

Introducción a la
E S T A D Í S T I C A

Jaime Vázquez • Lizbeth Naranjo • Ruth Fuentes • Margarita Chávez

Proyecto PAPIME UNAM PE107117
“Estadística para estudiantes de ciencias”

Índice general	1
Introducción	3
1. ¿Qué es la estadística?	4
1.1. La probabilidad y la estadística	5
1.2. Enfoques de la estadística	6
1.2.1. El concepto de muestra aleatoria	8
2. Estadística descriptiva	11
2.1. Introducción	11
2.2. Datos y variables	12
2.2.1. Tipos de variables	12
2.2.2. Escalas de medición	13
2.2.3. Formas de recolección	14
2.3. Los datos y R	14
2.3.1. Cargando los datos	15
2.4. Representación gráfica de los datos	16
2.4.1. Frecuencia, frecuencia relativa y frecuencia acumulada	16
2.4.2. Tablas de frecuencia	17
2.4.3. Gráficas de barras	17
2.4.4. Gráficas de pay o pastel	18
2.4.5. Diagrama de tallo y hojas	20
2.4.6. Histograma	21
2.4.7. Boxplot o diagrama de caja y brazos	23

2.5.	Medidas muestrales de tendencia central	26
2.5.1.	La media muestral	26
2.5.2.	La mediana muestral	26
2.5.3.	Percentiles	27
2.6.	Medidas de dispersión	28
2.6.1.	Varianza y desviación estándar muestral	28
2.6.2.	Rango muestral	29
2.6.3.	Rango intercuartil	30
2.7.	Otras medidas de resumen	30
2.8.	Relaciones lineales entre variables	31
2.9.	Ejercicio	33
3.	Estadísticas y distribuciones muestrales	34
3.1.	Introducción	34
3.2.	Distribución de las estadísticas muestrales bajo normalidad	38
3.2.1.	Distribución de la media muestral	39
3.2.2.	La distribución de la varianza muestral	39
3.2.3.	La distribución F de Fisher y el cociente de varianzas muestrales	48
3.2.4.	La distribución t de Student y algunas estadísticas relacionadas	51
3.3.	Estadísticas de orden	53
3.3.1.	r -ésima estadística de orden (Y_r)	54
3.3.2.	Distribución conjunta de las estadísticas de orden mínima y máxima	55
3.4.	Estadísticas suficientes	57
3.4.1.	El concepto de suficiencia	59
3.4.2.	El teorema de factorización	63
3.4.3.	La familia exponencial	69
3.4.4.	Suficiencia minimal	71
3.5.	Completez	76
3.6.	Algunas generalizaciones	79
3.7.	Estadísticas auxiliares	81
3.8.	Ejercicios	84
	Bibliografía	95

Introducción

Se puede decir que la estadística es una disciplina reciente con relación a otras ramas de las matemáticas, no obstante que desde la antigüedad existieron actividades relacionadas con el manejo de cifras, tales como las cuentas y datos vinculados con las poblaciones de las ciudades o, más generalmente, de los *estados*; situación que posiblemente es la génesis del término “estadística”. En Rao (1997), el estadístico hindú Radhakrishna Rao menciona que “la estadística tiene gran antigüedad pero escasa historia”, frase que refleja el hecho de que la estadística es joven como área formal, pero que prácticamente ha estado presente a lo largo de la historia de la humanidad.

Entre los científicos que más han aportado a la estadística para convertirla en la disciplina con el carácter matemático y de importante aplicación para ayudar a explicar fenómenos del mundo real, sobresalen Francis Galton (1822-1911), Karl Pearson (1857-1936), Charles Spearman (1863-1945), Ronald Fisher (1890-1962) y Jerzy Neyman (1894-1981), sólo por mencionar algunos.

La estadística inferencial es una disciplina que se basa en gran medida en la probabilidad y que ayuda a resolver problemas mediante inferencias de alguna característica de la población usando datos muestrales de la misma. Por ejemplo, los estadísticos pueden realizar estudios de opinión, en donde a través del punto de vista de algunos ciudadanos que componen una muestra suficientemente representativa, se puede medir el pulso de temas de interés para el país.

La estadística involucra conceptos y resultados que pueden resumirse en grandes temas: análisis exploratorio de datos, distribuciones muestrales, estimación puntual, estimación por intervalo y pruebas de hipótesis, los cuales son fundamentales en el estudio y la aplicación de esta disciplina. En esta parte se abordarán los dos primeros temas.

Para la lectura de este documento, es importante contar con conocimientos de teoría de la probabilidad, así como de cálculo diferencial e integral en una y varias variables.

CAPÍTULO 1

¿Qué es la estadística?

El progreso de la ciencia con frecuencia se adscribe a la experimentación. El investigador lleva a cabo un experimento, una encuesta o un conjunto de mediciones; obtiene datos y con base en ellos se busca sustentar una hipótesis o responder a una pregunta de investigación. Es decir, a partir de un experimento particular, es deseable generalizar hacia la clase de todos los experimentos similares.

La estadística no se refiere únicamente a la recolección de datos y a la presentación de cuadros y tablas resumen. Actualmente se comprende como la ciencia que basa la inferencia en datos observados y toma decisiones en términos de incertidumbre. Aunque en su estado actual no puede manejar todas las situaciones que se presentan alrededor de la incertidumbre, constantemente se desarrollan nuevas técnicas de análisis.

La estadística está presente en muchos ámbitos: el científico, el social y el empresarial, sólo por mencionar algunos. Por ejemplo, tanto en la iniciativa privada como en el ámbito gubernamental, es vital contar con un manejo adecuado de la información y el análisis de grandes bases de datos que sirva a los diferentes agentes de la economía y la política al momento de tomar decisiones. Actualmente existen profesionales de la estadística que utilizan diversas técnicas de este campo en empresas privadas, tales como casas encuestadoras que se dedican a hacer estudios de mercado, o bien en industrias relacionadas con diversos campos de la actividad económica.

Ejemplos de instituciones en donde el uso de la estadística es fundamental son: el INEGI, las secretarías de estado, el IMP, PEMEX, el Banco de México y las aseguradoras, sólo por mencionar algunas, pues el análisis y modelado de datos puede hacerse prácticamente en cualquier entidad en donde se cuente con información de cierto fenómeno o proceso. Adicionalmente, la estadística juega un papel importante en los estudios de mercado y otros

procesos en donde es necesario obtener datos para posteriormente analizarlos y llegar a conclusiones mediante un modelo que dependerá de la naturaleza de dicha información. Algunos estadísticos participan en el diseño y validación de encuestas y conteos rápidos relacionados con procesos electorales.

Se puede decir entonces que la *estadística* se ocupa de los métodos científicos para recolectar, organizar, resumir, presentar y analizar datos usando modelos, así como de obtener conclusiones válidas y tomar decisiones con base en ese análisis. Es la rama de la matemática que utiliza conjuntos de datos para obtener inferencias basadas en el *cálculo de probabilidades*.

En los siguientes párrafos se pretende explicar la relación entre la probabilidad y la estadística, así como establecer las diferencias entre los enfoques para analizar un conjunto de datos.

1.1. La probabilidad y la estadística

La teoría de probabilidad permite modelar ciertos fenómenos que ocurren en la naturaleza, siendo el modelo básico un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y una variable aleatoria X definida en ese espacio. En el contexto paramétrico, dada una variable aleatoria X , se tiene asociada una función de densidad $f(x; \theta)$, la cual actúa en función de características desconocidas llamadas parámetros.

Gracias al conocimiento de la función de densidad de probabilidad, y por lo tanto de la función de distribución de una variable aleatoria X , se facilita la construcción de espacios de probabilidad adaptados a fenómenos aleatorios concretos. Es decir, gracias a la teoría de la probabilidad se puede construir un modelo para describir una determinada variable aleatoria real a través de su función de distribución. Por ejemplo, gracias a la teoría de la probabilidad se pueden hacer aseveraciones como “el número de accidentes que ocurren en una importante intersección vial de la Ciudad de México durante un determinado periodo de tiempo es una variable aleatoria X que tiene una distribución $Poisson(\lambda)$ ” o “la vida (tiempo de supervivencia) de un cierto tipo de foco tiene una distribución Exponencial (λ) ”. Asimismo, se pueden contestar preguntas probabilísticas relacionadas con la variable aleatoria como: ¿cuál es la probabilidad de que no haya accidentados en ese cruce importante de avenidas en un día determinado? o ¿cuál es la probabilidad de que el foco dure más de un determinado periodo de tiempo?. Sin embargo, las respuestas a estas preguntas quedan en términos de parámetros; por ejemplo si X tiene distribución $Poisson(\lambda)$, $\mathbb{P}(X = 0) = e^{-\lambda}$, está en función de λ , donde $\lambda > 0$. Si se conociera el valor de λ , sería posible obtener un valor numérico para las preguntas probabilísticas asociadas a la variable aleatoria.

De esta manera, si se desea profundizar en la forma de adaptar un modelo probabilístico a cada fenómeno concreto, sería necesario llevar a cabo observaciones del fenómeno en cuestión con la finalidad de poder hacer conclusiones acerca de los parámetros de una población y llegar a resultados numéricos en el cálculo de probabilidades.

Es aquí donde la *estadística* juega un papel importante al analizar e interpretar la información obtenida de una cierta población con la finalidad de poder concluir sobre la ley de probabilidad que rige un fenómeno aleatorio. Este procedimiento se conoce como *inferencia estadística* o *estadística matemática*.

Cuando se tiene un conjunto de observaciones acerca del fenómeno considerado, se accede al terreno de la estadística con la finalidad de obtener información acerca de la población en estudio con base en un conocimiento parcial o no exhaustivo de dicha población, ya que, en la mayoría de los casos, una observación exhaustiva es imposible o muy costosa. Por ejemplo, si se desea saber la intención del voto para una elección presidencial, no se encuestará a todos los ciudadanos en edad de votar, sino a una muestra representativa de la población (cómo elegir una muestra representativa, el tamaño adecuado de la misma y la forma de hacer la encuesta es objeto de estudio de otra materia), y con las observaciones obtenidas y los métodos de la inferencia estadística se puede decir con cierta precisión el porcentaje de la población que en ese momento votaría por cada uno de los candidatos de la contienda electoral.

Con frecuencia el término *estadística* se entiende como el proceso de recolección de datos u observaciones, así como el tratamiento numérico que se le da a estos datos a través de gráficas y medidas que resumen la información, pero es necesario recalcar que el objetivo de la inferencia estadística es obtener conclusiones acerca de alguna característica de la población en estudio a través del análisis e interpretación de las observaciones. Sin embargo, el tratamiento numérico de los datos no es menos importante e idealmente un procedimiento inferencial va antecedido por un tratamiento descriptivo.

1.2. Enfoques de la estadística

La estadística se puede analizar desde dos enfoques: el descriptivo y el inferencial.

- Enfoque **descriptivo**: Resumen y descripción de un conjunto de datos mediante gráficas y medidas descriptivas.
- Enfoque **inferencial**: Análisis e interpretación de la información obtenida de una población para hacer conclusiones generales acerca de las características desconocidas de dicha población.

De acuerdo a lo que se desea conocer del parámetro, la *inferencia estadística* puede abordarse de la siguiente manera:

1. **Estimación puntual**: Consiste en obtener un valor numérico único para el parámetro.
2. **Estimación por intervalo**: Consiste en obtener un rango de valores que el parámetro puede tomar.

3. **Pruebas de hipótesis:** Consiste en contrastar dos aseveraciones acerca de la distribución de una variable aleatoria (usualmente representadas en términos de valores que el parámetro puede tomar).

Asímismo, la inferencia estadística se puede realizar desde una perspectiva diferente (no clásica) conocida como método *Bayesiano* o **Estadística Bayesiana**, el cual es un paradigma en el que se asocia una distribución inicial al parámetro y, junto con las observaciones, se obtiene una distribución final para dicho parámetro. Con base en este planteamiento, se hacen inferencias desde la óptica de estimación puntual o por intervalos y de pruebas bayesianas.

También puede hablarse de inferencia estadística **paramétrica** y **no paramétrica**. En el primer caso se conoce la forma de la distribución pero se desconocen los parámetros y es sobre éstos que se realizará la inferencia. En el segundo caso se desconocen la forma de la distribución y los parámetros y las inferencias se hacen sobre ciertas características que no tienen por qué corresponder a parámetros de una distribución.

Definición 1.1 *Al conjunto de valores que el parámetro θ puede tomar se le llama **espacio paramétrico** (o **parametral**) y se le denota por Θ .*

Ejemplos:

- En el caso de la distribución Bernoulli $\Theta = [0, 1]$.
- En el caso de la distribución Poisson $\Theta = (0, \infty)$.

La inferencia estadística también suele denominarse *estadística inductiva* o *inferencia inductiva*, pues va de lo particular a lo general al hacer conclusiones de una población a partir de una muestra representativa de la misma. La necesidad de trabajar con un conjunto reducido de la población está asociada a varios factores, siendo uno de los más importantes el económico. Se puede decir que la **población** es el conjunto de observaciones o individuos sobre los que se desea información y de ella se extrae un subconjunto al que se denomina **muestra**. El procedimiento para obtener las observaciones con las cuales se intenta disminuir el grado de desconocimiento de θ debe ser tal que la recolección de datos se realice siempre bajo las mismas condiciones y sin importar los resultados anteriores.

En un contexto experimental, la definición de la muestra puede depender de cuáles son los tratamientos que se desea comparar, bajo qué condiciones y si las comparaciones se harán con tamaños de muestra y repeticiones iguales. En la literatura sobre diseño de experimentos, se aborda este tipo de planteamiento. Si se hace referencia al estudio de poblaciones, hay diferentes formas de obtener una muestra representativa. Aunque la teoría del muestreo requiere una asignatura aparte, a continuación se resumen las principales características de los muestreos probabilísticos más usuales.

Muestreo aleatorio simple: se trata de un procedimiento de muestreo (sin reemplazo), en el que se seleccionan n unidades de las N en la población, de forma que cualquier posible muestra del mismo tamaño tenga la misma probabilidad de ser elegida.

Muestreo por conglomerados: se divide la población en grupos de acuerdo con su proximidad geográfica o de otro tipo (conglomerados). Se busca que cada grupo sea heterogéneo y que tenga representadas todas las características de la población. Se selecciona una muestra de conglomerados al azar y se toma el conglomerado completo o una muestra del mismo.

Muestreo estratificado: se divide la población en grupos homogéneos (estratos) de acuerdo con las características a estudiar. Por ejemplo, en un estudio de las características socio-económicas de una ciudad los estratos pueden ser las colonias de la misma, ya que las colonias suelen presentar características diferenciales. Se selecciona una muestra aleatoria de cada estrato tratando de que todos los estratos de la población queden representados.

En un muestreo estratificado se consideran todos los estratos y en cada uno se considera una muestra de individuos. En el muestreo por conglomerados se seleccionan al azar los conglomerados que serán considerados y una vez elegidos se estudian todos los individuos de cada conglomerado.

Existen otros tipos de muestreo, pero todo lo que se verá en los siguientes capítulos está pensado para poblaciones con muestreo aleatorio simple.

1.2.1. El concepto de muestra aleatoria

El muestreo aleatorio simple garantiza una muestra representativa de la población y la obtención de observaciones independientes. Esta terminología de muestreo puede entenderse si se considera una de las siguientes situaciones que aparecen frecuentemente en la estadística:

- (1) Se extraen los objetos, uno cada vez, de una colección finita llamada población y se determina una característica particular de interés de cada objeto extraído. Después de cada observación y antes de la siguiente extracción, se devuelve el objeto extraído y se mezcla la población de objetos.
- (2) Se extraen los objetos de una población finita como en (1), excepto que los objetos no se reemplazan.

La población de objetos puede ser una colección de personas y la característica observada puede ser el peso, color de ojos, preferencia política o alguna otra.

Si se supone que cada selección es aleatoria, el muestreo en (1) recibe el nombre de *muestreo aleatorio con reemplazo* y en (2) *muestreo aleatorio sin reemplazo* o *muestreo aleatorio simple*.

En cierto sentido, el muestreo aleatorio sin reemplazo es mejor que el muestreo aleatorio con reemplazo, ya que algunas veces la mezcla que se requiere con el reemplazo de los objetos no siempre es fácil de conseguir. Considerando los casos extremos, suponga que hay solamente

2 objetos en la población; cuando se extrae uno de ellos, la selección de un segundo objeto proporciona la información completa acerca de la población original si el primero no fue reemplazado. Por otro lado, si una población es muy grande con respecto al tamaño de la muestra que va a ser extraída, es prácticamente lo mismo si los objetos extraídos son o no son reemplazados; el muestreo sin remplazo se convierte en muestreo con remplazo cuando el tamaño de la población es infinito.

La diferencia básica entre los tipos de muestreo (1) y (2) no está en las distribuciones marginales de las observaciones individuales, porque en ambos casos estas observaciones son idénticamente distribuidas. Sin embargo, en el caso (1) el resultado de cualquier observación no está afectado por los resultados de cualesquiera otras observaciones; las observaciones son fenómenos aleatorios independientes. En el caso (2) las observaciones no son independientes.

Hay otro tipo de situación que ocurre continuamente, diferente de (1) y (2), en la cual los resultados son matemáticamente del mismo tipo que (1):

- (3) Las observaciones se obtienen como resultado de realizaciones independientes repetidas de un experimento, bajo condiciones que son idénticas con respecto a los factores que pueden ser controlados.

Esta descripción incluye a (1) como un caso especial, aunque no necesariamente se refiere a una “población” tangible de la cual se va a seleccionar un objeto. Sin embargo, es posible imaginar una población muy grande de posibles resultados, en donde en cada repetición del experimento se cuenta con la misma colección que estaba disponible en el primer ensayo. Esto es, repetir el experimento bajo condiciones idénticas significaría que el primer resultado es “reemplazado” y es nuevamente uno de los candidatos a ser “extraídos” la siguiente vez. En ambos tipos de muestreo (1) y (3), las observaciones son independientes e idénticamente distribuidas. El término *muestreo aleatorio* sin más especificación se referirá a tal proceso.

Suponga que cuando se extrae un objeto la característica a medir se denota por X . A la distribución de X se le llama la *distribución poblacional*.

Ejemplo 1.1 *Una urna contiene cuatro bolas numeradas del 1 al 4. Se extraen 2 aleatoriamente, una cada vez. Sea X_1 el número de la primera bola extraída y X_2 el número de la segunda bola extraída. Hay 12 posibles resultados del experimento:*

$$(1, 2) (1, 3) (1, 4) (2, 3) (2, 4) (3, 4)$$

$$(2, 1) (3, 1) (4, 1) (3, 2) (4, 2) (4, 3)$$

Cada uno de estos 12 resultados tiene probabilidad 1/12. A partir de esto, se pueden calcular las distribuciones de X_1 y X_2 . Por ejemplo,

$$\mathbb{P}(X_1 = 1) = \mathbb{P}[(1, 2), (1, 3) \text{ ó } (1, 4)] = \frac{1}{4}.$$

Similarmente:

$$\mathbb{P}(X_2 = 1) = \mathbb{P}[(2, 1), (3, 1) \text{ ó } (4, 1)] = \frac{1}{4}.$$

De la misma forma se encuentra que para X_1 cada uno de los valores posibles 1, 2, 3 y 4 tiene probabilidad $1/4$ y que X_2 tiene exactamente la misma distribución - la distribución poblacional.

De esta manera, se diseña un experimento y se lleva a cabo para proporcionar la observación X_1 de la característica observable X . El experimento se repite bajo las mismas condiciones proporcionando el X_2 . El proceso continúa hasta tener n observaciones X_1, X_2, \dots, X_n de la característica X ; a estas observaciones se les llaman los valores muestrales de X y se dice que constituyen una muestra aleatoria. Note que cuando se ha obtenido una muestra se tienen n observaciones numéricas (x_1, x_2, \dots, x_n) , pero cuando se está planificando el procedimiento de muestreo y diseñando el método para obtener inferencias aún no se sabe qué valores numéricos se obtendrán y deben considerarse n variables aleatorias, estas variables aleatorias serán, en el contexto que se discute, independientes e idénticamente distribuidas.

Definición 1.2 Si X_1, \dots, X_n es un conjunto de variables aleatorias, independientes e idénticamente distribuidas, entonces se dice que X_1, \dots, X_n es una **muestra aleatoria (m.a.)**.

Observación 1.1 Si X_1, \dots, X_n es una muestra aleatoria, entonces su función de densidad conjunta está dada por

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta).$$

Observación 1.2 La **distribución de la muestra aleatoria** (o simplemente **distribución de la muestra**) será la función de densidad conjunta de las variables aleatorias que componen la muestra aleatoria y, por la observación anterior, se calculará como el producto de las densidades marginales de dichas variables.

En el siguiente capítulo se estudiará el análisis exploratorio de datos y en el capítulo 3 se analizarán funciones de la muestra aleatoria y sus características con la finalidad de usarlas posteriormente en el proceso de estimación de parámetros.

Se ha planteado ya que la recolección de datos es un proceso que requiere planeación. Una vez que los datos están disponibles, su complejidad puede variar tanto en el número, como en el tipo de variables que se miden o registran. Un procedimiento de análisis comienza con la extracción de resúmenes útiles; así, cualquier procedimiento inductivo acerca de una población en estudio debe confirmar las hipótesis planteadas mediante el análisis exploratorio. Es por ello que la estadística descriptiva es el inicio de cualquier procedimiento de análisis estadístico. En este capítulo se proporcionan los elementos básicos para llevar a cabo un análisis exploratorio de datos.

2.1. Introducción

La *estadística descriptiva* se distingue de la estadística inferencial, en que la primera tiene como objetivo resumir un conjunto de datos en lugar de utilizarlos para hacer inferencias de la población que se cree que dichos datos representan. Esto generalmente significa que la estadística descriptiva, a diferencia de la estadística inferencial, no se desarrolla sobre la base de la teoría de la probabilidad. Su objetivo primario es entonces analizar un grupo determinado, resumiendo los datos de manera que se pueda enfatizar la información relevante que sea útil para el planteamiento de preguntas y modelos.

Este tema se puede abordar desde diferentes ángulos, pero considerando el objetivo del presente texto, que está más enfocado a la estadística inferencial, se hace una breve revisión de los conceptos más importantes de la estadística descriptiva. Se utilizan para ello datos que están disponibles en el paquete estadístico R, con el fin de ilustrar también el uso y resultados de este software. La razón principal para usar R es que este paquete es asequible por ser de

uso libre y que, además de la popularidad que ha adquirido en los últimos años en el ámbito educativo, se utiliza ampliamente en otros sectores.

También se exhiben los términos más importantes del análisis exploratorio de datos.

2.2. Datos y variables

Una premisa básica en el análisis estadístico es que el origen del conocimiento estadístico está en los datos. Cualquier procedimiento inductivo acerca de una población en estudio debe confirmar la fuerza de las diferencias encontradas mediante un procedimiento exploratorio. Es por ello que la estadística descriptiva debe anteceder a cualquier procedimiento inferencial.

Una *muestra* es una parte de la población obtenida y se entiende por *datos* a las mediciones u observaciones recolectadas de una muestra. Los *individuos* son las personas, animales o cosas descritos en un conjunto de datos; a veces se les llaman unidades experimentales. Una *variable* es cualquier característica de un individuo. Las variables pueden tomar distintos valores para distintos individuos.

Por ejemplo, los datos para un estudio del nivel socioeconómico y cultural de los estudiantes de la Facultad de Ciencias tienen que hacer referencia a todos los alumnos. Estos son los individuos descritos por el conjunto de datos. Para cada individuo, los datos contienen los valores de variables como la edad en años, el sexo (hombre o mujer), situación laboral (trabaja o no), tipo de horario (matutino, vespertino, mixto), ingreso mensual del principal sostén económico de su familia, tipo de entretenimientos (cine, TV, teatro, otro), práctica de deportes (sí, no), deporte preferido, entre otros.

No existe una única clasificación con respecto a la naturaleza de los datos. A continuación se darán dos clasificaciones de los datos considerando dos criterios: el tipo de escala en la que se miden los datos y la manera en la que se recolectó la información.

2.2.1. Tipos de variables

De acuerdo a su tipo, se puede clasificar a las variables en *cualitativas* y *cuantitativas*.

Los datos cuantitativos son aquellos en los que las observaciones se miden en una escala numérica. Los datos no-numéricos que sólo se pueden clasificar por categorías se conocen como datos cualitativos ó datos categóricos.

El número de hijos que tienen la familias en cierto poblado, el sueldo que perciben los trabajadores de cierta empresa son datos cuantitativos. Sin embargo, el tipo de sangre (O, A, B, AB) de los pacientes de cierto hospital y el estado civil de las personas (soltero, casado, divorciado) son datos categóricos.

2.2.2. Escalas de medición

Para considerar la precisión con la que se evalúan los diferentes valores, existen cuatro niveles de medición, los cuales se usarán de acuerdo al tipo de variable que se esté usando.

- Dato categórico en **escala nominal**. Como su nombre lo dice, este tipo de variables hace referencia a “nombres”. Básicamente es una clasificación de los datos con respecto a cierto criterio. En esta escala no hay ningún tipo de orden entre los datos. Son simples identificadores y son completamente arbitrarios. La única operación que se puede realizar con este tipo de escala es el de conteo de los datos con el mismo identificador.
- Dato categórico en **escala ordinal**. Desde un punto de vista formal no tienen un orden; sin embargo, se le puede asignar uno que haga sentido ya sea “creciente” ó “decreciente”. Por ejemplo, pequeño, mediano o grande; o mucho, poco o nada. Los valores de esta escala representan categorías con cierto orden asociado pero no en una cantidad específica, es decir, no se puede determinar la distancia entre las categorías, sólo es interpretable el orden entre sus valores. Se pueden hacer operaciones de igualdad y “orden de magnitud”. Generalmente representan una cualidad que se está “midiendo” y establecen si una observación tiene más de una cualidad que otra.
- Dato cuantitativo en **escala de intervalo**. Mediante esta escala se puede medir exactamente la intensidad con la que se posee una característica. Para ello se usa una unidad de medición cuyo origen se denomina “cero flotante” para enfatizar que cuando dicho origen se alcanza no implica la ausencia del atributo. Esta escala representa magnitudes, con la propiedad de igualdad de la distancia entre puntos de escala de la misma amplitud. Aquí se puede determinar el orden (formal) entre sus valores, hacerse comparaciones de igualdad, y medir la distancia existente entre cada valor de la escala. La igual distancia entre puntos de la escala significa que puede saberse cuántas unidades de más tiene una observación comparada con otra, con respecto a cierta característica analizada. El ejemplo por excelencia de este tipo de escala es cualquier escala de temperatura, ya que el cero en ellas no significa que no haya temperatura.
- Dato cuantitativo en **escala de razón**. Se dice que esta escala es la más completa. Tiene las mismas propiedades que la escala de intervalo, pero sí existe la noción de cero “absoluto”, ya que se sitúa un punto cero fijo que al ser alcanzado implica ausencia de un atributo. En este caso, es posible realizar operaciones aritméticas de producto y cociente, y así comparar mediante proporciones o razones. A iguales diferencias entre los números asignados corresponden iguales diferencias en el grado de atributo estudiado. Ejemplos de datos en esta escala son longitud, peso, distancia, ingresos, precios, etc.

