This midterm covers Chapters 6 and 7 in WMS (and the notes). The following problems are stratified by chapter.

Chapter 6 Problems

1. Suppose that $Y \sim \mathcal{U}(0,2)$ so that the probability density function (pdf) of Y is

$$f_Y(y) = \begin{cases} 1/2, & 0 < y < 2 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the pdf of $U = g(Y) = Y^4 + 1$ and compute $E(U)$.
(b) Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from $f_Y(y)$. Derive the pdf of $Y_{(n)}$, the maximum order statistic, and find $E(Y_{(n)})$.

2. Actuarial data are used to posit a probability model for Y, the amount of an insurance claim (in \$1000s) for a certain class of customers. Suppose that Y follows an exponential distribution with mean 5. What is the probability that the minimum of 4 independent claims Y_1, Y_2, Y_3, Y_4 exceeds \$2,000?

3. Suppose that Y_1 has a gamma distribution with parameters $\alpha = 2$ and $\beta = 1$. Suppose that Y_2 has an exponential distribution with mean 1. Suppose that Y_1 and Y_2 are independent. Derive the probability density function of

$$U = \frac{Y_1}{Y_2}.$$

Hint: Use either the distribution function technique or a bivariate transformation technique. Don't use mgfs.

4. In a medium-sized city, there are two hospitals. For a certain cohort of patients, assume that the length of stay in Hospital 1 is normally distributed with mean 4.6 days and standard deviation 0.9 days. Assume that the length of stay in Hospital 2 is normally distributed with mean 4.9 days and standard deviation 1.2 days. Assume additionally that the length of stay in Hospital 1 is independent of the length of stay in Hospital 2.
(a) Find the probability that a patient's stay at Hospital 1 would be longer that his/her stay at Hospital 2.
(b) Suppose that independent, iid samples of $n_1 = n_2 = 4$ patients will be taken from each hospital (that is, within each hospital, the 4 patient lengths of stay are iid, and the two samples from different hospitals are independent). Find the probability that the average stay at Hospital 1 would be longer than the average stay at Hospital 2.

5. You should know the following result:

$$Y \sim \mathcal{N}(0,1) \Longrightarrow U = Y^2 \sim \chi^2(1).$$

We have proven this result using the (cumulative) distribution function technique and the method of transformations.
(a) Prove this result using the method of moment generating functions.
(b) If $Y_1, Y_2, ..., Y_n$ is an iid $\mathcal{N}(0, 1)$ sample, what is the distribution of

$$T = \sum_{i=1}^{n} Y_i^2?$$

Explain (or prove) why your answer is correct.

6. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample from a $\mathcal{N}(\mu, \sigma^2)$ distribution.
(a) Under the normal model assumption (stated above), find the moment generating function of $\overline{Y}$, the sample mean. What is the distribution of $\overline{Y}$?
(b) In class, we looked at the (empirical) distribution of CD4 counts, observed from a sample of $n = 953$ Senegalese sex workers. Here is that distribution.

