STA 9715 - Applied Probability

In-Class Test 1

This is a closed-note, closed-book exam.
You may not use any external resources other than a (non-phone) calculator.

Name: ____

Instructions

This exam will be graded out of **100 points**.

All questions are worth the same amount (5 points), but individual questions within a section may vary in difficulty. You need to use your time wisely. Skip questions that are not easy to answer quickly and return to them later.

You have one hour to complete this exam from the time the instructor says to begin. The instructor will give time warnings at: 30 minutes, 15 minutes, 5 minutes, and 1 minute.

When the instructor announces the end of the exam, you must stop **immediately**. Continuing to work past the time limit may be considered an academic integrity violation.

Write your name on the line above *now* before the exam begins.

Each question is followed by a dedicated answer space. Place all answers in the relevant spot. Answers that are not clearly marked in the correct location **will not** receive full credit. Partial credit may be given at the instructor's discretion.

Mark and write your answers clearly: if I cannot easily identify and read your intended answer, you may not get credit for it.

Additional pages for scratch work are included at the end of the exam packet.

A formula sheet may be found on the first page of the exam packet.

This is a closed-note, closed-book exam.
You may not use any external resources other than a (non-phone) calculator.

STA9715 - Test 1 - Formula Sheet

Foundations:

- Probabilities are Non-Negative: $0 \leq \mathbb{P}(A) \leq 1$ for all *events* A
- Probability of Entire Sample Space: $\mathbb{P}(\Omega) = 1$
- Probability of Empty Set: $\mathbb{P}(\emptyset) = 0$
- 'Union Bound': $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$ with equality if A, B are *disjoint* $(A \cap B = \emptyset)$
- Complements: $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- Naive Probability: $\mathbb{P}(A) = |A|/|\Omega|$
- Counting: ${}_nP_k = \frac{n!}{(n-k)!}$ (permutations - *ordered* choice of k from n); ${}_nC_k = \binom{n}{k} = \frac{n!}{(n-k)!k!}$ (combinations - *unordered*)
- DeMorgan's Laws: $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$

Random Variables (Discrete):

- $\mathbb{P}(X = x) = f_X(x)$ is the *probability mass function* of X. Gives the probability of observing $X = x$ exactly
- $\sum_{x \in \mathsf{supp}(X)} \mathbb{P}(X = x) = 1$. $0 \leq f_X(x) \leq 1$

Random Variables (Continuous):

- *Probability density function* of X: $f_X(x)$ Integrates to give the probability X in an interval: $P(a \leq X \leq b) = \int_a^b f_X(x)\,\mathrm{d}x$
- $\int_{x \in \mathsf{supp}(X)} f_X(x)\,\mathrm{d}x = 1.f_X(x)$ may be greater than 1 for small ranges; never negative

Moments:

- Expectation: $\mathbb{E}[X] = \sum x * f_X(x)$ or $\int x * f_X(x)\,\mathrm{d}x$
- Linearity of Expectation: $\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$
- Expectation of Functions: $\mathbb{E}[g(X)] = \sum g(x)f_X(x)$ or $\int g(x)f_X(x)\,\mathrm{d}x$
- Variance: $\mathbb{V}[X] = \mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \geq 0$
- $\mathbb{V}[aX + bY + c] = a^2\mathbb{V}[X] + b^2\mathbb{V}[Y]$ only if X, Y are uncorrelated. (Independent implies uncorrelated)
- Expectation of Indicators: $\mathbb{E}[1_{\cdot \in A}(X)] = \mathbb{P}(X \in A)$. Useful to reduce probabilities to linear expectation calculations

Conditional Probabilities:

- $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$. Special case: $A \subseteq B \implies \mathbb{P}(A|B) = \mathbb{P}(A)/\mathbb{P}(B)$
- Bayes' Rule: $\mathbb{P}(A|B) = \mathbb{P}(B|A) * \mathbb{P}(A)/\mathbb{P}(B) = \mathbb{P}(B|A) * \mathbb{P}(A)/(\mathbb{P}(B|A) * \mathbb{P}(A) + \mathbb{P}(B|A^c) * \mathbb{P}(A))$
- Law of Total Probability: if $\{A_j\}$ are a disjoint partition of Ω then $\mathbb{P}(B) = \sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)$
- Law of Total Expectation: $\mathbb{E}[X] = \sum \mathbb{E}[X|A_i]\mathbb{P}(A_i)$ for partition $\{A_i\}$ or $\mathbb{E}_X[X] = \mathbb{E}_Y[\mathbb{E}_X[X|Y]]$
- Law of Total Variance: $\mathbb{V}[X] = \mathbb{E}_Y[\mathbb{V}_X[X|Y]] + \mathbb{V}_Y[\mathbb{E}_X[X|Y]]$
- Independence: A, B are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Equivalently $\mathbb{P}(B|A) = \mathbb{P}(B)$ and $\mathbb{P}(A|B) = \mathbb{P}(A)$

