Lisa Yan and Jerry Cain
CS109

Lecture Notes #9
October 2, 2020

Continuous Distributions

Based on a chapter by Chris Piech

Pre-recorded lecture: All sections (definitions only) except Section 2.
In-lecture: Section 2, Section 4's Riding the Bus example, Section 5's Earthquakes example.
Extra: Section 6 (Proofs), Section 5's Website visits and Laptop life examples.

So far, all random variables we have seen have been *discrete*. In all the cases we have seen in CS 109, this meant that our RVs could only take on integer values. Now it's time for *continuous random variables*, which can take on values in the real number domain ($\mathbb{R}$). Continuous random variables can be used to represent measurements with arbitrary precision (e.g., height, weight, or time).

1 Probability Density Functions

In the world of discrete random variables, the most important property of a random variable was its probability mass function (PMF), which told you the probability of the random variable taking on a certain value. When we move to the world of continuous random variables, we are going to need to rethink this basic concept. If I were to ask you what the probability is of a child being born with a weight of **exactly** 3.523112342234 kilograms, you might recognize that question as ridiculous. No child will have precisely that weight. Real values are defined with infinite precision; as a result, the probability that a random variable takes on a specific value is not very meaningful when the random variable is continuous. The PMF doesn't apply. We need another idea.

In the continuous world, every random variable has a *probability density function* (PDF), which says how likely it is that a random variable takes on a particular value, relative to other values that it could take on. The PDF has the nice property that you can integrate over it to find the probability that the random variable takes on values within a range (a, b).

X is a **continuous random variable** if there is a function $f(x)$ for $-\infty \le x \le \infty$, called the **probability density function** (PDF), such that:

$$P(a \le X \le b) = \int_a^b f(x)dx$$

To preserve the axioms that guarantee $P(a \le X \le b)$ is a probability, the following properties must also hold:

$$0 \le P(a \le X \le b) \le 1$$
$$P(-\infty < X < \infty) = 1$$

A common misconception is to think of $f(x)$ as a probability. It is instead what we call a probability density. It represents probability *divided by the units of X*. Generally this is only meaningful when we either take an integral over the PDF **or** we *compare* probability densities. As we mentioned when motivating probability densities, the probability that a continuous random variable takes on a specific value (to infinite precision) is 0.

$$P(X = a) = \int_a^a f(x)dx = 0$$

This is very different from the discrete setting, in which we often talked about the probability of a random variable taking on a particular value exactly.

2 Cumulative Distribution Function

Having a probability density is great, but it means we are going to have to solve an integral every single time we want to calculate a probability. To save ourselves some effort, for most of these variables we will also compute a *cumulative distribution function* (CDF). The CDF is a function which takes in a number and returns the probability that a random variable takes on a value *less than (or equal to)* that number. If we have a CDF for a random variable, we don't need to integrate to answer probability questions!

For a continuous random variable X, the **cumulative distribution function** is:

$$F_X(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

This can be written $F(a)$, without the subscript, when it is obvious which random variable we are using.

Why is the CDF the probability that a random variable takes on a value *less than* (or equal to) the input value as opposed to greater than? It is a matter of convention. But it is a useful convention. Most probability questions can be solved simply by knowing the CDF (and taking advantage of the fact that the integral over the range $-\infty$ to ∞ is 1). Here are a few examples of how you can answer probability questions by just using a CDF:

Probability Query	**Solution**	**Explanation**
$P(X \leq a)$	$F(a)$	This is the definition of the CDF
$P(X < a)$	$F(a)$	Note that $P(X = a) = 0$
$P(X > a)$	$1 - F(a)$	$P(X \leq a) + P(X > a) = 1$
$P(a < X < b)$	$F(b) - F(a)$	$F(a) + P(a < X < b) = F(b)$

As we mentioned briefly earlier, the cumulative distribution function can also be defined for discrete random variables, but there is less utility to a CDF in the discrete world, because with the exception of the geometric random variable, none of our discrete random variables had "closed form" (that is, without any summations) functions for the CDF:

$$F_X(a) = \sum_{i=0}^{a} P(X = i)$$

Example: PDF

Let X be a continuous random variable with PDF:

$$f(x) = \begin{cases} C(4x - 2x^2) & \text{when } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

In this function, C is a constant. What value is C? Since we know that the PDF must sum to 1:

$$\int_0^2 C(4x - 2x^2)dx = 1$$
$$C\left(2x^2 - \frac{2x^3}{3}\right)\bigg|_{x=0}^{2} = 1$$
$$C\left(\left(8 - \frac{16}{3}\right) - 0\right) = 1$$

Solving this equation for C gives $C = 3/8$.

