CHAPTER 1:

1.1. Define the sequence of sets $A_n = \{\omega : \frac{1}{n} < \omega < 1 + \frac{1}{n}\}$.
(a) Is $\{A_n\}$ monotone?
(b) Find

$$\bigcap_{n=1}^{\infty} A_n \quad \text{and} \quad \bigcup_{n=1}^{\infty} A_n.$$

(c) Find $\limsup_n A_n$ and $\liminf_n A_n$. Does $\lim_{n\to\infty} A_n$ exist? Why or why not?
(d) Repeat the parts above with $A_n = \{\omega : -n \leq \omega < 1 - \frac{1}{n}\}$.

1.2. Let $(S, \mathcal{B})$ be a measurable space and suppose $C \in \mathcal{B}$. Show the collection of sets defined by $\mathcal{B}_C = \{C \cap A : A \in \mathcal{B}\}$ is a σ-algebra on C. *Hint:* The complement of a set relative to $C \subseteq S$ is $C \cap A^c$, where A^c is the complement of A relative to S. Also if $C \in \mathcal{B}$, then $\mathcal{B}_C \subseteq \mathcal{B}$.

1.3. Suppose $(S, \mathcal{B})$ is a measurable space. Suppose P_1 and P_2 are probabilities on $(S, \mathcal{B})$. Suppose $0 \leq \alpha \leq 1$. For any $A \in \mathcal{B}$, define the set function

$$P(A) = \alpha P_1(A) + (1 - \alpha)P_2(A).$$

Show that P also defines a probability on $(S, \mathcal{B})$.

1.4. Suppose $S = \{\omega \in \mathbb{R} : -1 < \omega < 2\}$ and let $\mathcal{B} = \{B \cap S : B \in \mathcal{B}(\mathbb{R})\}$ denote the collection of Borel sets on S. Define the sequence of events

$$A_n = \begin{cases} (-1/n, 0], & \text{if } n \text{ is odd} \\ (0, 1 + 1/n], & \text{if } n \text{ is even.} \end{cases}$$

(a) True or False: $\{A_n, n = 1, 2, ...,\}$ partitions S.
(b) Does $\lim_{n\to\infty} A_n$ exist? If so, show that it does. If not, show why not.
(c) Suppose P is a set function on $(S, \mathcal{B})$ defined by

$$P(A) = \int_A \frac{2}{9}(x + 1)dx$$

for any $A \in \mathcal{B}$. Show that P satisfies the Kolmolgorov Axioms.

1.5. Suppose $(S, \mathcal{B}, P)$ is a probability space and that $\{A_n\}$ is a sequence of measurable events; i.e., $A_n \in \mathcal{B}$, for $n = 1, 2, ...,$. Show that

$$P(A_n) = 1 \ \forall n \implies P\left(\bigcap_{n=1}^{\infty} A_n\right) = 1.$$

Is the converse true? If so, prove it. If not, give a counterexample.

1.6. Let f_n denote the number of ways of tossing a coin n times such that successive heads never appear.

(a) First show $f_n = f_{n-1} + f_{n-2}$, for $n \geq 2$, where $f_0 = 1$ and $f_1 = 2$.
(b) If P_n denotes the probability successive heads never appear when a coin is tossed n times, find P_n (in terms of f_n) when all possible outcomes of the n tosses are assumed to be equally likely. Calculate P_{10}.
(c) Show that

$$\frac{f_n}{f_{n-1}} \rightarrow \frac{1+\sqrt{5}}{2}, \quad \text{as } n \rightarrow \infty.$$

(d) Show that, for large n,

$$P_n \approx \frac{P_{n-1}}{2}\left(\frac{1+\sqrt{5}}{2}\right).$$

1.7. Take $S = \{1, 2, ..., n\}$ and suppose A and B, independently, are equally likely to be any of the 2^n subsets of S, including $\emptyset$ and S itself.
(a) Show that

$$P(A \subseteq B) = \left(\frac{3}{4}\right)^n.$$

Hint: Let $|B|$ denote the number of elements in B. Write $P(A \subseteq B)$ using LOTP where you condition on $\{|B| = i\}$, for $i = 0, 1, 2, ..., n$. When I write $A \subseteq B$, I emphasize that A and B could be equal.
(b) Under identical conditions, what is $P(A \subset B)$? When I write $A \subset B$, I emphasize that A and B can not be equal; i.e., A is a proper subset of B.

1.8. Suppose $(S, \mathcal{B}, P)$ is a probability space and let A and B be measurable events; i.e., $A \in \mathcal{B}$ and $B \in \mathcal{B}$.
(a) If $P(A \cap B) = 0$ and $P(A) = 1$, show $P(B) = 0$.
(b) If $P(A) > 0$, $P(B) > 0$, and $P(A) < P(A|B)$, show $P(B) < P(B|A)$.
(c) If A and B are independent, show A^c and B^c are independent.

1.9. Suppose X is a continuous random variable with cdf F_X. Suppose $a \geq 0$. Prove that

$$\int_{-\infty}^{\infty} \{F_X(x+a) - F_X(x)\} dx = a.$$

1.10. A random variable X is said to be *symmetric* if X and $-X$ have the same probability distribution, that is, if $X \stackrel{d}{=} -X$.
(a) Show that X is symmetric if and only if $P_X(X \leq -x) = P_X(X \geq x)$, for all $x \in \mathbb{R}$.
(b) Obtain an equivalent condition for symmetry of X in terms of its cdf F_X.
(c) If X is symmetric, show $P_X(|X| \leq x) = 2F_X(x) - 1$, for all $x \geq 0$.
(d) If X is symmetric, show $P_X(X \geq 0) = 0.5$. Assume X is continuous.
(e) Suppose X has pdf f_X. Prove X is symmetric if and only if f_X is an even function.

1.11. Suppose F_1, F_2, ..., F_k are cdfs. Also let $p_1, p_2, ..., p_k$ be positive real numbers with $\sum_{i=1}^{k} p_i = 1$. Define $G_X(x) = \sum_{i=1}^{k} p_i F_i(x)$. G_X refers to a *mixture distribution.*
(a) Prove G_X is a valid cdf.

(b) For specificity, take $k = 2$, $F_1(x) = (1 - e^{-x^2})I(x > 0)$, $F_2(x) = (1 - e^{-x/10})I(x > 0)$, and $p_1 = p_2 = 0.5$. That is, the distribution of X is a mixture of 2 distributions (with equal weighting). Derive the pdf of X and show it integrates to 1. Graph the pdf. What happens to the pdf when the weights (p_1 and p_2) change?

1.12. Suppose $F_0(x)$ is a cumulative distribution function (cdf). Suppose X is a random variable with

$$F_X(x) = P_X(X \leq x) = (1 + \delta)F_0(x) - \delta[F_0(x)]^2,$$

for $x \in \mathbb{R}$, where δ satisfies $-1 \leq \delta \leq 1$.
(a) Show $F_X(x)$ is a valid cdf. You may assume $F_0(x)$, the so-called "base cdf," is valid.
(b) For this part, take the base cdf to be $F_0(x) = 1 - e^{-x/\lambda}$, for $x > 0$, where $\lambda > 0$. For $x \leq 0$, take $F_0(x) = 0$. Show the probability density function (pdf) of X is

$$f_X(x) = \frac{1}{\lambda}e^{-x/\lambda}(1 - \delta + 2\delta e^{-x/\lambda})I(x > 0).$$

1.13. The rules for the game of craps are as follows. A player rolls two dice and computes the total of the spots showing.

- If the player's first toss is a 7 or 11, the player wins the game.
- If the first toss is a 2, 3, or 12, the player loses the game.
- If the player rolls anything else (4, 5, 6, 8, 9, or 10) on the first toss that value becomes the player's "point." If the player does not win or lose on the the first toss he tosses the dice repeatedly until he obtains either his point or a 7. He wins if he tosses his point before tossing a 7 and loses if he tosses a 7 before his point.

(a) A player decides to play 10 games of craps. Define the random variable X to be the number of games that s/he wins (out of 10). Find the probability mass function (pmf) and cumulative distribution function (cdf) of X and graph them side by side. Be sure to list any assumptions that you make.
(b) A stubborn player keeps playing games until he wins. Let Y denote the number of games he will play. Find the pmf and cdf of Y and graph them side by side. Be sure to list any assumptions that you make.

1.14. There are M different chocolate bars for Ann to choose from. The bars are labeled from 1 to M ($M \geq 4$), and the one with label i is worth i dollars, for $i = 1, 2, ..., M$.
(a) Ann chooses one chocolate bar from the M bars. Suppose the probability Ann chooses bar i is proportional to the dollar value of this bar, for $i = 1, 2, ..., M$. Let X denote the dollar value of the selected chocolate bar. Provide the probability mass function of X.
(b) As in part (a), Ann chooses one chocolate bar with the probability of choosing bar i proportional to the dollar value of this bar, for $i = 1, 2, ..., M$. Now, seeing the value of the selected bar, Ann hesitates whether or not she should get a second bar. Assume the probability Ann selects a second bar is $(1/2)^i$, for $i = 1, 2, ..., M$ (so the cheaper the first bar is, the more likely Ann will get one more bar). Find the probability Ann chooses a second bar.
(c) Consider a different experiment where Ann chooses three chocolate bars from this collection

of M bars with replacement (i.e., after Ann chooses a bar, she places it back into the collection before choosing again). Assume the outcome of one choice does not affect her other choices. Use the pmf from (a) to find the probability that the \$1 chocolate bar is chosen at least once.

1.15. A pinball machine has 7 holes through which a ball can drop. Five balls are played and we observe which hole each ball goes down. For example, the first ball could go down hole 1, hole 2, ..., or hole 7 (similarly for the other 4 balls).
(a) Provide a probability model $(S, \mathcal{B}, P)$ for this random experiment. Clearly describe what S, $\mathcal{B}$, and P are; e.g., describe what an outcome $\omega \in S$ "looks like."
(b) On each play, assume the ball is equally likely to go down any one of the 7 holes. Find the probability that more than one ball goes down at least one of the holes.
(c) Let X count the number of balls that go down hole 1. Pick two or three outcomes $\omega \in S$ and calculate $X(\omega)$ for them. Also, find the distribution of X.

1.16. Suppose $(S, \mathcal{B}, P)$ is a probability space. In the parts that follow, assume A, B, and C are events in $\mathcal{B}$. For parts (b) and (c), assume that $P(A) > 0$, $P(B) > 0$, and $P(C) > 0$.
(a) The *symmetric difference* of A and B, denoted by $A \triangle B$, is the set of all outcomes belonging to A or B but not both events. Prove that

$$P(A \triangle B) = P(A) + P(B) - 2P(A \cap B).$$

(b) Prove: If B and C are independent, then

$$P(A|B) = P(A|B \cap C)P(C) + P(A|B \cap C^c)P(C^c).$$

(c) The converse to (b) is not true. However, if $P(A|B) = P(A|B \cap C)P(C) + P(A|B \cap C^c)P(C^c)$ and $P(A|B \cap C) \neq P(A|B)$, then B and C are independent. Prove this.

