\right] \geq \frac{1}{E[X]}$ si $X > 0$, ya que $g(x) = \frac{1}{x}$ es convexa para $x > 0$.
- $E[X^2] \geq (E[X])^2 \Rightarrow E[X^2] - (E[X])^2 = Var[X] \geq 0$, ya que $g(x) = x^2$ es convexa.

Note cómo si $g(x)$ es función convexa, entonces $h(x) = -g(x)$ es función cóncava, entonces a partir de la Desigualdad de Jensen es posible establecer el siguiente corolario.

Corolario.

Si $g(x)$ es una función cóncava, entonces $E[g(X)] \le g(E[X])$, considerando que los valores esperados existen.

Proposición

Si $g(x)$ es una función lineal, entonces $E[g(X)] = g(E[X])$, considerando que los valores esperados existen.

La proposición anterior establece que el caso de igualdad de la Desigualdad de Jensen se alcanza cuando $g(x)$ es función lineal.

Demostración. Suponga que $g(x)$ es función lineal, es decir, que existen a, $b \in \mathbf{R}$ tales que $g(x) = a + bx$. Evaluando g en $E[X]$ se obtiene $g(E[X]) = a + bE[X]$. Por otro lado, por propiedades de la esperanza se sabe que $E[g(X)] = E[a + bX] = a + bE[X]$, concluyendo así que $E[g(X)] = g(E[X])$.

Ejercicios $\boxed{\text{E89}}$ y $\boxed{\text{E90}}$.

2.6. Distribución de una Transformación de una Variable Aleatoria

2.6.1. Distribución de Transformaciones de Variables Aleatorias Discretas

Teorema.

Sea X una variable aleatoria discreta con soporte x_1, x_2,… y función de masa de probabilidad $f_X(x)$. Si $Y = g(X)$, g: $\mathbf{R} \to \mathbf{R}$, es una transformación de la variable aleatoria X entonces $f_Y(y) = \sum_{\{x_i : g(x_i) = y\}} f_X(x_i)$

El teorema anterior establece que $P[Y = y]$, para cada y del soporte de la variable aleatoria discreta Y, se determina sumando las probabilidades $P[X = x_i]$ correspondientes a la o a las masas x_i, que hacen que $g(x_i) = y$.

Ejercicio $\boxed{\text{E91}}$.

2.6.2. Distribución de Transformaciones de Variables Aleatorias Continuas

Se presentarán 3 métodos para obtener la distribución de la variable aleatoria $Y = g(X)$ a partir de la distribución de la variable aleatoria X:

- Método de la f.d.a
- Método de la transformación monótona
- Método de la f.g.m.

MÉTODO DE LA F.D.A.

Consiste en calcular $F_Y(y)$ a partir de $F_X(x)$, donde $Y = g(X)$ y $F_X(x)$ es conocida. Una vez que se tiene $F_Y(y)$ es posible calcular $f_Y(y) = \frac{d}{dy} F_Y(y)$.

Ejercicios $\boxed{\text{E92}}$ y $\boxed{\text{E93}}$.

MÉTODO DE LA TRANSFORMACIÓN MONÓTONA

Teorema. Distribución de la Transformación Monótona (DTM)

Sea X variable aleatoria continua con función de densidad de probabilidad $f_X(x)$ y sea $g(x)$ una función estrictamente monótona (creciente o decreciente) y diferenciable (consecuentemente continua) en x. Si $Y = g(X)$ entonces:

i) $f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy} g^{-1}(y)\right|$ si $y = g(x)$ para alguna x; y

ii) $f_Y(y) = 0$ si $y \neq g(x)$ para toda x,

donde $g^{-1}(y)$ es el valor de x tal que $g(x) = y$.

Ejercicios $\boxed{\text{E94}}$, $\boxed{\text{E95}}$ y $\boxed{\text{E96}}$.

MÉTODO DE LA F.G.M.

Teorema. Igualdad en Distribución vía f.g.m.

Si X y Y son variables aleatorias entonces $M_X(t) = M_Y(t)$ si y sólo si X y Y tienen la misma distribución de probabilidad.

La demostración de este teorema se omitirá ya que requiere del manejo de las funciones características de X y Y y del uso de probabilidad avanzada.

Ejercicio $\boxed{\text{E97}}$.

3. Distribuciones Importantes

3.1. Distribuciones Bernoulli y Binomial.

3.1.1. Distribución Bernoulli.

Def. Experimento Bernoulli o Ensayo Bernoulli. Es un experimento aleatorio que cumple con dos características:

i) Sus resultados pueden ser clasificados como **éxito**, con probabilidad p, o **fracaso** con probabilidad $q = 1 - p$; y
ii) El resultado de cada uno de los experimentos es **independiente** de los demás.

Def. Distribución Bernoulli.

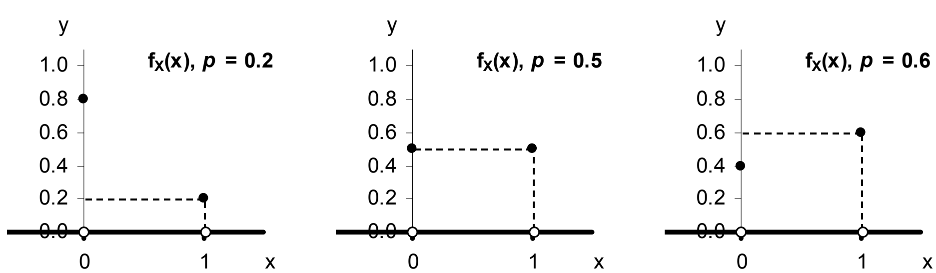
Se dice que $X \sim Bernoulli(p)$ si

$$f_X(x) = p^x (1-p)^{1-x} I_{\{0,1\}}(x) = \begin{cases} 1-p & si \;\; x=0 \\ p & si \;\; x=1 \\ 0 & e.o.c. \end{cases}, \qquad 0 \le p \le 1$$

Aplicaciones. Esta distribución recibe su nombre en honor del matemático suizo Jacques Bernoulli (1654-1705) y permite modelar Experimentos Bernoulli. Son Experimentos Bernoulli: el lanzamiento de un volado, el nacimiento de un niño, la perforación de un pozo exitoso, la efectividad de una vacuna, etc.

La Figura 3.1 muestra la gráfica de la función de masa de probabilidad Bernoulli para algunos valores del parámetro p.

Figura 3.1



Propiedades de la Distribución Bernoulli

Si $X \sim Bernoulli(p)$, entonces

i) $E[X] = p$ (Ver E84)

ii) $Var[X] = p(1-p)$ (Ver E84)

iii) $M_X(t) = pe^t + (1-p),\ t \in \mathbf{R}$ (Ver E84)

iv) $F_X(x) = (1-p)I_{[0,1)}(x) + I_{[1,\infty)}(x)$

Note cómo **esperanza y varianza son función del parámetro *p***. Así, por ejemplo, analizando $g(p) = Var[X] = p - p^2$ para $0 \le p \le 1$ se verifica de $p = \frac{1}{2}$ maximiza la varianza de una Distribución Bernoulli.

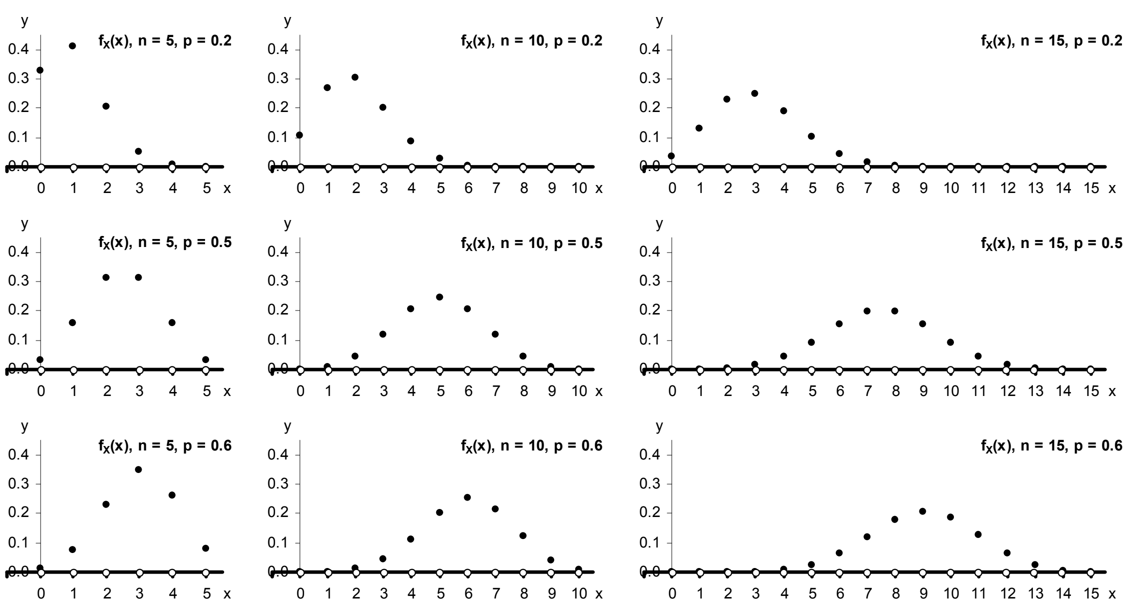
3.1.2. Distribución Binomial.

Def. Distribución Binomial.

Se dice que $X \sim Bin(n, p)$ si

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} I_{\{0,1,\ldots,n\}}(x), \quad n \in \mathbf{Z}^+, \quad 0 \le p \le 1$$

Figura 3.2



La figura 3.2 muestra la f.m.p. Binomial para distintos valores de los parámetros *n* y *p*. Observe cómo la distribución es simétrica cuando $p = 0.5$ y asimétrica cuando $p \neq 0.5$.

Aplicaciones. Si $X \sim Bin(n, p)$ entonces X cuenta el número de éxitos al realizar n Experimentos Bernoulli, cada uno con probabilidad de éxito p, e independiente de los demás.

Propiedades de la Distribución Binomial

Si $X \sim Bin(p)$, entonces
i) $E[X] = np$ (Ver E98)
ii) $Var[X] = np(1-p)$ (Ver E98)
iii) $M_X(t) = [(1-p) + pe^t]^n$, $t \in \mathbf{R}$ (Ver E98)
iv) $F_X(x)$ en **Tablas** para n = 1, 2,..., 25 y p = 0.05, 0.10,..., 1.00 (Ver p. 2-17)
v) $Mo = \lfloor (n+1)p \rfloor$ (y $\lfloor (n+1)p \rfloor - 1$ si n es impar y p = 0.5)

Proposición. Transformaciones de la Distribución Binomial

i) Si $X \sim Bin(1, p)$, entonces $X \sim Bernoulli(p)$
ii) Si $X \sim Bin(n, p)$, entonces $Y = n - X \sim Bin(n, 1-p)$

Ejercicios E98 y E99 .

3.2. Distribución Poisson.

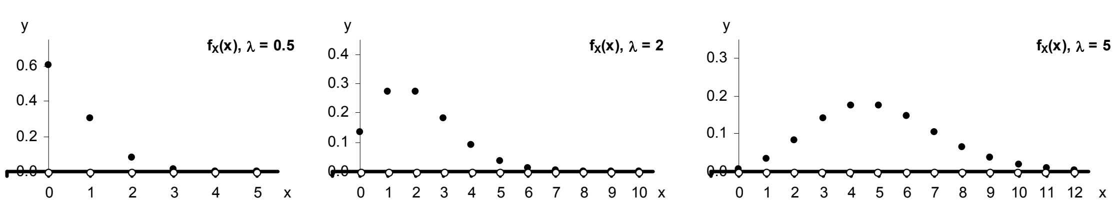
Def. Distribución Poisson.

Se dice que $X \sim Po(\lambda)$ si

$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} I_{\{0,1,\ldots\}}(x), \quad \lambda > 0$$

La figura 3.3 muestra la gráfica de la función de masa de probabilidad Poisson para algunos valores del parámetro λ.

Figura 3.3



Aplicaciones. Esta distribución recibe su nombre en honor del físico y matemático francés Siméon Denis Poisson (1781-1840), y permite modelar eventos que ocurren de acuerdo con los postulados de la *Ley Poisson*.

Postulados de la Ley Poisson.

i) El espacio muestral se genera por infinitos Experimentos Bernoulli con probabilidades de éxito muy pequeñas (*Ley de Eventos Raros*).
ii) El número de éxitos en intervalos de tiempo mutuamente excluyentes es independiente.
iii) La probabilidad de dos o más éxitos en el mismo punto es cero.
iv) El número promedio de éxitos en un intervalo es constante (λ).

Algunos ejemplos de variables aleatorias que usualmente obedecen a la Ley Poisson son: el número de errores en una página de un libro, el número de personas de una comunidad que viven más de 100 años, el número de teléfonos incorrectos que se marcan en un día, el número de clientes que llegan a una oficina postal, etc.

En algunas aplicaciones, el parámetro λ puede ser interpretado como una **tasa de ocurrencia** (# eventos / tiempo).

Propiedades de la Distribución Poisson

Si $X \sim Po(\lambda)$, entonces:

i) $E[X] = \lambda$ (Ver T2 I.8)

ii) $Var[X] = \lambda$ (Ver T2 I.8)

iii) $M_X(t) = \exp\{\lambda(e^t - 1)\}$, $t \in \mathbf{R}$ (Ver T2 I.8)

iv) $F_X(x)$ en **Tablas** para $\lambda =$ 0.02, 0.04,..., 0.10, 0.15,...,1.0, 1.1,...,2.0, 2.2,..., 10.8, 11, 12,..., 25 (Ver p. 18-21)

v) $Mo = \lfloor \lambda \rfloor$ (y $\lfloor \lambda \rfloor - 1$ si $\lambda \in \mathbf{Z}^+$)

Teorema. Aproximación Binomial por Poisson

Si $X \sim Bin(n, p)$, $n \to \infty$, $p \to 0$ y $np \to \lambda$ entonces $X \sim Po(\lambda)$ con $\lambda = np$.

Este Teorema se puede demostrar por la convergencia de la función de masa de probabilidad (E101) o por la convergencia de la función generadora de momentos (T4 I.1).

Ejercicios $\boxed{\text{E100}}$, $\boxed{\text{E101}}$ y $\boxed{\text{E102}}$.

3.3. Distribución Uniforme Continua.

Def. Distribución Uniforme Continua.

Se dice que $X \sim U(\theta_1, \theta_2)$ si
$$f_X(x) = \frac{1}{\theta_2 - \theta_1} I_{[\theta_1, \theta_2]}(x), \quad -\infty < \theta_1 < \theta_2 < \infty$$

Propiedades de la Distribución Uniforme Continua.

Si $X \sim U(\theta_1, \theta_2)$, entonces:

i) $E[X] = \dfrac{\theta_1 + \theta_2}{2}$ (Ver E70)

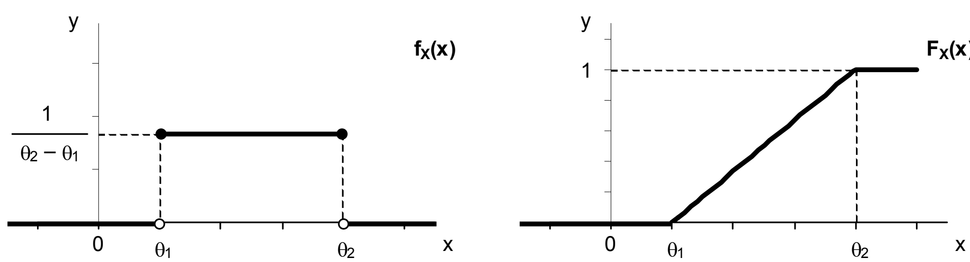
ii) $Var[X] = \dfrac{(\theta_2 - \theta_1)^2}{12}$ (Ver E82)

iii) $M_X(t) = \dfrac{e^{t\theta_2} - e^{t\theta_1}}{t(\theta_2 - \theta_1)}$, $t \neq 0$

iv) $F_X(x) = \dfrac{x - \theta_1}{\theta_2 - \theta_1} I_{[\theta_1, \theta_2]}(x) + I_{(\theta_2, \infty)}(x)$ (Ver E103)

La figura 3.4 muestra las gráficas de las funciones de densidad y de distribución acumulada de la Distribución Uniforme Continua con parámetros θ_1 y θ_2.

Figura 3.4



Aplicaciones. Permite modelar fenómenos equiprobables sobre intervalos de igual longitud.

Ejercicios $\boxed{\text{E103}}$ y $\boxed{\text{E104}}$.

La Distribución Uniforme Continua sobre el intervalo [0, 1] es de particular interés en la **Teoría de Simulación**, gracias al siguiente Teorema.

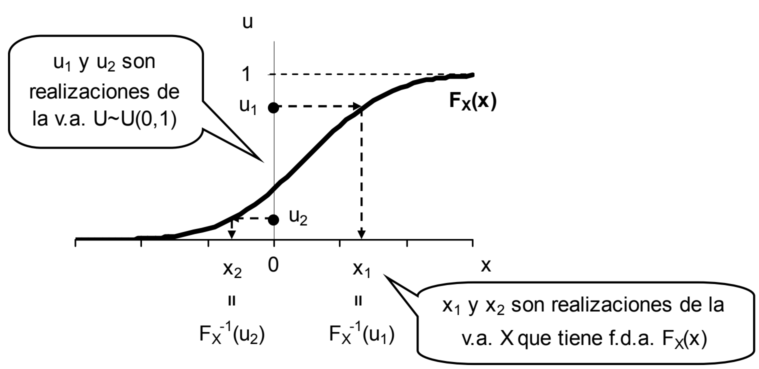
Teorema. Método de la Transformación Inversa

Sea $U \sim U(0, 1)$. Para cualquier función de distribución acumulada continua F, si se define a la variable aleatoria $X = F^{-1}(U)$, entonces X tiene función de distribución acumulada F.

Demostración. Considere los siguientes hechos: (i) como F es f.d.a. continua entonces F es función estrictamente creciente; (ii) si $U \sim U(0, 1)$ entonces $F_U(u) = u$ para $0 \le u \le 1$; y en particular (iii) como F es f.d.a. $0 \le F(x) \le 1$ para toda $x \in \mathbf{R}$. Si $X = F^{-1}(U)$ entonces al aplicar el método de la f.d.a. considerando (i), (ii) y (iii) se obtiene que: $F_X(x) = P[X \le x] = P[F^{-1}(U) \le x] = P[U \le F(x)] = F_U(F(x)) = F(x)$. Es decir, X tiene función de distribución acumulada F.

Este Teorema permite generar números aleatorios provenientes de cualquier distribución continua con f.d.a. inversa $F^{-1}(\cdot)$ a partir de números aleatorios distribuidos uniformemente en el intervalo [0, 1], como se muestra en la figura 3.5. Muchos paquetes de cómputo y calculadoras generan números aleatorios distribuidos uniformemente entre 0 y 1, por ejemplo, la función "=ALEATORIO()" de Excel o la tecla **Ran#** de algunas calculadoras.

Figura 3.5



Proposición. Simulación de variables aleatorias Uniformes Continuas.

Si $U \sim U(0, 1)$ y $\theta_1 < \theta_2$ entonces $X = \theta_1 + (\theta_2 - \theta_1)U \sim U(\theta_1, \theta_2)$ (Ver T4 I.2)

3.4. Distribuciones Gamma, Exponencial y Ji Cuadrada.

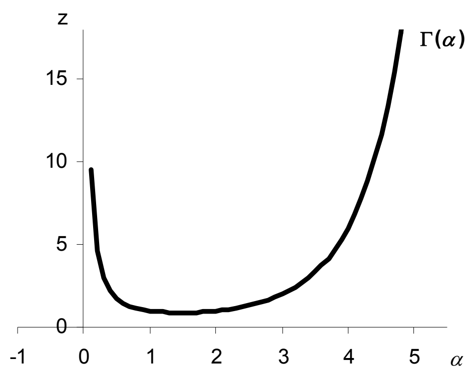
3.4.1. Distribución Gamma.

Def. Función Gamma.

$$\Gamma(\alpha)=\int_0^\infty y^{\alpha-1}e^{-y}dy\,, \quad \alpha>0$$

La figura 3.6 muestra la gráfica de la Función Gamma. Note cómo la función comienza decreciendo hasta alcanzar un mínimo entre 1 y 2 pero luego crece en forma acelerada.

