Clase 3

miércoles, 22 de agosto de 2018 13:14

Principio de Verosimilitud Formal

Seguiremos a Casella y Berger que trabajan con variables aleatorias discretas. Para quien este interesado en un tratamiento más profundo del tema (que incluye a las variables aleatorias continuas) lo referimos a Berger y Wolpert (1984).

Definición 25 Un experimento E es una terna $(X, \theta, \{f(x/\theta)\})$ donde X es un vector aleatorio con función de probabilidad $f(x/\theta)$ para algún θ en el espacio de parámetros Θ

Un ~~investigador~~ investigador que conoce que experimento E ha realizado y que observa una muestra $X = x$ entonces puede realizar inferencia sobre θ. A esta conclusión la notaremos como $Ev(E, x)$ y la llamaremos evidencia sobre θ obtenida a partir de E y x.

Ejemplo: Un experimento E consiste en observar $X_1, \ldots, X_n$ iid con distribución $N(\mu, \sigma^2)$ con σ^2 conocido.
$\bar{X}$ es un estadístico suficiente para μ y $E(\bar{X}) = \mu$.
Podemos considerar $EV(E, x) = (\bar{x}, \sigma/\sqrt{n})$. observemos que $\bar{x}$ depende de la muestra observada x y $\sigma/\sqrt{n}$ depende del conocimiento de E.

Definición 26: Principio de Suficiencia Formal

Consideremos un experimento $E=(X, \Theta, \{f(x/\theta)\})$ y sea $T(X)$ un estadístico suficiente para θ. Si x e y son muestras que satisfacen $T(x)=T(y)$ entonces $Ev(E,x)=Ev(E,y)$

Definición 27: Principio de Condicionalidad

Sean $E_1=(X_1, \Theta, \{f_1(x_1|\theta)\})$ y $E_2=(X_2, \Theta, \{f_2(x_2|\theta)\})$ dos experimentos (Θ común a los dos experimentos).

Consideremos el experimento mixto en el cual se observa una variable aleatoria J con $P(J=1)=P(J=2)=\frac{1}{2}$ y J es independiente de X_1, X_2 y Θ. Luego se realiza el experimento E_J.

Formalmente se realiza el experimento E^* con $E^*=(X^*, \Theta, \{f^*(x^*|\theta)\})$ con $X^*=(J, X_J)$ y

$$f^*(x^*|\theta)=f^*((j,x_j)|\theta)=\frac{1}{2}f_j(x_j|\theta)$$

Entonces $Ev(E^*,(j,x_j))=Ev(E_j,x_j)$

es decir si uno de dos experimentos es elegido al azar y luego se realiza entonces la evidencia sobre θ solo depende del experimento que se hizo.

Definición 28: Principio de Verosimilitud formal

Consideremos $E_1=(X_1, \Theta, \{f_1(x_1|\theta)\})$ y $E_2=(X_2, \Theta, \{f_2(x_2|\theta)\})$

Supongamos x_1^* y x_2^* son muestras de E_1 y E_2 respectivamente y que se cumple

$$L(\theta|x_2^*)=c\,L(\theta|x_1^*)$$

para todo θ y para alguna constante C la cual puede depender de x_1^* y x_2^* pero no de θ.

Entonces $\quad Ev(E_1, x_1^*) = Ev(E_2, x_2^*)$

Obs Si $E_2 = E_1$ entonces el principio de verosimilitud se puede deducir del principio de verosimilitud formal

Si $E = (X, \theta, \{f(x|\theta)\})$ es un experimento entonces $Ev(E, x)$ solo depende de E y x a través de $L(\theta/x)$

El resultado anterior es conocido como el corolario del principio de verosimilitud. Se omite su demostración

Un teorema muy importante debido a Birnbaum es el siguiente

Teorema 29: (Teorema de Birnbaum)

El principio de verosimilitud formal se deduce a partir del principio de suficiencia formal y del principio de condicionalidad. El recíproco también se cumple

Dem: Ver capítulo 6 de Casella y Berger.

Principio de Invarianza

Hay dos tipos de consideraciones sobre la invarianza, llamados invarianza de medida e invarianza formal.
La invarianza de medida refiere a que la inferencia realizada no debe depender de la escala de medida.
La invarianza formal establece que si dos problemas de inferencia tienen la misma estructura formal en términos del modelo matemático usado, entonces el mismo procedimiento de inferencia puede usarse en ambos problemas. Los elementos del problema que deben ser los mismos son Θ, $\{f(x|\theta) : \theta \in \Theta\}$ y el conjunto de todas las inferencias que se pueden hacer y las consecuencias de las inferencias equivocadas.

Definición 30: (Principio de Invarianza)
Si $Y = g(X)$ es un cambio de la escala de medida tal que el modelo para Y tiene la misma estructura formal que el modelo X, entonces el procedimiento de inferencia será invariante por medida y formalmente invariante.

