STA 9715 - Applied Probability

In-Class Test 2

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.

(Blank page for scratch work - not graded)

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: $\mathbb{V}[X] = \text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \geq 0$
- $\mathbb{V}[aX + bY + c] = a^2\mathbb{V}[X] + b^2\mathbb{V}[Y]$ 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: $X \sim \text{Bern}(p)$. $\mathbb{P}(X = 1) = p; \mathbb{P}(X = 0) = 1 - p$. $\mathbb{E}[X] = p$; $\mathbb{V}[X] = p(1 - p)$. '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. $X \sim \text{Pois}(\mu)$. $\mathbb{P}(X = x) = \mu^x e^{-\mu}/x!$. $\mathbb{E}[X] = \mathbb{V}[X] = \mu$
- Geometric: IID Bernoulli 'until' 1st success: $X \sim \text{Geom}(p)$. $\mathbb{P}(X = x) = p(1 - p)^{x-1}$. $\mathbb{E}[X] = 1/p$. $\mathbb{V}[X] = (1 - p)/p^2$. 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: $X \sim \text{Exp}(\lambda)$. $f_X(x) = \lambda e^{-\lambda x}$. $\mathbb{E}[X] = \lambda^{-1}$. $\mathbb{V}[X] = \lambda^{-2}$. Continuous geometric
- Uniform (Discrete and Continuous). $\mathbb{E}[\text{DUnif}\{a, \dots, b\}] = \mathbb{E}[\text{CUnif}([a, b])] = (a + b)/2$. $\mathbb{V}[\text{CUnif}([a, b])] = (b - a)^2/12$

STA9715 - Test 2 - Formula Sheet

Inequalities

- If $\mathsf{support}(X)$ is non-negative, $\mathbb{P}(X > x) \leq \mathbb{E}[X]/x$ (Markov)
- For any $X \sim (\mu, \sigma^2)$, $\mathbb{P}(|X - \mu| \geq k\sigma) \leq 1/k^2$ or $\mathbb{P}(|X - \mu| \geq k) \leq \sigma^2/k^2$ (Chebyshev)

Vector Arithmetic and Linear Algebra

- Vector addition $\boldsymbol{x}+\boldsymbol{y} = (x_1+y_1, x_2+y_2, \dots, x_n+y_n)$ Vector-times-scalar $\alpha\boldsymbol{x} = (\alpha x_1, \alpha x_2, \dots, \alpha x_n)$. (Both elementwise)
- Two-vector (dot / inner) product yields scalar: $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x} \cdot \boldsymbol{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n$
- Vector norm: $\|\boldsymbol{x}\| = \sqrt{\sum_i x_i^2}$ - generalizes length or absolute value. Angle between vectors: $\cos\angle(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{x}, \boldsymbol{y} \rangle / \|\boldsymbol{x}\|\|\boldsymbol{y}\|$
- Matrix-vector multiplication: yields a vector: $\boldsymbol{Ax}$ element i is dot product of row i of $\boldsymbol{A}$ with $\boldsymbol{x}$.
- Matrix-matrix multiplication: yields a matrix: $\boldsymbol{AB}$ element (i,j) is dot product of row i of $\boldsymbol{A}$ with column j of $\boldsymbol{B}$.
- Quadratic form: $\langle \boldsymbol{x}, \boldsymbol{Ax} \rangle = \|\boldsymbol{x}\|_{\boldsymbol{A}}^2 = \sum_{(i,j)} A_{ij}x_ix_j$. $\boldsymbol{A}$ is *positive-definite* if all quadratic forms are positive (for $\boldsymbol{x} \neq \boldsymbol{0}$)
- Identity matrix $\boldsymbol{I}$ is ones on diagonal; zeros elsewhere. $\boldsymbol{Ix} = \boldsymbol{x}$ and $\boldsymbol{AI} = \boldsymbol{IA} = \boldsymbol{A}$ for all $\boldsymbol{x}, \boldsymbol{A}$

Random Vectors

- Expectation is coordinate-wise: $\mathbb{E}[\boldsymbol{X}] = (\mathbb{E}[X_1], \mathbb{E}[X_2], \dots, \mathbb{E}[X_n])$
- Linear transforms: $\mathbb{E}[\boldsymbol{a} + \alpha\boldsymbol{X} + \beta\boldsymbol{Y}] = \boldsymbol{a} + \alpha\mathbb{E}[\boldsymbol{X}] + \beta\mathbb{E}[\boldsymbol{Y}]$ and $\mathbb{E}[\langle \boldsymbol{a}, \boldsymbol{X} \rangle] = \langle \boldsymbol{a}, \mathbb{E}[\boldsymbol{X}] \rangle$ Does not assume independence
- PDFs work via *multiple* integrals: $\mathbb{P}(\boldsymbol{X} \in A) = \iiint_A f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$. CDFs are difficult
- If joint PDF factorizes $f_{(X,Y)}(x,y) = f_X(x)f_Y(y)$ then $X \perp\!\!\!\perp Y$ (independence)
- Marginal PDF: $f_X(x) = \int_{-\infty,\infty} f_{(X,Y)}(x,y)\,\mathrm{d}y$
- Conditional PDF: $f_{X|Y=y}(x) = f_{(X,Y)}(x,y)/f_Y(y)$. General form: $f_{X|Y\in A}(x) = \int_A f_{(X,Y)}(x,y)\mathrm{d}y/\mathbb{P}(Y \in A)$

