ESTAD _ T22. MÉTODOS de ESTIMACIÓN.

MT. de los MOMENTOS.

MT. de la MÍNIMA χ^2.

MT. de la MÍNIMA VARIANZA.

MT. de los MÍNIMOS CUADRADOS.

MT. BAYESIANOS.

1_ MÉTODOS de ESTIMACIÓN.

Anteriormente, hemos estudiado las propiedades deseables de un buen estimador, (insesgadez, consistencia, suficiencia, eficiencia, robustez, etc).

Ahora necesitamos disponer de un criterio "objetivo" que permita obtener de entre infinitos estimadores el "más razonable" para luego deducir la bondad del estimador a partir de las propiedades que verifique.

Cada uno de los procedimientos tiene una base de partida distinta:

- el de los momentos utiliza el conocimiento de la distrib. de probabilidad poblacional,

COMPLETAR

0. INFERENCIA. INTRODUCCIÓN
~~1. ESTIMACIÓN PUNTUAL I.~~

Recordemos brevemente que la Inferencia Estadística consiste en ~~generalizar~~ (sacar conclusiones) sobre la poblac. que nos interesa estudiar a partir ~~de~~ la información que nos proporcione una muestra aleatoria basándonos en la Teoría de la Probabilidad.

Si estamos interesados en ~~con~~ estudiar el valor de una característica poblacional θ, la Inferencia que podemos utilizar es:
- Estimación
- Contrastación

La Estimación consiste en dar un valor aproximado del parám. poblacional a partir de la información ~~proporcionada~~ muestral.
La Contrastación consiste es formular una conjetura (hipótesis) sobre el valor del parámetro poblacional y utilizar la inform. muestral para aceptar ó rechazar dicha hipótesis.

En el caso de la estimación, se puede aproximar:
- Estim. puntual → valor concreto
- Estim. por intervalos de confianza → intervalo

En ambos métodos se utiliza un estadístico (función real de la muestra) para estimar.
La ventaja de la estimación puntual es que da un valor concreto del estimador que puede sustituirse directamente en el parámetro, pero sólo cuenta con las propiedades como garantía.
La ventaja de la estimación por intervalos es que el intervalo va acompañado de un grado de confianza de que el verdadero valor del parámetro se encuentre dentro del intervalo, pero su inconveniente es que ofrece infinitas soluciones.

1 – ESTIMACIÓN PUNTUAL I.

La estimación puntual consiste en obtener un único número, calculado a partir de las observaciones muestrales, y que es utilizado como estimación del parámetro poblacional θ. Se le llama estimación puntual porque a ese nº, que se utiliza como estimación puntual de θ, se le puede asignar un punto de la recta real.

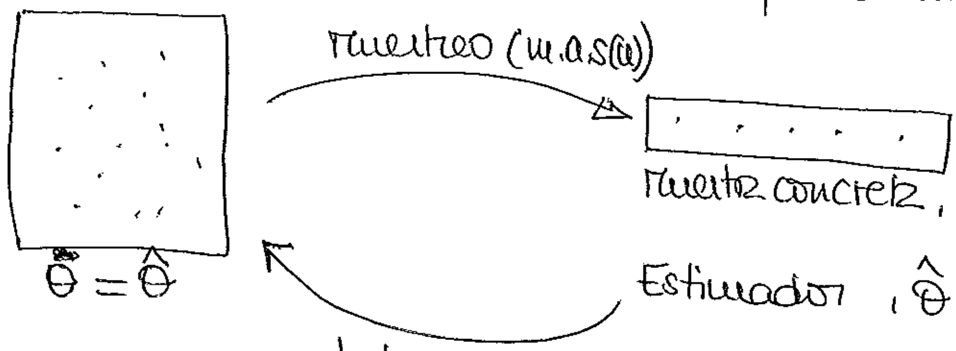
Esquema de la estimac. puntual:

Población

$\xi \longrightarrow F(x,\theta)$ θ desconocido

Muestra

Espacio muestral. $\mathbb{R}^n = \{X_1 \ldots X_n\}$



$\hat{\theta} = \hat{\theta}$

Muestra concreta, $\bar{x} = \{x_1 \ldots x_n\}$

Estimador, $\hat{\theta} = g(x_1 \ldots x_n)$

Inferencia
(Estim. puntual)

