CSE 312 : Spring 2024 Final Exam: Form A

Name:	**NetID:** @uw.edu

Instructions

- You have 110 minutes to complete this exam.
- You are permitted one piece of 8.5x11 inch paper with handwritten notes (notes on both sides allowed).
- You are also provided a reference sheet with the exam.
- You may not use a calculator or any other electronic devices during the exam.
- We will be scanning your exams before grading them. Please write legibly, and avoid writing up to the edge of the paper. If you run out of room, you may also use the last page for extra space, but tell us where to find your answer if it's not right below the problem.
- **Please put final answers in provided boxes on longer problems.**
- In general, show us the work you used to get to an answer; explanations will help us award partial credit, but we do not expect explanations at the level we usually require on homeworks.

Simplification Expectations

- Since you don't have a calculator for this exam, you do not have to do simplifications that could be done easily with a calculator. For example, the expression below is simplified enough to be a final answer.

$$\frac{\binom{5}{3} \cdot 17^2}{1-p} + 5^3$$

- However, answers which are much more complicated than the expected answer may receive deductions. For example: $\sum_{i=0}^{n} \binom{n}{i}$, or $\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n}$ are **not** simplified sufficiently. Generally summation notation or "..." are not sufficiently simple, but a few plus signs are fine.

Advice

- Writing a few words about where an expression came from is often very helpful for awarding partial credit.
- Remember to take deep breaths.

Question	**Max points**
Multiple Choice	30
Small Problems	10
Gamblers' Fallacy (Random Variables)	15
Puzzles! (Conditional Expectations)	16
Time to Cook (Concentration)	14
Mario's Mushrooms (MLE)	15
Morale	1
Total	**101**

This page was intentionally left blank.

1. [Multiple Choice] $0.25^{10} \approx 0.000001$ [30 points]

For the questions below,

- Questions with ◯ Circles have exactly one best answer. **Fully fill in** the circle for the one best answer.
- Questions with ☐ squares are "mark all that apply" questions. **Fully fill in** the square for all correct options.

(a) Let A, B be events with non-zero probability. Which of the following must be true for independent events (mark all that apply)?

☐ $\mathbb{P}(A|B) = \mathbb{P}(B|A)$
☐ $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
☐ $\mathbb{P}(A|B) = \mathbb{P}(A)$

(b) Suppose you play a game with a friend where you flip a coin. If the coin is heads, your friend pays you a dollar. If the coin is tails, you pay your friend a dollar. Let X be your profit and Y be your friend's profit. What can you say about the covariance of X and Y?

◯ Cov $(X, Y) \geq 0$
◯ Cov $(X, Y) \leq 0$
◯ Cov (X, Y) could be positive, negative, or 0 depending on the probability that the coin comes up heads.

(c) Suppose we want to distribute all 15 candies in a bag to 15 children. Let p_i be the probability that the ith child gets assigned at least 1 candy. If all such p_i are equal, then the number of children who get at least 1 candy must be a Binomial random variable with parameters $n = 15$ and $p = p_1$.

◯ True
◯ False

(d) Which random variable best models the number of car accidents that happen in Seattle in a day?

◯ A Poisson random variable, with parameter equal to the average number of car accidents in a day.
◯ A Poisson random variable, with parameter equal the the number of drivers on the road in a day.
◯ An Exponential random variable, with parameter equal to the average number of car accidents in a day.
◯ An Exponential random variable, with parameter equal the the number of drivers on the road in a day.

(e) $n + 1$ friends enter a casino. The first friend is very rich and wants to see if she can make more profit than the sum of the other n friends. Each friend plays at separate tables. The first friend makes large bets at her table and gets a profit with mean n and variance n. The other friends make small bets and get profits with mean 1 and variance 1 (each). Which of the following **must** be true? Mark **all** that apply.

☐ The mean of the profit of the first friend is the same as the mean of the sum of the other friends' profits.
☐ The variance of the profit of the first friend is the same as the variance of the sum of the other friends' profits.
☐ The profit of the first friend and the sum of the profits of the other friends are identically distributed.
☐ The probability that the first friend makes more profit than the sum of the other friends is 0.5

- Questions with ◯ Circles have exactly one best answer. **Fully fill in** the circle for the one best answer.
- Questions with ☐ squares are "mark all that apply" questions. **Fully fill in** the square for all correct options.