Ahora, suponga que se va a realizar un estudio médico y para ello se dispone de un conjunto de variables referentes a un grupo de pacientes. Las variables son:

- (a) Género (hombre o mujer).
- (b) Edad (en años).
- (c) Raza (asiática, blanca, negra u otras).
- (d) Fumador (sí o no).
- (e) Presión sanguínea sistólica (milímetros de mercurio).
- (f) Nivel de calcio en la sangre (microgramos por mililitro).
- (g) Practica algún deporte o Actividad Deportiva (1 a 4 días de la semana, 5 a 7 días de la semana, ningún día de la semana).

Como variables categóricas se distinguen a Género (en escala nominal), Raza (en escala nominal), Fumador (en escala nominal) y Actividad Deportiva (en escala ordinal). Como variables cuantitativas se tienen a Edad (en escala de razón), Presión Sanguínea (en escala de intervalo) y Nivel de Calcio (en escala de razón).

2.2.3. Formas de recolección

También se puede clasificar a los datos con respecto a la manera en la que se recolectaron.

- Los datos en sección cruzada son datos que se recolectan de diferentes elementos ó variables en el mismo punto del tiempo (ó en el mismo intervalo de tiempo).
- Los datos en serie de tiempo son datos que se recolectan de un mismo elemento ó variable en diferentes puntos en el tiempo (ó para diferentes periodos de tiempo).

2.3. Los datos y R

R es un lenguaje y entorno de programación para análisis estadístico y gráfico. Se trata de un proyecto de software libre, resultado de la implementación GNU del lenguaje S. R y S-Plus -versión comercial de S- son, probablemente, los dos lenguajes más utilizados en investigación por la comunidad estadística, siendo además muy populares en el campo de la investigación biomédica, la bioinformática y las matemáticas financieras. R se puede descargar gratuitamente en la página oficial del proyecto <http://www.r-project.org/>.

Para ilustrar los conceptos fundamentales de la estadística descriptiva, se considerará un conjunto de datos, `crabs` que se encuentre en la librería de R `MASS`. Los datos corresponden a

un marco de datos (*data frame*) de 200 renglones y 8 columnas, describiendo 5 medidas morfológicas de 50 cangrejos cada uno de dos colores y de ambos sexos, de la especie *Leptograpsus* recolectada en Fremantle, W. Australia¹.

2.3.1. Cargando los datos

Se cargarán los datos desde *R* de la siguiente manera:

```
> library(MASS)
> data(crabs)
> help(crabs)
> attach(crabs)
```

Los datos cargados están en el formato `data.frame` que es un formato que está compuesto de varios campos. La ventaja de este tipo de formato, es que se pueden agrupar variables de varios formatos en una sola. Para saber los campos que contiene el `data.frame` se puede utilizar la instrucción `ls()` de la siguiente manera:

```
> ls(crabs)
[1] "BD"    "CL"    "CW"    "FL"    "index" "RW"    "sex"    "sp"
```

Este conjunto de datos contiene las siguientes columnas:

- `sp` especie - “B” o “O” para blue u orange.
- `sex` sexo del cangrejo.
- `index` índice 1 a 50, dentro de cada unos de los cuatro grupos.
- `FL` tamaño del lóbulo frontal (mm).
- `RW` ancho trasero (mm).
- `CL` longitud del caparazón (mm).
- `CW` ancho del caparazón (mm).
- `BD` profundidad del cuerpo (mm).

En este punto se tienen las variables listas para realizar el análisis descriptivo.

¹Campbell, N.A. and Mahon, R.J. (1974). A multivariate study of variation in two species of rock crab of genus *Leptograpsus*. *Australian Journal of Zoology* 22, 417-425.

2.4. Representación gráfica de los datos

Una vez que se obtiene una muestra, para entender con más claridad el tipo de información que se está analizando, es muy común representar dicha información a través de tablas ó gráficas. Estas representaciones gráficas son muy útiles ya que dan un recurso visual que muchas veces facilita el análisis de la información al mostrar algunos patrones con respecto al comportamiento de las variables que se están estudiando. En un primer curso de estadística se estudian datos asociados a una sola variable. Las representaciones gráficas más comunes son:

- Tabla de frecuencias
- Gráfica de pastel
- Gráfica de barras
- Histograma

La información categórica generalmente se resume en tablas o gráficamente con gráficas de barras, diagramas de puntos y gráficas de pastel.

2.4.1. Frecuencia, frecuencia relativa y frecuencia acumulada

Suponga que se tienen los n elementos de una muestra de tamaño n , obtenida de una población, $x_1, x_2, x_3, \dots, x_k$, con k valores asociados a los eventos observados al realizar el experimento aleatorio que define la muestra, y que de ésta hay f_i resultados idénticos a x_i ($i \in \{1, \dots, k\}$).

Los números $f_1, f_2, f_3, \dots, f_k$ se conocen como frecuencias de ocurrencia de los valores $x_1, x_2, x_3, \dots, x_k$, respectivamente; y satisfacen

$$f_1 + \dots + f_k = n.$$

Al cociente de una frecuencia f_i entre el total de observaciones n (el tamaño de la muestra), se le conoce como frecuencia relativa de ocurrencia del valor x_i correspondiente. Representando la frecuencia relativa de x_i con f_i^* , se tiene que

$$f_i^* = \frac{f_i}{n}, \quad i \in \{1, \dots, k\}.$$

A partir la definición de frecuencia relativa se obtienen de inmediato las condiciones para que un conjunto de números sean frecuencias relativas de los valores de una muestra. Éstas son

$$0 \leq f_1^* \leq \dots \leq f_k^* \leq 1 \quad \text{y} \quad f_1^* + \dots + f_k^* = 1.$$

Se conoce como frecuencia relativa acumulada de un valor x_i , a la suma de frecuencias relativas de todos los valores anteriores o iguales al valor x_i . Si F_i representa la frecuencia relativa acumulada de x_i entonces

$$F_i = f_1^* + \dots + f_i^*.$$

2.4.2. Tablas de frecuencia

Las tablas son muy usadas para resumir información. En R la función para crear tablas es la función `table()`. En su uso más simple, `table(x)` encuentra todos los valores únicos en el vector `x` y tabula las frecuencias y su ocurrencia.

La variable sexo (`sex`) puede ser tabulada como

```
> table(sex)
sex
  F   M
100 100
```

Si se quisieran tabular conjuntamente las variables sexo (`sex`) y la especie del cangrejo (`sp`), se haría de la siguiente manera:

```
> table(sex,sp)
      sp
sex  B  O
  F 50 50
  M 50 50
```

2.4.3. Gráficas de barras

Los datos categóricos también se pueden resumir de manera gráfica. La representación más común quizá sea la gráfica de barras. Una gráfica de barras es un arreglo simple que acomoda los niveles de la variable en algún orden y representa su frecuencia con una barra.

Una gráfica de barras es una gráfica en la que las “barras” representan las frecuencias (ó frecuencias relativas) de las diferentes categorías.

En R, las gráficas de barras se hacen con la función `barplot()`. Esta usa un resumen de la información, frecuentemente el que hace la función `table()`. Los datos resumidos pueden estar en frecuencias o en proporciones. El resultado graficamente será el mismo, sin embargo el escalado del eje Y será distinto.

```
> barplot(table(sp),xlab="Especie",ylab="Frecuencia")
```

A continuación se presenta la gráfica resultante de esta instrucción en la Figura 2.1.

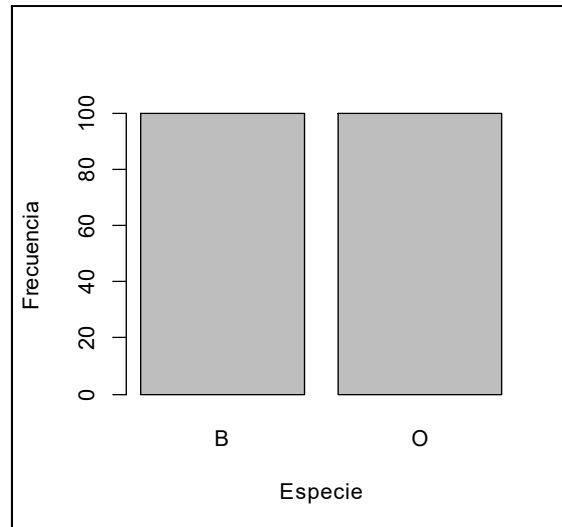


Figura 2.1: Gráfica de barras correspondiente a las especies B (blue) u O (orange) de los cangrejos.

2.4.4. Gráficas de pay o pastel

La gráfica de pay o pastel se utiliza para representar las frecuencias relativas o proporciones de las distintas posibles respuestas de una variable categórica. Esta gráfica, así como la gráfica de barras es ampliamente utilizada en los medios de difusión de información.

Para hacer una gráfica de pay en R se utiliza la función `pie()`, que utiliza argumentos similares al histograma para cambiar colores y agregar nombres:

```
> pie(table(sp),radius = 0.68, cex=1.1,
      col = c("cornflowerblue","coral4"),
      labels=NA, main="Especies")
> text(0.1,-0.3,"50%", cex=1.2)
> text(0.1,0.3,"50%", cex=1.2)
> legend(-1.1,-0.56,c("Orange", "Blue"),
      fill=c("cornflowerblue","coral4"),cex=0.8)
```

El resultado se puede ver en la Figura 2.2.

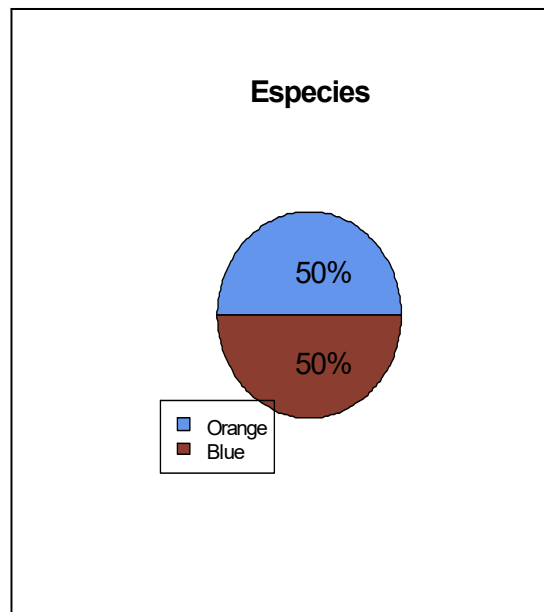


Figura 2.2: Gráfica de pay para las especies de cangrejos.

Para ejemplificar el caso de más categorías, considere una encuesta con 5 posibles respuestas A, B, C, D y E, codificadas, respectivamente como 1, 2, 3, 4 y 5. Las respuestas obtenidas fueron

4, 2, 3, 2, 1, 2, 3, 3, 3, 3, 4, 3, 5, 3, 3, 2, 2, 4, 3, 2

Para generar el diagrama de pay, se realiza lo siguiente:

```
> respuesta<-scan()
1: 4 2 3 2 1 2 3 3 3 3 4 3 5 3 3 2 2 4 3 2
21:
Read 20 items
```

La tabla de los resultados obtenidos en la encuesta es:

```
> (Tabla=table(respuesta))
respuesta
1 2 3 4 5
1 6 9 3 1
> names(Tabla) = c("A","B","C","D","E")
> pie(Tabla, main="Encuesta")
```

La gráfica resultante es

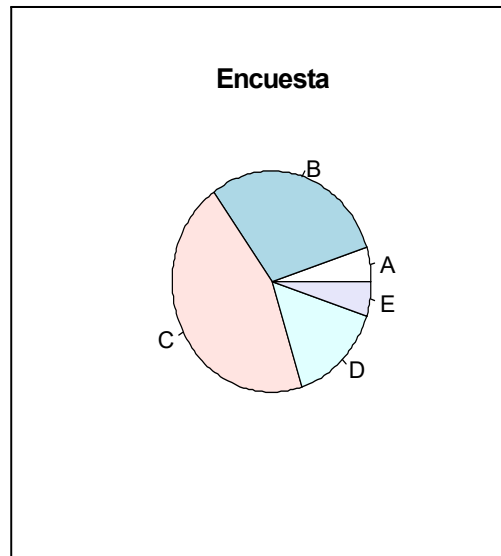


Figura 2.3: Diagrama de pay para las 5 respuestas de una encuesta.

2.4.5. Diagrama de tallo y hojas

El diagrama de tallo y hojas es una variante del histograma para desplegar la información muestral, especialmente cuando las observaciones tienen dos dígitos y no son conjuntos muy grandes de datos. Algo conveniente de este diagrama es el hecho de que se conserva la información de las observaciones en lugar de solo agruparlas en intervalos. Esta manera de desplegar los datos se obtiene al ordenar las observaciones de acuerdo a su dígito principal. Para ilustrar mejor este diagrama, antes de hacerlo en R se construirá el siguiente ejemplo. Supóngase que se tienen las siguientes observaciones:

75	98	42	75	84	87	65	59	63	86	78	37	99	66	90	79	80	89
68	57	95	55	79	88	76	60	77	49	92	83	71	78	53	81	77	58
93	85	70	62	80	74	69	90	62	84	64	73	48	72				

Ordenando los datos se tiene lo siguiente:

37	42	48	49	53	55	57	58	59	60	62	62	63	64	65	66	68	69
70	71	72	73	74	75	75	76	77	77	78	78	79	79	80	80	81	83
84	84	85	86	87	88	89	90	90	92	93	95	98	99				

Primero se listan los dígitos principales a la izquierda de la línea vertical. Después, para cada observación se anota el segundo dígito a la derecha de la línea vertical en el renglón de su dígito principal. Por último, se ordenan los dígitos de cada renglón y a la derecha de la línea para que estén en orden ascendente. Así, el diagrama queda de la siguiente manera:

```
> stem(x)
```

```

3 | 7
4 | 289
5 | 35789
6 | 022345689
7 | 01234556778899
8 | 00134456789
9 | 0023589

```

2.4.6. Histograma

Un histograma es la representación visual de la distribución de un conjunto de datos. Es decir, se intenta tener una idea acerca de cómo se comportan pensando en una función de densidad empírica. El histograma tiene algunas similitudes con la gráfica de barras (ver la función `barplot()`), en el sentido que también utiliza barras para indicar una frecuencia, pero a diferencia del diagrama de barras, cada barra en el histograma representa la frecuencia de un intervalo sobre el rango de las observaciones que se tienen. Cuando se elabora un histograma, se toma una decisión acerca de cómo se va a dividir el rango de la muestra en intervalos y cuán altas se dibujarán las barras, dado que únicamente tienen que estar la proporción correcta. *R* tiene varios métodos para la selección de estos intervalos (Sturges, Scott y Freedman–Diaconis). Las dos maneras de establecer la altura de las barras es la frecuencia absoluta del intervalo o aquella que haga al área de la barra igual a la frecuencia relativa del intervalo. Bajo este último método, el área total de las barras sumará 1, lo cual es conveniente cuando se está pensando en ajustar el modelo de una distribución de probabilidad.

En la función `hist()`, la longitud de los intervalos está controlada por el parámetro `breaks=`. Este puede ser especificado por el nombre de un algoritmo que los genere, el número de intervalos deseados o la localización exacta de los extremos de los intervalos deseados (`breaks`).

La siguiente instrucción imprime gráficas: 2 renglones, 2 columnas:

```
> par(mfrow=c(2,2))
```

O, un renglón y 2 columnas:

```
> par(mfrow=c(1,2))
```

Para graficar los histogramas la instrucción de *R* es:

```

> hist(CW,breaks="Sturges",freq=TRUE,col="lightgray",main=
  "Ancho del caparazon",ylab="Frecuencia absoluta")
> hist(CW,breaks=30,probability=TRUE,col="gray",main="Ancho
  del caparazon",ylab="Frecuencia relativa")

```

Siendo el resultado la Figura 2.4:

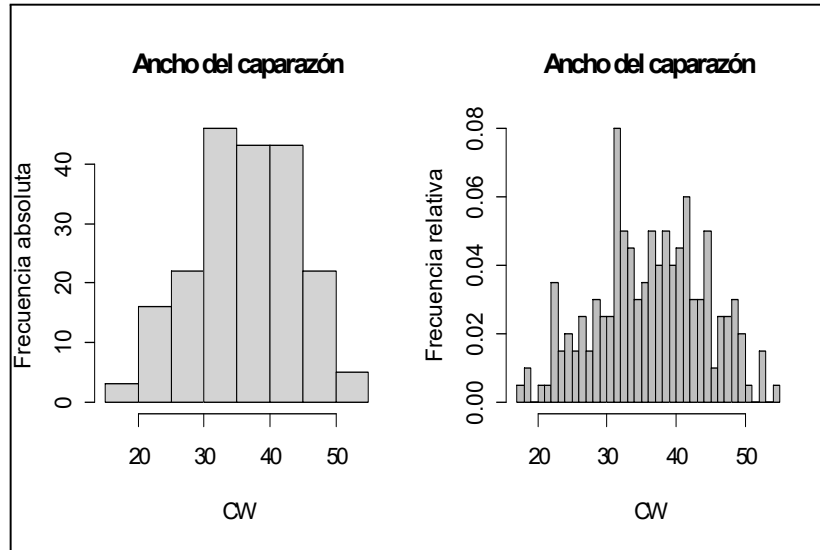


Figura 2.4: Histogramas para la característica *ancho del caparazón* de los cangrejos.

Puede notarse que los histogramas anteriores se graficaron de tal modo que muestran la frecuencia absoluta (`freq=TRUE` ó `probability=FALSE`) y la frecuencia relativa (`freq=FALSE` ó `probability=TRUE`) de los intervalos dados, respectivamente.

Procedimiento y consideraciones para la construcción de una tabla de frecuencias y un histograma

1. Determinar los valores mínimo y máximo de las observaciones y calcular la diferencia entre estos valores. A este número se le conoce como rango.
2. Seleccionar el número de clases M de tal forma que estas clases abarquen toda la información. Usualmente, una forma aproximada para elegir el número de clases es $M = \sqrt{n}$ ó $M = \log(n) + 1$, donde n es el número total de observaciones. La idea es utilizar suficientes clases para mostrar la variación de los datos pero no tantas como para que haya pocos datos en algunas de las clases. Una regla es que la longitud de las clases debe ser ligeramente mayor que el cociente $\frac{\max - \min}{M}$ donde M es el número de clases.
3. El primer intervalo debe tener extremo inferior ligeramente menor que el mínimo de los datos y el último intervalo debe tener extremo superior ligeramente mayor que el máximo de los datos. Los límites de la clase son los valores mínimo y máximo en cada clase. La marca de clase es el punto medio del intervalo de clase.

4. Ninguno de los datos debe estar en las fronteras de las clases.
5. Para una tabla de frecuencias se deben enlistar los intervalos de clase y escribir el número de datos en cada clase, f_i y también la frecuencia relativa $f_i^* = \frac{f_i}{n}$.
6. La base de cada barra será la longitud de la clase y la altura será la correspondiente frecuencia de dicha clase (es decir, el número de datos que pertenecen a dicha clase).

2.4.7. Boxplot o diagrama de caja y brazos

El boxplot o **diagrama de caja y brazos** es una manera de representar los datos de una muestra a través de la información de sus cuartiles. Estos diagramas tienen unas líneas que indican la variabilidad presente fuera del intervalo intercuartil. Es una gráfica que suministra información sobre los valores mínimo y máximo, los cuartiles Q_1 , Q_2 (o mediana) y Q_3 , y sobre la existencia de valores atípicos y la simetría de la distribución.

Los diagramas de caja muestran la variación de una muestra sin hacer suposiciones de la distribución probabilística de la cual provienen, es decir, tienen un enfoque no-paramétrico.

En R existe la instrucción `boxplot()` para dibujar este diagrama. En el caso de los datos que se han estado utilizando (`crabs`):

```
> boxplot(CW,ylab="Ancho del caparazon")
```

Resultando la gráfica que se muestra en la Figura 2.5:

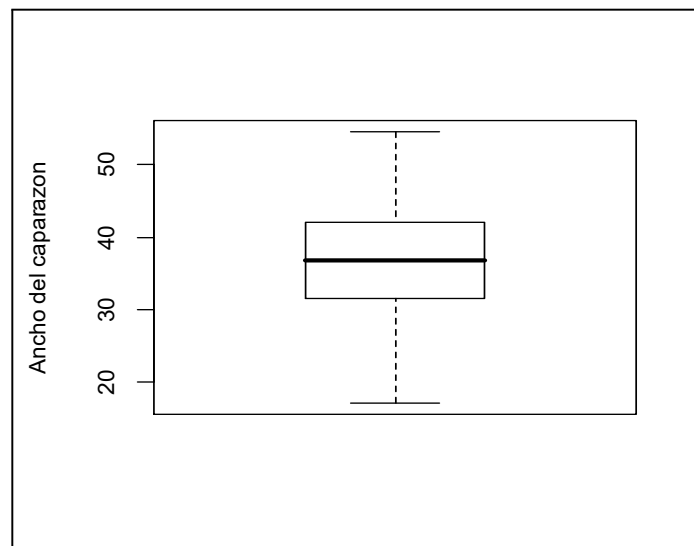


Figura 2.5: Diagrama de caja y brazos para la característica *ancho del caparazón* de la base de datos de los cangrejos *crabs*.

También se usan los diagramas de cajas y brazos cuando se busca comparar una variable cuantitativa con una variable cualitativa:

```
> par(mfrow=c(1,2))
> boxplot(CW~sp, xlab="Especie",ylab="Ancho del caparazon")
> boxplot(CW~sex, xlab="Sexo",ylab="Ancho del caparazon")
```

De estas instrucciones se obtiene la Figura 2.6.

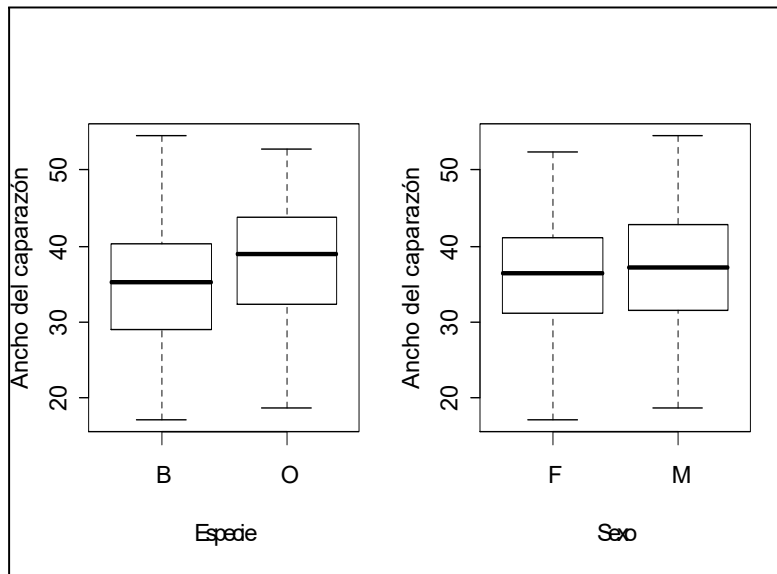


Figura 2.6: Diagramas de caja y brazos para la característica *ancho del caparazón* por *especie* y por *sexo* para los datos de los cangrejos.

Inclusive se puede comparar una variable cuantitativa con más de una variable cualitativa de manera simultánea:

```
> boxplot(CW~sp+sex, xlab="Especie y Sexo",ylab="Ancho del
  caparazon")
```

Resultando la Figura 2.7.

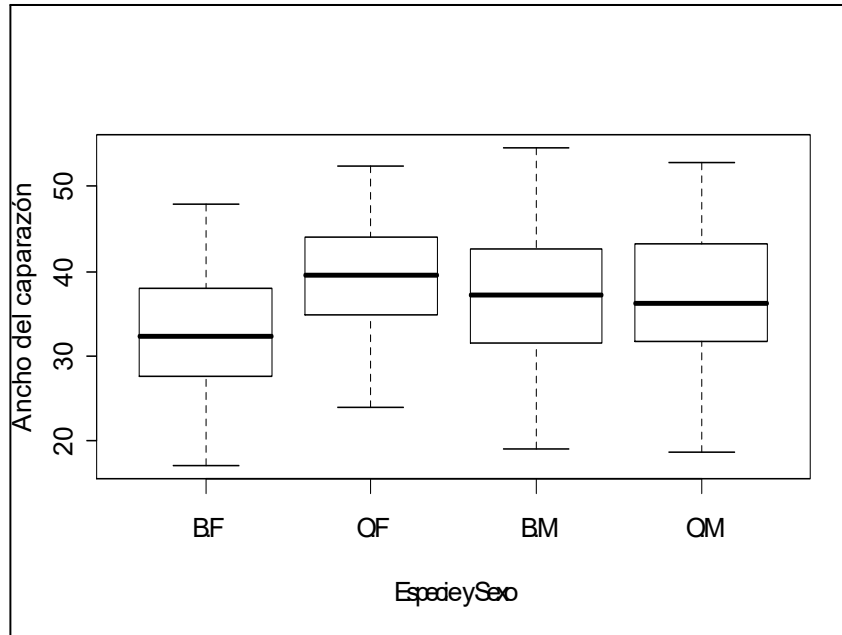


Figura 2.7: Gráfica de caja y brazos para la característica *ancho del caparazón* para las diferentes variables asociadas a *especie y sexo*.

Procedimiento para construir un diagrama de caja y brazos

1. Dibujar un eje de medida vertical y marcar Q_1 , Q_2 (la mediana) y Q_3 en este eje.
2. Construir una caja rectangular cuya base inferior es el cuantil Q_1 (primer cuartil) y su base superior es el cuantil Q_3 (tercer cuartil).
3. Dibujar una línea horizontal dentro de la caja a la altura de la mediana Q_2 .
4. Sea R_I el rango intercuartil, es decir, $R_I = Q_3 - Q_1$. Dibujar un segmento de recta del punto medio de la base inferior de la caja hacia abajo de longitud $1.5 \cdot R_I$.
5. Dibujar un segmento de recta del punto medio de la base superior de la caja hacia arriba de longitud $1.5 \cdot R_I$. A estos dos últimos segmentos se les conoce como “brazos”.
6. Marque en el diagrama con puntos aquellas observaciones que estén por encima y por debajo en una distancia de a lo más $1.5 \cdot R_I$. A estas observaciones se les conoce como observaciones atípicas moderadas.
7. Marque en el diagrama con asteriscos aquellas observaciones que estén por encima y por debajo en una distancia de al menos $1.5 \cdot R_I$. A estas observaciones se les conoce como observaciones atípicas extremas.

