
Capítulo 1

Estadística Descriptiva

1.1. Introducción

Una característica del ser humano que lo distingue de otros seres vivos es que tiene la capacidad de interpretar los fenómenos que lo rodean, aprender del mundo a partir de lo que se observa y de su experiencia a lo largo del tiempo. A partir de estas experiencias el hombre aprende a hacer deducciones útiles del mundo en que vive. Como sabemos, existe una gran variedad de fenómenos que quisieramos describir de forma matemática y exacta para poder hacer pronósticos 100 % certeros, sin embargo la misma naturaleza de ciertos fenómenos nos ha obligado a crear modelos matemáticos que permitieran la interacción con la incertidumbre y el azar con base en la teoría de las probabilidades.

Definición 1.1.1 (Fenómenos deterministas). *Un fenómeno determinista es aquel que, cuando se reproduce en las mismas condiciones, podemos predecir con certeza cuál va a ser el resultado, en otras palabras se rige bajo leyes causales. Este tipo de fenómenos no son parte de nuestro estudio.*

Definición 1.1.2 (Fenómenos aleatorios). *Por otro lado, el fenómeno aleatorio es el que cada vez que se realiza, aun bajo condiciones casi idénticas, el resultado no se conoce con certeza, además que el resultado sólo se sabe después de realizado el experimento.*

El estudio de los fenómenos deterministas quedarán fuera del estudio de este curso, por lo que nos concentraremos en desarrollar teoría para poder atacar los fenómenos aleatorios.

Las herramientas con la que contamos para estudiar los fenómenos aleatorios son:

1. La probabilidad, la cual se suele definir de tres formas:

- a) Como un grado de confianza o fundada apariencia de que algo suceda.

- b) La razón entre el número de casos favorables y el número de casos posibles. (Juegos de azar)
 - c) Como una medida basada en planteamiento axiomático de Kolmogorov en 1933. (Teoría de la Medida)
2. La estadística, la cual la podemos definir como sigue:
- a) Es el estudio de los datos cuantitativos de la población, además que es la rama de la matemática que utiliza grandes conjuntos de datos numéricos para obtener inferencias basadas en el cálculo de probabilidades. La estadística clásica o frecuentista se basa en la regularidad estadística, es decir que, al repetir un fenómeno aleatorio un número grande de veces en condiciones constantes, las proporciones en las que ocurren los posibles resultados son muy estables.

1.2. Concepto de medición y de variable

Para cuantificar o clasificar lo que percibimos de un fenómeno aleatorio necesitamos hacer mediciones u observaciones que nos ayuden a investigar una o varias características de interés sobre el fenómeno en estudio. Para un correcto manejo de nuestras mediciones, las observaciones deben ser registradas tomando en cuenta su tipo, para poder saber que tipo de operaciones podemos hacer con ellas.

Cada variable tiene una escala de medición, como vemos a continuación:

Variables	Escalas de medición
Categóricas	$\left\{ \begin{array}{l} \text{Nominales} \\ \text{Ordinales} \end{array} \right.$
Numéricas	$\left\{ \begin{array}{l} \text{De intervalo} \\ \text{De razón} \\ \text{Absoluta} \end{array} \right.$

Una variable es categórica cuando el registro de la medición es un elemento o una categoría, es muy importante que dichas categorías deben de cumplir con ser: mutuamente excluyentes (No hay un elemento que pertenezcan a dos o más categorías a la vez) y exhaustivas (Todo elemento pertenece a una categoría); con este tipo de variables podemos calcular, frecuencia de ocurrencia en cada categoría, la(s) moda(s), proporciones, porcentajes y tablas de contingencia. (Análisis de datos categóricos). La escala de medición que utilizamos en este tipo de variables son básicamente las siguientes dos:

1. Nominales: Cuando las categorías sólo se les da un nombre pero no tienen un orden entre ellas, ejemplos:
 - ¿Está de acuerdo con las obras de continuación del segundo piso del Periférico? Sí , No.
 - Sexo: Masculino, Femenino.
2. Ordinales: Cuando el registro de la medición se expresa en grados de intensidad que tienen un orden, pero no se puede determinar el incremento entre los grados, ejemplo:
 - Grados de satisfacción en un servicio Muy bueno, Bueno, Regular y Malo.
 - Nivel socio económico: Bajo, Medio, Alto

Por otro lado, tenemos una variable numérica cuando el resultado de nuestra medición son valores numéricos. Con este tipo de variables podemos calcular promedios o medias, desviaciones estándar, modas, correlaciones y serán las variables que más estaremos trabajando en este curso. La escala de medición que utilizamos en este tipo de variables son básicamente las siguientes tres:

1. Escala de intervalo: hay un orden entre observaciones, y la distancia entre las mismas tiene significado. En esta escala hay un cero, pero no indica ausencia de medición. Un ejemplo típico de este tipo de variable con esta escala, es la temperatura cuando se mide en grados Fahrenheit o en grados Centígrados. Sabemos que la diferencia entre 30 C y 35 C es la misma que entre 45 C y 50 C y si se dice que un líquido se encuentra a 0 C, no significa que no tiene temperatura.
2. Escala de razón: el cero sí indica una ausencia de la variable, es decir una completa ausencia de medición. Las variables peso, altura son de este tipo.
3. Escala absoluta: se usa para variables discretas o conteos. Ejemplo: Número de hijos en una familia

1.3. Conceptos estadísticos: población y muestra

Parte esencial de un análisis estadístico es tener bien definido el alcance que tienen nuestras inferencias, para ello surgen conceptos básicos como población y muestra que definimos a continuación:

Definición 1.3.1 (Poblacion y muestra). *Definamos como población a todos los elementos presentan una característica común que estamos estudiando (de interés), acerca de la cual intentamos sacar alguna conclusión. Y entenderemos como una muestra a un subconjunto de elementos de la población.*

Nuestro principal estudio en el curso será el estudio de muestras, pero surge la pregunta: ¿Por qué estudiamos muestras en vez de la población? La respuesta es simple, porque en ocasiones es poco factible o hasta imposible observar la totalidad de los individuos, es por esto que en lugar de examinar a toda la población, se estudia una pequeña parte, sin embargo esto no es tan fácil de hacer pues surgen preguntas muy interesantes respecto al proceso de muestreo, como por ejemplo el tamaño de la muestra y la forma en que la obtendremos para que sea **representativa**.

Una muestra de tamaño n de una población en general, denotaremos como:

$$\underline{X} = \{X_1, X_2, \dots, X_n\}$$

Debe de observarse que dicha muestra está formada por variables aleatorias, esto ocurre así porque estamos suponiendo que la muestra aun no la observamos y por lo tanto se considera variable aleatoria a los posibles valores que estaremos observando una vez realizado el experimento. Por otro lado, cuando ya hemos observado los valores de la muestra, entonces la denotaremos como sigue:

$$\underline{x} = \{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$$

Si además suponemos que la variable que analizamos es numérica, entonces denotaremos a la muestra ordenada como:

$$\{X_{(1)}, X_{(2)}, \dots, X_{(n)}\}$$

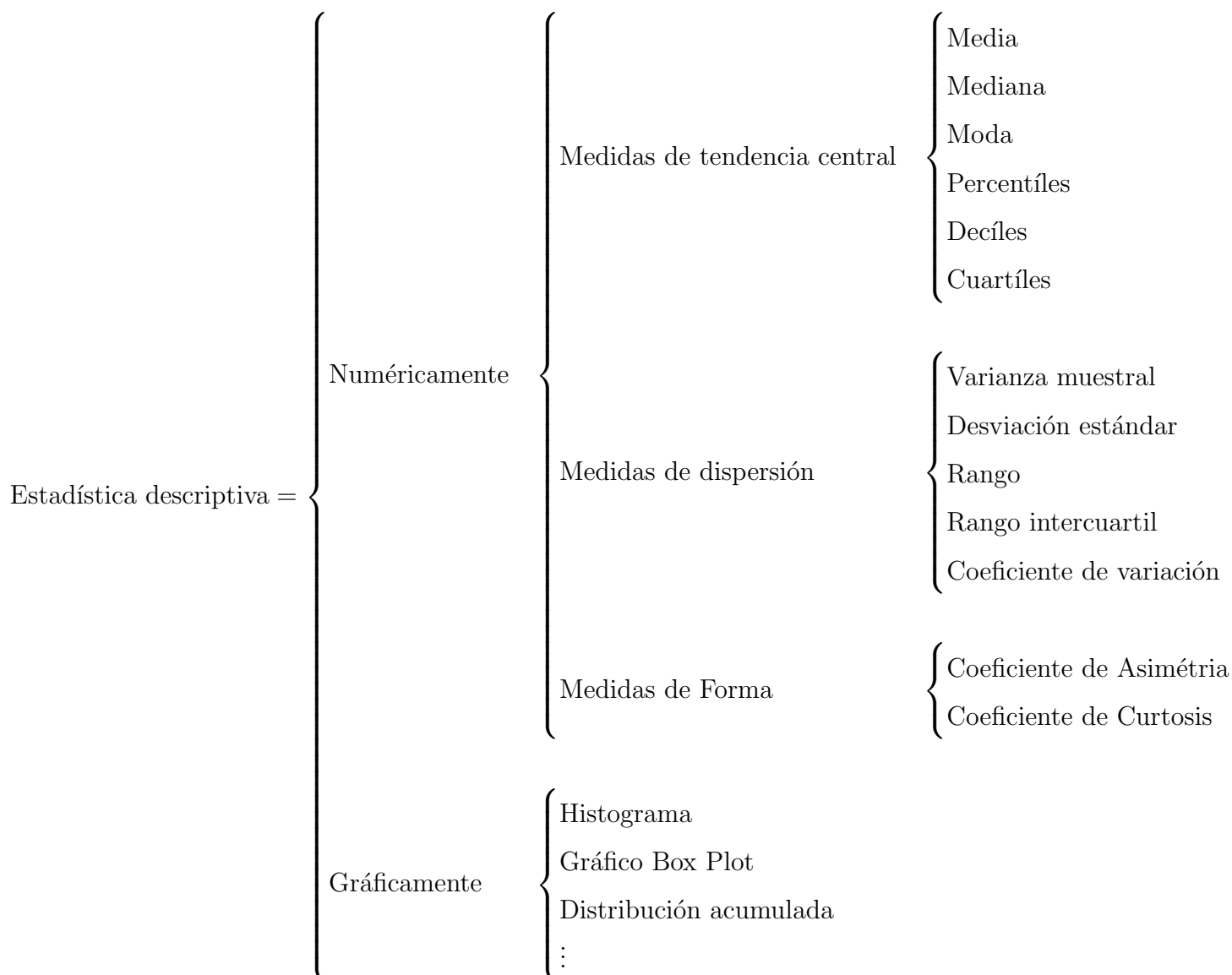
Donde $X_{(1)}$ es la observación mas chica, $X_{(2)}$ es la segunda observación más chica, y así sucesivamente hasta que $X_{(n)}$ representa la observación mayor, es decir:

$$X_{(1)} = \min \{X_1, X_2, \dots, X_n\} \quad ; \quad \dots \quad ; \quad X_{(n)} = \max \{X_1, X_2, \dots, X_n\}$$

Notemos entonces que $X_{(i)}$ con $i \in \{1, \dots, n\}$ son funciones de variables aleatorias y por tanto $X_{(i)}$ también es un variable aleatoria a la cual le podremos calcular su distribución.

1.4. Estadística Descriptiva

La estadística descriptiva tiene como fin presentar resúmenes de un conjunto de datos $\mathbf{X} = \{X_1, X_2, X_3, \dots, X_n\}$ y poner de manifiesto sus características, mediante representaciones numéricas y/o gráficas. Los datos se usan para fines comparativos, y no se usan principios de probabilidad. El interés se centra en describir un conjunto dado de datos y no se plantea el extender las conclusiones a otros datos diferentes o a una población, es decir, solo tiene como fin dar una **descripción** de los datos mediante un resumen



1.4.1. Medidas de tendencia central

Las medidas de tendencia central pretende resumir la tendencia o localización de los datos por medio de un sólo número.

- **Media**, el promedio de los datos : $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$
- **Mediana** \tilde{X} es el valor tal que el 50 % de los datos son menores que él y el 50 % son mayores. Aquí hay que distinguir entre dos casos: Si el tamaño de la muestra n es par entonces:

$$\tilde{X} = \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}$$

Por otro lado si el tamaño de la muestra es impar

$$\tilde{X} = X_{(\frac{n+1}{2})}$$

- **Moda** es el valor o categoría más frecuente (Por lo general es útil para datos categóricos y no funciona para variables continuas)
- El cuantíl o percentil de $\alpha\%$, $P_\alpha\%$ es aquel valor tal que un $\alpha\%$ de los datos son menores a él y un $(1 - \alpha)\%$ de ellos es mayor a él, es decir:

$$\underbrace{\underbrace{X_{(1)}, X_{(2)}, \dots, X_{(p-1)}}_{25\%} \underbrace{X_{(p)}, \dots, X_{(q)}}_{25\%}}_{50\%} \underbrace{\underbrace{X_{(q+1)}, \dots, X_{(s-1)}}_{25\%} \underbrace{X_{(s)}, \dots, X_{(n-1)}, X_{(n)}}_{25\%}}_{75\%}$$

1.4.2. Medidas de dispersion

Las medidas de disepersión pretenden darnos una idea de que tan variables son nuestros datos, se podría decir que sirve para medir el grado de dispersión que tienen los datos que analizamos.

- **Varianza muestral** se define como:

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

- **Desviación estándar** se define como:

$$S = \sqrt{S^2}$$

- **Rango** se define como:

$$R = X_{(n)} - X_{(1)}$$

- **Rango intercuantílico** se define como:

$$RIC = P_{75\%} - P_{25\%}$$

- **Coefficiente de variación**

$$cv = \frac{S}{\bar{X}} \times 100$$

1.4.3. Medidas de Forma

Las medidas de forma pretender darnos una idea de la características de la distribución de la población.

- **Coefficiente de Asimetría** se define como:

$$CA_F = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{nS^3}$$

Cuando $CA_F < 0$ la distribución tiene una asimetría negativa y decimos que los datos están sesgados a la derecha, por otro lado cuando $CA_F > 0$ la distribución tiene una asimetría positiva y decimos que los datos están sesgados a la izquierda. Finalmente cuando $CA_F = 0$ los datos son simétricos.

- **Coefficiente de Curtosis** se define como:

$$Curtosis = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{nS^4} - 3$$

La curtosis (o apuntamiento) es una medida de forma que mide cuán escarpada o achatada está una curva o distribución. Este coeficiente indica la cantidad de datos que hay cercanos a la media, de manera que a mayor grado de curtosis, más escarpada (o apuntada) será la forma de la curva. Cuando $Curtosis > 0$ la distribución tiene un apuntamiento superior a la distribución normal y decimos que la distribución es *leptocurtica* mientras que cuando $Curtosis < 0$, la distribución es *platicurtica*

1.4.4. Medidas de dependencia entre dos muestras

Muchas veces nos enfrentaremos al problema de analizar dos variables de forma simultánea, en cuyo caso será necesario definir métricas que nos ayuden a analizar de forma conjunta a las dos variables. Una pregunta de interés en este caso podría ser identificar si existe algún tipo de

dependencia o relación entre ambas variables, para responder esto se define la covarianza para variables numéricas:

Definición 1.4.1 (Covarianza). *Supongamos que tenemos dos muestras del mismo tamaño, X_1, \dots, X_n , Y_1, \dots, Y_n , definimos la covarianza muestral como:*

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

La covarianza no es mas que una medida de la variabilidad conjunta entre las dos variables para entender mejor este concepto consideremos lo siguiente:

Supongamos que tenemos n observaciones que consta de una variable respuesta Y y X la variable explicativa. Se desea medir la dirección y que tan fuerte es la relación entre Y y X . Si hacemos un gráfico de dispersión entre Y contra X , después trazamos una línea vertical en \bar{x} y una horizontal en \bar{y} . Las dos líneas dividen el gráfico en cuatro cuadrantes. Para cada punto i en el gráfico, calculamos lo siguiente:

- $y_i - \bar{y}$, la desviación de cada observación y_i con respecto a su media muestral \bar{y}
- $x_i - \bar{x}$, la desviación de cada observación x_i con respecto a su media muestral \bar{x}
- El producto de estas dos cantidades, $(y_i - \bar{y})(x_i - \bar{x})$

Es claro que:

Cuadrante	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
1	+	+	+
2	-	+	-
3	-	-	+
4	+	-	-

Luego entonces, si la relación entre Y y X es positiva (cuando X aumenta Y también aumenta), entonces hay mas puntos en el cuadrante 1 y 3 que en el cuadrante 2 y 4 y si la relación entre Y y X es negativa (cuando X aumenta Y disminuye), entonces hay más puntos en los cuadrantes 2 y 4 que en los cuadrantes 1 y 3. Por lo tanto

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

Toma un numero positivo cuando hay relación positiva entre X y Y , toma un valor negativo cuando hay una relación negativa y 0 cuando no hay ningún tipo de relación. En general la covarianza entre X y Y , nos indica la dirección de la relación lineal entre Y y X .

Por desgracia, $Cov(Y, X)$ no es una medida muy informativa de que tan fuerte es la relación, puesto que es afectada por cambios en las unidades de medida. Para evitar esta desventaja de la covarianza, lo que se estila hacer es estandarizar los datos antes de calcular la covarianza de la siguiente forma:

$$\frac{y_i - \bar{y}}{s_y} \quad , \quad \frac{x_i - \bar{x}}{s_x}$$

Donde:

$$s_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}} \quad , \quad s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}.$$

Finalmente, a la covarianza entre las variables estandarizadas la conocemos como coeficiente de correlación entre Y y X y está dada por:

$$r_{XY} = \frac{1}{n - 1} \sum_{i=1}^n \frac{(x_i - \bar{x})}{s_x} \frac{(y_i - \bar{y})}{s_y} = \frac{Cov(X, Y)}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Algunas observaciones importantes que podemos hacer sobre la correlación son:

- r_{XY} mide la correlación lineal entre dos conjuntos de datos $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$.
- Se puede probar que $-1 \leq r_{XY} \leq 1$ (Tarea)
- Si $r_{XY} \approx 1$ ó $r_{XY} \approx -1$ entonces podríamos escribir $y_i \approx \beta_0 + \beta_1 x_i$, para $i = 1, \dots, n$, donde $\beta_0, \beta_1 \in \mathbb{R}$. Más aún, si $r_{XY} \approx -1$ entonces $\beta_1 < 0$, en cambio, si $r_{XY} \approx 1$ entonces $\beta_1 > 0$
- Si $r_{XY} \approx 0$, lo único que podríamos afirmar es que nuestras muestras no guardan ninguna asociación lineal. No podemos afirmar que las muestras sean independientes. El único caso en el que $r_{XY} = 0$ implica independencia es cuando las dos muestras sigan una distribución normal bi-variada (Tarea)

1.4.5. Estadística descriptiva vía gráficas

Sin duda, una de las herramientas mas poderosas de la estadística es la generación de gráficos descriptivos que nos ayuden a visualizar el comportamiento de una población o muestra, a continuación presentamos los gráficos comunmente utilizados en la práctica:

Histograma

Un histograma es una gráfica en forma de barras, donde las bases de las barras son una partición del rango muestral, $R = X_{(n)} - X_{(1)}$, es decir, $X_{(1)} = a_0 < a_1 < \dots < a_{k-1} < a_k = X_{(n)}$ y esto forma las siguientes marcas de clase:

$$[a_0, a_1] (a_1, a_2] (a_2, a_3] \dots (a_{k-2}, a_{k-1}] (a_{k-1}, a_k]$$

Luego, la altura de cada barra es la frecuencia o número de elementos que cae en cada marca del clase. Un histograma se usa cuando se estudia una variable continua, por ejemplo para ver las franjas de edades o alturas de una muestra. Existen varios criterios para determinar el número de marcas de clases o barras, una de ellas es la regla de Sturges que establece que $k = \sqrt{n}$, aunque otras personas recomiendan tomar $k = \log(n) + 1$. Siempre será recomendable experimentar con varios valores de k dependiendo de cómo estén los datos y cuantos sean.

Frecuencia acumulada o distribución empírica acumulada

Otra de las gráficas mas importantes con las que contamos es aquella a la que denominamos distribución empírica, este gráfico tiene como fin aproximar a la verdadera función de distribución de donde vinieron los datos y la construcción es la siguiente: supongamos que tenemos una muestra $X = \{X_1, X_2, X_3, \dots, X_n\}$, ahora al ordenar la muestra tenemos $X_{(1)}, X_{(2)}, X_{(3)}, \dots, X_{(n)}$. Entonces la función de distribución empírica acumulada o frecuencia acumulada (cuando no hay empates) se define como la proporción de los datos menores o iguales a x .

$$F_n(x) = \begin{cases} 0 & \text{si } x \in (-\infty, X_{(1)}) \\ \frac{u}{n} & \text{si } x \in [X_{(u)}, X_{(u+1)}) \text{ y } u \in \{1, \dots, n-1\} \\ 1 & \text{si } x \in [X_{(n)}, \infty) \end{cases} \quad (1.1)$$

Los diagramas de caja

Uno de los gráficos más pulvares para visualizar la forma en como se distribuye una variable es utilizando el denominado diagrama de caja y brazos, esta técnica de análisis exploratorio de datos nos puede servir para

- ver que tan dispersos están los datos

- si hay simetría entre los datos o no
- la detección de valores atípicos

Un diagrama de caja y brazos se elabora como sigue:

1. De la muestra ordenada $X_{(1)}, X_{(2)}, X_{(3)}, \dots, X_{(n)}$ encontramos el primer cuartil $P_{25\%}$ y el tercer cuartil $P_{75\%}$ y dibujamos la caja o rectángulo cuyos extremos son $(P_{25\%}, P_{75\%})$, dentro del rectángulo dibujamos con una línea la posición de la mediana, $P_{50\%}$
2. El límite inferior del brazo, L_i será la primera observación mayor o igual al número

$$P_{25\%} - 1.5RIC$$

donde $RIC = (P_{75\%} - P_{25\%})$

3. El límite superior del brazo, L_s será la primera observación menor o igual al número

$$P_{75\%} + 1.5RIC$$

4. Consideraremos como valores atípicos a los valores situados fuera del intervalo (L_i, L_s)

Ejercicio 1.4.1. ■ Indica qué variables son numéricas y cuales categóricas, además indica la escala de medición :

- Comida Favorita.
- Profesión que te gusta.
- Número de goles marcados por tu equipo favorito en la última temporada.
- Número de alumnos de tu Instituto.
- El color de los ojos de tus compañeros de clase.
- Coeficiente intelectual de tus compañeros de clase.

- Demuestre que $-1 \leq r_{XY} \leq 1$, donde

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

- Demuestre que $|r_{XY}| = 1$ si y solo si existen una relación lineal entre las variables X y Y , es decir, $y_i = \alpha x_i + \beta$ con $\alpha \neq 0$

- Las calificaciones de 50 alumnos en Estadística han sido las siguientes:

0, 1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 6,
6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 10

Calcular las medidas de tendencia central y de dispersión, además encuentre el Percentil 0.25 y 0.75 y elabore un gráfico de caja y brazos.

¿La distribución de estos datos es Sesgada a la Izquierda?

Hint: $n = 50$, $\sum_{i=1}^n x_i = 274$, $\sum_{i=1}^n x_i^2 = 1700$, $\sum_{i=1}^n x_i^3 = 11356$

- Los 40 alumnos de una clase han obtenido las siguientes puntuaciones, sobre 50, en un examen de Física.

48, 47, 44, 42, 41, 39, 39, 38, 38, 38, 37, 36, 36, 35, 35, 34, 34, 34, 33,
32, 32, 31, 29, 28, 28, 27, 26, 25, 24, 23, 22, 20, 17, 15, 15, 13, 13, 11, 7, 3

Calcular las medidas de tendencia central y de dispersión, además encuentre el Cuartil 1 y 3 así como los Deciles 1 y 9. Elabore un gráfico de caja y brazos, así como el gráfico de distribución empírica.

¿ Los datos tiene una distribución platicurtica?

Hint: $n = 40$, $\sum_{i=1}^n x_i = 1169$, $\sum_{i=1}^n x_i^2 = 38929$, $\sum_{i=1}^n x_i^3 = 1388327$, $\sum_{i=1}^n x_i^4 = 51786709$

Capítulo 2

Repaso de Probabilidad

2.1. Funciones de distribución

2.1.1. Introducción

Supongamos que tenemos una muestra de un cierto fenómeno o experimento aleatorio

$$\{X_1, X_2, X_3, \dots, X_n\}$$

Como ya hemos visto una primera aproximación que podemos hacer es un análisis descriptivo de estos datos, sin embargo hagamos un paso mas de abstracción y preguntémonos si existe algún modelo matemático–probabilístico que describa este fenómeno aleatorio.

Resulta entonces interesante revisar los modelos probabilísticos más importantes que existen, para ello recordemos el concepto más importante de la teoría de probabilidad, la variable aleatoria.

2.1.2. Variable Aleatoria

El modelo matemático que explica el comportamiento de los resultados de los experimentos aleatorios está compuesto por dos elementos:

- El espacio de estados o espacio muestral Ω
- Una función de probabilidad, que denotaremos como \mathbb{P} que toma valores en el intervalo $[0, 1]$, esta función asigna probabilidades a los sucesos.

Los elementos que integran el espacio muestral Ω son eventos y aquí se nos dificulta hacer operaciones con estos elementos. Para resolver este problema se recurre a la asignación de números

a los elementos de Ω . Para llevar a cabo las transformaciones de sucesos en números reales se introduce el concepto de variable aleatoria.

Definición 2.1.1 (Variable Aleatoria). *Una variable aleatoria (v.a.) X es una función (medible) que manda elementos de Ω a \mathbb{R} (o un subconjunto de \mathbb{R}), denotado por*

$$X : \Omega \rightarrow \mathbb{R},$$

y al conjunto de los valores reales que puede tomar X le llamamos rango o recorrido. Diremos que una variable aleatoria X es discreta cuando el conjunto de valores que toma es finito o numerable. Por otro lado diremos que una variable aleatoria X es continua cuando los valores que puede tomar es un conjunto no numerable.

Ejemplo 2.1.1. *Pensemos en el experimento de lanzar tres volados con una moneda honesta y queremos calcular la probabilidad de que el número de águilas sea k , obviamente $k \in \{0, 1, 2, 3\}$. Entonces nuestro espacio muestral será:*

$$\Omega = \{(a, a, a), (a, a, s), (a, s, a), (s, a, a), (s, s, a), (s, a, s), (a, s, s), (s, s, s)\}$$

Luego, denamos una variable X como el número de águilas en los tres volados. Entonces el rango de la variable aleatoria, o los valores que puede tomar son $X \in \{0, 1, 2, 3\}$. Y las probabilidades de cada valor es:

$$\mathbb{P}(X = 0) = \mathbb{P}(\{(s, s, s)\}) = \frac{1}{8} \quad (2.1)$$

$$\mathbb{P}(X = 1) = \mathbb{P}(\{(s, s, a)\}) + \mathbb{P}(\{(s, a, s)\}) + \mathbb{P}(\{(a, s, s)\}) = \frac{3}{8} \quad (2.2)$$

$$\mathbb{P}(X = 2) = \mathbb{P}(\{(a, a, s)\}) + \mathbb{P}(\{(a, s, a)\}) + \mathbb{P}(\{(s, a, a)\}) = \frac{3}{8} \quad (2.3)$$

$$\mathbb{P}(X = 3) = \mathbb{P}(\{(a, a, a)\}) = \frac{1}{8} \quad (2.4)$$

Para resumir todas estas asignaciones que hace la variable aleatoria así como las probabilidades asociadas se deben el concepto de **función de distribución**.

Definición 2.1.2 (Función de distribución). *La función de distribución acumulada de una variable aleatoria X evaluada en un real x la denotamos como $F_X(x)$ y representa:*

$$F_X(x) = \mathbb{P}(X \leq x)$$

Notemos que $0 \leq F_X(x) \leq 1$ por ser la probabilidad de un evento.

Propiedades de la función de distribución $F_X(\cdot)$

- $F_X(-\infty) = \mathbb{P}(X \leq -\infty) = 0$ y $F_X(\infty) = \mathbb{P}(X \leq \infty) = 1$
- $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$
- Es monótona no decreciente $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) \geq 0$
- La función de distribución es continua por la derecha.

2.1.3. Tipos de Variables

Variable aleatoria discreta

Una variable aleatoria es discreta cuando su rango es un conjunto finito o numerable de puntos. Cada punto tiene una masa de probabilidad de ocurrir

$$\mathbb{P}(X = x_i) = p_i,$$

a $\mathbb{P}(X = x_i)$ se le conoce como función de masa de probabilidad. Además si sumamos sobre todo el rango de la variable aleatoria los pesos deben ser igual a uno, es decir,

$$\sum \mathbb{P}(X = x_i) = \sum p_i = 1.$$

Entonces para una v.a. discreta la función de distribución $F_X(\cdot)$ queda definida como:

$$F_X(u) = \mathbb{P}(X \leq u) = \sum_{x_i \leq u} \mathbb{P}(X = x_i), \quad \forall u \in \mathbb{R}.$$

Observación: Cuando tenemos una variable aleatoria discreta en general $\mathbb{P}(X \leq u) \neq \mathbb{P}(X < u)$.

Variables aleatorias continuas Una variable aleatoria X es continua si su función de distribución está definida como,

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

donde $f_X(\cdot)$ se le conoce como la función de densidad y cumple con ser:

1. una función continua
2. $f_X(x) \geq 0$
3. $F_X(\infty) = \int_{-\infty}^{\infty} f_X(x) dx = 1$

Algunas observaciones que podemos hacer son las siguientes:

1. Como $f_X(\cdot)$ es continua entonces $F_X(\cdot)$ será también continua.
2. A partir de $F_X(\cdot)$ podemos obtener $f_X(\cdot)$ usando el Teorema Fundamental del Cálculo, puesto que

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

3. El conjunto de puntos donde $f_X(\cdot) > 0$ será el rango o soporte de la variable aleatoria.
4. Notemos que

$$\mathbb{P}(X = x) = \int_x^x f_X(u) du = 0$$

y por lo tanto en este caso

$$\mathbb{P}(X \leq u) = \mathbb{P}(X < u)$$

2.2. Esperanza

A continuación estudiaremos una de las características más importantes de las variables aleatorias.

La esperanza matemática (o sus sinónimos: Esperanza, valor esperado, media poblacional, media, primer momento) de una v.a. X es un promedio ponderado de acuerdo a la distribución teórica de probabilidades del fenómeno estudiado. O también lo podemos ver como el valor hacia el que tendería la media aritmética \bar{x} si se tuvieran un número suficientemente grande de observaciones del fenómeno. La esperanza de una v.a. X , lo denotaremos por $\mathbb{E}(X)$ y lo calcularemos como sigue:

$$\mathbb{E}(X) = \begin{cases} \sum_{x \in \text{Rango}(X)} x \cdot \mathbb{P}(X = x) & \text{si la v.a. } X \text{ es discreta,} \\ \int_{-\infty}^{\infty} x \cdot f_X(x) dx & \text{si la v.a. } X \text{ es continua.} \end{cases}$$

Las propiedades más importantes de la Esperanza son las siguientes:

- Supongamos que k es constante, entonces la esperanza de una constante es igual a la misma constante, es decir $\mathbb{E}(k) = k$.
- Sean X_1, X_2, \dots, X_n , v.a.'s entonces la esperanza matemática de la suma (o resta) de variables aleatorias es igual a la suma (o resta) de las esperanzas de cada una de esas variables

aleatorias, es decir

$$\mathbb{E}(X_1 \pm X_2 \pm \dots \pm X_n) = \mathbb{E}(X_1) \pm \mathbb{E}(X_2) \pm \dots \pm \mathbb{E}(X_n)$$

- Sean X_1, X_2, \dots, X_n , v.a.'s, independientes, entonces la esperanza de un producto de v.a.'s es igual al producto de las esperanzas de cada una de las v.a.'s, si y sólo si son X_1, X_2, \dots, X_n independientes

$$\mathbb{E}(X_1 X_2 \dots X_n) = \mathbb{E}(X_1) \mathbb{E}(X_2) \dots \mathbb{E}(X_n)$$

- Sea X una v.a. y b una constante real, entonces la esperanza de una v.a. más una constante es igual a la esperanza de la v.a. más la constante, es decir:

$$\mathbb{E}(X + b) = \mathbb{E}(X) + b.$$

- Sea X una v.a. y a una constante real, entonces la esperanza matemática de una constante por una v.a. es igual a la constante por la esperanza de la v.a.

$$\mathbb{E}(aX) = a\mathbb{E}(X).$$

Como una consecuencia inmediata de lo anterior diremos que la esperanza es un operador lineal, es decir que abre sumas y saca escalares. Es decir que si X_1, X_2, \dots, X_n son v.a.'s, y a_1, a_2, \dots, a_n, b son escalares, entonces:

$$\mathbb{E}(a_1 X_1 + a_2 X_2 + \dots + a_n X_n + b) = a_1 \mathbb{E}(X_1) + a_2 \mathbb{E}(X_2) + \dots + a_n \mathbb{E}(X_n) + b.$$

Muchas veces estaremos interesados en calcular la esperanza de una función evaluada en una v.a., la definición de esperanza nos obligaría entonces a primero encontrar la distribución de la transformación de la v.a. sin embargo el siguiente teorema nos brinda una manera mas fácil de llevar a cabo el cálculo de la esperanza:

Teorema 2.2.1. *Sea X una v.a. g una función (medible), entonces $g(X)$ también es una variable aleatoria y su esperanza puede ser calculada como sigue:*

$$\mathbb{E}(g(X)) = \begin{cases} \sum_{x \in \text{Rango}(X)} g(x) \mathbb{P}(X = x) & \text{si la v.a. } X \text{ es discreta,} \\ \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{si la v.a. } X \text{ es continua.} \end{cases}$$

En particular si en el teorema anterior, proponemos la transformación $g(X) = X^r$, entonces a la cantidad $\mathbb{E}(g(X)) = \mathbb{E}(X^r)$ se le conoce como el r -ésimo momento con respecto a el origen, y en forma compacta lo denotamos por $\mathbb{E}(X^r) = \alpha_r$.

Un resultado muy importante nos dice que si el momento de orden t existe, entonces todos los momentos de orden inferior existen. En símbolos esto se escribe así:

$$\alpha_t = \mathbb{E}(X^t) < \infty \text{ entonces } \alpha_r = \mathbb{E}(X^r) < \infty, \text{ con } r \leq t.$$

Los momentos con respecto al origen más usados son:

- $\alpha_0 = \mathbb{E}(X^0) = 1$
- $\alpha_1 = \mathbb{E}(X) = \mu$
- $\alpha_2 = \mathbb{E}(X^2)$
- $\alpha_3 = \mathbb{E}(X^3)$
- $\alpha_4 = \mathbb{E}(X^4)$

Por otro lado, si definimos $g(X) = (X - \mu)^r$, donde $\mathbb{E}(X) = \mu$. Entonces a $\mathbb{E}((X - \mu)^r)$ lo llamamos el r -ésimo momento respecto a la media y lo denotaremos por $\mu_r = \mathbb{E}((X - \mu)^r)$.

Los momentos con respecto a la media más usados son:

- $\mu_1 = \mathbb{E}((X - \mu)^1) = 0$
- $\mu_2 = \mathbb{E}((X - \mu)^2)$, así definiremos la varianza
- $\mu_3 = \mathbb{E}((X - \mu)^3)$, se usa para calcular el coeficiente de asimetría poblacional.
- $\mu_4 = \mathbb{E}((X - \mu)^4)$, se usa para calcular la kurtosis poblacional.

Notemos que los momentos con respecto a la media, μ_r , se pueden calcular a partir de los momentos con respecto al origen α_r . Por ejemplo

$$\begin{aligned} \mu_2 &= \mathbb{E}((X - \mu)^2) \\ &= \mathbb{E}(X^2 - 2\mu X + \mu^2) \\ &= \mathbb{E}(X^2) - 2\mu \mathbb{E}(X) + \mu^2 \\ &= \mathbb{E}(X^2) - 2\mu^2 + \mu^2 \\ &= \alpha_2 - \alpha_1^2 \\ &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \end{aligned}$$

De manera análoga podemos calcular $\mu_k = \mathbb{E}((X - \mu)^k)$ usando el binomio de Newton, la única condición que necesitamos es que los momentos con respecto al origen $\alpha_j = \mathbb{E}(X^j)$ para $j \in \{1, 2, \dots, k\}$ existan, que de hecho es equivalente a decir que $\alpha_k = \mathbb{E}(X^k)$ exista.

2.3. Varianza

El segundo momento respecto a la media $\mu_2 = \mathbb{E}((X - \mu)^2)$ lo conoceremos como la varianza y lo denotado por $\text{Var}(X)$ ó σ^2 y es una medida que refleja que tan dispersos esperamos que estén los valores que toma la v.a. con respecto de la media μ .

Propiedades de la varianza

- La varianza siempre es mayor o igual a cero, puesto que estamos calculando la esperanza de la v.a. $(X - \mu)^2 \geq 0$
- La desviación típica o estándar es la parte positiva de la raíz cuadrada de la varianza, y la denotaremos por σ . Esta medida también representa que tanta dispersión hay en la v.a., pero σ está en las mismas unidades que la media μ de la v.a.
- La varianza de una variable aleatoria que no muestra dispersión será cero, es decir es constante.
- Si definimos el Error Cuadrático Medio como, $ECM_u(X) = \mathbb{E}((X - u)^2)$, con $u \in \mathbb{R}$. Es decir, $ECM_u(X)$ representa la dispersión de la v.a. al rededor de algún número real u . Entonces se puede probar que

$$\min_{u \in \mathbb{R}} \mathbb{E}((X - u)^2) = \mathbb{E}((X - \mu)^2) = \text{Var}(X).$$

- Si X es una v.a. con segundo momento y c es una constante real, entonces

$$\text{Var}(cX) = c^2 \text{Var}(X)$$

- Si X es una v.a. con segundo momento y b es una constante real, entonces

$$\text{Var}(X + b) = \text{Var}(X)$$

- Sean X y Y dos v.a. y el primer momento de cada v.a. lo representaremos como sigue:

$\mathbb{E}(X) = \mu_X$ y $\mathbb{E}(Y) = \mu_Y$, entonces como:

$$\begin{aligned}\mathbb{E}(X \pm Y) &= \mathbb{E}(X) \pm \mathbb{E}(Y) \\ &= \mu_X \pm \mu_Y.\end{aligned}$$

Entonces la varianza de la v.a. $X \pm Y$ es por definición:

$$\begin{aligned}\text{Var}(X \pm Y) &= \mathbb{E}(\{(X \pm Y) - (\mu_X \pm \mu_Y)\}^2) \text{ definición} \\ &= \mathbb{E}(\{(X - \mu_X) \pm (Y - \mu_Y)\}^2) \text{ reordenando} \\ &= \mathbb{E}((X - \mu_X)^2) + \mathbb{E}((Y - \mu_Y)^2) \pm 2 \cdot \mathbb{E}((X - \mu_X)(Y - \mu_Y)) \\ &= \text{Var}(X) + \text{Var}(Y) \pm 2 \cdot \text{Cov}(X, Y).\end{aligned}$$

- A la expresión $\mathbb{E}((X - \mu_X)(Y - \mu_Y))$ la conocemos como la covarianza entre las v.a.'s X y Y , y la denotamos por el símbolo $\text{Cov}(X, Y)$ o como σ_{XY}
- Si $\text{Cov}(X, Y) > 0$ significa que cuando la v.a. X crece (disminuye) también la v.a. Y crece (disminuye).
- Si $\text{Cov}(X, Y) < 0$ significa que cuando la v.a. X crece (disminuye) la v.a. Y disminuye (crece).
- Si $\text{Cov}(X, Y) = 0$ entonces $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$

Algunas observaciones de la varianza son las siguiente:

- Si X y Y son v.a. independientes entonces la covarianza es cero pues

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}((X - \mu_X)(Y - \mu_Y)) \text{ por independencia} \\ &= \mathbb{E}(X - \mu_X) \mathbb{E}(Y - \mu_Y) \\ &= 0 \cdot 0 = 0\end{aligned}$$

Y por lo tanto

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$$

Pero es muy importante notar que $\text{Cov}(X, Y) = 0$ **no implica** necesariamente independientes entre las v.a.'s. Para poder afirmar que $\text{Cov}(X, Y) = 0$ implica independencia se debe de añadir la hipótesis de Normalidad conjunta del vector (X, Y) .

El siguiente ejercicio muestra la necesidad de pedir Normalidad Conjunta para poder afirmar la independencia cuando $\text{Cov}(X, Y) = 0$

Ejercicio 2.3.1. Considere X v.a. con función de densidad dada por:

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

Defina además a la v.a. discreta W tal que $\mathbb{P}(W = 1) = \mathbb{P}(W = -1) = \frac{1}{2}$. Suponga que X es independiente de W . Ahora defina la v.a. Y como:

$$Y = WX$$

Demuestre entonces que

- Y sigue la misma distribución de $X \sim \text{Normal}(0, 1)$
- $\text{Cov}(X, Y) = 0$
- X no es independiente de Y

Algunas funciones de distribución discretas

Nombre	Parámetro	$\mathbb{P}(X = x)$	Rango	$\mathbb{E}(X)$	$\text{Var}(X)$
Bernoulli	$p \in (0, 1)$	$p^x (1 - p)^{1-x}$	$x \in \{0, 1\}$	p	$p(1 - p)$
Binomial	$n \in \mathbb{N}, p \in (0, 1)$	$\binom{n}{x} p^x (1 - p)^{n-x}$	$x \in \{0, 1, \dots, n\}$	np	$np(1 - p)$
Binomial Negativa	$p \in (0, 1), k \in \mathbb{N}$	$\binom{x-1}{k-1} p^k (1 - p)^{x-k}$	$x \in \{k, k + 1, \dots\}$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$
Poisson	$\lambda \in (0, \infty)$	$\frac{e^{-\lambda} \lambda^x}{x!}$	$x \in \{0, 1, 2, \dots\}$	λ	λ

Algunas funciones de distribución continuas

Nombre	Parámetro	$f_X(x)$	Rango	$\mathbb{E}(X)$	$\text{Var}(X)$
Uniforme	$a < b, a, b \in \mathbb{R}$	$\frac{1}{b-a}$	$x \in (a, b)$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
Exponencial	$\lambda \in \mathbb{R}^+$	$\lambda e^{-\lambda x}$	$x \in (0, \infty)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma	$\alpha, \beta \in \mathbb{R}^+$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$	$x \in (0, \infty)$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
Beta	$\alpha \in \mathbb{R}^+, \beta \in \mathbb{R}^+$	$\frac{x^{\alpha-1} (1-x)^{\beta-1}}{\text{Beta}(\alpha, \beta)}$	$x \in (0, 1)$	$\frac{\alpha}{\beta + \alpha}$	$\frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$
Normal	$\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$	$x \in \mathbb{R}$	μ	σ^2

Donde:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt; \quad ; \text{Beta}(\alpha, \beta) = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

Dentro de la teoría estadística, surgen otras distribuciones las cuales enunciamos a continuación:

Definición 2.3.1 (Densidad χ^2). *Decimos que X es una v.a. continua con distribución χ^2 con n grados de libertad si su función de densidad está dada por:*

$$f_X(x) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x}; \quad x > 0; \quad n > 0;$$

y lo denotamos como $X \sim \chi_{(n)}^2$. Se puede además probar que:

- $\mathbb{E}(X) = n$
- $\text{Var}(X) = 2n$

Observe que la distribución $\chi_{(n)}^2$ es un caso particular de la densidad *Gamma* con parámetros $\alpha = \frac{n}{2}$ y $\beta = \frac{1}{2}$

Definición 2.3.2 (Densidad t-student). *Decimos que X es una v.a. continua con distribución t-student con n grados de libertad si su función de densidad está dada por:*

$$f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{(n+1)}{2}}; \quad x > 0; \quad n > 0$$

y lo denotamos como $X \sim t_{(n)}$. Se puede además probar que:

- $\mathbb{E}(X) = 0; \quad n > 1$
- $\text{Var}(X) = \frac{n}{n-2}; \quad n > 2$

Definición 2.3.3 (Densidad F-snedecor). *Decimos que X es una v.a. continua con distribución F-snedecor con n_1 y n_2 grados de libertad si su función de densidad está dada por:*

$$f_X(x) = \frac{\sqrt{\frac{(n_1 x)^{n_1} n_2^{n_2}}{(n_1 x + n_2)^{n_1 + n_2}}}}{x \text{Beta}\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \quad x > 0; \quad n_1 > 0; \quad n_2 > 0$$

y lo denotamos como $X \sim F_{(n_1, n_2)}$. Se puede además probar que:

- $\mathbb{E}(X) = \frac{n_2}{n_2-2}; \quad n_2 > 2$
- $\text{Var}(X) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)} \quad n_2 > 4$

Diremos que la v.a. X sigue cierta una distribución $F_X(x)$, de las siguientes formas

- $X \sim f_X(x)$
- $X \sim F_X(x)$
- $X \sim$ nombre de la v.a. y sus parámetros

Ejercicio 2.3.2. ■ Sea X una v.a. continua tal que tiene por función de densidad dada por $f_X(x) = a + bx$ si $x \in [-1, 1]$ y 0 fuera de dicho intervalo. Se pide:

- Calcular a y b sabiendo que $\mathbb{E}(X) = \frac{1}{6}$
- Calcular la varianza de X
- Calcular la distribución de la variable X
- Encontrar la esperanza de X^2 y la $\text{Var}(|X|^{\frac{1}{2}})$

- Encuentre la esperanza y varianza para el modelo Binomial Negativo con función de masa de probabilidad dada por:

$$\mathbb{P}(X = x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}; \quad x \in \{k, k+1, k+2, \dots\}$$

- Encuentre la esperanza y varianza para el modelo Gamma con función de densidad dada por:

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

2.3.1. Función Generadora de Momentos

Supongamos que tenemos 2 variables aleatorias X y Y tales que se cumple que:

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(Y) \\ \mathbb{E}(X^2) &= \mathbb{E}(Y^2) \\ \mathbb{E}(X^3) &= \mathbb{E}(Y^3) \\ &\vdots \\ \mathbb{E}(X^k) &= \mathbb{E}(Y^k) \end{aligned}$$

Si ocurre lo anterior para un k muy grande pareciera mucha coincidencia y uno empezaría a sospechar que posiblemente X y Y tienen la misma distribución. ¿Pero cómo calculo todos los

momentos de una distribución?, precisamente aquí es donde entra la necesidad de definir una función que nos ayude a generar los momentos de las distribuciones que estamos estudiando.

