

UNIVERSIDAD DE BUENOS AIRES
Facultad de Ingeniería

Curso de Probabilidad y Estadística

Ing. Jemina M. García

Buenos Aires, 2020.

Índice general

1. Inferencia estadística	1
1.1. Concepto de muestra	1
1.2. Modelos paramétricos	2
1.3. Familias exponenciales	3
1.4. Estadísticos	4
1.4.1. Estadísticos suficientes	4
1.4.2. Estimadores	5
1.5. Método de máxima verosimilitud	6
1.6. Bondad de los estimadores	7
1.7. Estimadores asintóticamente normales	8
1.8. Algunas distribuciones importantes	9
2. Test de hipótesis	11
2.1. Test para hipótesis simple vs. hipótesis simple	15
2.2. Tests uniformemente mas potentes para hipótesis unilaterales	18
2.3. Test del cociente de máxima verosimilitud	20
2.4. Test con nivel de significación asintótico	22
2.4.1. Distribución asintótica del test del cociente de máxima verosimilitud	22
3. Intervalos de confianza	23
3.1. Intervalos de confianza de nivel exacto $1 - \alpha$	24
3.1.1. Método del pivote	24
3.2. Intervalos de confianza de nivel asintótico $1 - \alpha$	24
3.3. Relación entre regiones de confianza y test	25
4. Estimación Bayesiana	27
4.1. Otros estimadores de Bayes	30
4.2. Estimación de probabilidades	30

Capítulo 1

Inferencia estadística

En la primera parte del curso desarrollamos los principios básicos de la teoría de probabilidad, ofreciendo un conjunto de distribuciones probabilísticas particularmente útiles en la resolución de ciertos tipos de problemas. En la teoría de probabilidad se presupone que la naturaleza del fenómeno bajo estudio es conocida y se describe como se comportarían (es decir, como están distribuidas) ciertas características que pueden ser observadas en los experimentos. En la estadística matemática las cosas suceden al revés: el punto de partida son los datos experimentales (generalmente las observaciones de variables aleatorias). El proceso de construir un modelo sobre la base de datos experimentales y de extraer conclusiones de él se llama inferencia inductiva. Cuando la inferencia inductiva se aplica a problemas estadísticos se la llama inferencia estadística. El objetivo del trabajo de los estadísticos es hacer inferencias estadísticas por medio de datos experimentales. Cierta experimentación aleatoria se realiza bajo ciertas condiciones. La totalidad de los resultados experimentales posibles se llama población. El conjunto de datos que se obtiene al realizar el experimento una cierta cantidad de veces se denomina muestra de la población. Un problema básico de la estadística matemática es el de cómo extraer información de las muestras para utilizarla al estudiar las poblaciones de las que se obtuvieron dichas muestras.

1.1. Concepto de muestra

Analicemos en detalle el concepto de muestra. El conjunto de resultados de las observaciones sirve como punto de partida para toda investigación estadística. Estos resultados son los valores experimentales de cierta variable aleatoria X definida sobre un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$, cuya función de distribución, $F_X(x) = \mathbf{P}(X \leq x)$, se desconoce por lo menos parcialmente.

Supongamos que tenemos un experimento aleatorio relacionado con la variable aleatoria X . En tal caso la variable aleatoria X representa un *observable* del experimento aleatorio. Dicho de otro modo, X representa una magnitud física que se puede medir cuando realizamos el experimento. Los valores de X constituyen la población sujeta a estudio. Lo que se quiere saber es como se comporta la población (es decir, cual es la distribución de la variable aleatoria X).

Un **Presupuesto básico** es que se pueden examinar tantas repeticiones independientes del experimento aleatorio como se desee. Formalmente, esto significa que se dispone de una muestra aleatoria de volumen infinito. Cuando hablamos de muestra aleatoria nos referimos a una sucesión de variables aleatorias independientes $X_1, X_2, \dots, X_n, \dots$ todas idénticamente distribuidas a X .

Por lo tanto, tendremos que

$$\begin{aligned} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= \mathbf{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \\ &= \prod_{i=1}^n \mathbf{P}(X \leq x_i) = \prod_{i=1}^n F_X(x_i) \end{aligned} \quad (1.1)$$

para todo $n \in \mathbb{N}$ y para todo vector n -dimensional $(x_1, x_2, \dots, x_n) \in \mathbb{R}$.

Para modelar n repeticiones independientes del experimento aleatorio usaremos el vector aleatorio n -dimensional $\underline{X} = (X_1, \dots, X_n)$ compuesto por las primeras n coordenadas de X que sera llamado muestra aleatoria de tamaño n de la población con distribución $F_X(x)$.

El conjunto de las n observaciones obtenidas al examinar n repeticiones independientes del experimento sera designado por $\underline{x} = (x_1, x_2, \dots, x_n)$ (se llama muestra de tamaño n y corresponden a los valores observados de X)

1.2. Modelos paramétricos

Presupuesto basico. La distribucion de la poblacion sometida a estudio es parcialmente desconocida. En otras palabras, se sabe de antemano que la distribucion de la poblacion no puede ser arbitraria, sino que pertenece a una determinada familia de distribuciones \mathfrak{F} que dependen de un paramtero desconocido.

Descripción de la familia paramétrica de distribuciones. En todo lo que sigue $\mathfrak{F} = \{F_\theta(x) : \theta \in \Theta\}$ sera una familia de distribuciones de probabilidad parametrizadas por un subconjunto no vacío $\Theta \subset \mathbb{R}^p$ llamado el espacio parametrico. (sí...la mayúscula

de tita se llama titón). Se supone que existe una correspondencia uno a uno entre el espacio paramétrico Θ y la familia de distribuciones \mathfrak{F} (es decir., $F_{\theta_1}(x) \neq F_{\theta_2}(x)$ cuando $\theta_1 \neq \theta_2$). Ya sea que se trate del caso continuo o del caso discreto, usaremos la notación $f_{\theta}(x)$ para designar la función densidad de probabilidad o la función de probabilidad de la distribución $F_{\theta}(x)$, respectivamente.

Ejemplos para ver si entendemos de que estoy hablando: ¿qué familias paramétricas conocemos? Bernoulli, Binomial, Exponencial, Gamma, Normal, Uniforme.

Para un planteo del problema. Lo que se quiere hacer es resolver el problema de “adivinar de manera científica” cual es el valor de $\theta \in \Theta$ que determina la distribución de la población a partir de los datos experimentales \underline{x} obtenidos mediante n repeticiones independientes del experimento aleatorio.

A la función de probabilidad o función de densidad de probabilidad, vista como una función de θ la llamaremos función de verosimilitud. Dependiendo si el vector aleatorio \underline{X} es discreto o continuo, la función de verosimilitud será

$$L(\theta, \underline{x}) = \prod_{i=1}^n f_{\theta}(x_i), \quad \text{o} \quad L(\theta, \underline{x}) = \prod_{i=1}^n p_{\theta}(x_i)$$

Se usa la letra L del inglés likelihood.

Dentro de las familias paramétricas, tendremos lo que llamamos **Familias regulares**. Se dice que una familia paramétrica es regular si:

1. El soporte de $f_{\theta}(x)$ no depende de θ
2. $f_{\theta}(x)$ es derivable con respecto a θ para todo x
3. El conjunto paramétrico $\Theta \in \mathbb{R}^p$ es abierto

1.3. Familias exponenciales

Se dice que una familia de distribuciones continuas o discretas en \mathbb{R}^q , con distribución $F_{\theta}(x)$, $\theta \in \Theta \subset \mathbb{R}^k$ es una familia exponencial a k parámetros, si la función de densidad (o probabilidad) se puede escribir como:

$$f_{\theta}(x) = A(\theta) \cdot e^{\sum_{i=1}^k c_i(\theta) r_i(x)} \cdot h(x) \quad (1.2)$$

Donde $c_i(\theta)$ son funciones de $\Theta \rightarrow \mathbb{R}$, $A(\theta)$ es función de $\Theta \rightarrow \mathbb{R}^+$, $r_i(x)$ son funciones de $\mathbb{R}^q \rightarrow \mathbb{R}$, y $h(x)$ es función de $\mathbb{R}^q \rightarrow \mathbb{R}^+$.

