THE UNIVERSITY OF SYDNEY

STAT3023 Statiscal Inference

Lab Week 5

Tutor: Wen Dai

SID: 470408326

School of Mathematics and Statistics

The University of Sydney

Semester 2, 2021

We shall compare three different estimators of a binomial success probability. If $Y \sim B(2, \theta)$ then we have: $P(Y = 0) = (1 - \theta^2)$, $P(Y = 1) = 2\theta(1 - \theta)$, $P(Y = 2) = \theta^2$. Moreover, if we have an iid sample $Y_1, Y_2, ..., Y_n$ then if we define:

- $N_0 = \sum_{i=1}^n 1_{\{Y_i=0\}}$ as the number of 0's $\implies N_0 \sim B(n, (1-\theta)^2)$
- $N_1 = \sum_{i=1}^n 1_{\{Y_i=1\}}$ as the number of 1's $\implies N_1 \sim B(n, 2\theta(1-\theta))$
- $N_2 = \sum_{i=1}^n 1_{\{Y_i=2\}}$ as the number of 2's $\implies N_2 \sim B(n, \theta^2)$

The usual estimator of θ based on an iid sample $Y_1, Y_2, ..., Y_n$ is a function of $\overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$

1. Determine an unbiased estimator of θ which is a *linear* function of $\overline{Y}$. Call it $\hat{\theta_1}$

Solution

To find an unbiased estimator of θ we first note that:

$$\begin{aligned}\mathbb{E}[\overline{Y}] &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[Y_1] \\ &= \mathbb{E}[Y_1] \\ &= 2\theta\end{aligned}$$

Hence we should define an unbiased estimator $\hat{\theta_1}$ by:

$$\hat{\theta_1} = \frac{1}{2}\overline{Y}$$

2. Determine an unbiased estimator of θ which is a *nonlinear* function of N_0 (hint: use method of moments, i.e. set equal to expectation and solve for θ). Call it $\hat{\theta_0}$

Solution

To find an unbiased estimator of θ we recall the method of moments. We have that $1_{\{Y_i=0\}} \sim$ *bernoulli*$((1-\theta)^2)$. Hence we have that $\mathbb{E}\left(1_{\{Y_i=0\}}\right) = (1-\theta)^2$. With the random sample $1_{\{Y_1=0\}}, 1_{\{Y_2=0\}}, ..., 1_{\{Y_n=0\}}$. We have that (by the method of moments) $(1-\theta)^2 = \frac{1}{n}\sum_{i=1}^n 1_{\{Y_i=0\}} \implies (1-\theta)^2 = \frac{1}{n}N_0 \implies \hat{\theta_0} = 1 - \sqrt{\frac{N_0}{n}}$

3. Determine an unbiased estimator of θ which is a *nonlinear* function of N_2 (hint: use method of moments, i.e. set equal to expectation and solve for θ). Call it $\hat{\theta_2}$

Solution

To find an unbiased estimator of θ we recall the method of moments. We have that $1_{\{Y_i=2\}} \sim$ *bernoulli*(θ^2). Hence we have that $\mathbb{E}\left(1_{\{Y_i=2\}}\right) = \theta^2$. With the random sample $1_{\{Y_1=2\}}, 1_{\{Y_2=2\}}, ..., 1_{\{Y_n=2\}}$. We have that (by the method of moments) $\theta^2 = \frac{1}{n}\sum_{i=1}^n 1_{\{Y_i=2\}} \implies \theta^2 = \frac{1}{n}N_2 \implies \hat{\theta_2} = \sqrt{\frac{N_2}{n}}$

4. We shall simulate a sample if $n = 100$ iid such Y_is and compute the values of these three estimators and then compare their mean squared errors, for a fine grid of θ values.

Solution

We want to now compare the variance of $\hat{\theta}_0, \hat{\theta}_1$ and $\hat{\theta}_2$ with the CRLB of $\frac{\theta[1-\theta]}{2n}$.

```
# sample size
n = 100
# Number of simulation iterations
N = 1000
# theta values and length of this vector of values
thetaVals = (1:39)/40
len = length(thetaVals)
# Declare the mean squared error vectors to be plotted against the theta values
mse0 = vector(mode = "numeric", length = len)
mse1 = vector(mode = "numeric", length = len)
mse2 = vector(mode = "numeric", length = len)

# Declare temporary variables to be used in the loop
temp0 = vector(mode = "numeric", length = N)
temp1 = vector(mode = "numeric", length = N)
temp2 = vector(mode = "numeric", length = N)

for (i in 1:len) {
  # We set the current value of theta.
  currThetaValue = thetaVals[i]

  # For the current theta value, we calculate the mean squared error for all
  # three estimators
  for (j in 1:N) {

    # step 1: Draw a sample of n binomal observations
    obs = rbinom(n = n, size = 2, prob = currThetaValue)

    # calculate the value of theta0, theta1 and theta2
    temp0[j] = 1 - sqrt(sum(obs == 0)/n)
    temp1[j] = 1/2 * mean(obs)
    temp2[j] = sqrt(sum(obs == 2)/n)

  }

  temp0 = (temp0 - currThetaValue)^2
  temp1 = (temp1 - currThetaValue)^2
  temp2 = (temp2 - currThetaValue)^2

  mse0[i] = mean(temp0)
  mse1[i] = mean(temp1)
  mse2[i] = mean(temp2)
}
```

After all this processing, we can now plot the values of the mean squared error for each value of θ_k (where $k \in \{1, 2, 3, ..., 39\}$). We also plot the Cramer Rao Lower Bound and show the result below:

```
# returns the min and max value for the range
yRange = range(c(mse0, mse1, mse2))
# plot the best estimator theta1 hat
plot(x = thetaVals, y = mse1, col = "red", ylim = yRange,
     type = "l", main = "Empirical MSE's",
     xlab = "Theta values",
     ylab = "mean squared error")
# plot theta0 hat
lines(x = thetaVals, y = mse0, col = "blue")
# plot theta2 hat
lines(x = thetaVals, y = mse2, col = "DarkGreen")
# plot the CRLB
curve(0.5 * x * (1 - x)/n, add = TRUE, lty = 2)
# legend
legend(x = "top",
       legend = c("using average", "using N0", "using N2", "CRLB"),
       col = c("red", "blue", "DarkGreen", "black"),
       lty = c(1, 1, 1, 2))
```