Problem Set 6

Due Friday November 6, 10 am

Comments

- This covers material in Unit 9.
- It's due at 10 am (Pacific) on Friday Nov. 6, both submitted as a PDF to Pensive as well as committed to your GitHub repository.
- As usual, your electronic solution should be in the form of an Quarto file with Python code chunks. If you want to initially work in a Jupyter notebook, that is fine, but you'll need to run `quarto convert` to generate a qmd file before rendering to PDF and submitting.

Problems

1. Suppose I am interested in using a specific causal inference estimator on a dataset of interest. If I apply the estimator to that dataset I get an estimate of θ, say $\hat{\theta} = 2.342$, with an estimated standard error of, say, $\widehat{\text{se}}(\hat{\theta}) = 1.439$. Suppose I want to estimate the bias and MSE of the causal inference estimator via Monte Carlo simulation. I assume a true value, $\theta = 2.5$ (which I use when simulating the data), and I have m simulated datasets.

    a. How do I estimate the bias and the MSE using Monte Carlo. Write out the expectations of interest and the Monte Carlo estimators of those expectations.
    b. How do I estimate the simulation (Monte Carlo) uncertainty of my estimates of the bias and MSE, as mathematical expressions?
    c. Given the relevant values in the problem statement and the expressions from (a) and (b), discuss how one would decide if one should increase m. While your answer will involve the relevant quantities/expressions, I'm looking for an explanation of your judgment, not a mathematical result. You should include hypothetical numerical values to illustrate your reasoning.

2. Experimenting with importance sampling.

    a. Use importance sampling to estimate the mean (i.e., $\phi = E_f X$) of a truncated t distribution with 3 degrees of freedom, truncated such that $X < -4$. Have your sampling density be a normal distribution centered at -4 and then truncated so you only sample values less than -4 (this is called a half-normal distribution). You should be able to do this without discarding any samples (how?). Use $m = 10000$ samples. Create histograms of the weights $f(x)/g(x)$ and the summand $h(x)f(x)/g(x)$ to get a sense for whether $\text{Var}(\hat{\phi})$ is large. Note if there are any extreme weights that would have a very strong influence on $\hat{\phi}$. Estimate $\text{Var}(\hat{\phi})$. Hint:

remember that your $f(x)$ needs to be appropriately normalized or you need to adjust the weights per the class notes. For comparison, based on using numerical integration, which is feasible in this simple one-dimensional case, but increasingly infeasible in higher dimensions, the mean is -6.216.

b. Now use importance sampling to estimate the mean of the same truncated t distribution with 3 degrees of freedom, truncated such that $X < -4$, but have your sampling density be a t distribution, with 1 degree of freedom (not 3), centered at -4 and truncated so you only sample values less than -4. Again you shouldn't have to discard any samples. Respond to the same questions as above in part (a). In addition, compute an approximate 95% (simulation) uncertainty interval for your estimate, using the Monte Carlo simulation error, $\sqrt{\widehat{\text{Var}}(\hat{\phi})}$.