INFERENCIA ESTADÍSTICA

ESTIMACIÓN PUNTUAL DE PARÁMETROS

IE I

Concepto de inferencia estadística

a. Sea un experimento aleatorio existente en el mundo físico. Se dirá que dicho experimento presenta permanencia estadística si, a lo largo de sucesivas realizaciones del experimento permanece inalterado el modelo probabilístico que es "reflejo fiel" de la situación.
Supóngase que dicho modelo sea "a priori" desconocido (totalmente o en parte) por un operador que necesita conocer o evaluar alguna característica del mismo.
Se llamará inferencia estadística al conjunto de técnicas que, en base a experimentación directa, conducen a dicho conocimiento o evaluación.
Por ejemplo, sea una producción de bombas de luz de 100w nominales. Sea X la variable que asume el valor real del consumo de energía de un ejemplar de dicha producción tomado al azar. Evidentemente, la F de D correspondiente a X es "a priori" desconocida, pero tomando una cantidad suficientemente grande de bombas y midiendo su consumo individual, se puede, mediante un proceso de cálculo adecuado, evaluar las características que interesen de la F. de D. antedicha.

b. Las respuestas que da la inferencia estadística no son nunca verdades definitivas e inapelables. Por ejemplo, no se obtendrá nunca una respuesta tal como: "La F. de D. de la variable aleatoria asociada al experimento cumple tal condición". Lo mejor que se puede obtener es algo como: "En un 95% de los casos en que se recojan los datos que resultaron de la experimentación, la F. de D. cumplirá con tal condición".
Es decir que las respuestas dadas por la inferencia estadística estarán afectadas por una cierta incertidumbre crónica. Por lo general, el grado de aproximación y/o fiabilidad de los resultados obtenidos es función del volumen de la experimentación realizada. A mayor experimentación, mayor precisión y fiabilidad.

c. Vaya una palabra de prevención:
En su libro "Hechos y Estadísticas", M. J. Moroney dice lo siguiente:
"En gran parte, la inferencia estadística es un método de investigación al cual se acude cuando todos los demás métodos fracasan; es a menudo un último recurso o una última esperanza. Un análisis estadístico bien hecho es una delicada disección de incertidumbres, una cirugía de suposiciones. El cirujano debe estar bien en guardia contra las falsas incisiones de su escalpelo. Muy a menudo debe declarar al paciente como inoperable."
En otras palabras, la inferencia estadística es una técnica de uso peligroso que, para dar resultados fidedignos, debe ser manejada con un máximo de cautela y " escepticismo".

d. La inferencia estadística es un árbol sumamente frondoso. Para ser un especialista en la materia es necesario un caudal de conocimientos enorme.
En lo que sigue se darán únicamente los conceptos de algunas de las técnicas más elementales.

IE II

Repetición reiterada de un mismo experimento – Valor medio observado y varianza observada

a. Sea un experimento aleatorio con permanencia estadística, al cual se asocia una variable aleatoria X con F. de D.:

$$P(\mathrm{X} \le \mathrm{x}) = \mathrm{F}^{\mathrm{X}}(\mathrm{x}) \qquad \textbf{[1]}$$

Sean m y σ el valor medio y la varianza de X.
Supóngase que este experimento se repita n veces. Supóngase que a los resultados a obtener en las repeticiones 1ª, …, nª se asocien respectivamente las variables aleatorias $X_1, \ldots, X_n$.
Se define entonces que:
1º) $X_1, \ldots, X_n$ tienen la misma F. de D. y por lo tanto:

$$m = m_{\mathrm{X}_1} = \ldots = m_{\mathrm{X}_n} \qquad \textbf{[2]}$$

$$\sigma^2 = \sigma^2_{\mathrm{X}_1} = \ldots = \sigma^2_{\mathrm{X}_n} \qquad \textbf{[3]}$$

2º) $X_1, \ldots, X_n$ son independientes.

