UNIVERSIDAD NACIONAL DE TUCUMAN
Facultad de Ciencias Económicas
Instituto de Investigaciones Estadísticas (INIE)

CLASES DE

ESTADÍSTICA INFERENCIAL

LICENCIATURA EN ADMINISTRACIÓN
LICENCIATURA EN ECONOMÍA

Dra. Viviana Lencina

I. INTRODUCCIÓN

I.1 Introducción

En el proceso de conocer la realidad, la obtención, descripción y análisis de la información se vuelve de vital importancia para el diseño de cualquier tipo de política, toma de decisiones, implementación de estrategias, programas o proyectos empresariales.

La Estadística es una disciplina que brinda métodos y procedimientos para obtener, resumir, describir, analizar e interpretar un conjunto de datos, y se constituye como una herramienta fundamental cuando se intenta alcanzar el conocimiento científico de una realidad político/socio/económica, y en el proceso de poner a prueba algunas hipótesis de trabajo, mediante la formulación de diversos diseños de investigación, el ajuste y especificación de modelos teóricos complejos, permitiendo dar base sólidas para la elaboración de interpretaciones útiles y veraces que ayuden a comprender los fenómenos político-sociales de interés.

I. 2 ¿Por qué se estudia Estadística?

Es reconocido por todos en el mundo actual la aplicabilidad cada vez mayor de la Estadística en el área de negocios y la utilidad de este conocimiento para los profesionales. Consecuentemente, esta disciplina se encuentra en los planes de estudio de todas las carreras de Ciencias Económicas, debido a la necesidad creciente de su aplicación en la vida profesional de los egresados.

Anderson et al (2008)[1] manifiestan que "*especialmente en los negocios y en la economía, una razón básica para esa recopilación, análisis, presentación e interpretación de datos es proporcionar a los administradores y a quienes toman decisiones una mejor comprensión del entorno comercial y económico, para permitirles tomar mejores decisiones y más informadas*".

En 1810, Belgrano escribió en el Correo del Comercio "*Nada más importante que tener un conocimiento exacto de la riqueza y fuerza de los Estados; éste es el objeto de la ciencia Estadística, y su fin para proceder con acierto en todas las disposiciones que se dirijan al orden económico a efecto de fomentar la Agricultura, animar la Industria y proteger el Comercio, como que son los arcos torales de la felicidad pública. Esos datos son necesarios, son útiles, y en vano es creer que sin ellos se puedan tratar con acierto, según que lo desean nuestros jefes, las materias interesantes a la causa común del Estado; sin conocimientos de la fortuna pública, de las necesidades y recursos de estas Provincias, no es posible que se dicten las providencias más convenientes a la felicidad general*".

En su discurso presidencial de 1951, el estadístico matemático Samuel S. Wilks[2] (1906-1964) ante la Asociación Estadounidense de Estadística dijo: *¡Un día el pensamiento estadístico será tan necesario para una ciudadanía eficiente como la capacidad de leer y escribir!* Con el continuo progreso tecnológico, en las comunicaciones y la electrónica, el

[1] Anderson, Sweeney & Williams (2008): Estadística para Administración y Economía, 10ª edición.

[2] Wilks (1951): Discurso Presidencial. JASA, vol. 46, n.º 253, págs. 1-18.

uso de métodos estadísticos como ayuda para la toma de decisiones, frente a información incierta y con muestras imperfectas, ha tenido un crecimiento dinámico y ha de continuar creciendo, siendo la estadística quien brinda los fundamentos para los modelos que se aplican en inteligencia artificial.

I.3. Algunos problemas que resuelve la Estadística

La Estadística es el cuerpo integrado de métodos para adoptar decisiones correctas en situaciones de **incertidumbre**. Esta toma de decisiones se basa en **datos observados**. Al usar las observaciones para tomar decisiones se está expuesto al riesgo de equivocarse puesto que se cuenta con información limitada. La Estadística como disciplina se utiliza en dos tipos de situaciones, la adquisición de conocimientos, Inferencia, donde en base a datos que en general son incompletos, insuficientes o parciales, se pretende lograr un conocimiento general, y la acción práctica, donde se pretende adoptar una decisión.

El creciente uso de datos estadísticos en la Administración de Empresas es parte de la tendencia a basar las decisiones administrativas en un fundamento tan objetivo como resulte posible.

En esto ha jugado un papel preponderante la facilidad de acceso a equipos de computación. En el proceso de la toma de decisiones para fines gerenciales, es imprescindible la utilización de técnicas estadísticas dirigidas a la recopilación, presentación y análisis de datos que ayuden en el manejo de los registros de las empresas.

Descripción de datos: Una empresa puede estar interesada en estudiar el tiempo que sus empleados invierten en realizar un determinado conjunto de tareas. Tener esta información permitirá evaluar y mejorar el manejo del tiempo en la empresa.

Otros ejemplos: Auditoria de cuentas por cobrar para verificar la exactitud de las mismas, control de calidad de un producto ingresado para ser empaquetado, resúmenes del total de venta de ciertos productos, etc.

Análisis de Muestras: Se utiliza para elegir una muestra representativa y para hacer inferencias respecto a la población a partir de lo observado en la muestra. Decidir si un proceso industrial funciona o no adecuadamente de acuerdo con el cumplimiento de las especificaciones en un lote seleccionado al azar, es un ejemplo de aplicación de este tipo de análisis. Otros ejemplos: Juzgar la demanda potencial de un producto mediante un estudio de mercado, dar una medida de la variación de los precios en ciertos rubros, determinar los horarios de mayor atención en un banco para decidir la cantidad de cajeros atendiendo.

Contrastación de Hipótesis: Poner a prueba alguna hipótesis ante la evidencia suministrada por los datos observados. Permite responder preguntas del tipo ¿Ha mejorado un proceso de fabricación después de calibrarlo? ¿Las ventas aumentaron después de las actividades de promoción realizadas? ¿Son efectivos el cinturón de seguridad o la limitación de velocidad para reducir las muertes por accidente? ¿El ausentismo entre las docentes mujeres es mayor al de los varones?

Medición de Relaciones: Evaluar la existencia o no de relaciones estadísticas entre dos o más características de interés, para poder responder interrogantes del tipo: ¿Cuáles son las características de los viajes o traslados que más influyen en el consumo de gasoil de los camiones de una empresa de transporte? ¿Cómo se relaciona el rendimiento laboral con variables sociológicas y ambientales de empresas instaladas en parques industriales? ¿Cuál es la relación entre el desempleo y la inflación?

Predicción: Teniendo en cuenta la estructura presente en los datos, y pensando que ésta refleja la estructura de una población, se puede predecir la probabilidad de ocurrencia de ciertos acontecimientos en la población. Por ejemplo, la demanda de un producto en los meses de enero y febrero; los gastos partidarios en los meses próximos a elecciones; el costo de una canasta navideña para el próximo fin de año, el consumo de combustible en traslados bajo determinadas características.

Específicamente, en el área de Economía y Administración podemos mencionar algunos ejemplos que aparecen en el libro de Anderson, Sweeney y Williams (2008).

Finanzas

Los analistas financieros emplean una diversidad de información estadística como guía para sus recomendaciones de inversión. En el caso de acciones, el analista revisa diferentes datos financieros como la relación precio/ganancia y el rendimiento de los dividendos. Al comparar la información sobre una determinada acción con la información sobre el promedio en el mercado de acciones, el analista empieza a obtener conclusiones para saber si una determinada acción está sobre o subvaluada. Por ejemplo, Barron's (12 de septiembre de 2005) informa que la relación promedio precio/ganancia de 30 acciones del promedio industrial Dow Jones fue 16.5. La relación precio/ganancia de JPMorgan es 11.8. En este caso la información estadística sobre las relaciones precio/ganancia indican un menor precio en comparación con la ganancia para JPMorgan que el promedio en las acciones Dow Jones. Por tanto el analista financiero concluye que JPMorgan está subvaluada. Ésta y otras informaciones acerca de JPMorgan ayudarán al analista a comprar, vender o a recomendar mantener las acciones.

Marketing

Escáneres electrónicos en las cajas de los comercios minoristas recogen datos para diversas aplicaciones en la investigación de mercado. Por ejemplo, proveedores de datos como ACNielsen e Information Research Inc. compran estos datos a las tiendas de abarrotes, los procesan y luego venden los resúmenes estadísticos a los fabricantes; quienes gastan cientos de miles de dólares por producto para obtener este tipo de datos. Los fabricantes también compran datos y resúmenes estadísticos sobre actividades promocionales como precios o displays promocionales. Los administradores de marca revisan estas estadísticas y las propias de las actividades promocionales para analizar la relación entre una actividad promocional y las ventas. Estos análisis suelen resultar útiles para establecer futuras estrategias de marketing para diversos productos.

Producción

La importancia que se le da actualmente a la calidad, hace del control de calidad una aplicación importante de la Estadística a la producción. Para vigilar el resultado de los procesos de producción se usan diversas gráficas de control estadístico de calidad. En particular, para vigilar los resultados promedio se emplea una gráfica x-barra. Suponga, por ejemplo, que una máquina llena botellas con 12 onzas de algún refresco. Periódicamente un empleado del área de producción toma una muestra de botellas y mide el contenido promedio de refresco. Este promedio o valor x-barra se marca como un punto en una gráfica x-barra. Si este punto queda arriba del límite de control superior de la gráfica, hay un exceso en el llenado, y si queda debajo del límite de control inferior de la gráfica hay falta de llenado. Se dice que el proceso está "bajo control" y puede continuar, siempre que los valores x-barra se encuentren entre los límites de control inferior y superior. Con una interpretación adecuada, una gráfica de x-barra ayuda a determinar si es necesario hacer algún ajuste o corrección a un proceso de producción.

Economía

Los economistas suelen hacer pronósticos acerca del futuro de la economía o sobre algunos de sus aspectos. Usan una variedad de información estadística para hacer sus pronósticos. Por ejemplo, para pronosticar las tasas de inflación, emplean información estadística sobre indicadores como el índice de precios al consumidor, la tasa de desempleo y la utilización de la capacidad de producción. Estos indicadores estadísticos se utilizan en modelos computarizados de pronósticos que predicen las tasas de inflación.

I.4. Estadística y Probabilidades

En el proceso de adquirir conocimiento con respecto a un fenómeno, en general se hace necesario conocer qué reglas moderan su comportamiento. Por reglas se interpreta, por ejemplo, qué posibilidades tiene de ocurrir cada uno de los posibles resultados de ese fenómeno. Una forma de describir el fenómeno es adoptando un modelo probabilístico para todos los posibles resultados. De esta manera al estudiar Estadística se hace necesario conocer Teoría de Probabilidades, puesto que ésta brinda la base matemática para el desarrollo de diversas técnicas estadísticas.

Para comparar el tipo de razonamiento utilizado en Probabilidades con el que se usa en Estadística se pueden observar la Figura I.1 y la Figura I.2.

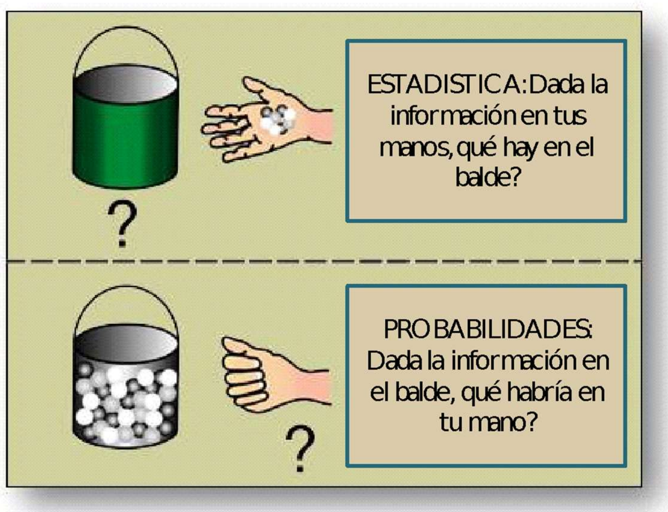


Figura I.1: Preguntas de interés en Probabilidades y Estadística

En Probabilidades se asume conocido un modelo probabilístico y se puede inducir el comportamiento que tendrán los datos cuando el esquema en que fueron seleccionados corresponde con un muestreo probabilístico o al azar. En Estadística, se suponen conocidos los datos y se centra el interés en la población de donde éstos fueron seleccionados.

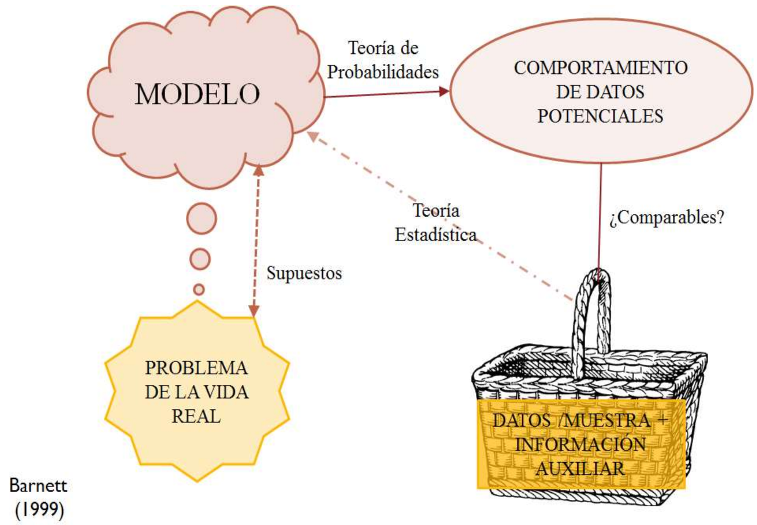


Figura I.2: Esquema del razonamiento utilizado en Estadística

Generalmente, las preguntas de interés en el problema de la vida real que se pretenda analizar, son traducidas o interpretadas desde la óptica de los parámetros o propiedades del modelo probabilístico asumido. El aplicar una metodología estadística, en el fondo consiste en contrastar lo que se observó al tomar una muestra con lo que se observaría con el modelo asumido, o datos potenciales. Si hay coherencia entre ambos se piensa que el modelo es adecuado para representar el proceso bajo estudio y se usa la información de ese modelo para intentar responder las preguntas planteadas. En caso contrario, se duda del modelo y se intenta buscar otro más adecuado.

1. Experimento Aleatorio

1.1. Espacio muestral

Un **Experimento Aleatorio** es un proceso que repetido bajo un mismo conjunto de condiciones controladas no conduce siempre a un mismo resultado.

Por experimento se entiende tanto a una experiencia que sirva para hacer operaciones destinadas a descubrir, comprobar o demostrar determinados fenómenos o principios científicos, como a cualquier situación o acontecimiento donde se quiera conocer algo. Por ejemplo, cuando se quiere calcular el precio medio de una canasta de útiles escolares en un mes, los precios de los productos de la canasta pueden variar según el tipo de negocio donde se vende el producto, la forma de pago, la posición relativa del momento en que se averigua el precio con relación al inicio de clases, etc.

En cualquier experimento aleatorio, si bien no se conoce cuál será el resultado, se puede definir el conjunto de todos los posibles resultados, siendo que, en cualquier repetición simple del experimento, ocurrirá uno y sólo uno de ellos. Ese conjunto se denomina **Espacio Muestral** y se denota con la letra $\mathcal{S}$.

Espacio Muestral ($\mathcal{S}$): Es el conjunto de todos los posibles resultados de un experimento aleatorio.

A modo de ejemplo se puede observar lo presentado en la Tabla 1.1. Cada uno de los experimentos allí mencionados puede considerarse un experimento aleatorio puesto que no se sabe exactamente cuál será el resultado en cualquier repetición simple del experimento.

Tabla 1.1: Ejemplos de experimentos y posibles espacios muestrales

	Experimento		Espacio Muestral $\mathcal{S}$
1	Se lanza una moneda 3 veces y se cuenta el número de caras.		$\mathcal{S}_1 = \{0,1,2,3\}$
2	Se observa el número de clientes atendidos en una sucursal del Banco A, entre las 9:30 y las 10:00.		$\mathcal{S}_2 = \{0,1,2,3,...\} = \mathbb{N}_0$
3	Se seleccionan tres alumnos egresados 2010 de la FCE/UNT y se les pregunta qué posibilidades tienen conseguir trabajo en el primer hasta julio de 2011.		$\mathcal{S}_3 = \{(x_1, x_2, x_3) : x_i \in [0,1], i = 1,2,3\}$

Puede haber más de una manera de especificar el espacio muestral de un experimento aleatorio, y, en general, estos pueden diferir en el nivel de información que presentan. Por ejemplo, en el caso del experimento 1 presentado en la Tabla 1, podría haberse considerado el espacio muestral $\mathcal{S}_1' = \{(x_1, x_2, x_3) : x_i \in \{\text{cara}, \text{sello}\}, i = 1,2,3\}$, el nivel de información de $\mathcal{S}_1'$ difiere de la información provista en $\mathcal{S}_1$, se puede observar que en $\mathcal{S}_1'$ hay mayor nivel de información, y que cada elemento de $\mathcal{S}_1$ se corresponde con al menos un elemento de $\mathcal{S}_1'$.

