Section 6

Slides by Khanh Nguyen and Nathan Akkaraphab

Administrivia

- Pset 5 due next Wednesday (Feb. 15th) at 11:59 pm PDT
- Midterm on Monday (Feb. 13th) 9:30-10:20 am
    - Will cover everything up till and including Lecture 11 and everything covered in the first 4 sections.
    - **Tentatively, 2 numerical + 4 True/False questions**
    - See Anna's posts on EdStem

Agenda

- **Section problems**
  - Task 1: “Continuous RV example” (or Task 3: “Create the distribution”)
  - Task 6: “Throwing a dart”
  - Task 2: “Will the battery last?” (if time permitted)
- **Midterm review**

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

Review

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

2) **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 $\mathrm{Var}(X) = \frac{1}{\lambda^2}$. $F_X(x) = 1 - e^{-\lambda x}$ 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.

3) 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: $\Phi(-z) = 1 - \Phi(z)$.

Question 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?

Part a)

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

For $f_X(x)$ to be a valid PDF, $f_X(x)$ must be non-negative and the area under the graph must be 1. For $0 \leqslant x \leqslant 2$, we have $2x - x^2 = x(2-x) \geqslant 0$ so we only need $C \geqslant 0$ for f_X to be non-negative everywhere. Computing the area under the graph as a function of C gives us

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^2 2C(2x - x^2)dx = 2C\left(x^2 - \frac{1}{3}x^3\Big|_0^2\right) = 2C\frac{4}{3} = \frac{8}{3}C$$

Setting this equation equal to 1, and solving for C gives use $C = \frac{3}{8}$.

Part b)

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

Part b)

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

The $\mathbb{P}(X > 1) = \int_1^\infty f_X(t)dt$. Using our value for C that we found in the previous part we can compute this integral as follows:

$$\int_1^\infty f_X(t)dt = \int_1^2 \frac{6}{8}\left(2x - x^2\right) dt = \frac{6}{8}\left(x^2 - \frac{1}{3}x^3\right)\Big|_1^2 = \frac{1}{2}$$

Question 3: “Create the distribution”

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

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

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

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

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

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

The fact that X is uniform on each of the intervals means that its PDF is constant on each. So, Suppose X is a continuous random variable that is uniform on $[0, 1]$ and uniform on $[1, 2]$, but

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

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

The fact that X is uniform on each of the intervals means that its PDF is constant on each. So,

$$f_X(x) = \begin{cases} c & 0 < x \leq 1 \\ d & 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

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

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

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

The fact that X is uniform on each of the intervals means that its PDF is constant on each. So,

$$f_X(x) = \begin{cases} c & 0 < x \leq 1 \\ d & 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Note that $F_X(1) - F_X(0) = c$ and $F_X(2) - F_X(1) = d$. The area under the PDF must be 1, so

$$1 = F_X(2) - F_X(0) = F_X(2) - F_X(1) + F_X(1) - F_X(0) = d + c$$

Additionally,

$$d = F_X(2) - F_X(1) = \mathbb{P}(1 \leq X \leq 2) = 2 \cdot \mathbb{P}(0 \leq X \leq 1) = 2 \cdot (F_X(1) - F_X(0)) = 2c$$

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

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

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

Additionally,

$$d = F_X(2) - F_X(1) = \mathbb{P}(1 \leq X \leq 2) = 2 \cdot \mathbb{P}(0 \leq X \leq 1) = 2 \cdot (F_X(1) - F_X(0)) = 2c$$

To solve for c and d in our PDF, we need only solve the system of two equations from above: $d + c = 1$, $d = 2c$. So, $d = \frac{2}{3}$ and $c = \frac{1}{3}$. Taking the integral of the PDF yields the CDF, which looks like

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

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

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

Additionally,

$$d = F_X(2) - F_X(1) = \mathbb{P}(1 \leq X \leq 2) = 2 \cdot \mathbb{P}(0 \leq X \leq 1) = 2 \cdot (F_X(1) - F_X(0)) = 2c$$

To solve for c and d in our PDF, we need only solve the system of two equations from above: $d + c = 1$, $d = 2c$. So, $d = \frac{2}{3}$ and $c = \frac{1}{3}$. Taking the integral of the PDF yields the CDF, which looks like

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{3}x & 0 < x \leq 1 \\ \frac{2}{3}x - \frac{1}{3} & 1 < x \leq 2 \\ 1 & x > 2 \end{cases}$$

Question 6: “Throwing a dart”

Consider the closed unit circle of radius r, i.e., $S = \{(x, y) : x^2 + y^2 \leq r^2\}$. 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, is equal to $\frac{\text{A}}{\text{Area of whole circle}}$.
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 $Var(X)$?

Consider the closed unit circle of radius r, i.e., $S = \{(x, y) : x^2 + y^2 \leq r^2\}$. 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, is equal to $\frac{\text{A}}{\text{Area of whole circle}}$.
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 $Var(X)$?

Since $F_X(x)$ is the probability that the dart lands inside the circle of radius x, that probability is the area of a circle of radius x divided by the area of the circle of radius r (i.e., $\pi x^2/\pi r^2$). Thus, our CDF looks like

Consider the closed unit circle of radius r, i.e., $S = \{(x, y) : x^2 + y^2 \leq r^2\}$. 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, is equal to $\frac{\text{A}}{\text{Area of whole circle}}$.
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 $Var(X)$?

