STA261 (Summer 2022) - Assignment 5

These problems are meant to test your understanding of the concepts in Module 5. They are *not* to be handed in. Some of these have been modified (or in some cases taken directly) from questions in the *Additional Resources* listed in the course syllabus, and no claims of originality are made.

---

1. If $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ and $Y_n = X_n + 1$ for each $n \in \mathbb{N}$, then justify the following:
   (a) $\bar{X}_n \xrightarrow{d} \mu$.
   (b) $S_n^2 \xrightarrow{d} \sigma^2$.
   (c) $\bar{X}_n + S_n^2 \xrightarrow{d} \mu + \sigma^2$.
   (d) $\bar{X}_n / S_n^2 \xrightarrow{p} \mu/\sigma^2$.
   (e) $S_n^2/\bar{X}_n \xrightarrow{p} \sigma^2/\mu$, provided that $\mu \neq 0$.
   (f) $X_n/n \xrightarrow{p} 0$.
   (g) $\bar{Y}_n - 1 \xrightarrow{d} \mu$.
   (h) $\bar{X}_n - \bar{Y}_n \xrightarrow{p} -1$.
   (i) $\sqrt{n}(\bar{X}_n - \mu)\bar{Y}_n \xrightarrow{d} \mathcal{N}\left(0, (\mu+1)^2\sigma^2\right)$.
   (j) $\sqrt{n}(\bar{X}_n - \mu) + \bar{Y}_n \xrightarrow{d} \mathcal{N}\left(\mu+1, \sigma^2\right)$.
   (k) $\sin(\bar{Y}_n) \xrightarrow{d} \sin(\mu+1)$.
   (l) $\bar{X}^2/\overline{X^2} \xrightarrow{p} \frac{\mu^2}{\mu^2+\sigma^2}$.

2. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$. Show that the following are consistent estimators of λ:
   (a) $\bar{X}_n$.
   (b) S_n^2.
   (c) $\alpha\bar{X}_n + (1-\alpha)S_n^2$ for any $\alpha \in [0,1]$.
   (d) $\frac{n}{n+1}\bar{X}_n + \frac{1}{n^2}$.
   (e) $\sum_{i=1}^n 2^{-i}\bar{X}_n$.
   (f) $\sin\left(\frac{1}{n}\right) \cdot \sum_{i=1}^n X_i + e^{-n}$.
   (g) $\frac{6}{\pi^2}\sum_{i=1}^n (S_n/i)^2 + a\mathbb{1}_{n \leq b}$ for any $a, b \in \mathbb{R}$.[1]

3. Show that if T_n is asymptotically efficient for $\tau(\theta)$, then it must be consistent for $\tau(\theta)$.

4. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$, where $p \in (0,1)$, and let $\tau(p) = \log\left(\frac{p}{1-p}\right)$ be the log-odds of p.
   (a) Find the MLE of $\tau(\theta)$, and call it T_n.
   (b) Show that T_n is asymptotically normal, and find its limiting distribution.

---

[1] You can use the fact that $\sum_{i=1}^\infty i^{-2} = \pi^2/6$.

(c) Show that T_n is asymptotically efficient for $\tau(\theta)$.

5. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, where $\mu \in \mathbb{R} \setminus \{0\}$ and $\sigma^2 > 0$. Let $\tau(\mu, \sigma^2) = \sigma/\mu$, which is called the *coefficient of variation*. Let's try and estimate that using $T_n = S_n/\bar{X}_n$.

   (a) Show that
$$\sqrt{n}\left(T_n - \tau(\mu,\sigma^2)\right) = \frac{1}{\bar{X}_n}\left(\sqrt{n}(S_n - \sigma) - \sqrt{n}\left[\frac{\sigma}{\mu}\bar{X}_n - \sigma\right]\right).$$

   (b) Show that $\sqrt{n}\left(\frac{\sigma}{\mu}\bar{X}_n - \sigma\right) \sim \mathcal{N}\left(0, \frac{\sigma^4}{\mu^2}\right)$.

   (c) By writing $\sqrt{n}(S_n - \sigma) = \frac{\sqrt{n}(S_n^2 - \sigma^2)}{S_n + \sigma}$, show that $\sqrt{n}(S_n - \sigma) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2}{2}\right)$ for all $n \in \mathbb{N}$. You'll have to use the fact that for the $\mathcal{N}(\mu, \sigma^2)$ model, $\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} \mathcal{N}(0, 2\sigma^4)$.[2]

   (d) Put the pieces together to show that T_n is asymptotically normal, and find its limiting distribution.

6. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$, where $\lambda > 0$. Find the UMVUE of $\tau_j(\lambda) := \mathbb{P}_\lambda(X_1 = j)$, and show that it's consistent for $j = 0$. (*Hint*: Look at Example 2.30).

