Intervalos de confianza para la binomial

Supongamos que el vector $\mathbf{x}$ en R^n tiene distribución $p(\mathbf{x}, \boldsymbol{\theta})$ donde $\boldsymbol{\theta}$ $=(\theta_1, ..., \theta_k)$. Sea $r : R^n \to R$ y $T = r(\mathbf{x})$.

Consideremos $A(\boldsymbol{\theta})$ y $B(\boldsymbol{\theta})$ tales que

$$P_{\boldsymbol{\theta}}(A(\boldsymbol{\theta}) \leq T \leq B(\boldsymbol{\theta})) = 1 - \alpha. \tag{0.1}$$

Luego $S(\mathbf{x})$ definido por

$$S(\mathbf{x} = \{\boldsymbol{\theta} : \ A(\boldsymbol{\theta}) \leq T \leq B(\boldsymbol{\theta})\} \tag{0.2}$$

es una región de confianza de nivel $1 - \alpha$. En efecto

$$P_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in S(\mathbf{x})) = P_{\boldsymbol{\theta}}(A(\boldsymbol{\theta}) \leq T \leq B(\boldsymbol{\theta})) = 1 - \alpha.$$

Si T es discreta pueden no existir $A(\boldsymbol{\theta})$ y $B(\boldsymbol{\theta})$ satisfaciendo exactamente (0.1). En este caso se podrán tomar de manera que

$$P_{\boldsymbol{\theta}}(A(\boldsymbol{\theta}) \leq T \leq B(\boldsymbol{\theta})) \geq 1 - \alpha. \tag{0.3}$$

Entonces, si se define $S(\mathbf{x})$ por (0.2), se tendrá

$$P_{\boldsymbol{\theta}}(\boldsymbol{\theta} \in S(\mathbf{x})) = P_{\boldsymbol{\theta}}(A(\boldsymbol{\theta}) \leq T \leq B(\boldsymbol{\theta})) \geq 1 - \alpha. \tag{0.4}$$

Por supuesto que $A(\boldsymbol{\theta})$ y $B(\boldsymbol{\theta})$ se deberán tomar de manera que $P_{\boldsymbol{\theta}}(A(\boldsymbol{\theta}) \leq T \leq B(\boldsymbol{\theta}))$ se aproxime lo más posible a $1 - \alpha$.

El siguiente Lema será necesario para construir un intervalo para el parámetro de una distribución binomial.

Lemma: Si T tiene distribución $\mathrm{Bi}(\theta, n)$, entonces

(i)Si $0 \leq k < n$ entonces $P_\theta(T > k)$ es una función continua estrictamente creciente de θ. Además $\lim_{\theta \to 1} P_\theta(T > k) = 1$ y $\lim_{\theta \to 0} P_\theta(T > k) = 0$

(ii) Si $0 < k \leq n$ entonces $P_\theta(T < k)$ es una función continua estrictamente decreciente de θ. Además $\lim_{\theta\to 1} P_\theta(T < k) = 0$ y $\lim_{\theta\to 0} P_\theta(T < k) = 1$

Demostracion. Para ver que $P_\theta(T > k)$ es continua, obsérvese que

$$P_\theta(T > k) = \sum_{i=k+1}^{n} P_\theta(T = i)$$

como $P_\theta(T = i)$ es continua como función de θ, lo mismo ocurre con $P_\theta(T > k)$.

Para demostrar que $\lim_{\theta\to 1} P_\theta(T < k) = 0$, basta demostrar que $\lim_{\theta\to 1} P_\theta(T = n) = 1$. Esto resulta del hecho que $P_\theta(T = n) = \theta^n$. Para demostrar que $\lim_{\theta\to 0} P_\theta(T < k) = 1$, basta ver que $\lim_{\theta\to 0} P_\theta(T = 0) = \lim_{\theta\to 0}(1-\theta)^n = 1$.

