Note: Please see the Midterm Practice Problems for additional problems explicitly from Chapters 6 and 7 (WMS). A majority of these problems (clearly not all) are from Chapters 8 and 9 (WMS).

1. Suppose that $Y_1, Y_2, ..., Y_{36}$ is an iid sample of size $n = 36$ from a geometric distribution with $p = 0.75$.
(a) Derive the moment generating function of

$$T = Y_1 + Y_2 + \cdots + Y_{36}.$$

Does T have a distribution that you recognize? If so, which one?
(b) Use the Central Limit Theorem to approximate $P(1.25 < \overline{Y} < 1.50)$.

2. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of Poisson observations with mean $\theta > 0$.
(a) Prove that $\overline{Y}$ is the maximum likelihood estimator (MLE) of θ and find its variance.
(b) We know that $\overline{Y}$ is an unbiased estimator of θ, but so is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

Explain or prove why.
(c) It turns out that

$$V(S^2) = \frac{\theta(2n\theta + n - 1)}{n(n-1)}.$$

You do not have to prove this fact. Which estimator has smaller variance, $\overline{Y}$ or S^2? Are you surprised? Why or why not?
(d) Use the Rao-Blackwell Theorem to establish the surprising identity $E(S^2|\overline{Y}) = \overline{Y}$.

3. A continuous random variable Y is said to have a (standard) logistic distribution if its probability density function (pdf) is given by

$$f_Y(y) = \begin{cases} e^{-y}(1+e^{-y})^{-2}, & -\infty < y < \infty \\ 0, & \text{otherwise.} \end{cases}$$

(a) Prove that

$$U = g(Y) = \frac{1}{1+e^{-Y}}$$

has a $\mathcal{U}(0,1)$ distribution. *Hint:* Use the transformation technique or the distribution function technique.
(b) Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from $f_Y(y)$. Find the pdf of $Y_{(n)}$.

4. An engineering component has a lifetime Y which follows a shifted exponential distribution, in particular, the probability density function (pdf) of Y is

$$f_Y(y;\theta) = \begin{cases} e^{-(y-\theta)}, & y > \theta \\ 0, & \text{otherwise.} \end{cases}$$

The unknown parameter $\theta > 0$ measures the magnitude of the shift. From an iid sample of component lifetimes $Y_1, Y_2, ..., Y_n$, we would like to estimate θ.
(a) Show that the pdf of $Y_{(1)}$, the minimum order statistic, is

$$f_{Y_{(1)}}(y;\theta) = \begin{cases} ne^{-n(y-\theta)}, & y > \theta \\ 0, & \text{otherwise.} \end{cases}$$

(b) It follows that $Y_{(1)}$ is a (complete) sufficient statistic for θ (you do not need to prove this). Find the minimum variance unbiased estimator (MVUE) of θ.

5. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from a $\mathcal{N}(0, \sigma^2)$ distribution where $\sigma^2 > 0$ is unknown.
(a) Argue that

$$Q = \sum_{i=1}^{n} \left(\frac{Y_i}{\sigma}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n} Y_i^2$$

is a pivotal quantity and use Q to derive a $100(1-\alpha)$ percent confidence interval for σ^2. Define any notation that you use.
(b) I used R to generate an iid sample of normal zero-mean data. Use the data to compute a 95 percent confidence interval for σ^2. Use your interval from part (a).

4.59	−4.14	3.02	9.16	0.24
0.32	4.07	−0.51	−6.50	−13.16

(c) If, instead, you wanted a 95 percent confidence interval for the population standard deviation, σ, how could you use the interval you derived in part (a)? Explain and report your interval using the data in part (b).

6. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from a beta distribution with $\alpha = 1$ and $\beta = \theta$ so that

$$f_Y(y;\theta) = \begin{cases} \theta(1-y)^{\theta-1}, & 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the method of moments estimator (MOM) of θ.
(b) Find the maximum likelihood estimator (MLE) of θ.
(c) Find the MLE of the variance of the beta$(1, \theta)$ population distribution.

