MATH/STAT 355: Problem Set 4

Prof. Taylor Okonek

Due: April 4, 2025

Correlated Data: Part 2

Recall the set-up from the previous Problem Set: Consider the following autoregressive model

$$X_n = \beta X_{n-1} + e_n, \quad n = 1, 2, 3, \ldots$$

with $-1 < \beta < 1$, $X_0 = 0$, and $e_1, \ldots, e_n$ i.i.d. with $E[e_i] = 0$ and $Var[e_i] = 1$ (Note that we have not assumed a pdf for e_i, so in this problem, this process is actually *semiparametric*). Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $\overline{e}_n = \frac{1}{n}\sum_{i=1}^n e_i$.

In our last problem set, we showed that $Var(X_n) = \sum_{i=0}^{n-1} \beta^{2i}$ and used Chebyshev's Inequality to show that $X_n/\sqrt{n} \xrightarrow{p} 0$. We'll now find the asymptotic distribution of $\sqrt{n}\overline{X}_n$!

1. We'll show that $\overline{X}_n \xrightarrow{p} 0$, in a few steps.
   (a) Show that $(1-\beta)\overline{X}_n - \overline{e}_n = -\frac{\beta X_n}{n}$.
   (b) Apply the Continuous Mapping Theorem, the WLLN, and what we showed in the Problem 2 to show that $\overline{X}_n \xrightarrow{p} 0$.
2. Show that $\sqrt{n}\overline{X}_n \xrightarrow{d} N(0, a)$ for some constant $a > 0$, and determine the value of a. (Hint: use Slutsky's, and again apply $X_n/\sqrt{n} \xrightarrow{p} 0$)

Asymptotics & the CLT

1. Suppose $\sqrt{n}(Y_n - \mu) \xrightarrow{d} N(0, \sigma^2)$. Find the asymptotic distribution of $\sqrt{n}(\log(Y_n) - \log(\mu))$, assuming $\mu > 0$.
2. Show that if $X_1, \ldots, X_n$ are iid random variables with $E[X_i] = \mu$ and $Var[X_i] = \sigma^2$, then

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{s_n}\right) \xrightarrow{d} N(0,1)$$

   where s_n is the sample standard deviation, given by $s_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2}$.
3. Suppose $Y_1, \ldots, Y_n \overset{iid}{\sim} Exponential(\lambda)$, where

$$f_Y(y_i \mid \lambda) = \lambda e^{-\lambda y_i} \quad y_i > 0, \lambda > 0$$

(a) Recall that $\sum_{i=1}^{n} Y_i \sim Gamma(n, \lambda)$. Explain why you cannot directly use the quantiles of *this* Gamma distribution to construct a confidence interval for λ.

(b) Show that $W = \frac{\lambda}{n} \sum_{i=1}^{n} Y_i \sim Gamma(n, n)$, and explain why you *can* use the quantiles of this distribution to construct a confidence interval for λ.

(c) Make a probability statement involving W, and use it to derive a formula for a $100(1-\alpha)\%$ confidence interval for λ. Use the notation Φ_a to represent the ath quantile of a $Gamma(n, n)$ distribution.

4. This problem walks you through comparing the asymptotic efficiency of the sample mean and the sample median as estimators for μ, for observations $X_i, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. (Note that, for a symmetric distribution, either seems like a "reasonable" estimator for μ!)

(a) Suppose $Y_1, \dots, Y_n$ are an iid sample from an *unknown* probability distribution with support $(-\infty, \infty)$, and our goal is to estimate the theoretical CDF $F(y)$ based on what we observe. Write an equation for the *empirical* (sample) CDF $F_n(y)$ based on the observations $Y_1, \dots, Y_n$. (Hint: probabilities are just expectations of indicator functions)

(b) Determine the distribution of $nF_n(y)$.

(c) Using what you solved for in part (b), determine the asymptotic distribution of $F_n(y)$ by applying the central limit theorem.

This set-up will allow us to determine the asymptotic distribution of sample quantiles (necessary for determining the asymptotic efficiency of the sample median). We'll skip a good bit of the math here and save it for a further course in statistical theory.

(d) For $0 < p < 1$, the asymptotic distribution of the pth sample quantile $X_{([n+1]p)}$ is given by

$$\sqrt{n}(X_{([n+1]p)} - F^{-1}(p)) \xrightarrow{d} N\left(0, \frac{p(1-p)}{[f(F^{-1}(p))]^2}\right)$$

Confirm that when $p = 1/2$, $X_{([n+1]p)}$ is indeed the sample median (recall that this is defined in the MLE section of our course notes), assuming n is odd.

(e) Let $m = F^{-1}(1/2)$ denote the *population* median, and simplify the asymptotic distribution in part (d) when $X_{([n+1]p)}$ is the sample median. For notation, you may let the sample median be denoted $\tilde{X}$.

(f) Recall that $X_i, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Calculate the relative asymptotic efficiency of the sample median to the sample mean when $\sigma^2 = 1$. (Hint: First calculate the asymptotic variance of the sample median and the sample mean, using the CLT and plugging things in)

*Fun Fact: Parts (a) through (c) are some of the building blocks of nonparametric statistics!