Teoría Estadística - Distribución Binomial

Gonzalo Barrera Borla y Jesús Zapata

2 de Septiembre de 2018

Introducción

Sea $\mathbb{X} = (X_1, X_2, ..., X_n)$ un vector aleatorio tal que $X_i \sim Bi(k, p) \; \forall \, i \in \{1, ..., n\}$, y $X_i \perp X_j \; \forall \, i, j \in \{1, ..., n\}$. Es decir, que todas las X_i son independientes entre sí, están idénticamente distribuidas, y su función de distribución pertenece a la "familia binomial"

Si escribimos $\theta = (k, p)$ la función de probabilidad puntual queda dada por

$$f(x; \theta) = \binom{k}{x} \; p^x \; (1-p)^{k-x}$$

La función de densidad conjunta, llamémosla $\mathbf{p}(x_1, x_2, ..., x_n; \theta)$ será entonces

$$\mathbf{p}(x_1, x_2, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) = \left[\prod_{i=1}^{n} \binom{k}{x_i}\right] \times p^{\sum_{i=1}^{n} x_i} \times (1-p)^{nk - \sum_{i=1}^{n} x_i}$$

Si convenimos en la notación $\mathbf{x} = (x_1, x_2, ..., x_n)$ y $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, podemos reescribir

$$\mathbf{p}(\mathbf{x}; \theta) = \left[\prod_{i=1}^{n} \binom{k}{x_i}\right] \times p^{n\overline{x}} \times (1-p)^{n(k-\overline{x})}$$

Estimadores

Consideraremos tres tipos de situaciones: - estimación del número de pruebas k con probabilidad de éxito p conocida, - estimación de p con k conocido, y - estimación de $\theta = (k, p)$, sin ningún parámetro conocido.

De $q(\theta) = k$

Asumiendo que la probabilidad de éxito p de un ensayo individual es conocida, es razonable querer estimar el parámetro desconocido k, el número de ensayos efectuados, siempre el mismo para cada variable aleatoria X_i

Lo intentaremos según el método de los momentos, usando los momentos de primer y segundo orden, y según el método de máxima verosimilitud.

Momentos con $g(x) = x$

Sea g una función de $\mathbb{R}$ en $\mathbb{R}$, luego el método de los momentos estima θ, por el valor $\hat{\theta} = \delta(\mathbb{X})$ que satisface la ecuación

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) = E(g(X_1) \mid \theta = \hat{\theta}) = E_{\hat{\theta}}(g(X_1))$$

Si usamos $g(x) = x$, y escribimos $\theta = (\hat{k}, p)$ y sabiendo que $E(X_i) = k \times p \; \forall \; i \in \{1, ..., n\}$, podemos afirmar que el estimador de k será el $\hat{k}$ que satisfaga

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}(X_1 \mid \theta = (\hat{k}_{m_1}, p)) \Rightarrow \overline{X} = \hat{k}_{m_1} \times p \Rightarrow \boxed{\hat{k}_{m_1} = \frac{\overline{X}}{p}}$$

Momentos con $g(x) = x^2$

Si en lugar de usar la función identidad para g utilizamos $g(x) = x^2$ obtenemos un estimador relacionado al momento $\hat{k}_{m_2}$ de segundo orden. Aquí será útil recordar que para toda variable aleatoria X con media y varianza finitas, $Var(X) = E(X^2) - [E(X)]^2$:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} X_i^2 &= E(X_1^2 \mid \theta = (\hat{k}_{m_2}, p)) = E_{\hat{k}_{m_2}}(X_1^2) \\
\overline{X^2} &= Var_{\hat{k}_{m_2}}(X_1) + [E_{\hat{k}_{m_2}}(X_1)]^2 \\
\overline{X^2} &= \hat{k}_{m_2} \; p \; (1-p) + [\hat{k}_{m_2} \; p]^2 \\
0 &= \hat{k}_{m_2}^2 \; p^2 + \hat{k}_{m_2} \; p(1-p) - \overline{X^2}
\end{aligned}$$

Que es una ecuación cuadrática con coeficientes $a = p^2, \; b = p(1-p), \; c = -\overline{X^2}$. La expresión general de sus raíces nos da

$$\hat{k}_{m_2} = \frac{p(p-1) \pm \sqrt{p^2(1-p)^2 + 4p^2\overline{X^2}}}{2p^2} = \boxed{\frac{p - 1 + \sqrt{(1-p)^2 + 4\overline{X^2}}}{2p}}$$

La conversión de $\pm$ en $+$ se justifica ya que k debe ser positivo, y $p - 1 < 0$.

