STA 2101/442 Assignment Three[1]

The questions are just practice for the quiz, and are not to be handed in. You may use R as a calculator, but do not bring printouts to the quiz. But **Please bring a calculator to the quiz.**

1. Suppose $X_1, \ldots, X_n$ are a random sample from a distribution with mean μ and variance σ^2. The central limit theorem says $\sqrt{n}\left(\overline{X}_n - \mu\right) \stackrel{d}{\rightarrow} T \sim N(0, \sigma^2)$. One version of the delta method says that if $g(x)$ is a function whose derivative is continuous in a neighbourhood of $x = \mu$, then $\sqrt{n}\left(g(\overline{X}_n) - g(\mu)\right) \stackrel{d}{\rightarrow} g'(\mu)T$. In many applications, both μ and σ^2 are functions of some parameter θ.

   (a) Let $X_1, \ldots, X_n$ be a random sample from a Bernoulli distribution with parameter θ. Find the limiting distribution of
   $$Z_n = 2\sqrt{n}\left(\sin^{-1}\sqrt{\overline{X}_n} - \sin^{-1}\sqrt{\theta}\right).$$
   Hint: $\frac{d}{dx}\sin^{-1}(x) = \frac{1}{\sqrt{1-x^2}}$.

   (b) In the same old coffee taste test example, suppose 60 out of 100 consumers prefer the new blend of coffee beans. Using your answer to the first part of this question, test the null hypothesis using a variance-stabilized test statistic. Give the value of the test statistic (a number), and state whether you reject H_0 at the usual $\alpha = 0.05$ significance level.

   (c) If the probability of an event is p, the *odds* of the event is (are?) defined as $p/(1-p)$. Suppose again that $X_1, \ldots, X_n$ are a random sample from a Bernoulli distribution with parameter θ. In this case the *log odds* of $X_i = 1$ would be estimated by
   $$Y_n = \log \frac{\overline{X}_n}{1 - \overline{X}_n}.$$
   That's the natural log, of course. Find the approximate large-sample distribution (that is, the asymptotic distribution) of Y_n. It's norrmal. Your job is to give the approximate mean and variance Y_n.

   (d) Again using the Taste Test data, give a 95% confidence interval for the log odds of preferring the new brand. Your answer is a pair of numbers.

   (e) Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with parameter θ, so that $E(X_i) = \theta$ and $Var(X_i) = \theta^2$.

      i. Find a variance-stabilizing transformation. That is, find a function $g(x)$ such that the limiting distribution of
      $$Y_n = \sqrt{n}\left(g(\overline{X}_n) - g(\theta)\right)$$
      does not depend on θ.

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

ii. According to a Poisson process model for calls answered by a service technician, service times (that is, time intervals between taking 2 successive calls; there is always somebody on hold) are independent exponential random variables with mean θ. In 50 successive calls, one technician's mean service time was 3.4 minutes. Test whether this technician's mean service time differs from the mandated average time of 3 minutes. Use your answer to the first part of this question.

(f) Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution on $(0, \theta)$.

i. What is the limiting distribution of $\sqrt{n}\left(\overline{X}_n - \mu\right)$? Just give the answer; there is no need to show any work.

ii. What is the limiting distribution of $2\sqrt{n}\left(\overline{X}_n - \mu\right)$? Just give the answer; there is no need to show any work. But what Slutsky Lemma are you using? Check the lecture slides if necessary.

iii. Find a variance-stabilizing transformation that produces a standard normal distribution. That is, letting $T_n = 2\overline{X}_n$, find a function $g(x)$ such that the limiting distribution of

$$Y_n = \sqrt{n}\left(g(T_n) - g(\theta)\right)$$

is standard normal.

(g) The label on the peanut butter jar says peanuts, partially hydrogenated peanut oil, salt and sugar. But we all know there is other stuff in there too. There is very good reason to assume that the number of rat hairs in a 500g jar of peanut butter has a Poisson distribution with mean λ, because it's easy to justify a Poisson process model for how the hairs get into the jars. A sample of 30 jars of Brand A yields $\overline{X} = 6.8$, while an independent sample of 40 jars of Brand B yields $\overline{Y} = 7.275$.

i. State the model for this problem.

ii. What is the parameter space Θ?

iii. State the null hypothesis in symbols.

iv. Find a variance-stabilizing transformation for the Poisson distribution.

v. Using your variance-stabilizing transformation, derive a test statistic that has an approximate standard normal distribution under H_0.

vi. Calculate your test statistic for these data. Do you reject the null hypothesis at $\alpha = 0.05$? Answer Yes or No.

vii. In plain, non-statistical language, what do you conclude? Your answer is something about peanut butter and rat hairs.

2. If the $p \times 1$ random vector $\mathbf{X}$ has variance-covariance matrix $\mathbf{\Sigma}$ 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:

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

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

4. 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.

5. 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) 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.

6. 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.

7. 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.

8. 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.

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

10. Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. Show $Cov(\overline{X}, (X_j - \overline{X})) = 0$ for $j = 1, \ldots, n$. This is the key to showing $\overline{X}$ and S^2 independent, a fact you may use without proof in the next problem.

11. Recall that the chi-squared distribution with ν degrees of freedom is just Gamma with $\alpha = \frac{\nu}{2}$ and $\beta = 2$. So if $X \sim \chi^2(\nu)$, it has moment-generating function $M_X(t) = (1-2t)^{-\nu/2}$.

    (a) 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 $X_2 \sim \chi^2(\nu_2)$.

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

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

    Hint: $\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n \left(X_i - \overline{X} + \overline{X} - \mu\right)^2 = \ldots$