2.5. Medidas muestrales de tendencia central

Los métodos gráficos vistos en la parte anterior ayudan a visualizar los patrones de un conjunto de observaciones. Para obtener un resumen más objetivo, el siguiente paso será obtener valores numéricos para saber dónde están centrados los datos y la variabilidad presente en ellos. Las dos medidas de tendencia central más comúnmente utilizadas son la media y la mediana.

2.5.1. La media muestral

La **media muestral** de un conjunto de n observaciones x_1, x_2, \dots, x_n es la suma de estas observaciones divididas entre n . La media muestral se denota como \bar{x} . Es decir,

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}.$$

Para calcular la media muestral del ancho del caparazón en R , se puede hacer de la siguiente manera

```
> (sumaCW=sum(CW))
[1] 7282.9
> nCW<-length(CW)
> (mediaCW<-sumaCW/nCW)
[1] 36.4145
```

Otra manera es utilizar la función `mean()` que calcula la media muestral de un conjunto de datos

```
> mean(CW)
[1] 36.4145
```

2.5.2. La mediana muestral

Otra medida de tendencia central muestral utilizada es la mediana definida de la siguiente manera.

La **mediana muestral** de un conjunto de n observaciones x_1, x_2, \dots, x_n , es el valor m_c definido en los siguiente dos casos:

- Si n es impar, m_c es el valor que ocupa la posición $\frac{n+1}{2}$ una vez que los datos han sido ordenados, es decir, $m_c = x_{(n+1)/2}$.

- Si n es par, la mediana es la media aritmética de los dos valores centrales. Cuando n es par, los datos que están en el centro de la muestra ocupan las posiciones $\frac{n}{2}$ y $\frac{n}{2} + 1$, así, $m_c = \frac{x_{(n/2)} + x_{((n/2)+1)}}{2}$.

Para calcular la mediana muestral en R , se puede construir una función, de acuerdo a la definición anterior, como se especifica a continuación:

```
mediana<-function(x){
  n<-length(x)
  x<-sort(x)          # sort() ordena el conjunto de datos
  if(n%%2==0){       # n%%2 es n modulo 2
    med<-(x[n/2]+x[1+(n/2)])/2
  }else{
    med<-x[ceiling(n/2)] # ceiling() es la funcion techo
  }
  return(med)
}
```

Aplicando esta función, se tiene lo siguiente:

```
> mediana(CW)
[1] 36.8
```

Existe una función predefinida en R para el cálculo de la mediana muestral, que es `median()`:

```
> median(CW)
[1] 36.8
```

2.5.3. Percentiles

Si el tamaño de la muestra es razonablemente grande, es útil muchas veces extender el concepto de la mediana y dividir los datos ordenados en cuartos. Así, el punto que divide a la muestra en dos partes es la mediana y los puntos que dividen a la muestra en cuartos son llamados cuartiles. De manera aún más general se tiene la siguiente definición.

El $100 \times p$ **percentil** es el valor tal que, una vez que se han ordenado los datos de menor a mayor, al menos una proporción del $100 \times p\%$ de las observaciones son menores o iguales a este valor.

Existen varias maneras equivalentes de calcular los percentiles, aquí se proporciona una de ellas:

1. Se ordena la muestra de menor a mayor.

2. Se calcula el producto (tamaño de la muestra) \times (proporción) = np .

Si np no es entero, el percentil $p \times 100$ será el valor $x_{(\lceil np \rceil)}$, donde $\lceil \cdot \rceil$ es la función techo.

Si np es entero, digamos, k , el percentil $p \times 100$ será el valor $\frac{x_{(k)} + x_{(k+1)}}{2}$.

Claramente, los cuartiles son los percentiles 25, 50 y 75.

Para el cálculo de los percentiles en R , existe la función `quantile()`, que recibe la muestra, el valor de p y el método `type=`, la definición dada aquí corresponde a `type=2`, así, para calcular el tercer cuartil `CW`, se hace de la siguiente manera:

```
> quantile(CW,0.75,type=2)
75%
42
```

Pueden darse diversos valores de p , agrupándolo como un vector, para calcular los cuartiles 1, 2 y 3 de ambas series se hará de la siguiente manera

```
> quantile(CW,c(0.25,0.5,0.75),type=2)
25% 50% 75%
31.5 36.8 42.0
```

2.6. Medidas de dispersión

Las medidas de dispersión, también llamadas medidas de variabilidad, muestran la variabilidad de una distribución, indicando por medio de un número, si las diferentes puntuaciones de una variable están muy alejadas de la media. Cuanto mayor sea ese valor, mayor será la variabilidad, cuanto menor sea, más homogénea será a la media. Así se sabe si todos los casos son parecidos o varían mucho entre ellos.

2.6.1. Varianza y desviación estándar muestral

Varianza muestral

Para calcular la variabilidad de una distribución respecto de su media, se calcula la media de las desviaciones de las puntuaciones respecto a la media aritmética. Pero la suma de las desviaciones es siempre cero, así que lo que usualmente se toma es la media de las desviaciones al cuadrado. Es decir, la varianza muestral para un conjunto de observaciones x_1, x_2, \dots, x_n está definida por

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Para hacer el cálculo en R , se puede hacer lo siguiente:

```
> difs.CW=CW-mediaCW          # Diferencias respecto a la media
> difs.cuad.CW=difs.CW^2      # Diferencias cuadradas
> sum.difs.cuad.CW=sum(difs.cuad.CW) # Suma
> var.CW=(sum.difs.cuad.CW)/(nCW-1) # Varianza muestral
> var.CW                       # Imprime el resultado
[1] 61.96768
```

Haciendo el cálculo de la varianza muestral para CM con menos variables auxiliares

```
> x=CW                          # Para hacer la notacion menos pesada
> var.CW=sum((x-mean(x))^2)/(length(x)-1) # Varianza muestral
> var.CW                          # Imprimir el resultado
[1] 61.96768
```

Ahora, utilizando la función `var()`, la cual calcula automáticamente la varianza muestral

```
> var(CW)
[1] 61.96768
```

Desviación estándar muestral

La varianza a veces no se interpreta claramente, ya que se mide en unidades cuadráticas. Para evitar ese problema se define otra medida de dispersión, que es la desviación típica, o desviación estándar, que se halla como la raíz cuadrada positiva de la varianza. La desviación típica informa sobre la dispersión de los datos respecto al valor de la media; cuanto mayor sea su valor, más dispersos estarán los datos. Así, la desviación estándar muestral estará dada por

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Haciendo el cálculo en *R*. La función `sqrt()` calcula la raíz cuadrada del valor dado

```
> (desv.est.CW=sqrt(var.CW))
[1] 7.871955
```

2.6.2. Rango muestral

Se denomina rango estadístico o recorrido estadístico al intervalo entre el valor máximo y el valor mínimo, por ello comparte unidades con los datos. El rango muestral es el tamaño del intervalo más pequeño que contiene a todas las observaciones. Permite obtener una idea de la dispersión de los datos: cuanto mayor es el rango, más dispersos están los datos de un conjunto. Para un conjunto de observaciones $\{x_1, x_2, \dots, x_n\}$, el rango está definido por:

$$\text{Rango} = \max\{x_1, x_2, \dots, x_n\} - \min\{x_1, x_2, \dots, x_n\}.$$

Haciendo el cálculo en R ,

```
> (rango.CW<-max(CW)-min(CW))
[1] 37.5
```

2.6.3. Rango intercuartil

El rango intercuartil es la diferencia entre el tercer y el primer cuartil de una muestra. Es una medida de la dispersión estadística; a diferencia del rango, no se ve afectada de la misma manera por datos atípicos. Sea Q_1 y Q_3 los cuartiles 1 y 3 respectivamente de una muestra $\{x_1, x_2, \dots, x_n\}$, entonces el rango intercuartil estará dado por:

$$\text{rango intercuartil} = Q_3 - Q_1.$$

Haciendo el cálculo en R :

```
> CW.Q1.Q3<-quantile(CW,c(0.25,0.75),type=2)
> (CW.R.I<-diff(CW.Q1.Q3))
10.5
```

Resumen de la información y comparativo

Se hace ahora un resumen de los resultados obtenidos para el ancho del caparazón de los cangrejos (CW) estudiado. Las estadísticas muestrales son:

		CW
Tendencia central	mín	17.1
	Q_1	31.5
	\bar{x}	36.4145
	m_e	36.8
	Q_3	42.0
	máx	54.6
Dispersión	Rango muestral	37.5
	Rango intercuartil	10.5
	Desv. Estándar	7.871955
	Varianza	61.96768

2.7. Otras medidas de resumen

Curtosis

La curtosis se define como:

$$\frac{\mu_4}{\sigma^4},$$

donde μ_4 es el cuarto momento alrededor de la media y σ es la desviación estándar.

En ocasiones se mide con referencia a la distribución normal que tiene una curtosis igual a 3,

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3.$$

La distribución normal tiene entonces una medida $\gamma_2 = 0$ y se llama *mesocúrtica*. Las distribuciones con una curtosis positiva se llaman *leptocúrticas* y son muy picudas y con colas pesadas. Las distribuciones con curtosis negativa se llaman *platicúrticas* y tienen picos menores o son aplanadas y con colas ligeras.

La curtosis muestral se define como:

$$K = \frac{m_4}{m_2^2} - 3 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^2} - 3.$$

Sesgo

- Se habla de *sesgo negativo o hacia la izquierda* si la cola izquierda es más larga, es decir, si la distribución está concentrada a la derecha.
- Se habla de *sesgo positivo o hacia la derecha* si la cola derecha es más larga, es decir, si la distribución está concentrada a la izquierda.

Se define el sesgo como:

$$Sesgo = \frac{\mu_3}{\sigma^3}.$$

En este caso, μ_3 es el tercer momento alrededor de la media.

2.8. Relaciones lineales entre variables

Considere las variables aleatorias continuas para la base `crabs`; si observamos una gráfica de dispersión para ellas notamos que existe una relación lineal entre los pares de variables.

```
> pairs(crabs[,4:8])
```

El resultado de esta instrucción se presenta en la gráfica de dispersión de la Figura 2.8:

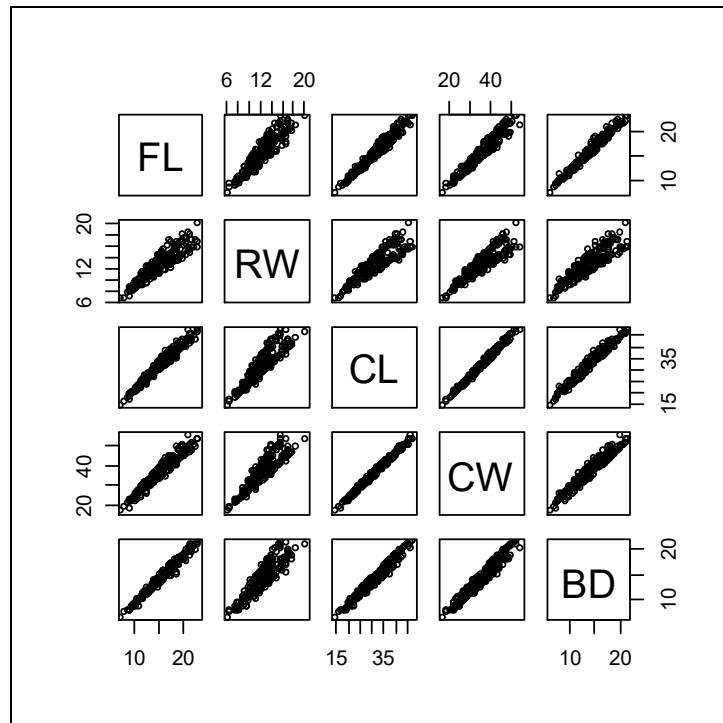


Figura 2.8: Gráfica de dispersión para la base *crabs* y que ilustra la relación entre las diferentes medidas morfológicas para los cangrejos.

Coefficiente de correlación de Pearson

Para un par de variables aleatorias continuas X y Y , se define el coeficiente de correlación de Pearson como:

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{(\sigma_X^2 \sigma_Y^2)}}.$$

Note que $-1 \leq \rho_{XY} \leq 1$, así valores cercanos a -1 y 1 indican una fuerte relación lineal con pendiente negativa y positiva, respectivamente.

Para una muestra aleatoria de variables (x_i, y_i) , $i = 1, 2, \dots, n$; se define el coeficiente de correlación muestral como:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2] [\sum_{i=1}^n (y_i - \bar{y})^2]}}.$$

La información muestral para las variables continuas en los datos *crabs*, puede resumirse en la gráfica de la Figura 2.9, obtenida con la instrucción:

```
> corrplot.mixed(cor(crabs[,4:8]), lower="number", upper="color")
```

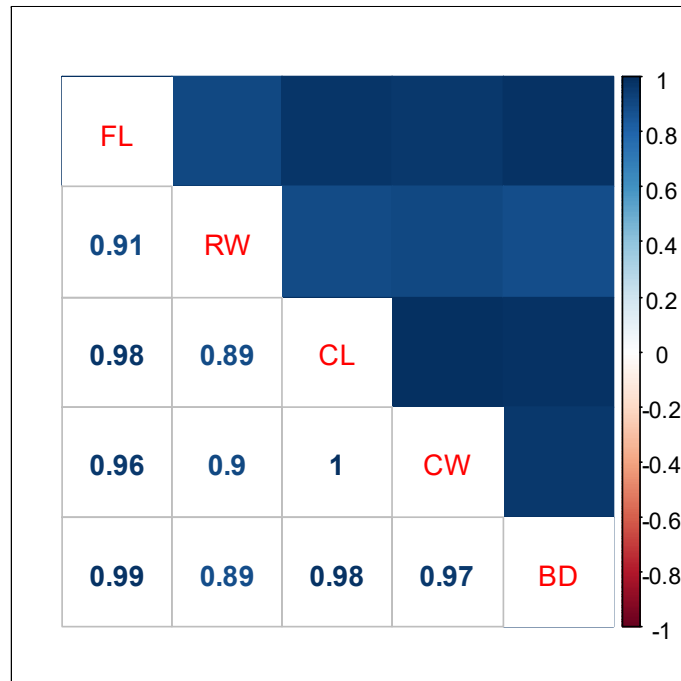


Figura 2.9: Gráfica de correlación para las medidas morfológicas de los cangrejos.

Note que, como se ha observado en el diagrama o gráfica de dispersión, la asociación lineal es fuerte entre las variables.

2.9. Ejercicio

Realice un análisis exploratorio para las variables incluidos en la base **Cars93**, incluida el la librería MASS de R, la base considera autos seleccionados aleatoriamente de entre los autos de pasajeros disponibles en el mercado en 1993², listados por el *Consumer Reports issue* y el *PACE Buying Guide*.

²Lock, R. H. (1993). 1993 New Car Data. *Journal of Statistics Education* (1).

3.1. Introducción

De acuerdo a la Definición 1.2, una muestra aleatoria representa un conjunto de variables aleatorias X_1, \dots, X_n , las cuales son independientes e idénticamente distribuidas. En este capítulo se estudiarán las propiedades de funciones de la muestra aleatoria que serán fundamentales en el proceso de estimación.

Definición 3.1 *El espacio de las muestras o espacio muestral es el conjunto de valores que puede tomar la muestra aleatoria X_1, \dots, X_n y se denotará por \mathfrak{X} .*

Definición 3.2 *Una estadística es cualquier función $T(X_1, \dots, X_n)$ de la muestra aleatoria que no depende de parámetros desconocidos.*

Se denotará una estadística por $T(X_1, \dots, X_n)$ o simplemente $T(\underline{X})$, donde $\underline{X} = X_1, \dots, X_n$.

Observación 3.1 *Note que una estadística, al ser función de variables aleatorias, es a su vez una variable aleatoria. De manera formal, se está pidiendo implícitamente que la función $T : \mathbb{R}^n \rightarrow \mathbb{R}^k$ sea Borel medible (medible en la σ -álgebra \mathcal{B}^n), donde k es la dimensión de la estadística; no obstante, la condición de medibilidad suele omitirse debido a que las funciones de la muestra aleatoria que de manera habitual se utilizan, son claramente medibles. Pero la importancia de señalar la medibilidad en este momento es enfatizar el hecho de que las estadísticas son variables aleatorias, puesto que no es posible anticipar el valor que tomarán al obtener observaciones concretas. Lo anterior tiene la implicación de que será necesario considerar la función de densidad (o de distribución) de las estadísticas, lo que a*

su vez permitirá calcular sus características tales como la esperanza o la varianza. En este contexto, es común hablar de la distribución muestral de una estadística $T(\underline{X})$ al referirse a la distribución de la variable aleatoria $T(\underline{X})$.

Entre las estadísticas más comunes que se utilizarán en procedimientos inferenciales destacan la media muestral, la varianza muestral y las estadísticas de orden, las cuales se estudiarán con mayor detalle en este capítulo. A continuación se enlistan algunas de estas funciones de la muestra aleatoria con su respectiva notación.

- $T(\underline{X}) = \frac{\sum_{i=1}^n X_i}{n} := \bar{X}$ *media muestral.*
- $T(\underline{X}) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} := S^2$ *varianza muestral.*
- $T(\underline{X}) = \min\{X_1, X_2, \dots, X_n\} := X_{(1)}$ *mínima estadística de orden.*
- $T(\underline{X}) = \max\{X_1, X_2, \dots, X_n\} := X_{(n)}$ *máxima estadística de orden.*
- $T(\underline{X}) = \frac{1}{n} \sum_{i=1}^n X_i^r := M_r$ *r-ésimo momento muestral.*

A continuación se probará que, en general, la esperanza de la media muestral es igual a la media poblacional, mientras que la varianza de la media muestral es la varianza poblacional dividida entre el tamaño de la muestra. Asimismo, la esperanza de la varianza muestral es igual a la varianza poblacional.

Proposición 3.1 Si X_1, \dots, X_n es una muestra aleatoria de $f(x; \theta)$, tal que $\mathbb{E}(X_i) = \mu$ y $Var(X_i) = \sigma^2$, para toda i , entonces

$$\mathbb{E}(\bar{X}) = \mu \quad y \quad Var(\bar{X}) = \frac{\sigma^2}{n}.$$

Demostración:

$$\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n} \cdot n\mu = \mu.$$

Ahora,

$$Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

■

A continuación se da una identidad que será de utilidad más adelante.

Observación 3.2 Si X_1, \dots, X_n es una muestra aleatoria de $f(x; \theta)$, entonces

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2. \quad (3.1)$$

Ahora se comprobará (3.1):

$$\begin{aligned} \sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n (X_i - \bar{X} + \bar{X} - \mu)^2 \\ &= \sum_{i=1}^n [(X_i - \bar{X})^2 + 2(X_i - \bar{X})(\bar{X} - \mu) + (\bar{X} - \mu)^2] \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \bar{X}) + n(\bar{X} - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2. \end{aligned}$$

■

En el siguiente resultado se probará que, en general, $\mathbb{E}(S^2) = \sigma^2$.

Proposición 3.2 Si X_1, \dots, X_n es una muestra aleatoria de $f(x; \theta)$, tal que $\mathbb{E}(X_i) = \mu$ y $\text{Var}(X_i) = \sigma^2$, para toda i , entonces

$$\mathbb{E}(S^2) = \sigma^2.$$

Demostración:

$$\begin{aligned} \mathbb{E}(S^2) &= \frac{1}{n-1} \mathbb{E} \left[\sum_{i=1}^n (X_i - \bar{X})^2 \right] \\ &= \frac{1}{n-1} \mathbb{E} \left[\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \right] \\ &= \frac{1}{n-1} \left\{ \sum_{i=1}^n \mathbb{E} [(X_i - \mu)^2] - n \mathbb{E} [(\bar{X} - \mu)^2] \right\} \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n \text{Var}(X_i) - n \cdot \text{Var}(\bar{X}) \right] \\ &= \frac{1}{n-1} \left(n\sigma^2 - \frac{n\sigma^2}{n} \right) \\ &= \sigma^2. \end{aligned}$$

■

Ejemplo 3.1 Se quiere conocer la probabilidad θ de obtener sol con una moneda, es decir, se está estudiando la variable aleatoria que toma valores $X = 1$ si se obtiene sol y $X = 0$ si se obtiene águila, donde $\theta = \mathbb{P}(X = 1)$, y para ello se realizarán tres lanzamientos de la moneda. En este caso, $f(x; \theta) = \theta^x(1 - \theta)^{1-x}$ con $x = 0, 1$; por lo tanto se está considerando una muestra aleatoria de tamaño tres de la distribución Bernoulli(θ). En la tabla siguiente se muestran los posibles resultados, la distribución de la muestra especificada a través de su función de densidad, así como los valores que pueden tomar las estadísticas \bar{X} y S^2 bajo los diferentes escenarios.

Resultados	x_1, x_2, x_3	Distribución	\bar{x}	s^2
s, s, s	$1, 1, 1$	θ^3	1	0
a, a, s	$0, 0, 1$	$\theta(1 - \theta)^2$	$1/3$	$1/3$
a, s, a	$0, 1, 0$	$\theta(1 - \theta)^2$	$1/3$	$1/3$
s, a, a	$1, 0, 0$	$\theta(1 - \theta)^2$	$1/3$	$1/3$
s, s, a	$1, 1, 0$	$\theta^2(1 - \theta)$	$2/3$	$1/3$
a, s, s	$0, 1, 1$	$\theta^2(1 - \theta)$	$2/3$	$1/3$
s, a, s	$1, 0, 1$	$\theta^2(1 - \theta)$	$2/3$	$1/3$
a, a, a	$0, 0, 0$	$(1 - \theta)^3$	0	0

Entonces, la función de densidad conjunta está dada por

$$\begin{aligned} f_{X_1, X_2, X_3}(x_1, x_2, x_3) &= \prod_{i=1}^3 f(x_i; \theta) = \prod_{i=1}^3 \theta^{x_i} (1 - \theta)^{1-x_i} \\ &= \theta^{\sum_{i=1}^3 x_i} (1 - \theta)^{3 - \sum_{i=1}^3 x_i}, \end{aligned}$$

y las funciones de densidad de la media y la varianza muestrales son:

$$f_{\bar{X}}(\bar{x}) = \begin{cases} (1 - \theta)^3 & \text{si } \bar{x} = 0; \\ 3\theta(1 - \theta)^2 & \text{si } \bar{x} = 1/3; \\ 3\theta^2(1 - \theta) & \text{si } \bar{x} = 2/3; \\ \theta^3 & \text{si } \bar{x} = 1. \end{cases}$$

$$f_{S^2}(s^2) = \begin{cases} \theta^3 + (1 - \theta)^3 & \text{si } s^2 = 0; \\ 3\theta(1 - \theta)^2 + 3\theta^2(1 - \theta) & \text{si } s^2 = 1/3. \end{cases}$$

Se obtendrán ahora las características numéricas de \bar{X} y S^2 para ilustrar las propiedades que

se presentaron previamente.

$$\begin{aligned}
 \mathbb{E}(\bar{X}) &= \theta^3(1) + 3\theta(1-\theta)^2\left(\frac{1}{3}\right) + 3\theta^2(1-\theta)\left(\frac{2}{3}\right) + (1-\theta)^3(0) \\
 &= \theta^3 + \theta(1-2\theta+\theta^2) + 2\theta^2(1-\theta) \\
 &= \theta^3 + \theta - 2\theta^2 + \theta^3 + 2\theta^2 - 2\theta^3 \\
 &= \theta.
 \end{aligned}$$

Por otro lado,

$$\begin{aligned}
 \mathbb{E}(\bar{X}^2) &= \theta^3(1) + 3\theta(1-\theta)^2\left(\frac{1}{3}\right)^2 + 3\theta^2(1-\theta)\left(\frac{2}{3}\right)^2 + (1-\theta)^3(0) \\
 &= \theta^3 + \frac{1}{3}\theta(1-2\theta+\theta^2) + \frac{4}{3}(\theta^2-\theta^3) \\
 &= \theta^3 + \frac{1}{3}\theta - \frac{2}{3}\theta^2 + \frac{1}{3}\theta^3 + \frac{4}{3}\theta^2 - \frac{4}{3}\theta^3 \\
 &= \frac{2}{3}\theta^2 + \frac{1}{3}\theta.
 \end{aligned}$$

Y

$$\begin{aligned}
 \text{Var}(\bar{X}) &= \frac{2}{3}\theta^2 + \frac{1}{3}\theta - \theta^2 \\
 &= -\frac{1}{3}\theta^2 + \frac{1}{3}\theta \\
 &= \frac{1}{3}\theta(1-\theta),
 \end{aligned}$$

es decir, la varianza poblacional dividida entre el tamaño de muestra.

Finalmente,

$$\begin{aligned}
 \mathbb{E}(S^2) &= \left(\frac{1}{3}\right) [3\theta(1-\theta)^2 + 3\theta^2(1-\theta)] \\
 &= \theta - 2\theta^2 + \theta^3 + \theta^2 - \theta^3 \\
 &= \theta - \theta^2 \\
 &= \theta(1-\theta).
 \end{aligned}$$

3.2. Distribución de las estadísticas muestrales bajo normalidad

Suponiendo ahora que las variables de la muestra aleatoria tienen una distribución normal, se analizará la distribución y las propiedades de la media (\bar{X}) y la varianza (S^2) muestrales.