Definición 2.3.4 (Función Generadora de Momentos F.G.M). *Sea X una v.a. aleatoria, tal que $\mathbb{E}(e^{tX}) < \infty$, definimos la función generadora de momentos para X como:*

$$M_X(t) = \mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

Pero, ¿por qué le decimos función generadora de momento? Notemos lo siguiente, sabemos utilizando el polinomio de Taylor que:

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

Por lo tanto :

$$M_X(t) = \mathbb{E}(e^{tX}) = \mathbb{E}\left(\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right) = \sum_{n=0}^{\infty} \mathbb{E}\left(\frac{(tX)^n}{n!}\right) = \sum_{n=0}^{\infty} \frac{t^n}{n!} m_n$$

Donde $m_n = \mathbb{E}(X^n)$ es el momento de orden n de la variable aleatoria X . Entonces:

$$\begin{aligned} \left. \frac{\partial}{\partial t} M_X(t) \right|_{t=0} &= \sum_{n=1}^{\infty} \frac{t^{n-1}}{(n-1)!} m_n \Big|_{t=0} = m_1 = \mathbb{E}(X) \\ \left. \frac{\partial^2}{\partial t^2} M_X(t) \right|_{t=0} &= \sum_{n=2}^{\infty} \frac{t^{n-2}}{(n-2)!} m_n \Big|_{t=0} = m_2 = \mathbb{E}(X^2) \\ &\vdots \\ \left. \frac{\partial^k}{\partial t^k} M_X(t) \right|_{t=0} &= \sum_{n=k}^{\infty} \frac{t^{n-k}}{(n-k)!} m_n \Big|_{t=0} = m_k = \mathbb{E}(X^k) \end{aligned}$$

Luego entonces, la función $M_X(t)$ es tal que su k -ésima derivada evaluada en 0 genera el k -ésimos momento de de la variable aleatoria X . Por lo tanto si resulta que dos variables aleatorias es tal que $M_X(t) = M_Y(t)$ entonces se concluye que ambas variables tendrán todos los momentos iguales, lo que nos haría sospechar que siguen la misma distribución. Esto se ve reflejado en el siguiente teorema.

Teorema 2.3.1. *Sea X y Y dos variables aleatorias tales que $M_X(t) = M_Y(t)$ entonces, X y Y*

tienen la misma distribución lo que denotamos como:

$$X \stackrel{d}{=} Y$$

Ejemplo 2.3.1. Supongamos que tenemos $X \sim N(0, 1)$. Entonces:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = e^{\frac{1}{2}t^2}$$

Luego si derivamos y valuando en cero verificamos que en efecto genera los primeros momentos de la distribución normal.

$$\begin{aligned} \left. \frac{\partial}{\partial t} M_X(t) \right|_{t=0} &= \left. e^{\frac{1}{2}t^2} t \right|_{t=0} = 0 = \mathbb{E}(X) \\ \left. \frac{\partial^2}{\partial t^2} M_X(t) \right|_{t=0} &= \left. e^{\frac{1}{2}t^2} + e^{\frac{1}{2}t^2} t^2 \right|_{t=0} = 1 = \mathbb{E}(X^2) \end{aligned}$$

Algunas propiedades del F.G.M son la siguientes:

- Si X y Y son independientes entonces:

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

- Si $a \in \mathbb{R}$ y X es variable aleatoria entonces:

$$M_{aX}(t) = M_X(at)$$

- Si $a, b \in \mathbb{R}$ y X es variable aleatoria entonces:

$$M_{aX+b}(t) = e^{tb} M_X(at)$$

Ejercicio 2.3.3. ■ Sea $X \sim N(\mu, \sigma^2)$ muestre:

$$M_X(t) = e^{\frac{1}{2}t^2\sigma^2 + t\mu}$$

- Sea $X \sim \text{Exp}(\lambda)$ muestre:

$$M_X(t) = \frac{\lambda}{\lambda - t}$$

Observe entonces que por el ejercicio anterior y el teorema (2.3.1) sabemos que si encontramos

Z una v.a. tal que

$$M_Z(t) = \int_{-\infty}^{\infty} e^{tz} f_Z(z) dz = e^{\frac{1}{2}t^2\sigma^2 + t\mu}$$

Entonces podríamos concluir que Z sigue una distribución Normal de parámetros μ , σ^2 .

Ejercicio 2.3.4. ■ Utilizando la función generadora de momentos pruebe que si $U \sim U(0, 1)$ entonces $Y = -\frac{1}{\lambda} \log U$ sigue una distribución exponencial

■ Sea $X \sim \text{Gamma}(\alpha, \beta)$ con función de densidad dada por:

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

1. Pruebe que la función generadora de momentos de X está dada por:

$$M_X(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha} \quad t < \beta$$

2. Obtenga la función generadora de momentos de la variable aleatoria $\chi_{(k)}^2$ Chi-cuadrado con k grados de libertad. (Recuerde que la densidad χ^2 es un caso especial de la densidad gamma).

3. Sea $X_i \sim \chi_{k_i}^2$ con $i \in \{1, \dots, n\}$ independientes. Defina $Y := X_1 + \dots + X_n$, pruebe entonces (usando la función generadora de momentos) que $Y \sim \chi_{(\sum_{i=1}^n k_i)}^2$

4. Obtenga la función generadora de momentos de la variable aleatoria $\text{Exp}(\lambda)$. (Recuerde que la densidad $\text{Exp}(\lambda)$ es un caso especial de la densidad gamma).

5. Sea $X_i \sim \text{Exp}(\lambda)$ con $i \in \{1, \dots, n\}$ independientes. Defina $Y := X_1 + \dots + X_n$, pruebe entonces (usando la función generadora de momentos) que $Y \sim \text{Gamma}(n, \lambda)$

■ Sea $X \sim \text{Binomial}(n, p)$ con función de densidad de masa de probabilidad dada por:

$$f_X(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x \in \{0, 1, 2, \dots, n\} \quad p \in (0, 1)$$

1. Pruebe que la función generadora de momentos de X está dada por:

$$M_X(t) = (1 - p + pe^t)^n$$

2. Obtenga la función generadora de momentos de la variable aleatoria $\text{Bernulli}(p)$. (Recuerde que la densidad $\text{Bernulli}(p)$ es un caso especial de la densidad Binomial).

3. Sea $X_i \sim \text{Bernulli}(p)$ con $i \in \{1, \dots, n\}$ independientes. Defina $Y := X_1 + \dots + X_n$, pruebe entonces (usando la función generadora de momentos) que $Y \sim \text{Binomial}(n, p)$

- Sea X_1, \dots, X_n variables aleatorias independientes con distribución $N(\mu, \sigma^2)$, pruebe que:

$$\sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$$

2.3.2. Función Característica

Un problema que tiene la función generadora de momentos es que no siempre existe y por tanto no siempre podrá ser utilizada para encontrar la distribución de transformaciones de variables aleatorias. Surge entonces la necesidad de definir una función que de igual manera nos ayude a identificar a las distribuciones.

Definición 2.3.5 (Función Característica). Sea X una v.a. aleatoria, definimos la función característica de momentos para X como la función compleja dada por

$$\phi_X(t) = \mathbb{E}(e^{itX}) = \mathbb{E}(\cos(tX) + i\sin(tX)) = \mathbb{E}(\cos(tX)) + i\mathbb{E}(\sin(tX))$$

En este curso no utilizaremos tanto a esta función por lo que no desarrollaremos la teoría que requiere el cálculo de las funciones características sin embargo el uso y aplicación que tiene es bastante grande, de hecho el famoso Teorema Central de Limite basa su demostración en esta función para concluir sus resultados.

Ejercicio 2.3.5. Suponga que X es v.a. simétrica, es decir que su función de densidad es simétrica respecto de 0, esto es:

$$f_X(x) = f_X(-x)$$

Demuestre que $\phi_X(t)$ no tiene parte imaginaria.

2.4. Vectores Aleatorios

Supongamos ahora que tenemos un experimento aleatorio que arroja un par de valores aleatorios, por ejemplo, lanzamos un dardo y el resultado de nuestro experimento será las coordenadas (x, y) donde cayó el dardo, en este caso ahora el modelo del experimento será lo que denominamos **vector aleatorio**.

Definición 2.4.1 (Vector Aleatorio). *Formalmente un vector aleatorio es un función (medible) $\underline{X} : \Omega \rightarrow \mathbb{R}^n$. Sin embargo se puede probar que $\underline{X} := (X_1, \dots, X_n)^T$ es vector aleatorio si cada X_i es variable aleatoria. (Observación, en esta parte asumiremos que tenemos vectores columna)*

En este caso podrán existir vectores aleatorios absolutamente continuos o discretos pero también puede existir el caso donde unas variables sean continuas y la otras discretas en cuyo caso decimos que el vector es mixto. En este curso solo trabajaremos con vectores discretos y continuos dejando a los vectores mixtos fuera de nuestro estudio.

Las definiciones de distribución se generaliza facilmente:

Definición 2.4.2 (Distribución de un vector aleatorio). *Sea $\underline{X} = (X_1, \dots, X_n)^T \in \mathbb{R}^n$ un vector aleatorio, definimos la función de distribución del vector aleatorio como:*

$$F_{\underline{X}}(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

La función de distribución cumple con las siguientes propiedades:

- Es monótona no decreciente en cada entrada.
- $0 \leq F_{\underline{X}}(x_1, \dots, x_n) \leq 1$
- $\lim_{x_i \rightarrow -\infty} F_{\underline{X}}(x_1, \dots, x_n) = 0 \quad i \in \{1, 2, \dots, n\}$
- $\lim_{\underline{x} \rightarrow \infty} F_{\underline{X}}(x_1, \dots, x_n) = 1$
- Es continua por la derecha en cada variable
- Se satisface la ley del paralelogramo, que en el caso de dos variables es lo siguiente: Para cualesquiera $a < b, c < d$

$$F_{\underline{X}}(b, d) - F_{\underline{X}}(a, d) - F_{\underline{X}}(b, c) + F_{\underline{X}}(a, c) \geq 0$$

- A partir de la función de distribución conjunta se pueden obtener las funciones de distribución marginales de la siguiente manera:

$$F_{X_i}(x_i) = \lim_{x_1 \rightarrow \infty} \dots \lim_{x_{i-1} \rightarrow \infty} \lim_{x_{i+1} \rightarrow \infty} \dots \lim_{x_n \rightarrow \infty} F_{\underline{X}}(x_1, \dots, x_n)$$

A partir de la función de distribución conjunta podemos definir la independencia de variables aleatorias como sigue:

Definición 2.4.3 (Independencia de Variables Aleatorias). *Se dice que las variables X y Y son independientes si:*

$$F_{(X,Y)}(x, y) = F_X(x) F_Y(y) \quad \text{para todo } x, y \in \mathbb{R}$$

El concepto de la función de densidad se generaliza como sigue:

Definición 2.4.4 (Función de Densidad). *Sea $\underline{X} = (X_1, \dots, X_n)$ un vector aleatorio, supongamos que el vector está formado por variables aleatorias continuas, decimos que $f_{\underline{X}}(x_1, \dots, x_n) > 0$ es función de densidad del vector si:*

$$F_{\underline{X}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f_{\underline{X}}(t_1, \dots, t_n) dt_1 \dots dt_n$$

De aquí queda claro usando el T.F.C que:

$$f_{\underline{X}}(x_1, \dots, x_n) = \frac{\partial}{\partial x_1 \partial x_2 \dots \partial x_n} F_{\underline{X}}(x_1, \dots, x_n)$$

Por otro lado si el vector contiene variables aleatorias discretas entonces, la función de densidad de probabilidad del vector se define como:

$$f_{\underline{X}}(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

De esto se prueba entonces que cuando el vector es discreto que:

$$F_{\underline{X}}(x_1, \dots, x_n) = \sum_{\{i_1: x_{i_1} \leq x_1\}; \dots; \{i_n: x_{i_n} \leq x_n\}} f_{\underline{X}}(x_{i_1}, \dots, x_{i_n})$$

Con la definición de independencia y de función de densidad podemos probar que X y Y son independientes si y solo si:

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

Esto se generaliza para n variables fácilmente diciendo que X_1, \dots, X_n son variables aleatorias independientes si y solo si:

$$f_{\underline{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n) = \prod_{i=1}^n f_{X_i}(x_i)$$

Además se pueden encontrar la funciones de densidad marginales integrando o sumando respecto

al resto de las variables como se muestra a continuación:

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{\underline{X}}(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n$$

Ejercicio 2.4.1. ■ Supongamos que las variables (X, Y) pueden tomar los valores

$$(0, 0); (1, 1); (-1, 1); (1, -1); (-1, -1)$$

cada uno con probabilidad $\frac{1}{5}$. Encuentre la función de distribución y determine si estas variables son independientes

- Supongamos que las variables (X, Y) pueden tomar los valores $(1, 1); (2, 1); (1, 2); (3, 1)$ tal que:

$$\mathbb{P}(X = i, Y = j) = C(i + j)$$

Donde C es una constante, determine el valor de C y obtenga la función de densidad de probabilidad marginal correspondiente a la primera variable.

- Suponga que

$$\mathbb{P}(X = i, Y = j) = C\alpha^i\beta^j \quad i, j \in \mathbb{N}; \quad 0 < \alpha, \beta < 1$$

Halle el valor de C para que $\mathbb{P}(X = i, Y = j)$ sea una función de densidad de probabilidad.

- Se lanzan dos tetraedros con caras numeradas del 1 al 4.
 - Escriba el espacio muestral Ω de este experimento
 - Sea X la v.a que indica el número obtenido en el primer tetraedro y Y la v.a. que indica el mínimo de las dos caras obtenidas, encuentre la función de densidad de probabilidad $\mathbb{P}(X = x, Y = y)$ con $x \in \{1, 2, 3, 4\}$, $y \in \{1, 2, 3, 4\}$. Finalmente encuentre la función de distribución $F_{XY}(x, y)$.

2.4.1. Probabilidades Condicionales

Dentro del estudio de probabilidades surgió la necesidad de medir la probabilidad de los eventos dado que ocurrió otro cierto evento, a esto se le llama una medida de probabilidad condicional y se define como sigue:

Definición 2.4.5 (Probabilidad Condicional). Sea A y B dos evento tal que $\mathbb{P}(B) > 0$, definimos la probabilidad condicional como:

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

Ahora supongamos que tenemos dos variables aleatorias discretas X y Y , se desea obtener la función de densidad de probabilidad para la variable aleatoria X condicionada a que ocurrió el evento $Y = y$ (Suponga que $\mathbb{P}(Y = y) > 0$), resulta natural entonces definir a dicha función como:

$$f_{X|Y}(x|y) = \frac{\mathbb{P}(X = x \cap Y = y)}{\mathbb{P}(Y = y)} = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Finalmente este último resultado se extiende para variables aleatorias continuas de la misma forma

Definición 2.4.6 (Densidad Condicional). Sean X y Y v.a. aleatorias, definimos la función de densidad condicional de X dado $Y = y$ tal que $f_Y(y) > 0$ como:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Observe que cuando X y Y son independientes se tiene que:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_X(x) f_Y(y)}{f_Y(y)} = f_X(x)$$

Generalizando lo anterior se puede probar que la función de distribución condicional se puede obtener sumando o integrando a la respectiva función de densidad condicional, en el caso continuo

$$F_{X|Y}(x|y) = \mathbb{P}(X \leq x | Y = y) = \int_{-\infty}^x f_{X|Y}(t|y) dt$$

Mientras que en el caso discreto:

$$F_{X|Y}(x|y) = \mathbb{P}(X \leq x | Y = y) = \sum_{\{i: x_i \leq x\}} f_{X|Y}(x_i|y)$$

Por otro lado el concepto de esperanza también puede ser extendido al caso condicional el cual se define de forma natural como:

Definición 2.4.7 (Esperanza Condicional evaluada). Sea X y Y variables aleatorias tal que $f_Y(y) > 0$ entonces definimos la esperanza condicional de X dado $Y = y$ como:

Si X y Y son continuas:

$$\mathbb{E}(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$$

Si X y Y son discretas:

$$\mathbb{E}(X|Y = y) = \sum_{x \in \text{Rango}(X)} x f_{X|Y}(x|y)$$

Se podrá además probar que la esperanza condicional de una transformación de variables aleatorias está dado por:

$$\mathbb{E}(h(X)|Y = y) = \int_{-\infty}^{\infty} h(x) f_{X|Y}(x|y) dx$$

A partir de la varianza condicional podremos definir a la varianza condicional:

Definición 2.4.8 (Varianza Condicional evaluada). Sea X y Y variables aleatorias tal que $f_Y(y) > 0$ entonces definimos la varianza condicional de X dado $Y = y$ como:

$$\text{Var}(X|Y = y) = \mathbb{E}(X^2|Y = y) - (\mathbb{E}(X|Y = y))^2$$

La interpretación que daremos a $\mathbb{E}(X|Y = y)$ es nuevamente la de un promedio ponderado y nos indica el valor que en promedio estará obteniendo la variable aleatoria X condicionada a que la v.a. Y en todas las repeticiones siempre tomó el valor de $Y = y$. Por otro lado $\text{Var}(X|Y = y)$ se interpreta como una medida de la variabilidad que tiene la variable X condicionada a que la v.a. Y tomó el valor de $Y = y$.

Observe además que $\mathbb{E}(X|Y = y)$ es una función de y , muchas veces escribimos entonces $\mathbb{E}(X|Y = y) = g(y)$, sin embargo a veces será necesario calcular la esperanza de X condicionada a Y sin evaluar a la variable aleatoria, en este caso también vamos a poder definir una esperanza condicional de la siguiente manera:

Definición 2.4.9 (Esperanza Condicional). Sea X y Y variables aleatorias tal que existe

$$\mathbb{E}(X|Y = y) = g(y)$$

$\forall y$ en el soporte de Y , definimos a $\mathbb{E}(X | Y)$ como a la variable aleatoria dada por:

$$\mathbb{E}(X | Y) = g(Y)$$

Definición 2.4.10 (Varianza Condicional). Sea X y Y variables aleatorias tal que existe

$$\text{Var}(X|Y = y) = g(y)$$

$\forall y$ en el soporte de Y , definimos a $\text{Var}(X|Y)$ como a la variable aleatoria dada por:

$$\text{Var}(X|Y) = g(Y)$$

Dado que $\mathbb{E}(X|Y)$ es una v.a. resulta entonces interesante también poder calcularle sus momentos, en este caso se prueba que:

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(\mathbb{E}(X|Y)) \\ \text{Var}(X) &= \mathbb{E}(\text{Var}(X|Y)) + \text{Var}(\mathbb{E}(X|Y))\end{aligned}$$

Ejercicio 2.4.2. ■ *Considera la siguientes densidad conjunta:*

$$f(x, y) = \frac{1}{8y} e^{-\left(\frac{x}{2y} + \frac{y}{4}\right)} \quad x, y > 0$$

Encuentra la densidad marginal $f_Y(y)$ y la densidad condicional $f_{X|Y}(x|y)$, finalmente encuentre $\mathbb{E}(Y)$ y $\text{Var}(Y)$ así como $\mathbb{E}(X|Y)$ y $\text{Var}(X|Y)$

■ *Considera la siguiente densidad conjunta:*

$$f(x, y) = 4xye^{-x^2-y^2} \quad x, y > 0$$

Encuentre las marginales $f_X(x)$, $f_Y(y)$ así como las condicionales $f_{X|Y}(x|y)$, $f_{Y|X}(y|x)$, finalmente encuentre $\mathbb{E}(Y)$ y $\text{Var}(Y)$ así como $\mathbb{E}(X|Y)$ y $\text{Var}(X|Y)$

■ *Considera la densidad Normal Multivariada.*

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right)$$

Prueba que la densidad marginal $f_X(x)$ corresponde a la densidad Normal de parámetros $\mu = \mu_1$ y $\sigma^2 = \sigma_1^2$, $N(\mu_1, \sigma_1^2)$

Finalmente encuentre la densidad condicional $f_{Y|X}(x|y)$ y prueba que corresponde a una densidad $N(\mu = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \sigma^2 = \sigma_2^2(1 - \rho^2))$

■ *Sea (X, Y) un vector aleatorio con función de densidad dada por:*

$$f_{XY}(x, y) = 2e^{-\frac{y}{x}} \quad 0 \leq x \leq 1; \quad y > 0$$

Encuentre $\mathbb{E}(Y)$ y $\text{Var}(Y)$ sin utilizar la densidad de Y .

2.5. Teorema de Cambio de Variable

Dentro del análisis de la muestra que haremos en nuestro estudio será necesario llevar a cabo transformaciones de variables aleatorias y por tanto debemos conocer herramientas que nos ayuden a encontrar las distribuciones de dichas transformaciones. Como ya vimos una forma de atacar el problema es utilizar la función generadora de momentos para deducir la distribución de la transformación, sin embargo a veces dicha técnica no funcionará.

El caso mas sencillo es el siguiente, supongamos que tenemos una v.a. X y construimos Y otra v.a. tal que Y es una transformación monótona (creciente o decreciente) de X , es decir:

$$Y = g(X)$$

Suponiendo X continua sabemos que Y también será continua por lo tanto podemos preguntarnos por encontrar a $F_Y(y)$ el cual nos dice el comportamiento distribucional de Y .

$$F_Y(y) := \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

Luego entonces, si queremos a la función de densidad de Y bastaría con derivar respecto a y a la función de distribución y entonces obtener:

$$f_Y(y) = \frac{\partial}{\partial y} F_Y(y) = \frac{\partial}{\partial y} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \frac{\partial}{\partial y} g^{-1}(y)$$

en la última igualdad tenemos el problema que no necesariamente $\frac{\partial}{\partial y} g^{-1}(y)$ es positivo, (pues depende de si g es creciente o decreciente) luego entonces para garantizar que $f_Y(y)$ sea densidad debemos colocar valor absoluto

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{\partial}{\partial y} g^{-1}(y) \right|$$

Con la formula anterior es fácil por ejemplo , obtener la densidad de transformaciones lineales de variables aleatorias.

Ejemplo 2.5.1. Sea $X \sim N(\mu, \sigma^2)$, muestre que $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

$$F_Z(z) = F_X(z\sigma + \mu)$$

Derivando:

$$f_Z(z) = f_X(z\sigma + \mu) \sigma = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{1}{2\sigma^2}(z\sigma + \mu - \mu)^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Esté método también puede ser utilizado en funciones que son invertibles por partes, esto lo podemos ver en el siguiente ejemplo:

Ejemplo 2.5.2. Sea $Z \sim N(0, 1)$, muestre que $Y = Z^2 \sim \chi_1^2$. Observemos primero que Y es una variable aleatoria que solo puede tomar valores positivos

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(Z^2 \leq y) = \mathbb{P}(|Z| \leq \sqrt{y}) = \mathbb{P}(-\sqrt{y} \leq Z \leq \sqrt{y}) \\ &= \mathbb{P}(Z \leq \sqrt{y}) - \mathbb{P}(Z < -\sqrt{y}) = F_Z(\sqrt{y}) - F_Z(-\sqrt{y}) \end{aligned}$$

Derivando obtenemos que la densidad de Y está dada por:

$$\begin{aligned} f_Y(y) &= \frac{\partial}{\partial y} (F_Z(\sqrt{y}) - F_Z(-\sqrt{y})) = f_Z(\sqrt{y}) \frac{1}{2\sqrt{y}} + f_Z(-\sqrt{y}) \frac{1}{2\sqrt{y}} \\ &= f_Z(\sqrt{y}) \frac{1}{\sqrt{y}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \frac{1}{\sqrt{y}} = \frac{1}{2^{\frac{1}{2}} \Gamma(\frac{1}{2})} y^{\frac{1}{2}-1} e^{-\frac{1}{2}y} \end{aligned}$$

Lo que demuestra que Y sigue una distribución $\chi_{(1)}^2$.

El teorema de cambio puede generalizarse para transformaciones de \mathbb{R}^n a \mathbb{R}^n invertibles de la siguiente forma:

Teorema 2.5.1 (Teorema de Cambio de Variable - Caso Absolutamente Continuo). Sea $\underline{X} = (X_1, \dots, X_n)^T$ un vector aleatorio con función de densidad dada por $f_{\underline{X}}(x_1, \dots, x_n)$. Sea $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ una transformación invertible.

$$g(x_1, \dots, x_n) = (y_1 = g_1(x_1, \dots, x_n), \dots, y_n = g_n(x_1, \dots, x_n))$$

Como g es invertible entonces sabemos existen w_1, \dots, w_n funciones de \mathbb{R}^n a \mathbb{R} tales que:

$$\begin{aligned} x_1 &= w_1(y_1, \dots, y_n) \\ x_2 &= w_2(y_1, \dots, y_n) \\ &\vdots \\ x_n &= w_n(y_1, \dots, y_n) \end{aligned}$$

Definamos al vector aleatorio $\underline{Y} = (Y_1, \dots, Y_n)^T$ obtenido de transformar a \underline{X} por medio de g , es decir $\underline{Y} = g(\underline{X})$. Entonces la densidad conjunta del vector aleatorio \underline{Y} está dado por:

$$f_{\underline{Y}}(y_1, \dots, y_n) = f_{\underline{X}}(w_1(y_1, \dots, y_n), \dots, w_n(y_1, \dots, y_n)) |\det J|$$

Donde J es la matriz jacobiana.

$$(J)_{ij} = \frac{\partial}{\partial y_j} w_i$$

Consideremos por ejemplo una transformación lineal de \mathbb{R}^3 a \mathbb{R}^3 dada por:

$$g(x_1, x_2, x_3) = (3x_1, x_1 - 4x_2, x_3)^T$$

Ahora supongamos que tenemos un vector aleatorio continuo en R^3 con función de densidad $f_{\underline{X}}(x_1, x_2, x_3)$, definamos al vector aleatorio $\underline{Y} = (Y_1, Y_2, Y_3)^T$, tal que:

$$\underline{Y} = g(\underline{X}) = (3X_1, X_1 - 4X_2, X_3)^T$$

Se desea encontrar la función de densidad del vector \underline{Y} . Esto lo podemos resolver utilizando el teorema de cambio de variable muy fácilmente, primero notemos que al ser g una transformación lineal, entonces existe una matriz que construye a dicha transformación lineal.

$$\underline{Y} = g(\underline{X}) = A\underline{X} = (3X_1, X_1 - 4X_2, X_3)^T$$

Con A una matriz de la forma:

$$A = \begin{pmatrix} 3 & 0 & 0 \\ 1 & -4 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Observe que como A es de rango completo, entonces existe su inversa lo cual era de esperarse al ser g invertible, luego entonces:

$$\underline{X} = A^{-1}\underline{Y} = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ \frac{1}{12} & -\frac{1}{4} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \left(\frac{1}{3}Y_1, \frac{1}{12}Y_1 - \frac{1}{4}Y_2, Y_3 \right)^T$$

En este caso entonces las funciones w_i están dadas por:

$$\begin{aligned} x_1 &= w_1(y_1, y_2, y_3) = \frac{1}{3}y_1 \\ x_2 &= w_2(y_1, y_2, y_3) = \frac{1}{12}y_1 - \frac{1}{4}y_2 \\ x_3 &= w_3(y_1, y_2, y_3) = y_3 \end{aligned}$$

De donde queda claro que la matriz Jacobiana está dada precisamente por A^{-1} , luego entonces,

por el teorema de cambio de variable la densidad del vector aleatorio Y está dado por:

$$f_Y(y_1, \dots, y_n) = f_X\left(\frac{1}{3}y_1, \frac{1}{12}y_1 - \frac{1}{4}y_2, y_3\right) |\det J| = f_X\left(\frac{1}{3}y_1, \frac{1}{12}y_1 - \frac{1}{4}y_2, y_3\right) \frac{1}{12}$$

Ejercicio 2.5.1. Sea $\underline{Z} = (Z_1, \dots, Z_n)^T$ un vector aleatorio continuo tal que:

$$f_{\underline{Z}}(z_1, \dots, z_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n z_i^2\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2} \underline{z}^T \underline{z}\right)$$

Sea Σ una matriz definida positiva y $\underline{\mu} \in \mathbb{R}^p$ un vector de números reales. Demuestre que el vector aleatorio $\underline{X} = \Sigma^{\frac{1}{2}} \underline{Z} + \underline{\mu}$ tiene por densidad:

$$f_{\underline{X}}(x_1, \dots, x_n) = f_{\underline{X}}(\underline{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \det\left(\Sigma^{-\frac{1}{2}}\right) \exp\left(-\frac{1}{2} (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu})\right)$$

Observacion: Cuando un vector aleatorio \underline{X} tiene la densidad anterior, decimos que \underline{X} sigue una distribución multivariada y lo denotamos por:

$$\underline{X} \sim N_n(\underline{\mu}, \Sigma)$$

Hint: Recuerde que como Σ es definida positiva entonces puede ser expresada según la descomposición espectral como:

$$\Sigma = \Gamma \Delta \Gamma^T$$

Luego entonces defina:

$$\Sigma^\alpha = \Gamma \Delta^\alpha \Gamma^T \quad \alpha \in \mathbb{R}$$

Donde:

$$\Delta^\alpha = \text{diag}(\lambda_1^\alpha, \dots, \lambda_n^\alpha)$$

Ahora utilizaremos el teorema de cambio de variable para obtener uno de los resultados mas utilizados en la teoría de la estadística:

Teorema 2.5.2. Sea $Z \sim N(0, 1)$ y $X \sim \chi_{(n)}$, suponga Z y X independientes. Entonces:

$$T = \frac{Z}{\sqrt{\frac{X}{n}}} \sim t_{(n)} \quad (t - \text{student})$$

Demostración. Por la definición (2.3.2), tenemos que demostrar que:

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{(n+1)}{2}}; \quad t > 0; \quad n > 0$$

Como $Z \sim N(0, 1)$ y $X \sim \chi_{(n)}$ por independencia tendríamos que:

$$f_{XZ}(x, z) = f_X(x) f_Z(z) = \frac{x^{\frac{n}{2}-1}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right) \sqrt{2\pi}} \exp\left(-\frac{1}{2}(z^2 + x)\right) \quad z \in \mathbb{R}; x > 0$$

Definamos la siguiente transformación del vector aleatorio $(X, Z)^T$

$$\begin{pmatrix} V \\ T \end{pmatrix} = \begin{pmatrix} X \\ \frac{Z}{\sqrt{\frac{X}{n}}} \end{pmatrix} \Rightarrow \begin{pmatrix} X \\ Z \end{pmatrix} = \begin{pmatrix} V \\ T\sqrt{\frac{V}{n}} \end{pmatrix}$$

Note que:

$$g(x, z) = \left(x, \frac{z}{\sqrt{\frac{x}{n}}}\right) = (v, t)$$

Por lo tanto las funciones inversas son:

$$\begin{aligned} x &= w_1(v, t) = v \\ z &= w_2(v, t) = t\sqrt{\frac{v}{n}} \end{aligned}$$

Esta transformación tiene por Jacobiano:

$$J = \begin{pmatrix} \frac{\partial w_1}{\partial v} & \frac{\partial w_1}{\partial t} \\ \frac{\partial w_2}{\partial v} & \frac{\partial w_2}{\partial t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \frac{\partial}{\partial v} t\sqrt{\frac{v}{n}} & \sqrt{\frac{v}{n}} \end{pmatrix}$$

Por lo tanto:

$$\det(J) = \sqrt{\frac{v}{n}}$$

Finalmente según el teorema de cambio de variable sabemos que:

$$f_{VT}(v, t) = f_{XZ}\left(v, t\sqrt{\frac{v}{n}}\right) |\det(J)| = \frac{v^{\frac{n-1}{2}}}{2^{\frac{n}{2}} \sqrt{2n\pi} \Gamma\left(\frac{n}{2}\right)} \exp\left(-\frac{1}{2}\left(1 + \frac{t^2}{n}\right)v\right); \quad v > 0; t \in \mathbb{R}$$

$$\begin{aligned}
f_T(t) &= \int_0^\infty f_{VT}(v, t) dv \\
&= \int_0^\infty \frac{v^{\frac{n-1}{2}}}{2^{\frac{n}{2}} \sqrt{2n\pi} \Gamma\left(\frac{n}{2}\right)} \exp\left(-\frac{1}{2} \left(1 + \frac{t^2}{n}\right) v\right) dv \\
&= \frac{1}{2^{\frac{n}{2}} \sqrt{2n\pi} \Gamma\left(\frac{n}{2}\right)} \int_0^\infty v^{\frac{n+1}{2}-1} \exp\left(-\frac{1}{2} \left(1 + \frac{t^2}{n}\right) v\right) dv
\end{aligned}$$

Haciendo $\alpha = \frac{n+1}{2}$ y $\beta = \frac{1}{2} \left(1 + \frac{t^2}{n}\right)$ Obtenemos que:

$$\begin{aligned}
f_T(t) &= \frac{1}{2^{\frac{n}{2}} \sqrt{2n\pi} \Gamma\left(\frac{n}{2}\right)} \int_0^\infty v^{\alpha-1} \exp(-\beta v) dv \\
&= \frac{\Gamma(\alpha)}{2^{\frac{n}{2}} \sqrt{2n\pi} \Gamma\left(\frac{n}{2}\right) \beta^\alpha} \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} v^{\alpha-1} \exp(-\beta v) dv \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{(n+1)}{2}}
\end{aligned}$$

Lo que concluye que T en efecto sigue una distribución t-student con n grados de libertad. \square

Se deja como ejercicios demostrar lo siguiente:

Ejercicio 2.5.2. Sea $X \sim \chi_{(n_1)}$ y $Y \sim \chi_{(n_2)}$ independientes, demuestre que:

$$F = \frac{\frac{X}{n_1}}{\frac{Y}{n_2}} = \frac{X n_2}{Y n_1} \sim F_{(n_1, n_2)}$$

Hint: Considere la transformación:

$$\begin{pmatrix} G \\ F \end{pmatrix} = \begin{pmatrix} Y \\ \frac{X n_2}{Y n_1} \end{pmatrix} \Rightarrow \begin{pmatrix} Y \\ X \end{pmatrix} = \begin{pmatrix} G \\ \frac{n_1}{n_2} F G \end{pmatrix}$$

Luego por el teorema de cambio de variable, encuentre la densidad conjunta del vector (G, F) y después encuentre la marginal de F y concluya

Capítulo 3

Inferencia Estadística

3.1. Planteamiento del problema

Se tiene un fenómeno aleatorio el cual queremos modelar matemáticamente para hacer inferencias sobre el comportamiento futuro del fenómeno estudiado.

3.2. Hipótesis

Para dar solución a nuestros problemas supongamos lo siguiente:

- Supondremos que el fenómeno aleatorio sigue una distribución de probabilidad conocida salvo ciertos parámetros $f_X(x; \underline{\theta})$ con $\underline{\theta} := (\theta_1, \dots, \theta_p)$ vector de parámetros desconocidos.
- Bajo el supuesto de que conocemos el modelo adecuado para nuestro problema, toda nuestra tarea se centrará en estimar al vector de parámetros $\underline{\theta}$

Debe de notarse que una vez determinado $\underline{\theta}$ se tiene un modelo completamente definido y podemos hacer cualquier tipo de inferencia distribucional de los datos, por ejemplo obtener probabilidades de caer en ciertas regiones o los momentos poblacionales como la esperanza y varianza.

Definición 3.2.1 (Espacio parametral). *Al conjunto de posibles valores que puede tomar un parámetro se le conoce como espacio parametral y se le denota como Θ*

Ejemplo 3.2.1. *Supongamos que tenemos un fenómeno aleatorio que se puede modelar con una distribución $N(1, \sigma^2)$ con σ^2 desconocido. Es decir la función de densidad de nuestra población está dada por:*

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

Entonces:

$$\Theta = \{\sigma^2 \in \mathbb{R} \quad : \quad \sigma^2 > 0\}$$

Observe que si conociéramos σ^2 el modelo estaría completamente especificado y podríamos **estimar** por ejemplo $\mathbb{P}(X > 1)$

Dado que nuestro objetivo es estimar a los parámetros, necesitaremos herramientas que nos ayuden a obtener una buena estimación, en estadística nuestra principal fuente de información será una muestra de la población la cual, bajo el supuesto de que aun no hemos llevado a cabo el experimento, serán variables aleatorias. Formalmente definiremos a una muestra aleatoria como:

Definición 3.2.2 (Muestra Aleatoria). *Supongamos que tenemos una población que es modelada por una función de densidad dada por $f_X(x; \underline{\theta})$ con $\underline{\theta}$ un vector de parámetros desconocidos. Decimos que X_1, \dots, X_n es una muestra aleatoria (denotada por m.a.) de tamaño n si:*

- Cada X_i sigue la misma distribución dada por la función de densidad $f_X(x; \underline{\theta})$
- X_i es independiente de X_j con $i \neq j$

En inferencia clásica $\underline{\theta}$ se considera un vector de parámetros fijos pero desconocidos y por lo tanto no tiene asociada ningún tipo de distribución de probabilidad, en cambio en estadística bayesiana $\underline{\theta}$ al ser desconocido se le asocia una distribución inicial de forma subjetiva.

Nuestro objetivo es entonces encontrar métodos eficientes que nos ayuden a estimar de la mejor manera a los parámetros desconocidos, para ello definamos lo siguiente:

Definición 3.2.3 (Estadística). *Sea X_1, \dots, X_n m.a. de cierto fenómeno aleatorio, una estadística es cualquier función de la muestra que no depende de parámetros desconocidos*

$$t_1 = t_1(X_1, \dots, X_n)$$

Ejemplo de estadísticas pueden ser:

$$\begin{aligned} t_1 &= \frac{\sum_{i=1}^n X_i}{n} \\ t_2 &= \max \{X_1, \dots, X_n\} \\ t_3 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \\ t_4 &= \left(\frac{\sum_{i=1}^n X_i}{n}, \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \right) \end{aligned}$$

Definición 3.2.4 (Estimador). *Un estimador, es una estadística, que esperamos cumpla con ciertas propiedades que ayuden a la estimación de un vector de parámetros desconocidos.*

La notación que hacemos para identificar a un estimador de cierto vector de parámetros $\theta \in \mathbb{R}^p$ es: $\hat{\theta}$, esto de cierta forma es la abreviación de:

$$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n) = \left(\hat{\theta}_1(X_1, \dots, X_n), \dots, \hat{\theta}_p(X_1, \dots, X_n) \right)$$

El problema general implica estimar varios parámetros, sin embargo supondremos en la mayor parte de este curso que solo tenemos un parámetro desconocido en cuyo caso:

$$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n) \text{ es una v.a. Real}$$

Una de las primeras propiedades que deberíamos pedir al estimador es que al menos su contradominio sea un subconjunto del espacio parametral, ya que la idea es que la distribución de dicho estimador esté centrada al rededor del parámetro que queremos estimar.

Para saber cuál estimador es mejor que otro, es necesario conocer las propiedades de cada uno y elegir aquél que más se adecue a nuestras necesidades. Con el objeto de conocer mejor estas propiedades es necesario saber cuando menos cuáles son sus momentos centrales y si es posible saber qué tipo de distribución está siguiendo.

Podemos motivar el estudio de la distribución de un estadístico de la siguiente manera. Sea X una variable aleatoria que nos mide alguna característica numérica de algún fenómeno aleatorio. En muchas ocasiones nos enfrentamos al problema de no saber completamente cuál es la función de distribución de esta variable aleatoria. Supongamos que sabemos que la variable aleatoria sigue una distribución $F_X(x; \theta)$ pero con el parámetro θ desconocido. Para obtener más información acerca de la distribución de X nosotros podemos repetir el experimento n veces de manera independiente y definir X_i la v.a. de la i -ésima repetición del experimento. Entonces X_1, \dots, X_n será una muestra aleatoria de la distribución en estudio con la cual se pretenderá estimar el valor de θ a través de una función de ella. Ahora suponga que definimos un estadístico $Y = u(X_1, \dots, X_n)$ para el cual podemos encontrar su función de densidad de probabilidad $G_Y(u)$ y supongamos que esta función nos muestra que existe una probabilidad considerable de que Y , al realizar los experimentos, tenga un valor cercano al parámetro desconocido. Así, una vez que el experimento ha sido repetido en n ocasiones se obtiene una muestra observada x_1, \dots, x_n la cual al ser evaluada en la función de la muestra $u(x_1, \dots, x_n)$ nos dará un número conocido al cual esperamos, bajo las condiciones arriba expuestas, sea cercano al valor desconocido del parámetro θ .

Observe que de lo anterior para saber qué tan cercano será el valor estimado al parámetro desconocido θ , es necesario conocer la forma de la distribución del estadístico que lo define, para así tener una medida de probabilidad de qué tan cercana será nuestra estimación.

Ejemplo 3.2.2. *Supongamos que tenemos X_1, \dots, X_n m.a. del modelo exponencial de parámetro λ desconocido, primero observemos que en este caso el espacio parametral está dado por $\Theta = \mathbb{R}^+$, queremos definir una función de la muestra que nos ayude a determinar cual de todos los valores que están en el espacio parametral es el mas apropiado para estimar a λ . Posibles estimadores para λ pueden ser:*

- $\hat{\lambda}(X_1, \dots, X_n) = \frac{n}{\sum_{i=1}^n X_i}$
- $\hat{\lambda}(X_1, \dots, X_n) = \min X_1, \dots, X_n$
- $\hat{\lambda}(X_1, \dots, X_n) = \frac{\sum_{i=1}^n X_i}{n}$

Algo que no puede ser llamado estimador para λ seria :

- $\hat{\lambda}(X_1, \dots, X_n) = \lambda \frac{n}{\sum_{i=1}^n X_i}$
- $\hat{\lambda}(X_1, \dots, X_n) = -\frac{\sum_{i=1}^n X_i}{n}$

En el primer caso no es estimador pues la expresión de la función involucra al parámetro desconocido λ y en el segundo caso, la contradominio de la función que determina al estimador no cae en el espacio parametral ($\Theta = \mathbb{R}^+$).

Notemos que en nuestra teoría que estamos desarrollando, los estimadores $\hat{\theta}$ al ser funciones de variables aleatorias, entonces también heredan esta propiedad y por tanto el estimador $\hat{\theta}$ también es v.a., luego entonces surge de manera natural preguntarnos por algunas propiedades como su distribución o sus momentos. En ocasiones es difícil y ambicioso conocer completamente la distribución que sigue un estimador por lo que a veces solo nos conformamos conociendo algunos momentos como su esperanza y varianza.

Por ejemplo, supongamos que tenemos X_1, \dots, X_n m.a. de una cierta distribución F con media μ y varianza σ^2 , supongamos además que μ es un parámetro de interés de la población que estamos estudiando, por lo que definimos el siguiente estimador:

$$\hat{\mu} = \hat{\mu}(X_1, \dots, X_n) = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$$

Ciertamente desconocemos la distribución de $\hat{\mu}$ pues de entrada no conozco ni siquiera como se distribuye cada X_i , sin embargo si puedo conocer los dos primeros momentos de mi estimador

con la información que cuento, pues:

$$\begin{aligned}\mathbb{E}(\hat{\mu}) &= \mathbb{E}(\overline{X}) = \frac{1}{n} \mathbb{E}\left(\sum_{i=1}^n X_i\right) = \mu \\ \text{Var}(\hat{\mu}) &= \text{Var}(\overline{X}) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{\sigma^2}{n}\end{aligned}$$

Observe que en la última igualdad utilizamos fuertemente el hecho de la independencia de las variables aleatorias X_1, \dots, X_n . Debe de quedar claro entonces que no conozco la distribución de $\hat{\mu}$ sin embargo el hecho de conocer los dos primeros momentos nos da mucha información pues ahora sabemos en promedio $\hat{\mu}$ tomará valores cercanos a μ y que además la varianza de dicho estimador se va haciendo cada vez mas chica conforme tenemos mas muestra, esto quiere decir que $\hat{\mu}$ es un estimador que conforme tiene mas muestra ofrece mejores aproximaciones hacia el valor desconocido μ .

Definición 3.2.5 (Estimador Insesgado). *Decimos que $\hat{\theta}$ es un estimador **insesgado** para θ si*

$$\mathbb{E}(\hat{\theta}) = \theta$$

*Y cuando no pasa la igualdad decimos que el estimador es **sesgado***

Los estimadores insesgados entonces tiene la propiedad de que sus esperanza es el valor del parámetro desconocido, lo que nos quiere decir esto entonces es que si repetimos muchas veces el experimento y evaluamos el estimador en cada una de las muestras que vamos obteniendo, entonces en promedio el valor que vamos a obtener es precisamente el valor desconocido θ .