Para una población ξ (realización de una v.a.), representada a partir de su función de distribución $F(x,\theta)$, que depende de un parámetro θ cuyo valor concreto se desconoce, se toma una muestra aleatoria (normalmente en poblac. infinitas se utiliza m.a.s.) con n elementos que, al ser el muestreo aleatorio tb será una v.a. A partir de la muestra se construye un estadístico (no depende del parámetro) que seguirá siendo una v.a. que se utiliza para estimar el parámetro por eso se llama estimador.

Para una muestra concreta se obtendrá un valor concreto ~~estimación~~ del estimador, que recibe el nombre de estimación puntual del parámetro poblacional.
Antes de continuar, merece la pena detenerse en los tres conceptos mencionados anteriormente para no confundirlos.

Parámetro → Constante con valor desconocido. Característica ~~Relacionada~~ poblacional, no depende de la muestra y es único (≠ muestras, = parám).

Estimador → Estadístico → función de la muestra que se utiliza para ofrecer una aproximación del parámetro desconocido. Es una variable aleatoria, no una cte → tiene dis

Estimación → Valor de la v.a. estimador para una muestra concreta → es una cte.
≠ muestras, = estimador ⇒ ≠ estimaciones

Para seleccionar el estadístico que utilizaremos como estimador del parámetro poblacional tendremos en cuenta las propiedades de la distribución muestral del estadístico, por lo que:

1º. Estudiar distribución de los estad. en el muestreo.
2º. Conocer las propiedades deseables de los estimadores puntuales, para estudiar su bondad.
3º. Métodos de obtención

Afortunadamente, las propiedades y los métodos nos sugieren los estimadores que se nos ocurren de manera natural.
Para la media poblacional μ, la media muestral $\bar{x}$
Para la varianza poblacional σ^2, una corrección de la varianza muestral, la cuasivarianza muestral (en tamaños infinitos da igual).
Para la proporción poblacional P, la proporción muestral p.

2_ MÉTODO de los MOMENTOS

Fue introducido por K. Pearson y ~~es~~ el método ~~más general~~ general más antiguo y sencillo para la obtención de estimadores de parámetros poblacionales.

En algunas ocasiones se suele utilizar para obtener una primera aproximación de los estimadores.

El mt. consiste en igualar tantos momentos muestrales como parámetros haya que estimar, a los correspondientes momentos poblacionales (m. respecto al origen), y resolviendo el sistema de ecuaciones resultante tendremos los estimadores de los parámetros.

K. Pearson se basó en el tma. de Khintchine → convergencia de los momentos muestrales a los momentos poblacionales, y en que $E[a_r] = \alpha_r$, es decir, cada momento muestral es un estimador insesgado de su su momento poblacional.

De manera formal:

Sea ξ con distrib. de probabilidad $F(x, \theta_1 \dots \theta_K) < \begin{matrix} P(\xi = x_i; \theta_1 \dots \theta_K) \\ f(x; \theta_1 \dots \theta_K) \end{matrix}$

en la que aparecen K parámetros desconocidos.

El momento poblacional de orden r con respecto al origen, α_r se define como:

$$\alpha_r(\theta_1 \dots \theta_K) = E[\xi^r] = \begin{cases} \sum_{i=1}^{\infty} x_i^r \cdot P(\xi = x_i) \\ \int_{-\infty}^{+\infty} x \cdot f(x; \theta_1 \dots \theta_K)\,dx \end{cases}$$

De la población tomamos una m.a.s.(n), $X = \{x_1 \dots x_n\}$ de la que obtenemos los momentos muestrales respecto al origen, a_r:

$$a_r = \sum_{i=1}^{n} \frac{x_i^r}{n}$$

Planteamos el sistema de k ecuaciones con k incógnitas, igual al nº de parámetros, y los estimadores de los parámetros serán la solución del sistema.

$$\left.\begin{array}{l}\alpha_1(\theta_1\dots\theta_K)=a_1\\ \dots\dots\dots\\ \alpha_K(\theta_1\dots\theta_K)=a_K\end{array}\right\} \xrightarrow{\text{Soluc.}} \begin{array}{l}\hat{\theta}_1=\hat{\theta}_1(a_1\dots a_K)\\ \dots\dots\dots\\ \hat{\theta}_K=\hat{\theta}_K(a_1\dots a_K)\end{array}$$

Este mt. no utiliza toda la información poblacional, prescinde de la distrib. poblacional.