Máxima Verosimilitud

Recordemos que el estimador de máxima verosimilitud se define como aquél que maximiza la función de probabilidad conjunta p ("función de verosimilitud") de la muestra observada. En otras palabras, diremos que $\delta(\mathbb{X}) = \widehat{g(\theta)}_{mv} = \hat{k}_{mv}$ es el estimador de máxima verosimilitud de $g(\theta) = k$ si se cumple que

$$\mathbf{p}(\mathbf{x} \mid \theta = (\hat{k}_{mv}, p)) = \max_{\theta \; \in \Theta} \mathbf{p}(\mathbf{x} \mid \theta)$$

La forma tradicional de maximizar una expresión así, consiste en buscar $\hat{k} : \frac{\partial \mathbf{p}(\mathbf{x}, \hat{k}, p)}{\partial k} = 0$. El problema, es que $\mathbf{p}$ incluye términos combinatorios de la forma $\binom{k}{x_i}$, con lo cual su derivada parcial respecto a k no es amistosa.

No es lo mismo afirmar que "el EMV no existe", a decir que "el EMV existe pero es difícil de tratar algebraicamente", que es una afirmación más precisa. Haciendo uso del lenguaje **R**, podemos escribir una función equivalente a lo siguiente:

$$EMV(k) = \boxed{\hat{k}_{mv} = \underset{k \in \mathbb{N}_{\geq 1}}{\text{argmax}} \; \mathbf{p}(\mathbb{X} \mid \theta)}$$

De $q(\theta) = p$

Momentos con $g(x) = x$

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}(X_1 \mid \theta = (k, \hat{p}_{m_1})) \Rightarrow \overline{X} = k \times \hat{p}_{m_1} \Rightarrow \boxed{\hat{p}_{m_1} = \frac{\overline{X}}{k}}$$

Momentos con $g(x) = x^2$

Seguiremos un procedimiento muy similar al utilizado para calcular $\hat{k}_{m_2}$, buscando$\hat{p}_{m_2}$ que satisfaga:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} X_i^2 &= E(X_1^2 \mid \theta = (k, \hat{p}_{m_2})) = E_{\hat{p}_{m_2}}(X_1^2) \\
\overline{X^2} &= Var_{\hat{p}_{m_2}}(X_1) + [E_{\hat{p}_{m_2}}(X_1)]^2 \\
\overline{X^2} &= k\ \hat{p}_{m_2}\ (1 - \hat{p}_{m_2}) + [k\ \hat{p}_{m_2}]^2 \\
0 &= k(k-1)\ \hat{p}_{m_2}^2 + k\ \hat{p}_{m_2} - \overline{X^2}
\end{aligned}$$

Tomando coeficientes $a = k(k-1),\ b = k,\ c = -\overline{X^2}$, la expresión general de sus raíces nos da

$$\boxed{\hat{p}_{m_2} = \frac{-\ k\ +\ \sqrt{k^2\ \ + 4k(k-1)\overline{X^2}}}{2k(k-1)}}$$

Al igual que en la estimación de $\hat{k}_{m_2}$, reemplazar $\pm$ por $+$ se justifica ya que $p \in [0, 1]$. También vale notar que este estimador sólo sirve si $k > 1$.

Máxima Verosimilitud

A diferencia de $\hat{k}_{mv}$, $\hat{p}_{mv}$ es más ameno a la manipulación simbólica.

