

Dpto. de Economía Cuantitativa
Universidad Complutense de Madrid
Introducción a la Econometría

Tema 4 — Estimación y muestreo

Marcos Bujosa y Gustavo A. Marrero

Material de apoyo para el curso *Introducción a la Econometría* de la licenciatura en Economía de la Universidad Complutense de Madrid.

© 2003–2007 Marcos Bujosa y Gustavo A. Marrero
Actualizado el: 10 de abril de 2007

Versión 4.0

Copyright © 2003–2007 Marcos Bujosa y Gustavo A. Marrero



Algunos derechos reservados. Esta obra está bajo una licencia Reconocimiento-CompartirIgual de Creative Commons. Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by-sa/2.5/es/deed.es> o envíe una carta a Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

Puede encontrar la última versión de este material en:

<http://www.ucm.es/info/ecocuan/mbb/index.html#ietria>

Índice

Índice	1
1. Introducción	2
2. Inferencia paramétrica	3
3. Estimación puntual y por intervalos (repaso)	5
4. Distribuciones asociadas al muestreo (repaso)	7
4.1. Distribución de la media muestral	7
4.2. Distribución de la varianza y la cuasi-varianza muestral	9
4.3. Distribución de la media muestral cuando la varianza NO es conocida	11
4.4. Otras distribuciones	11
5. Bibliografía	13
A. Relación entre la distribución de Poisson y la Binomial.	13

Este es un material de apoyo a las clases. En ningún caso sustituye a los libros de texto que figuran en el programa de la asignatura; textos que el alumno debe estudiar para afrontar el examen final con ciertas garantías de éxito.

El programa se cubre con los siguientes capítulos de libro de texto [Novales \(1997\)](#)¹:

Capítulos 1 a 3: Estos temas han sido cubiertos en asignaturas anteriores, y debido a su bajo nivel de complejidad no se verán en clase (aunque forman parte del programa).

Capítulos 4 a 6: Estos temas han sido cubiertos en las asignaturas [Estadística I](#) y [II](#). Se realizará un breve repaso en clase (una semana o semana y media como máximo), asumiendo que el alumno es capaz de preparar por su cuenta esta parte.

Capítulos 7 y 8: completos

Capítulo 9: secciones 9.4 a 9.6

¹Otros excelentes manuales en castellano son [Peña \(2001\)](#), [Peña \(2002\)](#) y [Peña y Romo \(1997\)](#).

Capítulos 10 y 12: completos**Tema 4. Repaso a las Distribuciones de Funciones de Variables Aleatorias. Distribuciones Asociadas al Muestreo. Distribuciones Asintóticas**

(1 semana como mucho, se supone que los alumnos ya conocen parte o todo)

Capítulos 8 y 9 del manual de Alfonso Novales “Estadística y Econometría”

1. Introducción.
2. Distribución de combinaciones lineales de variables aleatorias
3. Distribución de la media muestral. El caso Normal.
4. Distribución de combinaciones lineales de variables Normales independientes.
5. Distribuciones asociadas a Normales tipificadas. Las distribuciones chi-cuadrado y t de Student.
6. Distribución de la varianza y cuasivarianza muestrales en poblaciones Normales
7. Distribución de diferencias de medias muestrales
8. Distribución F de Fisher-Snedecor
9. Convergencia de sucesiones de variables aleatorias. Ley de los grandes números, teorema de Slutsky y Khintchine y Teorema Central del límite.

Capítulos 8 y 9 de [Novales \(1997\)](#), también puede consultar [Peña \(2001, Capítulos 5, 7 y 8\)](#); [López Cachero \(1992, Capítulos 23, 24, 25, 26, y 29\)](#); [Mittelhammer \(1996, Capítulos 5, y 6\)](#); [Casella y Berger \(2002, Capítulo 5\)](#); [Spanos \(1986, Capítulos 11 y 12\)](#); [Spanos \(1999, Capítulos 11 y 12\)](#); [Rao \(2002, Sección 3.b\)](#)

1. Introducción

↑	<u>De $(S, \mathfrak{B}, P(\cdot))$ a un modelo de probabilidad: historia en símbolos</u>	1
$S \rightarrow \mathbb{R}_X$ $(\mathfrak{B}, P(\cdot)) \rightarrow \{f_x(x; \theta), \theta \in \Theta\}$		
<p>Llamamos Soporte al conjunto \mathbb{R}_X de posibles valores de la v.a. Llamamos Modelo de Probabilidad a la familia o colección de funciones de densidad</p> $\Phi = \{f_x(x; \theta), \theta \in \Theta, x \in \mathbb{R}_X\}$ <p>Llamamos Espacio Paramétrico al conjunto $\theta \in \Theta$ de posibles valores de los parámetros de las funciones de densidad.</p>		