1.2. Eventos

Acontecimiento, evento o suceso: Es un subconjunto del espacio muestral.

Se denomina **evento**, **acontecimiento** o **suceso** a cualquier subconjunto del espacio muestral, y se los denotan con letras mayúsculas de molde. Los eventos pueden definirse por extensión o por comprensión. Por ejemplo, $A = \{0,1,2,3,4,5\} \subset \mathcal{S}_2$ y el conjunto $B = \{\text{el número de clientes atendidos no supera a 5}\} \subset \mathcal{S}_2$, son los mismos eventos enunciados de diferentes maneras.

De la definición de eventos, se deduce que el vacío o evento nulo $\varnothing$ y el propio espacio muestral, $\mathcal{S}$, también son eventos.

Otros eventos especiales a mencionar, son aquellos conjuntos formados por cada elemento del espacio muestral, $\mathcal{S}$. Estos eventos se denominan eventos elementales.

Acontecimiento o suceso elemental: Se denomina así al conjunto formado por un único elemento del espacio muestral.

Debido a que el espacio muestral $\mathcal{S}$ y los diversos acontecimientos son conjuntos, para trabajar con eventos se pueden utilizar todos los conceptos (ver Tabla 1.2) y propiedades (ver Tabla 1.4) de la Teoría de Conjuntos.

Tabla 1.2: Conceptos de la Teoría de Conjuntos

	Concepto	**Notación**	**Definición**
1	Complemento de un evento	A^c ó $\overline{A}$	$A^c = \{x \in \mathcal{S} : x \notin A\}$
2	Intersección de dos eventos	$A \cap B$	$A \cap B = \{x \in \mathcal{S} : x \in A \text{ y } x \in B\}$
3	Unión de dos eventos	$A \cup B$	$A \cup B = \{x \in \mathcal{S} : x \in A \text{ o } x \in B\}$
4	Diferencia entre dos eventos	$A - B$	$A - B = \{x \in \mathcal{S} : x \in A \text{ y } x \notin B\}$

Ejemplo 1.1: Una universidad está pensando en adherirse a un proceso de evaluación institucional que consta de dos etapas, la primera etapa denominada de "Autoevaluación" y la segunda de "Evaluación Externa". De experiencias en otras universidades se sabe que la etapa de autoevaluación puede demorar 4, 5 o 6 meses y la etapa de evaluación externa demora 1, o 2 meses. Las duraciones posibles en este tipo de proceso son entre 5 y 8 meses. La Tabla 1.3 presenta todos los resultados posibles al realizar este proceso.

Tabla 1.3: Resultados posibles en el proceso de evaluación institucional

Autoevaluación	Evaluación externa	Resultados posibles	Duración proceso
4 meses	1 mes	(4,1)	5 meses
	2 meses	(4,2)	6 meses
5 meses	1 mes	(5,1)	6 meses
	2 meses	(5,2)	7 meses
6 meses	1 mes	(6,1)	7 meses
	2 meses	(6,2)	8 meses

Sea A:“ El proceso dura a lo sumo 6 meses”, entonces A={(4,1),(4,2),(5,1)}, el complemento de A, A^c={(5,2),(6,1),(6,2)}. Sea B:”se invierte al menos 5 meses en la autoevaluación, entonces $A \cap B$ ={(5,1)}.

Para facilitar el trabajo con eventos, es importante recordar las propiedades clásicas de los conjuntos que se presentan en la Tabla 1.4.

Tabla 1.4: Propiedades de Conjuntos

Propiedades del evento vacío o nulo	$A \cap \varnothing = \varnothing$
	$A \cup \varnothing = A$
Propiedades de los complementos	$A \cap A^c = \varnothing$
	$A \cup A^c = \mathcal{S}$
	$\mathcal{S}^c = \varnothing$
	$\left(A^c\right)^c = A$
	$\varnothing^c = \mathcal{S}$
Leyes de Morgan	$\left(A \cup B\right)^c = A^c \cap B^c$
	$\left(A \cap B\right)^c = A^c \cup B^c$

En algunas situaciones es importante identificar eventos que no pueden ocurrir simultáneamente. Por ejemplo, el evento donde todos los egresados 2010 seleccionados piensan que la posibilidad de conseguir trabajo es menor que el 50%, con el evento de al menos uno de los egresados seleccionados piensa que seguramente conseguirá trabajo. Estos tipos de eventos se denominan eventos mutuamente excluyentes o disjuntos, puesto que la ocurrencia de uno implica la no ocurrencia del otro. Formalmente, se los define como sigue:

Dos eventos A y B son **mutuamente excluyentes** o **disjuntos** si y solo si $A \cap B = \varnothing$.

1.3. Conteo

Al intentar describir la ocurrencia de eventos o acontecimientos de un experimento aleatorio, con frecuencia es necesario decir cuántos resultados son compatibles con un determinado evento, definido conceptualmente. Para poder responder a la pregunta del "cuántos", es importante explorar herramientas y técnicas de conteo, que brindan las bases para contar el número de formas en que se pueden organizar objetos, seleccionar elementos de un conjunto o realizar una serie de tareas. Esto no solo es un ejercicio matemático interesante, sino que también es fundamental para comprender conceptos más avanzados en Probabilidad y Estadística.

Cuando se habla de la posibilidad de que ocurran eventos, es importante entender cómo contar el número total de resultados posibles. Cuando se quiere calcular la probabilidad de ganar la lotería, lo primero que se debería saber es cuántas combinaciones de números existen. La teoría de conteo proporciona las herramientas para hacer precisamente eso.

Además, el conteo no se limita a juegos de azar o acertijos matemáticos. Tiene aplicaciones prácticas en diversos campos:
Informática: El conteo es esencial para analizar la eficiencia de algoritmos y la complejidad de estructuras de datos.
Criptografía: La seguridad de las contraseñas y los sistemas de cifrado depende de la cantidad de combinaciones posibles.
Genética: El conteo se utiliza para determinar el número de posibles combinaciones genéticas.
Estadística: El conteo es necesario para calcular probabilidades y analizar datos. Que es el principal motivo para abordar ahora este tema.

En situaciones donde un experimento se realiza en dos o más etapas, la manera de cuantificar la cantidad de resultados posibles de ese experimento, depende de cuántos resultados posibles hay en cada etapa. Por ejemplo: Si se piensa en todas las posibles claves alfanuméricas que se pueden generar con una letra y un dígito.

La primera posición puede ser una letra mayúscula (26 opciones).

La segunda posición puede ser un dígito numérico (10 opciones).

La cantidad de claves posibles sería $26 \times 10 = 260$, puesto que cada una de las letras puede ir acompañada por cada uno de los dígitos. O sea, hay 260 contraseñas posibles en este sistema.

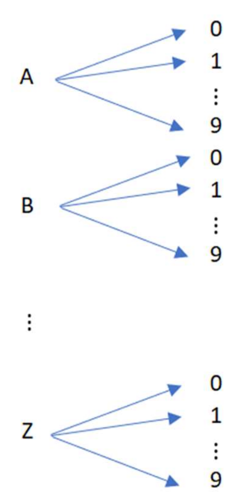


Figura 1.1: Esquema para contabilizar las claves posibles

Esta forma de razonar, se corresponde a lo que se denomina como principio de la multiplicación.

Principio de la multiplicación: En un experimento aleatorio que se realiza en dos partes o etapas, donde la primera parte tiene n_1 resultados posibles, y por cada resultado posible de la primera parte, la segunda parte tiene n_2 resultados posibles, el espacio muestral tendrá exactamente $n_1 \times n_2$ resultados.

Este principio se generaliza fácilmente cuando el experimento está compuesto por más de dos etapas, siendo que todas etapas deben completarse para completar el experimento aleatorio.

Existen situaciones donde un mismo experimento puede llevarse a cabo de dos o más maneras posibles, mutuamente excluyentes. Por ejemplo: Maneras de trasladarse entre dos ciudades, cuando se tienen las siguientes opciones de transporte:

Por tierra:
- Autobús: 3 rutas disponibles.
- Tren: 2 rutas disponibles.

Por aire:
- Vuelo directo: 4 opciones.

Dado que no se puede viajar por tierra y aire al mismo tiempo, para contabilizar las maneras de trasladarse entre esas dos ciudades serían 5(opciones por tierra)+4(opciones aéreas)=9 opciones.

En situaciones como estas, se aplica el denominado Principio de la Adición.

Principio de la adición: En un experimento aleatorio que se puede realizar de dos maneras mutuamente excluyente, donde la primera manera tiene n_1 resultados posibles, la segunda tiene n_2 resultados posibles, el espacio muestral tendrá exactamente n_1+n_2 resultados posibles.

Cuando se quiere contabilizar posibles resultados o elementos en un conjunto, es esencial distinguir entre combinaciones y permutaciones (o variaciones), ya que cada concepto responde a una pregunta diferente sobre la disposición de los elementos de un conjunto.

Combinación: Se denomina combinación a una colección o grupo de elementos (el orden no es importante).

Una combinación es un subconjunto de elementos tomado de un conjunto más grande, donde el orden no importa. En otras palabras, si se eligen los mismos elementos, sin importar el orden en que aparecen, se considera la misma combinación.

Ejemplo 1.2: Suponga que tiene tres letras: A, B y C, y se quiere formar grupos de dos letras, las combinaciones posibles son:

AB - AC - BC

Nótese que BA no se cuenta como una opción diferente a AB, ya que el orden no es relevante en las combinaciones.

Variación o permutación: Se denomina variación o permutación a cada secuencia u ordenación de elementos (el orden es importante).

Una variación (o permutación) es una secuencia u ordenación de elementos, donde el orden sí importa. Es decir, si los mismos elementos aparecen en un orden distinto, se considera una variación distinta.

Ejemplo 1.3: Con las mismas letras A, B y C, si quiere formar secuencias de dos letras, las permutaciones posibles son:

AB - BA - AC - CA - BC - CB

Aquí, AB y BA son diferentes porque el orden cambia.

Diferencias Claves:

- Dos combinaciones son diferentes si difieren en el número de elementos o en al menos un elemento.
- Dos variaciones (o permutaciones) son diferentes si difieren en el número de elementos, en algún elemento, o en el orden en que aparecen.

Este concepto es fundamental en muchas aplicaciones de la probabilidad, la estadística y la teoría de conjuntos, ya que permite calcular de manera eficiente el número de formas en que se pueden seleccionar o disponer elementos en distintas situaciones

Basándose en los principios antes mencionados se puede probar que:

1. El número de combinaciones posibles de n elementos tomados de a r, es
$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$
2. El número de variaciones posibles con n objetos distintos es $P_n = n!$.
3. El número de variaciones posibles de n objetos distintos tomados de r a la vez es $V_{n,r} = n!/((n-r)!)$.
4. El número de variaciones posibles con n objetos distintos ubicados en un círculo es $(n-1)!$.
5. El número de variaciones posibles con n objetos no todos distintos, de los cuales $\mathbf{n_1}$ son de un tipo, $\boldsymbol{n_2}$ son de un segundo tipo,…, $\boldsymbol{n_k}$ son de un k-ésimo tipo es
$$\frac{\boldsymbol{n}!}{\boldsymbol{n_1}!\,\boldsymbol{n_2}!\dots\boldsymbol{n_k}!}.$$
6. . El número de formas posibles de partir un conjunto de n objetos en r celdas con$\boldsymbol{n_1}$ elementos en la primera celda,$\boldsymbol{n_2}$ en la segunda,…, $\boldsymbol{n_r}$ elementos en la r-ésima celda es
$$\frac{\boldsymbol{n}!}{\boldsymbol{n_1}!\,\boldsymbol{n_2}!\dots\boldsymbol{n_r}!}.$$

2. Probabilidad

2.1 Introducción

Aún sin conocerse formalmente la definición de probabilidad, la idea de incertidumbre o azar es muy antigua. Hay antecedentes de juegos de azar desde 3500 AC. En los juegos de azar el concepto de chances o posibilidades de ganar se vuelve de importancia en el momento de decidir si conviene o no jugar.

Por ejemplo, en un juego de dados, si se tiene la posibilidad de apostar a favor del evento “sale un número par” o del evento “sale un múltiplo de 3”, claramente, en presencia de un dado justo, se decidirá por la ocurrencia de un número par pensando en que se dispondría de más posibilidades de ganar en ese caso.

Intentar introducir el concepto de probabilidad con un ejemplo de este tipo no es arbitrario. De hecho, en la historia, se considera como el inicio de la teoría matemática de probabilidad a una serie de cartas entre Blaise Pascal y Pierre Fermat, de 1654 a 1660, donde se discutían varios problemas relacionados con apuestas en juegos de azar, planteados a Pascal por Antoine Gombaud, caballero de Méré, un noble francés con un interés en juegos de azar y apuestas.

Uno de los problemas abordados en esta serie de cartas planteaba cómo dividir el monto de una apuesta entre dos jugadores cuando se interrumpe antes de terminar un juego que consistía en arrojar repetidamente una moneda balanceada, uno obtenía un punto cuando salía cara y el otro cuando salía ceca, y siendo que ganaba el juego aquel que obtenga primero cinco puntos.

Otro problema planteado estaba relacionado con una aparente contradicción respecto de un juego que consistía en lanzar un par de dados 24 veces. El interrogante era decidir si convenía o no apostar dinero a la ocurrencia de al menos un "doble seis".

La serie de cartas entre Pascal y Fermat conducen a la formulación de los principios fundamentales de la teoría de la probabilidad por primera vez. Aprendiendo de esta correspondencia, el científico holandés Christian Huygens, profesor de Leibniz, publica en el año 1657 el que puede ser considerado el primer libro sobre probabilidad; denominado *De Ratiociniis en Ludo Aleae*, (El cálculo de las chances en juegos de azar). En 1812, Pierre de Laplace (1749-1827) introdujo una serie de nuevas ideas y técnicas matemáticas en su libro, *Théorie Analytique des Probabilités*, donde aplica las ideas probabilísticas a muchos problemas científicos y prácticos. La teoría de la mecánica estadística, matemática actuarial y errores son ejemplos de algunas de las importantes aplicaciones de la teoría de las probabilidades usadas por primera vez en el siglo l9th.

Una de las dificultades en el desarrollo de una teoría matemática de la probabilidad ha sido llegar a una definición de probabilidad de que sea lo suficientemente precisa, y a su vez lo suficientemente abarcativa para que resulte aplicable a una amplia gama de fenómenos. La búsqueda de una definición ampliamente aceptable tomó cerca de tres siglos y fue marcada por mucha controversia. El asunto se resolvió finalmente en el siglo XX cuando en 1933 el matemático ruso Andrei Kolmogorov propone un enfoque axiomático para la definición de probabilidad, que forma la base para la teoría moderna.

2.2 Definición de Probabilidad

En la historia moderna de la Probabilidad se han conocido varias propuestas para definir el concepto de Probabilidad, que han sido criticadas por expertos transformando la discusión prácticamente en algo filosófico. Entre las definiciones propuestas se pueden mencionar las siguientes.

Definición de Probabilidad de Laplace: Sea un espacio muestral $\mathcal{S}$ y acéptese "a priori" que, por la simetría del experimento aleatorio, una persona razonable se sentiría indiferente con respecto a la preferencia de uno cualquiera de los acontecimientos elementales. Entonces la probabilidad de un acontecimiento es proporcional al número de acontecimientos elementales que lo forman.

Observaciones:

1. Esta definición también lleva el nombre de definición clásica o a priori.
2. Esta definición indica la manera de calcular las probabilidades en el caso de simetría del experimento.
3. Cuando se expresa indiferencia con relación a la preferencia de uno cualquiera de los acontecimientos elementales, se entiende que los acontecimientos elementales tienen la misma posibilidad de ocurrir. Para que esto sea posible, es necesario que los espacios muestrales donde se aplique esta definición sean finitos.
4. Es conveniente poner en evidencia que el conjunto de todos los acontecimientos elementales constituye la más fina partición del espacio muestral. Por partición se entiende que la unión de todos ellos resulta en el propio espacio muestral (completos y exhaustivos) y que la ocurrencia de uno de ellos implica la no ocurrencia de cualquiera del resto (mutuamente excluyentes).
5. Para poder comparar probabilidades de eventos provenientes de espacios muestrales con diferentes tamaños es conveniente proceder a una estandarización. Se puede mostrar, si se conviene en que la probabilidad varíe entre 0 y 1, que la constante de proporcionalidad debe ser la inversa del número de elementos del espacio muestral.
6. Si se considera como $n(A)$ al número de elementos de A y n el número de elementos en el espacio muestral, $P(A)= n(A)/n$.
7. Las probabilidades asignadas usando esta definición, con la normalización correspondiente, satisfacen la idea de que (i) $P(A) \geq 0$, para todo acontecimiento A, (ii) $P(\mathcal{S}) \geq 0$, y (iii) para cualquier par de eventos A y B mutuamente excluyentes $P\left(A \cup B\right) = P(A) + P(B)$
8. Este tipo de definición no es adecuada para asignar probabilidades en experimentos que carezcan de simetría o para responder preguntas del tipo, cuál es la probabilidad de que llueva mañana o cuál es la probabilidad de que el primer nacimiento registrado en el mes de abril de 2015 sea de un varón.
9. La definición provista por Laplace en 1810 en su trabajo "Théorie analytique des probabilités", decía:
   The probability of an event is the ratio of the number of cases favorable to it, to the number of all cases possible when nothing leads us to expect that any one of these cases should occur more than any other, which renders them, for us, equally possible.

