Section 6

Review

1) Cumulative Distribution Function (cdf): For any random variable (discrete or continuous) X, the cumulative distribution function is defined as

$$F_X(x) = \mathbb{P}(X \leqslant x).$$

Notice that this function must be monotonically nondecreasing: if $x < y$ then $F_X(x) \leqslant F_X(y)$, because $\mathbb{P}(X \leqslant x) \leqslant \mathbb{P}(X \leqslant y)$. Also notice that since probabilities are between 0 and 1, that $0 \leqslant F_X(x) \leqslant 1$ for all x, with $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to+\infty} F_X(x) = 1$.

2) Continuous Random Variable: A continuous random variable X is one for which its cumulative distribution function $F_X(x) : \mathbb{R} \to \mathbb{R}$ is continuous everywhere. A continuous random variable has an uncountably infinite number of values.

3) Probability Density Function (pdf or density): Let X be a continuous random variable. Then the probability density function $f_X(x) : \mathbb{R} \to \mathbb{R}$ of X is defined as $f_X(x) = \frac{d}{dx}F_X(x)$. Turning this around, it means that $F_X(x) = \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} f_X(t)\,dt$. From this, it follows that $\mathbb{P}(a \leqslant X \leqslant b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx$ and that $\int_{-\infty}^{\infty} f_X(x)dx = 1$. From the fact that $F_X(x)$ is monotonically nondecreasing it follows that $f_X(x) \geqslant 0$ for every real number x.

If X is a continuous random variable, note that in general $f_X(a) \neq \mathbb{P}(X = a)$, since $\mathbb{P}(X = a) = F_X(a) - F_X(a) = 0$ for all a. However, the probability that X is close to a is proportional to $f_X(a)$: for small δ, $\mathbb{P}\left(a - \frac{\delta}{2} < X < a + \frac{\delta}{2}\right) \approx \delta f_X(a)$.

4) i.i.d. (independent and identically distributed): Random variables $X_1, \ldots, X_n$ are i.i.d. (or iid) if they are independent and have the same probability mass function or probability density function.

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

6) Uniform: $X \sim \text{Uniform}(a, b)$ 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 $\text{Var}(X) = \frac{(b-a)^2}{12}$. This represents each real number from $[a, b]$ to be equally likely.

7) Exponential: $X \sim \text{Exponential}(\lambda)$ 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 $\text{Var}(X) = \frac{1}{\lambda^2}$. $F_X(x) = 1 - e^{-\lambda x}$ 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:

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

The geometric random variable also has this property.

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

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

$\mathbb{E}[X] = \mu$ 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 $\Phi\left(z\right) = F_Z\left(z\right) = \mathbb{P}\left(Z \leqslant z\right)$. Note from symmetry of the probability density function about $z = 0$ that: $\Phi\left(-z\right) = 1 - \Phi(z)$.

Task 1 – Continuous r.v. example

Suppose that X is a random variable with pdf

$$f_X(x) = \begin{cases} 2C(2x - x^2) & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

where C is an appropriately chosen constant.

a) What must the constant C be for this to be a valid pdf?

b) Using this C, what is $\mathbb{P}(X > 1)$?

Task 2 – 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 3 – Create the distribution

Suppose X is a continuous random variable that is uniform on $[0,1]$ and uniform on $[1,2]$, but

$$\mathbb{P}(1 \leqslant X \leqslant 2) = 2 \cdot \mathbb{P}(0 \leqslant X < 1).$$

Outside of $[0,2]$ the density is 0. What is the PDF and CDF of X?

Task 4 – Max of uniforms

Let $U_1, U_2, \ldots, U_n$ be mutually independent Uniform random variables on $(0,1)$. Find the CDF and PMF for the random variable $Z = \max(U_1, \ldots, U_n)$.

Task 5 – 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 6 – Throwing a dart

Consider the closed unit circle of radius r, i.e., $S = \{(x, y) : x^2 + y^2 \leqslant r^2\}$. Suppose we throw a dart onto this circle and are guaranteed to hit it, but the dart is equally likely to land anywhere in S. Concretely this means that the probability that the dart lands in any particular area of size A (that is entirely inside the circle of radius R), is equal to $\frac{\text{A}}{\text{Area of whole circle}}$. The density outside the circle of radius r is 0.

Let X be the distance the dart lands from the center. What is the CDF and pdf of X? What is $\mathbb{E}[X]$ and $\text{Var}(X)$?

Task 7 – A square dartboard?

You throw a dart at an $s \times s$ square dartboard. The goal of this game is to get the dart to land as close to the lower left corner of the dartboard as possible. However, your aim is such that the dart is equally likely to land at any point on the dartboard. Let random variable X be the length of the side of the smallest *square* B in the lower left corner of the dartboard that contains the point where the dart lands. That is, the lower left corner of B must be the same point as the lower left corner of the dartboard, and the dart lands somewhere along the upper or right edge of B. For random variable X, find the CDF, PDF, $\mathbb{E}[X]$, and $\text{Var}(X)$.

Task 8 – Normal questions at the table

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 $\text{Var}(X)$?

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

Task 9 – Batteries and exponential distributions

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 law of total probability.)

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.)