1.17. A random variable X measures the time to failure for a transistor and has cumulative distribution function (cdf)

$$F_X(x) = \begin{cases} 1 - e^{-x^2}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

(a) Verify $F_X(x)$ is a valid cdf.
(b) Find the probability density function $f_X(x)$.
(c) Find the value of m which satisfies $P_X(X \leq m) = 0.5$. This is called the *median* of X.
(d) The transistor will be replaced for free if it fails at or before time 1. Otherwise, the cost incurred from the transistor failing is $10(X - 1)$. Therefore, the cost of replacement is $Y = 10(X - 1)I(X > 1)$. Find the cdf of Y and graph it.

1.18. Suppose $(S, \mathcal{B}, P)$ is a probability space.
(a) Describe clearly what a sigma-algebra $\mathcal{B}$ is and what properties define a sigma-algebra.
(b) Using appropriate notation, list the three axioms that a probability set function P satisfies.
(c) Suppose that $A_1, A_2, ..., A_n$ is a finite sequence of events. Describe clearly what it means for $A_1, A_2, ..., A_n$ to be mutually independent.
(d) Define mathematically what it means for X to be a random variable over $(S, \mathcal{B}, P)$. Also, describe how probabilities of events involving X can be calculated over $(S, \mathcal{B}, P)$.

1.19. A diagnostic test used to detect the presence of a disease D is known to give a false positive result 1% of the time; it correctly gives a positive result when the disease is present 98% of the time. The disease is known to be present in 1 out of every 1000 people in a large population.
(a) If a randomly selected person tests positively, what is the probability the person has the disease?
(b) If a randomly selected person tests positively on each of three test applications (given in sequence, say), what is the probability the person has the disease?

1.20. Suppose $(S, \mathcal{B}, P)$ is a probability space. Assume all sets below (e.g., A, B, C, $A_1, A_2, ...,$, etc.) are measurable with respect to $\mathcal{B}$.
(a) Describe what it means for a sequence of sets $A_1, A_2, ...,$ to "converge." What does this imply about $\lim_{n\to\infty} P(A_n)$?
(b) If $A \subseteq B$, prove that $P(A) \leq P(B)$. Then, for the set C with $P(C) > 0$, show that $P(A|C) \leq P(B|C)$.
(c) True or False. Boole's Inequality

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq P\left(\sum_{n=1}^{\infty} A_n\right)$$

only holds when $A_1, A_2, ...,$ are pairwise disjoint.

1.21. An individual tried by a 3-judge panel is declared guilty if at least 2 judges cast votes of guilty. Suppose

- when the defendant is guilty, each judge will independently vote guilty with probability 0.8.
- when the defendant is not guilty, each judge will independently vote guilty with probability 0.2.

Suppose 70 percent of all defendants are guilty (thus, 30 percent of all defendants are not guilty).
(a) Find the probability that a defendant who is not guilty will be declared guilty by the panel.
(b) Find the probability that judge number 3 votes guilty given that judges 1 and 2 vote guilty.

CHAPTER 2:

2.1. Suppose X is a discrete random variable with probability mass function

$$f_X(x) = \begin{cases} (\frac{1}{2})^{x+1}, & x = 0, 1, 2, 3, ... \\ 0, & \text{otherwise.} \end{cases}$$

(a) Derive $E(X)$ and $\text{var}(X)$. Don't just state the answers.
(b) Derive the probability mass function of $Y = (X-2)^2$.

2.2. A continuous random variable X is said to have a *Cauchy distribution* if its probability density function is

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty.$$

Define $Y = -X$ and $Z = 1/X$. Show that X, Y, and Z all have the same distribution. *Hint:* Derive the cdf of each random variable.

2.3. A random variable X is said to be *bounded* if there exists $M > 0$ such that

$$P_X(|X| \leq M) = P_X(-M \leq X \leq M) = 1.$$

(a) Suppose X is a bounded continuous random variable with probability density function (pdf) $f_X(x)$. Prove that $E(|X|) \leq M$.
(b) Suppose X is a bounded continuous random variable with pdf $f_X(x)$. Prove that the moment generating function of X exists for all $t \in \mathbb{R}$.
Hint: For both parts, you can assume that $f_X(x) = 0$ if $x \notin [-M, M]$. Explain why.

2.4. Suppose $F_X(x)$ is the cdf of a continuous random variable X. We have learned $Y = F_X(X)$ follows a uniform distribution on $[0, 1]$. What can we say about the distribution of $W = f_X(X)$, the pdf of X? Let us focus on two cases where X follows a distribution with which we are familiar. Derive the distribution of W in each case.
(a) X follows an exponential distribution with pdf given by

$$f_X(x) = \frac{1}{\beta} e^{-x/\beta} I(x > 0),$$

where $\beta > 0$.
(b) X follows the standard normal distribution with pdf given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} I(x \in \mathbb{R}).$$

2.5. Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$, that is, the pdf of X is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} I(x \in \mathbb{R}),$$

where $-\infty < \mu < \infty$ and $\sigma^2 > 0$. Let $Y = e^X$. Derive the pdf of Y and calculate $E(Y)$ and $\text{var}(Y)$. The distribution of Y is called a *lognormal distribution.*

2.6. In Chapter 3, we will study many named probability distributions. While useful in their own right, these distributions sometimes do not describe exactly what we see in real life. Sometimes statisticians "adjust" familiar distributions to construct "closer-to-reality" distributions. Suppose X is a continuous random variable with pdf $f_X(x)$ which is symmetric about zero, that is, $f_X(x) = f_X(-x)$ for all $x \in \mathbb{R}$. However, after observing different values of X, one suspects that the true pdf of X may not be symmetric. Motivated by this new information, one adjusts $f_X(x)$ in the following way to define a new pdf that can be skewed

$$f^*_X(x) = 2 f_X(x)\pi(x),$$

where $\pi(x)$ is a *skewing function.* The function $\pi(\cdot)$ maps $\mathbb{R}$ to $[0,1]$ and satisfies $\pi(-x) = 1-\pi(x)$ for all $x \in \mathbb{R}$.
(a) Show that $f_X^*(x)$ is a valid pdf. *Hint:* Write the integral over $\mathbb{R}$ as the sum of two integrals. Let $u=-x$ in the first integral.
(b) Take X to have a standard normal distribution so that $E(X)=0$ and $\text{var}(X)=1$. Plot the pdf of X using your favorite software. Now choose two different skewing functions $\pi(x)$ and graph the corresponding revised pdfs. For each skewing function, calculate your adjusted $E(X)$ and $\text{var}(X)$. You may have to calculate these numerically.

2.7. Suppose X follows an exponential distribution with mean $\beta = 1$, that is, the pdf of X is $f_X(x) = e^{-x}I(x>0)$. For $-\infty < \alpha < \infty$ and $\beta > 0$, define

$$Y = g(X) = \alpha - \beta \ln X.$$

(a) Show the pdf of Y is

$$f_Y(y) = \frac{1}{\beta}\exp\left\{-\left(\frac{y-\alpha}{\beta}\right) - e^{-(y-\alpha)/\beta}\right\},$$

for all $y \in \mathbb{R}$. The random variable Y is said to follow a *Gumbel distribution.* This distribution is useful in modeling extreme values (e.g., maximum temperature, etc.).
(b) Define

$$Z = \frac{Y-\alpha}{\beta}.$$

Show $E(Z)=\gamma$, where γ is the *Euler-Mascheroni* constant defined via

$$\gamma = -\int_0^\infty e^{-s}\log s\, ds.$$

(c) Show the mgf of Z in part (b) is given by $M_Z(t) = \Gamma(1-t)$, for $t<1$, where $\Gamma(\cdot)$ is the gamma function.
(d) Derive the mgf of Y. Derive the mean and variance of Y using the mgf.

2.8. Suppose $X_n \sim \text{gamma}(n/2, 2)$ and let

$$W_n = \frac{X_n - n}{\sqrt{2n}}.$$

(a) Derive $M_{W_n}(t)$, the mgf of W_n.
(b) Show that $M_{W_n}(t) \to e^{t^2/2}$ for all t, as $n\to\infty$. Note $e^{t^2/2}$ is the mgf of a standard normal random variable. Therefore, $W_n \xrightarrow{d} \mathcal{N}(0,1)$.
(c) Perform a simulation study in R to see how good the standard normal approximation is in part (b) for n large. Experiment with different values of n.

2.9. An insurance company models the number of "serious injury" claims per day, X, as a random variable with pmf

$$f_X(x) = \begin{cases} \dfrac{1}{(x+1)(x+2)}, & x = 0,1,2,... \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show this is a valid pmf; that is, show that

$$\sum_{x=0}^{\infty} f_X(x) = 1.$$

(b) Show $E(X)$ does not exist. *Hint:* Try to find a function $g(x)$ such that

$$\frac{x}{(x+1)(x+2)} > g(x),$$

for all $x \geq 1$, where the function $g(x)$ is not integrable over $[1, \infty)$; i.e., $\int_1^\infty g(x)dx = +\infty$. The result then follows from the integral comparison test from calculus.
(c) Does the mgf of X exist? If so, derive it. If not, show that it does not.

2.10. Suppose X is a continuous random variable with pdf

$$f_X(x) = \beta^{-1}e^{-x/\beta}I(x > 0),$$

where $\beta > 0$. Derive the distribution of each of the following functions of X and derive the mean and variance of each function.
(a) $U = aF_X(X) + b$, where $F_X(x)$ is the cdf of X and $a, b \in \mathbb{R}$.
(b) $V = X + \theta$, where $\theta \in \mathbb{R}$.
(c) $W = \cos(X)$.
(d) $Y = \lceil X \rceil$, where $\lceil a \rceil$ is the *ceiling function*, that is, the smallest integer that is greater than or equal to a. For example, $\lceil 5.5 \rceil = 6$ and $\lceil 6 \rceil = 6$.
(e) $R = \lceil X \rceil - X$.
Note: I think part (c) can be done analytically, but I am not sure (note that this is not a 1:1 transformation over $\mathbb{R}^+$). If you can not make sufficient progress analytically, could you investigate this part using simulation?

2.11. In each part, derive the moment generating function (mgf) of X and then use the mgf to find $E(X)$ and $\text{var}(X)$.
(a) X has pmf

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

where $0 < \theta < 1$.
(b) X has pdf

$$f_X(x) = \begin{cases} \frac{1}{3}, & -1 < x < 2 \\ 0, & \text{otherwise.} \end{cases}$$

2.12. The random variable X has the following probability density function

$$f_X(x) = \begin{cases} \dfrac{1}{2\beta}e^{-|x-\alpha|/\beta}, & -\infty < x < \infty \\ 0, & \text{otherwise,} \end{cases}$$

for values of $-\infty < \alpha < \infty$ and $\beta > 0$.
(a) Find $M_X(t)$, the moment generating function of X. Make sure to note for which values of t the mgf is defined.

