TEMA 3: Estimación Puntual

Contents

3.1 Introducción . . . 2
  3.1.1 Conceptos generales . . . 2
3.2 Distribución del estimador en el muestreo . . . 3
  3.2.1 Distintos métodos para determinar la distribución de un estadístico . . . 3
3.3 Distribuciones relacionadas con la distribución Normal . . . 4
  3.3.1 Distribución χ^2 Chi-Cuadrado . . . 4
  3.3.2 Distribución t-student . . . 6
  3.3.3 Distribución Fisher-Snedecor . . . 7
3.4 Propiedades de los estimadores . . . 8
  3.4.1 Estimadores insesgados . . . 8
  3.4.2 Estimadores eficientes . . . 10
  3.4.3 Error Cuadrático Medio . . . 10
  3.4.4 Estimadores consistentes . . . 12
3.5 Casos de interés: . . . 12
  3.5.1 Estimador de la media poblacional. . . . 12
  3.5.2 Estimador de la varianza poblacional . . . 13
  3.5.3 Estimación de una proporción para muestras grandes . . . 16
  3.5.4 Estimación de la diferencia de medias de 2 muestras independientes . . . 16
  3.5.5 Estimación del cociente de varianzas de 2 muestras independientes . . . 18
  3.5.6 Estimación para la diferencia de dos proporciones para muestras grandes . . . 18
3.6 Métodos para la obtención de estimadores . . . 18
  3.6.1 Método de los momentos . . . 19
  3.6.2 Método de Máxima Verosimilitud . . . 19

Universidade Vigo

Departamento de Estatística e Investigación Operativa

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

El objetivo de esta unidad es asignar (a través de estimadores) un valor puntual a una característica desconocida de la población a partir de una muestra aleatoria, para ello hay que estudiar las propiedades de los estimadores y dar métodos para encontrar estimadores. Casos de especial relevancia son los estimadores relacionados con la distribución normal.

3.1 Introducción

En este tema se hacen inferencias sobre una población basándonos en información contenida en una muestra. Por ejemplo se puede inferir sobre el tiempo medio de acceso a internet de un determinado proveedor, la proporción de conexiones fallidas a la red internet facilitada por una empresa, el número medio de peticiones a un determinado servidor de páginas web o número medio de usarios conectados a un determinado chat, ... Para obtener tal información se extrae una muestra de la población base y se realiza una estimación del parámetro que nos interese. Por ejemplo, para obtener una estimación del tiempo medio de acceso a internet , se obtienen, mediante muestreo aleatorio, 100 tiempos correspondientes a 100 conexiones a internet y se calcula la media de estos tiempos. Este *valor* es una estimación del verdadero tiempo medio.

Las observaciones de una muestra las denotaremos por $x_1, \ldots, x_n$. Sin embargo, antes de hacer un muestreo o de experimentar, cualquier observación en particular estará sujeta a incertidumbre. Debido a esta incertidumbre, antes de que se disponga de datos numéricos concretos, consideramos las observaciones como variables aleatorias y las denotamos por letras mayúsculas $X_1, \ldots, X_n$. Esto a su vez implica que hasta que se hayan obtenido los datos, cualquier función de las observaciones (media muestral, varianza de la muestra, etc.) son funciones de variables aleatorias, y por tanto con distribución de probabilidad propia, llamada *distribución en el muestreo de la función estadística* que se considera.

3.1.1 Conceptos generales

Se está interesado en el estudio de una variable aleatoria X cuya función de distribución $F_\theta(x)$ es en mayor o menor medida desconocida. En muchas situaciones esta función se supondrá que pertenece a una familia de distribuciones conocida $\mathcal{F} = \{F_\theta(x) / \theta \in \Theta\}$ que depende únicamente de un parámetro desconocido $\theta \in \Theta$ (a Θ se le llama *espacio paramétrico)* de modo que el objetivo de la estimación puntual es emplear una muestra para calcular un número que represente en algún sentido el verdadero valor de ese parámetro.

Se supondrá con carácter general que se dispone de un conjunto de valores numéricos $x_1, \ldots, x_n$ que constituyen una *muestra aleatoria* de X, aunque antes de que se disponga de la información, las observaciones muestrales deben ser consideradas variables aleatorias $X_1, X_2, \ldots, X_n$ independientes y con la misma función de distribución $F_\theta(x)$, donde θ es desconocido (la distribución $F_\theta(x)$ puede ser una normal y entonces los parámetros son $\theta = (\mu, \sigma^2)$, puede ser una Bernoulli y el parámetro será $\theta = p$, puede ser una Poisson y el parámetro será $\theta = \lambda$, etc.).

La manera de proceder consiste en seleccionar una función llamada *estadístico o* **estimador** $T(X_1, X_2, \ldots, X_n)$ y tomar como estimación del parámetro θ, el valor de T calculado a partir de la muestra obtenida.

Definición 1. *Un* ***estadístico o estimador*** *es cualquier función $T(X_1, X_2, \ldots, X_n)$ de la muestra $X_1, X_2, \ldots, X_n$, por tanto es también una variable aleatoria con una distribución de probabilidad, llamada* ***distribución en el***

Universidade Vigo

Departamento de Estatística e Investigación Operativa

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

muestreo de T.

Algunos ejemplos de estadísticos son los siguientes:

a) $T(X_1, X_2, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$, *media muestral*, que suele designarse por $\overline{X}$, variable aleatoria frente a $\bar{x}$, estimación construida a partir de una realización particular de $\{X_1, X_2, \cdots, X_n\} = \{x_1, x_2, \cdots, x_n\}$.

b) $T(X_1, X_2, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$ *varianza muestral* que suele designarse por S^2 frente a la estimación s^2.

c) $T(X_1, X_2, \ldots, X_n) = \min(X_1, X_2, \ldots, X_n) = X_{(1)}$, el menor de los valores muestrales, frente a la estimación $x_{(1)}$.

d) $T(X_1, X_2, \ldots, X_n) = \max(X_1, X_2, \ldots, X_n) = X_{(n)}$, el mayor de los valores muestrales, frente a la estimación $x_{(n)}$.

Pueden existir varios estadísticos que estimen el verdadero valor del parámetro, por ejemplo la media y la mediana de una muestra son estimadores de la tendencia central. El problema radica en la selección del estimador adecuado a las características del modelo estadístico en estudio.

3.2 Distribución del estimador en el muestreo

Por lo general, la distribución del estadístico T es difícil de calcular: depende de la distribución de la variable de interés, del muestreo empleado y de la función T. Esta distribución es la que permite juzgar las cualidades del estadístico, comparar sus ventajas e inconvenientes frente a otros estimadores y evaluar los resultados que proporciona.

3.2.1 Distintos métodos para determinar la distribución de un estadístico

Aplicación del cambio de variable

Este método consiste en estudiar, utilizando el cambio de variable de vectores aleatorios, la distribución del estadístico en cuestió n. La principal ventaja de este método consiste en la obtención de la distribución exacta del estadístico y su principal inconveniente su complejidad analítica y que no siempre se obtiene solución.