Ejemplo: Analicemos si las siguientes distribuciones pertenecen a una familia exponencial: Bernoulli (p), Pareto ($2, \theta$), Uniforme ($2, \theta$)

1.4. Estadísticos

Definición: Un estadístico es cualquier función medible $T_n = T(\underline{X})$ con valores en un espacio euclídeo de dimensión finita.

Dada una muestra aleatoria \underline{X} , un estadístico es una función de la muestra aleatoria, que evaluada en los valores observados debe poder resultar en un valor numérico. Por lo tanto, esta función No puede depender de parámetros desconocidos!!.

Ordenemos nuestras ideas. Nos paremos frente a un experimento aleatorio, definimos la variable (o las variables) que queremos estudiar. Comenzamos a realizar el experimento y recolectamos muestras, es decir observaciones del experimento. Ahora, cuando obtengo una muestra, obtengo mucha información. MUCHA. ¿Quiero guardarla toda? o tal vez alcance con guardar solamente algunos números, que resulten de evaluar funciones en la muestra. Entonces la pregunta que me hago es ¿Que estadístico tengo que guardar para tener toda la información que necesito, en el menor espacio posible?

1.4.1. Estadísticos suficientes

Definición: Sea $\underline{X} = (X_1, \dots, X_n)$ un vector aleatorio de dimensión n cuya distribución es $F_\theta(\underline{x}), \theta \in \Theta$, se dice que un estadístico $T = r(\underline{X})$ es suficiente para θ si la distribución de \underline{X} condicionada a que $T = t$ es independiente de θ para todo t .

Esto significa que si conozco a T y a la distribución de $\underline{X}|T = t$ entonces puedo reconstruir una muestra con la misma distribución que la muestra original.

Encontrar estadísticos suficientes será muy útil al momento en el que busquemos conocer aproximaciones de nuestros parámetros desconocidos, y también al momento de tomar decisiones.

Ejemplo: Queremos encontrar un estadístico suficiente para el parámetro p de una población de Bernoulli. Entonces, nos proponen probar con la suma de todos los valores de la muestra. ¿Será un estadístico suficiente?

Si trato de usar la definición pero con variables aleatorias continuas, me doy cuenta que resolver ese problema es muy difícil y a veces casi imposible. Entonces buscamos algún

método para poder encontrar estadísticos suficientes sin tener que acudir a la definición.

Teorema de factorización:

Sea \underline{X} un vector aleatorio con función de densidad o función de probabilidad conjunta $f_{\theta}(\underline{x}), \theta \in \Theta$. Entonces el estadístico $T = r(\underline{X})$ es suficiente para θ si y solo si existen dos funciones h y g tal que:

$$f_{\theta}(\underline{x}) = g(r(\underline{x}), \theta)h(\underline{x}) \quad (1.3)$$

Ejemplo: Usando el teorema de factorización, hallar un estadístico suficiente para el parámetro θ de una población $\mathcal{U}(0, \theta)$, basado en una muestra aleatoria de tamaño n .

Teorema: Una familia exponencial a k parámetros tiene como estadísticos suficientes para $\theta \in \mathbb{R}^k$ el vector $T = (r_1(X), r_k(X), \dots, r_k(X))$.

Teorema: Sea X_1, \dots, X_n una muestra aleatoria de una distribución perteneciente a una familia exponencial a k parámetros, entonces el vector aleatorio (X_1, \dots, X_n) tiene una distribución perteneciente a una familia exponencial a k parámetros y el estadístico suficiente para θ basado en la muestra aleatoria es:

$$T = (T_1^*(X), \dots, T_k^*(X)) \text{ donde } T_i^* = \sum_{j=1}^n r_i(X_j).$$

Ejemplo: A partir de la familia exponencial, hallar el estadístico suficiente del parámetro λ de una población Gamma, con una muestra de tamaño n .

1.4.2. Estimadores

Un estimador es un estadístico cuyos valores se consideran medidas experimentales de un parámetro desconocido. Es decir,

Un estimador es una función de la muestra (estadístico) que provee un valor aproximado de un parámetro o característica desconocida.

Lo que Busco es construir estimadores para θ basados en la muestra aleatoria \underline{X}

Dado que un estimador es un valor que se calcula a partir de los datos de una muestra, si tomamos otra muestra su valor va a cambiar. Un estimador es una Variable aleatoria. ¿Cómo podemos construir estas funciones?

1.5. Método de máxima verosimilitud

Es un método para construir estimadores puntuales. Se basa en que, en los experimentos aleatorios, los resultados observados deben tener alta probabilidad de ocurrir.

Definición: Diremos que $\hat{\theta}(X)$ es un Estimador de Máxima Verosimilitud de θ si se cumple

$$f(X, \hat{\theta}) = \max_{\theta} f_{\theta}(x)$$

Es decir, buscamos el valor de θ que hace que la función de verosimilitud $L(\theta)$ sea máxima.

$$\hat{\theta}_n = \arg \max_{\theta} L(\theta)$$

La idea de base es muy simple, y lleva a estimadores con funciones intuitivamente deducibles en la mayoría de los casos. A demás, provee estimadores con características óptimas

Si Θ es un subconjunto abierto tal que el soporte de $f_{\theta}(x)$ no depende de θ (familia regular), como la función logaritmo es monótona creciente, maximizar f es lo mismo que maximizar $\log(f)$. Luego el EMV debe verificar

$$\frac{d}{d\theta} \log L(\theta) = 0$$

Si observamos la estructura de las familias exponenciales, basta con hallar la función que maximiza el exponente de la exponencial.

Ejemplos

1. Hallar el EMV del parámetro p de una población con distribución de Bernoulli
2. Hallar el EMV del parámetro θ de una población con distribución $\mathcal{U}(0, \theta)$
3. Hallar el EMV del parámetro θ de una población con distribución Pareto de parámetros $(2, \theta)$

Principio de invariancia: Supongamos $\lambda = q(\theta)$ es una función biunívoca de θ , si $\hat{\theta}$ es el EMV de θ entonces $\hat{\lambda} = q(\hat{\theta})$ es el EMV de λ

Uno de los grandes problemas de la estadística es construir estimadores razonables para el parámetro desconocido θ . Buscaremos entre ellos ciertas propiedades deseables para la estimación.

1.6. Bondad de los estimadores

$X_1, X_2, \dots, X_n \sim F_\theta(x), \theta \in \Theta$ una muestra aleatoria. Estimamos θ por $\hat{\theta}$. Llamamos función de riesgo (error cuadrático medio):

$$R(\theta, \hat{\theta}) = ECM(\hat{\theta}) = \mathbf{E}_\theta[(\theta - \hat{\theta})^2]$$

Un estimador óptimo para θ será $\hat{\theta}^*$, tal que:

$$ECM(\hat{\theta}^*) \leq ECM(\hat{\theta}), \forall \hat{\theta}$$

Definición: Se dice que $\hat{\theta}$ es un estimador **insesgado** para θ si

$$\mathbf{E}_\theta[\hat{\theta}] = \theta, \forall \theta \in \Theta$$

En caso contrario, el estimador será sesgado, y definimos su sesgo como

$$B(\hat{\theta}) = \mathbf{E}_\theta[\hat{\theta}] - \theta.$$

Será una propiedad deseable para los estimadores que creamos el hecho de que sean insesgados. Y al momento de elegir entre distintos estimadores de un mismo parámetro, elegiremos el que tenga menor error cuadrático medio.

Propiedad: dado un estimador de θ , se tiene que

$$ECM(\hat{\theta}) = \mathbf{var}_\theta(\hat{\theta}) + B(\hat{\theta})^2$$

Observación: Si $B=0$ entonces $ECM(\hat{\theta}) = \mathbf{var}(\hat{\theta})$.

Estimador asintóticamente insesgado

$$\lim_{n \rightarrow \infty} \mathbf{E}_\theta[\hat{\theta}] = \theta$$

Una propiedad deseable para un estimador, es que cuando n es grande la sucesión $\hat{\theta}_n(X_1, \dots, X_n)$ se aproxime en algún sentido al valor que queremos estimar. Para precisar estas ideas introduciremos el concepto de consistencia.