(f) Using maximum likelihood estimation, you find an estimator $\hat{\theta}$ of an unknown parameter θ, based on n independent data points. You discover $\mathbb{E}[\hat{\theta}] = \frac{\theta}{2}$. How do you de-bias the estimator (make it unbiased)?

◯ Draw another n points, since the first batch of data points was low.
◯ Use $2\hat{\theta}$ as your estimator.
◯ Use $\hat{\theta}^2$ as your estimator.
◯ You cannot de-bias a biased estimator.

(g) Let $X = \sum_{i=1}^{n} X_i$, where X_i are all independent Bernoulli random variables. You want to compute the **exact** value of $\mathbb{P}(X \geq \mathbb{E}[X])$. You should

◯ Use Markov's Inequality
◯ Use Chernoff's Inequality
◯ Use the Central Limit Theorem
◯ All of the above work.
◯ None of the above work.

(h) Robbie is collecting Pokemon, which can only be normal, shiny, or super shiny. One of the TAs says that there is some hidden parameter $\theta < 0.25$ in the game, and there is probability 0.4 that Robbie gets a normal Pokemon, probability 2θ for a shiny, and probability θ for a super shiny. Could this be a valid probability distribution?

◯ True, there are multiple values of $\theta < 0.25$ that can make this a valid probability distribution.
◯ True, there exists exactly one $\theta < 0.25$ that can make this a valid probability distribution.
◯ False, this is not possible.

(i) Select all the true statements about finding an MLE:

☐ **Maximizing** the log-likelihood function gets us the same estimator as **maximizing** the likelihood function because log is a monotonically increasing function.
☐ *Minimizing* the log-likelihood function gets us the same estimator as **maximizing** the likelihood function because log is a monotonically increasing function.
☐ Computing the log-likelihood allows us to take the derivative on **products** rather than *sums* of terms.
☐ Computing the log-likelihood allows us to take the derivative on *sums* rather than **products** of terms.

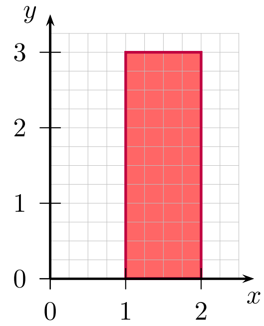
(j) Let $X \sim \mathcal{N}(0, 1)$. What is the smallest value of k for which we can say $\mathbb{P}(|X - \mathbb{E}[X]| \geq k) \leq 0.05$?

◯ 1.64
◯ 1.96
◯ 2.00
◯ This k cannot be computed exactly; it can only be bounded by concentration inequalities.

2. Small Problems

(a) The weight of an adult capybara is normally distributed with mean 100 pounds and variance 25 pounds. A researcher is only able to lift capybaras of 90 pounds or less. You give the researcher a random adult capybara, what is the probability they **cannot** lift it? Be sure to use the Φ-table provided with your reference sheet. [3 points]

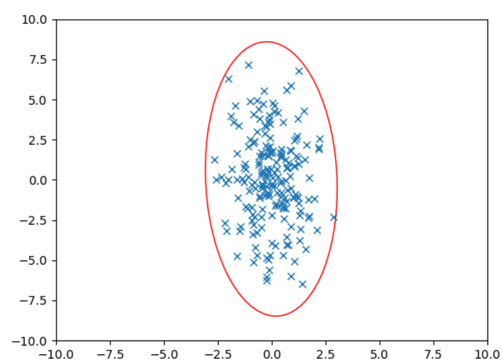
(b) Consider some joint PDF $f_{X,Y}(x, y)$ of random variables X and Y. Suppose the only area in the PDF with non-zero density is shaded below. In notation, the region is $1 \leq x \leq 2$ and $0 \leq y \leq 3$.



