Noah Arthurs
CS 109

Lecture #9
July 10, 2019

Continuous Distributions

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 From Discrete to Continuous

To make our transition from thinking about discrete random variable, to thinking about continuous random variables, lets start with a thought experiment: Imagine you are running to catch the bus. You know that you will arrive at 2:15pm but you don't know exactly when the bus will arrive, and want to think of the arrival time in minutes past 2pm as a random variable T so that you can calculate the probability that you will have to wait more than five minutes $P(15 < T < 20)$.

We immediately face a problem. For discrete distributions we would describe the probability that a random variable takes on exact values. This doesn't make sense for continuous values, like the time the bus arrives. As an example, what is the probability that the bus arrives at exactly 2:17pm and 12.12333911102389234 seconds? Similarly, if I were to ask you: what is the probability of a child being born with weight **exactly** = 3.523112342234 kilos, you might recognize that question as ridiculous. No child will have precisely that weight. Real values can have infinite precision and as such it is a bit mind boggling to think about the probability that a random variable takes on a specific value.

Instead, let's start by discretizing time, our continuous variable, by breaking it into 5 minute chunks. We can now think about something like, the probability that the bus arrives between 2:00p and 2:05 as an event with some probability (see figure 1, left). Five minute chunks seem a bit coarse. You could imagine that instead, we could have discretized time into 2.5minute chunks (figure 1, center). In this case the probability that the bus shows up between 15 mins and 20 mins after 2pm is the sum of two chunks, shown in orange. Why stop there? In the limit we could keep breaking time down into smaller and smaller pieces. Eventually we will be left with a derivative of probability at each moment of time, where the probability that $P(15 < T < 20)$ is the integral of that derivative between 15 and 20 (figure 1, right).

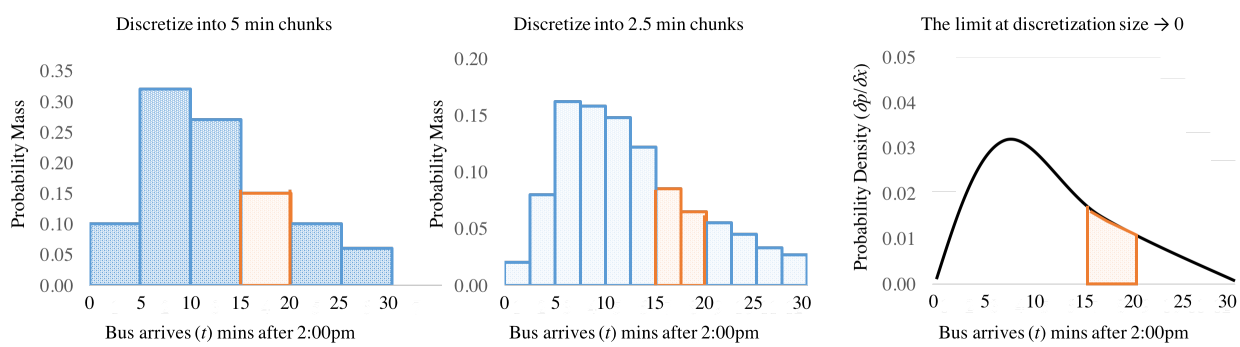

Figure 1: Bus thought process: Discrete to Continuous

2 Probability Density Functions

In the world of discrete random variables, the most important property of a random variable was its probability mass function (PMF) that would tell you the probability of the random variable taking on any value. When we move to the world of continuous random variables, we are going to need to rethink this basic concept.

In the continuous world, every random variable instead has a Probability *Density* Function (PDF) which defines the relative likelihood that a random variable takes on a particular value. Like in the bus example, the PDF is the derivative of probability at all points of the random variable. This means that the PDF has the important 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 $-\infty \leq x \leq \infty$, called the **probability density function** (PDF), such that:

$$P(a \leq X \leq b) = \int_a^b dx\, f_X(x)$$

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

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

A common misconception is to think of the PDF function $f_X(x)$ 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.

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

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.

3 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:

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

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 1

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:

$$\begin{aligned} \int_0^2 dx\, C(4x - 2x^2) &= 1 \\ C\left(2x^2 - \frac{2x^3}{3}\right)\Big|_{x=0}^{2} &= 1 \\ C\left(\left(8 - \frac{16}{3}\right) - 0\right) &= 1 \end{aligned}$$

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

What is $P(X > 1)$?

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

Example 2

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 $\int Ae^{Au} du = e^{Au}$:

$$\begin{aligned} \int_0^\infty dx\, \lambda e^{-x/100} &= 1 \\ -100\lambda \int_0^\infty dx\, \frac{-1}{100} e^{-x/100} &= 1 \\ -100\lambda \cdot e^{-x/100}\Big|_{x=0}^{\infty} &= 1 \\ 100\lambda \cdot 1 &= 1 \quad \Rightarrow \quad \lambda = 1/100 \end{aligned}$$

What is $P(X < 10)$?

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

4 Expectation and Variance

For continuous RV X:

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

For both continuous and discrete RVs:

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

5 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:

$$\begin{aligned}
P(a \leq X \leq b) &= \int_a^b dx\, f(x) = \frac{b-a}{\beta-\alpha} \text{ (for } \alpha \leq a \leq b \leq \beta) \\
E[X] &= \int_{-\infty}^{\infty} dx\, x \cdot f(x) = \int_\alpha^\beta dx\, \frac{x}{\beta-\alpha} = \left.\frac{x^2}{2(\beta-\alpha)}\right|_{x=\alpha}^{\beta} = \frac{\alpha+\beta}{2} \\
\text{Var}(X) &= \frac{(\beta-\alpha)^2}{12}
\end{aligned}$$

6 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 3

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 4

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: $X \sim \text{Exp}(\frac{1}{5000})$.

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