Figura 3.6



Propiedades de la Función Gamma.

i)	$\Gamma(1)=\Gamma(2)=1$	(Ver E105)
ii)	$\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$	(Ver E105)
iii)	Si $n\in\mathbf{Z}^+$ entonces $\Gamma(n)=(n-1)!$	(Ver E105)
iv)	$\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$	(Ver T4 I.3)

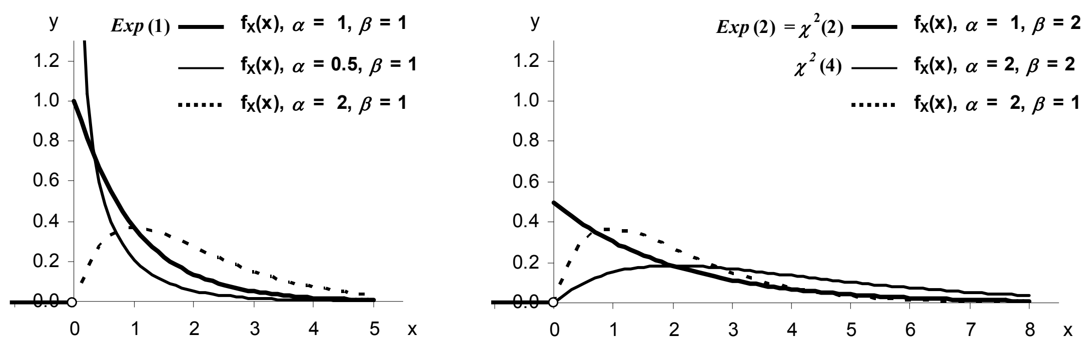
Def. Distribución Gamma.

Se dice que $X \sim Gamma(\alpha, \beta)$ si

$$f_X(x)=\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}}I_{(0,\infty)}(x), \quad \alpha>0, \quad \beta>0$$

La figura 3.7 muestra la gráfica de la función de densidad Gamma para distintos valores de los parámetros α y β. Los parámetros α y β son llamados **factores de forma y de escala**, respectivamente.

Figura 3.7



Propiedades de la Distribución Gamma.

Si $X \sim Gamma(\alpha, \beta)$, entonces:

i) $E[X] = \alpha\beta$ (Ver E106)

ii) $Var[X] = \alpha\beta^2$ (Ver E106)

iii) $M_X(t) = (1 - \beta t)^{-\alpha}$, $t < \dfrac{1}{\beta}$

iv) $F_X(x)$ se calcula mediante integración numérica.

v) $Mo = \beta(\alpha - 1)$, $\alpha > 1$

Aplicaciones. La Distribución Gamma es asimétrica con sesgo derecho y es muy versátil por los perfiles que pueden generar sus parámetros. Se utiliza frecuentemente para modelar tiempo de vida, tiempos de espera, ingreso familiar, pérdidas, etc.

Las Distribuciones Exponencial y Ji Cuadrada son casos particulares de la Distribución Gamma.

Proposición.

i) Si $X \sim Gamma(1, \beta)$, entonces $X \sim Exp(\beta)$

ii) Si $X \sim Gamma\left(\dfrac{\nu}{2}, 2\right)$, entonces $X \sim \chi^2(\nu)$, $\nu \in \mathbf{Z}^+$

Ejercicios $\boxed{\text{E105}}$ y $\boxed{\text{E106}}$.

3.4.2. Distribución Exponencial. $Exp(\beta) \equiv Gamma(1, \beta)$

Def. Distribución Exponencial.

Se dice que $X \sim Exp(\beta)$ si

$$f_X(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} I_{(0,\infty)}(x), \quad \beta > 0$$

Propiedades de la Distribución Exponencial.

Si $X \sim Exp(\beta)$, entonces:

i) $E[X] = \beta$ (Ver E74)

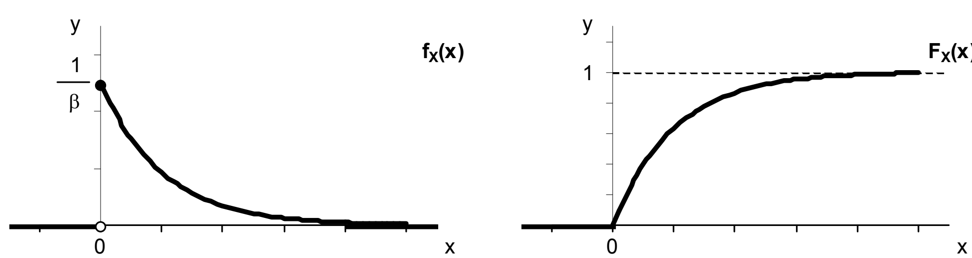
ii) $Var[X] = \beta^2$ (Ver E74)

iii) $M_X(t) = \dfrac{1}{1-\beta t}, \ t < \dfrac{1}{\beta}$ (Ver T2 I.7)

iv) $F_X(x) = \left(1 - e^{-\frac{x}{\beta}}\right) I_{(0,\infty)}(x)$ (Ver E59)

La figura 3.8 muestra las gráficas de las funciones de densidad y de distribución acumulada de la Distribución Exponencial con parámetro β.

Figura 3.8



Aplicaciones. La Distribución Exponencial permite modelar la distribución del tiempo hasta la ocurrencia de algún evento específico, por ejemplo, la ocurrencia de un terremoto, una llamada telefónica, la llegada de un tren, etc. Se utiliza con frecuencia en **Teoría de Colas** para modelar líneas de espera.

Proposición. Relación entre las Distribuciones Poisson y Exponencial.

El tiempo que transcurre entre la ocurrencia de dos eventos de una Distribución Poisson con parámetro λ tiene una Distribución Exponencial con parámetro $\beta = \dfrac{1}{\lambda}$.

Proposición. Pérdida de Memoria de la Distribución Exponencial.

Si $X \sim Exp(\beta)$ entonces $P[X > t + s | X > t] = P[X > s]$ para todo s, $t \geq 0$ (Ver E108)

Proposición. Simulación de variables aleatorias Exponenciales.

Si $U \sim U(0, 1)$, entonces $X = -\beta \ln(1-U) \sim Exp(\beta)$ (Ver E96)

Ejercicios $\boxed{\text{E107}}$, $\boxed{\text{E108}}$, y $\boxed{\text{E109}}$.

3.4.3. Distribución Ji Cuadrada. $\chi^2(\nu) = Gamma\left(\frac{\nu}{2}, 2\right)$

Def. Distribución Ji Cuadrada.

Se dice que $X \sim \chi^2(\nu)$ si

$$f_X(x) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right)2^{\frac{\nu}{2}}} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} I_{(0,\infty)}(x), \quad \nu = 1, 2, \ldots$$

Al parámetro ν se le denomina **grados de libertad**. Por ejemplo, si $X \sim \chi^2(8)$ se dice que la variable aleatoria X tiene una Distribución Ji Cuadrada con 8 grados de libertad.

Propiedades de la Distribución Ji Cuadrada.

Si $X \sim \chi^2(\nu)$, entonces:

i) $E[X] = \nu$

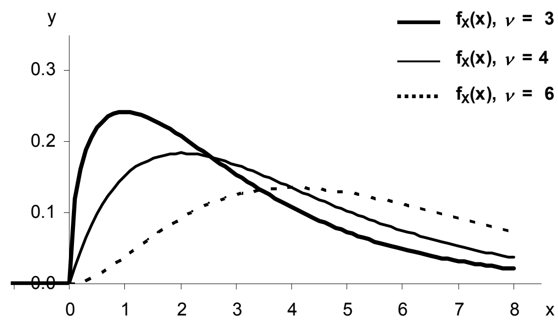
ii) $Var[X] = 2\nu$

iii) $M_X(t) = (1-2t)^{-\frac{\nu}{2}}$, $t < \frac{1}{2}$

iv) $F_X(x)$ mediante **cuantiles** 0.001, 0.005, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.975, 0.99, 0.995, 0.999 en **Tablas** para $\nu = 1, 2, \ldots, 30, 40, \ldots, 150$ (Ver p. 26-27)

La figura 3.9 muestra la gráfica de la función de densidad Ji Cuadrada para distintos valores del parámetro ν.

Figura 3.9



Aplicaciones. La Distribución Ji-Cuadrada es de gran utilidad para la **Inferencia Estadística** pues está asociada al comportamiento probabilístico de la *varianza muestral* y aparece en algunas *pruebas de hipótesis*.

Ejercicio $\boxed{\text{E110}}$.

3.5. Distribución Normal.

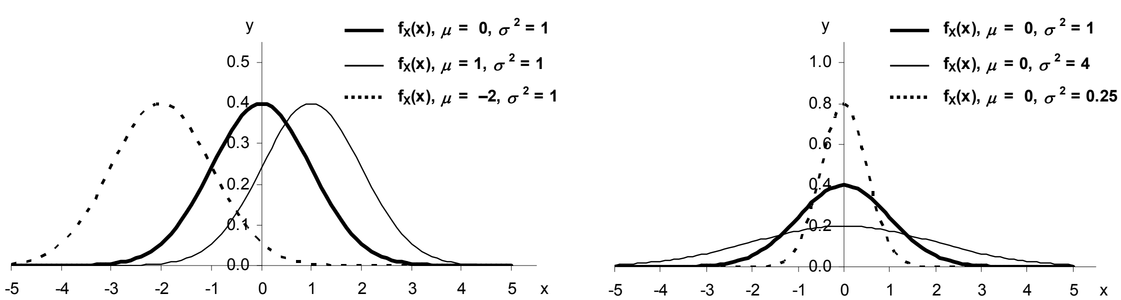
Def. Distribución Normal.

Se dice que $X \sim N(\mu,\ \sigma^2)$ si

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} I_{(-\infty,\infty)}(x), \quad \mu \in \mathbf{R}, \quad \sigma > 0$$

La **Distribución Normal** o **Distribución Gaussiana** es la distribución de probabilidad más importante en Estadística. La figura 3.10 muestra la gráfica de la función de densidad Normal para distintos valores de los parámetros μ y σ^2.

Figura 3.10



Note cómo la Distribución Normal siempre es simétrica respecto a μ. El parámetro μ es un **factor de localización** y modifica el centro de la función de densidad. El parámetro σ^2 es un **factor de forma** y modifica la picudez de la función de densidad.

A pesar de que σ^2 modifica su picudez, la Distribución Normal siempre es **mesocúrtica** (coeficiente de curtosis igual a 3) y sirve de estándar para clasificar la picudez del resto de las distribuciones.

La Distribución Normal con parámetros $\mu = 0$ y $\sigma^2 = 1$ se denomina **Normal Estándar** o **Normal Unitaria**.

Def. Distribución Estándar.

Se dice que $Z \sim$ *Normal Estándar* $\equiv N(0, 1)$ si

$$f_Z(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} I_{(-\infty,\infty)}(z)$$

La función ***f***$f_Z(z)$ **no tiene antiderivada**, es decir, no existe expresión analítica cerrada de una función $F_Z(z)$ cuya derivada sea $f_Z(z)$.

Para verificar que $A = \int_{-\infty}^{\infty} \phi(z)dz = 1$ hay que demostrar que $A^2 = 1$, es decir:

$$A^2 = \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx\right)\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dydx$$

El cálculo de la doble integral se puede realizar considerando el *cambio de variables a coordenadas polares*, es decir, $y = r$ sen(θ) y $x = r \cos(\theta)$ para $r > 0$ y $0 < \theta < 2\pi$.

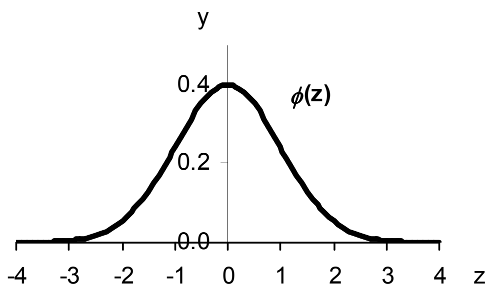
La **función de distribución acumulada** de la Distribución Normal Estándar se expresa a través de la *Función Fi*, pero sus valores se calculan mediante **integración numérica**.

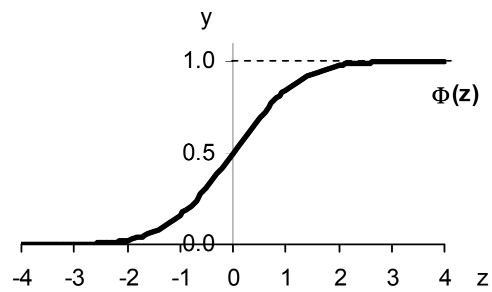
Def. Función Fi.

$$\Phi(z) = P[Z \le z] = \int_{-\infty}^{z} \phi(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad z \in \mathbf{R}$$

La figura 3.11 muestra las gráficas de las funciones de densidad y de distribución acumulada de la Distribución Normal Estándar.

Figura 3.11





Los valores de $\Phi(z)$ aparecen en **Tablas** para z = –2.99, –2.98,…, 3.09 (Ver p. 22-23).

Propiedades de la Función Fi.

i)	$\Phi(0) = \dfrac{1}{2}$
ii)	$\Phi(-x) = 1 - \Phi(x)$

Demostración. Suponga que Z tiene una Distribución Normal Estándar. Por la simetría respecto a cero y la continuidad de Z se tiene que $P[Z > x] = P[Z < -x] = P[Z \leq -x]$ para toda $x \in \mathbf{R}$, entonces: (i) $1 = P[Z \leq 0] + P[Z > 0] = P[Z \leq 0] + P[Z \leq 0] = 2\Phi(0)$ y en consecuencia $\Phi(0) = \dfrac{1}{2}$; y (ii) $\Phi(-x) = P[Z \leq -x] = P[Z > x] = 1 - P[Z \leq x] = 1 - \Phi(x)$.

Ejercicio $\boxed{\text{E111}}$.

El **cálculo de probabilidades de cualquier Distribución Normal** (μ y σ^2 arbitrarios) se realiza a través de $\Phi(z)$ gracias al siguiente Teorema.

Teorema. Estandarización de variables aleatorias Normales.

Si $X \sim N(\mu, \sigma^2)$, entonces $Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$

La transformación $Z = g(X) = \dfrac{X - \mu}{\sigma}$ se denomina **estandarización de la variable aleatoria *X***. La resta de $\mu = E[X]$ hace que el centro de la distribución de Z sea el cero, y la división entre $\sigma = \sigma[X]$ modifica la forma de la densidad para alcanzar la curvatura de la Normal Estándar.

La demostración del Teorema de Estandarización de variables aleatorias Normales se puede realizar mediante el Teorema de la Distribución de la Transformación Monótona (ver T3 I4) o mediante el Teorema de Igualdad en Distribución vía f.g.m. (ver E97).

Propiedades de la Distribución Normal.

Si $X \sim N(\mu, \sigma^2)$, entonces:

i) $E[X] = \mu$ (Ver E97)

ii) $Var[X] = \sigma^2$ (Ver E97)

iii) $M_X(t) = \exp\left\{\mu t + \dfrac{\sigma^2 t^2}{2}\right\},\ t \in \mathbf{R}$ (Ver E97)

iv) $F_X(x) = \Phi\left(\dfrac{x-\mu}{\sigma}\right)$ y **Tablas** para $z = -2.99, -2.98, \ldots, 3.09$ (Ver p. 22-23)

Aplicaciones. A pesar de que la Distribución Normal tiene soporte $(-\infty, \infty)$, la probabilidad de esta variable aleatoria esté a más de 3 desviaciones estándar de su media es muy cercana a cero. Se utiliza frecuentemente para modelar variables aleatorias que concentran su probabilidad alrededor de cierto valor y para las cuales es igualmente probable estar por arriba o por debajo de este valor (simetría respecto a la media).

Proposición. Regla del 68, 95 y 99%.

Si $X \sim N(\mu, \sigma^2)$, entonces:

i) $P[|X-\mu| < \sigma] = P[\mu - \sigma < X < \mu + \sigma] = 0.6827 \approx 68\%$ (Ver T3 I4)

ii) $P[|X-\mu| < 2\sigma] = P[\mu - 2\sigma < X < \mu + 2\sigma] = 0.9545 \approx 95\%$ (Ver T3 I4)

iii) $P[|X-\mu| < 3\sigma] = P[\mu - 3\sigma < X < \mu + 3\sigma] = 0.9973 \approx 99\%$ (Ver T3 I4)

Ejercicio $\boxed{\text{E112}}$.

La **importancia de la Distribución Normal en Estadística** está dada por el *Teorema Central del Límite*. Este Teorema establece que la suma o el promedio de un gran número de variables aleatorias siempre tienen una Distribución Normal, sin importar su distribución original.

Teorema de De Moivre-Laplace. Aproximación Binomial por Normal.

Si $X \sim Bin(n, p)$ y $n \to \infty$ entonces $X \sim_{aprox.} N(np, np(1-p))$

Este Teorema también se puede enunciar diciendo que si $X \sim Bin(n, p)$ y $n \to \infty$, entonces $P\left[a \le \frac{X - np}{\sqrt{np(1-p)}} \le b\right] \to \Phi(b) - \Phi(a)$, es decir, el cálculo aproximado de probabilidades asociadas una Distribución Binomial se puede realizar a través de una Distribución Normal.

El Teorema de De Moivre-Laplace es un caso particular del *Teorema Central del Límite*, ya que la Distribución Binomial surge de la suma de Distribuciones Bernoulli.

Al aproximar una distribución discreta (v.gr., Binomial) a través de una distribución continua (v.gr., Normal) es posible mejorar la aproximación considerando un **ajuste de corrección por continuidad** llamado *Ajuste de Yate*. Este ajuste consiste en sumar o restar 0.5 en los extremos del intervalo de la variable aleatoria discreta sobre el cual se desea calcular la probabilidad.

Ejercicio $\boxed{\text{E113}}$.

Teorema. Distribución de transformaciones lineales de la Distribución Normal.

Si $X \sim N(\mu, \sigma^2)$ y $Y = \alpha + \beta X$, $\beta \neq 0$, entonces $Y \sim N(\alpha + \beta\mu, \beta^2\sigma^2)$

Demostración. Si $X \sim N(\mu, \sigma^2)$ entonces $M_X(t) = \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$, $t \in \mathbf{R}$. Por su parte, la f.g.m. de $Y = \alpha + \beta X$, $\beta \neq 0$, es $M_Y(t) = E\left[e^{tY}\right] = E\left[e^{t(\alpha + \beta X)}\right] = e^{\alpha t} E\left[e^{(\beta t)X}\right] = e^{\alpha t} M_X(\beta t)$. Al sustituir en la expresión anterior la f.g.m. de la v.a. X evaluada en βt se obtiene $M_Y(t) = e^{\alpha t} \exp\left\{\mu(\beta t) + \frac{\sigma^2(\beta t)^2}{2}\right\} = \exp\left\{(\alpha + \beta\mu)t + \frac{(\beta^2\sigma^2)t^2}{2}\right\}$, que es la f.g.m. de una Distribución Normal con media $\alpha + \beta\mu$ y varianza $\beta^2\sigma^2$, i.e., $Y \sim N(\alpha + \beta\mu, \beta^2\sigma^2)$.

El Teorema anterior establece que **cualquier transformación lineal de una Distribución Normal también es Normal**. Note cómo los parámetros de Y pueden calcularse directamente utilizando propiedades de operador lineal de la esperanza:

- $E[Y] = E[\alpha + \beta X] = \alpha + \beta E[X] = \alpha + \beta\mu$; y
- $Var[Y] = Var[\alpha + \beta X] = \beta^2 Var[X] = \beta^2\sigma^2$.

Ejercicio $\boxed{\text{E114}}$.

Proposición. Otras transformaciones de la Distribución Normal.

Si $X \sim N(\mu, \sigma^2)$, entonces:

i) $Z = \dfrac{X-\mu}{\sigma} \sim N(0, 1)$ (Ver T4 I.4)

ii) $W = \left(\dfrac{X-\mu}{\sigma}\right)^2 \sim \chi^2(1)$ (Ver T4 I.5)

iii) $Y = e^X \sim LogN(\mu, \sigma^2)$ (Ver T4 I.5)

La distribución de la variable aleatoria Y del inciso iii) la proposición anterior se denomina **Distribución Lognormal** y se utiliza frecuentemente en Economía y Finanzas para modelar el rendimiento de acciones, índices y otros valores.

4. Distribuciones Multivariadas

4.1. Funciones de probabilidad conjunta y marginales

En forma análoga al caso univariado, para el caso multivariado se definen:

- *Función de masa de probabilidad conjunta* para el caso Discreto; y
- *Función de densidad de probabilidad conjunta* para el caso Continuo.

La mayoría de las definiciones de este tema se presentarán para el caso bivariado (variables aleatorias X y Y), pero estas expresiones se pueden generalizar intuitivamente para n variables aleatorias, $X_1, X_2,\ldots, X_n$, $n \in \mathbf{Z}^+$.

Def. Función de masa de probabilidad conjunta (f.m.p.c.)

