ST 501 Sample Midterm Exam 2

This is a list of sample problems for Midterm 2 for ST 501. The real exam is **closed book** and **closed notes**. However, you are allowed to bring one 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 (note that there are things on the fact sheet that will not be relevant to the current exam)

- The exam time is 90 minutes. There will be 5 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 Chapter 2, and sections 4.1, 4.2, and 4.5 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 varibles you are expected to know the **uniform**, **normal** and **lognormal**,**exponential** and **chi-squared** distributions.
- All answers to problems involving the normal distribution can (and should be) expressed in terms of the cumulative distribution function for the **standard** normal random variable $Z \sim \mathcal{N}(0,1)$. That is, your answer should be expressed in terms of the cdf

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

- Solution sketches are provided at the end of this document.

Problem 1

Suppose Mickey's pocket expenses per month are normally distributed with mean 1000 dollars and standard deviation 100 dollars. Suppose Minnie's pocket expenses per month are normally distributed with mean 600 dollars and standard deviation 50 dollars. Assume that Mickey's and Minnie's pocket expenses are independent.

- (a) Find the probability that Mickey and Minnie total pocket expense in one month exceeds 2000 dollars
- (b) Find the probability that Mickey spends **at least** twice as much as Minnie in pocket expenses in some random month.

Problem 2 (10pts)

Suppose U is uniformly distributed on $[-1, 1]$. Let $X = e^{-U^2}$. Find the probability density function for X.

Problem 3

Scores on a certain standardized test are approximately normally distributed with mean $\mu = 100$ and variance $\sigma^2 = 25$. Ten individual is selected at random. What is the probability that at least five individuals score above 110 on the exams ?

Problem 4

The waiting time at a teller's window in a bank is a random variable X with probability density function

$$f(x) = \begin{cases} \frac{1}{3}e^{-x/3} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$$

- (a) Find the average/expected waiting time and the standard deviation of the waiting time.
- (b) Find t^* such that the probability that you have to wait **longer** than t^* is **at most** 0.05.
- (c) Suppose you have waited for 3 minutes and had not yet been serviced. What is the probability that you will have to wait for **at least** 3 **additional** minutes ?

Problem 5

Let X be a **continuous** random variable with probability density function

$$f(x) = \frac{1}{2}e^{-|x|} \quad x \in \mathbb{R}.$$

Find the moment generating function for X.

Problem 6

Diameters of ball bearings made at a factory are **normally distributed** with mean 2 centimeters and standard deviation of 0.5 centimeter. Balls whose diameter exceed 3 cm or is less than 1 centimeter are discarded. Balls whose diameter are between 2.5 and 3 centimeters **or** between 1 and 1.5 centimeters are sent back to the machines to be fixed. The remaining balls are shipped for sale.

- (a) (10pts) What is the fraction of balls that are **shipped for sale** ?
- (b) (5pts) Given that a ball is **not shipped** for sale, what is the probability that it is discarded ? **Note:** All answers to this problem can (and should be) expressed in terms of the cumulative

distribution function for the **standard** normal random variable $Z \sim \mathcal{N}(0,1)$. That is, your answer should be expressed in terms of the cdf

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

Problem 7

Let X be a random variable with pdf of the form

$$f(x) = \begin{cases} \frac{\theta}{\tau}\Big(1 - \frac{x}{\tau}\big)^{\theta-1} & \text{if } 0 \leq x \leq \tau \\ 0 & \text{otherwise} \end{cases}$$

Here θ and τ are fixed constants.

(a) Write down an expression for the cdf F of X. Your expression should depend on θ and τ.
(b) Write down an expression for the quantile function of X, i.e., given a value $p \in (0,1)$, what is the value c such that $F(c) = p$? Your expression should depend on θ and τ.
(c) Suppose $\theta = 4$ and $\tau = 1$. Calculate $\mathbb{E}[X]$.

Problem 8

Given data from a dart-throwing experiment, a researcher proposed that the horizontal and vertical errors from aiming at a point target should be independent of one another, each with a normal distribution having mean 0 and variance 1. It can then be shown that the **pdf** of the distance D (measured in **centimeters**) from the target to the landing point is

$$f(d) = de^{-d^2/2}, \quad \text{for } d \geq 0.$$

(a) (5pts) What is the probability that a dart will land within 2 `cm` of the target ?
(b) (5pts) Suppose Alice and Bob **take turns** throwing at the target. Alice will win the game if, for any of her turn, she can get close to within 1 `cm` of the target. Bob will win the game if, for any of his turn, he can get close to within 2 `cm` of the target. Alice starts the game. What is the probability that Alice wins ?

