Section 7 – Solutions

Review

1) 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	$\mathbb{E}[g(X)] = \sum_x g(x) p_X(x)$	$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$

2) 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

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

3) 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.

4) 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.

5) 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.**

6) **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 $Var(Y) = 1$.

7) **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.

8) **'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.

9) **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.*

10) **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.

Task 1 – 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 $\mathbb{P}(X > t + s | X > s) = \mathbb{P}(X > t)$.

$$\begin{aligned}
\mathbb{P}(X > t + s | X > s) &= \frac{\mathbb{P}(X > t + s \cap X > s)}{\mathbb{P}(X > s)} \\
&= \frac{\mathbb{P}(X > t + s)}{\mathbb{P}(X > s)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} \\
&= e^{-\lambda t} \\
&= 1 - F_X(t).
\end{aligned}$$

Task 2 – 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?

Since the expected value of an exponential random variable is $1/\lambda$, we have $1/\lambda = 10$ minutes, so $\lambda = 1/10$. In addition, since the exponential distribution is memoryless (that is, it doens't matter how long the person being served has already been there), the time that you will have to wait is exponential with parameter $1/10$. Thus

$$\mathbb{P}(10 \leqslant T \leqslant 20) = \int_{10}^{20} \frac{1}{10} e^{-x/10} dx = e^{-1} - e^{-2}$$

Task 3 – 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?

We start with computing $\mathbb{P}(Y > y)$, by substituting in the definition of Y.

$$\mathbb{P}(Y > y) = \mathbb{P}(\min\{X_1, X_2\} > y)$$

The probability that the minimum of two values is above a value is the chance that both of them are above that value. From there, we can separate them further because X_1 and X_2 are independent.

$$\mathbb{P}(X_1 > y \cap X_2 > y) = \mathbb{P}(X_1 > y)\mathbb{P}(X_2 > y) = e^{-\lambda_1 y} e^{-\lambda_2 y}$$
$$= e^{-(\lambda_1 + \lambda_2) y} = e^{-\lambda y}$$

So $F_Y(y) = 1 - \mathbb{P}(Y > y) = 1 - e^{-\lambda y}$ and $f_Y(y) = \lambda e^{-\lambda y}$ so $Y \sim$ Exp(λ), since this is the same CDF and PDF as an exponential distribution with parameter $\lambda = \lambda_1 + \lambda_2$.

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

By the law of total probability,

$$\mathbb{P}(X_1 < X_2) = \int_0^\infty \mathbb{P}(X_1 < X_2 \mid X_1 = x) f_{X_1}(x) \, \mathrm{d}x = \int_0^\infty \mathbb{P}(X_2 > x) \lambda_1 e^{-\lambda_1 x} \, \mathrm{d}x =$$
$$\int_0^\infty e^{-\lambda_2 x} \lambda_1 e^{-\lambda_1 x} \, \mathrm{d}x = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

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.

Let T be the time until the end of the process. We are trying to find $\mathbb{E}[T]$. $T = Y_1 + ... + Y_{n-1}$ where Y_i is the time until we have to replace a battery from the ith pair. The reason it there are only $n-1$ RVs in the sum is because there are $n-1$ times where we have two batteries and wait for one to fail. By part (a), the time for one to fail is the min of exponentials, so $Y_i \sim$ Exponential(2λ). Hence the expected time for the first battery to fail is $\frac{1}{2\lambda}$. By linearity and memorylessness, $\mathbb{E}[T] = \sum_{i=1}^{n-1} \mathbb{E}[Y_1] = \frac{n-1}{2\lambda}$.

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.)

If there are two batteries i, j in the flashlight, by part (b), the probability each outlasts each other is $1/2$. Hence, the last battery n has probability $1/2$ of being the last one remaining. The second to last battery $n-1$ has to beat out the previous battery and the n^{th}, so the probability it lasts the longest is $(1/2)^2 = 1/4$. Work down inductively to get that the probability the i^{th} is the last remaining is $(1/2)^{n-i+1}$ for $i \geqslant 3$. Finally the first two batteries share the remaining probability as they start at the same time, with probability $(1/2)^{n-1}$ each.