Covariance

- Covariance of two scalars: $\mathbb{C}[X,Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ (can be positive or negative)
- Self-covariance is variance: $\mathbb{C}[X,X] = \mathbb{V}[X]$
- Linear transforms: $\mathbb{C}[aX+b, cY+d] = ac\,\mathbb{C}[X,Y]$. For random vector $\boldsymbol{X}$ and fixed matrix $\boldsymbol{A}$: $\mathbb{V}[\boldsymbol{\mu}+\boldsymbol{AX}] = \boldsymbol{A}\mathbb{V}[\boldsymbol{X}]\boldsymbol{A}^T$.
- Correlation: $\rho_{X,Y} = \mathbb{C}[X,Y]/\sqrt{\mathbb{V}[X]\mathbb{V}[Y]}$
- Variance of a random vector is a *(co)variance matrix*: $\mathbb{V}[\boldsymbol{X}]_{ij} = \mathbb{C}[X_i, X_j]$
- Covariance quadratic forms give variance of linear combinations: $\mathbb{V}[\langle \boldsymbol{a}, \boldsymbol{X} \rangle] = \langle \boldsymbol{a}, \mathbb{V}[\boldsymbol{X}]\boldsymbol{a} \rangle = \sum_{ij} a_ia_j\mathbb{C}[X_i, X_j] \geq 0$
- Independence implies uncorrelated, but not the other way: $X \perp\!\!\!\perp Y \implies \mathbb{C}[X,Y] = 0 \Leftrightarrow \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

Normal Distribution

- Standard normal distribution. $Z \sim \mathcal{N}(0,1)$. Mean Zero + Variance 1
- Standard normal PDF - $\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$. Standard normal CDF $\Phi(z) = \int_{-\infty}^{z} \phi(x)\,\mathrm{d}x$ - no closed form.
- General normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$ - generated by scale+shift of standard normal $X \stackrel{d}{=} \mu + \sigma Z$.
- Normal PDF via standardization (z-score): $f_X(x) = \phi(\frac{x-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/2\sigma^2}$. CDF: $\Phi(\frac{x-\mu}{\sigma})$.
- Multivariate normal parameterized by mean vector and (co)variance matrix: $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Standard multi-normal: $\boldsymbol{Z} \sim \mathcal{N}_n(\boldsymbol{0}_n, \boldsymbol{I}_n)$. PDF $f_{\boldsymbol{Z}}(\boldsymbol{z}) = (2\pi)^{-n/2}e^{-\|\boldsymbol{z}\|^2/2}$.
- General multi-normal $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}$ where $\boldsymbol{\Sigma}^{1/2}$ is a matrix square root (Cholesky or symmetric).
- Bivariate normal PDF $f_{(X,Y)}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2[1-\rho^2]}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right)$
- Multivariate normal: any linear combination (weighted sum) of X_i is normal.
- If $\mathbb{C}[X_i, X_j] = 0$, then $X_i \perp\!\!\!\perp X_j$ (for multi-normal, uncorrelated implies independent)
- If $\boldsymbol{Z}$ is a standard normal n-vector, $\|\boldsymbol{Z}\|^2 = \sum_{i=1}^n Z_i^2$ has a χ^2 distribution with n degrees of freedom

Q1: Let $X \sim \mathcal{N}(3, 1)$ be a Gaussian random variable. Additionally, conditional on $X = x$, let Y and Z be independent $\mathcal{N}(x, 1)$ random variables. What is $\mathbb{E}[X + Y + Z]$?

Answer to Q1:________________________________

Q2: Let $X \sim \mathcal{N}(3, 1)$ be a Gaussian random variable. Additionally, conditional on $X = x$, let Y and Z be independent $\mathcal{N}(x, 1)$ random variables. What is $\mathbb{V}[X + Y + Z]$?

Answer to Q2:________________________________

Q3: What is the variance of a χ^2 random variable with 5 degrees of freedom. You may use the fact that the fourth moment of a standard normal random variable is $\mathbb{E}[Z^4] = 3$.

Answer to Q3:________________________________

Q4: Let P be a point chosen uniformly at random on the surface of the Earth. What is the probability that P falls in both the Northern and the Western hemispheres?