Ejemplo 2.1*:* se arroja dos dados y se quiere calcular la probabilidad de que sumen 7.

Si lo que interesa es el resultado de la suma obtenida al arrojar los dos dados se podría usar el siguiente espacio muestral $\mathcal{S}=\{2,3,4,5,6,7,8,9,10,11,12\}$, el inconveniente que aparece al utilizar este espacio es que no podemos ser indiferentes a la ocurrencia de una suma que da 2 o 7, intuitivamente hay más resultados posibles al arrojar un dado que podrían dar como suma 7, $A=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$, mientras que una suma 2 solo se obtiene cuando se observa $B=\{(1,1)\}$. Para poder calcular probabilidades en el contexto de este experimento es conveniente trabajar con el espacio muestral

$$\mathcal{S}_*=\{(i,j): i,j=1,\ldots,6\}$$

$$=\left\{\begin{matrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{matrix}\right\}$$

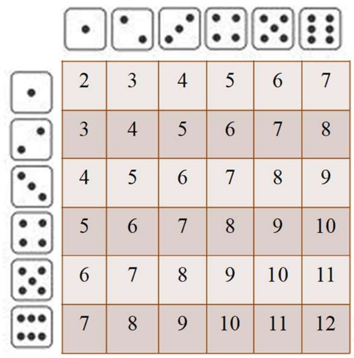


Figura 2.1

En la Figura 2.1 se muestra cómo se relacionan los puntos de $\mathcal{S}_*$ con los puntos de $\mathcal{S}$. Analizando los posibles eventos elementales de $\mathcal{S}_*$, se puede decir que no se tiene preferencia a priori por ninguno de ellos, lo que podría llevar a utilizar la definición de Laplace para adjudicar probabilidades en $\mathcal{S}_*$. Como son 36 resultados posibles, la probabilidad de cada evento elemental será 1/36, y la probabilidad del evento A, donde la suma da siete sería $P(A)=6/36=1/6$, por otro lado $P(B)=1/36$.

Como la definición de Laplace no puede aplicarse cuando no hay simetría, se puede considerar la siguiente definición frecuentista.

Definición Frecuentista de Probabilidad: Si cada vez que se realiza una serie suficientemente grande de repeticiones de un fenómeno aleatorio la razón (frecuencia relativa) del número de veces que un suceso *A* ocurre al número total de repeticiones es aproximadamente p, y si la frecuencia relativa es habitualmente más próxima a p cuando mayor es el número de repeticiones, entonces se acepta de antemano en definir $P(A)=p$.

Observaciones:

1. Esta definición a veces se conoce como a posteriori.
2. Si se considera como $n(A)$ al número de veces que ocurre A al repetir el experimento y n es el número de repeticiones, la frecuencia relativa viene dada por $n(A)/n$.
3. Esta definición presenta algunas limitaciones al no especificar el número de repeticiones que se considera como "suficientemente grande".
4. Para aplicar esta definición se parte del supuesto de que existe el límite.
5. Las probabilidades asignadas usando esta definición, con la normalización correspondiente, satisfacen la idea de que (i) $P(A)\geq 0$, para todo acontecimiento A, (ii) $P(\mathcal{S})\geq 0$, y (iii) para cualquier par de evento A y B mutuamente excluyentes $P\left(A\cup B\right)=P(A)+P(B)$.

6. Hay situaciones donde no es posible realizar una serie grande de repeticiones, y consecuentemente no es posible calcular el límite. Por ejemplo, no permite responder cuál es la probabilidad de que llueva mañana o cuál es la probabilidad de que el primer nacimiento registrado en el mes de abril de 2015 sea de un varón.

Ejemplo 2.2: Para abordar un ejemplo dónde aplicar la definición frecuentista de probabilidad, deberíamos simular la generación de números enteros con algún procedimiento de generación de números aleatorios de tres cifras. Se puede pensar en calcular la probabilidad de (i) A: "el último dígito del documento es impar", o (ii) B: "el número es múltiplo de 6" como el límite de la frecuencia relativa de ocurrencia de cada uno de los eventos *A* y *B*, como se muestra en la Figura 2.2.

n	Número aleatorio	A	n(A)	n(A)/n	B	n(A)	n(A)/n
1	425	Si	1	1.000	No	0	0.000
2	758	No	1	0.500	No	0	0.000
3	588	No	1	0.333	Si	1	1.000
4	547	Si	2	0.500	No	1	0.500
5	878	No	2	0.400	No	1	0.500
6	44	No	2	0.333	No	1	0.500
7	661	Si	3	0.429	No	1	0.333
8	136	No	3	0.375	No	1	0.333
9	112	No	3	0.333	No	1	0.333
10	91	Si	4	0.400	No	1	0.250
11	997	Si	5	0.455	No	1	0.200
12	625	Si	6	0.500	No	1	0.167
13	594	No	6	0.462	Si	2	0.333
14	537	Si	7	0.500	No	2	0.286
15	623	Si	8	0.533	No	2	0.250
16	261	Si	9	0.563	No	2	0.222
17	55	Si	10	0.588	No	2	0.200
18	752	No	10	0.556	No	2	0.200
19	501	Si	11	0.579	No	2	0.182
20	391	Si	12	0.600	No	2	0.167
21	694	No	12	0.571	No	2	0.167
22	520	No	12	0.545	No	2	0.167
23	333	Si	13	0.565	No	2	0.154
24	861	Si	14	0.583	No	2	0.143

Figura 2.2: Planilla de números simulados con frecuencias observadas de ocurrencia de A y B.

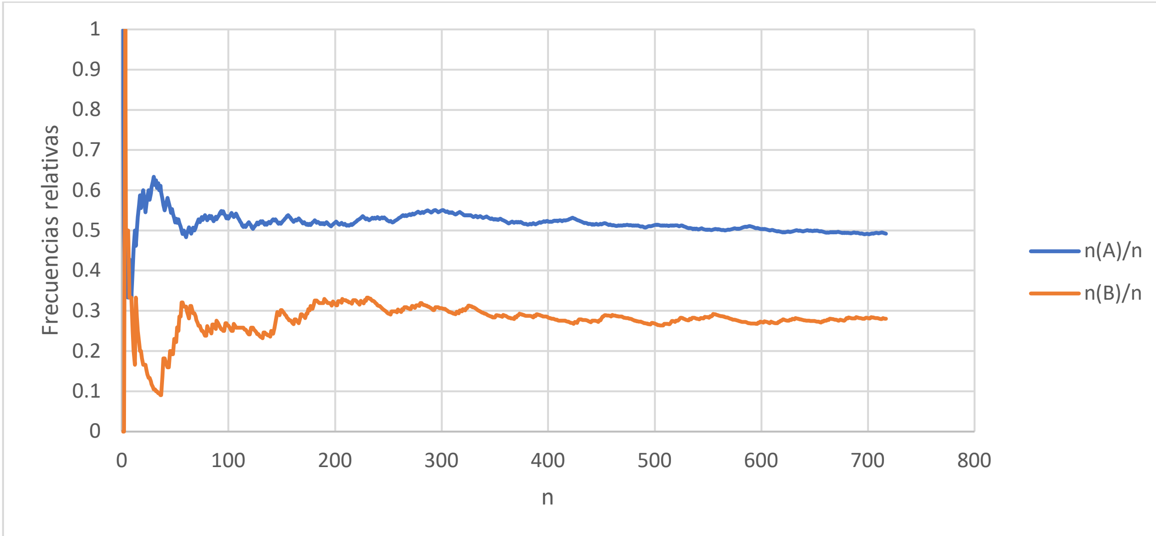


Figura 2.2: Convergencia de las frecuencias relativas de A y B.

Las definiciones antes mencionadas, permiten calcular las probabilidades bajo ciertos contextos, cuando todos los resultados elementales tienen las mismas chances de ocurrir (definición de Laplace), o cuando es posible repetir sucesivamente el experimento hasta alcanzar la convergencia de la frecuencia relativa (definición frecuentista). No necesariamente esto es factible siempre, consecuentemente ambas definiciones no permiten abordar todas las posibles situaciones, de esta manera, se llega a una tercera definición de probabilidad, la definición axiomática, que permite saber cuándo se está en presencia de una probabilidad, puesto que la caracteriza como una función sobre el conjunto de todos los posibles eventos posibles que cumplen ciertos axiomas, estando en presencia de una definición conceptual, mas no operacional, ya que no da la forma en que se pueden calcular las probabilidades.

Definición Axiomática de Probabilidad: Sea $\mathcal{S}$ el espacio muestral de un experimento aleatorio. Una probabilidad P es una función que asigna a cada acontecimiento A de $\mathcal{S}$ un número $P(A)$, llamado la probabilidad de A, que satisface los siguientes axiomas:

(i) $P(A) \geq 0$, para todo acontecimiento A,

(ii) $P(\mathcal{S}) = 1$, y

(iii) para cualquier sucesión de acontecimientos mutuamente excluyentes

$$A_1, A_2, \ldots \quad P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Observaciones:

1. Esta definición axiomática permite identificar si una dada función satisface o no los axiomas necesarios para ser probabilidad, pero no da las herramientas para asignar probabilidades.
2. Esta definición incluye a las probabilidades calculadas usando (i) definición de Laplace o clásica o a priori, (ii) Definición frecuentista o a posteriori y (iii) Definición subjetiva (siempre y cuando sea coherente con los axiomas).

2.3 Propiedades y reglas aditivas

Propiedades: Denotando por $\mathcal{A}$ el conjunto de todos los acontecimientos de $\mathcal{S}$, se tiene que:

1) $P(\varnothing)=0$	4) A y B$\in \mathcal{A},\ A\cap B=\varnothing \Rightarrow P(A\cup B)=P(A)+P(B)$
2) $A\in \mathcal{A},\ \ P(A^c)=1-P(A)$	5) A y $B\in \mathcal{A},\ \ P(B\cap A^c)=P(B)-P(A\cap B)$
3) $A\in \mathcal{A},\ \ P(A)\le 1$	6) A y $B\in \mathcal{A},\ \ P(B\cup A)=P(A)+P(B)-P(A\cap B)$

Dem. P.1

Sea $A_i=\varnothing,$ para todo $i\in\mathbb{N}$, entonces la secuencia $\{A_i\}_{i\in\mathbb{N}}$ es una secuencia de eventos mutuamente excluyentes, y por lo tanto se puede aplicar el axioma (iii) de la definición axiomática de probabilidad y se tiene que $P(\varnothing)=P\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i)$ $\sum_{i=1}^{\infty}P(\varnothing)=\lim_{n\to\infty}n\cdot P(\varnothing)$, consecuentemente $P(\varnothing)=0$.

Dem. P.2

Sean $A_1=A\in\mathcal{A},\ A_2=A^c\in\mathcal{A}$ y $A_i=\varnothing,$ para todo $i\in\mathbb{N}-\{1,2\}$, entonces la secuencia $\{A_i\}_{i\in\mathbb{N}}$ es una secuencia de eventos mutuamente excluyentes, y por lo tanto se puede aplicar el axioma (iii) de la definición axiomática de probabilidad y se tiene que $1=P(\mathcal{S})=P\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i)=P(A_1)+P(A_2)=P(A)+P(A^c)$, consecuentemente $P(A^c)=1-P(A)$.

Dem. P.3

De P.2. se sabe que $P(A^c)=1-P(A)$, por el axioma (1) de probabilidad se tiene que $P(A)\ge 0$, entonces $P(A^c)\le 1$.

Dem. P.4

Sean $A_1=A\in\mathcal{A},\ A_2=B\in\mathcal{A}$ y $A_i=\varnothing,$ para todo $i\in\mathbb{N}-\{1,2\}$, entonces la secuencia $\{A_i\}_{i\in\mathbb{N}}$ es una secuencia de eventos mutuamente excluyentes dado que $A\cap B=\varnothing$ y por lo tanto se puede aplicar el axioma (iii) de la definición axiomática de probabilidad y se tiene que $P(A\cup B)=P\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i)=P(A_1)+P(A_2)=P(A)+P(B)$, consecuentemente $P(A^c)=1-P(A)$.

Dem. P.5

Se puede observar que $B=A\cap B\cup A^c\cap B$, siendo $A\cap B$ y $A^c\cap B$ eventos mutuamente excluyentes, por lo tanto de P.4 se tiene $P(B)=P(A\cap B\cup A^c\cap B)=P(A\cap B)+P(A^c\cap B)$, consecuentemente $P(A^c\cap B)=P(B)-P(A\cap B)$.

Dem. P.6

Se puede observar que $A \cup B = A \cup A^c \cap B$, siendo A y $A^c \cap B$ eventos mutuamente excluyentes, por lo tanto, de P.4 y P.5 se tiene que $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Observaciones:

1. La propiedad P.6. puede generalizarse para cualquier número de unandos. Específicamente para tres eventos sería:
   $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$.
2. La propiedad P.5. puede generalizarse para cualquier número de unandos mutuamente excluyentes. Específicamente sería que para los siguientes acontecimientos mutuamente excluyentes $A_1, A_2, ..., A_n$, $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$.

Ejemplo 2.3: (De Groot pp 16) Si el 50% de las familias de cierta ciudad están suscritas al periódico matinal, el 65% de las familias al periódico vespertino y el 85% al menos uno de los dos periódicos, ¿Cuál es la proporción de familias que están suscritas a los dos periódicos?

Sean

A: familia suscrita al periódico matinal, donde $P(A)$=0,50

B: familia suscrita al periódico vespertino, donde $P(B)$=0,65

se sabe que $P(A \cup B) = 0,85$, y se debe calcular la probabilidad de $A \cap B$.

Se observa que $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, por lo tanto

$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &= 0,5 + 0,65 - 0,85 \\ &= 0,3. \end{aligned}$$

Ejemplo 2.4: (De Groot pp 16) Considere dos sucesos A y B con $P(A)$=0,4 y $P(B)$=0,7. Determine los posibles valores máximo y mínimo de $P(A \cap B)$.

Se observa que $A \cap B \subset A$ y $A \cap B \subset B$, por lo tanto $P(A \cap B) \le \min\{P(A), P(B)\} = 0,4$.

Por otro lado, $1 \ge P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1,1 - P(A \cap B)$, por lo tanto $P(A \cap B) \ge 1,1 - 1 = 0,1$. De esta manera el valor máximo posible para $P(A \cap B)$ es 0,4 y el mínimo posible es 0,1.

2.4 Probabilidad Condicional

El manejo de información adicional puede alterar la composición del espacio muestral y consecuentemente modificar el cálculo o asignación de las probabilidades.

Ejemplo 2.5: considere la situación en que 150 alumnos del último año de la secundaria fueron interrogados con relación a la facultad donde pretenden cursar su carrera universitaria y su habilidad en matemática. Los datos se presentan en la Tabla 2.1.

Tabla 2.1: Distribución de los alumnos según la facultad donde cursarán su carrera universitaria y su habilidad en matemática

		Habilidad en Matemática		*Total*
		Si	No	
Facultad	Ciencias Económicas	18	22	*40*
	Otra	42	68	*110*
Total		*60*	*90*	*150*

La probabilidad de tener habilidad para matemática es 60/150=0,40. ¿Cuánto será la probabilidad de tener habilidad en matemática entre los que quieren cursar una carrera en Ciencias Económicas? Si se restringe el espacio a los que quieren cursar una carrera en FCE, de los cuarenta, 18 presentan habilidad en matemática, por lo tanto la probabilidad será 18/40=0,45.

Formalmente, pensando en términos de experimento aleatorio y espacio muestral el ejemplo anterior puede pensarse de la siguiente manera:

Experimento:

Se selecciona aleatoriamente uno de los 150 alumnos y se observa si cursará o no una carrera en FCE y si tiene o no habilidad en matemática.

Espacio Muestral:

$$\mathcal{S}=\left\{(x,y): x\in\left\{FCE, otro\right\}; y\in\left\{S,N\right\}\right\}$$

Probabilidades de los eventos en $\mathcal{S}$:

$$P((FCE,S))=18/150 \quad P((FCE,N))=22/150$$
$$P((otro,S))=42/150 \quad P((otro,N))=68/150$$

Información adicional: Entre los que quieren carrera en FCE
Espacio Muestral actualizado:

$$\mathcal{S}_C=\left\{(FCE,y): y\in\left\{S,N\right\}\right\}$$

Probabilidades actualizadas de los eventos en $\mathcal{S}_C$:

$$P((FCE,S))=18/40 \quad P((FCE,N))=22/40$$

La información adicional, obliga a una actualización de las probabilidades. Las probabilidades actualizadas a la luz de esa información adicional es lo que se denominan probabilidades condicionales.

