ESTAD_T21. ESTIMACIÓN PUNTUAL II.

ESTIMADORES de MÍNIMA VARIANZA.
ESTIMADORES EFICIENTES
ESTIMADORES ROBUSTOS.
ESTIMADORES BAYESIANOS.

0. INFERENCIA. INTRODUCCIÓN
~~1. ESTIMACIÓN PUNTUAL I.~~ (Igual que T20)

Recordemos brevemente que la Inferencia Estadística consiste en ~~generalizar~~ (sacar conclusiones) sobre la poblac. que nos interesa estudiar a partir ~~de las~~ la información que nos proporciona 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

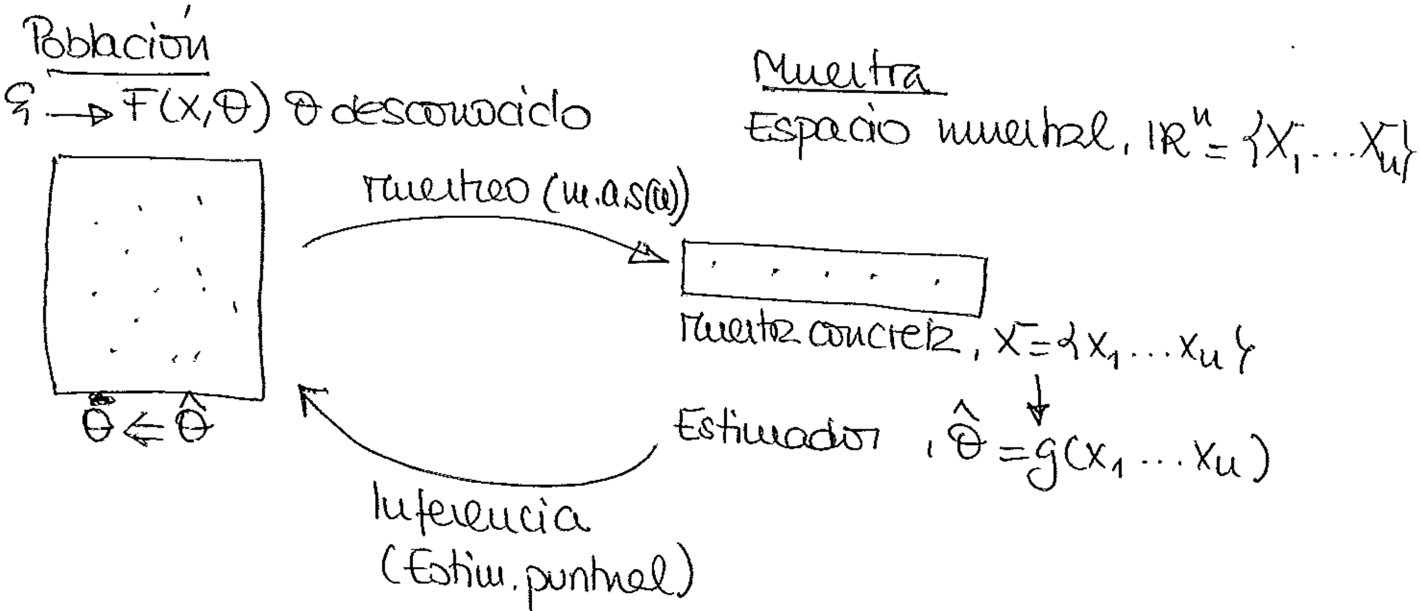
En ambos métodos se utiliza un estadístico (función real de la muestra) para estimar.

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:



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 (~~el~~ 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 ~~estimación~~ valor concreto 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.
~~Relacionado~~ Característica 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 distr.