3.2.1. Distribución de la media muestral

Proposición 3.3 Sea X_1, \dots, X_n una muestra aleatoria de una población con distribución $N(\mu, \sigma^2)$, entonces

$$\bar{X} \sim N(\mu, \sigma^2/n).$$

Demostración:

Se obtendrá la función generadora de momentos de \bar{X} . Para ello, se usará el hecho de que si $X_i \sim N(\mu, \sigma^2)$, la función generadora de momentos de X_i está dada por:

$$m_{X_i}(t) = \exp\left(t\mu + \frac{1}{2}t^2\sigma^2\right).$$

Así,

$$\begin{aligned} m_{\bar{X}}(t) &= \mathbb{E}(e^{t\bar{X}}) = \mathbb{E}\left(e^{t \cdot \frac{\sum_{i=1}^n X_i}{n}}\right) = \mathbb{E}\left(\prod_{i=1}^n e^{\frac{t}{n}X_i}\right) \\ &= \prod_{i=1}^n \mathbb{E}\left(e^{\frac{t}{n}X_i}\right) \quad \text{pues } X_1, \dots, X_n \text{ son independientes} \\ &= \prod_{i=1}^n m_{X_i}(t/n) \\ &= \prod_{i=1}^n e^{\frac{t}{n}\mu + \frac{1}{2} \cdot \frac{t^2}{n^2}\sigma^2} \quad \text{pues } X_1, \dots, X_n \text{ son idénticamente distribuidas} \\ &= e^{t\mu + \frac{1}{2}t^2 \frac{\sigma^2}{n}}, \end{aligned}$$

$$\therefore m_{\bar{X}}(t) = e^{t\mu + \frac{1}{2}t^2 \frac{\sigma^2}{n}},$$

de donde se concluye que $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. ■

3.2.2. La distribución de la varianza muestral

Enseguida se revisarán las distribuciones gamma y ji-cuadrada, las cuales serán de utilidad en los resultados que se analizarán más adelante.

La distribución Gamma

Se dice que la variable aleatoria continua X tiene distribución gamma con parámetros r y λ , si su función de densidad de probabilidad está dada por

$$f(x; r, \lambda) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, \quad x > 0,$$

donde $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx$ es conocida como la *función gamma*, la cual satisface:

$$\Gamma(r+1) = r\Gamma(r) \quad \text{y si } r \in \mathbb{N} \text{ entonces } \Gamma(r+1) = r!.$$

La notación $X \sim \text{Gamma}(r, \lambda)$, significa que X tiene esta distribución.

Las expresiones para su media, varianza y función generadora de momentos se resumen a continuación:

- $\mathbb{E}(X) = \frac{r}{\lambda}$,
- $\text{Var}(X) = \frac{r}{\lambda^2}$,
- $m_X(t) = \left(\frac{\lambda}{\lambda-t}\right)^r$.

La distribución Ji-cuadrada es un caso particular de la Gamma, de tal manera que las propiedades de la distribución Gamma pueden utilizarse para deducir las de la Ji-cuadrada, la cual se presenta a continuación.

Distribución Ji-cuadrada

Se dice que la variable aleatoria continua X tiene distribución Ji-cuadrada con k grados de libertad si función de densidad de probabilidad está dada por

$$f(x; k) = \frac{\left(\frac{1}{2}\right)^{k/2}}{\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{1}{2}x}, \quad x > 0.$$

Se denota como $X \sim \chi_{(k)}^2$. Además:

- $\mathbb{E}(X) = \frac{k/2}{1/2} = k$,
- $\text{Var}(X) = \frac{k/2}{1/4} = 2k$,
- $m_x(t) = \left(\frac{1/2}{1/2-t}\right)^{\frac{k}{2}} = \left(\frac{1}{1-2t}\right)^{k/2}$.

Estas propiedades pueden deducirse de las expresiones correspondientes para la distribución Gamma, notando que la distribución Ji-cuadrada es un caso particular con $r = \frac{k}{2}$ y $\lambda = \frac{1}{2}$.

El siguiente resultado establece una relación entre la distribución normal y la distribución Ji-cuadrada.

Teorema 3.4 Sean X_1, \dots, X_n variables aleatorias independientes tales que $\forall i \in \{1, \dots, n\}$, $X_i \sim N(\mu_i, \sigma_i^2)$. Sea $Z_i = \frac{X_i - \mu_i}{\sigma_i}$, entonces:

1. $Z_i^2 \sim \chi_{(1)}^2$.
2. $\sum_{i=1}^n Z_i^2 \sim \chi_{(n)}^2$.

Demostración:

1. Como $Z_i \sim N(0, 1)$, entonces

$$\begin{aligned}
 m_{Z_i^2}(t) &= \mathbb{E} \left(e^{tZ_i^2} \right) = \int_{-\infty}^{\infty} e^{tz_i^2} \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} dz_i \\
 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_i^2(1-2t)} dz_i \\
 &= \sqrt{\frac{1}{1-2t}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(\frac{1}{1-2t})}} e^{-\frac{1}{2}z_i^2(1-2t)} dz_i \\
 &= \sqrt{\frac{1}{1-2t}} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(\frac{1}{1-2t})}} e^{-\frac{1}{2} \cdot \frac{1}{1(1-2t)} z_i^2} dz_i}_1 \\
 &= \left(\frac{1}{1-2t} \right)^{1/2},
 \end{aligned}$$

la cual corresponde a la función generadora de momentos de una distribución $\chi_{(1)}^2$.

2. Para la segunda parte se utilizará el hecho de que la suma de variables aleatorias independientes con distribución Gamma tiene también distribución Gamma con parámetro de escala igual a la suma de los parámetros de escala de las variables que componen la suma; además de que la distribución Ji-cuadrada es un caso particular de la distribución Gamma. O bien, calculando directamente la función generadora de momentos de $\sum_{i=1}^n Z_i^2$:

$$m_{\sum_{i=1}^n Z_i^2}(t) = \prod_{i=1}^n m_{Z_i^2}(t) = \prod_{i=1}^n \left(\frac{1}{1-2t} \right)^{1/2} = \left(\frac{1}{1-2t} \right)^{n/2}.$$

■



Figura 3.1: Karl Pearson (1857-1936). Desarrolló la estadística Ji-cuadrada y estudió su distribución asintótica. Se le considera el padre de la Estadística. Imagen tomada de commons.wikipedia.org (public domain).

En el siguiente resultado se analizará la distribución de S^2 . La primera parte se refiere a la independencia entre \bar{X} y el vector de desviaciones $(X_1 - \bar{X}, \dots, X_n - \bar{X})$, la cual se justificará de dos formas: usando la función generadora de momentos y mediante una transformación. Otra manera de abordar este resultado es a través de la independencia entre una forma lineal y una forma cuadrática y las propiedades de la distribución normal multivariada (ver Rencher y Schaalje (2008)); sin embargo, en este texto no se estudiará ese enfoque.

Teorema 3.5 *Considere a X_1, \dots, X_n una muestra aleatoria de la distribución $N(\mu, \sigma^2)$. Entonces*

1. \bar{X} y el vector $(X_1 - \bar{X}, \dots, X_n - \bar{X})$ son independientes.
2. \bar{X} y S^2 son independientes.
3. $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{(n-1)}^2$.
4. $\mathbb{E}(S^2) = \sigma^2$ y $\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$.

Demostración:

1. Se considera la distribución conjunta de Y_1, Y_2, \dots, Y_n , donde:

$$Y_1 = \bar{X}, Y_2 = X_2 - \bar{X}, Y_3 = X_3 - \bar{X}, \dots, Y_n = X_n - \bar{X}.$$

Con este cambio de variable,

$$\begin{aligned} Y_1 - Y_2 - Y_3 - \dots - Y_n &= \bar{X} - (X_2 - \bar{X}) - \dots - (X_n - \bar{X}) \\ &= n\bar{X} - X_2 - \dots - X_n \\ &= X_1 + X_2 + \dots + X_n - X_2 - X_3 - \dots - X_n \\ &= X_1. \end{aligned}$$

Expresando las X_i 's en términos de las Y_i 's se obtiene:

$$X_1 = Y_1 - Y_2 - Y_3 - \dots - Y_n,$$

$$X_2 = Y_1 + Y_2,$$

$$X_3 = Y_1 + Y_3,$$

$$\vdots \quad \quad \quad \vdots$$

$$X_n = Y_1 + Y_n.$$

El Jacobiano de esta transformación es igual a n , pues la matriz $\left(\frac{\partial X_i}{\partial Y_j}\right)$ tiene la forma:

$$\begin{pmatrix} 1 & -1 & -1 & -1 & \dots & -1 \\ 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

la cual se puede transformar mediante operaciones elementales de renglones: para cada i , $2 \leq i \leq n$, se suma el renglón i -ésimo al primer renglón, obteniendo la matriz:

$$\begin{pmatrix} n & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

cuyo determinante es:

$$n \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{vmatrix} = n.$$

Por otro lado, de acuerdo a (3.1):

$$\begin{aligned} \sum_{i=1}^n (x_i - \mu)^2 &= \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2, \end{aligned}$$

la función de densidad conjunta de X_1, X_2, \dots, X_n (la muestra aleatoria de la distribución $N(\mu, \sigma^2)$) se puede escribir como:

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2\sigma^2} - \frac{n(\bar{x} - \mu)^2}{2\sigma^2} \right)},$$

con $-\infty < x_i < \infty$, $i = 1, 2, \dots, n$. Como $y_1 = \bar{x}$ y, por lo tanto, $x_1 - \bar{x} = -y_2 - y_3 - \dots - y_n$, la función de densidad conjunta de Y_1, Y_2, \dots, Y_n es:

$$(n) \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left[-\frac{(-y_2 - y_3 - \dots - y_n)^2}{2\sigma^2} - \frac{\sum_{i=2}^n y_i^2}{2\sigma^2} - \frac{n(y_1 - \mu)^2}{2\sigma^2} \right],$$

$-\infty < y_i < \infty$, $i = 1, 2, \dots, n$. El resultado obtenido es producto de la función de densidad de Y_1 , es decir, de

$$\frac{1}{\sqrt{2\pi\sigma^2/n}} \exp \left[-\frac{(y_1 - \mu)^2}{2\sigma^2/n} \right], \quad -\infty < y_1 < \infty,$$

y una función de y_2, \dots, y_n . De esta manera, Y_1 es independiente de las $n - 1$ variables aleatorias Y_2, Y_3, \dots, Y_n y la función de y_2, \dots, y_n es la función de densidad de Y_2, Y_3, \dots, Y_n . En conclusión, $Y_1 = \bar{X}$ es independiente de $-Y_2 - Y_3 - \dots - Y_n = X_1 - \bar{X}$, $Y_2 = X_2 - \bar{X}$, ..., $Y_n = X_n - \bar{X}$.

2. \bar{X} y S^2 son independientes porque $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$ es función de $(X_1 - \bar{X}, \dots, X_n - \bar{X})$. O bien, puede verificarse notando que la variable

$$W_1 = \frac{n(\bar{X} - \mu)^2}{\sigma^2} = \frac{n(Y_1 - \mu)^2}{\sigma^2}$$

es independiente de:

$$W_2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \frac{(-Y_2 - \dots - Y_n)^2 + \sum_{i=2}^n Y_i^2}{\sigma^2}.$$

3. Observe que

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}$$

Usando nuevamente la identidad

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$$

se tiene que

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2} \quad (3.2)$$

Pero note que $\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_{(n)}^2$ y también $\frac{n(\bar{X} - \mu)^2}{\sigma^2} = \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi_{(1)}^2$; y como \bar{X} es independiente de $X_1 - \bar{X}, \dots, X_n - \bar{X}$, entonces

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{(n-1)}^2,$$

debido a que bajo el supuesto de independencia y por (3.2), la función generadora de momentos de $\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}$ se puede escribir como el producto de las funciones generadoras de $W = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}$ y $\frac{n(\bar{X} - \mu)^2}{\sigma^2}$, es decir:

$$(1 - 2t)^{-n/2} = m_W(t) \times (1 - 2t)^{-1/2}, \quad t < 1/2$$

de donde,

$$m_W(t) = (1 - 2t)^{-n/2+1/2} = (1 - 2t)^{-(n-1)/2}.$$

$$\therefore \frac{(n-1)S^2}{\sigma^2} \sim \chi_{(n-1)}^2.$$

4. Anteriormente se probó que $\mathbb{E}(S^2) = \sigma^2$, sin embargo conociendo ya la distribución de $\frac{(n-1)S^2}{\sigma^2}$, se puede obtener de la siguiente manera, además de deducir la varianza correspondiente.

Como $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{(n-1)}^2$, entonces

$$\mathbb{E} \left[\frac{(n-1)S^2}{\sigma^2} \right] = n-1 \Rightarrow \frac{(n-1)}{\sigma^2} \mathbb{E}(S^2) = n-1 \Rightarrow \mathbb{E}(S^2) = \sigma^2.$$

También, como $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{(n-1)}^2$, se tiene que

$$\begin{aligned} \text{Var} \left(\frac{(n-1)S^2}{\sigma^2} \right) &= 2(n-1) \Rightarrow \frac{(n-1)^2}{\sigma^4} \text{Var}(S^2) = 2(n-1) \\ \Rightarrow \text{Var}(S^2) &= \frac{2(n-1)\sigma^4}{(n-1)^2} = \frac{2\sigma^4}{n-1}. \\ \therefore \text{Var}(S^2) &= \frac{2\sigma^4}{n-1}. \end{aligned}$$

■

Demostración alternativa para las partes 1, 2 y 3:

Se define la siguiente transformación:

$$\begin{aligned} Y_1 &= \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n), \\ Y_2 &= \frac{1}{\sqrt{2}}(X_1 - X_2), \\ Y_3 &= \frac{1}{\sqrt{6}}(X_1 + X_2 - 2X_3), \\ &\vdots \\ Y_n &= \frac{1}{\sqrt{n(n-1)}}(X_1 + X_2 + \cdots + X_{n-1} - (n-1)X_n). \end{aligned}$$

Para la cual, la matriz Jacobiana correspondiente es

$$J = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{-(n-1)}{\sqrt{n(n-1)}} \end{pmatrix}.$$

Esta matriz es una matriz ortogonal, esto significa que $JJ^t = J^tJ = I$, por lo tanto es posible escribir:

$$\begin{aligned} |\det(J)| &= |\det(J^t)| \\ &= |\det(JJ^t)^{1/2}| = 1. \end{aligned}$$

Ahora, la función de densidad conjunta de X_1, X_2, \dots, X_n es:

$$\begin{aligned} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu)^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i^2 - 2\mu x_i + \mu^2)\right\}, \end{aligned}$$

para $-\infty < x_1, x_2, \dots, x_n < \infty$. Entonces la función de densidad conjunta de Y_1, \dots, Y_n es:

$$\begin{aligned} g(y_1, y_2, \dots, y_n) &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n y_i^2 - 2\mu\sqrt{n}y_1 + n\mu^2\right)\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \left[\sum_{i=2}^n y_i^2 + (y_1 - \sqrt{n}\mu)^2\right]\right\}, \end{aligned}$$

donde $-\infty < y_1, y_2, \dots, y_n < \infty$. Además puede verse que $\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$, esto puede verificarse desarrollando para diferentes valores de n ; por ejemplo para $n = 2$:

$$\begin{aligned} Y_1^2 + Y_2^2 &= \frac{1}{2}(X_1 + X_2)^2 + \frac{1}{2}(X_1 - X_2)^2 \\ &= \frac{1}{2}X_1^2 + X_1X_2 + \frac{1}{2}X_2^2 + \frac{1}{2}X_1^2 - X_1X_2 + \frac{1}{2}X_2^2 \\ &= X_1^2 + X_2^2. \end{aligned}$$

Note que la función de densidad conjunta de Y_1, \dots, Y_n puede reescribirse como:

$$\left(\frac{1}{\sqrt{2\pi}\sigma}\right) \exp\left\{-\frac{1}{2\sigma^2}(y_1 - \sqrt{n}\mu)^2\right\} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n-1} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=2}^n y_i^2\right\},$$

donde se observa que Y_1 se distribuye como una $N(\sqrt{n}\mu, \sigma^2)$ y que Y_1, \dots, Y_n son variables aleatorias independientes con una distribución $N(0, \sigma^2)$ para Y_2, \dots, Y_n . De la transformación

definida, se puede escribir: $\bar{X} = n^{-1/2}Y_1$ y

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2 = \sum_{i=1}^n Y_i^2 - Y_1^2 = \sum_{i=2}^n Y_i^2,$$

concluyendo que \bar{X} es función de Y_1 únicamente y que $\sum_{i=1}^n (X_i - \bar{X})^2$ depende funcionalmente de Y_2, \dots, Y_n . Pero Y_1 y Y_2, \dots, Y_n son independientes, por lo que \bar{X} y $\sum_{i=1}^n (X_i - \bar{X})^2$ son independientes.

Como $\bar{X} = n^{-1/2}Y_1$ donde $Y_1 \sim N(\sqrt{n}\mu, \sigma^2)$ entonces la distribución de \bar{X} es $N(\mu, \frac{\sigma^2}{n})$. Ahora, $\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} = \sum_{i=2}^n \frac{Y_i^2}{\sigma^2}$, que es una suma de $(n-1)$ variables Ji-cuadradas independientes, por lo tanto tiene distribución χ_{n-1}^2 .

■

Ahora se presentarán dos distribuciones de probabilidad importantes en la inferencia estadística: la F de Fisher y la t de Student, así como las propiedades que serán de utilidad más adelante.

3.2.3. La distribución F de Fisher y el cociente de varianzas muestrales

Distribución F de Fisher

Se dice que una variable aleatoria X tiene la distribución **F de Fisher** con parámetros m y n (m y n grados de libertad), si su función de densidad está dada por:

$$f(x; m, n) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{m/2} \frac{x^{\frac{m-2}{2}}}{[1 + (\frac{m}{n})x]^{\frac{m+n}{2}}} \quad x > 0, \quad m, n \in \mathbb{N}^+.$$

Se denota como $X \sim F(m, n)$.



Figura 3.2: Ronald A. Fisher (1890-1962). *“Incluso los científicos necesitan sus héroes y Fisher fue sin duda el héroe de la estadística del siglo XX. Sus ideas transformaron nuestra disciplina de tal forma que hasta un César o un Alejandro hubieran envidiado”*: Efron (1998). *“Lo que es y fue importante para mí, es cómo Fisher, en los 1920..., hizo que los estadísticos reflexionáramos acerca de las ideas fundamentales”*: Egon Pearson (1974). Dos citas que hacen alusión a la importancia de las contribuciones de Fisher en la Estadística. La distribución que lleva su nombre, fue producto del trabajo que realizó con **George Snedecor** (1881-1974) de la Universidad de Iowa. Imagen tomada de commons.wikipedia.org (public domain).

Teorema 3.6 Si U y V son variables aleatorias independientes tales que $U \sim \chi_{(m)}^2$ y $V \sim \chi_{(n)}^2$, entonces

$$\frac{U/m}{V/n} \sim F(m, n).$$

Demostración:

Para probar este resultado se utilizará el teorema de cambio de variable. Sean

$$X = \frac{U/m}{V/n} = \frac{nU}{mV} \text{ y } Y = V.$$

La función de densidad conjunta de U y V está dada por

$$\begin{aligned}
f_{U,V}(u,v) = f_U(u)f_V(v) &= \frac{\left(\frac{1}{2}\right)^{m/2}}{\Gamma\left(\frac{m}{2}\right)} u^{\frac{m}{2}-1} e^{-\frac{1}{2}u} \frac{\left(\frac{1}{2}\right)^{n/2}}{\Gamma\left(\frac{n}{2}\right)} v^{\frac{n}{2}-1} e^{-\frac{1}{2}v} \\
&= \frac{\left(\frac{1}{2}\right)^{(m+n)/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} e^{-\frac{1}{2}(u+v)} u^{\frac{m}{2}-1} v^{\frac{n}{2}-1}.
\end{aligned}$$

Como $U = \frac{m}{n}XY$ y además $V = Y$, entonces el Jacobiano de la transformación está dado por

$$J = \begin{vmatrix} \frac{m}{n}y & \frac{m}{n}x \\ 0 & 1 \end{vmatrix} = \frac{m}{n}y.$$

La función de densidad conjunta de (X, Y) está determinada por

$$\begin{aligned}
f_{X,Y}(x,y) &= \frac{m}{n}y \cdot \frac{\left(\frac{1}{2}\right)^{\frac{m+n}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} e^{-\frac{1}{2}\left(\frac{m}{n}x+1\right)y} \left(\frac{m}{n}xy\right)^{\frac{m}{2}-1} y^{\frac{n}{2}-1} \\
&= \frac{\left(\frac{1}{2}\right)^{\frac{m+n}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} y^{\frac{m+n}{2}-1} x^{\frac{m}{2}-1} e^{-\frac{1}{2}\left(\frac{m}{n}x+1\right)y}.
\end{aligned}$$

Y como la densidad marginal de X está dada por

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy,$$

se tiene que:

$$\begin{aligned}
f_X(x) &= \frac{\left(\frac{1}{2}\right)^{\frac{m+n}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \int_0^{\infty} y^{\frac{m+n}{2}-1} e^{-\frac{1}{2}\left(\frac{m}{n}x+1\right)y} dy \\
&= \frac{\left(\frac{1}{2}\right)^{\frac{m+n}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\left[\frac{1}{2}\left(\frac{m}{n}x+1\right)\right]^{\frac{m+n}{2}}} \\
&= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \cdot \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m-2}{2}}}{\left(\frac{m}{n}x+1\right)^{\frac{m+n}{2}}},
\end{aligned}$$

que corresponde a la función de densidad de una variable aleatoria con distribución F , es decir, $X = \frac{U/m}{V/n} \sim F(m, n)$.

■

La distribución F en el marco del muestreo de la distribución normal

Sean X_1, X_2, \dots, X_{m+1} una muestra aleatoria de la distribución $N(\mu_x, \sigma_x^2)$ y Y_1, Y_2, \dots, Y_{n+1} una muestra aleatoria de la distribución $N(\mu_y, \sigma_y^2)$, de tal manera que ambas muestras son independientes entre sí.

En este caso $\bar{X} = \frac{1}{m+1} \sum_{i=1}^{m+1} X_i$, $\bar{Y} = \frac{1}{n+1} \sum_{j=1}^{n+1} Y_j$, $S_x^2 = \frac{1}{m} \sum_{i=1}^{m+1} (X_i - \bar{X})^2$ y $S_y^2 = \frac{1}{n} \sum_{j=1}^{n+1} (Y_j - \bar{Y})^2$.

Entonces,

$$\frac{mS_x^2}{\sigma_x^2} \sim \chi_{(m)}^2 \quad \text{y} \quad \frac{nS_y^2}{\sigma_y^2} \sim \chi_{(n)}^2.$$

Por el teorema anterior, se concluye que:

$$\frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} \sim F(m, n).$$

3.2.4. La distribución t de Student y algunas estadísticas relacionadas



Figura 3.3: William Sealy Gosset (1876-1937), quien usó el pseudónimo de *Student* para publicar su trabajo, desarrolló la distribución t como respuesta a problemas prácticos de variedades de cebada, trabajando en la cervecería Guinness. Imagen tomada de commons.wikimedia.org (public domain).

Distribución t de Student

Se dice que una variable aleatoria continua X tiene distribución t de Student con k grados de libertad, si su función de densidad está dada por

$$f(x; k) = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \frac{1}{\sqrt{k\pi}} \frac{1}{(1 + \frac{x^2}{k})^{\frac{k+1}{2}}}, \quad \text{con } k = 1, 2, \dots$$

Se denota como $X \sim t_{(k)}$.

Teorema 3.7 Si Z y U son variables aleatorias independientes tales que $Z \sim N(0, 1)$ y $U \sim \chi_{(k)}^2$, entonces $\frac{Z}{\sqrt{U/k}} \sim t_{(k)}$.

Demostración.

La prueba de este resultado es similar a la del Teorema 3.6, definiendo las variables

$$X = \frac{Z}{\sqrt{U/k}} \text{ y } Y = U.$$

■

La distribución t a partir de una muestra con distribución normal

Si X_1, \dots, X_n es una muestra aleatoria de la distribución normal $N(\mu, \sigma^2)$, se sabe que

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

de donde:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Por otra parte:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{(n-1)}^2.$$

Por el Teorema 3.7,

$$\frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}.$$

Ahora considere a X_1, \dots, X_m una muestra aleatoria de la distribución $N(\mu_X, \sigma^2)$ y a Y_1, \dots, Y_n una muestra aleatoria de la distribución $N(\mu_Y, \sigma^2)$, tales que X_i es independiente de Y_j , $i \in \{1, \dots, m\}$ y $j \in \{1, \dots, n\}$. Por los resultados vistos previamente, se puede concluir que

$$\bar{X} \sim N\left(\mu_X, \frac{\sigma^2}{m}\right) \quad \text{y} \quad \bar{Y} \sim N\left(\mu_Y, \frac{\sigma^2}{n}\right), \quad (3.3)$$

mientras que

$$\frac{(m-1)S_X^2}{\sigma^2} \sim \chi_{(m-1)}^2 \quad y \quad \frac{(n-1)S_Y^2}{\sigma^2} \sim \chi_{(n-1)}^2 \quad (3.4)$$

De (3.3) se tiene que

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right),$$

y por lo tanto,

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)}} \sim N(0, 1).$$

De (3.4) y las propiedades de la distribución Ji-cuadrada se sabe que:

$$\frac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2} \sim \chi_{(m+n-2)}^2.$$

Así que por el Teorema 3.7 se concluye que

$$\frac{\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)}}}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2(m+n-2)}}} = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{(m+n-2)}\left(\frac{1}{m} + \frac{1}{n}\right)}},$$

tiene distribución t de Student con $m + n - 2$ grados de libertad.

3.3. Estadísticas de orden

Sea X_1, \dots, X_n una muestra aleatoria de una población con función de densidad $f(x; \theta)$. Considérese la muestra ordenada en forma ascendente $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, sus elementos son llamados *estadísticas de orden*. Para facilitar la notación, sea $X_{(i)} := Y_i$.

Las Y_i 's no son independientes, pues si $Y_j \geq y$ entonces $Y_{j+1} \geq y$, y en general Y_i no tiene por qué coincidir con X_i . Así, por ejemplo, $Y_n = \max\{X_1, \dots, X_n\}$ puede ser cualquiera de las X_i 's :

$$Y_n = \begin{cases} X_1 & \text{si } X_1 \geq X_2, \dots, X_n. \\ X_2 & \text{si } X_2 \geq X_1, X_3, \dots, X_n. \\ \vdots & \vdots \\ X_n & \text{si } X_n \geq X_1, \dots, X_{n-1}. \end{cases}$$

En los siguientes párrafos se deducirán las funciones de distribución y densidad para las estadísticas de orden, únicamente para el caso continuo.

3.3.1. r -ésima estadística de orden (Y_r)

Sea X_1, \dots, X_n una muestra aleatoria de una distribución continua con función de densidad $f(x; \theta)$ y función de distribución $F(x; \theta)$.