Observacion Importante: Como ya hemos dicho un estimador es una v.a. pues es función de la muestra aleatoria, sin embargo una vez observada la muestra y evaluada en la función que determina el estimador se obtiene un número, a dicho número se le conoce como **estimación**. Debe entonces quedar claro la gran diferencia que hay entre un estimador y la estimación.

Ejercicio 3.2.1. *Conteste lo siguiente:*

- Suponga X_1, \dots, X_n m.a. de una distribución con media μ y varianza σ^2 . Pruebe que:

$$\hat{\sigma}_1^2 := \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n}$$

Es un estimador sesgado para σ^2 pero que:

$$\hat{\sigma}_2^2 := \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

si es insesgado para σ^2 , luego suponga que observamos la siguiente muestra: $X_1 = 2, X_2 = 1, X_3 = 4, X_4 = 2, X_5 = 1$, encuentre el valor estimado de σ^2 de ambos estimadores.

- Sea X_1, X_2, X_3, X_4, X_5 muestra aleatoria de tamaño $n = 5$ de una población con densidad dada por:

$$f_X(x; \theta) = \frac{\theta^x}{x!} e^{-\theta}; \quad \theta > 0; \quad x \in \{0, 1, 2, \dots\}$$

Considere las siguientes funciones de la muestra:

$$\hat{\theta}_1 = X_1 + X_2 + X_3 + X_4 + X_5$$

$$\hat{\theta}_2 = X_1 + \theta X_2$$

$$\hat{\theta}_3 = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$$

$$\hat{\theta}_4 = \frac{2X_1 + 3X_2}{5}$$

1. Indique cuáles funciones de la muestra pueden ser considerados estimadores para θ y cuales no
2. Indique cuáles estimadores son insesgados y cuáles sesgados

Como ya hemos dicho, encontrar la distribución exacta de nuestros estimadores a veces no es fácil sin embargo en algunas ocasiones si es posible encontrar algebraicamente cual es la distribución de nuestro estimador, en la sección anterior se definieron dos estimadores uno para μ y otro para σ^2 de una distribución genérica ($\hat{\mu} = \bar{X}$ y $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$), supongamos ahora que tenemos el modelo Normal y nos preguntamos ¿Cómo se distribuye $\hat{\mu}$ y $\hat{\sigma}^2$?. Afortunadamente en este caso si podremos encontrar la forma exacta de la distribución de estos dos estimadores utilizando la función generadora de momentos.

Ejercicio 3.2.2. ■ Sea X_1, \dots, X_n variables aleatorias Normales independientes tales que $X_i \sim N(\mu_i, \sigma_i^2)$, y defina la variable aleatoria W como:

$$W = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n$$

Demuestre que W sigue una distribución normal y encuentre los parámetros de media y varianza para esta nueva variable aleatoria.

Finalmente concluya que si X_1, \dots, X_n es m.a. de una población $N(\mu, \sigma^2)$ entonces:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Hasta ahora hemos encontrado la distribución exacta de \bar{X} cuando la muestra proviene de una población normal, sin embargo aun queda abierto el problema de conocer la distribución exacta del estimador para la varianza σ^2 dado por $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, veamos cómo encontrar esta distribución para el caso normal.

Por un lado sabemos $\frac{X_i - \mu}{\sigma} \sim N(0, 1)$ entonces, $\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_{(1)}^2$ y por lo tanto queda claro que cuando tenemos una m.a X_1, \dots, X_n de una población $N(\mu, \sigma^2)$, entonces:

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_{(n)}^2$$

De donde concluimos que:

$$\frac{n}{\sigma^2} \sum_{i=1}^n \frac{(X_i - \mu)^2}{n} \sim \chi_{(n)}^2 \quad \Rightarrow \quad \frac{n}{\sigma^2} \hat{\sigma}_1^{2*} \sim \chi_{(n)}^2$$

De donde $\hat{\sigma}_1^{2*} = \sum_{i=1}^n \frac{(X_i - \mu)^2}{n}$, y de aquí podemos encontrar la densidad exacta para $\hat{\sigma}_1^{2*}$ por medio del siguiente método, primero definamos a la v.a. U como $U = \hat{\sigma}_1^{2*}$ y V una v.a tal que $V \sim \chi_{(n)}^2$, queremos encontrar la distribución de U , entonces:

$$F_U(u) = \mathbb{P}(U \leq u) = \mathbb{P}(\hat{\sigma}_1^2 \leq u) = \mathbb{P}\left(\frac{n}{\sigma^2} \hat{\sigma}_1^2 \leq \frac{n}{\sigma^2} u\right) = \mathbb{P}\left(V \leq \frac{n}{\sigma^2} u\right) = F_V\left(\frac{n}{\sigma^2} u\right)$$

Todo lo anterior nos ayudo a encontrar la densidad de $\hat{\sigma}_1^{2*}$, sin embargo este no era nuestro objetivo, recordemos que queremos encontrar la densidad o distribución exacta para:

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

Para ello requeriremos un teorema muy importante para nuestra teoría.

Teorema 3.2.1 (Teorema de Cochran). Sea X_1, \dots, X_n m.a. de una población con densidad

$f(\cdot)$. Entonces las estadísticas dadas por:

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

Son independientes si y solo si lo m.a. proviene de una población normal

Demostración. Solo probaremos la implicación que utilizaremos en el curso, es decir supondremos que la muestra aleatoria proviene de una distribución Normal y probaremos que las estadísticas \bar{X} y S^2 son independientes.

Para la prueba utilizaremos algunos resultados importantes de vectores aleatorios Normales, lo primero que ocuparemos es el siguiente resultado.

Si $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$ y $\mathbf{A} \in \mathbb{R}^{k \times p}$ de rango completo, entonces:

$$\underline{Y} = \mathbf{A}\underline{X} \sim N_k(\mathbf{A}\underline{\mu}, \mathbf{A}\Sigma\mathbf{A}^T)$$

Es decir, la propiedad nos dice que cualquier transformación lineal de un vector aleatorio normal multivariado también es normal multivariado. En nuestro caso para la prueba supongamos que tenemos X_1, \dots, X_n m.a. de una población $N(\mu, \sigma^2)$. Con la muestra definamos al vector aleatorio $\underline{X} = (X_1, \dots, X_n)^T$, entonces al ser X_i variables aleatorias independientes e idénticamente distribuidas tenemos que:

$$\underline{X} \sim N_n(\mathbf{1}\mu, \sigma^2\mathbf{I})$$

Donde $\mathbf{1} = (1, 1, \dots, 1)^T$. Definamos ahora la matriz $\mathbf{T} \in \mathbb{R}^{n \times n}$ ortogonal tal que su primer renglón es del forma $(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$ es decir:

$$\mathbf{T} = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nn} \end{pmatrix}$$

Sabemos que esta matriz existe utilizando el algoritmo para ortogonalizar una base de \mathbb{R}^n . Definamos entonces el vector \underline{Y} como:

$$\underline{Y} = \mathbf{T}\underline{X}$$

Entonces como $\underline{X} \sim N_n(\mathbf{1}\mu, \sigma^2\mathbf{I})$, sabemos entonces que $\underline{Y} \sim N_n(\mathbf{T}\mathbf{1}\mu, \mathbf{T}\sigma^2\mathbf{I}\mathbf{T}^T)$, pero notemos que:

$$\mathbf{T}\sigma^2\mathbf{I}\mathbf{T}^T = \sigma^2\mathbf{T}\mathbf{T}^T = \sigma^2\mathbf{I}$$

Pues recordemos que \mathbf{T} es ortonormal y por lo tanto $\mathbf{T}\mathbf{T}^T = \mathbf{I}$.

Entonces hemos probado que $\underline{Y} \sim N_n(\mathbf{T}\mathbf{1}\mu, \sigma^2\mathbf{I})$. Lo anterior nos indica que Y_1, \dots, Y_n son independientes. Analicemos la primer entrada de este vector aleatorio.

$$Y_1 = \sum_{i=1}^n \frac{1}{\sqrt{n}} X_i = \frac{n}{\sqrt{n}} \sum_{i=1}^n \frac{1}{n} X_i = \sqrt{n}\bar{X}$$

Consideremos la suma de cuadrados del vector aleatorio \underline{X} y \underline{Y} .

$$\sum_{i=1}^n Y_i^2 = \underline{Y}^T \underline{Y} = (\mathbf{T}\underline{X})^T (\mathbf{T}\underline{X}) = \underline{X}^T \mathbf{T}^T \mathbf{T} \underline{X} = \underline{X}^T \underline{X} = \sum_{i=1}^n X_i^2$$

Es decir, la norma del vector \underline{X} y \underline{Y} son las mismas.

Por otro lado, ahora consideremos las variables aleatorias Y_2, \dots, Y_n y tomemos la suma de sus cuadrados, es decir:

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

Finalmente como Y_1, \dots, Y_n son independientes, entonces Y_1 es independientes de (Y_2, \dots, Y_n) y por tanto $g_1 \text{ par } Y_1$ es independiente de $g_2(Y_2, \dots, Y_n)$ con g_1 y g_2 funciones medibles. Tomando:

$$g_1(y) = \frac{1}{\sqrt{n}}y \quad y \quad g_2(y_2, \dots, y_n) = \sum_{i=2}^n y_i^2$$

Concluimos entonces que $g_1(Y_1)$ y $g_2(Y_2, \dots, Y_n)$ son independientes, pero como:

$$g_1(Y_1) = \frac{1}{\sqrt{n}}Y_1 = \frac{1}{\sqrt{n}}\sqrt{n}\bar{X} = \bar{X}$$

$$g_2(Y_2, \dots, Y_n) = \sum_{i=2}^n Y_i^2 = \sum_{i=1}^n (X_i - \bar{X})^2$$

Concluimos entonces que \bar{X} y $\sum_{i=1}^n (X_i - \bar{X})^2$ son independientes, lo que nos asegura la independencia entre \bar{X} y S^2

□

Luego, como:

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

Entonces:

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

Ahora definamos las variables aleatorias:

$$U := \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}; \quad V := \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}; \quad T := \frac{(\bar{X} - \mu)^2}{\frac{\sigma^2}{n}}$$

Como ya vimos anteriormente se sabe que cuando las variables aleatorias son normales que $V \sim \chi_{(n)}^2$ y como $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ entonces $T \sim \chi_{(1)}^2$. Sin embargo la pregunta es ¿Cómo se distribuye U? , hasta ahora lo que sabemos es que:

$$U = V - T \Rightarrow T + U = V$$

Pero usando el teorema (3.2.1) sabemos además que U es independiente de T por lo tanto:

$$M_V(t) = M_{T+U}(t) = M_T(t)M_U(t) \quad (3.1)$$

Pero como $V \sim \chi_{(n)}^2$ y $T \sim \chi_{(1)}^2$ entonces:

$$M_V(t) = (1 - 2t)^{-\frac{n}{2}}; \quad M_T(t) = (1 - 2t)^{-\frac{1}{2}}$$

Sustituyendo esto último en (3.1) y despejando obtenemos que:

$$M_U(t) = (1 - 2t)^{-\frac{n-1}{2}}$$

De donde finalmente concluimos que:

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

Con lo anterior resulta ahora fácil encontrar la distribución exacta de $\hat{\sigma}^2$ pues:

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

Ejercicio 3.2.3. Encontrar la densidad de $\hat{\sigma}^2$ y obtener su esperanza y varianza

Como resumen tenemos lo siguiente, cuando X_1, \dots, X_n es m.a. de una población $N(\mu, \sigma^2)$ entonces:

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

Debe de quedar claro que esto solo fue posible usando propiedades de la Normalidad y el teorema de Cochran por lo tanto lo anterior solo es válido bajo el supuesto de normalidad.

Hemos desarrollado una muy buena teoría bajo el supuesto de normalidad, sin embargo surge la pregunta, ¿Por qué estudiamos tanto al modelo Normal?, la respuesta es que este modelo surge como distribuciones límites via el Teorema de Límite Central.

3.2.1. Teorema del Límite Central

Teorema 3.2.2. [Teorema del Límite Central] Sea X_1, \dots, X_n m.a. de una distribución F_x (Puede ser continua o discreta), tal que, $\mathbb{E}(X) = \mu$ y $\text{Var}(X) = \sigma^2$ ambos finitos. Entonces:

$$\bar{X} := \frac{\sum_{i=1}^n X_i}{n} \underset{\text{aprox}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

O bien, otra forma de verlo es:

$$\sqrt{n} \left(\frac{\bar{X} - \mu}{\sigma} \right) \underset{\text{aprox}}{\sim} N(0, 1)$$

De hecho el teorema concluye que:

$$\sqrt{n} \left(\frac{\bar{X} - \mu}{\sigma} \right) \xrightarrow{n \rightarrow \infty} N(0, 1)$$

Algunas observaciones importantes del T.L.C

1. Es importante tener variables aleatorias independiente e idénticamente distribuidas y que el segundo momento exista (En series de tiempo no se aplica el T.L.C)
2. No importa de que distribución estemos muestreando, el teorema afirma que habrá una aproximación hacia la normalidad de \bar{X}
3. Una de las preguntas mas interesantes al momento de que vamos a aplicar el T.L.C es saber que n debo tener para que afirmar que tengo ya una buena aproximación. La respuesta a esta pregunta no es fácil y se puede encontrar por medio de simulaciones sin embargo se muestra que el tamaño de muestra necesario depende de la forma de F_X sin embargo en la práctica esta última distribución es desconocida.

En nuestra teoría de estimación el T.L.C. Juega un papel interesante pues de hecho nos dice lo siguiente:

Teorema 3.2.3. *Sea X_1, \dots, X_n m.a. de una población F_X tal que $\mathbb{E}(X) = \mu$ y $\text{Var}(X) = \sigma^2$, supongamos que queremos estimar a μ por medio de la estadística $\hat{\mu} = \bar{X}$ entonces, la distribución aproximada de nuestro estimador está dada por:*

$$\hat{\mu} \stackrel{\text{aprox}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

3.2.2. Estimación Puntual

En esta sección de nuestro trabajo nos enfocaremos a encontrar al mejor estimador para cierto parámetro desconocido, pero surge entonces la pregunta. ¿Cómo encontrar al mejor estimador? para responder esto primero necesitamos un criterio que nos ayude a determinar cuál estimador es mejor que otro.

Definición 3.2.6 (Error Cuadrático Medio). *Sea X_1, \dots, X_n una m.a. de una distribución $F_X(\cdot; \theta)$, con θ un parámetro desconocido, supongamos que tenemos $\hat{\theta}$ un estimador de θ , definimos el Error cuadrático medio de $\hat{\theta}$ como :*

$$E.C.M(\hat{\theta}) := \mathbb{E}\left(\left(\hat{\theta} - \theta\right)^2\right)$$

Observe que el E.C.M mide la discrepancia cuadrática promedio que tiene nuestro estimador $\hat{\theta}$ con respecto al verdadero valor del parámetro θ . Es fácil ver además que :

$$E.C.M(\hat{\theta}) = \text{Var}(\hat{\theta}) + \left(\mathbb{E}(\hat{\theta}) - \theta\right)^2$$

Donde $\left(\mathbb{E}(\hat{\theta}) - \theta\right)$ se le conoce como el sesgo del estimador, observe además también que cuando el estimador es insesgado, entonces $E.C.M(\hat{\theta}) = \text{Var}(\hat{\theta})$.

El E.C.M. es una medida de qué tanto nos alejamos en promedio del verdadero valor por lo que resulta lógico escoger siempre el estimador con menor $E.C.M$, parece entonces que si queremos encontrar al mejor estimador para θ necesitamos encontrar a que $\hat{\theta}$ tal que:

$$E.C.M(\hat{\theta}) \leq E.C.M(\hat{\theta}_1) \quad \forall \hat{\theta}_1 \text{ estimador de } \theta$$

sin embargo lo anterior no es tan fácil pues se puede probar que debido a que $E.C.M(\hat{\theta})$

depende de θ , entonces se pueden encontrar dos estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ tal que:

$$E.C.M(\hat{\theta}_1) < E.C.M(\hat{\theta}_2) \text{ si } \theta \in \Theta_0$$

$$E.C.M(\hat{\theta}_1) > E.C.M(\hat{\theta}_2) \text{ si } \theta \in \Theta_1$$

Donde $\Theta_0 \subset \Theta$ y $\Theta_1 \subset \Theta$

Ejemplo 3.2.3. Sea X_1, \dots, X_n m.a. de una población F_X tal que $\mathbb{E}(X) = \mu$ y $\text{Var}(X) = \sigma^2$ finitos, supongamos que queremos estimar a μ por medio de los siguientes estimadores:

$$\hat{\mu}_1 = \bar{X} \quad \hat{\mu}_2 = \mu_0$$

con μ_0 un numero conocido. Entonces:

$$E.C.M(\hat{\mu}_1) = \frac{\sigma^2}{n} \quad E.C.M(\hat{\mu}_2) = (\mu_0 - \mu)^2$$

Luego entonces si $\mu = \mu_0$ se tiene que $E.C.M(\hat{\mu}_2) = 0$ y por tanto $\hat{\mu}_2$ sería un mejor estimador que $\hat{\mu}_1$ sin embargo si μ está muy alejado de μ_0 tal que, $(\mu_0 - \mu)^2 > \frac{\sigma^2}{n}$, entonces $\hat{\mu}_1$ es mejor estimador que $\hat{\mu}_2$.

Para solucionar este inconveniente del E.C.M. lo que se hace es restringir la búsqueda del mejor estimador entre todos los **insesgados**. Recordemos ahora que si $\hat{\theta}$ es un estimador insesgado, entonces:

$$E.C.M(\hat{\theta}) = \text{Var}(\hat{\theta})$$

Por lo tanto, tenemos que desarrollar una teoría que nos ayude a encontrar los estimadores insesgados de mínima varianza.

Ejercicio 3.2.4. Responda los siguientes ejercicios:

1. Sea X_1, \dots, X_n muestra aleatoria de tamaño n de una población con distribución $F_X(x)$ tal que $\mathbb{E}(X_i) = \mu$ y $\text{Var}(X_i) = \sigma^2$. Pruebe que:

- Suponga μ conocida, entonces $S_1^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ es estimador insesgado para σ^2
- Suponga μ desconocida, entonces $S_2^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ es un estimador sesgado para σ^2
- Suponga μ desconocida, $S_3^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ es un estimador insesgado para σ^2

2. Sea X_1, \dots, X_n muestra aleatoria de tamaño n de una población con distribución $F_X(x)$ tal que $\mathbb{E}(X_i) = \mu$ y $\text{Var}(X_i) = \sigma^2$. Sea $\alpha_1 \dots \alpha_n \in \mathbb{R}$ tales que $\sum_{i=1}^n \alpha_i = 1$. Defina el siguiente estimador para μ :

$$\hat{\mu} = \sum_{i=1}^n \alpha_i X_i$$

Pruebe que:

- a) $\hat{\mu}$ es un estimador insesgado
 - b) La varianza de $\hat{\mu}$ se minimiza cuando $\alpha_i = \frac{1}{n}$ para toda $i \in \{1, \dots, n\}$. (Este inciso muestra entonces que \bar{X} es el mejor estimador lineal). Cuando ocurre esto se dice que el estimador es BLUE (Best Linear Unbiased Estimator).
3. Sea X_1, X_2, X_3, X_4, X_5 muestra aleatoria de tamaño $n = 5$ de una población con densidad dada por:

$$f_X(x; \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}; \quad \theta > 0; \quad x > 0$$

Considere las siguientes funciones de la muestra:

$$\hat{\theta}_1 = X_1 + X_2 + X_3 + X_4 + X_5$$

$$\hat{\theta}_2 = X_1 + \theta X_2$$

$$\hat{\theta}_3 = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$$

$$\hat{\theta}_4 = \frac{X_1 + 4X_2}{5}$$

Conteste lo siguiente:

- Indique cuales pueden ser considerados estimadores para θ y cuales no
 - Indique cuales estimadores son insesgados y cuales sesgados
 - Para los estimadores sesgados, calcule su sesgo.
 - Para los estimadores insesgados, indique cual tiene menor Error Cuadrático Medio y concluya cual estimador es mejor.
4. Sea X_1, X_2, X_3 muestra aleatoria de tamaño 3 de una población con distribución $F_X(x)$ tal que $\mathbb{E}(X_i) = \mu$ y $\text{Var}(X_i) = 1$. Calcular el error cuadrático medio para los siguientes estimadores de μ :

$$\hat{\mu}_1 = X_1 \quad ; \quad \hat{\mu}_2 = \frac{(3X_1 - 2X_2 + X_3)}{6}$$

Para cuáles valores de μ , se tiene que $\hat{\mu}_2$ es mejor estimador que $\hat{\mu}_1$?

5. Sea X_1, \dots, X_n muestra aleatoria de un modelo $N(\mu, \sigma^2)$ con μ y σ^2 desconocidos. Sabemos que $S_3^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ es un estimador insesgado para σ^2 . Demuestre que:

$$\text{Var}(S_3^2) = \frac{2\sigma^4}{n-1}$$

Esto muestra entonces que cuando $\lim_{n \rightarrow \infty} \text{Var}(S_3^2) = 0$ lo cual es una propiedad interesante del estimador pues implica que conforme la muestra tiende a infinito, la precisión es cada vez mejor.

3.2.3. Cota Inferior de Cramer Rao

Supongamos que podemos encontrar un número C_θ tal que:

$$\text{Var}(\hat{\theta}) \geq C_\theta \quad \forall \hat{\theta} \text{ estimador insesgado de } \theta$$

Entonces C_θ sería una cota inferior para la varianza de cualquier estimador insesgado de θ . Supongamos ahora que encontramos un estimador $\hat{\theta}^*$ tal que $\text{Var}(\hat{\theta}^*) = C_\theta$, entonces se verificaría que:

$$\text{Var}(\hat{\theta}) \geq C_\theta = \text{Var}(\hat{\theta}^*) \quad \forall \hat{\theta} \text{ estimador insesgado de } \theta$$

Entonces habríamos solucionado nuestro problema y $\hat{\theta}^*$ sería el mejor estimador insesgado que podríamos encontrar. (El de menor varianza).

Definición 3.2.7 (UMVUE). *Al estimador insesgado de mínima varianza se le llama UMVUE por sus siglas en ingles Uniformly Minimum-Variance Unbiased Estimator*

Resulta entonces que nuestro objetivo se centrará en encontrar estimadores que sean **UMVUES**. Una forma de atacar este problema es primero encontrar a C_θ conocida como la Cota Inferior para los estimadores insesgados de θ . Para ello definamos lo siguiente:

Definición 3.2.8 (Verosimilitud-Muestra no Observada). *Sea X_1, \dots, X_n m.a. de un modelo con densidad $f_X(x; \theta)$ y θ un parámetro desconocido tal que $\theta \in \Theta$, definimos la función de verosimilitud $\ell : \Theta \rightarrow \mathbb{R}$ como:*

$$\ell(\theta; \underline{X}) := f_{\underline{X}}(X_1, \dots, X_n) = \prod_{i=1}^n f_{X_i}(X_i; \theta)$$

Observe que como aun no observamos la muestra, entonces $\ell(\theta; \underline{X})$ sigue siendo v.a. pues está en función de variables aleatorias, cuando ya se observó la muestra y cada variable tomó ya un valor dado $X_1 = x_1, \dots, X_n = x_n$ entonces la función de verosimilitud ahora si toma un valor numérico para cada $\theta \in \Theta$ y se vuelve completamente una función real. Esta función la estaremos ocupando mas adelante cuando definamos métodos de estimación por máxima verosimilitud .

A partir de la verosimilitud definamos la función SCORE.

Definición 3.2.9 (Función SCORE). Sea X_1, \dots, X_n m.a. de un modelo con densidad $f_X(x; \theta)$ y θ un parámetro desconocido tal que $\theta \in \Theta$, definimos la función SCORE $S : \Theta \rightarrow \mathbb{R}$ como:

$$S(\theta; \underline{X}) = \frac{\partial}{\partial \theta} \text{Log}(\ell(\theta; \underline{X})) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \text{Log}(f_{X_i}(X_i; \theta))$$

En lo anterior definamos a la v.a. U_i como $U_i = \frac{\partial}{\partial \theta} \text{Log}(f_{X_i}(X_i; \theta))$ entonces se puede probar bajo condiciones de regularidad que:

1. $\mathbb{E}(U_i) = 0$
2. $\text{Var}(U_i) = \mathbb{E}(U_i^2) = -\mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \text{Log}(f_{X_i}(X_i; \theta))\right)$

Las condiciones de regularidad las utilizaremos mucho en la teoría de estimación a continuación exponemos de que hablan estas condiciones:

Definición 3.2.10 (Condiciones de Regularidad). Decimos que $f_X(., \theta)$ cumple con las condiciones de regularidad si:

- Es soporte de f no depende de θ .
- $\log f$ es una función dos veces diferenciable
- Los operadores de diferenciación e integración son intercambiables:

$$\int_{-\infty}^{\infty} \frac{\partial^2}{\partial \theta^2} f(x; \theta) dx = \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} f(x; \theta) dx = \frac{\partial^2}{\partial \theta^2} \int_{-\infty}^{\infty} f(x; \theta) dx$$

Definición 3.2.11 (Información de Fisher por Unidad Muestral). Dentro del contexto anterior definimos a **Información de Fisher por Unidad Muestral** denotada por $I(\theta)$ como:

$$I(\theta) = \text{Var}(U_i) = -\mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \text{Log}(f_{X_i}(X_i; \theta))\right)$$

De lo anterior se puede concluir lo siguiente:

$$\mathbb{E}(S(\theta, \underline{X})) = 0 \quad \text{Var}(S(\theta, \underline{X})) = nI(\theta)$$

Parentesis

Definición 3.2.12 (Coeficiente de Correlación). Sean X y Y dos v.a. definimos el Coeficiente de Correlación entre X y Y como:

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

Propiedades del Coeficiente de Correlación:

1. $-1 \leq \rho_{XY} \leq 1 \Rightarrow |\rho_{XY}| \leq 1$
2. $|\rho_{XY}| = 1$ si y solo si X y Y se relacionan de manera lineal, es decir, $X = aY + b$ o $Y = a^*X + b^*$ (con $a \neq 0$ y $a^* \neq 0$)

Dentro de toda nuestra teoría que hemos venido desarrollando, supongamos que estamos interesados en encontrar el coeficiente de correlación entre $\hat{\theta}$ y $S(\theta, \underline{X})$ entonces sabemos que:

$$-1 \leq \rho_{\hat{\theta}S(\theta; \underline{X})} = \frac{\text{Cov}(\hat{\theta}, S(\theta; \underline{X}))}{\sqrt{\text{Var}(\hat{\theta}) \text{Var}(S(\theta; \underline{X}))}} \leq 1$$

Por lo tanto:

$$\text{Cov}(\hat{\theta}, S(\theta; \underline{X}))^2 \leq \text{Var}(\hat{\theta}) nI(\theta) \Rightarrow \frac{\text{Cov}(\hat{\theta}, S(\theta; \underline{X}))^2}{nI(\theta)} \leq \text{Var}(\hat{\theta})$$

Nuevamente bajo condiciones de regularidad se puede probar que:

$$\text{Cov}(\hat{\theta}, S(\theta; \underline{X})) = \frac{\partial}{\partial \theta} \mathbb{E}(\hat{\theta})$$

De donde concluimos que:

$$\frac{\left(\frac{\partial}{\partial \theta} \mathbb{E}(\hat{\theta})\right)^2}{nI(\theta)} \leq \text{Var}(\hat{\theta})$$

Luego entonces, si suponemos $\hat{\theta}$ es un estimador insesgado se tendría que:

$$\frac{1}{nI(\theta)} \leq \text{Var}(\hat{\theta})$$

Esto da origen a la Cota Inferior de Cramer-Rao.

Definición 3.2.13 (Cota Inferior de Cramer-Rao). Sea X_1, \dots, X_n m.a. de un modelo con densidad $f_X(x; \theta)$ y θ un parámetro desconocido tal que $\theta \in \Theta$, definimos cota inferior de Cramer-Rao para los estimadores insesgados de θ como:

$$CICR(\theta) = \frac{1}{nI(\theta)} \leq \text{Var}(\hat{\theta})$$

Donde $\hat{\theta}$ es un estimador insesgado arbitrario

Debe quedar claro entonces que si encontramos un estimador insesgado $\hat{\theta}^*$ tal que $\text{Var}(\hat{\theta}^*) = CICR(\theta)$, entonces $\hat{\theta}^*$ es UMVUE, es decir, es el mejor estimador insesgado que podemos encontrar.

Ejemplo 3.2.4. Sea X_1, \dots, X_n m.a. de un modelo Bernoulli de parámetro $\theta \in (0, 1)$. Entonces sabemos que $f_{X_i}(x_i; \theta) = \theta^{x_i} (1 - \theta)^{1-x_i} \mathbf{1}_{\{0,1\}}(x_i)$, por lo que:

$$I(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \text{Log}(f_{X_i}(X_i; \theta))\right) = \frac{1}{\theta(1-\theta)}$$

Por lo tanto, para este modelo:

$$CICR(\theta) = \frac{\theta(1-\theta)}{n}$$

Lo que concluye que cualquier estimador insesgado de θ , digamos $\hat{\theta}$ es tal que:

$$\text{Var}(\hat{\theta}) \geq \frac{\theta(1-\theta)}{n}$$

No es difícil probar con esto entonces que:

$$\hat{\theta}^* := \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Es un UMVUE para θ en el modelo Bernoulli(θ)

Ejercicio 3.2.5. Resuelva lo siguiente:

1. Sea X_1, \dots, X_n m.a. de una población con densidad dada por:

$$f_X(x; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}x^2}, \sigma^2 > 0$$

a) Calcule la Cota Inferior de Cramer Rao para esta familia y verifique que:

$$CICR(\sigma^2) = \frac{2\sigma^4}{n}$$

b) Defina un estimador de σ^2 como $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$ y demuestre que este estimador es insesgado

c) ¿El estimador definido en el inciso anterior es UMVUE?

2. Sea X_1, \dots, X_n m.a. de una población con densidad poisson de parametro λ .

$$\mathbb{P}(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x \in \{0, 1, 2, \dots\}$$

a) Calcule la Cota Inferior de Cramer Rao para esta familia y verifique que:

$$CICR(\lambda) = \frac{\lambda}{n}$$

b) Defina el estimador $\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i$, pruebe que el estimador es insesgado

c) Calcule la $\text{Var}(\hat{\lambda})$. El estimador $\hat{\lambda}$ es UMVUE?

3. Sea X_1, \dots, X_n m.a. de una población con densidad Normal de parámetros μ, σ^2 (ambos desconocidos).

a) Calcule la Cota Inferior de Cramer Rao para los estimadores de σ^2 . Verifique que:

$$CICR(\sigma^2) = \frac{2\sigma^4}{n}$$

b) Defina un estimador de σ^2 como $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ y demuestre que este estimador es insesgado

c) El estimador definido en el inciso anterior es UMVUE?

Si nuestro objetivo es encontrar estimadores **UMVUES** para θ , surge de manera natural encontrar un método que nos ayude a encontrar dichos estimadores. Una forma de encontrar **UMVUES** es por medio de la función SCORE, pues recordemos que para que la CICR sea alcanzada por la varianza de un estimador requerimos que:

$$\left| \rho_{\hat{\theta}S(\theta; \underline{X})} \right| = 1$$

Es decir, necesita haber una relación lineal entre el estimador insesgado $\hat{\theta}$ y la función SCORE $S(\theta; \underline{X})$ de la forma:

$$\hat{\theta} = aS(\theta; \underline{X}) + b \quad \text{o} \quad S(\theta; \underline{X}) = a\hat{\theta} + b$$

Con a y b constantes reales.

Por ejemplo, en el modelo $Bernulli(\theta)$ se tiene que:

$$S(\theta; \underline{X}) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \text{Log}(f_{X_i}(X_i; \theta)) = \frac{1}{p(1-p)} \sum_{i=1}^n X_i - \frac{n}{p(1-p)}$$

Luego entonces si definimos $a = \frac{n}{p(1-p)}$ y $b = \frac{n}{p(1-p)}$ obtendríamos que:

$$S(\theta; \underline{X}) = a\bar{X} + b$$

y como \bar{X} es un estimador insesgado para θ , se sigue que $\hat{\theta} = \bar{X}$ es UMVUE.

Ejercicio 3.2.6. Resuelva lo siguiente:

1. Considere X_1, \dots, X_n m.a. del modelo poisson de parámetro λ :

$$f_X(x) = \mathbb{P}(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x \in \{0, 1, \dots, n\}$$

- a) Encuentre la función SCORE $S(X_1, \dots, X_n; \lambda) = S(\underline{X}; \lambda)$
- b) Encuentre con base en la pregunta anterior un estimador insesgado para λ que sea UMVUE

2. Considere X_1, \dots, X_n m.a. del modelo Normal con varianza σ^2 conocida y μ desconocida

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

- a) Encuentre la función SCORE $S(X_1, \dots, X_n; \mu) = S(\underline{X}; \mu)$
- b) Encuentre con base en la pregunta anterior un estimador insesgado para μ que sea UMVUE

Muchas veces los estimadores insesgados que encontremos para cierto parámetro no lograrán alcanzar la CICR, por lo que es necesario cuantificar que tan lejos estamos de la cota, para ello se define la eficiencia de un estimador como sigue:

Definición 3.2.14 (Eficiencia de un estimador). Sea $\hat{\theta}$ un estimador insesgado para θ , definimos la eficiencia de $\hat{\theta}$ como el numero:

$$eff(\hat{\theta}) = \frac{CICR(\theta)}{Var(\hat{\theta})}$$

Notamos entonces que $eff(\hat{\theta}) \leq 1$ y es si $eff(\hat{\theta}) = 1$ entonces $\hat{\theta}$ es UMVUE pero no al revés. Es decir, se puede demostrar que existen estimadores que son insesgados de mínima varianza pero que no alcanzarán la CICR.

Definición 3.2.15 (Eficiencia Relativa). Sean $\hat{\theta}_1, \hat{\theta}_2$ dos estimadores insesgados para θ . Definimos la eficiencia relativa entre $\hat{\theta}_1$ y $\hat{\theta}_2$ como:

$$eff.rel(\hat{\theta}_1, \hat{\theta}_2) = \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}$$

Notamos entonces que $eff.rel(\hat{\theta}_1, \hat{\theta}_2) < 1$ cuando $\hat{\theta}_2$ es mas eficiente que $\hat{\theta}_1$ y que $eff.rel(\hat{\theta}_1, \hat{\theta}_2) > 1$ cuando $\hat{\theta}_1$ es mas eficiente que $\hat{\theta}_2$ mientras que $eff.rel(\hat{\theta}_1, \hat{\theta}_2) = 1$ si y solamente si $\hat{\theta}_1$ es igual de eficiente que $\hat{\theta}_2$. Muchas veces las varianzas de algunos estimadores no alcanzarán la CICR sin embargo conforme la muestra va creciendo dicha cota se puede ir alcanzando, cuando eso ocurre decimos que nuestro estimador es asintóticamente eficiente.

Definición 3.2.16 (Asintóticamente eficiente). Sea $\hat{\theta}$ un estimador insesgado para θ , decimos que $\hat{\theta}$ es asintóticamente eficiente si:

$$\lim_{n \rightarrow \infty} \frac{CICR(\theta)}{Var(\hat{\theta})} = 1$$

Un comportamiento deseable de un estimador es que conforme vayamos obteniendo mas información (aumentado el tamaño de muestra) dicho estimador tenga mejores propiedades al momento de estimar, esta idea se resume en la propiedad de consistencia de un estimador.

Definición 3.2.17 (Estimador Consistente). Sea X_1, \dots, X_n m.a. de un fenómeno aleatorio con distribución $F_X(x; \theta)$ y θ un parámetro desconocido. Denotemos por $\hat{\theta}_n$ al estimador generado por la muestra de tamaño n . Decimos que $\hat{\theta}_n$ es consistente si $\forall \varepsilon > 0$:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \hat{\theta}_n - \theta \right| < \varepsilon \right) = 1$$

o equivalentemente, utilizando complementos:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \hat{\theta}_n - \theta \right| \geq \varepsilon \right) = 0$$

Demostrar la consistencia de un estimador es muy complicado, pues requiere demostrar la convergencia en probabilidad del estimador hacia el parámetro desconocido, sin embargo gracias a la desigualdad de *Tchebycheff* podemos encontrar una alternativa a esto.

Teorema 3.2.4 (Desigualdad de Tchebycheff). *Sea X una v.a. tal que $\mathbb{E}(|X|^k) < \infty$ entonces:*

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}(|X|^k)}{\varepsilon^k}$$

Luego entonces gracias a esta desigualdad utilizando $k = 2$ y suponiendo que $\mathbb{E}(|\hat{\theta}|^2) < \infty$ podemos afirmar que si:

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) = 0$$

Entonces $\hat{\theta}_n$ es consistente, sin embargo el hecho de que $\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) > 0$ no implicará que el estimador no sea consistente en cuyo caso tendríamos que ir directamente a la definición de consistencia del estimador para poder concluir.

Ejemplo 3.2.5. *Sea X_1, \dots, X_n m.a. de un población $N(\mu, \sigma^2)$ (Ambos parámetros desconocidos). Y definamos el estimador para σ^2 como:*

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

Entonces recordando que $\frac{(n-1)}{\sigma^2} \hat{\sigma}^2 \sim \chi_{(n-1)}^2$, se puede demostrar que el estimador es insesgado, es decir que: $\mathbb{E}(\hat{\sigma}^2) = \sigma^2$. Por otro lado:

$$\text{Var}(\hat{\sigma}^2) = \frac{\sigma^4}{(n-1)^2} \text{Var}\left(\frac{(n-1)}{\sigma^2} \hat{\sigma}^2\right) = \frac{2\sigma^4}{n-1}$$

Veamos si el estimador es eficiente, para ello calculemos la CICR para estimadores de σ^2

$$\text{CICR}(\sigma^2) = \frac{2\sigma^4}{n}$$

Luego entonces:

$$\text{eff}(\hat{\sigma}^2) = \frac{\frac{2\sigma^4}{n}}{\frac{2\sigma^4}{n-1}} = \frac{n-1}{n} = 1 - \frac{1}{n} < 1$$

Por lo tanto bajo la definición de eficiencia, $\hat{\sigma}^2$, no es eficiente, sin embargo se puede probar (mas adelante) que si es UMVUE. Por otro lado vemos que:

$$\lim_{n \rightarrow \infty} \frac{CICR(\sigma^2)}{\text{Var}(\hat{\sigma}^2)} = 1$$

Por lo tanto $\hat{\sigma}^2$ es asintóticamente eficiente. Supongamos ahora que definimos dos estimadores insesgados para σ^2 :

$$\hat{\sigma}_1^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad \hat{\sigma}_2^2 = \frac{(X_1 - X_2)^2}{2}$$

Se puede verificar rápidamente que ambos son estimadores insesgados, ahora calculemos su eficiencia relativa.

$$\text{eff.rel}(\hat{\sigma}_1^2, \hat{\sigma}_2^2) = \frac{\text{Var}(\hat{\sigma}_2^2)}{\hat{\sigma}_1^2} = \frac{2\sigma^4}{\frac{2\sigma^4}{n-1}} = n-1$$

Por lo tanto $\text{eff.rel}(\hat{\sigma}_1^2, \hat{\sigma}_2^2) > 1$ si $n > 2$ por lo que $\hat{\sigma}_1^2$ será mas eficiente que $\hat{\sigma}_2^2$ y son igual de eficientes solo cuando $n = 1$

Ejercicio 3.2.7. Resuelva lo siguiente:

1. Considere X_1, \dots, X_n m.a. del modelo con función de densidad dada por:

$$f_X(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \quad x > 0$$

Se definen dos estimadores para θ :

$$\hat{\theta}_1 = \sum_{i=1}^n \frac{X_i}{n} \quad \hat{\theta}_2 = \frac{X_1 + (n-1)X_2}{n}$$

Verifique que ambos estimadores son insesgados y calcule lo siguiente :

- a) $\text{eff}(\hat{\theta}_1)$
- b) $\text{eff}(\hat{\theta}_2)$
- c) $\text{eff.rel}(\hat{\theta}_1, \hat{\theta}_2)$
- d) ¿Es $\hat{\theta}_1$ consistente? (Justifique su Respuesta)
- e) ¿Es $\hat{\theta}_2$ consistente? (Justifique su Respuesta)

A partir de este momento, desarrollaremos la teoría suficiente para generar UMVUES por medio de un método al que llamaremos Rao-Blackwell, para ellos introduciremos el concepto de suficiencia.

Definición 3.2.18 (Estadística Suficiente). Sea X_1, \dots, X_n m.a. de población con función de distribución $F_X(x; \underline{\theta})$, con $\underline{\theta}$ un vector de parámetros desconocidos. Sea $T(X_1, \dots, X_n)$ una estadística. Decimos que T es suficiente para $\underline{\theta}$ si la densidad condicional conjunta:

$$f_{\underline{X}|T}(x_1, \dots, x_n | T(x_1, \dots, x_n) = t)$$

No depende de $\underline{\theta}$

Observación importante: T es una función de la muestra que no necesariamente es una función real, puede ser una función vectorial

Conceptualmente nos dice que una vez que sabemos el valor que tomó el estadístico T , entonces ya tenemos toda la información necesaria para estimar a θ .

Ejercicio 3.2.8. Probar con la definición anterior que si X_1, X_2, X_3 es una m.a. de una población Bernulli (θ), entonces:

$$T(X_1, X_2, X_3) = \sum_{i=1}^3 X_i$$

Es una estadística suficiente para θ

Teorema 3.2.5 (Principio de Suficiencia). Supongamos que tenemos dos muestras aleatorias independientes $\underline{X} = (X_1, \dots, X_n)$ y $\underline{Y} = (Y_1, \dots, Y_n)$ tal que:

$$T(\underline{X}) = T(\underline{Y})$$

Con T una estadística suficiente, entonces, con ambas muestras debo de llegar a la misma conclusión al momento de inferir

Demostrar que una estadística es suficiente no es fácil pues requiere el cálculo de densidades condicionales que involucran conocer la densidad del estadístico, afortunadamente contamos con el siguiente teorema que nos permite saber si una estadística es suficiente o no.

Teorema 3.2.6 (Teorema de Factorización de Neyman Pearson). Sea $\underline{X} = X_1, \dots, X_n$ m.a. de una población con densidad $f_X(x; \underline{\theta})$ y sea $T(x_1, \dots, x_n)$ un estadístico.

$T(X_1, \dots, X_n)$ es suficiente para $\underline{\theta}$ si y solo si existen h y g funciones positivas tales que:

$$\ell(\underline{\theta}; x_1, \dots, x_n) = f_{\underline{X}}(x_1, \dots, x_n; \underline{\theta}) = g(t(x_1, \dots, x_n); \underline{\theta})h(x_1, \dots, x_n)$$

Demostración. La prueba se basa en el hecho de que al ser T función de \underline{X} , entonces la densidad

conjunta de (\underline{X}, T) cumple con lo siguiente:

$$f_{\underline{X}T}(\underline{x}, t; \theta) = f_{\underline{X}}(\underline{x}; \theta) \mathbb{1}_{\{T(\underline{x})=t\}} \quad (3.2)$$

\Rightarrow

Supongamos que T es suficiente y demostremos que:

$$f(x_1, \dots, x_n; \theta) = f_{\underline{X}}(\underline{x}; \theta) = g(t(x_1, \dots, x_n); \theta)h(x_1, \dots, x_n)$$

Consideremos $f_{\underline{X}}(\underline{x}; \theta)$ y definamos $t = T(\underline{x})$, entonces por (3.2) tenemos que:

$$\begin{aligned} f_{\underline{X}}(\underline{x}; \theta) &= f_{\underline{X}}(\underline{x}; \theta) \mathbb{1}_{\{T(\underline{x})=t\}} = f_{\underline{X}T}(\underline{x}, t; \theta) \\ &= f_{\underline{X}|T}(\underline{x}|t; \theta) f_T(t; \theta) \end{aligned}$$

Como T es suficiente se sigue que $f_{\underline{X}|T}(\underline{x}|t; \theta)$ no depende de θ mientras que $f_T(t; \theta)$ si depende de θ y solo depende de \underline{x} a través de T , entonces haciendo:

$$\begin{aligned} h(x_1, \dots, x_n) &= h(\underline{x}) = f_{\underline{X}|T}(\underline{x}|t; \theta) \\ g(t; \theta) &= f_T(t; \theta) \end{aligned}$$

Concluimos entonces que:

$$f(x_1, \dots, x_n; \theta) = f_{\underline{X}}(\underline{x}; \theta) = g(t(x_1, \dots, x_n); \theta)h(x_1, \dots, x_n)$$

\Leftarrow

Supongamos que $f(x_1, \dots, x_n; \theta) = f_{\underline{X}}(\underline{x}; \theta) = g(t(x_1, \dots, x_n); \theta)h(x_1, \dots, x_n)$ y demostraremos que T es suficiente.

Consideremos la densidad de T :

$$\begin{aligned} f_T(t; \theta) &= \int_{\mathbb{R}^n} f_{\underline{X}T}(\underline{x}, t; \theta) d\underline{x} \\ &= \int_{\mathbb{R}^n} f_{\underline{X}}(\underline{x}; \theta) \mathbb{1}_{\{T(\underline{x})=t\}} d\underline{x} \\ &= \int_{\mathbb{R}^n} g(t; \theta)h(\underline{x}) \mathbb{1}_{\{T(\underline{x})=t\}} d\underline{x} \\ &= g(t; \theta) \int_{\mathbb{R}^n} h(\underline{x}) \mathbb{1}_{\{T(\underline{x})=t\}} d\underline{x} \end{aligned}$$

La segunda igualdad se debe a (3.2), la tercera por la hipótesis que tenemos mientras que la cuarta igualdad se debe a que integramos respecto de \underline{x} .

