STA 9715 - Applied Probability

Final Exam

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 **200 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 **two hours** to complete this exam from the time the instructor says to begin. The instructor will give time warnings at: 1 hour, 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 will 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.

Formula sheets may be found at the back of the exam packet. You may remove these.

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)

Q1: Suppose X is a continuous random variable with PDF proportional to $x^2(1-x)^5$ and support on $[0, 1]$. What is $\mathbb{E}[X]$?

Answer to Q1:________________________________

Q2: Let T be the time until a radioactive particle decays and suppose that $T \sim \text{Exponential}(10)$. 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). What is the half life of T?

Answer to Q2:________________________________

Q3: Suppose grades on a certain exam are normally distributed with mean 70 and standard deviation 4. Using the Gaussian Chernoff bound, approximate the probability that a given student gets a grade of 90 or higher.

Answer to Q3:________________________________

Q4: Let $X_1, X_2, \ldots$ be a series of IID χ^2_6 random variables and let $Y_1, Y_2, \ldots$ be a series of IID χ^2_4 random variables. (All X_i, Y_j are independent as well.) Let

$$\widetilde{S}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i}{Y_i}$$

What is the limiting value of $\widetilde{X}_n$ as $n \to \infty$?

Hint: What is the distribution of each term $T_i = \frac{X_i}{Y_i}$*? You may want to multiply by a constant.*

Answer to Q4:__________________________________

Q5: A certain family has 5 children, 2 girls and 3 boys. Assuming all birth orders are equally likely, what is the probability that the youngest and the oldest children are both girls?

Answer to Q5:__________________________________

Q6: Let Z_1, Z_2, Z_3 be three IID uniform random variables on the interval $[0, 1]$. What is the CDF of $Z_* = \min\{Z_1, Z_2, Z_3\}$?

Hint: Note that, for a single uniform random variable, $F_{Z_1}(z) = z$ for $z \in [0, 1]$.

Answer to Q6:________________________________

Q7: Suppose X has MGF $\mathbb{M}_X(t) = e^{5t+8t^2}$. What is the standard deviation of X?

Answer to Q7:________________________________

Q8: Four players, named A, B, C, and D, are playing a card game. A standard well-shuffled deck of cards is dealt to the players so each player receives a 5 card hand. How many possibilities are there for the hand that player A will get? (Within a hand, the order in which cards were received doesn't matter.)

Answer to Q8:________________________________

Q9: Let $X_1 = 4 - 2Z_1 + 4Z_2$ and $X_2 = 1 + 3Z_1 - 9Z_2$ where Z_1, Z_2 are independent standard normal variables. What is the covariance of X_1 and X_2, *i.e.*, $\mathbb{C}[X_1, X_2]$?

Answer to Q9:________________________________

Q10: The Rayleigh(β) distribution has CDF $F_Y(y) = 1 - e^{-y^2/2\beta}$. Suppose you have a source of uniform random variables U. Find a transform of U, $h(\cdot)$, such that $h(U) \sim \text{Rayleigh}(5)$.

Answer to Q10:________________________________

Q11: Let $P_1, \ldots, P_{40}$ be IID Poisson random variables, each with mean 2. Using the CLT, approximate the probability that $\mathbb{P}(\sum_{i=1}^{50} P_i > 100)$.

Answer to Q11:________________________________

Q12: Let $X_1, X_2, \ldots$ be IID random variables with mean μ and variance σ^2. Find a value of n (an integer) such that the sample mean $\overline{X}_n$ is within 1 standard deviation of the mean with probability 99% or greater. *Hint: Chebyshev*

Answer to Q12:________________________________

Q13: According to the numbers I just made up, students with dogs are 15 times more likely to loose their homework than students who don't. If 10% of students in a class have a dog, what is the probability that a student who lost their homework has a dog?

Answer to Q13:________________________________

Q14: The train I take to get to work averages 45 minutes. Use Markov's inequality to give an upper bound on the probability it takes me more than 2 hours to get to work.

