Facultad de Ciencias

Apuntes inferencia estadística

Autor: Alberto Campuzano Solórzano

4º Doble Grado Física y Matemáticas

2024-2025

Estas son unas notas que recogen lo visto en el curso 2024-2025 en la asignatura de inferencia estadística. Los apuntes completos están a disposición del alumno en la reprografía de la Facultad de Ciencias de la Universidad de Cantabria o en el curso virtual (si el profesor responsable los facilita en dicho momento). Por ello no pretende ser un objeto de autoría propia y todo lo expuesto aquí está directamente extraído de dichos apuntes. Espero aun así que os sirvan de ayuda.

Índice general

Capítulo 1: Muestreo Aleatorio Simple y Estadísticos **3**
- 1.1. Muestras aleatorias simples . . . 3
- 1.2. Distribución muestral . . . 3
- 1.3. Estadísticos y momentos muestrales . . . 4
- 1.4. Momentos muestrales . . . 4
- 1.5. Funciones características . . . 5
- 1.6. Distribuciones . . . 6

Capítulo 2: Estimación puntual **8**
- 2.1. Propiedades de los estimadores . . . 8
  - 2.1.1. Consistencia . . . 8
  - 2.1.2. Error cuadrático medio. Estimadores insesgados . . . 9
- 2.2. Método de máxima verosimilitud . . . 10
- 2.3. El método de Bayes . . . 12

Capítulo 3: Estimación por intervalos **14**
- 3.1. Construcción de intervalos de confianza. Cantidad pivotal . . . 14

Capítulo 1: Muestreo Aleatorio Simple y Estadísticos

Empezaremos introduciendo las nociones de *muestra aleatoria simple* (m.a.s) y *estadísticos*. A partir de ahí introduciremos los *momentos muestrales* y una serie de resultados asociados que nos serán de utilidad en los capítulos posteriores.

1.1. Muestras aleatorias simples

Definición 1.1. Llamamos *muestra aleatoria simple* (m.a.s) a cualquier conjunto de variables aleatorias independientes e igualmente distribuidas (v.a.i.i.d) con distribución P.

Supondremos de aquí en adelante que todas las v.a's son reales (aunque se pueden extender la mayor parte de resultados a espacios de dimensión finita) y que estarán definidas en el mismo espacio probabilístico (Ω, σ, ν). Una cuestión muy importante de notación es que, dada una m.a.s $X_1, \ldots, X_n$, una elección de un punto $\omega_0 \in \Omega$ nos conduce a los valores de nuestra muestra $x_1 = X_1(\omega_0), \ldots, x_n = X_n(\omega_0)$, aunque a menudo omitiremos ω_0. Esto es similar a lo que hacíamos en cálculo de probabilidades al denotar funciones de distribución o probabilidades como $P[X = x]$.

1.2. Distribución muestral

Supongamos un m.a.s $X_1, \ldots, X_n$ con distribución P. Podemos dar ahora la siguiente noción:

Definición 1.2. Tomando la σ-álgebra de Borel β en $\mathbb{R}$, definimos

$$P_n(A) := \frac{1}{n} \sum_{1 \leq i \leq n} I_A(x_i), A \in \beta.$$

donde $P_n = P_{(x_1, \ldots, x_n)}$. A esta probabilidad la llamamos *distribución muestral* o *distribución empírica* y llamaremos *función de distribución muestral* a su función de distribución F_n. El siguiente teorema nos muestra las propiedades asintóticas de esta distribución.

Teorema 1.3. *Supongamos que tenemos una m.a.s infinita $\{X_n\}$, entonces se verifica que:*

I Si $A \in \beta$, entonces $P_n(A) \xrightarrow{a.s} P(A)$.

II Si $A \in \beta$, entonces $n^{1/2}(P_n(A) - P(A)) \xrightarrow{\mathscr{L}} N(0, P(A)(1 - P(A)))$.

III $\sup_{x\in\mathbb{R}}\{|P_n(-\text{ínf}, x] - P(-\text{ínf}, x])|\} \xrightarrow{a.s} 0.$

donde $P(A)$ es la distribución teórica

Como resultado de este teorema tenemos que, con una m.a.s. lo suficientemente grande, es posible obtener información sobre la distribución teórica.

1.3. Estadísticos y momentos muestrales

Definición 1.4. Sea (χ, α) un espacio medible, llamamos *estadístico* a toda aplicación medible $T : \mathbb{R}^n \to \chi$ [1]

Con esta noción de estadístico introducimos el *estadístico ordenado*, una aplicación tal que $T(x_1, \dots, x_n) = (x_{(1)}, \dots, x_{(n)})$ donde se satisface que $x_{(1)} \leq \dots x_{(n)}$. En este sentido observamos que el estadístico ordenado es una permutación que ordena la muestra en orden creciente y es única salvo que existan valores iguales de x_i.