(b) Find $E(X)$ and $\text{var}(X)$.
(c) Derive the pdf of $Y = g(X) = (X - \alpha)/\beta$.

2.13. A discrete random variable X has probability mass function

$$f_X(x) = \begin{cases} \dfrac{c}{x2^x}, & x = 1, 2, 3, \ldots \\ 0, & \text{otherwise,} \end{cases}$$

where the constant $c = (\ln 2)^{-1}$.
(a) Sketch a graph of what the cumulative distribution function $F_X(x)$ looks like. Be precise.
(b) Show the moment generating function of X is

$$M_X(t) = -c \ln\left(1 - \frac{e^t}{2}\right), \quad \text{for } t < \ln 2.$$

Hint: Write out $h(a) = -\ln(1-a)$ in its Maclaurin series expansion. Where does the condition $t < \ln 2$ come from?
(c) Find $E(X)$ and $\text{var}(X)$.

2.14. A random variable X is said to have a *logistic distribution* if its cumulative distribution function is given by

$$F_X(x) = \frac{1}{1 + e^{-x}}, \quad x \in \mathbb{R}.$$

(a) Show that

$$\ln\left(\frac{F_X(x)}{1 - F_X(x)}\right)$$

is a linear function of x. This forms the basis for *logistic regression*.
(b) Show that X and $-X$ have the same distribution.
(c) What is the distribution of $Y = 1/(1 + e^{-X})$?
(d) Derive the distribution of $Z = 1 + e^X$. Show that $E(Z)$ does not exist.
(e) I had trouble deriving $E(|X|)$ so I used R to approximate $E(|X|) \approx 1.39$.

```
> sim = rlogis(1000000,0,1)  # simulate 1,000,000 observations from F(x)
> mean(abs(sim))             # sample mean of |x| values
[1] 1.385874
```

Write $E(|X|)$ as an integral. What do you think the pdf of $|X|$ looks like? Sketch a graph.

2.15. A continuous random variable X has probability density function

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

(a) Show that $f_X(x)$ can be written as follows:

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

Why do you think this density is called "a triangular density?"

(b) Derive $F_X(x)$, the cumulative distribution function of X.
(c) The moment generating function (mgf) for $t \neq 0$ is

$$M_X(t) = \frac{2e^t}{t^2}(\cosh t - 1).$$

How should the mgf be defined at $t = 0$ to make $M_X(t)$ a continuous function of t? *Hint:* Recall from calculus that $\cosh t = (e^t + e^{-t})/2$.
(d) Find $E[(X-1)^2]$.

2.16. Suppose $F_1(x)$ and $F_2(x)$ are cumulative distribution functions (cdfs). Suppose X is a random variable with

$$F_X(x) = P_X(X \leq x) = p_1F_1(x) + p_2F_2(x),$$

for $x \in \mathbb{R}$, where $0 < p_1 < 1$, $0 < p_2 < 1$, and $p_1 + p_2 = 1$.
(a) Prove $F_X(x)$ is a valid cdf. You may assume $F_1(x)$ and $F_2(x)$ are valid.
(b) For this part, take

$$\begin{aligned} F_1(x) &= (1 - e^{-x^2})I(x > 0) \\ F_2(x) &= (1 - e^{-x/5})I(x > 0). \end{aligned}$$

Derive the probability density function of X.
(c) In part (b), find $E(X)$.

2.17. Suppose X is a random variable with probability mass function (pmf)

$$f_X(x) = \begin{cases} \alpha p^x, & x = 1, 2, 3, ..., \\ 0, & \text{otherwise}, \end{cases}$$

where $0 < p < 1$ is a constant.
(a) What value of α makes $f_X(x)$ a valid pmf?

For parts (b) and (c), suppose $f_X(x)$ is a probability model for X, the number of children a family will have.

(b) What is the mean number of children the family will have?
(c) If each child is equally likely to be a girl or boy (independently of each other; no multiple births, etc.), what is the probability the family will have k girls (and any number of boys)?

2.18. Suppose X is a random variable with $E(X) = \mu$ and $\text{var}(X) = \sigma^2 < \infty$. Suppose the moment generating function (mgf) of X, $M_X(t)$, is defined for all $t \in (-h, h)$, where $h > 0$.
(a) What is $M_X(0)$?
(b) Suppose $a, b \in \mathbb{R}$. Derive the mgf of $Y = a + bX$ in terms of $M_X(t)$.
(c) Calculate

$$\lim_{t \to 0} \frac{\ln M_Y(t) - (a + b\mu)t}{t^2}.$$

2.19. Suppose $X_1, X_2, ...,$ is a sequence of random variables. The probability density function (pdf) of X_n is

$$f_{X_n}(x) = nx^{n-1}I(0 < x < 1).$$

(a) Derive the cumulative distribution function (cdf) of X_n. Call it $F_{X_n}(x)$.
(b) Find the pointwise limit of the sequence of functions $F_{X_1}, F_{X_2}, ...$ Call the limit F_X. Is $F_X(x)$ a valid cdf? Explain.
(c) The moment-generating function (mgf) of X_n is given by

$$M_{X_n}(t) = 1 + \sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{n+r}{n+r+1}\right)\frac{t^k}{k!}.$$

Suppose $M_{X_n}(t) \to M_X(t)$ pointwise for all t, where $M_X(t)$ is the mgf of $X \sim F_X(x)$, defined in part (b). Find $M_X(t)$.

2.20. Suppose X is a random variable with probability density function (pdf)

$$f_X(x) = \begin{cases} \frac{1}{2}(1+\theta x), & -1 < x < 1 \\ 0, & \text{otherwise,} \end{cases}$$

where θ is a constant satisfying $-1 < \theta < 1$.
(a) Find the pdf of $Y = e^X$.
(b) Find $E(Y)$.
(c) Repeat parts (a) and (b) with $f_X(x) = \alpha^{-1}e^{-x/\alpha}I(x > 0)$, where $\alpha > 0$.

CHAPTER 3:

3.1. Suppose X is a continuous random variable with probability density function (pdf)

$$f_X(x|\theta) = \theta x^{-(\theta+1)}I(x > 1),$$

where $\theta > 0$.
(a) Calculate an exact expression for the tail probability $P_X(X > c)$, for $c > 1$.
(b) Calculate the upper bound on $P_X(X > c)$, $c > 1$, provided by Markov's Inequality. Note any restrictions on θ that are needed for the upper bound to be applicable.

3.2. In each part below, give an example of a parametric family $\{f_X(x|\boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta\}$ that satisfies the stated conditions. Specify the family, the parameter space Θ, and the support $\mathcal{X}$. Prove any claims you make.
(a) a family that is a single parameter exponential family (i.e., $d = 1$) and also a scale family.
(b) a family that is a two-parameter full exponential family (i.e., $d = k = 2$) and also a location-scale family.
(c) a two-parameter full exponential family with $\boldsymbol{\theta} = (\theta_1, \theta_2)$ that contains a subfamily, with $\theta_2 = g(\theta_1)$, that is a full one-parameter exponential family.

3.3. Recall if X has a beta(α, β) distribution, then the probability density function (pdf) of X is

$$f_X(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I(0 < x < 1).$$

(a) Show $\{f_X(x|\alpha,\beta); \alpha > 0, \beta > 0\}$ is a two-parameter exponential family. Is this a full or curved family?
(b) For this part only, suppose that $\alpha > 1$ and $\beta > 1$. Find the mode of X, that is, find the value of x that maximizes $f_X(x|\alpha,\beta)$.
(c) Derive the pdf of $Y = 1 - X$.

3.4. Suppose $X \sim \mathcal{U}(a, b)$, where $a < b$.
(a) Show the pdf of X is a location-scale family generated by $f_Z(z) = I(0 < z < 1)$.
(b) Show that the pdf of $Y = Z(1 - Z)$ is

$$f_Y(y) = 2(1-4y)^{-1/2} I(0 < y < 1/4).$$

(c) Derive the pdf of $W = (X - a)(b - X)$.

3.5. Suppose $X \sim$ geometric(p), where $0 < p < 1$; i.e., the probability mass function (pmf) of X is

$$f_X(x|p) = \begin{cases} (1-p)^{x-1}p, & x = 1, 2, 3, ..., \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show

$$P_X(X > t + s | X > t) = P_X(X > s),$$

for positive integers t and s, that is, X has the *memoryless property.*
(b) Show X has pmf in the exponential family.

3.6. Suppose $X \sim$ exponential(β); i.e., the probability density function of X is

$$f_X(x|\beta) = \frac{1}{\beta} e^{-x/\beta} I(x > 0).$$

(a) Show $\{f_X(x|\beta); \beta > 0\}$ is a scale family.
(b) Derive the moment generating function (mgf) of $Y = 2X/\beta$, making certain to note the values at which this mgf is defined.

3.7. The folded normal distribution in Exercise 3.20 (CB, pp 131) can be derived from the standard normal distribution.
(a) Suppose $Z \sim \mathcal{N}(0, 1)$. Show that $X = |Z|$ has a distribution identified by the pdf in Exercise 3.20.
(b) Write an R simulation to approximate the distribution of

$$Y = g(X) = \log_{10} |\sin 2\sqrt{X}|$$

and also to approximate $E(Y)$ and var(Y). In this example, you can see the value of using simulation (as deriving the pdf of Y in closed form might be difficult or impossible).

3.8. Calculate the skewness ξ and kurtosis κ for
(a) $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$
(b) $X \sim \text{gamma}(\alpha, \beta)$, $\alpha, \beta > 0$.
Discuss the different pmf/pdf shapes for each family. Note that Casella and Berger denote the skewness and kurtosis by α_3 and α_4, respectively. See Exercise 2.28 (pp 79-80).

3.9. Suppose $X \sim \text{beta}(a, b)$, where $a > 0$ and $b > 0$. Define

$$Y = \frac{bX}{a(1-X)}.$$

(a) Derive the pdf of Y. We will see this distribution in Chapter 5.
(b) Suppose $b > 1$. Find $E(Y)$. Can this constraint on b be removed? Discuss why or why not.
(c) Show that the mgf of Y does not exist.

3.10. Suppose Z is a continuous random variable with pdf

$$f_Z(z) = \text{sech}(\pi z) I(z \in \mathbb{R}),$$

where $\text{sech}(\cdot)$ is the hyperbolic secant function.
(a) Show $f_Z(z)$ is a valid pdf.
(b) With $-\infty < \mu < \infty$ and $\sigma > 0$, derive the pdf of $X = \sigma Z + \mu$. Describe the collection of pdfs that arise.
(c) Pick two values of (μ, σ) and graph the corresponding pdfs on the same graph (one by changing the location parameter from zero and the other from changing the scale parameter from unity). Include the $(0, 1)$ case on the graph also.
(d) Calculate $E(X)$ and $\text{var}(X)$.

