This problem set covers material from Weeks 6-7, dates 3/28 - 4/04. Unless otherwise noted, all problems are taken from the textbook. Problems can be found at the end of the corresponding chapter.

Instructions: Write or type complete solutions to the following problems and submit answers to the corresponding Canvas assignment. Your solutions should be neatly-written, show all work and computations, include figures or graphs where appropriate, and include some written explanation of your method or process (enough that I can understand your reasoning without having to guess or make assumptions). A general rubric for homework problems appears on the final page of this assignment.

Friday 3/28

1. **(Change-of-Variables Formula)** Suppose X is a random variable with support S_X and CDF $F_X(x)$, and let $Y = g(X)$. If g is increasing (and hence, invertible) on S_X, then the CDF $H_Y(y)$ of Y is given by

$$H_Y(y) = F_X(g^{-1}(y)) \tag{1}$$

Moreover, if X is a continuous random variable with PDF $f_X(x)$, then the PDF $h_Y(y)$ of Y is

$$h_Y(y) = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}, \qquad S_Y = \{g(x) : x \in S_X\} \tag{2}$$

   (a) Prove the change-of-variables formula above. Note: there are two statements to prove: one about the CDF (1) and one about the PDF (2).
   Hint: Use the chain-rule from calculus, along with the formula for the derivative of an inverse function.

   (b) Use the Change-of-Variable formula from part (a), along with u-substitution from calculus, to prove LOTUS for continuous random variables (at least in the case when g is an increasing function):

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

   where X is a continuous random variable with density $f_X(x)$.

2. 5.4

3. 5.13

Monday 3/31

4. 5.22

5. 5.23

Wednesday 4/02

6. (adaption of Blitzstein 5.46): Let T be the lifetime of a person (how long that person lives), with CDF $F_T(t)$ and PDF $f_T(t)$ for $t > 0$. The *hazard function* of T is defined by:

$$h_T(t) = \frac{f_T(t)}{1 - F_T(t)}.$$

In common language, the *hazard* is the probability of the event (i.e. death) occurring during any given instant t. We will derive the hazard function of T.

(a) Find the conditional CDF of T given that the person has survived to at least time t_0. That is, find $P(T \leq t | T > t_0)$ for $0 < t_0 \leq t$.

(b) Using your conditional CDF from (a), find the conditional PDF of T given that the person has survived to at least time t_0. How does this relate to the hazard function? Briefly explain why this makes sense/give an interpretation to the hazard function.

(c) Show that an Exp(λ) random variable has a constant hazard function.

Friday 4/04

7. The Pareto distribution with parameter $\alpha > 0$ has PDF $f_X(x) = \alpha x^{-(\alpha+1)}$ for $x \geq 1$, and 0 otherwise. We will denote this as $X \sim \text{Pareto}(\alpha)$. This distribution is commonly used in statistical modeling.

(a) Find the CDF the Pareto(α) distribution.

(b) Write R code that generates 5000 i.i.d. Pareto(2) random variables, then use your simulations to approximate the mean of this distribution. Show any written work that was required to implement the code. You may submit a screenshot of your code, or a knitted `.Rmd` document.

General rubric

Points	Criteria
5	The solution is correct *and* well-written. The author leaves no doubt as to why the solution is valid.
4.5	The solution is well-written, and is correct except for some minor arithmetic or calculation mistake.
4	The solution is technically correct, but author has omitted some key justification for why the solution is valid. Alternatively, the solution is well-written, but is missing a small, but essential component.
3	The solution is well-written, but either overlooks a significant component of the problem or makes a significant mistake. Alternatively, in a multi-part problem, a majority of the solutions are correct and well-written, but one part is missing or is significantly incorrect.
2	The solution is either correct but not adequately written, or it is adequately written but overlooks a significant component of the problem or makes a significant mistake.
1	The solution is rudimentary, but contains some relevant ideas. Alternatively, the solution briefly indicates the correct answer, but provides no further justification.
0	Either the solution is missing entirely, or the author makes no non-trivial progress toward a solution (i.e. just writes the statement of the problem and/or restates given information).
Notes:	For problems with multiple parts, the score represents a holistic review of the entire problem. Additionally, half-points may be used if the solution falls between two point values above.
Notes:	For problems with code, well-written means only having lines of code that are necessary to solving the problem, as well as presenting the solution for the reader to easily see. It might also be worth adding comments to your code.