Continuous Probability Continued

In the Lecture 11, we introduced continuous random variables. For example, let's consider $X \sim Unif[0,1]$ which is a random variable that takes on continuous values. Then,

1. $\mathbf{P}(X = a) = 0 \; \forall a$
2. $\mathbf{P}(a < X \leq b) = b - a$, for $0 \leq a \leq b \leq 1$

In general, the probability that the random variable belongs to an interval of length Δ can be approximated by the probability density function (pdf) $f(x)$:

$$\mathbf{P}(x < X \leq x + \Delta) \approx f(x)\Delta.$$

Recall that probability values don't have units. However, as can be seen from the previous approximation, the density of the random variable does have a unit. Δ has the same unit as the random variable itself. Therefore, the unit of f should be the inverse of the unit of Δ so that the probability is unitless. For example, the unit of f is m^{-1} if the random variable is measured in meters.

The cumulative distribution function of a continuous random variable was defined as follows:

Definition 13.1 (Cumulative Distribution Function): The *cumulative distribution function* for a random variable X is a function $F : \mathbb{R} \to \mathbb{R}$ defined to be:

$$F(x) = \mathbf{P}(X \leq x). \tag{1}$$

Its relationship with the probability density function f of X is given by:

$$f(x) = \frac{d}{dx}F(x), \qquad F(x) = \int_{-\infty}^{x} f(a)da.$$

Therefore the cdf and pdf contain the same information since they can be obtained from each other. Note that the cumulative distribution applies to both continuous and discrete random variables. That is because the cumulative distribution is a well defined probability.

Expectation and variance of a continuous random variable

By analogy with the discrete case, we define the expectation of a continuous r.v. as follows:

Definition 13.2 (Expectation): The expectation of a continuous random variable X with probability density function f is

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

Note that the integral plays the role of the summation in the discrete formula $\mathbb{E}[X] = \sum_a a\mathbf{P}(X = a)$. Since variance is really just another expectation, we can immediately port its definition to the continuous setting as well:

Definition 13.3 (Variance): The variance of a continuous random variable X with probability density function f is

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \left(\int_{-\infty}^{\infty} x f(x)dx\right)^2.$$

Two more important continuous distributions

We have already seen one important continuous distribution, namely the uniform distribution. In this section we will see two more: the *exponential* distribution and the *normal* (or *Gaussian*) distribution. These three distributions cover the vast majority of continuous random variables arising in applications.

Exponential distribution: The exponential distribution is a continuous version of the geometric distribution, which we have already met. Recall that the geometric distribution describes the number of tosses of a coin until the first Heads appears; the distribution has a single parameter p, which is the bias (Heads probability) of the coin. Of course, in real life applications we are usually not waiting for a coin to come up Heads but rather waiting for a system to fail, a clock to ring, an experiment to succeed etc.

In such applications we are frequently not dealing with discrete events or discrete time, but rather with *continuous* time: for example, if we are waiting for an apple to fall off a tree, it can do so at any time at all, not necessarily on the tick of a discrete clock. This situation is naturally modeled by the exponential distribution, defined as follows:

Definition 13.4 (Exponential distribution): For any $\lambda > 0$, a continuous random variable X with pdf f given by

$$f(x) = \begin{cases} Ae^{-\lambda x} & \text{if } x \geq 0; \\ 0 & \text{otherwise.} \end{cases}$$

is called an *exponential* random variable with parameter λ.

Since we know that the pdf has to integrate to one, we can evaluate A:

$$\int_0^{\infty} Ae^{-\lambda x} = 1 \implies A = \lambda.$$

By direct calculations, $\mathbb{E}[X] = 1/\lambda, \mathrm{Var}(T) = 1/\lambda^2$.

We now revisit the call arrivals example from lecture 9, where the arrival rate λ represents the average number of call arrivals per minute. Let X be the random variable that counts the number of calls initiated within a particular minute. We showed that X follows a Poisson distribution with parameter λ. Suppose we now ask a different question. Starting at $t = 0$, we are interested in the distribution of the time we have to wait for the next call. This is a continuous random variable T. What is its pdf?