Problem 9

In a small town in an unnamed state there are 60 Republicans and 40 Democrats. Ten are selected at random for a council. What is the probability that there are more Democrats than Republicans in the council ?

Problem 10

Between the months of May and October you can see a shooting star at the rate of one every 15 minutes. Suppose you decide to sit on your porch for one hour each evening during the months of July and August (62 days). What is the probability that there is **at least** one evening during this period for which you see **ten or more** shooting stars during the one hour time window ?

Problem 11

Let X be a continuous random variable with pdf

$$f(x) = \begin{cases} \frac{1}{18}x^2 & \text{if } -3 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = -2$ if $X \leq -2$, $Y = 2$ if $X \geq 2$ and $Y = X$ if $-2 \leq X \leq 2$.

(a) Write down the **cumulative distribution function** (cdf) for Y.
(b) Is Y a continuous random variable ? Why or why not ?

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. A geometric rv is **memory-less**.

- 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

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

  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

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

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

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

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

Solutions Sketch

Problem 1 For part (a), as Mickey's and Minnie's expenses are both normally distributed and independent, their sum is also normally distributed. Hence the total expense is normally distributed with mean 1600 and variance $100^2 + 50^2 = 12500$.

Letting Z be a standard normal random variable, the probability that their expense exceeds 2000 dollars is

$$P(\mathcal{N}(1600, 12500) \geq 2000) = P(Z \geq \frac{400}{\sqrt{12500}}) = 1 - \Phi(400/\sqrt{12500}) = 1 - \Phi(8/\sqrt{5}) \approx 0.0002.$$

For part (b), we are interested in $P(X - 2Y \geq 0)$ where $X \sim \mathcal{N}(1000, 100^2)$ and $Y \sim \mathcal{N}(600, 50^2)$. As X and Y are independent and normally distributed, $X - 2Y$ is normally distributed with mean $1000 - 2 \times 600 = -200$ and variance $100^2 + 2^2 \times 50^2 = 20000$. We therefore have

$$P(X - 2Y > 0) = P(Z \geq \frac{200}{\sqrt{20000}}) = P(Z \geq \sqrt{2}) = 1 - \Phi(\sqrt{2}) \approx 0.08.$$

Problem 2 We use the method of cdf, but we need to be a bit careful about the function U^2 which is not monotone on $[-1, 1]$. See also Example C on page 61 in section 2.3 of your textbook.

Since $U \in [-1, 1]$, $U^2 \in [0, 1]$ and hence $e^{-U^2} \in [e^{-1}, 1]$. We therefore have, for $X \in [e^{-1}, 1]$,

$$\begin{aligned}\mathbb{P}[X \leq x] &= \mathbb{P}[-U^2 \leq \ln x] = \mathbb{P}[U^2 \geq -\ln x] = \mathbb{P}[|U| \geq \sqrt{-\ln x}] \\ &= 1 - \mathbb{P}[-\sqrt{-\ln x} \leq U \leq \sqrt{-\ln x}] = 1 - \int_{-\sqrt{-\ln x}}^{\sqrt{\ln x}} \frac{1}{2} \,\mathrm{d}u = 1 - \sqrt{-\ln x}.\end{aligned}$$

Taking the derivative of the cdf for X, we have

$$f_X(x) = -\frac{1}{2\sqrt{-\ln x}} \times \frac{-1}{x} = \frac{1}{2x\sqrt{-\ln x}} = \frac{1}{2x\sqrt{\ln \frac{1}{x}}}.$$

Problem 3 Let p be the probability that a randomly person has score exceeding 110 and let X be a normally distributed random variable with mean $\mu = 100$ and variance $\sigma^2 = 25$. Then

$$p = P(X \geq 110) = P((X - 100)/5 \geq 2) = 1 - \Phi(2) \approx 0.02275$$

where Φ denote the cdf of a standard normal. If we choose 10 people at random then Y, the number of people whose score exceed 110, can be modeled as a binomial random variable with $n = 10$ trials and success probability p. Therefore

$$P(Y \geq 5) = \sum_{k=5}^{10} \tbinom{10}{k} p^k (1-p)^{n-k} \approx 1.39 \times 10^{-6}.$$