Finalmente para probar que T es suficiente debemos verificar que $f_{\underline{X}|T}(\underline{x}|t; \underline{\theta})$ no depende de $\underline{\theta}$, veamos:

$$\begin{aligned} f_{\underline{X}|T}(\underline{x}|t; \underline{\theta}) &= \frac{f_{\underline{X}T}(\underline{x}, t; \underline{\theta})}{f_T(t; \underline{\theta})} \\ &= \frac{f_{\underline{X}}(\underline{x}; \underline{\theta}) \mathbb{1}_{\{T(\underline{x})=t\}}}{g(t; \underline{\theta}) \int_{\mathbb{R}^n} h(\underline{x}) \mathbb{1}_{\{T(\underline{x})=t\}} d\underline{x}} \\ &= \frac{g(t; \underline{\theta}) h(\underline{x}) \mathbb{1}_{\{T(\underline{x})=t\}}}{g(t; \underline{\theta}) \int_{\mathbb{R}^n} h(\underline{x}) \mathbb{1}_{\{T(\underline{x})=t\}} d\underline{x}} \\ &= \frac{h(\underline{x}) \mathbb{1}_{\{T(\underline{x})=t\}}}{\int_{\mathbb{R}^n} h(\underline{x}) \mathbb{1}_{\{T(\underline{x})=t\}} d\underline{x}} \end{aligned}$$

Lo que demuestra que $f_{\underline{X}|T}$ no depende de $\underline{\theta}$ por lo tanto T es suficiente. \square

Observe que según el teorema, en la factorización que estamos haciendo de la verosimilitud, la función g si puede depender de $\underline{\theta}$ y solo depende de la muestra a través del estadístico mientras que h es una función que depende de la muestra pero que no puede depender de parámetros desconocidos.

Ejemplo 3.2.6. Sea X_1, \dots, X_n m.a. de una población Bernulli (θ) entonces:

$$\ell(\theta; x_1, \dots, x_n) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} \prod_{i=1}^n \mathbb{1}_{\{0,1\}}(x_i)$$

Definamos entonces a $g : \mathbb{R} \rightarrow \mathbb{R}$ como

$$g(t; \theta) = \theta^t (1 - \theta)^{n-t}$$

y a $h : \mathbb{R}^n \rightarrow \mathbb{R}$ como

$$h(x_1, \dots, x_n) = \prod_{i=1}^n \mathbb{1}_{\{0,1\}}(x_i)$$

entonces queda claro que:

$$\ell(\theta; x_1, \dots, x_n) = g\left(\sum_{i=1}^n x_i; \theta\right) h(x_1, \dots, x_n)$$

De donde concluimos que $T(X_1, \dots, X_n) = \sum_{i=1}^n X_i$ es una estadística suficiente para θ

Analicemos otro ejemplo:

Ejemplo 3.2.7. Sea X_1, \dots, X_n m.a. de una población $\text{Gamma}(\alpha, \beta)$ y supongamos α conocido entonces

$$\ell(\beta; x_1, \dots, x_n) = \prod_{i=1}^n \frac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i} \mathbb{1}_{(0, \infty)}(x_i) = \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} e^{-\beta \sum_{i=1}^n x_i} \prod_{i=1}^n x_i^{\alpha-1} \mathbb{1}_{(0, \infty)}(x_i)$$

Definamos entonces a $g: \mathbb{R} \rightarrow \mathbb{R}$ como

$$g(t; \beta) = \beta^{n\alpha} e^{-\beta t}$$

y a $h: \mathbb{R}^n \rightarrow \mathbb{R}$ como

$$h(x_1, \dots, x_n) = \frac{1}{\Gamma(\alpha)^n} \prod_{i=1}^n x_i^{\alpha-1} \mathbb{1}_{(0, \infty)}(x_i)$$

entonces queda claro que:

$$\ell(\beta; x_1, \dots, x_n) = g\left(\sum_{i=1}^n x_i; \beta\right) h(x_1, \dots, x_n)$$

De donde concluimos que $T(X_1, \dots, X_n) = \sum_{i=1}^n X_i$ es una estadística suficiente para β

Ejercicio 3.2.9. Resuelva lo siguiente:

1. Considere una m.a. de tamaño n . Para cada una de las siguientes distribuciones encuentre un estadístico suficiente para el parámetro θ

- Bernoulli (θ)
- Binomial (k, θ), k conocida
- Poisson (θ)
- Geométrica (θ)
- Exponencial (θ)
- Gamma (k, θ), k conocida
- BinNeg (k, θ), k conocida
- Normal (μ, θ^2), μ conocida
- Normal (θ, σ^2), σ^2 conocida
- Beta (a, θ), a conocida

- Beta (θ, b) , b conocida
- Pareto (θ, b) , b conocida

Una observación interesante es que gracias al teorema de factorización, se puede ver que cualquier transformación biyectiva de un estadístico suficiente, es también un estadístico suficiente, en efecto pues consideremos T_1 suficiente y $T_2 = f(T_1)$ tal que f es biyectiva, entonces podemos escribir $T_1 = f^{-1}(T_2)$. Queremos probar que T_2 también es suficiente, esto último lo verificamos gracias al teorema de factorización como:

$$\begin{aligned}\ell(\theta; x_1, \dots, x_n) &= g(T_1; \theta) h(x_1, \dots, x_n) \\ &= g(f^{-1}(T_2); \theta) h(x_1, \dots, x_n) \\ &= g_1(T_2; \theta) h(x_1, \dots, x_n)\end{aligned}$$

Donde $g_1 = g \circ f^{-1}$, por lo que se concluye que T_2 también es suficiente.

Ejemplo 3.2.8. En unos de los ejemplos anteriores probamos que $T = \sum_{i=1}^n X_i$ es una estadística suficiente para θ en el modelo Bernulli (θ) , luego entonces dado que la función $f_1(t) = \frac{t}{n}$ es biyectiva se sigue que:

$$T_2(X_1, \dots, X_n) = f_1(T) = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$$

También es una estadística suficiente

Ejercicio 3.2.10. Sea X_1, \dots, X_n m.a. de una población con densidad $f_X(x; \underline{\theta})$ defina la estadística:

$$T(X_1, \dots, X_n) = (X_1, \dots, X_n)$$

Prueba que T es suficiente.

Veamos como funciona el teorema de factorización con dos parámetros desconocidos.

Ejemplo 3.2.9. Supongamos que tenemos X_1, \dots, X_n m.a. de $N(\mu, \sigma^2)$, mostrar que

$$T := \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$$

es una estadística suficiente para estimar a μ y a σ^2 . (Obs: cuando ocurre esto se dice que

$T_1 = \sum_{i=1}^n X_i$ y $T_2 := \sum_{i=1}^n X_i^2$ son estadísticas conjuntamente suficientes)

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

Luego entonces, definiendo $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ y $h : \mathbb{R}^n \rightarrow \mathbb{R}$ como:

$$g(t_1, t_2; \mu, \sigma^2) := (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(t_2 - 2\mu t_1 + n\mu^2)} \quad ; \quad h(x_1, \dots, x_n) = 1$$

Se tendría que $T_1 = \sum_{i=1}^n X_i$ y $T_2 = \sum_{i=1}^n X_i^2$ son estadísticas conjuntamente suficientes para μ y σ^2

Ejercicio: Demostrar que también $T_1 = \bar{X}$ y $T_2 = S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ son también estadísticas conjuntamente suficientes para μ y σ^2 .

Ejercicio 3.2.11. Resuelva lo siguiente:

1. Demuestre que:

$$T_1(X_1, X_2, \dots, X_n) := \prod_{i=1}^n X_i \quad T_2(X_1, X_2, \dots, X_n) := \sum_{i=1}^n X_i$$

Son estadísticas conjuntamente suficientes para (α, β)

Resulta entonces que podemos encontrar muchas estadísticas suficientes, por ejemplo en el caso $Bernulli(\theta)$ se tiene que tanto $T_1 = \sum_{i=1}^n X_i$ como $T_2 = (X_1, \dots, X_n)$ son ambas estadísticas suficientes para θ . ¿Cuál será mejor? ¿Cuál resume más la información? Observemos que T_1 es función de T_2 y que además T_1 resume mejor la información, nuestro objetivo es tener estadísticas suficientes que resuman de la mejor manera posible la información contenida en la muestra. A la estadística suficiente que mejor resume la información le llamamos **estadística suficiente minimal**.

Definición 3.2.19 (Estadístico Suficiente minimal). Dadas dos estadísticas suficientes T_1 y T_2 se dirá que T_1 es igual o más resumido que T_2 si existe una función g (no necesariamente biyectiva) tal que $T_1 = g(T_2)$; En este caso conociendo T_2 se puede conocer T_1 , pero no necesariamente al revés, es decir, el conocimiento de T_1 implica necesariamente el conocimiento de T_2 . Finalmente,

si un estadístico suficiente T_1^* es más resumido que cualquier otro, se dice que T_1^* es un estadístico suficiente minimal. Formalmente, T_1^* es un estadístico suficiente minimal si pasa lo siguiente:

- T_1^* es suficiente
- Si T_2 es cualquier otra estadística suficiente, entonces existe una función g tal que $T_1^* = g(T_2)$

Encontrar estadísticas suficientes minimales usando la definición no es fácil, afortunadamente contamos con el siguiente teorema:

Teorema 3.2.7 (Método para encontrar estadísticas suficientes minimales). Sea $f_X(x; \theta)$ una función de densidad, supongamos que tenemos T_1 una estadística

Si para cualquiera dos muestras $\underline{x} = (x_1, \dots, x_n)$, y $\underline{y} = (y_1, \dots, y_n)$, la razón de verosimilitudes $\frac{f_X(\underline{x}; \theta)}{f_X(\underline{y}; \theta)}$ es constante para θ si y solo si $T_1(\underline{x}) = T_1(\underline{y})$, entonces $T_1(x)$ es suficiente minimal.

Apliquemos este teorema al siguiente ejemplo, supongamos que tenemos $\underline{x} = (x_1, \dots, x_n)$ y $\underline{y} = (y_1, \dots, y_n)$, m.a. del modelo *Bernulli* (θ). Desarrollando el cociente de verosimilitudes tenemos:

$$\begin{aligned} \frac{f_X(\underline{x}; \theta)}{f_X(\underline{y}; \theta)} &= \frac{\theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}}{\theta^{\sum_{i=1}^n y_i} (1 - \theta)^{n - \sum_{i=1}^n y_i}} \\ &= \theta^{\sum_{i=1}^n x_i - \sum_{i=1}^n y_i} (1 - \theta)^{n - \sum_{i=1}^n x_i - n + \sum_{i=1}^n y_i} \\ &= \theta^{\sum_{i=1}^n x_i - \sum_{i=1}^n y_i} (1 - \theta)^{\sum_{i=1}^n y_i - \sum_{i=1}^n x_i} \end{aligned}$$

Si queremos que el cociente de verosimilitudes no dependa de θ , necesitamos que $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ luego entonces, es claro que: $T_1^* = \sum_{i=1}^n X_i$ es una estadística suficiente minimal para θ .

3.2.4. Familia Exponencial de un Parámetro

Dentro de las distribuciones que conocemos existe una clase que es muy utilizada en estadística y que tiene propiedades interesantes a la que denominamos familia exponencial

Definición 3.2.20 (Familia exponencial Uniparamétrica). Sea f una función de densidad, decimos que $f_X(x; \theta)$ pertenece a la familia exponencial uniparamétrica si:

$$f_X(x; \theta) = a(\theta) b(x) e^{c(\theta)d(x)} \quad x \in \mathbb{R}$$

Donde, $a(\theta), c(\theta)$ son funciones reales de θ y no depende de x , mientras que $b(x), d(x)$ son funciones reales de x que no dependen de θ .

Verifiquemos por ejemplo que el modelo bernulli pertenece a la familia exponencial.

$$\begin{aligned}
 f_X(x; \theta) &= \theta^x (1 - \theta)^{1-x} \mathbb{1}_{\{0,1\}}(x) \\
 &= e^{\ln(\theta^x (1-\theta)^{1-x})} \mathbb{1}_{\{0,1\}} \\
 &= e^{x \ln(\theta) + (1-x) \ln(1-\theta)} \mathbb{1}_{\{0,1\}} \\
 &= e^{x \ln(\theta) + \ln(1-\theta) - x \ln(1-\theta)} \mathbb{1}_{\{0,1\}} \\
 &= e^{x \ln\left(\frac{\theta}{1-\theta}\right)} (1 - \theta) \mathbb{1}_{\{0,1\}}
 \end{aligned}$$

Definiendo:

$$a(\theta) = (1 - \theta); \quad b(x) = \mathbb{1}_{\{0,1\}}(x); \quad c(\theta) = \ln\left(\frac{\theta}{1-\theta}\right); \quad d(x) = x$$

Obtenemos que:

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

Por lo tanto la densidad Bernulli pertenece a la familia exponencial.

Ejercicio 3.2.12. Demuestre que:

1. El modelo $Poisson(\lambda)$ pertenece a la familia exponencial. (Identifique adecuadamente quien es $a(\lambda), c(\lambda), b(x), d(x)$)

$$f_X(x) = \mathbb{P}(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x \in \{0, 1, 2, \dots, \}$$

2. El modelo $Binomial(k, p)$ pertenece a la familia exponencial, suponga para ello que el parámetro k es conocido. (Identifique adecuadamente quien es $a(p), c(p), b(x), d(x)$)

$$f_X(x) = \mathbb{P}(X = x) = \binom{k}{x} p^x (1-p)^{k-x} \quad x \in \{0, 1, 2, \dots, k\} \quad p \in (0, 1)$$

3. El modelo $Exponencial(\lambda)$ pertenece a la familia exponencial. (Identifique adecuadamente quien es $a(\lambda), c(\lambda), b(x), d(x)$)

$$f_X(x) = \lambda e^{-\lambda x} \quad x \in (0, \infty)$$

4. El modelo Geométrico(p) pertenece a la familia exponencial. (Identifique adecuadamente quien es $a(p)$, $c(p)$, $b(x)$, $d(x)$)

$$f_X(x) = \mathbb{P}(X = x) = (1 - p)^{x-1} p \quad x \in \{1, 2, 3, \dots\}$$

5. El modelo Normal(μ, σ^2) pertenece a la familia exponencial. Suponga σ^2 conocida. (Identifique adecuadamente quien es $a(\mu)$, $c(\mu)$, $b(x)$, $d(x)$)

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad x \in \mathbb{R}$$

6. El modelo Gamma(α, β) pertenece a la familia exponencial. Suponga α conocida. (Identifique adecuadamente quien es $a(\beta)$, $c(\beta)$, $b(x)$, $d(x)$)

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

La familia exponencial tiene la siguiente propiedad.

Teorema 3.2.8 (Estadística Suficiente Minimal para la Familia Exponencial). *En la familia exponencial uniparamétrica, la estadística:*

$$T(X_1, \dots, X_n) = \sum_{i=1}^n d(X_i)$$

es una estadística suficiente minimal.

Demostración. Haciendo el cociente de verosimilitudes obtenemos que:

$$\begin{aligned} \frac{f_X(\underline{x}; \theta)}{f_X(\underline{y}; \theta)} &= \frac{a(\theta)^n \prod_{i=1}^n b(x_i) e^{c(\theta) \sum_{i=1}^n d(x_i)}}{a(\theta)^n \prod_{i=1}^n b(y_i) e^{c(\theta) \sum_{i=1}^n d(y_i)}} \\ &= \frac{\prod_{i=1}^n b(x_i)}{\prod_{i=1}^n b(y_i)} e^{c(\theta) (\sum_{i=1}^n d(x_i) - \sum_{i=1}^n d(y_i))} \end{aligned}$$

Para que esto no depende de θ necesitamos que

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

Se sigue por el teorema (3.2.7) que $\sum_{i=1}^n d(x_i)$ es una estadística suficiente minimal. □

3.2.5. Familia Exponencial varios parámetros

Ahora veremos la generalización de la familia exponencial cuando tenemos un vector de parámetros.

Definición 3.2.21 (Familia exponencial). Sea f una función de densidad, decimos que $f_X(x; \underline{\theta})$ con $\underline{\theta} := (\theta_1, \dots, \theta_k)$ un vector de parámetros, pertenece a la familia exponencial si la densidad puede expresarse como:

$$f_X(x; \underline{\theta}) = a(\underline{\theta}) b(x) e^{\underline{c}(\underline{\theta}) \cdot \underline{d}(x)} \quad x \in \mathbb{R}$$

Donde:

- $a(\underline{\theta})$ es una función de \mathbb{R}^k a \mathbb{R} que solo depende de $\underline{\theta}$ o constantes conocidas
- $b(x)$ es una función de \mathbb{R} a \mathbb{R} que solo depende de x o de constantes conocidas
- $\underline{c}(\underline{\theta}) = (c_1(\underline{\theta}), \dots, c_m(\underline{\theta}))$ es una función vectorial de \mathbb{R}^k a \mathbb{R}^m que solo depende de $\underline{\theta}$ o constantes conocidas
- $\underline{d}(x) = (d_1(x), \dots, d_m(x))$ es una función vectorial de \mathbb{R} a \mathbb{R}^m que solo depende de x o de constantes conocidas
- $\underline{c}(\underline{\theta}) \cdot \underline{d}(x)$ debe considerarse como producto punto, y por tanto:

$$\underline{c}(\underline{\theta}) \cdot \underline{d}(x) = \sum_{j=1}^m c_j(\underline{\theta}) d_j(x)$$

Por lo tanto la densidad se puede expresar como:

$$f_X(x; \underline{\theta}) = a(\underline{\theta}) b(x) e^{\sum_{j=1}^m c_j(\underline{\theta}) d_j(x)} \quad x \in \mathbb{R}$$

Cuando $m = k$ se dice que la densidad exponencial es no singular mientras que cuando $m \neq k$ se dice que la densidad es singular

Teorema 3.2.9 (Estadística Suficiente en la familia exponencial). En la familia de densidades exponencial la estadística

$$T = \sum_{i=1}^n \underline{d}(X_i)$$

es conjuntamente suficiente para $\underline{\theta}$

Demostración. Utilizando el teorema de factorización (3.2.6) tenemos que:

$$f_X(x_1, \dots, x_n; \underline{\theta}) = a(\underline{\theta})^n \prod_{i=1}^n b(x_i) e^{\underline{c}(\underline{\theta}) \cdot \sum_{i=1}^n \underline{d}(x_i)}$$

Por lo tanto definiendo las funciones:

$$\begin{aligned} g\left(\sum_{i=1}^n \underline{d}(x_i); \underline{\theta}\right) &= a(\underline{\theta})^n e^{\underline{c}(\underline{\theta}) \cdot \sum_{i=1}^n \underline{d}(x_i)} \\ h(x_1, \dots, x_n) &= \prod_{i=1}^n b(x_i) \end{aligned}$$

Por lo que aplicando el teorema (3.2.6) concluimos que en efecto $T = \sum_{i=1}^n \underline{d}(X_i)$ es suficiente o conjuntamente suficiente para $\underline{\theta}$ \square

Otra propiedad importante es que en esta familia que la estadística $T = \sum_{i=1}^n \underline{d}(X_i)$ es también suficiente minimal bajo ciertas condiciones.

Teorema 3.2.10 (Estadística Suficiente Minimal en la familia exponencial). *En la familia de densidades exponencial la estadística*

$$T = \sum_{i=1}^n \underline{d}(X_i)$$

es conjuntamente suficiente minimal para $\underline{\theta}$ siempre y cuando se garantice que la función $\underline{c}(\underline{\theta})$ sea tal que la siguiente ecuación no dependa de $\underline{\theta}$ si y solo si $\sum_{i=1}^n \underline{d}(x_i) = \sum_{i=1}^n \underline{d}(y_i)$, para todo $\underline{\theta}$

$$\underline{c}(\underline{\theta}) \cdot \left(\sum_{i=1}^n \underline{d}(x_i) - \sum_{i=1}^n \underline{d}(y_i) \right)$$

Demostración. Se deja como ejercicio la demostración haciendo uso de (3.2.7) \square

Ejemplo 3.2.10. *Encuentra una estadística suficiente minimal en el modelo Normal con ambos parámetros desconocidos: $\underline{\theta} = (\mu, \sigma^2)$*

$$f_X(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (x - \mu)^2\right)$$

Primero veamos que la densidad pertenece a la familia exponencial.

$$\begin{aligned}
 f_X(x; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}x^2 + \frac{1}{2\sigma^2}x\mu - \frac{1}{2\sigma^2}\mu^2\right) \\
 &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\mu^2\right) \exp\left(-\frac{1}{2\sigma^2}x^2 + \frac{1}{2\sigma^2}x\mu\right) \\
 &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\mu^2\right) \exp\left(\left(-\frac{1}{2\sigma^2}, \frac{\mu}{2\sigma^2}\right) \cdot (x^2, x)\right)
 \end{aligned}$$

Definamos las siguiente funciones

- $a(\underline{\theta}) = a(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\mu^2\right)$
- $b(x) = 1$
- $\underline{c}(\underline{\theta}) = \underline{c}(\mu, \sigma^2) = \left(-\frac{1}{2\sigma^2}, \frac{\mu}{2\sigma^2}\right)$
- $\underline{d}(x) = (x^2, x)$

Enonces $f_X(x; \mu, \sigma^2)$ pertenece a la familia exponencial. Por lo que una estadística suficiente es:

$$T = \sum_{i=1}^n \underline{d}(X_i) = \left(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i \right)$$

Para verificar que la estadística es suficiente minimal tenemos que ver que la ecuación no dependa de (μ, σ^2) si y solo si $\sum_{i=1}^n \underline{d}(x_i) = \sum_{i=1}^n \underline{d}(y_i)$

$$\underline{c}(\mu, \sigma^2) \cdot \left(\sum_{i=1}^n \underline{d}(x_i) - \sum_{i=1}^n \underline{d}(y_i) \right) \quad (3.3)$$

Es fácil ver que cuando $\sum_{i=1}^n \underline{d}(x_i) = \sum_{i=1}^n \underline{d}(y_i)$ implica que la ecuación (3.3) no depende de θ , lo que realmente hay que probar es que si (3.3) no depende de los parámetros implica que $\sum_{i=1}^n \underline{d}(x_i) = \sum_{i=1}^n \underline{d}(y_i)$. Supongamos entonces que (3.3) no depende de (μ, σ^2) , entonces:

$$\left(-\frac{1}{2\sigma^2}, \frac{\mu}{2\sigma^2}\right) \cdot \left(\sum_{i=1}^n \underline{d}(x_i) - \sum_{i=1}^n \underline{d}(y_i) \right) = h(\underline{x}, \underline{y})$$

Lo anterior debe de ser valido para cualquier μ y σ^2 . Primero probaremos que $h(\underline{x}, \underline{y}) = 0$, para ello supongamos sin perdida de generalidad dos pares de posibles valores para los parámetros:

$(\mu = 0, \sigma^2 = 1)$, $(\mu = 0, \sigma^2 = 2)$, entonces:

$$\begin{aligned} \left(-\frac{1}{2}, 0\right) \cdot \left(\sum_{i=1}^n \underline{d}(x_i) - \sum_{i=1}^n \underline{d}(y_i)\right) &= h(\underline{x}, \underline{y}) \\ \left(-\frac{1}{4}, 0\right) \cdot \left(\sum_{i=1}^n \underline{d}(x_i) - \sum_{i=1}^n \underline{d}(y_i)\right) &= h(\underline{x}, \underline{y}) \end{aligned}$$

De estas dos ecuaciones se tiene entonces que:

$$\frac{1}{2}h(\underline{x}, \underline{y}) = h(\underline{x}, \underline{y}) \Rightarrow h(\underline{x}, \underline{y}) = 0$$

Entonces tenemos que

$$\left(-\frac{1}{2\sigma^2}, \frac{\mu}{2\sigma^2}\right) \cdot \left(\sum_{i=1}^n \underline{d}(x_i) - \sum_{i=1}^n \underline{d}(y_i)\right) = 0$$

La última ecuación es válida para todo μ, σ^2 , dado que los puntos en \mathbb{R}^2 que cumplen con la ecuación $\left(-\frac{1}{2\sigma^2}, \frac{\mu}{2\sigma^2}\right)$ no están sobre una recta, entonces se concluye que necesariamente que:

$$\left(\sum_{i=1}^n \underline{d}(x_i) - \sum_{i=1}^n \underline{d}(y_i)\right) = 0 \Rightarrow \sum_{i=1}^n \underline{d}(x_i) = \sum_{i=1}^n \underline{d}(y_i).$$

De donde se concluye que:

$$T = \sum_{i=1}^n \underline{d}(X_i) = \left(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i\right)$$

Es suficiente minimal.

Ejercicio 3.2.13. Resuelva lo siguiente:

- Encuentre una estadística suficiente minimal para el modelo $\text{Gamma}(\alpha, \beta)$ con ambos parámetros desconocidos
- Encuentre una estadística suficiente minimal para el modelo $\text{Beta}(\alpha, \beta)$ con ambos parámetros desconocidos

Ya sabemos como encontrar estadísticas suficientes que resumen mejor la información para estimar al parámetro desconocido, pero ¿Para qué nos sirve?. El siguiente teorema nos dice como podemos utilizar una estadística suficiente mínima.

Teorema 3.2.11 (Rao-Blackwell). *Sea una población con función de densidad $f_X(x; \theta)$ y $\hat{\theta}$ un estimador insesgado para θ y T_1 un estadístico suficiente, entonces la estadística $g(T_1)$ dada por:*

$$g(T_1) = \mathbb{E}(\hat{\theta} \mid T_1)$$

Cumple con:

- Es un estadístico función de la estadística suficiente
- $g(T_1)$ es un estimador insesgado ($\mathbb{E}(g(T_1)) = \theta$)
- $\text{Var}(g(T_1)) \leq \text{Var}(\hat{\theta})$

Es decir, el teorema nos dice que un estimador insesgado puede ser mejorado calculando la esperanza condicional del estimador $\hat{\theta}$ con respecto a un estadístico suficiente. Continuando con la teoría del teorema de Rao-Blackwell tenemos el siguiente ejemplo:

Ejemplo 3.2.11. *Sea X_1, \dots, X_n m.a. del modelo Bernulli(θ). Sabemos que $\hat{\theta} = X_1$ es un estimador insesgado y ya sabemos que $T_1 = \sum_{i=1}^n X_i$ es una estadística suficiente. Entonces según el Teorema 3.2.11 sabemos que:*

$$g(T_1) = \mathbb{E}(\hat{\theta} \mid T_1)$$

es un mejor estimador pues el teorema afirma dos cosas, primero que $g(T_1)$ es insesgado y luego que $g(T_1)$ tiene menor varianza. Calculemos entonces $g(T_1)$:

$$\begin{aligned} g(u) &= \mathbb{E}(\hat{\theta} \mid T_1 = u) = \mathbb{E}\left(X_1 \mid \sum_{i=1}^n X_i = u\right) = 0\mathbb{P}\left(X_1 = 0 \mid \sum_{i=1}^n X_i = u\right) + 1\mathbb{P}\left(X_1 = 1 \mid \sum_{i=1}^n X_i = u\right) \\ &= \frac{\mathbb{P}(X_1 = 1, \sum_{i=1}^n X_i = u)}{\mathbb{P}(\sum_{i=1}^n X_i = u)} = \frac{\mathbb{P}(X_1 = 1, \sum_{i=2}^n X_i = u-1)}{\mathbb{P}(\sum_{i=1}^n X_i = u)} = \frac{\mathbb{P}(X_1 = 1) \mathbb{P}(\sum_{i=2}^n X_i = u-1)}{\mathbb{P}(\sum_{i=1}^n X_i = u)} \\ &= \frac{\theta \binom{n-1}{u-1} \theta^{u-1} (1-\theta)^{n-1-(u-1)}}{\binom{n}{u} \theta^u (1-\theta)^{n-u}} = \frac{\binom{n-1}{u-1}}{\binom{n}{u}} = \frac{u}{n} \end{aligned}$$

Por lo tanto $g(u) = \frac{u}{n}$ por lo tanto $g(T_1) = \frac{T_1}{n}$ de donde concluimos que el estimador

$$\frac{T_1}{n} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$$

es mejor estimador que X_1 de hecho se probó que es UMVUE.

La idea entonces fue muy simple, dar un estimador insesgado inicial $\hat{\theta} = X_1$ y luego obtener la esperanza condicional de este estimador con respecto a una estadística suficiente. Lo que nos conviene siempre es condicionar respecto a una estadística suficiente minimal, lamentablemente condicionar respecto a la suficiente minimal no necesariamente genera al estimador UMVUE (de mínima varianza). Para que el teorema de Rao-Blackwell genere al UMVUE requerimos el concepto de completos.

Definición 3.2.22 (Familia de Densidades Completa). *La familia de densidades $\{f_X(x; \theta); \theta \in \Theta\}$ se dice que es una familia completa si **para todo** $\theta \in \Theta$ si tiene lo siguiente:*

$$\mathbb{E}(Z(X)) = 0 \quad \Rightarrow \quad \mathbb{P}(Z(X) = 0) = 1$$

Para Z una función real.

Luego entonces, para demostrar que una familia es completa debemos verificar que si $\mathbb{E}(Z(X)) = 0$ (**para todo** θ) esto implica que la función Z es idénticamente cero. Veamos un ejemplo:

Ejemplo 3.2.12. *Probaremos que la familia de densidades binomial es completa:*

$$\left\{ f_X(x; \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}; \theta \in (0, 1) \right\}$$

Tomemos entonces un $\theta \in (0, 1)$ arbitrario y Z cualquier función, por hipótesis vamos a suponer que la función Z es tal que: $\mathbb{E}(Z(X)) = 0$, queremos probar entonces que $Z(X) = 0$. Sabemos entonces que:

$$\begin{aligned} \mathbb{E}(Z(X)) &= \sum_{i=1}^n Z(j) \binom{n}{j} \theta^j (1 - \theta)^{n-j} = 0 \\ &= \sum_{j=0}^n Z(j) \binom{n}{j} \left(\frac{\theta}{1 - \theta} \right)^j (1 - \theta)^n = 0 \end{aligned}$$

Como $(1 - \theta) > 0$ conclumos que por hipótesis se tiene que:

$$\sum_{j=0}^n Z(j) \binom{n}{j} \left(\frac{\theta}{1 - \theta} \right)^j = 0$$

Definamos $\alpha = \frac{\theta}{1 - \theta}$, luego como $\theta \in (0, 1)$ se tiene que $\alpha \in (0, \infty)$, se sigue entonces que:

$$Z(0) \binom{n}{0} (\alpha)^0 + Z(1) \binom{n}{1} (\alpha)^1 + \cdots + Z(n) \binom{n}{n} (\alpha)^n = 0$$

Definamos al polinomio $P(\alpha)$ de grado n como:

$$P(\alpha) = Z(0) \binom{n}{0} (\alpha)^0 + Z(1) \binom{n}{1} (\alpha)^1 + \cdots + Z(n) \binom{n}{n} (\alpha)^n$$

Sin embargo sabemos que $P(\alpha) = 0$ para todo $\alpha > 0$ luego entonces este polinomio tiene mas de n raíces por lo tanto, la unica opción que queda es que $P(\alpha)$ sea el polinomio 0 es decir, es un polinomio tal que todos sus coeficientes son 0, se sigue entonces que:

$$Z(j) \binom{n}{j} = 0 \quad \forall \quad j \in \{0, 1, \dots, n\}$$

Como $\binom{n}{j} > 0$, se concluye que Z es una función tal que $Z(j) = 0$ con $j \in \{0, 1, \dots, n\}$, luego como X es una v.a. Binomial se sigue que:

$$\mathbb{P}(Z(X) = 0) = 1$$

Lo que demuestra la completitud de la familia binomial.

Definición 3.2.23 (Estadística Completa). Sea X_1, \dots, X_n m.a. de una población con función de densidad $f_X(x; \theta)$ y $T_1(X_1, \dots, X_n)$ una estadística. Decimos que T_1 es completa para θ si la familia dada por $\{f_{T_1}(t; \theta); \theta \in \Theta\}$ es completa

Mostrar que una estadística es completa se reduce a probar que la familia de densidades asociadas al estadístico es completa, sabemos por ejemplo ya que la familia binomial es completa, luego entonces, si tenemos X_1, \dots, X_n m.a. de $Benulli(\theta)$ se sigue que $T_1 = \sum_{i=1}^n X_i$ tiene una distribución binomial, luego entonces T_1 sería una estadística completa.

Necesitamos estadísticas completas por lo siguiente:

Teorema 3.2.12 (Teorema de Lehmann Scheffé (Versión 1)). Sea X_1, \dots, X_n m.a. de $f_X(x; \theta)$ y sea T_1 una estadística suficiente y completa y $\hat{\theta}$ cualquier estimador insesgado para θ . Entonces el estimador función de T_1 denotado por:

$$g(T_1) = \mathbb{E}(\hat{\theta} \mid T_1)$$

es *UMVUE!!!*

Teorema 3.2.13 (Teorema de Lehmann Scheffé (Versión 2)). Sea X_1, \dots, X_n m.a. de $f_X(x; \theta)$ y sea T_1 una estadística suficiente y completa, supongamos que existe una función g tal que el estimador $\hat{\theta}^* = g(T_1)$ es insesgado, entonces $\hat{\theta}^*$ es el *UMVUE*

Por ejemplo, ya sabemos que cuando tenemos X_1, \dots, X_n m.a. $Bernulli(\theta)$ entonces $T_1 = \sum_{i=1}^n X_i$ es una estadística completa, además sabemos que $\frac{\sum_{i=1}^n X_i}{n}$ es insesgado por lo tanto se concluye mediante el Teorema de Lehmann Scheffé que $\hat{\theta}^* = \frac{\sum_{i=1}^n X_i}{n}$ es UMVUE.

Toda esta teoría nos ha llevado a concluir que para encontrar UMVUE'S requerimos encontrar estadísticas completas, desafortunadamente encontrar estadísticas completas no es fácil. Sin embargo en la familia exponencial tenemos el siguiente teorema:

Teorema 3.2.14 (Complejidad en la Familia Exponencial). *En la familia exponencial:*

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

se tiene que la estadística $T_1 = \sum_{i=1}^n d(x_i)$ es completa y suficiente minimal para θ .

Veamos unos ejemplos de como aplicar esto en los modelos mas comunes:

Ejemplo 3.2.13. Sea X_1, \dots, X_n m.a. del modelo $Poisson(\lambda)$, buscamos un estimador UMVUE para λ . Primero verificaremos que el modelo poisson pertenece a la familia exponencial.

$$\begin{aligned} f_X(x; \lambda) &= \frac{\lambda^x}{x!} e^{-\lambda} \\ &= \frac{1}{x!} e^{-\lambda + x \ln(\lambda)} = \frac{1}{x!} e^{-\lambda} e^{\ln(\lambda)x} \end{aligned}$$

Por lo tanto definimos $a(\lambda) = e^{-\lambda}$; $b(x) = \frac{1}{x!}$; $c(\lambda) = \ln(\lambda)$; $d(x) = x$, luego entonces, sabemos que $\sum_{i=1}^n d(X_i) = \sum_{i=1}^n X_i$ es una estadística suficiente y completa para λ , por lo tanto si encontramos una función de $\sum_{i=1}^n X_i$ que sea insesgada para λ habremos encontrado el UMVUE. Como $\mathbb{E}(X_i) = \lambda$ se concluye entonces que $\hat{\lambda} = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ es un estimador insesgado función de una estadística completa y suficiente por lo tanto $\hat{\lambda} = \bar{X}$ es UMVUE.

Ejercicio 3.2.14. Responda lo siguiente:

1. Encuentra una estadística suficiente y completa para λ del modelo $Poisson(\lambda)$
2. Encuentra una estadística suficiente y completa para p del modelo $Binomial(k, p)$ Suponga k conocida
3. Encuentra una estadística suficiente y completa para λ del modelo $Exponencial(\lambda)$
4. Encuentra una estadística suficiente y completa para p del modelo $Geometrico(p)$
5. Encuentra una estadística suficiente y completa para μ del modelo $Normal(\mu, \sigma^2)$. Suponga σ^2 conocida

6. Encuentra una estadística suficiente y completa para β del modelo $\text{Gamma}(\alpha, \beta)$. Suponga α conocida

Ahora veremos un ejemplo muy interesante donde se encuentra el UMVUE directamente haciendo el calculo de la esperanza condicional.

Ejemplo 3.2.14. Imaginemos que en el modelo Poisson estamos interesados no por estimar a λ sino por una función, por ejemplo $\mathbb{P}(X = 0) = e^{-\lambda}$. Una forma muy común en estadística de llevar a cabo esto es utilizar el método del plug-in el cual consiste en estimar a λ en este caso seria a través de $\hat{\lambda} = \bar{X}$ y luego enchufar este estimador en la cantidad que queremos estimar y definir el estimador para $\mathbb{P}(X = 0) = e^{-\lambda}$ como:

$$\mathbb{P}(\widehat{X = 0}) = e^{-\hat{\lambda}} = e^{-\bar{X}}$$

Este sin duda seria un estimador, sin embargo una pregunta interesante es ver si este estimador es UMVUE, la respuesta desafortunadamente es que el estimador no es ni siquiera insesgado. Esto se prueba facilmente recordando que $\sum_{i=1}^n X_i \sim \text{Poisson}(n\lambda)$ entonces:

$$\mathbb{E}(e^{-\bar{X}}) = \mathbb{E}\left(e^{-\frac{1}{n} \sum_{i=1}^n X_i}\right) = \sum_{u=0}^{\infty} e^{-\frac{1}{n}u} \mathbb{P}(U = u)$$

Donde $U \sim \text{Poisson}(n\lambda)$ luego entonces:

$$\mathbb{E}(e^{-\bar{X}}) = \sum_{u=0}^{\infty} e^{-\frac{1}{n}u} e^{-n\lambda} (n\lambda)^u \frac{1}{u!} = e^{-n\lambda} \sum_{u=0}^{\infty} \left(e^{-\frac{1}{n}n\lambda}\right)^u \frac{1}{u!} = e^{-n\lambda} e^{e^{-\frac{1}{n}n\lambda}}$$

De donde concluimos que $\mathbb{E}(e^{-\bar{X}}) \neq e^{-\lambda}$ sin embargo se puede probar que este estimador si es asintoticamente insesgado, es decir:

$$\lim_{n \rightarrow \infty} \mathbb{E}(e^{-\bar{X}}) = e^{-\lambda} = \mathbb{P}(X = 0)$$

Dado que nuestro primer intento no pudo ser UMVUE utilizaremos el Teorema de Lehmann Scheffé para encontrarlo. Por un lado ya sabemos que $T_1 = \sum_{i=1}^n X_i$ es una estadística suficiente y completa, lo único que nos hace falta es encontrar un estimador insesgado para la cantidad a estimar $\mathbb{P}(X = 0) = e^{-\lambda}$.

Consideremos el siguiente estimador:

$$\mathbb{P}(\widehat{X = 0}) = \mathbf{1}_{\{X_1=0\}}$$

Dicho estimador cumple con ser función de la muestra y ser insesgado, en efecto pues

$$\mathbb{1}_{\{X_1=0\}} \sim \text{Bernulli}(\mathbb{P}(X=0))$$

Entonces

$$\mathbb{E}\left(\widehat{\mathbb{P}(X=0)}\right) = \mathbb{E}(\mathbb{1}_{\{X_1=0\}}) = \mathbb{P}(X=0)$$

Ya tenemos los dos ingredientes para llevar a cabo el proceso de Lehmann Scheffé, tenemos un estimador insesgado y una estadística suficiente y completa. Calculamos entonces la esperanza condicional:

$$\begin{aligned} g(u) &= \mathbb{E}\left(\widehat{\mathbb{P}(X=0)} \mid T_1 = u\right) = \mathbb{E}\left(\mathbb{1}_{\{X_1=0\}} \mid \sum_{i=1}^n X_i = u\right) \\ &= 0\mathbb{P}\left(\mathbb{1}_{\{X_1=0\}} = 0 \mid \sum_{i=1}^n X_i = u\right) + 1\mathbb{P}\left(\mathbb{1}_{\{X_1=0\}} = 1 \mid \sum_{i=1}^n X_i = u\right) \\ &= \mathbb{P}\left(\mathbb{1}_{\{X_1=0\}} = 1 \mid \sum_{i=1}^n X_i = u\right) = \frac{\mathbb{P}(X_1 = 0, \sum_{i=1}^n X_i = u)}{\mathbb{P}(\sum_{i=1}^n X_i = u)} \\ &= \frac{\mathbb{P}(X_1 = 0, \sum_{i=2}^n X_i = u)}{\mathbb{P}(\sum_{i=1}^n X_i = u)} = \frac{\mathbb{P}(X_1 = 0) \mathbb{P}(\sum_{i=2}^n X_i = u)}{\mathbb{P}(\sum_{i=1}^n X_i = u)} \end{aligned}$$

Como $\sum_{i=2}^n X_i \sim \text{Poisson}((n-1)\lambda)$ y $\sum_{i=1}^n X_i \sim \text{Poisson}(n\lambda)$ tenemos que:

$$g(u) = \frac{e^{-\lambda} e^{-(n-1)\lambda} ((n-1)\lambda)^u \frac{1}{u!}}{e^{-n\lambda} (n\lambda)^u \frac{1}{u!}} = \left(\frac{n-1}{n}\right)^u$$

Finalmente por Teorema de Lehmann Scheffé sabemos que $g(T_1)$ es UMVUE, por lo tanto:

$$\mathbb{P}(\widehat{X=0})^* = g(T_1) = \frac{(n-1)^u}{n} = \left(\frac{n-1}{n}\right)^{\sum_{i=1}^n X_i}$$

Es UMVUE para $\mathbb{P}(X=0)$. De ejercicio adicional se deja probar de forma directa que $\mathbb{P}(\widehat{X=0})^*$ es insesgado

Ejercicio 3.2.15. Resuelva lo siguiente:

- Sea X_1, \dots, X_n m.a. del modelo Bernulli (θ) . Encuentre el UMVUE para la cantidad $\theta(1-\theta)$

- Sea X_1, \dots, X_n m.a. del modelo Poisson (λ) . Encuentre el UMVUE para la cantidad:

$$\mathbb{P}(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

- Sea X_1, \dots, X_n m.a. del modelo Normal (r, σ^2) , $r > 0$ ambos desconocidos donde r es el radio de un círculo, obtener el UMVUE para el área del círculo

$$A = \pi r^2$$

- Sea X_1, \dots, X_n m.a. del modelo Normal $(\theta, 1)$, encuentre el UMVUE para θ^2
- Sea X_1, \dots, X_n m.a. del modelo Bernulli (θ) , encuentre el UMVUE para θ^n .
(Hint: Considere $\hat{\theta}^n = \mathbb{1}_{\{X_1=1, \dots, X_n=1\}}$ como un estimador insesgado y recuerde que $\sum_{i=1}^n X_i$ es una estadística suficiente y completa para θ)
- Sea X_1, \dots, X_n m.a. de la densidad MAXWELL:

$$f_X(x) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \theta^{-\frac{3}{2}} x^2 e^{-\frac{1}{2\theta} x^2}$$

Demuestre que dicha densidad pertenece a la familia exponencial, encuentre una estadística suficiente y completa y encuentre un UMVUE para θ

Dado que en la familia exponencial uniparamétrica siempre existe una estadística suficiente y completa, eso implica via el el proceso de LehmannScheffé que siempre existirá un UMVUE para alguna transformación del parámetro desconocido. En efecto, como en la familia exponencial $\sum_{i=1}^n d(X_i)$ es completa y suficiente entonces cualquier función de dicha estadística será UMVUE para cierta función de θ . Por ejemplo, consideremos $h(\theta) = \mathbb{E}(d(X_i))$ entonces $\frac{\sum_{i=1}^n d(X_i)}{n}$ es UMVUE para $h(\theta)$, pues

$$\mathbb{E}\left(\frac{\sum_{i=1}^n d(X_i)}{n}\right) = \frac{n\mathbb{E}(d(X_1))}{n} = \mathbb{E}(d(X_1)) = h(\theta)$$

Consideremos por ejemplo el modelo Exponencial:

$$f(x; \lambda) = \lambda e^{-\lambda x} > (x)$$

Es claro que pertenece a la familia exponencial pues $a(\lambda) = \lambda; b(x) = > (x); c(\lambda) = \lambda; d(x) = x$,

luego entonces como, $\mathbb{E}(X) = \frac{1}{\lambda}$ se sigue que la estadística:

$$\frac{\sum_{i=1}^n X_i}{n}$$

Es UMVUE para la cantidad $\frac{1}{\lambda}$, luego si reparametrizamos a la familia exponencial como $\theta = \frac{1}{\lambda}$ se sigue que el modelo:

$$f_X(x; \theta) = \frac{1}{\theta} e^{-\frac{1}{\theta}x}$$

La estadística $\frac{\sum_{i=1}^n X_i}{n}$ es UMVUE para θ

En la familia exponencial siempre podremos tener una reparametrización donde se alcanza la CICR, sin embargo surge la pregunta de saber qué pasa cuando no tenemos a la familia exponencial como modelo de nuestra población. A continuación veremos un ejemplo que no pertenece a la familia exponencial sin embargo se encontrará el UMVUE encontrando la estadística suficiente y completa, sin embargo requeriremos un poco de teoría de estadísticos de orden.

3.2.6. Estadísticos de Orden

En esta sección introduciremos el concepto de estadísticos de orden así como algunas definiciones importantes para la Estadística que sirven a consecuencia de éstos.

Los estadísticos de orden juegan sin duda un papel muy importante tanto para la estadística paramétrica como la no paramétrica. De hecho, algunas de sus propiedades no depende de la función de distribución de la población de donde fue obtenida la muestra aleatoria.

