ST 501 Sample Final Exam

This is a list of sample problems for the Final Exam for ST 501. The real exam is **closed book** and **closed notes**. However, you are allowed to bring **two** A4 page, double sided, formula/fact sheet. The real exam paper will also include a formula/fact sheet like the one at the end of this document.

- The exam time is 150 minutes. There will be 6 to 7 problems.
- A calculator is **not required** but you are welcome to bring one if you wish.
- These sample questions is meant to give you a rough idea of the format or type of questions you might see in the exam. The questions on the real exam are expected to be different, although we will try to make it so that the difficulty of the sample problems reflect, to some extent, the difficulty of the problems on the real exam (there will thus be a mix of easier problems and somewhat more complicated problems).
- Show all of your work, as detailed as you can. You do not need to provide exact numbers/answers, but the answers should be sufficiently detailed so that we can be sure that you know them. Note that if you do not show your work or explain your reasoning then it is much harder for us to grade your work; we might not be able to give you partial/full credits, even in cases where your answers are almost correct.
- The exam will cover Chapters 1 through 5 of the textbook. For discrete random variables you are expected to know the **binomial**, the **geometric**, the **Poisson**, **hypergeometric**, and **negative binomial** distributions. For continuous random variables you are expected to know the **uniform**, **normal** and **lognormal**, **special cases** of the **gamma** (including **exponential** and **chi-squared**).
- Solution sketches are provided at the end of this document.

Problem 1 (10pts)

A fair coin is tossed n times. Let X be the number of heads in this n toss. Given $X = x$, we generate a Poisson random variable Y with rate parameter $\lambda = x$. Find $\mathrm{Var}[Y]$. Your answer should only depends on n.

Hint Law of total variation. $\mathrm{Var}[Y] = \mathbb{E}[\mathrm{Var}[Y \mid X]] + \mathrm{Var}[\mathbb{E}[Y \mid X]]$.

Problem 2 (20pts)

Let (X, Y) be a bivariate random variable with **joint** probability density function (pdf)

$$f(x, y) = \begin{cases} 6x & \text{if } 0 \leq x \leq y \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

(a) (5pts) Write down a simple and **explicit** expression for the **marginal** pdf of X.
(b) (5pts) Are X and Y **independent** ? Justify your answer.
(c) (5pts) Calculate $P(X + Y \leq 0.6)$
(d) (5pts) Calculate $\mathbb{E}[\frac{X}{Y}]$.

Note Be careful of the limits of integrations, especially for part (c).

Problem 3 (15pts)

A factory that manufactures televisions operates in three shifts. Shift 1 produced 10% of the TVs, while Shift 2 and Shift 3 each produced 45% of the TVs. The rate of defective TVs produced by Shift 1, 2, and 3 are 2%, 4%, and 6%, respectively.

(a) (5pts) What proportion of all TVs produced are **defective** ?
(b) (5pts) You test a TV and found out that it is not defective. What is the probability it was made by Shift 1 ?
(c) (5pts) What proportion of TVs are made by Shift 1 **or** is not defective ? Here or is an **inclusive** or.

Problem 4 (10pts)

Dirty Harry's experience is that 10% of the packages he mails do not reach their destination. He has bought two books for 25 dollars apiece and wants to mail them to his sister. If he sends them in one package, the stamp is 6 dollars, while for separate packages the stamp is 4 dollars for each package. To minimize his expected total money lost (possible loss of books + postage), should he send one or two packages ?

Problem 5 (20pts)

Let $X_1, X_2, \ldots$ be independent and identically distributed **Bernoulli** random variables with the same success probability p. Let $Z_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$.

(a) (10pts) What is the **limiting** distribution of $n^{1/2}(Z_n - p)$ as $n \to \infty$?

(b) (10pts) Suppose $p \neq 1/2$. Find the limiting distribution of $Z_n(1 - Z_n)$ as $n \to \infty$? (Delta method)

Problem 6 (10pts)

Let X be a gamma random variable with shape parameters $\alpha = 4$ and $\beta = 1$. Let $Y = e^X$. Write down the expression for the **probability density function** of Y.

