Second Midterm Exam

Name

Please circle the section number, time, and instructor of your lecture:

SECTION 1	SECTION 2	SECTION 3
MWF 9:55am	MWF 2:25pm	TuTh 9:30am
Hao Shen	Gregory M. Shinault	Gregory M. Shinault

- **There are 5 problems on the exam, some of them have multiple parts.**
- Budget your time wisely! Read the problems carefully!
- You are not allowed to use a calculator or other handheld devices.
- Please present your solutions in a clear manner. Justify your steps.
- If you need extra room, use the back of the pages.
- You may use the PMF/PDF, expected value, and variance of the Bernoulli, Binomial, Geometric, Normal, and Poisson distributions without re-deriving the quantities. These can be treated as well known formulas.
- When there are 10 minutes left, you must remain seated until I dismiss you. This is to prevent disruptions for other students still working on their exams, and to make exam collection easier.

Problem	Points
1	/20
2	/20
3	/20
4	/20
5	/20
Total	/100

1. Consider a random variable X that has the cumulative distribution function (CDF)

$$F_X(t) = \begin{cases} 1 - 2te^{-2t} - e^{-2t} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0. \end{cases}$$

(a) Use the CDF to compute $P(1 \leq X \leq 2)$.
Solution. $F(2) - F(1) = -5e^{-4} - (-3e^{-2})$

(b) Find the PDF of X.
Hint: This should simplify to an expression with no addition or subtraction in it.
Solution. We differentiate the CDF. For $t \geq 0$,

$$\begin{aligned} f_X(t) &= 0 - (2e^{-2t} + 2te^{-2t}(-2)) - e^{-2t}(-2) \\ &= -2e^{-2t} + 4te^{-2t} + 2e^{-2t} \\ &= 4te^{-2t}. \end{aligned}$$

For $t < 0$, this derivative is always 0. So in total we have

$$f_X(t) = \begin{cases} 4te^{-2t} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

(c) Compute the expected value of X.

Solution. This is integration by parts.

$$\begin{aligned}
EX &= \int_0^\infty t \cdot 4te^{-2t}\ dt = \int_0^\infty 4t^2 e^{-2t}\ dt \\
&= [-2t^2 e^{-2t}]_{t=0}^\infty - \int_0^\infty -4te^{-2t}\ dt \\
&= \int_0^\infty 4te^{-2t}\ dt \\
&= [-2te^{-2t}]_{t=0}^\infty - \int_0^\infty -2e^{-2t}\ dt \\
&= \int_0^\infty 2e^{-2t}\ dt \\
&= [-e^{-2t}]_{t=0}^\infty \\
&= 1
\end{aligned}$$

(d) Compute the standard deviation of X. *Hint:* You can reuse your integral from part (c) to avoid some (but not all) computations.

Solution. First we need EX^2.

$$\begin{aligned}
EX^2 &= \int 4t^3 e^{-2t}\ dt \\
&= [-2t^3 e^{-2t}]_{t=0}^\infty - \int_0^\infty -6t^2 e^{-2t}\ dt \\
&= \int_0^\infty 6t^2 e^{-2t}\ dt \\
&= \frac{6}{4} \int_0^\infty 4t^2 e^{-2t}\ dt \\
&= \frac{3}{2} \cdot 1 \\
&= \frac{3}{2}
\end{aligned}$$

So

$$\mathrm{SD}(X) = \sqrt{\frac{3}{2} - 1^2} = \frac{1}{\sqrt{2}}$$

2. The *Benford distribution* gives the probability distribution for the first digit from the numbers in a naturally generated dataset. It can be useful to detect if data has been fraudulently generated. The probability mass function for the Benford distribution is given by

$$p_X(k) = \log_{10}\left(\frac{k+1}{k}\right) \qquad \text{for } k = 1, 2, \ldots, 8, 9.$$

(a) Verify that this is a valid PMF. In other words, show that

$$p_X(k) \geq 0 \quad \text{for all } k, \text{ and} \quad \sum_{k=1}^{9} p_X(k) = 1.$$

Solution. First,

$$\frac{k+1}{k} > 1 \quad \Rightarrow \quad \log_{10}\left(\frac{k+1}{k}\right) > \log_{10}(1) = 0$$

for $k = 1, 2, \ldots, 9$. So the first property is satsfied.

For the second,

$$\begin{aligned}
\sum_{k=1}^{9} \log_{10}\left(\frac{k+1}{k}\right) &= \log_{10}\left(\frac{2}{1}\right) + \log_{10}\left(\frac{3}{2}\right) + \cdots + \log_{10}\left(\frac{9}{8}\right) + \log_{10}\left(\frac{10}{9}\right) \\
&= \log_{10}\left(\frac{2 \cdot 3 \cdot 4 \cdots 9 \cdot 10}{1 \cdot 2 \cdot 3 \cdots 8 \cdot 9}\right) \\
&= \log_{10}(10) = 1.
\end{aligned}$$

(b) Prove that the expected value of the Benford distribution is $9 - \log_{10}(9!)$.

