STA 2101/442 Assignment Three[1]

The questions are just practice for the quiz, and are not to be handed in. Use R as necessary for Question 19, and **bring your printout to the quiz.**

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 $0 < F_X(x) < 1$, so that $F_X(x)$ has an inverse. Let U have a uniform distribution over the interval $(0, 1)$. Show that the random variable $Y = F_X^{-1}(U)$ has the same distribution as X. Hint: You will need an expression for $F_U(u) = Pr\{U \leq u\}$, where $0 \leq u \leq 1$.

2. 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 the maximum likelihood estimator of θ, and show that it is strongly consistent.

3. 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.

4. 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.

5. 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.

6. Independently for $i = 1, \ldots, n$, let
$$Y_i = \beta X_i + \epsilon_i,$$
where $E(X_i) = E(\epsilon_i) = 0$, $Var(X_i) = \sigma^2_X$, $Var(\epsilon_i) = \sigma^2_\epsilon$, and ϵ_i is independent of X_i. Let
$$\widehat{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}.$$
Is $\widehat{\beta}$ a consistent estimator of β? Answer Yes or No and prove your answer.

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

[1]Copyright information is at the end of the last page.

In Problem 6, 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 nice simple condition on the x_i values that guarantees $\widehat{\beta}_n$ will 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. Do you need a condition on the x_i values ?

(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.

8. Let X be a random variable with expected value μ and variance σ^2. Show $\frac{X}{n} \stackrel{p}{\rightarrow} 0$.

9. 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.

10. The ordinary univariate Central Limit Theorem says that if $X_1, \ldots, X_n$ are a random sample (independent and identically distributed) from a distribution with expected value μ and variance σ^2, then

$$Z_n^{(1)} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \stackrel{d}{\rightarrow} Z \sim N(0,1).$$

An application of some Slutsky theorems (see lecture slides) shows that also,

$$Z_n^{(2)} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\widehat{\sigma}_n} \stackrel{d}{\rightarrow} Z \sim N(0,1),$$

where $\widehat{\sigma}_n$ is any consistent estimator of σ. For this problem, suppose that $X_1, \ldots, X_n$ are Bernoulli(θ).

(a) What is μ?

(b) What is σ^2?

(c) Re-write $Z_n^{(1)}$ for the Bernoulli exanple.

(d) What about $Z_n = \frac{\sqrt{n}(\overline{X}_n - \theta)}{\sqrt{\overline{X}_n(1-\overline{X}_n)}}$? Does Z_n converge in distribution to a standard normal? Why or why not?

(e) What about the t statistic $T_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n}$, where S_n is the sample standard deviation? Does T_n converge in distribution to a standard normal? Why or why not?

11. 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 $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$. Start with the definition of a variance-covariance matrix:
$$V(\mathbf{Z}) = E(\mathbf{Z} - \boldsymbol{\mu}_z)(\mathbf{Z} - \boldsymbol{\mu}_z)'.$$

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

13. 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.

14. Let $\mathbf{X}$ be a $p \times 1$ random vector with mean $\boldsymbol{\mu}_x$ and variance-covariance matrix $\boldsymbol{\Sigma}_x$, and let $\mathbf{Y}$ be a $q \times 1$ random vector with mean $\boldsymbol{\mu}_y$ and variance-covariance matrix $\boldsymbol{\Sigma}_y$. Recall that $C(\mathbf{X}, \mathbf{Y})$ is the $p \times q$ matrix $C(\mathbf{X}, \mathbf{Y}) = E\left((\mathbf{X} - \boldsymbol{\mu}_x)(\mathbf{Y} - \boldsymbol{\mu}_y)'\right)$.
    (a) What is the (i, j) element of $C(\mathbf{X}, \mathbf{Y})$?
    (b) For this item, $p = q$. Find an expression for $V(\mathbf{X} + \mathbf{Y})$ in terms of $\boldsymbol{\Sigma}_x$, $\boldsymbol{\Sigma}_y$ and $C(\mathbf{X}, \mathbf{Y})$. Show your work.
    (c) Simplify further for the special case where $Cov(X_i, Y_j) = 0$ for all i and j.
    (d) Let $\mathbf{c}$ be a $p \times 1$ vector of constants and $\mathbf{d}$ be a $q \times 1$ vector of constants. Find $C(\mathbf{X} + \mathbf{c}, \mathbf{Y} + \mathbf{d})$. Show your work.

15. Denote the moment-generating function of a random variable Y by $M_Y(t)$. The moment-generating function is defined by $M_Y(t) = E(e^{Yt})$. Recall that the moment-generating function corresponds uniquely to the probability distribution.
    (a) Let a be a constant. Prove that $M_{aX}(t) = M_X(at)$.
    (b) Prove that $M_{X+a}(t) = e^{at} M_X(t)$.
    (c) Let X_1 and X_2 be *independent* random variables. Prove that
    $$M_{X_1+X_2}(t) = M_{X_1}(t)\, M_{X_1}(t).$$
    For convenience, you may assume that X_1 and X_2 are continuous, so you will integrate. This result extends to $M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n M_{X_i}(t)$, but you don't have to show it. (You could use induction.)