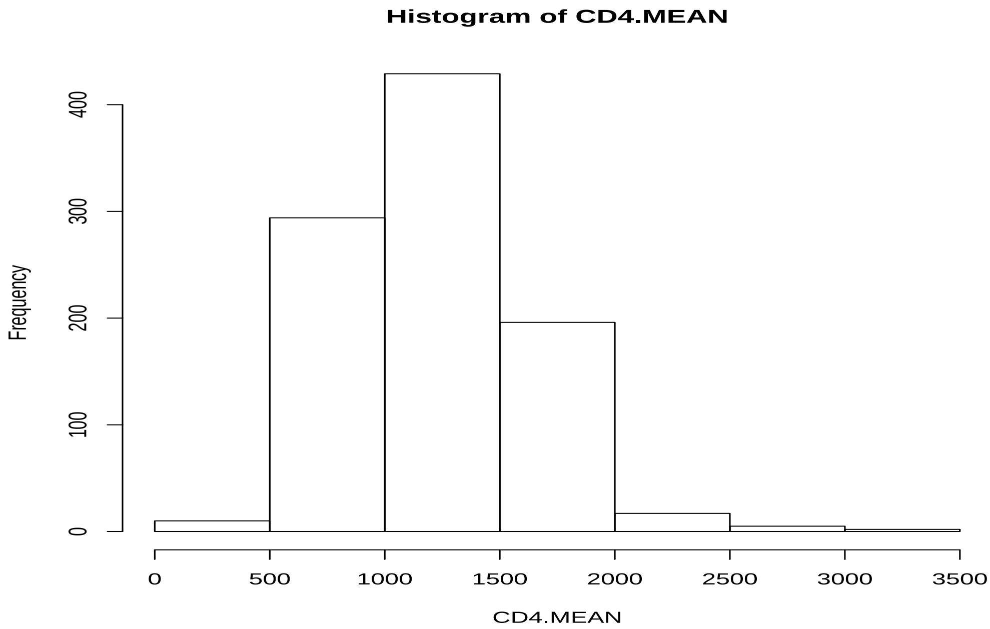

Figure 1: *Senegalese sex worker data. CD4 counts for* $n = 953$ *subjects.*

Do you think the normal model is a good model for these data? Why or why not? What about the iid assumption?
(c) If, instead of a normal model, we assume that the CD4 counts $Y_1, Y_2, ..., Y_{953}$ are iid gamma(a, b), find the moment generating function of $\overline{Y}$. What is the distribution of $\overline{Y}$ under this gamma assumption?

7. Prove the following three results. Your proofs should be rigorous (or at least very convincing).

(a) If $Y_1, Y_2, ..., Y_8$ are iid Poisson(θ), then

$$U = Y_1 + Y_2 + \cdots + Y_8 \sim \text{Poisson}(8\theta).$$

(b) If $Y_1, Y_2, ..., Y_8$ are independent $\mathcal{N}(\mu_i, \sigma_i^2)$, then

$$U = \sum_{i=1}^{8} \left(\frac{Y_i - \mu_i}{\sigma_i} \right)^2 \sim \chi^2(8).$$

(c) If $Y_1, Y_2, ..., Y_8$ are iid exponential(β) random variables, then

$$U = a(Y_1 + Y_2 + \cdots + Y_8) \sim \text{gamma}(8, a\beta),$$

for $a > 0$.

8. Actuarial data are used to posit a probability model for Y, the amount of an insurance claim (measured in \$10,000s) for a certain class of customers. Suppose that Y follows a Weibull distribution with cumulative distribution function (cdf)

$$F_Y(y) = \begin{cases} 0, & y \leq 0 \\ 1 - e^{-y^2}, & y > 0. \end{cases}$$

Suppose that $n = 5$ claims $Y_1, Y_2, ..., Y_5$ are made. Treat $Y_1, Y_2, ..., Y_5$ as an iid sample.
(a) Find the probability density function (pdf) of the minimum order statistic $Y_{(1)}$.
(b) What is the probability that the minimum of the claims exceeds \$5,000? That is, what is $P(Y_{(1)} > 0.5)$?

9. According to Newton's Law of Gravitation, if two bodies are a distance R apart, then the gravitational force U exerted by one body on the other is given by

$$U = g(R) = \frac{k}{R^2},$$

where k is a positive constant. Suppose that R is a random variable with probability density function (pdf)

$$f_R(r) = \begin{cases} 3r^2, & 0 < r < 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the pdf of U. Make sure you note the support.
(b) Find $E(U^{-1})$.

10. Suppose that Y_1, Y_2 is an iid sample of size $n = 2$ from a gamma$(2, 1)$ distribution. Recall that the gamma$(2, 1)$ probability density function (pdf) is given by

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

From the method of moment generating functions, we know that

$$U = Y_1 + Y_2 \sim \text{gamma}(4, 1).$$

Prove this result using either the cumulative distribution function (cdf) technique or the bivariate transformation technique. *Hint:* I think the bivariate transformation approach is easier, but both do work.