Ejemplos de este método se usan en la construcción de los estimadores de la media y varianza bajo poblaciones normales.

Aplicación del teorema Central del Límite o resultados similares

Usando el teorema Central del Limite se obtiene la distribución asintótica normal si el estadístico se puede escribir como suma de variables aleatorias independientes e idénticamente distribuidas. Hay otros teoremas que relajan esta suposición. La principal ventaja es que la mayoria de los estadísticos se pueden expresar, al menos de forma aproximada, por sumas. Una de sus principales desventajas es la lentitud de la velocidad de convergencia a la distribución normal.

Ejemplos de la utilización de este método se verán en el estadístico media muestral en muestras grandes.

Técnicas de simulación: Monte Carlo y Bootstrap

Si se conoce la distribución de los datos (método de Monte Carlo) o bién se puede aproximar (método Bootstrap) se puede determinar, de forma aproximada, la distribución del estadístico de interés. La principal ventaja de este tipo de técnicas es su no “complejidad técnica” y como principal desventaja

Universidade de Vigo
Departamento de Estatística e Investigación Operativa
Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense
Tel. 988 368 766
cotos@uvigo.es

el habitual desconocimiento de la distribución de los datos. Veremos ejemplos de uso de la técnica de Monte Carlo en las prácticas.

3.3 Distribuciones relacionadas con la distribución Normal

En la inferencia estadística que veremos en los siguientes temas surgen nuevas v.a. continuas que son transformaciones de v.a. normales: distribución χ^2, distribución t de Student y F de Snedecor.

3.3.1 Distribución χ^2 Chi-Cuadrado

Sean $Z_1, Z_2, \cdots, Z_n$ variables aleatorias independientes tales que $Z_i \sim N(0.1)$ $(i = 1, \ldots, n)$ entonces la variable aleatoria:

$$\chi_n^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2$$

se dice que tiene una distribución χ^2 de Pearson con n grados de libertad.
Sea $X \sim \chi_n^2$, se verifica que $E(X) = n$ y $Var(X) = 2n$.

Definición 2. *Se define el número $\chi^2_{n,\alpha}$ como el valor de la χ^2_n tal que $P\left(\chi^2_n \leq \chi^2_{n,\alpha}\right) = \alpha$.*

Observación 1. *La distribución χ^2 está tabulada y toma valores no negativos.*

Observación 2. *Usando, principalmente el Teorema Central de Límite, se sigue que para n grande, la distribución chi-cuadrado puede ser aproximada por la distribución normal de media n y varianza $2n$.*

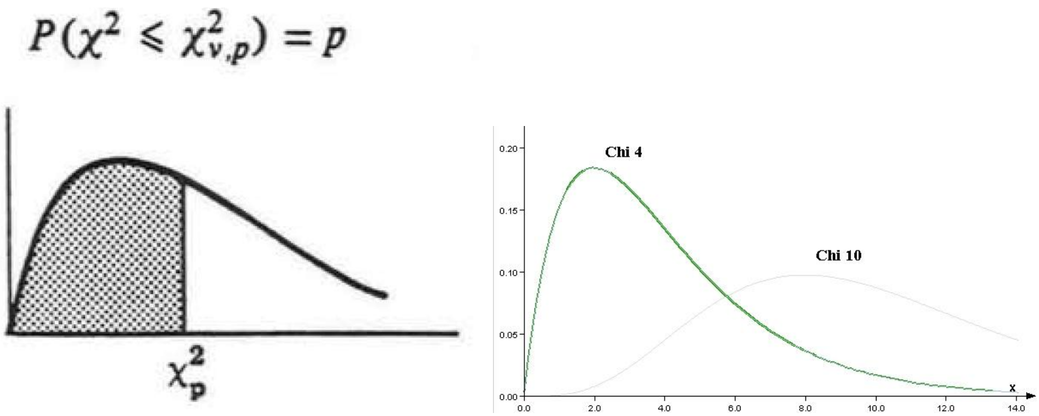

Figure 1:

Observación 3. *Sean $X \sim \chi^2_n$ y $Y \sim \chi^2_m$ independientes, entonces $X + Y \sim \chi^2_{n+m}$.*

Observación 4. *Para $n > 100$ se pueden obtener probabilidades y cuantiles la distribución χ^2_n usando la aproximación:*

$$\sqrt{2\chi_n^2 - \sqrt{2n-1}} \underset{aprox.}{\sim} N(0.1)$$

Universida de Vigo

Departamento de Estatística e Investigación Operativa

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

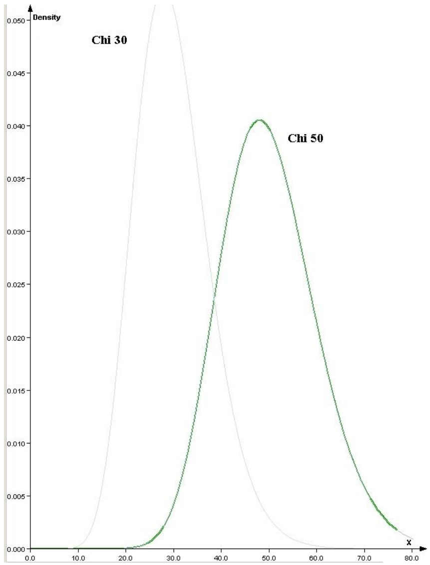

Figure 2:

Ejercicio 1. *Para un valor de n obtener:*

1. *Representar gráficamente la densidad de la variable.*

   $n < -3$
   # a) función de densidade
   curve(dchisq(x,df=n),from=0,to=qchisq(0.999,n),100, main="Densidade Chi-squared")

2. *Representar la función de distribución*
   curve(pchisq(x,df=n), from=0, to=qchisq(0.999,n),100, main="Distribución Chi-squared")

3. *Obtener las carácterísticas principales de la distribución (media y varianza).*

```
x_chisq <- function(x,df=1) dchisq(x,df)*x
esperanza_x <- integrate(x_chisq,lower=0,upper=Inf,n) \#Ollo devolve unha lista
x2_chisq $<-$ function(x,df=1) dchisq(x,df)*x*x
esperanza_x2 <- integrate(x2_chisq,lower=0,upper=Inf,n) \#Ollo devolve unha lista
var <- esperanza_x2$value - esperanza\_x$value**2
print(esperanza\_x)
print(var)
```

4. $P(X < 20)$ *con* $X \sim \chi^2_{18}$*; el valor de k tal que* $P(X > k) = 0.975$.
   print(pchisq(20,18)); print(qchisq(0.975,18))

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

Ejercicio 2. *Para valores $n > 100$, una $X \sim \chi^2_n$ se puede aproximar por $\frac{X-n}{\sqrt{2n}} \sim N(0,1)$. Dibujar ambas curvas de distribución de probabilidad para $n = 101$.*