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. ESTIMADORES de MÍNIMA VARIANZA

Al construir un estimador pretendemos que sus estimaciones puntuales se aproximen lo más posible al valor del parámetro desconocido, que equivale a pretender que el ECM sea mínimo.

$$ECM(\hat{\theta}) = E[(\hat{\theta}-\theta)^2] = E[\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2] = E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2 =$$

$$= \left(E[\hat{\theta}^2] - (E[\hat{\theta}])^2\right) + \left(E[\hat{\theta}]^2 - 2\theta E[\hat{\theta}] + \theta^2\right) =$$

$$= V(\hat{\theta}) + (\hat{\theta}-\theta)^2 = \text{Varianza}(\hat{\theta}) + \text{sesgo}^2(\hat{\theta})$$

En muchas ocasiones no resulta sencillo minimizar el ECM, pero sí es fácil obtener estimadores insesgados ($b(\hat{\theta})=0$), de modo que dados dos estimadores insesgados del mismo parámetro será mejor el que tenga varianza mínima, es decir el que tenga menor ECM.

Por tanto, debemos ~~obtener~~ intentar un estimador, de entre todos los estimadores insesgados, que tenga varianza mínima → estimador insesgado de varianza mínima.

Si además se verifica que la varianza es mínima para todos los valores del parámetro, entonces el estimador recibe el nombre de estimador insesgado uniforme de mínima varianza (UMVUE)

Definición: Diremos que el estimador $\hat{\theta}_0$ es insesgado uniforme de mínima varianza (UMVUE) para el parámetro θ, si dado cualquier otro estimador insesgado $\hat{\theta}$ de él, se verifica:

$$Var(\hat{\theta}_0) \leq Var(\hat{\theta})$$

para todos los posibles valores de θ.

La utilización de la mínima varianza como criterio para elegir el mejor estimador es válido tanto para estimadores insesgados como para estimadores con el mismo sesgo.

También se utiliza el concepto de eficiencia relativa de un estimador respecto a otro.

$$e(\hat{\theta}_1, \hat{\theta}_2) = \frac{ECM(\hat{\theta}_1)}{ECM(\hat{\theta}_2)} \overset{\hat{\theta}_1, \hat{\theta}_2 \text{ insesg}}{=} \frac{V(\hat{\theta}_1)}{V(\hat{\theta}_2)}$$

La eficiencia relativa es siempre >0, porque es un cociente de cantidades positivas.

Una manera de obtener de un EMVUE es con la cota de Frechet-Cramér-Rao, que proporciona una cota inferior de la varianza del estimador.

Cota de Frechet-Cramér-Rao

Sea ξ, población con $F(x,\theta) < \begin{matrix} P(x,\theta) \\ f(x,\theta) \end{matrix}$

Sea $X = (X_1, X_2 \dots X_n)$ m.a.s. de tamaño n.

1º. Calculamos la función de verosimilitud conjunta de la muestra, $\mathcal{L}(X,\theta) = f(x_1 \dots x_n, \theta) = \prod f(x_i, \theta)$.

2º. Calculamos el logaritmo neperiano de la función de verosimilitud conjunta, $\ln \mathcal{L}(X,\theta) = \ln \prod f(x_i,\theta) = \sum \ln f(x_i,\theta)$.

3º. Calculamos la derivada de esta expresión respecto al ~~valor del~~ parámetro θ.

$$\frac{\partial \ln \mathcal{L}(X,\theta)}{\partial \theta}$$

4º. La cantidad de información de Fisher se define como $I(\theta) = E\left[\frac{\partial \ln \mathcal{L}(X,\theta)}{\partial \theta}\right]^2$

Si se cumplen las condiciones de regularidad de Wolfowitz, la varianza del estimador está acotada inferiormente por $1/nI(\theta)$.

$$Var(\hat{\theta}) \geq \frac{1}{n E\left[\frac{\partial \ln \mathcal{L}(X,\theta)}{\partial \theta}\right]^2} = \frac{1}{nI(\theta)}$$

Las condiciones de regularidad de Fisher-Wolfowitz son:

1- El campo de variación, y por tanto la muestra $\bar{X}$, de donde procede la muestra no depende de θ.

2- La función $\ln \mathcal{L}(\bar{X}, \theta)$ admite al menos las dos primeras derivadas respecto a θ.

