CSE 312 :

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 scientific 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; $\int_1^3 x^3 dx$ or $(2t^3)\Big|_0^x$ are **not** simplified sufficiently.
- Generally derivatives, integrals, summations, or "..." are not sufficiently simple, unless otherwise indicated.

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**
Small Questions	22
Random Variables	17
Joint Distributions	12
Conditional Probability	10
Concentration Inequalities	16
MLE	10
Multiple Choice	22
Grading Morale	1
Total	**110**

1. Small Questions [22 points]

(a) There are 6 kids and 5 parents going to a play on a school field trip, and they want to take their seats (all in a row) at a baseball game. How many orderings are possible if...

(i) ... no kid can sit next to another kid (note this means they alternate kid, parent, kid, parent, ..., kid)? [2 points]

(ii) ... Stevie and Bobby refuse to sit next to each other, but there are no other restrictions? [2 points]

(b) Salt $\&$ Straw has 8 different flavors of ice cream available. You decide to make a mega-sampler (30 scoops, where order does not matter and repetitions are allowed), but you want to have have at most 1 scoop of Thanksgiving flavor (you think meat in ice cream is bad, but you are also a completionist). How many ways are there to make a mega-sampler according to your requirements? [4 points]

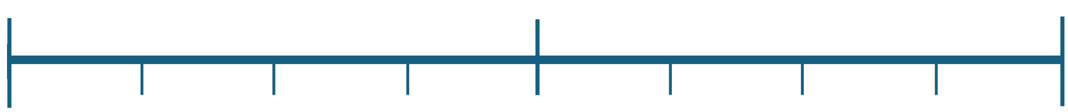
(c) Let W be a $\mathcal{N}(-10, 4)$ random variable. For what value of c does $\mathbb{P}[-16 < W < -4] = 2\Phi(c)$? [3 points]

(d) Let $X_1, \dots, X_n$ be continuous uniform random variables taking values in the range $[0, 20]$. Let $Z = \min\{X_1, \dots, X_n\}$. Compute $\mathbb{P}[Z > z]$ for a fixed $0 < z < 20$. [3 points]

(e) Suppose you are typing on a keyboard containing 14 distinct characters, one of which is 'x'. How many 4 letter strings can you type that contain at least two 'x' characters **in a row**? [3 points]
Hint: We know of at least two ways to do this problem: Inclusion/Exclusion or complementary counting.

(f) An ant is placed uniformly randomly on a stick of length 8. Let X be the distance to the center of the stick.

(i) On the line segment below, shade in the area(s) at which the ant could appear if $X \leq 1.5$. [1 point]



(ii) Give $F_X(x)$. [4 points] Be sure to include all cases.

$$F_X(x) = \begin{cases} \\ \\ \end{cases}$$

2. Random Variables [17 points]

You are playing a new video game and want to collect as many gems as possible. The goal of the game is trying to get to 1000 gems by the end of the game. You have the following methods of gem generation at your disposal.

- **Logic Puzzle**: If you solve a logic puzzle, you will be awarded 10 gems (and none if you get it wrong). Your logic skills are rusty, so you have only an 0.8 chance of getting the question right.
- **Coin Flips**: You can instead play a mini-game 3 times. Each of those 3 times, you flip a fair coin. If the coin is heads, you get two gems. However, if it's tails, you lose two gems. This might sounds harsh, but there's a bonus in this game! If you get all three heads, you get 24 (additional) gems.
- **Earthquakes**: You can stand by a fault line for one time unit, which awards you Poi(15) gems.

Let L, C, E be random variables representing the number of gems awarded each of these ways.

(a) You have taken CSE 312, so you wish to find the best method of gem generation. What is the expected number of gem you get from each method? [6 points]

$\mathbb{E}[L]$: ____________ $\mathbb{E}[C]$: ____________ $\mathbb{E}[E]$: ____________

(b) Suppose you have 500 gems at the moment and you really liked **Coin Flips**, so after this you will only play **Coin Flips**. What is the least number of rounds r you need to play such that your expected (total) number of gems you have is greater than or equal to 1000?
Write your answer in terms of μ_c, your answer for $\mathbb{E}[C]$ from the previous part. [3 points]

(c) You realize you'll need to get a bunch of experience points, and so will need to play these many times. You are therefore interested in the variances of each approach. Compute the variances of L and E. [4 points]

Var (L): ____________ Var (E): ____________

Here are the key facts of the problem from the last page

- **Logic Puzzle**: If you solve a logic puzzle, you will be awarded 10 gems (and none if you get it wrong). Your logic skills are rusty, so you have only an 0.8 chance of getting the question right.
- **Coin Flips**: You can instead play a mini-game 3 times. Each of those 3 times, you flip a fair coin. If the coin is heads, you get two gems. However, if it's tails, you lose two gems. This might sounds harsh, but there's a bonus in this game! If you get all three heads, you get 24 (additional) gems.
- **Earthquakes**: You can stand by a fault line for one time unit, which awards you Poi(15) gems.