Se dice que $f_{X,Y} : \mathbf{R}^2 \to [0, 1]$ es la función de masa de probabilidad conjunta de las variables aleatorias X y Y con soporte $S \subset \mathbf{R}^2$ si:

i) $$f_{X,Y}(x, y) = \begin{cases} P[X = x, Y = y] > 0 & \textit{para } (x, y) \in S \\ 0 & \textit{e.o.c.} \end{cases}$$

ii) $$\sum_x \sum_y f_{X,Y}(x, y) = 1$$

Cálculo de probabilidades vía f.m.p.c.

Si $A \subset \mathbf{R}^2$ y (X, Y) tienen f.m.p.c. $f_{X,Y}(x, y)$, entonces:

$$P[(X,Y) \in A] = \sum\sum_{\{(x,y):(x,y)\in A\}} f_{X,Y}(x, y)$$

Def. Funciones de masa de probabilidades marginales (f.m.p. marginal)

Si $f_{X,Y}(x, y)$ es la f.m.p.c de (X, Y), entonces:

i) $f_X(x) = \sum_y f_{X,Y}(x, y)$ es la f.m.p. marginal de X.

ii) $f_Y(y) = \sum_x f_{X,Y}(x, y)$ es la f.m.p. marginal de Y.

La forma más común y práctica de manejar las Distribuciones Bivariadas Discretas es a través de su **representación tabular**.

Ejercicios $\boxed{\text{E115}}$ y $\boxed{\text{E116}}$.

Def. Función de densidad de probabilidad conjunta (f.d.p.c.)

Se dice que $f_{X,Y}: \mathbf{R}^2 \to [0, \infty)$ es la función de densidad de probabilidad conjunta de las variables aleatorias X y Y con soporte $S \subset \mathbf{R}^2$ si:

i) $f_{X,Y}(x, y) \geq 0$ para todo $(x, y) \in S$

ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)dydx = 1$

Cálculo de probabilidades vía f.d.p.c.

Si $A \subset \mathbf{R}^2$ y (X, Y) tienen f.d.p.c. $f_{X,Y}(x, y)$, entonces:
$P[(X, Y) \in A] = \int\int_A f_{X,Y}(x, y)dydx$

Def. Funciones de densidad de probabilidades marginales (f.d.p. marginal)

Si $f_{X,Y}(x, y)$ es la f.d.p.c. de (X, Y), entonces:

i) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$ es la f.d.p. marginal de X.

ii) $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$ es la f.d.p. marginal de Y.

Ejercicios $\boxed{\text{E117}}$ y $\boxed{\text{E118}}$.

Def. Función de distribución acumulada conjunta (f.d.a. conjunta).

Se dice que $F_{X,Y}: \mathbf{R}^2 \to [0, 1]$ es la función de distribución acumulada conjunta de las variables aleatorias X y Y si $F_{X,Y}(x, y) = P[X \leq x, Y \leq y]$

Igual que en el caso univariado, la definición de f.d.a. conjunta es la misma para variables aleatorias Discretas y Continuas.

Propiedades de la f.d.a. conjunta.

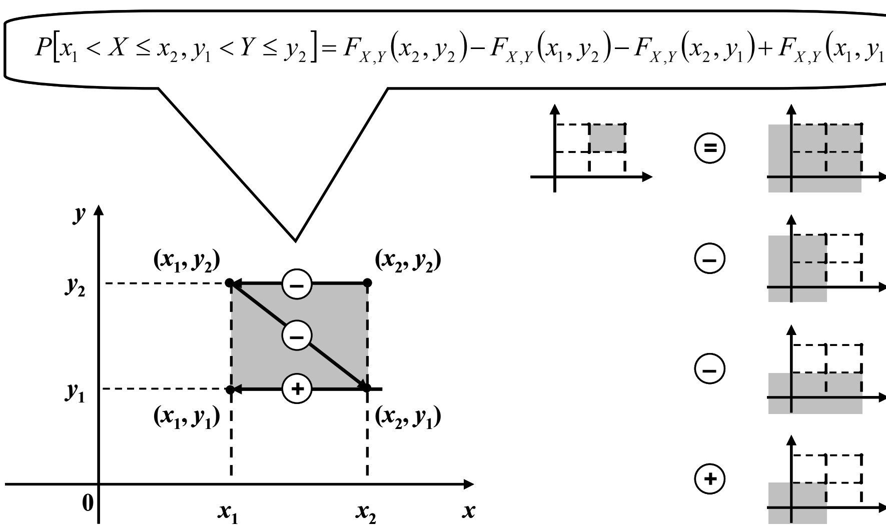
Si $F_{X,Y}(x, y)$ es la f.d.a. conjunta de (X, Y), entonces:

i) $F_{X,Y}(-\infty, y) = 0$, $F_{X,Y}(x, -\infty) = 0$ y $F_{X,Y}(\infty, \infty) = 1$

ii) Si $x_1 < x_2$ y $y_1 < y_2$, entonces (ver figura 4.1):
$$P[x_1 < X \le x_2, y_1 < Y \le y_2] = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1)$$

iii) $F_{X,Y}(x, y)$ es continua por la derecha en cada argumento, es decir:
$$\lim_{h\to 0} F_{X,Y}(x+h, y) = \lim_{h\to 0} F_{X,Y}(x, y+h) = F_{X,Y}(x, y)$$

Figura 4.1



Cálculo de $f_{X,Y}(x, y)$ vía $F_{X,Y}(x, y)$

Si $F_{X,Y}(x, y)$ es la f.d.a. conjunta de (X, Y), entonces:

i) $f_{X,Y}(x, y) = F_{X,Y}(x, y) - F_{X,Y}(x^-, y) - F_{X,Y}(x, y^-) + F_{X,Y}(x^-, y^-)$ si X y Y son variables aleatorias Discretas.

ii) $f_{X,Y}(x, y) = \dfrac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$ si X y Y son variables aleatorias Continuas.

Cálculo de $F_X(x)$ y $F_Y(y)$ vía $F_{X,Y}(x, y)$

Si $F_{X,Y}(x, y)$ es la f.d.a. conjunta de (X, Y), entonces:

i) $F_X(x) = F_{X,Y}(x, \infty)$

ii) $F_Y(y) = F_{X,Y}(\infty, y)$

La f.d.a. conjunta no es tan útil en el caso multivariado como en el univariado.

Modelos Paramétricos Multivariados

En el caso multivariado también existen Distribuciones de Probabilidad caracterizadas por parámetros que aparecen con frecuencia en algunas aplicaciones. Entre los principales modelos paramétricos multivariados están:

- *Distribución Multinomial* en el caso Discreto (ver Tarea 4); y
- *Distribución Normal Multivariada*, en el caso Continuo (ver Tema 5).

4.2. Funciones de probabilidad condicionales

Def. Funciones de masa / densidad de probabilidad condicionales

Sean X y Y variables aleatorias con función de masa / densidad conjunta $f_{X,Y}(x, y)$ y función de masa / densidad de probabilidad marginales $f_X(x)$ y $f_Y(y)$.

i) La función de masa / densidad condicional X dado $\{Y = y\}$ se define como $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$, si $f_Y(y) > 0$; y no se define si $f_Y(y) = 0$.

ii) La función de masa / densidad condicional de Y dado $\{X = x\}$, se define como $f_{Y|X}(y|x) = \dfrac{f_{X,Y}(x,y)}{f_X(x)}$, si $f_X(x) > 0$; y no se define si $f_X(x) = 0$.

Cada posible valor de X define una f.m.p. condicional distinta para Y. Por ejemplo, si X y Y son variables aleatorias discretas y X tiene soporte $\{x_1, x_2, \ldots, x_m\}$, entonces hay m f.m.p condicionales para Y.

Esperanza y Varianza Condicional

$\boldsymbol{f_{X|Y}(x|y)}$ es una f.m.p. (caso discreto) o f.d.p. (caso continuo) **univariada** y es posible calcular su esperanza y varianza. $\boldsymbol{E[X \mid Y = y]}$ es la **esperanza condicional** de X dado $\{Y = y\}$, y $\boldsymbol{Var[X \mid Y = y]}$ es la **varianza condicional** de X dado $\{Y = y\}$.

Esperanza y varianza condicionales de X dado $\{Y = y\}$, son **funciones de *y***, es decir, $E[X \mid Y = y] = g_1(y)$ y $Var[X \mid Y = y] = g_2(y)$. Es posible graficar $g_1(y)$ y $g_2(y)$ para los posibles valores de y.

Ejercicios $\boxed{\text{E119}}$ y $\boxed{\text{E120}}$.

Sin especificar un valor particular de Y, $E[X \mid Y]$ y $Var[X \mid Y]$ son **variables aleatorias**, y es posible calcular su esperanza y varianza. El cálculo de esperanza y varianza de la variable aleatoria X se puede realizar a través de esperanza y varianza condicional como se muestra en la siguiente proposición.

Proposición.

i) $E[X] = E[E[X \mid Y]]$
ii) $Var[X] = E[Var[X|Y]] + Var[E[X|Y]]$

Estas propiedades son de gran utilidad en el estudio de modelos de probabilidad más sofisticados como los *Procesos Estocásticos*. (v.gr., Proceso Poisson, Cadenas de Markov, Movimiento Browniano). También se utilizan en *Estadística Bayesiana* al suponer que los parámetros de una distribución de probabilidades son variables aleatorias.

4.3. Variables aleatorias independientes

Def. Variables aleatorias independientes.

Se dice que las variables aleatorias X y Y son independientes $\Leftrightarrow$ $f_{X,Y}(x,y) = f_X(x) f_Y(y)$

Es decir, sin importar si X y Y son variables discretas o continuas, X y Y son independientes si y sólo si la conjunta es igual al producto de las marginales.

Se dice que X y Y son **variables aleatorias dependientes** si no son independientes.

Proposición.

Si X y Y son variables aleatorias independientes, entonces:
i) $f_{X|Y}(x \mid y) = f_X(x)$
ii) $f_{Y|X}(y \mid x) = f_Y(y)$

Demostración: (i) Si X y Y son independientes entonces $f_{X,Y}(x,y) = f_X(x) f_Y(y)$, y por definición de función de masa / densidad de probabilidad condicional se obtiene que $f_{X|Y}(x \mid y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)} = \dfrac{f_X(x) f_Y(y)}{f_Y(y)} = f_X(x)$. (ii) Se demuestra por analogía.

La proposición $f_{X|Y}(x \mid y) = f_X(x)$ establece que la información de la variable aleatoria Y no modifica el comportamiento probabilístico de la variable aleatoria X y viceversa.

Teorema. Independencia vía factorización.

Las variables aleatorias X y Y son independientes sí y sólo si existen funciones h y g tales que $f_{X,Y}(x, y) = h(x)g(y)$ para $x \in \mathbf{R}$ y $y \in \mathbf{R}$.

Al aplicar este Teorema es importante asegurarse de la **independencia de los dominios**, es decir, que el dominio de la función $h(x)$ no dependa de y y que el dominio de la función $g(y)$ no dependa de x.

Def. Variables aleatorias independientes (generalización).

Se dice que las variables aleatorias $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes si $f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$

Las variables aleatorias conjuntas $X_1, X_2, \ldots, X_n$ suelen representarse a través del vector $\underline{X} = (X_1, X_2, \ldots, X_n)$, de modo que si son independientes $f_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$.

Def. Variables aleatorias idénticamente distribuidas. Se dice que X_i y X_j, $i \neq j$, son variables aleatorias idénticamente distribuidas si su distribución de probabilidad es la misma. Por ejemplo, si $X_i \sim Po(\lambda)$, $i = 1, 2$, entonces X_1 y X_2 son idénticamente distribuidas. Note que aunque X_1 y X_2 son idénticamente distribuidas no significa que sus valores sean iguales simultáneamente (puede ocurrir, pero no necesariamente).

Def. Muestra aleatoria. Es un conjunto de variables aleatorias $X_1, X_2, \ldots, X_n$, $n \in \mathbf{Z}^+$, independientes e idénticamente distribuidas (*iid*).

Ejercicios $\boxed{\text{E121}}$, $\boxed{\text{E122}}$, $\boxed{\text{E123}}$ y $\boxed{\text{E124}}$.

4.4. Valor esperado de una transformación de variables aleatorias

Def. Valor esperado de una transformación de variables aleatorias.

Sean X y Y variables aleatorias con función de masa / densidad de probabilidad conjunta $f_{X,Y}(x, y)$. Si g: $\mathbf{R}^2 \to \mathbf{R}$ es una transformación de las variables aleatorias X y Y, entonces su valor esperado se define como:

i) $E[g(X,Y)] = \sum_x \sum_y g(x,y) f_{X,Y}(x,y)$ si X y Y son Discretas.

ii) $E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dydx$ si X y Y son Continuas.

Son ejemplos de transformaciones $g(X, Y)$: $X + Y$, XY, $\dfrac{X}{Y}$, $(X - a)(Y - b)$, $e^{sX + tY}$, etc.

***E*[*X*] vía $f_{X,Y}(x,y)$.** Si $g(X, Y) = X$, y suponiendo que X y Y son continuas, entonces

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dydx = \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx = \int_{-\infty}^{\infty} x f_X(x)dx = E[X],$$

es decir, la esperanza de X se puede calcular a partir de la densidad conjunta, pero en su cálculo aparece siempre la densidad marginal. Lo mismo ocurre en el caso discreto.

Ejercicio $\boxed{\text{E125}}$.

4.5. Momentos conjuntos, covarianza y coeficiente de correlación

Def. Momentos conjuntos.

Sean X y Y variables aleatorias con $E[X] = \mu_X$ y $E[Y] = \mu_Y$. Si $j, k \in \mathbf{Z}^+$, entonces:

i) $E\left[X^j Y^k\right]$ es el momento conjunto de (X, Y) de orden (j, k).

ii) $E\left[(X - \mu_X)^j (Y - \mu_Y)^k\right]$ es el momento central conjunto de (X, Y) de orden (j, k).

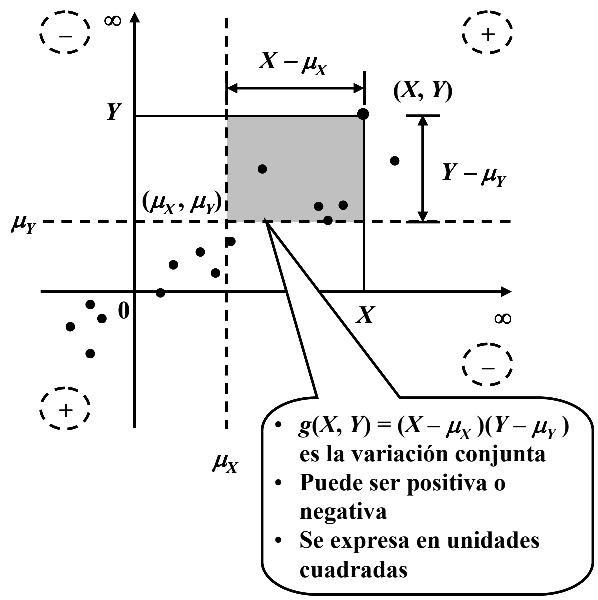
Note cómo $X^j Y^k$ y $(X - \mu_X)^j (Y - \mu_Y)^k$ son transformaciones de las variables X y Y, de modo que el cálculo de su valor esperado se realiza a través de $f_{X,Y}(x, y)$. Los momentos conjuntos por sí solos no tienen ninguna aplicación, sin embargo, aparecen en el cálculo de otras medidas poblacionales conjuntas de interés como la Covarianza.

Def. Covarianza.

Si X y Y son variables aleatorias con $E[X] = \mu_X$ y $E[Y] = \mu_Y$, entonces la covarianza entre X y Y se define por $Cov[X,Y] = \sigma[X,Y] = E[(X-\mu_X)(Y-\mu_Y)]$

La covarianza es el **momento central de orden (1, 1)**. La transformación $g(X,Y) = (X-\mu_X)(Y-\mu_Y)$, es la **variación conjunta** que hay entre la pareja aleatoria (X, Y) y el punto (μ_X, μ_Y) como se muestra en la figura 4.2.

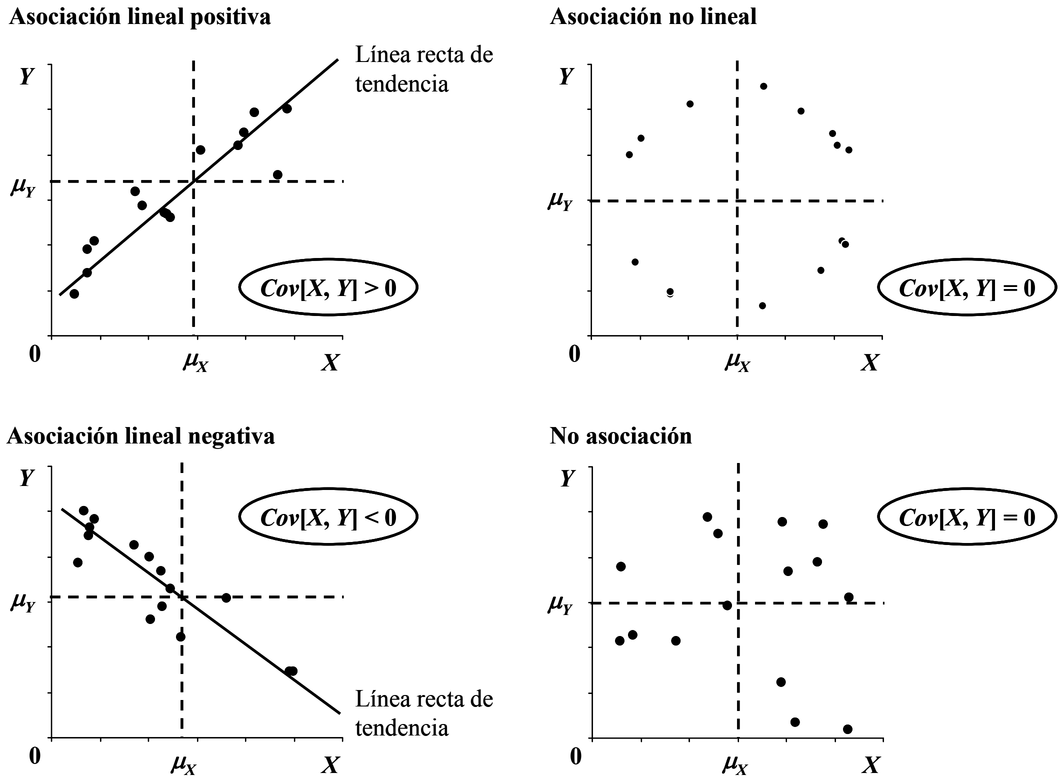
Figura 4.2



La covarianza es una medida de la variabilidad conjunta de X y Y. Como se expresa en unidades cuadradas sólo se interpreta su signo (positivo, negativo o cero), no su valor.

En la medida en que las realizaciones (x_i, y_i) de (X, Y), i = 1, 2,…., se concentran alrededor una línea recta que pasa por (μ_X, μ_Y), la covarianza incorpora más variaciones conjuntas positivas si la recta tiene pendiente positiva, o negativas si la recta tiene pendiente negativa. Consecuentemente, la **covarianza identifica asociación lineal** entre dos variables aleatorias, como se muestra en la Figura 4.3.

Figura 4.3



Interpretación de la covarianza

Si X y Y son variables aleatorias entonces:
i) $Cov[X, Y] < 0 \Rightarrow$ Asociación lineal negativa entre X y Y.
ii) $Cov[X, Y] > 0 \Rightarrow$ Asociación lineal positiva entre X y Y.
iii) $Cov[X, Y] = 0 \Rightarrow$ No asociación o asociación no lineal entre X y Y.

El cálculo de la covarianza mediante su definición puede ser complicado. Para facilitar su cálculo es posible considerar el siguiente Teorema.

Teorema. Cálculo de la covarianza.

$$Cov[X,Y] = E[XY] - E[X]E[Y]$$

Note cómo la **covarianza**:
- Es una **extensión de la definición de varianza**, de hecho,
$$Cov[X,X] = E[(X - E[X])(X - E[X])] = E[(X - E[X])^2] = Var[X].$$
- Es **simétrica**, es decir,
$$Cov[X,Y] = E[(X - E[X])(Y - E[Y])] = E[(Y - E[Y])(X - E[X])] = Cov[Y,X]$$

Con la finalidad de determinar si la asociación lineal identificada por la covarianza es fuerte o débil, se define otra **medida de asociación** denominada *coeficiente de correlación lineal*.

Def. Coeficiente de correlación lineal.

Si X y Y son variables aleatorias entonces su coeficiente de correlación lineal se define por $Corr[X,Y] = \rho_{X,Y} = \dfrac{Cov[X,Y]}{\sigma[X]\sigma[Y]}$

Note cómo el coeficiente de correlación lineal es **adimensional**, de hecho, se puede demostrar que es una cantidad entre –1 y 1 (Corolario de la Desigualdad de Cauchy-Schwarz).

