10-702 Homework 5 --Due in class on Wednesday, April 30.

This homework focuses on a comparison between a 1st order (so-called "Delta Method") estimate and a Bootstrap estimate of the variance of an MLE estimate of Bernoulli variance.

We have n *iid* Ber($\boldsymbol{p}$) observations: $\boldsymbol{X} = (X_1, \ldots, X_n)$.

Of course, the MLE $\hat{\boldsymbol{p}} = \sum_i x_i/n$ and $\mathrm{Var}_{\boldsymbol{p}}(\hat{\boldsymbol{p}}) = [\boldsymbol{p}(1-\mathbf{p})]/n$.
The MLE value of this variance with respect to the data is $[\hat{\boldsymbol{p}}(1-\hat{\boldsymbol{p}})]/n$.

What, however, is the $\mathrm{Var}_{\boldsymbol{p}}(\hat{\boldsymbol{p}}(1-\hat{\boldsymbol{p}}))$? That is what is the variance of this estimate of variance?

We contrast two techniques.

1) The ***Delta Method***. Let $h(\theta)$ be a function of interest of the parameter. Let $\hat{\theta}$ be its MLE based on n iid data, given θ. Using a first-order Taylor expansion and the Cramer-Rao lower bounds, we get that

$$\mathrm{Var}_\theta\,[h(\hat{\theta})] \approx [h'(\theta)]^2 / \boldsymbol{I}_n(\theta),$$

where $\boldsymbol{I}_n(\theta) = \mathbf{E}_\theta[\frac{\partial}{\partial\theta}\ln L(\theta \mid X)]^2$, which is the (expected) Fisher information in the sample, about the parameter θ.

So
$$[h'(\theta)]^2 / \boldsymbol{I}_n(\theta) = \frac{[\boldsymbol{h}'(\theta)]^2}{\boldsymbol{E}_\theta(-\frac{\partial^2}{\partial\theta^2}\ln \boldsymbol{L}(\theta|\boldsymbol{X}))}$$

$$\approx \frac{[\boldsymbol{h}'(\theta)]^2|_{\theta=\hat{\theta}}}{-\frac{\partial^2}{\partial\theta^2}\ln \boldsymbol{L}(\theta|\boldsymbol{X})|_{\theta=\hat{\theta}}}$$

where the denominator in this term is the observed Fisher information, $\hat{\boldsymbol{I}}_{\boldsymbol{n}}(\hat{\theta})$.

Problem 1: Apply this Delta Method to get an estimate, $\underline{\mathrm{Var}}(\hat{\boldsymbol{p}}(1-\hat{\boldsymbol{p}}))$ of $\mathrm{Var}_{\boldsymbol{p}}(\hat{\boldsymbol{p}}(1-\hat{\boldsymbol{p}}))$ based on the sample $\boldsymbol{X}$ of n, iid Ber($\boldsymbol{p}$) observations.
Calculate the values of $\underline{\mathrm{Var}}(\hat{\boldsymbol{p}}(1-\hat{\boldsymbol{p}}))$ for $\hat{\boldsymbol{p}} = 1/4$ and for $\hat{\boldsymbol{p}} = 1/3$ and $n = 24$.

Aside: Note what happens to this approximation at the distinguished value $\boldsymbol{p} = ½$!

2) The **Bootstrap** Method.
From the observed $\boldsymbol{x} = (x_1, \ldots, x_n)$ identify the Bootstrap estimate of the variance of $\hat{\boldsymbol{p}}(1-\hat{\boldsymbol{p}})$, as follows. That is, with a Bootstrap resampling of the given sample of n, the population of all n-fold Bootstrap draws from the sample of n yields the exact Bootstrap variance estimate of our quantity of interest:

$$\text{Var}^*(\hat{p}(1-\hat{p})) = (1/[n^n-1])\sum_{i=1}^{n^n}[\hat{p}(1-\hat{p})_i^* - \{\hat{p}(1-\hat{p})^*\}]^2$$

where $\hat{p}(1-\hat{p})_i^*$ is the i^{th} bootstrap sample and $\{\hat{p}(1-\hat{p})^*\}$ is the grand mean of these n^n samples.

(***Aside on notation***: the *electronic* version of this assignment lacks the conventional "bar" sign over the term within the brackets, needed to represent "average." So, I've added the extra bracket symbols to help remind you when reading the posted version of the assignment.)

Of course, this is infeasible to calculate for even modest sizes of n.

Problem 2: Instead, **sample** this Bootstrap population*. That is, let $n = 24$ in the original sample. You will carry out this Bootstrap sampling analysis ***twice***,

First use $\hat{p} = 1/4$. That is, first set $\hat{p} = 1/4$ in a sample of size 24, i.e. 6 of 24 trials are Bernoulli "successes." Of course, the order of the sequence of outcomes does not matter for the ***EDF***. From this sample (= population*) of size 24, draw $m = 1{,}000$ Bootstrap resamples each of size 24 and calculate

$$\text{Var}^*(\theta) = (1/[m-1]) \sum_{i=1}^{m}[\hat{\theta}_i^* - \{\hat{\theta}^*\}]^2, \text{ where } \theta = (\hat{p}(1-\hat{p})).$$

Again, $\hat{\theta}_i^*$ is the MLE estimate from the i^{th} bootstrap sample, and $\{\hat{\theta}^*\}$ is the average of the $m = 1{,}000$ samples.

Second, use $\hat{p} = 1/3$ again in a sample of 24 (so there are 8 "successes") and draw $m = 1{,}000$ Bootstrap resamples each of size 24, and calculate the Bootstrap estimate of variance of the same quantity $(\hat{p}(1-\hat{p}))$.

Problem 3) Compare the two pairs of estimates that you get for the values $\hat{p} = 1/4$ and $\hat{p} = 1/3$ of $\text{Var}_p(\hat{p}(1-\hat{p}))$. These are the values you get of $\underline{\text{Var}}(\hat{p}(1-\hat{p}))$ and of $\text{Var}^*(\hat{p}(1-\hat{p}))$ in the work above, each evaluated at the values $\hat{p} = 1/4$ and $\hat{p} = 1/3$.

Assuming that $p = 1/4$ and $n = 24$, the exact value for $\text{Var}_p(\hat{p}(1-\hat{p})) = .00484$,
and when $p = 1/3$ and $n = 24$, the exact value for $\text{Var}_p(\hat{p}(1-\hat{p})) = .00519$.

Which set of estimates is closer to the exact values?

Remembering that the Cramer-Rao bound is a ***lower*** bound, can you explain the pattern of estimates that you got?