El mt. de los momentos permite obtener estimaciones de una función de los momentos poblacionales → Varianza,
$\sigma^2=\mu_2=\alpha_2-\alpha_1^2 \Rightarrow \hat{\sigma}^2=\hat{\mu}_2=\hat{\alpha}_2-\hat{\alpha}_1^2=a_2-a_1^2=S^2$.

PROPIEDADES

P1 _ Insesgadez : Cuando el parámetro poblacional es un momento poblacional, entonces el estimador es insesgado.
*No ocurre con $\theta=\mu_r$.

Dem: $E[\hat{\alpha}_r]=E[a_r]=E\left[\frac{\sum X_i^r}{n}\right]=\frac{n}{n}E[X_i^r]=\alpha_r$.

P2 _ Consistencia : Bajo condiciones bastante generales, los estimador por el mt. momentos son consistentes.

Dem: Por el tma de Khintchine, los momentos muestrales son estimadores consistentes de los momentos poblac.
Por el tma. de Slutsky, una función de los mom. muestrales es un estim. consistente de la corresp. función poblacional

P3 _ Normalidad asintótica : Si los parámetros a estimar son los momentos poblacionales, entonces los estimadores serán asintóticamente normales.

Dem: $E[a_K]=\alpha_K$, $V[a_K]=\frac{\alpha_{2K}-\alpha_K^2}{n}$

Por el tma. de Lindeberg-Lévi:

$$a_K \xrightarrow[n\to\infty]{} N\left(\alpha_K, \sqrt{\frac{\alpha_{2K}-\alpha_K^2}{n}}\right)$$

3_ MÉTODO de la ~~χ^2~~ MÍNIMA χ^2

Es un método general para la obtención de estimadores puntuales. Es de menor aplicación que el mt. de los momentos y que el mt. de máxima verosimilitud.
Solamente se aplica cuando hay una gran cantidad de datos, ya que necesita datos agrupados por intervalos, tanto para V.a. discretas como para V.a. continuas.
Se basa en la minimización de la distancia entre las frecuencias observadas y las frecuencias teóricas, distancia definida mediante la χ^2.

Procedimiento:
Sea ξ, población con función de distrib $F(x,\theta_1 \dots \theta_K) <^{P}_{f}$ que depende de k parámetros desconocidos. ~~Dividi~~ Hacemos una partición del campo de variación de ξ en $S_1 \dots S_r$ subconjuntos excluyentes con probab. $p_1 \dots p_r$.

$$p_i(\theta_1 \dots \theta_K) = P(S_i) = P(\xi \in S_i) \quad \begin{cases} p_i > 0 \\ \sum p_i = 1 \end{cases}$$

$p_i \equiv$ probab. teóricas

Para estimar los parámetros desconocidos se toma una mas(n) y se ordenan sus observaciones según su pertenencia a $S_1 \dots S_r$ con número $n_1 \dots n_r$ / $\sum n_i = n$

Valores de ξ	Frec. absolutas	Frec. relativas	Frec. teóricas
S_1	n_1	n_1/n	p_1
S_2	n_2	n_2/n	p_2
---	---	---	---
S_r	n_r	n_r/n	p_r
	n		

El objetivo es minimizar las ~~diferencias~~ discrepancias entre las frec. observadas y las esperadas.