16. Recall that if $X \sim N(\mu, \sigma^2)$, it has moment-generating function $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$.
    (a) Let $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, where a and b are constants. Find the distribution of Y. Show your work.
    (b) Let $X \sim N(\mu, \sigma^2)$ and $Z = \frac{X-\mu}{\sigma}$. Find the distribution of Z.
    (c) Let $X_1, \ldots, X_n$ be random sample from a $N(\mu, \sigma^2)$ distribution. Find the distribution of $Y = \sum_{i=1}^n X_i$.
    (d) Let $X_1, \ldots, X_n$ be random sample from a $N(\mu, \sigma^2)$ distribution. Find the distribution of the sample mean $\overline{X}$.
    (e) Let $X_1, \ldots, X_n$ be random sample from a $N(\mu, \sigma^2)$ distribution. Find the distribution of $Z = \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}$.

17. A Chi-squared random variable X with parameter $\nu > 0$ has moment-generating function $M_X(t) = (1-2t)^{-\nu/2}$.

    (a) Let $X_1, \ldots, X_n$ be independent random variables with $X_i \sim \chi^2(\nu_i)$ for $i = 1, \ldots, n$. Find the distribution of $Y = \sum_{i=1}^n X_i$.

    (b) Let $Z \sim N(0,1)$. Find the distribution of $Y = Z^2$. For this one, you need to integrate. Recall that the density of a normal random variable is $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

    (c) Let $X_1, \ldots, X_n$ be random sample from a $N(\mu, \sigma^2)$ distribution. Find the distribution of $Y = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)^2$.

    (d) Let $Y = X_1 + X_2$, where X_1 and X_2 are independent, $X_1 \sim \chi^2(\nu_1)$ and $Y \sim \chi^2(\nu_1 + \nu_2)$, where ν_1 and ν_2 are both positive. Show $X_2 \sim \chi^2(\nu_2)$.

    (e) Let $X_1, \ldots, X_n$ be random sample from a $N(\mu, \sigma^2)$ distribution. Show

    $$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

    where $S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$. Hint: $\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n \left(X_i - \overline{X} + \overline{X} - \mu\right)^2 = \ldots$
    You may use the independence of $\overline{X}$ and S^2 without proof, for now.

18. Recall the definition of the t distribution. If $Z \sim N(0,1)$, $W \sim \chi^2(\nu)$ and Z and W are independent, then $T = \frac{Z}{\sqrt{W/\nu}}$ is said to have a t distribution with ν degrees of freedom, and we write $T \sim t(\nu)$.

    As in the last question, let $X_1, \ldots, X_n$ be random sample from a $N(\mu, \sigma^2)$ distribution. Show that $T = \frac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t(n-1)$. Once again, you may use the independence of $\overline{X}$ and S^2 without proof for now.

19. Fine machine screws are manufactured so as to have a diameter of one millimetre, but of course nothing is perfect. The screws have an *expected* diameter of one millimetre, and if the manufacturing process is running properly, they also have a very small standard deviation. As long as the standard deviation is three micrometres (thousandths of a millimetre) or less, virtually all the screws will fit properly. The industrial quality control process involves taking repeated samples of screws, measuring them, and determining whether the standard deviation is greater than three.

    Three features of this application are a bit unusual. First, the data really *are* normal. Variation from screw to screw is driven by a large number of separate tiny influences that more or less add up, and so the Central Limit Theorem applies. Second, nobody pays much attention to the mean; it's virtually always about one millimetre. When the manufacturing process starts to go goes wrong, what happens is that the variance goes up. Third, nobody cares if the standard deviation is *less* than three micrometres. They only worry if it's too big, because then they have to stop the assembly line and service the machines. So, a one-tailed test really is appropriate.

    It's a pain to measure those screws, so the engineers take samples of size ten. The most recent sample yields a sample mean of 1002.687 and a sample standard deviation of 4.51.

    (a) What is the model?

    (b) What is the null hypothesis, in symbols? What is the alternative hypothesis?

    (c) An earlier problem suggests a test statistic. Write down the formula.

    (d) Use R to calculate the p-value. The answer is a number. Do you reject H_0 at $\alpha = 0.05$? Do you stop the assembly line?

    (e) Derive a $(1-\alpha)100\%$ confidence interval for σ (not σ^2). Show your work.

(f) Calculate your confidence interval for the numerical data given above, using $\alpha = 0.05$. Your answer is a set of two numbers.

(g) Don't you think a *one-sided* confidence interval would be better here? Derive the formula for a statistic (say L, for upper limit) such that $Pr\{\sigma < L\} = 1 - \alpha$.

(h) Calculate your one-sided confidence interval for the numerical data given above, using $\alpha = 0.05$. Your answer is a single number.

(i) I hope that you are at least a little uncomfortable with that sample size of $n = 10$. Is it enough? The answer to such questions is always another question: "Enough for what?" Suppose that if the true value of σ is 4 or more, the quality control engineers want to be able to detect it with probability at least 0.90, using the usual $\alpha = 0.05$ significance level. What's the smallest sample size they can get away with? Please approach the problem this way.

   i. First, derive a formula for the power of the test, for general n, α, σ_0 and true σ.

   ii. What is the power for an n of exactly 10 (the engineers' intuitive choice) when the true value of σ is 4? The answer is a number

   iii. Then, plug in all the numbers except n. Starting with a nice small sample size (one lower than 10), increase n, calculating the power each time, until the power exceeds 0.90. Your final answer is a number.