1.4. Momentos muestrales

Estos son los estadísticos más empleados. Daremos la definición general y luego detallaremos los utilizados en el curso.

Definición 1.5. Dada la muestra $X_1, \dots, X_n$ y $k \in \mathbb{N}$, se denomina *momento muestral de orden k* al valor

$$m_k := \frac{1}{n}\sum_{i=1}^{n}(X_i)^k.$$

Y se llama *momento muestral centrado de orden k* a

$$M_k := \frac{1}{n}\sum_{i=1}^{n}(X_i - m_1)^k.$$

En particular haremos uso de la *media muestral* m_1 y la *varianza muestral* M_2, que se suelen representar por:

$$\overline{X_n} = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} X_i.$$
$$S^2 = \tfrac{1}{n}\textstyle\sum_{i=1}^{n}(X_i - \overline{X_n})^2.$$

Comenzaremos ahora a dar una serie de propiedades y resultados relacionados con estos dos momentos.

Teorema 1.6. *Dada una m.a.s $X_1, \dots X_n$, si las variables de partida tienen esperanza finita, se verifica que*

$$E[\overline{X_n}] = E[X_1] \quad y \quad \overline{X_n} \xrightarrow{a.s} E[X_1]. \tag{1.1}$$

Y si también es finita $\sigma^2 = Var(X_1)^2$, entonces

$$Var(\overline{X_n}) = \frac{Var(X_1)}{n} \quad y \quad n^{1/2}(\overline{X_n} - E[X_1]) \xrightarrow{\mathscr{L}} N[0, \sigma^2]. \tag{1.2}$$

---

[1]Aplicación medible hace referencia a una función que preserva la medida.

[2]Recordemos que $Var(X)$ podemos calcularla como $Var(X) = E[(X - E[X])^2]$ o $E[X^2] - (E[X])^2$

Teorema 1.7. *Dada una m.a.s $X_1, \dots X_n$ tomada de una distribución con varianza finita σ^2, se verifica que*

$$E[S^2] = \frac{n-1}{n}\sigma. \tag{1.3}$$

Además también se tiene que

$$S^2 \xrightarrow{a.s} \sigma^2. \tag{1.4}$$

Observando la Ecuación 1.3 vemos que si queremos estimar la varianza con S^2 el valor que obtenemos está sesgado inferiormente en función del tamaño de la muestra. Para ello definimos la *cuasi-varianza*:

$$\hat{S^2} = \frac{n}{n-1}S^2. \tag{1.5}$$

de modo que se verifica de manera directa que la media coincide con σ^2

1.5. Funciones características

En este pequeño apartado definiremos las funciones características, que tienen ciertas simulitudes con las funciones de distribución y presentan una serie de resultados que nos serán de utililidad.

Definición 1.8. Sea P una distribución de probabilidad definida en $\mathbb{R}^d, d \geq 1$, con función de densidad f. Se llama función característica de P a:

$$\phi_P(t) := \int e^{itx} P(dx) = \int e^{i\sum_{i=1}^{d} t_j x_j} f(x_1, \dots, x_n) dx_1 \dots dx_n, t \in \mathbb{R}^d$$

Como cuestión de notación, de manera similiar a las probabilidades, funciones etc. denotaremos ϕ_X a la función característica de una v.a X con distribución P. La propiedades siguientes de las funciones de característica son las que nos serán de gran utilidad.

Proposición 1.9. *Se verigican las siguientes propiedades*

I Dadas las probabilidades P y Q, se tiene que $\phi_P = \phi_Q \iff P = Q$

II Si $h \in \mathbb{R}$, entonces $\phi_{hX}(t) = \phi_X(ht), t \in \mathbb{R}^d$

III Las v.a $X_1, \dots, X_n$ son independientes $\iff \forall (t_1, \dots, t_n) \in \mathbb{R}^n$ se cumpe que

$$\phi_{X_1,\dots,X_n}(t_1, \dots, t_n) = \prod_{i=1}^{n} \phi_{X_i}(t_i)$$

IV Si las v.a $X_1, \dots, X_n$ son independientes y llamamos $S_n = \sum_{i=1}^{n} X_i$, entonces

$$\phi_{S_n}(t) = \prod_{i=1}^{n} \phi_{X_i}(t), t \in \mathbb{R}$$

V Si X es una v.a real con distribución $N(\mu, \sigma^2)$, entonces

$$\phi_X(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}, t \in R \tag{1.1}$$

VI La sucesión de probabilidades $\{P_n\} \xrightarrow{\mathscr{L}} P \iff \lim_n \phi_{P_n}(t) = \phi_P(t), \forall t \in \mathbb{R}^d$

1.6. Distribuciones

A continuación, pasaremos a definir una serie de variables y funciones que nos serán de utilidad, relacionadas con la normal y m.a.s. que siguen distribuciones conocidas. Comenzamos con el siguiente resultado.