Answer to Q14:________________________________

Q15: Let $X \sim \text{Gamma}(4, 2)$ and $Y \sim \text{Gamma}(1, 5)$. What is $\mathbb{V}[XY]$?

Answer to Q15:__________________________________

Q16: Suppose that (X, Y) are binary random variables generated according to the following joint distribution. What is the *covariance* of X and Y?

		X	
	$\mathbb{P}$	0	1
Y	0	$\frac{1}{4}$	$\frac{1}{3}$
	1	$\frac{1}{6}$	$\frac{1}{4}$

Answer to Q16:__________________________________

Q17: The *Poisson-Gamma* compound distribution makes draws from a Poisson random variable whose mean is itself Gamma distributed. For example, let $X \sim \text{Gamma}(5, 2)$ and let $Y|X \sim \text{Poisson}(X)$. Then the variable Y has a Poisson-Gamma distribution. What is the (unconditional) variance of Y?

Answer to Q17:__________________________________

Q18: Let (X, Y) come from a multivariate normal distribution with mean $\boldsymbol{\mu} = (5, 2)$ and variance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 36 & -8 \\ -8 & 16 \end{pmatrix}$$

What is the correlation of X and Y?

Answer to Q18:__________________________________

Q19: Let X, Y have a joint distribution with PDF of the form

$$f_{(X,Y)}(x, y) = c \exp\left\{-\frac{x^2}{6} - \frac{y^2}{54} + \frac{xy}{18}\right\} = c \exp\left\{-\frac{2}{3}\left(\frac{x^2}{4} + \frac{y^2}{36} - \frac{xy}{12}\right)\right\}$$

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

Answer to Q19:__________________________________

Q20: A vineyard produces two types of wine: white and red. White wine sells at \$25 per bottle and comprises 75% of their production, with the rest being red selling at \$50 per bottle. Historically, 10% of the white wine is 'spoiled' in transit to America, while 20% of the red is 'spoiled'. If a customer calls asking for a refund for a spoiled bottle, what is the expected size of the refund? (Assume that the purchase price is fully refunded.)

Answer to Q20:____________________________

Q21: Let P be a point chosen uniformly at random on the surface of the Earth. What is the probability that P falls in the Northern or Western hemispheres?

Answer to Q21:____________________________

Q22: Let $\boldsymbol{Z}$ be a vector of 6 independent $\mathcal{N}(1,1)$ random variables. What is $\mathbb{E}[\|\boldsymbol{Z}\|^2]$?

Answer to Q22:____________________________

Q23: Suppose a random variable X takes continuous values between 1 and 4 and that its PDF is of the form $f_X(x) = cx^3$ for some unknown c. What is c?

Answer to Q23:__________________________________

Q24: Suppose that grades in a given course are distributed as follows:

Grade	A	B	C	D	F
GPA Points	4	3	2	1	0
Fraction of Class	25%	40%	25%	5%	5%

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

Answer to Q24:__________________________________

Q25: In a dice game, Player A rolls 2 6-sided dice and takes the higher score; Player B rolls one 6-sided dice. If Player A's score (max of two) is higher than Player B's (one roll), then Player A wins; if Player B's roll is greater than or equal to both of Player A's rolls, then Player B wins (that is, a 'tie' goes to Player B). What is the probability that Player A wins this game?

Answer to Q25:______________________________

Q26: Suppose X follows a Pareto distribution with CDF $F_X(x) = 1 - 1/x^4$ supported on $[1, \infty)$. What is $\mathbb{E}[X|X > 5]$?
Hint: The Pareto distribution satisfies a 'restarting' property: $F_{X|X>x_0}(x) = 1-(x_0/x)^4$ *supported on* $[x_0, \infty)$.