7. Consider a toxicology study with k groups of animals who are given a drug at distinct dose levels $d_1, d_2, ..., d_k$, respectively (these are fixed by the experimenter; not random). The animals are monitored for a reaction to the drug. In group i, let n_i (fixed) denote the total number of animals dosed, and let Y_i denote the number of animals that respond to the drug. The observations $Y_1, Y_2, ..., Y_k$ are treated as independent random variables,

where $Y_i \sim b(n_i, p_i)$; $i = 1, 2, ..., k$, where p_i is the probability that an individual animal responds to dose d_i. Recall that the pmf of Y_i is given by

$$f_{Y_i}(y_i; p_i) = \binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i - y_i},$$

for $y_i = 0, 1, ..., n_i$. A standard assumption in such toxicology studies is that

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 d_i,$$

for $i = 1, 2, ..., k$, where β_0 and β_1 are real parameters (this is merely logistic regression using dose as a predictor).

(a) Use algebra to show that p_i, when written as a function of β_0, β_1, and d_i, satisfies

$$p_i = \frac{\exp(\beta_0 + \beta_1 d_i)}{1 + \exp(\beta_0 + \beta_1 d_i)}.$$

(b) Find a two-dimensional sufficient statistic for (β_0, β_1). *Hint:* First, substitute the expression for p_i from part (a) into $f_{Y_i}(y_i; p_i)$ and write the likelihood function in terms of β_0 and β_1. Recall that $Y_1, Y_2, ..., Y_k$ are independent, so

$$L(\beta_0, \beta_1 | y_1, y_2, ..., y_k) = \prod_{i=1}^{k} f_{Y_i}(y_i; p_i).$$

Now, simplify using algebraic properties of exponential functions and use the Factorization Theorem for the multiple parameter case.

(c) After the data are collected, suppose that a 95 percent confidence interval for β_1 is computed to be $(-0.84, 1.31)$. What can you say about the relationship between between dosing and response to the drug?

8. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of size n from an exponential distribution with mean β.

(a) Is $\overline{Y}$ a sufficient statistic for β? If so, prove it. If not, explain/prove why not.

(b) Derive (using mgfs) the distribution of $\overline{Y}$.

(c) Find a function of $\overline{Y}$ that converges in distribution to a standard normal distribution as $n \to \infty$. Explain why your answer is correct.

(d) Find two consistent estimators of $\sigma^2 = \beta^2$. Explain why your answers are correct.

9. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from $f_Y(y; \theta)$, where

$$f_Y(y;\theta) = \begin{cases} 2y/\theta^2, & 0 < y < \theta \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show that $\widehat{\theta} = 3\overline{Y}/2$ is an unbiased estimator of θ.

(b) Find $\text{MSE}(\widehat{\theta})$. Graph this MSE function for values of $\theta > 0$ and $n = 10$.

10. Suppose that Y is a discrete random variable that can take on values 0, 1, and 2, according to the following distribution:

y	0	1	2
$p_Y(y)$	p	$3p$	$1-4p$

Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from this distribution. Here, $0 < p < 1/4$ is unknown.
(a) Show that

$$\widehat{p} = \frac{2 - \overline{Y}}{5}$$

is an unbiased estimator of p.
(b) Show that

$$V(\widehat{p}) = \frac{p(7 - 25p)}{25n}.$$

(c) For this part, suppose that $n = 100$ and $\overline{y} = 1.0$. Find an estimated 2 standard error bound for the error in estimation $\epsilon = |\widehat{p} - p|$. Your final answer should be numerical.

11. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from a $\mathcal{N}(\mu, \sigma^2)$ population, where both μ and σ^2 are unknown parameters. To estimate σ^2, we will use an estimator of the form $\widehat{\sigma}^2 = cS^2$, where S^2 is the usual sample variance; that is,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2,$$

and $c > 0$ is a strictly positive constant. Prove that the value of c that minimizes $\text{MSE}(\widehat{\sigma}^2) = \text{MSE}(cS^2)$ is given by

$$c = \frac{n-1}{n+1}.$$

12. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid Bernoulli(p) sample, where $0 < p < 1$, and let $\widehat{p}$ denote the usual sample proportion. In class, we have shown that

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile from the standard normal distribution, is an approximate $100(1-\alpha)$ percent confidence interval for p. The inconsistencies associated with this interval are well documented, so we will find an alternative interval for p. The construction of the alternative interval is based on the fact that, when n is large,

$$h(\widehat{p}) \equiv \arcsin(\widehat{p}^{1/2}) \ \sim \mathcal{AN}\left[\arcsin(p^{1/2}), \frac{1}{4n}\right],$$