Proposición 1.10. *Si $\overline{X_n}$ es la media muestral de una m.a.s $X_1, \dots, X_n$ obtenida de una distribución de $N(\mu, \sigma^2)$, entonces la distribución de $\overline{X_n}$ es $N(\mu, \frac{\sigma^2}{n})$.*

Ahora, definiremos una distribución

Definición 1.11. Sea $n \in \mathbb{N}$, llamamos *t con n grados de libertad* a aquella que admite como función de densidad

$$f_t(x) = \frac{\Gamma(\frac{n+1}{2})}{(n\pi)^{1/2}\Gamma(\frac{n}{2})}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, x \in \mathbb{R} \tag{1.1}$$

donde Γ es la función Gamma de Euler.

Teorema 1.12. *Sea $X_1, \dots, X_n$ una m.a.s obtenida de una distribución $N(\mu, \sigma^2)$. Entonces la distribución de la variable*

$$t := (n-1)^{1/2}\frac{\overline{X_n} - \mu}{S}$$

es una t de Student con n-1 grados de libertad

Este teorema será de gran utilidad en la construcción de intervalos dxactos para la estimación de una normal, puesto que nos permitirá utilizar esta variable como *cantidad pivotal* [3] Finalmente, la última distribución que vamos a definir es la χ^2:

Definición 1.13. Se llama distribución ji cuadrado de k grados de libertad (χ^2_k) a la suma de k v.a.i.i.d's con distribución $N(0, 1)$. Admite como función de densidad:

La función característica de la χ^2_k es $\phi = (1 - 2it)^{-\frac{k}{2}}$. A partir de dicha expresión podemos probar

Proposición 1.14. *Si $X_1, \dots, X_n$ son unas v.a.'s independientes, donde la distribución de X_i es$\chi^2_{k_i}$, entonces $X_1 + \dots + X_n$ sigue una distribución χ^2 con $k_1 + \dots + k_n$ grados de libertad. Además, si X e Y son v.a's independientes y la dsitribución de X es χ^2_k y la de $X + Y$ es χ^2_m, entonces la de Y es χ^2_{m-k}.*

Finalmente, vamos a dar 3 resultados orientados a la independencia de $\overline{X_n}$ y S^2, las dos variables que caracterizan la distribución normal y que nos servirá más adelante.

Lema 1.15. *Sea (U, V) un vector aleatorio, supongamos que su función característica se puede escribir como*

$$\phi_{(U,V)}(s, t) = \phi_U(s)h(t)$$

donde ϕ_U es la función característica de U. Entonces h es la función característica de V.

Teorema 1.16. *En una m.a.s de una distribución normal $X_1, \dots, X_n$ se cumple que el estadística $\overline{X_n}$ y el vector aleatorio $(X_1 - \overline{X_n}, \dots, X_n - \overline{X_n})$ son independientes*

[3]La *cantidad pivotal* es una v.a p$(X_1, \dots, X_n, \theta)$ cuya distribución no depende de nuestro estimador θ

Como corolario de este teorema podemos deducir el siguiente resultado

Corolario 1.17. *En las condiciones del Teorema 1.16 se tiene que* $\overline{X_n}$ *y* S^2 *son independientes. ADemás, la distribución de* $\frac{nS^2}{\sigma^2}$ *es* χ^2_{k-1}

En último lugar definimos una distribución que de la que hemos hecho uso una sola vez pero no está mal conocer

Definición 1.18. Supongamos que U y V son dos v.a's independientes con distribución χ^2_n y χ^2_m respectivamente. Se llama *distribución F con (n,m) grados de libertad* a la de la variable

$$F_{(n,m)} = \frac{U/n}{V/m}$$

Podemos obtener la expresión de su función de densidad a partir de la de las ji cuadrado, a saber

$$f_F(x) = \frac{(m/n)^{m/2}}{B(m/2, n/2)} x^{\frac{m}{2}-1}\Big(1 + \frac{m}{n}x\Big)^{-\frac{m+n}{2}}, x > 0$$

donde $B(s,t)$ es la función beta.