Answer to Q26:______________________________

Q27: A group of 40 students are comparing their birthdays; as usual, assume their birthdays are independent, are not February 29^{th}, *etc.*. Find the expected number of pairs of people with birthdays in the same month. (To simplify calculations, you may assume the year has 12 months of 30 days each.)

Answer to Q27:______________________________

Q28: Let $\overline{X}_n$ be a random variable with a $\mathcal{N}(16, 4/n)$ distribution. Using the Delta Method, what is the approximate distribution of $\sqrt{\overline{X}_n}$?

Answer to Q28:______________________________

Q29: Suppose Stephen is playing a guessing game, where he has a 20% chance of answering a given question correctly (IID). If he answers a question correctly, he gets one point. Let Q be the total number of questions required for him to get 4 points. What is $\mathbb{E}[Q]$?

Answer to Q29:______________________________

Q30: X is a mixture distribution defined as follows: $X_1 \sim \mathcal{N}(2, 5^2)$, $X_2 \sim \text{Poisson}(5)$, 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 expected value of X.

Answer to Q30:________________________________

Q31: Suppose it takes a student 2 minutes on average to answer a question on an exam. Furthermore, assume that it never takes the student less than 1 minute or more than 5 minutes to answer a single question. Assuming the exam has 40 questions total and that the time taken on each question is IID, use the Chernoff inequality for *means of bounded random variables* to give an upper bound on the probability that it takes more than two hours to finish the exam.

Answer to Q31:________________________________

Q32: Let X be the number of Heads in 10 fair coin tosses, assumed IID. Find the conditional variance of X, conditional that the first two tosses do not match (*i.e.*, no HH or no TT).

Answer to Q32:________________________________

Q33: Suppose that a mother and daughter are selected from a population whose heights are $\mathcal{N}(66, 4^2)$; suppose further that the correlation between heights within a family is $\rho = 25\%$. What is the probability that the mother is more than 4 inches taller than her daughter?

Answer to Q33:________________________________

Q34: The Skellam(μ_1, μ_2) distribution is given by $X = N_1 - N_2$ where N_1, N_2 are independent Poisson random variables, with mean μ_1, μ_2 respectively. What is the variance of a Skellam$(4, 3)$ distribution?

Answer to Q34:________________________________

Q35: In a *mixed-effects* regression model, the response is determined by a combination of a noise term, a non-random 'fixed effect' term, and a subject specific 'random effect' term. Symbolically, the vector of responses $\boldsymbol{y}$ is distributed as

$$\boldsymbol{y}|\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{Z}\boldsymbol{b} + \boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}) \text{ where } \boldsymbol{b} \sim \mathcal{N}(\boldsymbol{0}, \nu^2\boldsymbol{I})$$

What is the (unconditional) variance matrix of $\boldsymbol{y}$? Here $\boldsymbol{Z}, \boldsymbol{X}$ are fixed (non-random) matrices and $\boldsymbol{\beta}$ is a fixed (non-random) vector.

Answer to Q35:______________________________

Q36: Suppose that the random variables (X, Y) have joint PDF $f_{(X,Y)} = 1$ with support on the unit square $[0, 1]^2$. What is the probability that $X > Y$ given $Y > 0.5$?

Answer to Q36:______________________________

Q37: The Laplace distribution can be formed as $L = S * E$ where S is a discrete uniform random variable taking values $\{-1, +1\}$ and E is an exponential random variable with mean 3 and S and E are independent. What are $\mathbb{E}[L]$ (2 points) and $\mathbb{V}[L]$ (3 points)?

Answer to Q37:________________________________

Q38: Suppose $Z_1, \ldots, Z_7$ are IID standard normal random variables. What is the probability that $\mathsf{median}(Z_1, \ldots, Z_7)$ is positive? (Recall that the median of a set of 7 values is the middle value, *i.e.*, the fourth largest when sorted.)

Answer to Q38:________________________________

Q39: Let $X \sim \mathcal{N}(1, 4)$ be a Gaussian random variable. Additionally, conditional on $X = x$, let Y, Z be independent $\mathcal{N}(3x, 1)$ random variables. What is $\mathbb{C}[X, Y + Z]$?