Suppose we wanted to compute $\mathbb{P}(Y > 2X)$. What would be the correct setup of the integral? Your answer must not integrate over any region with 0 density. Fill in each box. [4 points]

$$\int_{\square}^{\square} \int_{\square}^{\square} f_{X,Y}(x, y)\, d\square\, d\square$$

(c) The image below is 200 points drawn independently from a multidimensional Gaussian random variable with covariance matrix Σ and mean $[0, 0]$. The first entry of each vector is on the x-axis, and the second entry of each vector is on the y-axis. The figure goes from -10 to 10 in each dimension.
Which of the following options is Σ? [3 points]



○ $\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$ ○ $\begin{bmatrix} 3 & 2 \\ 2 & 7 \end{bmatrix}$ ○ $\begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix}$ ○ $\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$

3. [Random Variables] Gambler's Fallacy

You are working on a new game for a casino that combines dice and cards.

Here's the game:

1) A player first rolls a fair 3-sided die with faces $\{1, 2, 3\}$. Let X be the result of the die roll.
2) Then, the player gets to draw the first X cards from a shuffled standard 52-card deck (13 ranks of 4 suits).
   - If they draw at least one face card (Jack, Queen, or King of any suit), then they win \$$m$.
   - If they draw at least one Ace (any suit), they win \$$2m$.
   - If they draw both at least one Ace AND at least one face card, they win \$$3m$.
   - If none of these happen, they win $\$0$

Let R be the player's winning in dollars. Answer the following questions based on this setup.

(a) What is the probability of winning $\$3m$ if $X = 1$? [2 points]

(b) What is the probability of winning $\$3m$ if $X = 2$? [2 points]

(c) What is the probability of winning $\$3m$ if $X = 3$? [2 points]

(d) Let a, b, c be the correct answers to the previous three parts respectively. Give an expression in terms of a, b, c that represents the probability of winning $\$3m$. You do not have to simplify your expression. [2 points]

(e) Suppose you keep playing the game (reshuffling the deck before each round, and rerolling the die each round) until you win $\$3m$ in each of two games. What is the expected number of games needed? Give an expression in terms of d, the correct answer to part (d). [3 points]

The problem from the last page continues below, here's the setup again:

Here's the game:

1) A player first rolls a fair 3-sided die with faces $\{1, 2, 3\}$. Let X be the result of the die roll.
2) Then, the player gets to draw the first X cards from a shuffled standard 52-card deck (13 ranks of 4 suits).
   - If they draw at least one face card (Jack, Queen, or King of any suit), then they win $\$m$.
   - If they draw at least one Ace (any suit), they win $\$2m$.
   - If they draw both at least one Ace AND at least one face card, they win $\$3m$.
   - If none of these happen, they win $\$0$

Let R be the player's winning in dollars. Answer the following questions based on this setup.

(f) Let S be the event: "you rolled a 2, and the first card is the Ace of Hearts." What is the expected winnings given S? That is, what is $\mathbb{E}[R|S]$? [4 points]

4. [Conditional PRobbie-bility] Puzzles!

The puzzle game *Connections* shows the user 16 words, which should be grouped into 4 groups of 4. There is exactly one correct answer (i.e., each of the 16 words belongs in exactly one group of 4). The user selects 4 items they think form a group. If a correct group of 4 is selected, that entire group is removed from the puzzle; if it is not exactly a group of 4, then all items remain in the puzzle.

You get two groups immediately, so 8 items (2 groups of 4) remain. You wish to know the expected number of **incorrect** guesses required until you guess all the groups correctly.