Como la función $\ln(\mu)$ es monótona creciente, maximizar $\mathbf{p}(\mathbf{x}, \theta)$ será equivalente a maximizar $\ln \mathbf{p}(\mathbf{x}, \theta)$. Luego el $EMV(p)$ debe verificar $\frac{\partial \ln \mathbf{p}(\mathbf{x}, k, \hat{p})}{\partial p} = 0$. Calculemos la derivada parcial de $\ln \mathbf{p}$:

$$\begin{aligned}
\frac{\partial \ln \mathbf{p}(\mathbf{x}, \theta)}{\partial p} &= \frac{\partial}{\partial p} \ln \left[\left(\prod_{i=1}^{n} \binom{k}{x_i} \right) \times p^{n\overline{x}} \times (1-p)^{n(k-\overline{x})} \right] \\
&= \frac{\partial}{\partial p} \left[\left(\sum_{i=1}^{n} \ln \binom{k}{x_i} \right) + n\ \overline{x} \ln p + n\ (k - \overline{x}) \ln (1-p) \right] \\
&= \frac{n\overline{x}}{p} - \frac{n(k-\overline{x})}{1-p} = \frac{n\overline{x}(1-p) - np(k-\overline{x})}{p(1-p)}
\end{aligned}$$

Igualando a cero y reemplazando $p = \hat{p}_{mv},\ \mathbf{x} = \mathbb{X}$

$$\begin{aligned}
0 &= \frac{\partial \ln \mathbf{p}(\mathbb{X}, (k, \hat{p}_{mv}))}{\partial p} = \frac{n\overline{X}(1 - \hat{p}_{mv}) - n\hat{p}_{mv}(k - \overline{X})}{\hat{p}_{mv}(1 - \hat{p}_{mv})} \\
0 &= \frac{n}{\hat{p}_{mv}(1 - \hat{p}_{mv})} \times \overline{X}(1 - \hat{p}_{mv}) - \hat{p}_{mv}(k - \overline{X}) \\
0 &= \overline{X} - \overline{X}\hat{p}_{mv} - \hat{p}_{mv}\ k + \overline{X}\hat{p}_{mv} \\
&\Rightarrow \boxed{\hat{p}_{mv} = \frac{1}{k}\overline{X}}
\end{aligned}$$

Que coincide con el estimador de momentos de primer orden, $\hat{p}_{m_1} = \hat{p}_{mv}$

De $q(\theta) = \theta = (k, p)$

Intentaremos ahora estimar ambos parámetros de la distribución a la vez

Método de los momentos

Indicaremos por $\hat{\theta}_{m_{12}} = (\hat{k}_{m_{12}}, \hat{p}_{m_{12}})$ el estimador de momentos para ambos parámetros, que tendrá que cumplir:

$$
\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X} = E_{\hat{\theta}_{m_{12}}}(X_1) = \hat{k}_{m_{12}} \times \hat{p}_{m_{12}} \\
&\frac{1}{n} \sum_{i=1}^{n} X_i^2 = \overline{X^2} = E_{\hat{\theta}_{m_{12}}}(X_1^2) = Var_{\hat{\theta}_{m_{12}}}(X_1) + [E_{\hat{\theta}_{m_{12}}}(X_1)]^2 = \hat{k}_{m_{12}}\ \hat{p}_{m_{12}}\ (1 - \hat{p}_{m_{12}}) + [\hat{k}_{m_{12}}\ \hat{p}_{m_{12}}]^2
\end{aligned}
$$

Cuya solución es

$$ $\boxed{\hat{k}_{m_{12}} = \frac{\overline{X}^2 + \overline{X} - \overline{X^2}}{\overline{X}} \quad ; \quad \hat{p}_{m_{12}} = \frac{\overline{X}^2}{\overline{X}^2 + \overline{X} - \overline{X^2}}}$

$$

Máxima Verosimilitud

Al igual que con $\hat{k}_{mv}$, no hay una "expresión cerrada" de $\hat{\theta}_{mv}$ que podamos reportar. Sin embargo, podemos definirlo (y computarlo en **R**) según:

$$
EMV(\theta) = \boxed{\hat{\theta}_{mv} = \underset{\theta \in \Theta}{\text{argmax}}\ \mathbf{p}(\mathbb{X} \mid \theta)} \quad , \ \Theta = \{1, 2, ...\} \times [0, 1]
$$