Task 4 – Normal questions at the table (from Section 6)

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

Let $\frac{X-10}{6} = Z$. By the scale and shift properties of normal random variables $Z \sim \mathcal{N}(0,1)$.

$$\mathbb{P}(4 < X < 16) = \mathbb{P}\left(\frac{4-10}{6} < \frac{X-10}{6} < \frac{16-10}{6}\right) = \mathbb{P}(-1 < Z < 1)$$

$$= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.68268$$

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

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

$$\mathbb{P}(X > 9) = \mathbb{P}\left(\frac{X-5}{\sigma} > \frac{9-5}{\sigma}\right) = 1 - \Phi\left(\frac{4}{\sigma}\right) = 0.2$$

So, $\Phi\left(\frac{4}{\sigma}\right) = 0.8$. Looking up the phi values in reverse lets us undo the Φ function, and gives us $\frac{4}{\sigma} = 0.845$. Solving for σ we get $\sigma \approx 4.73$, which means that the variance is about 22.4.

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

$$\mathbb{P}(X > c) = \mathbb{P}\left(\frac{X-12}{2} > \frac{c-12}{2}\right) = 1 - \Phi\left(\frac{c-12}{2}\right) = 0.1$$

So, $\Phi\left(\frac{c-12}{2}\right) = 0.9$. Looking up the phi values in reverse lets us undo the Φ function, and gives us $\frac{c-12}{2} = 1.29$. Solving for c we get $c \approx 14.58$.

Task 5 – 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 $|X - Y| > 3$?

Let $X = \sum_{i=1}^{100} X_i$, and $Y = \sum_{i=1}^{100} r(X_i)$, where $r(X_i)$ is X_i rounded to the nearest integer. Then, we have

$$X - Y = \sum_{i=1}^{100} X_i - r(X_i)$$

Note that each $X_i - r(X_i)$ is simply the round off error, which is distributed as $\mathrm{Unif}(-0.5, 0.5)$. Since $X - Y$ is the sum of 100 i.i.d. random variables with mean $\mu = 0$ and variance $\sigma^2 = \frac{1}{12}$, $X - Y \approx W \sim \mathcal{N}(0, \frac{100}{12})$ by the Central Limit Theorem. For notational convenience let $Z \sim \mathcal{N}(0, 1)$

$$\begin{aligned}
\mathbb{P}(|X-Y| > 3) &\approx \mathbb{P}(|W| > 3) && \text{[CLT]} \\
&= \mathbb{P}(W > 3) + \mathbb{P}(W < -3) && \text{[No overlap between } W > 3 \text{ and } W < -3\text{]} \\
&= 2\,\mathbb{P}(W > 3) && \text{[Symmetry of normal]} \\
&= 2\,\mathbb{P}\left(\frac{W}{\sqrt{100/12}} > \frac{3}{\sqrt{100/12}}\right) \\
&\approx 2\,\mathbb{P}(Z > 1.039) && \text{[Standardize } W\text{]} \\
&= 2\,(1 - \Phi(1.039)) \approx 0.29834
\end{aligned}$$

Task 6 – 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!).

Let X be the total number of characters tweeted by a twitter user in a week. Let $X_i \sim Unif(10, 140)$ be the number of characters in the ith tweet (since the start of the week). Since X is the sum of 350 i.i.d. rvs with mean $\mu = 75$ and variance $\sigma^2 = 1430$, $X \approx N \sim \mathcal{N}(350 \cdot 75, 350 \cdot 1430)$. Thus,

$$\mathbb{P}(26,000 \leqslant X \leqslant 27,000) \approx \mathbb{P}(26,000 \leqslant N \leqslant 27,000)$$

Now, we apply continuity correction:

$$\mathbb{P}(26,000 \leqslant N \leqslant 27,000) \approx \mathbb{P}(25,999.5 \leqslant N \leqslant 27,000.5)$$