Para obtener $F_{Y_r}(y) = \mathbb{P}(Y_r \leq y)$, observe que el evento $\{Y_r \leq y\}$ ocurre si y sólo si al menos r de las X_i 's son menores o iguales a y , es decir si $\sum_{i=1}^n Z_i \geq r$, donde

$$Z_i = \begin{cases} 1 & \text{si } X_i \leq y, \\ 0 & \text{si } X_i > y. \end{cases}$$

Note que $Z_i \sim \text{Bernoulli}[\underbrace{\mathbb{P}(X_i \leq y)}_{F(y)}]$, entonces $\sum_{i=1}^n Z_i \sim \text{Bin}(n, F(y))$. Por lo tanto,

$$\begin{aligned} F_{Y_r}(y) &= \mathbb{P}(Y_r \leq y) = \mathbb{P}\left(\sum_{i=1}^n Z_i \geq r\right) = \sum_{j=r}^n \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j}. \\ \therefore F_{Y_r}(y) &= \sum_{j=r}^n \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j}. \end{aligned}$$

Usando este resultado se obtiene la distribución de Y_1 , la mínima estadística de orden, de la siguiente manera:

$$\begin{aligned} F_{Y_1}(y) &= \sum_{j=1}^n \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j} \\ &= [F(y) + 1 - F(y)]^n - [1 - F(y)]^n \\ &= 1 - [1 - F(y)]^n. \end{aligned}$$

Y diferenciando se deduce la función de densidad correspondiente:

$$f_{Y_1}(y) = n[1 - F(y)]^{n-1} f(y).$$

De la misma forma, se obtiene la distribución de la máxima estadística de orden Y_n :

$$F_{Y_n}(y) = [F(y)]^n. \quad (3.5)$$

De donde:

$$f_{Y_n}(y) = n[F(y)]^{n-1} f(y).$$

3.3.2. Distribución conjunta de las estadísticas de orden mínima y máxima

En teoría de la probabilidad se estudia que una propiedad de la función de distribución de dos variables $F_{X,Y}(x, y)$ es la siguiente:

$$\mathbb{P}[a < X \leq b, c < Y \leq d] = F_{X,Y}(a, c) + F_{X,Y}(b, d) - F_{X,Y}(a, d) - F_{X,Y}(b, c).$$

Lo que, junto con las otras propiedades:

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

y

$$\lim_{y \rightarrow \infty} F_{X,Y}(x, y) = F_X(x)$$

conduce a:

$$\mathbb{P}[y_1 < Y_1, Y_n \leq y_n] = F_{Y_n}(y_n) - F_{Y_1, Y_n}(y_1, y_n). \quad (3.6)$$

Por otro lado, se tiene que, por ser Y_1 y Y_n la mínima y la máxima estadísticas de orden, se satisface:

$$\begin{aligned} \mathbb{P}(y_1 < Y_1, Y_n \leq y_n) &= \mathbb{P}(y_1 < X_1 \leq y_n, y_1 < X_2 \leq y_n, \dots, y_1 < X_n \leq y_n) \\ &= \prod_{i=1}^n \mathbb{P}[y_1 < X_i \leq y_n] = \prod_{i=1}^n (F(y_n) - F(y_1)) \\ &= (F(y_n) - F(y_1))^n. \end{aligned}$$

Por lo tanto,

$$\mathbb{P}(y_1 < Y_1, Y_n \leq y_n) = (F(y_n) - F(y_1))^n. \quad (3.7)$$

Igualando (3.6) y (3.7) se obtiene que:

$$F_{Y_n}(y_n) - F_{Y_1, Y_n}(y_1, y_n) = (F(y_n) - F(y_1))^n.$$

De esta manera, $F_{Y_1, Y_n}(y_1, y_n) = F_{Y_n}(y_n) - (F(y_n) - F(y_1))^n$. Y por (3.5) se llega a que:

$$F_{Y_1, Y_n}(y_1, y_n) = [F(y_n)]^n - (F(y_n) - F(y_1))^n.$$

Para obtener $f_{Y_1, Y_n}(y_1, y_n)$ se usará la propiedad que relaciona a las funciones de densidad y de distribución para el caso de dos variables, a saber,

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

así:

$$\frac{\partial}{\partial y_1} F_{Y_1, Y_n}(y_1, y_n) = n(F(y_n) - F(y_1))^{n-1} f(y_1),$$

y

$$\frac{\partial^2}{\partial y_n \partial y_1} F_{Y_1, Y_n}(y_1, y_n) = n f(y_1) (n-1) (F(y_n) - F(y_1))^{n-2} f(y_n).$$

De esta forma se concluye que:

$$f_{Y_1, Y_n}(y_1, y_n) = n(n-1)(F(y_n) - F(y_1))^{n-2} f(y_1) f(y_n). \quad (3.8)$$

Ejemplo 3.2 Sea X_1, \dots, X_n una muestra aleatoria de una población con distribución $U(0, 1)$.

Obtener: $f_{Y_1}(y_1)$, $f_{Y_n}(y_n)$ y $f_{Y_1, Y_n}(y_1, y_n)$.

En este caso, $f(x) = \mathbb{I}_{(0,1)}^{(x)}$ y $F(y) = \int_0^y dx = y$. Entonces

$$F_{Y_1}(y) = 1 - [1 - F(y)]^n = 1 - [1 - y]^n \text{ y por tanto } f_{Y_1}(y) = n(1 - y)^{n-1}.$$

$$F_{Y_n}(y) = y^n \text{ y por tanto } f_{Y_n}(y) = n y^{n-1}.$$

$$f_{Y_1, Y_n}(y_1, y_n) = n(n-1)(y_n - y_1)^{n-2} \mathbb{I}_{(0,1)}^{(y_1)} \mathbb{I}_{(0,1)}^{(y_n)}.$$

Algunas aplicaciones que tienen las expresiones obtenidas previamente en esta sección, incluyen el cálculo de densidades correspondientes a funciones de las estadísticas de orden, por ejemplo, el **rango** $R := Y_n - Y_1$ y el **promedio** $T := \frac{Y_n + Y_1}{2}$.

Para obtener $f_{R,T}(r, t)$ se utilizará la siguiente igualdad:

$$f_{R,T}(r, t) = |J| f_{Y_1, Y_n}(y_1^{-1}(r, t), y_n^{-1}(r, t)).$$

Para encontrar la distribución conjunta del rango y el promedio, note lo siguiente:

- $Y_n = R + Y_1$ y sustituyendo en la expresión para T , se tiene que $T = \frac{Y_1 + R + Y_1}{2} = Y_1 + \frac{R}{2}$. Por lo tanto $Y_1 = T - \frac{R}{2}$, entonces $Y_n = R + Y_1 = T + \frac{R}{2}$.
- $(R, T) \mapsto (T - \frac{R}{2}, T + \frac{R}{2}) = (Y_1, Y_n)$ es la transformación involucrada.
- $J = \begin{vmatrix} -1/2 & 1 \\ 1/2 & 1 \end{vmatrix} = -\frac{1}{2} - \frac{1}{2} = -1$.

Entonces al sustituir se obtiene

$$f_{R,T}(r, t) = n(n-1) \left[F\left(t + \frac{r}{2}\right) - F\left(t - \frac{r}{2}\right) \right]^{n-2} f\left(t - \frac{r}{2}\right) f\left(t + \frac{r}{2}\right). \quad (3.9)$$

Otra expresión relacionada con las estadísticas de orden que puede ser de interés es la densidad conjunta de Y_1, \dots, Y_n , suponiendo nuevamente que estas variables corresponden a las estadísticas de orden de una muestra aleatoria X_1, \dots, X_n . Así,

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta),$$

en cualquier punto (x_1, \dots, x_n) de \mathbb{R}^n . Pero la densidad de Y_1, \dots, Y_n estaría dada por

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n; \theta) = n! \prod_{i=1}^n f(y_i; \theta),$$

debido a que, al ordenar la muestra, cada punto (y_1, \dots, y_n) acumula la densidad

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) \tag{3.10}$$

de cada punto (x_1, \dots, x_n) obtenido al permutar de forma arbitraria las coordenadas de (y_1, \dots, y_n) . Lo anterior se hace porque $f_{Y_1, \dots, Y_n}(y_1, \dots, y_n; \theta)$, a diferencia de (3.10), está concentrada en la región de \mathbb{R}^n en donde $y_1 \leq y_2 \leq \dots \leq y_n$ y vale cero fuera de esa región; es decir, (y_1, \dots, y_n) se obtiene de permutar la muestra, así que $n!$ considera las posibles permutaciones.

3.4. Estadísticas suficientes

Una muestra aleatoria contiene información del parámetro desconocido θ en $f(x; \theta)$, por lo que para lograr el objetivo de disminuir el grado de desconocimiento de dicho parámetro, se usará la muestra aleatoria. Como ya se ha señalado, una estadística es una función de la muestra aleatoria y, como se verá utilizando métodos de estimación, las estadísticas resultan ser los estimadores de los parámetros. Usualmente las estadísticas representan una reducción de dimensión con respecto a la muestra original X_1, \dots, X_n , el caso más común es una función T que va de \mathbb{R}^n a \mathbb{R} . Una pregunta que puede plantearse en este momento, es si estas estadísticas contienen la misma información (con respecto a θ) que la muestra original o, si en el proceso de transformación de la muestra aleatoria, se pierde información acerca del parámetro desconocido. Es claro que no se desea perder la información original que contiene la muestra aleatoria, por lo que si se tuviera que elegir entre una estadística que contiene la misma información que la muestra contra otra que no, la decisión sería quedarse con la primera. Este razonamiento lleva a un concepto muy importante en estadística, el de *suficiencia*.

Se puede decir que una estadística $S(X_1, X_2, \dots, X_n)$ es **suficiente** si conserva toda la información que contiene la muestra aleatoria X_1, \dots, X_n acerca de θ . En otras palabras, es **suficiente** conocer la estadística

$$S(X_1, X_2, \dots, X_n)$$

para saber del parámetro lo mismo que con X_1, \dots, X_n ; de ahí el nombre de *estadística suficiente*. Una forma alternativa de parafrasearlo es la siguiente: una estadística suficiente para un parámetro θ es aquella que usa toda la información contenida en la muestra con respecto a θ ; no obstante, no es claro cómo se podría saber si una estadística es suficiente o no utilizando esta descripción del concepto, al menos de que se tratara de un caso en donde sea posible “recuperar” los datos a través de la estadística que se esté analizando.

Como ilustración de estas ideas, suponga que se tiene el caso descrito en el ejemplo 3.1, en donde se quiere estudiar la probabilidad θ de obtener sol con una cierta moneda. En este ejemplo, se considera una muestra aleatoria de tamaño 3: X_1, X_2, X_3 , de una distribución Bernoulli(θ) y es un ejercicio en donde se pueden escribir de manera sencilla los ocho posibles valores de la muestra aleatoria. A continuación se reproduce parcialmente el cuadro obtenido en el ejemplo 3.1, incluyendo los valores para la estadística $S(X_1, X_2, X_3) = \bar{X}$.

Resultados	x_1, x_2, x_3	Distribución	\bar{x}
s,s,s	1, 1, 1	θ^3	1
a,a,s	0, 0, 1	$\theta(1 - \theta)^2$	1/3
a,s,a	0, 1, 0	$\theta(1 - \theta)^2$	1/3
s,a,a	1, 0, 0	$\theta(1 - \theta)^2$	1/3
s,s,a	1, 1, 0	$\theta^2(1 - \theta)$	2/3
a,s,s	0, 1, 1	$\theta^2(1 - \theta)$	2/3
s,a,s	1, 0, 1	$\theta^2(1 - \theta)$	2/3
a,a,a	0, 0, 0	$(1 - \theta)^3$	0

Si se propone ahora otra estadística definida como $T(X_1, X_2, X_3) = X_1X_2 + X_3$, los posibles valores que puede tomar son (en el orden de la tabla anterior): 2, 1, 0, 0, 1, 1, 1, 0.

Regresando a la discusión sobre la suficiencia, es claro que con

$$T(X_1, X_2, X_3) = X_1X_2 + X_3$$

no se puede obtener o “recuperar” la información original, pues el primer sumando siempre dará cero cuando al menos uno de X_1 o X_2 sean cero, mientras que \bar{X} sí permite reconstruir la información original en el sentido de que si se sabe que, por ejemplo, $\bar{x} = \frac{1}{3}$, se puede concluir que se obtuvieron dos águilas y un sol (o dos ceros y un uno). En otras palabras, $S(X_1, X_2, X_3)$ es suficiente y $T(X_1, X_2, X_3)$ no.

Observe también que una estadística, en general, condensa la información muestral en el sentido de que para cada valor $t(\underline{x})$ de $T(\underline{X})$ se obtiene un subconjunto del conjunto de las posibles muestras. Sean $\{A_t\}$ los elementos de la *partición generada por la estadística T* (asociada a T o inducida por T), definida como el conjunto de *clases de equivalencia* formadas por

$$A_t = \{(x_1, x_2, \dots, x_n) \in \mathfrak{X} : T(x_1, x_2, \dots, x_n) = t\}.$$

Así, la estadística $\bar{X} = \frac{X_1+X_2+X_3}{n}$ genera la siguiente partición de las posibles muestras, especificada por los valores que toma la estadística:

$$\begin{aligned} A_1 &= \{(1, 1, 1)\}, & A_{\frac{1}{3}} &= \{(0, 0, 1), (0, 1, 0), (1, 0, 0)\}, \\ A_{\frac{2}{3}} &= \{(1, 1, 0), (0, 1, 1), (1, 0, 1)\} & \text{y} & A_0 = \{(0, 0, 0)\}, \end{aligned} \quad (3.11)$$

mientras que la estadística $T(X_1, X_2, X_3) = X_1X_2 + X_3$ genera la siguiente partición:

$$\begin{aligned} A_0 &= \{(0, 1, 0), (1, 0, 0), (0, 0, 0)\}, \\ A_1 &= \{(0, 0, 1), (1, 1, 0), (0, 1, 1), (1, 0, 1)\}, \text{ y} \\ A_2 &= \{(1, 1, 1)\}. \end{aligned} \quad (3.12)$$

Puede decirse entonces que una estadística suficiente condensa la información muestral (en clases de equivalencia) sin perder información relevante. Note que otras estadísticas que sean funciones uno a uno de estadísticas suficientes, también son suficientes. En este ejemplo puede verse fácilmente que la estadística $\sum_{i=1}^3 X_i$ es suficiente y, de hecho, genera la misma partición del conjunto de posibles muestras que generó $S(X_1, X_2, X_3) = \bar{X}$, sólo que en este caso, los valores de $X_1 + X_2 + X_3$ (en el orden de la tabla) son: 3, 1, 2, 0. Así, lo que resulta de interés para hacer inferencia sobre el parámetro es saber la clase de equivalencia en la que está \underline{x} .

A continuación se da una definición formal de suficiencia.

3.4.1. El concepto de suficiencia

Definición 3.3 Sea X_1, \dots, X_n una muestra aleatoria de una población con función de densidad $f(x; \theta)$. La estadística $S(\underline{X})$ es **suficiente** si y sólo si la función de densidad condicional de X_1, \dots, X_n dada $S(\underline{X}) = s$ no depende de θ para cualquier valor s .

Note que esta definición concuerda con la idea intuitiva discutida antes, pues lo que dice es que, dado el conocimiento de $S(\underline{X})$, la muestra aleatoria no proporciona información adicional acerca de θ , es decir, toda la información está contenida en $S(\underline{X})$.

Usando la definición de estadística suficiente, se analizará la suficiencia de $S(\underline{X})$ y $T(\underline{X})$ del ejemplo descrito en los párrafos anteriores. Así,

$$\begin{aligned} \mathbb{P}(X_1 = 0, X_2 = 0, X_3 = 0 | T(\underline{X}) = 0) &= \frac{\mathbb{P}(X_1 = 0, X_2 = 0, X_3 = 0, T = 0)}{\mathbb{P}(T = 0)} \\ &= \frac{(1 - \theta)^3}{2\theta(1 - \theta)^2 + (1 - \theta)^3} \\ &= \frac{1 - \theta}{2\theta + 1 - \theta} = \frac{1 - \theta}{1 + \theta}. \end{aligned}$$

Es decir $\mathbb{P}(X_1 = 0, X_2 = 0, X_3 = 0 | T(\underline{X}) = 0)$ depende de θ , por lo tanto, el estimador T no es suficiente. Esto puede verificarse para otros valores.

Por otro lado,

$$\begin{aligned} \mathbb{P}(X_1 = 0, X_2 = 0, X_3 = 0 | S(\underline{X}) = 0) &= \frac{\mathbb{P}(X_1 = 0, X_2 = 0, X_3 = 0, S = 0)}{\mathbb{P}(S = 0)} \\ &= \frac{(1 - \theta)^3}{(1 - \theta)^3} = 1, \end{aligned}$$

que no depende de θ . En este último caso, debe verificarse para todos los posibles valores y llegar a la conclusión de que esta probabilidad condicional no depende de θ , para cualquier valor s (lo cual efectivamente ocurre).

El concepto de estadística suficiente enfatiza el hecho de que cualquier conocimiento adicional al valor de la estadística, no da mayor información acerca de θ . En el ejemplo, si $S(\underline{X}) = \frac{2}{3}$, se puede decir que se obtuvieron dos soles y un águila, pero no sólo eso: es posible concluir que cualquier $\underline{X} = (X_1, \dots, X_n)$ que satisfaga que $S(\underline{X}) = s$, tiene la misma distribución de probabilidad. Para verificarlo, suponga que sólo se tiene conocimiento de que $S(\underline{X}) = s$, pero no se conoce específicamente el valor muestral que generó este resultado (en el ejemplo, hay tres posibles valores muestrales que hacen que $S(\underline{X}) = \frac{2}{3}$, a saber $(0, 1, 1)$, $(1, 0, 1)$ y $(1, 1, 0)$). Conociendo la densidad condicional $\mathbb{P}(\underline{X} = \underline{x} | S(\underline{X}) = s)$, donde

$$\underline{x} \in A_s = \{(x_1, x_2, \dots, x_n) \in \mathfrak{X} | S(x_1, x_2, \dots, x_n) = s\},$$

se puede usar un proceso de simulación para generar un vector \underline{X}' tal que

$$\mathbb{P}(\underline{X}' = \underline{x} | S(\underline{X}) = s) = \mathbb{P}(\underline{X} = \underline{x} | S(\underline{X}) = s).$$

Así

$$\begin{aligned} \mathbb{P}(\underline{X} = \underline{x}) &= \mathbb{P}(\underline{X} = \underline{x}, S(\underline{X}) = s) \\ &= \mathbb{P}(\underline{X} = \underline{x} | S(\underline{X}) = s) \mathbb{P}(S(\underline{X}) = s) \\ &= \mathbb{P}(\underline{X}' = \underline{x} | S(\underline{X}) = s) \mathbb{P}(S(\underline{X}) = s) \\ &= \mathbb{P}(\underline{X}' = \underline{x}, S(\underline{X}) = s) \\ &= \mathbb{P}(\underline{X}' = \underline{x}), \end{aligned}$$

de tal manera que \underline{X} y \underline{X}' tienen la misma distribución. En conclusión, toda la información acerca de θ está contenida en el conocimiento de que $S(\underline{X}) = s$.

Los siguientes ejemplos ilustran la definición de suficiencia.

Ejemplo 3.3 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con distribución Bernoulli (θ). ¿Es $\sum_{i=1}^n X_i$ una estadística suficiente?

En este caso, $f(x_i; \theta) = \theta^{x_i} (1 - \theta)^{1-x_i}$, por lo que, de acuerdo a la definición de suficiencia:

$$\begin{aligned} \mathbb{P} \left(X_1 = x_1, \dots, X_n = x_n \mid \sum_{i=1}^n X_i = s \right) &= \frac{\theta^{\sum_{i=1}^n X_i} (1 - \theta)^{n - \sum_{i=1}^n X_i}}{\binom{n}{s} \theta^s (1 - \theta)^{n-s}} \\ &= \frac{\theta^s (1 - \theta)^{n-s}}{\binom{n}{s} \theta^s (1 - \theta)^{n-s}} \\ &= \frac{1}{\binom{n}{s}}, \end{aligned}$$

que no depende de θ , por lo que $\sum_{i=1}^n X_i$ es una estadística suficiente para θ . Note que se usó el hecho de que $\sum_{i=1}^n X_i$ tiene distribución $\text{Bin}(n, \theta)$, así como la condición de que $\sum_{i=1}^n X_i = s$.

Ejemplo 3.4 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con distribución $\text{Poisson}(\theta)$. ¿Es $\sum_{i=1}^n X_i$ una estadística suficiente?

Recordando que si $X \sim \text{Poisson}(\theta)$ entonces $f(x; \theta) = \frac{e^{-\theta} \theta^x}{x!}$, se tiene que:

$$\begin{aligned} \mathbb{P}[X_1 = x_1, \dots, X_n = x_n \mid \sum_{i=1}^n X_i = s] &= \frac{e^{-n\theta} \theta^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n x_i!}; y \sum_{i=1}^n X_i = s \\ &= \frac{e^{-n\theta} \theta^s}{\prod_{i=1}^n x_i!} \\ &= \frac{e^{-n\theta} (n\theta)^s}{s!} \\ &= \frac{n^s \prod_{i=1}^n x_i!}{s!}, \end{aligned}$$

que no depende de θ ; por lo tanto, $\sum_{i=1}^n X_i$ es una estadística suficiente para θ . En este caso se usó el hecho de que $\sum_{i=1}^n X_i \sim \text{Poisson}(n\theta)$.

Observe que en los ejemplos anteriores, a no ser que la suma de los enteros x_1, x_2, \dots, x_n sea igual a s , la probabilidad condicional es igual a cero.

Observación 3.3 En general, sea $k(s; \theta)$ la función de densidad de la estadística

$$S(X_1, X_2, \dots, X_n) := S(\underline{X}),$$

donde X_1, X_2, \dots, X_n es una muestra aleatoria de una población con función de densidad $f(x; \theta)$, $\theta \in \Theta$. La probabilidad condicional de $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$, dado que $S(\underline{X}) = s$, es igual a

$$\mathbb{P}[X_1 = x_1, \dots, X_n = x_n \mid S(\underline{X}) = s] = \frac{f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)}{k(s; \theta)},$$

siempre que x_1, \dots, x_n , sean tales que $S(x_1, x_2, \dots, x_n) = s$, y esta probabilidad condicional vale cero en otro caso. Se dice que $S(X_1, X_2, \dots, X_n)$ es una estadística suficiente para θ si y sólo si esta razón no depende de θ .

Con las distribuciones del tipo continuo, no se puede hablar de la probabilidad de que $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$. En este caso, se establece que si la razón

$$\frac{f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \theta)}{k(s; \theta)} = \frac{f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)}{k(s; \theta)},$$

no depende de θ , entonces la distribución condicional de X_1, X_2, \dots, X_n dado $S(\underline{X}) = s$, no depende de θ . En general, las matemáticas para probar que una estadística es suficiente en una distribución continua, pueden representar una dificultad mayor que para el caso discreto.

Ejemplo 3.5 Sea X_1, \dots, X_n una muestra aleatoria de una población con distribución normal $N(\mu, \sigma^2)$, donde σ^2 es conocida. La media muestral, $S(\underline{X}) = \bar{X} = (X_1 + \dots + X_n)/n$, es una estadística suficiente para μ . Para comprobarlo, se verificará que el cociente

$$\frac{f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \mu, \sigma^2)}{k(s; \mu, \sigma^2)} \quad (3.13)$$

no depende de μ (σ^2 es conocida), donde $k(s; \mu, \sigma^2)$ es la función de densidad de \bar{X} . La función de densidad conjunta de la muestra es

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n; \mu, \sigma^2) &= \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\left[\frac{-(x_i - \mu)^2}{2\sigma^2}\right] \\ &= (2\pi\sigma^2)^{-n/2} \exp\left[\frac{-\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right] \\ &= (2\pi\sigma^2)^{-n/2} \exp\left[\frac{-\sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2}{2\sigma^2}\right], \end{aligned}$$

al desarrollar el cuadrado en el exponente, esta expresión es equivalente a

$$(2\pi\sigma^2)^{-n/2} \exp\left[\frac{-\left(\sum_{i=1}^n (x_i - \bar{x})^2 + 2(\bar{x} - \mu) \sum_{i=1}^n (x_i - \bar{x}) + n(\bar{x} - \mu)^2\right)}{2\sigma^2}\right],$$

y como $\sum_{i=1}^n (x_i - \bar{x}) = 0$, se tiene que

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp \left[\frac{-\left(\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2\right)}{2\sigma^2} \right].$$

Recordando que la media muestral \bar{X} tiene distribución $N(\mu, \sigma^2/n)$, entonces el cociente (3.13) está dado por

$$\begin{aligned} \frac{f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \mu, \sigma^2)}{k(s; \mu, \sigma^2)} &= \frac{(2\pi\sigma^2)^{-n/2} \exp \left[\frac{-\left(\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2\right)}{2\sigma^2} \right]}{(2\pi\sigma^2/n)^{-1/2} \exp \left[\frac{-n(\bar{x} - \mu)^2}{2\sigma^2} \right]} \\ &= n^{-1/2} (2\pi\sigma^2)^{-\frac{(n-1)}{2}} \exp \left[\frac{-\sum_{i=1}^n (x_i - \bar{x})^2}{2\sigma^2} \right], \end{aligned}$$

que no depende de μ . Por lo tanto, la media muestral es una estadística suficiente para μ .

3.4.2. El teorema de factorización

El siguiente resultado conocido como el *teorema de factorización de Neyman* o simplemente teorema de factorización, permite encontrar una estadística suficiente sin hallar la función de densidad de la estadística de interés y, más aún, sin tener que proponer dicha estadística.

Teorema 3.8 Sea X_1, \dots, X_n una muestra aleatoria de una población con función de densidad $f(x; \theta)$; $S(\underline{X})$ es suficiente si y sólo si, la función de densidad conjunta de X_1, \dots, X_n puede factorizarse como:

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = g(S(\underline{x}); \theta) \cdot h(x_1, \dots, x_n),$$

donde g y h son funciones no negativas tales que $g(S(\underline{x}); \theta)$ depende de la muestra sólo a través de $S(\underline{x})$ y también depende de θ ; y $h(x_1, \dots, x_n)$ no depende de θ .

Es decir, el teorema de factorización establece que

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = \underbrace{g(S(\underline{x}); \theta)}_{\text{depende de } \theta \text{ y de la muestra sólo a través de } S} \cdot \underbrace{h(x_1, \dots, x_n)}_{\text{depende sólo de la muestra}}$$

Demostración (caso discreto):

$\Rightarrow S(\underline{X})$ es suficiente.