(a) If you do the puzzle while **focused**, you will identify 5 words, of which 4 form a group. You will guess a new set of 4 among the 5 words until you find the group (with each not-yet-guessed set equally likely). What is the expected number of incorrect guesses until you get the group of 4, assuming you are focused? [4 points] (Hint: our answer doesn't need combinations)

(b) If you do the puzzle while **distracted**, you will choose a set of 4 uniformly at random from among the 8 remaining. Since you are distracted, each guess is independent (and therefore you might repeat a previous guess). What is the expected number of incorrect guesses, assuming you are distracted? [4 points]

(c) Let a be your answer from (a) and b be your answer from (b), and let p be the probability you are focused (all other times you are distracted). Give a formula for the average number of **incorrect** guesses in terms of a, b, p. [2 points]

(d) Next you play the *Wordle* (a word-guessing game), with starting word ADIEU.
When the correct answer is **vowel-heavy**, you do well: half the time you get a score of 4 or better; the other half the time you get a score of 5 or worse.
When the correct answer is not **vowel-heavy**, you don't do as well: $1/10$ of the time you get a score of 4 or better; the other $9/10$ of the time you get a score of 5 or worse.
You know that $1/3$ of words are vowel-heavy.

- Let V be the event "the correct answer is vowel-heavy."
- Let $\overline{V}$ be the event "the correct answer is not vowel-heavy."
- Let E be the event you got a score of 4 or better.

(i) You wish to find the probability that the correct answer is vowel heavy, assuming that you got a score of 4 or better. Fill in the blank below for the expression you're seeking, using the notation above. [2 points]

$$\mathbb{P}(\underline{\hspace{4cm}})$$

(ii) Compute the expression from the last part (i.e., find the probability the correct answer is vowel-heavy, assuming you got a score of 4 or better). [4 points]

5. [Concentration] Time to Cook

Having finally defeated the Red Baron, Snoopy has set aside his role as a pilot in pursuit of his passions as a world famous chef. He decides to start by cooking many batches of his fanciest dish, french onion soup.

In the kitchen, Snoopy has 5 bags of infinite onions: each bag has only one type from red onions, yellow onions, white onions, sweet onions, and shallots. Snoopy cooks a batch of french onion soup by selecting a bag uniformly at random and taking an onion out. He repeats this process for each of the 45 onions required (independently each time). Let a batch of soup look **reddish** if there are between 2 and 16 (inclusive) red onions in the batch.

Answer the questions below. Be sure to use the bounds as stated on the provided reference sheet. Additionally, for each bound, write both the **direction** of the inequality (i.e., $\leq, \geq, <, >$) and the **bound** (i.e., the number).

(a) Consider Snoopy's first batch of french onion soup. Of the 45 onions required in the batch, compute the expected number of red onions. [3 points]

(b) Now use **Chernoff's bound** to bound the probability that Snoopy's first batch of french onion soup looks **reddish**. [5 points]

$$\mathbb{P}\,(\text{reddish}) \;\square\; \boxed{}$$

inequality — bound

(c) Suppose Snoopy makes n batches of french onion soup. He will be confident that he is a good cook if **all** n of his batches are **reddish**. Let b be the bound you got from part (b). Use the **Union bound** to bound the probability that Snoopy thinks he is a good cook. [3 points]

$$\mathbb{P}\,(\text{good cook}) \;\square\; \boxed{}$$

inequality — bound

(d) Now Snoopy wants feedback from his eccentric little bird friend, Woodstock, who rates dishes using an integer in $(-\infty, +\infty)$. More negative numbers represent stronger dislike and more positive numbers the opposite. Suppose that, independent of the soup quality, Woodstock gives a rating with expectation 0.5 and variance 0.2. Let a **favorable rating** be one that is greater than or equal to 1.

What is the best bound to apply on the probability that Woodstock gives Snoopy's first batch of soup a **favorable rating**? [3 points]

◯ Markov's Inequality
◯ Chebyshev's Inequality
◯ Chernoff Bound
◯ Union Bound

6. [MLE] Mario's Mushrooms

Mario has forgotten the four-digit PIN (personal identification number) to unlock one of his lunchboxes of red and green mushrooms. Mario asks Luigi to help him guess his PIN. Luigi proposes the following model of PIN generation:

- For each PIN, Mario chooses the first digit uniformly at random (all 10 digits in 0-9 are allowed).
- Then for every digit afterwards:
  - with probability m, Mario repeats the previous digit.
  - otherwise, Mario picks a digit from 0-9 *excluding* the previous digit uniformly at random (so each of the 9 other digits is equally likely).
- Each decision is independent of the others.

