

# POR QUÉ LA INFERENCIA ESTADÍSTICA BAYESIANA

\*

M. A. Gómez Villegas

## 1 Introducción

En este artículo se pretende justificar por qué se debe utilizar la aproximación bayesiana a la inferencia estadística.

Como anunciaba Lindley en el primer Congreso Internacional de Estadística Bayesiana, falta menos para el 2021 año en el que el adjetivo bayesiana para la estadística será superfluo al ser todas las aproximaciones a la estadística bayesianas.

## 2 La precisión en los modelos clásicos

La manera de medir la precisión en los procedimientos clásicos o frecuentistas no funciona adecuadamente. Se va a empezar por ponerlo de manifiesto en el caso de la estimación por punto. Supongamos que se pretende determinar un estimador por punto para el parámetro  $e^{-3\theta}$  para una población de Poisson de parámetro  $\theta$ , es fácil obtener, véase Gómez Villegas (2005), que el citado

---

\*Subvencionado por el MEC, proyecto MTM2005-05462

estimador ha de ser  $T(i) = (-2)^i$ ; un estimador que es negativo para los valores impares de  $i$  y que por lo tanto, en esos casos, es absurdo para estimar  $e^{-3\theta}$  que es una cantidad positiva.

Los inconvenientes se repiten, si se utilizan los intervalos de confianza. Para una m.a.s. de tamaño  $n$  de una población uniforme en el intervalo  $(\theta - 1/2, \theta + 1/2)$  un intervalo de confianza al 95% basado en el estadístico suficiente  $(X_{(1)}, X_{(n)})$  viene dado por

$$IC_{.95}(\theta) = (\hat{\theta}_n - 0.056, \hat{\theta}_n + 0.056)$$

con  $\hat{\theta}_n = (X_{(1)} + X_{(n)})/2$ .

Para la muestra con  $X_{(1)} = 3$  y  $X_{(n)} = 3.96$  el intervalo de confianza es  $(3.424, 3.536)$ .

Por otra parte, si se trabaja directamente a partir de la distribución uniforme, ha de ser

$$\theta - 1/2 < X_{(1)} < X_{(n)} < \theta + 1/2$$

por lo que despejando

$$X_{(n)} - 1/2 < \theta < X_{(1)} + 1/2$$

es decir que forzosamente ha de ser  $3.4 < \theta < 3.46$ . Por lo tanto directamente se obtiene un intervalo de amplitud menor que el intervalo de confianza.

Los contrastes de hipótesis tampoco escapan a las críticas. Si se quiere determinar el test de tamaño  $\alpha$  para contrastar la hipótesis nula  $H_0 : \theta = 0$  frente a la hipótesis alternativa  $H_1 : \theta \neq 0$  para una población  $Normal(\theta, 1)$ , es fácil comprobar que el test tiene de región crítica  $RC = \{\sqrt{n}|\bar{x}| > 1.96\}$ . Para valores de  $n$  suficientemente grandes el test rechaza siempre, aunque la muestra provenga de una población  $Normal(0, 1)$ , lo que no parece nada razonable.

El contraste de hipótesis puede realizarse también en función del p-valor, o probabilidad bajo la hipótesis nula del suceso cola observado, así para el contraste anterior, si se observa la muestra  $\mathbf{x} = (x_1, \dots, x_n)$  el p-valor es  $p - \mathbf{x} = P\{|\bar{X}| > \bar{x} | \theta = 0\}$ . Es decir es la probabilidad de que en una población  $N(0, 1)$  la media muestral sea mayor que la media muestral correspondiente a la muestra observada  $(x_1, \dots, x_n)$ . El p-valor se interpreta diciendo que si éste es menor que 0.05 o 0.01 o 0.001 se rechaza la hipótesis nula, en primer lugar ¿no es esto interpretar el factor bayes como  $P\{H_0 | T(\mathbf{x})\}$  –lo que técnicamente se llama falacia del fiscal–? y si es así ¿por qué no utilizar la aproximación bayesiana que permite hablar de las probabilidades finales de las hipótesis?

En segundo lugar si se admite el p-valor, ¿por qué un p-valor de 0.049 es estadísticamente significativo y uno de 0.051 no lo es? Como Jeffreys (1961) señala, el 0.05 fue elegido por Fisher debido a la coincidencia matemática de que en la  $Normal(0, 1)$  un bajo porcentaje del area total bajo la distribución queda más allá de 1.96 veces la desviación típica; es decir  $P\{|Z| \geq 1.96\} = 0.05$

cuando  $Z \sim Normal(0, 1)$ .