Problem 4 The random variable X follows an exponential distribution with rate parameter $\lambda = \frac{1}{3}$. Therefore, $\mathbb{E}[X] = \frac{1}{\lambda} = 3$ minutes and $\mathrm{Var}[X] = \frac{1}{\lambda^2} = 9$ and so the standard deviation is also 3 minutes. For part (b), the cumulative distribution function for X is

$$F(t) = P[X \leq t] = \int_0^t \frac{1}{3} e^{-x/3} \,\mathrm{d}x = -e^{-x/3}|_0^t = 1 - e^{-t/3}.$$

We are thus interested in finding t^* such that

$$P[X \geq t^*] = 1 - F(t^*) \leq 0.05 \Longrightarrow e^{-t^*/3} \leq 0.05.$$

We therefore have $t^* \geq -3\log(0.05) \approx 9$ minutes. For part (c), we are interested in the **conditional probability**

$$P[X \geq 6 \mid X \geq 3] = \frac{P[X \geq 6]}{P[X \geq 3]} = \frac{e^{-6/3}}{e^{-3/3}} = e^{-1} \approx 0.37.$$

Problem 5 The moment generating function for X is

$$\mathbb{E}[e^{tX}] = \frac{1}{2}\int_{-\infty}^{\infty} e^{tx}e^{-|x|}\,\mathrm{d}x = \frac{1}{2}\int_{-\infty}^{0} e^{tx+x}\,\mathrm{d}x + \frac{1}{2}\int_{0}^{\infty} e^{tx-x}\,\mathrm{d}x.$$

We now evaluate the first integral, i.e.,

$$\frac{1}{2}\int_{-\infty}^{0} e^{tx+x}\,\mathrm{d}x = \frac{1}{2}\int_{-\infty}^{0} e^{x(t+1)}\,\mathrm{d}x = \frac{1}{2}\frac{1}{(t+1)}e^{x(t+1)}\Big|_{-\infty}^{0} = \frac{1}{2(1+t)}.$$

We have used the fact that $|t| < 1$ so that the integral is finite. The second integral is similar, i.e.,

$$\frac{1}{2}\int_{-\infty}^{0} e^{tx-x}\,\mathrm{d}x = \frac{1}{2}\int_{-\infty}^{0} e^{x(t-1)}\,\mathrm{d}x = \frac{1}{2}\frac{1}{(t-1)}e^{x(t-1)}\Big|_{0}^{\infty} = \frac{1}{2(1-t)}.$$

We therefore have

$$\mathbb{E}[e^{tX}] = \frac{1}{2(1+t)} + \frac{1}{2(1-t)} = \frac{1}{1-t^2}.$$

The distribution of X is known as the Laplace distribution.

Problem 6 For part (a), a ball is shipped for sale if its diameter is between 1.5 to 2.5 centimeters.

We thus have

$$P(\text{shipped for sale}) = P(1.5 \leq X \leq 2.5) = P\Big(\frac{1.5-2}{0.5} \leq \frac{X-2}{0.5} \leq \frac{2.5-2}{0.5}\Big) = P(-1 \leq Z \leq 1)$$

where $Z = (X-2)/0.5$ is a standard normal random variable. We therefore have

$$P(\text{shipped for sale}) = \Phi(1) - \Phi(-1) \approx 0.6826895$$

For part (b), a ball is discarded if its diameter is less than 1 cm or larger than 3 centimeters. By transforming to standard normal, we have

$$P(\text{ball is discarded}) = P(Z \leq -2) + P(Z \geq 2) = \Phi(-2) + 1 - \Phi(2) \approx 0.0455003$$

We therefore have

$$P(\text{ball is discared} \mid \text{ball is not shipped}) = \frac{P(\text{ball is discarded})}{P(\text{ball is not shipped})} = \frac{1-\Phi(2)+\Phi(-2)}{1-\Phi(1)+\Phi(-1)} \approx 0.1433935.$$

Problem 7 The cdf for X is $F(x) = 0$ if $x \leq 0$ and $F(x) = 1$ if $x \geq \tau$. For $0 \leq x \leq \tau$ we can integrate the pdf to obtain

$$F(x) = \int_{0}^{x} \frac{\theta}{\tau}\Big(1 - \frac{y}{\tau}\Big)^{\theta-1}\,\mathrm{d}y = \Big(1 - \frac{y}{\tau}\Big)^{\theta}\Big|_{0}^{x} = 1 - \Big(1 - \frac{x}{\tau}\Big)^{\theta}.$$

Now for part (b) the quantile function is

$$F(c) = p \Longrightarrow \Big(1 - \frac{c}{\tau}\Big)^{\theta} = 1 - p \Longrightarrow c = \tau(1 - (1-p)^{1/\theta}).$$