El estadístico que vamos a utilizar es la suma de los cuadrados de las diferencias entre las frecuencias relativas observadas y las probab. teóricas para cada grupo, ponderadas por el cociente del tamaño muestral y las probab. teóricas.

$$\chi^2 = \sum_{i=1}^{r} \frac{n}{p_i}\left(\frac{n_i}{n} - p_i\right)^2 = \sum \frac{\cancel{n}}{p_i \cdot n^{\cancel{2}}}(n_i - np_i)^2 =$$

$$= \frac{(n_i - np_i)^2}{np_i} \longrightarrow \chi^2_{r-k-1}$$

$$\frac{(O - E)^2}{E}$$

nº clases nº parám.

El mt. de la mínima χ^2 escoge los estimadores de los parámetros θ_i que minimizan la expresión del estadístico χ^2.

$$\min_{\theta_1 \dots \theta_k} \chi^2 \Rightarrow CN \rightarrow \frac{\partial \chi^2}{\partial \theta_i} = 0 \rightarrow \hat{\theta}_1 \dots \hat{\theta}_k$$

soluc. del sistema

Minimizar esta expresión presenta dificultades a la hora de resolver el sistema.

Los estimadores de mínima χ^2 son asintóticamente equiv. a los estimadores de máxima verosimilitud (n pequeño no). Generalmente son estimadores sesgados y no eficientes.

Para facilitar la resolución del sistema, se utiliza el mt. modificado de la mínima χ^2 que ofrece estimadores asintóticamente equivalentes a los anteriores y coinciden con los obtenidos por el mt. de máxima verosimilitud.

4_ MÉTODO de la MÍNIMA VARIANZA

Es un método analítico que consiste en hacer mínima la varianza del estimador.

La técnica que se utiliza para minimizar el estimador condicionado por una serie de restricciones (linealidad, insesgadez, etc.) es la de los multiplicadores de Lagrange.

Aplicaciones:

(1) Estimador lineal, insesgado y de mínima varianza para la media poblacional, $\hat{\mu}$.

Sea ξ / $E[\xi] = \mu$ y $Var[\xi] = \sigma^2 \Longrightarrow \hat{\mu} = \overline{X}$.

Dem:

Lineal $\Rightarrow \hat{\mu} = a_1 X_1 + \dots + a_n X_n$

Insesgadez $\Rightarrow E[\hat{\mu}] = a_1 E[X_1] + \dots + a_n E[X_n] = \mu \sum_{i=1}^{n} a_i$

$E[\hat{\mu}] = \mu \Leftrightarrow \sum a_i = 1$

$Var(\hat{\mu}) = Var[a_1 X_1 + \dots + a_n X_n] = a_1^2 \sigma^2 + \dots + a_n^2 \sigma^2 = \sigma^2 \sum a_i^2$

Problema: $\min_{a_i} \sigma^2 \sum_{i=1}^{n} a_i^2$

s.a. $\sum_{i=1}^{n} a_i = 1$

Definimos el lagrangiano $\phi = \sigma^2 \sum a_i^2 + \lambda (\sum a_i - 1)$ que es la función a minimizar.

CN $\rightarrow \frac{\partial \phi}{\partial a_i} = 0 \Leftrightarrow 2 a_i \sigma^2 + \lambda = 0$, $\forall i = 1 \dots n$

Sumando todas las restricciones, $2\sigma^2 \underbrace{\sum a_i}_{=1} + n\lambda = 0$

luego $\lambda = -\frac{2\sigma^2}{n}$

Volviendo a la CN individual: $2 a_i \sigma^2 - \frac{2\sigma^2}{n} = 0 \Rightarrow a_i = 1/n$

y sustituyendo $\hat{\mu} = \frac{\sum X_i}{n} = \overline{X}$.