11. An insurance company sells an automotive insurance policy that covers all losses incurred by a policy holder. Let $Y_1, Y_2, ..., Y_8$ denote the losses (measured in \$1000) reported by a sample of $n = 8$ customers, and suppose that we model $Y_1, Y_2, ..., Y_8$ as an iid sample from an exponential distribution with mean $\beta = 2$. Let

$$T = Y_1 + Y_2 + \cdots + Y_8$$

denote the sum of the losses. Providing sufficient explanation, derive the moment generating function of T (don't just state the answer). What is the distribution of T?

12. Suppose that $Y_1, Y_2, ..., Y_5$ is an iid sample of size $n = 5$ from $f_Y(y)$, where

$$f_Y(y) = \begin{cases} 3y^2, & 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find the probability density function of the minimum order statistic $Y_{(1)}$ and compute $P(Y_{(1)} > 0.9)$.

13. Heparin is an anti-clotting medication, and its use is common among premature babies receiving care in neonatal intensive care units (NICU). The idea behind its use is to prevent blood clotting among babies who are receiving important fluids through a catheter (tube injected into the baby). Clotting at the site of catheter injection can be hazardous to the baby. While heparin can be successful at preventing blood clotting, its use has also been associated with increased infection (which is bad). To assess whether a heparin/antibiotic combination is helpful in preventing ancillary infections, a researcher plans to perform a clinical trial to compare premature babies randomised to the following treatment groups:

Group 1 : Heparin
Group 2 : Heparin + Antibiotic.

An important measurement to be studied is the length of time (measured in days) a baby is hospitalized in the NICU. Denote by Y_1 and Y_2 the length of stay for babies assigned to Groups 1 and 2, respectively. For this problem, we will assume that

- $Y_1 \sim \mathcal{N}(30, 64)$

- $Y_2 \sim \mathcal{N}(25, 36)$
- Y_1 and Y_2 are independent.

Compute the probability that a baby assigned to Group 2 has a longer NICU stay than a baby assigned to Group 1. That is, compute $P(Y_2 > Y_1)$.

14. Suppose that Y has an exponential distribution with mean $\beta = 1$; that is, the probability density function (pdf) of Y is given by

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

Derive the pdf of $U = g(Y) = \ln Y$.

15. Suppose that Y_1 and Y_2 are numbers generated independently and at random from the interval $(0, 1)$ so that the joint probability density function (pdf) of Y_1 and Y_2 is given by

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} 1, & 0 < y_1 < 1, 0 < y_2 < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Define $U = Y_1Y_2$, the product of Y_1 and Y_2. Derive the pdf of U. *Hint:* Use either the distribution function technique or a bivariate transformation technique. Don't use mgfs.

16. A triathlon is an athletic event made up of three contests (usually swimming, cycling, and running). Participants are timed for each contest, and the sum of the contest times represents the total time for the participant. For a certain triathlon, where time is measured in minutes, let

$$\begin{aligned} Y_1 &= \text{time for event 1} \\ Y_2 &= \text{time for event 2} \\ Y_3 &= \text{time for event 3.} \end{aligned}$$

In addition, for this triathlon, there is a sufficient amount of time between each event so that Y_1, Y_2, and Y_3 are independent.
(a) Suppose that $Y_1 \sim \mathcal{N}(120, 100)$, $Y_2 \sim \mathcal{N}(70, 50)$, and $Y_3 \sim \mathcal{N}(30, 20)$. What is the distribution of $U = Y_1 + Y_2 + Y_3$? *Hint*: U has a normal distribution, so first I want you to explain or prove why. Then, find the mean and variance of U.
(b) What is the probability that a participant in this triathlon will take longer than 250 minutes to complete all three events; that is, what is $P(U > 250)$?

17. A machine produces spherical containers whose radii, Y, vary according to the probability density function (pdf) given by

$$f_Y(y) = \begin{cases} 3y^2, & 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the pdf of U, the surface area of the containers. You will recall that the surface area of a sphere is given by $U = g(Y) = 4\pi Y^2$.
(b) Find $E(U)$.

18. Suppose that $Y_1, Y_2, ..., Y_n$ are independent random variables, where Y_i follows a Poisson distribution with mean λ_i; $i = 1, 2, ..., n$. That is,

$$\begin{aligned} Y_1 &\sim \text{Poisson}(\lambda_1) \\ Y_2 &\sim \text{Poisson}(\lambda_2) \\ &\vdots \\ Y_n &\sim \text{Poisson}(\lambda_n). \end{aligned}$$

Prove that $U = Y_1 + Y_2 + \cdots + Y_n$ has a Poisson distribution with mean $\lambda = \sum_{i=1}^n \lambda_i$.

