CS 547 Lecture 8: Continuous Random Variables

Daniel Myers

The Probability Density Function

The distribution of a continuous random variable is given by its *probability density function* (pdf), denoted $f(x)$. Questions about the behavior of a continuous RV can be answered by integrating over the pdf. For example,

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

All of the results we previously showed for discrete random variables also apply to continuous random variables, with necessary change of the discrete summation to a continuous integration. For example, total probabilty guarantees that the integral of $f(x)$ taken over the entire range of random variable will equal 1:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Here's a somewhat surprising result. Suppose we wish to evaluate $P(X = a)$. Using the pdf, we have,

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

Thus, the probability of a continuous random variable taking on any particular fixed value is zero (think of this as the probability of picking one distinct outcome out of an infinite number of possible outcomes). This implies that $P(X \leq a) = P(X < a)$, which does seem a bit odd.

The Cumulative Distribution Function

The *cumulative distribution function* (CDF) is defined as

$$F(x) = P(X \leq x).$$

For a continuous random variable, $F(x)$ can be found by integrating over the pdf for all values less than x:

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

Given the CDF, we can derive the pdf by differentiating,

$$f(x) = \frac{\partial}{\partial x} F(x).$$

The complementary cumulative distribution function (CCDF) of X is defined to be

$$\overline{F}(x) = P(X > x) = 1 - F(x) = \int_x^{\infty} f(x)\,dx$$

Thus, we can obtain the CCDF from either the CDF or the pdf[1].

[1] I follow Kleinrock's practice of capitalizing the abbreviation for the cumulative distribution but not the density function. I think it helps keep the two similar abbreviations conceptually distinct.

Expected Value and Variance

The definitions of expected value and variance apply to continuous random variables, with the appropriate conversion of summation to integration:

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

$$\sigma^2 = E[X^2] - E[X]^2$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2\, f(x)\, dx$$

All of the properties of the expected value that we previously showed are still valid for the continous case.

The Exponential Random Variable

The *exponential random variable* is the most important continuous random variable in queueing theory. Its pdf has a single parameter, $\lambda > 0$, and is given by

$$f(x) = \lambda e^{-\lambda x} \quad \text{if } x \geq 0$$

and $f(x) = 0$ if $x < 0$.

It is easy to verify that

$$\int_0^{\infty} \lambda e^{-\lambda x}\, dx = 1,$$

as it must by total probability.

We can derive the cumulative distribution $F(a)$ by integrating from 0 to a,

$$\begin{aligned} F(a) &= P(X \leq a) \\ &= \int_0^a \lambda e^{\lambda x}\, dx \\ &= 1 - e^{-\lambda a} \end{aligned}$$

Differentiating $F(a)$ with respect to a will recover the original density function, as expected.

The expected value of the exponential random variable is given by

$$E[X] = \int_0^{\infty} x\, \lambda e^{-\lambda x}\, dx$$

Using integration by parts ($u = x$ and $dv = e^{-\lambda x}\, dx$) to evaluate the integral, we obtain

$$E[X] = \frac{1}{\lambda}$$

The expected value of an exponentially distributed random variable is simply the inverse of the paramter λ.

It can be shown, using integration by parts twice, that

$$E[X^2] = \frac{2}{\lambda^2}$$

The variance is therefore

$$\sigma^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

The squared coefficient of variation is simply $c_X^2 = 1$. The exponential distribution is the special case of a continuous random variable with standard deviation equal to its mean.

The Uniform Random Variable

A random variable that is uniformly distributed over the range $[a, b]$ and has pdf given by

$$f(x) = \frac{1}{b-a} \quad \text{if } a \le x \le b,$$

and $f(x) = 0$ otherwise.

It is straightforward to derive the expected value,

$$E[X] = \frac{b+a}{2},$$

which is simply the midpoint of the interval $[a, b]$.

Generating Random Variables

In simulation studies, we frequently need to generate random variables from arbitrary distributions. The *inverse CDF method* is the simplest means of doing this.

Recall that the CDF is defined to be $F(x) = P(X \le x)$, which is a number between 0 and 1. Therefore we can generate a random variable having the desired CDF using the following procedure:

- choose a random probability u between 0 and 1
- solve for the unique value of x such that $F(x) = u$

All modern programming languages support a means of generating pseudo-random values between 0 and 1, so we can use the built-in pseudo-RNG to pick a value of u, then calculate the corresponding x for the desired distribution.

For the exponential distribution, we want to find the value of x that satisfies

$$F(x) = 1 - e^{-\lambda x} = u$$

Solving, we obtain,

$$x = \frac{\ln(1-u)}{-\lambda}$$

The following MATLAB function will generate a single exponentially distributed random variable.

```
function [x] = generate_exponential(lambda)
  u = rand;
  x = -log(1 - u) / lambda;
end
```