August 2020 Name:________________________

Comprehensive written exam – STOR655 Mathematical Statistics

Unless otherwise noted, all problem parts have equal weight. In budgeting your time expect that some part will take longer than others. When solving multi-part problems feel free to use results of earlier parts even if you cannot solve them in proving later parts.

1. Let $X_n \xrightarrow{\mathcal{D}} X > 0$ and $X_n/Y_n \xrightarrow{P} 1$. Show that $Y_n \xrightarrow{\mathcal{D}} X$. Could you drop the $X > 0$ condition?

2. Let us consider two integrals

$$I = \int_0^1 y^{-1}\sin(2\pi y)\,dy = -0.159\ldots \text{ and } J = \int_0^1 y^{-1}\sin(2\pi y^{-1})\,dy = 0.153\ldots.$$

For numerical calculations one can consider a Monte Carlo approximation by simulating $Y_1,\ldots,Y_n$ i.i.d. $U(0,1)$ and computing

$$\hat{I}_n = n^{-1}\sum_{i=1}^{n} Y_i^{-1}\sin(2\pi Y_i) \text{ and } \hat{J}_n = n^{-1}\sum_{i=1}^{n} Y_i^{-1}\sin(2\pi Y_i^{-1}).$$

Jan simulated $n = 10^9$ observations on his computer and obtained $\hat{I}_{10^9} = -0.159\ldots$ and $\hat{J}_{10^9} = 2.926\ldots$.

   (a) What does the SLLN says about $\hat{I}_n$? Estimate $P(|\hat{I}_n - I| > 10^{-3})$. (Hint: Hoeffding's inequality might be useful.)

   (b) What does the SLLN says about $\hat{J}_n$? Explain the result of the computer experiment.

3. Assume $\begin{pmatrix}X_1\\Y_1\end{pmatrix},\ldots\begin{pmatrix}X_n\\Y_n\end{pmatrix}$ are i.i.d. bivariate normal$\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1,\rho\\\rho,1\end{pmatrix}\right)$

   (a) Prove that $T = (\sum_{i=1}^n X_i^2 + Y_i^2, \sum_{i=1}^n X_iY_i)$ is a minimal sufficient statistics.

   (b) Consider $R_n(\mathbb{X}) = \frac{1}{n}\sum_{i=1}^n X_iY_i$. Is it a consistent estimator of ρ. Find the asymptotic variance of $R_n(\mathbb{X})$

(c) Next consider the traditional correlation coefficient

$$C_n = \frac{R_n - \bar{X}_n \bar{Y}_n}{\sqrt{n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2 \cdot n^{-1}\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}}.$$

Is it a consistent estimator of ρ. Find its asymptotic variance. Show as much of the calculation as you can.

(d) Is any of these estimators asymptotically efficient? If no could you suggest an asymptotically efficient estimator?

4. Consider the linear regression model $Y_i = \alpha + \beta x_i + e_i,\ i = 1, \ldots, n$.

(a) Show that the MLE of the coefficient $\hat{\beta}_{MLE}$ is asymptotically normal when e_i are iid $N(0, \sigma^2)$.

(b) Devise a bootstrap procedure for setting confidence intervals for the coefficient β.

(c) Show consistency of the bootstrap confidence interval using appropriate assumptions. Hint: Recall that for consistency of the bootstrap CIs it is enough to show

$$P\left(\frac{\hat{\beta} - \beta}{\hat{\sigma}_n} \leq x \mid P_0\right) \to F(x) \text{ and } P\left(\frac{\beta^\star - \hat{\beta}}{\sigma_n^\star} \leq x \mid \hat{P}_n\right) \xrightarrow{P} F(x).$$