Supongamos que tenemos un experimento aleatorio que sigue una distribución de probabilidad especificada por la función $F_X(\cdot)$. Enseguida repetimos el experimento n veces obteniendo con

ello una muestra aleatoria X_1, \dots, X_n a partir de esto definimos lo siguiente:

$$\begin{aligned}
 Y_{1:n} &= \min \{X_1, \dots, X_n\} \\
 Y_{2:n} &= \min \{\{X_1, \dots, X_n\} \setminus \{Y_{1:n}\}\} \\
 Y_{3:n} &= \min \{\{X_1, \dots, X_n\} \setminus \{Y_{2:n}, Y_{1:n}\}\} \\
 Y_{3:n} &= \min \{\{X_1, \dots, X_n\} \setminus \{Y_{3:n}, Y_{2:n}, Y_{1:n}\}\} \\
 Y_{4:n} &= \min \{\{X_1, \dots, X_n\} \setminus \{Y_{4:n}, Y_{3:n}, Y_{2:n}, Y_{1:n}\}\} \\
 &\vdots \\
 Y_{n-1:n} &= \min \{\{X_1, \dots, X_n\} \setminus \{Y_{n-2:n}, \dots, Y_{1:n}\}\} \\
 Y_{n:n} &= \min \{\{X_1, \dots, X_n\} \setminus \{Y_{n-1:n}, \dots, Y_{1:n}\}\} = \max \{X_1, X_2, \dots, X_n\}
 \end{aligned}$$

Dado la anterior nos vamos cuenta de tres propiedades importantes para estas funciones definidas a partir de la muestra aleatoria:

1. Cara $Y_{i:n}$ es un estadístico según la definición anteriormente dada.
2. Estos estadísticos definidos son tales que ordenan la muestra de menor a mayor magnitud, logrando con ellos que $Y_{k:n}$ sea la k -ésima observación de la muestra aleatoria ordenándola según sus magnitudes.
3. Debe de ser claro que no hay independencia entre los estadísticos ya que si $Y_{j:n} > k$ entonces $Y_{j+a:n} > k$.

La definición formal es la siguiente:

Definición 3.2.24 (Estadístico de Orden). Sea X_1, \dots, X_n una muestra aleatoria de tamaño n de un fenómeno aleatorio especificado por la función de distribución $F_X(x; \theta)$. Ordenemos cada colección de valores x_1, \dots, x_n de X_1, \dots, X_n en order decreciente de magnitudes tales que obtenemos los valores $y_{1:n}, \dots, y_{n:n}$ (conocida como la muestra observada ordenada), donde $y_{1:n} = \min \{x_1, \dots, x_n\}$, $y_{2:n}$ es el siguiente x_i en orden de magnitud, \dots , y $y_{n:n} = \max \{x_1, \dots, x_n\}$. La variable aleatorio $Y_{k:n}$ $k \in \{1, 2, \dots, n\}$ función de X_1, \dots, X_n la cual toma un valor y_{k_n} en cada posible secuencia de x_1, x_2, \dots, x_n , es llamado el k -ésimo estadístico de orden (Al número k se le conoce como la magnitud de dicho estadístico de orden)-

El siguiente problema es encontrar la distribución de los estadísticos de orden, hasta el momento no hemos hecho distinción entre la variable aleatoria continuo y discreta y que lo que hemos

definido puede ser aplicado a ambos tipos de variables aleatorias. A continuación enunciamos un teorema el cual nos da la distribución del estadístico $Y_{\alpha:n}$ conociendo la función de distribución de la población de donde fue obtenida la muestra aleatoria y el cual será válido para ambos tipos de variables aleatorias. (Absolutamente continuas y discretas).

Teorema 3.2.15 (Distribución del estadístico $Y_{\alpha:n}$). *Sea $Y_{1:n} \leq Y_{2:n} \leq \dots \leq Y_{n:n}$ los estadísticos de orden de una muestra aleatoria X_1, \dots, X_n de una función de distribución $F_X(x)$ conocida. Entonces la función de distribución acumulativa de $Y_{\alpha:n}$, $\alpha = 1, 2, \dots, n$ está dada por:*

$$F_{Y_{\alpha:n}}(y) = \sum_{j=\alpha}^n \binom{n}{j} F_X(y)^j (1 - F_X(y))^{n-j}$$

Demostración. Sea y fija y arbitraria y definamos la v.a. $Z_i = \mathbb{1}_{(-\infty, y]}(X_i)$, observemos lo siguiente:

- Z_i es una v.a. por estar en función de X_i que es v.a
- $\sum_{i=1}^n Z_i =$ número de X_i 's tales que son menores o iguales a y
- Z_i sigue una distribución *Bernulli* ($F(y)$)
- $\sum_{i=1}^n Z_i$ sigue una distribución *Binomial* ($n, F(y)$)

Ahora calculemos $F_{Y_{\alpha:n}}(y)$

$$F_{Y_{\alpha:n}}(y) = \mathbb{P}(Y_{\alpha:n} \leq y) = \mathbb{P}\left(\sum_{i=1}^n Z_i \geq \alpha\right) = \sum_{j=\alpha}^n \binom{n}{j} F_X(y)^j (1 - F_X(y))^{n-j}$$

Observe que la clave de la prueba anterior es la equivalencia entre los eventos $\{Y_{\alpha:n} \leq y\}$ y $\{\sum_{i=1}^n Z_i \geq \alpha\}$. Esto pasa porque si el α estadístico de orden es menor o igual a y , entonces necesitamos que el número de X_i 's menores o iguales a y sea más grande o incluso igual a α , y viceversa.

□

Como consecuencia inmediata de lo anterior tenemos que:

$$F_{Y_{n:n}}(y) = F_X(y)^n$$

$$F_{Y_{1:n}}(y) = 1 - (1 - F_X(y))^n$$

Luego entonces las densidades para el caso en que la distribución es absolutamente continua viene dado por:

$$f_{Y_{1:n}}(y) = \frac{d}{dy} F_{Y_{1:n}}(y) = n(1 - F_X(y))^{n-1} f_X(y) \quad (3.4)$$

$$f_{Y_{n:n}}(y) = \frac{d}{dy} F_{Y_{n:n}}(y) = n(F_X(y))^{n-1} f_X(y) \quad (3.5)$$

$$f_{Y_{\alpha:n}}(y) = \frac{d}{dy} F_{Y_{\alpha:n}}(y) = \frac{n!}{(\alpha-1)!(n-\alpha)!} F_X(y)^{\alpha-1} (1 - F_X(y))^{n-\alpha} f(y) \quad (3.6)$$

Finalmente encontramos la densidad conjunta mediante el siguiente teorema:

Teorema 3.2.16 (Densidad conjunta de los estadísticos de orden $Y_{1:n}, \dots, Y_{n:n}$). *La función de densidad conjunta de los estadísticos de orden $Y_{1:n}, Y_{2:n}, \dots, Y_{n:n}$ definidos a partir de una muestra aleatoria proveniente de una población con función de distribución $F(\cdot)$ y densidad $f(\cdot)$ está dada por:*

$$f_{Y_{1:n}, \dots, Y_{n:n}}(y_1, \dots, y_n) = n! f(y_1) f(y_2) \dots f(y_n) \mathbb{1}_{-\infty < y_1 < \dots < y_n < \infty}$$

El resultado anterior es de gran importancia pues gracias a la densidad conjunta de todos los estadísticos de orden seremos capaces de obtener cualquier densidad conjunta o marginal simplemente integrando respecto a las variables restantes. Por ejemplo, a densidad conjunta de dos estadísticos de orden con $\alpha < \beta$ está dada por

$$f_{Y_{\alpha:n}, Y_{\beta:n}}(y_\alpha, y_\beta) = \frac{n! F_X(y_\alpha)^{\alpha-1} (F_X(y_\beta) - F_X(y_\alpha))^{\beta-\alpha-1} (1 - F_X(y_\beta))^{n-\beta} f_X(y_\alpha) f_X(y_\beta)}{(\alpha-1)!(\beta-\alpha-1)!(n-\beta)!}$$

Finalmente si queremos la densidad conjunta del mínimo y el máximo lo único que debemos hacer es sustituir $\alpha = 1$ y $\beta = n$ obteniendo lo siguiente:

$$f_{Y_1, Y_{n:n}}(y_1, y_n) = n(n-1)(F_X(y_n) - F_X(y_1))^{n-2} f_X(y_1) f_X(y_n)$$

Ahora consideremos el siguiente ejemplo:

Ejemplo 3.2.15. Sea X_1, \dots, X_n m.a. del modelo $U(0, \theta)$, el objetivo es encontrar un UMVUE para θ , sin embargo no contamos con que el modelo uniforme pertenezca a la familia exponencial y además no podemos aplicar la teoría de CICR pues esta densidad no cumple con las condiciones de regularidad, luego entonces la única esperanza de encontrar un UMVUE es por la vía de encontrar un estadística suficiente y completa. Primero buscaremos una estadística suficiente por medio del teorema de factorización, primero recordemos que la densidad de nuestro modelo

está dada por:

$$f_X(x; \theta) = \frac{1}{\theta} \mathbb{1}_{(0, \theta)}(x)$$

Entonces la verosimilitud está dada por:

$$f_{\underline{X}}(x_1, \dots, x_n) = \prod_{i=1}^n f_X(x_i; \theta) = \frac{1}{\theta^n} \prod_{i=1}^n \mathbb{1}_{(0, \theta)}(x_i)$$

Analicemos la cantidad $\prod_{i=1}^n \mathbb{1}_{(0, \theta)}(x_i)$ notemos que esta cantidad toma el valor de 1 si toda $x_i \in (0, \theta)$ y esto ocurre cuando $Y_{n:n} < \theta$ mientras que vale 0 si existe una x_i tal que $x_i > \theta$ lo que se traduce a decir que $Y_{n:n} \geq \theta$, luego entonces tenemos la siguiente igualdad:

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

Por lo tanto la verosimilitud puede ser escrita como:

$$f_{\underline{X}}(x_1, \dots, x_n) = \frac{1}{\theta^n} \mathbb{1}_{(y_{n:n}, \infty)}(\theta)$$

De donde se concluye via el teorema de factorización (3.2.6) que $Y_{n:n} = \max\{X_1, \dots, X_n\}$ es una estadística suficiente. Ahora verifiquemos que dicha estadística es completa, para ellos requerimos la densidad del máximo estadístico de orden el cual según la ecuación (3.5) esta dada por:

$$f_{Y_{n:n}}(y) = \frac{d}{dy} F_{Y_{n:n}}(y) = n (F_X(y))^{n-1} f_X(y) = n \left(\frac{y}{\theta}\right)^{n-1} \frac{1}{\theta} \mathbb{1}_{(0, \theta)}(y)$$

Lo anterior es válido pues recordemos que la distribución de la v.a. aleatoria $U(0, \theta)$ está dada por:

$$F_X(x; \theta) = \int_{-\infty}^x \frac{1}{\theta} \mathbb{1}_{(0, \theta)}(y) dy = \frac{x}{\theta} \mathbb{1}_{(0, \theta)}(x)$$

Luego entonces, para probar que $Y_{n:n}$ es completa necesitamos verificar que la siguiente familia es completa.

$$\left\{ f_{Y_{n:n}}(x) = n \left(\frac{x}{\theta}\right)^{n-1} \frac{1}{\theta} \mathbb{1}_{(0, \theta)}(x) : \theta \in \mathbb{R}^+ \right\}$$

Para ello debemos verificar lo que dice la definición (3.2.22), sea entonces Z una función y $\theta \in \mathbb{R}^+$ arbitrarios, debemos demostrar que si para todo $\theta \in \mathbb{R}^+$ se tiene que :

$$\int_0^\theta Z(x) n \left(\frac{x}{\theta}\right)^{n-1} \frac{1}{\theta} \mathbb{1}_{(0, \theta)}(x) dx = 0 \Rightarrow \mathbb{P}(Z(Y_{n:n}) = 0) = 1$$

Entonces tenemos lo siguiente:

$$\begin{aligned}
 \int_0^\theta Z(x) n \left(\frac{x}{\theta}\right)^{n-1} \frac{1}{\theta} \mathbb{1}_{(0,\theta)}(x) dx &= 0 \\
 \frac{1}{\theta} \int_0^\theta Z(x) n x^{n-1} dx &= 0 \\
 \int_0^\theta Z(x) n x^{n-1} dx &= 0 \\
 \frac{\partial}{\partial \theta} \int_0^\theta Z(x) n x^{n-1} dx &= \frac{\partial}{\partial \theta} 0 \\
 Z(\theta) n \theta^{n-1} dx &= 0 \\
 Z(\theta) &= 0
 \end{aligned}$$

Lo último fue válido para todo $\theta \in \mathbb{R}^+$ y para cualquier Z función arbitraria además como $Y_{n:n} > 0$ se concluye entonces que $\mathbb{P}(Z(Y_{n:n}) = 0) = 1$ lo que demuestra la completitud de la familia y por tanto la completitud del estadístico $Y_{n:n}$. Tenemos entonces que $Y_{n:n}$ es un estadístico suficiente y completo, luego entonces solo bastaría encontrar una función de esta estadística que sea insesgada para θ . Calculemos la esperanza de nuestro estadístico:

$$\mathbb{E}(Y_{n:n}) = \int_0^\theta x n \frac{x^{n-1}}{\theta^n} dx = \frac{n}{\theta^n} \int_0^\theta x^n dx = \frac{n}{n+1} \theta$$

Desafortunadamente nuestro estadístico es sesgado, sin embargo se puede hacer insesgado muy fácilmente ya que solo hay que multiplicar por el inverso multiplicativo de $\frac{n}{n+1}$. Se concluye finalmente que el estimador:

$$\hat{\theta}^* = \frac{n+1}{n} Y_{n:n}$$

Es UMVUE para θ en este modelo.

Ejercicio 3.2.16. 1. Sea X_1, \dots, X_n m.a. del modelo $\text{Uniforme}(\theta, 0)$ con $\theta < 0$.

$$f_X(x) := -\frac{1}{\theta} \mathbf{1}_{(\theta, 0)}(x)$$

El objetivo del ejercicio es encontrar un UMVUE para θ

a) Encuentre la función de distribución del modelo $\text{Uniforme}(\theta, 0)$ y pruebe que:

$$F_X(x) = \mathbb{P}(X \leq x) = \begin{cases} 0 & x \leq \theta \\ (1 - \frac{x}{\theta}) & x \in (\theta, 0) \\ 1 & x \geq 0 \end{cases}$$

b) Muestre que la densidad de $X_{(1)} := \min\{X_1, \dots, X_n\}$ esta dada por:

$$f_{X_{(1)}}(x) := -n \frac{x^{n-1}}{\theta^n} \mathbf{1}_{(\theta, 0)}(x)$$

c) Muestre que $X_{(1)} := \min\{X_1, \dots, X_n\}$ es una estadística suficiente para θ

d) Muestre que $X_{(1)} := \min\{X_1, \dots, X_n\}$ es una estadística completa para θ

e) Muestre que:

$$\mathbb{E}(X_{(1)}) = \int_{\theta}^0 x f_{X_{(1)}}(x) dx = \frac{n}{n+1} \theta$$

f) Concluya entonces que:

$$\hat{\theta} = \frac{n+1}{n} X_{(1)}$$

Es UMVUE para θ .

3.3. Construcción de Estimadores

A continuación describiremos dos métodos muy conocidos para generar estimadores que gozan con ciertas propiedades.

3.3.1. Método de Momentos

Este método se basa en el teorema de *Glivenko – Cantelli* el cual es conocido por muchos como el Teorema Fundamental de la Estadística, el cual nos dice lo siguiente:

Teorema 3.3.1 (Teorema Glivenko-Cantelli (1933)). Sea X_1, X_2, \dots una sucesión de variables aleatorias idénticamente distribuidas con función de distribución comun dada por $F(x)$. Sea $F_n(x)$ la función de distribución empírica formada por las primeras n variables aleatorias definida por:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(-\infty, x]}(X_i)$$

. La ley fuerte de los grandes números afirma que la v.a. $F_n(x)$ converge puntualmente y casi seguramente a $F(x)$, sin embargo Glivenko-Cantelli demuestra que no solo hay convergencia puntual sino que hay una convergencia uniforme, es decir:

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0$$

El teorema nos dice entonces que conforme va creciendo la muestra se tiene que F_n se parece mucho a F para toda $x \in \mathbb{R}$ entonces:

$$\begin{aligned} F_n &\approx F \\ \int_{\mathbb{R}} dF_n &\approx \int_{\mathbb{R}} dF \\ \sum_{i=1}^n \frac{x_i^r}{n} = \int_{\mathbb{R}} x^r dF_n &\approx \int_{\mathbb{R}} x^r dF = \mathbb{E}(X^r) \end{aligned}$$

Luego entonces por el teorema (3.3.1) se tiene que:

$$\sum_{i=1}^n \frac{x_i^r}{n} \approx \mathbb{E}(X^r)$$

A la cantidad $\mathbb{E}(X^r)$ se le conoce como el r -ésimo momento poblacional mientras que $\sum_{i=1}^n \frac{x_i^r}{n}$ se le conoce como el r -ésimo momento muestral, el teorema nos afirma entonces que los momentos población y muestral son muy parecidos cuando la muestra es grande. Ahora supongamos que $X \sim F_X(\cdot; \underline{\theta})$ con $\underline{\theta} = (\theta_1, \dots, \theta_p)$ un vector de parámetros. El método de momentos nos dice que las siguientes igualdades son aproximadas:

$$\begin{aligned} \mathbb{E}(X) = a_1(\theta_1, \dots, \theta_p) &\approx \sum_{i=1}^n \frac{x_i}{n} \\ \mathbb{E}(X^2) = a_2(\theta_1, \dots, \theta_p) &\approx \sum_{i=1}^n \frac{x_i^2}{n} \\ &\vdots \\ \mathbb{E}(X^p) = a_p(\theta_1, \dots, \theta_p) &\approx \sum_{i=1}^n \frac{x_i^p}{n} \end{aligned}$$

Lo cual genera un sistema de p ecuaciones con p incógnitas dadas por el vector de parámetros

$\underline{\theta} = (\theta_1, \dots, \theta_p)$, el cual al resolverlo obtenemos:

$$\begin{aligned}\hat{\theta}_1 &= f_1(x_1, \dots, x_p) \\ &\vdots \\ \hat{\theta}_p &= f_p(x_1, \dots, x_p)\end{aligned}$$

Los cuales serán los estimadores encontrados $\hat{\underline{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_p)$. Las propiedades de estos estimadores es que siempre son consistentes pero en general son insertados. Consideremos los siguientes ejemplo:

Ejemplo 3.3.1. Consideremos X_1, \dots, X_n una muestra aleatoria del modelo Bernulli (θ), entonces sabemos que el primer momento poblacional está dad por : $\mathbb{E}(X) = \theta$ luego entonces la ecuación que nos queda para encontrar el estimador es:

$$\mathbb{E}(X) = \theta \approx \sum_{i=1}^n \frac{X_i}{n}$$

Luego entonces el estimador por momentos está dado por $\hat{\theta} = \sum_{i=1}^n \frac{X_i}{n}$.

Ejemplo 3.3.2. Consideremos X_1, \dots, X_n m.a. del modelo $U(0, \theta)$, entonces sabemos que $\mathbb{E}(X) = \frac{\theta}{2}$ luego entonces igualando el primer momento poblacional y muestral se tiene que:

$$\mathbb{E}(X) = \frac{\theta}{2} \approx \sum_{i=1}^n \frac{X_i}{n}$$

Luego resolviendo la ecuación obtenemos que el estimador es:

$$\hat{\theta} = 2 \sum_{i=1}^n \frac{X_i}{n} = 2\bar{X}$$

Es fácil ver que este estimador es insesgado y consistente viendo que $\text{Var}(\hat{\theta}) \rightarrow 0$

Ejemplo 3.3.3. Consideremos X_1, \dots, X_n m.a. del modelo $N(\mu, \sigma^2)$, en este caso tenemos 2 parámetros desconocidos por lo que necesitamos generar y resolver un sistemas de 2 ecuaciones

con los dos primeros momentos:

$$\begin{aligned}\mathbb{E}(X) &= \mu \approx \sum_{i=1}^n \frac{X_i}{n} \\ \mathbb{E}(X^2) &= \mu + \sigma^2 \approx \sum_{i=1}^n \frac{X_i^2}{n}\end{aligned}$$

Resolviendo el sistema obtenemos:

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

3.3.2. Método de Máxima Verosimilitud

Como ya hemos visto, la verosimilitud es una función de θ definida como sigue:

Definición 3.3.1 (Verosimilitud-Muestra Observada). Sea X_1, \dots, X_n m.a. de un modelo con densidad $f_X(x; \theta)$ y θ un parámetro desconocido tal que $\theta \in \Theta$, definimos la función de verosimilitud $\ell : \Theta \rightarrow \mathbb{R}$ como:

$$\ell(\theta; \underline{x}) := f_{\underline{X}}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i; \theta)$$

Método de máxima verosimilitud busca encontrar $\hat{\theta}(x_1, \dots, x_n)$ tal que maximice la función de verosimilitud, la idea que hay de tras de este método es buscar aquel parámetro que maximice la probabilidad o plausibilidad o verosimilitud de haber observado la muestra ($X_1 = x_1, \dots, X_n = x_n$), por ejemplo supongamos que en el modelo Bernulli de parámetro θ observamos la siguiente muestra con $n = 5$

$$(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 0)$$

entonces sería inverosímil suponer que $\theta \approx 1$.

Maximizar a $\ell(\theta; \underline{x})$ como función de θ por lo general no es fácil, pues involucra el producto de densidades por lo que en general se acostumbra no maximizar directamente a la verosimilitud sino mejor trabajar con el logaritmo de ésta debido a que el producto es transformado en suma:

$$L(\theta; \underline{x}) = \log \ell(\theta; \underline{x}) := \log f_{\underline{X}}(x_1, \dots, x_n) = \sum_{i=1}^n \log f_{X_i}(x_i; \theta)$$

Es sabido que una forma de encontrar el máximo de una función es derivando (No es la única forma de hacerlo) por lo que muchas veces para encontrar el estimador máximo verosímil se resuelve

la siguiente ecuación:

$$\frac{\partial}{\partial \theta} L(\theta; \underline{x}) = 0$$

Sin embargo en el caso general, cuando tenemos k parámetros $\underline{\theta} = (\theta_1, \dots, \theta_k)$ el problema de encontrar los estimadores máximo verosímiles se convierte en encontrar la solución al siguiente sistema de ecuaciones:

$$\begin{aligned} \frac{\partial}{\partial \theta_1} L(\underline{\theta}; \underline{x}) &= 0 \\ \frac{\partial}{\partial \theta_2} L(\underline{\theta}; \underline{x}) &= 0 \\ &\vdots \\ \frac{\partial}{\partial \theta_k} L(\underline{\theta}; \underline{x}) &= 0 \end{aligned}$$

Estos estimadores son muy utilizados en los modelos estadísticos que se han desarrollado mas que nada por las siguientes propiedades:

- Son consistentes
- Son asintoticamente eficientes
- En general los estimadores no siempre son insesgado pero son asintoticamente insesgados
- Los estimadores máximo verosímiles siempre son funciones de estadísticas suficientes
- Los estimadores máximo verosímiles gozan la propiedad de invarianza, supongamos que tenemos $\hat{\theta}_{MV}$ pero imaginemos que estamos interesados en un estimador para alguna transformación (no necesariamente biyectiva) $g(\theta)$ entonces $g(\hat{\theta})$ es el estimador máximo verosímil de $g(\theta)$. (Por ejemplo el estimador máximo verosímil para $\mathbb{P}(X=1) = e^{-\lambda}$ en el modelo poisson esta dado por $\mathbb{P}(\hat{X}=1) = e^{-\hat{\lambda}}$)
- Bajo ciertas condiciones de regularidad (3.2.10) se puede probar que:

$$\hat{\theta}_{MV} \sim N\left(\theta, \frac{1}{nI(\theta)}\right)$$

donde $I(\theta)$ es la información de Fisher por unidad muestral, es decir:

$$I(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \log(f_X(x; \theta))\right)$$

- Una generalización de lo anterior nos dice que bajos las mismas condiciones de regularidad (3.2.10) se tiene que:

$$f(\hat{\theta}_{MV}) \sim N\left(f(\theta), \frac{(f'(\theta))^2}{nI(\theta)}\right)$$

Demostración. Haremos un esbozo de la prueba de la aproximación hacia la normalidad de los estimadores máximo verosímiles.

Primero, sabemos que $\hat{\theta}_{MV}$ maximiza tanto a la verosimilitud como a la log-verosimilitud esto implica que bajo las condiciones de regularidad que:

$$L'(\hat{\theta}_{MV}; \underline{X}) = \frac{\partial}{\partial \theta} L(\hat{\theta}_{MV}; \underline{X}) = \frac{\partial}{\partial \theta} \log(f(X_1, \dots, X_n; \hat{\theta}_{MV})) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log(f(X_i; \theta_{MV})) = 0$$

Consideremos a la función

$$L'_n(\theta; \underline{X}) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log(f(X_i; \theta))$$

Entonces queda claro que al evaluar en $\hat{\theta}_{MV}$ se tiene que:

$$L'_n(\hat{\theta}_{MV}; \underline{X}) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log(f(X_i; \theta_{MV})) = 0$$

Ahora hagamos la aproximación de Taylor de la función $L'_n(\theta; \underline{x})$ alrededor de θ_{MV} :

$$L'_n(\theta; \underline{X}) \approx L'_n(\hat{\theta}_{MV}; \underline{X}) + L''_n(\hat{\theta}_{MV}; \underline{X}) (\theta - \hat{\theta}_{MV}) = L''_n(\hat{\theta}_{MV}; \underline{X}) (\theta - \hat{\theta}_{MV})$$

Por lo tanto:

$$\begin{aligned} L'_n(\theta; \underline{X}) &\approx L''_n(\hat{\theta}_{MV}; \underline{X}) (\theta - \hat{\theta}_{MV}) \\ \frac{-L'_n(\theta; \underline{X})}{L''_n(\hat{\theta}_{MV}; \underline{X})} &\approx (\hat{\theta}_{MV} - \theta) \\ \frac{-\sqrt{n}L'_n(\theta; \underline{X})}{L''_n(\hat{\theta}_{MV}; \underline{X})} &\approx \sqrt{n}(\hat{\theta}_{MV} - \theta) \end{aligned}$$

Analicemos el numerador $-\sqrt{n}L'_n(\theta; \underline{X})$.

$$-\sqrt{n}L'_n(\theta; \underline{X}) = -\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log(f(X_i; \theta)) \right)$$

Por el T.LC. sabemos que:

$$-\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log(f(X_i; \theta)) \right) \rightarrow^d N \left(\mathbb{E} \left(-\frac{\partial}{\partial \theta} \log(f(X; \theta)) \right), \text{Var} \left(-\frac{\partial}{\partial \theta} \log(f(X; \theta)) \right) \right)$$

Pero recordemos que bajo las condiciones de regularidad

$$\begin{aligned} \mathbb{E} \left(-\frac{\partial}{\partial \theta} \log(f(X; \theta)) \right) &= -\mathbb{E} \left(\frac{\partial}{\partial \theta} \log(f(X; \theta)) \right) = 0 \\ \text{Var} \left(-\frac{\partial}{\partial \theta} \log(f(X; \theta)) \right) &= \text{Var} \left(\frac{\partial}{\partial \theta} \log(f(X; \theta)) \right) = -\mathbb{E} \left(\frac{\partial^2}{\partial \theta^2} \log(f(X; \theta)) \right) = I(\theta) \end{aligned}$$

Concluimos entonces que:

$$-\sqrt{n}L'_n(\theta; \underline{X}) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n -\frac{\partial}{\partial \theta} \log(f(X_i; \theta)) \right) \rightarrow^d N(0, I(\theta))$$

Por otro lado el denominador $L''_n(\hat{\theta}_{MV}; \underline{X})$ tiene la siguiente propiedad gracias a la ley fuerte de los grandes números.

$$L''_n(\theta; \underline{X}) = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log(f(X_i; \theta)) \rightarrow \mathbb{E} \left(\frac{\partial^2}{\partial \theta^2} \log(f(X; \theta)) \right) = -I(\theta)$$

Luego como $\hat{\theta}_{MV}$ es consistente sabemos que $\hat{\theta}_{MV} \rightarrow \theta$ lo que concluye que:

$$L''_n(\hat{\theta}_{MV}; \underline{X}) \rightarrow -I(\theta)$$

Aquí necesitaremos un Teorema de Probabilidad:

Teorema 3.3.2 (Slutsky's). Sea $\{X_n\}$ y $\{Y_n\}$ sucesión de variables aleatorias, suponga que $X_n \rightarrow^d X$ y $Y_n \rightarrow c$ entonces:

- $X_n + Y_n \rightarrow^d X + c$
- $X_n Y_n \rightarrow^d cX$

$$\blacksquare \quad \frac{X_n}{Y_n} \rightarrow^d \frac{X}{c}$$

Aplicando el Teorema de Slutsky's concluimos que:

$$\frac{-\sqrt{n}L'_n(\theta; \underline{X})}{L''_n(\hat{\theta}_{MV}; \underline{X})} \rightarrow^d \frac{Z}{-I(\theta)}$$

Donde $Z \sim N(0, I(\theta))$. Entonces:

$$\frac{Z}{-I(\theta)} \sim N\left(0, \frac{1}{I(\theta)}\right)$$

Entonces:

$$\frac{-\sqrt{n}L'_n(\theta; \underline{X})}{L''_n(\hat{\theta}_{MV}; \underline{X})} \rightarrow^d N\left(0, \frac{1}{I(\theta)}\right)$$

Lo que concluye que:

$$\sqrt{n}(\hat{\theta}_{MV} - \theta) \rightarrow^d N\left(0, \frac{1}{I(\theta)}\right)$$

Por lo que para n suficientemente grande:

$$\hat{\theta}_{MV} \sim^{approx} N\left(\theta, \frac{1}{nI(\theta)}\right)$$

□

A continuación ponemos dos ejemplos de como encontrar estimadores máximo verosímiles en dos familias paramétricas.

Ejemplo 3.3.4. Sea X_1, \dots, X_n m.a. aleatoria del modelo $N(\mu; \sigma^2)$, entonces la función de verosimilitud esta dada por:

$$\ell((\mu, \sigma^2); \underline{x}) := f_{\underline{X}}(x_1, \dots, x_n; (\mu; \sigma^2)) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

Luego entonces la Log-Verosimilitud es:

$$L((\mu, \sigma^2); \underline{x}) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Al derivar parcialmente e igual a cero obtenemos el siguiente sistema de ecuaciones:

$$\begin{aligned}\sum_{i=1}^n (x_i - \mu) &= 0 \\ -\frac{n}{2} + \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 &= 0\end{aligned}$$

Resolviendo dicho sistema de ecuaciones obtenemos los siguientes estimadores:

$$\begin{aligned}\hat{\mu}_{MV} &= \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \\ \hat{\sigma}_{MV}^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\end{aligned}$$

Ahora consideraremos otro ejemplo que no requiere el uso de herramientas de cálculo para la maximización de la verosimilitud,

Ejemplo 3.3.5. Sea X_1, \dots, X_n m.a. aleatoria del modelo $U(0; \theta)$, entonces la función de verosimilitud esta dada por:

$$\ell(\theta; \underline{x}) := f_{\underline{X}}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n \frac{1}{\theta} \mathbb{1}_{(0, \theta)}(x_i) = \frac{1}{\theta^n} \prod_{i=1}^n \mathbb{1}_{(0, \theta)}(x_i)$$

Sin embargo sabemos que:

$$\mathbb{1}_{(0, \theta)}(x_i) = \mathbb{1}_{(y_{n:n}, \infty)}(\theta)$$

Lo que conculimos es que entonces la verosimilitud está dada por:

$$\ell(\theta; \underline{x}) = \frac{1}{\theta^n} \mathbb{1}_{(y_{n:n}, \infty)}(\theta)$$

Donde queda claro que $\frac{1}{\theta^n}$ es una función decreciente, por lo que el máximo se obtiene cuando $\theta = y_{n:n}$, luego entonces:

$$\hat{\theta}_{MV} = Y_{n:n} = \max \{X_1, \dots, X_n\}$$

Ejercicio 3.3.1. 1. Sea X_1, \dots, X_n m.a. del modelo $\text{Uniforme}(\theta, 0)$ con $\theta < 0$.

$$f_X(x) := -\frac{1}{\theta} \mathbf{1}_{(\theta, 0)}(x)$$

Encuentre el estimador de momentos y el de máxima verosimilitud y responda lo siguiente:

a) ¿Los estimadores encontrados son insesgados?

b) ¿El estimador es consistente?

2. Sea X_1, \dots, X_n m.a. del modelo Geométrico(p)

$$f_X(x) = \mathbb{P}(X = x) = (1 - p)^{x-1} p \quad x \in \{1, 2, 3, \dots\}$$

Encuentre el estimador de momentos y el de máxima verosimilitud y responda lo siguiente:

a) ¿Los estimadores encontrados son insesgados?

3. Sea X_1, \dots, X_n m.a. del modelo Exponencial(λ)

$$f_X(x) = \lambda e^{-\lambda x} \quad x \in (0, \infty)$$

Encuentre el estimador de momentos y el de máxima verosimilitud para λ y responda lo siguiente:

a) ¿Los estimadores encontrados son insesgados?

4. Sea X_1, \dots, X_n m.a. de una población con función de densidad dada por:

$$f_X(x; \theta) = \frac{2}{\theta^2}(\theta - x), x \in [0, \theta]$$

a) Aplica el método de momentos para encontrar un estimador para θ . (Denotemos a dicho estimador como $\hat{\theta}_M$)

b) Demuestre que $\hat{\theta}_M$ es insesgado

c) Encuentre el Error cuadrático medio de $\hat{\theta}_M$

d) Pruebe que

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_M) = 0$$

Concluya entonces que el estimador es consistente.

5. Sea X_1, \dots, X_n m.a. de una población con densidad dada por:

$$f_X(x; \theta) = \frac{1}{\theta} x^{\frac{1}{\theta}-1}, x \in [0, 1] \quad \theta > 0$$

■ Demuestre que la densidad pertenece a la familia exponencial

■ Demuestre que $\sum_{i=1}^n \text{Log}(X_i)$ es una estadística suficiente y completa para θ

- Aplica el método de máxima verosimilitud para encontrar un estimador para θ . (Denotemos a dicho estimador como $\hat{\theta}_{MV}$)
- Demuestre que $\hat{\theta}_{MV}$ es insesgado.
- ¿El estimador $\hat{\theta}_{MV}$ es UMVUE? (Justifique su respuesta)

En general sabemos que una de las ventajas de este tipo de estimadores es que tiene una distribución asintótica normal, esto se también se puede generalizar al caso de varios parámetros según el siguiente teorema:

Teorema 3.3.3. Consideremos el modelo $f_X(\cdot; \underline{\theta})$ con $\underline{\theta} = (\theta_1, \dots, \theta_k)$ un vector de parámetros. Entonces los estimadores máximo verosimiles tienen la siguiente propiedad:

$$\underline{\hat{\theta}}_{MV} \stackrel{aprox}{\sim} N_k\left(\underline{\theta}, \frac{1}{n} \mathbf{I}_{\theta}^{-1}\right)$$

Donde $\mathbf{I}_{\theta} \in \mathbb{R}^{k \times k}$ se le conoce como la matriz de información de Fisher por unidad muestral y donde:

$$\mathbf{I}_{\theta} = (I_{ij}) = -\mathbb{E}\left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(f(x; \underline{\theta}))\right)$$

Consideremos el siguiente ejemplo:

Ejemplo 3.3.6. Consideremos al modelo Gamma (α, β) :

$$f_X(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{(0, \infty)}(x)$$

Se puede probar que:

$$\mathbb{E}(X) = \frac{\alpha}{\beta} \quad \text{Var}(X) = \frac{\alpha}{\beta^2} \quad \mathbb{E}(X^2) = \frac{\alpha(1 + \alpha)}{\beta^2}$$

Luego los estimadores de momentos están dados por:

$$\begin{aligned} \frac{\alpha}{\beta} &= \bar{X} \Rightarrow \alpha = \bar{X}\beta \\ \frac{\alpha(1 + \alpha)}{\beta^2} &= \frac{1}{n} \sum_{i=1}^n X_i^2 \Rightarrow \frac{\bar{X}\beta(1 + \bar{X}\beta)}{\beta^2} = \sum_{i=1}^n X_i^2 \Rightarrow \beta = \frac{\bar{X}}{S^2} \end{aligned}$$

Entonces los estimadores son:

$$\hat{\alpha} = \frac{\bar{X}^2}{S^2}; \quad \hat{\beta} = \frac{\bar{X}}{S^2}$$

Por otro lado, el estimador máximo verosímil lo encontramos derivando la Log-verosimilitud:

$$\ell((\alpha; \beta); \underline{x}) := f_{\underline{X}}(x_1, \dots, x_n; (\alpha, \beta)) = \prod_{i=1}^n \frac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i} \mathbb{1}_{(0, \infty)}(x_i)$$

$$L((\alpha; \beta); \underline{x}) = n\alpha \log(\beta) - n \log(\Gamma(\alpha)) + (\alpha - 1) \sum_{i=1}^n \log(X_i) - \beta \sum_{i=1}^n X_i$$

Desafortunadamente el sistema de ecuaciones que se genera no se puede resolver de forma algebraica por lo que en la práctica se utilizan métodos numéricos para encontrar las estimaciones que maximizan a la log-verosimilitud, sin embargo la gran ventaja que tienen los estimadores es que tiene una distribución asintótica normal.

$$\left(\hat{\alpha}_{MV}, \hat{\beta}_{MV}\right) \stackrel{approx}{\sim} N_2\left((\alpha, \beta), \frac{1}{n} \mathbf{I}^{-1}\right)$$

A continuación algunos ejemplos de la distribución asintótica de los estimadores máximos verosímiles

Ejemplo 3.3.7. Algunas distribuciones de la familia exponencial:

- En el modelo Bernulli (θ):

$$\hat{\theta}_{MV} = \bar{X}; \quad I(\theta) = \frac{1}{\theta(1-\theta)}; \Rightarrow \hat{\theta}_{MV} = \bar{X} \stackrel{approx}{\sim} N\left(\theta, \frac{\theta(1-\theta)}{n}\right)$$

- En el modelo Poisson (λ):

$$\hat{\lambda}_{MV} = \bar{X}; \quad I(\lambda) = \frac{1}{\lambda}; \Rightarrow \hat{\lambda}_{MV} = \bar{X} \stackrel{approx}{\sim} N\left(\lambda, \frac{\lambda}{n}\right)$$

- En el modelo Exponencial (λ):

$$\hat{\lambda}_{MV} = \frac{1}{\bar{X}}; \quad I(\lambda) = \frac{1}{\lambda^2}; \Rightarrow \hat{\lambda}_{MV} = \frac{1}{\bar{X}} \stackrel{approx}{\sim} N\left(\lambda, \frac{\lambda^2}{n}\right)$$

Recuerde que para que funcione la aproximación se deben de cumplir las condiciones de regularidad, por ejemplo en el caso del modelo $U(0, \theta)$ el soporte de la densidad depende del parámetro desconocido, por lo tanto no se cumple las condiciones y es incorrecto afirmar que $\hat{\theta}_{MV} = Y_{n:n} \stackrel{approx}{\sim} N\left(\theta, \frac{1}{nI(\theta)}\right)$.

A continuación un ejemplo de invarianza

Ejemplo 3.3.8. Consideremos X_1, \dots, X_n m.a. del modelo *Poisson* (λ), sabemos que $\hat{\lambda}_{MV} = \bar{X}$, sin embargo consideremos que estamos interesados en estimar $\mathbb{P}(X = 0) = e^{-\lambda}$. Observemos que $e^{-\lambda}$ es una transformación biyectiva, entonces:

$$\mathbb{P}(\widehat{X = 0})_{M.V.} = e^{-\bar{X}}$$

Además por las propiedades asintóticas sabemos que:

$$\mathbb{P}(\widehat{X = 0})_{M.V.} = e^{-\bar{X}} \stackrel{\text{aprox}}{\sim} N\left(e^{-\lambda}, \frac{\lambda e^{-2\lambda}}{n}\right)$$

Ejercicio 3.3.2. Resuelva los siguientes ejercicios:

1. Considere una m.a. de tamaño n . Para cada una de las siguientes distribuciones, si es que existe, encuentre el vector de estimadores $\hat{\underline{\theta}}_M$ y $\hat{\underline{\theta}}_{MV}$ para $\underline{\theta}$ dado. En caso de que el estimador exista numéricamente mencione el método que se utiliza para encontrarlo. Justifique su respuesta.

- *Bernoulli* (θ)
- *Binomial* (k, θ), k conocida
- *Binomial* (θ, p), p conocida
- *Binomial* (θ_1, θ_2)
- *Poisson* (θ)
- *Geométrica* (θ)
- *Exponencial* (θ)
- *BinNeg* (k, θ), k conocida
- *BinNeg* (θ, p), p conocida
- *BinNeg* (θ_1, θ_2)
- *Hipergeométrica* (M, K, θ), M y K conocidas
- *Hipergeométrica* (M, θ, m), M y m conocidas
- *Hipergeométrica* (θ, K, m), K y m conocidas
- *Hipergeométrica* (M, θ_1, θ_2), M conocida
- *Hipergeométrica* (θ_1, K, θ_2), K conocida
- *Hipergeométrica* (θ_1, θ_2, m), m conocida

- *Hipergeométrica* $(\theta_1, \theta_2, \theta_3)$
- *Uniforme* $(0, \theta)$
- *Uniforme* $(-\theta, \theta)$
- *Uniforme* (a, θ) , a conocida
- *Uniforme* (θ, b) , b conocida
- *Uniforme* (θ_1, θ_2)
- *Normal* (μ, θ^2) , μ conocida
- *Normal* (θ, σ^2) , σ^2 conocida
- *Normal* (θ_1, θ_2)
- *Log-Normal* (μ, θ^2) , μ conocida
- *Log-Normal* (θ, σ^2) , σ^2 conocida
- *Log-Normal* (θ_1, θ_2)
- *Gamma* (α, θ) , k conocida
- *Gamma* (θ, β) , β conocida
- *Gamma* (θ_1, θ_2)
- *Beta* (a, θ) , a conocida
- *Beta* (θ, b) , b conocida
- *Beta* (θ_1, θ_2)
- *Weibull* (α, θ) , k conocida
- *Weibull* (θ, β) , β conocida
- *Weibull* (θ_1, θ_2)
- *Pareto* (a, θ) , a conocida
- *Pareto* (θ, b) , b conocida
- *Pareto* (θ_1, θ_2)
- *Ji-cuadrada* (θ) Recuerde que esta distribución puede verse como una *Gamma* $(\frac{\theta}{2}, \frac{1}{2})$
- *Rayleigh* (θ) Recuerde que esta distribución puede verse como *Weibull* $(2, \theta)$

Hint: Recuerde que en el método de máxima verosimilitud debe considerarse la indicadora, si esta depende de los parámetros desconocidos no se puede derivar y debe recurrir a otros métodos de maximización como los gráficos o numéricos.

2. Considere una m.a. de tamaño n Uniforme $(0, \theta)$ Encuentre los estimadores MV para la media y la varianza.
3. Considere una m.a. de tamaño n Bin $(1, \theta)$ sabiendo que \bar{X} es un estimador suficiente y completo para θ encuentre el UMVUE para: a) 3θ ; b) $3\theta - 1$; c) Encuentre el estimador MV para θ y $\theta(1 - \theta)$; d) Proponga un estimador insesgado para θ y use el Teorema de Rao-Blackwell para mejorarlo
4. Considere una m.a. de tamaño n , estimar el parámetro θ por el método de momentos. La función de densidad es la siguiente:

$$f(x; \theta) = \frac{2(\theta - x)}{\theta^2} \quad l_{0 < x < \theta} \quad (3.7)$$

Verifique si el estimador es insesgado y suficiente

5. Sea X_1, \dots, X_n una m.a. con densidad común

$$f(x; \theta) = \theta x^{\theta-1} \quad l_{0 < x < 1}, \quad \theta > 0 \quad (3.8)$$

Encuentre una estadística suficiente y diga si es completa. Justifique su respuesta.

6. Sea X_1, \dots, X_n una m.a. con densidad común

$$f(x; \theta) = e^{-(x-\theta)} \quad l_{\theta < x < \infty} \quad -\infty < \theta < \infty \quad (3.9)$$

- Encuentre una estadística suficiente
- Encuentre el estimador de MV de θ
- Encuentre el estimador por momentos de θ
- Si existe una estadística suficiente y completa, encuéntrala
- Encuentre el UMVUE para θ si es que existe

7. Sea X_1, \dots, X_n una m.a. con densidad común

$$f(x; \theta) = \theta x^{-2} \quad l_{\theta < x < \infty} \quad 0 < \theta < \infty \quad (3.10)$$

Encuentre el estimador MV para θ

8. Sea X una sola observación de la distribución $N(0, \theta)$. a) Es X una estadística suficiente?; b) Es X^2 un estimador insesgado para θ ?

-
9. Sea X_1, \dots, X_n una m.a. con densidad común $N(\mu, \sigma^2)$. Encuentre el UMVUE para: a) $6\mu + 4\sigma^2$; b) $\mu^2 - 5\sigma^2$
10. Sea μ el verdadero valor del cociente intelectual de cierto estudiante. Para medir su C.I. realiza un test y se sabe que las puntuaciones obtenidas se distribuye normalmente, con media μ y desviación estándar 5. El estudiante realiza el test y alcanza la puntuación 130. Cuál es es estimador MV de μ ?