↑	<u>Modelado empírico</u>	2
<ol style="list-style-type: none"> 1. Postulamos a priori una familia de densidades como mecanismo estocástico subyacente. <ol style="list-style-type: none"> a) debemos elegir aquel Modelo de Probabilidad más adecuado a los datos <ul style="list-style-type: none"> ▪ histograma $\rightarrow f_x(x; \theta)$ ▪ datos pueden restringir \mathbb{R}_X y espacio paramétrico, Θ b) Elegimos una familia (no una función en particular) $\Phi = \{f_x(x; \theta), \theta \in \Theta, x \in \mathbb{R}_X\}$ <p style="margin-left: 20px;">donde los parámetros son desconocidos.</p> 2. Inferencia estadística \Rightarrow modelo de probabilidad 		
		(Tema 4 y siguientes)

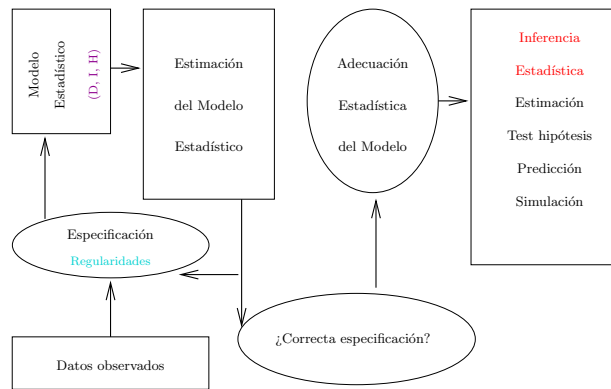


Figura 1: Esquema del proceso de modelado empírico

Inferencia estadística 3

Estudiar propiedades y/o relaciones estocásticas de poblaciones

X : Renta familia Madrileña
 Y : Renta familia Barcelonesa

Podemos preguntarnos por

$E(Y) = 10000?$; $E(Y) \in (8000, 15000)?$;
 $E_{Y|X}(Y | 5000) > E(Y)?$;
 $Var(Y) > Var(X)$; $P(Y > X)$; $f_Y(y) \sim N(\mu_Y, \sigma_Y^2)$

Conociendo $f_{XY}(x, y; \theta)$ respuesta inmediata.

Pero...

A veces sólo conocemos $f_{XY}(x, y)$ pero no θ (Inferenc. paramétrica)

A veces ni siquiera conocemos $f_{XY}(x, y)$ (Inferenc. no paramétrica)

Esquema Parte de Inferencia 4

- Estimación puntual vs Intervalo de Confianza
- Contraste de hipótesis (“Tª de la decisión”)
 - Paramétricos: θ en $f_X(x; \theta)$
 - No paramétricos: (Distribución, Homogeneidad, Independencia)

Inferencia estadística obtiene conclusiones a partir de los datos observados (la información muestral). Mediante distintos procedimientos deducimos la ley estocástica que ha generado los datos observados.

2. Inferencia paramétrica

Parámetros y momentos: ¿Qué relación hay? 5

Momentos Teóricos (o momentos de una distribución): valor esperado de funciones $g(X)$

$$E(g(X)) = \int_{\mathbb{R}_X} g(x) \cdot f_X(x; \theta) dx$$

por tanto son funciones $h(\theta; x)$ de los parámetros

$$E(g(X)) = h(\theta; x)$$

Inferencia: sobre valor de θ mediante *momentos muestrales* (dada x).

Los momentos de una distribución se definen en función de la esperanza matemática de funciones $g(\mathbf{X})$; es decir, en el caso continuo:

$$E(g(\mathbf{X})) = \int_{x \in \mathbb{R}_{\mathbf{X}}} g(x) \cdot f_{\mathbf{X}}(x; \boldsymbol{\theta}) dx,$$

y en el caso discreto;

$$E(g(\mathbf{X})) = \sum_{x \in \mathbb{R}_{\mathbf{X}}} g(x) \cdot f_{\mathbf{X}}(x; \boldsymbol{\theta}) \equiv \sum_{x \in \mathbb{R}_{\mathbf{X}}} g(x) \cdot P_{\mathbf{X}}(x; \boldsymbol{\theta}).$$

Por tanto, los *momentos teóricos* son función de los parámetros $\boldsymbol{\theta}$ de las funciones de densidad.

A su vez, los *momentos teóricos* están relacionados con los *momentos muestrales* de la estadística descriptiva (media, varianza, etc.)

Todo esto nos permite indagar sobre los valores de los parámetros *desconocidos* $\boldsymbol{\theta}$ a través de las medidas de estadística descriptiva.

Eligiendo formas específicas para $g(\mathbf{X})$:

$$g(\mathbf{X}) = \mathbf{X}^r; \quad g(\mathbf{X}) = (\mathbf{X} - E(\mathbf{X}))^r; \quad g(\mathbf{X}) = e^{t\mathbf{X}}; \dots$$

donde $r = 1, 2, \dots$ y $t \in (-k, k)$; obtenemos distintas $h(\boldsymbol{\theta})$ relacionadas con **momentos** de $f_{\mathbf{X}}(x; \boldsymbol{\theta})$.