Hint The pdf of a Gamma random variable X with shape parameter $\alpha = n$ and scale parameter $\beta = 1$ is

$$f(x) = \frac{x^{n-1}e^{-x}}{(n-1)!}, \quad \text{for } x \geq 0$$

Problem 7 (10pts)

The probability that an individual randomly selected from a particular population has a certain disease is $p = 0.05$. A diagnostic test correctly detects the presence of the disease 98% of the time and correctly detects the absence of the disease 99% of the time. If the test is **applied twice**, the two test results are **independent**, and **both** are **positive**, what is the probability that the selected individual has the disease?

Problem 8 (10pts)

Let X and Y have **joint** probability density function

$$f(x, y) = 6(y - x) \quad \text{for } 0 \le x \le y \le 1$$

Compute $\mathbb{E}[X - Y]$.

Problem 9 (20pts)

Let (X, Y) has a uniform density in the half circle, i.e.,

$$f(x, y) = \frac{2}{\pi}, \qquad y \ge 0, x^2 + y^2 \le 1.$$

and $f(x, y) = 0$ otherwise.

(a) (10pts) Find $\mathbb{E}[Y]$.
(b) (10pts) Find $\mathbb{E}[X \mid Y]$.

Problem 10 (10pts)

There are two urns; call these urn A and urn B. There are 5 balls in urn A with 3 of them being red and 2 being black. There are 6 balls in urn B with 2 of them being red and 4 being black. We flip a **fair** coin once and if the coin landed head we draw a ball **at random** from urn A, while if the coin landed tail we choose a ball **at random** from urn B. **Given** that we drawn a **red** ball, what is the probability that the coin landed head ?

Problem 11 (15pts)

Sales delay is the elapsed time between the manufacture of a product and its sale. Suppose we can model sales delay as a **log normal** random variable X with parameters $\mu = 2$ and $\sigma = 0.2$ (the unit for X here is in months).

(a) (5pts) What is the probability that delay time exceeds 8 months ?
(b) (5pts) What is the probability that delay time exceeds 10 months **given** that it already exceeds 6 months ?
(c) (5pts) Suppose we selected 10 items at random. What is the **expected** number of items (among these 10 items) that have delay time exceeding 8 months ?

Note: All answers to this problem can (and should be) expressed in terms of the **cumulative distribution function** (cdf) for the **standard normal** random variable $Z \sim \mathcal{N}(0, 1)$. That is, your answer should be expressed in terms of the cdf

$$\Phi(z) = P(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} \, dz.$$

For example, a valid answer to part (a) is $\Phi(c)$ where c is some number.

Problem 12 (10pts)

Write down, for each of the following statements (statements (a) through (e)) whether the statement is **true** or **false**.

(a) Given **only** the **marginal** distributions of two random variables X and Y, we can find $\mathbb{E}[XY]$.

(b) Suppose X and Y are two random variables with $\mathbb{E}[XY] = \mathbb{E}[X] \times \mathbb{E}[Y]$. Then $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$.

(c) Suppose X and Y are independent random variables. Then $\mathbb{E}[XY] = \mathbb{E}[X] \times \mathbb{E}[Y]$.

(d) Let A, B, and C be events. If $P(A \cap B) = P(A) \times P(B)$ and $P(B \cap C) = P(B) \times P(C)$ then $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$.

(e) Let A and B be events. Then $P(A \mid B) \geq P(A)$ if and only if $P(\bar{A} \mid B) \leq P(\bar{A})$.

Problem 13 (15pts)

Three players play 10 **independent** rounds of a game. The first player has probability 0.5 of winning each round, the second player has probability 0.25 of winning each round and the third player has probability 0.25 of winning each round.

(a) (10pts) Write down the expression for the **joint probability mass function** of the numbers of games won by each of the three players.

(b) (5pts) What is the probability that the first player won **at least** 3 out of the 10 rounds ?

Problem 14 (15pts)

Suppose your waiting time for a bus in the morning is uniformly distributed on $[0, 8]$ (in minutes), whereas waiting time in the evening is uniformly distributed on $[0, 10]$ (in minutes) independent of morning waiting time.

(a) (5pts) If you take the bus each morning and evening for a week (seven days). What is your total expected waiting time ?

(b) (5pts) What is the variance of your total waiting time ?

(c) (5pts) What are the expected value and variance of the **difference** between **average** morning waiting time and **average** evening waiting time for a particular week ? (seven days).