(d) We now analyze the variance of C. Let random variable F be the number of gems won (or lost) from just the flips, and let B be the number of gems won from the bonus (or 0 if you get no bonus). For example, if you get three tails: $F = -6, B = 0$; if you get three heads: $F = 6$, $B = 24$.

In the lines below, write the variances of F and B. [2 points]

Var (F): ____________ Var (B): ____________

(e) Which of the following best describes Var (C) in terms of Var (B) and Var (F)? [2 points]

- ◯ Var (C) = Var (B) + Var (F)
- ◯ (Var (C))^2 = (Var (B))^2 + (Var (F))^2
- ◯ Var (C) is neither of the formulas above; it is greater than Var (B) + Var (F)
- ◯ Var (C) is neither of the formulas above; it is less than Var (B) + Var (F)

3. Joint Distributions [12 points]

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{5} & 0 \le x, y \le 2 \\ \frac{1}{5} & -1 \le x, y \le 0 \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute $f_X(x)$. Be sure to include all cases. [5 points]

$$f_X(x) = \begin{cases} \\ \\ \end{cases}$$

(b) Compute $\mathbb{E}\left[Y|(X = 1.374)\right]$. [3 points]

(c) Write down an integral expression for $\mathbb{P}(X + Y < 0)$. You should fill in the bounds and function we are integrating, but you do not need to evaluate the integral. [4 points]
Hint: You only need to think about one of the two main cases; this problem is much easier if you draw a doodle of the event.

$$\int_{\square}^{\square} \int_{\square}^{\square} \underline{\hspace{6cm}} \, d\square \, d\square$$

4. Select All Images With Bots [10 points]

(a) You are working on a CAPTCHA system[1] to detect whether a request came from a human or a bot. Let T be the event that the request passes your CAPTCHA test, and let B be the event that the request came from a bot. From testing, you have the following probabilities:

- If the request came from a bot, the probability that it passes the test is 0.2.
- If the request came from a human, the probability that they pass the test is 0.9.
- The probability of any given user being a bot is 0.25.

Find the probability a request came from a bot, given that it has passed the test. [4 points]

(b) You've started up your CAPTCHA system on your website, but some bots are still getting through and leaving comments – exactly 100 bot-comments appear each day.

On days where you are busy with other tasks, you briefly skim the comments and delete each bot-comment independently with probability 0.5. On a slow day, you can check much more carefully, and delete each bot-comment independently with probability 0.8. Each day is busy with probability 0.6, and slow otherwise.

What is the expected number of **undeleted** comments at the end of the day? [4 points]

(c) Suppose that you attempt the CAPTCHA test, where each attempt takes you Exponential$(1/30)$ seconds. Furthermore, the first time you attempt a CAPTCHA, it will take you 20 additional seconds to read the directions (and after the first time, you ignore the directions). What is the expected amount of time that it takes for you to attempt 10 CAPTCHAs, including the initial time spent on reading the directions once? [2 points]

[1] CAPTCHAs are things like "pick which pictures have a traffic light" tests to tell whether you're a human.

5. Knitting Requires Concentration, Again [16 points]

It's winter, and Robbie is working steadily on knitting a scarf. On day i, he knits R_i rows of the scarf, where R_i is a random variable with expectation 5 and variance 3. Treat the R_i variables as continuous (i.e., you could make a fractional part of a row).

(a) Use Markov's inequality to bound the probability that $R_i < 7.5$. [4 points]

$$\mathbb{P}\left(R_i < 7.5\right) \; \square_{\text{inequality}} \; \boxed{}_{\text{bound}}$$

(b) Use Chebyshev's inequality to bound the probability that $R_i < 2$ or $R_i > 8$. [4 points]

$$\mathbb{P}\left(R_i < 2 \text{ or } R_i > 8\right) \; \square_{\text{inequality}} \; \boxed{}_{\text{bound}}$$

(c) Robbie's getting much closer to completing the scarf. He decides to focus: now each day he knits with probability 0.6 (independently of other days). If he knits on a given day, he'll complete exactly 10 lines of knitting. If he doesn't knit that day, he knits zero lines.

Over the next 30 days, he hopes to knit at least 100 lines. Let X be the total number of lines he knits. In the following parts, we will use Chernoff to bound the probability that he knits strictly less than 100 lines.