3.11. Suppose X is a continuous random variable with pdf

$$f_X(x|\eta, \psi) = \exp\left\{ \frac{x\eta - D(\eta)}{\psi} + c(x, \psi) \right\} I(x \in \mathcal{X}),$$

where the support $\mathcal{X}$ is free of all parameters, η and ψ are parameters, and $D(\eta)$ and $c(x, \psi)$ are real-valued functions. Assume that the conditions under which one can interchange derivatives and integrals are satisfied. Show that $E(X) = D'(\eta)$ and $\text{var}(X) = \psi D''(\eta)$, assuming these derivatives exist.

3.12. Suppose the continuous random variable X has probability density function

$$f_X(x|\mu, \sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x - \mu|}{\sigma} \right), \quad x \in \mathbb{R},$$

where $-\infty < \mu < \infty$ and $\sigma > 0$.
(a) Show that $\{f_X(x|\mu, \sigma); -\infty < \mu < \infty,\ \sigma > 0\}$ is a location-scale family and identify the standard density.

(b) Does X have pdf in the exponential family? Prove any claims you make.
(c) For $r > 0$, calculate

$$P_X\left(\frac{|X-\mu|}{\sigma} > r\right)$$

exactly. Compare this calculation with Markov's upper bound for this probability.

3.13. Suppose X is a continuous random variable with probability density function (pdf)

$$f_X(x|\theta) = \frac{1}{\theta}e^{-x/\theta}\exp(-e^{-x/\theta}), \ -\infty < x < \infty,$$

where $\theta > 0$ is an unknown parameter.
(a) Show $\{f_X(x|\theta); \theta > 0\}$ is a scale family. Identify the scale parameter and the standard pdf.
(b) Derive the pdf of

$$Y = g(X) = e^{-X/\theta}$$

and identify the distribution of Y.

3.14. Suppose X is a continuous random variable with probability density function (pdf)

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

where $\theta > 0$ is an unknown parameter.
(a) Show that $\{f_X(x|\theta); \theta > 0\}$ is an exponential family. Is the family full or curved?
(b) Find $E(X)$ and $\text{var}(X)$.
(c) Show that $Y = X/\theta$ has a gamma distribution. Identify the shape and scale parameters α and β.

3.15. Suppose X is a discrete random variable with probability mass function $f_X(x)$ given by

$$f_X(x) = pf_1(x) + (1-p)f_2(x),$$

where

$$\begin{aligned} f_1(x) &= I(x=0) \\ f_2(x) &= \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, ... \end{aligned}$$

$0 < p < 1$, and $\lambda > 0$. In other words, X is a mixture of a degenerate distribution (at zero) and the usual Poisson distribution with mean λ.
(a) Show that

$$\begin{aligned} P(X=0) &= p + (1-p)e^{-\lambda} \\ P(X=x) &= (1-p)\frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 1, 2, ... \end{aligned}$$

(b) Derive $E(X)$ and $\text{var}(X)$.
(c) Derive $M_X(t)$, the moment generating function of X. What happens to $M_X(t)$ when $p \to 0$? Comment on what this implies.

(d) I used R to generate an iid sample of size $n = 30$ from $f_X(x)$:

```
0 17  0  0 25  0  0 15  0  0  0  0 15  0  0  0  0  0  0  0  0 18  0 16 20  0 23  0  0  0
```

What values of p and λ do you think I used to perform the simulation? Just guess and justify your answers. Don't do anything fancy here (you should be able to make an educated guess).

3.16. Suppose $X \sim \text{nib}(r, p)$, where $r \geq 1$ is known and p is unknown. Recall that the probability mass function (pmf) of X is

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

where $0 < p < 1$, and that the moment generating function (mgf) of X is

$$M_X(t) = \left(\frac{pe^t}{1 - qe^t} \right)^r,$$

for $t < -\ln q$, where $q = 1 - p$. In this model, X records the number of Bernoulli trials until the rth success is observed.
(a) Show that X has pmf in the exponential family.
(b) Find the pmf of $Y = X - r$.
(c) Define $Z_p = 2pY$, and let $M_{Z_p}(t)$ denote the mgf of Z_p. Show that

$$\lim_{p \to 0} M_{Z_p}(t) = \left(\frac{1}{1 - 2t} \right)^r,$$

for all $t < 1/2$. In the light of this result, describe what happens to the cumulative distribution function (cdf) of Z_p as $p \to 0$.

CHAPTER 4:

4.1. Suppose $\mathbf{X} = (X_1, X_2)'$ is a bivariate random vector with pdf

$$f_{X_1,X_2}(x_1, x_2) = x_1 e^{-x_2} I(0 < x_1 < x_2 < \infty).$$

(a) Find $M_{X_1,X_2}(t_1, t_2)$, the joint moment generating function of X_1 and X_2. Does

$$M_{X_1,X_2}(t_1, t_2) = M_{X_1,X_2}(t_1, 0) M_{X_1,X_2}(0, t_2)$$

for all t_1 and t_2 where $M_{X_1,X_2}(t_1, t_2)$ exists?
(b) Find the marginal distributions of X_1 and X_2.
(c) Are X_1 and X_2 independent?
(d) Repeat the parts above with

$$f_{X_1,X_2}(x_1, x_2) = \left(\frac{1}{2} \right)^{x_1 + x_2} I(x_1 = 1, 2, 3, ...) I(x_2 = 1, 2, 3, ...).$$

4.2. Suppose $(X, Y)'$ is a bivariate random vector with pdf

$$f_{X,Y}(x, y) = 2\exp\{-(x + y)\}I(0 < x < y < \infty).$$

(a) Explain why X and Y are dependent.
(b) Find $E(Y|X = x)$ and $\text{var}(Y|X = x)$. Your answers should depend on x.
(c) Derive the distribution of $E(Y|X)$.
(d) Derive the conditional moment generating function of Y, given $X = x$. Differentiate this mgf and evaluate it at $t = 0$; you should get $E(Y|X = x)$.

4.3. Consider the random vector $(X, Y)'$ with joint pdf

$$f_{X,Y}(x, y) = |x - y|e^{-(x+y)}I(x > 0)I(y > 0).$$

(a) Verify this is a valid (joint) pdf.
(b) Calculate $P(0 < Y < X < 1)$.
(b) Derive the (joint) mgf of $(X, Y)'$.
(c) Are X and Y independent?

4.4. Suppose $W_1 \sim \chi^2(\nu_1)$ and $W_2 \sim \chi^2(\nu_2)$. If W_1 and W_2 are independent, show the pdf of

$$W = \frac{W_1/\nu_1}{W_2/\nu_2}$$

is given by

$$f_W(w) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} w^{(\nu_1-2)/2}I(w > 0)}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})\left(1 + \frac{\nu_1 w}{\nu_2}\right)^{(\nu_1+\nu_2)/2}}.$$

We say W has an F *distribution* with ν_1 (numerator) and ν_2 (denominator) degrees of freedom. This is denoted by $W \sim F(\nu_1, \nu_2)$.

4.5. Suppose X_1 and X_2 are continuous random variables with joint pdf $f_{X_1,X_2}(x_1, x_2)$, for $x_1, x_2 \in \mathbb{R}$. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_2$.
(a) Show the marginal pdf of $Y = X_1 + X_2$ can be written as

$$f_{Y_1}(y_1) = \int_{\mathbb{R}} f_{X_1,X_2}(y_1 - y_2, y_2)dy_2.$$

This is called the *convolution formula*. Compare this with Theorem 5.2.9 (CB, pp 215). What's the difference between these two results?
(b) Use the convolution technique to derive the pdf of $Y_1 = X_1 + X_2$ if

$$f_{X_1,X_2}(x_1, x_2) = \exp\{-(x_1 + x_2)\}I(x_1 > 0, x_2 > 0).$$

You should see that $Y_1 \sim \text{gamma}(2, 1)$.
(c) How many other ways can you prove the result in part (b)? I can immediately think of 3 other ways (maybe more?). Derive the result in part (b) in as many different ways that you can think of.

4.6. Consider the following hierarchical model:

$$\begin{aligned} X|Y &\sim \text{exponential}(1/Y) \\ Y &\sim \text{gamma}(a, b). \end{aligned}$$

(a) Find the pdf of X.
(b) Calculate $E(X)$ and $\text{var}(X)$. Note any restrictions on the values of a and b.
(c) Derive $f_{Y|X}(y|x)$ and $E(Y|X = x)$.

4.7. Suppose X_1, X_2, X_3 are (mutually) independent and identically distributed random variables, each with common pdf $f_X(x) = e^{-x}I(x > 0)$. Define

$$\begin{aligned} Y_1 &= \frac{X_1}{X_1 + X_2} \\ Y_2 &= \frac{X_1 + X_2}{X_1 + X_2 + X_3} \end{aligned}$$

and $Y_3 = X_1 + X_2 + X_3$. Show that Y_1, Y_2, and Y_3 are mutually independent and determine each marginal distribution.

4.8. Suppose $\mathbf{X} = (X_1, X_2)'$ is a bivariate random vector with (joint) moment-generating function $M_{X_1,X_2}(t_1, t_2)$.
(a) Show that the covariance of X_1 and X_2 can be computed using

$$\begin{aligned} \text{cov}(X_1, X_2) = &\left.\frac{\partial^2 M_{X_1,X_2}(t_1, t_2)}{\partial t_1 \partial t_2}\right|_{t_1=0,t_2=0} \\ &- \left[\left.\frac{\partial M_{X_1,X_2}(t_1, t_2)}{\partial t_1}\right|_{t_1=0,t_2=0}\right] \left[\left.\frac{\partial M_{X_1,X_2}(t_1, t_2)}{\partial t_2}\right|_{t_1=0,t_2=0}\right]. \end{aligned}$$

(b) Let $\psi(t_1, t_2) = \ln M_{X_1,X_2}(t_1, t_2)$. Show that

$$\text{cov}(X_1, X_2) = \left.\frac{\partial^2 \psi(t_1, t_2)}{\partial t_1 \partial t_2}\right|_{t_1=0,t_2=0}.$$

(c) Suppose $\mathbf{X} = (X_1, X_2)'$ has a bivariate normal distribution with marginal means μ_1 and μ_2, marginal variances σ_1^2 and σ_2^2, and correlation ρ. The moment-generating function of $\mathbf{X}$ is

$$M_{\mathbf{X}}(\mathbf{t}) = \exp(\mathbf{t}'\boldsymbol{\mu} + \mathbf{t}'\mathbf{V}\mathbf{t}/2),$$

where $\mathbf{t} = (t_1, t_2)'$, $\boldsymbol{\mu} = (\mu_1, \mu_2)'$, and

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix},$$

where $\sigma_{12} = \sigma_{21} = \text{cov}(X_1, X_2)$. Show that, in fact, one does get $\text{cov}(X_1, X_2) = \sigma_{12}$ using the result in part (b).