(a) Use this fact to construct a large-sample $100(1-\alpha)$ percent confidence interval for

$$h(p) \equiv \arcsin(p^{1/2}).$$

Define any notation you use. *Hint:* Consider

$$Q = \frac{h(\widehat{p}) - h(p)}{\sqrt{1/4n}}.$$

(b) Transform the endpoints of the interval in (a) to produce a large-sample $100(1-\alpha)$ percent confidence interval for the parameter p. You do this by finding the inverse function $h^{-1}(p)$ and applying this inverse rule to the endpoints of the $h(p)$ interval. Recall that $\arcsin(\cdot)$ is another symbol for $\sin^{-1}(\cdot)$.

13. It is common in biological applications to assume a $\mathcal{N}(\theta, \theta^2)$ model for a continuous measurement Y. Note that in this model, the variance is a function of the mean and increases as θ increases. Suppose that we observe an iid sample of measurements of Y, say, $Y_1, Y_2, ..., Y_n$ (e.g., wildlife growth rates, plant hormone concentrations, etc.) and suppose that our goal is to estimate θ. Define the two estimators

$$\widehat{\theta}_1 = \overline{Y} \qquad \text{and} \qquad \widehat{\theta}_2 = cS,$$

where $\overline{Y}$ and S denote the sample mean and sample standard deviation, respectively, and

$$c = \frac{\sqrt{n-1}\Gamma[(n-1)/2]}{\sqrt{2}\Gamma(n/2)}.$$

Both are unbiased estimators of θ.
(a) Prove that any convex combination of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ is also unbiased. That is, for any $a \in (0, 1)$, show that

$$\widehat{\theta} = a\widehat{\theta}_1 + (1-a)\widehat{\theta}_2$$

is an unbiased of estimator of θ.
(b) Find the value of a that minimizes $\text{MSE}(\widehat{\theta})$.

14. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of size n from a Poisson distribution with mean $\lambda > 0$. In class, we proved that $\overline{Y}$ is the maximum likelihood estimator of λ.
(a) Find the maximum likelihood estimator of $P(Y_1 \leq 1) = g(\lambda) = e^{-\lambda} + \lambda e^{-\lambda}$.
(b) Find a consistent estimator of λ.
(c) Find a consistent estimator of $g(\lambda)$.
(d) Use the Central Limit Theorem and Slutsky's Theorem to prove that, as $n \to \infty$,

$$Z_n = \frac{\overline{Y} - \lambda}{\sqrt{\frac{S^2}{n}}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Note: In parts (a-c), provide sufficient justification as to why your answers are correct.

15. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from a shifted-exponential distribution with probability density function (pdf)

$$f_Y(y) = \begin{cases} e^{-(y-\theta)}, & y > \theta \\ 0, & \text{otherwise.} \end{cases}$$

Derive the pdf of the minimum order statistic $Y_{(1)}$.

16. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from a beta distribution with parameters $\alpha = \theta$ and $\beta = 1$, so that the common pdf is

$$f_Y(y) = \begin{cases} \theta y^{\theta-1}, & 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show that

$$T = -\sum_{i=1}^{n} \ln Y_i$$

is a sufficient statistic for θ.
(b) Find the distribution of T. *Hint:* First, use a transformation or distribution technique to show that $W = -\ln Y$ has an exponential distribution with mean $1/\theta$. Then, T is the sum of iid exponentials. Note that you can do this part even if you couldn't do part (a).

17. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from a $\mathcal{N}(\mu, \sigma^2)$ distribution. Define

$$Q \equiv Q(\mu, \sigma^2) = \left(\frac{\overline{Y} - \mu}{\sigma/\sqrt{n}}\right)^2 + \frac{(n-1)S^2}{\sigma^2},$$

where $\overline{Y}$ and S^2 denote the usual sample mean and sample variance, respectively.
(a) Carefully argue that $Q \sim \chi^2(n)$, thus, demonstrating that Q is a pivot.
(b) If $\mu = 0$, use the result from part (a) to show that

$$\left(\frac{n\overline{Y}^2 + (n-1)S^2}{\chi^2_{n,\alpha/2}}, \frac{n\overline{Y}^2 + (n-1)S^2}{\chi^2_{n,1-\alpha/2}}\right)$$

is an exact $100(1-\alpha)$ percent confidence interval for σ^2. The symbols $\chi^2_{n,\alpha/2}$ and $\chi^2_{n,1-\alpha/2}$ denotes the upper and lower $\alpha/2$ quantiles, respectively, of the χ^2 distribution with n degrees of freedom. Note that you can do this part even if you couldn't do part (a).

18. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from $f_Y(y; \theta)$, a probability density function (pdf), where

$$f_Y(y; \theta) = \begin{cases} \theta e^{-\theta y}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show that

$$W = \sum_{i=1}^{n} Y_i$$

is a sufficient statistic for θ.
(b) Derive the moment generating function (mgf) of $U = \theta W$. What is the distribution of U?

19. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of size n from a population with pdf

$$f_Y(y;\theta) = \begin{cases} \frac{1}{\theta}e^{-(y-1)/\theta}, & y > 1 \\ 0, & \text{otherwise.} \end{cases}$$

where θ is a positive unknown parameter.
(a) Show that $\widehat{\theta} = \overline{Y} - 1$ is the minimum-variance unbiased estimator (MVUE) for θ. *Hint:* First show that

$$U = \sum_{i=1}^{n} Y_i.$$

is a sufficient statistic.
(b) Find the MVUE for $\tau(\theta) = \theta^2$. *Hint:* First try $E(\widehat{\theta}^2)$.

20. The Weibull distribution is a popular probability model in engineering applications. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of size n from a Weibull distribution with pdf

$$f_Y(y;\theta) = \begin{cases} \alpha\theta y^{\alpha-1}e^{-\theta y^{\alpha}}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

where α is known and $\theta > 0$ is unknown; that is, θ is the only unknown parameter.
(a) Find a formula for the method of moments estimator for θ.
(b) Find a formula for the the maximum likelihood estimator for θ.

21. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from a gamma(α, β) distribution, where both α and β are unknown, and let

$$U = \sum_{i=1}^{n} Y_i.$$

(a) Is U a sufficient statistic for $\boldsymbol{\theta} = (\alpha, \beta)$? If so, prove it. If not, explain/prove why not.
(b) Derive (using mgfs) the distribution of U.
(c) Find a function of U that converges in distribution to a standard normal distribution as $n \to \infty$. Explain why your answer is correct.
(d) Find a function of U that converges in probability to $\mu = \alpha\beta$ as $n \to \infty$. Explain why your answer is correct.

22. Suppose that Y is a continuous random variable with probability density function $f_Y(y)$ and cumulative distribution function $F_Y(y)$.
(a) Use the cdf technique to prove that the pdf of $U = Y^2$, where nonzero, is given by

$$f_U(u) = \frac{1}{2\sqrt{u}}\left[f_Y(\sqrt{u}) + f_Y(-\sqrt{u})\right].$$

(b) Use the result in part (a) to find the pdf of $U = Y^2$, where $Y \sim \mathcal{U}(-1, 1)$. Make sure to state the support of U.

23. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from an exponential distribution with mean θ. Recall that the exponential(θ) probability density function (pdf) is given by

$$f_Y(y) = \begin{cases} \frac{1}{\theta}e^{-y/\theta}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

and that the cumulative distribution function (cdf) is given by

$$F_Y(y) = \begin{cases} 0, & y \leq 0 \\ 1 - e^{-y/\theta}, & y > 0. \end{cases}$$

Consider the two estimators

$$\widehat{\theta}_1 = nY_{(1)} \qquad \widehat{\theta}_2 = \overline{Y},$$

where $Y_{(1)}$ is the minimum of $Y_1, Y_2, ..., Y_n$, and $\overline{Y}$ is the sample mean. Recall that the pdf of the minimum order statistic $Y_{(1)}$, where positive, is given by

$$f_{Y_{(1)}}(y) = nf_Y(y)[1 - F_Y(y)]^{n-1}.$$

(a) Show that $Y_{(1)}$ follows an exponential distribution with mean θ/n.
(b) Show that both $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are unbiased estimators of θ.
(c) Compute the variance of both estimators. Which estimator would you prefer?

24. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from a Poisson distribution with mean θ.
(a) Give two unbiased estimators of θ.
(b) Find an unbiased estimator of $\tau(\theta) = \theta^2$.

25. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of size n from the pdf

$$f_Y(y;\theta) = \begin{cases} \frac{2y}{\theta}e^{-y^2/\theta}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find a sufficient statistic for θ.
(b) Let U denote your sufficient statistic from part (a), and suppose that $\widehat{\theta}$ is any unbiased estimator of θ. Let

$$\widehat{\theta}^* = E(\widehat{\theta}|U).$$

Show that $\widehat{\theta}^*$ is unbiased and that $V(\widehat{\theta}^*) \leq V(\widehat{\theta})$. Essentially, I am asking you to prove the Rao-Blackwell Theorem here. *Hint:* Recall the iterated formulas for mean and variance from STAT 511. We also did this proof in class.
(c) What is the main implication of the result in part (b)? Talk about MVUE's and sufficiency. Note that you can answer this part even if you could not do part (b).

26. Suppose that we have two independent samples:

$$\begin{aligned} Y_{11}, Y_{12}, ..., Y_{1n} &\sim \text{iid } \mathcal{N}(\mu_1, \sigma^2) \\ Y_{21}, Y_{22}, ..., Y_{2n} &\sim \text{iid } \mathcal{N}(\mu_2, \sigma^2). \end{aligned}$$

Notice that the two population variances are the same and that the sample sizes are equal. The population means μ_1 and μ_2 are unknown parameters. The common population variance σ^2 is also an unknown parameter. Define

$$\begin{aligned} \overline{Y}_{1+} = \frac{1}{n}\sum_{j=1}^{n} Y_{1j} &= \text{sample mean for sample 1} \\ \overline{Y}_{2+} = \frac{1}{n}\sum_{j=1}^{n} Y_{2j} &= \text{sample mean for sample 2} \\ S_1^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_{1j} - \overline{Y}_{1+})^2 &= \text{sample variance for sample 1} \\ S_2^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_{2j} - \overline{Y}_{2+})^2 &= \text{sample variance for sample 2.} \end{aligned}$$

(a) Find the distribution of $U = \overline{Y}_{1+} - \overline{Y}_{2+}$.
(b) Find the distribution of

$$W = \frac{(n-1)S_1^2}{\sigma^2} + \frac{(n-1)S_2^2}{\sigma^2} = \frac{(n-1)S_1^2 + (n-1)S_2^2}{\sigma^2}.$$

(c) Explain why U and W are independent.
(d) If $\mu_1 - \mu_2 = 0$, find a statistic, which is a function of U and W, that has an F distribution. What are its degrees of freedom? Remember that a statistic can not depend on unknown parameters.

27. In an environmental study, Y denotes the proportion of particulate matter that is deemed hazardous. Empirical evidence from data collected by the EPA reveals that the probability density function (pdf) of Y is

$$f_Y(y) = \begin{cases} 4(1-y)^3, & 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

A total of $n = 36$ particulate matter measurements will be taken, yielding the sample $Y_1, Y_2, ..., Y_{36}$. Treating these measurements as an iid sample from $f_Y(y)$, use the Central

Limit Theorem to approximate $P(\overline{Y} < 0.175)$.

28. In an early-phase clinical trial, two drugs are administered to patients with hypertension. On each patient, we measure a point reduction of hypertension, which from past studies is known to follow a normal distribution. We have two independent samples

$$Y_{11}, Y_{12}, ..., Y_{1n_1} \text{ iid } \mathcal{N}(\mu_1, \sigma^2)$$
$$Y_{21}, Y_{22}, ..., Y_{2n_2} \text{ iid } \mathcal{N}(\mu_2, \sigma^2).$$

Physicians assume a common variance between the two drug populations. Here are some summary statistics from the trial (sample sizes, sample means, and sample standard deviations):

Drug 1	Drug 2
$n_1 = 11$	$n_2 = 13$
$\overline{y}_{1+} = 11.3$	$\overline{y}_{2+} = 14.6$
$s_1 = 6.3$	$s_2 = 7.7$

(a) Using this information, compute a 95 percent confidence interval for $\mu_1 - \mu_2$, the difference in the mean point reduction for the two drugs (under the common variance assumption). What does this confidence interval tell you about the two drugs?
(b) Suppose that a physician associated with the trial wanted to determine whether the two population variances were truly equal. Suggest a strategy on how she could accomplish this, and then execute your strategy with the information above.