7. Under the conditions of Example 5.17, we know that the asymptotic variance of $\tau(\bar{X}_n)$ is given by $(\tau'(\mu))^2\sigma^2/n$. In some applications, it's important to choose $\tau(\cdot)$ so that the asymptotic variance is free of μ. Such a transformation is called a *variance-stabilizing transformation.*

   (a) If $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$, show that $\tau(\lambda) = \sqrt{\lambda}$ is variance-stabilizing.

   (b) If $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$, show that $\tau(p) = \arcsin(\sqrt{p})$ is variance-stabilizing.

   (c) If $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$ where f_θ is supported on a subset of $(0, \infty)$ and $\text{Var}_\theta(X_i)$ is proportional to $\mathbb{E}_\theta[X_i]^2$, show that $\tau(\theta) = \log(\theta)$ is variance-stabilizing. Name at least two familiar distributions which satisfy this property.

8. Given the discussion preceding Definition 5.6, it's tempting to think that $[\tau'(\theta)]^2/I_1(\theta)$ is a kind of asymptotic Cramér-Rao Lower Bound, so that the asymptotic variance of any estimator of $\tau(\theta)$ can't get lower than $[\tau'(\theta)]^2/I_1(\theta)$. Unfortunately, that isn't true! Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\theta, 1)$. Let $|a| < 1$ and define the estimator
$$T_n = \begin{cases} \bar{X}_n, & |\bar{X}_n| \geq n^{-1/4} \\ a\bar{X}_n, & |\bar{X}_n| < n^{-1/4} \end{cases}.$$

   (a) Show that
$$\sqrt{n}(T_n - \theta) = \frac{\bar{X}_n - \theta}{1/\sqrt{n}} \cdot \mathbb{1}_{\left|\frac{\bar{X}_n - \theta + \theta}{1/\sqrt{n}}\right| \geq n^{1/4}} + \left(a\frac{\bar{X}_n - \theta}{1/\sqrt{n}} + \sqrt{n}\theta(a-1)\right) \cdot \mathbb{1}_{\left|\frac{\bar{X}_n - \theta + \theta}{1/\sqrt{n}}\right| < n^{1/4}}.$$

   (b) If $Z \sim \mathcal{N}(0,1)$, explain why
$$\sqrt{n}(T_n - \theta) \overset{d}{=} Z \cdot \mathbb{1}_{|Z + \sqrt{n}\theta| \geq n^{1/4}} + \left(aZ + \sqrt{n}\theta(a-1)\right) \cdot \mathbb{1}_{|Z + \sqrt{n}\theta| < n^{1/4}}.$$

[2] If you *really* like these kinds of manipulations, you can show this yourself. First show that
$$\sqrt{n}(S_n^2 - \sigma^2) = \frac{n}{n-1}\left[\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - \sigma^2\right)\right] + \frac{\sqrt{n}}{n-1}\sigma^2 - \frac{n}{n-1}\sqrt{n}(\bar{X}_n - \mu)^2.$$
For the two terms on the right, use Slutsky to show they both converge in distribution to 0. For the big term in brackets on the left, use the CLT to show that it converges in distribution to $\mathcal{N}\left(0, \text{Var}\left((X_i - \mu)^2\right)\right)$. Then figure out what $\text{Var}\left((X_i - \mu)^2\right) = \mathbb{E}\left[(X_i - \mu)^4\right] - \mathbb{E}\left[(X_i - \mu)^2\right]^2$ is. For the fourth central moment, $\mathbb{E}\left[(X_i - \mu)^4\right]$, using mgfs might be easiest.

(c) Thus, show that

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} \begin{cases} \mathcal{N}(0,1), & \theta \neq 0 \\ \mathcal{N}\left(0, a^2\right), & \theta = 0 \end{cases}.$$

(d) Conclude that when $\theta = 0$, the asymptotic variance of T_n is strictly lower than $[\tau'(\theta)]^2/I_1(\theta)$.

This devastating counterexample was published by Joseph Hodges in 1951, and is known as *Hodges' estimator*. Estimators whose asymptotic variances beat the Cramér-Rao Lower Bound are called *superefficient* estimators. In 1953, Lucien Le Cam proved that superefficiency can only happen when the true parameter θ lives in a subset of Θ of Lebesgue measure zero,[3] so it's usually not a practical concern.

9. Theorem 5.8 fails when $g'(\theta) = 0$, but all hope is not lost. Using the same proof strategy as in Theorem 5.8, establish a second-order delta method for the case that $g'(\theta) = 0$, under suitable conditions. (*Hint*: the limiting distribution involves a $\chi^2_{(1)}$ somewhere).