Definición: Dada una sucesión de estimadores $\hat{\theta}_n$ de θ , decimos que $T = \hat{\theta}$ es (débilmente) consistente si para todo $\epsilon > 0$, $\mathbf{P}_\theta(|T - \theta| > \epsilon) \rightarrow 0$

(Lo mínimo que puede pedirse a un estimador es que se aproxime al verdadero valor del parámetro a medida que el tamaño de muestra aumenta)

Teorema: Sea una sucesión de estimadores $\hat{\theta}_n$ de θ . Si $\mathbf{var}(\hat{\theta}) \rightarrow 0$ y $\mathbf{E}_\theta(\hat{\theta}) \rightarrow \theta$ entonces $\hat{\theta}_n$ es débilmente consistente.

Demostración. Usando Markov:

$$\mathbf{P}(|X| \geq t) \leq \frac{\mathbf{E}[h(X)]}{h(t)}$$

$$\mathbf{P}_\theta(|\hat{\theta} - \theta| > \epsilon) \leq \frac{\mathbf{E}[(\hat{\theta} - \theta)^2]}{\epsilon^2} = \frac{\mathbf{var}(\hat{\theta}) + (\mathbf{E}[\hat{\theta}] - \theta)^2}{\epsilon^2} \rightarrow 0$$

Observación: $h(y) = y^2$, $\hat{\theta} - \theta = X$, $\epsilon = t$.

Consistencia en media cuadrática

$$\lim_{n \rightarrow \infty} ECM[\hat{\theta}] = 0, \forall \theta \in \Theta$$

1.7. Estimadores asintóticamente normales

Sea X_1, \dots, X_n una muestra aleatoria de una población con distribución perteneciente a una familia $F_\theta(x)$, $\theta \in \Theta$ (perteneciente además a una familia regular), y sea $\hat{\theta}_n = \hat{\theta}_n(\underline{X})$ un estimador insesgado de θ , se tiene

$$\begin{aligned} \mathbf{E}_\theta[\hat{\theta}_n(\underline{X})] &= q(\theta) \\ \mathbf{var}_\theta[\hat{\theta}_n(\underline{X})] &\geq \frac{q'(\theta)^2}{nI(\theta)} \end{aligned}$$

Definición: Se dice que $\hat{\theta}_n$ es una sucesión de estimadores asintóticamente normal si: $\sqrt{n}(\hat{\theta}_n - q(\theta))$ converge en distribución a una normal con media cero y varianza $(q'(\theta))^2/I(\theta)$

$I(\theta)$ es el número de información de Fisher y se calcula como:

$$I(\theta) = \mathbf{E} \left[\left(\frac{d}{d\theta} \ln f_\theta(X) \right)^2 \right] = -\mathbf{E} \left[\frac{d^2}{d\theta^2} \ln f_\theta(X) \right]$$

¿Para qué nos sirve todo esto? Bueno, la respuesta está en el siguiente teorema:

Teorema : Bajo ciertas condiciones muy generales, sea $\hat{\theta}_n$ un EMV de θ consistente, y sea $q(\theta)$ derivable con $q'(\theta) \neq 0$ para todo θ , entonces $q(\hat{\theta}_n)$ es asintóticamente normal para estimar $q(\theta)$.

Propiedad: Si $\sqrt{nI(\theta)}(\hat{\theta}_n - \theta) \sim^{(a)} \mathcal{N}(0, 1)$ y $\hat{\theta}$ es un estimador consistente para θ , entonces vale que:

$$\sqrt{nI(\hat{\theta})}(\hat{\theta}_n - \theta) \sim^{(a)} \mathcal{N}(0, 1)$$

Método Delta:

Lema: Sea X_1, \dots, X_n una sucesión de variables aleatorias tal que:

$$\sqrt{n}(X_n - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$$

Sea g una función continua y derivable tal que $g'(\mu) \neq 0$ y $g'(\mu)$ continua en μ , entonces:

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{D} \mathcal{N}(0, g'(\mu)^2 \sigma^2)$$

Ejemplo: Hallar la distribución asintótica del EMV del parámetro p de una población con distribución bernoulli. Nuestro ejemplo 1

1.8. Algunas distribuciones importantes

Distribución Chi-cuadrado

La variable aleatoria X tiene una distribución Chi-cuadrada de parámetro ν (grados de libertad) si su función de densidad está dada por:

$$f_X(x) = \frac{(1/2)^{\nu/2} x^{\nu/2-1} e^{-x/2}}{\Gamma(\nu/2)} \mathbf{1}\{x > 0\}$$

Se puede observar que la variable Chi-cuadrado coincide con una variable con distribución gamma de parámetros $\lambda = 0,5$ y $k = \nu/2$. Se deduce fácilmente que $\mathbf{E}(X) = \nu$, y $\mathbf{var}(X) = 2\nu$.

Mediante un cambio de variable (ejercicio 2.10) se puede demostrar que si Z es una V.A. $\mathcal{N}(0, 1)$, la variable $Y = Z^2$ tiene distribución Chi-cuadrada con 1 grado de libertad.

Teorema: Si X_1, \dots, X_n son variables aleatorias independientes con distribución chi-cuadrada de ν_1, \dots, ν_n grados de libertad respectivamente, entonces la V.A. $Y = X_1 + X_2 + \dots + X_n$ tiene distribución chi-cuadrada con $\nu = \nu_1 + \dots + \nu_n$ (se puede probar fácilmente viendo a la chi como una gamma).

Corolario: Se llama distribución Chi-cuadrado con $\nu = n$ grados de libertad χ_n^2 a la distribución de $\sum_{i=1}^n Z_i$ donde Z_i son variables normales estándar independientes.

Distribución t de Student

Sean $Z \sim \mathcal{N}(0, 1)$ y $U \sim \chi_n^2$ entonces, si Z y U son independientes, $T = \frac{Z}{\sqrt{U/n}}$ tiene distribución t de Student con n grados de libertad.

Distribución F de Fisher

Sean U y V dos variables aleatorias independientes con distribución Chi-cuadrada de ν_1 y ν_2 grados de libertad respectivamente, entonces $F = \frac{U/\nu_1}{V/\nu_2}$ tiene distribución F de Fisher con ν_1 y ν_2 grados de libertad.

Teorema: Sea (X_1, \dots, X_n) una muestra aleatoria de una población con distribución $\mathcal{N}(\mu, \sigma^2)$.

1. $Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim \mathcal{N}(0, 1)$
2. $W = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$
3. W y Z son independientes
4. Si $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$ entonces $T = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$

Capítulo 2

Test de hipótesis

Una hipótesis es una suposición o conjetura sobre la naturaleza, cuyo valor de verdad o falsedad no se conoce. Una hipótesis estadística es una hipótesis sobre una o más poblaciones estadísticas o sobre un fenómeno aleatorio.

Se llama hipótesis estadística a cualquier suposición o enunciado provisorio que se realiza sobre una población o fenómeno aleatorio, cuyo valor de verdad o falsedad no se conoce.

La hipótesis estadística se refiere a la población. Sin embargo el análisis para tratar de comprobar la verdad o falsedad de la hipótesis se realizará sobre una muestra.

Se llama **ensayo de hipótesis** al procedimiento que se sigue para tratar de averiguar si una hipótesis es verdadera o falsa.

Veamos un ejemplo para entender que significa. Nos centramos en el ejercicio 1.17 (para que veas que todo está relacionado). En ese ejercicio teníamos que una planta de ensamblaje recibe una partida de 25 piezas de precisión que incluye exactamente k defectuosas. Al parecer, revisar todas esas piezas era muy dificultoso, por lo que la división de control de calidad elegía 5 piezas al azar para controlarlas.