Comparación y análisis

Error cuadrático medio

Recordemos que dado el estimador $\widehat{q(\theta)} = \delta(\mathbb{X})$, su error cuadrático medio es

$$
ECM_\theta(\delta(\mathbb{X})) = E_\theta(\delta(\mathbb{X}) - q(\theta))^2 = Var_\theta(\delta(\mathbb{X})) + Sesgo^2 con \quad Sesgo = b(\delta(\mathbb{X})) = E(\delta(\mathbb{X})) - q(\theta)
$$

Para determinar la bondad del estimador se comparan los ECM de cada estimador y el que tiene menor ECM es el mejor estimador, es decir, δ^* será mejor estimador que δ sí y sólo si:

$$
ECM_\theta(\delta^*) \leq ECM_\theta(\delta) \quad \forall \theta \in \Theta
$$

De todos los estimadores que consideramos hasta ahora, los únicos que permiten calcular sencillamente la esperanza y varianza son $\{\hat{p}_{m_1}, \hat{p}_{mv},\ \hat{k}_{m_1}\}$. Además, no se puede comparar los estimadores de p porque el de momentos como el de maxima verosimilitud son idénticos. De cualquier manera se puede determinar el ECM de los mismos.

Estimadores de $q(\theta) = p$

Calcularemos primero el sesgo del estimador $\delta = \hat{p}_{m_1} = \hat{p}_{mv}$:

En nuestro caso:

$$b(\delta, q) = E_\theta\left(\delta(\mathbb{X}) - q(\theta)\right) = E_\theta\left[\frac{1}{k}\left(\frac{1}{n} \times \sum_{i=1}^{n} X_i\right) - p\right] = \frac{1}{kn} n\ E(X_1) - p = \frac{kp}{k} - p = 0$$

Por lo tanto el estimador es insesgado, y su ECM es igual a la varianza del estimador. En este caso,

$$Var_\theta(\delta) = Var_\theta\left[\frac{1}{k} \times \frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{(kn)^2} Var_\theta\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{(kn)^2}\sum_{i=1}^{n} Var_\theta(X_i) = \frac{1}{(kn)^2} n\ k\ p\ (1-p)$$

$$\boxed{ECM_\theta(\delta) = Var_\theta(\delta) = \frac{p(1-p)}{kn}}$$

Estimadores de $q(\theta) = k$

De manera análoga a lo expuesto para $\hat{p}_{m_1}$, podemos ver que $\hat{k}_{m_1}$ es insesgado ya que $E_\theta(\hat{k}_{m_1}) = k = q(\theta)$. Similarmente, obtendremos que

$$\boxed{ECM_\theta(\hat{k}_{m_1}) = Var_\theta(\hat{k}_{m_1}) = \frac{k(1-p)}{np}}$$

Consistencia

Sabemos que por la ley fuerte de los grandes números,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \to E(X_1) \quad c.t.p.$$

Luego,

$$\frac{1}{k}\overline{X} \to \frac{1}{k} E(X_1) = q(\theta) = p \quad c.t.p.$$

y por lo tanto $\frac{1}{k}\overline{X} = \hat{p}_{m_1} = \hat{p}_{mv}$ es un estimador fuertemente consistente de $q(\theta) = p$.

De manera análoga, pero reemplazando $\frac{1}{k}$ por $\frac{1}{p}$ podemos probar que $\hat{k}_{m_1}$ es fuertemente consistente para $q(\theta) = k$.

Estadísticos suficientes

Sea $\mathbb{X}$ un vector aleatorio de dimensión n cuya distribución es$F(\mathbf{x}, \theta)$ con $\theta \in \Theta$ Se dice que un estad?stico $T = r(X)$ es suficiente para todo θ si la distribución de X condicional a que $T = t$ es independiente de θ para todo t