10. Using Theorem 5.9, sketch a proof of Theorem 5.12.

11. Suppose we are in the setup of Assignment 4 Q14, except this time we want to test $H_0 : p_1 = p_2$ versus $H_A : p_1 \neq p_2$.

(a) Show that a test can be based on the statistic

$$T = \frac{(\hat{p}_1 - \hat{p}_2)^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{p}(1-\hat{p})},$$

where $\hat{p}_1 = S_1/n_1$, $\hat{p}_2 = S_2/n_2$, and $\hat{p} = (S_1 + S_2)/(n_1 + n_2)$, and find its distribution as $n_1, n_2 \to \infty$. (*Hint*: What's the MLE of the common value of $p_1 = p_2$ under H_0?)

(b) An alternative choice of test statistic is

$$T' = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}.$$

Find the distribution of $(T')^2$ as $n_1, n_2 \to \infty$.

(c) We can use these test statistics to test for independence in contingency tables, using the same interpretations of the S_i's and n_i's as in Assignment 4; such a test is called a *chi-squared test for independence*. In the late 1800s, Joseph Lister, a British surgeon and pioneer of antiseptic surgery, conjectured that the use of carbolic acid as a disinfectant would help reduce mortality associated with surgery. He recorded the following data based on 75 amputations over several years:

	Carbolic acid	No carbolic acid
Patient lived	34	19
Patient died	6	16

Use each of the two test statistics to test whether the use of carbolic acid is associated with patient mortality. If you know how, also use a computer (or an internet calculator) to test the same hypothesis using Fisher's exact test.

12. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_\theta$, where $\mathbb{E}_\theta[X_n] = \theta$ and $\text{Var}_\theta(X_n) < \infty$. Show that

$$\left(\bar{X}_n - z_{1-\alpha/2}\sqrt{\frac{S^2}{n}}, \bar{X}_n + z_{1-\alpha/2}\sqrt{\frac{S^2}{n}}\right)$$

[3] Don't worry if you don't know what this means.

is an approximate $(1-\alpha)$-confidence interval for θ.[4]

13. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$ where μ is known and $\sigma^2 > 0$. Derive a Wald statistic and a score statistic for testing $H_0 : \sigma^2 = \sigma_0^2$ versus $H_A : \sigma^2 \neq \sigma_0^2$, and write down the approximate $(1-\alpha)$-confidence regions associated with each, simplifying as much as possible.

14. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \Gamma(\alpha, \beta)$ where α is known and $\beta > 0$. Derive a Wald statistic and a score statistic for testing $H_0 : \beta = \beta_0$ versus $H_A : \beta \neq \beta_0$, and write down the approximate $(1-\alpha)$-confidence regions associated with each, simplifying as much as possible.

15. Suppose $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}\,(p)$, where $p \in (0,1)$. Suppose we observe 15 successes in $n = 32$ trials. Test $H_0 : p = \frac{1}{2}$ versus $H_A : p \neq \frac{1}{2}$ using an asymptotic LRT, a Wald test, and a score test, and compare the results. Then do the same for $H_0 : p = 0.8$ versus $H_A : p \neq 0.8$.

16. Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}\,(p)$, where $p \in (0,1)$. In Example 5.29, we found a nice-looking confidence interval for p based on the Wald statistic. We can also construct one based on the score statistic, although it looks a bit nastier.

    (a) Show that the approximate $(1-\alpha)$-confidence region obtained from the score statistic is
    $$\left\{ p \in (0,1) : \left| \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \right| \leq -z_{\alpha/2} \right\}.$$

    (b) Square both sides of the inequality and rearrange to get
    $$\left\{ p \in (0,1) : \left(1 + \frac{z_{\alpha/2}^2}{n}\right) p^2 - \left(2\hat{p} + \frac{z_{\alpha/2}^2}{n}\right) + \hat{p}^2 \leq 0 \right\}.$$

    (c) This is a quadratic inequality. Since the coefficient on p^2 is positive, the quadratic opens upward, and so the inequality is satisfied if p lies between the two roots of the quadratic. So this region is, in fact, an interval. Find its endpoints to produce our approximate $(1-\alpha)$-confidence interval for p.

    (d) Repeat the same process to find an approximate $(1-\alpha)$-confidence interval based on the score statistic for λ in the Poisson (λ) model.

    It turns out that in many ways, this interval is much better than the Wald interval for p. It has a special name: the *Wilson score interval.* The only reason that most introductory statistics courses teach the Wald interval instead of the score interval is that the former looks less intimidating.

[4] Clearly this gives a very easy way to construct an approximate confidence interval for the mean of any distribution with a finite variance. However, it tends to underperform compared to other approximate intervals that take into account information specific to f_θ.