What is $P(X > 1)$?

$$\int_1^\infty f(x)dx = \int_1^2 \frac{3}{8}(4x - 2x^2)dx = \frac{3}{8}\left(2x^2 - \frac{2x^3}{3}\right)\bigg|_{x=1}^{2} = \frac{3}{8}\left[\left(8 - \frac{16}{3}\right) - \left(2 - \frac{2}{3}\right)\right] = \frac{1}{2}$$

Example: Disk crashes

Let X be a RV representing the number of days of use before your disk crashes, with PDF:

$$f(x) = \begin{cases} \lambda e^{-x/100} & \text{when } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

First, determine λ. Recall that $\int Ae^{Au}du = e^{Au}$:

$$\int_0^\infty \lambda e^{-x/100}dx = 1$$
$$-100\lambda \int_0^\infty \frac{-1}{100} e^{-x/100}dx = 1$$
$$-100\lambda \cdot e^{-x/100}\bigg|_{x=0}^{\infty} = 1$$
$$100\lambda \cdot 1 = 1 \quad \Rightarrow \quad \lambda = 1/100$$

What is $P(X < 10)$?

$$F(10) = \int_0^{10} \frac{1}{100} e^{-x/100} dx = -e^{-x/100}\Big|_{x=0}^{10} = -e^{-1/10} + 1 \approx 0.095$$

3 Expectation and Variance

For continuous RV X:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx$$
$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x) dx$$
$$E[X^n] = \int_{-\infty}^{\infty} x^n \cdot f(x) dx$$

For both continuous and discrete RVs:

$$E[aX + b] = aE[X] + b$$
$$\text{Var}(X) = E[(X - \mu)^2] = E[X^2] - (E[X])^2 \qquad \text{(with } \mu = E[X]\text{)}$$
$$\text{Var}(aX + b) = a^2 \, \text{Var}(X)$$

4 Uniform Random Variable

The most basic of all the continuous random variables is the uniform random variable, which is equally likely to take on any value in its range (α, β).

X is a **uniform random variable** ($X \sim \text{Uni}(\alpha, \beta)$) if it has PDF:

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{when } \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

Notice how the density $1/(\beta - \alpha)$ is exactly the same regardless of the value for x. That makes the density uniform. So why is the PDF $1/(\beta - \alpha)$ and not 1? That is the constant that makes it such that the integral over all possible inputs evaluates to 1.

The key properties of this RV are:

$$P(a \leq X \leq b) = \int_a^b f(x) dx = \frac{b - a}{\beta - \alpha} \text{ (for } \alpha \leq a \leq b \leq \beta\text{)}$$
$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_\alpha^\beta \frac{x}{\beta - \alpha} dx = \frac{x^2}{2(\beta - \alpha)}\Big|_{x=\alpha}^{\beta} = \frac{\alpha + \beta}{2}$$
$$\text{Var}(X) = \frac{(\beta - \alpha)^2}{12}$$

Example: Riding the Bus

You want to get on the bus. The bus stops in front of your building at 15-minute intervals (2:00, 2:15, etc.). Suppose you arrive at the stop uniformly between 2:00pm and 2:30pm. What is the probability that you wait less than 5 minutes for the bus?

Solution : Let X be the minutes (continuous time) after 2:00pm that we arrive at the bus. $X \sim \text{Uni}(0, 30)$. There are two events to consider: Either our arrival time is in the interval $10 < X \leq 15$, where we take the 2:15pm bus, or our arrival time is in the interval $25 < X \leq 30$, where we take the 2:30pm bus. Integrating the PDF of the Uniform RV, we get:

$$P(10 < X \leq 15) + P(25 < X \leq 30) = \int_{10}^{15} \frac{1}{30} dx + \int_{25}^{30} \frac{1}{30} dx = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}$$

5 Exponential Random Variable

An **exponential random variable** ($X \sim \text{Exp}(\lambda)$) represents the time until an event occurs. It is parametrized by $\lambda > 0$, the (constant) rate at which the event occurs. This is the same λ as in the Poisson distribution; a Poisson variable counts the number of events that occur in a fixed interval, while an exponential variable measures the amount of time until the next event occurs.

(Example 2 sneakily introduced you to the exponential distribution already; now we get to use formulas we've already computed to work with it without integrating anything.)

Properties

The probability density function (PDF) for an exponential random variable is:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases}$$

The expectation is $E[X] = \frac{1}{\lambda}$ and the variance is $\text{Var}(X) = \frac{1}{\lambda^2}$.

