From Casella and Berger, do the following problems from Chapter 10:

Homework 12: 9, 31, 36, 38, 40, and 47.

These are extra problems that I have given on past exams (in STAT 713 or in related courses). You do not have to turn these in.

10.1. Suppose that $X_1, X_2, ..., X_n$ is an iid sample from

$$f_X(x|\theta,\nu) = \frac{1}{\Gamma(\theta+1)\nu^{(\theta+1)/2}} x^{\theta} e^{-x/\sqrt{\nu}} I(x>0),$$

where $\theta > -1$ and $\nu > 0$. Note that $f_X(x|\theta,\nu)$ is the pdf of $X \sim \text{gamma}(\theta+1, \sqrt{\nu})$.
(a) For this part only, suppose that both θ and ν are unknown. Find the method of moments estimators of θ and ν.
(b) For this part, suppose that θ is known. Show that the maximum likelihood estimator (MLE) of ν is

$$\widehat{\nu} = \left(\frac{\overline{X}}{\theta+1}\right)^2,$$

and derive the large-sample distribution of $\widehat{\nu}$, properly centered and scaled. *Hint:* Recall that, under regularity conditions (which hold here), a maximum likelihood estimator $\widehat{\nu}$ satisfies $\sqrt{n}(\widehat{\nu}-\nu) \xrightarrow{d} \mathcal{N}(0, v(\nu))$, as $n \to \infty$, where $v(\nu) = 1/I_1(\nu)$ and

$$I_1(\nu) = E_\nu\left\{\left[\frac{\partial \log f_X(X|\nu)}{\partial \nu}\right]^2\right\}$$

is the Fisher information based on a single observation.
(c) If θ is known, as in part (b), find the Wald test statistic and approximate level α rejection region for testing $H_0: \nu = \nu_0$ versus $H_1: \nu \neq \nu_0$.

10.2. Consider the density function

$$f_X(x|\theta) = \frac{1}{2}(1+\theta x) I(|x|<1),$$

where $|\theta| < 1$, which describes the decay distribution of electrons from muon decay when $X = \cos(W)$, and W is the angle measured in an experiment. The parameter θ, which is related to polarization, is to be estimated using $X_1, X_2, ..., X_n$, an iid sample from this distribution.
(a) Find the method of moments (MOM) estimator $\widetilde{\theta}$, and also calculate $E_\theta(\widetilde{\theta})$ and $\text{var}_\theta(\widetilde{\theta})$.
(b) Write out the likelihood function and also the score equation that would be solved to find the maximum likelihood estimator (MLE).
(c) Recall that, under certain regularity conditions, maximum likelihood estimators $\widehat{\theta}$ satisfy $\sqrt{n}(\widehat{\theta}-\theta) \xrightarrow{d} \mathcal{N}(0, v(\theta))$, where $v(\theta) = 1/I_1(\theta)$ and

$$I_1(\theta) = E_\theta\left\{\left[\frac{\partial \log f_X(X|\theta)}{\partial \theta}\right]^2\right\}$$

is the Fisher information based on a single observation. Show that

$$\sigma^2_{\widehat{\theta}} = \frac{2\theta^3}{\log\left(\frac{1+\theta}{1-\theta}\right) - 2\theta}.$$

(d) For small θ and large n, compare $\sigma^2_{\widehat{\theta}}$ with $n \times \text{var}_\theta(\widetilde{\theta})$ using the fact that for $|\theta| < 1$,

$$\log\left(\frac{1+\theta}{1-\theta}\right) = 2\left(\theta + \frac{\theta^3}{3} + \frac{\theta^5}{5} + \cdots\right).$$

10.3. Suppose that $X_1, X_2, ..., X_n$ is an iid sample with density $f_X(x|\theta)$, where θ is a scalar parameter. Assume that all the regularity conditions hold for the MLE $\widehat{\theta}$ to be asymptotically normal with

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, I(\theta)^{-1})$$

where $I(\theta)$ is the Fisher information based on one observation. Define $\tau = g(\theta)$, where g is a differentiable, monotone increasing function with inverse h. Show that you get the same asymptotic distribution for $g(\widehat{\theta})$ by using the Delta Method as you would by transforming the problem to one with density $f_Y(y|\tau)$ and appealing to the asymptotic results for maximum likelihood estimators. Recall that, by invariance, $\widehat{\tau} = g(\widehat{\theta})$ is the MLE of τ.