4.9. Suppose X_1 and X_2 are independent $\mathcal{U}(0,1)$ random variables.
(a) Derive the pdf of $\overline{X} = (X_1 + X_2)/2$ and calculate $P(\overline{X} > 0.75)$.
(b) Calculate

$$E\left(\frac{X_1}{\overline{X}}\right).$$

(c) Calculate $E(X_1|\overline{X})$, $\text{var}(X_1|\overline{X})$, and $\text{cov}(X_1, \overline{X})$.

4.10. Consider the random vector $(X,Y)'$ with probability density function (pdf)

$$f_{X,Y}(x,y) = 2x^{-1}e^{-2x}I(0 < y < x < \infty).$$

(a) Verify this is a valid pdf.
(b) Calculate $P(Y > X/2, X < 1)$.
(c) Calculate the correlation of X and Y.

4.11. Suppose $(X,Y)'$ has the following probability density function (pdf)

$$f_{X,Y}(x,y) = (xy)^{-2}I(x > 1)I(y > 1).$$

(a) Are X and Y independent? Explain.
(b) Find the joint pdf of $U = XY$ and $V = X/Y$.
(c) Find the marginal pdfs of U and V. Are U and V independent?

4.12. Suppose $X_1, X_2, ..., X_n$ are mutually independent random variables. Each random variable, marginally, follows a standard normal distribution.
(a) Derive the distribution of $S = n^{-1}(X_1 + X_2 + \cdots + X_n)$.
(b) Derive the distribution of $T = X_1^2 + X_2^2 + \cdots + X_n^2$.
(c) Derive the moment-generating functions (mgf) of

$$U = \sqrt{n}S \qquad \text{and} \qquad V = \frac{T-n}{\sqrt{2n}}.$$

(Derive each one separately; I am not asking for a "joint" mgf here). What does each mgf converge to as $n \to \infty$?

4.13. Suppose X and Y are random variables with finite means and variances. We want to predict Y as a linear function of X. We are interested only in functions of the form $Y = \beta_0 + \beta_1 X$, for fixed constants β_0 and β_1. Define the *mean squared error of prediction* by

$$Q(\beta_0, \beta_1) \equiv E\{[Y - (\beta_0 + \beta_1 X)]^2\}.$$

(a) Show that $Q(\beta_0, \beta_1)$ is minimized when

$$\beta_1 = \rho\left(\frac{\sigma_Y}{\sigma_X}\right)$$

and

$$\beta_0 = E(Y) - \beta_1 E(X),$$

where $\rho = \text{corr}(X, Y)$.

(b) Calculate the values of β_0 and β_1 for the bivariate distribution in Problem 4.10 (last page). Does $\beta_0 + \beta_1 X$ equal $E(Y|X)$ in that problem? Comment.

4.14. A philanthropist decides to choose s persons at random and to invite them to a series of parties over the year. A party is given on a particular day if at least one person (among the s) has a birthday on that day (multiple parties cannot occur on the same day). Assume there are exactly 365 days in the year and that birthdays occur at random across days. Define

$$X_i = \begin{cases} 1, & \text{if a party is given on day } i \\ 0, & \text{otherwise,} \end{cases}$$

for $i = 1, 2, ..., 365$. Therefore, $X_1, X_2, ..., X_{365}$ is a sequence of 0-1 random variables.
(a) What is the distribution of X_1?
(b) Compute $\text{cov}(X_i, X_j)$, where $i \neq j$.
(c) Let T denote the number of parties that will be given over the year. Find the mean and variance of T.

4.15. Suppose U_1 and U_2 are independent random variables with $U_1 \sim \chi^2_{n_1}(\lambda)$ and $U_2 \sim \chi^2_{n_2}$. That is, U_1 is distributed as noncentral χ^2 with degrees of freedom $n_1 > 0$ and noncentrality parameter $\lambda > 0$, and U_2 is distributed as (central) χ^2 with degrees of freedom $n_2 > 0$.
(a) Recall that the distribution of U_1 arises as a mixture in the following hierarchy:

$$\begin{aligned} U_1|Y &\sim \chi^2_{n_1+2Y} \\ Y &\sim \text{Poisson}(\lambda). \end{aligned}$$

Show that $E(U_1) = n_1 + 2\lambda$ and $\text{var}(U_1) = 2n_1 + 8\lambda$.
(b) We know the random variable

$$W = \frac{U_1/n_1}{U_2/n_2}$$

has a non-central F distribution, seen in Chapter 5. Show that

$$E(W) = \frac{n_2}{n_2 - 2}\left(1 + \frac{2\lambda}{n_1}\right)$$

for $n_2 > 2$. Make sure you explain why $n_2 > 2$.

4.16. Let $F_X : \mathbb{R} \to [0,1]$ and $F_Y : \mathbb{R} \to [0,1]$ be univariate cumulative distribution functions (cdfs) and suppose $-1 \leq \alpha \leq 1$. Define $F^{(\alpha)}_{X,Y} : \mathbb{R}^2 \to [0,1]$ by

$$F^{(\alpha)}_{X,Y}(x,y) = F_X(x)F_Y(y)\{1 + \alpha[1 - F_X(x)][1 - F_Y(y)]\}.$$

The collection $\{F^{(\alpha)}_{X,Y} : -1 \leq \alpha \leq 1\}$ is called the *Farlie-Morgenstern family* of bivariate cdfs corresponding to F_X and F_Y.
(a) Starting with the expression for $F^{(\alpha)}_{X,Y}(x,y)$, show the marginal cdfs of X and Y are given by $F_X(x)$ and $F_Y(y)$, respectively.

(b) What value of α corresponds to X and Y being independent? Explain.
(c) As a special case, suppose X and Y are each distributed as exponential with mean 1. In this case, for any $\alpha \in [-1, 1]$, show that the joint probability density function of $(X, Y)'$ is

$$f_{X,Y}^{(\alpha)}(x, y) = \{1 + \alpha[(1 - 2e^{-x})(1 - 2e^{-y})]\}e^{-(x+y)}I(x > 0)I(y > 0).$$

4.17. Let α be a positive constant. Let $(X, Y)'$ be a bivariate random vector that is uniformly distributed on the circle with center $(0, 0)$ and radius α. That is, the probability density function (pdf) of $(X, Y)'$ is

$$f_{X,Y}(x, y) = \frac{1}{\pi\alpha^2} I(x^2 + y^2 < \alpha^2).$$

Define $R = \sqrt{X^2 + Y^2}$ and $S = Y/X$.
(a) Derive the cumulative distribution function (cdf) of R.
(b) Find the pdf of R.
(c) Show that R and S are independent.
(d) What is the geometrical interpretation of the two variables R and S?

4.18. Suppose the random vector $(X, Y)'$ has probability density function:

$$f_{X,Y}(x, y) = \frac{e^{-x/y}e^{-y}}{y} I(x > 0, y > 0).$$

(a) Find $E(X|Y = y)$.
(b) Find $\text{corr}(X, Y)$.

4.19. Let X denote the concentration of a certain substance based on one trial of an experiment, and let Y denote the concentration of the substance based on a second trial. Assume that the joint probability density function of (X, Y) is given by $f_{X,Y}(x, y) = 4xy$, if $0 < x < 1$ and $0 < y < 1$, and $f_{X,Y}(x, y) = 0$ otherwise.
(a) Derive the joint cumulative distribution function of (X, Y).
(b) Find the probability that the average concentration from the two trials is less than 0.5.
(c) Find the probability that the concentration from the first trial is less than 0.5.

4.20. Suppose that

$$X|Y = y \sim \text{binomial}(y, p),$$

where Y denotes the number of trials in a binomial experiment and p is the probability of "success" for each trial. Assume that, marginally, $Y \sim \text{nib}(r, \pi)$, where Y is defined as the number of failures before the rth success, and π is the probability of "success" for each trial in the negative binomial experiment. Show that, marginally,

$$X \sim \text{nib}(r, \pi/\{1 - (1 - p)(1 - \pi)\}).$$

That is, show that X records the number of failures before the rth success and the (marginal) probability of success is $\pi/\{1 - (1 - p)(1 - \pi)\}$. *Hint:* One can try to derive the marginal probability mass function of X (this should be doable). You could also try to derive the marginal moment generating function (mgf) of X. Since the mgf is an expectation, you could use the iterated expectation technique.

4.21. Suppose X is a random variable with $E(X) = -1$, $\text{var}(X) = 1$ and $P(X < 0) = 1$. Argue that $E(X^3) < -1$. You must argue that the inequality is strict.

4.22. Suppose X has a geometric distribution; i.e., the probability mass function (pmf) of X is

$$f_X(x|p) = \begin{cases} (1-p)^{x-1}p, & x = 1, 2, 3, ..., \\ 0, & \text{otherwise}, \end{cases}$$

where $0 < p < 1$.
(a) Show that $\{f_X(x|p), 0 < p < 1\}$ is a one-parameter exponential family.
(b) Derive the moment-generating function of X.
(c) For this part only, suppose that, conditional on $P = p$, the random variable X follows a geometric distribution. In other words, the first level of the corresponding hierarchical model is

$$X|P = p \sim f_{X|P}(x|p),$$

where the conditional pmf, now written $f_{X|P}(x|p)$, is given above. The second level of the hierarchical model is $P \sim \text{beta}(a, b)$, where $a > 1$ is known and $b = 1$. Derive the marginal pmf of X and find $E(X)$ under this hierarchical model.
(d) Describe a real application where the hierarchical model in part (c) might be useful.

4.23. Consider the random vector $(X, Y)'$ with joint probability density function

$$f_{X,Y}(x, y) = \begin{cases} c(1 + xy), & -1 < x < 1, -1 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show that $c = 1/4$.
(b) Calculate $P(Y > X^2)$.
(c) Derive the marginal probability density functions of X and Y. Show that X and Y are not independent.
(d) Even though X and Y are not independent, it turns out that X^2 and Y^2 are independent! Prove this. *Hint*: Work with $P(X^2 \leq u, Y^2 \leq v)$, the joint cdf of (X^2, Y^2).

4.24. Consider the random vector $(X, Y)'$ with joint probability density function

$$f_{X,Y}(x, y) = \begin{cases} xe^{-x(1+y)}, & x > 0, y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show that $E(Y)$ does not exist but that $E(Y|X = x) = 1/x$.
(b) Find the variance of $E(X|Y)$.