There is a closed form for the cumulative distribution function (CDF):

$$F(x) = 1 - e^{-\lambda x} \text{ where } x \geq 0$$

Example: Earthquakes

Major earthquakes (magnitude 8.0+) independently occur on average once every 500 years.

1. What is the probability of a major earthquake happening in the next 30 years?

   Solution: Define X to be the amount of time until the next earthquake happens. $X \sim \text{Exp}(\lambda = 1/500 = 0.002)$, since $E[X] = 1/500$. We can compute $P(X < 30) = \int_0^{30} 0.002 e^{-0.002x} dx = 0.002 \left[\frac{1}{0.002} e^{-0.002x}\right]_0^{30} = -(e^{-0.06} - e^0) \approx 0.058$.

2. What is the standard deviation of years until the next earthquake?

Solution: Using the same definition of X as before, $\text{Var}(X) = \frac{1}{\lambda^2} = \frac{1}{0.002^2} = 250,000$ years^2, and therefore $\text{SD}(X) = \sqrt{\text{Var}(X)} = 500$ years.

3. What is the probability of zero major earthquakes next year?

   Solution 1: Using the same definition of X as before, we would like to compute $P(X > 1) = 1 - F(1)$, where $F(x)$ is the CDF of X at x. $P(X > 1) = 1 - (1 - e^{-\lambda \cdot 1}) = e^{-\lambda} \approx 0.998$.

   Solution 2: Given that earthquake occurrences are independent, we could also define a random variable N to be the number of earthquakes next year, where $N \sim \text{Poi}(\lambda = 0.002)$ and $E[N] = \lambda = 1/500$. We then compute $P(N = 0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda} \approx 0.998$.

Example: Website visits

Let X be a random variable that represents the number of minutes until a visitor leaves your website. You have calculated that on average a visitor leaves your site after 5 minutes, and you decide that an exponential distribution is appropriate to model how long a person stays before leaving the site. What is the $P(X > 10)$?

We can compute $\lambda = \frac{1}{5}$ either using the definition of $E[X]$ or by thinking of how many people leave every minute (answer: "one-fifth of a person"). Thus $X \sim \text{Exp}(1/5)$.

$$\begin{aligned} P(X > 10) &= 1 - F(10) \\ &= 1 - (1 - e^{-\lambda \cdot 10}) \\ &= e^{-2} \approx 0.1353 \end{aligned}$$

Example: Laptop life

Let X be the number of hours of use until your laptop dies. On average laptops die after 5000 hours of use. If you use your laptop for 7300 hours during your undergraduate career (assuming usage = 5 hours/day and four years of university), what is the probability that your laptop lasts all four years?

As above, we can find λ either using $E[X]$ or thinking about laptop deaths per hour: $X \sim \text{Exp}(\frac{1}{5000})$.

$$\begin{aligned} P(X > 7300) &= 1 - F(7300) \\ &= 1 - (1 - e^{-7300/5000}) \\ &= e^{-1.46} \approx 0.2322 \end{aligned}$$

6 Proofs

6.1 Expectation of the Exponential RV

Let $X \sim \text{Exp}(\lambda)$. Then $E[X] = 1/\lambda$.

Proof: We would like to evaluate $E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x\lambda e^{-\lambda x} dx$.

Let $u = x$ and $dv = \lambda e^{-\lambda x} dx$. Then $du = dx$ and $v = -e^{-\lambda x}$. Recall that integration by parts states that $\int u \cdot dv = u \cdot v - \int v \cdot du$, and therefore $\int x\lambda e^{-\lambda x} dx = -xe^{-\lambda x} - \int (-e^{-\lambda x}) dx$.

$$E[X] = \int_0^{\infty} x\lambda e^{-\lambda x} dx = -xe^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx$$
$$= -xe^{-\lambda x}\Big|_0^{\infty} - \frac{1}{\lambda} e^{-\lambda x}\Big|_0^{\infty} = [0 - 0] + \left[0 - \left(\frac{-1}{\lambda}\right)\right] = \frac{1}{\lambda}$$

6.2 CDF of the Exponential RV

Let $X \sim \text{Exp}(\lambda)$. The CDF of X is $F(x) = 1 - e^{-\lambda x}$, for $x \geq 0$.

Proof:

$$F(x) = P(X \leq x) = \int_{y=-\infty}^{x} f(y) dy = \int_{y=0}^{x} \lambda e^{-\lambda y} dy$$
$$= \lambda \frac{1}{-\lambda} e^{-\lambda y}\Big|_0^{x} = -1(e^{-\lambda x} - e^{-\lambda 0}) = 1 - e^{-\lambda x}.$$