b. Sean las variables aleatorias $\overline{\mathrm{X}}$ y S^2, definidas en base a las variables $X_1, \ldots, X_n$ según las siguientes expresiones:

$$\overline{\mathrm{X}} = \frac{1}{n}(\mathrm{X}_1 + \ldots + \mathrm{X}_n) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{X}_i \qquad \textbf{[4]}$$

$$\mathrm{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{X}_i - \overline{\mathrm{X}})^2 \qquad \textbf{[5]}$$

Se tiene que:
1º) Según visto en VAM XVI:

$$m_{\overline{\mathrm{X}}} = m \qquad \textbf{[6]}$$

$$\sigma_{\overline{\mathrm{X}}} = \frac{\sigma}{\sqrt{n}} \qquad \textbf{[7]}$$

2º) Puede demostrarse (pero no se intentará hacerlo) que:

$$m_{\mathrm{S}^2} = \left(1 - \frac{1}{n}\right)\sigma^2 \qquad \textbf{[8]}$$

c. Supóngase que, según indicado en **a**, se efectúen *n* repeticiones de un experimento aleatorio. Supóngase que los resultados de esos *n* experimentos hayan sido los números $x_1, \ldots, x_n$ respectivamente.

Esto implica que las variables $X_1, \ldots, X_n$ definidas en **a** hayan asumido los valores $x_1, \ldots, x_n$ respectivamente. Entonces, las variables $\overline{X}$ y S^2 asumirán los valores (ver **[4]** y **[5]**):

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad \textbf{[9]}$$

Al valor $\overline{x}$ se lo llamará <u>valor medio observado</u> y al valor s^2 se lo llamará <u>varianza observada</u>.

Notar que ni $\overline{x}$ es un valor medio ni s^2 es una varianza. Son valores aleatorios asumidos por $\overline{X}$ y S^2 (al revés de los valores medios y varianzas reales, que son magnitudes fijas e inmutables). Los términos "valor medio observado" y "varianza observada" no son más que convencionalismos.

IE III

<u>Estimación puntual de parámetros</u>

La estimación puntual de parámetros es uno de los renglones más "populares" de la inferencia estadística.
A continuación se tratará de dar una idea general de esta técnica.

IE III. 1

a. Sea un experimento aleatorio con permanencia estadística al cual se asocia una variable aleatoria X <u>cuya F. de D. se supone conocida</u> salvo en lo que se respecta a uno o más de sus parámetros.
<u>El objetivo de la estimación puntual de parámetros es obtener, en base a experimentación reiterada, aproximaciones lo mejor posibles de estos parámetros incógnita</u>.
Un ejemplo de esta situación puede ser la producción de tornillos fabricados por un torno que se mantiene en perfectas condiciones de ajuste. Dada esta última circunstancia, puede suponerse que las variaciones de longitud entre los tornillos son debidas a un sinnúmero de pequeñas causas entre las cuales no hay ninguna dominante.
Por lo tanto, una variable X que se asocie a la longitud de los tornillos tendrá una F. de D. normal (o muy cercana). Supóngase que no se conozca ni el valor medio ni la varianza de esta variable X, y que se desea tener estimaciones de estos valores.

b. En general, definir un proceso de estimación puntual equivale a definir para cada uno de los parámetros incógnita una función:

$$\hat{\Omega} = \varphi\,(X_1,\ldots,X_n) \qquad \textbf{[1]}$$

en la cual:

$\hat{\Omega}$: variable aleatoria que asumirá el valor estimado del parámetro.
φ : función estimadora del parámetro propia del proceso de estimación usado.
$X_1, \ldots, X_n$: variables aleatorias correspondientes a las reiteraciones del experimento.