⬆
Inferencia estadística
6

Llamamos **modelo muestral** de tamaño n , \mathbf{X} , a

$$\mathbf{X} \equiv (X_1, X_2, \dots, X_n)$$

y llamamos **muestra** a una realización de \mathbf{X}

$$\mathbf{x} \equiv (x_1, x_2, \dots, x_n) \in \mathbb{R}_{\mathbf{X}}^n.$$

Supondremos los datos observados como realizaciones del **modelo muestral** (i.e., una muestra).

⬆
Inferencia estadística
7

Modelo estadístico

1. Modelo probabilístico

$$f_{\mathbf{X}_i}(x_i; \boldsymbol{\theta})$$

2. Modelo muestral

$$\mathbf{X} = (X_1, X_2, \dots, X_n)$$

\Rightarrow

Distribución conjunta de \mathbf{X}

$$f_{\mathbf{X}}(\mathbf{x}, \boldsymbol{\theta})$$

\Downarrow

Datos observados (muestra)

$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$

Realización de \mathbf{X}
(un punto del soporte $\mathbb{R}_{\mathbf{X}}^n$)

\Rightarrow

Inferencia estadística

- Estimación
- Contrastación
- Predicción

Recuérdese el ejemplo de la geometría euclídea (*modelo estadístico*), la pared los datos) y la indagación sobre la longitud de su diagonal (*valor de ciertos parámetros*) del **Tema 1**.

Estadístico 8

Llamamos *estadístico* a una función del modelo muestral

$$g(X_1, X_2, \dots, X_n) \equiv g(\mathbf{X});$$

es por tanto *v.a.*

Sus momentos respecto al origen son

$$E([g(\mathbf{X})]^r) = \int \cdots \int_{\mathbb{R}_x^n} [g(\mathbf{x})]^r f_{\mathbf{X}}(\mathbf{x}, \boldsymbol{\theta}) dx_1 \cdots dx_n$$

donde $\mathbf{x} \equiv (x_1, x_2, \dots, x_n) \in \mathbb{R}_x^n$.

Muestreo aleatorio simple 9

El *muestreo aleatorio simple (m.a.s.)* supone que X_i son idéntica e independientemente distribuidas

$$X_i \sim \text{iid}; \quad i = 1, \dots, n.$$

Por tanto, $f_{X_1}(x) = f_{X_2}(x) \cdots = f_{X_n}(x) = f_X(x)$;
es decir, cada x_i proviene de una misma población.
Además (por la independencia)

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

y

$$M_{X_1, \dots, X_n} = \prod_{i=1}^n M_{X_i} = (M_X)^n$$

3. Estimación puntual y por intervalos (repaso)

Estimación 10

Supongamos una muestra aleatoria simple (*m.a.s.*)

$$\mathbf{x} \equiv (x_1, x_2, \dots, x_n);$$

realización del muestreo aleatorio simple (modelo muestral)

$$\mathbf{X} \equiv (X_1, X_2, \dots, X_n) \text{ con } f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta}).$$

Buscamos un valor para el vector $\boldsymbol{\theta}$ que “*de algún modo*” es más acorde con la muestra \mathbf{x} .

Estimador puntual 11

\mathbf{x} es *una* de las posibles realizaciones de \mathbf{X} .
Buscamos una regla que elija un valor de $\boldsymbol{\theta}$ para cada \mathbf{x} .

estimación: $\hat{\boldsymbol{\theta}} = g(x_1, x_2, \dots, x_n) = g(\mathbf{x})$

- (*Resumen de algún aspecto de la información muestral*)

estimador: $\hat{\boldsymbol{\theta}} = g(X_1, X_2, \dots, X_n) = g(\mathbf{X})$

Estimador puntual de la esperanza 12

Primera muestra: 10, 12, 11 → $\bar{x} = 11$
 Segunda muestra: 9, 13, 10 → $\bar{x} = 10.6$
 Tercera muestra: 10, 14, 12 → $\bar{x} = 12$

- Modelo muestral: (X_1, X_2, X_3)
- Estimador: $\hat{\theta} = (X_1 + X_2 + X_3)/3$
- Primera muestra: $\mathbf{x} = (10, 12, 11)$
- Estimación con primera muestra: $\hat{\theta} = 11$
(es un valor puntual, una realización de $\hat{\theta}$)

Boxmodels (Sumas y medias)

Estimador puntual 13

Existen infinidad de posibles estimadores.

- problema de la elección (*Insegadez, eficiencia*).
 - Máxima verosimilitud
 - Mínimos cuadrados
 - Método de los momentos
 - Otros.

Estimación puntual es *tan sólo* una realización de $\hat{\theta}$.
 Estimación puntual es *poco* informativa.
 Podemos decir más si conocemos la distribución de $\hat{\theta}$.

Estimador por intervalos 14

Buscamos una regla que elija, para cada \mathbf{x} , un conjunto de valores que pueda contener el verdadero valor de θ con cierta probabilidad.