Proof.

```
ff <- function( n=101) return( (rchisq(1,n)-n)/sqrt(2*n) )
dd <- replicate( 10000, ff(n=101))
hist(dd,freq=FALSE)
xx <- seq(-3,3,length=200)
yy <- dnorm(xx,mean=0,sd=1)
lines(xx,yy,col="red")
```

□

Ejercicio 3. *1. Comprobar por simulación la relación (definición) entre la distribución normal y la Chi-cuadrado*

Ejercicio 4. *Obtener*

1. $P\left(\chi^2_{10} > 3.9\right)$
   1-pchisq(3.9,10)
2. $P\left(3.25 < \chi^2_{10} < 16\right)$
   pchisq(16,10)-pchisq(3.25,16)
3. $P\left(\chi^2_{15} > 15\right)$
   1-pchisq(15,15)
4. $a \,/\, P\left(\chi^2_8 > a\right) = 0.25$
   qchisq(0.75,8)
5. $a \,/\, P\left(\chi^2_8 < a\right) = 0.05$
   qchisq(0.05,8)

Ejercicio 5. *Obtener por simulación las correspondientes aproximaciones de las cantidades anteriores.*

3.3.2 Distribución t-student

Sean $X \sim N(0,1)$ y $Y \sim \chi^2_n$ independientes, entonces la distribución $t-$student con n grados de libertad es la distribución de la transformación:

$$t_n = \frac{X}{\sqrt{\frac{Y}{n}}}$$

Como principales características destacar que es una v.a. simétrica con respecto al cero $(E(X) = 0)$ y que $Var(X) = \frac{n}{n-2}$ si $n > 2$.

Definición 3. *Se define el número $t_{n,\alpha}$ como el valor de la t_n tal que $P(t_n \leq t_{n,\alpha}) = \alpha$.*

Observación 5. *La distribución está tabulada y habitualmente aparecen valores positivos y negativos.*

Observación 6. *Al aumentar los grados de libertad, la distribución t-Student se aproxima cada vez más a la distribución $N(0,1)$.*

Observación 7. *Presenta una mayor dispersión que la distribución normal y una menor curtosis.*

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

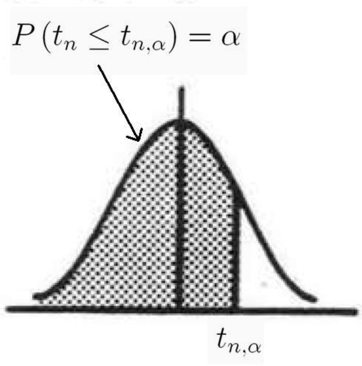

Figure 3:

Ejercicio 6. *Obtener (a través de las tablas)*

1. $P(t_{10} > 0.26)$
   1-pt(0.26,10)
2. $P(-0.26 < t_{10} < 0.26)$
   pt(0.26,10)-pt(-0.26,10)
3. $P(t_{15} > 0.75)$
   1-pt(0.75,15)
4. $a \;/\; P(t_8 > a) = 0.25$
   qt(0.75,8)
5. $a \;/\; P(t_8 < a) = 0.5$
   qt(0.5,8)

Ejercicio 7. *Obtener por simulación las correspondientes aproximaciones de las cantidades anteriores.*

3.3.3 Distribución Fisher-Snedecor

Dadas dos variables $X \sim \chi^2_n$ y $Y \sim \chi^2_m$ independientes, se define la distribución de Fisher-Snedecor con el par (n, m) de grados de libertad a la distribución de la v.a.

$$F_{n,m} = \frac{\frac{X}{n}}{\frac{Y}{m}}.$$

Como principales características destacar que es una v.a. que toma valores no negativos con $E(X) = \frac{n}{n-2}$ $n > 2$.

Definición 4. *Si $X \sim F_{n,m}$ entonces $\frac{1}{X} \sim F_{m,n}$. Usando esta relación se demuestra que $F_{m,n,\alpha} = \frac{1}{F_{n,m,1-\alpha}}$.*

Universidade Vigo

Departamento de Estatística e Investigación Operativa

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

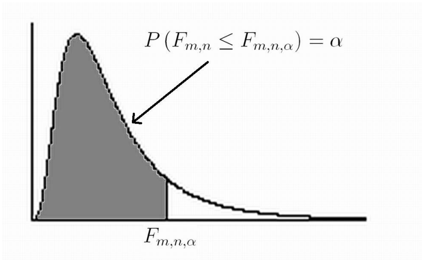

Figure 4:

Observación 8. *La distribución está tabulada habitualmente para probabilidades próximas a uno.*

Observación 9. *Se define el número $F_{n,m,\alpha}$ como el valor de la $F_{n,m}$ tal que $P\left(F_{n,m} \leq F_{n,m,\alpha}\right) = \alpha$.*

Ejercicio 8. *Obtener (a través de las tablas)*

1. $a \; / \; P\left(F_{5,10} < a\right) = 0.9$
   qf(0.9,5,10)

2. $a \; / \; P\left(F_{5,10} < a\right) = 0.1$
   qf(0.1,5,10)

3. $a \; / \; P\left(F_{10,10} > a\right) = 0.05$
   qf(0.95,10,10)

4. $a \; / \; P\left(F_{10,10} < a\right) = 0.05$
   qf(0.05,10,10)

Ejercicio 9. *Obtener por simulación las correspondientes aproximaciones de las cantidades anteriores.*

En 'http://www.kuleuven.ac.be/ucs/java/index.htm' –> 'Basics' –> 'Characteristics of distributions' se muestran gráficamente (Applets) las funciones de densidad de las v. anterirores, permitiendo el cálculo de los cuantiles y/o probabilidades

3.4 Propiedades de los estimadores

A continuación se estudiarán diversos criterios de comparación de estimadores y las propiedades que son deseables para que un estadístico produzca buenas estimaciones. Muchas de estas cuestiones tendrán que basarse, no en los valores que se obtengan de los estimadores para unas muestras concretas, sino en la distribución en el muestreo de los estadísticos. Al margen de las preferencias que pueda haber entre los estimadores, existen propiedades globales que resultan convenientes para considerar un estimador como "bueno".

3.4.1 Estimadores insesgados

Intuitivamente, un estimador es insesgado para un parámetro si al seleccionar un gran número de muestras de tamaño n, la media de todas las estimaciones está próxima al valor del verdadero parámetro θ.