Proposición.

$-1 \le \rho_{X,Y} \le 1$

Interpretación del coeficiente de correlación lineal.

Si X y Y son variables aleatorias entonces:

i) $\rho_{X,Y} \to -1 \Rightarrow$ Alto grado de asociación lineal negativa entre X y Y.
ii) $\rho_{X,Y} \to 0 \Rightarrow$ Bajo grado de asoc. lineal negativa (0^-) o positiva (0^+) entre X y Y.
iii) $\rho_{X,Y} \to 1 \Rightarrow$ Alto grado de asociación lineal positiva entre X y Y.

Def. Variables aleatorias no correlacionadas. Se dice que las variables aleatorias X y Y son no correlacionadas si $Cov[X, Y] = 0$ o $\rho_{X,Y} = 0$.

Teorema. Independencia implica no correlación.

Si X y Y son variables aleatorias independientes entonces $Cov[X, Y] = 0$

Importante: independencia implica no correlación, sin embargo, *no correlación no necesariamente implica independencia*. Puede ocurrir que $Cov[X, Y] = 0$ pero ser X y Y variables aleatorias dependientes (ver T4 II 4.17).

Al considerar n variables aleatorias, X_1, X_2,…, X_n, es posible calcular covarianza y coeficiente de correlación para cada pareja de variables aleatorias. Sean

$$\sigma_{ij} = Cov\left[X_i, X_j\right] \quad \text{y} \quad \rho_{ij} = Corr\left[X_i, X_j\right]$$

para $i, j = 1, 2,\ldots, n$, entonces, es posible definir las siguientes matrices:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix} \quad \text{y} \quad \mathrm{P} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{pmatrix}$$

Σ es la **matriz de varianza-covarianza** y tiene las siguientes características:

- Es simétrica, ya que $\sigma_{ij} = Cov\left[X_i, X_j\right] = Cov[X_j, X_i] = \sigma_{ji}$; y
- En la diagonal principal aparecen las varianzas ya que $\sigma_{ii} = Cov\left[X_i, X_i\right] = Var\left[X_i\right]$.

P es la **matriz de correlaciones** y tiene las siguientes características:

- Es simétrica, ya que $\rho_{ij} = Corr\left[X_i, X_j\right] = Corr[X_j, X_i] = \rho_{ji}$; y
- En la diagonal principal sólo aparecen unos ya que:
  $\rho_{ii} = Corr\left[X_i, X_i\right] = \dfrac{Cov[X_i, X_i]}{\sigma\left[X_i\right]\sigma\left[X_i\right]} = \dfrac{Var\left[X_i\right]}{Var\left[X_i\right]} = 1$.

Ejercicios $\boxed{\text{E126}}$ y $\boxed{\text{E127}}$.

4.6. Propiedades de Esperanza, Varianza y Covarianza

Teorema. Esperanza y Varianza de una combinación lineal.

Si X y Y son variables aleatorias y $\alpha, \beta \in \mathbf{R}$, entonces:

i) $E\left[\alpha X \pm \beta Y\right] = \alpha E\left[X\right] \pm \beta E\left[Y\right]$ (Operador Lineal)

ii) $Var\left[\alpha X \pm \beta Y\right] = \alpha^2 Var\left[X\right] + \beta^2 Var\left[Y\right] \pm 2\alpha\beta Cov\left[X, Y\right]$

Corolario. Varianza una combinación lineal de v.a.'s independientes.

Si X y Y son variables aleatorias independientes y $\alpha, \beta \in \mathbf{R}$, entonces:
$Var\left[\alpha X \pm \beta Y\right] = \alpha^2 Var\left[X\right] + \beta^2 Var\left[Y\right]$

Ejercicios $\boxed{\text{E128}}$, $\boxed{\text{E129}}$ y $\boxed{\text{E130}}$.

Teorema. Esperanza como Operador Lineal.

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias, $\underline{X} = (X_1, X_2, \ldots, X_n)$, $\alpha_i \in \mathbf{R}$ y g_i es una función tal que g_i: $\mathbf{R}^n \to \mathbf{R}$, $i = 1, 2, \ldots, m$, entonces $E\left[\sum_{i=1}^{m} \alpha_i g_i(\underline{X})\right] = \sum_{i=1}^{m} \alpha_i E[g_i(\underline{X})]$

La aplicación de este Teorema es muy intuitiva. Permite, por ejemplo, determinar que $E\left[(3X - Ye^X)^2\right] = E\left[9X^2 - 6XYe^X + Y^2 e^{2X}\right] = 9E\left[X^2\right] - 6E\left[XYe^X\right] + E\left[Y^2 e^{2X}\right]$.

Teorema. Esperanza del producto de transformaciones de v.a.'s independientes.

Si X y Y son variables aleatorias independientes, y g y h son funciones arbitrarias de $\mathbf{R}$ en $\mathbf{R}$, entonces $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$

En el caso particular en que g y h son funciones identidad, es decir, $g(X) = X$ y $h(y) = Y$, entonces $E[XY] = E[X]E[Y]$.

Ejercicios $\boxed{\text{E131}}$, $\boxed{\text{E132}}$ y $\boxed{\text{E133}}$.

Teorema. Propiedades de la Covarianza.

Si $X_1, X_2, \ldots, X_m$; $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias y α_i , $\beta_j \in \mathbf{R}$, para $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$; entonces:

i) $Cov[\alpha_1 X_1 + \beta_1, \alpha_2 X_2 + \beta_2] = \alpha_1 \alpha_2 Cov[X_1, X_2]$

ii) $Cov\left[\sum_{i=1}^{m} \alpha_i X_i, \sum_{j=1}^{n} \beta_j Y_j\right] = \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i \beta_j Cov[X_i, Y_j]$ (Operador Bilineal)

Teorema. Esperanza y Varianza de una combinación lineal (generalización).

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias y $\alpha_i \in \mathbf{R}$, $i = 1, 2, \ldots, n$, entonces:

i) $E\left[\sum_{i=1}^{n} \alpha_i X_i\right] = \sum_{i=1}^{n} \alpha_i E[X_i]$

ii) $Var\left[\sum_{i=1}^{n} \alpha_i X_i\right] = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j Cov[X_i, X_j] = \sum_{i=1}^{n} \alpha_i^2 Var[X_i] + 2\sum_{i<j} \alpha_i \alpha_j Cov[X_i, X_j]$

Ejercicio $\boxed{\text{E134}}$.

Teorema. Coeficiente de correlación de una transformación lineal.

$\left|\rho_{X,Y}\right| = 1 \Leftrightarrow$ existen α, $\beta \in$ **R** tales que $Y = \alpha + \beta X$ (Ver Tarea 4)

4.7. Función generadora de momentos conjunta

Def. Función generadora de momentos conjuntos.

Sean X y Y variables aleatorias la función generadora de momentos conjuntos de (X, Y) se define por $M_{X,Y}(s,t) = E\left[e^{sX+tY}\right]$ para (s, t) tales que $|s| < b$ y $|t| < b$ para $b > 0$.

Es decir, la función generadora de momentos conjuntos es el valor esperado de una función de variables aleatorias $g(X,Y) = e^{sX+tY}$. La condición $|s| < b$ y $|t| < b$ para $b > 0$ se establece para asegurar que las derivadas parciales de $M_{X,Y}(s, t)$ puedan ser evaluadas en (0, 0) o en el límite cuando (s, t) tiende a (0, 0). Como su nombre lo indica, $M_{X,Y}(s, t)$ se utiliza para calcular momentos conjuntos de (X, Y) gracias al siguiente Teorema.

Teorema. Generación de momentos conjuntos.

$$E[X^j Y^k] = \left.\frac{\partial^{j+k}}{\partial s^j \partial t^k} M_{X,Y}(s,t)\right|_{(s,t)=(0,0)}$$

Proposición.

Si X y Y son variables aleatorias con f.g.m. conjuntos $M_{X,Y}(s,t)$, entonces:

i) $M_{X,Y}(0,t) = M_Y(t)$ y $M_{X,Y}(s,0) = M_X(s)$

ii) X y Y son independientes $\Leftrightarrow M_{X,Y}(s,t) = M_X(s) M_Y(t)$

Ejercicio $\boxed{\text{E135}}$.

4.8. Distribución de transformaciones de variables aleatorias

4.8.1. Distribución de transformaciones de variables aleatorias discretas

Proposición.

Si X y Y son variables aleatorias discretas con f.m.p.c. $f_{X,Y}(x, y)$ y $W = g(X, Y)$, entonces:

$$f_W(w) = P[W = w] = P[g(X,Y) = w] = \sum\sum_{\{(x,y):\, g(x,y)=w\}} f_{X,Y}(x,y)$$

Por sus aplicaciones, es de particular interés obtener la distribución de la transformación $S = g(X, Y) = X + Y$ cuando X y Y son variables aleatorias independientes, es decir, la distribución de la **suma de variables aleatorias independientes**.

Ejercicio $\boxed{\text{E136}}$.

4.8.2. Distribución de transformaciones de variables aleatorias continuas

Para determinar la distribución de $Z = g(X, Y)$, una función de variables aleatorias continuas, es posible considerar **tres métodos**:

- Método de la función de distribución acumulada.
- Método de la función generadora de momentos; y
- Teorema de cambio de variable para funciones vectoriales.

MÉTODO DE LA FUNCIÓN DE DISTRIBUCIÓN ACUMULADA

Considerando que X y Y son variables aleatorias continuas con función de densidad de probabilidad conjunta $f_{X,Y}(x, y)$, este método consiste en calcular:

$$F_Z(z) = P[Z \le z] = P[g(X,Y) \le z] = \int\int_A f_{X,Y}(x,y)\,dy\,dx \text{ , donde } A = \{(x,y) : g(x,y) \le z\}.$$

Ejercicios $\boxed{\text{E137}}$, $\boxed{\text{E138}}$ y $\boxed{\text{E139}}$.

MÉTODO DE LA FUNCIÓN GENERADORA DE MOMENTOS

Este método es aplicable a **variables aleatorias discretas o continuas** y permite calcular la distribución de $S = g(X, Y) = X + Y$ a través del siguiente Teorema.

Teorema.

Si X y Y son variables aleatorias independientes entonces $M_{X+Y}(t) = M_X(t)M_Y(t)$

Demostración: Sabemos que si X y Y son independientes entonces también los son $g(X)$ y $h(Y)$ para g y h funciones arbitrarias. Sean $g(X) = e^{tX}$ y $h(Y) = e^{tY}$, entonces, $M_{X+Y}(t) = E\left[e^{t(X+Y)}\right] = E\left[e^{tX}\, e^{tY}\right] = E\left[e^{tX}\right]E\left[e^{tY}\right] = M_X(t)M_Y(t)$.

Ejercicios $\boxed{\text{E140}}$, $\boxed{\text{E141}}$ y $\boxed{\text{E142}}$.

TEOREMA DE CAMBIO DE VARIABLE PARA FUNCIONES VECTORIALES.

Teorema.

Sean X_1 y X_2 variables aleatorias continuas con función de densidad conjunta $f_{X_1,X_2}(x_1, x_2)$ y soporte $S \subset \mathbf{R}^2$; $Y_1 = g_1(X_1, X_2)$ y $Y_2 = g_2(X_1, X_2)$ funciones uno a uno de X_1 y X_2; y T: $\mathbf{R}^2 \to \mathbf{R}^2$ la transformación $T(x_1, x_2) = (g_1(x_1, x_2), g_2(x_1, x_2)) = (y_1, y_2)$.

i) Si la solución única de las ecuaciones $y_1 = g_1(x_1, x_2)$ y $y_2 = g_2(x_1, x_2)$ está dada por $x_1 = h_1(y_1, y_2)$ y $x_2 = h_2(y_1, y_2)$; y

ii) Si las funciones y_1 y y_2 tienen derivadas parciales continuas y $J(x_1, x_2) \neq 0$ para todo (x_1, x_2), donde $J(x_1, x_2) = \begin{vmatrix} \dfrac{\partial g_1}{\partial x_1} & \dfrac{\partial g_1}{\partial x_2} \\ \dfrac{\partial g_2}{\partial x_1} & \dfrac{\partial g_2}{x_2} \end{vmatrix} = \left(\dfrac{\partial g_1}{\partial x_1}\right)\left(\dfrac{\partial g_2}{x_2}\right) - \left(\dfrac{\partial g_1}{\partial x_2}\right)\left(\dfrac{\partial g_2}{\partial x_1}\right)$ es el determinante Jacobiano de g_1 y g_2;

entonces $f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2)\left|J(x_1, x_2)\right|^{-1}$ con soporte $T(S) \subset \mathbf{R}^2$.

Este teorema es la generalización del Teorema de la Distribución de la Transformación Monótona (DTM), y surge de la aplicación directa del Teorema de Cambio de Variable del Cálculo Vectorial.

Este Teorema requiere considerar siempre dos transformaciones de las variables aleatorias X_1 y X_2. Aunque la distribución de interés sea sólo $Y_1 = g_1(X_1, X_2) = X_1 + X_2$, es común definir una *transformación dummy*, por ejemplo $Y_2 = g_2(X_1, X_2) = X_1$, para aplicar el Teorema anterior.

Ejercicios $\boxed{\text{E143}}$ y $\boxed{\text{E144}}$.

4.8.3. Distribución de suma de variables aleatorias independientes

A continuación se resumen los principales resultados al determinar la distribución de la suma de variables aleatorias independientes (v.a.'s ind.).

Teorema. Suma de Distribuciones Binomiales independientes.

Si $X \sim Bin(m, p)$ y $Y \sim Bin(n, p)$ son v.a.'s ind., entonces $X + Y \sim Bin(m + n, p)$

Teorema. Suma de Distribuciones Poisson independientes.

Si $X \sim Po(\lambda_1)$ y $Y \sim Po(\lambda_2)$ son v.a.'s ind., entonces $X + Y \sim Po(\lambda_1 + \lambda_2)$

Teorema. Suma de Distribuciones Gamma independientes.

Si $X \sim Gamma(\alpha_1, \beta)$ y $Y \sim Gamma(\alpha_2, \beta)$ son variables aleatorias independientes, entonces $X + Y \sim Gamma(\alpha_1 + \alpha_2 , \beta)$

Corolario. Suma de Distribuciones Exponenciales y Ji-Cuadradas independientes.

Sean $X_1, X_2,\ldots, X_n$ variables aleatorias independientes:

i) Si $X_i \sim Exp(\beta)$, $i = 1, 2,\ldots, n$, entonces $\sum_{i=1}^{n} X_i \sim Gamma(n, \beta)$

ii) Si $X_i \sim \chi^2(\nu_i)$, $i = 1, 2,\ldots, n$, entonces $\sum_{i=1}^{n} X_i \sim \chi^2\left(\sum_{i=1}^{n} \nu_i\right)$

Teorema. Transformaciones lineales de Distribuciones Normales.

Si $X_1, X_2,\ldots, X_n$ son variables aleatorias independientes con $X_i \sim N(\mu_i, \sigma_i^2)$ y $\alpha_i, \in \mathbf{R}$, $i = 1, 2,\ldots, n$; entonces $Y = \sum_{i=1}^{n} \alpha_i X_i \sim N(\mu, \sigma^2)$ donde $\mu = \sum_{i=1}^{n} \alpha_i \mu_i$ y $\sigma^2 = \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2$.

Corolario. Suma de Distribuciones Normales.

Si $X_1, X_2,\ldots, X_n$ son variables aleatorias independientes con $X_i \sim N(\mu_i, \sigma_i^2)$, para $i = 1, 2,\ldots, n$; entonces $S = \sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$

5. Distribución Normal Multivariada

5.1. Funciones de densidad conjunta, marginales y condicionales de la Distribución Normal Bivariada.

Def. Distribución Normal Bivariada

Se dice que $(X_1, X_2) \sim N\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\right)$, $\mu_i \in \mathbf{R}$, $\sigma_i > 0$, $i = 1, 2$; y $-1 < \rho < 1$ si su función de densidad de probabilidad conjunta está dada por:

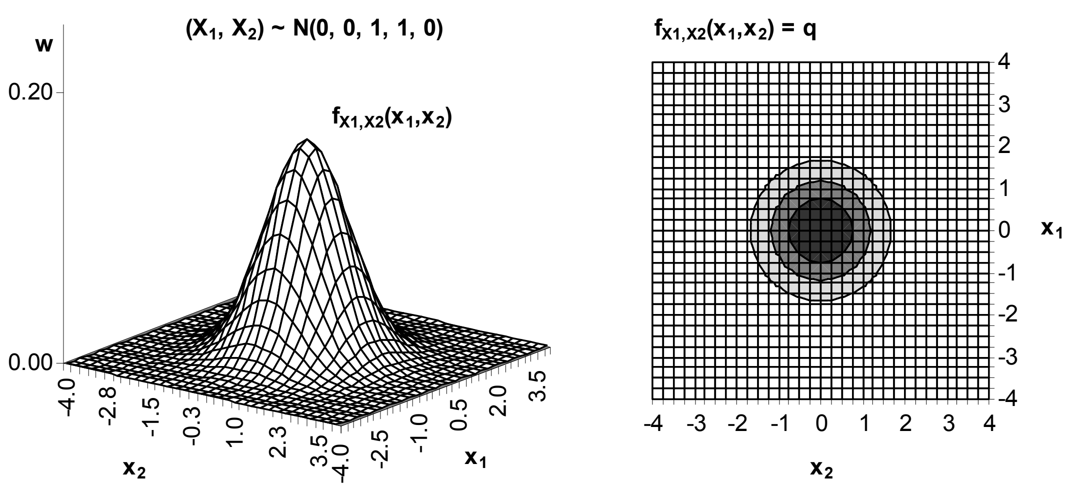
$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\left(1-\rho^2\right)} Q(x_1, x_2)\right\}, \quad -\infty < x_i < \infty, \quad i = 1, 2;$$

donde $$Q(x_1, x_2) = \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2.$$

Def. Distribución Normal Bivariada Estándar. Se dice que (Z_1, Z_2) tienen una Distribución Normal Bivariada Estándar si $(Z_1, Z_2) \sim N(0, 0, 1, 1, 0)$.

La figura 5.1 muestra la gráfica y las curvas de nivel de la Distribución Normal Bivariada Estándar. La gráfica de $f_{X_1,X_2}(x_1, x_2)$ tiene forma de una campana que alcanza su máximo en $(x_1, x_2) = (\mu_1, \mu_2) = (0, 0)$. Las curvas de nivel en este caso son circunferencias con centro en $(\mu_1, \mu_2) = (0, 0)$.

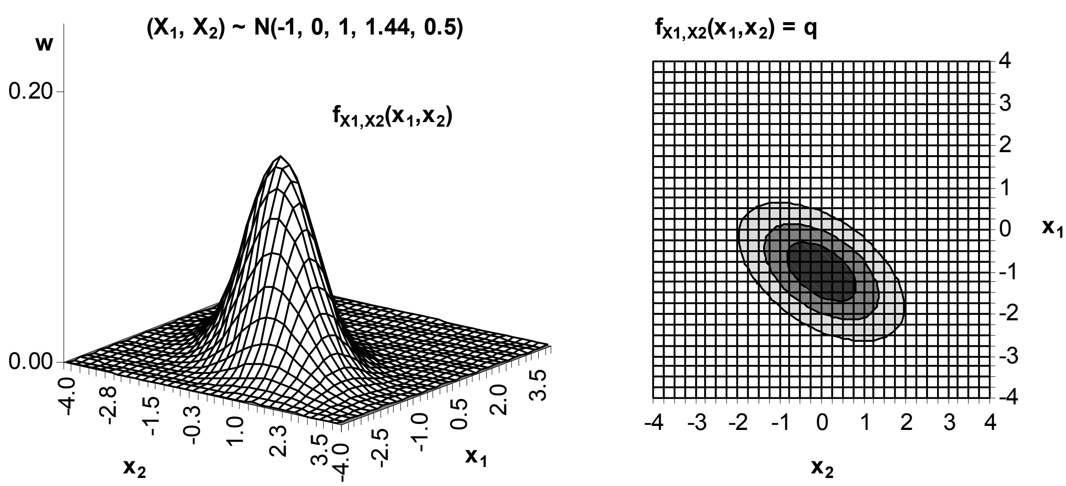
Figura 5.1



Efecto de los parámetros. La figura 5.2 muestra la gráfica y las curvas de nivel de $f_{X_1,X_2}(x_1,x_2)$ para el caso en que $(X_1, X_2) \sim N(-1, 0, 1, 1.44, 0.5)$. En este caso las curvas de nivel se convierten en elipses con entro en $(\mu_1, \mu_2) = (-1, 0)$. Note cómo:

- Los parámetros $\boldsymbol{\mu_1}$ y $\boldsymbol{\mu_2}$ trasladan el centro de la densidad conjunta.
- Por su parte, los parámetros $\boldsymbol{\sigma_1}$ y $\boldsymbol{\sigma_2}$ determinan la longitud de los ejes de las elipses (eje mayor y eje menor), reflejo de la mayor o menor concentración de probabilidad alrededor de (μ_1, μ_2).
- Finalmente, el efecto del parámetro $\boldsymbol{\rho}$ es una rotación de los ejes de las elipses, reflejo de una mayor o menor asociación lineal entre las variables X_1 y X_2.