10.4. Suppose that $X_1, X_2, ..., X_n$ is an iid sample from the probability mass function (pmf) given by

$$f_X(x|\theta) = \begin{cases} (1-\theta)\theta^x, & x = 0, 1, 2, ... \\ 0, & \text{otherwise,} \end{cases}$$

where $0 < \theta < 1$. I have calculated $E_\theta(X) = \theta/(1-\theta)$ and $\text{var}_\theta(X) = \theta/(1-\theta)^2$.
(a) Find the maximum likelihood estimator $\widehat{\theta}$.
(b) Find the large-sample distribution of $\widehat{\theta}$, properly centered and scaled. You may assume that the "regularity conditions" hold for this distribution (because they do).
(c) Calculate the Wald statistic Z_n^W to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

10.5. Suppose that $X_1, X_2, ..., X_n$ are iid exponential random variables with mean $\theta > 0$. Derive large-sample size α Wald, score, and likelihood ratio tests for $H_0 : \theta = \theta_0$ versus a two-sided alternative. Do not derive a finite-sample test. Perform a simulation study to examine the size properties of the three tests. How do they compare?

10.6. Suppose that $X_1, X_2, ..., X_n$ is an iid sample from

$$f_X(x|\theta) = \frac{2x}{\theta} e^{-x^2/\theta} I(x > 0).$$

where $\theta > 0$. Derive large-sample $1 - \alpha$ Wald, score, and likelihood ratio confidence intervals. Do not derive a finite-sample interval. Perform a simulation study to examine the coverage properties of the three intervals. How do they compare?

10.7. For a random variable X with $E(X^j) \equiv \mu'_j$, the **cumulant generating function** of X is $\kappa_X(t) = \ln m_X(t)$, where $m_X(t)$ denotes the moment generating function of X. The j**th cumulant** is defined by

$$\kappa_j = \frac{\partial^j}{\partial t^j}\, \kappa_X(t)\bigg|_{t=0}.$$

Suppose that $X_1, X_2, ..., X_n$ is an iid sample, where the assumed model is $\mathcal{N}(\mu, \sigma^2\mu^2)$, and σ^2 is **known**. Let $\widetilde{\mu}$ and $\widehat{\mu}$ denote the MOM and MLE estimators of μ, respectively.
(a) Show that

$$\widetilde{\mu} = \overline{X} \qquad \text{and} \qquad \widehat{\mu} = \frac{\sqrt{\overline{X}^2 + 4\sigma^2 m_2'} - \overline{X}}{2\sigma^2},$$

where, recall, $m_2' = n^{-1}\sum_{i=1}^n X_i^2$ is the second sample (uncentered) moment.
(b) Assuming that the $\mathcal{N}(\mu, \sigma^2\mu^2)$ model is **correct**, prove that both estimators $\widetilde{\mu}$ and $\widehat{\mu}$ are consistent for μ.
(c) Assuming that the $\mathcal{N}(\mu, \sigma^2\mu^2)$ model is **correct**, show that $\sqrt{n}(\widetilde{\mu} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\widetilde{\mu}})$ and $\sqrt{n}(\widehat{\mu} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\widehat{\mu}})$. Find formulae for $\sigma^2_{\widetilde{\mu}}$ and $\sigma^2_{\widehat{\mu}}$ and compute

$$\text{ARE}(\widehat{\mu} \text{ to } \widetilde{\mu}) = \frac{\sigma^2_{\widehat{\mu}}}{\sigma^2_{\widetilde{\mu}}}.$$

Which estimator is more efficient under the $\mathcal{N}(\mu, \sigma^2\mu^2)$ model assumption?
(d) Under the $\mathcal{N}(\mu, \sigma^2\mu^2)$ model, show that $\kappa_1 = \mu$, $\kappa_2 = \sigma^2\mu^2$, and that $\kappa_3 = \kappa_4 = 0$.
(e) Now, let's suppose that we have our two estimators $\widetilde{\mu}$ and $\widehat{\mu}$ (as stated above), but we allow for departures from normality. That is, let's assume that $X_1, X_2, ..., X_n$ is an iid sample with $E(X_1^j) \equiv \mu_j' < \infty$, for $j = 1, 2, 3, 4$. Results for the remaining parts will be stated in terms of the first four cumulants κ_j, $j = 1, 2, 3, 4$; that is, **we now examine the properties of $\widetilde{\mu}$ and $\widehat{\mu}$ when $\kappa_1 = \mu$, but when κ_2, κ_3, and κ_4 are arbitrary;** i.e., κ_2 is no longer necessarily equal to $\sigma^2\mu^2$, and κ_3 and κ_4 are not necessarily zero. First, show that, in general,