Let us split time starting at time 0 into small intervals of length $\frac{1}{n}$, where X_i indicated whether or not there is an arrival in interval i. Within an interval, there are two possibilities:

$$X_i = \begin{cases} 1 & \text{with probability } \frac{\lambda}{n}; \\ 0 & \text{with probability } 1 - \frac{\lambda}{n}; \end{cases}$$

We are discretizing the continuous time into intervals. Moreover, we are assuming as before arrivals are independent from interval to interval. Thus, the number of intervals we have to wait until we have an arrival has a geometric distribution. Let S denote this random variable. Then $S \sim Geom(\frac{\lambda}{n})$. How is the continuous arrival time T related to S?

Now note that S is a discrete random variable while T is continuous. We will use the cdf to connect them because the cdf can be used for both continuous and discrete random variables. We start by computing the cdf of T:

$$\begin{aligned}
F_T(t) &= \mathbf{P}(T \leq t) \\
&\approx \mathbf{P}(T \leq \frac{k}{n}), \text{ where } k = [tn] \\
&= \mathbf{P}(S \leq k) \\
&= 1 - \mathbf{P}(S > k) \\
&= 1 - (1 - \frac{\lambda}{n})^k \\
&\approx 1 - (1 - \frac{\lambda}{n})^{nt}
\end{aligned}$$

The first approximation in the derivation above comes from the fact k is the interval where t falls in, and is accurate when n is large.

As $n \to \infty$,

$$1 - (1 - \frac{\lambda}{n})^{nt} \to 1 - e^{-\lambda t} \text{ with } t > 0,$$

and by differentiating the above cdf we get:

$$f(t) = \lambda e^{-\lambda t} \text{ for } t > 0.$$

Like the geometric, the exponential distribution has a single parameter λ. We have shown that the first arrival time has an exponential distribution. Like the geometric distribution, the exponential distribution is decreasing in t. What is the distribution of the time between the nth arrival and the $(n+1)$th arrival? It is also exponential with the same parameter. Note that the mean interarrival time is $1/\lambda$, which makes sense because λ is the arrival *rate*; the higher the arrival rate, the smaller the mean inter arrival time.

Suppose we are looking at the arrivals of buses, where the inter-arrival times are exponentially distributed with mean $\lambda = 1$ minute. We arrive at a time τ at the bus stop, and ask about the time the last bus left the stop which we call X. Then, we wait until the next bus arrives which occurs at time Y. We consider the gap between the arrivals $Y - X$, which is the observed inter-arrival time. How does the mean observed inter-arrival time compare to 1 minute, the expected inter arrival time?

The time between τ and the next arrival time Y is exponentially distributed with mean 1. The previous arrival time X also has a mean that is strictly positive. So the expected gap between the previous arrival and the next arrival is greater than one minute. Why? This is a surprising result which is due to conditioning. We are more likely to arrive in a longer gap between arrivals of buses than a smaller gap. So when we observe the inter-arrival times, the average inter-arrival time we observe is longer than a minute because of this conditioning.

If the buses arrived regularly every minute (i.e if the inter-arrival time is exactly one minute), then the answer would be one minute regardless of when we arrive.

Normal Distribution: The last continuous distribution we will look at, and by far the most prevalent in applications, is called the *normal* or *Gaussian* distribution. It has two parameters, μ and σ.

Definition 13.5 (Normal distribution): For any μ and $\sigma > 0$, a continuous random variable X with pdf f given by

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

is called a *normal* random variable with parameters μ and σ. In the special case $\mu = 0$ and $\sigma = 1$, X is said to have the *standard normal* distribution and the density becomes

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

A plot of the pdf f reveals a classical "bell-shaped" curve, centered at (and symmetric around) $x = \mu$, and with "width" determined by σ. (The precise meaning of this latter statement will become clear when we discuss the variance below.)

Let's run through the usual calculations for this distribution. We first check that the density integrates to 1:

$$\int_{-\infty}^{\infty} f(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx = 1. \tag{2}$$

The fact that this integral evaluates to 1 is a routine exercise in integral calculus, and is left as an exercise (or feel free to look it up in any standard book on probability or on the internet).

The normal distribution is ubiquitous throughout the sciences and the social sciences, because it is the standard model for any aggregate data that results from a large number of independent observations of the same random variable (such as the heights of females in the US population, or the observational error in a physical experiment). Such data, as is well known, tends to cluster around its mean in a "bell-shaped" curve, with the correspondence becoming more accurate as the number of observations increases. A theoretical explanation of this phenomenon is the Central Limit Theorem, which we next discuss.