Figura 5.2



Para entender mejor el impacto de los parámetros de la Distribución Normal Bivariada es posible consultar algunos modelos animados que aparecen en Internet, por ejemplo: http://demonstrations.wolfram.com/TheBivariateNormalDistribution/.

Proposición.

Si $(X_1, X_2) \sim N\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\right)$ entonces:

i) $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$

ii) $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$

Este resultado indica que las **Distribuciones Marginales** de la Distribución Normal Bivariada son Distribuciones Normales univariadas.

Con base en el resultado anterior y considerando la función de densidad conjunta de la Distribución Normal Bivariada es posible determinar sus propiedades.

Propiedades de la Distribución Normal Bivariada.

Si $(X_1, X_2) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ entonces:

i) $E[X_1] = \mu_1$ y $E[X_2] = \mu_2$

ii) $Var[X_1] = \sigma_1^2$ y $Var[X_2] = \sigma_2^2$

iii) $Corr[X_1, X_2] = \rho$ y $Cov[X_1, X_2] = \sigma_{12} = \rho\sigma_1\sigma_2$

Combinando la función de densidad conjunta y las funciones de densidad marginales de la Distribución Normal Bivariada es posible calcular las funciones de densidad condicionales de $\{X_1 | X_2 = x_2\}$ y de $\{X_2 | X_1 = x_1\}$ para demostrar que las **Distribuciones Condicionales de la Distribución Normal Bivariada** también son Distribuciones Normales univariadas, como se establece en la siguiente proposición.

Proposición.

Si $(X_1, X_2) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ entonces:

i) $\{X_1 | X_2 = x_2\} \sim N\left(\mu_1 + \rho\dfrac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1-\rho^2)\right)$

ii) $\{X_2 | X_1 = x_1\} \sim N\left(\mu_2 + \rho\dfrac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1-\rho^2)\right)$

De este resultado se concluye que (ver Formulario) para $\{X_1 | X_2 = x_2\}$:

$$E[X_1 | X_2 = x_2] = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2) \quad \text{y} \quad Var[X_1 | X_2 = x_2] = \sigma_1^2(1-\rho^2);$$

En forma análoga (ver Formulario), para $\{X_2 | X_1 = x_1\}$:

$$E[X_2 | X_1 = x_1] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1) \quad \text{y} \quad Var[X_2 | X_1 = x_1] = \sigma_2^2(1-\rho^2).$$

Ejercicio $\boxed{\text{E145}}$.

5.2. Función generadora de momentos de la Distribución Normal Bivariada.

Proposición.

Si $(X_1, X_2) \sim N\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\right)$ entonces:

$$M_{X_1,X_2}(t_1, t_2) = \exp\left\{\mu_1 t_1 + \mu_2 t_2 + \frac{1}{2}\left(\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2\right)\right\},\ |t_i| > b,\ b > 0,\ i = 1, 2.$$

Este resultado se puede demostrar directamente considerando la definición de función generadora de momentos conjuntos.

5.3. Independencia de variables aleatorias con Distribución Normal Bivariada.

Proposición.

Si $(X_1, X_2) \sim N\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\right)$ y $\rho = 0$ entonces X_1 y X_2 son independientes.

Importante: En general, independencia implica $\rho = 0$; sin embargo, $\rho = 0$ no necesariamente implica independencia. **En el caso particular de la Distribución Normal Bivariada $\rho = 0$ si implica independencia**.

Ejercicio $\boxed{\text{E146}}$.

5.4. Combinaciones lineales de variables aleatorias con Distribución Normal Bivariada.

Teorema.

Si $(X_1, X_2) \sim N\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\right)$ y $T = \alpha_1 X_1 + \alpha_2 X_2$, $\alpha_i \in \mathbf{R}$, $i = 1, 2$, es una combinación lineal arbitraria de X_1 y X_2, entonces $T \sim N\left(E[T], Var[T]\right)$, donde:

i) $E[T] = \alpha_1\mu_1 + \alpha_2\mu_2$; y

ii) $Var[T] = \alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 + 2\alpha_1\alpha_2\rho\sigma_1\sigma_2$.

Ejercicio $\boxed{\text{E147}}$.

5.5. Distribución Normal Multivariada.

La Distribución Normal Bivariada se puede generalizar para más de dos variables y su manejo resulta más simple utilizando vectores y matrices.

Def. Distribución Normal Multivariada.

Se dice que $\underline{X} = (X_1, X_2, \ldots, X_n)^t \sim N_n(\underline{\mu}, \Sigma)$, es decir, el vector aleatorio $\underline{X}$ tiene una Distribución Normal Multivariada de dimensión n con vector de medias $\underline{\mu}$ y matriz de varianzas-covarianzas Σ definidos por

$$\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \quad \text{y} \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix},$$

si su función de densidad de probabilidad conjunta está dada por:

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\det(\Sigma)|} \exp\left\{ -\frac{1}{2} (\underline{x} - \underline{\mu})^t \Sigma^{-1} (\underline{x} - \underline{\mu}) \right\}, \quad \underline{x} \in \mathbf{R}^n$$

donde $\det(\Sigma)$ y Σ^{-1} son el determinante y la inversa de la matriz Σ, respectivamente.

Las principales propiedades de la Distribución Normal Bivariada se pueden generalizar para el caso de la Distribución Normal Multivariada.

Teorema.

Si $\underline{X} = (X_1, X_2, \ldots, X_n)^t \sim N_n(\underline{\mu}, \Sigma)$ y $\mathrm{A} \in \mathbf{R}^{m \times n}$, entonces $\mathrm{A}\underline{X} \sim N_m(\mathrm{A}\underline{\mu}, \mathrm{A}\Sigma\mathrm{A}^t)$

El Teorema anterior establece que **cualquier transformación lineal de una Distribución Normal Multivariada continua siendo Normal Multivariada**.

Por ejemplo, sea $\underline{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N_3(\underline{\mu}, \Sigma)$ donde $\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}$ y $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{pmatrix}$. Si $\mathrm{A} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$, entonces la transformación A aplicada a $\underline{X}$ por la izquierda intercambia o permuta las variables X_1 y X_2 y elimina a la variable X_3, y se puede pensar que $(X_2, X_1)^t$ debe tener una Distribución Normal Bivariada.

Aplicando el Teorema anterior se observa que:

$$\mathrm{A}\underline{X} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} X_2 \\ X_1 \end{pmatrix}, \quad \mathrm{A}\underline{\mu} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} \mu_2 \\ \mu_1 \end{pmatrix}, \quad \text{y}$$

$$\mathrm{A}\Sigma\mathrm{A}^t = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_1^2 & \sigma_{12} & \sigma_{13} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \sigma_2^2 & \sigma_{21} \\ \sigma_{12} & \sigma_1^2 \end{pmatrix};$$

es decir, $\mathrm{A}\underline{X} = \begin{pmatrix} X_2 \\ X_1 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_2 \\ \mu_1 \end{pmatrix}, \begin{pmatrix} \sigma_2^2 & \sigma_{21} \\ \sigma_{12} & \sigma_1^2 \end{pmatrix}\right)$. Alternativamente, considerando la notación particular de la Normal Bivariada, $(X_2, X_1) \sim N_2\left(\mu_2, \mu_1, \sigma_2^2, \sigma_1^2, \dfrac{\sigma_{12}}{\sigma_1\sigma_2}\right)$.

La transformación $\mathrm{A} \in \mathbf{R}^{m\times n}$ puede ser elegida apropiadamente para concluir que la distribución marginal de la distribución de cualquiera de las variables X_i, $i = 1, 2,\ldots, n$, es Normal univariada y que cualquier subconjunto de variables $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$, $k \le n$, es Normal multivariada de dimensión k.

Corolario. Distribuciones Marginales y Subconjuntos de la Normal Multivariada.

Si $\underline{X} = (X_1, X_2, \ldots, X_n)^t \sim N_n(\underline{\mu}, \Sigma)$ entonces:

i) $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2,\ldots, n$; y

ii) Cualquier subconjunto $(X_{i_1}, X_{i_2}, \ldots, X_{i_k})^t \sim N_k(\underline{\mu}_k, \Sigma_k)$, $k \le n$, donde $\underline{\mu}_k$ y Σ_k corresponden a sus medias y varianzas-covarianzas, respectivamente.

Otra transformación de interés es $\mathrm{A} = (\alpha_1, \alpha_2, \ldots, \alpha_n) = \underline{\alpha}^t \in \mathbf{R}^{1\times n}$ que permite concluir que la combinación lineal $T = \underline{\alpha}^t\,\underline{X} = \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n = \sum_{i=1}^{n} \alpha_i X_i$ tiene una distribución Normal univariada.

El cálculo de $E[T]$ y $Var[T]$ se puede realizar fácilmente a partir de las propiedades de esperanza y varianza:

- $E[T] = E\left[\sum_{i=1}^{n} \alpha_i X_i\right] = \sum_{i=1}^{n} \alpha_i E[X_i] = \sum_{i=1}^{n} \alpha_i \mu_i$
- $Var[T] = Var\left[\sum_{i=1}^{n} \alpha_i X_i\right] = \sum_{i=1}^{n} \alpha_i^2 Var[X_i] + 2\sum_{i<j} \alpha_i \alpha_j Cov[X_i, X_j] = \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2 + 2\sum_{i<j} \alpha_i \alpha_j \sigma_{ij}$

Corolario. Combinaciones lineales de la Distribución Normal Multivariada.

Si $\underline{X} = (X_1, X_2, \ldots, X_n)^t \sim N_n(\underline{\mu}, \Sigma)$ y $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^t$, $\alpha_i, \in$ **R**, i = 1, 2,…, n; entonces $T = \underline{\alpha}^t \, \underline{X} = \sum_{i=1}^{n} \alpha_i X_i \sim N(E[T], Var[T])$ donde:

i) $E[T] = \underline{\alpha}^t \, \underline{\mu} = \sum_{i=1}^{n} \alpha_i \mu_i$; y

ii) $Var[T] = \underline{\alpha}^t \Sigma \underline{\alpha} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \sigma_{ij} = \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2 + 2 \sum_{i<j} \alpha_i \alpha_j \sigma_{ij}$

Finalmente, igual que en el caso de la Normal Bivariada, en el caso de la Normal Multivariada es posible hablar de distribuciones condicionales.

Si $\underline{X} \sim N_n(\underline{\mu}, \Sigma)$ se puede descomponer en 2 subconjuntos arbitrarios $\underline{X}_1$ y $\underline{X}_2$, entonces es posible determinar la distribución condicional $\{\underline{X}_1 | \underline{X}_2 = \underline{x}_2\}$.

Teorema. Distribuciones Condicionales de la Normal Multivariada.

Si $\begin{pmatrix} \underline{X}_1 \\ \underline{X}_2 \end{pmatrix} \sim N_n\left(\begin{pmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$, $\underline{X}_1 \in \mathbf{R}^{k\times 1}$, $\underline{X}_2 \in \mathbf{R}^{(n-k)\times 1}$, $k \le n$, y Σ_{22} es no singular, entonces $\{\underline{X}_1 | \underline{X}_2 = \underline{x}_2\} \sim N_k(\underline{\mu}_{1.2}, \Sigma_{11.2})$ donde:

i) $\underline{\mu}_{1.2} = \underline{\mu}_1 + \Sigma_{12} \Sigma_{22}^{-1} (\underline{x}_2 - \underline{\mu}_2)$; y

ii) $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$

Ejercicio E148 .

Material de Apoyo para el Curso de Probabilidad

David Ruelas Rodríguez, Agosto – Diciembre 2012

E1. Utilice la definición de igualdad de conjuntos para demostrar que $\left(A^c\right)^c = A$.

E2. Suponga que el conjunto universo está dado por $\Omega = \{x \in \mathbf{Z}^+ \mid x < 12\}$ y que se definen los conjuntos $A = \{1, 2, 3, 4, 5\}$, $B = \{2, 4, 6, 8, 10\}$ y $C = \{8, 9, 10\}$.

a) Exprese Ω por extensión. ¿Cuál es su cardinalidad?

b) Determine $A \cup B$ y determine su cardinalidad.

c) Determine $A \cap B$, $A \cap C$ y $B \cap C$.

d) Represente mediante un Diagrama de Venn los conjuntos A, B y C.

e) Se define $D = \{x \in A \mid x \text{ es par}\}$. Calcule D^c y $A - D$. Compruebe mediante el diagrama del inciso anterior que $D = A \cap B$.

E3. Considere los conjuntos del ejercicio E2 y verifique que $(A \cup B)^c = A^c \cap B^c$ (Ley de De Morgan).

E4. **Leyes de De Morgan**. Sean $A_1, A_2, ..., A_n$, $n \in \mathbf{N}$, subconjuntos de Ω. Demuestre que:

a) $$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c$$

b) $$\left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c$$

E5. Considere los siguientes conjuntos de $\Omega = \mathbf{R}$ (la recta real): $A = \{x \in \mathbf{R} \mid x > 0\}$, $B = \{x \in \mathbf{R} \mid x \leq 5\}$, $C = \{x \in \mathbf{R} \mid -10 < x \leq 5\}$ y $D = \{x \in \mathbf{R} \mid |x + 2| \geq 2\}$.

a) Represente cada conjunto en la recta real. ¿Es posible representar estos conjuntos por extensión?

b) ¿Se puede afirmar que $A \subset D$?

c) Calcule $A \cap B$ y $C \cap D$.

d) Calcule $\#(C \cap B^c)$

E6. Sean $\Omega = \mathbf{R}^2$, $M = \{(x, y) : x^2 + y^2 \leq 16\}$ y $N = \{(x, y) : |x - 1| < 2\}$. Represente geométricamente $M \cap N^c$.

E7. Construya el espacio muestral para cada uno de los siguientes experimentos aleatorios. Determine en cada caso si el espacio muestral es discreto o continuo.

a) Se lanza al aire una moneda justa cuyas caras son "Águila" (A) o "Sol" (S).

b) Se pregunta a dos personas si la primera letra de su apellido es vocal o consonante.

c) Se pregunta a tres personas si el día del mes en que nacieron es número par o non.

d) Se observa el comportamiento de 10 acciones de la Bolsa Mexicana de Valores y se registra cuáles de éstas cerraron ayer a la baja y cuáles no.

e) Se observa el comportamiento de 10 acciones de la Bolsa Mexicana de Valores y se registra cuántas de éstas cerraron ayer a la baja.

f) Se observa la temperatura ambiente (en grados centígrados).

g) Se calcula el saldo promedio de las cuentas de ahorro de un banco.

h) Se eligen dos coches al azar y se observa su kilometraje.

i) Se pregunta a una pareja el número de años completos de escuela primaria cursados por cada uno.

E8. Considere el experimento aleatorio en el que se lanza al aire una moneda y se tira un dado. La moneda puede tener resultados "Águila" (A) o "Sol" (S), mientras que el dado puede caer en 1, 2, 3, 4, 5 o 6.

a) Defina el espacio muestral asociado a este experimento por comprensión y por extensión. Determine su cardinalidad.

b) Se definen E, el evento en el que cae "Sol"; y F, el evento en el que caen "Águila" o un número mayor a 3. Determine el evento $G = E - F^c$ y calcule $\#G$.

E9. Considere el experimento aleatorio en el que se lanzan dos dados honestos, cada uno numerado del 1 al 6, uno seguido del otro.

a) Determine el espacio muestral y su cardinalidad.

b) Sean A, el evento en que el primer dado es mayor al segundo; y B, el evento en que ambos dados caen en el mismo valor. Exprese A y B por extensión y por comprensión. Determine sus cardinalidades.

c) Calcule $\#(A \cup B)$ y compruebe que $\#(A \cup B) = \#A + \#B$ por ser eventos mutuamente excluyentes.

d) Sea C, el evento en que la suma de los dados es 7. Determine $\#C$.

E10. Considere el experimento aleatorio y los eventos del ejercicio E9. Bajo el enfoque clásico de la probabilidad:

a) Determine la probabilidad de que el valor del primer dado sea mayor al del segundo.

b) Calcule la probabilidad de que ambos dados caigan en el mismo valor.

c) Determine la probabilidad de que al valor del primer dado sea mayor o igual al del segundo.

d) Demuestre que $P(A \cup B) = P(A) + P(B)$ por ser eventos mutuamente excluyentes.

E11. Suponga que un punto es escogido al azar en el cuadrado unitario, es decir, $\Omega = \{(x,y): 0 \le x \le 1, 0 \le y \le 1\}$. Sea *A* el evento de que el punto está en el triángulo delimitado por las rectas $y = 0$, $x = 1$ y $x = y$; y sea *B* el evento de que el punto está en el rectángulo con vértices en (0,0), (1,0), $\left(1, \frac{1}{2}\right)$ y $\left(0, \frac{1}{2}\right)$. Calcule:

a) $P(A \cup B)$.

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

E12. Considere el experimento aleatorio en que se lanza un volado. Sea *A*, el evento en que sale “Águila”.

a) Realice al menos 1,000 repeticiones de este experimento aleatorio y determine, para cada lanzamiento, la frecuencia relativa del evento *A*.

b) Construya una gráfica que muestre el valor de la frecuencia relativa del evento *A* para cada una de las repeticiones del experimento y determine, bajo el enfoque frecuentista, la probabilidad de obtener águila.

E13. Una urna contiene 3 pelotas con las letras “a”, “m” y “o”. Se extraen de la urna al azar y en forma consecutiva 3 pelotas, formando palabras en el orden que salen. Utilizando diagramas de árbol determine la probabilidad de que la palabra comience y termine en vocal si:

a) Las extracciones se hacen sin reemplazo.

b) Las extracciones se hacen con reemplazo.

c) ¿Cuántos casos totales habría si se agregara una cuarta pelota con la letra “s” y las extracciones se hicieran nuevamente con reemplazo?

E14. Suponga que las placas de los coches se integran por 3 dígitos seguidos de 3 letras correspondientes a un alfabeto de 26 letras. Si se elige una placa al azar:

a) ¿Cuál es la probabilidad de que el número de la placa (integrado por los 3 dígitos) sea par?

b) ¿Cuál es la probabilidad de que el número de la placa (integrado por los 3 dígitos) termine en 0 o 9 y su última letra sea “Z”?

E15. Considere el experimento aleatorio en que se lanzan 3 dados sucesivamente.

a) ¿Cuál es la probabilidad de que exactamente uno de ellos sea 2?

b) ¿Cuál es la probabilidad de obtener al menos un 2?

E16. Se desea colocar 2 libros de Matemáticas, 2 de Economía y uno de Historia sobre un librero. Si los libros se colocan aleatoriamente, ¿cuál es la probabilidad de que en el extremo izquierdo quede el libro de Historia?

E17. Para $x = 0, 1, \ldots, 10$, calcule $x!$ ¿Qué conclusiones se pueden obtener? ¿Existe alguna función continua que genere los valores de los factoriales?

E18. Calcule $\frac{20!}{18!}$ y $\frac{20!}{2!(18!)}$

E19. Suponga que el número de identificación personal (NIP) de una tarjeta de débito debe estar integrado por 4 dígitos distintos. ¿Cuál es la probabilidad de que una persona eligiendo 4 dígitos distintos al azar adivine el NIP de una persona?

E20. “Melate” es un juego de Pronósticos para la Asistencia Pública que consiste en seleccionar 6 números de una boleta en la que aparecen impresos números del 1 al 56. Posteriormente se desarrolla un sorteo en el que se extraen al azar 6 pelotas numeradas de una urna con las 56 pelotas. El ganador del primer lugar es aquel que haya registrado los primeros 6 números que se extraen en el sorteo sin importar el orden. Una variante de este sorteo es “Melate Retro” que sigue la misma mecánica pero utiliza sólo 39 números.

a) Calcule la probabilidad de ganar el “Melate” y el “Melate Retro”.

b) ¿Cuál es el impacto de aumentar de 39 a 56 el total de números del sorteo?

E21. **Modelo Hipergeométrico**. El Consejo de Administración de una compañía está integrado por 7 hombres y 5 mujeres, y ha decidido constituir un Comité de Auditoría formado por 4 personas: presidente, secretario, asesor y vocal. Si la selección de los miembros del Comité de Auditoría se realiza al azar:

a) ¿Cuál es la probabilidad de que el Comité se integre sólo por hombres?

b) Sea C el evento en que figura al menos una mujer en el Comité de Auditoría. Calcule $P(C)$.

c) Determine la probabilidad de que presidente y secretario sean mujeres.