Capítulo 2: Estimación puntual

En esta sección nos centraremos en la estimación de un parámetro, ya sea de una proporción, la media, varianza... Para ello supondremos que tenemos una familia de espacios probabilísticos $(\Omega, \sigma, \nu_\theta)$, $\theta \in \Theta$. Estos son los espacios candidatos a ser el real. Además, contaremos con una sucesión de v.a.i.i.d's $\{X_n\}$ de las que conoceremos los valores de una muestra finita de tamaño n. Así, con esta información daremos una serie de resultados y procedimientos que nos permitan identificar el valor de θ. Nótese que en nuestros espacios, la distribución de probabilidad depende del parámetro. El siguiente ejemplo clarifica un poco esta idea.

Ejemplo 2.1. Supongamos que queremos conocer la probabilidad de que al tirar una moneda, esta sea cara. En este caso, el conjunto de posibles valores de nuestro parámetro θ es $\Theta = [0, 1]$ y a su vez podemos tomar Ω como el formado por las posibles combinaciones de cara o cruz, indentificando cara con 1 y cruz con 0, de modo que:

$$\nu_\theta[(\omega_1, \ldots, \omega_n) \times \{0,1\} \times \{0,1\} \ldots] = \theta^{\sum_i \omega_i}(1-\theta)^{n-\sum_i \omega_i}$$

Volveremos sobre este mismo ejemplo a la hora de estudiar el *Estimador de Máxima Verosimilitud* ya que es muy ilustrativo.

Denotaremos por $P^\theta_{X_1}$ a la dsitribución de la v.a X_1 cuando θ es el verdadero valor del parámetro y por

$$E_\theta[X_1] = \int x P^\theta_{X_1}(dx) = \int_\Omega X_1(\omega)\nu_\theta(dw)$$

a la *esperanza* de la v.a cuando θ es el verdadero valor del parámetro. Para estimar, una vez elegido el tamaño de la muestra, nuestro problema queda reducido, desde un punto de vista teórico, a la elección de una aplicación medible $T : \mathbb{R}^n \to \Theta$. Llamaremos *estimador* a cualquiera de estas aplicaciones (previamente hemos tomado el σ- álgebra de Borel en el espacio de llegada $\Theta \subset \mathbb{R}^n$), y denotaremos por T_n o $\hat{\theta_n}$ a esta aplicación. El subíndice n aparece para dejar clara la dependencia con el tamaño muestral, aunque frecuentemente prescindiremos de él.

2.1. Propiedades de los estimadores

A la hora de elegir nuestro estimador será necesario distinguir entre aquellos que parecen conducir a resultados razonables o algunos que no sean lógicos. Incluso si existiese, nos gustaría encontrar uno que fuera *el mejor*.

2.1.1. Consistencia

Ya hemos comentado que el tamaño de la muestra será muy relevante para la cantidad de información de la que disponemos de nuestro parámetro. Por ello parece razonable pensar que a medida que nuestra muestra crece el estimador se acerque al verdadero valor del parámetro. Introducimos por ello las siguientes nociones: Dado una sucesión de estimadores $\{T_n\}$, se dice que

Consistente en probabilidad si	$T_n(X_1,\ldots,X_n) \xrightarrow{c.p} \theta$
Consistente casi seguro si	$T_n(X_1,\ldots,X_n) \xrightarrow{a.s.} \theta$
Consistente en norma cuadrática si	$E_\theta[(T_n-\theta)^2] \to 0$

Es importante remarcar que estas propiedades nos sirven para cribar los malos estimadores pero no para escoger uno bueno.

2.1.2. Error cuadrático medio. Estimadores insesgados

Para entrar en el problema de elegir un *mejor* estimador debemos tener en cuenta que éste depende de la muestra y en ocasiones se encontrará más *lejos* y otras más *cerca* de θ. Por ello trabajaremos en *promedio* eligiendo una función creciente $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ llamda *función de pérdida*. De este modo mediremos la *calidad* de un estimador como

$$E_\theta[\phi|(T(X_1,\ldots,X_n)-\theta)|] \tag{2.1}$$

Eligiendo $\phi(t)=t^2$ nos limitamos al caso de la *pérdida cuadrática*, que nos conduce a que la Ecuación 2.1 se corresponda con el *error cuadrático*. Es claro que, de existir el estimador óptimo será aque que minimice Ecuación 2.1 uniformemente en θ. El problema radica en que puede no existir. Esto se debe a que ciertos estimadores prestan demasiada atención a ciertos valores del parámetro mientras que otros no. Esta condición de *imparcialidad* es lo que llamamos *insesgadez*.