(a) Luigi finds that Mario locked up a lunchbox using the PIN 1234.
Under Luigi's model, what is the likelihood of Mario's PIN number in terms of m? [4 points]

$$\mathcal{L}(\text{Mario's PIN}; m) = \underline{\hspace{6cm}}$$

likelihood

(b) Luigi does some more snooping and finds that Mario has 12 locked lunchboxes.

- Mario has 10 PINs that are 0000.
- Mario has 2 PINs that are 1981.

Under Luigi's model, what is the likelihood of these 12 PIN numbers in terms of m? [4 points]

$$\mathcal{L}(\text{Mario's PINs}; m) = \underline{\hspace{6cm}}$$

likelihood

(c) Suppose that m is close to 1, i.e., $0.99 < m \leq 1$. Of these four PINs below, which is the most likely? [2 points]

◯ 1234
◯ 1991
◯ 2228
◯ 3333

(d) Suppose that m is close to 0, i.e., $0 \leq m < 0.01$. Of these four PINs below, which is the most likely? [2 points]

◯ 6062
◯ 4999
◯ 5574
◯ 0883

(e) Let $\widehat{m_a}$ be the MLE estimate of m you would get from the data in part (a). Let $\widehat{m_b}$ be the MLE estimate of m you would get from the data in part (b). Which of these estimates is higher?[3 points]
Hint: You don't need to do a computation; use intuition from parts (c) and (d).

◯ $\widehat{m_a} > \widehat{m_b}$
◯ $\widehat{m_a} = \widehat{m_b}$
◯ $\widehat{m_a} < \widehat{m_b}$

7. Morale

Draw a picture, write a poem (or anything you want!) that incorporates your favorite story from one of the questions on this final.
You will get the point for this problem as long as this page is not blank.

Use this page for extra space if needed. Be sure to tell us to look here on the original problem.

Reference Sheet: Counting, Discrete Probability

Theorem: Binomial Theorem

Let $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$ a positive integer. Then: $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

Theorem: Principle of Inclusion-Exclusion (PIE)

2 events: $|A \cup B| = |A| + |B| - |A \cap B|$
3 events: $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$
k events: singles - doubles + triples - quads + ...

Theorem: Pigeonhole Principle

If there are n pigeons we want to put into k holes (where $n > k$), then at least one pigeonhole must contain at least 2 (or to be precise, $\lceil n/k \rceil$) pigeons.

Definition: Key Probability Definitions

The **sample space** is the set Ω of all possible outcomes of an experiment.
An **event** is any subset $E \subseteq \Omega$.
Events E and F are **mutually exclusive** if $E \cap F = \emptyset$.

Definition: Probability space

A *probablity space* is a pair $(\Omega, \mathbb{P})$, where Ω is the sample space
$\mathbb{P} : \Omega \to [0, 1]$ is a *probability measure* such that $\sum_{x \in \Omega} \mathbb{P}(x) = 1$.
The probability of an event $E \subseteq \Omega$ is $\mathbb{P}(E) = \sum_{x \in E} \mathbb{P}(x)$.

Definition: Conditional Probability

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

Theorem: Bayes Theorem

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[B \mid A]\,\mathbb{P}[A]}{\mathbb{P}[B]}$$

Definition: Partition

Non-empty events $E_1, \dots, E_n$ **partition** the sample space Ω if:

- **(Exhaustive)** $E_1 \cup E_2 \cup \dots \cup E_n = \bigcup_{i=1}^{n} E_i = \Omega$ (they cover the entire sample space).
- **(Pairwise Mutually Exclusive)** For all $i \neq j$, $E_i \cap E_j = \emptyset$ (none of them overlap)

Theorem: Law of Total Probability (LTP)

If events $E_1, \dots, E_n$ partition Ω, then for any event F:

$$\mathbb{P}[F] = \sum_{i=1}^{n} \mathbb{P}[F \cap E_i] = \sum_{i=1}^{n} \mathbb{P}[F \mid E_i]\,\mathbb{P}[E_i]$$