Finally, if $\theta = 4$ and $\tau = 1$ then the pdf is $f(x) = 4(1-x)^3$ for $0 \leq x \leq 1$ and hence the expectation is

$$\mathbb{E}[X] = \int_{0}^{1} 4x(1-x)^3\,\mathrm{d}x = \frac{4\Gamma(2)\times\Gamma(4)}{\Gamma(6)} = 0.2.$$

Problem 8 The cdf for D is

$$F(d) = \int_0^d xe^{-x^2/2}\,\mathrm{d}x = -e^{-x^2/2}\Big|_0^d = 1 - e^{-d^2/2}$$

for $d \geq 0$ and $F(d) = 0$ otherwise. Note that we have done a u-substitution when evaluating the above integral, i.e., we set $u = x^2/2$ so that $\mathrm{d}u = x\mathrm{d}x$.

We thus have $P(D \leq 2) = F(2) = 1 - e^{-2}$. For part (b) let $p_A = 1 - e^{-1/2}$ be the probability that Alice hits within 1 cm of the target for her turn and $p_B = 1 - e^{-2}$ be the probability that Bob hits within 2 cm of the target for his turn. Then Alice wins with probability

$$\begin{aligned}P(\text{Alice wins}) &= p_A + (1-p_A)(1-p_B)p_A + (1-p_A)(1-p_B)(1-p_A)(1-p_B)p_A + \dots \\ &= p_A \sum_{k=0}^{\infty}(1-p_A)^k(1-p_B)^k = \frac{p_A}{1-(1-p_A)(1-p_B)} = \frac{1-e^{-1/2}}{1-e^{-5/2}} \approx 0.43\end{aligned}$$

Note that the solution for part (b) is similar to problem 2.14 in your textbook.

Problem 9 Let X be the number of Democrats selected among the 10 people in the council. Then X has a hypergeometric distribution with $M = 40, N = 100$ and $n = 10$ (as the population size is finite (100 people) and the sampling is without replacement). For this problem we are interested in $P(X \geq 6)$ and thus, by the form of the pmf for a hypergeometric, we have

$$P(X \geq 6) = \frac{\binom{40}{6}\times\binom{60}{4} + \binom{40}{7}\times\binom{60}{3} + \binom{40}{8}\times\binom{60}{2} + \binom{40}{9}\times\binom{60}{1} + \binom{40}{10}}{\binom{100}{10}} \approx 0.154$$

Problem 10 Let X_i be the number of shooting stars we see in 1 hour during the ith evening where $i \in \{1, 2, \dots, 62\}$ spanning the month of July and August. Then from the problem statement each X_i has a Poisson distribution with rate parameter $\lambda = 4$.

Let Y_i be the event that $\{X_i \geq 10\}$, i.e., $Y_i = 1$ if $X_i \geq 10$ and $Y_i = 0$ otherwise. We can now assume that the Y_i are mutually independent and identically distributed. Let $p = P(Y_i = 1)$. We then have

$$p = P(X_i \geq 10) = \sum_{k=10}^{\infty} e^{-4}4^k/k! \approx 0.0081$$

Now the number of evenings for which we see ten or more shooting stars can be viewed as a binomial distribution with $n = 62$ trials and success probability p. We thus have

$$P(\text{at least one evening}) = 1 - P(\text{no evening}) = 1 - (1-p)^{62} \approx 0.397$$

Problem 11 The cdf for Y is any function F such that $F(y) = P(Y \leq y)$. From the problem description we see that

$$P(Y < -2) = 0, \quad P(Y \leq -2) = P(X \leq -2) = \int_{-3}^{-2} \frac{1}{18}x^2\,\mathrm{d}x = \frac{x^3}{54}\Big|_{-3}^{-2} = \frac{19}{54}$$

and thus Y cannot be a continuous random variable as there is a jump in the cdf at $Y = -2$. Using the same reasoning we then obtain

$$\begin{aligned}P(Y \leq y) = P(Y = -2) + P(-2 < Y \leq y) = \frac{19}{54} + \int_{-2}^{y} \frac{1}{18}x^2\,\mathrm{d}x = \frac{y^3}{54} + \frac{1}{2} \quad \text{if } -2 \leq y < 2 \\ P(Y \leq 2) = P(Y < 2) + P(Y = 2) = P(Y < 2) + P(X \geq 2) = 1\end{aligned}$$