Definición 2.2. Dado un estimador T se llama sesgo de T en $\theta \in \Theta$ a la cantidad

$$b_T(\theta) = E_\theta[T] - \theta$$

Y diremos que T es insesgado si $b_T(\theta)=0, \forall\theta\in\Theta$.

Del Teorema 1.6 tenemos que $T_n = \overline{X_n}$ es siempre insesagado y por la construcción de *cuasi-varianza* dada por la Ecuación 1.5, ésta lo es respecto de la varianza. Además, dado un estimador T_n se verifica que:

$$E_\theta[T_n-\theta]^2 = Var(T_n) + b^2_{T_n}(\theta) \tag{2.2}$$

Por lo que un estimador con sesgo muy grande no puede ser cero. Además, necesitaremos que el sesgo de un estimador T_n tienda a cero para ser consistente. Sin embargo que sea insesgado no garantiza que sea un buen estimador. Ahora bien, de existir un estimador óptimo entre los insesgados será aque que tenga menor varianza, lo que conocemos como *U.M.V.U.E, (uniformly, minimum variance, umbiased estimator)*. En este sentido nos será de utilidad la *desigualdad de Cramer-Rao* aunque previamente introduciremos una hipótesis muy importante.

Definición 2.3. Llamamos *cantidad de información de Fisher* a $I(\theta) := E_\theta\left\{\left[\frac{d}{d\theta}\log f_\theta(X_1)\right]^2\right\}$

Teorema 2.4. *Sea $X_1, \dots, X_n$ una m.a.s tomada de una distribución perteneciente a una familia que satisface que $P^\theta_{X_1}$ admite función de densidad f_θ y existe $I(\theta)$. Sea T un estimador insesgado de θ. Bajo las condiciones de regularidad adecuadas se cumple que*

$$Var_\theta(T) \geq \frac{1}{nI(\theta)} \tag{2.3}$$

Definimos además la eficiencia como

$$e_\theta(T) = \frac{1/[nI(\theta)]}{Var_\theta(T)}$$

de modo que un estimador que alcanza la cota de Cramer-Rao se llama eficiente.

Esta definición no es del todo buena dado que hay casos donde no se alcanza la cota de Cramer-Rao pero sí existe U.M.V.U.E. y estaríamos diciendo que dicho estimador, el mejor, no es eficiente.

2.2. Método de máxima verosimilitud

Recuperando la idea del Ejemplo 2.1 veamos a qué hacemos referencia con el título de este apartado.

Ejemplo 2.5. Estábamos con el problema de estimar el parámetro θ de una Bernouilli para el caso de tirar una moneda. Supongamos que solo puede tomar dos valores, 0,01 y 0,15. Para realizar una estimación, tomamos una muestra de tamaño 20 y supongamos que no hay ninguna cara, es decir, tenemos que $x = (0, \stackrel{18}{\dots}, 0)$. Calculando ahora la probabilidad de obtener nuestra muestra para ambos casos tenemos:

$$\nu_{0{,}15}(0, \dots, 0) = (1 - 0{,}15)^2 0 \simeq 0{,}04$$

mientras que

$$\nu_{0{,}01}(0, \dots, 0) = (1 - 0{,}01)^2 0 \simeq 0{,}0818$$

Por lo tanto parece más *creíble* que el parámetro sea 0,01. Aquí es donde introducimos el concepto de *verosimilitud* y estimamos que $\hat{\theta} = 0{,}01$

En el Ejemplo 2.5 limitamos nuestro parámetro a dos valores, pero supongamos ahora que estamos en el caso general en e lque $\theta \in [0, 1]$. Podemos entonces, si n_1 han sido el número de caras, definir la siguiente aplicación:

$$\theta :\longrightarrow L(\theta) = \binom{n}{n_1} \theta^{n_1} (1 - \theta)^{n - n_1}$$

y tomando logaritmos y derivando llegar a la siguiente igualdad

$$\frac{n_1(1 - \theta) - (n - n_1)p}{p(1 - p)} = 0 \Longleftrightarrow \theta = n_1/n$$

de modo que ese valor de θ sera un estimador razonable.

Desarrollamos esta idea para un caso genral

Definición 2.6. Llamamos *función de verosimilitud* a la aplicación

$$\begin{aligned} L : \Theta &\longrightarrow \mathbb{R} \\ \theta &\longrightarrow \textstyle\prod_{i=1}^{n} f_\theta(x_i) \end{aligned}$$

donde $f_\theta(x_i)$ será $\nu_\theta[X_i = x_i]$ en el caso discreto.