3- Las operaciones de derivación e integración (suma para las v.a. discretas) son intercambiables.

Observaciones:

O1 → $I(\hat{\theta}) \equiv$ cantidad de información de Fisher, representa la cantidad de información que la muestra proporciona sobre el parámetro.
Cuanto menor es $V(\hat{\theta})$, mayor es la información que proporciona $\hat{\theta}$ de θ.

O2 → La CCR depende del estimador concreto utilizado, a través de su sesgo y de la distrib. estudiada. En el caso de estimadores insesgados únicamente depende de la función de densidad poblacional y del tamaño muestral.

O3 → La CCR permite determinar si un estimador es eficiente ($V(\hat{\theta}) = CCR$), pero puede no existir.

O3 → El valor de la CCR no tiene por qué ser cercano a 0, solamente mínimo.

3_ ESTIMADORES EFICIENTES

El inconveniente de la propiedad de insesgadez es que solamente exige que el valor esperado del parámetro coincida con el estimador, $E[\hat{\Theta}] = \Theta$, pero no requiere que mucho, ni alguna estimación tomen valores próximos al parámetro, que sería muy deseable.

Por eso, además de insesgado, se le pide que tenga la varianza pequeña.

Así, en un grupo de estimadores, el estimador más eficiente será el que tenga menor varianza.

Definición: $\hat{\Theta}$ es eficiente si : 1_ $\hat{\Theta}$ es insesgado
2- $V(\hat{\Theta}) = CCR$

Luego un estimador eficiente es un estimador insesgado uniforme de mínima varianza, UMVUE, cuya varianza coincide con el límite inferior, la Cota de Frechet-Cramér-Rao.

Pero puede que un UMVUE no sea eficiente pues, aunque tenga varianza mínima no alcanza el límite.

Definición: Se define la eficiencia absoluta de un estimador insesgado como el cociente entre la CCR y su varianza.

$$ef.a = \frac{CCR}{V(\hat{\Theta})}$$

alcanzando su máximo (=1) cuando el estimador es eficiente.

Definición: $\hat{\Theta}$ es asintóticamente eficiente si se verifica

$$\lim_{n \to \infty} V(\hat{\Theta}) = CCR$$

- En $\xi \to N(\mu, \sigma)$, $\overline{X}$ es eficiente de μ.
  s^2 no es eficiente de σ^2
- En $\xi \to B(n, p)$, $\hat{p} = \frac{\overline{X}}{n}$ es eficiente
- En $\xi \to \mathcal{P}(\lambda)$, $\hat{\lambda} = \overline{X}$ es eficiente

Propiedades de los estimadores eficientes:

P1 → El estimador eficiente es único.
Eficiente ⇒ suficiente

4_ ESTIMADORES ROBUSTOS

Diremos que un procedimiento estadístico es robusto si su comportamiento es relativamente insensible a desviaciones de las hipótesis iniciales sobre las que se había planteado el procedimiento.

Definición: Un estimador es robusto cuando pequeños cambios en las hptsis de partida del procedimiento de estim. considerado no producen ~~diferencias~~ variaciones significativas en los resultados obtenidos.

Habitualmente, se considera una v.a. con cierta función de distribución, se extrae una m.a.s(n) y se obtiene un estimador $\hat{\theta}$ que depende de F.

Si la distrib. ~~de~~ poblacional es ligeramente distinta, la distrib. del estimador en el muestreo tb. lo será. Lo deseable es que no difiera mucho del primer estimador.

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 $\bar{X}$ 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 $\bar{X}$.

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

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

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

El <u>estimador de Bayes</u> es una función de decisión (función de la información muestral $\bar{X}$ sobre el espacio de las posibles decisiones a adoptar) establecida sobre el espacio paramétrico θ, que produzca el menor riesgo esperado.

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

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

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:

$$\min E_{\Theta/X} \{\ell(\Theta, d(X))\} = \int_\Theta \ell(\Theta, d(X))\, g(\Theta X)\, d\Theta$$

El estimador será aquel que

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

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

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.