4.25. Suppose X_1, X_2, and X_3 are random variables with joint probability density function (pdf)

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = 6e^{-x_1-x_2-x_3} I(0 < x_1 < x_2 < x_3 < \infty).$$

Define the random variables

$$U_1 = X_1, \quad U_2 = X_2 - X_1, \quad \text{and} \quad U_3 = X_3 - X_2.$$

(a) Derive the joint pdf of $\mathbf{U} = (U_1, U_2, U_3)'$. Identify the marginal distributions of U_1, U_2, and U_3 separately.
(b) Find the mean and variance of $W = -2U_1 + 3U_2 - U_3$.

4.26. Suppose $X_1, X_2, ..., X_n$ are mutually independent $b(m_i, p_i)$ random variables; i.e., the probability mass function of X_i is given by

$$f_{X_i}(x_i|p_i) = \binom{m_i}{x_i} p_i^{x_i}(1-p_i)^{m_i - x_i}, \quad x_i = 0, 1, 2, ..., m_i,$$

for $i = 1, 2, ..., n$. The success probabilities $0 < p_i < 1$ are unknown parameters. The number of trials m_i are fixed and known.
(a) If $p_1 = p_2 = \cdots = p_n = p$, derive the distribution of $Z = \sum_{i=1}^n X_i$. What can you say about the distribution of Z is the p_i's are different?
(b) Suppose the success probability associated with X_i is

$$p_i = \frac{\exp(a + bw_i)}{1 + \exp(a + bw_i)}, \quad i = 1, 2, ..., n,$$

where the w_i's are fixed constants and $a \in \mathbb{R}$ and $b \in \mathbb{R}$ are unknown parameters. Show that the pmf of X_i is a two-parameter exponential family.
(c) For this part only, suppose $n = 2$ and $p_1 = p_2 = p$. Derive the (conditional) probability mass function of X_1 given $Z = X_1 + X_2 = r$, where $r > 0$.

4.27. Suppose $(X, Y)'$ is a continuous random vector with joint probability density function (pdf)

$$f_{X,Y}(x, y) = cx^2 y \; I(0 < 2x < y < 1).$$

(a) Find c that makes $f_{X,Y}(x, y)$ a valid pdf.
(b) Calculate $P(X^2 > Y)$.
(c) Find both marginal pdfs.
(d) Find $E(X|Y)$.

4.28. Consider the hierarchical model

$$\begin{aligned} X|Y = y &\sim \text{geometric}(y) \\ Y &\sim \text{beta}(a, b), \end{aligned}$$

where $a > 0$ and $b > 0$. In the first model, conditional on $Y = y$, the random variable X counts the number of Bernoulli trials until the first success is observed. That is,

$$f_{X|Y}(x|y) = \begin{cases} (1-y)^{x-1} y, & x = 1, 2, 3, ..., \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the (marginal) probability mass function (pmf) of X.
(b) Calculate $E(X)$ and $\text{var}(X)$. Note any additional restrictions on the values of a and b.

4.29. Suppose $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$, and $X \perp\!\!\!\perp Y$. Define

$$\begin{aligned} U &= X+Y \\ V &= X^2+Y^2. \end{aligned}$$

(a) Show the joint moment-generating function of U and V is given by

$$M_{U,V}(s,t) = E(e^{sU+tV}) = \frac{1}{1-2t}\exp\left(\frac{s^2}{1-2t}\right), \quad \text{for } t < 1/2.$$

(b) Find $\text{corr}(U,V)$.
(c) Are U and V independent? Explain.

4.30. Particles are subject to collisions that cause them to split into two parts with each part a fraction of the parent. Suppose that this fraction X is uniformly distributed over $(0,1)$. Following a single particle through several splits, we obtain a fraction of the original particle

$$Z = \prod_{i=1}^{n} X_i.$$

Assume $X_1, X_2, ..., X_n$ are mutually independent and $X_i \sim \mathcal{U}(0,1)$, for $i = 1, 2, ..., n$.
(a) Show that the probability density function (pdf) of Z is given by

$$f_Z(z) = \frac{1}{(n-1)!}(-\ln z)^{n-1} I(0 < z < 1).$$

(b) Calculate $\text{var}(Z)$.

4.31. Consider the random vector $(X,Y)'$ with joint probability density function

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{e^{-y}}{y}, & 0 < x < y < \infty \\ 0, & \text{otherwise.} \end{cases}$$

(a) Write $P(Y < X^2)$ as a double integral. You do not have to calculate the integral, but I want to see a good picture explaining where the limits come from.
(b) Find $E(X)$ and $E(Y)$.
(c) Calculate the covariance of X and Y.

4.32. Suppose $X \sim \text{beta}(\alpha, \beta)$, where $\alpha > 0$ and $\beta > 0$. Recall that the probability density function (pdf) of X is

$$f_X(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1} I(0 < x < 1).$$

Throughout this problem, we will consider the subfamily of beta distributions where $\alpha = \beta$.
(a) Describe, geometrically, the parameter space for members of the subfamily; e.g., draw a picture, etc.

(b) Suppose $k > 0$. Show that the kth central moment of X is

$$\mu_k = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)} \sum_{j=0}^{k} \binom{k}{j} \frac{\Gamma(\alpha + j)}{\Gamma(2\alpha + j)} \left(-\frac{1}{2}\right)^{k-j}.$$

(c) Show that $X \stackrel{d}{=} 1 - X$, where recall "$\stackrel{d}{=}$" means "equal in distribution."
(d) Find $\text{corr}(X, 1 - X)$.

4.33. Suppose the continuous random variable X, conditional on $P = p$, has a two-component normal mixture distribution; specifically, $X|P = p \sim f_X(x|p)$, where the probability density function (pdf)

$$f_X(x|p) = pf_0(x) + (1 - p)f_1(x),$$

and where the component densities

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{and} \quad f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2},$$

where $-\infty < \mu < \infty$. Both $f_0(x)$ and $f_1(x)$ are defined for all $x \in \mathbb{R}$. In turn, the random variable $P \sim \text{beta}(\alpha, 1)$, where $\alpha > 0$ is known.
(a) Find the marginal pdf of X.
(b) Using the marginal pdf of X in part (a) or otherwise, show that $E(X) = \mu/(\alpha + 1)$.
(c) Suppose that, after observing X, you need to make a decision about which distribution X comes from: $f_0(x)$ or $f_1(x)$. Formulate a rule that will enable you to do this, discuss why your rule makes sense, and then investigate the characteristics of your rule (e.g., how often you make the correct decision, etc.).

4.34. Suppose $(X, Y)'$ is a continuous random vector with joint probability density function (pdf)

$$f_{X,Y}(x, y) = \begin{cases} 2y, & 0 < x < 1,\ x < y < 2x \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the marginal pdf of X and find $P(X > 1/2)$. Can you calculate this probability by using the joint distribution $f_{X,Y}(x, y)$?
(b) Find the marginal pdf of Y.
(c) Find $\text{cov}(X, Y)$.

CHAPTER 5:

5.1. *Gamma distribution fun.*
(a) Suppose $X \sim \text{gamma}(\alpha, \beta)$. Show that

$$Y = \frac{2X}{\beta} \sim \chi^2_{2\alpha}.$$

(b) Suppose $X_1 \sim \text{gamma}(\alpha_1, \beta_1)$, $X_2 \sim \text{gamma}(\alpha_2, \beta_2)$, and X_1 and X_2 are independent. Find a function of X_1 and X_2 that has an F distribution. What are the degrees of freedom?
(c) Suppose $X_1, X_2, ..., X_n$ are iid exponential(β_1) and $Y_1, Y_2, ..., Y_m$ are iid exponential(β_2).

Suppose the samples are independent. Under the assumption that $\beta_1 = \beta_2$, find a function of $\overline{X}$ and $\overline{Y}$ that has an F distribution. What are the degrees of freedom?

5.2. *Normal distribution fun.* Suppose $X_1, X_2, ..., X_n$ is an iid sample from a $\mathcal{N}(\mu, \sigma^2)$ distribution. Let $\overline{X}$ and S^2 denote the sample mean and sample variance, respectively.
(a) Find the distribution of

$$Q = \frac{(\overline{X} - \mu)^2}{\sigma^2/n} + \frac{(n-1)S^2}{\sigma^2}.$$

(b) Determine the value of c that satisfies

$$P\left(-c < \frac{S}{\overline{X}} < c\right) = 1 - \alpha,$$

for $\alpha \in (0, 1)$. The statistic $T(\mathbf{X}) = S/\overline{X}$ is called the *coefficient of variation* (it is a measure of variation, relative to the mean of the distribution).
(c) Suppose X_{n+1} is an additional observation, which is distributed as $\mathcal{N}(\mu, \sigma^2)$ and is independent of $X_1, X_2, ..., X_n$ (i.e., X_{n+1} is a "new" observation). Prove that

$$T = \frac{X_{n+1} - \overline{X}}{S\sqrt{1 + 1/n}} \sim t(n-1).$$

5.3. *Noncentral F distribution fun.* The univariate random variable W is said to have a noncentral F distribution with degrees of freedom $n_1 > 0$ and $n_2 > 0$ and noncentrality parameter $\lambda > 0$ if it has the pdf

$$f_W(w|n_1, n_2, \lambda) = \sum_{j=0}^{\infty} \left(\frac{e^{-\lambda}\lambda^j}{j!}\right) \frac{\Gamma(\frac{n_1+2j+n_2}{2}) \left(\frac{n_1+2j}{n_2}\right)^{(n_1+2j)/2} w^{(n_1+2j-2)/2}}{\Gamma(\frac{n_1+2j}{2})\Gamma(\frac{n_2}{2}) \left(1 + \frac{n_1 w}{n_2}\right)^{(n_1+2j+n_2)/2}} I(w > 0).$$

We write $W \sim F_{n_1,n_2}(\lambda)$. When $\lambda = 0$, the noncentral F distribution reduces to the central F distribution.
(a) If U_1 and U_2 are independent random variables with $U_1 \sim \chi^2_{n_1}(\lambda)$ and $U_2 \sim \chi^2_{n_2}$, show that

$$W = \frac{U_1/n_1}{U_2/n_2} \sim F_{n_1,n_2}(\lambda).$$

Hint: Introduce $U = U_1 + U_2$ as a dummy variable and perform a bivariate transformation.
(b) If $W \sim F_{n_1,n_2}(\lambda)$, show that

$$E(W) = \frac{n_2}{n_2 - 2}\left(1 + \frac{2\lambda}{n_1}\right)$$

and

$$\text{var}(W) = \frac{2n_2^2}{n_1^2(n_2-2)}\left\{\frac{(n_1+2\lambda)^2}{(n_2-2)(n_2-4)} + \frac{n_1+4\lambda}{n_2-4}\right\}.$$

Note that $E(W)$ exists only when $n_2 > 2$ and $\text{var}(W)$ exists only when $n_2 > 4$. The moment generating function for the noncentral F distribution does not exist in closed form.
(c) Suppose that $W \sim F_{n_1,n_2}(\lambda)$. For fixed n_1, n_2, and $c > 0$, show the quantity $P(W > c)$

is a strictly increasing function of λ. That is, if $W_1 \sim F_{n_1,n_2}(\lambda_1)$ and $W_2 \sim F_{n_1,n_2}(\lambda_2)$, where $\lambda_2 > \lambda_1$, then $P(W_2 > c) > P(W_1 > c)$. *Note:* That the noncentral F distribution tends to be larger than the central F distribution is the basis for many tests used in linear models. Typically, test statistics will have a central F distribution when H_0 is true and a noncentral F distribution when H_1 is true. Because the noncentral F distribution tends to be larger, large values of a test statistic are consistent with H_1. Thus, the form of an appropriate rejection region is to reject H_0 for large values of the test statistic. The power is simply the probability of rejection region (defined under H_0) when the probability distribution is noncentral F.