Problem 15 (10pts)

In repeated throws of a **fair** die (so all six faces of the die are equally likely to appear), let X be the throw in which the first six is obtained and Y the throw in which the second six is obtained. For example, $X = 3$ and $Y = 7$ indicates that the first six appeared in the third throw and the second six appeared in the seventh throw.

(a) (5pts) Write down the **joint** probability mass function for X and Y. That is, write down a simple expressions for the function $p(x, y) = P(X = x, Y = y)$

(b) (5pts) Compute $\mathbb{E}[Y - X]$.

Hint Think memory-less. Especially for part (b).

Problem 16 (20pts)

Let X denote the temperature at which a certain chemical reaction takes place. Suppose that X has pdf

$$f(x) = \begin{cases} \frac{1}{9}(4 - x^2) & \text{if } -1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) (10pts) Write down a simple expression for the cdf (cumulative distribution function) of X.
(b) (5pts) What is the probability that the temperature is **positive** when the reaction occurs ?
(c) (5pts) Compute $\mathbb{E}[X]$, the **expected temperature**, at which a reaction will take place.

Problem 17 (10pts)

You are to play a game involving three tennis sets. You win the game if you win two sets **consecutively**, otherwise you lose the game. More specifically you win the game if you win both sets 1 and 2 or both sets 2 and 3. For each set you will play either your father or Djovak Nokovic. Let F denote your father and N denote Djovak. You get to choose whether you want to play the sequence FNF (you play your father in the first and third sets and Djovak in the second set) or the sequence NFN (you play Djovak in the first and third sets and your father in the second set). The probability that you win a set against your father is $p_F = 0.95$ and the probability that you win a set against Djovak is $p_N = 0.01$.

(a) (5pts) Assuming that the results of the three sets are **independent**, which sequence is **more advantageous** for you ?

(b) (5pts) Suppose the rule is changed so that you only have to win **at least** two out of the three sets (not necessarily consecutively). Which sequence is now **more advantageous** for you ?

Problem 18 (10pts)

Suppose that when the `pH` of a certain chemical compound is 5.00, the `pH` measured by a randomly selected beginning chemistry student is a **normal** random variable with mean 5.00 and standard deviation 0.3. A large batch of the compound is subdivided and a sample given to each student in a morning lab and each student in an afternoon lab. Let $\bar{X}$ be the average `pH` as determined by the morning students and $\bar{Y}$ be the average `pH` as determined by the afternoon students. Assuming that there are 25 students in each lab, compute $P(-0.1 \leq \bar{X} - \bar{Y} \leq 0.1)$.

Hint A linear combination of independent normal random variables is normally distributed.

Basic results in probability

Let Ω be a sample space and P a probability measure on Ω.

- $P(A) = 1 - P(\bar{A})$ for any event $A \subset \Omega$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ for any events $A, B \subset \Omega$.
- $P(A) = P(A \cap B) + P(A \cap \bar{B})$ for any events $A, B \subset \Omega$.
- Inclusion-exclusion for three events $A, B, C \subset \Omega$.

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

- Law of total probability

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A \mid B_i) \times P(B_i)$$

  for any **partition** $\{B_i\}$ of the sample space Ω.

- Bayes Law:

$$P(B_i \mid A) = \frac{P(A \mid B_i) \times P(B_i)}{\sum_j P(A \mid B_j) \times P(B_j)}$$

- Independence: If A and B are independent then $P(A \cap B) = P(A) \times P(B)$
- De Morgan's law.

$$(A \cup B)' = A' \cap B', \quad (A \cap B)' = A' \cup B'$$

- Distributive law.

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Basic facts about discrete random variables

- If X is a **binomial** random variable with n trials and success probability p then

$$P(X = x) = \tbinom{n}{x} p^x (1-p)^{n-x}, \quad \mathbb{E}[X] = np, \quad \mathrm{Var}[X] = np(1-p).$$

- If X is a **negative** binomial random variable for the number of trials at which the rth success **first** occurs and p is the success probability for each trial then

$$P(X = x) = \tbinom{x-1}{r-1} p^r (1-p)^{x-r}, \quad \mathbb{E}[X] = \frac{r}{p}, \quad \mathrm{Var}[X] = \frac{r(1-p)}{p^2}.$$

  The case when $r = 1$ is known as the **geometric** distribution.

