Section 6

Review

- **Discrete to Continuous:**

	Discrete	Continuous
PMF/PDF	$p_X(x) = \mathbb{P}(X = x)$	$f_X(x) \neq \mathbb{P}(X = x) = 0$
CDF	$F_X(x) = \sum_{t \leqslant x} p_X(t)$	$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$
Normalization	$\sum_x p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$
Expectation	$\mathbb{E}[X] = \sum_x x p_X(x)$	$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$
LOTUS	$\mathbb{E}[g(X)] = \sum_x g(x) p_X(x)$	$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$

- **Uniform:** $X \sim$ Uniform(a, b) iff X has the following probability density function:

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

  $\mathbb{E}[X] = \frac{a+b}{2}$ and $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$. This represents each real number from $[a, b]$ to be equally likely.

- **Exponential:** $X \sim$ Exponential(λ) iff X has the following probability density function:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

  $\mathbb{E}[X] = \frac{1}{\lambda}$ and $\mathrm{Var}(X) = \frac{1}{\lambda^2}$. $F_X(x) = 1 - e^{-\lambda x}$ for $x \geqslant 0$. The exponential random variable is the continuous analog of the geometric random variable: it represents the waiting time to the next event, where $\lambda > 0$ is the average number of events per unit time. Note that the exponential measures how much time passes until the next event (any real number, continuous), whereas the Poisson measures how many events occur in a unit of time (nonnegative integer, discrete). The exponential random variable X is memoryless:

$$\text{for any } s, t \geqslant 0,\ \mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t)$$

  The geometric random variable also has this property.

- **Normal (Gaussian, "bell curve"):** $X \sim \mathcal{N}(\mu,\ \sigma^2)$ iff X has the following probability density function:

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

  $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$. The "standard normal" random variable is typically denoted Z and has mean 0 and variance 1: if $X \sim \mathcal{N}(\mu,\ \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$. The CDF has no closed form, but we denote the CDF of the standard normal as $\Phi(z) = F_Z(z) = \mathbb{P}(Z \leqslant z)$. Note from symmetry of the probability density function about $z = 0$ that: $\Phi(-z) = 1 - \Phi(z)$.

  Here is the **Standard normal table (link found on the course website).**

- **Standardizing**: Let X be any random variable (discrete or continuous, not necessarily normal), with $\mathbb{E}[X] = \mu$ and $Var(X) = \sigma^2$. If we let $Y = \frac{X-\mu}{\sigma}$, then $\mathbb{E}[Y] = 0$ and $Var(Y) = 1$.

- **Closure of the Normal Distribution**: Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then, $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$. That is, linear transformations of normal random variables are still normal.

- **'Reproductive" Property of Normals**: Let $X_1, \ldots, X_n$ be independent normal random variables with $\mathbb{E}[X_i] = \mu_i$ and $Var(X_i) = \sigma_i^2$. Let $a_1, \ldots, a_n \in \mathbb{R}$ and $b \in \mathbb{R}$. Then,

$$X = \sum_{i=1}^{n}(a_i X_i + b) \sim \mathcal{N}\left(\sum_{i=1}^{n}(a_i\mu_i + b), \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

  There's nothing special about the parameters – the important result here is that the resulting random variable is still normally distributed.

- **Central Limit Theorem (CLT)**: Let $X_1, \ldots, X_n$ be iid random variables with $\mathbb{E}[X_i] = \mu$ and $Var(X_i) = \sigma^2$. Let $X = \sum_{i=1}^{n} X_i$, which has $\mathbb{E}[X] = n\mu$ and $Var(X) = n\sigma^2$. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, which has $\mathbb{E}[\overline{X}] = \mu$ and $Var(\overline{X}) = \frac{\sigma^2}{n}$. $\overline{X}$ is called the *sample mean*. Then, as $n \to \infty$, $\overline{X}$ approaches the normal distribution $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$. Standardizing, this is equivalent to $Y = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ approaching $\mathcal{N}(0,1)$. Similarly, as $n \to \infty$, X approaches $\mathcal{N}(n\mu, n\sigma^2)$ and $Y' = \frac{X-n\mu}{\sigma\sqrt{n}}$ approaches $\mathcal{N}(0,1)$.