E22. Considere los las siguientes sucesiones:

i	x_i	y_i
1	65	4.3
2	84	5.6
3	109	7.3
4	67	4.5
	325	21.7

Realice los cálculos que se piden a continuación.

a) $\sum_{i=1}^{4} 2i$

b) $\sum_{i=1}^{4} x_i$

c) $\prod_{i=1}^{4} y_i$

d) $\sum_{i=1}^{4} x_i y_i$

e) $\sum_{i=1}^{4} x_i \sum_{i=1}^{4} y_i$

f) $\dfrac{\left(\sum_{i=1}^{4} y_i\right)^2 - \sum_{i=1}^{4} y_i^2}{\sum_{i=3}^{4} \ln(x_i)}$

E23. Demuestre que si $x_1, x_2,..., x_n$ y $y_1, y_2,..., y_n$ son conjuntos de n números reales, y c es una constante, entonces

a) $\sum_{i=1}^{n} c = nc$

b) $\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i$

c) $\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

E24. Se sabe que $\sum_{i=1}^{4} x_i = 325$ y que $\sum_{i=1}^{4} y_i = 21.7$. Calcule $\sum_{i=1}^{4}\left(\frac{x_i}{10} - 2y_i + 5\right)$ utilizando las propiedades de la suma.

E25. Se sabe que si $n \in \mathbf{N}$, entonces $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$ y $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$.

a) Compruebe la validez de estas fórmulas para $n = 5$

b) Calcule $1 + 2 + 3 + \cdots + 1{,}000$

c) A partir de estas fórmulas y las propiedades de la suma calcule $\sum_{i=1}^{50} (2i-3)^2$

E26. Utilice el Teorema Binomial para demostrar que a partir de un conjunto de n elementos, $n \geq 1$, se pueden formar 2^n subconjuntos.

E27. Una baraja de 52 cartas contiene 13 corazones. Suponga que se barajan las cartas y se distribuyen entre cuatro jugadores: A, B, C y D, de forma que cada jugador recibe 13 cartas (esta es la forma en que se inicia una partida de *Bridge*). Calcule la probabilidad de que los jugadores A, B, C y D reciban 6, 4, 2 y 1 corazones, respectivamente.

E28. Sea $(\Omega, \mathscr{A}, P)$ un espacio de probabilidad y suponga que A y B son eventos en $\mathscr{A}$. Utilizando los Axiomas de Probabilidad demuestre que:

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

b) $P(\varnothing) = 0$.

c) Si $B \subset A$ entonces $P(A - B) = P(A) - P(B)$

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

E29. Estudios recientes muestran que en cierta población de México la probabilidad de que un habitante sea mayor de 40 años o tenga calvicie es de 0.40. La probabilidad de que sea mayor de 40 años es de 0.20 y la probabilidad de que tenga calvicie es de 0.30. Se elige un individuo al azar. Calcule la probabilidad de que el individuo…

a) … tenga 40 años o menos.

b) … sea mayor de 40 años con calvicie.

c) … sea mayor de 40 años sin calvicie.

d) … tenga 40 años o menos con calvicie.

e) … tenga 40 años o menos sin calvicie.

E30. **El problema del cumpleaños**. Suponga que en un grupo de 20 personas sus cumpleaños no están relacionados (v.gr., no hay gemelos), y que los 365 días del año tienen las mismas posibilidades de ser el cumpleaños de cualquier persona (i.e., tasa de natalidad uniforme a lo largo del año y los nacidos el 29 de febrero festejan su cumpleaños el 1 de marzo). ¿Cuál es la probabilidad de que al menos dos personas en el grupo tengan el mismo cumpleaños?

E31. La compañía “La Casita” produce y vende cera para pisos en el mercado de productos para cuidado del hogar. La fábrica produce 10,000 litros de cera semanalmente y, generalmente, tiene en inventario 5,000 litros. Si las ventas exceden la producción la compañía cubre el exceso de demanda con el inventario y si las ventas son menores a la producción se incrementa el inventario. El economista de la compañía tiene la siguiente información de eventos de las ventas semanales en litros de cera: $A = [0, 5{,}000]$, $B = (5{,}000, 10{,}000]$, $C = [2{,}500, 7{,}500]$ y $D = (5{,}000, 7{,}500]$, con probabilidades $P(A) = 0.25$, $P(B) = 0.65$, $P(C) = 0.35$ y $P(D) = 0.20$

a) ¿Cuál es la probabilidad de que se tenga que utilizar el inventario para satisfacer las ventas de una semana?

b) ¿Cuál es la probabilidad de que se vendan menos de 2,500 litros de cera en una semana?

c) ¿Cuál es la probabilidad del evento $G = A \cup C$?

E32. Sobre una pared de 2 metros de alto y 2.25 metros de ancho se coloca una diana con tres círculos concéntricos cuyos diámetros son 20 cm., 40 cm. y 60 cm. Si se lanza un dardo al azar sobre esta pared:

a) Determine la probabilidad de que caiga dentro de la diana.

b) ¿Cuál es la probabilidad condicional de que el dardo caiga en el círculo del centro dado que cayó dentro de la diana?

E33. Sean A y B eventos un espacio de probabilidad dado con $P(A) > 0$ y $P(B) > 0$. Demuestre que si B es favorable a A, entonces A es favorable a B. Es decir, que si $P(A|B) > P(A)$, entonces $P(B|A) > P(B)$.

E34. A 100 estudiantes del ITAM, hombres y mujeres, se les pregunta si utilizan coche propio o algún otro medio de transporte para llegar al ITAM. Con base en los resultados de la encuesta se obtuvo la siguiente *distribución de frecuencias absolutas*:

Sexo	**Transporte de llegada**		**Total**
	Coche propio	**Otro**	
Hombre	40	22	62
Mujer	29	9	38
Total	69	31	100

Fuente: Encuesta ITAM

¿Cuál es la probabilidad de que al elegir un estudiante del ITAM al azar…

a) … sea mujer?

b) … sea hombre y utilice coche propio?

c) … sea mujer o use coche propio?

d) … no utilice coche propio dado que es mujer?

e) … no sea mujer dado que no utiliza coche propio?

f) … no sea hombre ni utilice coche propio?

E35. **Regla de la Multiplicación (Probabilidad)**. Demuestre que si $(\Omega, \mathscr{A}, P)$ es un espacio de probabilidad con $A_1, A_2, ..., A_n$, $n \geq 2$, eventos en $\mathscr{A}$ para los cuales $P(A_1 A_2 \cdots A_j) > 0$, $j = 1, 2, \ldots, n - 1$, entonces:

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 A_2)\cdots P(A_n | A_1 A_2 \ldots A_{n-1}).$$

E36. Una urna contiene 10 bolas, de las cuales 3 son negras y 7 blancas. Considere el siguiente experimento: en cada turno una bola es elegida al azar, se anota su color, y se regresa a la urna junto con 2 bolas de su mismo color.

a) ¿Cuál es la probabilidad de obtener 3 bolas negras en los tres primeros turnos de experimento?

b) ¿Cuál es la probabilidad de obtener 2 bolas negras en los tres primeros turnos de experimento?

E37. Comente si intuitivamente A y B son eventos independientes en los siguientes casos:

a) A, color de cabello de un individuo; y B, sabor favorito de helado del individuo.

b) A, persona con una estatura de más de 1.83 m; y B, su padre mide más de 1.83 m.

c) A, edad de una persona; y B, tipo de música favorita de esta persona.

d) A, persona que es jugador de basketball; y B, la estatura de esta persona es mayor a 1.83 m.

E38. Se lanzan dos dados justos. Sea A el evento en que la suma es impar, B el evento en que aparece un uno en el primer dado, y C el evento en que la suma es siete.

a) ¿Son A y B independientes?

b) ¿Son A y C independientes?

c) ¿Son B y C independientes?

E39. Sean A y B dos eventos mutuamente excluyentes tales que $P(A) > 0$ y $P(B) > 0$. Demuestre que A y B no pueden ser independientes.

E40. Se lanza un tetraedro (cuerpo geométrico de 4 caras del mismo tamaño) con una cara verde, otra blanca, otra roja y la otra tricolor (verde, blanco y rojo). Sean V el evento en que la cara abajo aparece el color verde, B si aparece el blanco, R si aparece el rojo. ¿Son V, B y R *eventos completamente independientes*? ¿Qué puede concluir de este ejemplo acerca de la independencia de eventos?

E41. **Modelo Binomial**. Alrededor del 15% de la población se marea mientras toma paseos en lancha. Suponga que 4 turistas seleccionados aleatoriamente abordan una lancha.

a) ¿Cuál es la probabilidad de que ninguno se maree?

b) ¿Cuál es la probabilidad de que al menos uno se maree?

E42. **Regla de Probabilidades Totales**. Demuestre que si (Ω, $\mathscr{A}$, P) es un espacio de probabilidad dado con $B_1, B_2,..., B_n$, una partición de Ω y $P(B_i) > 0$ para $i = 1, 2,..., n$, entonces para todo $A \in \mathscr{A}$, $P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$.

E43. Una urna que contiene 10 pelotas, de las cuales 5 son negras. Considere el experimento en que primero se escoge al azar un número n del conjunto {1, 2, 3} y después se selecciona una muestra de n pelotas sin reemplazo de la urna. Encuentre la probabilidad de que todas las pelotas en la muestra sean negras.

E44. **Teorema de Bayes**. Demuestre que si (Ω, $\mathscr{A}$, P) es un espacio de probabilidad dado con $B_1, B_2,..., B_n$, una partición de Ω y $P(B_i) > 0$ para $i = 1, 2,..., n$, entonces para todo $A \in \mathscr{A}$, con $P(A) > 0$, $P(B_j|A) = \dfrac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$, $j = 1, 2,..., n$.

E45. Una empresa petrolera planea perforar un pozo en una zona donde no se tiene la seguridad de que exista petróleo. Los geólogos, de acuerdo con su experiencia, creen que la probabilidad de que exista petróleo en la zona es de 0.1. Se tiene la opción de hacer una prueba preliminar antes de tomar una decisión. La prueba no es concluyente, puesto que hay casos en que da resultados erróneos. Si existe petróleo, la prueba es positiva el 90% de las veces, pero aún si no existe petróleo, la prueba es positiva el 20% de las veces. Determine la probabilidad de que exista petróleo dado que la prueba resulta positiva.

E46. Cierta compañía produce objetos con tres tipos de máquinas con diferentes tecnologías. La máquina A, elabora el 30% de la producción, la máquina B elabora el 50% de la producción y la máquina C el resto de la producción. Se sabe que la máquina A tiene probabilidad de fabricar objetos defectuosos de 0.1, la máquina B de 0.12 y la máquina C de 0.04.

a) ¿Cuál es la probabilidad de que un objeto tomado al azar sea defectuoso?

b) ¿Cuál es la probabilidad de que un objeto tomado al azar haya sido producido por la máquina A si se sabe que es defectuoso?

E47. Considere el experimento aleatorio en que se lanzan 3 volados sucesivamente. Sean X, la variable aleatoria que indica el número de águilas que aparecen durante el experimento; y Y, la variable aleatoria que toma el valor 1 si aparece al menos un sol y 0 en otro caso.

a) Defina el espacio muestral del experimento por extensión.

b) Para cada punto del espacio muestral determine su probabilidad y los valores que toman las variables aleatorias X y Y.

c) Calcule $P[X = 2]$ y $P[Y = 0]$.

E48. Suponga que un banco presta servicios de banca múltiple de 9:00 a 17:00 hrs. El banco cuenta con 3 cajas y una ventanilla para realizar trámites administrativos de los cuentahabientes. En las operaciones de un día en particular de este banco se han identificado algunas variables aleatorias que se definen a continuación. Indique en cada caso si se trata de una variable discreta o continua e indique sus posibles valores.

a) X = Número de personas que visitan el banco durante el día.

b) Y = Caja en la que será atendida una persona que desea realizar un depósito.

c) Z = Número de trámites administrativos que realiza un cuentahabiente.

d) T = Tiempo que tarda en ser atendida una persona que llega a las 15:00 hrs.

e) S = Monto total de los depósitos que recibe el banco en este día.

f) R_i = El tiempo que tarda en atender a una persona el cajero i, $i = 1, 2, 3$.

E49. Suponga que X es una variable aleatoria discreta para la cual

$$f_X(x) = \begin{cases} \dfrac{x}{c} & si \;\; x = 2, 3 \\ \dfrac{7-x}{c} & si \;\; x = 4, 5 \\ 0 & e.o.c. \end{cases}$$

a) Determine el valor de c de modo que $f_X(x)$ sea función de masa de probabilidad.

b) Grafique $f_X(x)$ y compruebe que esta función satisface las condiciones que caracterizan a las funciones de masa de probabilidad.

c) Calcule $P[3 < X < 5]$ y $P[3 \leq X \leq 5]$

d) Calcule $P[X < 5 \mid X > 2]$

E50. Reescriba la siguiente función de masa de probabilidad utilizando funciones indicadoras:

$$f_X(x) = \begin{cases} \dfrac{x}{10} & si \;\; x = 2, 3 \\ \dfrac{7-x}{10} & si \;\; x = 4, 5 \\ 0 & e.o.c. \end{cases}$$

E51. Una persona tiene la siguiente estrategia de juego en un casino de Las Vegas. Apuesta 100 dólares a que la ruleta caerá en número rojo y si gana, se retira. Si pierde, entonces hace la misma apuesta pero con 200 dólares e independientemente del resultado se retira. Suponiendo que la ruleta tiene 36 números, que la mitad de los números son rojos y que X es la variable aleatoria de la ganancia o pérdida de esta persona, determine la función de masa de probabilidad de X.

E52. Se tira un dado consecutivamente hasta que aparece un 6.

a) Si se define a la variable aleatoria Y como el número de tiros necesarios hasta que aparezca un 6, obtenga su función de masa de probabilidad.

b) ¿Cuál es la probabilidad de que a lo más el dado se tire 3 veces?

c) ¿Cuántos tiros se necesitan para asegurar que la probabilidad de obtener un 6 sea al menos 0.5?

E53. La *Distribución Poisson* con parámetro λ, $\lambda > 0$, se caracteriza por la siguiente función de masa de probabilidad: $f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} I_{\{0,1,\ldots\}}(x)$.

a) Verifique que, efectivamente, $f_X(x)$ es función de masa de probabilidad.

b) Calcule $P[X \geq 2]$.

E54. Considere que X es una variable aleatoria continua para la cual

$$f_X(x) = \begin{cases} 0.2 & si \;\; 0 < x \leq 1 \\ 0.3 & si \;\; 1 < x \leq 2 \\ -0.5 + 0.4x & si \;\; 2 < x \leq 3 \\ 0 & e.o.c. \end{cases}$$

a) Grafique $f_X(x)$ y exprésela en términos de funciones indicadoras.

b) Verifique que $f_X(x)$ es función de densidad de probabilidad.

c) Calcule $P[0.5 \leq X < 1]$ y $P[0.5 < X \leq 1]$. ¿Hay alguna diferencia?

d) Determine $P[X = 2]$ y comente su resultado.

e) Calcule $P[X > 2]$, $P[X < 0.5]$ y $P[X \geq 0.5]$.

f) Calcule $P[X < 1 \mid X \geq 0.5]$.

E55. La *Distribución Exponencial* con parámetro β, $\beta > 0$, se caracteriza por la siguiente función de densidad de probabilidad: $f_X(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} I_{(0,\infty)}(x)$.

a) Verifique que, efectivamente, $f_X(x)$ es función de densidad de probabilidad.

b) Calcule $P[a < X < b]$ para $0 < a < b$.

E56. Suponga que $f_1(x)$ y $f_2(x)$ son funciones de densidad de probabilidad. Demuestre que la función $f(x) = \alpha f_1(x) + (1 - \alpha) f_2(x)$, $0 < \alpha < 1$, también es función de densidad de probabilidad.

E57. Considere el experimento aleatorio en que se lanza un dado. Se define la variable aleatoria X como el valor que toma el dado.

a) Determine la función de masa de probabilidad de X y grafíquela.

b) Construya la función de distribución acumulada de X y grafíquela.

c) Compruebe que $F_X(x)$ satisface las condiciones que caracterizan a cualquier función de distribución acumulada.

d) A partir de la función de distribución acumulada de X calcule $P[X < 3]$, $P[X \leq 2]$, $P[X > 2]$, $P[3 < X < 6]$ y $P[3 \leq X < 6]$.

E58. Una estación de servicio tiene dos bombas y cada una puede surtir hasta 10,000 galones por mes. La cantidad total de gasolina demandada en esta estación por mes es una variable aleatoria X (medida en diez miles de galones) con función de densidad de probabilidad

$$f_X(x) = \begin{cases} x & si \;\; 0 < x < 1 \\ 2 - x & si \;\; 1 \leq x < 2 \\ 0 & e.o.c. \end{cases}$$

a) Construya la función de distribución acumulada de X y grafíquela.

b) Compruebe que $F_X(x)$ satisface las condiciones que caracterizan a cualquier función de distribución acumulada.

c) Calcule la probabilidad de que la estación surta entre 8,000 y 12,000 galones en un mes en particular.

d) Dado que durante un mes la estación surtió más de 10,000 galones, ¿cuál es la probabilidad de que haya surtido más de 15,000 galones?

E59. Se sabe que si X tiene una *Distribución Exponencial* con parámetro β, $\beta > 0$, entonces $f_X(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} I_{(0,\infty)}(x)$. Demuestre que $F_X(x) = \left(1 - e^{-\frac{x}{\beta}}\right) I_{(0,\infty)}(x)$.

E60. El número de solicitudes de apertura de crédito que se reciben diariamente en un banco es una variable aleatoria W con función de distribución acumulada

$$F_W(w) = \begin{cases} 0 & si \;\; w < 0 \\ 0.1 & si \;\; 0 \leq w < 1 \\ 0.3 & si \;\; 1 \leq w < 2 \\ 0.7 & si \;\; 2 \leq w < 4 \\ 1 & si \;\; w \geq 4 \end{cases}$$

a) Encuentre la función de masa de probabilidad de la variable aleatoria W.

b) Si en un día en particular se ha recibido una solicitud, ¿cuál es la probabilidad de que se reciban más de 3 solicitudes?

c) Calcule la probabilidad de que se reciban 3 solicitudes en un día particular.

E61. El período de funcionamiento de cierto transmisor hasta su primera falla (en cientos de horas) es una variable aleatoria T con función de distribución acumulada

$$F_T(t) = \begin{cases} 0 & si\ \ t < 0 \\ 1 - e^{-t^2} & si\ \ t \geq 0 \end{cases}$$

a) Calcule la probabilidad de que el transmisor trabaje por lo menos durante 200 horas hasta tener su primera falla.

b) Determine la función de densidad de probabilidad de la variable aleatoria T.

E62. La *Distribución Weibull* con parámetros α y β, $\alpha > 0$ y $\beta > 0$, se utiliza frecuentemente para modelar el tiempo de vida (X) de algunas máquinas. Si la función de supervivencia de X está dada por $S_X(x) = e^{-\left(\frac{x}{\alpha}\right)^{\beta}}$ para $x > 0$, calcule la función de densidad de X.

E63. Suponga que X es la variable aleatoria del número de veces que llueve por día en cierta región y que su función de masa de probabilidad está dada por:

X	0	1	2
$P[X = x]$	0.1	0.6	0.3

a) Determine la expresión de $f_X(x)$ y construya su gráfica.

b) ¿Cuántas veces se espera que llueva por día en esta región? Grafique este valor en la gráfica del inciso anterior.

E64. La función de masa de probabilidad de la variable aleatoria X está definida por

$$f_X(x) = \begin{cases} \dfrac{2x}{n(n+1)} & si\ \ x = 1, 2, \ldots, n \\ 0 & e.o.c. \end{cases}, \quad n \in \mathbf{Z}^+$$

Calcule la esperanza de X.

E65. El tiempo en horas (Z) requerido por los estudiantes de un curso para resolver un examen final es una variable aleatoria con función de densidad:

$$f_Z(z) = \begin{cases} \dfrac{z(4-z)}{9} & si\ \ 0 \leq z \leq 3 \\ 0 & e.o.c. \end{cases}$$

a) Realice un bosquejo de la gráfica de $f_Z(z)$.

b) Calcule e interprete la esperanza matemática de la variable aleatoria Z.

E66. La variable aleatoria Y tiene una función de densidad dada por

$$f_Y(y) = \begin{cases} a + by & si\ \ 0 < y < 1 \\ 0 & e.o.c. \end{cases}$$

Encuentre los valores de a y b tales que la media de Y sea $\dfrac{2}{3}$.