- If X is a **Poisson** random variable with rate parameter λ then

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad \mathbb{E}[X] = \mathrm{Var}[X] = \lambda.$$

- If X is a **hypergeometric** random variable with parameters N, M and n then

$$\mathbb{P}(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \quad \mathbb{E}[X] = \frac{nM}{N}, \quad \mathrm{Var}[X] = \frac{nM}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}.$$

  Here N denote the total number of items, of which M are **special items** and we wish to sample n items, without replacement, from the N items.

Basic facts about continuous random variables

- If X is **normally distributed** with mean μ and variance σ^2 then $X = \mu + \sigma Z$ where Z is **standard normal**. Thus all probability calculations for X can be converted to that for Z. The mean and variance of a standard normal is 0 and 1, respectively.
- The $100p$-th percentile of a continuous random variable X with cdf F is any $c = c(p)$ such that $F(c) = p$.
- If X is a uniform random variable on $[a, b]$ with $a < b$ then

$$f(x) = \frac{1}{b-a}, \quad \text{for } a \le x \le b, \quad \mathbb{E}[X] = \frac{a+b}{2}, \quad \mathrm{Var}[X] = \frac{(b-a)^2}{12}.$$

- If X is an exponential random variable with rate parameter λ then

$$f(x) = \lambda e^{-\lambda x}, \quad \text{for } 0 \le x, \quad \mathbb{E}[X] = \frac{1}{\lambda}, \quad \mathrm{Var}[X] = \frac{1}{\lambda^2}.$$

  In addition, X is memory-less, i.e., $P(X \ge s + t \mid X \ge t) = P(X \ge s)$.

Basic facts about bivariate continuous random variables

- (X, Y) is a bivariate continuous random variables if there exists a function f with $f(x, y) \ge 0$ for all (x, y) and that

$$P(A) = \int_A f(x, y)\, \mathrm{d}x\mathrm{d}y.$$

  for **any** set $A \subset \mathbb{R}^2$. The function f is known as the **joint** probability density function (pdf)
- Given (X, Y) with **joint** pdf f, the marginal pdfs for X and Y are

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}y, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}x$$

- The conditional pdfs for Y given X and X given Y is then given by

$$f(y \mid x) = \frac{f(x, y)}{f_X(x)}, \quad f(x \mid y) = \frac{f(x, y)}{f_Y(y)}$$

  provided that $f_X(x) > 0$ or $f_Y(y) > 0$ as appropriate, with $f(y \mid x) = 0$ or $f(x \mid y) = 0$ otherwise.
- Conditional expectations and conditional cdf can be calculated using the conditional pdf, e.g.,

$$P(Y \le a \mid X = x) = \int_{-\infty}^{a} f(y \mid x)\, \mathrm{d}y, \quad \mathbb{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y f(y \mid x)\, \mathrm{d}y$$

- The covariance between X and Y is

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

- If $X_1, X_2, \ldots, X_m$ are random variables then

$$\mathrm{Var}[X_1 + X_2 + \cdots + X_m] = \sum_i \mathrm{Var}[X_i] + 2 \sum_{i<j} \mathrm{Cov}(X_i, X_j)$$

  Furthermore, expectation is **linear**.

Solutions Sketch

Problem 1 We applied the law of total variation, i.e., $\mathrm{Var}[Y] = \mathbb{E}[\mathrm{Var}[Y \mid X]] + \mathrm{Var}[\mathbb{E}[Y \mid X]]$.

Now, given $X = x$, Y is Poisson with rate parameter x and hence $\mathbb{E}[Y \mid X] = X$ and $\mathrm{Var}[Y \mid X] = X$.

As $X \sim \mathrm{Bin}(n, 1/2)$, we have

$$\mathbb{E}[\mathrm{Var}[Y \mid X]] = \mathbb{E}[X] = \frac{n}{2}, \quad \mathrm{Var}[\mathbb{E}[Y \mid X]] = \mathrm{Var}[X] = \frac{n}{4}, \quad \mathrm{Var}[Y] = \frac{3n}{4}.$$

Problem 2 For part (a) the marginal pdf of X is

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \int_x^1 6x\,\mathrm{d}y = 6xy\Big|_x^1 = 6x(1-x)$$