Sea X en número de piezas defectuosas en la muestra. Entonces, si k es la cantidad de defectuosas en la partida, tenemos que la distribución de X es hipergeométrica, y su función de probabilidad está dada por

$$p_k(x) = \frac{\binom{k}{x} \binom{25-k}{5-x}}{\binom{25}{5}}, \max(0, k-20) \leq x \leq \min(5, k)$$

y k puede tomar valores en el conjunto $\Theta = \{0, 1, 2, \dots, 25\}$.

supongamos que se hubiese convenido con el vendedor que la partida debía tener no

mas de k_0 piezas defectuosas. Entonces, en base a la variable X , que es la que observa el control de calidad, debe decidir si la partida cumple con los requisitos convenidos. Tendrán que decidir entre dos alternativas:

$$k \in \Theta_1 = \{0, 1, \dots, k_0\} \quad o \quad k \in \Theta_2 = \{k_0 + 1, \dots, 25\}$$

Esto puede ser expresado como que el área de control de calidad debe decidir entre dos hipótesis

$$H_0 : k \in \Theta_1 \quad Vs. \quad H_1 : k \in \Theta_2$$

Y esta decisión debe hacerse sobre los valores observados de X .

Generalizando, en un ensayo de hipótesis se plantean dos hipótesis:

- H_1 : El investigador suele tener una hipótesis de investigación que es la predicción que se deriva de la teoría que está analizando o que sustenta su trabajo. Esta hipótesis de investigación, cuando se formula con una proposición operacional, se llama hipótesis alternativa
- H_0 : Se llama hipótesis nula a la hipótesis objeto del ensayo. Esta se formula con el propósito expreso de ser rechazada. Un ensayo puede rechazar o no la hipótesis nula. Cuando el ensayo no rechaza, diremos que no hay pruebas concluyentes en contra de la hipótesis, pero no quedará probado que la hipótesis sea verdadera. En cambio si se rechaza la hipótesis nula quedará probado que es falsa, y se concluirá que la hipótesis alternativa es correcta. (La conclusión fuerte en un ensayo de hipótesis únicamente se tiene en el rechazo de la hipótesis nula)

Comienzos: Supongamos que nos encontramos en un juicio, donde una persona será sometida a juicio para decidir si enviarle o no a la cárcel por un crimen. Al comienzo se cuenta con dos hipótesis: la persona es culpable o es inocente. Para tomar una decisión se debe encontrar evidencia suficiente que respalde dicha decisión. La presunción de inocencia que goza el acusado es la hipótesis nula que ensaya el juez. Si el juez encuentra evidencia suficiente de que la hipótesis nula es falsa, entonces rechaza la hipótesis y decide enviarle a la cárcel. Ahora, al momento de tomar una decisión podría llegar a cometer distintos errores: Enviar a la cárcel a un inocente o dejar libre a un culpable.

Definición:

Existen dos tipos de errores que pueden cometerse al tomar una decisión:

- Error de tipo I: Es el error que se comete cuando se rechaza una hipótesis nula que en realidad era verdadera. Se trata de que este suceso tenga muy baja probabilidad.

- Error de tipo II: Es el error que se comete cuando no se rechaza una hipótesis nula que en realidad es falsa.

La teoría clásica de test de hipótesis considera que el error de tipo 1 es mucho más grave que el error de tipo 2. Es decir, la situación de las hipótesis H_0 y H_1 no es simétrica; es mucho más grave rechazar H_0 cuando es cierta que aceptarla cuando es falsa. Esto significa que se debe tener mucha evidencia sobre que H_0 es falsa antes de rechazarla. Se observa en el Ejemplo, donde es más grave enviar a la cárcel a un inocente que dejar libre a un culpable.

Ahora bien, el Juez también podría tomar una decisión correcta.

La **potencia de un test** es la probabilidad de rechazar una hipótesis nula en función del parámetro desconocido sobre el cual se plantea la hipótesis.

Un **test de hipótesis** es una regla de decisión entre H_0 y H_1 y la expresamos como una función de la muestra aleatoria $\delta(\underline{X})$ que puede tomar los valores 0 o 1. Si $\delta(\underline{X}) = 1$ se rechaza H_0 , y en caso contrario no se rechaza.

Volvamos a nuestro ejemplo de las piezas y el lote. El enunciado del ejercicio 1.17 decía que se rechazaba el lote si encontraban 1 o más piezas defectuosas en las 5 revisadas. El Test de hipótesis es la regla de decisión basada en X . Entonces en este ejercicio, un posible test está dado por

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } X > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Y de acuerdo con este test, se rechaza la partida (se devuelve), o sea que se decide que $k \in \Theta_2$ si se observa en la muestra 1 o más piezas defectuosas. Este test no es único. Por ahora no tenemos ningún criterio para elegir entre dos tests, ni entre los muchos otros que podrían definirse. Trataremos de definir criterios para encontrar tests óptimos.

Definición: Sea $\underline{X} = (X_1, \dots, X_n)$ una muestra aleatoria de una población con distribución $F_\theta(x)$, $\theta \in \Theta$. Sean Θ_1 y Θ_2 tal que $\Theta_1 \cup \Theta_2 = \Theta$ y $\Theta_1 \cap \Theta_2 = \emptyset$. Un test para este problema es una regla de decisión basada en \underline{X} para decidir entre 2 hipótesis:

$$H_0 : \theta \in \Theta_1 \quad Vs. \quad H_1 : \theta \in \Theta_2$$

Entonces, de acuerdo a lo definido antes,

$$\mathbf{P}(\text{error I}) = \mathbf{P}_\theta(\text{rechazar } H_0), \text{ con } \theta \in \Theta_1$$

$$\mathbf{P}(\text{error II}) = \mathbf{P}_\theta(\text{NO rechazar } H_0), \text{ con } \theta \in \Theta_2$$

La potencia del test es una función del parámetro desconocido:

$$\pi_{\delta}(\theta) = \mathbf{P}_{\theta}(\text{Rechazar } H_0) = \mathbf{P}_{\theta}(\delta(\underline{X}) = 1) = E_{\theta}(\delta(\underline{X}))$$

Porque la función $\delta(\underline{X})$ es una variable que vale 0 o 1, su esperanza será la probabilidad de que valga 1.

Se puede observar que:

$$\mathbf{P}(\text{error I}) = \pi_{\delta}(\theta), \text{ con } \theta \in \Theta_1$$

$$\mathbf{P}(\text{error II}) = 1 - \pi_{\delta}(\theta), \text{ con } \theta \in \Theta_2$$

Un buen test deberá tener errores de tipo 1 y 2 pequeños, y por lo tanto debe tener una función de potencia $\pi_{\delta}(\theta)$ que tome valores cercanos a 0 para $\theta \in \Theta_1$ y valores cercanos a 1 $\theta \in \Theta_2$.

En realidad, no podemos hacer ambos errores pequeños simultáneamente. Más aún, para un tamaño de muestra dado, para que decrezca la probabilidad de que ocurra un error de tipo 1, debemos aumentar la probabilidad de que ocurra un error de tipo 2 (o sea disminuir la potencia). Si queremos que ambos sean pequeños debemos aumentar la cantidad de observaciones.

Se llama **Nivel de significación del test** a la máxima probabilidad de cometer un error de tipo I, es decir:

$$\alpha = \sup_{\theta \in \Theta_1} \pi_{\delta}(\theta)$$

Se llama **p-valor** de un test al menor nivel de significación para el cual se rechaza H_0 para una observación dada. Es la probabilidad de encontrar un valor tan o mas extremo que el que se encontró con la muestra observada. Esta definición es muy importante pero podremos entenderla mucho mejor cuando hagamos un ejemplo y podamos graficar lo que significa

Llamaremos $\beta(\theta) = \mathbf{P}_{\theta}(\text{Error II}) = 1 - \pi_{\delta}(\theta)$, con $\theta \in \Theta_2$, que en muchos casos se conoce como la curva característica operativa.

Definición: Diremos que δ es un test uniformemente más potente, de nivel menor o igual que α para $H_0 : \theta \in \Theta_1$ Vs. $H_1 : \theta \in \Theta_2$ si es el más potente de nivel menor o igual que α para todo $\theta_2 \in \Theta_2$, es decir, si el mismo test es óptimo cualquiera sea la alternativa fija considerada. Lo ideal sería encontrar (cuando existan) tests uniformemente más potentes de nivel menor o igual que α . Estudiaremos casos donde estos tests existen y otros

donde no. En estos últimos habrá que elegir otros criterios para seleccionar el test a usar.

