CC-0288 - Inferência Estatística I

Exemplo 1.3.9 - 17/03/2023

Prof. Maurício Mota

1. Vamos comentar o Exemplo 1.3.9 da página 15 do livro do Heleno e da Mônica.

   Consideremos uma urna com N bolas idênticas marcadas com os números $1, 2, \ldots, N$. O objetivo é estimação de N , o número de bolas numeradas da urna. Este problema está muitas vezes associado ao problema de estimação do número N de táxis em uma cidade , em que os táxis estão numerados de 1 até N.

   Portanto Uma determinada quantidade n de bolas(táxis) é observada, com reposição.

   Associada à i-ésima observação temos a variável aleatória:

   $X_i =$ número da i-ésima bola retirada da urna, $i = 1, 2, \ldots, n$.

   Neste caso temos:

$$P(X_i = k) = \frac{1}{N}\, I_A(k), \;\; A = \{1, 2, \ldots, N\}.$$

   Assim

   $X_1, X_2, \ldots, X_n$ é uma amostra aleatória de tamanho n de

$$X \sim U_d(A), \quad A = \{1, 2, \ldots, N\}.$$

   Vamos calcular a esperança de X:

$$\mu = E(X) = \sum_{x=1}^{N} x\, P(X = x)$$

$$\mu = E(X) = \frac{1}{N} \sum_{x=1}^{N} x = \frac{1}{N} \times \frac{N(N+1)}{2} = \frac{N+1}{2}.$$

   O valor esperado de X^2 é dado por:

$$E(X^2) = \frac{1}{N} \sum_{x=1}^{N} x^2 = \frac{1}{N} \times \frac{N(N+1)(2N+1)}{6} = \frac{(N+1)(2N+1)}{6}.$$

   A variância de X é dada por:

$$V(X) = E(X^2) - E^2(X) = \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4} =$$

$$V(X) = \frac{2(N+1)(2N+1) - 3(N+1)^2}{12} = \frac{(N+1)\,(4N+2-3N-3)}{12}$$

$$\sigma^2 = V(X) = \frac{(N+1)(N-1)}{12} = \frac{N^2-1}{12}.$$

   O livro usa como dois estimadores inicialmente:

$$\hat{N}_1 = \bar{X} \quad e \quad \hat{N}_2 = Y_n = max(X_1, X_2, \ldots, X_n).$$

Note que:

$$E[\hat{N}_1] = E[\bar{X}] = \mu = \frac{N+1}{2} \neq N$$

que é um estimador viciado para N.

A partir de

$$E(\bar{X}) = \frac{N+1}{2}$$

podemos criar um terceiro estimador que é conhecido como estimador pelo método dos momentos:

$$E(2\bar{X}) = N + 1$$

$$E(2\bar{X} - 1) = N.$$

Assim

$$\hat{N}_3 = 2\bar{X} - 1$$

é um estimador não viciado de N.

A variância de $\hat{N}_1$ é dada por:

$$Var[\hat{N}_1] = Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{N^2 - 1}{12n}.$$

A variância de $\hat{N}_3$ é dada por:

$$Var[\hat{N}_3] = Var(2\bar{X} - 1) = 4Var(\bar{X}) = \frac{N^2 - 1}{3n}.$$

Vamos estudar o estimador Y_n.

A função de probabilidade de Y_n é dada por:

$$P(Y_n = y) = P(Y_n \leq y) - P(Y_n \leq y - 1)$$

$$P(Y_n \leq y) = [P(X \leq y)]^n = \frac{y^n}{N^n}.$$

Assim

$$P(Y_n = y) = \frac{y^n}{N^n} - \frac{(y-1)^n}{N^n} = \frac{y^n - (y-1)^n}{N^n} \, I_A(y), \quad A = \{1, 2, \ldots, N\}.$$

Será que

$$\sum_{y=1}^{N} P(Y_n = y) = 1?$$

$$\sum_{y=1}^{N} P(Y_n = y) = \sum_{y=1}^{N} N^{-n} \left(y^n - (y-1)^n\right) = N^{-n} \sum_{y=1}^{N} \left(y^n - (y-1)^n\right)$$

Esta serie na literatura é conhecida como série telescópica.

