Note: This homework assignment covers Chapter 4.

Disclaimer: If you use R, include all R code and output as attachments. Do not just "write in" the R code you used. Also, don't just write the answer and say this is what R gave you. If my grader can't see how you got an answer, it is wrong. I want to see your code and your answers accompanying your code (like in the notes).

1. UPS ships millions of packages in a specific 1-ft^3 packing container. Let

$$Y = \text{amount of space (in ft}^3\text{) occupied in this container.}$$

The probability density function (pdf) of Y is given by

$$f_Y(y) = \begin{cases} 90y^8(1-y), & 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Graph the pdf of Y. If you want, you can use this R code below:

```
y = seq(0,1,0.01)
pdf = 90*y^8*(1-y)
plot(y,pdf,type="l",xlab="y",ylab="PDF")
abline(h=0) # puts horizontal line at 0
abline(v=1,lty=2) # puts vertical line at 1; "lty" option alters line type
```

(b) Find the probability that the contents of a randomly selected 1-ft^3 packing container will occupy less than 1/2 of the volume of the container; i.e., calculate $P(Y < 0.5)$. Shade in the corresponding region on your pdf in part (a) showing this probability.
(c) Find the mean and variance of Y. Indicate the units of each quantity. Place an "$\times$" on the horizontal axis of your pdf in part (a) indicating where $E(Y)$ falls.
(d) I calculated the median of this distribution to be 0.8377 ft^3. Write out an equation (involving an integral) that when solved will give this answer. Make sure you say "what is being solved for" in your equation. You don't have to try to solve the equation! The equation would involve a 10 degree polynomial with only 1 real root in $(0, 1)$.

2. We say that Y follows a **uniform distribution** (from a to b) if the probability density function of Y is given by

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

In this model, probability (area) is allocated uniformly across the interval $[a, b]$.
(a) Derive the cumulative distribution function of Y and graph it. *Hint:* $F_Y(y) = 0$ when $y < a$ and $F_Y(y) = 1$ when $y > b$. Now find $F_Y(y)$ when $y \in [a, b]$.
(b) The size of particles used in sedimentation experiments often have a uniform distribution. Suppose that particles have diameters uniformly distributed between 0.01 and 0.05 cm.

(i) Find the mean and variance of this distribution.

(ii) Find the probability that a particle has diameter larger than 0.045 cm.

3. For a certain class of jet engines, the time until a major overhaul is needed, Y (measured in years), varies according to the following probability density function:

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

(a) What is the value of c? *Hint:* Is this a special exponential density?
(b) Find $E(Y)$ and $\text{var}(Y)$.
(c) What is the probability that the time to major overhaul is longer than 3 years? Would you consider this event unusual? Why or why not?
(d) Find the 95th percentile of this distribution and interpret what it means.
(e) Let t be a fixed constant. Show that, for $t < 1/2$,

$$E(e^{tY}) = \int_0^\infty e^{ty} f_Y(y)dy = \frac{1}{1-2t}.$$

Note that if you take the derivative of $E(e^{tY})$ with respect to t and then evaluate this derivative at $t = 0$, you get an answer that matches the value of $E(Y)$. Verify this. This function $E(e^{tY})$ is called the *moment-generating function* of Y. How do you think you could calculate $E(Y^2)$ using the moment-generating function? $E(Y^3)$? Some of you will note that $E(e^{tY})$ is essentially the LaPlace transform of the density $f_Y(y)$.

4. Alkalinity is a measure of the capacity of water or any solution to neutralize acids (e.g., acid rain). The alkalinity concentration in ground water (Y, measured in mg/L) near a waste disposal landfill in Wisconsin is modeled using a gamma distribution with $\alpha = 9.37$ and $\lambda = 0.16$.
(a) Graph the pdf of Y and the cdf of Y side by side (like in the notes).
(b) What is the mean alkalinity concentration? the median alkalinity concentration?
(c) What proportion of ground water samples near this landfill will have alkalinity concentrations less than 80 mg/L? You can use R to get this answer quickly; however, I want you to show on your pdf and on your cdf in part (a) where this answer comes from.
(d) Find the 99th percentile of this distribution. Interpret. Again, show on your pdf and on your cdf in part (a) where this answer comes from.

5. Human papillomavirus (HPV) infection has been established as the cause of virtually all forms of cervical cancer. Large cohort studies incorporate active follow-up with multiple visits and the collection of cervical specimens for HPV DNA testing. This is an important part of disease assessment and allows researchers to collect important covariate information that is linked to disease status. In a recent study, the Guanacaste Project, one of the covariates recorded (for all eligible subjects) was Y, the age at which the subject gave birth to her first child. This variable is possibly linked to HPV infectivity because HPV is a sexually transmitted disease. A normal probability model is attached to Y with mean $\mu = 26.3$ years and standard deviation $\sigma = 3.75$ years. Use this probability model to answer the following questions.
(a) Graph the pdf of Y and the cdf of Y side by side (like in the notes).
(b) Find the probability that a mother gives birth to her first child after she is 35 years old.
(c) Find the proportion of first-time mothers between 20 and 30 years.
(d) Ten percent of first-time mothers give birth before what age?
(e) In each of parts (b)-(d), you can get these answers quickly in R. Show how to get each answer from the pdf and the cdf in part (a); e.g., just hand annotate your graphs.