Gov 2002: Problem Set 7

Problem Set Instructions

This problem set is due on **November 8, 11:59 pm** Eastern time. Please upload a PDF of your solutions to Gradescope. We will accept hand-written solutions but we strongly advise you to typeset your answers in Rmarkdown. Please list the names of other students you worked with on this problem set.

Question 1 (25 points)

Assume $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, v^2)$.

(a) Use the Delta method to find the asymptotic distribution of $\hat{\theta}^4$.

(b) Use the Delta method to find the asymptotic distribution of $\frac{1}{1+\exp(-\hat{\theta})}$.

(c) We learned in class a geometric interpretation of the Delta method. In fact, you can relate the Delta method to Taylor polynomials. Explain how you can (approximately) derive the Delta method using the Taylor polynomials. (**Hint**: The first-order Taylor approximation of a function $f(x)$ at value c is $f(x) = f(c) + (x - c)f'(c)$.)

Question 2 (25 points)

Delta method is a very powerful tool for analyzing asymptotic properties of random variables. In this question, we will analyze what the distribution of quantiles would look like.

Let $X_1, ..., X_n$ be i.i.d. (continuous random variables). Let $F_X(x)$ be the CDF of X_i. Also consider a random variable $Y_n(x)$, which is defined as

$$Y_n(x) = \frac{1}{n} \sum_{i=1}^{n} Z_i$$

where $Z_i = I\{X_i \leq x\}$. Here, I is a usual indicator function, so $Z_i = 1$ if $X_i \leq x$.

(a) What distribution does Z_i follow? Name the distribution. What is its mean and variance? Express the mean and variance as a function of $F_X(x)$. (**Hint**: The fundamental bridge.)

(b) What is the asymptotic distribution of $Y_n(x)$?

(c) Apply the Delta method and identify the asymptotic distribution of $F_X^{-1}(Y_n(x))$. You may want to use the fact that $\frac{d}{dx}F_X^{-1}(x) = \frac{1}{f_X(F_X^{-1}(x))}$, where $f_X(x)$ is the PDF of X_i.

(d) Let $F^{-1}(p) = q_{X,p}$ be the p-th quantile of the distribution of X. Given the results from (c), show that the asymptotic distribution of p-th sample quantile of X_i, denoting it as $Q_{X,p}$, can be expressed as

$$\sqrt{n}(Q_{X,p} - q_{X,p}) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{\{f_X(q_{X,p})\}^2}\right)$$

(**Hint**: You can assume the following fact: $F_X^{-1}(Y_n(x)) = Q_{X,p}$.)

Question 3 (25 points)

Whereas we often work on the mean instead of the median, a practically useful quantity is often the median because it is not so much affected by the existence of outliers. Let's investigate what the asymptotic distributions of the median look like and how we can work on the median instead of the mean.

(a) Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_1, X_2, ..., X_n$ be n i.i.d. observations each following the same distribution as X. Let δ_1 denote the (true) median of X and let X_{med} be the median of the n observations. What is the asymptotic distribution of the sample median X_{med}? You can use the result from Question 2. The obtained variance in your answer can be a function of σ_1 but should not include f_X, δ_1, or μ_1.

(b) Just like comparing the mean, you come up with an idea to compare the median of two different distributions. Suppose that you have $n = N$ samples of $X_1, ..., X_n$, just like part (a). And then suppose that you have another set of samples $Y_1, ..., Y_m$, where $m = M$ observations are i.i.d. and follow $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. The true median of Y is δ_2. Assume that N and M are sufficiently large (but still finite). Also assume that X and Y are independent. In approximation, what would the distribution of $X_{med} - Y_{med}$ look like?

(c) How do the results from part (b) differ from the approximated distribution of the difference in means $(\overline{X} - \overline{Y}$, where $\overline{X} = \frac{\sum_{i=1}^N X_i}{N}$ and $\overline{Y} = \frac{\sum_{i=1}^M Y_i}{M}$)? (**Hint**: Use CLT.)

Question 4 (25 points)

Using the techniques we've learned in class, we can now work on a seemingly complex problem step by step.

Let $Y_1, \ldots, Y_n$ be i.i.d. random variables such that $E(Y_i) = 0, \text{Var}(Y_i) = 1$, and the fourth moment $E(Y_i^4)$ exists. Also, define

$$S_n = \frac{1}{n}\sum_{i=1}^n Y_i^2, \qquad V_n = \frac{1}{n}\sum_{i=1}^n \left(Y_i^2 - 1\right)^2.$$

The goal is to identify a constant c such that

$$c \cdot \frac{\sqrt{n}\left(\exp(S_n) - e\right)}{\sqrt{V_n}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Let's see how CLT, LLN, CMT, Slutsky, and Delta method are used in practice.

(a) What is the approximated distribution of S_n as $n \to \infty$? Your answer can include $Var(Y_i^2)$ but should not include $E[Y_i^2]$. (**Hint**: Use CLT...)

(b) Show that $\sqrt{V_n} \xrightarrow{p} \sqrt{Var(Y_i^2)}$. (**Hint**: LLN and CMT...)

(c) Given parts (a) and (b), prove the following asymptotic result. (**Hint**: Slutsky...)

$$\frac{\sqrt{n} \cdot (S_n - 1)}{\sqrt{V_n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

(d) Find the constant c such that

$$c \cdot \frac{\sqrt{n}\left(\exp\left(S_n\right) - e\right)}{\sqrt{V_n}} \xrightarrow{d} \mathcal{N}(0, 1)$$

(**Hint**: and Delta method!)