Probabilidad Condicional: Sean A y $B\in\mathcal{A}$, con $P(B)\neq 0$, la probabilidad condicional de A dado B, que se indica por $P(A|B)$, se define como

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

En el contexto del ejemplo 2.5, si *A* es el poseer habilidad para matemática y *B* es el querer cursar una carrera en la FCE, se tiene que la probabilidad de tener habilidad para matemática entre los que quieren cursar una carrera en la FCE viene dada por

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{18/150}{40/150}=18/40.$$

Ejemplo 2.6: Se dispone de información de 280 productores de limones en Tucumán, según tipo de productor y canales de comercialización por ellos utilizados (Ver Tabla 2.2). Al seleccionar aleatoriamente un productor se quiere saber la probabilidad de que use el empaque si se sabe que se eligió un productor grande.

Tabla 2.2: Canales de comercialización utilizados según tipo de productores

Tipo de Productor	Canales				*Total*
	Empaque	**Industria**	**Exportación propia**	**Mercado Interno**	
Chico (hasta 10.000 tns)	61	45	2	4	*112*
Grande (más de 10.000 tns)	48	42	66	12	*168*
Total	*109*	*87*	*68*	*16*	*280*

Sean

A: Usa el empaque como canal de comercialización

B: Comercializa más de 10.000 tns

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{48/280}{168/280} = 48/168 .$$

Se puede mostrar que dado B, con $P(B)>0$, la probabilidad condicional define una probabilidad. Específicamente, si para todo evento A de $\mathcal{S}$, es decir $A \in \mathcal{A}$, conjunto de eventos de $\mathcal{S}$, se define $P^*(A) = P(A|B)$, se puede probar que $P^*(A) = P(A|B)$ cumple con todos los axiomas de la definición axiomática de probabilidad.

(i) Para todo acontecimiento A, $P^*(A) = \dfrac{P(A \cap B)}{P(B)} \geq 0$, puesto que $P(A \cap B) \geq 0$ y $P(B) > 0$.

(ii) $P^*(\mathcal{S}) = \dfrac{P(\mathcal{S} \cap B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1$, y

(iii) para cualquier sucesión de acontecimientos mutuamente excluyentes $A_1, A_2, \ldots$

$$P^*\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{P\left(\left[\bigcup_{i=1}^{\infty} A_i\right] \cap B\right)}{P(B)} = \frac{P\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{P(B)} = \frac{\sum_{i=1}^{\infty} P\left(A_i \cap B\right)}{P(B)} = \sum_{i=1}^{\infty} P^*\left(A_i\right).$$

Con esto queda probado que la probabilidad condicional efectivamente define una nueva función de probabilidad en $\mathcal{A}$.

Asociado al concepto de probabilidad condicional se introduce el concepto de independencia entre dos eventos.

***Eventos independientes*: A y B son independientes sii $P(A \cap B) = P(A)P(B)$.**

Intuitivamente se puede afirmar que dos eventos son independientes si la probabilidad de ocurrencia de uno de ellos no se modifica por la ocurrencia del otro. Específicamente con relación a esta idea se presenta el siguiente teorema.

Teorema: Sean A y $B \in \mathcal{A}$, con $P(B) \neq 0$, entonces A y B son independientes sii $P(A|B) = P(A)$.

Dem: Sean A y $B \in \mathcal{A}$, con $P(B) \neq 0$.

$(\Rightarrow)$ Si $P(B)\neq 0$ y A y B son independientes entonces $P(A\cap B)=P(A)P(B)$, por lo tanto $P(A|B)=\dfrac{P(A\cap B)}{P(B)}=\dfrac{P(A)P(B)}{P(B)}=P(A)$.

$(\Leftarrow)$ Si $P(A|B)=P(A)$, entonces $P(A)=\dfrac{P(A\cap B)}{P(B)}$, por lo tanto $P(A\cap B)=P(A)P(B)$.

2.5 Reglas multiplicativas y Teorema de Bayes

A partir de la definición de probabilidad condicional se puede enunciar una regla multiplicativa de probabilidades que permite calcular la probabilidad de la intersección de dos eventos.

Regla multiplicativa de probabilidades: Sean $A \text{ y } B\in\mathcal{A}$, con $P(B)\neq 0$, entonces $P(A\cap B)=P(A|B)P(B)$.

Ejemplo 2.7: Un pueblo tiene dos carros de bomberos que operan independientemente. La probabilidad de que un vehículo específico esté disponible cuando se necesite es de 0,96.
(a) ¿Cuál es la probabilidad de que ninguno esté disponible en caso que se los necesite?, y
(b) ¿Cuál es la probabilidad de que algún carro esté disponible cuando se lo necesita?

Sea $B(i)$ el evento donde el i-ésimo carro de bombero está disponible, para i=1 y 2. El evento en que ninguno está disponible en caso que se los necesite es $\overline{B}(1)\cap\overline{B}(2)$ y que alguno esté disponible es $B(1)\cup B(2)$. De esta manera,

$$\begin{aligned}
P\left(\overline{B(1)}\cap\overline{B(2)}\right)&=P\left(\overline{B(1)\cup B(2)}\right)=1-P\left(B(1)\cup B(2)\right)\\
&=1-P\left(B(1)\right)-P\left(B(2)\right)+P\left(B(1)\cap B(2)\right)\\
&=1-P\left(B(1)\right)-P\left(B(2)\right)+P\left(B(1)\right)P\left(B(2)\right)\\
&=1-P\left(B(1)\right)+P\left(B(2)\right)\left[1-P\left(B(1)\right)\right]\\
&=\left[1-P\left(B(1)\right)\right]\left[1-P\left(B(2)\right)\right]\\
&=0,04\times 0,04=0,0016,
\end{aligned}$$

y $P\left(B(1)\cup B(2)\right)=1-0,0016=0,9984$.

Observación: Sean $A \text{ y } B\in\mathcal{A}$, si A y B son independientes entonces también son independientes, $A \text{ y } \overline{B}$, $\overline{A} \text{ y } B$ y $\overline{A} \text{ y } \overline{B}$.

Generalización de la regla multiplicativa: Si en un experimento los eventos $A_1, A_2,..., A_k$ pueden ocurrir, entonces

$$P(A_1\cap A_2\cap...\cap A_k)=P(A_1)P(A_2|A_1)P(A_3|A_1\cap A_2)...P(A_k|A_1\cap A_2\cap...\cap A_{k-1}).$$

Si los eventos son independientes, entonces $P(A_1\cap A_2\cap...\cap A_k)=P(A_1)P(A_2)...P(A_k)$.

Ejemplo 2.8: Suponga que se distribuyen pelotas de colores en tres cajas idénticas de la siguiente manera.

	Caja 1	Caja 2	Caja 3
Roja	3	4	3
Blanca	2	1	4
Azul	5	3	3

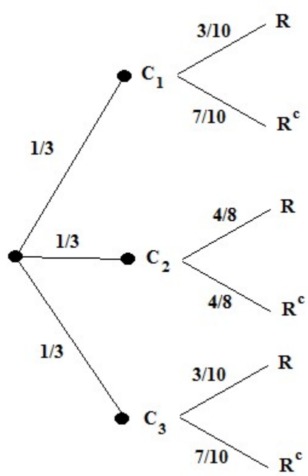
Si se selecciona aleatoriamente una caja, y de ella se selecciona aleatoriamente una pelota. Calcule (a) La probabilidad de que la pelota escogida sea roja, y (b) siendo roja la pelota seleccionada, calcule la probabilidad de que haya sido escogida la caja 3.

Sea

R: "la pelota escogida es roja"

C_i : "se escoge la caja i", i=1, 2, 3.

Para calcular probabilidades en el contexto de este ejemplo, es conveniente recurrir a la utilización de un diagrama de árbol.



Las probabilidades de los diferentes eventos se pueden calcular usando la regla multiplicativa de probabilidades, de la siguiente forma:

Evento	Probabilidad
$R \cap C_1$	$1/3 \times 3/10 = 1/10$
$R^c \cap C_1$	$1/3 \times 7/10 = 7/30$
$R \cap C_2$	$1/3 \times 4/8 = 4/24$
$R^c \cap C_2$	$1/3 \times 4/8 = 4/24$
$R \cap C_3$	$1/3 \times 3/10 = 1/10$
$R^c \cap C_3$	$1/3 \times 7/10 = 7/30$

(a)

$$\begin{aligned} P(R) &= P(R \cap C_1) + P(R \cap C_2) + P(R \cap C_3) \\ &= P(R|C_1)P(C_1) + P(R|C_2)P(C_2) + P(R|C_3)P(C_3) \\ &= \frac{3}{10}\frac{1}{3} + \frac{4}{8}\frac{1}{3} + \frac{3}{10}\frac{1}{3} = \frac{11}{30}. \end{aligned}$$

(b)

$$P(C_3|R) = \frac{P(R \cap C_3)}{P(R)} = \frac{P(R|C_3)P(C_3)}{P(R)} = \frac{\frac{3}{10}\frac{1}{3}}{\frac{11}{30}} = 3/11.$$

El razonamiento empleado para resolver este ejemplo, se aplica en muchas situaciones en que existe una partición del espacio muestral.

Partición del Espacio Muestral: Los eventos $C_1, C_2, ..., C_k$ constituyen una partición del espacio muestral $\mathcal{S}$ sii (i) $\forall i, j \; i \neq j, \; C_i \cap C_j = \varnothing$ y (ii) $\bigcup_{i=1}^{k} C_i = \mathcal{S}$.

Teorema de la Probabilidad Total: Si $C_1, C_2, ..., C_k$ constituye una partición del espacio muestral $\mathcal{S}$, con $P(C_i) \neq 0$ para todo $i = 1, ..., k$, entonces para cualquier evento $A \in \mathcal{A}$

$$P(A) = \sum_{i=1}^{k} P(C_i) P(A|C_i) .$$

Dem: Si $C_1, C_2, ..., C_k$ constituye una partición del espacio muestral $\mathcal{S}$, entonces $\bigcup_{i=1}^{k} C_i = \mathcal{S}$ y $A = A \cap \mathcal{S} = A \cap \bigcup_{i=1}^{k} C_i = \bigcup_{i=1}^{k} (A \cap C_i)$, siendo $\{A \cap C_i\}_{i=1,...,k}$ eventos mutuamente excluyentes. Consecuentemente, $P(A) = \sum_{i=1}^{k} P(A \cap C_i) = \sum_{i=1}^{k} P(A|C_i) P(C_i)$.

Observación: La selección de las cajas 1, 2 ó 3 constituía una partición del espacio muestral, en el apartado (b) se interrogaba sobre la ocurrencia de un evento de la partición cuando se sabe la ocurrencia de otro evento. Situaciones como éstas se resuelven aplicando directamente el Teorema de Bayes.

Teorema de Bayes: Si $C_1, C_2, ..., C_k$ constituye una partición del espacio muestral $\mathcal{S}$, con $P(C_i) \neq 0$ para todo $i = 1, ..., k$, entonces para cualquier evento $A \in \mathcal{A}$, con $P(A) \neq 0$

$$P(C_r|A) = \frac{P(C_r) P(A|C_r)}{\sum_{i=1}^{k} P(C_i) P(A|C_i)}, \quad para \;\; r = 1, ..., k.$$

Dem: Por definición de probabilidad condicional,

$$P(C_r|A) = \frac{P(C_r \cap A)}{P(A)} = \frac{P(C_r) P(A|C_r)}{\sum_{i=1}^{k} P(C_i) P(A|C_i)}.$$

Ejemplo 2.9: Tres fábricas proveen de equipos de precisión para el laboratorio de química de una universidad, que con una pequeña probabilidad pueden subestimar o sobreestimar las medidas efectuadas, como se muestra en la siguiente tabla.

Tabla N°2.3: Probabilidades de subestimar o sobreestimar las medidas efectuadas

Origen del equipo	Subestima	Exacto	Sobreestima
Fábrica 1	0,01	0,98	0,01
Fábrica 2	0,005	0,98	0,015
Fábrica 3	0	0,99	0,01

Las fábricas 1, 2 y 3 proveen, respectivamente, el 20%, 30% y 50% de los equipos utilizados. Si se escoge un equipo al azar, calcule la probabilidad de que:

a) Se sobreestime la medida.
b) No se subestime la medida.
c) Haya sido fabricado por la fábrica 3 si la medición exacta.
d) Haya sido producido por la fábrica 1 si no subestima la medida.

Sean los eventos

F_i: "El equipo es producido por la fábrica i"
B: "Subestima la medida"
E: "da la medida exacta"
M: "Sobrestima la medida"

a)

$$\begin{aligned} P(M) &= \sum_{i=1}^{3} P(F_i)P(M|F_i) \\ &= 0,2\cdot 0,01 + 0,3\cdot 0,015 + 0,5\cdot 0,01 \\ &= 0,002 + 0,0045 + 0,005 = 0,0115 \end{aligned}$$

b)

$$\begin{aligned} P(B^c) &= 1 - P(B) = 1 - \sum_{i=1}^{3} P(F_i)P(B|F_i) \\ &= 1 - (0,2\cdot 0,01 + 0,3\cdot 0,005 + 0,5\cdot 0) \\ &= 1 - (0,002 + 0,0015) = 1 - 0,005 = 0,9965 \end{aligned}$$

c)

$$P(F_3|E) = \frac{P(F_3)P(E|F_3)}{\sum_{i=1}^{3} P(F_i)P(E|F_i)} = \frac{0,5\cdot 0,99}{0,2\cdot 0,98 + 0,3\cdot 0,98 + 0,5\cdot 0,99}$$

$$= \frac{0,495}{0,98\cdot 0,05 + 0,495} = \frac{0,495}{0,985} = 0,503$$

d)

$$P(F_1|B^c) = \frac{P(F_3)P(B^c|F_3)}{\sum_{i=1}^{3} P(F_i)P(B^c|F_i)} = \frac{0,2\cdot 0,99}{0,9965} = \frac{0,198}{0,9965} = 0,199$$

3. Variables Aleatorias

3.1. Introducción

Considere el siguiente ejemplo.

Ejemplo 3.1: El directorio de una empresa está compuesto por 5 personas, 3 de Córdoba y 2 de Tucumán. Si se seleccionan aleatoriamente 3 personas del directorio, calcule la probabilidad de: (a) no seleccionar ninguno de Tucumán, (b) seleccionar solo uno de Tucumán, y (c) seleccionar a los dos de Tucumán.

Experimento: "se selecciona aleatoriamente 3 personas del directorio y se registra son o no de Tucumán".

Espacio muestral $\mathcal{S} = \{CCC, CCT, CTC, CTT, TCC, TCT, TTC\}$

Tabla 3.1: Asignación de probabilidades para los elementos del espacio muestral y cálculo de la variable "Número de tucumanos en la comisión"

s	$P(\{s\})$	Número de tucumanos en la comisión
CCC	1/10	0
CCT	1/5	1
CTC	1/5	1
CTT	1/10	2
TCC	1/5	1
TCT	1/10	2
TTC	1/10	2

Se observa que el número de tucumanos en la comisión define una función de $\mathcal{S}$ en los números reales.

$$\begin{array}{rccc} X: & \mathcal{S} & \to & \mathbb{R} \\ & s & \mapsto & X(s) \end{array}$$

Las probabilidades de los valores asumidos por la función *x* serían de acuerdo a lo que se presenta en la siguiente tabla

Tabla 3.2: Asignación de probabilidades para los valores asumidos por la variable "Número de tucumanos en la comisión"

Nro de personas de Tucumán x	$P(X{=}x)$
0	1/10
1	3/5
2	3/10

Por lo tanto, la probabilidad en el espacio muestral $\mathcal{S}$ induce probabilidades en $\text{Rec}(X) \subset \mathbb{R}$..

3.2 Variables aleatorias

Variable Aleatoria: Sea un espacio muestral $\mathcal{S}$ de un experimento aleatorio. Una variable aleatoria X es una función cuyo dominio es $\mathcal{S}$ y cuyo recorrido es un conjunto de números reales, es decir:

$$\begin{aligned} X : \mathcal{S} &\rightarrow \mathbb{R} \\ s &\mapsto X(s). \end{aligned}$$

Observaciones:

1. Las probabilidades estaban definidas en subconjuntos de $\mathcal{S}$. Una variable aleatoria induce ahora probabilidades en $\mathbb{R}$, de manera tal que $P(A) = P(\{s \in \mathcal{S} : X(s) \in A\})$, donde A es un subconjunto de $\mathbb{R}$.
2. Las características del recorrido de X dependerán de la definición de X y de las características del espacio muestral $\mathcal{S}$.
   a. Se dice que un espacio muestral es discreto si él contiene un número finito o infinito numerable de elementos.
   b. Se dice que un espacio muestral es continuo cuando contiene un número infinito de posibles resultados, igual al número de puntos en un segmento de línea.

