COMP680 Midterm Exam

Name:
NetID:

Instructions:

- Please show steps of your work to maximize partial credit.
- You do not need to have a fully computed numerical answer to get full credit. For example $\frac{4}{9} + (\frac{2}{5})^2$ will be considered correct.
- Attempt every problem, but do not spend too much time on one problem.
- You can leave any standard normal percentile as z-score notation, for example 75% percentile of standard normal would be $z_{75\%}$. Here are some z-scores you may find useful if you choose to compute the numerical answer:
  - $-z_{2.5\%} = z_{97.5\%} = 1.96$
  - $-z_{5\%} = z_{95\%} = 1.64$
- Common distribution reference:

Distribution Name	PDF (PMF)	CDF	Mean	Variance
Bernoulli(p)	$\mathbb{P}(X = x) = p^x(1-p)^{1-x}$	$F_X(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$	p	$p(1-p)$
Binomial(n, p)	$\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$	a piece-wise step function	np	$np(1-p)$
Poisson(λ)	$\mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$	a piece-wise step function	λ	λ
Uniform$[a, b]$	$f_X(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$	$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal(μ, σ^2)	$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp[\frac{-(x-\mu)^2}{2\sigma^2}]$	$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp[\frac{-(t-\mu)^2}{2\sigma^2}]dt$	μ	σ^2
Exponential	$f_X(x) = \begin{cases} 0 & x < 0 \\ \lambda \exp(-\lambda x) & x \ge 0 \end{cases}$	$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - \exp(-\lambda x) & x \ge 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

1 Multiple Choices: select one correct answer only! 3 points each.

1.1 The outcome of throwing a fair die once is a random variable. What distribution does it follow?

A. Discrete uniform distribution.

B. Continuous uniform distribution.

C. Bernoulli distribution.

D. Binomial distribution.

E. Multinomial distribution.

1.2 The number of times you get the outcome "5" when throwing a die 5 times is a random variable. What distribution does it follow?

A. Discrete uniform distribution.

B. Continuous uniform distribution.

C. Bernoulli distribution.

D. Binomial distribution.

E. Multinomial distribution.

1.3 You throw a special die with 7 faces, and take the outcome number divided by 2, the remainder is a random variable. What distribution does it follow?

A. Discrete uniform distribution.

B. Continuous uniform distribution.

C. Bernoulli distribution.

D. Binomial distribution.

E. Multinomial distribution.

1.4 You throw a die 20 times, and count how many times you get the outcome larger than 3. This count is a random variable. What distribution does it follow?

A. Discrete uniform distribution.

B. Continuous uniform distribution.

C. Bernoulli distribution.

D. Binomial distribution.

E. Multinomial distribution.

1.5 You throw a die 100 times, and keep a tab of how many times you get each of the 6 outcomes. These counts are random variables. What distribution do they follow?

A. Discrete uniform distribution.

B. Continuous uniform distribution.

C. Bernoulli distribution.

D. Binomial distribution.

E. Multinomial distribution.

1.6 $X_1, X_2, \cdots X_n$ **is i.i.d from an unknown population distribution with mean μ and variance σ^2. What can you conclude about its sample mean $\bar{X}_n$?**

A. $\bar{X}_n$ **is a statistic that has a sampling distribution depends on sample size n.**

B. $\bar{X}_n$ **converges to μ in distribution.**

C. $\bar{X}_n$ **converges to μ in probability.**

D. The sampling distribution of $\bar{X}_n$ converges to a normal distribution centered at μ.

E. All of the above.

1.7 If two random variables X and Y are independent, which of the following statement is INCORRECT ?

A. We can derive the joint density function $f_{X,Y}(x, y)$ if we know both marginal density functions $f_X(x)$ and $f_Y(y)$.

B. We can conclude there is no overlap in the support of X and support of Y.

C. We can calculate the expectation $\mathbb{E}[XY]$ as long as we know $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

D. We can calculate the variance of $X + Y$ if we know the variance of X and Y respectively.

E. We can conclude the correlation coefficient between X and Y is 0.

1.8 You estimate a 90% confidence interval(CI) for the sample median to estimate population median, which of the following statement is CORRECT ?

A. If you collect 100 datasets with the same sample size and construct 100 CIs, one for each dataset, about 90 of the CIs should contain the population median.

B. If you randomly sample one data point from the population, there is 90% probability the data is in the CI.

C. If you collect a new dataset with the same sample size, there is 90% probability the median of your new dataset is in the CI .

D. The population median must be in the CI .

E. There is 90% probability that the population median is in the CI.

1.9 You observe i.i.d data $X_1, X_2, \cdots X_n$, and you use nonparametric bootstrap technique to generate one bootstrap re-sample. What is the probability that X_1 appears in this bootstrap re-sample?

A. $\frac{1}{n}$. **B.** $\frac{n-1}{n}$. **C.** $(1 - \frac{1}{n})^n$. **D.** $1 - (1 - \frac{1}{n})^n$. **E.** $1 - (\frac{1}{n})^n$.

1.10 A box contains two coins: a regular fair coin and one fake two-headed coin. I choose a coin at random and toss it once and result in a head. What is the probability that I toss the same coin again and it will also result in a head?

A. 1/2. B. 1/3. C. 2/3. D. 3/4 E. 5/6.

2 The joint probability mass function (PMF) of X and Y are giving in the following table:

$\mathbb{P}(X, Y)$	X = 1	X=2	X=3
Y=1	0.1	0.2	0.1
Y=2	0.05	0.05	0.4
Y=3	0.02	0.03	0.05

2.1 Find the marginal distribution for X and Y respectively. (10 points)

2.2 Find the conditional distribution for X conditioned on $Y = 1$. (5 points)

2.3 Find the conditional expectation of X conditioned on $Y = 1$. (5 points)

2.4 Find the conditional variance of X conditioned on $Y = 1$. (5 points)

3 A Poisson random variable is a discrete random variable that can take any integer values. The PMF of a Poisson random variable X is:

$$\mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, 2, \cdots$$

where λ is the single parameter.

3.1 Write down the likelihood function assuming i.i.d data $x_1, x_2, \cdots x_n$. (10 points)

3.2 Find the maximum likelihood estimate for λ. (5 points)

4 You observe data $X_1 = x_1, X_2 = x_2, \cdots X_n = x_n \overset{\text{i.i.d.}}{\sim}$ Uniform$[0, 1]$.

4.1 **Define a new random variable $Y = \min(X_1, X_2, \cdots X_n)$. Find the PDF of Y. (5 points)**

4.2 **Your boss told you the correct population distribution is not Uniform$[0, 1]$ but instead Uniform$[\theta, 1]$ where $0 \leq \theta < 1$ is unknown, and would like you to estimate it. So you go ahead and find the maximum likelihood estimate of θ. (10 points)**

4.3 **Your boss do not understand your maximum likelihood estimate, but knows that for a Uniform$[a, b]$, the expectation is $(a+b)/2$ and the variance is $(b-a)^2/12$. Your boss insists to estimate θ by matching the mean, i.e.**

$$\mathbb{E}[X] = \frac{1+\theta}{2} \approx \bar{X}_n \quad \Rightarrow \hat{\theta} = 2\bar{X}_n - 1$$

Your boss asks you to assess this estimate by calculating the means square error for $\hat{\theta}$. (10 points)

4.4 **You decide to go with the estimate your boss suggested. However, in addition to the point estimate $\hat{\theta} = 2\bar{X}_n - 1$, you strongly recommend to report a confidence interval as well. Give a 95% confidence interval using sample standard deviation s. (5 points)**