Standardizing this gives the following formula

$$\begin{aligned}
\mathbb{P}(25,999.5 \leqslant N \leqslant 27,000.5) &\approx \mathbb{P}\left(-0.3541 \leqslant \frac{N - 350 \cdot 75}{\sqrt{350 \cdot 1430}} \leqslant 1.0608\right) \\
&= \mathbb{P}\left(-0.3541 \leqslant Z \leqslant 1.0608\right) \\
&= \mathbb{P}(Z \leqslant 1.0608) - \mathbb{P}(Z \leqslant -0.3541) \\
&= \Phi(1.0608) - \Phi(-0.3541) \\
&\approx 0.4923
\end{aligned}$$

So the probability that this user tweets between 26,000 and 27,000 characters in a particular week is approximately 0.4923.

Task 7 – 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 $\Phi^{-1}(0.985) = 2.17$.

Our goal is to find n such that μ lies within 0.11 of $\bar{X}$ 97% of the time. This is equivalent to finding n such that the probability that μ lies outside the range is less than 3%.

$$\mathbb{P}(|\bar{X} - \mu| > 0.11) \leqslant 0.03$$

Let us define $Z = \frac{\bar{X}-\mu}{\sigma}$. We can solve for σ by using the Properties of Variance. Since

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

we can say that

$$\text{Var}(\bar{X}) = \text{Var}(\frac{1}{n}\sum_{i=1}^{n} X_i)$$

Using the Properties of Variance and the fact that X_i's are i.i.d., $\text{Var}(\bar{X}) = \frac{1}{n^2} \cdot n \cdot 36 = \frac{36}{n}$, so $\sigma = \frac{6}{\sqrt{n}}$.

$$\begin{aligned}
\mathbb{P}(|\bar{X} - \mu| > 0.11) &\leqslant 0.03 \\
\mathbb{P}(|Z| \cdot \sigma > 0.11) &\leqslant 0.03 && \text{[Definition of } Z\text{]} \\
\mathbb{P}\left(|Z| > \frac{0.11}{6}\sqrt{n}\right) &\leqslant 0.03 \\
\mathbb{P}\left(Z < -\frac{0.11}{6}\sqrt{n}\right) &\leqslant 0.015 && \text{[Symmetry of Normal Dist.]} \\
\Phi\left(-\frac{0.11}{6}\sqrt{n}\right) &\leqslant 0.015 && \text{[CDF of Standard Norm.]} \\
-\frac{0.11}{6}\sqrt{n} &\leqslant -\Phi^{-1}(0.985) \\
\sqrt{n} &\geqslant \frac{6 \cdot \Phi^{-1}(0.985)}{0.11} \\
n &\geqslant \left(\frac{6 \cdot \Phi^{-1}(0.985)}{0.11}\right)^2 \\
&\approx 14009.95
\end{aligned}$$

Then n must be at least 14010.

Task 8 – 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?

Let S be the total amount of noise. We want to find $\mathbb{P}(|S| < 10) = \mathbb{P}(-10 < S < 10)$. Let X_i be the noise from source i. Then, we have

$$S = \sum_{i=1}^{100} X_i.$$

Since the X_i are uniformly distributed, we have that $\mathbb{E}[X_i] = \frac{a+b}{2} = 0$ and $\mathrm{Var}(X_i) = \frac{(b-a)^2}{12} = \frac{1}{3}$. Since the X_i are i.i.d, by the Central Limit Theorem, we find that S is approximately distributed according to $N\left(0, 100 \cdot \frac{1}{3}\right)$. Now, we standardize to get

$$\begin{aligned} \mathbb{P}(-10 < S < 10) &= \mathbb{P}\left(\frac{-10-0}{\sqrt{100/3}} < \frac{S-0}{\sqrt{100/3}} < \frac{10-0}{\sqrt{100/3}}\right) \\ &= 2\Phi(\sqrt{3}) - 1 \approx 0.91 \end{aligned}$$