Función de Distribución Acumulada (fda): Sea X una variable aleatoria. La función de distribución acumulada F_X es la función definida para todo número real x por $F_X(x) = P(X \leq x)$, es decir

$$\begin{aligned} F_X : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto P(\{s \in \mathcal{S} : X(s) \leq x\}). \end{aligned}$$

Ejemplo 3.1 (cont): Continuando con el ejemplo del directorio, el cálculo de la función de distribución acumulada se realiza de la siguiente manera.

Si $x < 0$ $F_X(x) = P(X \leq x) = P(\{s \in \mathcal{S} : X(s) \leq x\}) = 0$

$0 \leq x \leq 1$ $F_X(x) = P(X \leq x) = P(\{s \in \mathcal{S} : X(s) \leq x\}) = 1/10$

$1 \leq x < 2$ $F_X(x) = P(X \leq x) = P(\{s \in \mathcal{S} : X(s) \leq x\}) = 1/10 + 3/5 = 7/10$

$x \geq 2$ $F_X(x) = P(X \leq x) = P(\{s \in \mathcal{S} : X(s) \leq x\}) = 1$

De esta manera

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/10 & 0 \leq x \leq 1 \\ 7/10 & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}.$$

Notación: Las variables aleatorias se denotan con letras de molde mayúsculas (X), las letras de molde minúsculas (x) denotan valores de su recorrido.

Propiedades de la función de distribución acumulada:

(1) F_X es una función no decreciente

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

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

(4) $P(a < X \le b) = F_X(b) - F_X(a)$, cuando $b \ge a$

Para visualizar que F_X es una función no decreciente, se observa que si $x_1 \le x_2$, entonces $\{s \in \mathcal{S} : X(s) \le x_1\} \subseteq \{s \in \mathcal{S} : X(s) \le x_2\}$, por lo tanto $P\left(\{s \in \mathcal{S} : X(s) \le x_1\}\right) \le P\left(\{s \in \mathcal{S} : X(s) \le x_2\}\right)$ y así $F_X(x_1) \le F_X(x_2)$.

3.2.1 Variables aleatorias discretas

Variable Aleatoria Discreta: Se dice que una variable aleatoria X es discreta cuando el conjunto de valores posibles $\left(\mathrm{Rec}(X)\right)$ es numerable (finito o infinito semejante a los números naturales).

Intuitivamente, se asocian las variables discretas, a aquellas variables que se generan en el proceso de contabilizar la ocurrencia de algunas características. También se usan para modelar situaciones donde los datos surgen de eventos que ocurren en valores puntuales discretos.

Distribución de Probabilidad (para una variable aleatoria discreta): Sea X una variable aleatoria discreta con $\mathrm{Rec}(X) = \left\{x_i : i \in I \subset \mathbb{N}\right\}$, entonces **la función de probabilidad o función de masa** se denota por p_X y es tal que

$$\begin{aligned} p_X &: \mathbb{R} \to \mathbb{R} \\ x &\mapsto p_X(x) = P(X = x) \end{aligned}$$

Las funciones de masa sirven para modelar las probabilidades puntuales.

Observaciones:

- Si x no pertenece al recorrido de X, $p_X(x)$ es cero.
- La probabilidad de cualquier evento se calcula como la suma de las probabilidades individuales de los puntos que componen el evento. Es decir

$$\boldsymbol{P(A) = \sum_{x\in A} p_X(x)}.$$

Propiedades de la función de masa:

(1) $p_X(x) \ge 0$ (2) $\sum_{i\in I} p_X(x_i) = 1$

.

Ejemplos 3.2: (1) Directorio: La función de masa para el número de tucumanos entre los tres miembros del directorio viene dada por:

x	$p_X(x)$
0	1/10
1	3/5
2	3/10

(2) Arrojar un dado hasta obtener un nro 1: El número de veces que se arroja el dado hasta obtener un número 1 es una variable aleatoria, que asume valores naturales (1, 2, …).

La probabilidad de tener que hacer x tiros, viene dada por $p_X(x) = \left(5/6\right)^{x-1} \times 1/6$, donde $\left(5/6\right)^{x-1}$ es la probabilidad de no obtener ningún uno en los primeros x-1 tiros. Consecuentemente la función de masa viene dada por:

$$p_X(x) = \left(5/6\right)^{x-1} \times 1/6, \qquad x \in \mathbb{N}.$$

Se debe notar que $\mathrm{P}(X = x) = \left(\frac{5}{6}\right)^{x-1} \frac{1}{6} \geq 0$ y que

$$\sum_{i=1}^{\infty} \mathrm{P}(X = i) = \sum_{i=1}^{\infty} \left(\frac{5}{6}\right)^{i-1} \frac{1}{6} = \frac{1}{6} \sum_{i=0}^{\infty} \left(\frac{5}{6}\right)^{i} = \frac{1}{6} \lim_{n\to\infty} \sum_{i=0}^{n} \left(\frac{5}{6}\right)^{i} = \frac{1}{6} \lim_{n\to\infty} \frac{1-\left(\frac{5}{6}\right)^{n+1}}{1-\left(\frac{5}{6}\right)} = \frac{1}{6} \times 6 = 1,$$

con lo que se verificarían las condiciones que debe cumplir una función para ser una función de masa.

3.2.2 Variables aleatorias continuas

Ejemplo del Motor de Movimiento Continuo: Considere la siguiente rueda continua. Suponga que es absolutamente simétrica y se la hace girar hasta que se detenga. Se le propone elegir para apostar cualquiera de los siguientes tres eventos:
A: "La rueda se detiene en el número 1/6"
B: "La rueda se detiene entre los números 1/12 y 4/12"
C: "La rueda se detiene entre los números 4/12 y 8/12"
¿A cuál evento usted querría apostar y por qué?

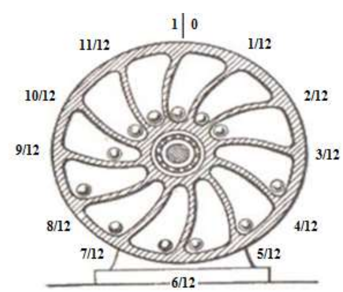


Figura 4.1: Esquema de un motor de movimiento continuo

Cuando se hace girar la rueda, ella se puede detener en cualquier número entre 0 y 1. Si se tiene que elegir apostar por cualquier evento, convendría elegir aquel que sea más amplio, pensando que la probabilidad de ganar es proporcional a la longitud del intervalo. Contextualizando el problema en términos de los conceptos de experimento aleatorio y espacio muestral, se tiene que:

Experimento:

Se hace girar la rueda y se registra el número donde se detiene.

Espacio Muestral:

$$\mathcal{S} = \{x : x \in \mathbb{R}, 0 \le x < 1\}$$

Probabilidades de los eventos en $\mathcal{S}$:

$$P(\mathcal{S}) = 1,\ P(\varnothing) = 0 \text{ y para todo } 0 \le a \le b < 1 : P((a,b)) = k(b-a).$$

En virtud de que $P((0,1)) = 1 \Rightarrow k = 1$, y de esta manera se tiene que

$$P(A) = 0,\ P(B) = 4/12 - 1/12 = 1/4 \text{ y } P(C) = 8/12 - 4/12 = 1/3.$$

En general, usando resultados del análisis se tiene que $\forall A \subset [0,1) : P(A) = \int_A dx$, siempre y cuando esa integral exista.

La variable aleatoria definida por

X: Número obtenido al hacer girar la rueda

Es una variable aleatoria continua, pues asume cualquier valor entre [0,1), conjunto no numerable. Formalmente se define la variable aleatoria continua como sigue.

Variable Aleatoria Continua: Se dice que una variable aleatoria X es continua cuando el conjunto de valores posibles $\left(\mathrm{Rec}(X)\right)$ es no-numerable (semejantes a un intervalo en $\mathbb{R}$).

Asociada a una variable aleatoria continua, generalmente, se encuentra una función denominada función de densidad que permite el cálculo de probabilidades para eventos en $\mathbb{R}$.

Distribución de Probabilidad (para una variable aleatoria continua): Sea X una variable aleatoria continua. Si existe f_X tal que:

(1) $f_X(x) \ge 0$, para todo *x* real

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

(3) $P(a < X < b) = \int_a^b f_X(x)dx$,

entonces se dice que f_X es la **función de probabilidad o función de densidad** de X .

La integral de una función no negativa representa el área bajo la curva definida por esa función. Dado que la función de densidad de probabilidad es no negativa y su integral total es igual a uno, el área bajo su curva corresponde a la probabilidad.

En el caso de variables aleatorias continuas, la función de densidad no asigna probabilidades a valores puntuales, sino que describe cómo la probabilidad se distribuye a lo largo del conjunto continuo de valores posibles. Esto se debe a que, en un continuo, la probabilidad de que la variable tome un valor exacto es cero. En su lugar, la función de densidad$f(x)$ proporciona una medida relativa de la concentración de probabilidad en torno a un valor dado, y la probabilidad de que la variable caiga dentro de un intervalo se obtiene integrando dicha función sobre ese rango.

Ejemplo 3.3: La cantidad de gasto anualmente, en millones de pesos, para el mantenimiento del asfalto en una ciudad es representada por una variable aleatoria con densidad dada por:

$$f_Y(y) = \begin{cases} \frac{8}{9}y - \frac{4}{9} & si \;\; 0,5 \le y < 2 \\ 0 & caso \;\; contrario \end{cases}$$

a) Verifique que *f* es una función de densidad
b) Calcule la probabilidad de que el gasto en mantenimiento del asfalto sea inferior a \$800.000.
c) Grafique la función de distribución acumulada.

Resolución:

a) Para verificar que *f* es una función de densidad, se debe probar que:

$f_Y(y) \ge 0$, para todo *y* real y que $\int\limits_{-\infty}^{\infty} f_Y(y)dy = 1$.

Se observa que cuando y<0,5 ó $y \ge 2$, la función *f* asume el valor 0, y que

$$\begin{aligned} y \in [0,5,\; 2) &\Leftrightarrow 0,5 \le y < 2 \\ &\Leftrightarrow \frac{1}{2}\times\frac{8}{9} \le \frac{8}{9}y < 2\times\frac{8}{9} \\ &\Leftrightarrow \frac{4}{9}-\frac{4}{9} \le \frac{8}{9}y - \frac{4}{9} < 2\times\frac{8}{9}-\frac{4}{9} \\ &\Leftrightarrow 0 \le f(y) < \frac{12}{9} = \frac{4}{3}. \end{aligned}$$

Con lo cual queda demostrado que $f_Y(y) \ge 0$.

Por otro lado,

$$\begin{aligned} \int_{-\infty}^{\infty} f_Y(y)\,dy &= \int_{-\infty}^{0,5} f_Y(y)\,dy + \int_{0,5}^{2} f_Y(y)\,dy + \int_{2}^{\infty} f_Y(y)\,dy \\ &= \int\limits_{-\infty}^{0,5} 0\,dy + \int\limits_{0,5}^{2}\left(\frac{8}{9}y - \frac{4}{9}\right)dy + \int\limits_{0,5}^{2} 0\,dy \\ &= \int_{0,5}^{2} \tfrac{8}{9}y - \tfrac{4}{9}\,dy = \left(\tfrac{4}{9}y^2 - \tfrac{4}{9}y\right)\Big|_{x=0,5}^{x=2} \\ &= \left(\tfrac{4}{9}2^2 - \tfrac{4}{9}2\right) - \left(\tfrac{4}{9}0,5^2 - \tfrac{4}{9}0,5\right) \\ &= \frac{16}{9} - \frac{8}{9} - \frac{1}{9} + \frac{2}{9} = 1 \end{aligned}$$

Por lo que se trata de una función de densidad de probabilidades.

b) Para calcular la probabilidad de que el gasto en mantenimiento del asfalto sea inferior a \$800.000, se debe integrar *f* entre 0,5 y 0,8. El límite superior es 0,8 porque en el enunciado se menciona que la cantidad de gasto anual está medida en millones de pesos.

$$\int_{-\infty}^{0,8} f_Y(y)\,dy = \int_{0,5}^{0,8} f_Y(y)\,dy$$
$$= \int_{0,5}^{0,8} \left(\tfrac{8}{9}y - \tfrac{4}{9}\right)dy = \left(\tfrac{4}{9}y^2 - \tfrac{4}{9}y\right)\Big|_{x=0,5}^{x=0,8}$$
$$= \left(\tfrac{4}{9}\left(\tfrac{4}{5}\right)^2 - \tfrac{4}{9}\tfrac{4}{5}\right) - \left(\tfrac{4}{9}0,5^2 - \tfrac{4}{9}0,5\right)$$
$$= \frac{64}{225} - \frac{16}{45} - \frac{1}{9} + \frac{2}{9} = \frac{64-80-25+50}{225} = \frac{9}{225} = \frac{1}{25}$$

c) Para calcular la función de distribución acumulada en el punto y se calcula $\int_{-\infty}^{y} f_Y(a)\,da$.

Si $y \le 0,5$, $\int_{-\infty}^{y} f_Y(a)\,da = \int_{-\infty}^{y} 0\,da = 0$

Si $0,5 < y \le 2$,

$$\int_{-\infty}^{y} f_Y(a)\,da = \int_{0,5}^{y} \tfrac{8}{9}a - \tfrac{4}{9}\,da = \left(\tfrac{4}{9}a^2 - \tfrac{4}{9}a\right)\Big|_{a=0,5}^{a=y}$$
$$= \left(\tfrac{4}{9}y^2 - \tfrac{4}{9}y\right) - \left(\frac{1}{9} - \frac{2}{9}\right)$$
$$= \tfrac{4}{9}y^2 - \tfrac{4}{9}y + \tfrac{1}{9}$$

Si $y > 2$, $\int_{-\infty}^{y} f_Y(a)\,da = \int_{-\infty}^{2} f_Y(a)\,da + \int_{-\infty}^{y} 0\,da = 1 + 0 = 1$

De esta manera,

$$F_Y(y) = \begin{cases} 0 & \text{si } \; y < 0,5 \\ \dfrac{4}{9}y^2 - \dfrac{4}{9}y + \dfrac{1}{9} & \text{si } \; 0,5 \le y < 2 \\ 1 & \text{si } \; y \ge 2 \end{cases}$$

Y graficando la función de densidad y la distribución acumulada serían

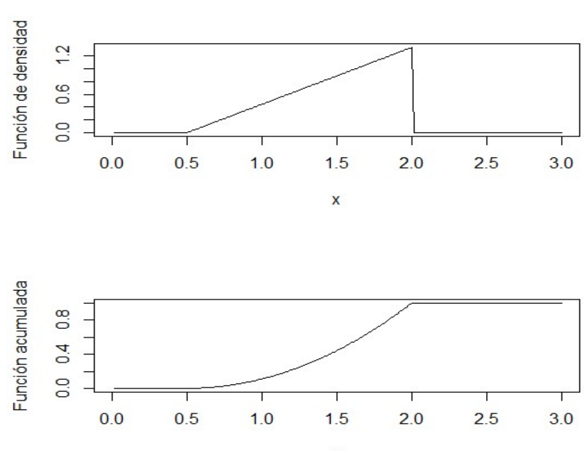


Figura 3.2: Representaciones gráficas de las funciones de densidad y de la distribución acumulada

Observaciones: Cuando se trabaja con variables aleatorias continuas**:**

1. La función de distribución acumulada es una función continua.
2. La probabilidad de un evento en un subconjunto A de los números reales se calcula integrando la función de densidad sobre ese conjunto A, $P(A) = \int_A f(x)dx.$
3. Las probabilidades en un punto son siempre nulas, por lo que en esta situación $F_X(x) = P(X < x)$**.**
4. La función de distribución acumulada es la integral de la función de densidad, con lo que la densidad se puede obtener de la acumulada derivándola, $f_X(x) = \dfrac{dF_X(x)}{dx}$.

Para cerrar la discusión sobre funciones de masa y de densidad en la Tabla 3.3 se ponen en evidencia algunas similitudes y diferencias.

Tabla 3.3: Similitudes y diferencias entre la función de masa y la función de densidad.

Características	**Función de masa**	**Función de densidad**
Definición	Los valores que asume se corresponden con las probabilidades puntuales en cada valor x	Los valores que asumen no son una probabilidad sino una densidad que debe integrarse para producir una probabilidad.
P(X=x)	Puede ser >0 en algunos valores posibles de la v.a.	Siempre es igual a =0
Suma vs. Integral	$\sum_{i\in I} p_X(x_i) = 1$	$\int_{-\infty}^{\infty} f_X(x)dx = 1$
Cálculo de probabilidades	$\boldsymbol{P(A) = \sum_{x\in A} p_X(x)}$.	$P(A) = \int_A f(x)dx.$

3.3 Distribuciones de Probabilidad Conjunta

Hasta ahora se ha trabajado con espacios muestrales donde se definía una variable aleatoria cuyo comportamiento probabilístico resultaba de interés describir. Existen situaciones donde el interés se centra en más de una variable aleatoria simultáneamente. Por ejemplo:

1. Se lanzan dos dados y se observa:

   X: Número de dados donde aparece el 1.

   Y: Número de dados donde aparece el 5 ó el 6.