Ejemplo: La producción diaria de cierta empresa es aleatoria con distribución $\mathcal{N}(10; 1,5^2)$. Se propone comprar una maquina que promete una distribución $\mathcal{N}(11; 1,5^2)$. Plantear un las hipótesis correspondientes al problema y el test que permita decidir.

¿cómo encuentro mi regla de decisión (test)?

Idea: Necesito decidir si esta nueva máquina es igual a la que tengo o realmente produce con una media mejor (igual a 11), solamente si encuentro evidencia de que es mejor es cuando voy a gastar mucha plata y comprarla, esa es la decisión fuerte que debo tomar, por lo tanto las hipótesis que debo plantear son:

Observación: las hipótesis siempre se plantean sobre un parámetro desconocido del cual nunca voy a conocer su verdadero valor.

Un criterio de decisión puede ser: Tomo una muestra de 16 elementos con la nueva máquina, Si lo que observo tiene mayor probabilidad de ocurrir cuando la media es 10 entonces no compraría la maquina (no rechazo $H_0 : \mu = 10$) pero si tiene mayor probabilidad de ocurrir cuando la media es 11 entonces debe ser que la maquina nueva es mejor, y eso debería llevarme a decidir comprarla (rechazar H_0).Entonces rechazaría H_0 cuando

$$T = \frac{f_{\mu=11}(X)}{f_{\mu=10}(X)} > k_\alpha$$

donde k_α es una constante que depende del nivel de significación. Fijado el valor de α , el test será:

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } \frac{f_{\mu=11}(X)}{f_{\mu=10}(X)} > k_\alpha \\ 0 & \text{en otro caso} \end{cases}$$

De manera que cumpla:

$$\mathbf{E}_{\mu=10}[\delta(\underline{X})] = \alpha$$

Un test de esta forma se llama test del cociente de verosimilitud.

2.1. Test para hipótesis simple vs. hipótesis simple

El caso más simple de problema de test de hipótesis es la situación donde Θ_1 y Θ_2 contengan cada uno un elemento. En este caso, se dice que las hipótesis son simples:

$$H_0 : \theta = \theta_1 \quad Vs. \quad H_1 : \theta = \theta_2$$

Por lo tanto, parece razonable plantear la regla de decisión:

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } \frac{f_{\theta_2}(X)}{f_{\theta_1}(X)} > k_\alpha \\ 0 & \text{en otro caso} \end{cases}$$

Que para un nivel α dado, debemos hallar el valor de k_α que cumpla

$$\mathbf{E}_{\theta_1}[\delta(\underline{X})] = \alpha = \mathbf{P}_{\theta_1}(\delta(\underline{X}) = 1)$$

El teorema de Newman-Pearson dice que el test de arriba es el uniformemente más potente para hipótesis simples. Desarrollar este teorema ya excede lo que buscamos en la materia, pero puede buscarse en la bibliografía.

Volviendo al ejemplo. Si las poblaciones son normales, con igual varianza, es decir que $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ independientes,

$$H_0 : \mu = \mu_1 \quad Vs. \quad H_1 : \mu = \mu_2$$

entonces al hacer el cociente de verosimilitud y despejando, si suponemos $\mu_1 < \mu_2$, resulta la siguiente regla de decisión ([Desarrollo en el video.](#)):

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } \sum_{i=1}^n X_i > k'_\alpha \\ 0 & \text{si no} \end{cases}$$

Para hallar el valor de k' uso la segunda condición:

$$\alpha = \mathbf{P}_{\mu_1}(\delta(\underline{X}) = 1) = \mathbf{P}_{\mu_1} \left(\sum_{i=1}^n X_i > k' \right) = \mathbf{P}_{\mu_1} \left(\frac{\sum_{i=1}^n X_i - n\mu_1}{\sqrt{n\sigma^2}} > z_{1-\alpha} \right)$$

Resultando el test que cumple con las condiciones pedidas:

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } \frac{\sum_{i=1}^n X_i - n\mu_1}{\sqrt{n\sigma^2}} > z_{1-\alpha} \\ 0 & \text{si no} \end{cases}$$

De esta manera observamos que la regla de decisión depende de α , n , y el valor que toma el parámetro bajo H_0 (es decir, cuando H_0 es verdadera). La regla de decisión no depende del valor que pusimos en H_1 , lo único que afecta en su definición es que el valor que elegimos en H_1 es mayor que el que pusimos en H_0 (si hubiese sido menor, al despejar

la desigualdad se hubiese invertido). Por lo tanto la regla de decisión hubiese sido la misma para una alternativa $H_1 : \mu > \mu_1$. Se observa también que la variable aleatoria de la cual depende el test es el estadístico suficiente para μ . A la función de la muestra aleatoria de la cual depende el test la llamamos Estadístico del test

Estudiemos ahora la función potencia en este caso ($\mu_1 < \mu_2$).

$$\begin{aligned}\pi_\delta(\mu) &= \mathbf{P}_\mu(\delta(\underline{X}) = 1) = \mathbf{P}\left(\frac{\sum_{i=1}^n X_i - n\mu_1}{\sqrt{n\sigma^2}} > z_{1-\alpha}\right) \\ &= \mathbf{P}_\mu\left(\sum_{i=1}^n X_i > z_{1-\alpha}\sqrt{n\sigma^2} + n\mu_1\right) \\ &= 1 - \Phi\left(z_{1-\alpha} - \frac{\sqrt{n}}{\sigma}(\mu - \mu_1)\right)\end{aligned}$$

Propiedades de la función potencia:

- Si μ aumenta entonces $\pi_\delta(\mu)$ aumenta
- $\pi_\delta(\mu)$ es creciente
- $\pi_\delta(\mu_1) = \alpha$
- $\lim_{\mu \rightarrow \infty} \pi_\delta(\mu) = 1, \lim_{\mu \rightarrow -\infty} \pi_\delta(\mu) = 0$

A partir de este análisis es fácil observar que el mismo test vale para

$$H_0 : \mu \leq \mu_1 \quad Vs. \quad H_1 : \mu > \mu_1$$

Otro ejemplo teórico interesante para hacer en clase o practicar en casa: Si se tiene una población con distribución exponencial λ , hallar el test para

$$H_0 : \lambda = \lambda_1 \quad Vs. \quad H_1 : \lambda = \lambda_2$$

basado en una muestra aleatoria de tamaño n .

Ejemplo: Supongamos que se mide el grado de impurezas de un producto químico. El método de medición está afectado por un error que se supone $\mathcal{N}(0, 0.01)$. Además los errores correspondientes a diferentes mediciones son independientes entre sí. Se sabe que el producto es aceptable si el grado de impurezas es menor a 0.7. Se hacen 64 mediciones y se quiere encontrar un test de modo que la probabilidad de aceptar el producto cuando este no satisfaga las condiciones sea menor que 0.05. Luego, si la muestra observada arrojó una suma de 47.71, ¿qué decisión toma?

Desarrollo en el video 1 del capítulo 10.

Se puede demostrar (pero no lo haremos en este curso) que el test de cociente de verosimilitud presentado es el test uniformemente mas potente para hipótesis unilaterales (no solo para hipótesis simples).

2.2. Tests uniformemente mas potentes para hipótesis unilaterales

En la sección anterior vimos como crear un test de hipótesis cuando tenemos hipótesis simples, y también vimos como ese test nos servía en el caso de tener hipótesis unilaterales (esto es por ejemplo cuando tenemos $H_0 : \theta \leq \theta_1$ Vs. $H_1 : \theta > \theta_1$) Para encontrar dicho test, planteábamos el cociente de verosimilitudes, y a partir del dato del valor de α podíamos encontrar la regla de decisión. Observamos en los ejemplos que el test nos quedaba en función del estadístico suficiente.

Definición: Una familia de distribuciones discretas o continuas con función $f_\theta(x), \theta \in \Theta \subset \mathbb{R}$ se dice de cociente de verosimilitud monótono (CVM) en el estadístico $T = r(\underline{X})$ donde toma valores reales, si para todo par $\theta_1 < \theta_2$ se tiene:

1. Las distribuciones correspondientes a $f_{\theta_1}(x)$ y $f_{\theta_2}(x)$ son distintas
2. $f_{\theta_2}(x)/f_{\theta_1}(x) = g(t)$, donde $g(t)$ es una función no decreciente en el conjunto $S = \{t : t = t(x) \text{ con } f_{\theta_2}(x) > 0 \text{ o } f_{\theta_1}(x) > 0\}$

Propiedad: Las familias exponenciales a un parámetro con $c(\theta)$ estrictamente monótona son de CVM.