Capítulo 4

Intervalos de Confianza

4.1. Introducción

El objetivo principal para encontrar un intervalo de confianza es encontrar dos estadísticas $L(\underline{X})$ y $U(\underline{X})$ tales que:

$$\mathbb{P}(L(\underline{X}) \leq \theta \leq U(\underline{X})) = 1 - \alpha$$

Donde α toma valores pequeños (0.01, 0.05, 0.1). Debe observarse que el límite inferior y superior al ser función de variables aleatorias entonces también son aleatorios, luego entonces debido a que θ es considerado un parámetro fijo, la interpretación que debemos dar a la expresión anterior es **la probabilidad de que el intervalo cubra al parámetro desconocido θ** . Por otro lado una vez observada la muestra y dado que ya no tenemos variables aleatorias, entonces al intervalo generado por las estadísticas evaluadas en la muestra observada se le conoce como intervalo al $1 - \alpha \times 100\%$ de confianza para θ .

A continuación describimos el método tradicional para encontrar los estadísticos $L(\underline{X})$ y $U(\underline{X})$ que dan origen a un intervalo de confianza.

Definición 4.1.1 (Cantidad Pivotal para θ). *Una cantidad pivotal es una función de la muestra X_1, \dots, X_n cuya regla de correspondencia debe de involucrar a θ pero que su función de distribución **no** depende de parámetros desconocidos*

4.2. Intervalos de confianza: Caso Normal

4.2.1. Intervalo para la media μ , σ^2 conocida

Supongamos que tenemos X_1, \dots, X_n m.a. del una población con distribución $N(\mu, \sigma^2)$ y analicemos el caso en donde σ^2 es conocida. Entonces definamos:

$$Q_1(\underline{X}) := \frac{X_1 - \mu}{\sigma} \sim N(0, 1)$$

Debe de quedar claro que Q_1 es entonces una cantidad pivotal para μ . Luego entonces, por sus características distribuciones de Q_1 podemos afirmar que:

$$\mathbb{P}\left(-Z_{(1-\alpha/2)} \leq \frac{X_1 - \mu}{\sigma} \leq Z_{(1-\alpha/2)}\right) = 1 - \alpha \quad (4.1)$$

Donde $Z_{(1-\alpha/2)}$ como siempre, representa el cuantíl $(1-\alpha/2)$ de una distribución normal estándar. Por otro lado también podemos afirmar que:

$$\mathbb{P}\left(-\infty \leq \frac{X_1 - \mu}{\sigma} \leq Z_{(1-\alpha)}\right) = 1 - \alpha \quad (4.2)$$

La cantidad pivotal Q_1 que hemos definido arriba tiene el pequeño defecto que no toma en cuenta a toda la muestra, luego entonces resultaría interesante tener otra cantidad pivotal que si hiciera uso de toda la información de la muestra, definamos entonces.

$$Q_2(\underline{X}) := \sqrt{n} \left(\frac{\bar{X} - \mu}{\sigma} \right) \sim N(0, 1)$$

Q_2 es entonces una cantidad pivotal para μ de donde afirmamos también que:

$$\mathbb{P}\left(-Z_{(1-\alpha/2)} \leq \sqrt{n} \left(\frac{\bar{X} - \mu}{\sigma} \right) \leq Z_{(1-\alpha/2)}\right) = 1 - \alpha \quad (4.3)$$

En cada uno de las tres últimas expresiones (4.1, 4.2, 4.3) podemos llevar a cabo un despeje del parámetro desconocido μ obteniendo las siguientes expresiones:

De la ecuación (4.1) obtenemos:

$$\mathbb{P}\left(X_1 - \sigma Z_{(1-\alpha/2)} \leq \mu \leq X_1 + \sigma Z_{(1-\alpha/2)}\right) = 1 - \alpha \quad (4.4)$$

De la ecuación (4.2) obtenemos:

$$\mathbb{P}(X_1 - \sigma Z_{(1-\alpha)} \leq \mu < \infty) = 1 - \alpha \quad (4.5)$$

De la ecuación (4.3) obtenemos:

$$\mathbb{P}\left(\bar{X} - \frac{\sigma}{\sqrt{n}} Z_{(1-\alpha/2)} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{(1-\alpha/2)}\right) = 1 - \alpha \quad (4.6)$$

Hemos encontrado entonces 3 intervalos que cumplen con la propiedad deseada de cubrir al parámetros desconocido con una probabilidad $1 - \alpha$, surge entonces la pregunta de saber con cual intervalo quedarnos. La respuesta lógica es escoger aquel intervalo de longitud mínima. Veamos entonces la longitud de los tres intervalos que hemos generado a partir de las cantidades pivotaes:

1. *Longitud* = $2\sigma Z_{(1-\alpha/2)}$
2. *Longitud* = ∞
3. *Longitud* = $2\frac{\sigma}{\sqrt{n}} Z_{(1-\alpha/2)}$

Es claro entonces que el mejor intervalo de los tres posibles es el último, esto era de esperarse pues el intervalo está en función de una estadística suficiente. En general los intervalos óptimos estarán en función de estadísticas con propiedades importantes como suficiencia y completos.

A continuación hacemos un breve paréntesis sobre el tamaño de muestra necesario para hacer una estimación con cierto error haciendo uso de la teoría desarrollada en la construcción de intervalos de confianza

Tamaño de muestra

Vimos entonces que la mejor intervalo es el proporcionado en la ecuación (4.6) , imaginemos ahora que queremos construir un intervalo de longitud $2d$ y con una confianza de $1 - \alpha$. Surge de manera natural la pregunta de saber cuál deberá ser el tamaño de muestra necesario para cumplir este objetivo. La respuesta se obtiene despejando entonces a n de la siguiente igualdad:

$$2\frac{\sigma}{\sqrt{n}} Z_{(1-\alpha/2)} = 2d \quad \Rightarrow \quad n = \frac{\sigma^2 Z_{(1-\alpha/2)}^2}{d^2}$$

Ahora notemos lo siguiente si $\frac{\sigma}{\sqrt{n}} Z_{(1-\alpha/2)} = d$ entonces por (4.6) obtenemos que:

$$\mathbb{P}(\bar{X} - d \leq \mu \leq \bar{X} + d) = 1 - \alpha \Rightarrow \mathbb{P}(|\bar{X} - \mu| \leq d) = 1 - \alpha$$

Luego entonces, esto nos dice que hemos encontrado el tamaño de muestra necesario para que el estimador $\hat{\mu} = \bar{X}$ tenga un error menor a d unidades con una probabilidad de $1 - \alpha$.

Ejemplo 4.2.1. *Supongamos que queremos inferir la estatura promedio de un mexicano mayor de 18 años. Para resolver este problema asumiremos la hipótesis distribucionales de que la estatura de los mexicanos adultos se pueden modelar por medio del modelo normal. Supongamos además para fines prácticos que conocemos la desviación estándar del modelo $\sigma = 10$. La pregunta es: ¿Qué tamaño de muestra requiero para poder estimar μ si quiero tener un error menor a 1 cm con una confianza del 95 %?. La respuesta según la teoría desarrollada nos dice que la n requerida es:*

$$n = \frac{10^2 * (1.96)^2}{1} = 384.16$$

Debido a que el tamaño de muestra solo toma valores enteros, entonces tomamos $n = 385$. Lo que nos afirma entonces que debemos tener una muestra de ese tamaño para que nuestro intervalo de confianza tenga una longitud de 2cm, lo que se traduce en un error menor a 1cm en nuestra estimación con una confianza del 95 %

4.2.2. Intervalo para la media μ , σ^2 desconocida

Cuando σ^2 es desconocida entonces la expresión:

$$\sqrt{n} \left(\frac{\bar{X} - \mu}{\sigma} \right) \sim N(0, 1)$$

Deja de ser cantidad pivotal para μ , pues σ^2 es desconocida y recordemos que la única cantidad desconocida debe de ser μ , por lo tanto debemos de buscar alguna otra cantidad, para ello recordemos que:

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

Además recordemos que si $Z \sim N(0, 1)$ y $Y \sim \chi_{(n-1)}^2$ (independientes) entonces

$$\frac{Z}{\sqrt{\frac{Y}{n-1}}} \sim t_{n-1} \text{ T student}$$

Por lo tanto concluimos que: (Obs: Recordar la independencia entre \bar{X} y S^2)

$$\frac{\sqrt{n} \left(\frac{\bar{X} - \mu}{\sigma} \right)}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \frac{1}{n-1}}} = \sqrt{n} \left(\frac{\bar{X} - \mu}{\hat{\sigma}} \right) \sim t_{n-1}$$

Donde $\hat{\sigma}^2 := \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$, la cual ya puede ser considerada cantidad pivotal para μ , entonces nuestro intervalo de confianza después de llevar a cabo el despeje estaría dado por:

$$\mathbb{P} \left(\bar{X} - \frac{\hat{\sigma}}{\sqrt{n}} t_{n-1}^{1-\alpha/2} \leq \mu \leq \bar{X} + \frac{\hat{\sigma}}{\sqrt{n}} t_{n-1}^{1-\alpha/2} \right) = 1 - \alpha$$

Donde $t_{n-1}^{1-\alpha/2}$ es el cuantil $1 - \alpha/2$ de una distribución t -student con $n - 1$ grados de libertad.

4.2.3. Intervalo para σ^2 , μ conocida

Suponiendo μ conocida, consideremos la siguiente cantidad pivotal.

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$$

La anterior es claro que cumple las condiciones de ser llamada cantidad pivotal pues recordemos que μ es conocida, luego entonces considerando a y b dos cantidades tales que:

$$\mathbb{P} \left(a \leq \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \leq b \right) = 1 - \alpha \quad (4.7)$$

Al despejar a σ^2 de esta expresión se obtiene el siguiente intervalo de confianza:

$$\mathbb{P} \left(\sum_{i=1}^n \frac{(X_i - \mu)^2}{b} \leq \sigma^2 \leq \sum_{i=1}^n \frac{(X_i - \mu)^2}{a} \right) = 1 - \alpha \quad (4.8)$$

Donde a y b se escogen de tal manera que se cumpla la ecuación (4.7) y que se obtenga la longitud mínima en la ecuación (4.8). Este problema desafortunadamente no arroja una solución algebraica para a y b por lo que su solución se obtiene por métodos numéricos. Algunos libros por facilidad optan por definir:

$$a = \chi_n^{2(\alpha/2)}; \quad y \quad b = \chi_n^{2(1-\alpha/2)};$$

Quedando el intervalo de confianza como sigue:

$$\mathbb{P} \left(\sum_{i=1}^n \frac{(X_i - \mu)^2}{\chi_n^{2(1-\alpha/2)}} \leq \sigma^2 \leq \sum_{i=1}^n \frac{(X_i - \mu)^2}{\chi_n^{2(\alpha/2)}} \right) = 1 - \alpha$$

Sin embargo debe de quedar claro que este intervalo no necesariamente genera al óptimo (de longitud menor) sin embargo es el más práctico para fines didácticos.

4.2.4. Intervalo para σ^2 , μ desconocida

Si ahora suponemos μ desconocida simplemente procederemos a estimar a μ por medio de \bar{X} y sustituyendolo en la cantidad pivotal anterior, de esta manera se obtiene la siguiente cantidad pivotal:

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

Y de forma análoga el intervalo de confianza queda como:

$$\mathbb{P} \left(\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{b} \leq \sigma^2 \leq \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{a} \right) = 1 - \alpha \quad (4.9)$$

Donde nuevamente a y b son cantidades que hacen que cumpla la confianza y se tenga la cobertura mínima, y como en el punto anterior, a veces a y b son calculados como los siguientes cuantiles para fines prácticos :

$$a = \chi_{n-1}^{2(\alpha/2)}; \quad y \quad b = \chi_{n-1}^{2(1-\alpha/2)};$$

A continuación construiremos intervalos de confianza para la diferencia de medias en 2 poblaciones Normales, este un punto muy importante ya que a través de estos intervalos se concluyen diferencias de medias significativas que muchas veces son utilizadas para identificar la eficacia de tratamientos.

4.2.5. Intervalo para la diferencia de Medias en 2 dos poblaciones; Caso 1: Varianzas conocidas

Supongamos que tenemos X_1, \dots, X_{n_1} y Y_1, \dots, Y_{n_2} dos muestras aleatorias de dos poblaciones normales $N(\mu_x, \sigma_x^2)$, $N(\mu_y, \sigma_y^2)$ respectivamente. Además supongamos que tanto σ_x^2 como σ_y^2 son conocidas. Nuestro objetivo es lograr construir un intervalo de confianza para la diferencia de medias $(\mu_x - \mu_y)$. Para ello primero buscaremos una cantidad pivotal de la siguiente manera,

sabemos que por ser v.a. normales que:

$$\bar{X} \sim N\left(\mu_x, \frac{\sigma_x^2}{n_1}\right) \quad y \quad \bar{Y} \sim N\left(\mu_y, \frac{\sigma_y^2}{n_2}\right)$$

Además por la independencia del muestreo entre las dos poblaciones se tiene que \bar{X} es independiente de \bar{Y} por lo que es fácil ver que:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}} \sim N(0, 1)$$

Lo anterior ya es una cantidad pivotal la diferencia de medias $(\mu_x - \mu_y)$, pues recordemos que estamos suponiendo que tanto σ_x^2 como σ_y^2 son conocidas. Luego entonces de esta cantidad pivotal obtenemos el siguiente intervalo:

$$\mathbb{P}\left((\bar{X} - \bar{Y}) - \sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}} Z_{1-\alpha/2} \leq \mu_x - \mu_y \leq (\bar{X} - \bar{Y}) + \sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}} Z_{1-\alpha/2}\right) = 1 - \alpha$$

Donde denotamos a $Z_{1-\alpha/2}$ como el cuantíl $1 - \alpha/2$ de la distribución Normal Estándar.

4.2.6. Intervalo para la diferencia de Medias en 2 dos poblaciones; Caso 2: Varianzas desconocidas pero iguales

Bajo el mismo contexto anterior pero asumiendo que $\sigma_x^2 = \sigma_y^2 = \sigma^2$ se obtiene que:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \sim N(0, 1)$$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$$

Sin embargo al ser σ^2 desconocida, la expresión anterior no puede ser considerada una cantidad pivotal para la diferencia de medias $\mu_x - \mu_y$, la forma en como atacaremos este problema es la siguiente, primero notemos que tenemos muestra de dos poblaciones normales, entonces:

$$\sum_{i=1}^{n_1} \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi_{(n_1-1)}^2 \quad \sum_{i=1}^{n_2} \frac{(Y_i - \bar{Y})^2}{\sigma^2} \sim \chi_{(n_2-1)}^2$$

Entonces por independencia entre ambas muestras podemos concluir que:

$$\sum_{i=1}^{n_1} \frac{(X_i - \bar{X})^2}{\sigma^2} + \sum_{i=1}^{n_2} \frac{(Y_i - \bar{Y})^2}{\sigma^2} = \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{\sigma^2} \sim \chi^2_{(n_1+n_2-2)}$$

Tenemos entonces una cantidad con distribución $N(0, 1)$ y otra con distribución χ^2 , suponiendo independencia entre estas dos cantidades se concluye que:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \bigg/ \sqrt{\frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{\sigma^2(n_1 + n_2 - 1)}} \sim t_{n_1+n_2-2}$$

Definiendo $\hat{\sigma}_p^2$ como:

$$\hat{\sigma}_p^2 = \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)\hat{\sigma}_x^2 + (n_2 - 1)\hat{\sigma}_y^2}{n_1 + n_2 - 2}$$

Obtemos que:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\hat{\sigma}_p^2 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Lo anterior ya es una cantidad pivotal la cual obtiene el siguiente intervalo de confianza para la diferencia de medias:

$$\mathbb{P} \left((\bar{X} - \bar{Y}) - \hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{(n_1+n_2-2)}^{(1-\alpha/2)} \leq \mu_x - \mu_y \leq (\bar{X} - \bar{Y}) + \hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{(n_1+n_2-2)}^{(1-\alpha/2)} \right) = 1 - \alpha$$

4.2.7. Intervalo para la diferencia de Medias en 2 dos poblaciones; Caso 3: Varianzas desconocidas y no iguales

Este quizás sea el caso mas común al que nos estaremos enfrentando, pero desafortunadamente no existe una cantidad pivotal que se pueda utilizar, por lo general para atacar este problema se utilizan los estimadores de σ^2 de cada población y se obtiene la siguiente cantidad pivotal:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\hat{\sigma}_x^2}{n_1} + \frac{\hat{\sigma}_y^2}{n_2}}} \underset{\text{aprox}}{\sim} t_\nu$$

Donde:

$$\nu = \left(\frac{\hat{\sigma}_x^2}{n_1} + \frac{\hat{\sigma}_y^2}{n_2} \right)^2 \bigg/ \left(\frac{\left(\frac{\hat{\sigma}_x^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{\hat{\sigma}_y^2}{n_2} \right)^2}{n_2 - 1} \right)$$

Luego entonces un intervalo de confianza aproximado viene dado por:

$$\mathbb{P} \left((\bar{X} - \bar{Y}) - \sqrt{\frac{\hat{\sigma}_x^2}{n_1} + \frac{\hat{\sigma}_y^2}{n_2}} t_{(\nu)}^{(1-\alpha/2)} \leq \mu_x - \mu_y \leq (\bar{X} - \bar{Y}) + \sqrt{\frac{\hat{\sigma}_x^2}{n_1} + \frac{\hat{\sigma}_y^2}{n_2}} t_{(\nu)}^{(1-\alpha/2)} \right) = 1 - \alpha$$

Hacer intervalos de confianza para la diferencia se puede utilizar como sigue, imaginemos que tenemos 2 poblaciones y queremos ver si $\mu_x > \mu_y$ a través una muestra de cada población, una forma de probarlo es construir un intervalo de confianza para la diferencia, luego entonces, si el número 0 se encuentra dentro del intervalo se dice que no hay diferencia significativa, pero si resulta que el 0 está a la izquierda del intervalo entonces $0 < (\mu_x - \mu_y)$ lo que concluye que estadísticamente $\mu_x > \mu_y$.

Supongamos ahora que estamos interesados en verificar la igualdad entre las varianzas de dos poblaciones, para atacar este problema se acostumbra construir intervalos para el cociente de varianzas, luego entonces si el 1 se encuentra en dicho intervalo se dice que no hay diferencias significativas, a continuación se exponen las cantidades pivotaes para generar los intervalos pertinentes.

4.2.8. Intervalo de confianza para el cociente de varianza Caso 1: Medias Conocidas

Si las medias son conocidas sabemos que :

$$\sum_{i=1}^{n_1} \frac{(X_i - \mu_x)^2}{\sigma_x^2} \sim \chi_{(n_1)}^2 \quad y \quad \sum_{i=1}^{n_2} \frac{(Y_i - \mu_y)^2}{\sigma_y^2} \sim \chi_{(n_2)}^2$$

Por la independencia de los muestreos se tiene que al hacer el cociente de las χ^2 's se tiene que:

$$\sum_{i=1}^{n_2} \frac{(Y_i - \mu_y)^2}{\sigma_y^2 n_2} \bigg/ \sum_{i=1}^{n_1} \frac{(X_i - \mu_x)^2}{\sigma_x^2 n_1} \sim F_{n_2, n_1}$$

Por lo tanto

$$\left(\frac{\sigma_x^2}{\sigma_y^2} \right) \left(\frac{\sum_{i=1}^{n_2} \frac{(Y_i - \mu_y)^2}{n_2}}{\sum_{i=1}^{n_1} \frac{(X_i - \mu_x)^2}{n_1}} \right) \sim F_{n_2, n_1}$$

La cual ya es una cantidad pivotal para el cociente de medias $\left(\frac{\sigma_x^2}{\sigma_y^2}\right)$, luego eligiendo a y b dos número reales tales que:

$$\mathbb{P} \left(a \leq \left(\frac{\sigma_x^2}{\sigma_y^2} \right) \left(\frac{\sum_{i=1}^{n_2} \frac{(Y_i - \mu_y)^2}{n_2}}{\sum_{i=1}^{n_1} \frac{(X_i - \mu_x)^2}{n_1}} \right) \leq b \right) = 1 - \alpha$$

Se obtiene el siguiente intervalo:

$$\mathbb{P} \left(\frac{\sum_{i=1}^{n_1} \frac{(X_i - \mu_x)^2}{n_1}}{\sum_{i=1}^{n_2} \frac{(Y_i - \mu_y)^2}{n_2}} a \leq \left(\frac{\sigma_x^2}{\sigma_y^2} \right) \leq \frac{\sum_{i=1}^{n_1} \frac{(X_i - \mu_x)^2}{n_1}}{\sum_{i=1}^{n_2} \frac{(Y_i - \mu_y)^2}{n_2}} b \right) = 1 - \alpha$$

Sin embargo la obtención de a y b se hace por medio de métodos numéricos para obtener el intervalo óptimo (de menor longitud), sin embargo en la literatura se acostumbra hacer:

$$a = F_{n_2, n_1}^{\alpha/2} \quad b = F_{n_2, n_1}^{1-\alpha/2}$$

Donde F_{n_2, n_1}^{α} es el cuantil α de la distribución F con n_2, n_1 grados de libertad.

Ejercicio 4.2.1. Sea X_1, \dots, X_{n_1} y Y_1, \dots, Y_{n_2} m.a. de una población normal con medias desconocidas μ_x y μ_y respectivamente. Se desea encontrar un intervalo de confianza para el cociente de las varianzas $\frac{\sigma_x^2}{\sigma_y^2}$. Encontrar una cantidad pivotal y construir un intervalo de confianza al $1 - \alpha$ de confianza para esta cantidad.

4.3. Intervalo de Confianza para poblaciones no Normales

Hasta ahora fuimos capaces de encontrar cantidades pivotaes para los parámetros de una distribución normal, sin embargo muchas veces nos enfrentaremos al problema de querer encontrar intervalos de confianza en familias distintas a la normal o pero aun, a veces estaremos interesados en construir intervalos de confianza para funciones de cierto parámetro. Hemos visto que la clave para encontrar el intervalo de confianza se centra en poder localizar una cantidad pivotal adecuada, sin embargo en muchas ocasiones esto es una tarea difícil mas que nada debido no hay una metodología para encontrarlos.

Ejemplo 4.3.1. Supongamos que tenemos X_1, \dots, X_n una m.a. de una población $U(0, \theta)$ se puede probar que:

$$T_1 = \frac{Y_{n:n}}{\theta}$$

Es una cantidad pivotal y por tanto:

$$\mathbb{P} \left(\frac{Y_{n:n}}{\xi_{1-\alpha/2}} \leq \theta \leq \frac{Y_{n:n}}{\xi_{\alpha/2}} \right) = 1 - \alpha$$

Es un intervalo de confianza al $(1 - \alpha)100$ % para θ , donde ξ_α es el cuantil α de la distribución asociada a $T_1 = \frac{Y_{n:n}}{\theta}$.

Ejemplo 4.3.2. Supongamos que tenemos X_1, \dots, X_n m.a. de una población $Exp(\lambda)$, entonces se puede probar que:

$$T_2 = \lambda \left(\sum_{i=1}^n X_i \right) \sim Gamma(n, 1)$$

Es cantidad pivotal y por tanto:

$$\mathbb{P} \left(\frac{\sum_{i=1}^n X_i}{\xi_{1-\alpha/2}} \leq \lambda \leq \frac{\sum_{i=1}^n X_i}{\xi_{\alpha/2}} \right) = 1 - \alpha$$

Es un intervalo de confianza al $(1 - \alpha)100$ % para λ , donde ξ_α es el cuantil α de la distribución asociada a $T_2 \sim Gamma(n, 1)$.

Cuando no podemos encontrar una cantidad pivotal adecuada, muchas veces debemos de recurrir a resultados asintóticos como el **Teorema del Limite Central** (3.2.2) o bien recordar la propiedad distribucional asintótica con la que cuentan los estimadores máximo verosímiles $\hat{\theta}_{MV} \stackrel{approx}{\sim} N \left(\theta, \frac{1}{nI(\theta)} \right)$.

Por ejemplo, gracias al *TLC* (3.2.2) podemos construir intervalos de confianza para la media μ de cualquier distribución F_x con segundo momento finito de la siguiente manera:

- Sabemos que bajo las condiciones del *TLC* (3.2.2) podemos afirmar que :

$$\bar{X} \stackrel{approx}{\sim} N \left(\mu, \frac{\sigma^2}{n} \right) \Rightarrow \sqrt{n} \left(\frac{\bar{X} - \mu}{\sigma} \right) \stackrel{approx}{\sim} N(0, 1)$$

- Si suponemos σ^2 conocida la última expresión puede utilizarse como cantidad pivotal y afirmar que:

$$\mathbb{P} \left(\bar{X} - \frac{\sigma}{\sqrt{n}} Z_{(1-\alpha/2)} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} Z_{(1-\alpha/2)} \right) \stackrel{approx}{=} 1 - \alpha$$

es un intervalo aproximado al $(1 - \alpha)100$ % de confianza para μ

- Si suponemos que σ^2 es desconocida, se hace una segunda aproximación y se estima σ^2 por medio del estimador:

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

Quedando el intervalo de confianza como:

$$\mathbb{P} \left(\bar{X} - \frac{\hat{\sigma}^2}{\sqrt{n}} Z_{(1-\alpha/2)} \leq \mu \leq \bar{X} + \frac{\hat{\sigma}^2}{\sqrt{n}} Z_{(1-\alpha/2)} \right) \stackrel{aprox}{=} 1 - \alpha$$

Sin embargo algunos autores aconsejan cambiar los cuantiles de la Normal por los de una t -student, no obstante si tenemos una muestra grande ambos cuantiles son muy similares y es indiferente entre usar una t o una Normal para construir el intervalo.

A continuación veremos el ejemplo de como construir un intervalo de confianza basados en estimadores máximo verosímiles, la idea del método recae en la siguiente aproximación:

$$\hat{\theta}_{MV} \stackrel{aprox}{\sim} N \left(\theta, \frac{1}{nI(\theta)} \right) \Rightarrow \sqrt{nI(\theta)} (\hat{\theta}_{MV} - \theta) \stackrel{aprox}{\sim} N(0, 1)$$

Debe de quedar claro que la última expresión ya es una cantidad pivotal (aproximada) para θ , sin embargo despejar θ de la ecuación anterior a veces puede llegar a ser difícil, en muchas ocasiones y para evitar el despeje de θ en la cantidad pivotal se acostumbra llevar a cabo una segunda aproximación y afirmar que:

$$\hat{\theta}_{MV} \stackrel{aprox}{\sim} N \left(\theta, \frac{1}{nI(\hat{\theta}_{MV})} \right) \Rightarrow \sqrt{nI(\hat{\theta}_{MV})} (\hat{\theta}_{MV} - \theta) \stackrel{aprox}{\sim} N(0, 1)$$

De donde claramente el intervalo de confianza generado es de la forma:

$$\mathbb{P} \left(\hat{\theta}_{MV} - \frac{1}{\sqrt{nI(\hat{\theta}_{MV})}} Z_{1-\alpha/2} \leq \theta \leq \hat{\theta}_{MV} + \frac{1}{\sqrt{nI(\hat{\theta}_{MV})}} Z_{1-\alpha/2} \right) \stackrel{aprox}{=} 1 - \alpha$$

Consideremos el siguiente ejemplo:

Ejemplo 4.3.3. Sea X_1, \dots, X_n una m.a. del modelo Poisson(λ). Construiremos un intervalo de confianza aproximado para $\lambda = \mathbb{E}(X)$, primero notemos que:

$$\hat{\lambda}_{MV} = \bar{X} \Rightarrow \bar{X} \stackrel{aprox}{\sim} N \left(\lambda, \frac{1}{nI(\lambda)} \right)$$

Como en este modelo:

$$I(\lambda) = -\mathbb{E}\left(\frac{\partial^2}{\partial \lambda^2} f_X(X; \lambda)\right) = \frac{1}{\lambda}$$

Se concluye entonces que la cantidad pivotal es:

$$\frac{\hat{\lambda}_{MV} - \lambda}{\sqrt{\frac{\lambda}{n}}} \stackrel{aprox}{\sim} N(0, 1)$$

Despejar λ no es sencillo pero se puede llevar a cabo:

$$\begin{aligned} \mathbb{P}\left(\left|\frac{\hat{\lambda}_{MV} - \lambda}{\sqrt{\frac{\lambda}{n}}}\right| \leq Z_{1-\alpha}\right) &= 1 - \alpha \\ \mathbb{P}\left(\left(\frac{\hat{\lambda}_{MV} - \lambda}{\sqrt{\frac{\lambda}{n}}}\right)^2 \leq Z_{1-\alpha}^2\right) &= 1 - \alpha \\ \mathbb{P}\left(\left(\hat{\lambda}_{MV} - \lambda\right)^2 \leq Z_{1-\alpha}^2 \frac{\lambda}{n}\right) &= 1 - \alpha \\ \mathbb{P}\left(\hat{\lambda}_{MV}^2 - 2\hat{\lambda}_{MV}\lambda + \lambda^2 \leq Z_{1-\alpha}^2 \frac{\lambda}{n}\right) &= 1 - \alpha \\ \mathbb{P}\left(n\hat{\lambda}_{MV}^2 - 2n\hat{\lambda}_{MV}\lambda + n\lambda^2 \leq Z_{1-\alpha}^2 \lambda\right) &= 1 - \alpha \\ \mathbb{P}\left(n\lambda^2 - \left(Z_{1-\alpha}^2 + 2n\hat{\lambda}_{MV}\right)\lambda + n\hat{\lambda}_{MV}^2 \leq 0\right) &= 1 - \alpha \end{aligned}$$

La última expresión es una ecuación de segundo grado para λ la cual puede ser resuelta via la formula tradicional para encontrar raíces, haciendo lo siguiente:

$$\begin{aligned} a &= n \\ b &= -\left(Z_{1-\alpha}^2 + 2n\hat{\lambda}_{MV}\right) \\ c &= n\hat{\lambda}_{MV}^2 \end{aligned}$$

Obteniendo finalmente el siguiente intervalo de confianza:

$$\mathbb{P}\left(\frac{-b - \sqrt{b^2 - 4ac}}{2a} \leq \lambda \leq \frac{-b + \sqrt{b^2 - 4ac}}{2a}\right) = 1 - \alpha$$

Otra forma de obtener un intervalo de confianza y para no pasar por todo el proceso de despeje

es hacer otra aproximación y utilizar el hecho que:

$$\frac{\hat{\lambda}_{MV} - \lambda}{\sqrt{\frac{\hat{\lambda}_{MV}}{n}}} \stackrel{aprox}{\sim} N(0, 1)$$

De donde fácilmente el intervalo de confianza obtenido está dado por:

$$\mathbb{P} \left(\hat{\lambda}_{MV} - \sqrt{\frac{\hat{\lambda}_{MV}}{n}} Z_{1-\alpha/2} \leq \lambda \leq \hat{\lambda}_{MV} + \sqrt{\frac{\hat{\lambda}_{MV}}{n}} Z_{1-\alpha/2} \right) = 1 - \alpha$$

Debe de quedar claro que ambos son intervalos aproximados, sin embargo el segundo requirió una segunda aproximación, lo cual sugiere que su cobertura esperada no sera tan buena como la del primer caso sin embargo via simulaciones se puede probar que ambos intervalos son iguales con n mayores a 30.

4.3.1. Intervalo de Confianza para transformaciones de los parámetros

Imaginemos que ahora no estamos interesados es construir un intervalo para θ sino que estamos interesados en construir un intervalo para alguna transformación $f(\theta)$, para ello podemos recurrir a la siguiente aproximación:

$$f(\hat{\theta}_{MV}) \stackrel{aprox}{\sim} N \left(f(\theta), \frac{(f'(\theta))^2}{nI(\theta)} \right) \Rightarrow \frac{f(\hat{\theta}_{MV}) - f(\theta)}{\sqrt{\frac{(f'(\theta))^2}{nI(\theta)}}} \stackrel{aprox}{\sim} N(0, 1)$$

La última expresión ya es una cantidad pivotal para $f(\theta)$ sin embargo, casi siempre esta ecuación será difícil de trabajar por lo que en general lo que se hace es hacer una segunda aproximación y utilizar la siguiente cantidad pivotal:

$$\frac{f(\hat{\theta}_{MV}) - f(\theta)}{\sqrt{\frac{(f'(\hat{\theta}_{MV}))^2}{nI(\hat{\theta}_{MV})}}} \stackrel{aprox}{\sim} N(0, 1)$$

De donde el intervalo de confianza aproximado está dado por:

$$\mathbb{P} \left(f(\hat{\theta}_{MV}) - \sqrt{\frac{(f'(\hat{\theta}_{MV}))^2}{nI(\hat{\theta}_{MV})}} Z_{1-\alpha/2} \leq f(\theta) \leq f(\hat{\theta}_{MV}) + \sqrt{\frac{(f'(\hat{\theta}_{MV}))^2}{nI(\hat{\theta}_{MV})}} Z_{1-\alpha/2} \right) = 1 - \alpha$$

Ejemplo 4.3.4. Sea X_1, \dots, X_n m.a. del modelo Gamma suponiendo α conocido, constituir un intervalo de confianza para $\mathbb{P}(X > x)$ dado por:

$$\mathbb{P}(X > x) = \int_x^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} dt$$

Como α es conocido debe de quedar claro que $\mathbb{P}(X > x)$ queda en función de β y por tanto lo que nos piden es construir un intervalo de confianza para una transformación del parámetro desconocido β .

Primero obtenemos el estimador máximo verosímil y la información de fisher por unidad muestral correspondiente.

$$\hat{\beta}_{MV} = \frac{\alpha}{\bar{X}}; \quad I(\beta) = -\mathbb{E} \left(\frac{\partial^2}{\partial \beta^2} f_X(X; \beta) \right) = \frac{\alpha}{\beta^2}$$

Definamos $f(\beta) := \mathbb{P}(X > x)$, entonces:

$$f(\hat{\beta}_{MV}) \stackrel{aprox}{\sim} N \left(f(\beta), \frac{(f'(\beta))^2 \beta^2}{n\alpha} \right)$$

De donde la cantidad pivotal nos queda como:

$$\frac{f(\hat{\beta}_{MV}) - f(\beta)}{\sqrt{\frac{(f'(\beta))^2 \beta^2}{n\alpha}}} \stackrel{aprox}{\sim} N(0, 1)$$

Debido a la complejidad que tiene la función f haremos una segunda aproximación obteniendo la siguiente cantidad:

$$\frac{f(\hat{\beta}_{MV}) - f(\beta)}{\sqrt{\frac{(f'(\hat{\beta}_{MV}))^2 \hat{\beta}_{MV}^2}{n\alpha}}} \stackrel{aprox}{\sim} N(0, 1)$$

De donde el intervalo de confianza esta dado por:

$$\mathbb{P} \left(f(\hat{\beta}_{MV}) - \sqrt{\frac{(f'(\hat{\beta}_{MV}))^2 \hat{\beta}_{MV}^2}{n\alpha}} Z_{1-\alpha/2} \leq f(\beta) \leq f(\hat{\beta}_{MV}) + \sqrt{\frac{(f'(\hat{\beta}_{MV}))^2 \hat{\beta}_{MV}^2}{n\alpha}} Z_{1-\alpha/2} \right) = 1-\alpha$$

Donde recordemos que:

$$f(\beta) := \mathbb{P}(X > x) = \int_x^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} dt$$

$$f(\hat{\beta}_{MV}) := \mathbb{P}(\widehat{X} > x) = \int_x^\infty \frac{\hat{\beta}_{MV}^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\hat{\beta}_{MV} t} dt$$

$$f'(\beta) := \frac{\partial}{\partial \beta} \mathbb{P}(\widehat{X} > x) = \frac{\partial}{\partial \beta} \int_x^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} dt$$

En la última igualdad podemos aplicar las condiciones de regularidad y llegar a que:

$$f'(\beta) = \int_x^\infty \frac{\partial}{\partial \beta} \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} dt$$

4.4. Intervalos via Simulación

La simulación de variables aleatorias permite hacer aproximaciones muy precisas del comportamiento distribucional de la población, por ejemplo imaginemos que sabemos que $X \sim F(\mu, \sigma^2)$ con F alguna función de distribución conocida. Supongamos que nos interesa saber el valor de $\mu = \mathbb{E}(X)$. Sabemos que para conocer dicho valor tenemos que llevar a cabo el cálculo de la esperanza lo cual en ocasiones nos lleva a problemas de integración difíciles. Supongamos que somos capaces de simular observaciones de la v.a. X , es decir, tenemos un algoritmo que no genera x_1, \dots, x_n observaciones de la v.a. X , entonces por la ley fuerte de los grandes numero sabemos que:

$$\mathbb{E}(X) \approx \bar{x}$$

De hecho sabemos que si n tiende a infinito entonces $\bar{x} \rightarrow \mu = \mathbb{E}(X)$, por lo tanto una forma de aproximar el valor desconocido μ es simular **muchas** observaciones de la v.a. X y luego asumir que $\mathbb{E}(X) \approx \bar{x}$.

Otra de las grandes ventajas de poder simular de una v.a. es que nos permite aproximar el comportamiento de transformaciones de variables aleatorias, imaginemos que nos interesa conocer el comportamiento distribucional de la v.a. $Y = \sin(U)$ donde $U \sim \text{Uniforme}(0, 2\pi)$.

Una forma de resolver el problema es realizar el cambio de variable correspondiente y encontrar $f_Y(y)$, sin embargo, via simulación podemos aproximar las cantidades de inters de Y , para ellos supongamos que podemos simular observaciones de la v.a. $U(0, 2\pi)$ de tal forma que tenemos u_1, \dots, u_n muestra observada de la v.a. U , para cada muestra observada procedemos a transformar $y_1 = \sin u_1, \dots, y_n = \sin(u_n)$, entonces se prueba que la muestra y_1, \dots, y_n son observaciones (simulaciones) de la v.a. Y . Entonces si estamos interesados por ejemplo en encontrar a $\mathbb{E}(Y) = \mathbb{E}(\sin(U))$, entonces bastaría con simular una gran cantidad de observaciones la v.a. U luego transformarla para obtener y_1, \dots, y_n finalmente sabemos que una buena aproximación para $\mathbb{E}(Y)$ es:

$$\mathbb{E}(Y) = \mathbb{E}(\sin(U)) \approx \bar{y}$$

En inferencia estadística la simulación juega un papel importante pues nos permite aproximar la distribución del estimador $\hat{\theta}$, como vimos, el hecho de conocer la distribución del estimador nos permite construir cantidades pivotaes. Para llevar a cabo la aproximación por medio de simulaciones veremos dos casos, el Bootstrap Paramétrico y el no paramétrico.

4.4.1. Bootstrap Paramétrico

4.4.2. Bootstrap no Paramétrico

Ejercicio 4.4.1. *Resuelva lo siguiente:*

1. Sea X_1, \dots, X_n m.a. del modelo $Uniforme(0, \theta)$ con $\theta > 0$.

$$f_X(x) := \frac{1}{\theta} \mathbf{1}_{(0, \theta)}(x)$$

El objetivo del ejercicio es encontrar un intervalo de confianza para θ

- a) Defina la variable aleatoria T como:

$$T = \frac{X_{(n)}}{\theta} = \frac{\max\{X_1, \dots, X_n\}}{\theta}$$

y muestre que la función de densidad de T está dada por:

$$f_T(t) = nt^{n-1} \mathbf{1}_{(0,1)}(t)$$

Hint: Recuerde que la densidad del máximo estadístico de orden en esta familia está da-

da por:

$$f_{X_{(n)}}(x) = \frac{nx^{n-1}}{\theta^n} \mathbf{1}_{(0,\theta)}(x)$$

Por lo tanto:

$$f_T(t) = \frac{\partial}{\partial t} F_T(t) = \frac{\partial}{\partial t} \mathbb{P}(T \leq t) = \frac{\partial}{\partial t} \mathbb{P}\left(\frac{X_{(n)}}{\theta} \leq t\right) = \frac{\partial}{\partial t} \mathbb{P}(X_{(n)} \leq t\theta) = \frac{\partial}{\partial t} \int_0^{t\theta} f_{X_{(n)}}(x) dx$$

b) Usando el ejercicio anterior sabemos que T es una cantidad pivotal, utilice esta cantidad pivotal para encontrar una intervalo al 95 % de confianza para θ

2. Sea X_1, \dots, X_n m.a. del modelo $\text{Exp}(\lambda)$ con $\lambda > 0$.

$$f_X(x) := \lambda e^{-\lambda x} \mathbf{1}_{(0,\infty)}(x)$$

a) Defina la variable aleatoria T como:

$$T = \lambda \sum_{i=1}^n X_i$$

Usando la función generadora de momentos prueba que $T \sim \text{Gamma}(n, 1)$

b) Usando el ejercicio anterior sabemos que T es una cantidad pivotal, utilice esta cantidad pivotal para encontrar una intervalo al $(1 - \alpha) \%$ de confianza para λ

3. Considera X_1, \dots, X_{n_x} m.a. de $N(\mu_x, \sigma_{n_x}^2)$ y Y_1, \dots, Y_{n_y} m.a. de $N(\mu_y, \sigma_{n_y}^2)$, suponga μ_x y μ_y desconocidas. Muestre que:

$$\left(\frac{S_x^2}{S_y^2} F_{n_y-1, n_x-1}^{\alpha/2}, \frac{S_x^2}{S_y^2} F_{n_y-1, n_x-1}^{1-\alpha/2} \right)$$

Es un intervalo al $(1 - \alpha) \%$ de confianza para el cociente $\frac{\sigma_x^2}{\sigma_y^2}$. Donde:

$$S_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (x_i - \bar{x})^2; \quad S_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (y_i - \bar{y})^2;$$

$F_{n_y-1, n_x-1}^{1-\alpha/2}$ y $F_{n_y-1, n_x-1}^{\alpha/2}$ son los cuantiles de una distribución F con $(n_y - 1, n_x - 1)$ grados de libertad

4. Sea X_1, \dots, X_n m.a. del modelo $\text{Exp}(\lambda)$ con $\lambda > 0$.

$$f_X(x) := \lambda e^{-\lambda x} \mathbf{1}_{(0, \infty)}(x)$$

- Encuentra un intervalo de confianza aproximado al $(1 - \alpha)\%$ de confianza para λ usando el hecho de que $\hat{\lambda}_{MV} \sim^* N\left(\lambda, \frac{1}{nI(\lambda)}\right)$, es decir, utilice la cantidad pivotal dada por:

$$\frac{\hat{\lambda}_{MV} - \lambda}{\sqrt{\frac{1}{nI(\lambda)}}} \sim^* N(0, 1)$$

- Encuentra un intervalo de confianza aproximado al $(1 - \alpha)\%$ de confianza para λ usando una segunda aproximación suponiendo ahora que $\hat{\lambda}_{MV} \sim^{**} N\left(\lambda, \frac{1}{nI(\hat{\lambda}_{MV})}\right)$, es decir, utilice la cantidad pivotal dada por:

$$\frac{\hat{\lambda}_{MV} - \lambda}{\sqrt{\frac{1}{nI(\hat{\lambda}_{MV})}}} \sim^{**} N(0, 1)$$

5. Se quiere verificar que la estatura promedio de las mujeres es más pequeña que la estatura promedio de los hombres. Suponga que la distribución de la estatura de tanto mujeres como hombres es modelada por una distribución Normal con misma varianza conocida para ambas poblaciones. Es decir

$$X \sim N(\mu_x, \sigma^2) \quad Y \sim N(\mu_y, \sigma^2)$$

Donde X modela la estatura de las mujeres y Y la estatura de los hombres.

Se llevo a cabo un muestreo aleatorio en las dos poblaciones de forma independiente, el tamaño de muestra en ambos muestreos fue de 100 y se obtuvieron los siguientes resultados:

$$\bar{x} = \frac{\sum_{i=1}^{100} x_i}{100} = 169\text{cm} \quad \bar{y} = \frac{\sum_{i=1}^{100} y_i}{100} = 175\text{cm}$$

Si la varianza de ambas poblaciones es $\sigma^2 = 25$. Construya un intervalo al 95 % de confianza para la diferencia de medias $\mu_x - \mu_y$. Concluya si en efecto tenemos evidencia como para decir que la estatura promedio de las mujeres es más pequeña que la estatura promedio de los hombres.

6. Se desea llevar a cabo un muestreo para estimar el peso promedio en kilogramos de los hombres mayores de 18 años en México, suponiendo que el peso de esta población sigue

una distribución $N(\mu, \sigma^2)$ con σ^2 conocida e igual a 36 ($\sigma^2 = 36$). Calcule el tamaño de muestra necesario para estimar la media μ con un error de $d = 0.5\text{kg}$ y una confianza de 95 %. ($\alpha = 0.05$).

7. Suponga que una aseguradora modela el número de siniestros semanales usando un modelo Poisson de parámetro desconocido λ . La aseguradora pretende encontrar un intervalo al 95 % de confianza para la probabilidad de observar 0 siniestros en una semana:

$$\mathbb{P}(X = 0) = e^{-\lambda}$$

Supongamos que la aseguradora observa la siguiente muestra:

$$(3, 6, 6, 2, 4, 6, 1, 5, 6, 6, 6, 7, 5, 5, 9, 2, 8, 6, 5, 10, 7, 4, 2, 5, 4, 5, 1, 4, 2, 2)$$

Encuentre el valor estimado de $\mathbb{P}(X = 0) = e^{-\lambda}$ y construya el intervalo de confianza (aproximado)

8. Sea X_1, \dots, X_n m.a. del modelo:

$$f_X(x; \lambda) := \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \mathbf{1}_{(0, \infty)}(x)$$

- Encuentra un intervalo de confianza aproximado al $(1 - \alpha)\%$ de confianza para λ usando el hecho de que $\hat{\lambda}_{MV} \sim^* N\left(\lambda, \frac{1}{nI(\lambda)}\right)$, es decir, utilice la cantidad pivotal dada por:

$$\frac{\hat{\lambda}_{MV} - \lambda}{\sqrt{\frac{1}{nI(\lambda)}}} \sim^* N(0, 1)$$

- Encuentra un intervalo de confianza **aproximado** al $(1 - \alpha)\%$ de confianza para $P(X > 1)$.