(2) Estimador lineal insesgado de b y de mínima varianza para el coeficiente de regresión

Sean Y_i v.a. indep con $E[Y_i] = a + bX_i$ y $Var[Y_i] = \sigma^2$,
entonces $\hat{b} = \dfrac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$

Dem:

Lineal $\rightarrow \hat{b} = c_1 Y_1 + \dots + c_n Y_n = \sum c_i Y_i$

Insesgado $\rightarrow E[\hat{b}] = \dots = \sum c_i (a + bX_i) = a\sum c_i + b \sum c_i X_i$

Insesgado $\Leftrightarrow E[\hat{b}] = b \Leftrightarrow \sum c_i = 0$
$\sum c_i X_i = 1$

$Var(\hat{b}) = Var \sum c_i Y_i = \sigma^2 \sum c_i^2$

$\min Var(\hat{b}) = \min\limits_{c_i} \sigma^2 \sum c_i^2 \sim \min \sum c_i^2$

s.a. (a) $\sum c_i = 0$
(b) $\sum c_i X_i = 1$

Lagrangiano: $\phi = \sum c_i^2 + \lambda_1 (\sum c_i) + \lambda_2 (\sum c_i X_i - 1)$

$\dfrac{\partial \phi}{\partial c_i} = 2c_i + \lambda_1 + \lambda_2 X_i = 0 \Leftrightarrow c_i = \dfrac{-\lambda_1 - \lambda_2 X_i}{2} \sim +\lambda_1 + \lambda_2 X_i$

(a) $\sum c_i = 0 \Rightarrow n\lambda_1 + \lambda_2 \sum X_i = 0 \rightarrow \lambda_1 = -\lambda_2 \dfrac{\sum X_i}{n} = -\lambda_2 \overline{X}$

luego $c_i = -\lambda_2 \overline{X} + \lambda_2 X_i = \lambda_2 (X_i - \overline{X})$.

(b) $\sum c_i X_i = 1 \Rightarrow \sum \lambda_2 (X_i - \overline{X}) \cdot X_i = 1$

$\lambda_2 (\sum X_i^2 - \overline{X} \sum X_i) = 1$

$\lambda_2 (\sum X_i^2 - n\overline{X}^2) = 1$

$\lambda^2 \sum (X_i - \overline{X})^2 = 1 . \Rightarrow \lambda_2 = \dfrac{1}{\sum (X_i - \overline{X})^2}$

$c_i = \dfrac{(X_i - \overline{X})}{\sum (X_i - \overline{X})^2}$

$\hat{b} = \sum c_i Y_i = \dfrac{\sum (X_i - \overline{X}) Y_i}{\sum (X_i - \overline{X})^2} = \dfrac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$

que coincide con el EMV y con EMC.

5_ MÉTODO de los MÍNIMOS CUADRADOS

Introducido por Gauss, generalmente se utiliza para estimar los parámetros de un modelo lineal.

Sea el modelo

$$Y = a + bX + e$$

donde Y es una v.a. con $E[Y] = a + bX$,

X es una var. observable con valores conocidos,

a y b parámetros desconocidos

e es una var. aleatoria o error.

Para cada valor X_i le corresponde un valor teórico Y_i, y el error cometido será:

$$e_i = Y_i - Y'_i = Y_i - a - bX_i \quad , \quad i = 1 \ldots n$$

y se verifica que:

- e_i están incorrelacionados entre sí, $Cov(e_i, e_j) = 0$
- $E[e_i] = 0$
- $Var[e_i] = \sigma^2$

Para estimar los parámetros a y b, se minimiza la distancia global de todos los puntos observados a la función teórica ajustada:

$$\min_{a,b} e_i^2 = \min_{a,b} \sum_{i=1}^{n} (Y_i - a - bX_i)^2$$

$$CN \longrightarrow \begin{cases} \dfrac{\partial \phi}{\partial a} = 0 \Leftrightarrow -2\sum(Y_i - a - bX_i) = 0 \\ \dfrac{\partial \phi}{\partial b} = 0 \Leftrightarrow -2\sum(Y_i - a - bX_i)X_i = 0 \end{cases}$$

$$\longrightarrow \begin{cases} \sum Y_i = na + b\sum X_i \\ \sum Y_i X_i = a\sum X_i + b\sum X_i^2 \end{cases}$$