Si las variables $X_1, \ldots, X_n$ asumen respectivamente los valores $x_1, \ldots, x_n$ (resultados de los experimentos 1º, ..., *n*º), la variable $\hat{\Omega}$ asumirá el valor:

$$\hat{\omega} = \varphi\,(x_1,\ldots,x_n) \qquad \textbf{[2]}$$

el cual constituye el valor estimado del parámetro.

c. Muy a menudo existen diversos procedimientos para estimar un mismo parámetro. En rigor, cualquier función de $X_1, \ldots, X_n$, puede ser considerada como función estimadora de un parámetro. Será una función "buena" si tiende a dar aproximaciones precisas del parámetro, y será una función "mala" en caso contrario. Así por ejemplo, a primera vista parece que:

1º) $\hat{M} = \dfrac{1}{n}(X_1 + \ldots + X_n)$
es una "buena" función estimadora de *m*.

2º) $\hat{M}' = X_1^2 + \ldots + X_n^2$
es una "mala" función estimadora de *m*.

3º) $\hat{M}'' = \dfrac{X_{max} + X_{min}}{2}$ (Promedio de los valores extremos obtenidos)
es una función estimadora cuyos méritos y defectos no se vislumbran bien.

Queda así abierta la cuestión de qué es lo que se entiende por una "buena" o "mala" función estimadora.
Para zanjar esta cuestión, se estudiarán algunas de las características que son deseables en una función estimadora.

IE III. 2

a. Sea una función estimadora $\hat{\Omega} = \varphi\,(X_1,\ldots,X_n)$ de un parámetro cuyo valor verdadero es ω. Se define:

$$\text{Sesgo de } \varphi = m_{\hat{\Omega}} - \omega \qquad \textbf{[3]}$$

Se dice que una función estimadora no tiene sesgo, o que es insesgada cuando su sesgo es nulo, es decir cuando el valor medio de la variable $\hat{\Omega}$ coincide con el verdadero valor del parámetro.
Por ejemplo, supóngase que las tres f. de d. indicadas en la figura IE III a corresponden a tres funciones estimadoras de un mismo parámetro cuyo valor verdadero es ω.

