6. Continous Distributions

Chris Piech and Mehran Sahami

May 2017

So far, all random variables we have seen have been *discrete*. In all the cases we have seen in CS109 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. They usually represent measurements with arbitrary precision (eg height, weight, time).

1 Probability Density Functions

X is a Continuous Random Variable if there is a Probability Density Function (PDF) $f(x)$ for $-\infty \leq x \leq \infty$ such that:

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

The following properties must also hold. These preserve the axiom that $P(a \leq X \leq b)$ is a probability:

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

A common misconception is to think of $f(x)$ as a probability. It is instead what we call a probability density. It represents probability/unit of X. Generally this is not particularly meaningful without either taking the interval over X or comparing it to another probability density. Of special note, the probability that a continuous random variable takes on a specific value (to infinite precision) is 0.

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

That is pretty different than in the discrete world where we often talked about the probability of a random variable taking on a particular value.

2 Cumulative Distribution Function

For a continuous random variable X the Cumulative Distribution Function, written $F(a)$ or as (CDF) is:

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

Example 1

Let X be a continuous random variable (CRV) 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:

$$\int_0^2 C(4x - 2x^2)dx = 1$$

$$C\left(2x^2 - \frac{2x^3}{3}\right)\Bigg|_0^2 = 1$$

$$C\left(\left(8 - \frac{16}{3}\right) - 0\right) = 1$$

And if you solve the equation for C you find that $C = 3/8$.

What is $P(X > 1)$

$$\int_1^\infty f(x)dx = \int_1^2 \frac{3}{8}(4x - 2x^2)dx = \frac{3}{8}\left(2x^2 - \frac{2x^3}{3}\right)\Bigg|_1^2 = \frac{3}{8}\left[\left(8 - \frac{16}{3}\right) - \left(2 - \frac{2}{3}\right)\right] = \frac{1}{2}$$

Example 2

Let X be a random variable which represents 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 $\int e^u du = e^u$:

$$\int \lambda e^{-x/100} dx = 1 \Rightarrow -100\lambda \int \frac{-1}{100} e^{-x/100} dx = 1$$

$$-100\lambda e^{-x/100}\Bigg|_0^\infty = 1 \Rightarrow 100\lambda = 1 \Rightarrow \lambda = \frac{1}{100}$$

What is the $P(X < 10)$?

$$F(10) = \int_0^{10} \frac{1}{100} e^{-x/100} dx = -e^{-x/100}\Bigg|_0^{10} = -e^{-1/10} + 1 \approx 0.095$$

3 Expectation and Variance

For continuous RV X:

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

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

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

For both continuous and discrete RVs:

$$E[aX + b] = aE[X] + b$$

$$Var(X) = E[(X - \mu)^2] = E[X^2] - (E[X])^2$$

4 Uniform Random Variable

X is a Uniform Random Variable $X \sim Uni(\alpha, \beta)$ if:

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

The key properties of this RV are:

$$P(a \le X \le b) = \int_a^b f(x)dx = \frac{b-a}{\beta-\alpha} \text{ (for } \alpha \le a \le b \le \beta)$$

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_\alpha^\beta \frac{x}{\beta-\alpha}dx = \left.\frac{x^2}{2(\beta-\alpha)}\right|_\alpha^\beta = \frac{\alpha+\beta}{2}$$

$$Var(X) = \frac{(\beta-\alpha)^2}{12}$$

5 Normal Random Variable

The single most important random variable type is the Normal (aka Gaussian) random variable, parametrized by a mean (μ) and variance (σ^2). If X is a normal variable we write $X \sim \mathcal{N}(\mu, \sigma^2)$. The normal is important for many reasons: it is generated from the summation of independent random variables and as a result it occurs often in nature. Many things in the world are not distributed normally but data scientists and computer scientists model them as Normal distributions anyways. Why? Because it is the most entropic (conservative) distribution that we can apply to data with a measured mean and variance.

Properties

The Probability Density Function (PDF) for a Normal is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

By definition a Normal has $E[X] = \mu$ and $Var(X) = \sigma^2$.

If X is a Normal such that $X \sim \mathcal{N}(\mu, \sigma^2)$ and Y is a linear transform of X such that $Y = aX + b$ then Y is also a Normal where $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

There is no closed form for the integral of the Normal PDF, however since a linear transform of a Normal produces another Normal we can always map our distribution to the “Standard Normal” (mean 0 and variance 1) which has a precomputed Cumulative Distribution Function (CDF). The CDF of an arbitrary normal is:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Where Φ is a precomputed function that represents that CDF of the Standard Normal.