29. The Pareto distribution is often used in economics to model income distributions. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of size n from a Pareto pdf of the form

$$f_Y(y;\theta) = \begin{cases} \theta\nu^\theta y^{-(\theta+1)}, & y > \nu \\ 0, & \text{otherwise,} \end{cases}$$

where $\theta > 1$. In this model, ν represents the minimum income level (in \$1,000's). In this problem, we will assume that $\nu = 65$; i.e., ν is a known value. The parameter θ is unknown.
(a) Find a formula for the method of moments estimator for θ.
(b) Find a formula for the the maximum likelihood estimator for θ.
(c) Compute your MOM and MLE estimates of θ with the following data:

```
87.10  68.55  123.48  77.77  110.88  100.54  98.78  84.32
```

30. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid Bernoulli(θ) sample; recall that the Bernoulli(θ) pmf is given by

$$f_Y(y;\theta) = \begin{cases} \theta^y(1-\theta)^{1-y}, & y = 0, 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the MVUE of θ.
(b) Find the MVUE of $\tau(\theta) = \theta^2$.

31. Prove any two of the following results.

- If $Y \sim \text{beta}(\theta, \theta)$, then $U = 1 - Y \sim \text{beta}(\theta, \theta)$.
- If $Y_1, Y_2, ..., Y_n$ are iid exponential with mean $\beta > 0$, then $Y_{(1)} \sim \text{exponential}(\beta/n)$.
- If $Y_1, Y_2, ..., Y_n$ are independent random variables, where $Y_i \sim \text{gamma}(\alpha_i, \beta)$, then $U = \sum_i Y_i \sim \text{gamma}(\sum_i \alpha_i, \beta)$.
- If $Y_1 \sim \text{gamma}(\alpha, 1)$, $Y_2 \sim \text{gamma}(\beta, 1)$, and Y_1 and Y_2 independent, then $U = Y_1/(Y_1 + Y_2) \sim \text{beta}(\alpha, \beta)$.

32. In an agricultural experiment, researchers are interested in comparing three different fertilizers. The researchers decide to model their yields as follows:

$$\begin{array}{ll} \text{Fertilizer 1:} & Y_{11}, Y_{12}, ..., Y_{1n_1} \sim \text{iid } \mathcal{N}(\mu_1, \sigma^2) \\ \text{Fertilizer 2:} & Y_{21}, Y_{22}, ..., Y_{2n_2} \sim \text{iid } \mathcal{N}(\mu_2, \sigma^2) \\ \text{Fertilizer 3:} & Y_{31}, Y_{32}, ..., Y_{3n_3} \sim \text{iid } \mathcal{N}(\mu_3, \sigma^2). \end{array}$$

The population means μ_1, μ_2, and μ_3 are unknown parameters. The common population variance σ^2 is also unknown. Fertilizers were randomly assigned to plots, so they assume that the samples are independent. The researchers are interested in the linear combination

$$\theta = a_1\mu_1 + a_2\mu_2 + a_3\mu_3,$$

where a_1, a_2, and a_3 are known constants. As an estimator for θ, they use

$$\widehat{\theta} = a_1\overline{Y}_{1+} + a_2\overline{Y}_{2+} + a_3\overline{Y}_{3+},$$

where $\overline{Y}_{i+}$ denotes the ith sample mean; $i = 1, 2, 3$.
(a) Argue that

$$\widehat{\theta} \sim \mathcal{N}\left(\theta, \sigma^2\left[\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}\right]\right).$$

(b) Define

$$S_{p:3}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + (n_3 - 1)S_3^2}{n_1 + n_2 + n_3 - 3}.$$

Show that $(n_1 + n_2 + n_3 - 3)S_{p:3}^2/\sigma^2$ has a χ^2 distribution with $n_1 + n_2 + n_3 - 3$ degrees of freedom.
(c) Show that

$$t = \frac{\widehat{\theta} - \theta}{S_{p:3}\sqrt{\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}}}$$

has a t distribution with $n_1 + n_2 + n_3 - 3$ degrees of freedom.
(d) Use the result in (c) to derive a $100(1-\alpha)$ percent confidence interval for θ.