Hint: Recuerde que por propiedades asintóticas de los estimadores máximo verosímiles se tiene que:

$$f(\lambda_{MV}) \sim^* N\left(f(\lambda), \frac{(f'(\lambda))^2}{nI(\lambda)}\right)$$

Con f una transformación diferenciable.

Capítulo 5

Pruebas de hipótesis

5.1. Preliminares

En esta sección estudiaremos una tercer forma de llevar a cabo inferencia sobre parámetros desconocidos de una población. Comencemos por definir que es una hipótesis.

Definición 5.1.1. *Una hipótesis es una aseveración o afirmación sobre el comportamiento distribucional de ciertas poblaciones de interés*

A forma de ejemplo supongamos que la v.a. X modela el tiempo de vida del componente 1 y Y modela el tiempo de vida de otro componente 2. Por definición sabemos que existe una función de distribución asociada a las v.a. aleatorias definidas, por lo que quizás una hipótesis inicial que podríamos tener es afirmar que X sigue una distribución Gamma con ciertos parámetros desconocidos. En términos de notación esto lo pondremos de la siguiente forma:

$$H_0 : X \sim \text{Gamma}(\alpha, \beta)$$

Ahora supongamos que ya tenemos cierto conocimiento del fenómeno y que conocemos que el modelo *Gamma* ajusta bien pero se desconoce los parámetros, entonces algunas hipótesis pueden ser:

$$H_0 : \alpha = 1$$

$$H_0 : \alpha > 1$$

$$H_0 : (\alpha, \beta) = (1, 2)$$

A veces también estaremos interesados en probar hipótesis sobre dos poblaciones, por ejemplo imaginando que $X \sim \text{Gamma}(\alpha_x, \beta_x)$ y $Y \sim \text{Gamma}(\alpha_y, \beta_y)$, una prueba de hipótesis muy interesante sería ver si el tiempo de vida de los dos componentes es el mismo en promedio, esto lo podríamos plantear de la siguiente forma:

$$H_0 : \mu_x = \mu_y \quad \Rightarrow \quad H_0 : \frac{\alpha_x}{\beta_x} = \frac{\alpha_y}{\beta_y}$$

La idea de una prueba de hipótesis es extraer una muestra del fenómeno y luego con esta información verificar la plausibilidad de la Hipótesis en cuyo caso a la **luz** de los datos tomar una decisión sobre si rechazo o no rechazo mi hipótesis.

En este curso nos enfocaremos a pruebas sobre los parámetros suponiendo entonces que el fenómeno estudiado ya tiene un modelo específico y solo se desconoce los parámetros.

Establecer una hipótesis sobre θ supone dividir el espacio parameral Θ en 2 subconjuntos Θ_0 y Θ_1 tal que:

$$\Theta = \Theta_0 \cup \Theta_1 \quad \Theta_0 \cap \Theta_1 = \emptyset$$

Donde:

$$\Theta_0 := \{\theta \in \Theta | H_0 \text{ es cierta}\} \quad \Theta_1 := \{\theta \in \Theta | H_0 \text{ no es cierta}\}$$

Así definido queda claro que entonces Θ_0, Θ_1 forman una partición del espacio parametral Θ . A la hipótesis que queremos contrastar la llamaremos hipótesis nula y la denotamos por H_0 , por otro lado llamaremos la hipótesis alternativa H_1 a la formada por los valores de Θ tales que están en Θ_1 .

Ejemplo 5.1.1. Suponga que $X \sim \text{Bernulli}(\theta)$ y desconocemos completamente a Θ , en ese caso queda claro que $\Theta := \{(0, 1)\}$. Si planteamos la siguiente hipótesis $H_0 : \theta \leq 0.3$, entonces quedaría claro que $\Theta_0 = \{(0, 0.3]\}$ y por completo entonces $\Theta_1 = \{(0.3, 1)\}$ por lo que $H_1 : \theta > 0.3$. Por fines prácticos en la literatura decimos que cuando planteamos la hipótesis $H_0 : \theta \leq 0.3$ entonces estamos contrastando las hipótesis H_0 y H_1 y lo escribimos así:

$$H_0 : \theta \leq 0.3 \quad vs \quad H_1 : \theta > 0.3$$

Imaginemos que tenemos cierta información adicional y sabemos que θ solo puede tomar dos posibles valores θ_0 y θ_1 . En este caso $\Theta := \{\theta_0, \theta_1\}$. Bajo este contexto tenemos entonces que

tomar una decisión sobre cuál parámetro es el correcto. Planteamos entonces la prueba:

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1$$

Notemos que en este caso, la prueba anterior se reduce a poner a competir dos modelos:

$$H_0 : X \sim \text{Bernulli}(\theta_0) \quad vs \quad H_1 : X \sim \text{Bernulli}(\theta_1)$$

Definición 5.1.2 (Hipótesis Simple). *Decimos que una hipótesis es simple si determina completamente a la función de distribución, en caso contrario decimos que la hipótesis es compuesta.*

Así por ejemplo en el caso anterior:

$$H_0 : X \sim \text{Bernulli}(\theta_0) \quad vs \quad H_1 : X \sim \text{Bernulli}(\theta_1)$$

Estamos probando dos hipótesis simples, sin embargo suponga que bajo el mismo contexto ahora se nos plantea que $\Theta = (0, 1)$ con la siguientes hipótesis:

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

Luego entonces, en este último caso estamos contrastando una hipótesis simple contra una compuesta y finalmente si se plantean la hipótesis:

$$H_0 : \theta \leq \theta_0 \quad vs \quad H_1 : \theta > \theta_0$$

se contrasta una hipótesis compuesta contra compuesta.

Una vez planteado tanto el espacio parametral como la hipótesis nula, la siguiente pregunta natural que surge en la teoría es saber como decidir si rechazamos o no rechazamos nuestra hipótesis. Resulta lógico que para tomar nuestra decisión tenemos que tener una muestra de la población que se esta estudiando.

Definición 5.1.3 (Espacio Muestral). *Sea X_1, X_2, \dots, X_n muestra aleatoria de un fenómeno aleatorio con función de densidad $f_X(x; \theta)$. El espacio muestral es un subconjunto de \mathbb{R}^n que contiene a todos los posibles valores de la muestra de tamaño n y lo denotamos como \mathfrak{X} .*

$$\mathfrak{X} = \{\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n | x_i \in \{\text{Soporte de } X_i\}\}$$

5.1.1. Caso: Simple vs Simple

Imaginemos que tenemos un fenómeno modelado por una densidad *Bernulli* (θ), supongamos que $\Theta = (0.1, 0.8)$ y nos plantemos la hipótesis $H_0 : \theta = 0.1$, entonces estaremos contrastando las hipótesis:

$$H_0 : \theta = 0.1 \quad vs \quad H_1 : \theta = 0.8$$

Supongamos ahora que observaremos una muestra aleatoria de tamaño 3, (X_1, X_2, X_3) . Listemos las posibles muestras que recibiremos del fenómeno.

$$\mathfrak{X} := \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0), (1, 1, 0), (0, 1, 1), (1, 0, 1), (1, 1, 1)\}$$

De estas muestras parece que el caso $(X_1 = 0, X_2 = 0, X_3 = 0)$ es mas compatible con la hipótesis nula es decir observar esa muestra nos debería llevar a no rechazar H_0 , mientras que el caso $(X_1 = 1, X_2 = 1, X_3 = 1)$ no es muy compatible con H_0 y por tanto observar esa muestra nos debe de llevar a rechazar nuestra hipótesis nula. Parece entonces que para crear una regla de decisión debemos llevar a cabo una partición de \mathfrak{X} . Supongamos entonces que participamos \mathfrak{X} en dos conjuntos denotados por:

$$\mathcal{C}^* = \{\underline{x} \in \mathfrak{X} | \text{Rechazo } H_0\} \quad y \quad \mathcal{C} = \{\underline{x} \in \mathfrak{X} | \text{No Rechazo } H_0\};$$

(Muchas veces a \mathcal{C}^* se le conoce como la región crítica)

En nuestro ejemplo supongamos que hacemos la siguiente partición:

$$\mathcal{C} = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0)\} \quad y \quad \mathcal{C}^* = \{(1, 1, 0), (1, 0, 1), (1, 1, 1)\};$$

Es decir:

$$\mathcal{C}^* = \left\{ \underline{x} \left| \sum_{i=1}^3 x_i > 1 \right. \right\} \quad y \quad \mathcal{C} = \left\{ \underline{x} \left| \sum_{i=1}^3 x_i \leq 1 \right. \right\}$$

Bajo esta partición tenemos la siguiente regla de decisión:

$$\text{No Rechazo } H_0 \text{ si } \sum_{i=1}^n x_i \leq 1 \text{ y Rechazar } H_0 \text{ si } \sum_{i=1}^n x_i > 1$$

Pero, ¿nuestra elección de \mathcal{C} es correcta?. Necesitamos desarrollar teoría que nos ayude a determinar si nuestra partición es correcta o no.

Cuando optamos por Rechazar H_0 existe una probabilidad de equivocarse en la decisión tomada o bien también existe una probabilidad de equivocarse por no rechazar H_0 . No parecería desca-

bellado entonces buscar aquellas regiones \mathcal{C} y \mathcal{C}^* que minimicen la probabilidad de equivocarse.

Decisión \ Realidad	H_0 cierta	H_1 cierta
No rechazo H_0	No hay error	Error tipo II
Rechazo H_0	Error tipo I	No hay error

Definición 5.1.4 (Tipos de Error). *Decimos que en una prueba de hipótesis cometemos el **error tipo I** si rechazamos H_0 cuando en realidad H_0 era cierta y decimos que cometemos el **error tipo II** si no rechazamos H_0 cuando en realidad era falsa.*

Supongamos que ya tenemos definido a \mathcal{C} y \mathcal{C}^* e imaginemos que aun no observamos la muestra, como la muestra no se ha observado entonces tiene sentido preguntarnos por las probabilidades del error tipo I y error tipo 2. Definamos α y β como sigue:

$$\begin{aligned}\alpha &:= \mathbb{P}(\text{Error tipo I}) = \mathbb{P}(\text{Rechazar } H_0 \mid H_0 \text{ Cierta}) = \mathbb{P}(\underline{X} \in \mathcal{C}^* \mid H_0 \text{ Cierta}) \\ \beta &:= \mathbb{P}(\text{Error tipo II}) = \mathbb{P}(\text{No Rechazar } H_0 \mid H_1 \text{ Cierta}) = \mathbb{P}(\underline{X} \in \mathcal{C} \mid H_1 \text{ Cierta})\end{aligned}$$

Nuestro objetivo es entonces encontrar \mathcal{C} y \mathcal{C}^* que minimicen α y β de manera simultánea.

Definición 5.1.5 (Tamaño de la prueba). *Sea \mathcal{C}^* una región de rechazo en una prueba de hipótesis (simple vs simple), decimos que la prueba con \mathcal{C}^* como su región de rechazo, tiene tamaño α si*

$$\mathbb{P}(\underline{X} \in \mathcal{C}^* \mid H_0 \text{ Cierta}) = \alpha$$

En el ejemplo que hemos venido revisando tenemos que:

$$H_0 : \theta = 0.1 \quad vs \quad H_1 : \theta = 0.8$$

$$\mathcal{C}^* = \left\{ \underline{x} \mid \sum_{i=1}^3 x_i > 1 \right\} \quad y \quad \mathcal{C} = \left\{ \underline{x} \mid \sum_{i=1}^3 x_i \leq 1 \right\}$$

entonces:

$$\begin{aligned}
 \alpha &= \mathbb{P}(\underline{X} \in \mathcal{C}^* \mid H_0 \text{ Cierta}) = \mathbb{P}\left(\sum_{i=1}^3 X_i > 1 \mid \theta = 0.1\right) \\
 &= \mathbb{P}\left(\sum_{i=1}^3 X_i > 1 \mid X_i \sim \text{Ber}(0.1)\right) = 0.028 \\
 \beta &= \mathbb{P}(\underline{X} \in \mathcal{C} \mid H_1 \text{ Cierta}) = \mathbb{P}\left(\sum_{i=1}^3 X_i \leq 1 \mid \theta = 0.8\right) \\
 &= \mathbb{P}\left(\sum_{i=1}^3 X_i \leq 1 \mid X_i \sim \text{Ber}(0.8)\right) = 0.104
 \end{aligned}$$

Definición 5.1.6 (Potencia de la prueba). *En una prueba de hipótesis simple vs simple, a la cantidad $1 - \beta$ se le conoce como la potencia de la prueba*

Por lo anterior la potencia de esta prueba (con \mathcal{C} y \mathcal{C}^* definidas) está dada por $1 - \beta = 0.896$. Ahora supongamos que modificamos la región crítica:

$$\mathcal{C}^* = \left\{ \underline{x} \mid \sum_{i=1}^3 x_i > 2 \right\} \quad y \quad \mathcal{C} = \left\{ \underline{x} \mid \sum_{i=1}^3 x_i \leq 2 \right\}$$

Bajo esta nueva configuración se puede verificar que ahora:

$$\begin{aligned}
 \alpha &= \mathbb{P}\left(\sum_{i=1}^3 X_i > 2 \mid \theta = 0.1\right) = 0.001 \\
 \beta &= \mathbb{P}\left(\sum_{i=1}^3 X_i \leq 2 \mid \theta = 0.8\right) = 0.488
 \end{aligned}$$

Observamos que con esta nueva región crítica hemos logrado reducir α sin embargo β aumentó considerablemente ocasionado una disminución en la potencia de la prueba.

En general se puede probar que minimizar α aumenta la probabilidad del error tipo 2 y viceversa luego entonces en la práctica se opta entonces por fijar un α y luego buscar la región crítica \mathcal{C}^* que minimice β (maximice la potencia) (Obs: **En el contraste siempre se controla el error tipo I fijando un α previo a observar la muestra**)

Ejemplo 5.1.2. Sea X_1 una m.a. del modelo $N(\theta, \sigma^2 = 1)$ y suponga que $\Theta = \{0, 3\}$. Suponga-

mos que se nos plantea la siguiente prueba simple vs simple:

$$H_0 : \theta = 0 \quad vs \quad H_1 : \theta = 3$$

Definamos la siguiente región crítica

$$\mathcal{C}^* = \{\underline{x} | x_1 > 2\} \quad y \quad \mathcal{C} = \{\underline{x} | x_1 \leq 2\}$$

Entonces:

$$\begin{aligned} \alpha &= \mathbb{P}(X_1 > 2 \mid \theta = 0) = \mathbb{P}(X_1 > 2 \mid X_1 \sim N(0, 1)) = 0.0227 \\ \beta &= \mathbb{P}(X_1 \leq 2 \mid \theta = 3) = \mathbb{P}(X_1 \leq 2 \mid X_1 \sim N(3, 1)) = 0.159 \\ 1 - \beta &= 0.841 \end{aligned}$$

Ahora definamos otra región crítica:

$$\mathcal{C}^* = \{\underline{x} | 1.5 < x_1 < 1.705\} \quad y \quad \mathcal{C} = \{\underline{x} | x_1 \leq 1.5 \text{ o } x_1 \geq 1.705\}$$

Entonces:

$$\begin{aligned} \alpha &= \mathbb{P}(1.5 < X_1 < 1.705 \mid \theta = 0) = \mathbb{P}(1.5 < X_1 < 1.705 \mid X_1 \sim N(0, 1)) = 0.0227 \\ \beta &= \mathbb{P}(X_1 \notin (1.5, 1.705) \mid \theta = 3) = \mathbb{P}(X_1 \notin (1.5, 1.705) \mid X_1 \sim N(3, 1)) = 0.969 \\ 1 - \beta &= 0.031 \end{aligned}$$

Notemos que ambas pruebas generan una probabilidad del error tipo 1 similar $\alpha = 0.0227$, sin embargo la segunda región tiene una potencia mucho menor que la primera, luego entonces entre las dos regiones aquí expuestas preferimos la primera por tener mayor potencia. El objetivo en nuestra teoría es entonces dada α encontrar la región con mayor potencia.

5.1.2. Teorema de Neyman Pearson

El Teorema de Neyman Pearson nos dice como encontrar la región de rechazo mas potente dada un α fijo y conocido en el caso de una prueba de hipótesis simple vs simple.

Teorema 5.1.1 (Teorema de Neyman Pearson). Sea $\underline{X} = (X_1, \dots, X_n)$ m.a. de $f_X(x; \theta)$ y $\alpha \in (0, 1)$. Supongamos que $\Theta = \{\theta_0, \theta_1\}$ y planteamos la prueba de hipótesis:

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1$$

Entonces la región de rechazo más potente \mathcal{C}^* tal que $\mathbb{P}(\underline{X} \in \mathcal{C}^* \mid H_0 \text{ Cierta}) = \alpha$ está dada por:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{f_X(x_1, \dots, x_n; \theta_0)}{f_X(x_1, \dots, x_n; \theta_1)} \leq k_\alpha \right. \right\}$$

Donde k_α es número real tal que $\mathbb{P}(\underline{X} \in \mathcal{C}^* \mid H_0 \text{ Cierta}) = \alpha$

Demostración. Sea \mathcal{C}_1^* cualquier otra región de rechazo tal que $\mathbb{P}(\underline{X} \in \mathcal{C}_1^* \mid H_0 \text{ Cierta}) = \alpha$
Entonces la potencia asociada a esta región está dada por:

$$\begin{aligned} \pi_1 = 1 - \beta_1 &= 1 - \mathbb{P}(\underline{X} \in \mathcal{C}_1^{*c} \mid H_1 \text{ Cierta}) = \mathbb{P}(\underline{X} \in \mathcal{C}_1^* \mid H_1 \text{ Cierta}) \\ &= \int \dots \int_{\mathcal{C}_1^*} f_X(\underline{x}; \theta_1) d\underline{x} \\ &= \int \dots \int_{\mathbb{R}^n} \mathbb{1}_{\mathcal{C}_1^*}(\underline{x}) f_X(\underline{x}; \theta_1) d\underline{x} \end{aligned}$$

Por otro lado, la potencia asociada a \mathcal{C}^* es:

$$\begin{aligned} \pi^* = 1 - \beta^* &= 1 - \mathbb{P}(\underline{X} \in \mathcal{C}^{*c} \mid H_1 \text{ Cierta}) = \mathbb{P}(\underline{X} \in \mathcal{C}^* \mid H_1 \text{ Cierta}) \\ &= \int \dots \int_{\mathcal{C}^*} f_X(\underline{x}; \theta_1) d\underline{x} \\ &= \int \dots \int_{\mathbb{R}^n} \mathbb{1}_{\mathcal{C}^*}(\underline{x}) f_X(\underline{x}; \theta_1) d\underline{x} \end{aligned}$$

Demostraremos que $\pi^* - \pi_1 \geq 0$ lo que indicaría que en efecto la región \mathcal{C}^* tiene una mayor potencia que la región \mathcal{C}_1^* .

$$\begin{aligned} \pi^* - \pi_1 &= \int \dots \int_{\mathbb{R}^n} \mathbb{1}_{\mathcal{C}^*}(\underline{x}) f_X(\underline{x}; \theta_1) d\underline{x} - \int \dots \int_{\mathbb{R}^n} \mathbb{1}_{\mathcal{C}_1^*}(\underline{x}) f_X(\underline{x}; \theta_1) d\underline{x} \\ &= \int \dots \int_{\mathbb{R}^n} (\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) f_X(\underline{x}; \theta_1) d\underline{x} \\ &= \int \dots \int_{\mathcal{C}^*} (\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) f_X(\underline{x}; \theta_1) d\underline{x} + \\ &\quad \int \dots \int_{\mathbb{R}^n \setminus \mathcal{C}^*} (\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) f_X(\underline{x}; \theta_1) d\underline{x} \end{aligned}$$

Concluimos entonces que:

$$\pi^* - \pi_1 = \int \dots \int_{\mathcal{C}^*} (\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) f_X(\underline{x}; \theta_1) d\underline{x} + \int \dots \int_{\mathbb{R}^n \setminus \mathcal{C}^*} (\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) f_X(\underline{x}; \theta_1) d\underline{x} \quad (5.1)$$

Ahora recordemos que $\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \left| \frac{f_X(\underline{x}; \theta_0)}{f_X(\underline{x}; \theta_1)} \right| \leq k_\alpha \right\}$ y por tanto para $\underline{x} \in \mathcal{C}^*$ se tiene que $f_X(\underline{x}; \theta_1) \geq \frac{1}{k_\alpha} f_X(\underline{x}; \theta_0)$ además de que $\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x}) \geq 0$, se sigue entonces que:

$$\int \dots \int_{\mathcal{C}^*} (\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) f_X(\underline{x}; \theta_1) d\underline{x} \geq \int \dots \int_{\mathcal{C}^*} (\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) \frac{1}{k_\alpha} f_X(\underline{x}; \theta_0) d\underline{x} \quad (5.2)$$

Por otro lado si $\underline{x} \in \mathbb{R}^n \setminus \mathcal{C}^*$ entonces $f_X(\underline{x}; \theta_1) \leq \frac{1}{k_\alpha} f_X(\underline{x}; \theta_0)$ sin embargo se tiene que la resta de indicadoras $\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x}) \leq 0$ entonces:

$$(\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) f_X(\underline{x}; \theta_1) \geq (\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) \frac{1}{k_\alpha} f_X(\underline{x}; \theta_0)$$

Por lo tanto:

$$\int \dots \int_{\mathbb{R}^n \setminus \mathcal{C}^*} (\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) f_X(\underline{x}; \theta_1) d\underline{x} \geq \int \dots \int_{\mathbb{R}^n \setminus \mathcal{C}^*} (\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) \frac{1}{k_\alpha} f_X(\underline{x}; \theta_0) d\underline{x} \quad (5.3)$$

Finalmente sustituyendo las ecuaciones (5.2) y (5.3) en (5.1) tenemos que:

$$\begin{aligned} \pi^* - \pi_1 &\geq \frac{1}{k_\alpha} \left(\int \dots \int_{\mathcal{C}^*} (\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) f_X(\underline{x}; \theta_0) d\underline{x} + \int \dots \int_{\mathbb{R}^n \setminus \mathcal{C}^*} (\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) f_X(\underline{x}; \theta_0) d\underline{x} \right) \\ &= \frac{1}{k_\alpha} \int \dots \int_{\mathbb{R}^n} (\mathbb{1}_{\mathcal{C}^*}(\underline{x}) - \mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) f_X(\underline{x}; \theta_0) d\underline{x} \\ &= \frac{1}{k_\alpha} \left(\int \dots \int_{\mathbb{R}^n} (\mathbb{1}_{\mathcal{C}^*}(\underline{x})) f_X(\underline{x}; \theta_0) d\underline{x} - \int \dots \int_{\mathbb{R}^n} (\mathbb{1}_{\mathcal{C}_1^*}(\underline{x})) f_X(\underline{x}; \theta_0) d\underline{x} \right) \\ &= \frac{1}{k_\alpha} \left(\int \dots \int_{\mathcal{C}^*} f_X(\underline{x}; \theta_0) d\underline{x} - \int \dots \int_{\mathcal{C}_1^*} f_X(\underline{x}; \theta_0) d\underline{x} \right) \\ &= \frac{1}{k_\alpha} (\mathbb{P}(\underline{X} \in \mathcal{C}^* \mid H_0 \text{ cierta}) - \mathbb{P}(\underline{X} \in \mathcal{C}_1^* \mid H_0 \text{ cierta})) = \frac{1}{k_\alpha} (0 - 0) = 0 \end{aligned}$$

Se concluye entonces que $\pi^* - \pi_1 \geq 0$ por lo tanto la región \mathcal{C}^* tiene mayor potencia □

Veamos un ejemplo de como utilizar este Teorema:

Ejemplo 5.1.3. Sea X_1, \dots, X_n m.a. de $\text{Exp}(\lambda)$, supongamos que $\Theta = \{1, 5\}$, y se plantea la hipótesis:

$$H_0 : \lambda = 1 \quad \text{vs} \quad H_0 : \lambda = 5$$

Se nos pide encontrar la mejor región de rechazo tal que $\alpha = 0.05$. El teorema de Neyman Pearson

nos indica que:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{f_X(x_1, \dots, x_n; \lambda = 1)}{f_X(x_1, \dots, x_n; \lambda = 5)} \leq k_\alpha \right. \right\}$$

Como:

$$\begin{aligned} f_X(x_1, \dots, x_n; \lambda = 1) &= \prod_{i=1}^n e^{-x_i} = e^{-\sum_{i=1}^n x_i} \\ f_X(x_1, \dots, x_n; \lambda = 5) &= \prod_{i=1}^n 5e^{-5x_i} = 5^n e^{-5\sum_{i=1}^n x_i} \end{aligned}$$

Al hacer el cociente obtenemos que:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{e^{-\sum_{i=1}^n x_i + 5\sum_{i=1}^n x_i}}{5^n} \leq k_\alpha \right. \right\}$$

Esta región es equivalente a:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \sum_{i=1}^n x_i \leq \frac{1}{4} \log(k_\alpha 5^n) \right. \right\}$$

Haciendo $k_\alpha^* = \frac{1}{4} \log(k_\alpha 5^n)$ obtenemos que la mejor región de rechazo de tamaño α está dada por:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \sum_{i=1}^n x_i \leq k_\alpha^* \right. \right\}$$

Supongamos que fijamos un $\alpha = 0.05$, entonces queremos que:

$$\mathbb{P}(\underline{X} \in \mathcal{C}^* \mid H_0 \text{ cierta}) = 0.05 \Rightarrow \mathbb{P}\left(\sum_{i=1}^n X_i \leq k_\alpha^* \mid H_0 \text{ cierta}\right) = 0.05$$

Lo que necesitamos entonces es encontrar k_α^* tal que se cumpla con lo siguiente:

$$\mathbb{P}\left(\sum_{i=1}^n X_i \leq k_\alpha^* \mid \lambda = 1\right) = 0.05$$

Como **bajo** $H_0 \quad \lambda = 1$ se sigue que $\sum_{i=1}^n X_i \sim \text{Gamma}(n, \lambda = 1)$, entonces el problema de encontrar k_α^* se reduce a encontrar un el cuantíl de la distribución Gamma de esos parámetros.

Sea γ_α el cuantíl α de la de la densidad Gamma $(n, 1)$, entonces sabemos que:

$$\mathbb{P}\left(\sum_{i=1}^n X_i \leq \gamma_\alpha \mid \lambda = 1\right) = 0.05$$

Por lo tanto queda claro que $k_\alpha^* = \gamma_\alpha$ por lo que finalmente la regla de decisión asociada a esta prueba es rechazar H_0 si

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \sum_{i=1}^n x_i \leq \gamma_\alpha \right\}$$

Luego entonces el Teorema de Neyman-Pearson asegura que esta es la mejor región de rechazo para la hipótesis planteada de tamaño α , es decir, es la región que tiene una probabilidad de error tipo I igual a α con mayor potencia. En este caso para calcular la potencia de la prueba una vez encontrado γ_α se calcula como:

$$1 - \beta = 1 - \mathbb{P}(\underline{X} \in \mathcal{C}^{*c} \mid H_1 \text{ cierta}) = \mathbb{P}(\underline{X} \in \mathcal{C}^* \mid H_1 \text{ cierta}) \quad (5.4)$$

$$= \mathbb{P}\left(\sum_{i=1}^n X_i \leq \gamma_\alpha \mid \lambda = 5\right) \quad (5.5)$$

Debe de observarse que la región de rechazo quedó en función de la estadística $\sum_{i=1}^n X_i$ a la cual se le conoce como **estadístico de prueba** y se espera tenga una distribución conocida bajo H_0 para poder determinar la región de rechazo de forma exacta.

En una prueba de hipótesis vista desde el punto de Neyman-Person, uno como decisor fija un α antes de observar la muestra y con ello se calcula la región con la que se rechazará o no la hipótesis, **Fisher** por otro lado atacó el problema de otra manera, para él no es necesario determinar un α antes de observar la muestra, para **Fisher** lo mas importante es observar primero la muestra y luego evaluar que tan **compatible** es esta muestra con la hipótesis. La forma en como se determina la compatibilidad de la muestra es evaluar que tan probable es observar esa muestra bajo la hipótesis nula basado en el estadístico obtenido por el Teorema de Neyman Pearson.

Definición 5.1.7 (P-Value). . En estadística clásica la forma en como medimos la compatibilidad de una muestra con la hipótesis nula es por medio del calculo de la probabilidad de observar la muestra bajo H_0 basados en algún estadístico de prueba, a dicho número lo denominamos **p-value**. Así pues, p-values chicos nos hablan de poca compatibilidad de la muestra con la hipótesis y por tanto se podría optar por rechazar la hipótesis, por otro lado, en caso de observar p-values altos, esto nos habla de que la muestra es compatible con nuestra hipótesis y por tanto no habría evidencia para rechazar.

En nuestro ejemplo anterior trabajamos con X_1, \dots, X_n m.a. de $Exp(\lambda)$, y nos planteamos la hipótesis :

$$H_0 : \lambda = 1 \quad vs \quad H_0 : \lambda = 5$$

El teorema de Neyman-Pearson nos llevó a concluir que la mejor región de rechazo de tamaño α es:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \sum_{i=1}^n x_i \leq \gamma_\alpha \right. \right\}$$

El estadístico de prueba asociado al contraste es $\sum_{i=1}^n X_i$ el cual bajo H_0 tiene una distribución completamente conocida $Gamma(n, \lambda = 1)$ Neyman-Person afirma que debemos rechazar H_0 si se observan valores pequeños de este estadístico, Por otro lado la forma en como lo ve Fisher, es primero observar la muestra y luego calcular la probabilidad bajo H_0 de que el estadístico tome un valor igual o más crítico que el de la estimación del estadístico con la muestra observada. Por ejemplo supongamos que observamos la muestra y tenemos que:

$$\sum_{i=1}^n x_i = 5$$

Entonces el **p-value** se podrá calcular como:

$$p - value = \mathbb{P} \left(\sum_{i=1}^n X_i \leq \sum_i x_i = 5 \mid \lambda = 1 \right)$$

Luego entonces debe de quedar claro que $p - values$ menores a α nos hacen rechazar, mientras que valores superiores a α nos hacen no rechazar la hipótesis nula.

Ejemplo 5.1.4. Sea X_1, \dots, X_{10} m.a del modelo Poisson de parámetro λ . Se plantea la hipótesis:

$$H_0 : \lambda = 1 \quad vs \quad H_0 : \lambda = 3$$

- Suponga que se fija $\alpha = 0.048$, encontrar la mejor región de rechazo
- Para la región de rechazo encontrada calcular su potencia
- Suponga que observa una muestra tal que $\sum_{i=1}^n x_i = 8$, calcule el p -value
- Concluya con la información anterior si rechaza o no H_0 .

Haciendo el cociente de verosimilitudes obtenemos que:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{f_X(x_1, \dots, x_n; \lambda = 1)}{f_X(x_1, \dots, x_n; \lambda = 3)} \leq k_\alpha \right. \right\}$$

Como:

$$\begin{aligned} f_X(x_1, \dots, x_{10}; \lambda = 1) &= \prod_{i=1}^{10} \frac{1^{x_i}}{x_i!} e^{-1} = e^{-10} \prod_{i=1}^{10} \frac{1}{x_i!} \\ f_X(x_1, \dots, x_{10}; \lambda = 3) &= \prod_{i=1}^{10} \frac{3^{x_i}}{x_i!} e^{-3} = e^{-30} \prod_{i=1}^{10} \frac{3^{x_i}}{x_i!} \end{aligned}$$

Al hacer el cociente obtenemos que:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{e^{-10} \prod_{i=1}^{10} \frac{1}{x_i!}}{e^{-30} \prod_{i=1}^{10} \frac{3^{x_i}}{x_i!}} \leq k_\alpha \right. \right\}$$

Esta región es equivalente a:

$$\begin{aligned} \mathcal{C}^* &= \left\{ \underline{x} \in \mathfrak{X} \left| 3^{-\sum_{i=1}^n x_i} \leq k_\alpha e^{20} \right. \right\} \\ \mathcal{C}^* &= \left\{ \underline{x} \in \mathfrak{X} \left| -\sum_{i=1}^n x_i \log(3) \leq \log(k_\alpha e^{20}) \right. \right\} \\ \mathcal{C}^* &= \left\{ \underline{x} \in \mathfrak{X} \left| \sum_{i=1}^n x_i \geq -\frac{\log(k_\alpha e^{20})}{\log(3)} \right. \right\} \end{aligned}$$

Finalmente, haciendo $k_\alpha^* = -\frac{\log(k_\alpha e^{20})}{\log(3)}$ llegamos que la región óptima de rechazo es:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \sum_{i=1}^{10} x_i \geq k_\alpha^* \right. \right\}$$

Como $\alpha = 0.048$, entonces buscamos k_α^* tal que se cumpla que:

$$\mathbb{P} \left(\sum_{i=1}^{10} X_i \geq k_\alpha^* \left| \lambda = 1 \right. \right) = \alpha = 0.048$$

Como bajo H_0 , $\lambda = 1$ se sigue que $\sum_{i=1}^{10} X_i \sim \text{Poisson}(10)$. Buscamos entonces un punto en el cual la distribución $\text{Poisson}(10)$ deje en su cola derecha 0.048 de probabilidad. Usando el

programa R sabemos que si $Y \sim \text{Poisson}(10)$ entonces

$$\mathbb{P}(Y \geq 16) = 0.048$$

Por lo que en nuestro caso $k_\alpha^* = 16$ y por tanto la región más potente de tamaño $\alpha = 0.048$ está dada por:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \sum_{i=1}^{10} x_i \geq 16 \right. \right\}$$

Ahora calculemos la potencia de esta región.

$$\begin{aligned} 1 - \beta &= 1 - \mathbb{P}(\underline{X} \in \mathcal{C}^{*c} \mid \lambda = 3) = \mathbb{P}(\underline{X} \in \mathcal{C}^* \mid \lambda = 3) \\ &= \mathbb{P}\left(\sum_{i=1}^n X_i \geq 16 \mid \lambda = 3\right) = 0.9980525 \end{aligned}$$

Suponiendo que observamos $\sum_{i=1}^n x_i = 8$ entonces calculamos el p -value

$$p\text{-value} = \mathbb{P}\left(\sum_{i=1}^n X_i \geq \sum_{i=1}^n x_i = 8 \mid \lambda = 1\right) = 0.7797794$$

Esto nos indica que hay mucha compatibilidad de la muestra observada con la hipótesis nula además como $\alpha = 0.048$ se concluye que no se debe de rechazar H_0 .

5.2. Generalización $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$

El objetivo es ahora poder generar regiones óptimas de rechazo para los distintos casos a los que nos podemos enfrentar al hacer inferencia sobre los parámetros de un modelo. Por ejemplo:

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta > \theta_0$$

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta < \theta_0$$

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

$$H_0 : \theta \leq \theta_0 \quad vs \quad H_1 : \theta > \theta_0$$

$$H_0 : \theta \geq \theta_0 \quad vs \quad H_1 : \theta < \theta_0$$

En general todo lo anterior pueden resumirse al caso:

$$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1$$

Donde $\Theta = \Theta_0 \cup \Theta_1$ (Una partición del espacio parametral de nuestro problema). En el caso simple vs simple sabemos que $\Theta_0 = \{\theta_0\}$ y $\Theta_1 = \{\theta_1\}$ y además basados en el Teorema de Neyman Pearson se conoce que la mejor región de Rechazo para la prueba es de la forma:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{f_X(x_1, \dots, x_n; \theta_0)}{f_X(x_1, \dots, x_n; \theta_1)} \leq k_\alpha \right. \right\}$$

Esta región es equivalente a:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{f_X(x_1, \dots, x_n; \theta_0)}{\max_{\theta \in \Theta} \{f_X(x_1, \dots, x_n; \theta)\}} \leq k_\alpha^* \right. \right\}$$

La forma en como vamos a generalizar esta idea de Neyman-Person es por medio de la generación de regiones via un cociente de verosimilitudes conocido en inglés como L.R.T (Likelihood Ratio Test), el cual hablaremos a continuación.

5.2.1. Region L.R.T Likelihood Ratio Test

Comencemos definiendo la región de rechazo via L.R.T.

Definición 5.2.1 (Región L.R.T (Cociente de verosimilitudes generalizado)). *Suponga que tenemos X_1, \dots, X_n m.a de $f_X(x; \theta)$. Imaginemos que planteamos la prueba de hipótesis*

$$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1$$

Entonces la región de rechazo formada por:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{\sup_{\theta \in \Theta_0} f_X(x_1, \dots, x_n; \theta)}{\sup_{\theta \in \Theta} \{f_X(x_1, \dots, x_n; \theta)\}} \leq k_\alpha^* \right. \right\}$$

Se le denomina la Región L.R.T

Para simplificar la expresión que aparece en la definición muchas veces se renombra a dicho cociente con la letra λ de tal manera que:

$$\lambda(x_1, \dots, x_n) := \frac{\sup_{\theta \in \Theta_0} f_X(x_1, \dots, x_n; \theta)}{\sup_{\theta \in \Theta} \{f_X(x_1, \dots, x_n; \theta)\}}$$

Obs: $0 \leq \lambda(x_1, \dots, x_n) \leq 1$

Notemos que en el denominador del cociente se esta maximizando a la verosimilitud sobre todo el espacio parámetral, si suponemos que tenemos un completo desconocimiento del parámetro θ entonces dicho supremo sabemos que se obtiene evaluando la verosimilitud en el estimador máximo verosímil, es decir:

$$\sup_{\theta \in \Theta} \{f_X(x_1, \dots, x_n; \theta)\} = f_X(x_1, \dots, x_n; \hat{\theta}_{M.V.})$$

Entonces:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{\sup_{\theta \in \Theta_0} f_X(x_1, \dots, x_n; \theta)}{f_X(x_1, \dots, x_n; \hat{\theta}_{M.V.})} \leq k_\alpha^* \right. \right\}$$

La idea de la región LRT en este caso es simple, debemos rechazar H_0 si la estimación máximo verosímil se aleja mucho de la región Θ_0 , es decir entre la estimación maximo verosímil se aleje mas de la región Θ_0 mas evidencia hay para rechazar la hipótesis nula $H_0 : \theta \in \Theta_0$. Ahora bien, ¿Como obtenemos k_α^* ?, la idea nuevamente será controlar el error tipo 1 y buscar una constante para k_α tal que se tenga una probabilidad del error tipo 1 de magnitud α . El problema al que nos enfrentamos ahora es que bajo H_0 no necesariamente se tiene completamente definida a la función de distribución, es por ello que debemos re-definir lo que entenderemos por **tamaño de la prueba**.

Definición 5.2.2 (Tamaño de la prueba). . Sea $\underline{X} = (X_1, \dots, X_n)$ m.a. de $f_X(x; \theta)$ y \mathcal{C}^* una región de rechazo para la prueba

$$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1$$

Definimos el tamaño de la prueba como:

$$\sup_{\theta \in \Theta_0} \mathbb{P}(\underline{X} \in \mathcal{C}^* | \theta) = \alpha$$

Observe que la la definición anterior hace que la maxima probabilidad de cometer el error tipo 1 sea precisamente α .

Otra forma en que comúnmente se define el tamaño de la prueba es mediante la denominada **función potencia** de la prueba, la cual definimos a continuación.

Definición 5.2.3 (Función Potencia). Sea $\underline{X} = (X_1, \dots, X_n)$ m.a. de $f_X(x; \theta)$ y \mathcal{C}^* una región de rechazo para la prueba

$$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1$$

Definimos la función potencia como:

$$\pi_{\mathcal{C}^*}(\theta) = \mathbb{P}(\underline{X} \in \mathcal{C}^* \mid \theta)$$

Entonces con estas dos definiciones, es claro que el tamaño de una prueba se calcula como:

$$\alpha = \sup_{\theta \in \Theta_0} \pi_{\mathcal{C}^*}(\theta)$$

Una observación importante es que dada \mathcal{C}^* buscamos una función potencia $\pi_{\mathcal{C}^*}(\theta)$ que se comporte de forma que:

$$\text{si } \theta \in \Theta_0 \quad \Rightarrow \quad \pi_{\mathcal{C}^*}(\theta) \approx 0$$

$$\text{si } \theta \in \Theta_1 \quad \Rightarrow \quad \pi_{\mathcal{C}^*}(\theta) \approx 1$$

Es decir, queremos una región de rechazo que haga que la probabilidad de rechazar H_0 cuando $\theta \in \Theta_0$ sea pequeña (para no rechazar equivocadamente) y que la probabilidad de rechazar H_0 cuando $\theta \in \Theta_1$ sea alta (para rechazar acertadamente). Parece entonces que entre dos regiones de rechazo \mathcal{C}_1^* , \mathcal{C}_2^* de tamaño α escogeremos aquella que genere una mejor función potencia, es decir parecida a la función potencia ideal.

Ejemplo 5.2.1. Supongamos que tenemos X_1 m.a. de tamaño 1 del modelo $N(\mu, \sigma_0^2 = 1)$ y planteamos la prueba de hipótesis:

$$H_0 : \mu \leq 0 \quad \text{vs} \quad H_1 : \mu > 0$$

Definamos la siguientes regiones de rechazo:

$$\mathcal{C}_1^* = \{x \in \mathfrak{X} \mid x > 2\}$$

$$\mathcal{C}_2^* = \{x \in \mathfrak{X} \mid 1.5 < x < 1.705\}$$

Entonces si graficamos ambas funciones potencia observaríamos que la mejor región de rechazo está dada por \mathcal{C}_1^* (Ver figura 5.1). Además se puede verificar que ambas regiones tienen el mismo tamaño $\alpha = 0.0227$

En resumen, sabemos que nuestra mejor **apuesta** para encontrar una región óptima de rechazo es llevar a cabo el cociente L.R.T. es decir, definir nuestra la región critica como:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sup_{\theta \in \Theta_0} f_X(x_1, \dots, x_n; \theta)}{\sup_{\theta \in \Theta} \{f_X(x_1, \dots, x_n; \theta)\}} \leq k_\alpha^* \right\}$$

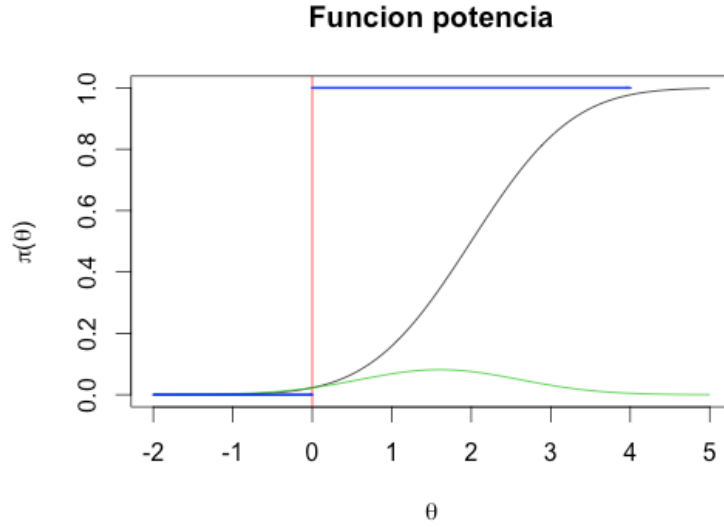


Figura 5.1: Función potencia

Donde k_α es una constante que asegure que el tamaño de la prueba sea α , es decir, el objetivo es encontrar k_α que haga que:

$$\sup_{\theta \in \Theta_0} \mathbb{P}(\underline{X} \in \mathcal{C}^* \mid \theta) = \sup_{\theta \in \Theta_0} \mathbb{P}\left(\frac{\sup_{\theta \in \Theta_0} f_X(X_1, \dots, X_n; \theta)}{\sup_{\theta \in \Theta} \{f_X(x_1, \dots, x_n; \theta)\}} \leq k_\alpha \mid \theta\right) = \alpha$$

$$\sup_{\theta \in \Theta_0} \mathbb{P}(\lambda(X_1, \dots, X_n) \leq k_\alpha \mid \theta) = \alpha$$

Como era de esperarse, los estadísticos de prueba que arroja este tipo de cocientes tienen el inconveniente de que son muy complejos y por tanto llegar a conocer su distribución es muy difícil, afortunadamente existe un resultado (Wilks - 1938) que aproxima la distribución de este cociente.

Teorema 5.2.1 (Teorema de Wilks - 1938). Sea X_1, \dots, X_n m.a. de $f_X(x; \underline{\theta})$, con $\underline{\theta} \in \mathbb{R}^k$. Suponga que planteamos la prueba de hipótesis:

$$H_0 : \underline{\theta} \in \Theta_0 \quad vs \quad H_a : \underline{\theta} \in \Theta_1$$

Con $\Theta = \Theta_0 \cup \Theta_1$ una partición del espacio parametral, sea p el número de parámetros libres bajo Θ_0 entonces:

$$\Lambda(\underline{X}) = -2 \log(\lambda(\underline{X})) \sim_{approx} \chi^2_{(k-p)}$$

Como la teoría nos dice que debemos rechazar H_0 si $\lambda(\underline{x}) \leq k^*$, esto equivale a rechazar H_0

cuando:

$$-2 \log (\lambda (\underline{x})) \geq k^{**}$$

Luego entonces por la propiedad distribucional asintótica de $\Lambda (\underline{X})$, sabemos que debemos rechazar H_0 a un nivel de significancia α si:

$$-2 \log (\lambda (\underline{x})) \geq \chi_{k-p}^{2(1-\alpha)}$$

Donde $\chi_{k-p}^{2(1-\alpha)}$ es el cuantíl $1 - \alpha$ de la distribución χ^2 con $k - p$ grados de libertad

Veamos algunos ejemplo de como obtener la L.R.T en el caso de una prueba de hipótesis para la distribución Normal.