For part (b) X and Y are not independent due to the constraint $x \leq y$. Another way to see this is that the **marginal** pdf of Y is

$$f_Y(y) = \int_0^y 6x\,\mathrm{d}x = 3x^2\Big|_0^y = 3y^2$$

and so $f_X(x) \times f_Y(y) \neq f(x,y)$. For part (c) if $x + y \leq 0.6$ and $0 \leq x \leq y$ then $0 \leq x \leq 0.3$ and $x \leq y \leq 0.6 - x$. Indeed we cannot have $x > 0.3$ as otherwise if $y > x$ then $x + y > 0.6$. We thus have

$$P(X + Y \leq 0.6) = \int_0^{0.3} \int_x^{0.6-x} 6x\,\mathrm{d}y\,\mathrm{d}x = \int_0^{0.3} 6x(0.6 - 2x)\,\mathrm{d}x = 1.8x^2 - 4x^3\Big|_0^{0.3} = 0.054$$

Note that the region of integration in part (c) is not $0 \leq y \leq 0.6 - x$. That is why we explicitly warn you to be careful of the limits of integration.

Finally for part (d) we have

$$\mathbb{E}[X/Y] = \int_0^1 \int_0^y \frac{x}{y} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 \int_0^y \frac{6x^2}{y}\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 2y^2\,\mathrm{d}y = \frac{2}{3}$$

Problem 3 For part (a), by the law of total probability, we have

$$P(\text{defective}) = 0.1 \times 0.02 + 0.45 \times 0.04 + 0.45 \times 0.06 = 0.047$$

For part (b), using Bayes' theorem we have

$$P(\text{Shift 1} \mid \text{not defective}) = \frac{P(\text{not defective} \mid \text{Shift 1}) \times P(\text{Shift 1})}{P(\text{not defective})} = \frac{0.98 \times 0.1}{1 - 0.047} \approx 0.103$$

For part (c), we have

$$\begin{aligned} P(\text{Shift 1} \cup \text{not defective}) &= P(\text{Shift 1}) + P(\text{not defective}) - P(\text{Shift 1 and not defective}) \\ &= 0.1 + (1 - 0.047) - (0.98 \times 0.1) = 0.955 \end{aligned}$$

Problem 4 Suppose Dirty Harry sent the two books in a single package. Then the probability that the package went missing is 0.1 and so his expected loss is

$$(25 \times 2 + 6) \times 0.1 = 5.6\,\text{dollars}$$

Suppose he now sent two packages. Then the probability that exactly one package went missing is $2 \times 0.1 \times 0.9 = 0.18$ and the probability that both packages went missing is $0.1^2 = 0.01$. Dirty Harry's expected loss is then

$$(25 + 4) \times 0.18 + 2 \times (25 + 4) \times 0.1^2 = 5.8\,\text{dollars}$$

So Dirty Harry should send a single package.

Problem 5 The mean and variance of Z_n is $\mathbb{E}[Z_n] = p$ and $\mathrm{Var}[Z_n] = p(1-p)/n$. For part (a), by the central limit theorem, we have

$$n^{1/2}(Z_n - p) \to \mathcal{N}(0, p(1-p))$$

as $n \to \infty$.

Meanwhile, for part (b), we let $g(x) = x(1-x)$. Then $g'(x) = 1 - 2x$ and hence, by the first order delta method for normal approximation, we have

$$n^{1/2}(g(Z_n) - g(p)) \to \mathcal{N}(0, (g'(p))^2 \times p(1-p))$$

That is to say,

$$n^{1/2}(Z_n(1-Z_n) - p(1-p)) \to \mathcal{N}(0, p(1-p)(1-2p)^2)$$

as $n \to \infty$.

Problem 6 By the method of cdf we have

$$P(Y \leq y) = P(e^X \leq y) = P(X \leq \ln y) = F_X(\ln y)$$

Taking the derivative with respect to y we have

$$f_Y(y) = \frac{d}{dy}F_X(\ln y) = f_X(\ln y) \times \Big(\frac{d}{dy}\ln y\Big) = f_X(\ln y) \times \frac{1}{y}$$

Now the pdf of X is $x^3e^{-x}/\Gamma(4) = x^3e^{-x}/6$ from the hint. Substituting $\ln y$ for x we obtain

$$f_Y(y) = \frac{1}{6}(\ln y)^3 e^{-\ln y} \times \frac{1}{y} = \frac{(\ln y)^3}{6y^2}$$