  It is no surprise that $\overline{X}$ has mean μ and variance σ^2/n – this can be done with simple calculations. The importance of the CLT is that, for large n, regardless of what distribution X_i comes from, $\overline{X}$ is *approximately normally distributed with mean μ and variance σ^2/n.*

- **Continuity Correction**: This is a technique for getting a better estimate when applying CLT to the sum $X = \sum_{i=1}^{n} X_i$ or the average of a set of random variables $X_1, \ldots, X_n$ that are discrete. Specifically, if asked to compute $\mathbb{P}(a \leqslant X \leqslant b)$ where $a \leqslant b$ are integers, you should compute $\mathbb{P}(a - 0.5 \leqslant X \leqslant b + 0.5)$ so that the width of the interval being integrated is the same as the number of terms you are summing over $(b-a+1)$. Note that if you applying the CLT to sums/averages of continuous RVs instead, you should not apply the continuity correction.

- **Continuous Law of Total Probability:**

  Suppose that E is an event, and X is a continuous random variable with density function $f_X(x)$. Then

$$\mathbb{P}(E) = \int_{-\infty}^{\infty} \mathbb{P}(E \mid X = x) f_X(x) dx$$

Task 1 – Content Review

a) True or False: For any random variable X, $\mathbb{P}(X = 5) = \mathbb{P}(X - 5 = 0)$.

b) True or False: For some continuous random variable X, $\mathbb{P}(X \leqslant 5) \neq \mathbb{P}(X < 5)$.

c) True or False: Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$. Then $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

d) Select one: For an event A and a continuous random variable X with density $f_X(x)$,

◯ $\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x)\mathbb{P}(X = x) dx$

◯ $\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x) f_X(x) dx$

◯ $\mathbb{P}(A) = \int_{-\infty}^{\infty} x f_X(x) dx$

◯ $\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x) dx$

e) Select one: Suppose we have n independent and identically distributed random variables $X_1, X_2, \ldots, X_n$, each with mean μ and variance σ^2. Let $X = \sum_{i=1}^{n} X$. Then as n grows large, the Central Limit Theorem tells us that X behaves similarly to which normal distribution?

◯ $X \sim \mathcal{N}(n\mu, n\sigma^2)$

- ◯ $X \sim \mathcal{N}(\mu, n\sigma^2)$
- ◯ $X \sim \mathcal{N}(n\mu, \sigma^2)$
- ◯ $X \sim \mathcal{N}(n\mu, n^2\sigma^2)$

f) Select one: Given two discrete random variables X and Y, the joint CDF is

- ◯ $F_{X,Y}(x, y) = \sum_{t<x} p_{X,Y}(t, y)$
- ◯ $F_{X,Y}(x, y) = \sum_{s<y} p_{X,Y}(x, s)$
- ◯ $F_{X,Y}(x, y) = \sum_{t<x} \sum_{s<y} p_{X,Y}(t, s)$
- ◯ $F_{X,Y}(x, y) = p_{X,Y}(x, y)$

Task 2 – The exponential distribution is memoryless (problem from lecture)

Show that the exponential distribution is memoryless. Specifically, suppose that X is exponential with parameter λ. Show that $\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t)$.

Task 3 – More practice with exponentials (problem from lecture)

The time it takes to check someone out at a grocery store is exponential with an expected value of 10 minutes. Suppose that when you arrive at a grocery store, there is one person in the middle of being served. What is the probability that you will have to wait between 10 and 20 minutes before that person is done being served?

Task 4 – Will the battery last?

The owner of a car starts on a 5000 mile road trip. Suppose that the number of miles that the car will run before its battery wears out is exponentially distributed with expectation 10,000 miles. After successfully driving for 2000 miles on the trip without the battery wearing out, what is the probability that she will be able to complete the trip without replacing the battery?

Task 5 – Batteries and exponential distributions (from Section 6)

Let X_1, X_2 be independent exponential random variables, where X_i has parameter λ_i, for $1 \leqslant i \leqslant 2$. Let $Y = \min(X_1, X_2)$.

a) Show that Y is an exponential random variable with parameter $\lambda = \lambda_1 + \lambda_2$. Hint: Start by computing $\mathbb{P}(Y > y)$. Two random variables with the same CDF have the same pdf. Why?

b) What is $\mathbb{P}(X_1 < X_2)$? (Use the continuous version of the law of total probability, conditioning on the probability that $X_1 = x$.)

c) You have a digital camera that requires two batteries to operate. You purchase n batteries, labelled $1, 2, \ldots, n$, each of which has a lifetime that is exponentially distributed with parameter λ, independently of all other batteries. Initially, you install batteries 1 and 2. Each time a battery fails, you replace it with the lowest-numbered unused battery. At the end of this process, you will be left with just one working battery. What is the expected total time until the end of the process? Justify your answer.

d) In the scenario of the previous part, what is the probability that battery i is the last remaining battery as a function of i? (You might want to use the memoryless property of the exponential distribution that has been discussed.)