Departamento de Estatística e Investigación Operativa

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

Definición 5. *Se denomina **sesgo** del estimador $\widehat{\theta} = T(X_1, X_2, \ldots, X_n)$ como estimador de θ a*

$$Sesgo\left(\widehat{\theta}\right) = E\left(\widehat{\theta}\right) - \theta$$

*En caso de que el sesgo sea nulo, es decir, $E\left(\widehat{\theta}\right) = \theta$ para cada $\theta \in \Theta$, el estimador $\widehat{\theta} = T(X_1, X_2, \ldots, X_n)$ se dice **centrado** o **insesgado** para θ.*

Esta propiedad no es estrictamente necesaria (depende de la variabilidad del estimador). Sin embargo, la siguiente debe verificarla todo estimador razonable:

Definición 6. *Un estimador $\widehat{\theta} = T(X_1, X_2, \ldots, X_n)$ es **asintóticamente insesgado** para θ si*

$$\lim_{n\to\infty} E\left(\widehat{\theta}_n\right) = \theta \quad \forall\theta \in \Theta$$

Ejemplo 1. *Dada una m.a. $\{X_1, X_2, \cdots, X_n\}$ demostrar que cualquier estimador de la forma $a_1X_1 + a_2X_2 + \ldots + a_nX_n$ es insesgado para $\mu = E(X_i)$, si $a_1 + a_2 + \ldots + a_n = 1$.*

La propiedad de que un estimador sea centrado es deseable pero no es definitiva. La idea se muestra en la siguiente figura

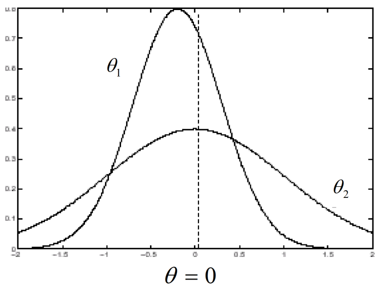

Figure 5:

Los dos estadísticos θ_1 y θ_2 estiman el valor $\theta = 0$. El primer estimador se observa que es centrado (función de densidad simétrica con respecto al 0) mientras que el segundo tiene un sesgo de -0.2 pero, a simple vista, ¿cuál parece mejor? A continuación se muestran de nuevo dos estadísticos θ_1 y θ_2 que estiman el valor $\theta = 0$. El primero sigue siendo centrado pero ahora el segundo tiene un sesgo de -1. Ahora, ¿cuál parece mejor?

Está claro que deberíamos combinar criterios. Interesan estimadores centrados pero no con demasiada varianza y viceversa, interesan estimadores con baja varianza pero no demasiado sesgados.

Universidade Vigo

Departamento de Estatística e Investigación Operativa

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

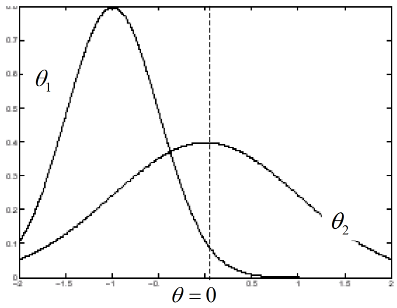

Figure 6:

3.4.2 Estimadores eficientes

Definición 7. *Se denomina **precisión** o **eficiencia** del estimador $\widehat{\theta} = T(X_1, \ldots, X_n)$ como estimador de θ a la inversa de la varianza, es decir*

$$\boxed{Efic\left(\widehat{\theta}\right) = \frac{1}{Var\left(\widehat{\theta}\right)}}$$

Se dice, por tanto, que un estimador $\widehat{\theta}_1$ es más eficiente o más preciso que $\widehat{\theta}_2$ si, para cualquier tamaño muestral

$$Efic\left(\widehat{\theta}_1\right) \geq Efic\left(\widehat{\theta}_2\right) \Leftrightarrow Var\left(\widehat{\theta}_1\right) \leq Var\left(\widehat{\theta}_2\right)$$

Definición 8. *Se llama **eficiencia relativa** de $\widehat{\theta}_1$ respecto a $\widehat{\theta}_2$ al cociente entre sus eficiencias:*

$$ER\left(\widehat{\theta}_1, \widehat{\theta}_2\right) = \frac{Efic\left(\widehat{\theta}_1\right)}{Efic\left(\widehat{\theta}_2\right)} = \frac{Var\left(\widehat{\theta}_2\right)}{Var\left(\widehat{\theta}_1\right)}$$

3.4.3 Error Cuadrático Medio

Una manera de evaluar la calidad de las estimaciones es "medir" de alguna manera la "distancia" del parámetro al estimador, por ejemplo $\left(\widehat{\theta} - \theta\right)^2$. Esta expresión depende de la muestra, por lo que es una variable aleatoria. Podemos considerar su esperanza, que nos dará la pérdida media.

Definición 9. *Se llama **Error Cuadrático Medio** de $\widehat{\theta}$ como estimador de θ a*

$$\boxed{ECM\left(\widehat{\theta}\right) = E\left[\left(\widehat{\theta} - \theta\right)^2\right] = \left(Sesgo\left(\widehat{\theta}\right)\right)^2 + Var\left(\widehat{\theta}\right)}$$

En general el Error Cuadrático Medio depende de θ y n y no suele ser fácil decidir si un estimador es mejor que otro. La solución es complementar de alguna forma el criterio del Error Cuadrático Medio.

Universida de Vigo

Departamento de Estatística e Investigación Operativa

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

Observación 10. *En ocasiones, cuando se trabaja con estimadores sesgados, se utiliza el error cuadrático medio en lugar de la varianza (p.e. se podría definir una eficiencia en términos de e.c.m.)*

Definición 10. *Una vez decidido limitarse a estimadores insesgados, evidentemente interesan aquellos que tengan la menor dispersión posible alrededor del parámetro . Estos estimadores se llaman* ***Estimadores Centrados Uniformemente de Mí-nima Varianza (ECUMV).***

En muchos casos es posible encontrar funciones $c(\theta)$ que acoten inferiormente $Var\left(\widehat{\theta}\right)$:

$$c(\theta) \leq Var\left(\widehat{\theta}\right) \quad \forall\, \widehat{\theta} \text{ estimador insesgado de } \theta$$

En tales casos, si $Var\left(\widehat{\theta}\right) = c(\theta)$, entonces $\widehat{\theta}$ es el estimador centrado uniformemente de mínima varianza (ECUMV) de θ.

DESIGUALDAD DE FRECHET-CRAMER-RAO

Bajo condiciones generales de regularidad:

- Θ es un intervalo abierto de $\mathbb{R}$.
- $\{(x_1, \ldots, x_n) \in \chi / f_\theta(x_1, \ldots, x_n) > 0\}$ no depende de θ.
- Existe $\frac{\partial}{\partial\theta} f_\theta(x_1, \ldots, x_n)$

Si $\widehat{\theta} = T(X_1, X_2, \ldots, X_n)$ es un estimador unidimensional centrado tal que

$$\int T(x_1, x_2, \ldots, x_n) \frac{\partial}{\partial\theta} f_\theta(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1$$

entonces

$$Var\left(\widehat{\theta}\right) \geq c(\theta) = \frac{1}{I(\theta)} = \textbf{Cota de Frechet-Cramer-Rao}$$

donde

$$I(\theta) = E_\theta\left(\left(\frac{\partial}{\partial\theta} \ln f_\theta(x_1, \ldots, x_n)\right)^2\right) = nE_\theta\left(\left(\frac{\partial}{\partial\theta} \ln f_\theta(x)\right)^2\right)$$

se llama **Información de Fisher**.