Así decimos que el *estimador máximo verosímil*, E.M.V. de θ será cualquier valor del parámetro para el que se alcance el máximo de la función de verosimilitud. Siempre que dicha función sea derivable buscaremos su máximo mediante la *ecuación de verosimilitud*

$$\frac{d}{d\theta} \log L(\theta) = 0$$

En ocasiones probar que una solución de la ecuación de verosimilitud es un máximo es complicado y se suele llamar E.M.V a cualquiera de las soluciones independientemente de que sean o no un máximo.

El E.M.V. satisface una serie de propiedades bajo las hipótesis de regularidad adecuadas que exponemos con el siguiente teorema.

Teorema 2.7. *Supongamos que tenemos un m.a.s de una distribución que satisface la hipótesis del Teorema 2.4, entonces se tiene que*

I El E.M.V es consistente

II Si existe un estimador insesgado T que alcanza la cota de Cramer-Rao y existe un E.M.V, entonces T es solución de la ecuación de verosimilitud

III Si $\hat{\theta_n}$ es el E.M.V., entonces

$$n^{1/2}(\hat{\theta_n} - \theta) \xrightarrow{\mathscr{L}} N(0, \frac{1}{I(\theta)}) \tag{2.1}$$

La más importante de las propiedades sea la Ecuación 2.1, ya que sumando θ y diviendo por $n^{1/2}$ tenemos que

$$\hat{\theta_n} \xrightarrow{\mathscr{L}} N(\theta, \frac{1}{nI(\theta)})$$

es decir, que la distribución de $\hat{\theta_n}$ es una normal de media θ y varianza la cota de Cramer-Rao. Además la misma Ecuación 2.1 implica que, asintóticamente, el E.M.V es insesgado y eficiente.

Finalmente, el E.M.V. tiene una propiedad muy importante, que es la invarianza respecto de transformaciones, en el siguiente sentido. Si pasamos de estimar θ a $\phi = \Phi(\theta)$, el problema de redefinir la función de verosimilitud si Φ es inyectiva puede reducirse a que $L(\phi) = L(\Phi^{-1}(\phi))$. Para aclarar la relevancia de esta propiedad, si nosotros queremos estimar la media al cuadrado μ^2 y obtenemos un E.M.V $\hat{\mu}$ para μ, entonces podemos tomar, en virtud de esta propiedad, $\hat{\mu_2} = \hat{\mu}^2$. Esto nos será de utilidad cuando tengamos sumas, divisiones o potencias en el parámetro a calcular, pues en ocasiones podremos *despejar* para estimar uno conocido o más sencillo y despúes por dichas aplicaciones obtener el deseado inicialmente.

Como apunte final, el E.M.V no es necesariamente un buen estimador, si bien es cierto que para unas condiciones de regularidad adecuadas y un tamaño muestral alto funciona bien.

2.3. El método de Bayes

La idea de estimación introducida a continuación es radicalmente diferente al resto. Ahora consideraremos que el parámetro a estimar es una v.a y que la muestra m.a.s tomada es información acerca de dicho parámetro. Aplicaremos entonces los resultados para probabilidades y distribuciones condicionadas de manera que obtengamos un estimador razonable.

El esquema es el siguiente. Supondremos que tenemos dos v.a.'s X y Θ, definidas en un espacio probabilístico (Ω, σ, ν). Supondremos ahora que Θ es una v.a real y X un vector aleatorio $X = (X_1, \ldots, X_n)$, que, cuando se conoce el valor de θ, es una m.a.s de dsitribución $P^{\theta}_{X_1}$, con $\theta \in \Theta$ (aquí el conjunto imagen y no la v.a). Así, supondremos que existe la función de densidad conjunta $f_{X,\Theta}(x, \theta)$. Aplicaremos entonces las relaciones para probabilidades condicionadas

$$g(\theta) = \prod_i \int_{\mathbb{R}} f(x_i, \theta) dx_i$$

$$f(x_1, \ldots, x_n) = \int_{\mathbb{R}} \prod_i f(x_i, \theta) d\theta$$

$$f(x/\theta) = \frac{f(x, \theta)}{g(\theta)}$$

$$g(\theta/x) = \frac{f(x, \theta)}{f(x)}$$

Y sustituyendo llegamos a la *forma continua del Teorema de Bayes*

$$g(\theta/x) = \frac{\prod_i f(x_i, \theta)}{\int_{\mathbb{R}} \prod_i f(x_i, \theta) d\theta} \tag{2.1}$$

Que tiene su versión discreta (y podemos combinarlas cuando sea necesario por la forma del problema a afrontar)

$$g(\theta/x) = \frac{\nu[X_i = x_i \mid \Theta = \theta] g(\theta)}{\int_{\mathbb{R}} \prod_i \nu[X_i = x_i \mid \Theta = \theta] g(\theta) d\theta} \tag{2.2}$$