$$\begin{aligned}
\kappa_1 &= \mu_1' \quad \text{(mean)} \\
\kappa_2 &= \mu_2' - \kappa_1^2 \quad \text{(variance)} \\
\kappa_3 &= \mu_3' - 3\kappa_1\kappa_2 - \kappa_1^3 \\
\kappa_4 &= \mu_4' - 4\kappa_1\kappa_3 - 3\kappa_2^2 - 6\kappa_1^2\kappa_2 - \kappa_1^4.
\end{aligned}$$

In part (e), note that κ_1 is the mean and that κ_2 is the variance (again, κ_2 is not necessarily equal to $\sigma^2\mu^2$ henceforth). **To reiterate, in the parts that follow, we are not assuming normality.** The value σ^2 continues to be a known constant (it is present in the MLE estimator).
(f) Show that, under our new (relaxed) model assumptions, $\sqrt{n}(\widetilde{\mu} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\widetilde{\mu}})$, and give a formula for $\sigma^2_{\widetilde{\mu}}$ in terms of the cumulants. Of course, this implies that the MOM estimator is consistent for μ regardless of whether or not we assume normality and regardless of whether or not we have correctly specified the correct variance function.
(g) Show that, under our new (relaxed) model assumptions,

$$\widehat{\mu} \xrightarrow{p} \frac{\sqrt{\mu^2 + 4\sigma^2\mu^2 + 4\sigma^2\kappa_2} - \mu}{2\sigma^2} = \mu^*, \text{ say.}$$

That is, under these relaxed model assumptions, the MLE isn't even consistent.
(h) Show that $\sqrt{n}(\widehat{\mu} - \mu^*) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\widehat{\mu}})$, and give a formula for $\sigma^2_{\widehat{\mu}}$ in terms of μ, σ^2 (a known constant), κ_2, κ_3, and κ_4. This will be messy; you might use MAPLE to help with the algebra.
(i) Show that the MLE is consistent if we correctly specify that $\kappa_2 = \sigma^2\mu^2$, even in the absence of normality; i.e., just substitute $\kappa_2 = \sigma^2\mu^2$ into the probability limit in part (g). This implies that the MLE **will be consistent** if we specify the correct variance structure (consistency

doesn't have anything to do with whether or not we assume normality).
(j) Using your values of $\sigma^2_{\widetilde{\mu}}$ and $\sigma^2_{\widehat{\mu}}$ in parts (f) and (h), respectively, show that when $\kappa_1 = \mu$ and $\kappa_2 = \sigma^2\mu^2$; *that is, we have correctly specified the variance function*, the asymptotic relative efficiency

$$\text{ARE}(\widehat{\mu} \text{ to } \widetilde{\mu}) = \frac{\sigma^2_{\widehat{\mu}}}{\sigma^2_{\widetilde{\mu}}} = \frac{1 + 2\sigma^2 + 2\sigma\gamma_3 + \sigma^2\gamma_4}{(1 + 2\sigma^2)^2},$$

where $\gamma_3 = \kappa_3/\sigma^3\mu^3$ and $\gamma_4 = \kappa_4/\sigma^4\mu^4$.
(k) Summarize your findings; i.e., describe how the MLE and MOM can compare in the settings we have examined. In particular, under what assumptions does the MLE outperform the MOM? Under what assumptions can the MOM outperform the MLE?

10.8. Suppose that $X_1, X_2, ..., X_n$ is an iid sample, each with probability p of being distributed as uniform over $(-1/2, 1/2)$ and with probability $1 - p$ of being distributed as uniform over $(0, 1)$.
(a) Find the maximum likelihood estimator (MLE) of p and determine its asymptotic distribution.
(b) Find another estimator of p using the method of moments (MOM). Determine its asymptotic distribution.
(c) Which of the two estimators (MLE or MOM) is more efficient? Prove your answer.