SDS 383C: Statistical Modeling I, Fall 2022
Homework 1, Due Sept 19, 12:00 Noon

Instructor: Abhra Sarkar (abhra.sarkar@utexas.edu)
Teaching Assistant: Preston Biro (prestonbiro@utexas.edu)
Department of Statistics and Data Sciences
The University of Texas at Austin
2317 Speedway D9800, Austin, TX 78712-1823, USA

All homework must be submitted typed-in as a single pdf file. Name the file "firstname-lastname-SDS383C-HW-1.pdf" Submit this file without compression such as zip or rar. Figures accompanying the solutions must be presented close to the actual solution. Computer codes will be rarely evaluated but must still be submitted separately from the main file. Codes must be commented properly and should run easily on other machines. Precise, concise, clear, innovative solutions may be rewarded with bonus points. Explain your answer with logic reasoning and/or mathematical proofs. Organize your solutions in the same order as they were presented. If you can solve a problem using multiple techniques, present only your best solution.

1. (15 points) Prove Slutsky's theorem (the version taught in class).

2. (10 points) For $y \sim \text{Poisson}(\lambda)$, show that $\mathbb{E}(y) = \text{var}(y) = \lambda$. Method of moments suggests $\overline{y}_n$, the sample mean, as well as s_n^2, the sample variance, could both be reasonable estimators of λ. Which one would you prefer? Why?

3. (10 points) For $(x_1, y_1), \ldots, (x_n, y_n) \overset{iid}{\sim} f_{x,y}$ with finite second order moments. Show that the sample correlation coefficient r_n converges in probability to the population correlation coefficient ρ.

4. (10 points) Notations having their usual significance, for $y_1, \ldots, y_n \overset{iid}{\sim} \text{Ga}(\alpha, \beta)$, a method of moment estimator of α is $\widehat{\alpha}_n = \overline{y}_n^2 / s_n^2$. Using the multivariate delta method, show that $SE(\widehat{\alpha}_n) \approx \sqrt{2\alpha(\alpha+1)/n}$ for large values of n.

5. (10 points) Now fix the values of α and β. For your chosen values of α and β, draw a random sample of size $n = 50$ from a $\text{Ga}(\alpha, \beta)$ distribution. Using method of moments and assuming α and β to now be unknown, estimate α and β. Plot the histogram of the samples, superimposed with the true density and the estimated density.

   Repeat the above procedure $B = 50, 500$ and 1000 times. Plot a histogram of $\sqrt{n}\frac{(\widehat{\alpha}_n - \alpha)}{\sqrt{2\widehat{\alpha}_n(\widehat{\alpha}_n+1)}}$ for each of the above values of B. In each case, superimpose a $\text{Normal}(0, 1)$ distribution over the histogram.

   For your final output, provide a very brief description of what you did, the plots, the codes, and your general comments, if any.

6. (5 points) For a $\text{Normal}(\mu, \sigma^2)$ distribution, show that the MGF is $M(t) = \exp(\mu t + \sigma^2 t^2/2)$.

7. (10 points) Show that a binomial random variable R with denominator m and probability π has cumulant generating function $K(t) = m\log(1 - \pi + \pi e^t)$. Find $\lim K(t)$ as $m \to \infty$, $\pi \to 0$ in a way so that $m\pi \to \lambda > 0$. Show that

$$\Pr(R = r) \to \frac{\lambda^r}{r!} e^{-\lambda},$$

and hence establish that $R \xrightarrow{d}$ Poisson(λ). Using your favorite programming language, provide a numerical illustration of the result.

[Hints: $\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x$.]

8. (5 points) If $Z \sim$ Normal$(0, 1)$, derive the density of $Y = Z^2$. Although Y is determined by Z, show that they are uncorrelated.

9. (5 points) Let $Y = X_1 + bX_2$ where the X_j are independent normals with means μ_j and variances σ_j^2. Show that conditional on $X_2 = x$, the distribution of Y is normal with mean $\mu_1 + bx$ and variance σ_1^2. Hence establish that

$$\int \frac{1}{\sigma_1}\phi\left(\frac{y - \mu_1 - bx}{\sigma_1}\right) \frac{1}{\sigma_2}\phi\left(\frac{x - \mu_2}{\sigma_2}\right) dx = \frac{1}{\sqrt{\sigma_1^2 + b^2\sigma_2^2}}.\phi\left(\frac{y - \mu_1 - b\mu_2}{\sqrt{\sigma_1^2 + b^2\sigma_2^2}}\right).$$

10. Read pages 62-75 (Section 3.2: Normal Model) from AC Davison's Statistical Models.