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

$$f_X(x|\mu) = e^{-(x-\mu)} I(x > \mu).$$

(a) Show that $\{f_X(x|\mu); -\infty < \mu < \infty\}$ is a location family.
(b) Show that the probability density function of the minimum order statistic $X_{(1)}$ is

$$f_{X_{(1)}}(x|\mu) = ne^{-n(x-\mu)} I(x > \mu).$$

(c) Show that $X_{(1)}$ converges in probability to μ.

5.5. Suppose $X_1, X_2, ..., X_n$ is an iid sample from $f_X(x) = e^{-x} I(x > 0)$.
(a) Show that

$$\begin{aligned} Z_1 &= nX_{(1)} \\ Z_2 &= (n-1)(X_{(2)} - X_{(1)}) \\ Z_3 &= (n-2)(X_{(3)} - X_{(2)}) \\ &\vdots \\ Z_n &= X_{(n)} - X_{(n-1)} \end{aligned}$$

are mutually independent and that each Z_i follows an exponential distribution marginally.
(b) Demonstrate that all linear combinations of $X_{(1)}, X_{(2)}, ..., X_{(n)}$ can be expressed as a linear combination of independent random variables.

5.6. Suppose $Z_1, Z_2, ..., Z_6$ is an iid sample from a $\mathcal{N}(0, 1)$ distribution. Suppose $Z_7 \sim \mathcal{N}(0, 1)$ 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) $V = (Z_1^2 + Z_2^2 + Z_3^2)/(Z_4^2 + Z_5^2 + Z_6^2)$.

5.7. Let $X \sim F_{r,s}$; i.e., the probability density function (pdf) of X is

$$f_X(x) = \frac{\Gamma(\frac{r+s}{2}) \left(\frac{r}{s}\right)^{r/2} x^{(r-2)/2}}{\Gamma(\frac{r}{2})\Gamma(\frac{s}{2}) \left(1 + \frac{rx}{s}\right)^{(r+s)/2}} I(x > 0).$$

(a) Derive the pdf of $Y = 1/X$. Do not just state the answer.
(b) Carefully argue that $X \xrightarrow{p} 1$, as $\min\{r, s\} \to \infty$.

5.8. 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 $3T_1 - 2T_2$?
(b) Find a statistic that has a t_8 distribution.
(c) Find a statistic that has an $F_{4,6}$ distribution.

5.9. Suppose that X and Y are independent random variables and let $Z = X + Y$. Use the method of convolution to derive the pdf of Z if
(a) X and Y are exponential(β).
(b) X and Y are $\mathcal{N}(0, 1)$.

5.10. Suppose X_1, X_2, X_3 are iid beta$(2, 1)$.
(a) Derive the probability density function (pdf) of $X_{(1)}$.
(b) Let m denote the median of the beta$(2, 1)$ distribution. Find $P(X_{(1)} > m)$.
(c) Calculate $\text{cov}(X_{(2)}, X_{(3)})$.

5.11. Suppose $X_1, X_2, ..., X_n$ are iid exponential(1). Define the statistics

$$\begin{aligned} U_n &= 2(X_1 + X_2 + \cdots + X_n) \\ V_n &= \sqrt{n}(\overline{X} - 1). \end{aligned}$$

(a) Derive the moment generating function (mgf) of U_n. What is the distribution of U_n?
(b) Show that the mgf of V_n, for $t < \sqrt{n}$, is

$$M_{V_n}(t) = \{e^{t/\sqrt{n}} - (t/\sqrt{n})e^{t/\sqrt{n}}\}^{-n}.$$

(c) For $t < \sqrt{n}$, find $\lim_{n\to\infty} M_{V_n}(t)$.

5.12. Suppose $X_1, X_2, ..., X_n$ are iid Poisson(θ) and let $T = \sum_{i=1}^n X_i$ denote the sample sum.
(a) Find the conditional distribution of X_1 given $T = t$.
(b) Show that T has pdf in the exponential family.
(c) Show that $\overline{X}$ has pdf in the exponential family.

5.13. Suppose $X_1, X_2, ..., X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$, where $-\infty < \mu < \infty$ and $\sigma^2 > 0$.
(a) Consider the statistic cS^2, where c is a constant and S^2 is the usual sample variance (denominator $= n-1$). Find the value of c that minimizes

$$\text{MSE}(cS^2) = E[(cS^2 - \sigma^2)^2] = \text{var}(cS^2) + [E(cS^2 - \sigma^2)]^2.$$

(b) Consider the normal subfamily where $\sigma^2 = \mu^2$, where $\mu > 0$. Let S denote the sample standard deviation. Find a linear combination $c_1\overline{X} + c_2 S$ whose expectation is equal to μ. Find the variance of this linear combination.

5.14. Suppose that $Z_0, Z_1, Z_2, ...,$ is a sequence of iid standard normal random variables, and define, for $i = 1, 2, ...,$

$$X_i = Z_0 + Z_i,$$

Show that $\overline{X}_n \xrightarrow{p} Z_0$, as $n \to \infty$.

5.15. *Monte Carlo approximation.* Suppose that you want to compute

$$I = E\left[Y \sin\left(2\pi \cos Y\right)\right],$$

where $Y \sim \mathcal{U}(-\pi/2, \pi/2)$. Approximate this expectation by computing

$$\widehat{I}(Y_1, Y_2, ..., Y_n) = \frac{1}{n}\sum_{i=1}^{n} Y_i \sin\left(2\pi \cos Y_i\right),$$

for large n, where $Y_1, Y_2, ..., Y_n$ is an iid $\mathcal{U}(-\pi/2, \pi/2)$ sample. What is your approximation of I? How good do you think the approximation is? Could you use the same method to approximate $E(\tan Y)$? Why or why not?

5.16. Suppose the bivariate vector $\mathbf{Y} = (Y_1, Y_2)'$ has the following density:

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

We observe n iid copies of $\mathbf{Y}$, say, $\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n$, where $\mathbf{Y}_j = (Y_{1j}, Y_{2j})'$, for $j = 1, 2, ..., n$. Define

$$\overline{\mathbf{Y}}_n = (\overline{Y}_{1+}, \overline{Y}_{2+})' \equiv \left(\frac{1}{n}\sum_{j=1}^{n} Y_{1j}, \frac{1}{n}\sum_{j=1}^{n} Y_{2j}\right)'$$

and $\boldsymbol{\mu} = (\mu_1, \mu_2)'$, where $\mu_i = E(Y_{i1})$, for $i = 1, 2$.
(a) Find a centered and scaled version of $\overline{\mathbf{Y}}_n$ that converges in distribution to $\mathbf{Y} \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma})$, as $n \to \infty$, and compute $\boldsymbol{\Sigma}$.
(b) When $n = 30$, approximate $P(\overline{Y}_{1+} - \overline{Y}_{2+} > 0.5)$.

5.17. Suppose $\mathbf{X} = (X_1, X_2, X_3)' \sim \text{mult}(n, \mathbf{p} = \frac{1}{3}\mathbf{1})$, where $\mathbf{1} = (1, 1, 1)'$.
(a) Show that

$$\mathbf{Y}_n = \frac{\mathbf{X} - n\mathbf{p}}{\sqrt{n}} \xrightarrow{d} \mathcal{N}_3(\mathbf{0}, \boldsymbol{\Sigma}),$$

as $n \to \infty$, and compute $\boldsymbol{\Sigma}$.

(b) For $n = 30$, first approximate $P(|X_1 - X_2| \leq 1)$ using the result in (a). Then, compute this probability exactly using the multinomial pmf. How close are the two answers?

5.18. Suppose $X_1, X_2, ..., X_n$ is an iid gamma$(\alpha, 1)$ sample, and let $\overline{X}_n$ denote the sample mean. Define

$$Z_n = \frac{\sqrt{n}\left(\overline{X}_n - \alpha\right)}{\sqrt{\overline{X}_n}}.$$

(a) Prove that $Z_n \xrightarrow{d} Z$, where $Z \sim \mathcal{N}(0,1)$.
(b) Argue that

$$\left(\overline{X}_n - 1.96\sqrt{\frac{\overline{X}_n}{n}},\ \overline{X}_n + 1.96\sqrt{\frac{\overline{X}_n}{n}}\right)$$

serves as a large-sample 95 percent confidence interval for α.
(c) Show that Z_n^2 converges in distribution, and find the distribution.

5.19. Suppose $X_1, X_2, ..., X_n$ is an iid exponential(θ) sample, where $E(X_1) = \theta$ is unknown, and define $Y_i = X_i^2$, for $i = 1, 2, ..., n$.
(a) Use the CLT to derive large-sample distribution of a properly centered and scaled version of $(\overline{X}, \overline{Y})'$.
(b) Find a consistent estimator of the covariance matrix in part (a). For all intents and purposes, "consistency" means "convergence in probability."

5.20. Suppose that $X_1, X_2, ..., X_n$ is an iid sample with mean $\theta > 0$. Under a Poisson model, $E(X_1) = \text{var}(X_1) = \theta$. Therefore, a standard test of the hypothesis that

$$H_0 : X_1, X_2, ..., X_n \text{ is an iid Poisson sample}$$

uses $T = T(\mathbf{X}) = S_n^2/\overline{X}$ and rejects H_0 if T is too large. This test is good against alternatives whose variance is greater than the mean, such as the negative binomial distribution or any other mixture of Poisson distributions.
(a) Derive the large-sample distribution of T, properly centered and scaled, under the Poisson assumption.
(b) Test the Poisson model assumption with the following data, which denote the number of airborne mold spores collected from specimens of air in my office (just kidding?).

4	8	10	6	14	7	5	11	5	17

Hint: To perform the test in (b), you'll want to find a suitably normalized version of T that converges to a standard normal distribution as n gets large. Remember that your statistic can not depend on θ.