1. Si $c(\theta)$ es estrictamente creciente, la familia dada es de CVM en $T(X)$
2. Si $c(\theta)$ es estrictamente decreciente, la familia dada es de CVM en $-T(X)$

Teorema: Sea \underline{X} un vector aleatorio con función de densidad o probabilidad perteneciente a la familia $f_\theta(x), \theta \in \Theta \subset \mathbb{R}$, que tiene la propiedad de ser CVM en $T = r(\underline{X})$, luego:
El test para

$$H_0 : \theta \leq \theta_1 \quad \text{Vs.} \quad H_1 : \theta > \theta_1$$

Será:

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } T > k_\alpha \\ 0 & \text{si } T \leq k_\alpha \end{cases}$$

2.2. TESTS UNIFORMEMENTE MAS POTENTES PARA HIPÓTESIS UNILATERALES 19

Que para, un nivel α dado, se tendrá:

$$\alpha = \mathbf{P}_{\theta_1}(\delta(\underline{X}) = 1) \quad (2.1)$$

Propiedad

Sea \underline{X} una muestra aleatoria con distribución perteneciente a una familia exponencial, luego:

1. Si $c(\theta)$ es creciente, el test para

$$H_0 : \theta \leq \theta_1 \quad V s. \quad H_1 : \theta > \theta_1$$

Será:

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } T > k_\alpha \\ 0 & \text{si } T \leq k_\alpha \end{cases}$$

Que para, un nivel α dado, se tendrá:

$$\alpha = \mathbf{P}_{\theta_1}(\delta(\underline{X}) = 1) \quad (2.2)$$

2. Si $c(\theta)$ es decreciente, el test para

$$H_0 : \theta \leq \theta_1 \quad V s. \quad H_1 : \theta > \theta_1$$

Será:

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } -T > k_\alpha \\ 0 & \text{si } -T \leq k_\alpha \end{cases}$$

Que para, un nivel α dado, se tendrá:

$$\alpha = \mathbf{P}_{\theta_1}(\delta(\underline{X}) = 1) \quad (2.3)$$

Ejemplo: Test de hipótesis para la varianza de una población $\mathcal{N}(\mu, \sigma^2)$ donde μ es un valor conocido. Entonces

$$H_0 : \sigma^2 \leq s \quad V s. \quad H_1 : \sigma^2 > s$$

Como estamos frente a una familia exponencial, sabemos que $T = \sum_{i=1}^n (X_i - \mu)^2$ y que $c(\sigma^2) = -\frac{1}{2\sigma^2}$, la cual es una función creciente, entonces el test correspondiente a las hipótesis planteadas será:

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } \sum_{i=1}^n (X_i - \mu)^2 > k_\alpha \\ 0 & \text{si no} \end{cases}$$

A demás $\sum_{i=1}^n (X_i - \mu)^2 / \sigma^2 \sim \chi_n^2$, entonces el test de nivel alfa resulta:

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } \frac{\sum_{i=1}^n (X_i - \mu)^2}{s} > \chi_{n,1-\alpha}^2 \\ 0 & \text{si no} \end{cases}$$

Ejemplo 2: Desarrollemos el caso uniforme

La estructura que estamos observando es la siguiente: Tenemos un juego de hipótesis sobre un parámetro desconocido, y luego una regla de decisión basada en la muestra aleatoria. Esta regla es función de un estadístico (función de la muestra aleatoria) y de un cuantil que depende de α y de la distribución del estadístico. Esos son los elementos que componen el ensayo de hipótesis. Los test no son únicos, pero que haya varios no significa que sean todos igual de buenos. Las herramientas vistas hasta ahora sirven para construir lo que se llama *tests uniformemente más potentes*, pero no siempre voy a poder conseguirlos, ya sea porque no conozca la distribución del estadístico, o porque la muestra que tenga no me lo permita, o simplemente porque no exista. Una técnica muy útil para crear test se llama test del cociente de máxima verosimilitud, que sirve para crear test en situaciones más complejas de las que tratamos hasta ahora.

2.3. Test del cociente de máxima verosimilitud

Dada una muestra aleatoria \underline{X} , perteneciente a una población con distribución $f_\theta(x)$, se quiere testear:

$$H_0 : \theta \in \Theta_1 \quad Vs. \quad H_1 : \theta \in \Theta_2$$

Podemos pensar que rechazaríamos H_0 si el valor mas probable de Θ_2 tiene probabilidad considerablemente más grande que el valor más probable de Θ_1

El test de CMV para estas hipótesis será:

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } L < K_\alpha \\ 0 & \text{si no} \end{cases}$$

Donde:

$$L = \frac{\sup_{\theta \in \Theta_1} f_\theta(\underline{X})}{\sup_{\theta \in \Theta_2} f_\theta(X^{(n)})} \quad (2.4)$$

Si la dimensión de Θ_1 es menor que la dimensión de Θ , entonces es equivalente plantear un test en función de L^* , con

$$L^* = \frac{\sup_{\theta \in \Theta_1} f_\theta(\underline{X})}{\sup_{\theta \in \Theta} f_\theta(X^{(n)})}$$

Que en general resulta mucho más fácil.

Ejemplo: Sea X_1, \dots, X_n una muestra aleatoria de una población con distribución $\mathcal{N}(\mu, \sigma^2)$, hallar el test para

$$H_0 : \mu = \mu_1 \quad Vs. \quad H_1 : \mu \neq \mu_1$$

Desarrollo completo en el video 2

Versión resumida:

Sea X_1, \dots, X_n una muestra aleatoria de una población con distribución $\mathcal{N}(\mu, \sigma^2)$, hallar el test para

$$H_0 : \mu = \mu_1 \quad Vs. \quad H_1 : \mu \neq \mu_1$$

Proponemos

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } T < K_1 \quad o \quad T > K_2 \\ 0 & \text{si no} \end{cases}$$

Con estadístico del test $T = \sqrt{n} \frac{\bar{X} - \mu_1}{S}$, ya que bajo H_0 , $T \sim t_{n-1}$.

Entonces,

$$\alpha = \mathbf{P}_{\mu_1}(\delta(\underline{X}) = 1) = \mathbf{P}_{\mu_1}(T < K_1 \quad o \quad T > K_2) \quad (2.5)$$

Como la distribución T es simétrica con respecto al cero, entonces resulta

$$\alpha = \mathbf{P}_{\mu_1}(|\sqrt{n} \frac{\bar{X} - \mu_1}{S}| > k) \quad (2.6)$$

con $k = t_{n-1, 1-\alpha}$

A partir de este resultado pueden deducirse los test unilaterales para el caso de poblaciones normales con varianza desconocida.

2.4. Test con nivel de significación asintótico

Para la mayoría de los test de hipótesis se requiere para encontrar k_α la distribución del estadístico T para $\theta \in \Theta_1$. Como en muchos casos esta distribución es muy compleja, se puede reemplazar por una asintótica. En este caso el test tendrá un nivel de significación aproximado al deseado para muestras grandes.

Definición: Sea $\underline{X} = (X_1, \dots, X_n)$ una muestra aleatoria de una población con distribución $F_\theta(x), \theta \in \Theta$. Se quiere testear:

$$H_0 : \theta \in \Theta_1 \quad Vs. \quad H_1 : \theta \in \Theta_2$$

Se dirá que una sucesión de test tiene nivel de significación asintótico alfa si:

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta_1} \pi_\delta(\theta) = \alpha$$

Ejemplo Test de hipótesis para $H_0 : p \in \Theta_1 \quad Vs. \quad H_1 : p \in \Theta_2$, cuando la población tiene distribución Bernoulli de parámetro p .