$I(\theta)$ es un índice de la información que proporciona la muestra para estimar θ.

Si $Var\left(\widehat{\theta}\right) = c(\theta) = \frac{1}{I(\theta)}$, entonces se dice que el estimador $\widehat{\theta} = T(X_1, \ldots, X_n)$ es un **estimador centrado uniformemente de mínima varianza (ECUMV)**.

Universidade Vigo

Departamento de Estatística e Investigación Operativa

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

3.4.4 Estimadores consistentes

La consistencia de un estimador hace referencia a su comportamiento cuando la muestra es muy grande. Intuitivamente, un estimador es consistente si se aproxima, al crecer el tamaño muestral, al verdadero valor del parámetro.

Definición 11. *Una sucesión de estimadores* $\widehat{\theta}_n = T_n(X_1, X_2, \ldots, X_n)$ *es* ***consistente en media cuadrática*** *para estimar* θ *si*

$$\lim_{n\to\infty} ECM\left(\widehat{\theta}_n\right) = 0 \ \ \forall\theta \in \Theta$$

O equivalentemente expresado en términos de media y varianza se pide que

$$\lim_{n\to\infty} Sesgo\left(\widehat{\theta}_n\right) = \lim_{n\to\infty} Var\left(\widehat{\theta}_n\right) = 0$$

3.5 Casos de interés:

3.5.1 Estimador de la media poblacional.

Sea $X_1, \ldots, X_n$ una m.a. de una variable aleatoria X con media $E(X) = \mu$ y varianza $Var(X) = \sigma^2$. Entonces el estimador de la media μ, es la **media muestral** que viene dado por:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

verifica las siguientes propiedades:

1. Es un **estimador insesgado** de μ.

$$E\left(\overline{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

2. Es un estimador **consistente en media cuadrática** de μ, puesto que es insesgado y su varianza tiende a 0 cuando $n \to \infty$ [1].

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

3. La distribución de exacta de $\overline{X}$ depende de la distribución de la población X. Por ejemplo, si X es normal la distribución de $\overline{X}$ también lo será.

Theorem 1. *Sea* $X_1, \ldots, X_n$ *una m.a. de una variable aleatoria* $X \sim N(\mu, \sigma)$*, entonces*

$$\boxed{\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

[1] Recordemos que si X_1 y X_2 son independientes, entonces $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$

Universida de Vigo

Departamento de Estatística e Investigación Operativa

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

Theorem 2. *Sea $X_1, \ldots, X_n$ una m.a. de una variable aleatoria $X \sim F$ tal que $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ $(i = 1, 2, \ldots, n)$ entonces*

$$\boxed{\overline{X} \overrightarrow{n \to +\infty} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

Observación 11. *Esta aproximación es buena para muestras grandes $(n \geq 30)$ y es una consecuencia del Teorema Central del Límite.*

Ejemplo 2. *Sea X una población con distribución $N(90, 20)$.*

a) Si se obtiene una m.a. de tamaño 16, ¿cuál es la probabilidad de que la media muestral $\overline{X}$ sea mayor o igual que 92?

b) Determinar el tamaño muestral para que la probabilidad de que la media muestral sea menor o igual que 98 sea $P(\overline{X} \leq 98) = 0.99$.

$X \,\text{-}\, N(\mu = 90, \sigma = 20)$

$$\left.\begin{array}{r} E(\overline{X}) = \mu = 90 \\ Var(\overline{X}) = \dfrac{\sigma^2}{n} = \dfrac{20^2}{16} \Rightarrow \sigma_{\overline{X}} = \sqrt{\dfrac{400}{16}} = \dfrac{20}{4} = 5 \end{array}\right\} \Rightarrow \overline{X} \,\text{-}\, N(90, 5)$$

a) $P(\overline{X} \geq 92) = P(N(90,5) \geq 92) = P\left(N(0,1) \geq \dfrac{92-90}{5}\right) =$

$= P\left(Z \geq \frac{2}{5}\right) = 1 - P\left(Z \leq \frac{2}{5}\right) = 1 - 0.6554 = \boxed{0.3446}$

b) $0.99 = P(\overline{X} \leq 98) = P\left(N\left(90, \dfrac{20}{\sqrt{n}}\right) \leq 98\right) = P\left(N(0,1) \leq \dfrac{98-90}{\frac{20}{\sqrt{n}}}\right) =$

$= P\left(Z \leq \sqrt{n}\frac{2}{5}\right) \Longrightarrow \sqrt{n}\dfrac{2}{5} = 2.326348 \Longrightarrow n = \left(\dfrac{2.326348 \cdot 5}{2}\right)^2 = 33.82434 \Longrightarrow \boxed{n \geq 34}$

3.5.2 Estimador de la varianza poblacional

Sea $X_1, \ldots, X_n$ una m.a. de una variable aleatoria X con media $E(X) = \mu$ y varianza $Var(X) = \sigma^2$. El estimador más razonable de la varianza poblacional σ^2 es la **varianza muestral**

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

que verifica las siguientes propiedades:

1. Es un **estimador consistente** de σ^2.
2. Es un estimador **asintóticamente insesgado** de σ^2 puesto que

$$\boxed{E(S^2) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 \underset{n\to\infty}{\longrightarrow} \sigma^2}$$

UniversidadeVigo

Departamento de Estatística e Investigación Operativa

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

Puesto que la varianza muestral S^2 es un estimador sesgado (aunque asintóticamente insesgado), para estimar la varianza σ^2 se usa la **cuasivarianza muestral**:

$$\widehat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{n}{n-1}S^2$$

que es un estimador **insesgado** y **consistente** de σ^2.

$$E\left(\widehat{S}^2\right) = E\left(\frac{n}{n-1}S^2\right) = \frac{n}{n-1}E\left(S^2\right) = \frac{n}{n-1}\left(\sigma^2 - \frac{\sigma^2}{n}\right) = \sigma^2$$

Proposición 1. *Bajo hipótesis de normalidad* $(X \sim N(\mu,\sigma))$ *, las variables aleatorias* $\overline{X}$ *y* $\widehat{S}^2$ *son independientes.*

Theorem 3. *(**Teorema de Fisher**) Bajo hipótesis de normalidad* $(X \sim N(\mu,\sigma))$ *se verifica que*

$$\boxed{\frac{nS^2}{\sigma^2} = \frac{(n-1)\widehat{S}^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\sigma^2} \sim \chi^2_{n-1}}$$

puesto que

1. $$\frac{nS^2}{\sigma^2} = \frac{n}{\sigma^2}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - (\overline{X}-\mu)^2\right) = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right) - \left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right) =$$

$$= \sum_{i=1}^{n} Z_i^2 - Z^2 = \sum_{i=1}^{n-1} Z_i^2 \sim \chi^2_{n-1}$$

Observación 12. *Si la* μ *es conocida entonces se verifica que*

$$\boxed{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2_n}$$

Ejemplo 3. *Dada una población* $X \sim N(6, 2.5)$*, y tomando una muestra aleatoria* $X_1, \ldots, X_n$ *de tamaño* $n = 12$*, calcular la probabilidad de que la varianza muestral sea mayor que* 4.9.