Task 6 – Grading on a curve

In some classes (not CSE classes) an examination is regarded as being good (in the sense of determining a valid spread for those taking it) if the test scores of those taking it are well approximated by a normal density function. The instructor often uses the test scores to estimate the normal parameters μ and σ^2 and then assigns a letter grade of A to those whose test score is greater than $\mu + \sigma$, B to those whose score is between μ and $\mu + \sigma$, C to those whose score is between $\mu - \sigma$ and μ, D to those whose score is between $\mu - 2\sigma$ and $\mu - \sigma$ and F to those getting a score below $\mu - 2\sigma$. If the instructor does this and a student's grade on the test really is normally distributed with mean μ and variance σ^2, what is the probability that student will get each of the possible grades A,B,C,D and F? (Use a table for anything you can't calculate.)

Task 7 – Normal questions at the table (from Section 6)

a) Let X be a normal random with parameters $\mu = 10$ and $\sigma^2 = 36$. Compute $\mathbb{P}(4 < X < 16)$.

b) Let X be a normal random variable with mean 5. If $\mathbb{P}(X > 9) = 0.2$, approximately what is $\mathrm{Var}(X)$?

c) Let X be a normal random variable with mean 12 and variance 4.
Find the value of c such that $\mathbb{P}(X > c) = 0.10$.

Central Limit Theorem Problems

The next few problems are CLT focused problems. Here's a general template for that! Sometimes we'll be trying to solve for the probability of something (e.g., $P(X \leqslant 10)$, and sometimes, we'll be trying to find a value of some parameter that will allow for the probability to be in a certain range (e.g., $P(X \leqslant 10) \leqslant 0.2$). Regardless, we still will want to apply CLT on X (the only difference is that we may be solving for different things).

1. Setup the problem - write event you are interested in, in terms of sum of random variables. (what do we want to solve for/what is the probability we want to be true?)
   - Write the random variable we're interested in as a sum of i.i.d., random variables
   - Apply CLT to $X = X_1 + X_2 + ... + X_n$ (we can approximate X as a normal random variable $Y \sim N(\mu, \sigma^2)$)
   - Write the probability we're interested in
2. *If the RVs are discrete*, apply continuity correction.
3. Normalize RV to have mean 0 and standard deviation 1: $Z = \frac{Y-\mu}{\sigma}$
4. Replace RV in probability expression with $Z \sim N(0, 1)$
5. Write in terms of $\Phi(z) = P(Z \leqslant z)$
6. Look up in the Phi table (or a reverse Phi table lookup if we're for a value of z that gives a certain probability)

Task 8 – Round-off error

Let X be the sum of 100 real numbers, and let Y be the same sum, but with each number rounded to the nearest integer before summing. If the roundoff errors are independent and uniformly distributed between -0.5 and 0.5, what is the approximate probability that $|X - Y| > 3$?

Task 9 – Tweets

A prolific Twitter user tweets approximately 350 tweets per week. Let's assume for simplicity that the tweets are independent, and each consists of a uniformly random number of characters between 10 and 140. (Note that this is a discrete uniform distribution.) Thus, the central limit theorem (CLT) implies that the number of characters tweeted by this user is approximately normal with an appropriate mean and variance. Assuming this normal approximation is correct, estimate the probability that this user tweets between 26,000 and 27,000 characters in a particular week. (This is a case where continuity correction will make virtually no difference in the answer, but you should still use it to get into the practice!).

Task 10 – Confidence interval

Suppose that $X_1, \ldots, X_n$ are i.i.d. samples from a normal distribution with unknown mean μ and variance 36. How big does n need to be so that μ is in

$$[\overline{X} - 0.11, \overline{X} + 0.11]$$

with probability at least 0.97?

Recall that

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

You may use the fact that $\Phi^{-1}(0.985) = 2.17$.

Task 11 – Normal Approximation of a Sum

Imagine that we are trying to transmit a signal. During the transmission, there are 100 sources independently making low noise. Each source produces an amount of noise that is uniformly distributed between $a = -1$ and $b = 1$. If the total amount of noise is greater than 10 or less than -10, then it corrupts the signal. However, if the absolute value of the total amount of noise is under 10, then it is not a problem. What is the approximate probability that the absolute value of the total amount of noise from the 100 signals is less than 10?

