The University of Sydney
School of Mathematics and Statistics

Computer Lab Week 5

STAT3023/3923/4023: Statistical Inference — Semester 2, 2021

Lecturers: Neville Weber and Michael Stewart

We shall compare three different estimators of a binomial success probability. If Y $B(2, \theta)$ then we have

$$
\begin{aligned}
P(Y = 0) &= (1 - \theta)^2 ; \\
P(Y = 1) &= 2\theta(1 - \theta) ; \\
P(Y = 2) &= \theta^2 .
\end{aligned}
$$

If we have an iid sample $Y_1, \ldots, Y_n$ then if we define $N_0 = \sum_{i=1}^n 1\{Y_i = 0\}$ as the number of 0's and $N_2 = \sum_{i=1}^n 1\{Y_i = 2\}$ as the number of 2's then we also have that

$$
\begin{aligned}
N_0 &\sim B(n, (1 - \theta)^2) \quad \text{and} \\
N_2 &\sim B(n, \theta^2) .
\end{aligned}
$$

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

1. Determine an unbiased estimator of θ which is a *linear* function of $\bar{Y}$. Call it $\hat{\theta}_1$.

2. Determine an 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 $\hat{\theta}_0$.

3. Determine an 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 $\hat{\theta}_2$.

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 find grid of θ values.

(a) Define a vector of θ-values: `th=(1:39)/40`.

(b) Set `n=100` for the sample size and `N=10000` for the number of simulation iterations.

(c) Define `MSE0=MSE1=MSE2=0`.

(d) Construct a double loop:

- For the outer loop, iterate `i` over values of `th`; for each such value of `i`, define `th.hat1=th.hat0=th.hat2=0` and then:
  - for the inner loop, iterate `j` over values `1,2,...,N`; for each such value of `j`,
    * generate a vector `Y` of `n` pseudo-random values from the $B(2, \texttt{th}[\texttt{i}]$ distribution;
    * compute values of the three estimators described above and save (respectively) as `th.hat1[j]`, `th.hat0[j]` and `th.hat2[j]`.
    * Save the average squared error for each in `MSE1[i]`, `MSE0[i]` and `MSE2[i]` respectively
- We are going to plot the `MSEs` against `th`:
  - Determine an appropriate range of y-values using `yrange=range(c(MSE0,MSE1,MSE2))`.
  - Plot `MSE1` against `th` (using `type='l',col='red',ylim=yrange`).
  - Add the other two MSE plots to this plot using `lines(...,col='blue')` and `lines(...,col='DarkGreen')` respectively.
  - Add an informative heading and legend.
- Finally, add the graph of the function $\frac{\theta(1-\theta)}{2n}$ to your plot using the command `curve(0.5*x*(1-x)/n,add=T,lty=2)`.

(e) Comment on the plots and how they may or may not agree with theory, in particular explain the importance of the final curve added to the plot. If necessary, modify the legend to appropriately indicate this last curve.