2. En el cajero automático de la FCE, a las 9:00hs del día posterior al depósito de los sueldos en la UNT, se observa:

   X: Número de personas esperando en la cola para ingresar al cajero.

   Y: Tiempo transcurrido hasta la llegada de la próxima persona a la fila.

3. Se selecciona una cuenta sueldo del Banco M, y se observa

   X: Monto disponible.

   Y: Tiempo transcurrido desde la última extracción realizada.

4. En hogares con tres hijos, y se observa

   X: Número de hijos varones.

   Y: primero hijo varón si (Y=1) o no (Y=0).

   Z: Número de veces en que hubo variación de sexo entre dos nacimientos consecutivos.

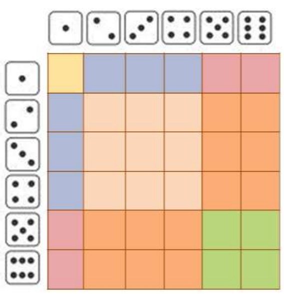


Figura 3.3: Esquema de resultados posibles y valores asumidos por las dos variables aleatorias definidas

En el caso del ejemplo 1, usando la Figura 4.3 que muestra el resultado de los dos dados, se puede saber los valores posibles que asumen X e Y y con qué probabilidad. Se observa que

$$\text{Rec}(X)=\{0,1,2\} \text{ y } \text{Rec}(Y)=\{0,1,2\},$$

de esta manera los pares (x,y), con $x \in \text{Rec}(X)$ e $y \in \text{Rec}(Y)$ ocurren con la probabilidad que se muestra en la siguiente tabla.

x \ y	0	1	2
0	9/36	12/36	4/36
1	6/36	4/36	0
2	1/36	0	0

Figura 3.4: Distribución de probabilidad conjunta

Lo que se construyó a partir de las probabilidades de los eventos elementales que se obtienen al arrojar los dos dados se denomina distribución de probabilidad conjunta, cuya definición formal se escribe a continuación.

Distribución de Probabilidad Conjunta (para variables aleatorias discretas): Sean X e Y variables aleatorias discretas, la función $f(x, y)$ es una **función de distribución de probabilidad conjunta o función de masa de probabilidad** de las variables aleatorias X e Y, si:

(1) $f(x,y) \geq 0$ para todo (x,y),

(2) $\sum_x \sum_y f(x,y) = 1$,

(3) $P(X = x, Y = y) = f(x,y)$.

Observación: Para cualquier región A del plano, se tiene que

$$P((X,Y) \in A) = \sum\sum_{(x,y) \in A} f(x,y).$$

En el caso del ejemplo 4 se puede ver que las secuencias de hijos con cada sexo posibles son las 8 que figuran en la siguiente Tabla 3.4, y puede suponerse que cada una de esas secuencias ocurre con la misma probabilidad (1/8). A partir de esas secuencias es posible estudiar el valor que asume cada una de las variables X, Y y Z.

Tabla 3.4: Secuencias posibles de nacimientos en familias con tres hijos, y cálculo de las variables aleatorias definidas en el ejemplo 4

Eventos	Probabilidades	X	Y	Z
HHH	1/8	3	1	0
HHM	1/8	2	1	1
HMH	1/8	2	1	2
HMM	1/8	1	1	1
MHH	1/8	2	0	1
MHM	1/8	1	0	2
MMH	1/8	1	0	1
MMM	1/8	0	0	0

Las funciones de probabilidad conjunta para (X, Z), (Y, X) e (Y, Z) se expresan a continuación.

		x			
	$f(y,x)$	0	1	2	3
y	0	1/8	2/8	1/8	0
	1	0	1/8	2/8	1/8

(a)

		z		
	$f(x,z)$	0	1	2
	0	1/8	0	0
x	1	0	2/8	1/8
	2	0	2/8	1/8
	3	1/8	0	0

(b)

		z		
	$f(y,z)$	0	1	2
y	0	1/8	2/8	1/8
	1	1/8	2/8	1/8

(c)

Figura 3.5: Tablas de funciones de masas conjuntas de las variables bidimensionales (a)(Y,X), (b) (X,Z), y (c) (Y,Z)

Distribución de Probabilidad Conjunta (para variables aleatorias continuas): Sean X e Y variables aleatorias continuas, la función $f(x,y)$ es una **función de distribución de probabilidad conjunta o función de densidad conjunta** de las variables aleatorias X e Y, si:

(1) $f(x,y) \geq 0$ para todo (x,y),

(2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$,

(3) $P((X,Y) \in A) = \iint_A f(x,y)\,dx\,dy$ para cualquier región A del plano.

Para ilustrar el manejo y cálculo de probabilidades en variables aleatorias bivariadas conjuntas se considera el siguiente ejemplo.

Ejemplo 3.4: Una instalación de servicio telefónico opera con dos líneas de servicio. En un día seleccionado aleatoriamente, sea X la proporción de tiempo que se utiliza la primera línea e Y la proporción de tiempo que se usa la segunda. Suponga que la función de densidad de probabilidad para (X,Y) es

$$f(x,y)=\begin{cases} k(x^2+y^2) & 0\le x\le 1, 0\le y\le 1,\\ 0 & \text{en cualquier otro caso.}\end{cases}$$

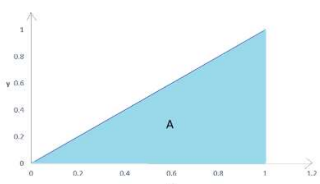
a) Determine el valor de k.
b) Encuentre la probabilidad de que la primera línea esté ocupada más tiempo que la segunda línea.

(a) Para que $f(x,y)$ sea una función de densidad conjunta, debe integrar 1 sobre todo $\mathbb{R}^2$, por lo tanto

$$1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy=\int_0^1\int_0^1 k(x^2+y^2)dxdy$$
$$=\int_0^1 k(\tfrac{1}{3}x^3+y^2x)\Big|_{x=0}^{x=1}dy=\int_0^1 k(\tfrac{1}{3}+y^2)dy=k(\tfrac{1}{3}y+\tfrac{1}{3}y^3)\Big|_{y=0}^{y=1}$$
$$=k(\tfrac{1}{3}+\tfrac{1}{3})=k\,\tfrac{2}{3}$$

Para que se trate de una función de densidad conjunta, k=3/2.

(b) Sea A:" $X>Y$", entonces



$$P(A)=\iint_A f(x,y)dxdy=\int_0^1\int_0^x \tfrac{3}{2}(x^2+y^2)dy\,dx=\int_0^1 \tfrac{3}{2}(x^2y+\tfrac{1}{3}y^3)\Big|_{y=0}^{y=x}dx$$
$$=\int_0^1 \tfrac{3}{2}(x^3+\tfrac{1}{3}x^3)dx=\tfrac{3}{2}\tfrac{4}{3}(\tfrac{1}{4}x^4)\Big|_{x=0}^{x=1}=1/2$$

A veces, aunque la información se brinda de manera conjunta, resulta de interés el hablar sobre el comportamiento probabilístico de una particular variable, en ese caso se puede trabajar con las distribuciones marginales.

Distribuciones Marginales: Sean X e Y variables aleatorias con función de distribución de probabilidad conjunta $f(x,y)$. Las distribuciones marginales de X e Y están dadas respectivamente por:

$$g(x)=\sum_y f(x,y) \quad \text{y} \quad h(y)=\sum_x f(x,y) \text{ para el caso discreto, y}$$

$$g(x)=\int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{y} \quad h(y)=\int_{-\infty}^{\infty} f(x,y)\,dx \text{ para el caso continuo.}$$

En el caso de la instalación del servicio telefónico las distribuciones marginales vienen dadas por

$$g(x) = \begin{cases} \int_0^1 \frac{3}{2}(x^2+y^2)dy & si\ x\in(0,1) \\ 0 & si\ x\notin(0,1) \end{cases} = \begin{cases} \frac{3}{2}(x^2+\frac{1}{3}) & si\ x\in(0,1) \\ 0 & si\ x\notin(0,1) \end{cases}$$

$$h(y) = \begin{cases} \int_0^1 \frac{3}{2}(x^2+y^2)dx & si\ y\in(0,1) \\ 0 & si\ y\notin(0,1) \end{cases} = \begin{cases} \frac{3}{2}(y^2+\frac{1}{3}) & si\ x\in(0,1) \\ 0 & si\ x\notin(0,1) \end{cases}$$

Distribuciones Condicionales: Sean X e Y variables aleatorias con función de distribución de probabilidad conjunta $f(x, y)$. La distribución condicional de Y dado $X = x$ está dada por:

$$f(y|x) = \frac{f(x,y)}{g(x)}, \quad g(x) > 0.$$

Similarmente, la distribución condicional de X dado $Y = y$ está dada por:

$$f(x|y) = \frac{f(x,y)}{h(y)}, \quad h(y) > 0.$$

Donde $g(x)$ y $h(y)$ son las distribuciones marginales de X e Y respectivamente.

Ejemplo 3.4 (cont): Líneas telefónicas.
Encuentre la probabilidad de que la línea 1 se use menos del cuarto del tiempo cuando la línea 2 se usa 1/2 del tiempo**.**
Por definición

$$f(x|y) = \frac{f(x,y)}{h(y)}, \quad h(y) > 0$$

$$= \begin{cases} \dfrac{\frac{3}{2}(x^2+y^2)}{\frac{3}{2}(y^2+\frac{1}{3})} & si\ x\in(0,1) \\ 0 & si\ x\notin(0,1) \end{cases}$$

$$= \begin{cases} \dfrac{(x^2+y^2)}{(y^2+\frac{1}{3})} & si\ x\in(0,1) \\ 0 & si\ x\notin(0,1) \end{cases}$$

En el ejemplo en particular, Y=1/2, entonces

$$f(x|y) = \begin{cases} \dfrac{(x^2+\frac{1}{4})}{(\frac{1}{4}+\frac{1}{3})} & si\ x\in(0,1) \\ 0 & si\ x\notin(0,1) \end{cases}$$

$$= \begin{cases} \dfrac{(x^2+\frac{1}{4})}{\frac{7}{12}} & si\ x\in(0,1) \\ 0 & si\ x\notin(0,1) \end{cases}$$

$$= \begin{cases} \frac{12}{7}x^2+\frac{3}{7} & si\ x\in(0,1) \\ 0 & si\ x\notin(0,1) \end{cases}$$

De esta manera

$$P(X < 1/4 | Y = 1/2) = \int_0^{1/4} \left(\tfrac{12}{7}x^2 + \tfrac{3}{7}\right)dx =$$

$$= \left(\tfrac{4}{7}x^3 + \tfrac{3}{7}x\right)\Big|_{x=0}^{x=1/4}$$

$$= \left(\tfrac{1}{112} + \tfrac{3}{28}\right) = \tfrac{13}{118}$$

Independencia Estadística: Sean X e Y variables aleatorias con función de distribución de probabilidad $f(x, y)$ y distribuciones marginales $g(x)$ y $h(y)$, respectivamente. Las variables X e Y se dicen estadísticamente independientes si y solo si

$$f(x, y) = g(x)h(y) \text{ para todo par } (x, y).$$

En el ejemplo de las líneas telefónicas se puede observar que

$$f(x, y) \neq g(x) \times h(y)$$

$$\tfrac{3}{2}(x^2 + y^2) \neq \tfrac{3}{2}(x^2 + \tfrac{1}{3}) \times \tfrac{3}{2}(y^2 + \tfrac{1}{3})$$

Por lo que se puede concluir que las variables que miden la proporción del tiempo de las dos líneas no son estadísticamente independientes.

En el ejemplo de los hogares con tres hijos se puede observar que $f(y, z) = g(y) \times h(z)$ para cualquier valor de y y z. Por lo que se puede concluir que las variables son estadísticamente independientes.

4. Esperanzas, Varianzas y Covarianzas

4.1. Introducción

Cuando se tira un dado muchas veces, ¿cuál sería el promedio de los valores obtenidos al arrojar sucesivamente el dado?

1	3	2	5	4	6	5	4	2	3	4	1
2	1	3	2	4	5	5	4	3	1	6	6
2	5	3	1	2	4	6	2	5	6	3	1

Denotando con f_1, f_2, f_3, f_4, f_5 y f_6 la frecuencia observada de cada valor i=1, 2,…,6, se calcula el promedio como

$$\text{promedio} = f_1 + 2f_2 + 3f_3 + 4f_4 + 5f_5 + 6f_6$$

Cuando el número de tiradas del dado es suficientemente grande la frecuencias observadas se aproximarán a la probabilidad de obtener cada resultado, que bajo el supuesto de simetría o de que el dado no está alterado, será 1/6, con lo que se tendrá:

$$\text{promedio} \to \frac{1}{6} + 2 \bullet \frac{1}{6} + 3 \bullet \frac{1}{6} + 4 \bullet \frac{1}{6} + 5 \bullet \frac{1}{6} + 6 \bullet \frac{1}{6}.$$

Así, el número promedio esperado a largo plazo sería 3.5.

Por otro lado, se tiran dos monedas sucesivamente, ¿Cuál es el número promedio de caras por lanzamiento? Los resultados posibles en cada lanzamiento son 0, 1 o 2. Denotando con f_0, f_1 y f_2 la frecuencia observada de cada valor i=0, 1 o 2, el número promedio de caras por lanzamiento será $\mathbf{0 \times f_0 + 1 \times f_1 + 2f_2}$.

A largo plazo, esas frecuencias relativas se aproximarán a las probabilidades de obtener 0, 1 y 2 caras respectivamente. Así el número promedio de caras por lanzamiento será a largo plazo $0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2\frac{1}{4} = 1$.

x	$p_X(x)$
0	1/4
1	1/2
2	1/4

En general si X es una v.a. con $\text{Rec}(X) = \{x_i : i \in I \subset \mathbb{N}\}$, con función de probabilidad f_X, el valor promedio de X por cada vez que se la observa será $\sum_{i \in I} x_i \times f_X(x_i)$.

4.2 Esperanza Matemática

Cuando se observan a largo plazo los valores asumidos por una variable aleatoria, por ejemplo al repetir el experimento aleatorio un número muy grande de veces, el valor promedio de lo observado en las repeticiones es una forma de describir el valor central de la distribución de probabilidad de la variable aleatoria analizada. Este valor será un valor asociado a la distribución de la variable aleatoria, que sirve para describir la posición central su distribución. También se suele interpretar con el centro de gravedad de la distribución de probabilidad. Ese valor central, centro de gravedad de la distribución, se denota **esperanza matemática** o **valor esperado** o **media** de la variable aleatoria *X*.

Esperanza Matemática: Sea *X* una variable aleatoria con distribución de probabilidad $f_X(x)$, la media (valor esperado o esperanza matemática) de *X* se denota con μ y es

$$\mu = E(X) = \sum_x x\, f_X(x) \text{, cuando } X \text{ es una variable discreta, ó}$$

$$\mu = E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx \text{, cuando } X \text{ es una variable continua.}$$

Ejemplo 4.1: Una empresa de mudanzas asigna 2, 3 o 4 ayudantes a las mudanzas contratadas dependiendo de la dimensión del trabajo solicitado. Históricamente, se ha necesitado 2 ayudantes en el 10% de las mudanzas, 3 en el 70% y 4 en el 20%. Si *X* es el número de ayudantes en la primera mudanza del próximo mes, se quiere encontrar el número esperado de ayudantes.

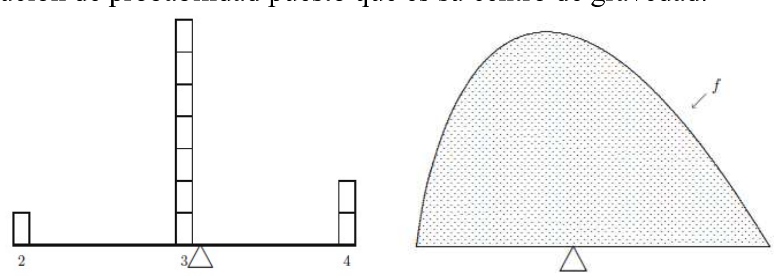
La función de probabilidad de *X* será

x	2	3	4
$p(x)$	0,1	0,7	0,2

El valor esperado de *X* se calcula haciendo $E(X) = 2\times 0{,}1 + 3\times 0{,}7 + 4\times 0{,}2 = 3{,}1$. Lo que significa que a largo plazo, en promedio se utilizan 3,1 ayudantes en cada mudanza

Observaciones:

1. Se dice que la esperanza matemática es una medida de posición central de la distribución de probabilidad puesto que es su centro de gravedad.