19. An engineering system consists of $n = 5$ components operating together. The component lifetimes, which we will denote by $Y_1, Y_2, ..., Y_5$, are modeled as iid Weibull random variables with common pdf

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

(a) First, I want you to show that the cumulative distribution function (cdf) for this Weibull distribution is given by

$$F_Y(y) = \begin{cases} 0, & y \leq 0 \\ 1 - e^{-y^2/\theta}, & y > 0. \end{cases}$$

(b) Use the result in (a) to derive the probability density function for $Y_{(1)}$, the minimum component lifetime. Note that you can do this part, even if you could not do part (a).

20. Suppose that Y_1, Y_2 is an iid sample of size $n = 2$ from a standard normal distribution. Recall that the standard normal pdf is given by

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

Now, define the random variables $U_1 = g_1(Y_1, Y_2)$ and $U_2 = g_2(Y_1, Y_2)$, where

$$\begin{aligned} g_1(Y_1, Y_2) &= Y_1 + Y_2 \\ g_2(Y_1, Y_2) &= Y_1 - Y_2. \end{aligned}$$

(a) Use a bivariate transformation to derive the joint distribution of U_1 and U_2.
(b) What is the (marginal) distribution of U_2?

Chapter 7 Problems

21. The lifetime of a certain brand of industrial light bulb, Y, is assumed to follow a gamma distribution with shape parameter $\alpha = 3$ and scale parameter $\beta = 6$. A random sample (i.e., an iid sample) of $n = 25$ light bulbs will be taken; denote these measurements by $Y_1, Y_2, ..., Y_{25}$. Approximate the probability that the sample mean $\overline{Y}$ will be larger than 21.

22. Suppose that $Z_1, Z_2, ..., Z_6$ is an iid sample from a $\mathcal{N}(0, 1)$ distribution. Suppose that $Z_7 \sim \mathcal{N}(0, 1)$ and that Z_7 is independent of $Z_1, Z_2, ..., Z_6$. Find the distribution of

(a) $\overline{Z} = \frac{1}{6}\sum_{i=1}^{6} Z_i$

(b) $T = \sum_{i=1}^{6}(Z_i - \overline{Z})^2$

(c) $U = \sqrt{3}Z_7/\sqrt{Z_1^2 + Z_2^2 + Z_3^2}$

(d) $V = (Z_1^2 + Z_2^2 + Z_3^2)/(Z_4^2 + Z_5^2 + Z_6^2)$.

23. Let $Y_1, Y_2, ..., Y_{953}$ denote CD4 count measurements for $n = 953$ subjects. Assume that $Y_1, Y_2, ..., Y_{953}$ are iid gamma random variables with shape $\alpha = 9$ and scale $\beta = 130$.
(a) Approximate the probability that the sample mean $\overline{Y}$ will be between 1150 and 1200. That is, approximate $P(1150 < \overline{Y} < 1200)$.
(b) Your answer in part (a) is an approximation. Explain how you could compute $P(1150 < \overline{Y} < 1200)$ exactly without using an approximation. *Hint:* I'm not necessarily looking for a numerical answer here. I'm more interested in your explanation and mathematical approach. You could even leave your final answer as a definite integral.

24. I have four statistics T_1, T_2, T_3, and T_4. I have determined that

- $T_1 \sim \mathcal{N}(0, 1)$
- $T_2 \sim \mathcal{N}(-3, 4)$
- $T_3 \sim \chi^2(3)$
- $T_4 \sim \chi^2(5)$
- T_1, T_2, T_3, and T_4 are mutually independent.

(a) What is the distribution of $T_1 - T_2$?
(b) Find a function of T_1, T_2, T_3, and T_4 that has a t distribution with 8 degrees of freedom.
(c) Find a function of T_1, T_2, T_3, and T_4 that has an F distribution with 4 (numerator) and 6 (denominator) degrees of freedom.