Answer to Q4:________________________________

Q5: Suppose X follows a Pareto distribution with CDF $F_X(x) = 1 - 1/x^2$ supported on $[1, \infty)$. What is $\mathbb{E}[X|X > 8]$?
Hint: The Pareto distribution satisfies a 'restarting' property: $F_{X|X>x_0}(x) = 1-(x_0/x)^2$ *supported on* $[x_0, \infty)$.

Answer to Q5:________________________________

Q6: Let X and Y have joint PDF $f_{(X,Y)}(x, y) = cxy$ for $0 \leq x \leq y \leq 1$. What is c?

Answer to Q6:________________________________

Q7: On average, Baruch students have to wait 5 minutes to catch an elevator. Use Markov's inequality to give an upper bound on the probability that it takes more than 15 minutes to get an elevator.

Answer to Q7:______________________________

Q8: Let $X_1 = 3 + 2Z_1 + 4Z_2$ and $X_2 = 5 + 4Z_1 - 6Z_2$ where Z_1, Z_2 are indpendent standard normal variables. What is the correlation of X_1 and X_2, *i.e.*, ρ_{X_1,X_2}?

Answer to Q8:______________________________

Q9: Let X, Y have a joint distribution with PDF of the form

$$f_{(X,Y)}(x, y) = c \exp\left\{-\frac{x^2}{8} - \frac{y^2}{50}\right\}$$

supported on $\mathbb{R}^2$ (that is, X, Y can both be any real number). What is c?

Answer to Q9:______________________________

Q10: Suppose that a father and son are selected from a population whose heights are $\mathcal{N}(72, 6^2)$; suppose further that the correlation between heights within a family is $\rho = 1/3$. What is the probability that the son is more than 3 inches taller than his father?

Answer to Q10:________________________________

Q11: Let $X \sim \mathcal{N}(3, 5^2)$ and $Y \sim \mathcal{N}(-5, 3^2)$ be independent random variables. What is $\mathbb{V}[XY]$?

Answer to Q11:________________________________

Q12: Let X have a log-normal distribution given by $X = e^{3+2Z}$ where Z is a standard normal distribution. What is the *mode* of X?
Hints: Compute the CDF of X first and use it to derive a PDF.
Remember the derivative of $\ln(x)$ *is* $1/x$ *and apply the chain rule carefully.*
The mode of a continuous RV is the value x *that maximizes the PDF* $f_X(x)$.

Answer to Q12:________________________________

Q13: Let (X, Y) be a point selected uniformly at random from (the interior of) the triangle with corners at $(-1, -1)$, $(+1, -1)$ and $(0, +1)$. What is the probability that X is positive?
Hint: Remember that the area of a triangle is $1/2 * \mathsf{base} * \mathsf{height}$ *and draw a picture.*

Answer to Q13:________________________________

Q14: Let $\boldsymbol{Z}$ be a vector of 5 independent standard normal random variables. What is $\mathbb{E}[\|\boldsymbol{Z}\|^2]$?

Answer to Q14:________________________________

Q15: Let (X, Y) be bivariate normal with $X \sim \mathcal{N}(0, 3^2)$ and $Y \sim \mathcal{N}(0, 4^2)$ marginally and correlation $\rho_{X,Y} = 50\%$. What is the value of c such that $Y - cX$ is independent of X?

Answer to Q15:________________________________

Q16: Suppose X has a Pareto distribution with parameter 5 and a PDF given by $f_X(x) = 5/x^6$ supported on $[1, \infty)$. What is the *median* of X?

Answer to Q16:____________________________________

Q17: Suppose that (X, Y) are binary random variables generated according to the following joint distribution. What is the *covariance* of X and Y?

	$\mathbb{P}$	X = 0	X = 1
Y	0	$\frac{1}{3}$	$\frac{1}{12}$
	1	$\frac{1}{2}$	$\frac{1}{12}$

Answer to Q17:____________________________________

Q18: Let (X, Y) come from a multivariate normal distribution with mean $\boldsymbol{\mu} = (3, 4)$ and variance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 25 & 9 \\ 9 & 9 \end{pmatrix}$$

What is the correlation of X and Y?

Answer to Q18:______________________________

Q19: The *trace* of a matrix is the sum of the elements on its diagonal. Suppose (X, Y) is a random vector with variance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} 25 & 1 \\ 1 & 4 \end{pmatrix}$. What is the *trace* of the variance matrix of the vector $(Z, W) = (2X + 3Y, X - Y)$?

Hint: Note that $\begin{pmatrix} Z \\ W \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$.

Answer to Q19:______________________________

Q20: Suppose $X_1, \ldots, X_{16}$ are IID samples from a distribution with variance 25 and unknown mean. Using Chebyshev's inequality, give an upper bound on the probability that the sample mean $\overline{X} = \frac{1}{16} \sum_{i=1}^{16} X_i$ is more than 3 units away from the true mean?

Answer to Q20:______________________________

(Blank page for scratch work - not graded)

(Blank page for scratch work - not graded)

(Blank page for scratch work - not graded)