2.4.1. Distribución asintótica del test del cociente de máxima verosimilitud

Definición: Sea $\underline{X} = (X_1, \dots, X_n)$ una muestra aleatoria de una población con distribución $F_\theta(x), \theta \in \Theta$. Se quiere testear:

$$H_0 : \theta \in \Theta_1 \quad Vs. \quad H_1 : \theta \in \Theta_2$$

Supongamos que Θ es un abierto en \mathbb{R}^p y supongamos que Θ_1 tiene dimensión j -p, entonces bajo H_0 , $-2\log L^* \sim \chi_j^2$, y el test de nivel asintótico α resulta:

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } -2\log L^* \geq \chi_{j,1-\alpha}^2 \\ 0 & \text{si no} \end{cases}$$

Capítulo 3

Intervalos de confianza

En el capítulo 9 nos focalizamos en encontrar estimadores para parámetros desconocidos. Una vez obtenida la muestra, podíamos evaluar esos estimadores y conseguir de esa forma una estimación puntual de dichos parámetros. ¿Que encontramos por ejemplo?

- Que si tenemos una muestra aleatoria de una población con distribución $\mathcal{N}(\mu, \sigma^2)$, los estimadores de máxima verosimilitud de cada parámetro eran $\hat{\mu} = \bar{X}$ y $\hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$, pero este último no es insesgado por lo que en general usamos como estimador para la varianza a S^2
- Que si tenemos una muestra aleatoria de una población con distribución $\mathcal{B}(p)$, el EMV también resulta el promedio de la muestra. (¿sorprende?)
- Que si tenemos una muestra aleatoria de una población con distribución $\mathcal{E}(\lambda)$, el EMV resulta la inversa del promedio, resultado muy coherente si pensamos en todo lo que aprendimos hasta ahora

Una vez que evaluamos el estimador en la muestra, nos da un numero que es una estimación del valor desconocido, pero bien sabemos que ese valor no es el verdadero valor del parámetro. Es más, si tomamos otra muestra ese valor va a cambiar. Ahora, a mi me gustaría tener cierta certeza en mi estimación, o al menos me gustaría tener una idea del error o cuan lejos de verdadero valor podría llegar a estar.

La idea entonces será encontrar un intervalo que contenga al parámetro desconocido con una alta probabilidad (lamentablemente, la certeza absoluta no la tendremos). Este intervalo será aleatorio, y luego para cada muestra podremos evaluarlo.

3.1. Intervalos de confianza de nivel exacto $1 - \alpha$

Definición: Dada X_1, X_2, \dots, X_n una m.a. de una población con distribución $F_\theta(x), \theta \in \Theta$, un intervalo de confianza para θ de nivel $1 - \alpha$ es un intervalo de extremos aleatorios $[a(\underline{X}), b(\underline{X})]$ tal que

$$\mathbf{P}(\theta \in [a(\underline{X}), b(\underline{X})]) = 1 - \alpha$$

A $1 - \alpha$ se lo llama nivel de confianza, y α nivel de riesgo.

Definición: Dada X_1, X_2, \dots, X_n una m.a. de una población con distribución $F_\theta(x), \theta \in \Theta$, una región de confianza para θ de nivel $1 - \alpha$ es un conjunto $S(X)$ tal que $\mathbf{P}(\theta \in S(X)) = 1 - \alpha$

Ejemplo: Intervalo de confianza para la media μ de una población normal con desvío σ conocido

3.1.1. Método del pivote

Dada X_1, X_2, \dots, X_n una m.a. de una población con distribución $F_\theta(x), \theta \in \Theta$ y sea $U = r(\underline{X}, \theta)$ una variable aleatoria cuya distribución no dependa de θ . Sean A y B tales que $\mathbf{P}(A < U < B) = 1 - \alpha$.

Luego la región $S(\underline{X}) = \{\theta : A < r(\underline{X}, \theta) < B\}$: es una región de confianza de nivel $1 - \alpha$ para θ (a U se lo llama **Pivote**).

Observación: un pivote NO es un estadístico, porque depende de parámetros desconocidos y nunca vas a poder evaluarlo y conocer su valor solo con la muestra.

Ejemplo: Intervalo de confianza para la media μ de una población normal con desvío σ desconocido

Ahora, muchas veces logramos encontrar un pivote pero no conocemos su distribución, lo que si podremos conocer es su distribución asintótica. Para esos casos, definimos lo que sigue.

3.2. Intervalos de confianza de nivel asintótico $1 - \alpha$

Def: X_1, X_2, \dots, X_n una m.a. de una población con distribución $F_\theta(x), \theta \in \Theta$. Se dice que $S_n(X)$ es una sucesión de regiones de confianza para θ de nivel asintótico $1 - \alpha$ si

$$\lim_{n \rightarrow \infty} \mathbf{P}(\theta \in S_n(X)) = 1 - \alpha, \quad \theta \in \Theta$$

Teorema: X_1, X_2, \dots, X_n una m.a. de una población con distribución $F_\theta(x), \theta \in \Theta$. Supongamos que para cada n se tiene una variable aleatoria $U_n = r_n(X, \theta)$, donde U es una variable aleatoria cuya distribución no depende de θ . Entonces si A y B son tales que $\mathbf{P}(A < U < B) = 1 - \alpha$ se tiene que:
 $S_n(X) = \{\theta : A < r_n(X, \theta) < B\}$ es una región de confianza de nivel asintótico $1 - \alpha$ para θ .

Lo utilizaremos para hallar intervalos de confianza basados en estimadores de máxima verosimilitud, ya que gracias al capítulo 9 conocemos su distribución asintótica.

Ejemplo: Intervalo de confianza para los parámetros de una población de bernoulli y una población de Pareto.

Ocurre que algunas veces no podrás encontrar el pivote. Para esos casos es muy útil entender la relación entre regiones de confianza y test.

3.3. Relación entre regiones de confianza y test

Sea X una variable aleatoria con distribución perteneciente a la familia $F_\theta(x)$ y que para cada θ_0 se tiene un test de nivel α , para $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. Se puede definir una región de confianza de nivel $1 - \alpha$ para θ como

$$S(\underline{X}) = \{\theta : \delta(\underline{X}) = 0\}$$

Es el conjunto de todos los $\theta \in \Theta$ tales que la hipótesis de que el valor verdadero es θ no es rechazada se observa \underline{X} , S es una región de confianza de nivel $1 - \alpha$. Esto quiere decir que para construir intervalos de confianza siempre podemos conseguirlo si “invertimos” la región de aceptación del test para hipótesis bilaterales.

Por otro lado, si se tiene una región de confianza de nivel $1 - \alpha$, $S(X)$, para θ se puede construir un test de nivel α :

$$\delta(\underline{X}) = \begin{cases} 1 & \text{si } \theta_0 \notin S(X) \\ 0 & \text{si } \theta_0 \in S(X) \end{cases}$$

Esto último es usado para comparaciones de parámetros, uno podrá decir que dos medias son significativamente diferentes, si el intervalo de confianza para la diferencia de medias no contiene al valor cero. No siempre puede construirse un test a partir de un intervalo, pero siempre puede construirse un intervalo a partir de un test, invirtiendo la región de aceptación.

Capítulo 4

Estimación Bayesiana

Lo estudiado hasta ahora se conoce como el enfoque frecuentista, en el que buscamos estimar un parámetro θ del cual no teníamos ningún conocimiento previo. Este parámetro lo suponíamos como una constante desconocida a la cual buscamos estimar mediante muestras.

Consideremos nuevamente un problema estadístico de estimación paramétrico. Vamos a tener un vector aleatorio $\underline{X} = (X_1, \dots, X_n)$, que puede ser, por ejemplo, aunque no necesariamente, una muestra aleatoria de cierta distribución.

El enfoque Bayesiano, a diferencia del frecuentista, supone que se tiene alguna información previa sobre θ , expresada por medio de una distribución sobre θ , es decir que considera al parámetro desconocido como una variable aleatoria, con función de densidad (o probabilidad). La variable aleatoria la llamaremos Θ .

Esta información previa, denominada distribución a priori, tiene una densidad $f_{\Theta}(\theta)$ si se trata de una variable continua, o $p_{\Theta}(\theta)$ si se trata de una variable discreta. La distribución a priori puede tener distintas interpretaciones según el problema. Se pueden dar las siguientes alternativas

- La distribución a priori está basada en experiencias previas similares.
- La distribución a priori expresa una creencia subjetiva.