25. A nerdy professor who teaches mathematical statistics observes

$$Y = \text{ the number of absent students}$$

each day that class meets. There are $n = 25$ days of class throughout the semester, so he observes $Y_1, Y_2, ..., Y_{25}$. Treat these observations as an iid sample from a Poisson distribution with mean $\lambda = 4$.
(a) Derive the moment generating function of

$$U = Y_1 + Y_2 + \cdots + Y_{25},$$

the total number of absences he observes throughout the semester. What is the distribution of U?
(b) Approximate the probability that U will not exceed 125. *Hint:* Use CLT.

26. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid $\mathcal{N}(\mu, \sigma^2)$ sample. Let $\overline{Y}$ and S^2 denote the sample mean and sample variance, respectively.
(a) Find the moment generating function of $\overline{Y}$. What is the distribution of $\overline{Y}$?
(b) Find the moment generating function of S^2. What is the distribution of S^2?

27. Suppose that

- $Y \sim \mathcal{N}(1, 4)$; i.e., Y has a normal distribution with $\mu = 1$ and $\sigma^2 = 4$
- $U \sim \chi^2(6)$
- $V \sim F(6, 1)$
- Y and U are independent.

(a) Find a function of Y and U that has a t distribution with 6 degrees of freedom.
(b) Find a function of Y and U that has the same distribution as $1/V$.

28. Body mass index (BMI) is a useful measure to estimate a healthy body weight based on how tall a person is. Suppose that a random sample of $n = 26$ severely obese patients is observed and that we model the BMI measurements $Y_1, Y_2, ..., Y_{26}$ as an iid sample from

a $\mathcal{N}(40, 25)$ distribution. Let $\overline{Y}$ and S^2 denote the sample mean and sample variance, respectively.
(a) What are the sampling distributions of $\overline{Y}$ and S^2?
(b) Compute $P(\overline{Y} > 41,\ S^2 < 16.5)$ *Hint:* You will use two probability tables.

29. Suppose that $Y_1, Y_2, ..., Y_n$ is an iid sample of Bernoulli observations with mean p, where $0 < p < 1$. Recall that the sample proportion is

$$\widehat{p} = \frac{X}{n},$$

where $X = Y_1 + Y_2 + \cdots + Y_n$. Prove that $E(\widehat{p}) = p$ and that

$$V(\widehat{p}) = \frac{p(1-p)}{n}.$$

30. Suppose that $Y_1, Y_2, ..., Y_5$ is an iid sample of size $n = 5$ from a $\mathcal{N}(0, 1)$ distribution. Let $\overline{Y}$ and S^2 denote the sample mean and sample variance, respectively. Define the statistics

$$\begin{aligned} T_1 &= \frac{Y_1^2}{(Y_2^2 + Y_3^2)/2} \\ T_2 &= \frac{\overline{Y}}{\sqrt{S^2/5}} \\ T_3 &= 4S^2 + (\sqrt{5}\overline{Y})^2 \end{aligned}$$

Give the precise distribution of each statistic. You need not derive anything here rigorously; you can use sampling distribution results we discussed/proved in class. However, your arguments should be convincing. A correct answer without explanation receives almost no credit.

31. Let Y denote the time (in minutes) that it takes for a customer representative to respond to a telephone inquiry. It is assumed that Y follows a uniform distribution from 0.5 to 3.5. That is, $Y \sim \mathcal{U}(0.5, 3.5)$. In a 2-hour shift, the customer representative responds to $n = 15$ calls with response times $Y_1, Y_2, ..., Y_{15}$. Treating these times as an iid sample, approximate the probability that the sample mean response time $\overline{Y}$ is less than 2.15 minutes. That is, approximate $P(\overline{Y} < 2.15)$.

32. Suppose that $U \sim \chi^2(64)$, that is, U has a χ^2 distribution with $\nu = 64$ degrees of freedom.
(a) Use the Central Limit Theorem (CLT) to approximate $P(60 < U < 70)$. State clearly how you are using the CLT.
(b) If $V \sim \chi^2(8)$, independent of U, find a function of U and V that has an F distribution. What are the degrees of freedom associated with your distribution?