Si x_1, x_2, \dots, x_n , son tales que $S(x_1, x_2, \dots, x_n) = s$, la densidad conjunta de X_1, \dots, X_n puede escribirse como

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n; S(\underline{X}) = s).$$

Usando la definición de probabilidad condicional, la expresión anterior es equivalente a

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n | S(\underline{X}) = s) \cdot \mathbb{P}[S(\underline{X}) = s]. \quad (3.14)$$

Por la suficiencia de $S(\underline{X})$, la probabilidad condicional en (3.14) no depende de θ , por lo que la densidad conjunta de X_1, \dots, X_n se puede expresar como:

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = h(x_1, \dots, x_n) \cdot g(S(\underline{x}); \theta),$$

donde

$$h(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n | S(\underline{X}) = s)$$

y

$$g(S(\underline{x}); \theta) = \mathbb{P}[S(\underline{X}) = s].$$

⊞ Por demostrar que $S(\underline{X})$ es suficiente, es decir, se debe probar que

$$\mathbb{P}[X_1 = x_1, \dots, X_n = x_n | S(\underline{X})]$$

no depende de θ , bajo el supuesto de que la factorización es válida.

Se tiene que

$$\mathbb{P}[X_1 = x_1, \dots, X_n = x_n | S(\underline{X}) = s] = \frac{\mathbb{P}[X_1 = x_1, \dots, X_n = x_n, S(\underline{X}) = s]}{\mathbb{P}[S(\underline{X}) = s]}$$

pero

$$\mathbb{P}[S(\underline{X}) = s] = \sum_{A_s} \mathbb{P}[X_1 = x_1, \dots, X_n = x_n], \quad (3.15)$$

donde:

$$A_s = \{(x_1, x_2, \dots, x_n) \in \mathfrak{X} : S(x_1, x_2, \dots, x_n) = s\}$$

y \mathfrak{X} representa el conjunto de todos los posibles valores de (X_1, \dots, X_n) . Es decir, A_s consta de todos aquellos (x_1, x_2, \dots, x_n) tales que $S(\underline{x}) = s$ (ver por ejemplo (3.11) y (3.12)).

Por la hipótesis, $\mathbb{P}[X_1 = x_1, \dots, X_n = x_n]$ se puede factorizar como el producto $g(S(\underline{x}); \theta)h(\underline{x})$, de tal manera que (3.15) se puede escribir como:

$$\mathbb{P}[S(\underline{X}) = s] = \sum_{A_s} g(S(\underline{x}); \theta)h(\underline{x}) = g(s; \theta) \sum_{A_s} h(\underline{x})$$

y

$$\begin{aligned}\mathbb{P}[X_1 = x_1, \dots, X_n = x_n | S(\underline{X}) = s] &= \frac{g(S(\underline{x}); \theta) h(\underline{x})}{g(s; \theta) \sum_{A_s} h(\underline{x})} \\ &= \frac{g(s; \theta) h(\underline{x})}{g(s; \theta) \sum_{A_s} h(\underline{x})} \\ &= \frac{h(\underline{x})}{\sum_{A_s} h(\underline{x})},\end{aligned}$$

si (x_1, x_2, \dots, x_n) es tal que $S(\underline{x}) = s$ y vale cero en otro caso. Es decir,

$$\mathbb{P}[X_1 = x_1, \dots, X_n = x_n | S(\underline{X}) = s]$$

no depende de θ , por lo que $S(\underline{X})$ es una estadística suficiente. ■

Demostración (caso continuo):

Se supone primero la factorización, es decir,

$$\begin{aligned}f_{X_1, \dots, X_n}(x_1, x_2, \dots, x_n; \theta) &= \prod_{i=1}^n f(x_i, \theta) \\ &= g(S(\underline{x}); \theta) h(x_1, x_2, \dots, x_n).\end{aligned}$$

Sea $V_1 = S(X_1, \dots, X_n)$ y considere la transformación uno a uno definida de la siguiente manera

$$\begin{aligned}v_1 &= S(x_1, x_2, \dots, x_n), \\ v_i &= u_i(x_1, x_2, \dots, x_n), \text{ para } i = 2, \dots, n;\end{aligned}$$

con funciones inversas

$$x_i = w_i(v_1, v_2, \dots, v_n) \text{ para } i = 1, 2, \dots, n$$

y Jacobiano $J = [\partial w_i / \partial v_j]$, que denota la entrada (i, j) de la matriz.

Entonces la densidad conjunta de las variables V_1, V_2, \dots, V_n está dada por:

$$k_{V_1, \dots, V_n}(v_1, \dots, v_n; \theta) = |J| g(v_1; \theta) h(w_1(v_1, v_2, \dots, v_n), \dots, w_n(v_1, v_2, \dots, v_n)).$$

Así, la densidad de V_1 puede obtenerse como

$$\begin{aligned}k_{V_1}(v_1; \theta) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} k_{V_1, \dots, V_n}(v_1, \dots, v_n; \theta) dv_2 \cdots dv_n \\ &= g(v_1; \theta) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |J| h(w_1(v_1, \dots, v_n), \dots, w_n(v_1, \dots, v_n)) dv_2 \cdots dv_n.\end{aligned}$$

Observe que θ no está involucrada ni en el Jacobiano, ni en los límites de integración, además de que la función h no depende de θ . Por lo tanto, la integral múltiple es una función exclusivamente de v_1 . Sea

$$m(v_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |J| h(w_1(v_1, v_2, \dots, v_n), \dots, w_n(v_1, v_2, \dots, v_n)) dv_2 \cdots dv_n,$$

con lo que

$$k_{V_1}(v_1; \theta) = g(v_1; \theta) m(v_1).$$

Note que si $m(v_1) > 0$, es posible escribir:

$$g(v_1; \theta) = \frac{k_{V_1}(v_1; \theta)}{m(v_1)},$$

o

$$g(S(\underline{x}); \theta) = \frac{k_{V_1}(S(\underline{x}); \theta)}{m[S(\underline{x})]},$$

con lo que la factorización que se ha supuesto como hipótesis, se puede reescribir como:

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, x_2, \dots, x_n; \theta) &= g(S(\underline{x}); \theta) h(x_1, x_2, \dots, x_n) \\ &= \frac{k_{V_1}(S(\underline{x}); \theta)}{m[S(\underline{x})]} h(x_1, x_2, \dots, x_n) \\ &= k_{V_1}(S(\underline{x}); \theta) \frac{h(x_1, x_2, \dots, x_n)}{m[S(\underline{x})]}, \end{aligned}$$

de donde:

$$\frac{f_{X_1, \dots, X_n}(x_1, x_2, \dots, x_n; \theta)}{k_{V_1}(S(\underline{x}); \theta)} = \frac{h(x_1, x_2, \dots, x_n)}{m[S(\underline{x})]}$$

no depende de θ , lo cual implica que $V_1 = S(\underline{X})$ es una estadística suficiente (ver Observación 3.3).

Inversamente, si $V_1 = S(\underline{X})$ es suficiente, la factorización se puede hacer tomando la función g como la densidad de $S(\underline{X})$, es decir, como $k_{V_1}(S(\underline{x}); \theta)$.

■

Ejemplo 3.6 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con distribución $Poisson(\theta)$. Hallar una estadística suficiente para θ .

Usando el teorema de factorización:

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) &= \frac{e^{-n\theta} \theta^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n x_i!} \prod_{i=1}^n \mathbb{I}_{\{0,1,\dots\}}^{(x_i)} \\ &= \underbrace{e^{-n\theta} \theta^{\sum_{i=1}^n X_i}}_{g(S(\underline{x}); \theta)} \underbrace{\frac{1}{\prod_{i=1}^n x_i!} \prod_{i=1}^n \mathbb{I}_{\{0,1,\dots\}}^{(x_i)}}_{h(x_1, \dots, x_n)} \end{aligned}$$

$\therefore S(\underline{X}) = \sum_{i=1}^n X_i$ es una estadística suficiente para θ .

Ejemplo 3.7 Sea X_1, \dots, X_n una muestra aleatoria de una población con distribución Uniforme $(0, \theta)$. Para encontrar una estadística suficiente para θ , se tiene que:

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) = \frac{1}{\theta^n} \prod_{i=1}^n \mathbb{I}_{(0, \theta)}^{(x_i)}.$$

Como todas las observaciones son positivas, es decir $x_i > 0$, para $i = 1, 2, \dots, n$, se tiene que

$$\prod_{i=1}^n \mathbb{I}_{(0, \theta)}^{(x_i)} = 1$$

siempre y cuando $y_n = \max\{x_1, \dots, x_n\} < \theta$. Por lo tanto, la densidad conjunta puede escribirse como

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = \frac{1}{\theta^n} \mathbb{I}_{(0, \theta)}^{(y_n)},$$

concluyendo que, por el teorema de factorización, $Y_n = \max\{X_1, \dots, X_n\}$ es suficiente, con $g(s(x_1, x_2, \dots, x_n); \theta) = \frac{1}{\theta^n} \mathbb{I}_{(0, \theta)}^{(y_n)}$ y $h(x_1, x_2, \dots, x_n) = 1$.

Observe que esta factorización no es única, pueden proponerse otras; pero Y_n será la estadística suficiente.

Ejemplo 3.8 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con función de densidad:

$$f(x; \theta) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1, 0 < \theta \\ 0, & \text{en otro caso.} \end{cases}$$

Se usará el teorema de factorización para probar que el producto

$$S(X_1, X_2, \dots, X_n) = X_1 X_2 \cdots X_n$$

es una estadística suficiente para θ . La función de densidad conjunta de X_1, X_2, \dots, X_n es:

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) &= f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta) = \prod_{i=1}^n \theta x_i^{\theta-1} \\ &= \theta^n \prod_{i=1}^n x_i^{\theta-1} \\ &= \left[\theta^n (\prod_{i=1}^n x_i)^\theta \left(\frac{1}{\prod_{i=1}^n x_i} \right) \right] \end{aligned}$$

donde $0 < x_i < 1$, $i = 1, 2, \dots, n$. En el teorema de factorización, sean:

$$g(s(x_1, x_2, \dots, x_n); \theta) = \theta^n (\prod_{i=1}^n x_i)^\theta$$

y

$$h(x_1, x_2, \dots, x_n) = \frac{1}{\prod_{i=1}^n x_i}.$$

Ya que $h(x_1, x_2, \dots, x_n)$ no depende de θ , el producto $X_1 X_2 \cdots X_n = \prod_{i=1}^n X_i$ es una estadística suficiente para θ . Observe que si se hubiera querido usar la definición para este ejercicio, era necesario obtener la densidad de la estadística dada por el producto $S(\underline{X}) = X_1 X_2 \cdots X_n = \prod_{i=1}^n X_i$, de ahí lo valioso del teorema de factorización.

Observación 3.4 La muestra ordenada (Y_1, \dots, Y_n) siempre es una estadística suficiente, pues en el esquema de muestreo aleatorio que se está considerando, la función de densidad conjunta no depende del orden de las variables. A pesar de que la muestra ordenada tiene la misma dimensión que la muestra original, al ordenar se elimina la información irrelevante con relación a la posición con la que se produce cada observación. Una forma de comprobarlo es usando el teorema de factorización. La función de densidad conjunta de X_1, \dots, X_n es:

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) &= f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta) \\ &= n! f(y_1; \theta) f(y_2; \theta) \cdots f(y_n; \theta). \end{aligned}$$

En el teorema de factorización, sean:

$$g(s(x_1, x_2, \dots, x_n); \theta) = f(y_1; \theta) f(y_2; \theta) \cdots f(y_n; \theta)$$

y

$$h(x_1, x_2, \dots, x_n) = n!.$$

Ya que $h(x_1, x_2, \dots, x_n)$ no depende de θ , la muestra ordenada (Y_1, \dots, Y_n) es una estadística suficiente.

Ejemplo 3.9 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con distribución Logística(μ, σ):

$$f(x; \mu, \sigma) = \frac{e^{-(x-\mu)/\sigma}}{\sigma[1 + e^{-(x-\mu)/\sigma}]^2}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0.$$

La función de densidad conjunta de X_1, X_2, \dots, X_n es:

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) &= f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta) \\ &= \prod_{i=1}^n \frac{e^{-(x_i-\mu)/\sigma}}{\sigma[1 + e^{-(x_i-\mu)/\sigma}]^2} \\ &= n! \prod_{i=1}^n \frac{e^{-(y_i-\mu)/\sigma}}{\sigma[1 + e^{-(y_i-\mu)/\sigma}]^2}, \end{aligned}$$

donde y_1, \dots, y_n es la muestra ordenada. Debido a que no es posible factorizar la función de densidad conjunta de otra manera tal que se pueda aplicar el teorema de factorización, entonces la muestra ordenada (Y_1, \dots, Y_n) es una estadística suficiente.

Cabe aclarar que la muestra ordenada se usará como la estadística suficiente sólo cuando no sea posible obtener otra estadística suficiente.

3.4.3. La familia exponencial

Hay un conjunto de familias paramétricas de distribuciones que, por sus propiedades, tiene gran relevancia dentro de la inferencia estadística. Este conjunto se conoce como la *clase exponencial* o la *familia exponencial* y su importancia en este momento es que las densidades de esta clase tienen una relación con el concepto de suficiencia, en el sentido de que si se sabe que una densidad pertenece a la familia exponencial, es posible encontrar una estadística suficiente de una forma sencilla. En esta parte se revisará el caso de densidades con un sólo parámetro, mientras que en la sección 3.6 se verá el caso de dos o más parámetros. Algunas de las distribuciones más conocidas como la binomial, la geométrica, la binomial negativa, la Poisson, la Gamma, la Normal y la Beta, pertenecen a la familia exponencial.

Definición 3.4 *Se dice que $f(x; \theta)$ pertenece a la familia exponencial (o clase exponencial) si puede factorizarse como:*

$$f(x; \theta) = a(\theta)b(x)e^{c(\theta)d(x)}, \quad \theta \in \Theta, \quad -\infty < x < \infty,$$

donde $a(\theta)$ y $c(\theta)$ son funciones de θ , y $b(x)$ y $d(x)$ son funciones de x .

Ejemplo 3.10 *Las siguientes distribuciones pertenecen a la familia exponencial:*

1. Si X tiene distribución exponencial con parámetro θ , es decir

$$X \sim \exp(\theta),$$

su función de densidad está dada por

$$f(x; \theta) = \theta \cdot e^{-\theta \cdot x} \cdot \mathbb{I}_{(0, \infty)}^{(x)}.$$

En este caso, $a(\theta) = \theta$, $b(x) = \mathbb{I}_{(0, \infty)}^{(x)}$, $c(\theta) = -\theta$ y $d(x) = x$.

2. Si X tiene distribución Poisson con parámetro θ , es decir

$$X \sim \text{Poisson}(\theta),$$

su función de densidad es

$$f(x; \theta) = \frac{e^{-\theta} \theta^x}{x!} \mathbb{I}_{\{0, 1, \dots\}}^{(x)} = e^{-\theta} \cdot \frac{1}{x!} \mathbb{I}_{\{0, 1, \dots\}}^{(x)} \cdot e^{x \cdot \ln(\theta)},$$

de donde puede verse que pertenece a la familia exponencial.

3. Si X tiene distribución Bernoulli con parámetro θ , es decir

$$X \sim \text{Bernoulli}(\theta),$$

con

$$f(x; \theta) = \theta^x (1 - \theta)^{1-x} = (1 - \theta) \cdot \mathbb{I}_{\{0,1\}}^{(x)} \cdot e^{x \cdot \ln(\frac{\theta}{1-\theta})},$$

se tiene que pertenece a la familia exponencial.

4. Si X tiene distribución Geométrica con parámetro θ , es decir

$$X \sim \text{Geométrica}(\theta),$$

su función de densidad se puede escribir como

$$f(x; \theta) = \theta (1 - \theta)^x \mathbb{I}_{\{0,1,\dots\}}^{(x)} = \theta \cdot \mathbb{I}_{\{0,1,\dots\}}^{(x)} e^{x \cdot \ln(1-\theta)},$$

con lo que se concluye que la distribución Geométrica pertenece a la familia exponencial.

Note que la distribución Uniforme en el intervalo $(0, \theta)$ no pertenece a la familia exponencial.

El resultado que se enuncia a continuación proporciona la relación entre los miembros de la familia exponencial y la suficiencia.

Proposición 3.9 Sea X_1, \dots, X_n una muestra aleatoria de una población con función de densidad de la forma

$$f(x; \theta) = a(\theta)b(x)e^{c(\theta)d(x)},$$

es decir, $f(x; \theta)$ pertenece a la familia exponencial, entonces $\sum_{i=1}^n d(x_i)$ es una estadística suficiente.

Demostración: Usando el teorema de factorización, se tiene que:

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) &= \prod_{i=1}^n a(\theta)b(x_i)e^{c(\theta)d(x_i)} \\ &= \underbrace{\left[a(\theta)^n e^{c(\theta) \sum_{i=1}^n d(x_i)} \right]}_{g(S(\underline{X}); \theta)} \cdot \underbrace{\prod_{i=1}^n b(x_i)}_{h(X_1, \dots, X_n)} \end{aligned}$$

$\therefore S(\underline{X}) = \sum_{i=1}^n d(X_i)$ es una estadística suficiente para θ . ■

En la siguiente tabla se muestran algunas de las distribuciones más conocidas que pertenecen a la familia o clase exponencial, aunque en la sección 3.6 se ampliará esta lista.

Algunos miembros de la Familia Exponencial

$$f(x; \theta) = a(\theta) b(x) \exp [c(\theta) d(x)]$$

Distribución	$f(x; \theta)$	$a(\theta)$	$b(x)$	$c(\theta)$	$d(x)$
Bernoulli (θ)	$\theta^x (1 - \theta)^{1-x}$	$1 - \theta$	1	$\ln \left(\frac{\theta}{1-\theta} \right)$	x
Binomial (n, θ)	$\binom{n}{x} \theta^x (1 - \theta)^{n-x}$	$(1 - \theta)^n$	$\binom{n}{x}$	$\ln \left(\frac{\theta}{1-\theta} \right)$	x
Geométrica (θ)	$\theta (1 - \theta)^x$	θ	1	$\ln(1 - \theta)$	x
Poisson (θ)	$\frac{e^{-\theta} \theta^x}{x!}$	$e^{-\theta}$	$\frac{1}{x!}$	$\ln \theta$	x
Exponencial (θ)	$\theta e^{-\theta x}$	θ	1	$-\theta$	x
Rayleigh (θ)	$\frac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}}$	$\frac{1}{\theta^2}$	x	$-\frac{1}{2\theta^2}$	x^2

3.4.4. Suficiencia minimal

La idea de la suficiencia es no perder información relevante del parámetro (contenida en la muestra) al usar una estadística. Pero puede haber diferentes estadísticas que satisfagan esa condición y lo ideal sería encontrar la que condense más la información. A tales estadísticas se les llama suficientes minimales.

Intuitivamente, una estadística es *suficiente minimal* si cualquier reducción de la misma da lugar a estadísticas que ya no son suficientes. Es decir, ya no se pueden reducir más los datos, puesto que se perdería información contenida en la muestra acerca del parámetro. Lo anterior también está asociado con la noción de clases de equivalencia o partición del espacio muestral inducida por cada estadística.

Para ilustrar esta situación, considere una muestra aleatoria X_1, \dots, X_n de la distribución Bernoulli con parámetro θ , así como las siguientes estadísticas

$$S_1(\underline{X}) = (X_{(1)}, \dots, X_{(n)}) = (Y_1, \dots, Y_n),$$

$$S_2(\underline{X}) = \left(\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} X_i, \sum_{i=\lfloor \frac{n}{2} \rfloor + 1}^n X_i \right),$$

donde $\lfloor v \rfloor$ indica la parte entera de v , y

$$S_3(\underline{X}) = \sum_{i=1}^n X_i.$$

Considerando que

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) &= \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} \prod_{i=1}^n \mathbb{I}_{\{0,1\}}(x_i) \\ &= (1 - \theta)^n \left(\frac{\theta}{1 - \theta} \right)^{\sum_{i=1}^n x_i} \prod_{i=1}^n \mathbb{I}_{\{0,1\}}(x_i), \end{aligned}$$

por el teorema de factorización, las tres estadísticas son suficientes en la distribución Bernoulli, pues la densidad conjunta también puede escribirse como

$$\begin{aligned} &(1 - \theta)^n \left(\frac{\theta}{1 - \theta} \right)^{x(1)} \left(\frac{\theta}{1 - \theta} \right)^{x(2)} \cdots \left(\frac{\theta}{1 - \theta} \right)^{x(n)} \prod_{i=1}^n \mathbb{I}_{\{0,1\}}(x_i) \\ &= (1 - \theta)^n \left(\frac{\theta}{1 - \theta} \right)^{\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} x_i} \left(\frac{\theta}{1 - \theta} \right)^{\sum_{i=\lfloor \frac{n}{2} \rfloor + 1}^n x_i} \prod_{i=1}^n \mathbb{I}_{\{0,1\}}(x_i). \end{aligned}$$

La diferencia entre las tres estadísticas radica en el distinto grado de simplificación obtenido por eliminación de información que es irrelevante para el conocimiento de θ . $S_1(\underline{X})$ tiene dimensión n , $S_2(\underline{X})$ tienen dimensión 2 y $S_3(\underline{X})$ tiene dimensión 1. S_3 es más resumida que S_1 y S_2 , mientras que S_2 es más resumida que S_1 . Observe entonces que conociendo S_1 o S_2 se puede conocer S_3 , pero no al revés, y conociendo S_1 se puede conocer S_2 , pero no al revés. Si una estadística suficiente es más resumida que cualquier otra, se le llama *suficiente minimal*. En el ejemplo, note que S_3 puede escribirse como función de S_1 o de S_2 , y el hecho de que S_3 sea más resumida que S_1 significa que existe una función (medible) φ tal que S_3 puede escribirse como $S_3 = \varphi(S_1)$. Lo anterior se formaliza en la Definición 3.5.

Siguiendo la idea del párrafo anterior, suponga que $S'(\underline{X})$ es más resumida que $S(\underline{X})$. En términos de la partición inducida por una estadística, si $\{A_{s'}\}$ son los elementos de la partición asociados con $S'(\underline{X})$ y $\{A_s\}$ son los elementos de la partición asociados con $S(\underline{X})$, se tendría que cada A_s es un subconjunto de algún $A_{s'}$. Si en el ejemplo de la distribución Bernoulli, se toma $n = 3$, se tienen los siguientes valores de S_1, S_2 y S_3 , de donde se pueden revisar las particiones asociadas:

(X_1, X_2, X_3)	$S_1(\underline{X})$	$S_2(\underline{X})$	$S_3(\underline{X})$
(0, 0, 0)	(0, 0, 0)	(0, 0)	0
(0, 0, 1)	(0, 0, 1)	(0, 1)	1
(0, 1, 0)	(0, 0, 1)	(0, 1)	1
(1, 0, 0)	(0, 0, 1)	(1, 0)	1
(1, 1, 0)	(0, 1, 1)	(1, 1)	2
(1, 0, 1)	(0, 1, 1)	(1, 1)	2
(0, 1, 1)	(0, 1, 1)	(0, 2)	2
(1, 1, 1)	(1, 1, 1)	(1, 2)	3

Note que S_1 y S_3 inducen la misma partición (de 4 elementos), aún cuando en S_3 hay una reducción en la dimensión; mientras que S_2 induce una partición de 6 elementos, a saber:

$$\begin{aligned} A_1 &= \{(0, 0, 0)\}, \\ A_2 &= \{(0, 0, 1), (0, 1, 0)\}, \\ A_3 &= \{(1, 0, 0)\}, \\ A_4 &= \{(1, 1, 0), (1, 0, 1)\}, \\ A_5 &= \{(0, 1, 1)\} \end{aligned}$$

y

$$A_6 = \{(1, 1, 1)\},$$

siendo cada uno de éstos un subconjunto de alguno de los elementos de la partición inducida por S_3 . Por ejemplo, A_2 es un subconjunto de

$$\{(0, 0, 1), (0, 1, 0), (1, 0, 0)\},$$

que corresponde al subconjunto del espacio muestral tal que $S_3 = 1$.

Una forma de referirse a una partición inducida por $S(\underline{X})$, tal que $A_s \subseteq A_{s'}$ para alguna estadística $S'(\underline{X})$, es como una *partición más fina* (es decir, $S(\underline{X})$ induce una partición más fina que $S'(\underline{X})$ o $S'(\underline{X})$ induce una partición menos fina que $S(\underline{X})$). En estos términos, si $S'(\underline{X})$ es más resumida que $S(\underline{X})$, entonces $S(\underline{X})$ genera una partición más fina que $S'(\underline{X})$. En el ejemplo, S_2 induce una partición más fina que S_3 .

En este contexto, la suficiencia minimal está asociada con *la partición menos fina que conserva la suficiencia*.

Definición 3.5 *Se dice que una estadística es suficiente minimal si y sólo si es función de cualquier otra estadística suficiente; es decir, $S'(\underline{X})$ es suficiente minimal si y sólo si existe φ tal que $S' = \varphi(S)$, donde $S(\underline{X})$ es cualquier otra estadística suficiente.*

En términos de las clases de equivalencia, cualquier estadística con una partición más fina que la minimal, será suficiente. El que $S'(\underline{X})$ sea función de $S(\underline{X})$ significa que si $S(\underline{x}) = S(\underline{x}')$, entonces $S'(\underline{x}) = S'(\underline{x}')$.

Observación 3.5 *Las estadísticas suficientes minimales no son únicas, pues al igual que las estadísticas suficientes, cualquier función biunívoca resulta ser también una estadística suficiente minimal.*

El siguiente resultado proporciona un método para encontrar estadísticas suficientes minimales.

Teorema 3.10 Sea X_1, \dots, X_n una muestra aleatoria de una población con función de densidad $f(x; \theta)$. Sean $\underline{x} = (x_1, \dots, x_n)$ y $\underline{x}' = (x'_1, \dots, x'_n)$ dos valores muestrales y $S(\underline{X})$ una estadística, tal que el cociente

$$\frac{f_{X_1, \dots, X_n}(\underline{x}; \theta)}{f_{X_1, \dots, X_n}(\underline{x}'; \theta)}$$

no depende de θ si y sólo si $S(\underline{x}) = S(\underline{x}')$. Entonces $S(\underline{X})$ es una estadística suficiente minimal.

Demostración:

Para simplificar la prueba, suponga que $f_{X_1, \dots, X_n}(\underline{x}; \theta) > 0$ para todo valor $\underline{x} \in \mathfrak{X}$ y para toda θ .