Hasta aquí parece que el título del artículo debiera haber sido *por qué no se debe utilizar la estadística frecuentista*, se pasa ahora a dar una justificación muy resumida de los fundamentos de la aproximación bayesiana.

### 3 La aproximación bayesiana

La inferencia bayesiana se basa en el uso de una distribución de probabilidad para describir todas las cantidades desconocidas relevantes a un problema de estimación, la concreción técnica de este resultado consiste en lo siguiente.

Si se dispone de una colección de variables aleatorias intercambiables  $(x_1, \dots, x_n)$  es decir que su distribución sólo depende del valor de esas variables y no del orden en que han sido observadas, entonces la distribución de probabilidad

$$f(x_1, \dots, x_n) = \int_{\Theta} \prod_{i=1}^n f(x_i|\theta)\pi(\theta)d\theta$$

donde  $f(x_i|\theta)$  denota el modelo de probabilidad,  $\theta$  es el límite de alguna función de las observaciones y  $\pi(\theta)$  es una distribución de probabilidad sobre  $\Theta$  (la distribución inicial). El concepto de intercambiabilidad es más débil que el de muestra aleatoria simple.

Por ejemplo, si las variables intercambiables  $x_i$  toman el valor 0 o 1, el teorema de representación toma la forma

$$f(x_1, \dots, x_n) = \int_{\Theta} \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} \pi(\theta) d\theta,$$

donde  $\theta = \lim_{n \rightarrow \infty} (\sum_{i=1}^n x_i) / n$

Es importante notar, que lo que quiere decir el anterior resultado es que siempre que se tenga una colección de variables intercambiables, y una m.a.s. lo son, existe una distribución inicial sobre el parámetro  $\theta$ . Además, el valor del parámetro puede obtenerse como límite de las frecuencias relativas.

Los aspectos técnicos pueden consultarse en Bernardo y Smith (1994) pág. 35.

La aproximación bayesiana implica entonces, que la información muestral y la distribución inicial se actualizan mediante el teorema de Bayes para dar lugar a la distribución final.

$$\pi(\theta|x_1, \dots, x_n) = \frac{\pi(\theta)f(x_1, \dots, x_n|\theta)}{\int_{\Theta} \pi(\theta)f(x_1, \dots, x_n|\theta)d\theta}$$

Ahora todas las inferencias, la estimación por punto, la estimación por regiones de confianza y los contrastes de hipótesis, se realizan mediante la distribución final.

Un resumen actualizado del desarrollo de la aproximación bayesiana a la inferencia puede verse en Bernardo (2003).

## 4 Conclusiones

Los procedimientos basados en la distribución en el muestreo son ad hoc para prácticamente cada aplicación o grupo de aplicaciones con los que se esté trabajando. En contraposición, los procedimientos bayesianos siempre funcionan

de la misma manera; hay que determinar una distribución inicial que recoja la información que se tenga del problema, construir la distribución final y esta es la que recoge, en forma de una distribución de probabilidad, la información suministrada por la muestra.

Una crítica que suele hacerse a la aproximación bayesiana es que está influenciada por la distribución inicial, pero es hoy perfectamente factible examinar el problema con una variedad de distribuciones iniciales, o bien emplear distribuciones iniciales objetivas, y en todo caso se debe tener en cuenta que para tamaños muestrales grandes la verosimilitud domina a la distribución inicial por lo que las inferencias se ven poco afectadas por la distribución inicial. A cambio, los métodos bayesianos siempre tratan la incertidumbre mediante la probabilidad y la precisión de los mismos se mide siempre en términos de probabilidad.

## 5 Referencias

Bernardo, J.M. (2003) *Bayesian Statistics. Encyclopedia of Life Support Systems. Probability and Statistics*. (R. Viertl. ed) Oxford: UNESCO.

Bernardo, J.M. and Smith, A.F.M. (1994) *Bayesian Theory*. New York: Wiley.

Gómez Villegas, M.A. (2005) *Inferencia Estadística*. Madrid: Díaz de Santos.

Jeffreys, H. (1939) *Theory of Probability (3 ed. 1961)*. Londres: Oxford University Press.