Solution. We can compute directly using the same approach as part (a).

$$\begin{aligned}
\sum_{k=1}^{9} k \log_{10}\left(\frac{k+1}{k}\right) &= \sum_{k=1}^{9} \log_{10}\left(\frac{(k+1)^k}{k^k}\right) \\
&= \log_{10}\left(\frac{2^1 \cdot 3^2 \cdot 4^3 \cdots 9^8 \cdot 10^9}{1^1 \cdot 2^2 \cdot 3^3 \cdots 8^8 \cdot 9^9}\right) \\
&= \log_{10}\left(\frac{10^9}{1 \cdot 2 \cdot 3 \cdots 8 \cdot 9}\right) \\
&= \log_{10}(10^9) - \log_{10}(9!) \\
&= 9 - \log_{10}(9!)
\end{aligned}$$

3. A fair coin is flipped 10,000 times. Let S denote the number of heads observed.

(a) Are the following events have the same probabilities or not? If not, which of the following events have the largest probability?

$$\{1000 \leq S \leq 1100\}, \qquad \{3000 \leq S \leq 3100\}, \qquad \{5000 \leq S \leq 5100\}$$

Solution. $\{5000 \leq S \leq 5100\}$ has the largest probability. All three events have the same interval size, but this is closest to the mean of the distribution.

(b) Approximately compute the probability of the event that has the largest probability (If you think that they have the same probabilities then please compute this probability).

(Hint: you may (or may not) need these values $\Phi(1) \approx 0.8413$, $\Phi(2) \approx 0.9772$)

Solution. Note that $S \sim \text{Bin}(10,000, 1/2)$. So

$$\mu = 5000, \qquad \sigma = \sqrt{10000 \cdot \frac{1}{2} \cdot \frac{1}{2}} = \sqrt{5000 \cdot \frac{1}{2}} = \sqrt{2500} = 50.$$

So

$$\begin{aligned} P(5000 \leq S \leq 5100) &= P\left(\frac{5000-5000}{50} \leq \frac{S-5000}{50} \leq \frac{5100-5000}{50}\right) \\ &\approx P(0 \leq Z \leq 2) \\ &= \Phi(2) - \Phi(0) \\ &\approx 0.9772 - 0.5 \\ &= 0.4772. \end{aligned}$$

(c) Now we change the experiment slightly. A coin is still flipped 10,000 times and S denotes the number of heads observed. However, this coin is not fair. It is heavily weighted so that the probability of heads on a single flip is $\frac{1}{2000}$. With this change, approximate the probability that we observe strictly less than 3 heads in 10,000 coin flips.

Solution. In this case, we have $\mu = 10,000/2000 = 5$. We will use this for the approximation.

$$\begin{aligned}
P(S < 3) &= P(S = 0) + P(S = 1) + P(S = 2) \\
&\approx e^{-5} + e^{-5}\frac{5^1}{1!} + e^{-5}\frac{5^2}{2!} \\
&= \frac{37}{2} \cdot e^{-5}
\end{aligned}$$

4. In the class we computed expectation and variance for Binomial distribution. Those calculations involved clever tricks. This question asks you to use **moment generating function** to re-derive the expectation and variance for Binomial distribution.

   (a) Prove that the moment generating function for a random variable X that follows the Binomial(n, p) distribution is $M_X(t) = (1 - p + pe^t)^n$.
   Hint: Recall that $P(X = k) = \binom{n}{k} p^k (1-p)^k$. When finding moment generating function, you might apply the binomial theorem $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$.
   Solution. Using the hint,

$$\begin{aligned} M_X(t) &= E[e^{tX}] \\ &= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\ &= (pe^t + 1 - p)^n \end{aligned}$$

   (b) Use the moment generating function to find expectation and variance for a random variable X that follows the Binomial(n, p) distribution.
   Solution. First we need the derivative of the MGF.

$$M'(t) = n(pe^t + 1 - p)^{n-1} pe^t$$

   So $E(X) = M'(0) = np$.
   For variance, we need the second derivative of the MGF.

$$M''(t) = n(n-1)(pe^t + 1 - p)^{n-1} p^2 e^{2t} + n(pe^t + 1 - p)^{n-1} pe^t$$

   Now compute

$$E(X^2) = M''(0) = n(n-1)p^2 + np.$$

   Therefore $Var(X) = E(X^2) - E(X)^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)$

5. Suppose $X \sim \text{Exp}(\lambda)$ and $Y = \ln X$. Find the probability density function of Y.

   Solution. $\lambda e^{t-\lambda e^t}$. This is Exercise 5.7 from Homework 8.