Espacio muestral Espacio paramétrico

Deseamos que:

- el conjunto sea pequeño (precisión)
- la probabilidad de que *el verdadero* θ pertenezca al conjunto sea grande (confianza)

↑
Estimador por intervalos
15

$$\left[h(\hat{\theta}_1), h(\hat{\theta}_2) \right] = g(\mathbf{X}_{(n)})$$

de tal manera que

$$P\left(h(\hat{\theta}_1) \leq \theta \leq h(\hat{\theta}_2) \right) = (1 - \alpha)$$

donde

- $(1 - \alpha)$ es el nivel de confianza.

Buscamos que $(1 - \alpha)$ sea grande; y $(\hat{\theta}_2 - \hat{\theta}_1)$ sea mínimo.

(Generalmente se fija $(1 - \alpha)$ y se busca que $(\hat{\theta}_2 - \hat{\theta}_1)$ sea mínimo.)

Necesitamos la distribución de $g(\mathbf{X}_{(n)})$

4. Distribuciones asociadas al muestreo (repaso)

4.1. Distribución de la media muestral

↑
Distribución de la media muestral: bajo normalidad
16

ESTIMADOR DE LA ESPERANZA CUANDO LA VARIANZA ES CONOCIDA

Sea $X \sim N(\mu, \sigma^2)$ con σ^2 conocida.

Emplearemos como estimador de $E(X)$ la media aritmética

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

donde

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

o tipificando

$$\frac{\hat{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Ejemplo 1.

DISTRIBUCIÓN DEL ESTIMADOR DE MEDIA MUESTRAL BAJO NORMALIDAD CUANDO LA VARIANZA ES CONOCIDA

Sea \mathbf{x} *m.a.s.* entonces $X_i \sim$ i.i.d. Vamos a derivar la distribución por dos procedimientos

1. Mediante la función generatriz de momentos

$$\begin{aligned} M_{\hat{x}}(t) &= E\left(e^{t\hat{x}}\right) = E\left(e^{t(X_1 + \dots + X_n)/n}\right) = E\left(e^{\frac{t}{n}X_1 + \dots + \frac{t}{n}X_n}\right) \\ &= M_{X_1, \dots, X_n}(t/n) \underset{\substack{\uparrow \\ \text{indep.}}}{=} \prod_{i=1}^n M_{X_i}(t/n) \underset{\substack{\uparrow \\ \text{indent. dist.}}}{=} (M_X(t/n))^n \end{aligned}$$

Si, además, $X \sim N(\mu, \sigma^2)$, entonces $M_X(t) = e^{t\mu + \frac{1}{2}(t^2\sigma^2)}$ y

$$M_{\hat{x}}(t) = (M_X(t/n))^n = \left(e^{\frac{t}{n}\mu + \frac{1}{2}\left(\frac{t}{n}\right)^2\sigma^2}\right)^n = e^{t\mu + \frac{1}{2}(t^2\frac{\sigma^2}{n})}$$

por tanto

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

2. Como combinación lineal de Normales

Por una parte, $E(X_i) = \mu$ y $\text{Var}(X_i) = \sigma^2$ ya que \mathbf{x} es *m.a.s.*; entonces

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

y

$$\text{Var}(\hat{\bar{x}}) = \frac{1}{n^2} \sum \text{Var}(X_i) = \frac{1}{n^2} n \sigma^2 = \frac{1}{n} \sigma^2$$

Por otra, puesto que $\hat{\bar{x}}$ es combinación lineal de normales, $\hat{\bar{x}}$ se distribuye normal.

Por tanto

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



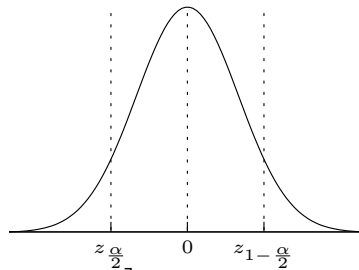
Distribución de la media muestral: bajo normalidad

17

DADO UN NIVEL DE CONFIANZA CALCULAMOS EL INTERVALO

Puesto que $\frac{\hat{\bar{x}} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$.

$$P\left(a \leq \frac{\hat{\bar{x}} - \mu}{\sigma/\sqrt{n}} \leq b\right) = 1 - \alpha$$



$$\mu \in \left[\hat{\bar{x}} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \right]; \quad \text{con probabilidad } 1 - \alpha.$$

La simetría de la función de densidad de una variable aleatoria con distribución $N(0, 1)$ implica que $z_{\alpha/2} = -z_{1-(\alpha/2)}$.