Distributions:

- Bernoulli: $X \sim \mathrm{Bern}(p)$. $\mathbb{P}(X = 1) = p; \mathbb{P}(X = 0) = 1 - p$. $\mathbb{E}[X] = p$; $\mathbb{V}[X] = p(1 - p)$. 'Coin Flip'
- Binomial: Sum of n IID Bernoulli: $P(X = x) = \binom{n}{x}p^x(1 - p)^{n-x}$. $\mathbb{E}[X] = np$. $\mathbb{V}[X] = np(1 - p)$
- Poisson: Limit of $n \to \infty, p \to 0$ Binomial. $X \sim \mathrm{Pois}(\mu)$. $\mathbb{P}(X = x) = \mu^x e^{-\mu}/x!$. $\mathbb{E}[X] = \mathbb{V}[X] = \mu$
- Geometric: IID Bernoulli 'until' 1st success: $X \sim \mathrm{Geom}(p)$. $\mathbb{P}(X = x) = p(1 - p)^{x-1}$. $\mathbb{E}[X] = 1/p$. $\mathbb{V}[X] = (1 - p)/p^2$. Memoryless
- Hypergeometric: Population N w/K successes & n total draws. $\mathbb{P}(X = k) = \binom{K}{k}\binom{N-K}{n-k}/\binom{N}{n}$. $\mathbb{E}[X] = nK/N$. $\mathbb{V}[X] = n * K/N * (N - K)/N * (N - n)/(N - 1)$
- Normal: $X \sim \mathcal{N}(\mu, \sigma^2)$. $f_X(x) = e^{-(x-\mu)^2/2\sigma^2}/\sqrt{2\pi\sigma^2}$. $\mathbb{E}[X] = \mu$. $\mathbb{V}[X] = \sigma^2$. 'Bell curve'
- Exponential: $X \sim \mathrm{Exp}(\lambda)$. $f_X(x) = \lambda e^{-\lambda x}$. $\mathbb{E}[X] = \lambda^{-1}$. $\mathbb{V}[X] = \lambda^{-2}$. Continuous geometric
- Uniform (Discrete and Continuous). $\mathbb{E}[\mathrm{DUnif}\{a, \dots, b\}] = \mathbb{E}[\mathrm{CUnif}([a, b])] = (a + b)/2$. $\mathbb{V}[\mathrm{CUnif}([a, b])] = (b - a)^2/12$

Q1: In an MLB Divisional Series, two teams play a sequence of games against each other, and the first team to win three games wins the series. Let p the probability that Team A wins an individual game, and assume that the games are independent. What is the probability that A wins the series?

Answer to Q1:__________________________________

Q2: For a group of 7 people, find the probability that all 4 seasons (winter, spring, summer fall) occur at least once each among their birthdays, assuming that all seasons are equally likely.

Answer to Q2:__________________________________

Q3: Suppose that Ashley is playing a guessing game, where she has a 25% chance of answering a given question correctly (IID). If she answers a question correctly, she gets one point. Let Q be the total number of questions required for her to get five total points. What is the variance of Q?

Answer to Q3:__________________________________

Q4: A group of 50 people are comparing their birthdays; as usual, assume their birthdays are independent, are not February 29, *etc.* Find the expected number of pairs of people with the same birthday.

Answer to Q4:________________________________

Q5: A certain family has 6 children, consisting of 3 boys and 3 girls. Assuming that all birth orders are equally likely, what is the probability that the 3 eldest children are the 3 girls?

Answer to Q5:________________________________

Q6: Suppose $X \sim \text{Poisson}(\lambda)$, that is X is a Poisson random variable with mean $\lambda = \mathbb{E}[X]$. What is the second (non-central) moment of X, *i.e.*, $\mathbb{E}[X^2]$?

Answer to Q6:________________________________

Q7: Suppose that course grades are distributed as follows:

Grade	A	B	C	D	F
GPA Points	4	3	2	1	0
Fraction of Class	30%	30%	15%	5%	20%

Given that a student did not receive an F, what is the probability they receive an A or a B in the course?