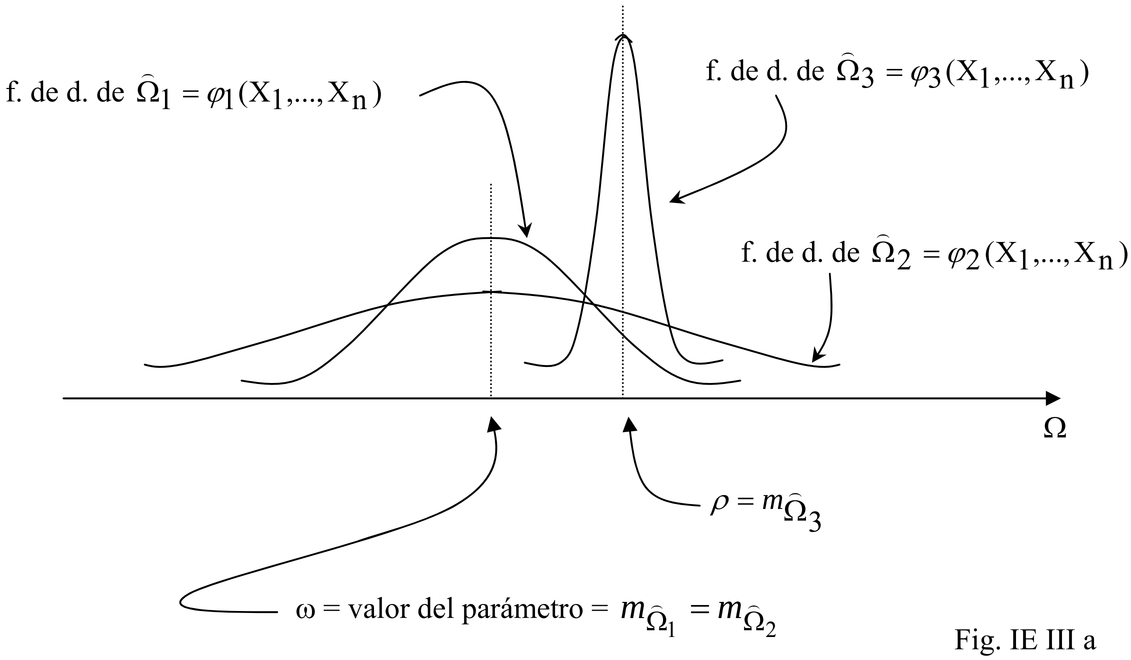

Fig. IE III a

Puede observarse que φ_1 y φ_2 no tienen sesgo y que φ_3 tiene un sesgo igual a ρ - ω.
Como ejemplo adicional, puede verificarse que si se toma a las fórmulas **[4]** y **[5]** de IE II como funciones estimadoras de m y σ^2 (es decir considerando que $\overline{X} = \hat{M}$ y $S^2 = \hat{\sigma}^2$), resulta que $\overline{X}$ no tiene sesgo y que S^2 tiene un sesgo igual a $\frac{\sigma^2}{n}$, ya que (ver **[6]** y **[8]** de IE II):

$$m_{\overline{X}} = m \qquad m_{S^2} = \left(1 - \frac{1}{n}\right)\sigma^2$$

Las desventajas de una función estimadora sesgada son obvias. Puede ser expresada de una manera intuitiva diciendo que efectuando una gran cantidad de estimaciones se obtendrá un valor promedio que tiende al de $m_{\hat{\Omega}}$ en vez de tender al verdadero valor ω del parámetro.

b. El no tener sesgo es una característica deseable en las funciones estimadoras, pero no constituye el total de la historia.
Por ejemplo, sean las funciones estimadoras φ_2 y φ_3 indicadas en la figura IE III a. Las variables aleatorias correspondientes son $\hat{\Omega}_2$ y $\hat{\Omega}_3$. Puede verse en dicha figura que φ_2 es no sesgada, mientras que φ_3 tiene un sesgo igual a ρ - ω. Si este sesgo es pequeño, es más conveniente el uso de φ_3 que el de φ_2 debido a que la varianza de $\hat{\Omega}_3$ es mucho menor que la de $\hat{\Omega}_2$, lo que implica que, con sesgo y todo, φ_3 da una mejor probabilidad que φ_2 de conseguir una buena estimación del parámetro.
Un criterio, más o menos empírico, de comparar funciones estimadoras es calcular la magnitud:

$$W_{\hat{\Omega}} = \sigma^2_{\hat{\Omega}} + \text{sesgo}^2$$

siendo una función estimadora tanto mejor cuanto menor sea el $W_{\hat{\Omega}}$ correspondiente.

Evidentemente:

$$W_{\hat{\Omega}} = \sigma^2_{\hat{\Omega}} \quad \text{para funciones estimadoras no sesgadas.}$$

El inconveniente que presenta este criterio es que muy a menudo es muy difícil calcular la magnitud $\sigma^2_{\hat{\Omega}}$.

IE IV

Estimación de parámetros por el método de máxima verosimilitud

IE IV. 1

Hasta ahora se ha visto el concepto de lo que son las funciones estimadoras y algunas propiedades muy generales de las mismas. Ahora se tratará el problema de su diseño, es decir de cómo construir "buenas" funciones estimadoras.
Existe un método general para encarar este problema, llamado "Método de máxima verosimilitud", el cual da resultados satisfactorios en una gran mayoría de los casos, a pesar de que a menudo conduce a funciones estimadoras sesgadas.
La idea básica de este método consiste en tomar como estimación de un parámetro desconocido el valor que mejor justifique los resultados obtenidos en la experimentación, es decir el valor que de una probabilidad máxima "a priori" de que ocurra lo que realmente ocurrió.

IE IV. 2

a. Es bastante largo y complicado introducirse en todos los vericuetos del método de máxima verosimilitud.
Por lo tanto, se indicarán únicamente algunas de las funciones estimadoras suministradas por este método.