Para encontrar el estad?stico suficiente se utilizar? el siguiente teorema

Teorema de Factorización

Sea X un vector aleatorio con funci?n de densidad o funci?n de probabilidad puntural $p(x, \theta), \theta \in \Theta$. Entonces, el estad?stico $T = r(X)$ es suficiente para θsi y s?lo si existen dos funciones g y h tales que:

$$p(\mathbf{x}, \theta) = g(r(\mathbf{x}), \theta)h(\mathbf{x})$$

Anexo 1: Código de simulaciones

```
library(tidyverse)
verosimilitud <- function(X) {
  function(tita) {
      sum(log(dbinom(X, size = tita[["n"]], prob = tita[["p"]])))
        }
    }

tita <- list(n = 50, p = 0.23)
muestra <- rbinom(100, size = tita[["n"]], prob = tita[["p"]])
ver1 <- verosimilitud(muestra)
espacio <- cross(list(
  n = 1:100,
    p = (1:100)/100
    ))

tibble(titon = espacio,
       vero = map_dbl(titon, ver1)) %>%
         transmute(
         n = map_dbl(titon, "n"),
             p = map_dbl(titon, "p"),
             vero) -> resumen

resumen %>% mutate(E = n*p) %>% arrange(desc(vero))
resumen %>%
  filter(near(n*p, 0.6)) %>%
    arrange(desc(vero))

resumen %>%
  ggplot(aes(n, p, z = vero)) +
    geom_contour(binwidth = 0.1)

resumen %>%
  ggplot(aes(n, p)) +
    geom_density2d()

espacio_vero
max_n <- 10000
map(espacio, ver1)
ver1(list(n = 50, p = 0.1))
data <- tibble(
  x = seq_len(max_n),
    y = map_dbl(x, ver1))
```

```r
data %>%
  ggplot(aes(x, y)) +
    geom_line()

data %>% arrange(desc(y))
```

```r
estimadores <- list(
  m1_p = function(X, n, ...) { mean(X) / n },
    m2_p = function(X, n, ...) {
        (n + sqrt(n^2 + 4*n*(n-1) * mean(X^2))
         ) / (2 * n * (n - 1))
          },
            m1_n = function(X, p, ...) { mean(X) / p },
          m2_n = function(X, p, ...) {
              (sqrt(p^2 - 2*p + 4*mean(X^2) + 1) + p - 1
                   ) / (2 * p)
                 },
                   m12_n = function(X, ...) {mean(X)^2 / (mean(X)^2 + mean(X) - mean(X^2))},
                 m12_p = function(X, ...) {(mean(X)^2 + mean(X) - mean(X^2)) / mean(X)},
                   mv_p = function(X, p, ...) { mean(X) / p }
                   )

rango_p_final <- c(0, 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 0.5, 0.7, 0.9, 0.97, 0.99, 0.997, 0.999, 1)
rango_p <- c(0, 0.01, 0.1, 0.5, 0.9, 0.99, 1)
rango_n_final <- round(exp(seq_len(9)), 0)
rango_n <- round(exp(seq_len(4)), 0)
rango_k <-rango_n
sims <- 1000

estimaciones_a_realizar <- cross_df(list(
  k = rango_k, estimador = names(estimadores))
  )

df <- cross_df(list(n = rango_n, p = rango_p, sim_id = seq_len(sims))) %>%
  mutate(
     data = map2(n, p, ~rbinom(n = max(rango_k), size = .x, prob = .y)))

asistente_estimacion <- function(X, nombre_estimador, k, ...) {
  estimadores[[nombre_estimador]](X[seq_len(k)], ...)
  }

resultados <- crossing(df, estimaciones_a_realizar) %>%
  mutate(valor = pmap_dbl(
      .l = list(X = data, nombre_estimador = estimador,k = k, n = n, p = p),
          .f = asistente_estimacion))

pryr::object_size(resultados)

resultados %>%
  filter(n == 20, p == 0.001, estimador %in% c("m1_p", "m2_p", "m12_p")) %>%
    ggplot(aes(valor, color = estimador)) +
      geom_freqpoly()
```

```r
resultados %>%
  select(-data) %>%
    write_csv("resultados.csv")

resultados %>%
  group_by(n, p, k, estimador) %>%
    summarise(media = mean(valor)) %>%
                #varianza = sd(valor)) %>%
          spread(estimador, media) %>%
            View()

  ggplot(aes(valor, color = k)) +
    geom_freqpoly() +
        facet_grid(n ~ p)
```