Primero se exhibirá que $S(\underline{X})$ es una estadística suficiente. Sea

$$\mathcal{S} = \{s : S(\underline{x}) = s \text{ para algún } \underline{x} \in \mathfrak{X}\},$$

la imagen de \mathfrak{X} bajo $S(\underline{x})$. Como antes, la partición de conjuntos inducida por $S(\underline{x}) = s$ es

$$A_s = \{\underline{x} : S(\underline{x}) = s\}.$$

Para cada $\underline{x} \in \mathfrak{X}$, sea \underline{x}_s un elemento que está en el mismo conjunto A_s , que \underline{x} . Puesto que \underline{x} y \underline{x}_s están en el mismo conjunto A_s , $S(\underline{x}) = S(\underline{x}_s)$ y usando la hipótesis, $f_{X_1, \dots, X_n}(\underline{x}; \theta)/f_{X_1, \dots, X_n}(\underline{x}_s; \theta)$ no depende de θ . Sean $h(\underline{x}) = f_{X_1, \dots, X_n}(\underline{x}; \theta)/f_{X_1, \dots, X_n}(\underline{x}_s; \theta)$ y $g(s; \theta) = f_{X_1, \dots, X_n}(\underline{x}_s; \theta)$, una función en \mathcal{S} . Entonces, se puede ver que

$$f_{X_1, \dots, X_n}(\underline{x}; \theta) = \frac{f_{X_1, \dots, X_n}(\underline{x}_s; \theta)f_{X_1, \dots, X_n}(\underline{x}; \theta)}{f_{X_1, \dots, X_n}(\underline{x}_s; \theta)} = g(s; \theta)h(\underline{x})$$

y, por el teorema de factorización, $S(\underline{X})$ es una estadística suficiente para θ .

Para probar que $S(\underline{X})$ es minimal, sea $S^*(\underline{X})$ cualquier otra estadística suficiente. Por el teorema de factorización, existen funciones g^* y h^* tales que $f_{X_1, \dots, X_n}(\underline{x}; \theta) = g^*(S^*(\underline{x}); \theta)h^*(\underline{x})$. Sean \underline{x} y \underline{x}' cualesquiera dos valores muestrales que satisfacen que $S^*(\underline{x}) = S^*(\underline{x}')$, entonces

$$\frac{f_{X_1, \dots, X_n}(\underline{x}; \theta)}{f_{X_1, \dots, X_n}(\underline{x}'; \theta)} = \frac{g^*(S^*(\underline{x})|\theta)h^*(\underline{x})}{g^*(S^*(\underline{x}')|\theta)h^*(\underline{x}')} = \frac{h^*(\underline{x})}{h^*(\underline{x}')}.$$

Como h^* no depende de θ , este cociente no depende de θ y la hipótesis del teorema implica que $S(\underline{x}) = S(\underline{x}')$.

Por lo tanto, a partir de $S^*(\underline{x}) = S^*(\underline{x}')$, se deduce que $S(\underline{x}) = S(\underline{x}')$; lo cual significa que $S(\underline{x})$ es una función de $S^*(\underline{x})$, concluyéndose que $S(\underline{X})$ es minimal. ■

Ejemplo 3.11 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con distribución Bernoulli(θ), entonces:

$$\begin{aligned} \frac{f_{X_1, \dots, X_n}(\underline{x}; \theta)}{f_{X_1, \dots, X_n}(\underline{x}'; \theta)} &= \frac{\prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i}}{\prod_{i=1}^n \theta^{x'_i} (1-\theta)^{1-x'_i}} = \frac{\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}}{\theta^{\sum_{i=1}^n x'_i} (1-\theta)^{n-\sum_{i=1}^n x'_i}} \\ &= \left(\frac{\theta}{1-\theta} \right)^{\sum_{i=1}^n x_i - \sum_{i=1}^n x'_i}, \end{aligned}$$

el cual no depende de θ si y sólo si $\sum_{i=1}^n x_i = \sum_{i=1}^n x'_i$. Así, $\sum_{i=1}^n X_i$ es una estadística suficiente minimal.

Ejemplo 3.12 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con distribución Uniforme en el intervalo $(0, \theta)$. En este caso,

$$\begin{aligned} \frac{f_{X_1, \dots, X_n}(\underline{x}; \theta)}{f_{X_1, \dots, X_n}(\underline{x}'; \theta)} &= \frac{\frac{1}{\theta^n} \prod_{i=1}^n \mathbb{I}_{(0, \theta)}^{(x_i)}}{\frac{1}{\theta^n} \prod_{i=1}^n \mathbb{I}_{(0, \theta)}^{(x'_i)}} \\ &= \frac{\mathbb{I}_{(0, \theta)}^{(y_n)}}{\mathbb{I}_{(0, \theta)}^{(y'_n)}} = \frac{\mathbb{I}_{(y_n, \infty)}^{(\theta)}}{\mathbb{I}_{(y'_n, \infty)}^{(\theta)}}, \end{aligned}$$

cociente que no depende de θ si y sólo si $y_n = y'_n$, concluyendo que Y_n es una estadística suficiente minimal.

Suficiencia minimal y la familia exponencial

Se sabe que $f(x; \theta)$ pertenece a la familia exponencial si se cumple que:

$$f(x; \theta) = a(\theta)b(x)e^{c(\theta)d(x)}$$

Tomando $\underline{X} = (X_1, \dots, X_n)$ y $\underline{X}' = (X'_1, \dots, X'_n)$ y usando el Teorema 3.10, se tiene que:

$$\begin{aligned} \frac{f(\underline{x}; \theta)}{f(\underline{x}'; \theta)} &= \frac{\prod_{i=1}^n a(\theta)b(x_i) \exp\{c(\theta)d(x_i)\}}{\prod_{i=1}^n a(\theta)b(x'_i) \exp\{c(\theta)d(x'_i)\}} \\ &= \frac{(a(\theta))^n \prod_{i=1}^n b(x_i) \exp\{c(\theta) \sum_{i=1}^n d(x_i)\}}{(a(\theta))^n \prod_{i=1}^n b(x'_i) \exp\{c(\theta) \sum_{i=1}^n d(x'_i)\}} \\ &= \frac{\prod_{i=1}^n b(x_i)}{\prod_{i=1}^n b(x'_i)} \exp\{c(\theta) [\sum_{i=1}^n d(x_i) - \sum_{i=1}^n d(x'_i)]\}. \end{aligned}$$

Este cociente no depende de θ si sólo si $\sum_{i=1}^n d(x_i) = \sum_{i=1}^n d(x'_i)$.

$\therefore S(\underline{X}) = \sum_{i=1}^n d(X_i)$ es suficiente minimal.

Es decir, si $f(x; \theta)$ pertenece a la familia exponencial, $\sum_{i=1}^n d(X_i)$ es una estadística suficiente minimal.

3.5. Completez

El concepto de estadística completa se utilizará en uno de los resultados más importantes que servirán para encontrar estimadores insesgados de varianza mínima.

Definición 3.6 Sea X_1, \dots, X_n una muestra aleatoria de $f(x; \theta)$, $\theta \in \Theta$. Se dice que la estadística $T(\underline{X})$ es completa si y sólo si, para cualquier función g de T , se tiene que si $\mathbb{E}(g(T)) = 0 \forall \theta \in \Theta$, entonces

$$\mathbb{P}(g(T) = 0) = 1, \quad \forall \theta \in \Theta.$$

También se dice que la **familia de densidades de T es completa**.

Se puede pensar entonces que T es completa si una estadística, función de T , cuya esperanza es 0, vale 0 con probabilidad 1.

Ejemplo 3.13 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con distribución Bernoulli (θ), $0 < \theta < 1$, y $T(\underline{X}) = \sum_{i=1}^n X_i$. Para ver si T es completa, se considera $\mathbb{E}(g(T)) = 0$, que es equivalente a:

$$\mathbb{E}(g(T)) = \sum_{t=0}^n g(t) \binom{n}{t} \theta^t (1 - \theta)^{n-t} = 0,$$

de donde

$$(1 - \theta)^n \sum_{t=0}^n g(t) \binom{n}{t} \left(\frac{\theta}{1 - \theta} \right)^t = 0,$$

ó

$$\sum_{t=0}^n g(t) \binom{n}{t} \left(\frac{\theta}{1 - \theta} \right)^t = 0,$$

el cual es un polinomio en $\left(\frac{\theta}{1-\theta}\right)$, que vale cero si cada uno de los coeficientes de $\left(\frac{\theta}{1-\theta}\right)^t$ vale cero, es decir, si

$$g(t) \binom{n}{t} = 0, \quad \text{para toda } t = 0, 1, 2, \dots, n,$$

pero $\binom{n}{t} \neq 0$, así que $g(t) = 0, \forall t \in \{0, 1, 2, \dots, n\}$, lo que lleva a la conclusión de que $T(\underline{X}) = \sum_{i=1}^n X_i$ es una estadística completa.

Observe que si en esta distribución se toma $T(\underline{X}) = X_1 - X_2$ y $g(T) = T$, se tiene que $\mathbb{E}(X_1 - X_2) = \mathbb{E}(X_1) - \mathbb{E}(X_2) = 0$, pero $X_1 - X_2$ no vale cero con probabilidad 1, así que $X_1 - X_2$ no es completa.

Ejemplo 3.14 Sea X_1, \dots, X_n una muestra aleatoria de una población con distribución Uniforme continua en el intervalo $(0, \theta)$. Para ver si

$$T(\underline{X}) = Y_n = \max\{X_1, \dots, X_n\}$$

es completa, se hace

$$\mathbb{E}(g(T)) = \int g(t) f_T(t) dt = 0.$$

En el caso de la máxima estadística de orden,

$$f_T(t) = n \frac{t^{n-1}}{\theta^n} \mathbb{I}_{(0,\theta)}^{(t)}, \quad t = y_n.$$

Así

$$\begin{aligned} \mathbb{E}(g(T)) &= \int_0^\theta g(t) \frac{n}{\theta^n} t^{n-1} dt \\ &= \frac{n}{\theta^n} \int_0^\theta g(t) t^{n-1} dt = 0, \quad \forall \theta > 0, \end{aligned}$$

entonces

$$\int_0^\theta g(t) t^{n-1} dt = 0, \quad \forall \theta > 0.$$

Lo cual implica que

$$g(\theta) \theta^{n-1} = 0 \quad \forall \theta > 0,$$

y por lo tanto, $g(\theta) = 0, \forall \theta > 0$, lo que lleva a concluir que la máxima estadística de orden es una estadística completa.

Observación 3.6 En general, puede decirse que una familia paramétrica de distribuciones $f(x; \theta)$ es completa si $\mathbb{E}[g(X)] = 0$ implica que $g(x) = 0$ casi seguramente (c.s.). En este contexto, si $f(x; \theta)$ pertenece a la familia exponencial, entonces $f(x; \theta)$ es completa. Para verificarlo, se considerará el caso particular en el que $c(\theta) = \theta$ y $d(x) = x$, es decir: $f(x; \theta) = a(\theta) b(x) e^{\theta x}$. Ahora, si para toda θ , se cumple que:

$$\int_{-\infty}^{\infty} g(x) f(x; \theta) dx = 0,$$

se tiene que

$$\int_{-\infty}^{\infty} g(x) a(\theta) b(x) e^{\theta x} dx = 0,$$

o

$$\int_{-\infty}^{\infty} [g(x) b(x)] e^{\theta x} dx = 0,$$

la cual corresponde a la transformada de Laplace de la función $g(x)b(x)$. Por la propiedad de unicidad de la transformada, la única función que tiene una transformada igual a cero es la función que vale cero c.s., es decir, $g(x)b(x) = 0$ c.s., de donde se obtiene que $g(x) = 0$ c.s., considerando $b(x) \neq 0$. Así que $f(x; \theta)$ es completa.

Observación 3.7 Si $f(x; \theta)$ pertenece a la familia exponencial, entonces $\sum_{i=1}^n d(X_i)$ es completa.

En síntesis, una de las ventajas que se tiene al identificar a un miembro de la familia exponencial está explicada en el siguiente resultado.

Teorema 3.11 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con función de densidad $f(x; \theta)$ con $\theta \in \Theta \subseteq \mathbb{R}$, donde $f(x; \theta)$ pertenece a la familia exponencial, es decir, $f(x; \theta) = a(\theta)b(x) \exp\{c(\theta)d(x)\}$. Entonces, la estadística $\sum_{i=1}^n d(X_i)$ es **suficiente minimal y completa**.

Aunque este resultado es un resumen de lo que ya se justificó en 3.4.4 (suficiencia minimal y la familia exponencial) y de las observaciones anteriores, la parte correspondiente a la completez puede consultarse con mayor detalle en Zacks (1971), página 69, o Schervish (1995), páginas 108-110. ■

De esta manera, es posible encontrar estadísticas suficientes y completas fácilmente cuando la muestra proviene de un miembro de la familia exponencial. Por ejemplo, sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con distribución *Geométrica* (θ), con $\theta \in [0, 1]$; para encontrar una estadística suficiente y completa, observe que $f(x; \theta)$ pertenece a la familia exponencial, ya que como se vio antes,

$$\begin{aligned} f(x; \theta) &= \theta(1 - \theta)^x \mathbb{I}_{\{0,1,\dots\}}(x) \\ &= \theta \exp\{x \ln(1 - \theta)\} \mathbb{I}_{\{0,1,\dots\}}(x), \end{aligned}$$

con

$$\begin{aligned} a(\theta) &= \theta, \\ b(x) &= \mathbb{I}_{\{0,1,\dots\}}(x), \\ c(\theta) &= \ln(1 - \theta), \\ d(x) &= x. \end{aligned}$$

De donde, puede concluirse que $S(\underline{X}) = \sum_{i=1}^n X_i$ es una estadística suficiente (minimal) y completa.

3.6. Algunas generalizaciones

La primera generalización de los resultados vistos previamente, se refiere al teorema de factorización para un conjunto de estadísticas suficientes.

Teorema 3.12 *Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con función de densidad $f(x; \underline{\theta})$, donde $\underline{\theta}$ es un vector de parámetros. Las estadísticas $S_1(\underline{X}), S_2(\underline{X}), \dots, S_r(\underline{X}), r \geq k$, son conjuntamente suficientes si y sólo si existen dos funciones: $g(S_1, \dots, S_r; \underline{\theta})$ que depende de X_1, X_2, \dots, X_n sólo a través de S_1, \dots, S_r y de $\underline{\theta}$; y $h(\underline{X})$ cualquier función no negativa que sólo depende de la muestra, para las cuales la densidad conjunta $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_k)$ puede factorizarse como:*

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \underline{\theta}) = g(S_1, \dots, S_r; \underline{\theta}) h(\underline{x}).$$

■

Ejemplo 3.15 *Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con distribución $N(\mu, \sigma^2)$. Encontrar estadísticas suficientes para $\underline{\theta} = (\mu, \sigma^2)$.*

La densidad conjunta de X_1, X_2, \dots, X_n puede escribirse como

$$\begin{aligned} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i - \mu)^2} \\ &= \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} \\ &= \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2} (\sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2)}, \end{aligned}$$

la densidad conjunta depende de x_1, x_2, \dots, x_n sólo a través de $\sum_{i=1}^n x_i^2$ y $\sum_{i=1}^n x_i$, por lo que las estadísticas $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ son conjuntamente suficientes. Note que \bar{X} y S^2 son funciones uno a uno de $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$, por lo que también son suficientes para la distribución Normal.

Familias exponenciales k -paramétricas

Cuando la familia paramétrica tiene más de un parámetro, es decir, su función de densidad es de la forma $f(x; \underline{\theta})$ con $\underline{\theta} \in \Theta \subseteq \mathbb{R}^k$, se dice que pertenece a la familia exponencial k -paramétrica si y sólo si puede expresarse de la forma

$$f(x; \underline{\theta}) = a(\underline{\theta})b(x) \exp \left\{ \sum_{j=1}^k c_j(\underline{\theta})d_j(x) \right\}$$

Ejemplo 3.16 Sea $X \sim \text{Gamma}(r, \lambda)$, ¿pertenece a la familia exponencial k -paramétrica?.
En este caso,

$$\begin{aligned} f(x; r, \lambda) &= \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \mathbb{I}_{(0, \infty)}(x) \\ &= \frac{\lambda^r}{\Gamma(r)} e^{-\lambda x + (r-1) \ln(x)} \mathbb{I}_{(0, \infty)}(x), \end{aligned}$$

por lo que:

$$\begin{aligned} a(\boldsymbol{\theta}) &= \frac{\lambda^r}{\Gamma(r)}, & b(x) &= \mathbb{I}_{(0, \infty)}(x), \\ c_1(\boldsymbol{\theta}) &= -\lambda, & d_1(x) &= x, \\ c_2(\boldsymbol{\theta}) &= r - 1, & d_2(x) &= \ln x, \end{aligned}$$

por lo tanto, se concluye que la densidad Gamma pertenece a la familia exponencial.

Ejemplo 3.17 La distribución Normal (μ, σ^2) pertenece a la familia exponencial.

Para comprobarlo, la densidad puede escribirse como:

$$\begin{aligned} f(x; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x^2 - 2x\mu + \mu^2)} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\mu^2}{2\sigma^2}} e^{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x}, \end{aligned}$$

de donde: $a(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\mu^2}{2\sigma^2}}$, $b(x) = 1$, $c_1 = -\frac{1}{2\sigma^2}$, $d_1 = x^2$, $c_2 = \frac{\mu}{\sigma^2}$ y $d_2 = x$.

Ejemplo 3.18 La distribución Beta (θ_1, θ_2) pertenece a la familia exponencial, pues la densidad se puede escribir como:

$$\begin{aligned} f(x; \theta_1, \theta_2) &= \frac{1}{B(\theta_1, \theta_2)} x^{\theta_1-1} (1-x)^{\theta_2-1} \mathbb{I}_{(0,1)}(x) \\ &= \frac{1}{B(\theta_1, \theta_2)} \mathbb{I}_{(0,1)}(x) e^{(\theta_1-1) \ln x + (\theta_2-1) \ln(1-x)}. \end{aligned}$$

Por último, se enuncia una extensión del resultado que relaciona a la familia exponencial con la suficiencia.

Teorema 3.13 Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población con función de densidad $f(x; \boldsymbol{\theta})$, con $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$, que pertenece a la familia exponencial k -paramétrica, es decir:

$$f(x; \boldsymbol{\theta}) = a(\boldsymbol{\theta})b(x) \exp \left\{ \sum_{j=1}^k c_j(\boldsymbol{\theta})d_j(x) \right\}.$$

Entonces, el conjunto de estadísticas

$$\left(\sum_{i=1}^n d_1(X_i), \sum_{i=1}^n d_2(X_i), \dots, \sum_{i=1}^n d_k(X_i) \right)$$

son suficientes y completas. ■

Con este resultado y considerando el ejemplo 3.16, se puede afirmar que si $X \sim \text{Gamma}(r, \lambda)$, entonces $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n \ln(X_i) \right)$ son suficientes y completas. También puede concluirse lo mismo para $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ en la distribución Normal y para $(\sum_{i=1}^n \ln X_i, \sum_{i=1}^n \ln(1 - X_i))$ en la distribución Beta.

3.7. Estadísticas auxiliares

En esta sección se estudiará un tipo especial de estadísticas llamadas auxiliares. Como se verá, una estadística auxiliar no contiene información acerca de θ , es una variable aleatoria cuya distribución es fija y conocida, sin relación con θ . Sin embargo, cuando se usa en conjunto con otras estadísticas, puede contener información valiosa para hacer inferencias acerca del parámetro.

Definición 3.7 *A una estadística $T(\underline{X})$ cuya distribución no depende del parámetro θ , se le llama estadística auxiliar.*

Ejemplo 3.19 *Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución uniforme continua sobre el intervalo $(\theta, \theta + 1)$, $-\infty < \theta < \infty$. Sean $Y_1 = X_{(1)}, \dots, Y_n = X_{(n)}$ las estadísticas de orden de la muestra. La estadística Rango $R = Y_n - Y_1$, es una estadística auxiliar. Para verificarlo, se mostrará que la densidad de R no depende de θ .*

La función de distribución de cada X_i es

$$F(x; \theta) = \begin{cases} 0, & \text{si } x \leq \theta, \\ x - \theta, & \text{si } \theta < x < \theta + 1, \\ 1, & \text{si } x \geq \theta + 1, \end{cases}$$

mientras que la función de densidad está dada por

$$f(x; \theta) = \mathbb{I}_{(\theta, \theta+1)}^{(x)}.$$

Por lo tanto, usando la expresión (3.8) deducida en la sección 3.3.2 la función de densidad conjunta de Y_1 y Y_n es

$$\begin{aligned} f_{Y_1, Y_n}(y_1, y_n) &= n(n-1)(F(y_n) - F(y_1))^{n-2} f(y_1) f(y_n) \\ &= \begin{cases} n(n-1)(y_n - y_1)^{n-2} & \text{si } \theta < y_1 < y_n < \theta + 1, \\ 0 & \text{en otro caso.} \end{cases} \end{aligned}$$

Haciendo la transformación $R = Y_n - Y_1$ y $T = (Y_1 + Y_n)/2$, la cual tiene la transformación inversa $Y_1 = (2T - R)/2$ y $Y_n = (2T + R)/2$ con Jacobiano igual a -1 (ver Sección 3.3.2), y usando la expresión (3.9), la función de densidad conjunta de R y T está dada por

$$f_{R,T}(r, t) = \begin{cases} n(n-1)r^{n-2} & \text{si } 0 < r < 1; \theta + (r/2) < t < \theta + 1 - (r/2), \\ 0 & \text{en otro caso.} \end{cases}$$

Por lo tanto, la función de densidad de probabilidad para R es

$$\begin{aligned} f_R(r) &= \int_{\theta+(r/2)}^{\theta+1-(r/2)} n(n-1)r^{n-2} dt \\ &= n(n-1)r^{n-2}(1-r), \quad 0 < r < 1. \end{aligned}$$

Esta es la función de densidad de una variable aleatoria con distribución Beta con $\alpha = n-1$ y $\beta = 2$. Y así, la función de densidad es la misma para toda θ ; por lo tanto, la distribución de R no depende de θ , concluyéndose que $R = Y_n - Y_1$ es una estadística auxiliar.

El siguiente resultado, conocido como el teorema de Basu, permite verificar la independencia de dos estadísticas sin necesidad de encontrar su distribución conjunta.

Teorema 3.14 (Teorema de Basu). Si $S(\underline{X})$ es una estadística suficiente y completa y $T(\underline{X})$ es una estadística auxiliar, entonces $S(\underline{X})$ y $T(\underline{X})$ son independientes.

Demostración:

Se hará para el caso discreto. Como $T(\underline{X})$ es una estadística auxiliar, entonces $\mathbb{P}(T(\underline{X}) = t)$ no depende de θ . También la probabilidad condicional

$$\mathbb{P}(T(\underline{X}) = t | S(\underline{X}) = s) = \mathbb{P}(\underline{X} \in \{\underline{x} : T(\underline{x}) = t\} | S(\underline{X}) = s),$$

no depende de θ porque $S(\underline{X})$ es una estadística suficiente.

Por lo tanto, para demostrar que $S(\underline{X})$ y $T(\underline{X})$ son independientes, basta comprobar que

$$\mathbb{P}(T(\underline{X}) = t | S(\underline{X}) = s) = \mathbb{P}(T(\underline{X}) = t),$$

para todos los posibles valores de $s \in \mathcal{S}$. Observe que

$$\mathbb{P}(T(\underline{X}) = t) = \sum_{s \in \mathcal{S}} \mathbb{P}(T(\underline{X}) = t | S(\underline{X}) = s) \mathbb{P}(S(\underline{X}) = s). \quad (3.16)$$

Por otro lado, dado que $\sum_{s \in \mathcal{S}} \mathbb{P}(S(\underline{X}) = s) = 1$, se puede escribir

$$\begin{aligned} \mathbb{P}(T(\underline{X}) = t) &= \mathbb{P}(T(\underline{X}) = t) \sum_{s \in \mathcal{S}} \mathbb{P}(S(\underline{X}) = s) \\ &= \sum_{s \in \mathcal{S}} \mathbb{P}(T(\underline{X}) = t) \mathbb{P}(S(\underline{X}) = s). \end{aligned} \quad (3.17)$$

Sea $g(S)$ definida como

$$g(s) = \mathbb{P}(T(\underline{X}) = t | S(\underline{X}) = s) - \mathbb{P}(T(\underline{X}) = t),$$

la cual no depende de θ , pues como se había señalado, ni $\mathbb{P}(T(\underline{X}) = t | S(\underline{X}) = s)$ (por la suficiencia de S), ni $\mathbb{P}(T(\underline{X}) = t)$ (por ser T una estadística auxiliar) dependen de θ ; así que $g(S)$ es una estadística.

Por (3.16) y (3.17), se deduce que

$$\begin{aligned} \mathbb{E}[g(S)] &= \sum_{s \in \mathcal{S}} g(s) \mathbb{P}(S(\underline{X}) = s) \\ &= \sum_{s \in \mathcal{S}} [\mathbb{P}(T(\underline{X}) = t | S(\underline{X}) = s) - \mathbb{P}(T(\underline{X}) = t)] \mathbb{P}(S(\underline{X}) = s) \\ &= \sum_{s \in \mathcal{S}} \mathbb{P}(T(\underline{X}) = t | S(\underline{X}) = s) \mathbb{P}(S(\underline{X}) = s) \\ &\quad - \sum_{s \in \mathcal{S}} \mathbb{P}(T(\underline{X}) = t) \mathbb{P}(S(\underline{X}) = s) \\ &= \mathbb{P}(T(\underline{X}) = t) - \mathbb{P}(T(\underline{X}) = t) \\ &= 0, \text{ para toda } \theta, \end{aligned}$$

y como $S(\underline{X})$ es una estadística completa, se tiene que $g(s) = 0$, para $s \in \mathcal{S}$; por lo que

$$\mathbb{P}(T(\underline{X}) = t | S(\underline{X}) = s) = \mathbb{P}(T(\underline{X}) = t),$$

concluyéndose que $S(\underline{X})$ y $T(\underline{X})$ son independientes. ■

En el teorema de Basu se hace el supuesto de que $S(\underline{X})$ es suficiente y completa, esto implica que $S(\underline{X})$ también es suficiente minimal (ver Schervish (1995), sección 2.1).