(i) In English, detail the random variable that you will bound using Chernoff, i.e. give it a name (like X, Y, Z, etc.) and describe what it represents in terms of the problem statement. [1 point]

________ := ______________________________

RV name — English description

(ii) Calculate the expectation of your random variable in (i). [2 points]

(iii) Finally, bound the probability that Robbie knits less than 100 lines. [5 points]

$\mathbb{P}$ (less than 100 lines) [ ] inequality [ ] bound

6. Maximum Knitting [10 points]

Alice and Bob start a knitting business. They each are capable knitters, but they have to knit quickly to meet demand, which means they sometimes make mistakes. Suppose that Alice makes Poi(3) mistakes per project, while Bob makes Poi(2) mistakes per project.

They each have a pile of completed projects, but you aren't sure whose pile is whose. You decide to use maximum likelihood estimation to determine whether Alice or Bob is the creator of the pile right in-front of you (i.e., $\theta \in \{\text{Alice}, \text{Bob}\}$ is the true parameter, and you will use MLE to find $\hat{\theta}$, a prediction of θ).

(a) You observe that the first pile has 5 projects, where each project has 2, 3, 2, 1, and 3 errors respectively. Write the likelihood that this is Alice's pile. [4 points]

(b) Suppose that you picked up 5 projects and all had 3 or more errors. Which of the following is the best description of the MLE for this sample? [2 points]

- ◯ $\hat{\theta} = \text{Alice}$
- ◯ $\hat{\theta} = \text{Bob}$
- ◯ $\hat{\theta}$ is more likely to be Alice than Bob, but we can't be sure which it is.
- ◯ $\hat{\theta}$ is more likely to be Bob than Alice, but we can't be sure which it is.

(c) When finding an MLE, why do we take the log of the likelihood function? Mark ALL that apply [2 points]

- ☐ It (usually) makes taking derivatives easier while calculating the MLE.
- ☐ It (usually) makes integrating easier while calculating the MLE.
- ☐ It is required (i.e., your MLE may be incorrect if you skip the step).

(d) Suppose you are working with a continuous distribution. When finding an MLE for a relevant parameter, when do we take the natural logarithm ln instead of $\log_2$ or $\log_{10}$? [2 points]

- ◯ You must always use ln (not $\log_2$ or $\log_{10}$), or you will get the wrong MLE.
- ◯ You cannot use ln (you must use $\log_2$ or $\log_{10}$) in order to get the correct MLE.
- ◯ You cannot take the log (natural or otherwise) at all when finding an MLE.
- ◯ Any log base can be used; we pick whichever is most convenient.

7. Multiple Choice [22 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.

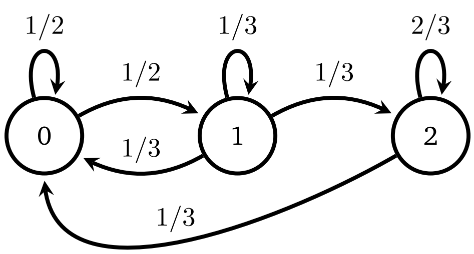
(a) Suppose I roll a fair 6-sided die, and then my friend rolls a fair k-sided die, where k was the outcome of my die roll. Let X be the outcome of my die roll, and Y be the outcome of my friend's die roll. Select ALL that apply. [3 points]

☐ Y is a discrete uniform random variable
☐ $p_{X,Y}(x,y) = \frac{1}{6x}$ for all integers x, y such that $0 \leq y \leq x \leq 6$
☐ $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$
☐ $X|(Y = 3)$ is a uniform random variable
☐ $Y|(X = 3)$ is a uniform random variable

(b) Let X be a random variable. Then $\mathbb{E}\left[(X-4)^2\right] \geq (\mathbb{E}[X-4])^2$. [3 points]

◯ Always
◯ Sometimes
◯ Never

(c) Consider the following Markov Chain:



What is the probability we are in state 0 at timestep two, if we were in state 1 at timestep zero? [3 points]

◯ $(1/3)^2 + (1/3)(1/2) + (1/2)^2$
◯ $(1/3)^2 + (1/3)(1/2)$
◯ $2(1/3)^2 + (1/3)(1/2)$
◯ $2(1/3)^2 + (1/2)^2$

(d) Using a trigram model trained on the following sentence, what is the probability of $\mathbb{P}$ [probability|I love]? [3 points]

"START START I love probability almost as much as I love TikTok and I love probability more than I love Roblox STOP"

◯ 1/2
◯ 1/4
◯ 1/10
◯ 1/11

(e) You repeatedly roll a fair 20-sided die until you roll the first 20. Let t be the probability that you see your first 20 on the 8th roll.
Suppose that your first two rolls are 1 and 2 respectively. Knowing this, what is the probability that you see your first 20 on the 10th roll? [3 points]

◯ t^2
◯ t
◯ $t^2(1-t)$
◯ $t(1-t)^9$
◯ The events described are disjoint, so you can't express one in terms of the other.