Para trabajar en este capítulo, usaremos todo lo aprendido en el capítulo 5, porque lo que buscaremos es que ocurre con la distribución de Θ una vez observada la muestra aleatoria.

Una vez observada la muestra, podremos calcular la distribución que llamamos *a posteriori* de Θ dada $\underline{X} = \underline{x}$ usando la fórmula de Bayes (por ejemplo si las variables aleatorias son continuas) de la siguiente forma:

$$f_{\Theta|\underline{X}=\underline{x}}(\theta) = \frac{f_{\underline{X}|\Theta=\theta}(\underline{x}) * f_{\Theta}(\theta)}{\int_{-\infty}^{\infty} f_{\underline{X}|\Theta=\theta}(\underline{x}) * f_{\Theta}(\theta) d\theta}$$

Si las variables son discretas usaremos las funciones de probabilidad, y si tenemos el caso de una discreta y una continua, tal como lo aprendimos en el capítulo 5, usaremos mezclas de variables aleatorias y Bayes para mezclas.

Observación: Si Θ es una variable aleatoria discreta, en lugar de integrar debemos sumar. Notar que el denominador corresponde a la distribución marginal de \underline{X} .

Resumiendo: Tendremos como dato la distribución a priori de Θ , la distribución de $\underline{X}|\Theta = \theta$, y buscaremos la distribución de $\Theta|\underline{X} = \underline{x}$ con todo lo que aprendimos en el capítulo 5, nada nuevo para aprender más que el significado estadístico.

Ahora, a partir de todo este desarrollo, no olvidemos que lo que nos interesa es buscar estimadores para nuestros parámetros desconocidos.

Una de las ventajas de este método es que se pueden definir en forma natural estimadores óptimos.

Supongamos que queremos estimar θ . Si consideramos una función pérdida $l(\theta, d)$, que equivale al costo de estimar θ con d , si $\hat{\theta} = \varphi(\underline{X})$, entonces la pérdida será una variable aleatoria $l(\Theta, \varphi(\underline{X}))$.

La pérdida esperada es lo que se llama *riesgo de Bayes*: ¿Qué significa? dada la distribución a priori γ , un estimador de Bayes será el que minimice el riesgo de Bayes, definido como:

$$r(\varphi, \gamma) = \mathbf{E}(l(\Theta, \varphi(\underline{X})))$$

Si por ejemplo se considera la función de pérdida cuadrática: $l(\theta, d) = (\theta - d)^2$, el estimador de Bayes será la función $\varphi(\underline{X})$ que minimiza el error cuadrático medio, por lo tanto (otra vez el capítulo 5) ¿Quién será el estimador? Fijate que busco la función de \underline{X} que minimiza el ECM, entonces:

$$\varphi(\underline{X}) = \mathbf{E}(\Theta|\underline{X})$$

con respecto a la distribución a posteriori.

Hagamos ahora 4 ejemplos.

Ejemplo 1: Harvey tiene una bolsa con 10 monedas. La distribución a priori de la cantidad de monedas de dos caras en la bolsa es equiprobable sobre $\{1, 2, 3, 4\}$. El resto de las monedas con cara y ceca equilibradas. Harvey elige tres monedas al azar de las cuales 2 resultan ser de dos caras. Hallar la distribución a posteriori de la cantidad de monedas de dos caras que había inicialmente en la bolsa.

Ejemplo 2: Se quiere estudiar la proporción de componentes electrónicos defectuosos (θ) en un lote muy grande. Se supone que θ es cercano a 0.04. Se propone *a priori* $\Theta \sim \beta(1, 24)$. Entonces a priori:

$$f_{\Theta}(\theta) = \frac{\Gamma(25)}{\Gamma(1) * \Gamma(24)} \theta^0 (1 - \theta)^{23} \mathbf{1}_{\{0 < \theta < 1\}}$$

$$\mathbf{E}(\Theta) = \frac{a}{a+b} = 0,04, \mathbf{var}(\Theta) = \frac{ab}{(a+b)^2(a+b+1)} = 0,0015$$

Tomo una muestra aleatoria de n componentes, con $X_1, \dots, X_n | \Theta = \theta \sim Ber(\theta)$, entonces:

$$p_{X_i | \Theta = \theta}(x_i) = \theta^{x_i} (1 - \theta)^{1-x_i}$$

A posteriori, si $X = \underline{X}$:

$$f_{\Theta | X = x}(\theta) = \frac{\theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} * 24(1 - \theta)^{23}}{24 \int_0^1 \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i + 23} d\theta}$$

Se puede observar entonces que

$$\Theta | X = x \sim \beta \left(\sum_{i=1}^n x_i + 1, n - \sum_{i=1}^n x_i + 24 \right)$$

El estimador de Bayes para θ será

$$\varphi(\underline{X}) = \mathbf{E}(\Theta | \underline{X}) = \frac{\sum_{i=1}^n X_i + 1}{n + 25}$$

Recordemos que $\hat{\theta} = \bar{X}$, entonces:

$$\mathbf{E}(\Theta | X) = \frac{n}{n + 25} \bar{X} + \frac{25}{n + 25} \frac{1}{25}$$

Donde \bar{X} es el EMV de θ y $1/25$ es la esperanza de Θ a priori, por lo tanto el estimador de Bayes es un promedio ponderado entre el EMV y la esperanza a priori. Cuando n tiende a infinito, el EMV tiene mucho mas peso que lo que espero de θ a priori.

Ejemplo 3: El tiempo de espera (medido en horas) hasta que se produce la segunda falla en cierto tipo de sistemas es una variable aleatoria con distribución $\Gamma(2, \lambda)$. A priori se supone que λ tiene una distribución exponencial de media 1. Se observó que en uno de

tales sistemas el tiempo hasta la segunda falla fue 3/4 de hora. En virtud de la información muestral hallar la probabilidad de que el tiempo de espera hasta que se produzcan las próximas dos fallas sea mayor a 1 hora.

Ejemplo 4: El tiempo (en minutos) entre fallas de cierto tipo de sistemas es una variable aleatoria con distribución exponencial de parámetro λ . A priori se supone que λ se distribuye de acuerdo con la función de probabilidad $\mathbf{P}(\lambda = 1/120) = \mathbf{P}(\lambda = 1/20) = 0,5$. Si los tiempos entre las primeras 2 fallas fueron 20 y 30 minutos, estimar la probabilidad de que el tiempo entre la segunda y la tercera sea mayor a 30 minutos.

Otro ejemplo para hacer en clase: Muestra aleatoria de una población con distribución $Poi(\theta)$ donde *a priori*, $\Theta \sim \Gamma(\alpha, \lambda)$

4.1. Otros estimadores de Bayes

Para la pérdida $l(\theta, d) = w(\theta)(q(\theta) - d)^2$, con $w(\theta) > 0$ y $\mathbf{E}(w(\theta)) > \infty$, el estimador de Bayes será:

$$\varphi(x) = \frac{\mathbf{E}(q(\Theta)w(\Theta)|X = x)}{\mathbf{E}(w(\Theta)|X = x)}$$

Para la pérdida $l(\theta, d) = |q(\theta) - d|$, el estimador de Bayes será la mediana de la distribución a posteriori de $q(\Theta)$ condicionada a $X = x$.

En algunos casos, se estima θ con la moda de la distribución a posteriori de $\Theta|X = x$ (MAP)

4.2. Estimación de probabilidades

Para poder estimar probabilidades a partir de las distribuciones a posteriori de los parámetros aleatorios, usaremos la fórmula de probabilidad total. Dependiendo del tipo de variables que tengamos, usaremos alguna de sus versiones.

Por ejemplo: si tengo X_1, \dots, X_n una muestra aleatoria con $F_{\underline{X}|\Theta=\theta}(\underline{x})$ y la distribución *a posteriori* de $\Theta|\underline{X} = \underline{x}$, entonces si todas las variables son continuas:

$$\mathbf{P}(X_{n+1} > a) = \int_a^\infty \int_{-\infty}^\infty f_{X_{n+1}|\Theta=\theta}(x) f_{\Theta|\underline{X}=\underline{x}}(\theta) d\theta dx$$