Intervalo de confianza: ejemplo

18

X_i : Renta familia Madrileña i
 $n = 16$; $\sum x_i = 2536$; $X_i \sim N(\mu, 250)$;
 $(1 - \alpha) = 0.95 \implies z_{0.975} = 1.96$

Entonces

$$\left[\frac{2536}{16} \pm \frac{\sqrt{250}}{\sqrt{16}} 1.96 \right] = [151,025; 165,975]$$

Cálculo con muestras A, B, C, ...

$$\begin{aligned} \hat{\bar{x}}^A &= 150,5 && \rightarrow [143,025; 157,975] \\ \hat{\bar{x}}^B &= 164 && \rightarrow [156,525; 171,475] \\ \hat{\bar{x}}^C &= 155 && \rightarrow [147,525; 162,475] \\ \dots &&& \dots \end{aligned}$$

Boxmodels (Intervalos de confianza)

↑
Intervalo de confianza: ejemplo
19

En economía generalmente se dispone de una única muestra (no hay laboratorio)
Si queremos aumentar el nivel de confianza...

$$(1 - \alpha) = 0.99 \implies z_{0.995} = 2.57$$

entonces

$$\left[\hat{x} \pm \frac{\sqrt{\sigma^2}}{\sqrt{n}} 2.57 \right] = [153, 81; 174, 158]$$

menor precisión!!

Dado $1 - \alpha$ podemos aumentar la precisión de dos formas

1. disminuir σ^2 (política de distribución de rentas)
2. aumentar n (tamaño de muestra)

4.2. Distribución de la varianza y la cuasi-varianza muestral

Distribución chi cuadrado: Sea $X \sim N(0, 1)$. ¿Cómo es la distribución de $Y = X^2$?

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\ &= F_X(\sqrt{y}) - (1 - F_X(\sqrt{y})) \\ &= 2F_X(\sqrt{y}) - 1 \end{aligned}$$

y derivando

$$f_Y(y) = 2f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} y^{-\frac{1}{2}}; \quad y \geq 0$$

Que es la función de densidad de una chi cuadrado con un grado de libertad

$$Y \sim \chi_{(1)}^2$$

En general, la función de distribución de una chi cuadrado con r grados de libertad es:

$$f_X(x) = \frac{1}{2^{r/2} \Gamma(r/2)} x^{(r/2)-1} e^{-x/2}; \quad x \geq 0$$

cuyos momentos son

$$\mu = r; \quad \sigma^2 = 2r$$

y cuya función generatriz de momentos es

$$M(t) = (1 - 2t)^{-r/2} \quad \text{para } t < 1/2$$

EJERCICIO 1. Sean X_1, X_2, \dots, X_n independientes y con distribuciones $\chi_{(r_1)}^2, \chi_{(r_2)}^2, \dots, \chi_{(r_n)}^2$ respectivamente. Demuestre que

$$Y = \sum_{i=1}^n X_i \sim \chi_{(r)}^2;$$

donde $r = \sum_{i=1}^n r_i$.

Solución:

Pista. use la función generatriz de momentos de Y .

Ejercicio 1

EJERCICIO 2. Sean X_1, X_2, \dots, X_n independientes y con distribuciones $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2), \dots, N(\mu_n, \sigma_n^2)$ respectivamente. Demuestre que

$$Y = \sum_{i=1}^n \frac{(X_i - \mu_i)^2}{\sigma_i^2} \sim \chi_{(n)}^2.$$

Solución: Es suma de n cuadrados de $X_i \sim N(0, 1)$.

Ejercicio 2

Distribución de la varianza muestral: bajo normalidad 20

ESTIMADOR DE LA VARIANZA
 Sean $X_i \sim \text{iid. } N(\mu, \sigma^2)$, con μ y σ^2 desconocidos.
 Emplearemos como estimador de $\text{Var}(X_i)$ el estadístico:

$$g(X_1, \dots, X_n) = \sum_{i=1}^n \frac{(X_i - \hat{x})^2}{\sigma^2} = \frac{n\hat{s}^2}{\sigma^2} = \frac{(n-1)\hat{s}^2}{\sigma^2} \sim \chi_{(n-1)}^2$$

donde

$$\hat{s}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{x})^2,$$

y

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{x})^2 \quad (\text{insesgado}).$$

Sabemos que $\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi_{(n)}^2$; por lo tanto

$$\begin{aligned} \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 &= \sum_{i=1}^n \left(\frac{X_i - \hat{x} + \hat{x} - \mu}{\sigma} \right)^2 = \\ &= \sum_{i=1}^n \left(\frac{X_i - \hat{x}}{\sigma} \right)^2 + \sum_{i=1}^n \left(\frac{\hat{x} - \mu}{\sigma} \right)^2 + \overbrace{2 \sum_{i=1}^n \frac{(X_i - \hat{x})(\hat{x} - \mu)}{\sigma}}^{=0} = \\ &= \underbrace{\sum_{i=1}^n \left(\frac{X_i - \hat{x}}{\sigma} \right)^2}_{2 \text{ v.a.s indep. (Spanos, 1999, pp. 623)}} + n \left(\frac{\hat{x} - \mu}{\sigma} \right)^2 = \underbrace{\sum_{i=1}^n \left(\frac{X_i - \hat{x}}{\sigma} \right)^2}_{\text{estadístico } g(\mathbf{X})} + \underbrace{\left(\frac{\hat{x} - \mu}{\frac{\sigma}{\sqrt{(n)}}} \right)^2}_{\sim \chi_{(1)}^2} \end{aligned}$$