Answer to Q39:________________________________

Q40: Consider the standard linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, but we now assume $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ instead of IID. What is the (co)variance matrix of the OLS estimator

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X} \boldsymbol{y}?$$

Answer to Q40:________________________________

(Blank page for scratch work - not graded)

(Blank page for scratch work - not graded)

(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: $\mathbb{V}[X] = \mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \geq 0$
- $\mathbb{V}[aX + bY + c] = a^2\mathbb{V}[X] + b^2\mathbb{V}[Y]$ 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). $\mathbb{E}[\mathrm{DUnif}\{a, \dots, b\}] = \mathbb{E}[\mathrm{CUnif}([a, b])] = (a + b)/2$. $\mathbb{V}[\mathrm{CUnif}([a, b])] = (b - a)^2/12$

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: $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x} \cdot \boldsymbol{y} = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$
- 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: $\langle \boldsymbol{x}, \boldsymbol{Ax} \rangle = \|\boldsymbol{x}\|_{\boldsymbol{A}}^2 = \sum_{(i,j)} A_{ij} x_i x_j$. $\boldsymbol{A}$ 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: $\mathbb{V}[\langle \boldsymbol{a}, \boldsymbol{X} \rangle] = \langle \boldsymbol{a}, \mathbb{V}[\boldsymbol{X}]\boldsymbol{a} \rangle = \sum_{ij} a_i a_j \mathbb{C}[X_i, X_j] \geq 0$
- 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

STA9715 - Test 3 - Formula Sheet

Advanced Inequalities

- Chernoff - Gaussian: if $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathbb{P}(|X-\mu| > t) \leq 2e^{-t^2/2\sigma^2}$ and $\mathbb{P}(X > \mu + t) \leq e^{-t^2/2\sigma^2}$
- Chernoff - Bounded: if X takes values in the range $[a, b]$ with mean μ, then $\mathbb{P}(|X-\mu| > t) \leq 2e^{-2t^2/(b-a)^2}$ and $\mathbb{P}(X > \mu + t) \leq e^{-2t^2/(b-a)^2}$. For means, $\mathbb{P}(|\overline{X}_n - \mu| > t) \leq 2e^{-2nt^2/(b-a)^2}$ and $\mathbb{P}(\overline{X}_n \geq \mu + t) \leq e^{-2nt^2/(b-a)^2}$

Moment Generating Functions

- Moment Generating Function: $\mathbb{M}_X(t) = \mathbb{E}[e^{tX}]$
- MGF to Moments: $\mathbb{E}[X^k] = \mathbb{M}_X^{(k)}(0)$
- MGF of linear transforms: $\mathbb{M}_{aX+b} = \mathbb{E}[e^{t(aX+b)}] = e^{tb}\mathbb{M}_X(ta)$
- MGF of sums of independent RVs: $\mathbb{M}_{X+Y}(t) = \mathbb{M}_X(t)\mathbb{M}_Y(t)$
- If $\mathbb{M}_X(t) = \mathbb{M}_Y(t)$, then X and Y have the same distribution.

