STA 2201/442 Assignment 2

1. This is about how to simulate from a continuous univariate distribution. Let the random variable X have a continuous distribution with density $f_X(x)$ and cumulative distribution function $F_X(x)$. Suppose the cumulative distribution function is strictly increasing over the set of x values where $F_X(x) > 0$, so that $F_X(x)$ has an inverse. Let U have a uniform distribution over the interval $(0, 1)$. Show that the random variable $Y = F^{-1}(U)$ has the same distribution as X. Hint: Start by finding $F_U(u) = Pr\{U \leq u\}$.

2. A random variable following the *Pareto distribution* has density $f(x) = \frac{\alpha}{x^{\alpha+1}}$ for $x > 1$ and zero otherwise, where the parameter $\alpha > 0$. The Pareto distribution is handy because it is right skewed with heavy tails, yielding observations that would be outliers relative to some other distribution. If you want to try some statistical method with data that are not normal, the Pareto definitely fits that description. Let X have a Pareto distribution with parameter $\alpha > 0$.

   (a) Find $E(X)$; show your work. For what values of α does this quantity exist?

   (b) Find $F(x) = Pr\{X \leq x\}$. Show your work.

   (c) Find $F^{-1}(x)$. Show your work.

   (d) Give a formula for the 0.50 quantile. The answer is an expression in α.

   (e) Using `R`, simulate 1000 values from a Pareto distribution with expected value 3. Calculate and display the sample mean. Bring the `R` printout to the quiz; it may be handed in.

   (f) Make a QQ plot to see whether the data come from a Gamma distribution with $\beta = 1$. Determine α based on the data. See `help(qgamma)`. Make sure to include the line $y = x$. Do you see evidence that the largest order statistics are much bigger than expected, suggesting a heavy tail? Display the plot and the code that generated it. Bring the `R` printout to the quiz; it may be handed in.

   (g) Make another QQ plot, this time to see whether the data come from a Pareto distribution; determine α based on the data. Make sure to include the line $y = x$. What do you conclude? Display the plot and the code that generated it. Bring the `R` printout to the quiz; it may be handed in.

3. Let $X_1, \ldots, X_n$ be a random sample from a Binomial distribution with parameters 3 and θ. That is,

$$P(X_i = x_i) = \binom{3}{x_i} \theta^{x_i} (1-\theta)^{3-x_i},$$

   for $x_i = 0, 1, 2, 3$. Find a reasonable estimator of θ, and prove that it is strongly consistent. Where you get your estimator does not really matter, but please state how you thought of it.

4. Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with density

$$f(x;\tau) = \frac{\tau^{1/2}}{\sqrt{2\pi}} e^{-\frac{\tau x^2}{2}},$$

where the parameter $\tau > 0$. Let

$$\widehat{\tau} = \frac{n}{\sum_{i=1}^n X_i^2}.$$

Is $\widehat{\tau}$ a consistent estimator of τ? Answer Yes or No and prove your answer. Hint: You can just write down $E(X^2)$ by inspection. This is a very familiar distribution.

5. Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean μ. Show that $T_n = \frac{1}{n+400}\sum_{i=1}^n X_i$ is a strongly consistent estimator of μ.

6. Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Prove that the sample variance $S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$ is a strongly consistent estimator of σ^2.

7. Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λ, and let X be a general random variable with that distribution. You know that $E(X) = Var(X) = \lambda$; there is no need to prove it.

   From the Strong Law of Large Numbers, it follows immediately that $\overline{X}_n$ is strongly consistent for λ. Let

$$\widehat{\lambda} = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{n-4}.$$

   Is $\widehat{\lambda}$ also a consistent estimator of λ? Answer Yes or No and prove your answer.

8. Independently for $i = 1, \ldots, n$, let

$$Y_i = \beta X_i + \epsilon_i,$$

where $E(X_i) = E(\epsilon_i) = 0$, $Var(X_i) = \sigma_X^2$, $Var(\epsilon_i) = \sigma_\epsilon^2$, and ϵ_i is independent of X_i. Let

$$\widehat{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}.$$

   (a) Show that $\widehat{\beta}$ is the least-squares estimator of β.

   (b) Is $\widehat{\beta}$ a consistent estimator of β? Answer Yes or No and prove your answer.

9. In this problem, you'll use (without proof) the *variance rule*, which says that if θ is a real constant and $T_1, T_2, \ldots$ is a sequence of random variables with

$$\lim_{n\to\infty} E(T_n) = \theta \text{ and } \lim_{n\to\infty} Var(T_n) = 0,$$

then $T_n \stackrel{P}{\rightarrow} \theta$.

In Problem 8, the independent variables are random. Here they are fixed constants, which is more standard (though a little strange if you think about it). Accordingly, let

$$Y_i = \beta x_i + \epsilon_i$$

for $i = 1, \ldots, n$, where $\epsilon_1, \ldots, \epsilon_n$ are a random sample from a distribution with expected value zero and variance σ^2, and β and σ^2 are unknown constants.

(a) What is $E(Y_i)$?

(b) What is $Var(Y_i)$?

(c) Find the Least Squares estimate of β by minimizing $Q = \sum_{i=1}^n (Y_i - \beta x_i)^2$ over all values of β. Let $\widehat{\beta}_n$ denote the point at which Q is minimal.

(d) Is $\widehat{\beta}_n$ unbiased? Answer Yes or No and show your work.

(e) Give a sufficient condition for $\widehat{\beta}_n$ to be consistent. Show your work. Remember, in this model the x_i are fixed constants, not random variables.