5.21. Suppose $X_1 \sim b(n, p_1)$, $X_2 \sim b(n, p_2)$, and that X_1 and X_2 are independent. Define the population odds ratio by

$$\theta = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

and the sample odds ratio by

$$\widehat{\theta} = \frac{\widehat{p}_1/(1-\widehat{p}_1)}{\widehat{p}_2/(1-\widehat{p}_2)}.$$

(a) Find the large-sample distribution of $\widehat{\theta}$, properly centered and scaled.
(b) Find the large-sample distribution of $\ln\widehat{\theta}$, properly centered and scaled.
(c) Use simulation to create qq plots of $\widehat{\theta}$ and $\ln\widehat{\theta}$ for different sample sizes n and different values of p_1 and p_2 (don't go overboard). Which statistic, $\widehat{\theta}$ or $\ln\widehat{\theta}$, seems to "converge to normality" faster?

5.22. Suppose $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ is an iid sample (with appropriate moments existing) from a distribution with correlation ρ.
(a) Show that the sample correlation coefficient r can be written as a function of (uncentered) sample moments

$$\overline{X},\ \overline{Y},\ \frac{1}{n}\sum_{i=1}^{n} X_i^2,\ \frac{1}{n}\sum_{i=1}^{n} Y_i^2,\ \text{and}\ \frac{1}{n}\sum_{i=1}^{n} X_iY_i.$$

(b) Explain how to get the asymptotic distribution of r using the multivariate Delta Method. Give as many details as possible, but you don't have to carry out the matrix multiplication.
(c) When the population distribution is bivariate normal,

$$\sqrt{n}(r-\rho) \xrightarrow{d} \mathcal{N}[0, (1-\rho^2)^2].$$

Derive a transformation g so that the asymptotic variance of $g(r)$ is free of ρ. This is called a *variance-stabilizing transformation.*

5.23. Let Y have an F distribution with r numerator and s denominator degrees of freedom; i.e., $Y \sim F(r, s)$.
(a) Show that $rY \xrightarrow{d} \chi^2(r)$, as $s \to \infty$.
(b) For specificity, assume that $Y \sim F(3, 96)$. Give an applied situation where this distribution would arise (e.g., think of one-way analysis of variance). Explain how you could use the result from part (a) to approximate $P(Y \geq 1.25)$. How good is this approximation?

5.24. Suppose $X_1, X_2, ..., X_n$ is an iid sample from a Poisson distribution with mean $\theta > 0$.
(a) Why does $\overline{X}_n \xrightarrow{p} \theta$, as $n \to \infty$? Just state the name of the result.
(b) Consider the function $g(\theta) = (\theta+1)\exp(-\theta)$. Prove that, as $n \to \infty$,

$$\sqrt{n}\left\{g(\overline{X}_n) - g(\theta)\right\} \xrightarrow{d} \mathcal{N}\left\{0, \theta^3\exp(-2\theta)\right\}.$$

(c) Find two statistics that converge in probability to $\theta^3\exp(-2\theta)$.

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

$$f_X(x|\theta) = \frac{1}{2\theta} I(0 < x < 2\theta),$$

where $\theta > 0$.
(a) Find a function of $X_{(n)}$ that converges in probability to θ, as $n \rightarrow \infty$. Prove your claim.
(b) Find a function of $\overline{X}$ that converges in probability to θ, as $n \rightarrow \infty$. Prove your claim.
(c) Find a function of $\overline{X}$ that converges in distribution to a normal distribution as $n \rightarrow \infty$. Find the mean and variance of the limiting distribution.

5.26. Suppose $X_1, X_2, ..., X_n$ is an iid sample from a $\mathcal{U}(0,1)$ distribution. Define the statistics

$$\begin{aligned}
\widehat{\theta}_A &= \frac{1}{n}\sum_{i=1}^{n} X_i \\
\widehat{\theta}_G &= \left(\prod_{i=1}^{n} X_i\right)^{1/n} \\
\widehat{\theta}_H &= \left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{X_i}\right)^{-1}.
\end{aligned}$$

(a) Which statistics converge in probability to a finite constant (as $n \rightarrow \infty$)? For those that do converge in probability, find the limiting constant. For those that do not, explain why.
(b) If possible, find a function of each statistic that converges to a standard normal distribution. Work with each statistic separately (do not combine them).

5.27. Suppose $X_1, X_2, ..., X_n$ are iid $\mathcal{N}(0, \sigma^2)$, where $\sigma^2 > 0$.
(a) Find a function of

$$T \equiv T(\mathbf{X}) = \sum_{i=1}^{n} X_i^2$$

that converges in distribution to a normal distribution as $n \rightarrow \infty$. State the mean and variance of your limiting normal distribution.
(b) Find a function of T whose large-sample variance is free of σ^2.

5.28. Suppose X is a random variable with cumulative distribution function (cdf)

$$F_X(x|\lambda, \beta) = \begin{cases} 1 - e^{-(\lambda x)^\beta}, & x > 0 \\ 0, & \text{otherwise,} \end{cases}$$

where the parameters $\lambda > 0$ and $\beta > 0$.
(a) Find the probability density function $f_X(x|\lambda, \beta)$. Does X have pdf in the exponential family? What named distribution does X have when $\beta = 1$?
(b) Derive $E(X)$ and $\text{var}(X)$.
(c) Suppose $X_1, X_2, ..., X_n$ are iid from $F_X(x|\lambda, \beta)$. Derive $F_{X_{(1)}}(x) = P(X_{(1)} \leq x)$, the cumulative distribution function of the minimum order statistic $X_{(1)}$, and calculate $E(X_{(1)})$.

5.29. Suppose X_1, X_2, X_3 are mutually independent random variables where $X_1 \sim \chi^2_{r_1}$, $X_2 \sim \chi^2_{r_2}$, and $X_3 \sim \chi^2_{r_3}$.
(a) Under what condition are X_1, X_2, X_3 iid?
(b) Derive the joint probability density function (pdf) of

$$U_1 = \frac{X_1}{X_2} \quad \text{and} \quad U_2 = X_1 + X_2.$$

(c) Show that U_1 and U_2 are independent and that $U_2 \sim \chi^2_{r_1+r_2}$.
(d) Argue that

$$\frac{X_1/r_1}{X_2/r_2} \quad \text{and} \quad \frac{X_3/r_3}{(X_1+X_2)/(r_1+r_2)}$$

are independent and that each quantity has an F distribution. Your argument can consist of a well written description (without a lot of mathematics).

5.30. Suppose $X_1, X_2, ..., X_n$ is an iid sample from a $\mathcal{N}(0, \sigma^2)$ distribution, where $\sigma^2 > 0$. Define the statistics

$$T_1 = \left| \frac{1}{n} \sum_{i=1}^{n} X_i \right| \quad \text{and} \quad T_2 = \frac{1}{n} \sum_{i=1}^{n} |X_i|.$$

(a) Derive the sampling distribution of T_1.
(b) Calculate $E(T_1)$ and $E(T_2)$ and establish an inequality between them.
(c) Calculate $\text{var}(T_2)$.

5.31. Suppose $X_1, X_2, ..., X_n$ are iid Pareto(α, β), where $\alpha > 0$ and $\beta > 0$. Recall that the Pareto probability density function (pdf) is given by

$$f_X(x) = \frac{\beta \alpha^\beta}{x^{\beta+1}}\, I(x > \alpha).$$

Let $X_{(1)}$ and $X_{(n)}$ denote the minimum and maximum order statistics, respectively.
(a) Show that $X_{(1)} \sim \text{Pareto}(\alpha, n\beta)$.
(b) Show that the cumulative distribution function (cdf) of $X_{(n)}$ is

$$F_{X_{(n)}}(x) = \left[1 - \left(\frac{\alpha}{x}\right)^\beta\right]^n, \quad x > \alpha.$$

(c) Define the sequence of real numbers $b_n = \alpha n^{1/\beta}$, for $n = 1, 2, ...,$. Determine

$$\lim_{n\to\infty} F_{X_{(n)}}(b_n x).$$

This limit turns out to be the cdf of the standard Fréchet distribution, a useful model in extreme value theory.

5.32. Suppose $X \sim \text{Poisson}(\mu)$, where $\mu > 0$. Define

$$Y = \frac{X - \mu}{\sqrt{\mu}}.$$

(a) Show that the moment generating function of Y is

$$M_Y(t) = E(e^{tY}) = \exp\{\mu e^{t/\sqrt{\mu}} - \mu - t\sqrt{\mu}\}.$$

(b) Show that you can write

$$\mu e^{t/\sqrt{\mu}} = \mu + t\sqrt{\mu} + \frac{t^2}{2} + o(1),$$

where $o(1)$ is a term that satisfies $o(1) \to 0$, as $\mu \to \infty$. *Hint:* Expand $e^{t/\sqrt{\mu}}$ in a Taylor series about $t = 0$.
(c) Combine parts (a) and (b) and show that

$$Y = \frac{X - \mu}{\sqrt{\mu}} \xrightarrow{d} \mathcal{N}(0, 1),$$

as $\mu \to \infty$.

5.33. Suppose $X_1, X_2, ..., X_n$ are iid random variables with population distribution

$$f_X(x|p) = \begin{cases} (1-p)^{x-1}p, & x = 1, 2, 3, ..., \\ 0, & \text{otherwise}, \end{cases}$$

where $0 < p < 1$. Note that this is a geometric distribution that describes the number of Bernoulli trials needed to find the first "success." Recall that $E(X) = 1/p$ and $\text{var}(X) = 1/p^2$.
(a) The statistic $\overline{X}$ converges in probability to a constant, as $n \to \infty$. Explain why and then identify the constant.
(b) The statistic $\overline{X}^2$ converges in probability to a constant, as $n \to \infty$. Explain why and then identify the constant.
(c) Carefully argue that

$$Z_n = \frac{\overline{X} - \frac{1}{p}}{\overline{X}/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1),$$

as $n \to \infty$.

5.34. Suppose $W_n \sim \chi^2_n$.
(a) Prove $W_n \stackrel{d}{=} X_1 + X_2 + \cdots + X_n$, where $X_1, X_2, ..., X_n$ are iid χ^2_1. Recall that "$\stackrel{d}{=}$" means "equal in distribution."
(b) Explain why $W_n/n \xrightarrow{p} 1$, as $n \to \infty$.
(c) Find sequences of constants a_n, b_n, and c_n such that

$$a_n(W_n/n - b_n) + c_n \xrightarrow{d} \mathcal{N}(-3, 5),$$

as $n \to \infty$. Prove your answer. *Note:* Sequences of constants can be constant sequences of constants.
(d) Suppose $V_n \sim F_{n,2n}$. Does V_n converge in probability to 1/2, 1, or 2? Or, does V_n not converge at all? Prove your claim.

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

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

where $\theta > 0$ is unknown.
(a) Argue that $e^{-\overline{X}_n}$ converges in probability to a constant and find the constant.
(b) Define the sequence

$$\widehat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$

Prove that $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \theta^2)$, as $n \to \infty$.
(c) Find a function of the $\widehat{\theta}_n$ sequence, say $g(\widehat{\theta}_n)$, whose large-sample variance is free of θ. *Hint:* Use the (first-order) Delta Method.