Así pues, $M_{\chi_{(n)}^2}(t) = E(e^{tg(\mathbf{X})} \cdot e^{t\chi_{(1)}^2}) = E(e^{tg(\mathbf{X})}) \cdot M_{\chi_{(1)}^2}$ por independencia; y entonces

$$(1 - 2t)^{-n/2} = M_{g(\mathbf{X})} \cdot (1 - 2t)^{-1/2} \Rightarrow M_{g(\mathbf{X})} = (1 - 2t)^{-(n-1)/2}$$

es decir, $g(\mathbf{X}) = \sum_{i=1}^n \left(\frac{X_i - \hat{x}}{\sigma} \right)^2 \sim \chi_{(n-1)}^2$

Mientras

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

resulta que

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

ello es debido a que las desviaciones con respecto a la media muestral no son todas independientes ya que

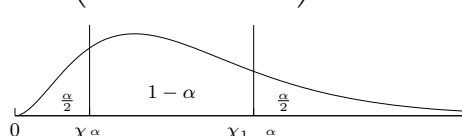
$$\sum (X_i - \hat{x}) = 0$$

Distribución de la varianza muestral: bajo normalidad 21

CÁLCULO DEL INTERVALO Puesto que

$$\frac{(n-1)\hat{s}^2}{\sigma^2} \sim \chi_{(n-1)}^2$$

Fijado $(1 - \alpha)$ buscamos a y b tales que

$$P\left(a \leq \frac{(n-1)\hat{s}^2}{\sigma^2} \leq b\right) = (1 - \alpha)$$


entonces $\sigma^2 \in \left[\frac{\hat{s}^2(n-1)}{\chi_{(n-1, 1-\alpha/2)}^2}, \frac{\hat{s}^2(n-1)}{\chi_{(n-1, \alpha/2)}^2} \right]$ con prob. = $(1 - \alpha)$

4.3. Distribución de la media muestral cuando la varianza NO es conocida

Distribución de la media muestral: bajo normalidad (varianza desconocida) 22

ESTIMADOR DE LA ESPERANZA CUANDO LA VARIANZA NO ES CONOCIDA

Sea X , y sean $\text{Var}(X)$ y $E(X)$ desconocidas.
Emplearemos como estimador de $E(X)$ el estadístico:

$$g(X_1, \dots, X_n) = \frac{\hat{x} - \mu}{\sqrt{\hat{s}^2/n}} \sim t_{n-1}.$$

Si $Z \sim N(0, 1)$ y $V \sim \chi_{(n-1)}^2$; entonces

$$T = \frac{Z}{\sqrt{V/(n-1)}} \sim t_{n-1}.$$

Por tanto,

$$\frac{\frac{\hat{x} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)\hat{s}^2}{\sigma^2(n-1)}}} = \frac{\hat{x} - \mu}{\sqrt{\hat{s}^2/n}} \sim t_{n-1}$$

ya que

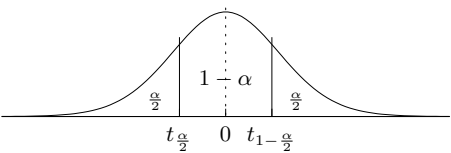
$$\frac{\hat{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1), \quad y \quad \frac{(n-1)\hat{s}^2}{\sigma^2} \sim \chi_{(n-1)}^2$$

Distribución de la media muestral: bajo normalidad (varianza desconocida) 23

CÁLCULO DEL INTERVALO Puesto que

$$\frac{\hat{x} - \mu}{\sqrt{\hat{s}^2/n}} \sim t_{n-1}$$

Fijado $(1 - \alpha)$ buscamos a y b tales que

$$P\left(a \leq \frac{\hat{x} - \mu}{\sqrt{\hat{s}^2/n}} \leq b\right) = (1 - \alpha)$$


entonces

$$\mu \in \left[\hat{x} \pm t_{n-1, \alpha/2} \cdot \sqrt{\hat{s}^2/n} \right] \text{ con Prob. } = (1 - \alpha).$$

La simetría de la función de densidad de una variable aleatoria con distribución t -Student con n grados de libertad, t_n , implica que $t_{n, \alpha/2} = -t_{n, 1-\alpha/2}$.

4.4. Otras distribuciones

teorema central del Límite 24

Sean X_1, X_2, \dots, X_n con $E(X_i) = \mu_i$, con $\text{Var}(X_i) = \sigma_i^2$ con $f_{X_i}(x_i)$ cualquiera; entonces si $Y = X_1 + X_2 + \dots + X_n$

$$\frac{Y - \sum \mu_i}{\sqrt{\sum \sigma_i^2}} \xrightarrow{d} N(0, 1);$$

por lo tanto

$$Y \xrightarrow{d} N\left(\sum \mu_i, \sum \sigma_i^2\right).$$

Distribución F: Sean $X \sim N(\mu_1, \sigma_1^2)$ y $Y \sim N(\mu_2, \sigma_2^2)$ independientes.