$$P\left(S^2 > 4.9\right) = 1 - P\left(S^2 \le 4.9\right) = 1 - P\left(\frac{nS^2}{\sigma^2} \le \frac{n\cdot 4.9}{\sigma^2}\right) =$$

$$= 1 - P\left(\chi^2_{n-1} \le \frac{12\cdot 4.9}{2.5^2}\right) = 1 - P\left(\chi^2_{11} \le 9.41\right) = 1 - pchisq(9.41, 11) = 0.5841031$$

Estadístico t de Student. Estimación de la media poblacional cuando σ^2 es desconocida

Universida de Vigo

Departamento de Estatística e Investigación Operativa

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

Bajo hipótesis de normalidad $(X \sim N(\mu, \sigma))$ se verifica que $\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, o equi-valentemente

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Este resultado puede ser de poca utilidad si la varianza poblacional σ^2 es desconocida, ya que entonces no se podrá usar esta conclusión para hacer previsiones acerca de $\overline{X}$. Cabe pensar que el resultado no será muy distinto si se sustituye σ por la cuasidesviación típica muestral $\widehat{S}$, puesto que, al menos para muestras grandes, σ^2 y $\widehat{S}^2$ tendrán valores semejantes. Tal idea llevó a Student (pseudónimo de W. Gosset) a considerar el estadístico

$$\frac{\overline{X}-\mu}{\frac{\widehat{S}}{\sqrt{n}}} = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n-1}}} \sim t_{n-1}$$

Ejemplo 4. *Dada una población $X \sim N(-1, \sigma)$, se extrae una m.a. de tamaño $n = 10$ con los siguientes resultados:*

$$1.08, \ -1.79, \ -2.54, \ 0.37, \ -0.6, \ 0.53, \ 0.28, \ -2.21, \ -2.66, \ 1.45$$

Calcular la probabilidad de que la media muestral $\overline{X}$ sea mayor que -1.2.

$$P\left(\overline{X} > -1.2\right) = 1 - P\left(\overline{X} \leq -1.2\right) = 1 - P\left(\frac{\overline{X}-\mu}{\frac{\widehat{S}}{\sqrt{n}}} \leq \frac{-1.2-\mu}{\frac{\widehat{s}}{\sqrt{n}}}\right)$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 = \frac{1}{10}(1.08^2 + 1.79^2 + 2.54^2 + 0.37^2 + 0.6^2 + 0.53^2 + 0.28^2$$
$$+2.21^2 + 2.66^2 + 1.45^2) - [\frac{1}{10}(1.08 + 1.79 + 2.54 + 0.37 + 0.6 + 0.53+$$
$$+0.28 + 2.21 + 2.66 + 1.45)]^2 = \frac{1}{10} \cdot 25.7405 - 0.609^2 = 2.20317$$

$$s = \sqrt{2.20317} = 1.4843 \Longrightarrow \widehat{s} = \sqrt{\frac{n}{n-1}s^2} = \sqrt{\frac{10}{9}2.20317} = 1.5646$$

o con R: sd(c(1.08,-1.79,-2.54,0.37,-0.6,0.53,0.28,-2.21,-2.66,1.45))=1.564598

$$P\left(\overline{X} > -1.2\right) = 1 - P\left(\frac{\overline{X}-\mu}{\frac{\widehat{S}}{\sqrt{n}}} \leq \frac{-1.2-\mu}{\frac{\widehat{s}}{\sqrt{n}}}\right) = 1 - P\left(t_{n-1} \leq \frac{-1.2-(-1)}{\frac{1.5646}{\sqrt{10}}}\right) =$$

$$= 1 - P\left(t_9 \leq \frac{-1.2-(-1)}{\frac{1.5646}{\sqrt{10}}}\right) = 1 - P(t_9 \leq -0.4042) = 1 - pt(-0.4042, 9) = 0.6522489$$

Universidade Vigo

Departamento de Estatística e Investigación Operativa

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

3.5.3 Estimación de una proporción para muestras grandes

Se desea estimar la proporción p de individuos de una población que tiene una determinada característica. Para ello se toma una m.a. de elementos de la población, anotando un 1 si dicho elemento tiene la característica, y 0 en otro caso, es decir, se tiene una m.a. $X_1, \ldots, X_n$ de una $B(1, p)$

$$X = \begin{cases} 1 & \text{(tiene la caract.) con probab. } p \\ 0 & \text{(no la tiene) con probab. } 1-p \end{cases}, \ \mathbf{E}(X) = p, \ Var(X) = p(1-p)$$

Un estimador razonable de p es la proporción de elementos de la muestra que tiene dicha característica, es decir

$$\boxed{\widehat{p} = \frac{X_1 + \cdots + X_n}{n}}$$

Propiedades

1. $\mathbf{E}(\widehat{p}) = \mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}(X_i) = \frac{1}{n}np = p$

2. $Var(\widehat{p}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} p(1-p) = \frac{p(1-p)}{n}$

3. La distribución de $\widehat{p}$ depende de la distribución de la población X, pero cuando n es grande

$$\hat{p} \xrightarrow[n\to+\infty]{} N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Ejemplo 5. *En el proceso de producción de una empresa, en una m.a. de tamaño $n = 25$, el 1% de los productos sale defectuoso y se desea estimar la probabilidad de que la proporción de productos defectuosos sea mayor que el 2%.*

$$\mathbf{P}(\widehat{p} > 0.02) = \mathbf{P}\left(\frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} > \frac{0.02 - p}{\sqrt{\frac{p(1-p)}{n}}}\right) \simeq \mathbf{P}\left(Z > \frac{0.02 - p}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

$$= \mathbf{P}\left(Z > \frac{0.02 - 0.01}{\sqrt{\frac{0.01(1-0.01)}{25}}}\right) = \mathbf{P}(Z > 0.5025) = 1 - \mathbf{P}(Z \leq 0.5025) =$$

$$= 1 - 0.6935 = 0.3065$$

$$\mathbf{P}(\widehat{p} > 0.02) = 0.3065$$

3.5.4 Estimación de la diferencia de medias de 2 muestras independientes

Sean $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ dos muestras aleatorias recogidas independientemente de dos poblaciones normales $X \sim N(\mu_X, \sigma_X)$ e $Y \sim N(\mu_Y, \sigma_Y)$. Se desea construir intervalos de confianza para la diferencia de medias $\mu_X - \mu_Y$.