Definición 31: Un conjunto de funciones $\{g(x) : g \in G\}$ donde $g : \mathcal{X} \longrightarrow \mathcal{X}$ se dice que es un grupo de transformaciones de $\mathcal{X}$ si:

i) Inverso. Para todo $g \in G$ existe $g' \in G$ tal que $g'(g(x)) = x$ para todo $x \in \mathcal{X}$.

ii) Composición. Para todo $g \in G$ y $g' \in G$ existe $g'' \in G$ tal que $g'(g(x)) = g''(x)$ para todo $x \in X$

iii) Identidad $e(x) = x$ es un elemento de G

(Notar que iii) es consecuencia de i) y ii))

Ejemplo $X \sim Bin(n, p)$. Sea $T(x)$ el estimador de p usado cuando $X = x$. Sea $Y = n - X$ (cantidad de fallos) $Y \sim Bin(n, q)$. Sea $T^*(y)$ es estimador de q cuando $Y = y$ luego $1 - T^*(y)$ es el estimador de p cuando $Y = y$, luego $1 - T^*(n-x)$ es un estimador de p.

La invarianza de medida requiere que

$$T(x) = 1 - T^*(n-x)$$

ya que el cambio de X a Y es un cambio de escala de medida. La estructura formal para los problemas de inferencia basados en X e Y es la misma, X e Y son ambas binomiales n, θ con $0 \leq \theta \leq 1$. Así pues la invarianza formal requiere que $T(z) = T^*(z)$ $\forall z = 0, 1, \dots, n$.

Así pues la invarianza de medida y la invarianza formal vistas conjuntamente implican

$$T(x) = 1 - T^*(n-x) = 1 - T(n-x)$$

Es claro que esta ecuación muestra como se reduce mucho el conjunto de estimadores se deben tener.

Si tenemos un estimador cualquiera se debe especificar $T(0), T(1), \ldots, T(n)$ mientras que si se cumple la ecuación anterior sólo debe especificarse $T(0), T(1), \ldots, T([n/2])$ y el resto se determina mediante la igualdad, por ejemplo $T(n) = 1 - T(0)$ etc.

Dos estimadores que son invariantes para este problema son $T_1(x) = x/n$ y $T_2(x) = 0{,}9\,(x/n) + 0{,}1\,(0{,}5)$. Se verifica fácilmente que

$$T_1(x) = 1 - T_2(n-x) = 1 - T_1(n-x)$$

Para este problema sólo dos transformaciones están involucradas $G = \{g_1(x), g_2(x)\}$ tal que $g_1(x) = n - x$ y $g_2(x) = x$.
Verifique el lector que se cumple la definición 31

Definición 32: Sea $\mathcal{F} = \{f(x|\theta) : \theta \in \Theta\}$ un conjunto de funciones de densidad o de probabilidad de X y sea G un grupo de transformaciones del espacio de muestras $\mathcal{X}$. Entonces $\mathcal{F}$ es invariante bajo el grupo G si para todo $\theta \in \Theta$ y $g \in G$ existe un único $\theta' \in \Theta$ tal que $Y = g(X)$ tiene la distribución $f(y|\theta')$ si X tiene la distribución $f(x|\theta)$

En el ejemplo anterior, si $X \sim Bin(n,p)$ entonces:

$$g_1(X) = n - X \sim Bin(n, 1-p) \text{ luego } p' = 1-p$$
$$g_2(X) = X \sim Bin(n,p) \text{ así } p' = p$$

O sea el conjunto de las funciones de probabilidad binomiales es invariante bajo el grupo $G = \{g_1, g_2\}$

Estimación Puntual

Definición 33: Un estimador puntual es toda función de una muestra $W(X_1, \ldots, X_n)$. Es decir todo estadístico es un estimador puntual.

Método de los momentos.

Sea $X_1, \ldots, X_n$ una muestra de una población con función de densidad o distribución $f(x/\theta_1, \ldots, \theta_k)$. El método de los momentos consiste en igualar los k momentos muestrales con los k momentos poblacionales. O sea para $j = 1, \ldots, k$ sean

$$m_1 = \frac{1}{n} \sum_{i=1}^{n} X_i \quad ; \quad \mu_1 = E(X)$$

$$m_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 \quad ; \quad \mu_2 = E(X^2)$$

$$\cdots\cdots\cdots$$

$$m_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k \quad ; \quad \mu_k = E(X^k)$$

Los momentos poblacionales generalmente son función de $\theta_1, \theta_2, \ldots, \theta_k$. El método de los momentos encuentra estimadores $\hat{\theta}_1, \ldots, \hat{\theta}_k$ resolviendo el sistema de ecuaciones

$$m_1 = \mu_1(\theta_1, \ldots, \theta_k)$$

$$m_2 = \varphi_2(\theta_1, \dots, \theta_k)$$

$$- - - - - - - - - -$$

$$m_k = \varphi_k(\theta_1, \dots, \theta_k)$$

Ejemplo: Sean $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ iid.

Tenemos que

$$\varphi_1 = E(X) = \mu$$

$$\varphi_2 = E(X^2) = \mu^2 + \sigma^2$$

además

$$m_1 = \overline{X}$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Se tiene el sistema

$$\overline{X} = \mu$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mu^2 + \sigma^2$$

luego $\hat{\mu} = \overline{X}$ y $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Ejemplo: Sea $X_1, \dots, X_n$ iid $\sim Bin(k, p)$.