Limit Theory

- Convergence in Probability: $X_n \xrightarrow{P} X_*$ means $\mathbb{P}(|X_n - X_*| > \epsilon) \xrightarrow{n\to\infty} 0$
- Convergence in Distribution: $X_n \xrightarrow{d} X_*$ means $\mathbb{E}[f(X_n)] \xrightarrow{n\to\infty} \mathbb{E}[f(X_*)]$ for 'reasonable' functions $f(\cdot)$.
- Law of Large Numbers: if $X_1, X_2, \dots$ are IID random variables with mean μ and finite variance, $\overline{X}_n \xrightarrow{P} \mu$ for $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.
- Central Limit Theorem: if $X_1, X_2, \dots$ are IID random variables with mean μ and finite variance σ^2, then $\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) \xrightarrow{d} \mathcal{N}(0,1)$. More usefully: $\overline{X}_n \xrightarrow{d} \mathcal{N}(\mu, \sigma^2/n)$
- Delta Method: if $X_n \xrightarrow{d} \mathcal{N}(\mu, \sigma^2)$, then $g(X_n) \xrightarrow{d} \mathcal{N}(g(\mu), \sigma^2 g'(\mu)^2)$ for any differentiable $g(\cdot)$
- Glivenko-Cantelli Theorem ('Fundamental Theorem of Statistics'): the sample CDF $\hat{F}_n(\cdot)$ converges to the true CDF $F_X(\cdot)$ at all points where $F_X(\cdot)$ is continuous
- Massart's Inequality: $\mathbb{P}(\max_x |\hat{F}_n(x) - F_X(x)| > \epsilon) \leq 2e^{-2n\epsilon^2}$ for any distribution and any $\epsilon > 0$

Key Statistical Distributions

- Gaussian / Normal - See above. Standardizing, CLT, Delta Method.
- χ^2_k sum of squares of k IID Standard Normals. Arises from 'goodness of fit' type statistics (*e.g.*, SSE in OLS)
- t_k - (Student's) t distribution with k degrees of freedom. $t_k \stackrel{d}{=} Z/\sqrt{\chi^2_k/k}$ where $Z \perp\!\!\!\perp \chi^2_k$. t_1 is a Cauchy; t_∞ is a standard normal. Can replace Z with other normal distributions. Arises in testing with unknown variance.
- χ^2_2 is an Expo(1/2) distribution with mean 2
- Gamma distribution: sum of n exponential distributions with mean $1/\theta$ is $\Gamma(n, \theta)$ distributed. $\Gamma(n/2, 2) \xrightarrow{d} \chi^2_n$
- Beta distribution = Gamma ratio. $X \sim \Gamma(\alpha, \theta), Y \sim \Gamma(\beta, \theta) \implies X/(X+Y) \sim B(\alpha, \beta)$. Support on $[0, 1]$
- F distribution: $\frac{\chi^2_{k_1}/k_1}{\chi^2_{k_2}/k_2}$

Name	Parameters	Density	Mean	Variance	MGF
Standard Normal	None	$\phi(z) = e^{-z^2/2}/\sqrt{2\pi}$	0	1	$e^{t^2/2}$
Normal	Mean μ, StdDev σ	$e^{-(x-\mu)^2/2\sigma^2}/\sqrt{2\pi\sigma^2}$	μ	σ^2	$e^{\mu t + \sigma^2 t^2/2}$
χ^2	Degrees of freedom k	$\propto x^{k/2-1}e^{-x/2}$	k	$2k$	$(1-2t)^{-k/2}$
Standard Student's t	Degrees of freedom k	$\propto (1+x^2/k)^{-(k+1)/2}$	0	$k/(k-2)$	NA
Gamma	Shape k, Scale θ	$\propto x^{k-1}e^{-x/\theta}$	$k\theta$	$k\theta^2$	$(1-\theta t)^{-k}$
Beta	Shapes α, β	$\propto x^{\alpha-1}(1-x)^{\beta-1}$	$\alpha/(\alpha+\beta)$	$(\alpha\beta)(\alpha+\beta)^{-2}(\alpha+\beta+1)^{-1}$	Hard
F	Deg. Freedom k_1, k_2	Hard	$k_2/(k_2-2)$	Hard	NA

Sampling (Probability Integral Transform): If X has CDF F_X, $F_X^{-1}(U) \stackrel{d}{=} X$ for $U \sim \mathcal{U}([0,1])$

Sampling (Box-Mueller): Let $R^2 \sim \chi^2_2$ and $\Theta \sim \mathcal{U}([0, 2\pi])$; then $X = R\cos\Theta, Y = R\sin\Theta$ are independent Z_1, Z_2