Ilustremos con par de ejemplos este método. En primer lugar uno para el caso discreto

Ejemplo 2.8. Sea $X_1, \ldots, X_5$ un m.as con distribución Bernouilli de parametro θ, sabemos que $X = (1, \ldots, 1)$. Apliquemos el método de Bayes para obtener un estimador

Sabemos que la distribución de θ es una uniforme $[0, 1]$ y por tanto

$$g(\theta) = \begin{cases} 1 & \text{si } 0 \leq \theta \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

Ahora nuestro problema reside en conocer $g(\theta \mid (1, \ldots, 1))$, bien pues aplicando la *forma discreta del teorema de Bayes* tenemos que

$$g(\theta \mid (1, \ldots, 1)) = \frac{\nu[X_i = x_i \mid \Theta = \theta] g(\theta)}{\int_{\mathbb{R}} \prod_i \nu[X_i = x_i \mid \Theta = \theta] g(\theta) d\theta} = \frac{\theta^5}{\int_0^1 \theta^5 d\theta} = 6\theta^5.$$

Así, si estuvieramos estimando la probabilidad de que el siguiente hijo fuera Varón por ejemplo, sabiendo que las 5 anteriores fueron mujeres (identificando 0 con varón 1 con mujer), entonces , al ser aleatorio podríamos considerar el estimador como la esperanza dada por la distribución calculada, de manera que:

$$\int_0^1 \theta g(\theta \mid (1, \ldots, 1))d\theta = \int_0^1 \theta 6\theta^5 d\theta = \frac{6}{7}$$

Veamos ahora uno para el continuo, algo más sofisticado que el anterior.

Ejemplo 2.9. Supongamos que tenemos una variable aleatoria $X = x$ que sigue una distribución uniforme $U[0, 1 - \theta]$ donde θ tiene función de densidad:

$$f(\theta) = \begin{cases} 2(1-t) & \text{si } 0 \leq \theta \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

Supongamos que $X = x$, queremos calcular el estimador de Bayes de θ. Siguiendo el razonamiento que hemos empleado anteriormente con la Ecuación 2.1, necesitamos calcular $g(\theta \mid X = x)$. Conocemos a su vez la función de densidad condicionada al parámetro θ:

$$f(X \mid \theta)^4 = \begin{cases} \frac{1}{1-\theta} & \text{si } 0 \leq x \leq 1 - \theta \\ 0 & \text{en otro caso} \end{cases}$$

Bien, nuestro problema ahora queda reducido a calcular $g(\theta)f(x \mid \theta)$ y después aplicar la forma continua de Bayes. Así pues tenemos que,

$$g(\theta)f(x \mid \theta) = \begin{cases} 2 & \text{si } 0 \leq x \leq 1 - \theta \wedge 0 \leq \theta \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

Observamos que $0 \leq x \leq 1 - \theta \wedge 0 \leq \theta \leq 1 \iff 0 \leq \theta \leq 1 - x \wedge 0 \leq x \leq 1$ y por tanto integrando tenemos que, sea $x \in [0, 1]$ [5]

$$\int_{\mathbb{R}} g(\theta)f(x \mid \theta)d\theta = \int_0^{1-x} 2d\theta = 2(1-x)$$

Así hemos obtenido la función de densidad condicionada que queríamos:

$$g(\theta \mid x) \begin{cases} \frac{1}{1-x} & \text{si } 0 \leq \theta \leq 1 - x \\ 0 & \text{en otro caso} \end{cases}$$

Nuestro interés estaba en calcular un estimador por el método de Bayes. En función del problema podríamos aplicar distintas estrategias pero en general podemos tomar como estimador de Bayes la esperanza de dicha probabilidad, si bien, como he dicho, podríamos tomar la mediana, la moda u otro estimador en función de la cuestión planteada. Así pues tenemos que:

$$\hat{\theta} = \int_0^{1-x} \theta g(\theta \mid x)d\theta = \frac{1-x}{2}$$

En resumen, el estimador de Bayes se comporta *bien* en condiciones generales, e sdecir que no hay otro estimador mejor que él. Además sus propiedades asinstóticas son razonablemente satifactorias y suele conducir a situaciones un tanto desastrosas cuando nuestro espacio métrico tiene dimensión infinita. Esta situación es más común de lo que podría parecer, como ejemplo imaginemos que no tenemos información sobre la distribución engendrada por nuestra muestra, entonces nuestro espacio Θ deberá ser el conjunto de todas las posibles, lo que es lógicamente infinito dimensionalmente.