Task 12 – Joint PMF's

Suppose X and Y have the following joint PMF:

X/Y	1	2	3
0	0	0.2	0.1
1	0.3	0	0.4

a) Identify the range of X (Ω_X), the range of Y (Ω_Y), and their joint range ($\Omega_{X,Y}$).

b) Find the marginal PMF for X, $p_X(x)$ for $x \in \Omega_X$.

c) Find the marginal PMF for Y, $p_Y(y)$ for $y \in \Omega_Y$.

d) Are X and Y independent? Why or why not?

e) Find $\mathbb{E}[X^3Y]$.

Task 13 – Joint PMF's

Suppose X and Y have the following joint PMF:

X/Y	1	2	3
0	0	0.2	0.1
1	0.3	0	0.4

a) Identify the range of X (Ω_X), the range of Y (Ω_Y), and their joint range ($\Omega_{X,Y}$).

b) Find the marginal PMF for X, $p_X(x)$ for $x \in \Omega_X$.

c) Find the marginal PMF for Y, $p_Y(y)$ for $y \in \Omega_Y$.

d) Are X and Y independent? Why or why not?

e) Find $\mathbb{E}[X^3Y]$.

Task 14 – Do You "Urn" to Learn More About Probability?

Suppose that 3 balls are chosen without replacement from an urn consisting of 5 white and 8 red balls. Let $X_i = 1$ if the i-th ball selected is white and let it be equal to 0 otherwise. Give the joint probability mass function of

a) X_1, X_2

b) X_1, X_2, X_3

Task 15 – Trinomial Distribution

A generalization of the Binomial model is when there is a sequence of n independent trials, but with three outcomes, where $\mathbb{P}(\text{outcome } i) = p_i$ for $i = 1, 2, 3$ and of course $p_1 + p_2 + p_3 = 1$. Let X_i be the number of times outcome i occurred for $i = 1, 2, 3$, where $X_1 + X_2 + X_3 = n$. Find the joint PMF $p_{X_1,X_2,X_3}(x_1, x_2, x_3)$ and specify its value for all $x_1, x_2, x_3 \in \mathbb{R}$.

Are X_1 and X_2 independent?

Task 16 – Successes

Consider a sequence of independent Bernoulli trials, each of which is a success with probability p. Let X_1 be the number of failures preceding the first success, and let X_2 be the number of failures after the first success but preceding the second success. Find the joint pmf of X_1 and X_2. Write an expression for $\mathbb{E}[\sqrt{X_1 X_2}]$. You can leave your answer in the form of a sum.

Task 17 – Who fails first?

Here's a question that commonly comes up in industry, but isn't immediately obvious. You have a disk with probability p_1 of failing each day. You have a CPU which independently has probability p_2 of failing each day. What is the probability that your disk fails *before* your CPU?

a) Compute the probability by summing over the relevant part of the probability space.

b) Try to provide an intuitive reason for the answer.

c) Recompute the probability using the law of total probability, conditioning on the value of X_1.

Task 18 – Continuous joint density

The joint density of X and Y is given by

$$f_{X,Y}(x, y) = \begin{cases} xe^{-(x+y)} & x > 0, y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

and the joint density of W and V is given by

$$f_{W,V}(w, v) = \begin{cases} 2 & 0 < w < v, 0 < v < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent? Are W and V independent?

Task 19 – Grades and homework turn-in time

Suppose we're currently trying to find a relationship between the time a student turns in their homework and the grade that they receive on the respective homework. Let T denote the amount of time *prior* to the deadline that the homework is submitted. We have observed that no student submits the homework more than 2 days earlier than the deadline, and also no student submits their assignment late, so $0 \leqslant T \leqslant 2$. Now let G be a random variable, indicating the percentage that the student receives on the homework assignment, that is, $0 \leqslant G \leqslant 1$. Suppose G and T are continuous random variables, and their joint pdf is given by

$$f_{G,T}(g, t) = \begin{cases} \frac{9}{10}g^2 t + \frac{1}{5} & \text{when } 0 \leqslant g \leqslant 1 \text{ and } 0 \leqslant t \leqslant 2 \\ 0 & \text{otherwise .} \end{cases}$$

For both parts, round your solution to three decimal places.

a) What is the probability that a randomly selected student gets a grade above 50% on the homework?

b) What is the probability that a student gets a grade above 50%, given that the student submitted less than a day before the deadline?