UniversidadeVigo

Departamento de Estatística e Investigación Operativa

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

- Caso 1: varianzas conocidas

Puesto que $\overline{X}$ e $\overline{Y}$ son independientes:

$$\left.\begin{array}{l}\overline{X} \sim N\left(\mu_X, \dfrac{\sigma_X}{\sqrt{n}}\right) \\ \overline{Y} \sim N\left(\mu_Y, \dfrac{\sigma_Y}{\sqrt{m}}\right)\end{array}\right\} \Rightarrow \overline{X}-\overline{Y} \sim N\left(\mu_X-\mu_Y, \sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}\right)$$

- Caso 2: varianzas desconocidas pero iguales

Ahora supongamos que las varianzas σ_X^2 y σ_Y^2 son desconocidas pero iguales (a σ^2). En este caso, se sabe que la variable

$$\overline{X}-\overline{Y} \sim N\left(\mu_X-\mu_Y, \sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}\right) \Leftrightarrow \overline{X}-\overline{Y} \sim N\left(\mu_X-\mu_Y, \sqrt{\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)}\right)$$

y por tanto

$$T=\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sigma\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \sim N(0,1)$$

Como en la expresión anterior no se conoce el valor de σ^2, éste se puede sustituir por el valor

$$\widehat{S}^2=\frac{(n-1)\,\widehat{S}_X^2+(m-1)\,\widehat{S}_Y^2}{n+m-2}$$

que es una ponderación de las dos cuasivarianzas muestrales $\widehat{S}_X^2$ y $\widehat{S}_Y^2$. Se puede demostrar que

$$T=\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\widehat{S}\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \sim t_{n+m-2}$$

- Caso 3: varianzas desconocidas y distintas

Si las varianzas no se pueden suponer iguales, el estadístico anterior no se puede utilizar. En el caso de que las varianzas fuesen conocidas, el estadístico a utilizar era

$$T=\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n}+\dfrac{\sigma_Y^2}{m}}} \sim N(0,1)$$

a) **Tamaños muestrales grandes (**$n, m \geq 30$)
En este caso se utiliza el estadístico pivote anterior estimando σ_X por $\widehat{S}_X$ y σ_Y por $\widehat{S}_Y$, es decir,

$$T=\frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\dfrac{\widehat{S}_X^2}{n}+\dfrac{\widehat{S}_Y^2}{m}}} \underset{A\sin t}{\sim} N(0,1)$$

Departamento de Estatística e Investigación Operativa
Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense
Tel. 988 368 766
cotos@uvigo.es

que tiene una distribución aproximadamente normal $N(0,1)$.

b) **Tamaños muestrales pequeños ($n, m < 30$)**

En este caso la aproximación normal no es nada precisa. La solución más usada habitualmente es la aproximación debida a Welch, según la cual el estadístico anterior sigue una distribución t de Student con $g = n + m - 2 - \delta$ grados de libertad, con δ el entero más próximo a

$$\Delta = \frac{\left[(m-1)\frac{\widehat{S}_X^2}{n} - (n-1)\frac{\widehat{S}_Y^2}{m}\right]^2}{(m-1)\left(\frac{\widehat{S}_X^2}{n}\right)^2 + (n-1)\left(\frac{\widehat{S}_Y^2}{m}\right)^2}$$

Se comprueba que $0 \leq \delta \leq \max(n-1, m-1)$

3.5.5 Estimación del cociente de varianzas de 2 muestras independientes

Teniendo en cuenta la normalidad e independencia de ambas poblaciones, se verifica que

$$\left.\begin{array}{l} \frac{(n-1)\widehat{S}_X^2}{\sigma_X^2} \sim \chi_{n-1}^2 \\ \frac{(m-1)\widehat{S}_Y^2}{\sigma_Y^2} \sim \chi_{m-1}^2 \end{array}\right\} \Rightarrow \frac{\frac{\widehat{S}_X^2}{\sigma_X^2}}{\frac{\widehat{S}_Y^2}{\sigma_Y^2}} = \frac{\widehat{S}_X^2}{\sigma_X^2}\frac{\sigma_Y^2}{\widehat{S}_Y^2} \sim F_{n-1,m-1}$$

3.5.6 Estimación para la diferencia de dos proporciones para muestras grandes

La manera de proceder es idéntica a la del caso de una sola proporción. Se recogen sendas muestras aleatorias simples $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ de las variables X e Y donde

$$X_i = \begin{cases} 1 & \text{con probabilidad } p_X \\ 0 & \text{con probabilidad } 1 - p_X \end{cases} \quad \text{e} \quad Y_i = \begin{cases} 1 & \text{con probabilidad } p_Y \\ 0 & \text{con probabilidad } 1 - p_Y \end{cases}$$

Supuesta aceptable la aproximación normal, se tiene que

$$\left.\begin{array}{l} \frac{\widehat{p}_X - p_X}{\sqrt{\frac{p_X(1-p_X)}{n}}} \sim N(0,1) \\ \frac{\widehat{p}_Y - p_Y}{\sqrt{\frac{p_Y(1-p_Y)}{m}}} \sim N(0,1) \end{array}\right\} \Rightarrow \widehat{p}_X - \widehat{p}_Y \sim N\left(p_X - p_Y, \sqrt{\frac{p_X(1-p_X)}{n} + \frac{p_Y(1-p_Y)}{m}}\right)$$

3.6 Métodos para la obtención de estimadores

Veremos dos técnicas para obtener estimacimadores puntuales, el método de los momentos y el de máxima verosimilitud.

3.6.1 Método de los momentos

Sea $(X_1, \ldots, X_n)$ una m.a. de una población X que se caracteriza por k parámetros $\theta_1, \ldots, \theta_k$ que se desean estimar. Los momentos poblacionales (cuando existan) serán, en general, función de ellos, es decir

$$\alpha_j = E\left(X^j\right) = g_j\left(\theta_1, \ldots, \theta_k\right) \quad j = 1, 2, \ldots$$

La idea del método de los momentos consiste en igualar los primeros momentos poblacionales a los correpondientes momentos muestrales

	Momentos Poblacionales	**Momentos Muestrales**
Centrados en el origen	$\alpha_j = \mathbf{E}\left(X^j\right)$	$a_j = \frac{1}{n}\sum\limits_{i=1}^{n} x_i^j$
Centrados en la media	$\mu_j = \mathbf{E}\left((X-\mu)^j\right)$	$m_j = \frac{1}{n}\sum\limits_{i=1}^{n} (x_i - \overline{x})^j$

y despejar del sistema de ecuaciones

$$\alpha_j\left(\theta_1, \ldots, \theta_k\right) = a_j \qquad j = 1, 2, \ldots, k$$

el valor de $(\theta_1, \ldots, \theta_k)$. Este método puede dar estimadores no razonables; la única propiedad de la que gozan estos estimadores es la de la consistencia.