Distribución	Parámetro a estimar	Función estimadora	
Poisson	λ	$\hat{\Lambda} = \overline{X}$	**[1]**
Binomial	p	$\hat{P}$ = frecuencia relativa	**[2]**
Normal	m	$\hat{M} = \overline{X}$	**[3]**
	σ^2	$\hat{\sigma}^2 = S^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2$	**[4]**
Exponencial: $f^X(x) = c \cdot e^{-cx}$ para $x \geq 0$; $= 0$ para $x < 0$	c	$\hat{C} = \frac{1}{\overline{X}}$	**[5]**

Notar que: Ver **[8]** de IE II

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

y por lo tanto $\widehat{S}^2$ como función estimadora de σ^2 tiene un sesgo (ver **[3]** de IE III) igual a:

$$\left(1 - \frac{1}{n}\right)\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n} \qquad \textbf{[6]}$$

Evidentemente, este sesgo tiende a 0 cuando $n \to \infty$.

b. Observación:
Es muy común el artificio de multiplicar una función estimadora sesgada por una constante para quitarle el sesgo. Así, si a la **[4]** se la multiplica por $n/_{n-1}$ se obtendrá una estimación de σ^2 no sesgada:

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad \textbf{[7]}$$

IE IV. 3

Se enunciarán a continuación algunas propiedades de las funciones estimadoras máximo verosímiles:

1º) Tal como se dijo en **a**, son a menudo sesgadas (ver por ejemplo la función estimadora indicada en **[4]**, aunque a veces pueden no serlo (ver por ejemplo la función estimadora indicada en **[3]**).

2º) A menudo son las funciones estimadoras mejores posibles según el criterio indicado en **b** de IE III 2.

3º) Las variables aleatorias $\hat{\Omega}$ correspondientes a las funciones estimadoras máximo verosímiles son aproximadamente normales para n grande. Entonces, si hay manera de calcular $\sigma_{\hat{\Omega}}$, se pueden usar las técnicas propias de la distribución normal para calcular la probabilidad de que la estimación hallada aproxime al valor verdadero del parámetro con una precisión convencional. Estas propiedades 2º y 3º son las responsables en gran parte de la popularidad de las estimaciones máximo verosímiles.

APÉNDICES

A. IE I

Funciones estimadoras sesgadas o insesgadas

a. Se definió en IE III.2 que dada una función estimadora $\hat{\Omega} = \varphi\,(x_1, \ldots, x_n)$ de un parámetro cuyo verdadero valor es ω se tiene que:

$$\text{Sesgo de } \hat{\Omega} = m_{\hat{\Omega}} - \omega \qquad \textbf{[1]}$$

b. Ejemplo de función estimadora no sesgada:

Función estimadora del valor medio *m* de una variable aleatoria X:

$$\hat{\mu} = \frac{1}{n}(x_1 + ... + x_n) = \overline{X} \qquad \text{por ser } m_{X_1} = ... = m_{X_n} = m$$

$$m_{\hat{\mu}} = \frac{1}{n}\left(m_{X_1} + ... + m_{X_n}\right) = \frac{1}{n} n\ m = m$$

y entonces:

$$\text{Sesgo de } \hat{\mu} = m_{\hat{\mu}} - m = m - m = 0$$

c. Ejemplo de función estimadora sesgada:

Función estimadora de la varianza σ^2 de una variable aleatoria X:

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

$$= \frac{1}{n}\left[\sum_{i=1}^{n}\left(X_i - m\right)^2 - 2\left(\overline{X} - m\right)\sum_{i=1}^{n}\left(X_i - m\right) + n\left(\overline{X} - m\right)^2\right] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - m\right)^2 - 2\left(\overline{X} - m\right)\underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - m\right)}_{=\overline{X}-m} + \left(\overline{X} - m\right)^2 =$$

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

y por lo tanto:

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

y por lo visto en A.VAU III:

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

por ser $\sigma^2_{X_1} = \ldots = \sigma^2_{X_n} = \sigma^2$ y $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$

y entonces:

$$\text{Sesgo de } \hat{S}^2 = \sigma^2\left(1-\frac{1}{n}\right) - \sigma^2 = -\frac{\sigma^2}{n}$$

Evidentemente, este sesgo tiende a cero cuando *n* tiende a infinito.