Problem 7 Let D be the event that a person has a disease and T_{pos} be the event that the test yield a positive result. From the problem statement we have $P(T_{\mathrm{pos}} \mid D) = 0.98$ and $P(T_{\mathrm{neg}} \mid \bar{D}) = 0.99$. Now if, for a randomly selected individual, we run the test twice then both test will be positive with probability 0.98×0.98 if the person indeed has the disease; otherwise both test will be positive with probability 0.01×0.01. By Bayes law we therefore have

$$\begin{aligned} P(D \mid \text{both test positive}) &= \frac{P(\text{both test positive} \mid D) \times P(D)}{P(\text{both test positive} \mid D) \times P(D) + P(\text{both test positive} \mid \bar{D}) \times P(\bar{D})} \\ &= \frac{0.98^2 \times 0.05}{0.98^2 \times 0.05 + 0.01^2 \times 0.95} = 0.9980256 \end{aligned}$$

Problem 8 Given the joint pdf, we can then compute

$$\mathbb{E}[X-Y] = \int_0^1 \int_0^y (x-y)f(x,y) = \int_0^1 \int_0^y -6(x-y)^2 \,\mathrm{d}x\,\mathrm{d}y = \int_0^1 \Big(-2(x-y)^3\Big|_0^y\Big)\,\mathrm{d}y = \int_0^1 -2y^3\,\mathrm{d}y = \frac{-y^4}{2}\Big|_0^1 = -\frac{1}{2}.$$

Problem 9 For part (a) we have

$$\mathbb{E}[Y] = \frac{2}{\pi}\int_0^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} y\,\mathrm{d}x\,\mathrm{d}y = \frac{2}{\pi}\int_0^1 2y\sqrt{(1-y^2)}\,\mathrm{d}y = \frac{4}{3\pi}(1-y^2)^{3/2}\Big|_0^1 = \frac{4}{3\pi}.$$

For part (b), we first find the conditional pdf of X given Y. We have

$$f_Y(y) = \frac{2}{\pi}\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \mathrm{d}x = \frac{4\sqrt{1-y^2}}{\pi}$$

and hence the conditional pdf is

$$f_{X|Y}(x) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{2\sqrt{1-y^2}}.$$

We therefore have

$$\mathbb{E}[X \mid Y = y] = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \frac{x}{2\sqrt{1-y^2}}\,\mathrm{d}x = \frac{x^2}{4\sqrt{1-y^2}}\Big|_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} = 0$$

which is to be expected as the random variable X is symmetric around 0.

Problem 10 We apply Bayes' theorem. We have

$$\begin{aligned}\mathbb{P}[\text{coin head} \mid \text{red ball}] &= \frac{\mathbb{P}[\text{red ball} \mid \text{coin head}] \times \mathbb{P}[\text{coin head}]}{\mathbb{P}[\text{red ball} \mid \text{coin head}] \times \mathbb{P}[\text{coin head}] + \mathbb{P}[\text{red ball} \mid \text{coin tail}] \times \mathbb{P}[\text{coin tail}]} \\ &= \frac{3/5 \times 1/2}{3/5 \times 1/2 + 2/6 \times 1/2} = \frac{9}{14}\end{aligned}.$$

Problem 11 Recall that if X is log-normal then $\ln(X)$ is normal,i.e., $\ln(X)$ with mean $\mu = 2$ and standard deviation $\sigma = 0.2$ We then have

$$P(X \geq 8) = P\Big(\frac{\ln X - 2}{0.2} \geq \frac{\ln 8 - 2}{0.2}\Big) \approx 1 - \Phi(0.3972) \approx 0.346$$

For part (b) we have

$$P(X \geq 10 \mid X \geq 6) = \frac{P((X \geq 10) \cap (X \geq 6))}{P(X \geq 6)} = \frac{P(X \geq 10)}{P(X \geq 6)} \approx \frac{1 - \Phi(1.513)}{1 - \Phi(-1.041)} \approx 0.076$$

Finally for part (c) we can model the number of events as a binomial rv with $n = 10$ trials and success probability p given in part (a). Therefore the **expected** number of items is $np \approx 3.46$ items.

Problem 12 (a) and (d) are false and (b), (c) and (e) are true. Indeed (a) is false as the marginal pdf does not define the joint pdf. (d) is also false as pairwise independence does not imply that the events are mutually independent. (e) is true as conditional probability is also a probability. (b) and (c) are both true from the basic definitions.