2. No necesariamente es un valor posible de la v.a.
3. En caso de simetría entre las probabilidades y el recorrido de la v.a., la media está en el punto medio.
4. **Notación:** Generalmente la esperanza matemática de una variable aleatoria se denota con la letra griega μ y en caso de ser necesario se indica con un subíndice el nombre la v.a. donde se calcula. Específicamente: E(*X*)=μ_X

Existen situaciones donde lo que se quiere calcular es el valor esperado de una función de una variable aleatoria. Como por ejemplo, en el contexto del ejemplo del directorio presentado en el capítulo 3 (Ejemplo 3.1), se podrá querer calcular el gasto medio en pasajes aéreos para los miembros de la comisión cuando la reunión se realiza en córdoba. Si *X* es el número de tucumanos en la comisión y el pasaje ida y vuelta a salta tiene un costo de $300.000, el gasto en pasajes para los miembros tucumanos de la comisión será una variable aleatoria \$300.000*X*. El gasto medio en pasajes para los miembros tucumanos puede calcularse usando el siguiente resultado.

Teorema: Sea X una v.a. con distribución de probabilidad $f_X(x)$. El valor esperado de la variable aleatoria $g(X)$ está dada por

$$\mu_{g(X)} = E\left[g(X)\right] = \sum_x g(x)\, f_X(x) \text{, si } X \text{ es discreta, y}$$

$$\mu_{g(X)} = E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx \text{, si } X \text{ es continua.}$$

En el contexto del ejemplo del directorio, donde la función de masa para el número de tucumanos entre los tres miembros del directorio viene dada por:

x	$p_X(x)$
0	1/10
1	3/5
2	3/10

se tiene que el gasto medio en pasajes aéreos será $\boldsymbol{E(G) = E(300.000X) = 0 \times \frac{1}{10} + 300.000 \times \frac{3}{5} + 600.000 \times \frac{3}{10} = 360.000}$.

Ejemplo 4.2: Sea X el volumen en miles de toneladas de limones comprados por una determinada empresa exportadora. Suponga que la función de densidad de la v.a. X es

$$f(x) = \begin{cases} \dfrac{x}{12}, & \text{si } 1 \le x \le 5 \\ 0, & \text{en cualquier otro punto} \end{cases}$$

Por cada tonelada exportada se obtiene una utilidad de 300 dólares, mientras que cada tonelada no exportada durante un año determinado tiene una utilidad de 100 dólares, porque tendrá que venderse en el mercado local. Si en un año tiene la posibilidad de exportar hasta 3.000 toneladas, ¿cuál es la utilidad esperada?

En este caso la variable aleatoria U (utilidad) es función del volumen de limones comprados, X, a través de la siguiente expresión

$$U(X) = \begin{cases} 300 \times X \times 10^3, & \text{si } 1 \le X \le 3 \\ 300 \times 3 \times 10^3 + 100 \times (X-3) \times 10^3, & \text{si } 3 < X \le 5 \end{cases}$$

$$= \begin{cases} 3 \times 10^5 \times X, & \text{si } 1 \le X \le 3 \\ 10^5 \times X + 6 \times 10^5, & \text{si } 3 < X \le 5 \end{cases}$$

Por lo tanto,

$$E(U) = E(U(X)) = \int_1^3 3 \cdot 10^5 \cdot x \cdot \frac{x}{12} dx + \int_3^5 (10^5 \cdot x + 6 \cdot 10^5) \cdot \frac{x}{12} dx$$

$$= \frac{10^5}{12} x^3 \Big|_1^3 + \frac{10^5}{12} \frac{x^3}{3} \Big|_3^5 + \frac{10^5}{2} \frac{x^3}{2} \Big|_3^5 = \frac{80 \cdot 10^5}{9} = 888.888{,}9$$

La utilidad esperada será de USD\$888.889.

Así como el resultado anterior permite el cálculo de la esperanza de una función de una variable aleatoria sin necesidad de tener que calcular la función de probabilidad de la nueva variable aleatoria, se puede actuar de la misma manera cuando la nueva variable aleatoria es función de un vector de variables aleatorias bidimensional, como se expresa en el siguiente teorema.

Teorema. Sean X e Y vs.as. con distribución de probabilidad conjunta $f(x,y)$. La media de la v.a. $g(X,Y)$ es

$$\mu_{g(X,Y)} = E\left[g(X,Y)\right] = \sum_x \sum_y g(x,y)\, f(x,y)$$, si X e Y son discretas, y

$$\mu_{g(X,Y)} = E\left[g(X,Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f(x,y)\, dx\, dy$$, si X e Y son continuas.

Ejemplo 4.3: El número de personas que trabajan en la familia (*T*) y el número de adolescentes (*X*) son variables aleatorias que tienen la siguiente función de probabilidad conjunta.

		Nro de adolescentes (*x*)					
		0	**1**	**2**	**3**	**4**	**Totales**
Nro de personas que trabajan (*t*)	**0**	5/80	4/80	2/80	3/80	1/80	**15/80**
	1	2/80	8/80	6/80	4/80	1/80	**21/80**
	2	4/80	8/80	8/80	5/80	2/80	**27/80**
	3	4/80	2/80	2/80	5/80	4/80	**17/80**
	Totales	**15/80**	**22/80**	**18/80**	**17/80**	**8/80**	**80/80**

Suponiendo que se define la razón *R*=(1+*X*)/(1+*T*), como indicador del número de adolescentes bajo la responsabilidad de cada trabajador. Encuentre la esperanza matemática de esta razón.

Se pueden calcular los valores observados de la variable aleatoria R como función de los valores asumidos por *X* y *T* en la siguiente tabla.

$\frac{1+x}{1+t}$		**Nro de adolescentes (*x*)**				
		0	**1**	**2**	**3**	**4**
Nro de personas que trabajan (*t*)	**0**	1	2	3	4	5
	1	1/2	2/2	3/2	4/2	5/2
	2	1/3	2/3	3/3	4/3	5/3
	3	1/4	2/4	3/4	4/4	5/4

Consecuentemente

$$\begin{aligned}
\mu_{R(T,X)} &= E\left[g(T,X)\right] = \sum_t \sum_x g(t,x)\, f(t,x) \\
&= \sum_{t=0}^{3} \sum_{x=0}^{4} \frac{1+x}{1+t} f(t,x) \\
&= 1\times\tfrac{5}{80} + 2\times\tfrac{4}{80} + 3\times\tfrac{2}{80} + 4\times\tfrac{3}{80} + 5\times\tfrac{1}{80} + \ldots + \tfrac{4}{4}\times\tfrac{5}{80} + \tfrac{5}{4}\times\tfrac{4}{80} \\
&= 1{,}28\widehat{3}.
\end{aligned}$$

4.3 Varianza

Cuando se informa el valor medio de una variable aleatoria, no se puede inferir si el conjunto de valores posibles que ella puede asumir se encuentra bien próximo del valor medio, o no. Para tener una idea de cuán próximos o dispersos se encuentran los valores posibles de la v.a. con relación a la esperanza matemática se define el concepto de varianza de una v.a. como una medida de variabilidad de la v.a., de distancia cuadrática media con relación al valor medio. Específicamente, es de esperar que cuanto más concentrada esté la distribución de la v.a. X alrededor de μ menor será la varianza de X, y viceversa.

Varianza*:* Sea X una v.a. con media μ. La varianza de X, $V(X)$, es.

$$\sigma_X^2 = E\left[(X-\mu)^2\right].$$

Dependiendo de si X es una variable aleatoria discreta o continua, la esperanza se calculará como una sumatoria o como una integral, respectivamente.

Teorema: Sea X una v.a. con distribución de probabilidad $f_X(x)$ y media μ. La varianza de X, $V(X)$, está dada por

$$\sigma_X^2 = \sum_x (x-\mu)^2 f_X(x), \text{ si } X \text{ es discreta, y}$$

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\, dx, \text{ si } X \text{ es continua.}$$

Se observa que la varianza es menor cuando la distribución está más concentrada alrededor de μ, y mayor cuando está más dispersa. Su expresión matemática puede interpretarse con la esperanza matemática de las distancias cuadráticas de los valores asumidos por la variable aleatoria a su media. Como se trata de una distancia cuadrática, su magnitud es el cuadrado de la magnitud de la media, por lo cual a veces conviene hablar de su raíz cuadrada positiva, denominada desviación estándar.

Desviación Estándar*:* La desviación estándar se denota por σ_X y es la raíz cuadrada positiva de σ_X^2.

Ejemplo 4.4: Se quiere abrir un nuevo negocio y de estadísticas anteriores en el rubro que incursionará se sabe que

Porcentaje	Ganancia primer año (USD)
20%	\$50.000 pérdida
30%	\$0
40%	\$50.000 ganancia
10%	\$150.000 ganancia

Usando la información de la tabla como si fueran las probabilidades de ganancias, se puede calcular el valor esperado de la ganancia al primer año como

$$\mu = \sum_x x_i \cdot p(x_i) = 25.$$

Para calcular una medida de variabilidad se tiene que

prob	Ganancia (x1000)	x_i p(x_i)	$(x_i - m)^2$ p(x_i)
0,2	-50	-10	1125
0,3	0	0	187,5
0,4	50	20	250
0,1	150	15	1562,5

Por lo que

$$\sigma_X^2 = \sum_i (x_i - \mu)^2 \; p(x_i) = 3125 \text{, y } \quad \sigma_X = \sqrt{\sum_i (x_i - \mu)^2 \; p(x_i)} = 55,9.$$

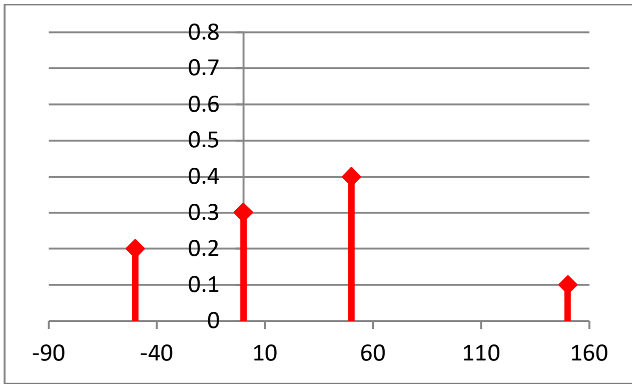
Si la información a priori sobre la ganancia fuera en realidad

prob	Ganancia primer año (USD)
0,1	\$50.000 pérdida
0,1	\$0
0,7	\$30.000 ganancia
0,1	\$90.000 ganancia

En este caso $\mu = \sum_x x_i \cdot p(x_i) = 25$, $\sigma_X^2 = \sum_i (x_i - \mu)^2 \; p(x_i) = 1065$ y

$$\sigma_X = \sqrt{\sum_i (x_i - \mu)^2 \; p(x_i)} = 32,6 \; .$$

Se observa que en esta segunda configuración la esperanza matemática es la misma que en el caso anterior (\$25.000), pero los valores posibles asumidos por esta última variable aleatoria están menos dispersos (ver gráfico).



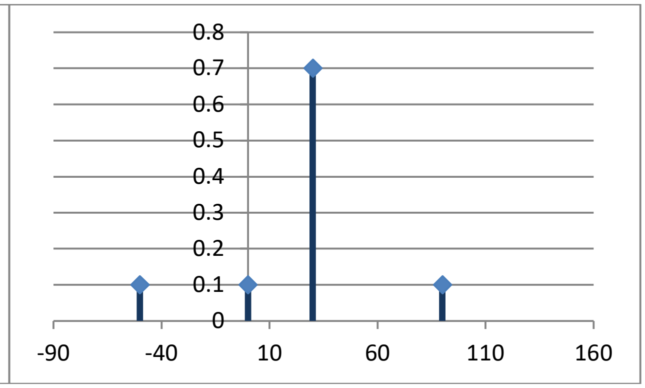


Figura 4.1: Funciones de masa de las utilidades en las dos configuraciones analizadas

Se observa que en esta segunda configuración la esperanza matemática es la misma que en el caso anterior ($25.000), pero los valores posibles asumidos por esta última variable aleatoria están menos dispersos (ver figura 4.1).

Cuando la varianza o el desvío estándar se debe calcular de manera manual, a veces es conveniente usar alguna expresión alternativa como la que se presenta en el siguiente teorema:

Teorema: Sea X una v.a. con media μ, entonces $V(X)=E(X^2)-\mu^2$.

Dem: Por definición $V(X)=E\left[(X-\mu)^2\right]$, por lo que

$V(X)=E\left[(X-\mu)^2\right]=E\left[X^2-2X\mu+\mu^2\right]$

Considerando que se trata de una variable aleatoria continua se tiene que

$$
\begin{aligned}
V(X) &= \int_{-\infty}^{\infty}\left(x^2-2x\mu+\mu^2\right)f(x)dx \\
&= \int_{-\infty}^{\infty} x^2 f(x)dx - \int_{-\infty}^{\infty} 2x\mu f(x)dx + \int_{-\infty}^{\infty} \mu^2 f(x)dx \\
&= E(X^2)-2\mu E(X)+\mu^2 = E(X^2)-\mu^2.
\end{aligned}
$$

Teorema: Sea X una v.a. con distribución de probabilidad $f_X(x)$. La varianza de la variable aleatoria $g(X)$ está dada por

$$V\left[g(X)\right]=E\left[\left\{g(X)-\mu_{g(X)}\right\}^2\right]=\sum_x\left\{g(x)-\mu_{g(X)}\right\}^2 f_X(x)\text{, si } X \text{ es discreta, y}$$

$$V\left[g(X)\right]=E\left[\left\{g(X)-\mu_{g(X)}\right\}^2\right]=\int_{-\infty}^{\infty}\left\{g(x)-\mu_{g(X)}\right\}^2 f_X(x)\,dx\text{, si } X \text{ es continua.}$$

Como se manifestó anteriormente la media es una medida de posición y la varianza o desvío estándar de variabilidad. Estas medidas descriptivas de las distribuciones de variables aleatorias no necesariamente siempre existen. Se pueden encontrar situaciones donde las integrales no convergen.

Por otro lado, cuando existen, ellas describen el comportamiento de las distribuciones de probabilidad pero no las caracterizan, con esto se quiere decir que pueden existir dos distribuciones de probabilidad diferentes con las mismas medias y varianzas.

4.4 Teorema de Chebyshev

Existe un teorema que usando la media y el desvío estándar da información acerca de con qué probabilidad la v.a. asume valores próximos a la media sin necesidad de conocer toda la función de probabilidad en su totalidad. Permite obtener cotas de probabilidad sin hacer suposiciones sobre la distribución de la variable aleatoria, excepto que tenga media y varianza finitas.

Teorema de Chebyshev: La probabilidad de que cualquier variable aleatoria X asuma un valor dentro de k desviaciones estándares de la media es al menos $1-1/k^2$, es decir

$$\mathrm{P}(\mu\text{-}k\sigma_{\mathrm{X}} < X < \mu + k\sigma_{\mathrm{X}}) \geq 1-1/k^2.$$

Las cotas pueden expresarse de diferentes maneras equivalente:

- $\mathrm{P}(\mu\text{-}k\sigma_{\mathrm{X}} < X < \mu + k\sigma_{\mathrm{X}}) \geq 1-1/k^2$
- $\mathrm{P}(|X-\mu|X < k\sigma_{\mathbf{X}}) \geq 1-1/k^2$
- $\mathrm{P}(|X-\mu|X \geq k\sigma_{\mathbf{X}}) < 1/k^2$

Del Teorema de Chebyshev se deduce que al menos una fracción $1-1/k^2$ de los valores de X se encuentran dentro de k desviaciones estándar de la media.

Ejemplo 4.5: Una variable aleatoria X tiene esperanza μ=2 y varianza σ^2= 4. Encuentre cotas para

a) $P(\ |X\text{-}2| \geq 4)$

b) $P(\ |X\text{-}2| < 5)$

c) $P(\text{-}1<X<5)$

d) Encuentre el valor de la constante c tal que $P(\ |X\text{-}2| \geq c) \leq 1/9$.

a) $P(\ |X\text{-}2| \geq 4)$

$$P\left(|X-2|\geq 4\right) = P\left(|X-2|\geq 2\cdot 2\right) = P\left(|X-\mu|\geq 2\cdot\sigma\right) < 1\big/2^2$$

b) $P(\ |X\text{-}2| < 5)$

$$P\left(|X-2|<5\right) = P\left(|X-2|<\frac{5}{2}\cdot 2\right) = P\left(|X-\mu|<\frac{5}{2}\cdot\sigma\right) \geq 1-\left(5\big/2\right)^{-2}$$

c) $P(\text{-}1<X<5)$

$$P(-1<X<5) = P(-1-2<X-2<5-2) = P(-3<X-2<3)$$

$$= P\left(|X-2|<3\right) = P\left(|X-2|<\frac{3}{2}\cdot 2\right) = P\left(|X-\mu|<\frac{3}{2}\cdot\sigma\right) \geq 1-\left(3\big/2\right)^{-2}$$

d) Encuentre el valor de la constante c tal que $P(|X\text{-}2| \geq c) \leq 1/9$.

Puesto que

$$\mathrm{P}(|X-\mu| \geq 3\sigma_{\mathrm{X}}) < 1/3^2$$

Se tiene que c=3σ=6.

Ejemplo 4.6 (caso continuo): El tiempo necesario para el cambio de un amortiguador de un auto de cierta marca bajo uso continuo o severo, puede ser considerado como una v.a. continua, medida en años. Suponga que su función de densidad está dada por:

$$f(x)=\begin{cases} \frac{1}{4}x & 0\le x\le 2 \\ \frac{1}{8} & 2<x\le 6 \\ 0 & c.c. \end{cases}$$

Usando Chebyshev indique entre qué tiempos al menos el 75% de los autos necesitarán cambio de amortiguadores.

