```
> knitr::opts_chunk$set(
+   comment = '', fig.width = 6, fig.height = 4, fig.align = "center",
+   tide = TRUE, size = "small"
+ )
```

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) $\theta^2 = \frac{1}{n}\sum_{i=1}^{n} 1_{\{Y_i=2\}} \implies \theta^2 = \frac{1}{n}N_0 \implies \hat{\theta_2} = \sqrt{\frac{N_0}{n}}$

4. We shall simulate a sample if $n = 100$ iid such Y_i's 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 $\frac{\theta[1-\theta]}{2n}$. To do this, we illustrate the idea with $\hat{\theta}_0$ first.

```
> # create a partiton of theta values
> theta_vals = (1:39)/40
> # create a vector to store the theta_0 values we obtain
> var_theta_0_for_each_theta = rep(0, length(theta_vals))
> # set the sample size n and the number of simulation iterations N
> n = 100
> N = 1000
> # Define the MSE0 to be zero
> MSE0 = 0
> # define a vector of length 1000 so store our simulation values
> simulate_N = rep(0, N)
> # for each value of theta_vals, we want to estimate the variance of theta_0.
> for (i in 1:length(theta_vals)) {
+   # We draw n times from Y ~ B(2, theta_vals[1]) and calculate theta_1_hat.
+   # We simulate this experiment N times.
+
+   # temp_vect = rbinom(n * N, 2, theta_vals[i])
+   # mat_sim = matrix(temp_vect, ncols = n, nrows = N)
+   # mat_sim = mat_sim == 0
+   # theta_0_estimates = apply(mat_sim, MARGIN = 1, FUN = sum)
+   # theta_0_estimates = sqrt(theta_0_estimates/n)
+   # var_theta_0_for_each_theta[i] = sum((theta_0_estimates - theta_vals[i])^2)/N
+
+   for (ii in 1:N) {
+     temp_vect = rbinom(n, 2, theta_vals[i])
+     temp_vect = temp_vect == 0
+     simulate_N[ii] = 1 - sqrt(sum(temp_vect)/n)
+   }
+
+   var_theta_0_for_each_theta[i] = sum((simulate_N - theta_vals[i])^2)/N
+
+
+ }
> var_theta_0_for_each_theta

 [1] 0.0001298333 0.0002541709 0.0003738888 0.0004728320 0.0005869841
 [6] 0.0007200118 0.0008558126 0.0008858230 0.0010000302 0.0010605262
[11] 0.0012217792 0.0012858581 0.0013935799 0.0015341275 0.0015389720
[16] 0.0017292130 0.0016570296 0.0017520842 0.0017844477 0.0019746090
[21] 0.0018957917 0.0021606012 0.0022098312 0.0022712450 0.0021134446
```

```
[26] 0.0023292737 0.0024023104 0.0023773007 0.0024837647 0.0026832055
[31] 0.0027801322 0.0031332682 0.0034612553 0.0037073586 0.0042293449
[36] 0.0047323146 0.0040348470 0.0026634800 0.0009129289

> plot(theta_vals, var_theta_0_for_each_theta)
```

```
>
```