Ejemplo 3.20 Sea X_1, \dots, X_n una muestra aleatoria de una población con distribución Uniforme en el intervalo $(0, \theta)$ y sean Y_1 y Y_n las estadísticas de orden mínima y máxima, respectivamente. Entonces las estadísticas $T(\underline{X}) = \frac{Y_1}{Y_n}$ y $S(\underline{X}) = Y_n$, son variables aleatorias independientes. Para verificarlo, recuerde que Y_n es una estadística suficiente y completa para θ (ver ejemplos 3.7 y 3.14). De acuerdo con el teorema de Basu basta mostrar que $T(\underline{X})$ es una estadística auxiliar, por lo que se calculará la función de distribución de $T(\underline{X})$ usando la expresión general para la función de densidad conjunta de Y_1 y Y_n dada por (3.8) e integrando sobre la región adecuada:

$$\begin{aligned} F_T(t) &= \mathbb{P} \left[\frac{Y_1}{Y_n} \leq t \right], & 0 < t < 1 \\ &= \mathbb{P} [Y_1 \leq tY_n] \\ &= \int_0^\theta \int_0^{ty_n} \frac{n(n-1)}{\theta^2} \left(\frac{y_n}{\theta} - \frac{y_1}{\theta} \right)^{n-2} dy_1 dy_n \\ &= [1 - (1-t)^{n-1}] I_{(0,1)}(t) + I_{[1,\infty)}(t). \end{aligned}$$

Por lo tanto, la función de densidad de $T(\underline{X})$ no depende de θ . Así, $T(\underline{X})$ y $S(\underline{X})$ son estadísticas (variables aleatorias) independientes.

3.8. Ejercicios

- Suponga que X es una variable aleatoria discreta con función de densidad dada por

$$f_X(x) = \begin{cases} 0.2 & \text{si } x = 0, \\ 0.3 & \text{si } x = 3, \\ 0.5 & \text{si } x = 12. \end{cases}$$

Sea X_1, X_2 y X_3 una muestra aleatoria de esta distribución. Obtenga:

- La distribución de la muestra (es decir, haga la lista de todos los posibles valores de la muestra aleatoria y sus respectivas probabilidades).
 - La distribución muestral de la media muestral \bar{X} .
 - La distribución muestral de la mediana muestral.
- Sea X una variable aleatoria discreta que toma los valores 1, 2 y 3 con probabilidades $\frac{1}{6}$, $\frac{1}{2}$ y $\frac{1}{3}$, respectivamente. Considere las posibles muestras aleatorias simples de tamaño dos.

- (a) Determine la distribución de la media muestral (es decir, $f_{\bar{X}}(\cdot)$).
- (b) Compruebe que la esperanza de la media muestral es igual a la media poblacional.
- (c) Compruebe que la varianza de la media muestral es igual a la varianza poblacional dividida entre el tamaño de muestra.
3. En un dado están marcados los resultados 1, 2 y X . Para hacer inferencias acerca de las probabilidades p_1 , p_2 y p_X de cada resultado, se lanza tres veces el dado.
- (a) Construya el espacio muestral y la distribución de la muestra (de tamaño 3).
- (b) Obtenga la distribución de T_1 , T_2 y T_X , siendo T_i la frecuencia de resultados iguales a i .
- (c) Determine la distribución de T_X , su media y su varianza.
- (d) Calcule la covarianza entre T_1 y T_2 .
4. De una población con distribución Poisson(θ), se obtiene una muestra aleatoria de tamaño n . Determine la distribución de la media muestral.
5. Sea X_1, \dots, X_n una muestra aleatoria de la población con función de densidad

$$f(x; \theta) = e^{-(x-\theta)} \exp \{-e^{-(x-\theta)}\},$$

donde θ es un número real.

- (a) Obtenga la distribución de $T = \sum_{i=1}^n e^{-X_i}$.
- (b) Calcule $\mathbb{E}[\ln(T)]$ y $Var[\ln(T)]$.
6. Sea X_1, \dots, X_n una muestra aleatoria de una población cuya función de densidad es

$$f(x; \theta) = \frac{\theta}{(1+x)^{1+\theta}}, \quad x > 0,$$

siendo θ un parámetro mayor que cero. Determine la distribución de la estadística $T = \sum_{i=1}^n \ln(1 + X_i)$.

7. Sea X_1, \dots, X_n una muestra aleatoria de una población con distribución Rayleigh, cuya función de densidad es

$$f(x; \theta) = \frac{2}{\theta} x e^{-x^2/\theta}, \quad \mathbb{I}_{(0,\infty)}(x),$$

con $\theta > 0$.

- (a) Obtenga la distribución de $T = \sum_{i=1}^n X_i^2$.

(b) Obtenga $\mathbb{E}(T)$ y $Var(T)$.

8. Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución $N(\mu, \sigma^2)$. Encuentre la media y la varianza de

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}.$$

9. Sea X una variable aleatoria con distribución $\chi_{(1)}^2$. Defina $Z_1 = \sqrt[3]{X}$ y $Z_2 = \sqrt{X}$. Obtenga las distribuciones de Z_1 y Z_2 .

10. Sea X una variable aleatoria con distribución $F(m, n)$. Demuestre que

(a) $Z = \frac{1}{X} \sim F(n, m)$.

(b) $Z = \frac{mX/n}{1+mX/n} \sim Beta(m/2, n/2)$.

11. Sea X una variable aleatoria con distribución $t_{(k)}$. Demuestre que

(a) La distribución de X se aproxima a una distribución $N(0, 1)$ cuando k crece.

(b) $Y = X^2 \sim F(1, k)$.

12. Demuestre el teorema 3.7. Es decir, si Z y U son variables aleatorias independientes con distribución $N(0, 1)$ y $\chi_{(k)}^2$, respectivamente, demuestre que

$$\frac{Z}{\sqrt{U/k}} \sim t_{(k)}.$$

13. Sea X_i , $i = 1, 2, 3$, una muestra aleatoria de la población con distribución $N(i, i^2)$, respectivamente. En cada una de las siguientes situaciones utilice las X_i 's para construir una estadística con la distribución indicada.

(a) χ^2 con 3 grados de libertad.

(b) t de Student con 2 grados de libertad.

(c) F con 1 y 2 grados de libertad.

14. Sea X_1, X_2 una muestra aleatoria de la población con distribución $N(0, 1)$. Para cada uno de los siguientes incisos obtenga las distribuciones de las estadísticas $T(\underline{X})$.

(a) $T(\underline{X}) = (X_2 - X_1)/\sqrt{2}$.

- (b) $T(\underline{X}) = (X_1 + X_2)^2 / (X_2 - X_1)^2$.
- (c) $T(\underline{X}) = X_1^2 / X_2^2$.
- (d) $T(\underline{X}) = (X_2 + X_1) / \sqrt{(X_1 - X_2)^2}$.
15. Sea Z_1, Z_2 una muestra aleatoria de la población con distribución $N(0, 1)$, y sea X_1, X_2 una muestra aleatoria de la población con distribución $N(1, 1)$. Suponga que las Z 's son independientes de las X 's. Para cada uno de los siguientes incisos obtenga las distribuciones de las estadísticas $T = T(\underline{X}, \underline{Z})$.
- (a) $T = \bar{X} + \bar{Z}$.
- (b) $T = (Z_1 + Z_2) / \sqrt{[(X_2 - X_1)^2 + (Z_2 - Z_1)^2] / 2}$.
- (c) $T = [(X_1 - X_2)^2 + (Z_1 - Z_2)^2 + (Z_1 + Z_2)^2] / 2$.
- (d) $T = (X_2 + X_1 - 2)^2 / (X_2 - X_1)^2$.
16. Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución $N(0, 1)$. Sean

$$\bar{X}_k = \frac{1}{k} \sum_{i=1}^k X_i, \quad \bar{X}_{n-k} = \frac{1}{n-k} \sum_{i=k+1}^n X_i.$$

Para cada uno de los siguientes incisos obtenga las distribuciones de las estadísticas $T(\underline{X})$.

- (a) $T(\underline{X}) = \frac{1}{2}(\bar{X}_k + \bar{X}_{n-k})$.
- (b) $T(\underline{X}) = k\bar{X}_k^2 + (n-k)\bar{X}_{n-k}^2$.
- (d) $T(\underline{X}) = X_1 / X_n$.
17. Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución $N(\mu, \sigma^2)$. Sean

$$\begin{aligned} \bar{X}_k &= \frac{1}{k} \sum_{i=1}^k X_i, & S_k^2 &= \frac{1}{k-1} \sum_{i=1}^k (X_i - \bar{X}_k)^2, \\ \bar{X}_{n-k} &= \frac{1}{n-k} \sum_{i=k+1}^n X_i, & S_{n-k}^2 &= \frac{1}{n-k-1} \sum_{i=k+1}^n (X_i - \bar{X}_{n-k})^2, \\ \bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i, & S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \end{aligned}$$

Para cada uno de los siguientes incisos obtenga las distribuciones de las estadísticas $T(\underline{X})$.

- (a) $T(\underline{X}) = [(k-1)S_k^2 + (n-k-1)S_{n-k}^2] / \sigma^2$.
 (b) $T(\underline{X}) = \frac{1}{2}(\bar{X}_k + \bar{X}_{n-k})$.
 (c) $T(\underline{X}) = (\bar{X} - \mu) / (S / \sqrt{n})$.
 (d) $T(\underline{X}) = S_k^2 / S_{n-k}^2$.

18. Sean X_1, \dots, X_n y Z_1, \dots, Z_n muestras aleatorias independientes entre sí de la distribución normal tales que $\mathbb{E}(X_i) = \mathbb{E}(Z_i) = \mu$, $Var(X_i) = \sigma^2$ y $Var(Z_i) = 2\sigma^2$ para cualquier $i \in \{1, 2, \dots, n\}$. Defina

$$U = \sum_{i=1}^n (X_i - \bar{X})^2, \quad V = \sum_{i=1}^n (Z_i - \bar{Z})^2.$$

Determine la distribución de $\frac{2U+V}{2\sigma^2}$.

19. Un inversionista bursátil compra o vende acciones de CEMEX, S. A. mediante el siguiente procedimiento: Selecciona al azar una muestra de días para los que determina el índice medio, \bar{X} de la empresa Edificaciones, S. A.; selecciona también al azar otra muestra de días, para los que determina el índice medio, \bar{Z} , de CEMEX, S. A. Compra acciones de CEMEX, S. A. cuando $\bar{Z} \leq \bar{X}$, vendiendo en caso contrario.

El inversionista supone que ambos índices bursátiles se distribuyen normal e independientemente con una diferencia de medias de 1,432 unidades en favor de la primera y que los índices bursátiles de ambas empresas se comportan de forma independiente todos los días.

Calcular la probabilidad de que el inversionista compre bajo el supuesto de que seleccionó 60 días para calcular el primer índice y obtuvo una varianza muestral de 23 y 50 días para el segundo con una varianza de 7.

20. Sea X_1, X_2 una muestra aleatoria de la población con distribución $N(0, 1)$. Defina $Y = \min(X_1, X_2)$. Demuestre que $Y^2 \sim \chi_{(1)}^2$.
21. Sea X_1, X_2 una muestra aleatoria de la población con distribución Pareto, cuya función de densidad es

$$f(x; \alpha) = \frac{\alpha}{x^{\alpha+1}} \mathbb{I}_{[1, \infty)}(x), \alpha > 0.$$

Sean Y_1 y Y_2 las correspondiente estadísticas de orden. ¿Qué puede decir de la distribución de las variables aleatorias presentadas en los siguientes incisos?

- (a) Y_2 .
 (b) $Y_2 - Y_1$.
 (c) Y_2 / Y_1 .

22. Sea X_1, X_2, X_3 una muestra aleatoria de la población con distribución $U(0, 1)$.

- (a) Obtenga la distribución conjunta de Y_1, Y_2, Y_3 (las correspondientes estadísticas de orden).
- (b) Obtenga la función de densidad de la mediana, es decir, de Y_2 .
- (c) Obtenga la función de densidad del rango $R = Y_3 - Y_1$.

23. Sean Y_1, Y_2, Y_3, Y_4 las estadísticas de orden de una muestra aleatoria de tamaño 4 de la población con función de densidad

$$f_X(x) = \exp(-x) \mathbb{I}_{(0, \infty)}(x).$$

Obtenga lo siguiente:

- (a) $\mathbb{P}(3 \leq Y_4)$.
- (b) $\mathbb{E}(Y_1)$.
- (c) La función de densidad conjunta de Y_1 y Y_4 .

24. Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución exponencial de parámetro $\theta = 1$, es decir $X_i \sim \exp(1)$. Y sean Y_1, \dots, Y_n las correspondientes estadísticas de orden. Demuestre que nY_n y $n(Y_2 - Y_1)$ son independientes.

25. Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución uniforme $U(\theta - 1/2, \theta + 1/2)$.

- (a) Obtenga la distribución de la mínima estadística de orden Y_1 , así como de la máxima Y_n .
- (b) Obtenga las esperanzas de Y_1 y Y_n .

26. Sea X_1, X_2, \dots, X_n una muestra aleatoria de la población con distribución $U(0, \theta)$. Demuestre que Y_1/Y_n y Y_n son variables aleatorias independientes, donde Y_1 y Y_n son las estadísticas de orden mínima y máxima respectivamente.

27. Sea X_1, \dots, X_n una muestra aleatoria de la población con función de densidad

$$f_X(x; a, \theta) = \frac{a}{\theta^a} x^{a-1}, \text{ si } 0 < x < \theta.$$

Sean Y_1, \dots, Y_n las correspondientes estadísticas de orden. Demuestre que $Y_1/Y_2, Y_2/Y_3, \dots, Y_{n-1}/Y_n$ son variables aleatorias mutuamente independientes y obtenga la distribución de cada una de ellas.

28. Sea U_i , $i = 1, 2, \dots$, una muestra aleatoria de la población con distribución $U(0, 1)$, y sea X una variable aleatoria con función de densidad

$$\mathbb{P}(X = x) = \frac{c}{x!}, \quad x = 1, 2, 3, \dots,$$

donde $c = 1/(e - 1)$. Obtenga la distribución de $Y = \min(U_1, \dots, U_X)$. [Nota: La distribución condicional de Y dada $X = x$, $Y|X = x$, es la distribución de la mínima estadística de orden de una muestra de tamaño x].

29. Sea X_1, X_2 una muestra aleatoria de la población con distribución $N(0, 1)$. Obtenga la distribución del Rango, $R = Y_2 - Y_1$.

30. Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución $U(0, 1)$. Sean Y_1, \dots, Y_n las correspondientes estadísticas de orden.

- Obtenga la media y varianza de $Y_n - Y_1$.
- Obtenga la media y varianza de $(Y_1 + Y_n)/2$.
- Obtenga la media y varianza de Y_{k+1} (mediana) considerando que $n = 2k + 1$, con $k = 0, 1, \dots$.
- Compare las medias y las varianzas de \bar{X} , Y_{k+1} , $(Y_1 + Y_n)/2$.

31. Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución $N(\mu, \sigma^2)$. Demuestre que:

- $S(\underline{X}) = \sum_{i=1}^n X_i^2$ es una estadística suficiente para σ^2 cuando $\mu = 0$.
- $S(\underline{X}) = \sum_{i=1}^n \frac{(x_i - \mu)^2}{n}$ es una estadística suficiente para σ^2 cuando μ es conocida.
- $S(\underline{X}) = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}$ no es una estadística suficiente para σ^2 cuando μ es desconocida.

32. Sea X una variable aleatoria con distribución $N(0, \theta)$.

- ¿Es $S(X) = X$ una estadística suficiente para θ ?
- ¿Es $S(X) = |X|$ una estadística suficiente para θ ?

33. *Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución $Gamma(\alpha, \beta)$ cuya función de densidad es

$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta},$$

con $\alpha > 0$, $\beta > 0$ y $0 < x < \infty$.

- (a) Demuestre que si $\alpha = 1$, $S(\underline{X}) = \sum_{i=1}^n X_i$ es una estadística suficiente para β .
- (b) Demuestre que si β es conocido, $S(\underline{X}) = \prod_{i=1}^n X_i$ es una estadística suficiente para α .
- (c) Si α y β son desconocidos, obtenga una estadística suficiente (bidimensional) para los parámetros (α, β) .
34. Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución $Beta(\alpha, 2)$, $\alpha > 0$, $\beta = 2$, cuya función de densidad es

$$f(x; \alpha) = \frac{\Gamma(\alpha + 2)}{\Gamma(\alpha)\Gamma(2)} x^{\alpha-1} (1-x)^1, \quad 0 < x < \infty,$$

cero en cualquier otro caso. Demuestre que $S(\underline{X}) = X_1 X_2 \cdots X_n = \prod_{i=1}^n X_i$ (el producto) es una estadística suficiente para α .

35. Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución $Cauchy(\theta)$ cuya función de densidad es

$$f(x; \theta) = \frac{1}{\pi(1+(x-\theta)^2)}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty.$$

- (a) ¿Puede escribirse la función de densidad conjunta de X_1, X_2, \dots, X_n como en el teorema de factorización?
- (b) ¿Existe una estadística suficiente para el parámetro θ ?

Sugerencia: al hacer la integral (para calcular la esperanza) considere el cambio de variable $x - \theta = \tan \eta$.

36. *Sea X_1, \dots, X_n una muestra aleatoria de una población cuya función de densidad es

$$f_X(x; \mu, \sigma) = \frac{1}{\sigma} \exp\left\{-\frac{(x-\mu)}{\sigma}\right\} \mathbb{I}_{(\mu, \infty)}(x),$$

con $\mu \in \mathbb{R}$ y $\sigma \in \mathbb{R}^+$.

- (a) Demuestre que $S(\underline{X}) = (Y_1, \frac{1}{n} \sum_{i=1}^n (X_i - Y_1))$ es una estadística suficiente para (μ, σ) si ambos parámetros son desconocidos.
- (b) Demuestre que si μ es conocido entonces $S(\underline{X}) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)$ es una estadística suficiente para σ .
- (c) Demuestre que si σ es conocido entonces $S(\underline{X}) = Y_1$ es una estadística suficiente para μ .

37. *Considere la siguiente función de densidad dependiente de tres parámetros α , p y σ ,

$$f(x; \alpha, p, \sigma) = \frac{1}{\sigma \Gamma(p)} \exp \left\{ - \left(\frac{x - \alpha}{\sigma} \right) \right\} \left(\frac{x - \alpha}{\sigma} \right)^{p-1},$$

con $p > 0$, $\sigma > 0$ y $\alpha \leq x < \infty$. Demuestre que:

- (a) Existe una estadística suficiente para p cuando α y σ son conocidos.
 - (b) Existe una estadística suficiente para σ cuando α y p son conocidos.
 - (c) De forma conjunta existe un par de estadísticas suficientes para (p, σ) cuando α es conocida.
 - (d) Si σ es conocida y $p = 1$, existe una estadística suficiente para α .
38. Sea X_1, \dots, X_n una muestra aleatoria de las distribuciones $f(x; \theta)$ especificadas a continuación. Encuentre una estadística suficiente minimal y completa para θ .
- (a) $f(x; \theta) = \theta x^{\theta-1} \mathbb{I}_{(0,1)}(x)$, $\theta > 0$.
 - (b) $f(x; \theta) = \frac{1}{6\theta^4} x^3 \exp(-x/\theta) \mathbb{I}_{(0,\infty)}(x)$, $\theta > 0$.
 - (c) *Binomial*(k, θ) (k fijo).
 - (d) *Binomial Negativa*(k, θ) (k fijo).

39. Sea X_1, \dots, X_n una muestra aleatoria de la población cuya función de densidad es

$$f(x; \theta) = \frac{\theta^2}{\theta + 1} (x + 1) e^{-\theta x} \mathbb{I}_{(0,\infty)}(x), \quad \theta > 0.$$

- (a) Demuestre que la función de densidad de X pertenece a la familia exponencial.
 - (b) Obtenga una estadística suficiente minimal y completa.
40. Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución Pareto, con función de densidad

$$f(x; x_0, \theta) = \theta x_0^\theta x^{-\theta-1}, \quad x_0 \leq x, \quad \theta > 1.$$

donde $x_0 > 0$. Obtenga una estadística suficiente para θ :

- (a) Usando el teorema de factorización.
- (b) Usando la propiedad de la familia exponencial.

¿Las estadísticas suficientes son las mismas?

41. *Sea X_1, \dots, X_n una muestra de aleatoria de la población con distribución Gaussiana Inversa, $IG(\mu, \lambda)$, cuya función de densidad es

$$f(x; \mu, \lambda) = \left[\frac{\lambda}{2\pi x^3} \right]^{1/2} \exp \left\{ -\frac{\lambda(x - \mu)^2}{2\mu^2 x} \right\} \mathbb{I}_{(0, \infty)}(x).$$

Demuestre que $S(\underline{X}) = \left(\bar{X}, \frac{n}{\sum_{i=1}^n \frac{1}{X_i} - \bar{X}^{-1}} \right)$ es una estadística suficiente y completa.

42. Para cada una de las siguientes distribuciones, sea X_1, \dots, X_n una muestra aleatoria. Obtenga una estadística suficiente minimal para θ , o en su caso muestra que no existe alguna.

(a) $f(x; \theta) = \frac{1}{2} e^{-|x-\theta|}, \quad -\infty < x < \infty, -\infty < \theta < \infty.$

(b) $f(x; \theta) = \frac{1}{2\sigma}, \quad \mu - \sigma < x < \mu + \sigma, \theta = (\mu, \sigma).$

(c) $f(x; \theta) = \theta x^{-2}, \quad \theta < x, \theta > 0.$

43. Sea X_1, X_2, X_3 una muestra aleatoria de la población con distribución $Bernoulli(\theta)$. Defina $T = \sum_{i=1}^3 X_i$, $T_1 = X_1$ y $T_2 = (T, T_1)$.

(a) Obtenga las particiones inducidas sobre el espacio muestra a partir de T , T_1 y T_2 , respectivamente.

(b) Muestre que T es una estadística suficiente minimal para θ , pero T_2 no lo es. (Nota: Primero muestre que T conduce a la partición suficiente minimal sobre el espacio muestral, pero que T_2 no conduce a tal partición).

44. Sea X una variable aleatoria con distribución uniforme $U(-\theta, \theta)$.

(a) ¿ $S(X) = X$ es una estadística suficiente minimal para θ ?

(b) ¿ $S(X) = X$ es una estadística completa?

45. *Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución uniforme $U(\theta, \theta + 1)$. Demuestre que $S(\underline{X}) = (Y_1, Y_n)$ es una estadística suficiente minimal y que no es completa.

46. *Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución uniforme $U(\theta_1, \theta_2)$, con $\theta_1 < \theta_2$.

(a) Demuestra que si θ_1 es conocido, entonces $S(\underline{X}) = Y_n$ es una estadística suficiente minimal y completa para θ_2 .

- (b) Demuestra que si θ_2 es conocido, entonces $S(\underline{X}) = Y_1$ es una estadística suficiente minimal y completa para θ_1 .
- (c) Si θ_1 y θ_2 son desconocidos, obtenga una estadística suficiente (bidimensional) minimal y completa para los parámetros (θ_1, θ_2) .
47. Para cada una de las siguientes distribuciones, sea X_1, \dots, X_n una muestra aleatoria. Obtenga una estadística suficiente y completa para θ , o en su caso muestre que no existe alguna.
- (a) $f(x; \theta) = \frac{2x}{\theta^2}, \quad 0 < x < \theta, \theta > 0.$
- (b) $f(x; \theta) = \frac{\ln(\theta)\theta^x}{\theta-1}, \quad 0 < x < 1, \theta > 1.$
- (c) $f(x; \theta) = \frac{\theta}{(1+x)^{1+\theta}}, \quad 0 < x < \infty, \theta > 0.$
- (d) $f(x; \theta) = e^{-(x-\theta)} \exp\{-e^{-(x-\theta)}\}, \quad -\infty < x, \theta < \infty.$
48. Sea T una estadística auxiliar para θ con función de densidad $f_T(t)$. Sea $g(\cdot)$ una función diferenciable uno-a-uno que no depende de θ . Demuestre que $T^* = g(T)$ también es una estadística auxiliar para θ .
49. Sea X_1, \dots, X_n una muestra aleatoria de la población con distribución $N(\mu, \sigma^2)$, donde σ^2 es conocido.
- (a) Demuestre que $S(\underline{X}) = \bar{X}$ es una estadística suficiente y completa para μ , y que $T(\underline{X}) = S^2$ es una estadística auxiliar.
- (b) Demuestre que \bar{X} y S^2 son estadísticas independientes.

Bibliografía

- [1] Campbell, N.A. and Mahon, R.J. (1974). “A multivariate study of variation in two species of rock crab of genus *Leptograpsus*”. *Australian Journal of Zoology* 22, 417-425.
- [2] Casella, G. y Berger, R. L. (2002). *Statistical Inference*. Duxbury Advanced Series. 2nd. ed.
- [3] Efron, B. (1998). “R. A. Fisher in the 21st Century”. *Statistical Science*, 13 (2), 95-122.
- [4] Hogg, R.V., McKean, J. W., Craig, A. T. (2014). *Introduction to Mathematical Statistics*. Pearson Education International. 7th. ed.
- [5] Johnson, R. A. y Bhattacharyya, G. K. (2010). *Statistics, Principles and Methods*. John Wiley and Sons, Inc. 6th. ed.
- [6] Kapadia, A.S., Chan, W. y Moyé, L. (2005). *Mathematical Statistics with Applications*. Chapman & Hall, Inc./CRC Press.
- [7] Lock, R. H. (1993). “1993 New Car Data”. *Journal of Statistics Education* 1.
- [8] Mood, A. M., Graybill, F. A. y Boes, D. C. (1974). *Introduction to the theory of statistics*. Mc Graw-Hill, Inc. 3rd. ed.
- [9] Moore, D. S. (2005). *Estadística Aplicada Básica*. Antoni Bosch. 2a. ed.
- [10] Mukhopadhyay, N. (2006). *Introductory Statistical Inference*. Chapman & Hall CRC Press.
- [11] Pearson, E. S. (1974). “Memories on the impact of Fisher’s work in the 1920’s”. *Int. Stat. Rev.* 42 (1).

- [12] Rao, R. (1997). *Statistics and Truth: Putting Chance to Work*. Singapore: World Scientific Publishing Co. 2nd. ed.
- [13] Rencher, A.C. y Schaalje, G.B. (2008). *Linear Models in Statistics*. John Wiley and Sons, Inc. 2nd. ed.
- [14] Schervish, M. J. (1995). *Theory of Statistics*. Springer Series in Statistics.
- [15] Velez I., R. y García, P. A. (2012). *Principios de Inferencia Estadística*. Universidad Nacional de Educación a Distancia (UNED). 2a. ed.
- [16] Yañez C., S. (2000). “La estadística una ciencia del siglo XX. R.A. Fisher: El Genio”. *Revista Colombiana de Estadística*, 23 (2), 1-14. Editorial Universidad Nacional de Colombia.
- [17] Zacks, S. (1971). *The Theory of Statistical Inference*. John Wiley and Sons, Inc.