[4] Notemos que todo el rato estamos haciendo este abuso de notación de $f(x \mid \theta) = f(X = x \mid \theta = \Theta)$.

[5] Esto es muy impotante ya que $g(\theta)f(x \mid \theta) \equiv 0$ si $x \notin [0, 1]$ y entonces $g(\theta \mid x)$ no está definida.

Capítulo 3: Estimación por intervalos

En la sección anterior nos hemos centrado en elegir la distribución (o parámetro más concretametne) que creemos corresponderse con la verdadera. Esto evidentemente es algo así como elegir justo la combinación que va a ganar el gordo de la lotería. Por ello introducimos el concepto de intervalo (o conjuntos) de confianza. Su finalidad será darnos un conjunto de valores entre los que, con alta probabilidad (o la deseada) se encuentra el verdadero valor del parámetro. Bien pues construiremos a continuación la teoría que sea necesaria para identificar dichos conjuntos de manera que sean razonablemente pequeños y a su vez buenos para que nuestro parámetro θ se encuentre en él

3.1. Construcción de intervalos de confianza. Cantidad pivotal

Bien, ahora que la idea parece más o menos clara introduciremos los conceptos necesarios. Diremos que, dado un conjunto $[a, b]$, es un *intervalo de confianza a nivel* α *si* $\nu_\theta\{\omega : a(\omega) \leq \theta \leq b(\omega)\} \geq \alpha, \forall \theta \in \Theta$.

Parce razonable que, conocido qué es el *nivel de confianza* α que queremos soportar, demos la definición de intervalo o conjunto de confianza.

Definición 3.1. Dada la $X_1, \ldots, X_n$, si $\mathcal{C}$ es la familia de los subconjuntos medibles de Θ, llamaremos conjunto de confianza a nivel α a cualquier aplicación medible

$$I : \mathbb{R}^n \longrightarrow \mathcal{C}$$

tal que $\nu_\theta\{\omega : \theta \in I(X_1, \ldots, X_n)\} \geq \alpha$

Observemos que esta definición, para el caso particular de $n = 1$ se corresponde precisamente con los intervalos de la forma $[a, b]$.

Ilustremos con un ejemplo el *modus operandi* para estimar este tipo de intervalos

Ejemplo 3.2. Sea $X_1, \ldots, X_n$ una m.a.s. de distribución normal $N(\mu, \sigma^2)$, quermos estimar $\mu = \theta$. Bien, en primer lugar necesitamos una cantidad pivotal que nos facilite el cálculo de los extremos del intervalo. En virtud del Teorema 1.12 sabmoes que la variable $t := \sqrt{n-1}\frac{\overline{X}_n - \mu}{S}$ tiene distribución *t de Student de n-1 grados de libertad*, que no depende de mi parámetro μ. Así, sea α el nivel de confianza que queremos, como la distribución es simétrica sabemos que

$$\alpha = \nu_\mu\{\omega \in \Omega : \frac{\alpha - 1}{2} \leq t(\omega) \leq \frac{1 - \alpha}{2}\}$$

Donde sustituyendo y despejando llegamos a que

$$\alpha = \nu_\mu\{\omega \in \Omega : \overline{X_n}(\omega) - \frac{\lambda S(\omega)}{\sqrt{n-1}} \leq \mu \leq \overline{X_n}(\omega) + \frac{\lambda S(\omega)}{\sqrt{n-1}}\}$$

donde hemos llamado $\lambda = \pm\frac{1-\alpha}{2}$

Supongamos ahora que hemos medido las alturas de una población para una muestra de $n = 26$ personas, obteniendo que $\overline{X_n} = 170$ y $S = 15$. Ahora, para $\alpha = 0{,}90$, tendríamos que, sistituyendo

$$0{,}90 = \nu_\mu\{\omega \in \Omega : 170 - \frac{1{,}706 \cdot 15}{5} \leq \mu \leq 170 + \frac{1{,}706 \cdot 15}{25}\}$$

de modo que nuestro *intervalo de confianza a nivel 0.90* es, aproximadamente $[165, 175]$.

Cabe recalcar en este punto que la cantidad θ (μ en el Ejemplo 3.2) es fija, está determinada ya y no estamos estudiando su pertenencia al intervalo dado, es decir, no hay una probabilidad α de que $\theta \in [a, b]$, de ahí que denominemos *intervalo al nivel* α. Podríamos elevar o disminuir α en función del interés del problema dado, pero θ no cambiaría, podría (aunque fuera desafortunado) haber quedado fuera del intervalo que hemos dado y nosotros no lo sabríamos y tampoco estaríamos cambiando esta situación por mucho que refinasemos los cálculos.