Since $F_X(x)$ is the probability that the dart lands inside the circle of radius x, that probability is the area of a circle of radius x divided by the area of the circle of radius r (i.e., $\pi x^2/\pi r^2$). Thus, our CDF looks like

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{r^2} & 0 < x \leq r \\ 1 & x > r \end{cases}$$

Consider the closed unit circle of radius r, i.e., $S = \{(x, y) : x^2 + y^2 \leq r^2\}$. 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, is equal to $\frac{\text{A}}{\text{Area of whole circle}}$.
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 $Var(X)$?

Since $F_X(x)$ is the probability that the dart lands inside the circle of radius x, that probability is the area of a circle of radius x divided by the area of the circle of radius r (i.e., $\pi x^2/\pi r^2$). Thus, our CDF looks like

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{r^2} & 0 < x \leq r \\ 1 & x > r \end{cases}$$

To find the PDF we just need to take the derivative of the CDF, which give us the following:

Consider the closed unit circle of radius r, i.e., $S = \{(x, y) : x^2 + y^2 \leq r^2\}$. 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, is equal to $\frac{\text{A}}{\text{Area of whole circle}}$.
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 $Var(X)$?

Since $F_X(x)$ is the probability that the dart lands inside the circle of radius x, that probability is the area of a circle of radius x divided by the area of the circle of radius r (i.e., $\pi x^2/\pi r^2$). Thus, our CDF looks like

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{r^2} & 0 < x \leq r \\ 1 & x > r \end{cases}$$

To find the PDF we just need to take the derivative of the CDF, which give us the following:

$$f_X(x) = \begin{cases} \frac{2x}{r^2} & 0 < x \leq r \\ 0 & \text{otherwise} \end{cases}$$

Consider the closed unit circle of radius r, i.e., $S = \{(x, y) : x^2 + y^2 \leq r^2\}$. 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, is equal to $\frac{\text{A}}{\text{Area of whole circle}}$.
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 $Var(X)$?

Using the definition of expectation we get

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^r x \frac{2x}{r^2} dx = \frac{2}{3r^2} \left(x^3 \Big|_0^r \right) = \frac{2}{3} r$$

Consider the closed unit circle of radius r, i.e., $S = \{(x, y) : x^2 + y^2 \leq r^2\}$. 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, is equal to $\frac{\text{A}}{\text{Area of whole circle}}$.
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 $Var(X)$?

Using the definition of expectation we get

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^r x \frac{2x}{r^2} dx = \frac{2}{3r^2} \left(x^3 \Big|_0^r \right) = \frac{2}{3} r$$

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

Consider the closed unit circle of radius r, i.e., $S = \{(x,y) : x^2 + y^2 \leq r^2\}$. 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, is equal to $\frac{\text{A}}{\text{Area of whole circle}}$.
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 $Var(X)$?

Using the definition of expectation we get

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^r x \frac{2x}{r^2} dx = \frac{2}{3r^2}\left(x^3\Big|_0^r\right) = \frac{2}{3}r$$

We know that $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^r x^2 \frac{2x}{r^2} dx = \frac{2}{4r^2}\left(x^4\Big|_0^r\right) = \frac{1}{2}r^2$$

Consider the closed unit circle of radius r, i.e., $S = \{(x,y) : x^2 + y^2 \leq r^2\}$. 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, is equal to $\frac{\text{A}}{\text{Area of whole circle}}$.
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 $Var(X)$?

Using the definition of expectation we get

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^r x \frac{2x}{r^2} dx = \frac{2}{3r^2}\left(x^3\Big|_0^r\right) = \frac{2}{3}r$$

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

$$\mathbb{E}\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^r x^2 \frac{2x}{r^2} dx = \frac{2}{4r^2}\left(x^4\Big|_0^r\right) = \frac{1}{2}r^2$$

Plugging this into our variance equation gives

$$Var(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \frac{1}{2}r^2 - \left(\frac{2}{3}r\right)^2 = \frac{1}{18}r^2$$

Midterm Review Session

- We are happy to answer any questions regarding the upcoming Midterm!

Question 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?

Question 2: “Will the battery last?”

Let N be a r.v. denoting the number of miles until the battery wears out. Then $N \sim$ exp$(10,000^{-1})$, because N measures the "time" (in this case miles) before an occurrence (the battery wears out) with expectation 10,000.

Question 2: “Will the battery last?”

Let N be a r.v. denoting the number of miles until the battery wears out. Then $N \sim \exp(10,000^{-1})$, because N measures the "time" (in this case miles) before an occurrence (the battery wears out) with expectation 10,000. Since this is an exponential distribution, and the expectation of an exponential distribution is $\frac{1}{\lambda}$, $\lambda = \frac{1}{10,000}$. Therefore, via the property of memorylessness of the exponential distribution:

Question 2: "Will the battery last?"

Let N be a r.v. denoting the number of miles until the battery wears out. Then $N \sim \exp(10,000^{-1})$, because N measures the "time" (in this case miles) before an occurrence (the battery wears out) with expectation 10,000. Since this is an exponential distribution, and the expectation of an exponential distribution is $\frac{1}{\lambda}$, $\lambda = \frac{1}{10,000}$. Therefore, via the property of memorylessness of the exponential distribution:

$$\mathbb{P}(N \geqslant 5000 \mid N \geqslant 2000) = \mathbb{P}(N \geqslant 3000) = 1 - \mathbb{P}(N \leqslant 3000) = 1 - \left(1 - e^{-\frac{3000}{10000}}\right) \approx 0.741$$