$$\sum_{y=1}^{N} \left(y^n - (y-1)^n\right) = 1^n - 0^n + 2^n - 1^n + \ldots + (N-1)^n - (N-2)^n + N^n - N^{n-1} = N^n$$

$$\sum_{y=1}^{N} P(Y_n = y) = N^{-n} \times N^n = 1.$$

Vamos calcular a esperança de Y_n é dada por:

$$E(Y_n) = \sum_{y=1}^{N} y \times P(Y_n = y)$$

$$E(Y_n) = \frac{1}{N^n} \times \sum_{y=1}^{N} y \times \left(y^n - (y-1)^n\right).$$

Note que:

$$y \times \left(y^n - (y-1)^n\right) = y^{n+1} - y(y-1)^n$$

Usando

$$y = y - 1 + 1$$

temos:

$$y(y-1)^n = (y-1+1)(y-1)^n = (y-1)^{n+1} - (y-1)^n.$$

Logo

$$y \times \left(y^n - (y-1)^n\right) = \left(y^{n+1} - (y-1)^{n+1}\right) - (y-1)^n.$$

$$E(Y_n) = \frac{1}{N^n} \times \left[\sum_{y=1}^{N} \left(y^{n+1} - y^{n+1}\right) - \sum_{y=1}^{N} (y-1)^n\right]$$

Note que:

$$E(Y_n) = \frac{1}{N^n} \times \left[N^{n+1} - \sum_{y=1}^{N} (y-1)^n\right] = N - N^{-n} \times \sum_{y=1}^{N} (y-1)^n.$$

Para N grande a soma $\sum_{y=1}^{N} (y-1)^n$ pode ser aproximada pela área sob a curva $y = x^n$ de $x = 0$ até $x = N$.

Logo

$$\int_0^N x^n dx = \frac{x^{n+1}}{n+1}\Big|_0^N = \frac{N^{n+1}}{n+1}$$

Finalmente

$$E(Y_n) \approx N - \frac{N}{n+1} = \frac{n}{n+1}N.$$

Daí surge o quarto estimador proposto $\bar{N}_4$

$$\bar{N}_4 = \frac{n=1}{n}\,Y_n.$$

Assim

$$E(\bar{N}_4) \approx N.$$

O melhor estimador para N é o quinto estimador proposto .

$$T = h(Y_n) = \bar{N}_5 = \frac{Y_n^{n+1} - (Y_n - 1)^{n+1}}{Y_n^n - (Y_n - 1)^n}$$

.

Vamos mostrar que T é um estimador não viciado para N.

$$E(T) = \sum_{y=1}^{N} h(y) \times P(Y_n = y).$$

Note que:

$$h(y) \times P(Y_n = y) = \frac{y^{n+1} - (y-1)^{n+1}}{y^n - (y-1)^n} \times \frac{y^n - (y-1)^n}{N^n},$$

$$h(y) \times P(Y_n = y) = \frac{y^{n+1} - (y-1)^{n+1}}{N^n}.$$

Logo

$$E(T) = \frac{1}{N^n} \times \sum_{y=1}^{N} \left(y^{n+1} - (y-1)^{n+1}\right) = \frac{1}{N^n} \times N^{n+1} = N.$$

O livro tira uma amostra aleatória de tamanho 8 e foram observados:

$$x_1 = 926; x_2 = 212; x_3 = 712; x_4 = 628; x_5 = 124; x_6 = 315; x_7 = 782; x_8 = 684.$$

Calcule as estimativas de N usando os 5 estimadores:

```
> A=c(926,212,712,628,124,315,782,684);A
[1] 926 212 712 628 124 315 782 684
> n=length(A);n
[1] 8
> AO=sort(A);AO
[1] 124 212 315 628 684 712 782 926
> S=sum(A)
> Xb=mean(A);S/n
[1] 547.875
> Xb=mean(A);Xb;S/n
[1] 547.875
[1] 547.875
> y_8=max(A);y_8
[1] 926
> 2*Xb-1
[1] 1094.75
> ((n+1)/n)*y_8
[1] 1041.75
> num=y_8^9 -(y_8-1)^9;num
[1] 4.844568e+24
> den=y_8^8 -(y_8-1)^8;den
[1] 4.652923e+21
> t=num/den;t
[1] 1041.188
>
```