Problem 13 For part (a) the joint pmf is that of a multinomial with $K = 3$ outcomes and success probabilities $(p_1, p_2, p_3) = (0.5, 0.25, 0.25)$. In other words

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{10!}{x_1!x_2!x_3!} 0.5^{x_1} 0.25^{x_2} 0.25^{x_3}, \quad 0 \leq x_1, 0 \leq x_2, 0 \leq x_3, x_1 + x_2 + x_3 = 10$$

For part (b) if we are only interested in the number of games won by the first player then this is a binomial with success probabilities $p = 0.5$. Thus

$$P(X_1 \geq 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2) = 1 - \sum_{k=0}^{2} \tbinom{10}{k} 0.5^k 0.5^{10-k} = 1 - \sum_{k=0}^{2} \tbinom{10}{k} 2^{-10}.$$

Problem 14 For part (a) the total waiting time is $T = M_1 + \cdots + M_7 + E_1 + \ldots E_7$ where M_i and E_i denote the waiting time in the morning and evening for each day. For part (a) we have

$$\mathbb{E}[T] = \mathbb{E}[M_1 + \cdots + M_7 + E_1 + \cdots + E_7] = \mathbb{E}[M_1] + \cdots + \mathbb{E}[M_7] + \mathbb{E}[E_1] + \cdots + \mathbb{E}[E_7] = 63$$

as $\mathbb{E}[M_i] = 4$ and $\mathbb{E}[E_i] = 5$ for each day i. For part (b), as the waiting times are independent, we have

$$\mathrm{Var}[T] = \mathrm{Var}[M_1 + \cdots + M_7 + E_1 + \cdots + E_7] = \mathrm{Var}[M_1] + \cdots + \mathrm{Var}[E_7] = 7 \times \Big(\frac{64}{12} + \frac{100}{12}\Big) \approx 95.67$$

as $\text{Var}[M_1] = (8-0)^2/12$ and $\text{Var}[E_1] = (10-0)^2/12$. Finally for part (c) the average morning waiting time is $\bar{M} = (M_1 + \cdots + M_7)/7$ and the average evening waiting time is $\bar{E} = (E_1 + \cdots + E_7)/7$. Recall the formula for the expected value and variance of a sample mean we have

$$\mathbb{E}[\bar{M}] = \mathbb{E}[M_1] = 4, \quad \text{Var}[\bar{M}] = \text{Var}[M_1]/7 = 64/84, \quad \mathbb{E}[\bar{E}] = \mathbb{E}[E_1] = 5, \quad \text{Var}[E_1] = \text{Var}[E_1]/7 = 100/84$$

We thus have

$$\mathbb{E}[\bar{E} - \bar{M}] = \mathbb{E}[\bar{M}] - \mathbb{E}[\bar{E}] = 4 - 5 = -1, \quad \text{Var}[\bar{M} - \bar{E}] = \text{Var}[\bar{M}] + \text{Var}[\bar{E}] = 164/84$$

Problem 15 For part (a) we can model X as a geometric rv with success probability $p = 1/6$ so $p(x) = (5/6)^{x-1} \times (1/6)$. Now $Y = Y - X + X$ and given X, $Y - X$ is another geometric rv with success probability also $p = 1/6$ so $p(y - x|x) = (5/6)^{y-x-1} \times (1/6)$. Combining these we obtain the joint pmf

$$p(x, y) = \frac{1}{6}(5/6)^{x-1} \times \frac{1}{6}(5/6)^{y-x-1} = (5/6)^{y-2}(1/6)^2, \quad \text{for } 1 \le x \le y - 1$$

Therefore, for part (b), we have $\mathbb{E}[Y - X] = 6$ using the formula for the expected value of a geometric rv. You can also calculate $\mathbb{E}[Y - X]$ exactly. The calculation is slightly tedious though.

Problem 16 For part (a) we have $F(x) = 0$ for $x \le -1$ and

$$F(x) = \int_{-1}^{x} \tfrac{1}{9}(4 - y^2)\,\mathrm{d}y = \tfrac{1}{9}\big(4y - \tfrac{y^3}{3}\big)\Big|_{-1}^{x} = \tfrac{1}{9}(4x - \tfrac{x^3}{3}) + \tfrac{11}{27}$$

for $-1 \le x \le 2$ and $F(x) = 1$ for $x \ge 2$.