(f) Suppose you have a binomial random variable $X \sim \text{Bin}(1000, 0.3)$; you wish to estimate $\mathbb{P}[X \geq 250]$ using the Central Limit Theorem. Which of the following is the event you should use to apply a continuity correction? [3 points]

◯ $\mathbb{P}[X \geq 250.5]$
◯ $\mathbb{P}[X \geq 249.5]$
◯ You should not apply a continuity correction in this case.
◯ The Central Limit Theorem cannot be used here as 250 is less than $\mathbb{E}[X]$.

(g) For any continuous random variable X and any constant $k > 0$ we have that $\mathbb{P}[X \leq k] = \mathbb{P}[X^2 \leq k^2]$. [2 points]

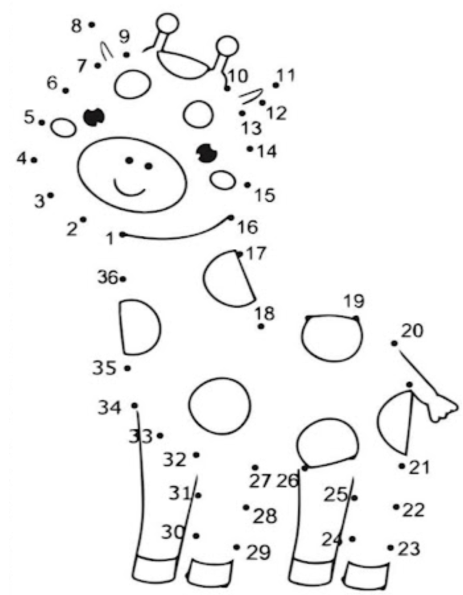
◯ True
◯ False

(h) For $X \sim \mathcal{N}(2, 100)$, we have that $\mathbb{P}[X \geq 10] \leq 1/5$ by Markov's inequality. [2 points]

◯ True
◯ False

8. Grading Morale [1 point]

Put something on this page, finish the image below, write a message to the teaching staff below, or do something else. Depending on your mood, this might be a poem, the statement of your favorite theorem, the statement of your least favorite theorem, or a piece of art. As long as you make some mark on this page, you will get the point. Looking at these helps keep the TAs happy while grading.



Have a relaxing winter break!

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 $E_1, \ldots, E_n$ **partition** the sample space Ω if:

- **(Exhaustive)** $E_1 \cup E_2 \cup \cdots \cup E_n = \bigcup_{i=1}^{n} E_i = \Omega$ (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 $E_1, \ldots, E_n$ 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 $E_1, \ldots, E_n$ 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 $A_1, \ldots, A_n$ be events with nonzero probabilities. Then:
$\mathbb{P}[A_1 \cap \cdots \cap A_n] =$
$\mathbb{P}[A_1]\, \mathbb{P}[A_2 \mid A_1]\, \mathbb{P}[A_3 \mid A_1 \cap A_2] \cdots \mathbb{P}[A_n \mid A_1 \cap \cdots \cap A_{n-1}]$

Definition: Mutual Independence (Events)

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

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

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 $F_X(t) = \mathbb{P}[X \le t]$.
If X is a *continuous* RV, $F_X(t) = \mathbb{P}[X \le t] = \int_{-\infty}^{t} f_X(w)\, dw$.

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:
$$f_{X,Y}(a, b) \ge 0$$
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 $X \ge 0$ be a **non-negative** RV, and let $k > 0$. Then:
$$\mathbb{P}[X \ge k] \le \frac{\mathbb{E}[X]}{k}$$
.

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,
$$\mathbb{P}[|X - \mu| \ge \alpha] \le \frac{\text{Var}(X)}{\alpha^2}$$
.

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 $1 > \delta > 0$:
$$\mathbb{P}(X \ge (1 + \delta)\mu) \le \exp(-\delta^2\mu/3)$$
$$\mathbb{P}(X \le (1 - \delta)\mu) \le \exp(-\delta^2\mu/2)$$

Theorem: The Union Bound

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

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 $\text{Var}(X) = p(1-p)$.

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 $\text{Var}(X) = np(1-p)$.

Definition: Uniform Random Variable (Discrete)

$X \sim$ Uniform(a, b) (Unif(a, b) for short), for integers $a \le b$, 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 $\text{Var}(X) = \frac{(b-a)(b-a+2)}{12}$.

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 $\text{Var}(X) = \frac{1-p}{p^2}$.

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 $\text{Var}(X) = \lambda$. 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 $\text{Var}(X) = \frac{(b-a)^2}{12}$.

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 $\text{Var}(X) = \frac{1}{\lambda^2}$.
$F_X(x) = 1 - e^{-\lambda x}$ for $x \ge 0$.

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 $\text{Var}(X) = \sigma^2$.

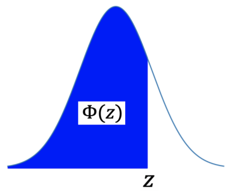
Theorem: Closure of the Normal Under Scale and Shift

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