Ejemplo 6. *Dada una m.a.* $(X_1, \ldots, X_n)$ *de una v.a.* $X \sim U\,[0, b]$, *estimar el parámetro por el método de los momentos. Calcular su valor exacto para la muestra 2, 3, 4, 10 y 1.*

En primer lugar se igualan los primeros momentos centrados en el origen

$$\left.\begin{array}{r} \alpha_1 = \mathbf{E}\left(X\right) = \dfrac{b-0}{2} = \dfrac{b}{2} \\ a_1 = \frac{1}{n}\sum_{i=1}^{n} n_i x_i = \overline{X} \end{array}\right\} \Rightarrow \frac{b}{2} = \overline{X} \Rightarrow \boxed{\widehat{b} = 2\overline{X}}$$

Para los valores numéricos concretos de la muestra, la estimación del parámetro b resulta ser

$$\widehat{b} = 2 \cdot \frac{2+3+4+10+1}{5} = 2 \cdot \frac{20}{5} = 2 \cdot 4 = 8$$

Ésta no es una buena estimación de b puesto que hay una observación, $x_4 = 10$ *que no está en el intervalo* $[0, b] = [0, 8]$.

3.6.2 Método de Máxima Verosimilitud

La idea de la estimación máximo verosímil de los parámetros $\theta_1, \ldots, \theta_k$ que caracterizan a la v.a. X es elegir el valor o valores del parámetro que hacen que la muestra observada $x_1, \ldots x_n$ sea la más verosímil.

Departamento de Estatística e Investigación Operativa
Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense
Tel. 988 368 766
cotos@uvigo.es

Ejemplo 7. *En una bolsa se tienen 6 bolas, unas blancas y otras negras, y se desea estimar la proporción de blancas que hay. Ésta puede ser*

$$p = \frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6} \text{ó } \frac{5}{6}$$

Se extraen dos bolas con reemplazamiento y se obtienen una de cada color. El valor de p que hace más verosímil este suceso (el que otorga más probabilidad al hecho de tener una bola de cada color en la muestra) es $p = 1/2$, de modo que la estimación de máxima verosimilitud de p es $\widehat{p} = 1/2$.

Definición 12. *Sea $(X_1, \ldots, X_n)$ una m.a. de una población X cuya distribución depende de un parámetro θ que se desea estimar. En el caso discreto, se denomina* ***función de verosimilitud*** *a*

$$L(X_1 = x_1, \ldots, X_n = x_n; \theta) = \mathbf{P}_\theta (X_1 = x_1, \ldots, X_n = x_n) \stackrel{iid}{=} \mathbf{P}_\theta (X_1 = x_1) \cdots \mathbf{P}_\theta (X_n = x_n)$$

Si la población X es continua, la ***función de verosimilitud*** *se define por*

$$L(X_1 = x_1, \ldots, X_n = x_n; \theta) = f_\theta (x_1, \ldots, x_n) \stackrel{iid}{=} f_\theta (x_1) \cdots f_\theta (x_n)$$

donde $f_\theta(\cdot)$ es la función de densidad de X.

Definición 13. *El* ***estimador de máxima verosimilitud*** *de θ es aquel que hace máxima la función de verosimilitud.*

Puesto que la función logaritmo es monótona creciente, resulta a menudo más cómodo trabajar con $\ln f_\theta (x_1, \ldots, x_n)$ en lugar de con $f_\theta (x_1, \ldots, x_n)$. Los **pasos a seguir**, cuando la función es diferenciable con repecto al parámetro θ, son los correspondientes al cálculo de extremos relativos de una función y son los siguientes:

- Calcular la función de verosimilitud $L(\theta) = f_\theta (x_1, \ldots, x_n)$.
- Aplicar logaritmos neperianos a la función de verosimilitud

$$\ln f_\theta (x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln f_\theta (x_i)$$

- Obtener la derivada parcial $\dfrac{\partial}{\partial \theta} \ln f_\theta (x_1, \ldots, x_n)$
- Igualar $\dfrac{\partial}{\partial \theta} \ln f_\theta (x_1, \ldots, x_n) = 0$ y despejar θ. Así obtenemos un valor θ_0 que tendremos que comprobar si es máximo.
- Verificar que $\left. \dfrac{\partial^2}{\partial \theta^2} \ln f_\theta (x_1, \ldots, x_n) \right|_{\theta=\theta_0} < 0.$

Si se cumplen estas condiciones al estimador para θ obtenido se denota por $\hat{\theta}_{MV}$

PROPIEDADES

Para distribuciones cuyo rango posible de valores es conocido a priori y no depende de ningún parámetro θ, puede demostrarse que los estimadores obtenidos por el método de máxima verosimilitud verifican:

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

- **Insesgadez asintótica:** los estimadores no tienen por qué ser insesgados, aunque siempre son asintóticamente insesgados
- **Normalidad asintótica:** los estimadores tienen una distribución asintóticamente normal.
- **Suficiencia:** son función de un estimador suficiente, aunque no tienen por qué ser ellos mismos suficientes.
- **Eficiencia:** si existe un estimador insesgado y eficiente, entonces ese estimador es el único estimador de máxima verosimilitud.
- **Invarianza:** si $T = \widehat{\theta}$ es el estimador de MV de θ y $g(x)$ es una función biyectiva, entonces $g(T)$ es el estimador de MV de $g(\theta)$.
- **Consistencia:** bajo condiciones generales, los estimadores de MV son consistentes (si no son insesgados, son asintóticamente consistentes).

Universidade de Vigo

Departamento de Estatística e Investigación Operativa

Grao en Enxeñaría Informática
Campus de Ourense
E-32004 Ourense

Tel. 988 368 766
cotos@uvigo.es

Tabla Resumen

Estimadores para una muestra aleatoria $\{X_1, X_2, \cdots, X_n\}$ de una población Normal de media μ y desviación típica σ^2:

Parámetro	Estimador		Distribución	Estadístico Pivotal	Distribución
μ	$\overline{X}$	σ^2 conocida	$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$	$\sqrt{n}\frac{\overline{X}-\mu}{\sigma}$	N(0,1)
μ	$\overline{X}$	σ^2 desconocida		$\sqrt{n}\frac{\overline{X}-\mu}{\widehat{S}_X}$	t_{n-1}
σ^2	$\frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{n}$	μ conocida		$\frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}$	χ^2_n
σ^2	$\widehat{S}^2_X$	μ desconocida		$\frac{(n-1)\widehat{S}^2_X}{\sigma^2}$	χ^2_{n-1}

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

Estimadores para muestras grandes (generalmente $n > 30$) y por lo tanto con distribución del estimador aproximada:

Parámetro	Estimador	Distribución	Estadístico Pivotal	Distribución
μ	$\overline{X}$	$\overline{X} \underset{\text{approx}}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$	$\sqrt{n}\frac{\overline{X}-\mu}{S_X}$	N(0,1)
p	$\hat{p}$	$\hat{p} \underset{\text{approx}}{\sim} N\left(p, \sqrt{\frac{p(1-p)}{\sqrt{n}}}\right)$	$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$	N(0,1)
θ	$\hat{\theta}_{MV}$	$\hat{\theta}_{MV} \underset{\text{approx}}{\sim} N\left(\theta, \sqrt{var(\hat{\theta}_{MV})}\right)$	$\frac{\hat{\theta}-\theta}{\sqrt{var(\hat{\theta})}}$	N(0,1)

donde $\hat{\theta}_{MV}$ es el estimador obtenido por Máxima Verosimilitud.