33. Suppose that we have two independent samples:

$$\begin{aligned} \text{Sample 1}: &\quad Y_{11}, Y_{12}, ..., Y_{1n_1} \sim \text{iid } \mathcal{N}(\mu_1, \sigma^2) \\ \text{Sample 2}: &\quad Y_{21}, Y_{22}, ..., Y_{2n_2} \sim \text{iid } \mathcal{N}(\mu_2, \sigma^2) \end{aligned}$$

All parameters are unknown, but note that the variance in each population distribution is the same. Let S_1^2 and S_2^2 denote the sample variances from Sample 1 and Sample 2, respectively. Recall that the pooled variance estimator is

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

(a) Prove that

$$Q = \frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2}$$

is a pivotal quantity and give its distribution.
(b) Use the pivot in part (a) to derive a $100(1-\alpha)$ percent confidence interval for σ^2. Define any notation you use (e.g., notation for quantiles, etc.).

34. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of size n from a Weibull distribution with pdf

$$f_Y(y;\beta) = \begin{cases} \beta^{-1}\alpha_0 y^{\alpha_0 - 1}\exp(-y^{\alpha_0}/\beta), & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

where $\alpha_0 > 0$ is known and $\beta > 0$ is unknown.
(a) Prove that $Y_i^{\alpha_0} \sim \text{exponential}(\beta)$. Then, argue that $Q = 2T/\beta \sim \chi^2(2n)$, where

$$T = \sum_{i=1}^{n} Y_i^{\alpha_0}.$$

(b) Use the result in part (a) to derive a $100(1-\alpha)$ percent confidence interval for β. Draw pictures to help me understand your reasoning. Define all notation.
(c) The following data are $n = 20$ death times (in days) for rats injected with a toxin as part of an early phase clinical trial:

4.76	10.57	8.23	6.20	9.83
7.75	3.46	7.05	6.65	0.75
5.30	7.77	1.77	3.55	3.94
1.59	5.61	3.26	8.33	5.23

Under the iid Weibull model assumption with $\alpha_0 = 2$, use the data to find a 90 percent confidence interval for β.

35. Highly active antiretroviral therapy (HAART) refers to the administration of aggressive treatment regimens used to suppress HIV viral replication and to delay the onset of AIDS. HAART is a combination therapy that restores CD4+ T cell numbers in HIV-infected patients. While HAART dramatically improves the prognosis of HIV-infected patients, there has been a recent concern that taking HAART may increase the risk of cardiovascular disease. To examine this question, suppose that a clinical trial is to be performed with n_1 patients receiving a HAART regimen with a cardiovascular disease drug supplement (Enalapril) and n_2 receiving HAART but no Enalapril. Assume that patients are randomly assigned to the treatment groups:

- Group 1: HAART with Enalapril
- Group 2: HAART with no Eanlapril.

Patients will then be monitored for the development of cardiovascular disease (CVD). Define

$$\begin{aligned} Y_1 &= \text{the number of Group 1 patients who develop CVD} \\ Y_2 &= \text{the number of Group 2 patients who develop CVD.} \end{aligned}$$

Assume that $Y_1 \sim b(n_1, p_1)$, $Y_2 \sim b(n_2, p_2)$ and that Y_1 and Y_2 are independent. The investigators are interested in comparing the binomial probabilities p_1 and p_2.
(a) Prove that

$$\widehat{\theta} = \frac{Y_1}{n_1} - \frac{Y_2}{n_2}$$

is an unbiased estimator of $\theta = p_1 - p_2$.
(b) Prove that the variance of $\widehat{\theta}$ is given by

$$V(\widehat{\theta}) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}.$$

(c) The Central Limit Theorem guarantees that $\widehat{\theta}$ is approximately normally distributed when n_1 and n_2 are both "large." If $n_1 = 100$, $n_2 = 100$, $y_1 = 30$, and $y_2 = 20$, compute a 95 percent confidence interval for θ. Write an English sentence that describes the meaning of your interval.

36. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from a Rayleigh distribution with probability density function (pdf)

$$f_Y(y;\theta) = \begin{cases} \frac{2y}{\theta} e^{-y^2/\theta}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the method of moments (MOM) estimator of θ.
(b) Find the maximum likelihood estimator (MLE) of θ.