E67. Sea X una variable aleatoria con función de densidad $f_X(x) = x^{-2} I_{[1,\infty)}(x)$.

a) Grafique y verifique que $f_X(x)$ es función de densidad de probabilidad.

b) Demuestre que $E[X]$ no existe para esta variable aleatoria.

E68. Un puesto de periódicos vende un diario francés poco demandado. Si Y es el número de diarios vendidos en un día, su función de masa de probabilidad está dada por:

Y	5	6	7	8
$P[Y = y]$	0.2	0.4	0.3	0.1

Suponga que la utilidad (en pesos) para el puesto de periódicos por la venta del diario francés está dada por $U(Y) = 60\sqrt{Y-5}$.

a) Calcule la utilidad esperada.

b) Calcule la probabilidad de que la utilidad sea mayor a 60 pesos.

E69. Una casa de bolsa ofrece a sus clientes un fondo de renta variable con una *tasa de rendimiento anual* con la siguiente función de densidad:

$$f_X(x) = \begin{cases} \frac{2}{9}(x+1) & si \quad -1 \le x \le 2 \\ 0 & e.o.c. \end{cases}$$

Un inversionista puede ganar o perder en este fondo dependiendo si la tasa X es positiva o negativa, respectivamente.

a) Calcule la tasa de rendimiento anual esperada.

b) Si un cliente de la casa de bolsa decide invertir 50 mil pesos en este fondo, entonces la comisión del ejecutivo de cuenta será de $C = 5\sqrt{X+1}$ (en miles de pesos). Calcule el valor esperado de la comisión del ejecutivo de cuenta.

c) La *tasa de rendimiento continuo* del fondo está dada por $Y = \ln(1+X)$. Calcule la probabilidad de que la tasa de rendimiento continuo sea mayor a 100%.

d) Si otro cliente de la casa de bolsa decide invertir 90 mil pesos en este fondo, entonces el saldo de la inversión al final de 2 años será de $S = 90(1+X)^2$ (en miles de pesos). Calcule la probabilidad de que al final de los 2 años el saldo del inversionista sea menor al saldo final promedio de esta inversión.

E70. Se sabe que si X tiene una *Distribución Uniforme Continua* sobre el intervalo $[\theta_1, \theta_2]$, con $\theta_1, \theta_2 \in \mathbf{R}$ y $\theta_1 < \theta_2$, entonces $f_X(x) = \frac{1}{\theta_2 - \theta_1} I_{[\theta_1,\theta_2]}(x)$, es decir, la distribución asigna la misma probabilidad a intervalos de igual longitud en $[\theta_1, \theta_2]$.

a) Calcule el k-ésimo momento de la variable aleatoria X, $k = 1, 2,...$

b) Grafique e interprete el valor del primer momento ($k = 1$).

c) Calcule el k-ésimo momento central de la variable aleatoria X, $k = 1, 2,\ldots$

E71. Sea W el número que se obtiene al lanzar un dado honesto cuyas caras están numeradas del 1 al 6. Calcule esperanza, varianza y desviación estándar de la variable aleatoria W.

E72. **Fórmula para calcular la Varianza.** Si X es una variable aleatoria discreta, demuestre que $Var[X] = E[X^2] - (E[X])^2$.

E73. Considere nuevamente la variable aleatoria W que denota el número que se obtiene al lanzar un dado honesto con caras numeradas del 1 al 6. Verifique cómo la varianza de W calculada a partir del "segundo momento menos el primer momento al cuadrado" coincide con el valor calculado a partir de la definición de varianza.

E74. Se sabe que si X tiene una *Distribución Exponencial* con parámetro β, $\beta > 0$, entonces $f_X(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} I_{(0,\infty)}(x)$. Calcule la varianza y la desviación estándar de X.

E75. Un inversionista realiza dos inversiones. La inversión 1 tendrá una ganancia de \$1,000 con probabilidad 0.6 o una pérdida de \$400 con probabilidad de 0.4. La inversión 2 tendrá una ganancia de \$2,000 con probabilidad de 0.5 o una pérdida de \$500 con probabilidad de 0.5.

a) Grafique y compare las funciones de masa de probabilidad de la ganancia (pérdida) de estas inversiones.

b) Calcule esperanza y desviación estándar de cada inversión.

c) Si usted fuera analista financiero y considerara sólo la ganancia o pérdida esperada, ¿qué inversión recomendaría?

d) En Finanzas, la desviación estándar de una inversión es interpretada como su *riesgo financiero*. Si usted fuera analista financiero y considerara sólo el riesgo de las inversiones, ¿qué inversión recomendaría?

E76. **Propiedades del Valor Esperado y la Varianza.** Suponga que X es una variable aleatoria continua con función de densidad de probabilidad $f_X(x)$, que a, b y c son constantes, y que $g(\cdot)$ y $h(\cdot)$ son funciones. Demuestre que:

a) $E[c] = c$

b) $E[ag(X) \pm bh(X)] = aE[g(X)] \pm bE[h(X)]$

c) $Var[c] = 0$

d) $Var[cX] = c^2 Var[X]$

E77. La producción mensual de jarabe para la tos de un laboratorio (*X*, en miles de litros) es una variable aleatoria con función de densidad $f_X(x)=\dfrac{6-x}{8}I_{[2,6]}(x)$. Se sabe que para la elaboración de este jarabe, el laboratorio enfrenta *costos fijos* por $145,000 y una función de *costos variables* (en miles de pesos) dada por $C(X)=\dfrac{105(X-2)^2}{4}$.

a) Calcule esperanza y varianza de la producción mensual de jarabe.

b) Calcule el *costo total esperado* que debe pagar mensualmente el laboratorio.

c) Calcule el *costo medio esperado* que debe pagar mensualmente el laboratorio.

E78. Suponga que *Y* es el número de veces que falla una impresora diariamente en una oficina. La probabilidad de que la impresora no falle es de 0.1, pero la probabilidad de que falle 1, 2 o 3 veces es de 0.3, 0.4 y 0.2, respectivamente. Calcule media, mediana y moda de la variable aleatoria *Y*.

E79. En el ejercicio E65 se consideró que el tiempo en horas (*Z*) requerido por los estudiantes de un curso para resolver un examen final es una variable aleatoria con función de densidad:

$$f_Z(z)=\begin{cases}\dfrac{z(4-z)}{9} & si\ \ 0\le z\le 3\\ 0 & e.o.c.\end{cases}$$

Calcule moda y mediana de la variable aleatoria *Z*.

E80. El tiempo (en minutos) que pasa una persona en un verificentro es una variable aleatoria con función de distribución acumulada

$$F_T(t)=\begin{cases}0 & si\ \ t<0\\ 1-e^{-\frac{1}{20}t} & si\ \ t\ge 0\end{cases}$$

a) Calcule la función de densidad de probabilidad de la variable aleatoria *T*.

b) Determine el valor de la mediana de la distribución de *T*.

c) Calcule el percentil 0.10 de la distribución de probabilidad de la variable aleatoria *T*. Interprete dicho valor.

d) Determine el valor de p_α (el percentil α) para $0 < \alpha < 1$.

e) ¿Qué tiempo de espera máximo debiera comunicar el verificentro a sus clientes para asegurar que será correcto en el 95% de los casos?

f) Considere el intervalo [k, $p_{0.95}$]. ¿Qué valor debe tomar k para afirmar que el tiempo de espera se ubica en este intervalo con una probabilidad de 90%?

g) Calcule los cuartiles de la distribución de la variable aleatoria *T*. Ubique estos valores en la gráfica de la función de densidad de *T*.

E81. Se sabe que si X tiene una *Distribución Geométrica* con parámetro p , $0 < p < 1$, su función de masa de probabilidad está dada por

$$f_X(x) = \begin{cases} (1-p)^{x-1} p & si \;\; x = 1, 2, \ldots \\ 0 & e.o.c. \end{cases}$$

a) Calcule el coeficiente de variación de la variable aleatoria X.

b) Si en cierta población la probabilidad de que nazca un niño es 0.4 y una pareja de esta población decide tener hijos hasta tener un niño, ¿cuántos hijos se espera que tenga esta pareja? Calcule e interprete el coeficiente de variación en este caso.

E82. Se sabe que si X tiene una *Distribución Uniforme Continua* sobre el intervalo $[\theta_1, \theta_2]$, con $\theta_1, \theta_2 \in \mathbf{R}$ y $\theta_1 < \theta_2$, entonces $f_X(x) = \frac{1}{\theta_2 - \theta_1} I_{[\theta_1, \theta_2]}(x)$. Además, en el ejercicio E70 se demostró que $\mu_k = \frac{1}{(k+1)(\theta_2 - \theta_1)} \left[\left(\frac{\theta_2 - \theta_1}{2} \right)^{k+1} - \left(-\frac{\theta_2 - \theta_1}{2} \right)^{k+1} \right]$.

a) Calcule el coeficiente de asimetría de X. Interprete.

b) Calcule el coeficiente de curtosis de X. Interprete.

E83. **Cálculo de momentos vía f.g.m.** Asumiendo *condiciones de regularidad*, demuestre que si X es una variable aleatoria con función generadora de momentos $M_X(t)$, $|t| < b$, $b > 0$, entonces $E[X^k] \equiv \mu'_k = \left. \frac{d^k}{dt^k} M_X(t) \right|_{t=0}$ para $k = 1, 2, \ldots$

E84. Se dice que un experimento es un *Experimento Bernoulli* si su resultado sólo puede ser éxito con probabilidad p, $0 < p < 1$, o fracaso con probabilidad $q = 1 - p$, independientemente de otros experimentos. Si la variable X toma valores 1 o 0 considerando el éxito o fracaso de un Experimento Bernoulli, respectivamente, entonces se dice que X tiene una *Distribución Bernoulli*.

a) Construya la función de masa de probabilidad de la variable aleatoria X.

b) Calcule directamente esperanza y varianza de X.

c) Calcule la función generadora de momentos de X.

d) Calcule esperanza y varianza de X utilizando la función generadora de momentos.

E85. **Distribución Normal Estándar**. Por sus aplicaciones y sus propiedades, la distribución más importante en Estadística es la *Distribución Normal Estándar*. Se dice que Z tiene una *Distribución Normal Estándar* si su función de densidad de probabilidad es $f_Z(z) \equiv \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} I_{(-\infty,\infty)}(z)$.

a) Realice un bosquejo de la gráfica de $f_Z(z)$.

b) Demuestre que la función generadora de momentos de Z es $M_Z(t) = e^{\frac{t^2}{2}}$, $t \in \mathbf{R}$.

c) Calcule esperanza y varianza de Z utilizando la función generadora de momentos.

d) Calcule los coeficientes de asimetría y de curtosis de Z.

E86. Considere la variable aleatoria Y que representa el ingreso de las personas en cierta localidad. Una posible forma de estudiar el comportamiento de Y es suponiendo que tiene una *Distribución Pareto* con parámetros $\theta > 0$ y $\lambda > 0$, y con función de densidad de probabilidad $f_Y(y) = \frac{\theta\lambda^\theta}{y^{\theta+1}} I_{[\lambda,\infty)}(y)$.

a) Realice un bosquejo de la gráfica de $f_Y(y)$.

b) Demuestre que para esta variable aleatoria no existe un valor b tal que $M_Y(t)$ sea finita para $|t| < b$, $b > 0$.

c) Calcule esperanza y varianza de esta distribución.

E87. **Desigualdad de Markov y Desigualdad de Chebyshev.**

a) Demuestre la *Desigualdad de Markov*, es decir, que si Z es variable aleatoria no negativa ($Z \geq 0$), entonces para $a > 0$, $P[Z \geq a] \leq \frac{E[Z]}{a}$

b) Utilice la Desigualdad de Markov para demostrar la *Desigualdad de Chebyshev*, es decir, que si X es una variable aleatoria con media μ y varianza finita σ^2, entonces $P[|X - \mu| \geq k] \leq \frac{\sigma^2}{k^2}$ para toda $k > 0$.

E88. En un cine se venden refrescos de 300 mililitros, sin embargo, la máquina que surte el refresco en ocasiones vierte más o menos refresco del debido. Suponga que X es la cantidad (en mililitros) de refresco adicional ($X > 0$) o faltante ($X < 0$) por vaso, y que su función de densidad de probabilidad es $f_X(x) = \frac{1}{2}e^{-|x-2|}I_{(-\infty,\infty)}(x)$. Para esta distribución se puede demostrar que $\mu = E[X] = 2$ y que $\sigma^2 = Var[X] = 2$.

a) Calcule la probabilidad de que el refresco adicional o faltante se aleje del contenido promedio en más de 5 mililitros.

b) Utilice la Desigualdad de Chebyshev para acotar la probabilidad del inciso anterior.

c) Sin hacer el cálculo exacto, ¿qué probabilidad mínima se puede garantizar a los clientes de que el contenido de refresco por vaso estará entre 290 y 314 mililitros?

d) Sin hacer el cálculo exacto, ¿cuál es la probabilidad mínima de que el contenido de refresco se aleje menos de dos desviaciones estándar de su contenido promedio?

E89. **Desigualdad de Jensen**. Demuestre que si g es una función convexa, entonces $E[g(X)] \geq g(E[X])$, siempre que los valores esperados existan.

E90. Un inversionista enfrenta dos alternativas. *Alternativa 1*: Invertir su dinero en un instrumento riesgoso que le proporcionará un rendimiento aleatorio X con media m. *Alternativa 2*: Invertir su dinero en instrumentos libres de riesgo que le proporcionan un rendimiento m con probabilidad 1. Suponga que la decisión del inversionista la realiza maximizando el valor esperado de su función de utilidad $u(\cdot)$. ¿Qué alternativa debe elegir el inversionista si...

a) ... $u(\cdot)$ es convexa?

b) ... $u(\cdot)$ es cóncava?

E91. Un vendedor de autos estima las siguientes probabilidades para el número de autos de un modelo en particular (X) que venderá la próxima semana:

X	0	1	2	3
$P[X = x]$	0.2	0.3	0.4	0.1

El vendedor recibe un sueldo semanal de 4 mil pesos más un bono semanal si vende más de un auto. El monto del bono semanal, B (en miles de pesos) se determina por:

$$B = \max\{0, bp(X-1)\} = \begin{cases} 0 & si \;\; X < 1 \\ bp(X-1) & si \;\; X \geq 1 \end{cases}$$

Donde: b = porcentaje de bono semanal.
p = precio de venta del auto (en miles de pesos).

Si el *precio de venta* de este modelo de autos es de \$200,000 y el *porcentaje de bono* semanal es del 0.5%:

a) Construya la función de masa de probabilidad del ingreso semanal del vendedor.

b) ¿Cuál es el ingreso esperado semanal del vendedor?

E92. Un productor cuenta con un proceso para refinar azúcar que le permite producir hasta 1.5 toneladas de azúcar por día, pero la cantidad realmente producida X es una variable aleatoria debido a fallas mecánicas y otras contingencias. Suponga que X tiene la función de densidad:

$$f_X(x) = \begin{cases} x & si \;\; 0 \le x \le 1 \\ 1 & si \;\; 1 < x \le \dfrac{3}{2} \\ 0 & e.o.c. \end{cases}$$

Se sabe que el productor recibe 300 dólares por tonelada de azúcar refinada pero enfrenta costos fijos de 100 dólares por día.

a) Exprese la utilidad diaria del productor (en cientos de dólares).

b) Obtenga la función de distribución acumulada y la función de densidad de la utilidad diaria del productor.

E93. Si Y tiene función de densidad de probabilidad $f_Y(y) = (1-|y|)I_{[-1,1]}(y)$:

a) Obtenga $f_U(u)$ si $U = U(Y) = Y^2$.

b) Indique cómo se relacionan $E[U(Y)]$ y $U(E[Y])$.

E94. **Teorema de la Distribución de la Transformación Monótona (DTM)**. Suponga que X es una variable aleatoria continua con función de densidad $f_X(x)$. Si $g(x)$ es una función estrictamente creciente o decreciente y diferenciable en x, demuestre que la función de densidad de $Y = g(X)$ es $f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\dfrac{d}{dy}g^{-1}(y)\right|$ si $y = g(x)$ para alguna x.

E95. Suponga que X es una variable aleatoria continua con función de densidad $f_X(x)$ y función de distribución acumulada $F_X(x)$. Obtenga $f_Z(z)$ si $Z = g(X) = \dfrac{1}{X+1}$.

E96. Suponga que X tiene una *Distribución Uniforme* sobre el intervalo [0, 1], es decir, su función de densidad es $f_X(x) = I_{[0,1]}(x)$. Demuestre que $Y = -\beta \ln(X)$, $\beta > 0$, tiene una *Distribución Exponencial* con media β.

E97. **Estandarización de la Distribución Normal.** Se dice que X tiene una *Distribución Normal* con parámetros $\mu \in \mathbf{R}$ y $\sigma > 0$, si su función de densidad de probabilidad está dada por $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} I_{(-\infty,\infty)}(x)$.

a) Demuestre que $M_X(x) = \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$ para $t \in \mathbf{R}$.

b) Demuestre que $E[X] = \mu$ y que $Var[X] = \sigma^2$.

c) A la transformación $Z = \frac{X-\mu}{\sigma}$ se le denomina *estandarización de la variable aleatoria X*. Demuestre que Z tiene una *Distribución Normal Estándar* considerando su función generadora de momentos (calculada en el ejercicio E85).

E98. Suponga que $X \sim Bin(n, p)$

a) Demuestre que $M_X(t) = \left[(1-p) + pe^t\right]^n$, $t \in \mathbf{R}$.

b) Utilice la función generadora de momentos de la variable aleatoria X para demostrar que $E[X] = np$ y que $Var[X] = np(1-p)$.

E99. Suponga que cada acción de la Bolsa Mexicana de Valores (BMV) aumenta su valor diariamente con probabilidad de 45%, independientemente de las demás. Si se seleccionan aleatoriamente cinco acciones:

a) ¿Cómo se distribuye el número de acciones que aumentan su valor?

b) ¿Cuál es la probabilidad de que exactamente cuatro de ellas aumenten su valor?

c) ¿Cuál es la probabilidad de que al menos tres acciones aumenten su valor?

d) Suponga que el valor de las acciones siempre cambia, es decir, no puede permanecer constante. Si de las cinco acciones elegidas se sabe que dos de ellas disminuyen su valor, ¿cuál es la probabilidad de que el valor de las cinco acciones disminuya?

E100. La tasa de llegadas de los aviones a un aeropuerto es de 4 aviones por hora y se sabe que lo hacen de acuerdo con los postulados de la Ley Poisson.

a) Defina la función de masa de probabilidad de la variable aleatoria que permite modelar la llegada del número de aviones a dicho aeropuerto.

b) Calcule la probabilidad de que en una hora lleguen exactamente 4 aviones.

c) ¿Cuál es la probabilidad de que en una hora lleguen más de 4 aviones?

d) ¿Cuál es la probabilidad de que en un lapso de dos horas lleguen más de 8 aviones?

e) Calcule la moda de la distribución del número de aviones por hora que llegan a este aeropuerto.

E101. Demuestre que si $X \sim Bin(n, p)$, $n \to \infty$, $p \to 0$ y $np \to \lambda$ entonces $f_X(x) \to f_Y(y)$, donde $Y \sim Po(\lambda)$ con $\lambda = np$.

E102. Suponga que en una fábrica de palillos el 1% de la producción resulta defectuosa. Los palillos se venden en cajas de 200 piezas. ¿Cuál es la probabilidad de que al menos el 98% del contenido de una caja no resulte defectuoso?

E103. Suponga que $X \sim U(\theta_1, \theta_2)$, $\theta_1 < \theta_2$.

a) Demuestre que $F_X(x) = \dfrac{x-\theta_1}{\theta_2-\theta_1} I_{[\theta_1,\theta_2]}(x) + I_{(\theta_2,\infty)}(x)$

b) ¿Cómo es $F_X(x)$ si $\theta_1 = 0$ y $\theta_2 = 1$?

E104. En cierto club de paracaidismo, cada paracaidista cae en un sitio aleatorio de la línea recta entre los puntos A y B.

a) ¿Cuál es la probabilidad de que un paracaidista caiga más cerca del punto A que del punto B?

b) Calcule la probabilidad de que la distancia entre el punto en el que cae un paracaidista y el punto A sea más de tres veces la distancia con respecto al punto B.

c) Si tres paracaidistas caen en forma independiente sobre la línea recta que une los puntos A y B, ¿cuál es la probabilidad de que exactamente uno de los tres caiga más cerca de A que de B?