Sean dos muestras de tamaños m y n , entonces

$$\frac{\widehat{s}_1^2/\sigma_1^2}{\widehat{s}_2^2/\sigma_2^2} \sim F_{m-1, n-1}.$$

$Y \sim \text{Binomial}(n, p)$ es suma de n v.a. $X_i \sim \text{Bernulli}(1)p$ (con $E(X_i) = p$ y $\text{Var}(X_i) = pq$);² entonces

$$\text{Binomial}(n, p) \xrightarrow{d} N(np, npq)$$

o también

$$\frac{Y - np}{\sqrt{npq}} \xrightarrow{d} N(0, 1)$$

Para proporciones, si definimos $Z = \sum X_i/n$

$$\frac{\frac{\sum X_i}{n} - p}{\sqrt{pq/n}} = \frac{\widehat{x} - p}{\sqrt{pq/n}} \xrightarrow{d} N(0, 1)$$

siempre que n grande y p y q alejados de 0 y 1 aproximación válida

Tablas resumen de varios estadísticos

Parámetro	Estadístico pivote	Distribución	Intervalo de confianza
p	$\frac{\widehat{p} - p}{\sqrt{\frac{\widehat{p}\widehat{q}}{n}}}$	$N(0, 1)$	$\widehat{p} \pm z_{\alpha/2} \sqrt{\frac{\widehat{p}\widehat{q}}{n}}$
$p_1 - p_2$	$\frac{(\widehat{p}_1 - \widehat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\widehat{p}_1 \widehat{q}_1}{n_1} + \frac{\widehat{p}_2 \widehat{q}_2}{n_2}}}$	$N(0, 1)$	$\widehat{p}_1 - \widehat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\widehat{p}_1 \widehat{q}_1}{n_1} + \frac{\widehat{p}_2 \widehat{q}_2}{n_2}}$

Cuadro 1: Intervalos de confianza para proporciones; ($q = 1 - p$)

Parámetro	Estadístico pivote	Distribución	Intervalo de confianza
μ ; σ conocido	$\frac{(\bar{x} - \mu)}{\sqrt{\sigma^2/n}}$	$N(0, 1)$	$\bar{x} \pm z_{\frac{\alpha}{2}} \sqrt{\sigma^2/n}$
μ ; σ desconocido	$\frac{(\bar{x} - \mu)}{\sqrt{\widehat{s}^2/n}}$	t_{n-1}	$\bar{x} \pm t_{\frac{\alpha}{2}} \sqrt{\widehat{s}^2/n}$
σ^2	$(n-1)\widehat{s}^2/\sigma^2$	χ_{n-1}^2	$\frac{\widehat{s}^2(n-1)}{\chi_{1-\frac{\alpha}{2}}^2}, \frac{\widehat{s}^2(n-1)}{\chi_{\frac{\alpha}{2}}^2}$
$\mu_1 - \mu_2$; σ_1, σ_2 conocidas	$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$	$N(0, 1)$	$\bar{x}_1 - \bar{x}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$
$\mu_1 - \mu_2$; $\sigma_1 \neq \sigma_2$	$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\widehat{s}_1^2}{n} + \frac{\widehat{s}_2^2}{m}}}$	t_{n+m-2}	$\bar{x}_1 - \bar{x}_2 \pm t_{\frac{\alpha}{2}} \sqrt{\frac{\widehat{s}_1^2}{n} + \frac{\widehat{s}_2^2}{m}}$
$\mu_1 - \mu_2$; $\sigma_1 = \sigma_2$	$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n-1)\widehat{s}_1^2 + (m-1)\widehat{s}_2^2}{n+m-2} \left(\frac{1}{n} + \frac{1}{m}\right)}}$	t_{n+m-2}	$\bar{x}_1 - \bar{x}_2 \pm t_{\frac{\alpha}{2}}$ denominador
$\frac{\sigma_2^2}{\sigma_1^2}$	$\frac{\widehat{s}_1^2/\sigma_1^2}{\widehat{s}_2^2/\sigma_2^2} = \frac{\widehat{s}_1^2}{\widehat{s}_2^2} \frac{\sigma_2^2}{\sigma_1^2}$	$F_{n-1, m-1}$	$F_a \frac{\widehat{s}_2^2}{\widehat{s}_1^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq F_b \frac{\widehat{s}_2^2}{\widehat{s}_1^2}$