Theorem: Bayes Theorem with LTP

Let events $E_1, \dots, E_n$ partition the sample space Ω, and let F be another event. Then:

$$\mathbb{P}[E_1 \mid F] = \frac{\mathbb{P}[F \mid E_1]\,\mathbb{P}[E_1]}{\sum_{i=1}^{n} \mathbb{P}[F \mid E_i]\,\mathbb{P}[E_i]}$$

Definition: Independence (Events)

A and B are **independent** if any of the following equivalent statements hold:
1. $\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$
2. $\mathbb{P}[A \mid B] = \mathbb{P}[A]$
3. $\mathbb{P}[B \mid A] = \mathbb{P}[B]$

Theorem: Chain Rule

Let $A_1, \dots, A_n$ be events with nonzero probabilities. Then:
$\mathbb{P}[A_1 \cap \dots \cap A_n] =$
$\mathbb{P}[A_1]\,\mathbb{P}[A_2 \mid A_1]\,\mathbb{P}[A_3 \mid A_1 \cap A_2] \cdots \mathbb{P}[A_n \mid A_1 \cap \dots \cap A_{n-1}]$

Definition: Mutual Independence (Events)

We say n events $A_1, A_2, \dots, A_n$ are **(mutually) independent** if, for *any* subset $I \subseteq [n] = \{1, 2, \dots, n\}$, we have

$$\mathbb{P}\left[\bigcap_{i \in I} A_i\right] = \prod_{i \in I} \mathbb{P}[A_i]$$

This equation is actually representing 2^n equations since there are 2^n subsets of $[n]$.

Definition: Conditional Independence

A and B are **conditionally independent given an event** C if any of the following equivalent statements hold:
1. $\mathbb{P}[A \cap B \mid C] = \mathbb{P}[A \mid C]\,\mathbb{P}[B \mid C]$
2. $\mathbb{P}[A \mid B \cap C] = \mathbb{P}[A \mid C]$
3. $\mathbb{P}[B \mid A \cap C] = \mathbb{P}[B \mid C]$

Definition: Random Variable (RV)

A random variable X is a function of the outcome $X : \Omega \to \mathbb{R}$. The set of possible values X can take on is its **range/support**, denoted Ω_X.

Definition: Probability Mass Function (PMF)

For a discrete RV X, assigns probabilities to values in its range. That is $p_X : \Omega_X \to [0, 1]$ where: $p_X(k) = \mathbb{P}[X = k]$.

Definition: Expectation

The **expectation** of a discrete RV X is: $\mathbb{E}[X] = \sum_{k \in \Omega_X} k \cdot p_X(k)$.

Theorem: Linearity of Expectation (LoE)

For any random variables X, Y (possibly dependent):
$$\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$$

Theorem: Law of the Unconscious Statistician (LOTUS)

For a discrete RV X and function g, $\mathbb{E}[g(X)] = \sum_{b \in \Omega_X} g(b) \cdot p_X(b)$.

Definition: Variance

$\text{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$.

Theorem: Property of Variance

$\text{Var}(aX + b) = a^2 \text{Var}(X)$.

Definition: Independence (Random Variables)

Random variables X and Y are **independent** if for *all* $x \in \Omega_X$ and all $y \in \Omega_Y$:
$$\mathbb{P}[X = x, Y = y] = \mathbb{P}[X = x] \cdot \mathbb{P}[Y = y].$$

Theorem: Variance Adds for Independent RVs

If X, Y are independent, then $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

Definition: Standard Deviation (SD)

$\sigma_X = \sqrt{\text{Var}(X)}$.

Reference: Continuous and Multivariate Probability

Definition: Cumulative Distribution Function (CDF)

The **cumulative distribution function (CDF)** of ANY random variable is $F_X(t) = \mathbb{P}[X \leq t]$.
If X is a *continuous* RV, $F_X(t) = \mathbb{P}[X \leq t] = \int_{-\infty}^{t} f_X(w)\, dw$.

Theorem: Multiplicativity of expectation

For any independent random variables X, Y:
$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Definition: Expectation (Continuous)

The **expectation** of a continuous RV X is:
$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx$.

Theorem: Law of the Unconscious Statistician (LOTUS)

For a continuous RV X: $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$.