For part (b) we are interested in $P(X > 0) = 1 - P(X \le 0) = 1 - F(0) = \frac{16}{27}$.

Finally for part (c) we have

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = \int_{-1}^{2} \tfrac{1}{9}(4x - x^3) = \tfrac{1}{9}\big(x^2 - \tfrac{x^4}{4}\big)\Big|_{-1}^{2} = \tfrac{1}{4}$$

Problem 17 For part (a) let us first consider the sequence NFN. Let A_1 be the event that you win the first set against your Nokovic, A_2 be the event that you win the second set against your father and A_3 be the event that you win the last set against Nokovic. Then

$$\begin{aligned} P(\text{you win}) &= P((A_1 \cap A_2) \cup (A_2 \cap A_3)) \\ &= P(A_1 \cap A_2) + P(A_2 \cap A_3) - P(A_1 \cap A_2 \cap A_3) \\ &= 0.01 \times 0.95 + 0.95 \times 0.01 - 0.01 \times 0.95 \times 0.01 = 0.018905 \end{aligned}$$

Similarly, for the sequence FNF let B_1 and B_3 be the events that you win the first and third set (against your father) and B_2 be the event that you win the second set against Nokovic. We then have

$$\begin{aligned} P(\text{you win}) &= P((B_1 \cap B_2) \cup (B_2 \cap B_3)) \\ &= P(B_1 \cap B_2) + P(B_2 \cap B_3) - P(B_1 \cap B_2 \cap B_3) \\ &= 0.01 \times 0.95 + 0.95 \times 0.01 - 0.95 \times 0.01 \times 0.95 = 0.009975 \end{aligned}$$

Thus the sequence NFN is more advantageous. The main reason being that you need to win the middle set always and this is easier if you play against your father.

For part (b) we have for the sequence NFN that

$$\begin{aligned} P(\text{you win}) &= P((A_1 \cap A_2) \cup (A_1 \cap A_3) \cup (A_2 \cap A_3)) \\ &= P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_2 \cap A_3) - 2P(A_1 \cap A_2 \cap A_3) \\ &= 0.95 \times 0.01 + 0.01 \times 0.01 + 0.95 \times 0.01 - 2 \times 0.01 \times 0.95 \times 0.01 = 0.01891 \end{aligned}$$

and for the sequence FNF we have

$$\begin{aligned} P(\text{you win}) &= P((B_1 \cap B_2) \cup (B_1 \cap B_3) \cup (B_2 \cap B_3)) \\ &= P(B_1 \cap B_2) + P(B_1 \cap B_3) + P(B_2 \cap B_3) - 2P(B_1 \cap B_2 \cap B_3) \\ &= 0.95 \times 0.01 + 0.95 \times 0.95 + 0.95 \times 0.01 - 2 \times 0.95 \times 0.91 \times 0.95 = 0.90535 \end{aligned}$$

and so now the sequence FNF is more advantageous.

Problem 18 From the above problem description let X_1 through X_25 be the pH measurements for the students in the morning lab. Then, assuming the measurements are independent, $\bar{X} = \frac{1}{25}\sum_{i=1}^{25} X_i$ is normally distributed with mean $\mu = 5$ and variance $0.3^2/25$. Similarly, $\bar{Y}$ is normally distributed with mean $\mu = 5$ and variance $0.3^2/25$ and hence $\bar{X} - \bar{Y}$ is normally distributed with mean $\mu = 0$ and variance $2 \times 0.3^2/25$. Therefore, by transforming to standard normal, we have

$$\begin{aligned} P(-0.1 \leq \bar{X} - \bar{Y} \leq 0.1) &= P\Big(\frac{-0.1 - 0}{\sqrt{2 \times 0.3^2/25}} \leq Z \leq \frac{0.1 - 0}{\sqrt{2 \times 0.3^2/25}}\Big) \\ &= P\Big(-\frac{5}{3\sqrt{2}} \leq Z \leq \frac{5}{3\sqrt{2}}\Big) \approx \Phi(1.1785) - \Phi(-1.1785) \approx 0.761 \end{aligned}$$

where Z denote a standard normal and Φ denote its cdf.