Se tiene que $\varphi_1 = E(X) = kp$, como $V(X) = kp(1-p)$ entonces $\varphi_2 = E(X^2) = V(X) + E^2(X) = kp(1-p) + k^2p^2$

Resolvemos $\frac{1}{n}\sum_{i=1}^{n} X_i = kp$ (1)

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = kp(1-p) + k^2p^2 \quad (2)$$

De (1) $\hat{p} = \frac{\overline{X}}{\hat{k}}$ luego en (2)

De (1) $\hat{p} = \dfrac{\bar{X}}{\hat{k}}$ luego en (2)

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \cancel{\hat{k}}\,\frac{\bar{X}}{\cancel{\hat{k}}}\left(1-\frac{\bar{X}}{\hat{k}}\right) + \cancel{\hat{k}^2}\,\frac{\bar{X}^2}{\cancel{\hat{k}^2}} \quad \text{o sea}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \bar{X}\left(1-\frac{\bar{X}}{\hat{k}}\right) + \bar{X}^2 \quad \text{y}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \bar{X}\left(1-\frac{\bar{X}}{\hat{k}}\right)$$

$$\frac{(1/n)\sum_{i=1}^{n}(X_i-\bar{X})^2}{\bar{X}} = 1-\frac{\bar{X}}{\hat{k}} \quad \text{entonces}$$

$$\frac{1}{\hat{k}} = \frac{\bar{X} - (1/n)\sum_{i=1}^{n}(X_i-\bar{X})^2}{\bar{X}^2} \quad \text{de donde}$$

$$\hat{k} = \frac{\bar{X}^2}{\bar{X} - (1/n)\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

Estos estimadores no son los mejores, aún más podemos obtener valores negativos de $\hat{p}$ y $\hat{k}$ (discuta el lector cuando ocurre esto)

Ejemplo (Satterhwaite (1946)) "Moment Matching"

Si Y_i $i=1,\dots,k$ son independientes con $Y_i \sim \chi^2_{r_i}$

(Demuestre el lector que $\sum Y_i \sim \chi^2_{\sum r_i}$)

Si tenemos la variable $\sum a_i Y_i$ con a_i constantes conocidas. La distribución de esta suma es difícil de obtenerse, sin embargo parece razonable asumir que la distribución es χ^2_ν para algún valor de ν.

Satterthwaite estaba interesado en aproximar el denominador de un estadístico t y $\sum a_i Y_i$ representa el cuadrado del denominador de ese estadístico. De esto, dados $a_1, a_2, \dots, a_k$ se desea encontrar el valor de ν tal que aproximadamente

$$\sum_{i=1}^{k} a_i Y_i \sim \frac{\chi^2_\nu}{\nu}$$

Ya que $E(\chi^2_\nu/\nu) = 1$ entonces

$$E\left(\sum_{i=1}^{k} a_i Y_i\right) = \sum_{i=1}^{k} a_i E(Y_i) = \sum_{i=1}^{k} a_i r_i = 1$$

Tambien $E\left[\left(\chi^2_\nu/\nu\right)^2\right] = \frac{2}{\nu} + 1$ luego

$$E\left[\left(\sum_{i=1}^{k} a_i Y_i\right)^2\right] = \frac{2}{\nu} + 1$$

así eliminando la esperanza

$$\hat{\nu} = \frac{2}{\left(\sum_{i=1}^{k} a_i Y_i\right)^2 - 1}$$

$\hat{\nu}$ podría ser negativo por lo que Satterthwaite no lo usó, sino que continuó trabajando

$$E\left[\left(\sum a_i Y_i\right)^2\right] = V\left(\sum a_i Y_i\right) + \left(E\left(\sum a_i Y_i\right)\right)^2 =$$

$$= \left[\underbrace{E\left(\sum a_i Y_i\right)}\right]^2 \left[\frac{V\left(\sum a_i Y_i\right)}{\left[E\left(\sum a_i Y_i\right)\right]^2} + 1\right] =$$

$$= \Big[\underbrace{E\big(\sum a_i Y_i\big)}_{=1}\Big]^2 \left[\frac{V(\sum a_i Y_i)}{[E(\sum a_i Y_i)]^2} + 1\right] =$$

$$= \frac{V(\sum a_i Y_i)}{[E(\sum a_i Y_i)]^2} + 1$$

entonces $\frac{V(\sum a_i Y_i)}{[E(\sum a_i Y_i)]^2} + 1 = \frac{2}{\nu} + 1$

o sea $\nu = \frac{2[E(\sum a_i Y_i)]^2}{V(\sum a_i Y_i)}$

Como los Y_j son independientes entonces

$$V(\sum a_i Y_i) = \sum a_i^2 V(Y_i) = 2 \sum \frac{a_i^2 [E(Y_i)]^2}{r_i}$$

pues $V(Y_i) = 2[E(Y_i)]^2 / r_i$

Finalmente $\hat{\nu} = \frac{(\sum a_i Y_i)^2}{\sum \frac{a_i^2}{r_i} Y_i^2}$ donde se

eliminan las esperanzas