Definition: Independent and Identically Distributed (i.i.d.)

We say $X_1, \ldots, X_n$ are said to be **independent and identically distributed (i.i.d.)** if all the X_i's are independent of each other, and have the same distribution (PMF for discrete RVs, or CDF for continuous RVs).

Definition: Joint PMFs

The joint PMF of discrete RVs X and Y is:
$$p_{X,Y}(a, b) = \mathbb{P}[X = a, Y = b]$$
Their joint range is
$$\Omega_{X,Y} = \{(c, d) : p_{X,Y}(c, d) > 0\} \subseteq \Omega_X \times \Omega_Y$$
Note that $\sum_{(s,t)\in\Omega_{X,Y}} p_{X,Y}(s, t) = 1$.

Definition: Joint PDFs

The joint PDF of continuous RVs X and Y is:
$$f_{X,Y}(a, b) \geq 0$$
Their joint range is
$$\Omega_{X,Y} = \{(c, d) : f_{X,Y}(c, d) > 0\} \subseteq \Omega_X \times \Omega_Y$$
Note that $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(u, v) du dv = 1$.

Definition: Marginal PMFs

Let X, Y be discrete random variables. The marginal PMF of X is: $p_X(a) = \sum_{b\in\Omega_Y} p_{X,Y}(a, b)$.

Definition: Marginal PDFs

Let X, Y be continuous random variables. The marginal PDF of X is: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$.

Definition: Independence of RVs (Continuous)

Continuous RVs X, Y are independent, written $X \perp Y$, if for all $x \in \Omega_X$ and $y \in \Omega_Y$,
$$f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

Definition: Conditional Expectation

If X is discrete (and Y is either discrete or continuous), then we define the conditional expectation of $g(X)$ given (the event that) $Y = y$ as:
$$\mathbb{E}[\, g(X) \mid Y = y\,] = \sum_{x\in\Omega_X} g(x)\, \mathbb{P}(X = x \mid Y = y)$$
If X is continuous (and Y is either discrete or continuous), then
$$\mathbb{E}[\, g(X) \mid Y = y\,] = \int_{-\infty}^{\infty} g(x)\, \frac{f_{X,Y}(x, y)}{f_Y(y)} dx$$

Theorem: Law of Total Expectation (LTE)

Let X, Y be jointly distributed random variables.

If Y is discrete (and X is either discrete or continuous), then:
$$\mathbb{E}[\, g(X)\,] = \sum_{y\in\Omega_Y} \mathbb{E}[\, g(X) \mid Y = y\,]\; p_Y(y)$$
If Y is continuous (and X is either discrete or continuous), then
$$\mathbb{E}[\, g(X)\,] = \int_{-\infty}^{\infty} \mathbb{E}[\, g(X) \mid Y = y\,]\; f_Y(y) dy$$

Definition: Covariance

The Covariance between random variables X and Y is:
$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Theorem: Variance of sums

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X, Y)$$

Reference: Tail Bounds

Theorem: Markov's Inequality

Let $X \geq 0$ be a **non-negative** RV, and let $k > 0$. Then:
$$\mathbb{P}[X \geq k] \leq \frac{\mathbb{E}[X]}{k}$$
.

Theorem: Chebyshev's Inequality

Let X be any RV with expected value $\mu = \mathbb{E}[X]$ and finite variance $\text{Var}(X)$. Then, for any real number $\alpha > 0$. Then,
$$\mathbb{P}[|X - \mu| \geq \alpha] \leq \frac{\text{Var}(X)}{\alpha^2}$$
.