A. IE II

Funciones estimadoras de varianza mínima y funciones estimadoras insesgadas de varianza mínima

Una función estimadora $\hat{\Omega}$ de un parámetro ω es llamada de varianza mínima si para toda otra función estimadora $\hat{\Omega}^*$ se tiene que:

$$\sigma^2_{\hat{\Omega}} \langle \sigma^2_{\hat{\Omega}^*}$$

Si además $\hat{\Omega}$ es insesgada se tiene que será una función estimadora insesgada de varianza mínima.

Problemas sobre estimación puntual de parámetros

IE 1 Se desea estimar el valor medio de una distribución normal. Se proponen las siguientes funciones estimadoras:

$$\hat{M}_1 = \frac{4X_1 + X_2}{5} \qquad \text{y} \qquad \hat{M}_2 = \frac{3X_1 + X_2}{4}$$

Indicar cual de ellas es preferible.

IE 2 Sea una serie de *m* experiencias efectuadas sobre una población binomial y sea X la variable aleatoria correspondiente a las veces que apareció un cierto suceso. Sea otra serie de *n* experiencias efectuadas sobre la misma población y sea Y la variable aleatoria correspondiente a las veces que apareció el mismo suceso antedicho. Se proponen las siguientes funciones estimadoras para el parámetro *p*:

$$\hat{P}_1 = \frac{X/m + Y/n}{2} \qquad \text{y} \qquad \hat{P}_2 = \frac{X + Y}{m + n}$$

Indicar cual de ellos es preferible.

IE 3 Sean las siguientes funciones estimadoras del parámetro λ de una distribución de Poisson:

$$\hat{\Lambda}_1 = \frac{X_1 + X_2 + X_3}{3} \qquad \text{y} \qquad \hat{\Lambda}_2 = \frac{X_1 + 2X_2 + 3X_3}{6}$$

Indicar cual de ellos es preferible.

IE 4 Un regimiento de montaña transportado íntegramente a lomo de mula tiene 1000 soldados. Según datos, en el último año la cantidad mensual de accidentes por patadas de mula ha sido de:

1, 0, 2, 0, 1, 1, 0, 2, 1, 0, 1, 3

Indicar la estimación máximo verosímil del parámetro λ perteneciente a la f. de p. correspondiente a la cantidad de accidentes.

IE 5 En una bolsa hay bolillas azules, blancas y coloradas en cantidades desconocidas. Se hicieron 6 extracciones con reemplazamiento obteniéndose sucesivamente los resultados:

Blanca, azul, azul, blanca, colorada, azul

Hallar estimaciones máxima verosímiles de las probabilidades de obtener respectivamente una bolilla azul, una blanca y una colorada en una extracción aislada.

IE 6 Sea una variable X a la cual se le asocia una f. de d.

$$f^{X}(x) = \begin{cases} 0 & \text{para } x < 0 \text{ y } x \geq \alpha \\ \dfrac{1}{\alpha} & \text{para } 0 \leq x < \alpha \end{cases}$$

siendo α desconocida.
Se efectuaron 5 realizaciones del experimento al cual se asoció la variable X. Los valores asumidos por esta variable fueron:

1,8 3,2 1,75 5,21 y 0,9

Se pide hallar una estimación máximo verosímil del parámetro α.
(Ayudita: Tener bien en cuenta la idea básica del método de máximo verosimilitud, tal como indicada en IE IV. 1)