Para encontrar una cota inferior para k=2, pues

$$P(\boldsymbol{\mu}\text{-}\boldsymbol{k\sigma_X} < \boldsymbol{X} < \boldsymbol{\mu} + \boldsymbol{k\sigma_X}) \ge \mathbf{1} - \frac{\mathbf{1}}{\boldsymbol{k^2}} = \mathbf{0,75} = \mathbf{1} - \mathbf{0,25} = \mathbf{1} - \mathbf{1/4}$$

Consecuentemente, se debe calcular $\boldsymbol{\mu}$ y $\boldsymbol{\sigma_X}$.

$$\boldsymbol{\mu} = \int_0^2 \frac{1}{4}x^2\,dx + \int_2^6 \frac{1}{8}\,x\,dx = \frac{1}{4\cdot 3}x^3\Big|_0^2 + \frac{1}{8\cdot 2}\,x^2\Big|_2^6$$

$$\boldsymbol{\mu} = \frac{2}{3} + 2 = \frac{8}{3}$$

Para calcular la varianza se puede usar el resultado $\boldsymbol{V(X) = E(X^2) - \mu^2}$

$$E(X^2) = \int_0^2 \frac{1}{4}x^3\,dx + \int_2^6 \frac{1}{8}\,x^2\,dx = \frac{1}{4\cdot 4}x^4\Big|_0^2 + \frac{1}{8\cdot 3}\,x^3\Big|_2^6$$

$$E(X^2) = 1 + \frac{26}{3} = \frac{29}{3}$$

$$V(X) = E(X^2) - \mu^2 = \frac{29}{3} - \frac{64}{9} = \frac{87-64}{9} = \frac{23}{9}$$

Consecuentemente $\sigma=\sqrt{\frac{23}{9}} \cong \mathbf{1,6}$. Por lo tanto $k\sigma=3.2$, y se tiene que

$$P(\mathbf{2,66}\text{-}3{,}2 < \boldsymbol{X} < \mathbf{2,66} + \mathbf{3,2}) \ge \mathbf{0,75}$$

Como el tiempo no puede ser negativo y está en el rango $0\le x\le 6$, se tiene el intervalo $0\le X\le 5.87$, al menos el 75% de los autos necesitarán cambio de amortiguadores hasta los 5,87 años.

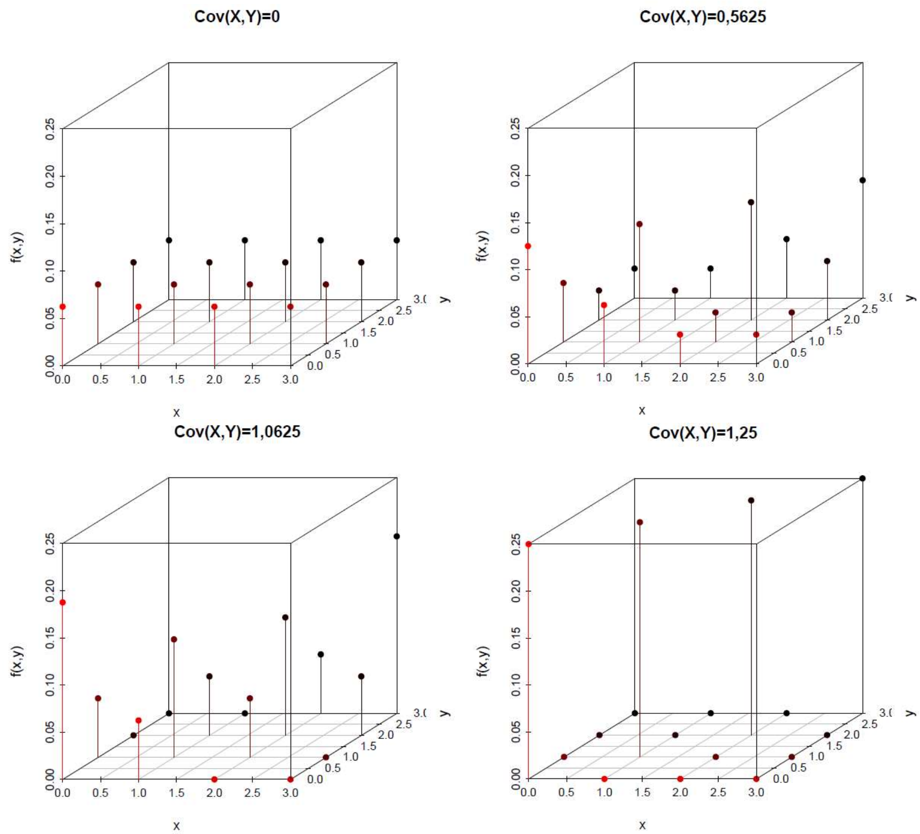
4.6 Covarianza

Covarianza: Sean X e Y vs.as. con distribución de probabilidad conjunta $f(x,y)$ y medias μ_X y μ_Y, respectivamente. La covarianza de X e Y, $Cov(X,Y)$ ó σ_{XY}, es

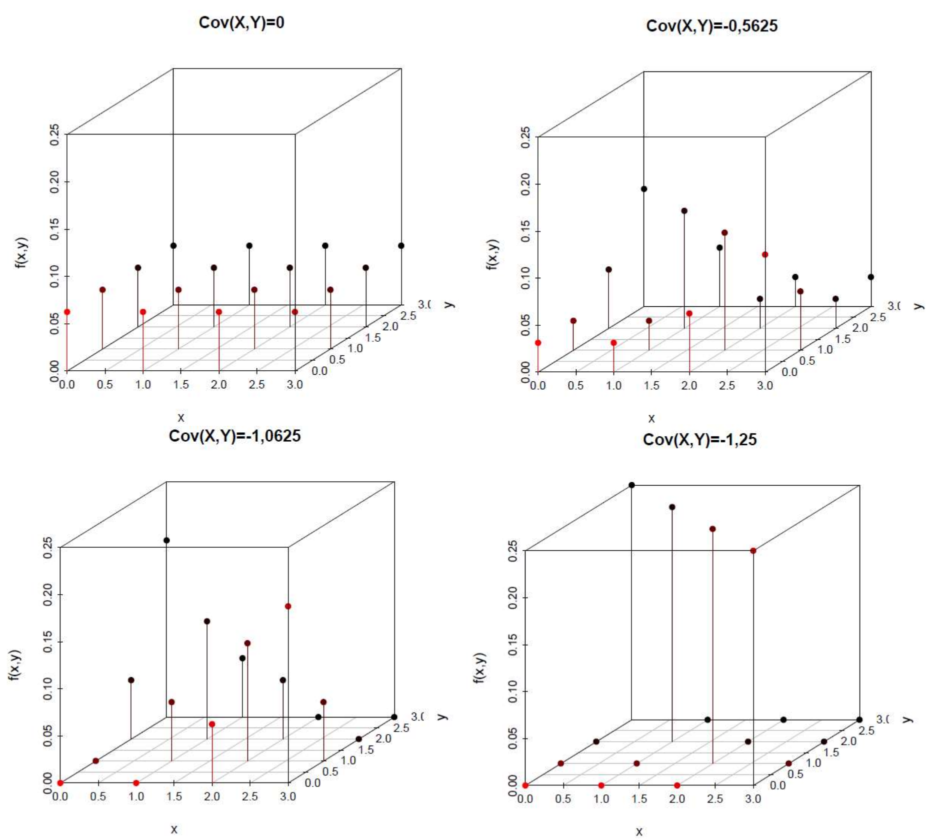
$$\sigma_{XY} = E\left[(X-\mu_X)(Y-\mu_Y)\right].$$

¿Qué mide la covarianza? La covarianza mide la dependencia lineal (tendencia de las variables a moverse en una misma dirección)

covarianza positiva: tendencia de alineación en el sentido de los cuadrantes impares



covarianza negativa: tendencia de alineación en los cuadrantes pares



Así como para la variancia, hay una fórmula equivalente para el cálculo de la covarianza que evita el centrar las variables para el cálculo de la esperanzas.

Teorema: $\sigma_{XY} = E(XY) - \mu_X \mu_Y$.

Ejemplo 4.7: Sean las variables U y V con dn. de masa conjunta

		v				
		2	**4**	**6**	**8**	**10**
u	**2**	0,1	0	0	0	0
	3	0	0,2	0	0,1	0
	4	0	0	0,2	0	0
	5	0	0,1	0	0,2	0
	6	0	0	0	0	0,1

Para calcular la Cov(U,V) se observa que:

- La distribución marginal de U viene dada por la función de masa

u	2	3	4	5	6
g(*u*)	0,1	0,3	0,2	0,3	0,1

- La distribución marginal de V viene dada por la función de masa

v	2	4	6	8	10
h(v)	0,1	0,3	0,2	0,3	0,1

Consecuentemente $E(U) = 4$ y $E(V) = 6$.

$$E(U \cdot V) = 4 \cdot 0{,}1 + 12 \cdot 0{,}2 + 24 \cdot 0{,}1 + 24 \cdot 0{,}2 +$$
$$+ 20 \cdot 0{,}1 + 40 \cdot 0{,}2 + 60 \cdot 0{,}1 = 26$$

De esta manera $\sigma_{UV} = E(UV) - \mu_U \mu_V = 26 - 4 \cdot 6 = 2$.

4.7 Medias y Varianzas de Combinaciones Lineales de Variables Aleatorias

Es importante recordar algunas propiedades de las medias, varianzas y covarianzas que simplifican el cálculo en algunas situaciones.

Propiedades de las medias, varianzas y covarianzas de combinaciones lineales de vs.as.:
Sean X e Y vs.as. y sean a, b, c y d constantes

(1) $E(aX + b) = a\,E(X) + b$

Dem:

$$\begin{aligned} E(aX+b) &= \int_{-\infty}^{\infty} (a \cdot x + b) f(x) dx \\ &= \int_{-\infty}^{\infty} a \cdot x \cdot f(x) dx + \int_{-\infty}^{\infty} b \cdot f(x) dx \\ &= a \cdot \int_{-\infty}^{\infty} x \cdot f(x) dx + b \cdot \int_{-\infty}^{\infty} f(x) dx \\ &= a \cdot E(X) + b \end{aligned}$$

(2) $E\left[g(X) \pm h(X)\right] = E\left[g(X)\right] \pm E\left[h(X)\right]$

Dem:

$$\begin{aligned}E(g(X)+h(X)) &= \int_{-\infty}^{\infty} (g(x)+h(x)) f(x)dx \\ &= \int_{-\infty}^{\infty} g(x)\cdot f(x)dx + \int_{-\infty}^{\infty} h(x)\cdot f(x)dx \\ &= E(g(X)) + E(h(X))\end{aligned}$$

(3) $E\left[g(X,Y) \pm h(X,Y)\right] = E\left[g(X,Y)\right] \pm E\left[h(X,Y)\right]$

Dem:

$$\begin{aligned}E(g(X,Y)+h(X,Y)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (g(x,y)+h(x,y)) f(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y)dxdy \\ &= E(g(X,Y)) + E(h(X,Y))\end{aligned}$$

(4) $E\left[g(X) \pm h(Y)\right] = E\left[g(X)\right] \pm E\left[h(Y)\right]$

Dem:

$$\begin{aligned}E(g(X)+h(Y)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (g(x)+h(y)) f(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x) f(x,y)dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(y) f(x,y)dxdy \\ &= E(g(X)) + E(h(Y))\end{aligned}$$

(5) Si X e Y son independientes entonces $E(X\cdot Y) = E(X)\cdot E(Y)$

Dem:

$$\begin{aligned}E(X\cdot Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\cdot y\cdot f(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\cdot y\cdot g(x)h(y)dxdy \\ &= \int_{-\infty}^{\infty} y\cdot h(y)\left(\int_{-\infty}^{\infty} x\cdot g(x)dx\right)dy = \int_{-\infty}^{\infty} y\cdot h(y)\cdot E(X)dy \\ &= E(X)\cdot \int_{-\infty}^{\infty} y\cdot h(y)dy = E(X)\cdot E(Y)\end{aligned}$$

(6) $\sigma^2_{aX+b} = a^2\sigma^2_X$

Dem:

$$\begin{aligned}V(aX+b) &= E\left[\left(aX+b-\mu_{aX+b}\right)^2\right] = E\left[\left(aX+b-a\mu_X-b\right)^2\right] \\ &= E\left[a^2\left(X-\mu_X\right)^2\right] = a^2E\left[\left(X-\mu_X\right)^2\right] = a^2\cdot V(X)\end{aligned}$$

(7) $\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}$

Dem:

$$V(aX+bY) = E\left[\left(aX+bY-\mu_{aX+bY}\right)^2\right] = E\left[\left(aX+bY-a\mu_X-b\mu_Y\right)^2\right]$$
$$= E\left[\left(aX-a\mu_X+bY-b\mu_Y\right)^2\right] = E\left[(aX-a\mu_X)^2+(bY-b\mu_Y)^2+2(aX-a\mu_X)(bY-b\mu_Y)\right]$$
$$= E\left[a^2(X-\mu_X)^2+b^2(Y-\mu_Y)^2+2ab(X-\mu_X)(Y-b\mu_Y)\right]$$
$$= E\left[a^2(X-\mu_X)^2\right]+E\left[b^2(Y-\mu_Y)^2\right]+2abE\left[(X-\mu_X)(Y-b\mu_Y)\right]$$
$$= a^2E\left[(X-\mu_X)^2\right]+b^2E\left[(Y-\mu_Y)^2\right]+2abE\left[(X-\mu_X)(Y-b\mu_Y)\right]$$
$$= a^2V(X)+b^2V(Y)+2abCov(X,Y)$$

(8) Si X e Y son independientes entonces $\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y$

Dem

Si X e Y son independientes entonces $\sigma_{XY} = 0 \Rightarrow \sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y$

(9) $Cov(aX+b, cY+d) = a\,cCov(X,Y)$

Dem:

$$Cov(aX+b,cY+d) = E\left[(aX+b-a\mu_X-b)(cY+d-c\mu_X-d)\right]$$
$$= E\left[(aX-a\mu_X)(cY-c\mu_X)\right]$$
$$= E\left[ac(X-\mu_X)(Y-\mu_X)\right] = acCov(X,Y)$$

(10) $Cov(\sum_{i=1}^{A} a_iX_i, \sum_{j=1}^{B} b_jY_j) = \sum_{i=1}^{A}\sum_{j=1}^{B} a_ib_jCov(X_i,Y_j)$, donde $a_1,...,a_A$ y $b_1,...,b_B$ son constantes y, $X_1,...,X_A$ y $Y_1,...,Y_B$ son vs.as.

Dem: Vamos a demostrar para A=B=2

$$Cov(a_1X_1+a_2X_2, b_1Y_1+b_2Y_2) = E\left[(a_1X_1+a_2X_2-a_1\mu_{X_1}-a_2\mu_{X_2})(b_1Y_1+b_2Y_2-b_1\mu_{Y_1}-b_2\mu_{Y_2})\right]$$
$$= E\left[(a_1(X_1-\mu_{X_1})+a_2(X_2-\mu_{X_2}))(b_1(Y_1-\mu_{Y_1})+b_2(Y_2-\mu_{Y_2}))\right]$$
$$= E\left[(a_1(X_1-\mu_{X_1})b_1(Y_1-\mu_{Y_1})+a_2(X_2-\mu_{X_2})b_1(Y_1-\mu_{Y_1}))+(a_1(X_1-\mu_{X_1})b_2(Y_2-\mu_{Y_2})+a_2(X_2-\mu_{X_2})b_2(Y_2-\mu_{Y_2}))\right]$$
$$= a_1b_1\,\mathrm{cov}(X_1,Y_1)+a_2b_1\,\mathrm{cov}(X_2,Y_1)+a_1b_2\,\mathrm{cov}(X_1,Y_2)+a_2b_2\,\mathrm{cov}(X_2,Y_2)$$

Ejemplos 4.8:

1. En promedio, los estudiantes utilizan las redes sociales en un día de fin de semana 8 veces. Si se seleccionan 3 estudiantes, para calcular el valor esperado de acceso a las redes sociales el próximo fin de semana entre ellos tres, se puede pensar en que X_1, X_2 y X_3 representan el número de veces que accesan a redes sociales cada uno de los tres estudiantes respectivamente. El número de acceso a las redes sociales entre los tres es $X_1+X_2+X_3$, y el número esperado de acceso entre los tres será $E(X_1+X_2+X_3) = E(X_1)+E(X_2)+E(X_3) = 8\times 3 = 24$.

2. Cuando el número esperado de errores por páginas en un libro es de 1,5, el número el número esperado de errores en un capítulo con 8 páginas será $E\left(\sum_{i=1}^{8} X_i\right) = \sum_{i=1}^{8} E(X_i) = 8 \times 1,5 = 12.$