Problem 1

Let X be a continuous random variable with CDF F and PDF f.

(a) Find the conditional CDF of X given $X > a$ (where a is a constant with $\mathsf{P}(X > a) \neq 0$). That is, find $\mathsf{P}(X \leq x \mid X > a)$ for all a, in terms of F.

(b) Find the conditional PDF of X given $X > a$.

(c) Check that the conditional PDF from (b) is a valid PDF, by showing directly that it is nonnegative and integrates to 1.

Problem 2

A circle with a random radius $R \sim \mathsf{Unif}(0, 1)$ is generated. Let A be its area.

(a) Use simulations in R (the statistical programming language) to numerically estimate the mean and the variance of A.

(b) Find the theoretical mean and the variance of A, without first finding the CDF or PDF of A. Compare with your numerical results from (a).

(b) Find the CDF and PDF of A.

Problem 3

A stick of length 1 is broken at a uniformly random point, yielding two pieces. Let X and Y be the lengths of the shorter and longer pieces, respectively, and let $R = X/Y$ be the ratio of the lengths of X and Y.

(a) Use simulations in R (the statistical programming language) to gain some understanding about the distribution of the random variable R. Numerically estimate the expected value of R and $1/R$.

(b) Find the CDF and PDF of R.

(c) Find the expected value of R (if it exists).

(d) Find the expected value of $1/R$ (if it exists).

Problem 4

Let $U_1, \ldots, U_n$ be i.i.d. $\mathsf{Unif}(0, 1)$, and $X = \max(U_1, \ldots, U_n)$.

(a) What is the PDF of X?

(b) What is $\mathsf{E}X$?

(c) Use simulations in R (the statistical programming language) to numerically estimate $\mathsf{E}X$.

Problem 5

(a) Find $\mathsf{P}(X < Y)$ for $X \sim \mathsf{N}(a, b)$, $Y \sim \mathsf{N}(c, d)$ with X and Y independent. You might need the following result which you should consider known: *if X_1 and X_2 are independent with $X_1 \sim \mathsf{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathsf{N}(\mu_2, \sigma_2^2)$, then $X_1 + X_2 \sim \mathsf{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

(b) Use simulations in R (the statistical programming language) to numerically estimate $\mathsf{P}(X < Y)$ for $X \sim \mathsf{N}(0, 1)$, $Y \sim \mathsf{N}(1, 5)$ with X and Y independent.

Problem 6

The heights of men in the United States are normally distributed with mean 69.1 inches and standard deviation 2.9 inches. The heights of women are normally distributed with mean 63.7 inches and standard deviation 2.7 inches. Let x be the average height of 100 randomly sampled men, and y be the average height of 100 randomly sampled women.

(a) What is the distribution of $x - y$?

(b) Using simulations in R, calculate the Monte Carlo estimates of the mean and standard deviation of the distribution of $x - y$. Compare these estimates with the exact values of the mean and standard deviation.

(c) What is the probability that a randomly sampled man is taller than a randomly sampled woman? Please do not answer this question by reporting a Monte Carlo estimate of this probability.

Problem 7

Suppose that Y is binomially distributed $Y \sim \mathsf{Bin}(n = 5, \theta)$ where θ is unknown. Furthermore, assume that

$$\theta \in \Theta = \{0.0, 0.1, \ldots, 0.9, 1.0\},$$

and that *a priori* we view each of these 11 possibilities as equally likely, i.e.

$$\mathsf{P}(\theta) = \frac{1}{11}, \text{ for each } \theta \in \Theta.$$

(a) Using the Bayes' rule, write down a formula for the posterior distribution $\mathsf{P}(\theta \mid y)$ in terms of θ_i, and simplify as much as possible.

(b) For $y = 0$, make a plot for $\mathsf{P}(\theta \mid y)$ for each $\theta \in \Theta = \{0.0, 0.1, \ldots, 0.9, 1.0\}$. In other words, make a plot with the horizontal axis representing the 11 values of θ and the vertical axis representing the corresponding values of $\mathsf{P}(\theta \mid y)$.

(c) Repeat (b) for each $y \in \{1, 2, 3, 4, 5\}$, so in the end you have six plots (including the one in (b)). Describe what you see in your plots and discuss whether or not they make sense.

Problem 8

(a) Starting from independent uniform random variables ($U \sim \mathsf{Uni}(0, 1)$), devise an algorithm to generate independent samples from a Logistic distribution, having density

$$f(x) = \frac{e^{-x}}{(1 + e^{-x})^2}, \quad \text{for } x \in \mathbb{R}$$

(b) Implement your sampling algorithm in R, and use your code to produce a Monte Carlo estimate of $\mathsf{P}(X \in (2, 3))$ where X is a random variable that has a Logistic distribution. [*Please note that using a function that already exists in an R library that samples from the Logistic distribution will not help you obtain any points for this question*.]