(f) Let $\widehat{\beta}_{2,n} = \frac{\overline{Y}_n}{\overline{x}_n}$. Is $\widehat{\beta}_{2,n}$ unbiased? Consistent? Answer Yes or No to each question and show your work.

(g) Prove that $\widehat{\beta}_n$ is a more accurate estimator than $\widehat{\beta}_{2,n}$ in the sense that it has smaller variance. Hint: The sample variance of the independent variable values cannot be negative.

10. Let $X_1, \ldots, X_n$ be a random sample from a Gamma distribution with $\alpha = \beta = \theta > 0$. That is, the density is

$$f(x;\theta) = \frac{1}{\theta^\theta \Gamma(\theta)} e^{-x/\theta} x^{\theta-1},$$

for $x > 0$. Let $\widehat{\theta} = \overline{X}_n$. Is $\widehat{\theta}$ a consistent estimator of θ? Answer Yes or No and prove your answer.

11. Let $X_1, \ldots, X_n$ be a random sample from a distribution with expected value μ and variance σ_x^2. Independently of $X_1, \ldots, X_n$, let $Y_1, \ldots, Y_n$ be a random sample from a distribution with the same expected value μ and variance σ_y^2. Let Let $T_n = \alpha \overline{X}_n + (1-\alpha)\overline{Y}_n$, where $0 \le \alpha \le 1$. Is T_n a consistent estimator of μ? Answer Yes or No and show your work.

12. Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Prove that the sample variance $S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$ is a consistent estimator of σ^2.

13. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a bivariate distribution with $E(X_i) = \mu_x$, $E(Y_i) = \mu_y$, $Var(X_i) = \sigma_x^2$, $Var(Y_i) = \sigma_y^2$, and $Cov(X_i, Y_i) = \sigma_{xy}$. Show that the sample covariance $S_{xy} = \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$ is a consistent estimator of σ_{xy}.

14. A sample of size $n = 200$ is drawn from a distribution with density $f(x) = \frac{1}{\theta}e^{-x/\theta}$ for $x > 0$, where $\theta > 0$. The sample mean is $\overline{X}_n = 11.6$. Give an approximate 95% confidence interval for θ. Your final answer is two numbers, a lower confidence limit and an upper confidence limit. Where do you use continuity?

15. An application of the Law of Large Numbers is Monte Carlo estimation of probabilities. It is good practice to accompany such estimates with confidence intervals. Suppose the probability being estimated is 1/3.

    (a) What Monte Carlo sample size is required for the 99% margin of error (half the width of the confidence interval) to be about 0.01? The answer is an integer. Show your work.

    (b) What Monte Carlo sample size is required for the 99% margin of error to be about 0.01 regardless of the probability being estimated? Show your work.

16. Let Z_1 and Z_2 be independent standard normal random variables, and let $Y = Z_1 Z_2$, but rounded to the nearest integer. Give a Monte Carlo estimate of the probability that $Y = 1$. Accompany your estimate with a 99% margin of error that is no greater than 0.01.

17. If the $p \times 1$ random vector $\mathbf{X}$ has variance-covariance matrix $\boldsymbol{\Sigma}$ and $\mathbf{A}$ is an $m \times p$ matrix of constants, prove that the variance-covariance matrix of $\mathbf{AX}$ is $\mathbf{A\Sigma A}'$. Start with the definition of a variance-covariance matrix:

$$V(\mathbf{Z}) = E(\mathbf{Z} - \boldsymbol{\mu}_z)(\mathbf{Z} - \boldsymbol{\mu}_z)'.$$

18. If the $p \times 1$ random vector $\mathbf{X}$ has mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, show $\boldsymbol{\Sigma} = E(\mathbf{XX}') - \boldsymbol{\mu\mu}'$.

19. Let the $p \times 1$ random vector $\mathbf{X}$ have mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, and let $\mathbf{c}$ be a $p \times 1$ vector of constants. Find $V(\mathbf{X} + \mathbf{c})$. Show your work.

20. Let $\mathbf{X} = (X_1, X_2, X_3)'$ be multivariate normal with

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 0 \\ 6 \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

    Let $Y_1 = X_1 + X_2$ and $Y_2 = X_2 + X_3$. Find the joint distribution of Y_1 and Y_2.

21. Let X_1 be Normal(μ_1, σ_1^2), and X_2 be Normal(μ_2, σ_2^2), independent of X_1. What is the joint distribution of $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$? What is required for Y_1 and Y_2 to be independent? Hint: Use matrices.

22. Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is an $n \times p$ matrix of known constants, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown constants, and $\boldsymbol{\epsilon}$ is multivariate normal with mean zero and covariance matrix $\sigma^2 \mathbf{I}_n$, where $\sigma^2 > 0$ is a constant. In the following, it may be helpful to recall that $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$.

   (a) What is the distribution of $\mathbf{Y}$?

   (b) The maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. What is the distribution of $\hat{\boldsymbol{\beta}}$? Show the calculations.

   (c) Let $\widehat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. What is the distribution of $\widehat{\mathbf{Y}}$? Show the calculations.

   (d) Let the vector of residuals $\mathbf{e} = (\mathbf{Y} - \widehat{\mathbf{Y}})$. What is the distribution of $\mathbf{e}$? Show the calculations. Simplify both the expected value (which is zero) and the covariance matrix.

23. Show that if $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $Y = (\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ has a chi-square distribution with p degrees of freedom.