Dividiendo por n la 1ª ecuación:

$$\hat{a} = \frac{\sum Y_i}{n} - \hat{b}\frac{\sum X_i}{n} = \bar{Y} - \hat{b}\bar{X}$$

y sustituyendo en la segunda ecuación:

$$\hat{b} = \frac{\sum Y_i X_i - \hat{a}\sum X_i}{\sum X_i^2} = \frac{\sum Y_i X_i - \left(\frac{\sum Y_i}{n} - \hat{b}\frac{\sum X_i}{n}\right)\sum X_i}{\sum X_i^2} =$$

$$= \frac{\sum Y_i X_i - \frac{\sum X_i Y_i}{n}}{\sum X_i^2} + \frac{\hat{b}\,\frac{\cancel{\sum X_i^2}}{n}}{\cancel{\sum X_i^2}} \to \left(1 - \frac{1}{n}\right)\hat{b} =$$

$$\hat{b} = \frac{\sum Y_i X_i - \frac{\sum X_i Y_i}{n}}{\sum X_i^2}$$ No me sale.

$$\hat{b} = \frac{\sum (Y_i - \bar{Y})(X_i - \bar{X})}{\sum (Y_i - \bar{Y})^2}$$

6_MÉTODOS ~~5 - ESTIMADORES~~ BAYESIANOS

En Inferencia estadística clásica se supone que los parámetros poblacionales son constantes desconocidas, no aleatorios.

El enfoque inferencial bayesiano considera el parámetro desconocido θ como variable aleatoria, y tiene por tanto una distribución de probabilidad

$g(\theta) \rightarrow$ distribución a priori de θ.

Bajo la óptica bayesiana, la información proporcionada por la muestra X puede cambiar la idea del comportamiento probabilístico de $\theta \Rightarrow$ cambia la distribución de probabilidad de θ

$g(\theta/x) \rightarrow$ distrib. a posteriori de θ cuando se dispone de la información muestral X.

Aplicando el teo. de Bayes pueden determinarse las probab. a posteriori para cada valor de θ:

$$P(\theta=\theta_i/x) = \frac{P(\theta=\theta_i)\cdot P(X/\theta=\theta_i)}{\sum_i P(\theta=\theta_i)\cdot P(X/\theta=\theta_i)} \quad \text{, caso discreto}$$

$$g(\theta/x) = \frac{g(\theta)\cdot f(X/\theta)}{\int_\Theta g(\theta) f(X/\theta)\, d\theta} \quad \text{, caso continuo}$$

El estimador de Bayes es una función de decisión (función de la información muestral X sobre el espacio de las posibles decisiones a adoptar) establecida sobre el espacio paramétrico Θ, que produzca el menor riesgo esperado.

El estimador será $\hat{\Theta} = d^*(X)$ (mejor decisión de Bayes) tal que

$$\min B(d) = \int_\Theta \int_X \ell(\Theta, d(X)) \cdot f(X/\Theta) \cdot g(\Theta)\, d\Theta\, dX$$

donde

$B(d) \equiv$ riesgo de Bayes = valor esperado de la función de riesgo

$\ell(\Theta, d(X)) \equiv$ función de pérdida

$R(\Theta, d) \equiv$ función de riesgo = esperanza de la función de pérdida

El estimador de Bayes se puede obtener minimizando la expresión:

$$\min_{\Theta/X} E\{\ell(\Theta, d(X))\} = \int_\Theta \ell(\Theta, d(X))\, g(\Theta X)\, d\Theta$$

El estimador será aquel que

CN $\rightarrow$ $$\frac{\partial E_{\Theta/X}\{\ell(\Theta, d(X))\}}{\partial d(X)} = 0 \Rightarrow \hat{\Theta} = d^*(X)$$

CS $\rightarrow$ $$\left.\frac{\partial^2 E_{\Theta/X}\{\ell(\Theta, d(X))\}}{\partial d^2(X)}\right|_{d(X)=d^*(X)} > 0$$

P1 $\rightarrow$ Si la función de riesgo es constante respecto a Θ, el estimador de Bayes coincide con el est. minimax

P2 $\rightarrow$ Si la función de pérdida es cuadrática, el estim. de Bayes es el valor medio de la distrib. a posteriori.

P3 $\rightarrow$ Si la función de pérdida es proporcional, $\ell(\Theta, d(X)) = |\Theta - d(X)|$ el estimador de Bayes es la mediana de la distrib. a posteriori de Θ dado X.