Theorem: Chernoff Bound

Let $X = X_1 + X_2 + \ldots + X_n$, where $X_1, X_2, \ldots, X_n$ are independent random variables, each taking values in $[0, 1]$. Also, let $\mu = \mathbb{E}[X]$.
For any $1 > \delta > 0$:
$$\mathbb{P}(X \geq (1 + \delta)\mu) \leq \exp\left(-\delta^2\mu/3\right)$$
$$\mathbb{P}(X \leq (1 - \delta)\mu) \leq \exp\left(-\delta^2\mu/2\right)$$

Theorem: The Union Bound

Let $E_1, E_2, \ldots, E_n$ be a collection of events. Then:
$$\mathbb{P}\left[\bigcup_{i=1}^{n} E_i\right] \leq \sum_{i=1}^{n} \mathbb{P}[E_i]$$

Reference: Zoo

Definition: Bernoulli/Indicator Random Variable

$X \sim$ Bernoulli(p) (Ber(p) for short) iff X has PMF:

$$p_X(k) = \begin{cases} p, & k = 1 \\ 1-p, & k = 0 \end{cases}$$

$\mathbb{E}[X] = p$ and Var$(X) = p(1-p)$.

Definition: Binomial Random Variable

$X \sim$ Binomial(n, p) (Bin(n, p) for short) iff X has PMF

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \Omega_X = \{0, 1, \dots, n\}$$

$\mathbb{E}[X] = np$ and Var$(X) = np(1-p)$.

Definition: Uniform Random Variable (Discrete)

$X \sim$ Uniform(a, b) (Unif(a, b) for short), for integers $a \leq b$, iff X has PMF:

$$p_X(k) = \frac{1}{b-a+1}, \quad k \in \Omega_X = \{a, a+1, \dots, b\}$$

$\mathbb{E}[X] = \frac{a+b}{2}$ and Var$(X) = \frac{(b-a)(b-a+2)}{12}$.

Definition: Geometric Random Variable

$X \sim$ Geometric(p) (Geo(p) for short) iff X has PMF:

$$p_X(k) = (1-p)^{k-1} p, \quad k \in \Omega_X = \{1, 2, 3, \dots\}$$

$\mathbb{E}[X] = \frac{1}{p}$ and Var$(X) = \frac{1-p}{p^2}$.

Definition: Poisson Random Variable

$X \sim$ Poisson(λ) (Poi(λ) for short) iff X has PMF:

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \Omega_X = \{0, 1, 2, \dots\}$$

$\mathbb{E}[X] = \lambda$ and Var$(X) = \lambda$. If $X_1, \dots, X_n$ are independent Poisson RV's, where $X_i \sim$ Poi(λ_i), then $X = X_1 + \dots + X_n \sim$ Poi$(\lambda_1 + \dots + \lambda_n)$.

Definition: Uniform Random Variable (Continuous)

$X \sim$ Uniform(a, b) (Unif(a, b) for short) iff X has PDF:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in \Omega_X = [a, b] \\ 0 & \text{otherwise} \end{cases}$$

$\mathbb{E}[X] = \frac{a+b}{2}$ and Var$(X) = \frac{(b-a)^2}{12}$.

Definition: Exponential Random Variable

$X \sim$ Exponential(λ) (Exp(λ) for short) iff X has PDF:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \in \Omega_X = [0, \infty) \\ 0 & \text{otherwise} \end{cases}$$

$\mathbb{E}[X] = \frac{1}{\lambda}$ and Var$(X) = \frac{1}{\lambda^2}$.

$F_X(x) = 1 - e^{-\lambda x}$ for $x \geq 0$.

Definition: Normal (Gaussian, "bell curve") Random Variable

$X \sim \mathcal{N}(\mu, \sigma^2)$ iff X has PDF:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, \quad x \in \Omega_X = \mathbb{R}$$

$\mathbb{E}[X] = \mu$ and Var$(X) = \sigma^2$.

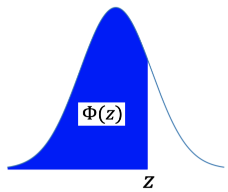
Theorem: Closure of the Normal Under Scale and Shift

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.
In particular, we can always scale/shift to get the standard Normal: $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$.

Theorem: Closure of the Normal Under Addition

If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ are *independent*, then

$$aX + bY + c \sim \mathcal{N}(a\mu_X + b\mu_Y + c, a^2\sigma_X^2 + b^2\sigma_Y^2)$$



Φ Table: $\mathbb{P}(Z \leq z)$ when $Z \sim \mathcal{N}(0, 1)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999