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:
  - $z_{2.5\%} = -1.96$
  - $z_{5\%} = -1.64$
  - $z_{50\%} = 0$
  - $z_{95\%} = 1.64$
  - $z_{97.5\%} = 1.96$
- Feel free to ask questions if you find any problem statement not clear.
- 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 Suppose you flip a coin until a head occurs, what is the sample space?

A. {H, T}.

B. {$\emptyset$, H, T}.

C. {$\emptyset$, H, T, {H or T}}.

D. {H, TH, TTH, TTTH, TTTTH, ...}.

E. {H, T, TH, HT, TTH, HHT, TTTH, HHHT, ...}.

1.2 Which of the following is a Bernoulli random variable?

A. Number of heads before getting the first tail when tossing a fair coin.

B. Number of tails before getting the first head when tossing a biased coin.

C. Number of heads when tossing a fair coin 20 times.

D. Number of tails when tossing a biased coin 20 times.

E. A discrete random variable that can only take values 0 and 1.

1.3 Which of the following is a Binomial random variable?

A. Number of heads before getting the first tail when tossing a fair coin.

B. Number of tails before getting the first head when tossing a biased coin.

C. Number of heads when tossing a fair coin 20 times.

D. Number of tails when tossing a biased coin 20 times.

E. Both C and D.

1.4 Which of the following is a discrete Uniform random variable?

A. Number of times needed to throw a fair die until first 1 appears.

B. Total number of 1s when throwing a fair die 100 times.

C. Outcome of throwing a fair die once.

D. Outcome of throwing a biased die once.

E. All of them.

1.5 Which of the following is a Multinomial random variable?

A. Number of times each outcome appears when throwing a fair die 100 times.

B. Number of times each outcome appears when throwing a biased die 100 times.

C. Number of students belonging to each residential college in a random sample of 100 Rice students.

D. Number of supporters for Democratic, Republican and Independent parties in a election poll of 100.

E. All of them.

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

A. $\bar{X}_n$ **is unbiased.**

B. The Mean Squared Error of $\bar{X}_n$ **is** σ^2/n.

C. The sampling variance of $\bar{X}_n$ **is** σ^2/n.

D. $\bar{X}_n$ **is consistent.**

E. All of the above.

1.7 Which of the following statement is true for two random variables X and Y without any further assumption?

A. $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

B. $\mathbb{V}[X+Y] = \mathbb{V}[X] + \mathbb{V}[Y]$.

C. $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

D. $\mathbb{V}[X+Y] = \mathbb{V}[X] + \mathbb{V}[Y] + \mathbf{Cov}(X,Y)$.

E. All of the above.

1.8 Which of the following statement is true about confidence intervals?

A. Given everything else equal, larger sample size leads to wider confidence intervals.

B. Given everything else equal, higher confidence level leads to wider confidence interval.

C. Given everything else equal, wider confidence intervals mean lower sampling variance.

D. Given everything else equal, wider confidence intervals are better estimate.

E. All of the above.

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 does NOT 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 psychological study consists participates of equal number of humans and monkeys. A multiple choice quiz question is answered correctly with probability 0.9 by human, and is answered by randomly picking one of the 5 possible answers by monkeys. If the question is answered correctly, what is the probability that it is answered by a monkey?

A. 0.9. B. 0.5. C. 0.25. D. 0.18 E. 0.125.

2 The joint probability density function (PDF) of X and Y are giving in the following:

$$f_{X,Y}(x,y) = \begin{cases} 2 & \text{for } 0 \leq x \leq 1, 0 \leq y \leq 1, x+y \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

2.1 Find $f_X(x)$ and $f_Y(y)$: the marginal PDF of X and Y respectively.

(10 points)

2.2 Find $f_{X|Y}(x|y)$: the conditional PDF of X conditioned on Y. What distribution is X conditioned on Y?

(10 points)

2.3 Find $\mathbb{E}[X|Y]$: the conditional expectation of X conditioned on Y.

(5 points)

2.4 Show that the $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$.

Hint: $\mathbb{E}[X|Y]$ should be a function of Y, therefore the outside expectation will be taken with respect to Y.

(5 points)

3 An exponential random variable is a non-negative continuous random variable, often used to model the amount of time until some specific event occurs. The exponential distribution family has a single parameter λ and its PDF $f_X(x;\lambda)$ is given in the common distribution reference.

3.1 Write down the likelihood function assuming i.i.d data $x_1, x_2, \cdots x_n \sim f_X(x;\lambda)$.

(5 points)

3.2 Find the maximum likelihood estimator $\hat{\lambda}^{\mathbf{MLE}}$.

(5 points)

3.3 Write pseudo code to estimate a 95% Confidence Interval for λ using bootstrap. Assume you have observed the data in 3.1, and correctly calculated $\hat{\lambda}^{\mathbf{MLE}}$.

(10 points)

4 You work in the department of quality control for a manufacturer. For a particular type of products, assume whether each item being defect follows independent and identical Bernoulli distribution.

4.1 **Your boss wants you to come up with a point estimate $\hat{p}$ for the population defect rate p with the following properties:**

$$\mathbb{P}(|\hat{p} - p| \leq \epsilon) \geq 1 - \alpha$$

Here ϵ is tolerance of error margin, and $1 - \alpha$ is confidence level. This means that the probability of your point estimate being within the tolerated error margin is guaranteed at the confidence level. What point estimate do you suggest and why?

(10 points)

4.2 **Your boss picks $\epsilon = 0.01$ and $\alpha = 0.05$, and asks you to estimate the minimum number of products to sample in order to come up with the point estimate that is within error of 0.01 at confidence level 0.95.**

Hint: The variance for Bernoulli(p) is $p(1 - p)$, and you can use $p(1 - p) \leq 1/4$.

(10 points)