Answer to Q7:__________________________________

Q8: Suppose that the number of Baruch students to win the lottery each year is Poisson distributed with mean 2. What is the probability that an above average (*i.e.*, above mean) number of Baruch students win the lottery next year?

Answer to Q8:__________________________________

Q9: According to the CDC, men who smoke are 23 times more likely to develop lung cancer than men who don't smoke. Also according to the CDC, 21.6% of men in the US smoke. What is the probability that a man in the US is a smoker, given that he develops lung cancer?

Answer to Q9:________________________________

Q10: Let T be the time until a radioactive particle decays and suppose that $T \sim \text{Exponential}(\lambda)$. The *half-life* of the particle is the time at which there is a 50% chance that the particle has decayed (*i.e.*, the median of T). Find the half-life of the particle in terms of λ Hint: recall $\int e^{ax}\,\mathrm{d}x = e^{ax}/a$.

Answer to Q10:________________________________

Q11: Suppose the random variables (X, Y) have joint PDF $f_{(X,Y)}(x, y) = 4xy$ with support on the unit square $[0, 1]^2$. (That is, both X and Y can take any value in $[0, 1]$). What is the conditional expectation of X given $Y > 0.5$

Answer to Q11:________________________________

Q12: Five students board the Baruch express elevators at the same time from the second floor. Assuming that their destinations are uniformly random (*i.e.*, that they exit at each floor with equal probability), what is the probability that no students exit on the 11th floor? (Recall that the Baruch express elevators stop at the 5th, 8th, and 11th floors.)

Answer to Q12:________________________________

Q13: Let $X \sim \mathcal{N}(3, \sqrt{2}^2)$ and $Y \sim \mathcal{N}(1, \sqrt{2}^2)$ be independent normal random variables. Calculate $\mathbb{P}(X < Y)$. You may leave your answer in terms of the standard normal CDF $\Phi(\cdot)$.
Hint: Use the fact that $Z_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Z_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ implies $aZ_1 + bZ_2 + c \sim \mathcal{N}(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2)$ for Z_1, Z_2 independent.

Answer to Q13:________________________________

Q14: Suppose a random variable X takes continuous values between 2 and 5. Suppose further that its PDF is $f_X(x) = cx^2$ for some unknown c. What is c?

Answer to Q14:________________________________

Q15: In the Gregorian calendar, each year has either 365 days (normal) or 366 (leap year). A year is randomly chosen with probability 3/4 of being a normal year and 1/4 of being a leap year. Find the mean and variance of the number of days in the chosen year.

Answer to Q15:________________________________

Q16: The random variable X has PDF $f(x) = 12x^2(1-x)$ and support $[0, 1]$. Compute $\mathbb{P}(0 \le X \le 1/2)$.

Answer to Q16:________________________________

Q17: X is a mixture distribution defined as follows: $X_1 \sim \mathcal{N}(0, 5^2)$, $X_2 \sim \text{Poisson}(5)$, and $X_3 \sim \text{ContinuousUniform}([-3, 3])$. Z is uniform on $\{1, 2, 3\}$ and $X = X_Z$, that is, X is the value of each mixture arm with equal probability. Z, X_1, X_2, X_3 are all independent. Compute the variance of X.

Answer to Q17:________________________________

Q18: Four players, named A, B, C, and D, are playing a card game. A standard well-shuffled deck of cards is dealth to the players so each player receives a 13 card hand. How many possibilities are there for the hand that play A will get? (Within a hand, the order in which cards were received doesn't matter.)

Answer to Q18:________________________________

Q19: Let X be the number of Heads in 10 fair coin tosses (IID). Find the conditional variance of X, given that the first two tosses both land Heads.

Answer to Q19:________________________________

Q20: Suppose that, before a major election, a polling company contacts a large number of likely voters and successfully asks 1000 voters who they intend to vote for. Candidate A's supporters have a 100% chance of answering the poll if contacted, while Candidate B's supporters have only a 50% chance of answering the poll. If 75% of respondents say they intend to vote for Candidate A, what is A's expected fraction of total votes cast on election day?

(You may assume that supporters of candidate A and B are equally likely to vote and that they only differ in their likelihood to respond to the poll. You may also assume only two candidates. You may also assume respondents don't lie and are otherwise fully representative, except for differential response rates.)

Answer to Q20:________________________________

(Blank page for scratch work - not graded)

(Blank page for scratch work - not graded)

(Blank page for scratch work - not graded)