Projection to Standard Normal

For any Normal X we can define a random variable $Z \sim \mathcal{N}(0,1)$ to be a linear transform

$$\begin{aligned} Z &= \frac{X-\mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma} \\ &\sim \mathcal{N}\left(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{\sigma^2}{\sigma^2}\right) \\ &\sim \mathcal{N}(0,1) \end{aligned}$$

Using this transform we can express $F_X(x)$, the CDF of X, in terms of the known CDF of Z, $F_Z(x)$. Since the CDF of Z is so common it gets its own Greek symbol: $\Phi(x)$

$$\begin{aligned} F_X(x) &= P(X \le x) \\ &= P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) \\ &= P\left(Z \le \frac{x-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

The values of $\Phi(x)$ can be looked up in a table. We also have an online calculator.

Example 1

Let $X \sim \mathcal{N}(3,16)$, what is $P(X > 0)$?

$$\begin{aligned} P(X > 0) &= P\left(\frac{X-3}{4} > \frac{0-3}{4}\right) = P\left(Z > -\frac{3}{4}\right) = 1 - P\left(Z \le -\frac{3}{4}\right) \\ &= 1 - \Phi(-\frac{3}{4}) = 1 - (1 - \Phi(\frac{3}{4})) = \Phi(\frac{3}{4}) = 0.7734 \end{aligned}$$

What is $P(2 < X < 5)$?

$$\begin{aligned} P(2 < X < 5) &= P\left(\frac{2-3}{4} < \frac{X-3}{4} < \frac{5-3}{4}\right) = P\left(-\frac{1}{4} < Z < \frac{2}{4}\right) \\ &= \Phi(\frac{2}{4}) - \Phi(-\frac{1}{4}) = \Phi(\frac{1}{2}) - (1 - \Phi(\frac{1}{4})) = 0.2902 \end{aligned}$$

Example 2

You send voltage of 2 or -2 on a wire to denote 1 or 0. Let X = voltage sent and let R = voltage received. $R = X + Y$, where $Y \sim \mathcal{N}(0,1)$ is noise. When decoding, if $R \ge 0.5$ we interpret the voltage as 1, else 0. What is $P(\text{error after decoding}|\text{original bit} = 1)$?

$$P(X+Y < 0.5) == P(2+Y < 0.5) = P(Y < -1.5) = \Phi(-1.5) = 1 - \Phi(1.5) \approx 0.0668$$

Binomial Approximation

You can use a Normal distribution to approximate a Binomial $X \sim Bin(n,p)$. To do so define a normal $Y \sim (E[X], Var(X))$. Using the Binomial formulas for expectation and variance, $Y \sim (np, np(1-p))$. This approximation holds for large n. Since a Normal is continuous and Binomial is discrete we have to use a continuity correction to discretize the Normal.

$$P(X = k) \sim P\left(k - \frac{1}{2} < Y < k + \frac{1}{2}\right) = \Phi\left(\frac{k - np + 0.5}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - np - 0.5}{\sqrt{np(1-p)}}\right)$$

Example 3

100 visitors to your website are given a new design. Let X = # of people who were given the new design and spend more time on your website. Your CEO will endorse the new design if $X \ge 65$. What is $P(\text{CEO endorses change}|\text{it has no effect})$?

$E[X] = np = 50$. $Var(X) = np(1-p) = 25$. $\sigma = \sqrt{(Var(X))} = 5$. We can thus use a Normal approximation: $Y \sim \mathcal{N}(50, 25)$.

$$P(X \geq 65) \approx P(Y > 64.5) = P\left(\frac{Y-50}{5} > \frac{64.5-50}{5}\right) = 1 - \Phi(2.9) = 0.0019$$

Example 4

Stanford accepts 2480 students and each student has a 68% chance of attending. Let X = # students who will attend. $X \sim Bin(2480, 0.68)$. What is $P(X > 1745)$?

$E[X] = np = 1686.4$. $Var(X) = np(1-p) = 539.7$. $\sigma = \sqrt{(Var(X))} = 23.23$. We can thus use a Normal approximation: $Y \sim \mathcal{N}(1686.4, 539.7)$.

$$P(X > 1745) \approx P(Y > 1745.5) = P\left(\frac{Y-1686.4}{23.23} > \frac{1745.5-1686.4}{23.23}\right) = 1 - \Phi(2.54) = 0.0055$$

6 Exponential Random Variable

An Exponential Random Variable $X \sim Exp(\lambda)$ represents the time until an event occurs. It is parametrized by $\lambda > 0$, the rate at which the event occurs. This is the same λ as in the Poisson distribution.

Properties

The Probability Density Function (PDF) for an Exponential 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 $Var(X) = \frac{1}{\lambda^2}$

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

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

Example 1

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 function is appropriate to model how long until a person leaves your 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 much of a person leaves every minute (one fifth of a person). Thus $X \sim Exp(1/5)$.

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

Example 2

Let X be the # 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?

We can compute $\lambda = \frac{1}{5000}$ either using the definition of $E[X]$. Thus $X \sim Exp(1/5000)$.

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