Sea $\theta_1 < \theta_{2..}$ Consideremos una muestra aleatoria $x_1, ..., x_n$ de una distribución $\mathrm{Bi}(\theta, n)$, y sean $u_1, ..., u_n$ variables con distribución $U[0,1]$ independientes de las x_i. Sea

$$b = \frac{\theta_2 - \theta_1}{1 - \theta_1}$$

y definamos $y_1, ..., y_n$ por

$$y_i = \begin{cases} 1 & \text{si} \quad x_i = 1 \text{ o } u_i \leq b \\ 0 & \text{si} \quad x_i = 0 \text{ y } u_i > b. \end{cases}$$

Luego

$$\begin{aligned} P(y_i &= 1) = P(x_i = 1) + P(u_i \leq b) - P(x_i = 1)P(u_i \leq b) \\ &= \theta_1 + b - \theta_1 b = \theta_1 + b(1-\theta_1) = \theta_2. \end{aligned}$$

Sea $z_i = I(u_i \leq b), 1 \leq i \leq n$ y definamos

$$T_x = \sum_{i=1}^{n} x_i,\ T_y = \sum_{i=1}^{n} y_i, Z = \sum_{i=1}^{n} z_i.$$

Luego T_x tiene distribución $\mathrm{Bi}(\theta_1, n)$ y T_y tiene distribución $\mathrm{Bi}(\theta_2, n)$. Además si $0 \leq k < n$

$$\{T_y > k\} \ \supset \{T_x > k\} \cup \{\{T_x = 0\} \cap \{Z = n\}\}.$$

Como lo dos eventos de la unión son disjuntos y los dos eventos de la intersección son independientes se tiene

$$P(T_y > k) \ \geq P(T_x > k) + P(T_x = 0)P(Z = n).$$

Como $P(T_x = 0) > 0$ y $P(Z = n) > 0$, resulta $P(T_y > k) \ > P(T_x > k)$ y (i) queda demostrado. (ii) se demuestra similarmente.

Consideremos ahora una muestra aleatoria $x_1, ..., x_n$ de una distribución $\mathrm{Bi}(\theta, 1)$. Es lógico definir $T = \sum_{i=1}^{n} x_i$, ya que este estadístico es suficiente. Sabemos que T tiene distribución $\mathrm{Bi}(\theta, n)$.

Definamos ahora

$$A(\theta) = \max\{z : \ P_\theta(T < z) \leq \alpha/2\} \tag{0.5}$$

y

$$B(\theta) = \min\{z : \ P_\theta(T > z) \leq \alpha/2\}. \tag{0.6}$$

Claramente, estas definiciones de $A(\theta)$ y $B(\theta)$ satisfacen (0.4). El siguiente Teorema nos permite encontrar la region de confianza $S(\mathbf{x})$ dado por (0.2) y que resulta un intervalo.

Teorema. Supongamos que el valor observado sea $T = t$. Entonces

(a) Sea $0 < t < n$, definamos $\underline{\theta}$ y $\overline{\theta}$ por

$$P_{\underline{\theta}}(T > t - 1) = \alpha/2, \tag{0.7}$$

y

$$P_{\overline{\theta}}(T < t + 1) = \alpha/2. \tag{0.8}$$

Obsévese que $\underline{\theta}$ y $\overline{\theta}$ esán bien definidos por el Lema. Luego en este caso la región $S(\mathbf{x})$ definida (0.2) con $A(\theta)$ y $B(\theta)$ dados por (0.5) y (0.6) respectivamente corresponde al intervalo $S(\mathbf{x}) = (\underline{\theta}, \overline{\theta})$.

(b) Si $t = 0$, definimos $\overline{\theta}$ por (0.8). Luego $S(\mathbf{x}) = [0, \overline{\theta})$.

(c) Supongamos que $t = n$. Definamos $\underline{\theta}$ por (0.7). Luego $S(\mathbf{x}) = (\underline{\theta}, 1]$.

Demostración de (a)

Sea $\theta > \underline{\theta}$, luego por el Lema

$$P_\theta(T > t - 1) > P_{\underline{\theta}}(T > t - 1) = \alpha/2$$

y luego $t \leq B(\theta)$.

Sea $\theta \leq \underline{\theta}$, luego por el Lema se tiene

$$P_\theta(T > t - 1) \leq P_{\underline{\theta}}(T > t - 1) = \alpha/2$$

y entonces $t > B(\theta)$. Luego

$$t \leq B(\theta) \Longleftrightarrow \underline{\theta} < \theta.$$

Similarmente se prueba que

$$A(\theta) \leq t \Longleftrightarrow \theta < \overline{\theta}. \tag{0.9}$$

Demostración de (b) Similarmente que en (a) se prueba (0.9).
Por otro lado para todo θ se tiene que $t = 0 \leq B(\theta)$. Luego $S(\mathbf{x}) =[0, \overline{\theta})$.
La demostración de (c) es idéntica.