Ejemplo 5.2.2 (L.R.T para μ , σ^2 conocida). Sea X_1, \dots, X_n m.a. de una población $N(\mu, \sigma_0^2)$ con σ_0^2 conocida. Encuentre la región L.R.T para la prueba:

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$$

En este caso $\Theta = \mathbb{R}$ y por tanto $\Theta_0 = \{\mu_0\}$ y $\Theta_1 = \mathbb{R} \setminus \{\mu_0\}$.

Sabemos que:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \lambda (\underline{x}) = \frac{\sup_{\mu \in \Theta_0} f_X (x_1, \dots, x_n; \mu)}{f_X (x_1, \dots, x_n; \hat{\mu}_{M.V.})} \leq k \right. \right\}$$

En este caso:

$$\sup_{\mu \in \Theta_0} f_X (x_1, \dots, x_n; \mu) = f_X (x_1, \dots, x_n; \mu_0) = (2\pi\sigma_0^2)^{-\frac{n}{2}} \exp \left(-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \mu_0)^2 \right) \right)$$

$$f_X (x_1, \dots, x_n; \hat{\mu}_{M.V.}) = (2\pi\sigma_0^2)^{-\frac{n}{2}} \exp \left(-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right) \right)$$

Por lo tanto la región de rechazo queda como:

$$\lambda (\underline{x}) = \frac{\exp \left(-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \mu_0)^2 \right) \right)}{\exp \left(-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right) \right)} \leq k$$

Despejando y tras unos pasos algebraicos

$$\frac{\exp \left(-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \mu_0)^2 \right) \right)}{\exp \left(-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right) \right)} \leq k \Leftrightarrow \exp \left(\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (x_i - \mu_0)^2 \right) \right) \leq k$$

$$\begin{aligned}
& \left(\sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (x_i - \mu_0)^2 \right) \leq 2\sigma_0^2 \log(k) \\
& \sum_{i=1}^n x_i^2 - n\bar{x}^2 - \sum_{i=1}^n x_i^2 + 2\mu_0 \sum_{i=1}^n x_i - n\mu_0^2 \leq 2\sigma_0^2 \log(k) \\
& -n\bar{x}^2 + 2\mu_0 \sum_{i=1}^n x_i \leq 2\sigma_0^2 \log(k) + n\mu_0^2 \Leftrightarrow -n\bar{x}^2 + 2n\mu_0\bar{x} \leq 2\sigma_0^2 \log(k) + n\mu_0^2 \\
& \bar{x}^2 - 2\mu_0\bar{x} \geq -\frac{1}{n} (2\sigma_0^2 \log(k) + n\mu_0^2) \Leftrightarrow \bar{x}^2 + 2\mu_0\bar{x} + \mu_0^2 \geq -\frac{1}{n} (2\sigma_0^2 \log(k) + n\mu_0^2) + \mu_0^2 \\
& (\bar{x} - \mu_0)^2 \geq -\frac{1}{n} (2\sigma_0^2 \log(k)) \Leftrightarrow |\bar{x} - \mu_0| \geq k^*
\end{aligned}$$

Finalmente la región L.T.R. queda de la siguiente forma:

$$\mathcal{C}_{LRT}^* = \{\underline{x} \in \mathfrak{X} \mid |\bar{x} - \mu_0| \geq k^*\}$$

Observe como la región de rechazo llega a algo intuitivamente obvio, debemos rechazar H_0 cuando la estimación máximo verosímil se aleje demasiado del valor μ_0 (El valor bajo H_0). El problema es ahora encontrar k^* tal que el error tipo 1 sea igual a α sin embargo para encontrar dicho valor necesitamos conocer exactamente la distribución de $|\bar{X} - \mu_0|$ bajo H_0 . Afortunadamente para el caso Normal sabemos que si H_0 es cierta se tiene que :

$$\bar{X} \sim N\left(\mu_0, \frac{\sigma_0^2}{n}\right) \Rightarrow \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma_0^2}{n}}} \sim N(0, 1)$$

Entonces podemos sacar provecho de esta característica distribucional redefiniendo nuestra región de rechazo como:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \left| \sqrt{n} \left(\frac{\bar{x} - \mu_0}{\sqrt{\sigma_0^2}} \right) \right| \geq \sqrt{n} \frac{k^*}{\sqrt{\sigma_0^2}} = k^{**} \right\}$$

Bajo esta nueva definición tenemos que encontrar el valor de k^{**} se simplifica pues:

$$\begin{aligned}
\alpha &= \mathbb{P}(\underline{X} \in \mathcal{C}_{LRT}^* \mid H_0 \text{ cierta}) = \mathbb{P}\left(\left| \sqrt{n} \left(\frac{\bar{X} - \mu_0}{\sqrt{\sigma_0^2}} \right) \right| \geq k^{**} \mid \mu = \mu_0 \right) \\
&= 1 - \mathbb{P}\left(-k^{**} \leq \sqrt{n} \left(\frac{\bar{X} - \mu_0}{\sqrt{\sigma_0^2}} \right) \leq k^{**} \mid \mu = \mu_0 \right)
\end{aligned}$$

Entonces:

$$\mathbb{P}\left(-k^{**} \leq \sqrt{n} \left(\frac{\bar{X} - \mu_0}{\sqrt{\sigma_0^2}} \right) \leq k^{**} \mid \mu = \mu_0\right) = 1 - \alpha$$

De donde concluimos que $k^{**} = Z_{(1-\alpha/2)}$, donde $Z_{(1-\alpha/2)}$ es el cuantil $1 - \alpha/2$ de la distribución $N(0, 1)$. Por lo tanto concluimos que la región de Rechazo L.R.T de tamaño α esta dada por:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \left| \sqrt{n} \left(\frac{\bar{x} - \mu_0}{\sqrt{\sigma_0^2}} \right) \right| \geq Z_{(1-\alpha/2)} \right\}$$

Notemos entonces que nuestro estadístico de prueba en este caso es:

$$T = \left| \sqrt{n} \left(\frac{\bar{X} - \mu_0}{\sqrt{\sigma_0^2}} \right) \right|$$

Ahora imaginemos que $\sigma_0^2 = 1$ y que estamos probando

$$H_0 : \mu = 1 \quad vs \quad H_1 : \mu \neq 1$$

y que recibimos una muestra x_1, \dots, x_{10} tal que $\sum_{i=1}^{10} x_i = 20$, entonces, el valor del estadístico evaluado en la muestra es igual a

$$t = \left| \sqrt{10} (2 - 1) \right| = 3.1622$$

Suponiendo $\alpha = 0.05$, sabemos que la región de rechazo queda de la forma:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \left| \sqrt{n} \left(\frac{\bar{x} - 1}{\sqrt{1}} \right) \right| \geq 1.96 \right\}$$

como $3.1622 > 1.96$ entonces concluimos que podemos rechazar H_0 , tenemos suficiente evidencia estadística como para rechazar H_0 y la probabilidad de equivocarse esta controlada y es menor a 0.05. Ahora, desde el punto de vista de Fisher, el p -value se encontraría de la siguiente manera:

$$\mathbb{P}\left(\left| \sqrt{n} \left(\frac{\bar{X} - 1}{\sqrt{1}} \right) \right| \geq 3.1622 \mid \mu = 1 \right) = 0.0007829099$$

De esta manera dado que $\alpha = 0.05 > 0.0007829099 = p\text{-value}$, se rechaza H_0 pues desde punto de vista de Fisher, la muestra recibida es incompatible con la hipótesis nula pues la probabilidad de observar una muestra así bajo el supuesto de que $\mu = 1$ es muy baja ($p\text{-value} = 0.0007829099$)

Ejercicio 5.2.1. Encuentre la región de rechazo L.R.T para una muestra X_1, \dots, X_n de $N(\mu, \sigma^2)$

para la prueba de hipótesis:

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$$

Con σ^2 desconocida. Verifique que dicha región de rechazo es equivalente a

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \left| \sqrt{n} \left(\frac{\bar{x} - \mu_0}{\sqrt{\hat{\sigma}_{M.V.}^2}} \right) \right| \geq K \right. \right\}$$

Luego como bajo el supuesto de normalidad se tiene que el estadístico:

$$T = \left| \sqrt{n} \left(\frac{\bar{x} - \mu_0}{\sqrt{\hat{\sigma}_{M.V.}^2}} \right) \right| \sim t - student_{(n-1)}$$

Entonces la región de Rechazo es:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \left| \sqrt{n} \left(\frac{\bar{x} - \mu_0}{\sqrt{\hat{\sigma}_{M.V.}^2}} \right) \right| \geq t_{n-1}^{1-\alpha/2} \right. \right\}$$

Donde $t_{(n-1)}^{1-\alpha/2}$, es el cuantil $1 - \alpha/2$ de una distribución $t - student$ con $n - 1$ grados de libertad.

Ejemplo 5.2.3 (L.R.T para σ^2 , μ conocida). Suponga que ahora en el caso normal nos interesa a hacer inferencia sobre la varianza σ^2 y se nos plantea la hipótesis.

$$H_0 : \sigma^2 = \sigma_0^2 \quad vs \quad H_1 : \sigma^2 \neq \sigma_0^2$$

Supondremos además que $\mu = \mu_0$ conocida. Se nos pide encontrar la región L.R.T. para esta prueba.

En este caso $\Theta = \mathbb{R}^+$ y por tanto $\Theta_0 = \{\sigma_0^2\}$ y $\Theta_1 = \mathbb{R}^+ \setminus \{\sigma_0^2\}$.

Sabemos que:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \lambda(\underline{x}) = \frac{\sup_{\mu \in \Theta_0} f_X(x_1, \dots, x_n; \sigma^2)}{f_X(x_1, \dots, x_n; \hat{\sigma}_{M.V.}^2)} \leq k \right. \right\}$$

En este caso:

$$\sup_{\sigma^2 \in \Theta_0} f_X(x_1, \dots, x_n; \sigma^2) = f_X(x_1, \dots, x_n; \sigma_0^2) = (2\pi\sigma_0^2)^{-\frac{n}{2}} \exp \left(-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \mu_0)^2 \right) \right)$$

$$f_X(x_1, \dots, x_n; \hat{\sigma}_{M.V.}^2) = (2\pi\hat{\sigma}_{M.V.}^2)^{-\frac{n}{2}} \exp \left(-\frac{1}{2\hat{\sigma}_{M.V.}^2} \left(\sum_{i=1}^n (x_i - \mu_0)^2 \right) \right)$$

Por lo tanto la región de rechazo queda como:

$$\lambda(\underline{x}) = \frac{(2\pi\sigma_0^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \mu_0)^2\right)\right)}{(2\pi\hat{\sigma}_{M.V.}^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\hat{\sigma}_{M.V.}^2} \left(\sum_{i=1}^n (x_i - \mu_0)^2\right)\right)} \leq k$$

Recordemos ahora que en este contexto:

$$\hat{\sigma}_{M.V.}^2 := \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2$$

Por lo tanto:

$$-\frac{1}{2\hat{\sigma}_{M.V.}^2} \left(\sum_{i=1}^n (x_i - \mu_0)^2\right) = -\frac{n}{2} \quad y \quad -\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \mu_0)^2\right) = -\frac{n}{2} \frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}$$

Entonces:

$$\lambda(\underline{x}) = \frac{(2\pi\sigma_0^2)^{-\frac{n}{2}} \exp\left(-\frac{n}{2} \frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right)}{(2\pi\hat{\sigma}_{M.V.}^2)^{-\frac{n}{2}} \exp\left(-\frac{n}{2}\right)} \leq k$$

Lo cual es equivalente a:

$$\left(\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right)^{\frac{n}{2}} \leq \exp\left(-\frac{n}{2}\right) k$$

Elevando a la $-\frac{n}{2}$ en ambos lados:

$$\left(\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right) \exp\left(-\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right) \leq \left(\exp\left(-\frac{n}{2}\right) k\right)^{-\frac{n}{2}} = k^*$$

Por lo tanto la la región de rechazo queda como:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \left(\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right) \exp\left(-\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right) \leq k^* \right\}$$

Ahora notemos que para que $\left(\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right) \exp\left(-\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right)$ sea pequeño, requerimos que:

$$\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2} < k_1 \quad \frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2} > k_2$$

Lo anterior se concluye analizando a la función $f(x) = xe^{-x}$ la cual toma valores cercanos a cero

cuando $x \rightarrow 0$ o $x \rightarrow \infty$. Concluimos que debemos rechazar H_0 si :

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2} < k_1 \quad o \quad \frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2} > k_2 \right. \right\}$$

Notemos además que bajo H_0 tenemos la siguiente propiedad distribucional:

$$\frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} \sim \chi_{(n)}^2 \Rightarrow \frac{n\hat{\sigma}_{M.V.}^2}{\sigma_0^2} \sim \chi_{(n)}^2$$

Por lo que nos conviene redefinir nuestra región de rechazo como:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{n\hat{\sigma}_{M.V.}^2}{\sigma_0^2} < k_1 \quad o \quad \frac{n\hat{\sigma}_{M.V.}^2}{\sigma_0^2} > k_2 \right. \right\}$$

Luego entonces, sacando uso de la propiedad distribucional de $\frac{n\hat{\sigma}_{M.V.}^2}{\sigma_0^2}$ la región de rechazo de tamaño α que obtenemos es :

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{n\hat{\sigma}_{M.V.}^2}{\sigma_0^2} < \chi_{(n)}^{2(\alpha/2)} \quad o \quad \frac{n\hat{\sigma}_{M.V.}^2}{\sigma_0^2} > \chi_{(n)}^{2(1-\alpha/2)} \right. \right\}$$

Donde $\chi_{(n)}^{2(1-\alpha/2)}$ y $\chi_{(n)}^{2(\alpha/2)}$ son los cuantiles $1 - \alpha/2$ y $\alpha/2$ respectivamente de una distribución $\chi_{(n)}^2$

Ejercicio 5.2.2. Encuentre la región de rechazo $L.R.T$ para una muestra X_1, \dots, X_n de $N(\mu, \sigma^2)$ para la prueba de hipótesis:

$$H_0 : \sigma^2 = \sigma_0^2 \quad vs \quad H_1 : \sigma^2 \neq \sigma_0^2$$

Con μ desconocida. Muestre que dicha región de rechazo de tamaño α es

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{n\hat{\sigma}_{M.V.}^2}{\sigma_0^2} < \chi_{(n-1)}^{2(\alpha/2)} \quad o \quad \frac{n\hat{\sigma}_{M.V.}^2}{\sigma_0^2} > \chi_{(n-1)}^{2(1-\alpha/2)} \right. \right\}$$

Donde $\chi_{(n-1)}^{2(1-\alpha/2)}$ y $\chi_{(n-1)}^{2(\alpha/2)}$ son los contiles el cuantil $1 - \alpha/2$ y $\alpha/2$ respectivamente de una distribución $\chi_{(n-1)}^2$

Ahora veamos un último ejemplo para el caso en que no conocemos exactamente la distribución de la región de rechazo y utilizaremos el teorema de *Wilks* para solucionarlo.

Ejemplo 5.2.4. Sea X_1, \dots, X_{n_1} m.a. de Poisson(λ_x) y Y_1, \dots, Y_{n_2} m.a de Poisson(λ_y) (Dos

muestras independientes una de otra). Se plantea la prueba de hipótesis:

$$H_0 : \lambda_x = \lambda_y \quad vs \quad H_1 : \lambda_y \neq \lambda_x$$

Encontrar la región de rechazo L.R.T. para esta prueba:

Primero notemos que $\Theta = R^{2+}$, $\Theta_0 = \{(x_1, x_2) \in R^{2+} : x_1 = x_2\}$. En este caso tenemos una muestra general de tamaño $n_1 + n_2$ $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$. Entonces por independencia la verosimilitud conjunta es:

$$f(x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}; \lambda_x, \lambda_y) = \prod_{i=1}^{n_1} \frac{\lambda_x^{x_i}}{x_i!} e^{-\lambda_x} \prod_{i=1}^{n_2} \frac{\lambda_y^{y_i}}{y_i!} e^{-\lambda_y}$$

Hay que maximizar respecto λ_x y λ_y de donde es claro (por estimación máxima verosímil) que el máximo se alcanza en:

$$\hat{\lambda}_{x_{M.V.}} = \bar{x}; \quad \hat{\lambda}_{y_{M.V.}} = \bar{y}$$

Por lo tanto:

$$\begin{aligned} \sup_{\Theta} f(x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}; \lambda_x, \lambda_y) &= \prod_{i=1}^{n_1} \frac{\bar{x}^{x_i}}{x_i!} e^{-\bar{x}} \prod_{i=1}^{n_2} \frac{\bar{y}^{y_i}}{y_i!} e^{-\bar{y}} \\ &= e^{-n_1 \bar{x}} \bar{x}^{\sum_{i=1}^{n_1} x_i} e^{-n_2 \bar{y}} \bar{y}^{\sum_{i=1}^{n_2} y_i} \prod_{i=1}^{n_1} \frac{1}{x_i} \prod_{i=1}^{n_2} \frac{1}{y_i} \end{aligned}$$

Por otro lado, bajo H_0 tenemos que $\lambda_x = \lambda_y = \lambda$ entonces:

$$\begin{aligned} f(x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}; \lambda) &= \prod_{i=1}^{n_1} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \prod_{i=1}^{n_2} \frac{\lambda^{y_i}}{y_i!} e^{-\lambda} \\ &= e^{-(\lambda(n_1+n_2))} \lambda^{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i} \prod_{i=1}^{n_1} \frac{1}{x_i} \prod_{i=1}^{n_2} \frac{1}{y_i}. \end{aligned}$$

Al maximizar respecto a λ en la última ecuación obtenemos el λ^* que maximiza está dado por

$$\lambda^* = \frac{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i}{n_1 + n_2} = \frac{n_1 \bar{x} + n_2 \bar{y}}{n_1 + n_2} := \hat{\lambda}_p$$

Donde $\hat{\lambda}_p$ se le conoce como el estimador de λ ponderado. Por lo tanto:

$$\sup_{\Theta_0} f(x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}; \lambda) = e^{-(\hat{\lambda}_p(n_1+n_2))} \hat{\lambda}_p^{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i} \prod_{i=1}^{n_1} \frac{1}{x_i} \prod_{i=1}^{n_2} \frac{1}{y_i}.$$

Finalmente al hacer el cociente de verosimilitudes obtenemos:

$$\lambda(\underline{x}, \underline{y}) = \frac{e^{-(\hat{\lambda}_p(n_1+n_2))} \hat{\lambda}_p^{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i}}{e^{-n_1 \bar{x}} \bar{x}^{\sum_{i=1}^{n_1} x_i} e^{-n_2 \bar{y}} \bar{y}^{\sum_{i=1}^{n_2} y_i}} = \frac{\hat{\lambda}_p^{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i}}{\bar{x}^{\sum_{i=1}^{n_1} x_i} \bar{y}^{\sum_{i=1}^{n_2} y_i}}$$

Finalmente la región de Rechazo será:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{\hat{\lambda}_p^{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i}}{\bar{x}^{\sum_{i=1}^{n_1} x_i} \bar{y}^{\sum_{i=1}^{n_2} y_i}} \leq k \right. \right\}$$

Desafortunadamente el estadístico de prueba asociado a este cociente no tiene una distribución conocida por lo que procedemos a aplicar el Teorema de Wilks que nos asegura que:

$$\Lambda = -2 \log(\lambda(\underline{x}, \underline{y})) \sim_{approx} \chi_1^2$$

En esta caso la χ^2 tiene 1 grado de libertad asociado por que bajo $\Theta \in \mathbb{R}^{2+}$ y bajo H_0 solo se tiene 1 parámetro libre por lo tanto según Wilks la χ^2 asociada tiene $2 - 1$ grados de libertad. Luego entonces rechazaremos H_0 si:

$$-2 \log(\lambda(\underline{x}, \underline{y})) > \chi_{(1)}^{2(1-\alpha)}$$

Supongamos un ejemplo numérico, una aseguradora sospecha que la tasa de accidentes diarios de motocicleta y de autos es la misma. La aseguradora supone el número de accidentes diarios de autos y motos siguen una distribución poisson. Sea λ_x = Tasa de accidente en auto, y λ_y = Tasa de accidente en motocicleta. La aseguradora se está planteando entonces la hipótesis:

$$H_0 : \lambda_x = \lambda_y \quad vs \quad H_1 : \lambda_y \neq \lambda_x$$

Para probar su hipótesis la aseguradora observa por 20 días los accidentes en motos y autos obteniendo la siguiente muestra: Número accidente en autos

$$\{4, 4, 1, 3, 7, 5, 1, 1, 1, 4, 2, 4, 1, 6, 6, 3, 3, 4, 3, 2\} \rightarrow \bar{x} = 3.25, \quad \sum_{i=1}^n x_i = 65$$

Número accidente en motos

$$\{6, 1, 5, 2, 1, 2, 4, 1, 2, 0, 0, 1, 2, 0, 2, 3, 2, 2, 2, 3\} \rightarrow \bar{y} = 2.05 \quad \sum_{i=1}^n y_i = 41$$

Con esta información tenemos que:

$$\hat{\lambda}_p = \frac{20 * 2.05 + 20 * 3.25}{40} = 2.65$$

Entonces

$$\lambda(\underline{x}, \underline{y}) = \frac{\hat{\lambda}_p^{\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i}}{\bar{x}^{\sum_{i=1}^{n_1} x_i} \bar{y}^{\sum_{i=1}^{n_2} y_i}} = \frac{2.65^{65+41}}{3.25^{65} 2.05^{41}} = 0.06452617$$

Entonces

$$\Lambda(\underline{x}, \underline{y}) = -2 \log(\lambda(\underline{x}, \underline{y})) = -2 \log(0.06452617) = 5.481369$$

La regla es rechazar H_0 a un nivel de significancia α si:

$$\Lambda(\underline{x}, \underline{y}) = 5.481369 > \chi_{(1)}^{2(1-\alpha)}$$

Si fijamos $\alpha = 0.05$ entonces $\chi_{(1)}^{2(1-\alpha)} = 3.841459$ por lo tanto se Rechaza H_0 , tenemos suficiente evidencia para decir que los accidentes de moto y de auto no tienen la misma tasa de incidencia. El p-value asociado a la prueba lo obtenemos calculando la siguiente probabilidad:

$$\mathbb{P}(W > 5.481369 \mid W \sim \chi_{(1)}^2) = 0.0192202$$

Quiere decir que la probabilidad de haber observado estas muestras bajo H_0 es muy baja, por lo tanto la muestra no es compatible con la hipótesis y por tanto se Rechaza H_0

5.3. Regiones Críticas para distribuciones Normales (L.R.T.)

A continuación escribimos las regiones críticas obtenidas por el procedimiento del cociente de verosimilitudes (L.R.T).

5.3.1. Contrastes para μ , suponiendo σ^2 conocida

Entrada: X_1, \dots, X_n m.a. de un modelo $N(\mu, \sigma^2)$

Estadístico de Prueba:

$$T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim N(0, 1)$$

H_0	H_1	\mathcal{C}_{LRT}^* de tamaño α
$\mu = \mu_0$	$\mu > \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} > Z_{(1-\alpha)} \right\}$
$\mu = \mu_0$	$\mu < \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} < Z_{(\alpha)} \right\}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \left \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \right > Z_{(1-\frac{\alpha}{2})} \right\}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} > Z_{(1-\alpha)} \right\}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} < Z_{(\alpha)} \right\}$

5.3.2. Contrastes para μ , suponiendo σ^2 desconocida

Entrada: X_1, \dots, X_n m.a. de un modelo $N(\mu, \sigma^2)$

Estadístico de Prueba:

$$T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\hat{\sigma}} \sim t_{(n-1)}; \quad \text{con} \quad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

H_0	H_1	\mathcal{C}_{LRT}^* de tamaño α
$\mu = \mu_0$	$\mu > \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{\sigma}} > t_{(n-1)}^{(1-\alpha)} \right\}$
$\mu = \mu_0$	$\mu < \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{\sigma}} < t_{(n-1)}^{(\alpha)} \right\}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \left \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{\sigma}} \right > t_{(n-1)}^{(1-\frac{\alpha}{2})} \right\}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{\sigma}} > t_{(n-1)}^{(1-\alpha)} \right\}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{\sigma}} < t_{(n-1)}^{(\alpha)} \right\}$

5.3.3. Contrastes para σ^2 , suponiendo μ conocida

Entrada: X_1, \dots, X_n m.a. de un modelo $N(\mu, \sigma^2)$

Estadístico de Prueba:

$$T = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma_0^2} \sim \chi_{(n)}^2$$

H_0	H_1	\mathcal{C}_{LRT}^* de tamaño α
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} > \chi_{(n)}^{2(1-\alpha)} \right\}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} < \chi_{(n)}^{2(\alpha)} \right\}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} < \chi_{(n)}^{2(\frac{\alpha}{2})} \quad o \quad \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} > \chi_{(n)}^{2(1-\frac{\alpha}{2})} \right\}$
$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} > \chi_{(n)}^{2(1-\alpha)} \right\}$
$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} < \chi_{(n)}^{2(\alpha)} \right\}$

5.3.4. Contrastes para σ^2 , suponiendo μ desconocida

Entrada: X_1, \dots, X_n m.a. de un modelo $N(\mu, \sigma^2)$

Estadístico de Prueba:

$$T = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma_0^2} \sim \chi_{(n-1)}^2$$

H_0	H_1	\mathcal{C}_{LRT}^* de tamaño α
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} > \chi_{(n-1)}^{2(1-\alpha)} \right\}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} < \chi_{(n-1)}^{2(\alpha)} \right\}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} < \chi_{(n-1)}^{2(\frac{\alpha}{2})} \quad o \quad \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} > \chi_{(n-1)}^{2(1-\frac{\alpha}{2})} \right\}$
$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} > \chi_{(n-1)}^{2(1-\alpha)} \right\}$
$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} < \chi_{(n-1)}^{2(\alpha)} \right\}$

5.3.5. Contrastes para la diferencia de medias $\mu_x - \mu_y$ suponiendo σ_x^2, σ_y^2 conocidas

Entrada: X_1, \dots, X_{n_1} m.a. de $N(\mu_x, \sigma_x^2)$ y Y_1, \dots, Y_{n_2} m.a. de $N(\mu_y, \sigma_y^2)$;

Estadístico de Prueba:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\left(\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}\right)}} \sim N(0, 1)$$

H_0	H_1	\mathcal{C}_{LRT}^* de tamaño α
$(\mu_x - \mu_y) = d_0$	$(\mu_x - \mu_y) > d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\frac{1}{n}(\sigma_x^2 + \sigma_y^2)}} > Z_{(1-\alpha)} \right\}$
$(\mu_x - \mu_y) = d_0$	$(\mu_x - \mu_y) < d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\frac{1}{n}(\sigma_x^2 + \sigma_y^2)}} < Z_{(\alpha)} \right\}$
$(\mu_x - \mu_y) = d_0$	$(\mu_x - \mu_y) \neq d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \left \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\frac{1}{n}(\sigma_x^2 + \sigma_y^2)}} \right > Z_{(1-\frac{\alpha}{2})} \right\}$
$(\mu_x - \mu_y) \leq d_0$	$(\mu_x - \mu_y) > d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\frac{1}{n}(\sigma_x^2 + \sigma_y^2)}} > Z_{(1-\alpha)} \right\}$
$(\mu_x - \mu_y) \geq d_0$	$(\mu_x - \mu_y) < d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\frac{1}{n}(\sigma_x^2 + \sigma_y^2)}} < Z_{(\alpha)} \right\}$

5.3.6. Contrastes para la diferencia de medias $\mu_x - \mu_y$ suponiendo σ_x^2 , σ_y^2 desconocidas pero iguales $\sigma_x^2 = \sigma_y^2$

Entrada: X_1, \dots, X_{n_1} m.a. de $N(\mu_x, \sigma^2)$ y Y_1, \dots, Y_{n_2} m.a. de $N(\mu_y, \sigma^2)$;

Estadístico de Prueba:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\hat{\sigma}_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t_{(n_1+n_2-2)}$$

Donde $\hat{\sigma}_p^2$ está dada por:

$$\hat{\sigma}_p^2 := \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1) \hat{\sigma}_x^2 + (n_2 - 1) \hat{\sigma}_y^2}{n_1 + n_2 - 2}$$

H_0	H_1	\mathcal{C}_{LRT}^* de tamaño α
$(\mu_x - \mu_y) = d_0$	$(\mu_x - \mu_y) > d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\hat{\sigma}_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} > t_{(n_1+n_2-2)}^{(1-\alpha)} \right\}$
$(\mu_x - \mu_y) = d_0$	$(\mu_x - \mu_y) < d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\hat{\sigma}_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} < t_{(n_1+n_2-2)}^{(\alpha)} \right\}$
$(\mu_x - \mu_y) = d_0$	$(\mu_x - \mu_y) \neq d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\hat{\sigma}_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} > t_{(n_1+n_2-2)}^{(1-\frac{\alpha}{2})} \right\}$
$(\mu_x - \mu_y) \leq d_0$	$(\mu_x - \mu_y) > d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\hat{\sigma}_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} > t_{(n_1+n_2-2)}^{(1-\alpha)} \right\}$
$(\mu_x - \mu_y) \geq d_0$	$(\mu_x - \mu_y) < d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\hat{\sigma}_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} < t_{(n_1+n_2-2)}^{(\alpha)} \right\}$

5.3.7. Contrastes para el cociente de varianzas $\frac{\sigma_x}{\sigma_y}$ suponiendo μ_x, μ_y conocidas

Entrada: X_1, \dots, X_{n_1} m.a. de $N(\mu_x, \sigma_x^2)$ y Y_1, \dots, Y_{n_2} m.a. de $N(\mu_y, \sigma_y^2)$;

Estadístico de Prueba:

$$F = \frac{\frac{1}{n_2} \sum_{i=1}^{n_2} (Y_i - \mu_y)^2}{\frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - \mu_x)^2} \sim F_{(n_2, n_1)}$$

H_0	H_1	\mathcal{C}_{LRT}^* de tamaño α
$\frac{\sigma_x^2}{\sigma_y^2} = 1$	$\frac{\sigma_x^2}{\sigma_y^2} > 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - \mu_y)^2}{\frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \mu_x)^2} > F_{(n_1, n_2)}^{1-\alpha} \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} = 1$	$\frac{\sigma_x^2}{\sigma_y^2} < 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - \mu_y)^2}{\frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \mu_x)^2} < F_{(n_1, n_2)}^\alpha \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} = 1$	$\frac{\sigma_x^2}{\sigma_y^2} \neq 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - \mu_y)^2}{\frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \mu_x)^2} < F_{(n_1, n_2)}^{\frac{\alpha}{2}} \quad o \quad \frac{\frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - \mu_y)^2}{\frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \mu_x)^2} > F_{(n_1, n_2)}^{1-\frac{\alpha}{2}} \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} \leq 1$	$\frac{\sigma_x^2}{\sigma_y^2} > 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - \mu_y)^2}{\frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \mu_x)^2} > F_{(n_1, n_2)}^{1-\alpha} \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} \geq 1$	$\frac{\sigma_x^2}{\sigma_y^2} < 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - \mu_y)^2}{\frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \mu_x)^2} < F_{(n_1, n_2)}^\alpha \right\}$

5.3.8. Contrastes para el cociente de varianzas $\frac{\sigma_x}{\sigma_y}$ suponiendo μ_x, μ_y desconocidas

Entrada: X_1, \dots, X_{n_1} m.a. de $N(\mu_x, \sigma_x^2)$ y Y_1, \dots, Y_{n_2} m.a. de $N(\mu_y, \sigma_y^2)$;

Estadístico de Prueba:

$$F = \frac{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2} \sim F_{(n_2-1, n_1-1)}$$

H_0	H_1	\mathcal{C}_{LRT}^* de tamaño α
$\frac{\sigma_x^2}{\sigma_y^2} = 1$	$\frac{\sigma_x^2}{\sigma_y^2} > 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2} > F_{(n_1-1, n_2-1)}^{1-\alpha} \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} = 1$	$\frac{\sigma_x^2}{\sigma_y^2} < 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2} < F_{(n_1-1, n_2-1)}^\alpha \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} = 1$	$\frac{\sigma_x^2}{\sigma_y^2} \neq 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2} < F_{(n_1-1, n_2-1)}^{\frac{\alpha}{2}} \quad o \quad \frac{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2} > F_{(n_1-1, n_2-1)}^{1-\frac{\alpha}{2}} \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} \leq 1$	$\frac{\sigma_x^2}{\sigma_y^2} > 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2} > F_{(n_1-1, n_2-1)}^{1-\alpha} \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} \geq 1$	$\frac{\sigma_x^2}{\sigma_y^2} < 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2} < F_{(n_1-1, n_2-1)}^\alpha \right\}$

5.3.9. Contrastes para la igualdad de medias de k poblaciones normales (varianzas conocidas)

Una de las pruebas más utilizadas en la práctica es la que nos permite ver si existe alguna diferencia en las medias de k poblaciones normales, por ejemplo, imaginemos que estamos realizando un experimento con 3 tipos de nuevos anti hipertensivos, para ello se forman 3 grupos de personas con hipetensión, a cada uno de los grupos se le aplica un tipo de anti hipertensivo y después de cierto tiempo se cuantifica la eficacia del medicamento midiendo la presión de cada individuo en todos los grupos. Imaginemos que se desea verificar si existe un anti hipertensivo que sea mejor que los demás. Para ello se decide modelar $X_{i,j}$ = la presión arterial del individuo j ($j \in \{1, \dots, n_i\}$) en la población i ($i \in \{1, 2, 3\}$). Si suponemos que

$$X_{i,j} \sim N(\mu_i, \sigma_i^2) \quad i \in \{1, 2, 3\} \quad j \in \{1, \dots, n_i\}$$

Entonces el problema se podría reducir a contrastar la siguientes hipótesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad vs \quad H_1 : \exists i, j \text{ tal que } \mu_i \neq \mu_j$$

La región de rechazo para este tipo de pruebas puede ser encontrada utilizando la técnica L.T.R., en el cual si suponemos varianzas conocidas se obtiene lo siguiente:

Contraste:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad vs \quad H_1 : \exists i, j \text{ tal que } \mu_i \neq \mu_j$$

Entrada:

$$\begin{array}{ll} X_{1,1}, \dots, X_{1,n_1} & \text{m.a. de } N(\mu_1, \sigma_1^2) \\ X_{2,1}, \dots, X_{2,n_2} & \text{m.a. de } N(\mu_2, \sigma_2^2) \\ & \vdots \\ X_{k,1}, \dots, X_{k,n_k} & \text{m.a. de } N(\mu_k, \sigma_k^2) \end{array}$$

Estadístico de Prueba:

$$T = \sum_{i=1}^k \frac{n_i}{\sigma_i^2} (\bar{X}_i - \bar{X}_{**})^2 \sim \chi_{(k-1)}^2$$

Donde:

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j} \quad \bar{X}_{**} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{i,j} \quad \sum_{i=1}^k n_i = n$$

Región de Rechazo:

$$\mathcal{C}_{LRT}^* = \left\{ \sum_{i=1}^k \frac{n_i}{\sigma_i^2} (\bar{x}_i - \bar{x}_{**})^2 > \chi_{(k-1)}^{2(1-\alpha)} \right\}$$

5.3.10. Contrastes para la igualdad de medias de k poblaciones normales (varianzas desconocidas pero iguales) ANOVA

El problema de la prueba anterior es que por lo general las varianzas van a ser desconocidas en el problema por lo que es necesario crear una regla de decisión para el caso en que las varianzas son desconocidas.

Atacar el problema de varianzas desconocidas es complejo y el estadístico de prueba se vuelve intratable cuando suponemos varianzas desiguales en cada población, afortunadamente cuando se supone homocedasticidad (igualdad de varianzas) los cálculos quedan mas ordenados y pueden ser presentados en una tabla llamada ANOVA (análisis de varianza)

Contraste:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad vs \quad H_1 : \exists i, j \text{ tal que } \mu_i \neq \mu_j$$

Entrada:

$$\begin{array}{ll} X_{1,1}, \dots, X_{1,n_1} & \text{m.a. de } N(\mu_1, \sigma^2) \\ X_{2,1}, \dots, X_{2,n_2} & \text{m.a. de } N(\mu_2, \sigma^2) \\ & \vdots \\ X_{k,1}, \dots, X_{k,n_k} & \text{m.a. de } N(\mu_k, \sigma^2) \end{array}$$

Obs: Observe que estamos suponiendo igualdad de varianzas en las poblaciones (homocedasticidad)

Estadístico de Prueba:

$$F = \frac{(n-k) \sum_{i=1}^k n_i (\bar{X}_i - \bar{X}_{**})^2}{(k-1) \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{i,j} - \bar{X}_i)^2} \sim F_{k-1, n-k}$$

Donde:

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j} \quad \bar{X}_{**} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{i,j} \quad \sum_{i=1}^k n_i = n$$

Región de Rechazo:

$$\mathcal{C}_{LRT}^* = \left\{ \frac{(n-k) \sum_{i=1}^k n_i (\bar{x}_i - \bar{x}_{**})^2}{(k-1) \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2} > F_{k-1, n-k}^{1-\alpha} \right\}$$

Una forma en como los paquetes estadísticos presentan esta prueba es mediante la tabla ANOVA:

Variación	Suma de Cuadrados	G. de Lib	Cuadrados Medios	Estadístico F
Tratamientos	$\sum_{i=1}^k n_i (\bar{x}_i - \bar{x}_{**})^2$	$k - 1$	$\frac{\sum_{i=1}^k n_i (\bar{x}_i - \bar{x}_{**})^2}{k-1}$	$\frac{(n-k) \sum_{i=1}^k n_i (\bar{x}_i - \bar{x}_{**})^2}{(k-1) \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2}$
Error	$\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2$	$n - k$	$\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2}{n-k}$	
Total	$\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_{**})^2$	$n - 1$	$\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_{**})^2}{n-1}$	

Ejercicio 5.3.1. Resuelva lo siguiente:

1. Sea X_1 una observación del modelo:

$$f_X(x) = \theta x^{\theta-1} \mathbf{1}_{(0,1)}(x); \quad \theta > 0$$

Se desea contrastar las siguientes hipótesis:

$$H_0 : \theta = 2 \quad vs \quad H_1 : \theta = 1$$

- Encontrar la región de rechazo mas potente suponiendo que deseamos una probabilidad de error tipo 1 $\alpha = 0.05$.
- Para la región de rechazo encontrada en el inciso anterior, calcular la probabilidad del error tipo 2 y encuentre la potencia de la prueba

2. Sea X_1, \dots, X_5 m.a. del modelo $\text{Normal}(0, \sigma^2)$. Se desea contrastar las siguientes hipótesis:

$$H_0 : \sigma^2 = 1 \quad vs \quad H_1 : \sigma^2 = 2$$

- Encontrar la región de rechazo mas potente suponiendo que deseamos una probabilidad de error tipo 1 $\alpha = 0.1$.
- Para la región de rechazo encontrada en el inciso anterior, calcular la probabilidad del error tipo 2 y encuentre la potencia de la prueba
- Suponga que observa la siguiente muestra: $(4.14, 0.78, -1.28, 1.30, 1.33)$, ¿rechazaría H_0 ?, encuentre el p-value asociado a esta prueba.

3. Sea X_1 una observación del modelo:

$$f_X(x) = (2\theta x + 1 - \theta) \mathbf{1}_{(0,1)}(x); \quad \theta \in [-1, 1]$$

Se desea contrastar las siguientes hipótesis:

$$H_0 : \theta = 0 \quad vs \quad H_1 : \theta = 1$$

- Encontrar la región de rechazo mas potente suponiendo que deseamos una probabilidad de error tipo 1 $\alpha = 0.05$.
- Para la región de rechazo encontrada en el inciso anterior, calcular la probabilidad del error tipo 2 y encuentre la potencia de la prueba
- Suponga que observa la siguiente muestra: (0.5) , ¿rechazaría H_0 ?, encuentre el p-value asociado a esta prueba.

4. Sea X_1, \dots, X_7 m.a. del modelo $\text{Gamma}(2, \beta)$

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

Se desea contrastar las siguientes hipótesis:

$$H_0 : \beta = 1 \quad \text{vs} \quad H_1 : \beta = 4$$

- Encontrar la región de rechazo mas potente suponiendo que deseamos una probabilidad de error tipo 1 $\alpha = 0.05$.
- Para la región de rechazo encontrada en el inciso anterior, calcular la probabilidad del error tipo 2 y encuentre la potencia de la prueba
- Suponga que observa la siguiente muestra: $(0.650, 1.048, 3.109, 0.526, 0.558, 1.228, 1.320)$, ¿rechazaría H_0 ?, encuentre el p-value asociado a esta prueba.

5. La UNAM utiliza bombillas especiales para poder iluminar los distintos salones de varias facultades. Actualmente la universidad utiliza bombillas fabricadas por la **compañía 1** la cual afirma que sus bombillas tienen una vida útil de 900 horas en promedio, la **compañía 2** afirma que ellos fabrican una nueva bombilla con una tecnología mas eficiente que hace que sus bombillas duren mas de 900 horas en promedio. La UNAM desea tomar una decisión sobre si cambia de proveedor o no, para ello optó por hacer un experimento sencillo, compró 20 bombillas a cada compañía y las colocó en distintos salones para medir su durabilidad.

Los datos de este experimento se encuentran la siguiente tabla. Con esta información responda lo siguiente suponiendo normalidad e igualdad de varianzas en ambas poblaciones

COMPAÑIA1	COMPAÑIA2
856.33	866.77
969.69	907.61
828.62	831.06
962.91	937.76
917.56	963.75
938.78	941.05
930.21	980.58
759.69	940.63
809.59	926.95
802.53	862.5
779.71	858.92
728.85	945.34
889.03	775.36
797.97	960.19
845.82	891.21
767.92	939.64
824.37	978.08
940.06	922.53
891.85	973.22
897.06	995.29

- a) ¿Se tiene evidencia estadística como para afirmar que las bombillas de la **compañía 1** en efecto tienen una vida promedio de 900 horas?. Para responder esto lleve a cabo la prueba de hipótesis:

$$H_0 : \mu_1 = 900 \quad \text{vs} \quad H_1 : \mu_1 \neq 900$$

Donde μ_1 representa el tiempo de vida promedio de las bombillas de la **compañía 1**. Calcule el p-value de la prueba y asumiendo un $\alpha = 0.05$ concluya.

- b) ¿Se tiene evidencia estadística como para afirmar que las bombillas de la **compañía 2** tienen una vida promedio mayor a 900 horas?. Para responder este lleve a cabo la prueba de hipótesis:

$$H_0 : \mu_2 \leq 900 \quad \text{vs} \quad H_1 : \mu_2 > 900$$

Donde μ_2 representa el tiempo de vida promedio de las bombillas de la **compañía 2**

. Calcule el p-value de la prueba y asumiendo un $\alpha = 0.05$ concluya.

- c) Finalmente, se desea tomar la decisión final para saber si se cambia o no de proveedor de bombillas. Para responder esto lleve a cabo la prueba de hipótesis.

$$H_0 : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 < \mu_2$$

Calcule el p-value de la prueba y asumiendo un $\alpha = 0.05$ concluya.

6. Una compañía de jugos compró una maquina que llena botellas. El manual de instructivo dice que la maquina está configurada para llenar las botellas en promedio con 1 litro de liquido. La compañía esta interesada en verificar esta afirmación, para ello hizo que la maquina llenara $n = 10$ botellas y luego midió la cantidad exacta (en litros) con la que fueron llenadas. La información se encuentra en la siguiente tabla.

BOTELLA	LITROS SURTIDOS
1	0.96
2	1.01
3	1.14
4	0.87
5	1.08
6	1.05
7	1.13
8	0.99
9	1.06
10	1.05

Para probar la aseveración y dado que la compañía esta preocupada de que la máquina este surtiendo mas líquido de lo esperado se planteó la siguiente prueba de hipótesis

$$H_0 : \mu \leq 1 \quad vs \quad H_1 : \mu > 1$$

Asumiendo normalidad y varianza conocida ($\sigma = 0.1$) La compañía propuso la siguiente región de rechazo:

$$C = \left\{ \sqrt{n} \left(\frac{\bar{X} - 1}{0.1} \right) > 1.8 \right\}$$

- a) Calcule el tamaño de la prueba ($\alpha = \max_{H_0} \pi(\mu)$)

b) Con los datos observados, calcule el p -value y concluya si rechaza o no la hipótesis nula. (utilice $\alpha = 0.5$ como su nivel de significancia para decidir)

c) Realice el contraste nuevamente pero ahora suponga σ^2 desconocida (utilice $\alpha = 0.5$ como su nivel de significancia para decidir)

7. Sean $X_1, \dots, X_n \stackrel{i.i.d}{\sim} \text{Bernoulli}(\theta)$.

Utilice el cociente de verosimilitudes generalizado (LRT) para probar que la región de rechazo asociada a la prueba de hipótesis

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0$$

está dada por:

$$C := \left\{ (X_1, \dots, X_n) \mid \sum_{i=1}^n X_i \geq k_1 \text{ ó } \sum_{i=1}^n X_i \leq k_2 \right\}$$