Cuadro 2: Intervalos de confianza para poblaciones normales

²donde $q = (1 - p)$

Lista de Trasparencias

- 1 De $(S, \mathfrak{B}, P(\cdot))$ a un modelo de probabilidad: historia en símbolos
- 2 Modelado empírico
- 3 Inferencia estadística
- 4 Esquema Parte de Inferencia
- 5 Parámetros y momentos: ¿Qué relación hay?
- 6 Inferencia estadística
- 7 Inferencia estadística
- 8 Estadístico
- 9 Muestreo aleatorio simple
- 10 Estimación
- 11 Estimador puntual
- 12 Estimador puntual de la esperanza
- 13 Estimador puntual
- 14 Estimador por intervalos
- 15 Estimador por intervalos
- 16 Distribución de la media muestral: bajo normalidad
- 17 Distribución de la media muestral: bajo normalidad
- 18 Intervalo de confianza: ejemplo
- 19 Intervalo de confianza: ejemplo
- 20 Distribución de la varianza muestral: bajo normalidad
- 21 Distribución de la varianza muestral: bajo normalidad
- 22 Distribución de la media muestral: bajo normalidad (varianza desconocida)
- 23 Distribución de la media muestral: bajo normalidad (varianza desconocida)
- 24 teorema central del Límite
- 25 Partes del temario

5. Bibliografía

- Casella, G. y Berger, R. L. (2002). *Statistical Inference*. Duxbury Advanced Series. Duxbury, USA, segunda ed. ISBN 0-534-24312-6. [2](#)
- López Cachero, M. (1992). *Fundamentos y métodos de estadística*. Ediciones Pirámide, Madrid. ISBN 84-368-0425-2. [2](#)
- Mittelhammer, R. C. (1996). *Mathematical Statistics for Economics and Business*. Springer-Verlag, New York, primera ed. ISBN 0-387-94587-3. [2](#)
- Novalés, A. (1997). *Estadística y Econometría*. McGraw-Hill, Madrid, primera ed. ISBN 84-481-0798-5. [1](#), [2](#), [14](#)
- Peña, D. (2001). *Fundamentos de Estadística*. Alianza Editorial, Madrid. ISBN 84-206-8696-4. [1](#), [2](#), [14](#)
- Peña, D. (2002). *Regresión y diseño de experimentos*. Alianza Editorial, Madrid. ISBN 84-206-8695-6. [1](#)
- Peña, D. y Romo, J. (1997). *Introducción a la Estadística para la Ciencias Sociales*. McGraw-Hill, Madrid. ISBN 84-481-1617-8. [1](#)
- Rao, C. R. (2002). *Linear Statistical Inference and Its Applications*. Wiley series in probability and statistics. John Wiley & Sons, Inc., New York, segunda ed. ISBN 0-471-21875-8. [2](#)
- Spanos, A. (1986). *Statistical foundations of econometric modelling*. Cambridge University Press. ISBN 0-521-26912-1. [2](#)
- Spanos, A. (1999). *Probability Theory and Statistical Inference. Econometric Modeling with Observational Data*. Cambridge University Press, Cambridge, UK. ISBN 0-521-42408-9. [2](#), [10](#)

A. Relación entre la distribución de Poisson y la Binomial.

Sea $X \sim \text{Poisson}(\lambda)$ una v.a. que mide el número de sucesos en un periodo fijo T (por ejemplo N° de llamadas recibidas en 10 horas). Podemos dividir el tiempo total T en intervalos cada vez más pequeños (10 periodos de una hora, 600 periodos de un minuto, 36000 periodos de un segundo, ...). Entonces podemos considerar X como una binomial donde p es la probabilidad del suceso en cada uno de los intervalos, y el número de repeticiones $n = 10$ en el primer caso, $n = 600$ en el segundo caso, ...; donde

$$E(X) = np$$

si llamo $\lambda = np$, entonces $p = \lambda/n$.

Por tanto, podemos aproximar una binomial a una poisson del siguiente modo:

$$P_x(x) = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

y tomando límites

$$\lim_{n \rightarrow \infty} P_x(x) = \frac{\lambda^x}{x!} \lim_{n \rightarrow \infty} \frac{n(n-1) \cdots (n-x+1)}{\left(1 - \frac{\lambda}{n}\right)^x n^x} \left(1 - \frac{\lambda}{n}\right)^n$$

donde

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

y

$$\lim_{n \rightarrow \infty} \frac{n}{(n-x)} \cdot \frac{(n-1)}{(n-\lambda)} \cdots \frac{(n-x+1)}{(n-\lambda)} = 1;$$

Así pues, en el límite

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

que es la función de cuantía de una Poisson (Demostación de [Peña, 2001](#), pp. 172) (Véase también [Novales, 1997](#), pp.210).

Partes del temario

- Tema 1 [IntEctr-T01](#)
- Tema 2 [IntEctr-T02](#)
- Tema 3 [IntEctr-T03](#)
- Tema 4 [IntEctr-T04](#)
- Tema 5 [IntEctr-T05](#)
- Tema 6 [IntEctr-T06](#)
- Tema 7 [IntEctr-T07](#)