E105. La Función Gamma se define por $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy$, $\alpha > 0$.

a) Demuestre que $\Gamma(1) = \Gamma(2) = 1$

b) Demuestre que $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$

c) Si $n \in \mathbf{Z}^+$ determine el valor de $\Gamma(n)$

E106. Suponga que $X \sim Gamma(\alpha, \beta)$.

a) Calcule el k-ésimo momento de la variable aleatoria X.

b) Demuestre que $E[X] = \alpha\beta$ y que $Var[X] = \alpha\beta^2$.

E107. Suponga que el tiempo (en horas) que dura encendida una batería es una variable aleatoria exponencial y que la probabilidad de que se funda en menos de 10 horas es de 0.4. Una lámpara utiliza 5 de estas baterías y se colocan en un circuito que permite que la lámpara permanezca encendida con que al menos 4 de ellas funcionen. ¿Cuál es la probabilidad de que la lámpara dure encendida más de 45 horas?

E108. Demuestre la propiedad de *Pérdida de Memoria de la Distribución Exponencial*, es decir, que $P[X > t+s \mid X > t] = P[X > s]$ para todo s, $t \geq 0$.

E109. La tasa de llegadas de los aviones a un aeropuerto es de 4 aviones por hora y se sabe que lo hacen de acuerdo con los postulados de la Ley Poisson.

a) Defina la función de densidad y la función de distribución acumulada de la variable aleatoria que permita modelar el tiempo que ocurre entre las llegadas de los aviones.

b) Calcule el coeficiente de variación para el tiempo entre ocurrencias.

c) ¿Cuál es la probabilidad de que pase más de media hora entre la llegada de dos aviones?

d) Si se sabe que ha pasado media hora desde que llegó el último avión, ¿cuál es la probabilidad de que el siguiente avión llegue después de otra media hora?

E110. Para cierto modelo de automóviles, el monto de la pérdida (en miles de pesos) en caso de choque tiene una Distribución Gamma con media 20 y varianza 40.

a) Si un automóvil de este modelo tiene un choque, ¿cuál es la probabilidad de que la pérdida sea de más de 10,850 pesos?

b) Calcule la probabilidad de que la pérdida sea menor a 40,000 pesos en caso de choque.

c) ¿Cuál es la pérdida máxima con 99.9% de probabilidad para este modelo de automóviles?

E111. Suponga que una compañía fabrica tornillos de una longitud específica, sin embargo, defectos del equipo de producción hacen que los tornillos sean ligeramente más grandes o más pequeños. Suponga que esta variación (en milímetros) tiene una distribución Normal Estándar.

a) ¿Cuál es la probabilidad de que la variación sea positiva?

b) Calcule la probabilidad de que la variación esté dentro del intervalo $[-1, 1]$ ¿Cómo se interpreta esta probabilidad?

c) ¿Qué valor debe tomar k ($k > 0$) para asegurar que la variación esté dentro del intervalo $[-k , k]$ el 95% de las veces?

E112. La estatura de los hombres adultos de la población A tiene una Distribución Normal con media de 1.75 metros y desviación estándar de 5 centímetros. En la población B la estatura de los hombres adultos también tiene una Distribución Normal con media de 1.75 metros pero con una desviación estándar de 10 centímetros.

a) Si se elige un hombre adulto de cada población, ¿cuál de ellos es más probable que supere una altura de 1.90 metros?

b) ¿Qué estatura mínima se puede establecer para describir a los hombres adultos de la población A y tener una certeza del 95%?

E113. En cierta localidad la mitad de las personas adultas está a favor de un proyecto municipal y la otra mitad está en contra. Si se toma una muestra de tamaño 100…

a) ¿Cómo se distribuye el número de personas de la muestra que está a favor del proyecto? Indique sus supuestos.

b) Aproxime la probabilidad de que 60 o más personas de la muestra estén a favor del proyecto.

E114. La temperatura al amanecer en cierta ciudad se distribuye normalmente con media y desviación estándar de 39.2 y 5.6 grados Fahrenheit, respectivamente. Si C denota la temperatura en grados Centígrados y F la temperatura en grados Fahrenheit, entonces la relación entre ambas escalas está dada por $C = \dfrac{5(F-32)}{9}$

a) ¿Cómo se distribuye la temperatura al amanecer de esta ciudad medida en grados Centígrados?

b) Si la temperatura en cada amanecer es independiente de las demás, ¿cuál es la probabilidad de que en más de 2 días de los 7 días de una semana elegida al azar, dicha temperatura esté por debajo de los cero grados Centígrados?

E115. Suponga que X y Y son variables aleatorias discretas con función de masa de probabilidad conjunta

$$f_{X,Y}(x,y) = \begin{cases} cxy & si \;\; x = 1,2,3; y = 2,4 \\ 0 & e.o.c. \end{cases}$$

a) Determine el valor de c que permita asegurar que $f_{X,Y}(x,y)$ sea función de masa de probabilidad conjunta.

b) Construya las distribuciones marginales de X y Y.

c) Calcule $P[X = Y]$

E116. Una urna contiene 8 pelotas, la mitad son negras y la otra mitad blancas. Considere el experimento aleatorio en que se lanza un dado justo con caras numeradas del 1 al 6. Si el dado cae en un número menor o igual a 4 se extrae una pelota de la urna, pero si el dado cae en un número mayor a 4 se extraen 2 pelotas sin reemplazo. Suponga que X es el número total de pelotas extraídas durante el experimento y que Y es el número de pelotas negras.

a) Calcule la función de masa de probabilidad conjunta de las variables aleatorias X y Y, así como sus respectivas funciones de masa de probabilidad marginales.

b) ¿Cuántas pelotas negras se espera obtener al final del experimento?

c) ¿Cuál es la probabilidad de que todas las pelotas extraídas durante el experimento sean negras?

d) Si sabemos que el dado cae en 5, ¿cuál es la probabilidad de obtener al menos una pelota negra?

e) Si por cada pelota blanca se reciben \$50 y por cada negra se pierden \$100, ¿cuál es la probabilidad de obtener una ganancia?

E117. Suponga que X es el tiempo (en minutos) que tarda en llegar un estudiante de su casa a la parada del autobús y que Y es el tiempo (en minutos) que tarda en llegar de la parada del autobús a la escuela. La función de densidad conjunta de X y Y está dada por

$$f_{X,Y}(x,y)=\begin{cases} k & si\ 10<x<15,\ 10<y<20 \\ 0 & e.o.c. \end{cases}$$

a) Determine el valor de k.

b) ¿Cuál es la probabilidad de que el estudiante tarde menos de 12 minutos en cada trayecto? Es decir, que de su casa a la parada de autobús tarde menos 12 minutos y que también de la parada de autobús a la escuela tarde menos de 12 minutos.

c) Suponga que llegando a la parada de autobús el estudiante inicia inmediatamente el segundo trayecto. ¿Cuál es la probabilidad de que tarde más de 30 minutos en llegar de su casa a la escuela?

d) Calcule las funciones de densidad marginales de X y Y. ¿Cómo se distribuyen marginalmente las variables aleatorias X y Y?

E118. Sean X y Y variables aleatorias con función de densidad de probabilidad conjunta dada por $f_{X,Y}(x,y)=2x(x-y)I_{(0,1)}(x)I_{(-x,x)}(y)$. Obtenga las funciones de densidad de probabilidad marginales de X y de Y.

E119. En la siguiente tabla se presentan las distribuciones conjuntas y marginales de las variables aleatorias X y Y, donde X es el número de años de estudio concluidos por el jefe de familia y Y es el estrato de ingreso del jefe de familia (según veces el salario mínimo general).

X	Y					Total
	1	2	3	4	5	
0	0.199	0.124	0.122	0.005	0	0.450
3	0.177	0.034	0.009	0.003	0	0.223
6	0.008	0.025	0.040	0.049	0.065	0.187
9	0.002	0.005	0.022	0.041	0.071	0.141
Total	0.386	0.188	0.193	0.098	0.136	1

a) Si se elige a un jefe de familia del estrato de ingreso más alto, ¿cuántos años de estudio se espera que haya concluido?

b) Calcule $Var[X \mid Y=5]$.

c) Calcule y grafique $E[Y \mid X=x]$ para $x = 0, 3, 6, 9$. ¿Qué puede concluir?

E120. Suponga que la función de densidad conjunta de X y Y está dada por

$$f_{X,Y}(x,y)=\begin{cases}\frac{1}{y}\exp\left\{-\frac{x+y^2}{y}\right\} & si \;\; x>0,\;\; y>0\\ 0 & e.o.c.\end{cases}$$

a) Obtenga la función de densidad condicional de X, dado $\{Y=y\}$.

b) Calcule $P[X>1 \mid Y=y]$

c) Grafique $E[X \mid Y=y]$ y $Var[X \mid Y=y]$

E121. Considere nuevamente las distribuciones conjuntas y marginales de las variables aleatorias X y Y del ejercicio E119. ¿Son X y Y variables aleatorias independientes?

E122. Suponga que X y Y son variables aleatorias independientes con $X\sim Po(\lambda)$ y $Y\sim Po(2\lambda)$.

a) Calcule la función de masa de probabilidad conjunta de X y Y.

b) Calcule $P[X+Y\leq 1]$

E123. Suponga que U, V, X y Y son variables aleatorias tales que:

$$f_{U,V}(u,v)=\begin{cases}\frac{1}{6}e^{-\left(\frac{3u+2v}{6}\right)} & si \;\; u>0,\;\; v>0\\ 0 & e.o.c.\end{cases} \quad \text{y}$$

$$f_{X,Y}(x,y)=\begin{cases}24xy & si \;\; 0<x<1,\;\; 0<y<1 \;\; y \;\; 0<x+y<1\\ 0 & e.o.c.\end{cases}$$

Utilice el Teorema de independencia vía factorización para determinar si…

a) …U y V son variables aleatorias independientes.

b) …X y Y son variables aleatorias independientes.

E124. Suponga que X_1, X_2,…, X_n es una muestra aleatoria de una Distribución Normal con media μ y varianza σ^2. Obtenga la función de densidad de probabilidad conjunta de esta muestra aleatoria.

E125. Sean Y_1 y Y_2 las proporciones del tiempo, en un día de trabajo, que los empleados I y II ocupan respectivamente en hacer sus tareas asignadas. El comportamiento de las frecuencias relativas conjuntas de Y_1 y Y_2 se representan por el modelo de la función de densidad

$$f_{Y_1,Y_2}(y_1,y_2)=\begin{cases} y_1+y_2 & si\ \ 0<y_1<1,\ 0<y_2<1 \\ 0 & e.o.c. \end{cases}$$

a) Calcule e interprete $E\left[\dfrac{Y_1+Y_2}{2}\right]$.

b) Calcule $E[Y_1]$ y E[Y_2] y verifique que $E\left[\dfrac{Y_1+Y_2}{2}\right]=\dfrac{E[Y_1]+E[Y_2]}{2}$.

E126. Una inmobiliaria ha determinado que si X es el número de habitaciones en los departamentos que maneja y Y el número de lugares de estacionamiento, entonces su distribución conjunta está dada por:

X	Y			Total
	0	1	2	
1	0.20	0.15	0.00	0.35
2	0.15	0.20	0.05	0.40
3	0.05	0.10	0.10	0.25
Total	0.40	0.45	0.15	1.00

a) Calcule la covarianza entre las variables aleatorias X y Y.

b) ¿Son X y Y variables aleatorias independientes?

c) Calcule el coeficiente de correlación lineal entre las variables aleatorias X y Y. Interprete su resultado.

E127. Las variables aleatorias X y Y tienen función de densidad conjunta dada por:

$$f_{X,Y}(x,y)=\begin{cases} \frac{2}{3}(x+y)e^{-x} & si\ \ x>0,\ 0<y<1 \\ 0 & e.o.c. \end{cases}$$

a) Calcule $Cov[X, Y]$.

b) Calcule e interprete $\rho_{X,Y}$.

c) Construya la matriz de varianza-covarianza y la matriz de correlaciones del vector aleatorio (X, Y).

E128. Suponga que X y Y son variables aleatorias continuas con función de densidad de probabilidad conjunta $f_{X,Y}(x, y)$, y que α, $\beta \in \mathbf{R}$. Demuestre que la esperanza es un operador lineal, es decir, que $E[\alpha X \pm \beta Y]=\alpha E[X]\pm \beta E[Y]$.

E129. Sabemos que, por definición, $Cov[X,Y] = E[(X - E[X])(Y - E[Y])]$. Demuestre mediante las propiedades de la esperanza que $Cov[X,Y] = E[XY] - E[X]E[Y]$.

E130. Suponga que X y Y son variables aleatorias y que $\alpha, \beta \in \mathbf{R}$. Utilice las propiedades de la esperanza para demostrar que $Var[\alpha X \pm \beta Y] = \alpha^2 Var[X] + \beta^2 Var[Y] \pm 2\alpha\beta\, Cov[X,Y]$.

E131. Suponga que X y Y son variables aleatorias continuas con función de densidad de probabilidad $f_{X,Y}(x, y)$. Si X y Y son independientes y g y h son funciones arbitrarias de $\mathbf{R}$ en $\mathbf{R}$, demuestre que $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.

E132. Demuestre que si X y Y son variables aleatorias independientes entonces son no correlacionadas, es decir, $Cov[X,Y] = 0$.

E133. Cierta universidad aplica pruebas de aptitudes en ciencias y humanidades a todos los alumnos de primer ingreso. Si X y Y son las proporciones de respuestas correctas que obtiene un estudiante en las pruebas de ciencias y humanidades, respectivamente, entonces $f_{X,Y}(x, y) = \frac{2}{5}(2x + 3y)I_{(0,1)}(x)I_{(0,1)}(y)$.

a) Demuestre que $E[X^j Y^k] = \frac{4}{5(k+1)(j+2)} + \frac{6}{5(k+2)(j+1)}$, $j, k = 0, 1, 2,\ldots$

b) Calcule e interprete el coeficiente de correlación entre X y Y.

c) Si un alumno se inscribe a una licenciatura en el área de humanidades, su calificación final en el examen de admisión está dada por $C = 30X + 70Y$. Calcule esperanza y varianza de la calificación C.

E134. Suponga que X_i y Y_j son variables aleatorias y que α_i, $\beta_j \in \mathbf{R}$, para i, j = 1 y 2. Demuestre que $Cov[\alpha_1 X_1 + \beta_1, \alpha_2 X_2 + \beta_2] = \alpha_1\alpha_2 Cov[X_1, X_2]$ ¿Qué puede concluir?

E135. La función de densidad conjunta de las variables aleatorias X y Y está dada por

$$f_{X,Y}(x, y) = \begin{cases} e^{-x} & si\ \ 0 < y < x < \infty \\ 0 & e.o.c. \end{cases}$$

a) Calcule la función generadora de momentos conjuntos de X y Y.

b) A partir de $M_{X,Y}(s, t)$ calcule la covarianza entre X y Y.

E136. Suponga que X y Y son variables aleatorias Poisson independientes con parámetros λ_1 y λ_2, respectivamente. Calcule la distribución de $X + Y$.

E137. Sean $U_i \sim U(0,1)$, $i = 1, 2$, variables aleatorias independientes.

a) Si $S = U_1 + U_2$, calcule $f_S(s)$. Compruebe que $E[S] = E[U_1] + E[U_2]$

b) Si $T = U_1U_2$, calcule $f_T(t)$. Compruebe que, como U_1 y U_2 son variables aleatorias independientes entonces, $E[T] = E[U_1]E[U_2]$.

E138. Demuestre que si $X \sim Gamma(s, \beta)$ y $Y \sim Gamma(t, \beta)$ son variables aleatorias independientes, entontes $X + Y \sim Gamma(s + t, \beta)$.

E139. Suponga que $X_1, X_2,\ldots, X_n$ son variables aleatorias independientes.

a) Si $X_i \sim Exp(\beta)$, $i = 1, 2,\ldots, n$, ¿cómo se distribuye $S = \sum_{i=1}^{n} X_i$?

b) Si $X_i \sim \chi^2(\nu_i)$, $i = 1, 2,\ldots, n$, ¿cómo se distribuye $S = \sum_{i=1}^{n} X_i$?

E140. Considere que $X \sim Bin(m, p)$ y que $Y \sim Bin(n, p)$ son variables aleatorias independientes. Determine la distribución de $X + Y$.

E141. Suponga que $X_1, X_2,\ldots, X_n$ son variables aleatorias independientes, que $X_i \sim N\left(\mu_i, \sigma_i^2\right)$, y que α_i, $\in \mathbf{R}$, $i = 0\ 1, 2,\ldots, n$.

a) Demuestre que $Y = \sum_{i=1}^{n} \alpha_i X_i \sim N\left(\mu, \sigma^2\right)$ donde $\mu = \sum_{i=1}^{n} a_i \mu_i$ y $\sigma^2 = \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2$.

b) Determine la distribución de $S = \sum_{i=1}^{n} X_i$.

E142. Una refinería ubicada en la costa norte del Golfo de México compra petróleo crudo de México para producir gasolinas. El crudo que compra es una mezcla crudo Maya (60%) y crudo Olmeca (40%). Suponga que el precio de estos crudos son variables aleatorias Normales e independientes con medias de 85 y 110 dólares por barril para los crudos Maya y Olmeca, respectivamente; y que en ambos casos la desviación estándar es de 5 dólares por barril. Si la refinería desea adquirir 200 mil barriles de crudo mexicano y para ello además requiere pagar un costo de transporte fijo de 2 millones de dólares:

a) Determine la distribución del costo que enfrenta la refinería.

b) ¿Cuál es la probabilidad de que el costo sea superior a 22 millones de dólares?

c) ¿Cuál es el costo mínimo que debe considerar la refinería que debe pagar para estar 95% seguro de esa cantidad?

E143. Suponga que X_1 y X_2 son variables aleatorias independientes y que $X_i \sim U(0, 1)$, $i = 1, 2$. Sean $Y_1 = X_1 + X_2$ y $Y_2 = X_1 - X_2$. Encuentre la función de densidad conjunta de las variables aleatorias Y_1 y Y_2.

E144. Considere que $X_1 \sim Gamma(\alpha_1, \beta)$ y que $X_2 \sim Gamma(\alpha_2, \beta)$ son variables aleatorias independientes. Suponga que $Y_1 = X_1 + X_2$ y que $Y_2 = \dfrac{X_1}{X_2}$.

a) Calcule la función de densidad conjunta de Y_1 y Y_2.

b) Verifique cómo $Y_1 \sim Gamma(\alpha_1 + \alpha_2, \beta)$.

E145. Suponga que $(X_1, X_2) \sim N\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\right)$.

a) Demuestre que $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$. Infiera la distribución de X_2.

b) Demuestre que $X_1 \mid X_2 = x_2 \sim N\left(\mu_1 + \rho\dfrac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2\left(1-\rho^2\right)\right)$. ¿Cuál es la distribución condicional de X_2 dado que $X_1 = x_1$?

E146. Si $(X_1, X_2) \sim N\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\right)$, demuestre que si $\rho = 0$, entonces X_1 y X_2 son variables aleatorias independientes. ¿Se puede generalizar este resultado para cualquier distribución?

E147. Suponga que X_i denota la tasa de rendimiento anual que otorga el fondo de inversión i, $i = 1, 2$. Sabemos que el vector aleatorio (X_1, X_2) tiene una Distribución Normal Bivariada con vector de medias $\underline{\mu}$ y matriz de varianzas-covarianzas Σ definidos por:

$$\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 0.06 \\ 0.08 \end{pmatrix} \quad \text{y} \quad \Sigma = \begin{pmatrix} 0.01 & -0.005 \\ -0.005 & 0.04 \end{pmatrix}$$

Un inversionista quiere invertir 200 mil pesos distribuidos entre los fondos 1 y 2.

a) ¿Qué fondo tiene mayor probabilidad de dar al inversionista una tasa de rendimiento mayor al 10%?

b) Si el fondo 1 otorga una tasa de rendimiento del 7%, ¿cuál es la tasa de rendimiento esperada del fondo 2?

c) Si el fondo 2 otorga una tasa de rendimiento del 10%, ¿cuál es la probabilidad de que el fondo 1 tenga una tasa de rendimiento menor al 6%?

d) El inversionista decide invertir 40% de su capital en el fondo 1 y el resto en el fondo 2. ¿Cuál es la probabilidad de que el capital del inversionista aumente más de 10% en un año?

E148. Suponga que $\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N_3\left(\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 & 0 & -1 \\ 0 & 2 & 1 \\ -1 & 1 & 4 \end{pmatrix}\right)$, $Y_1 = 2X_1 - X_2$ y $Y_2 = 3X_1 - 2X_3$.

a) Encuentre las distribuciones marginales y conjunta de Y_1 y Y_2.

b) Si $\underline{Z} = \left(X_1, X_3\right)^t$, encuentre la distribución condicional de $\left\{\underline{Z} \middle| X_2 = 1\right\}$.