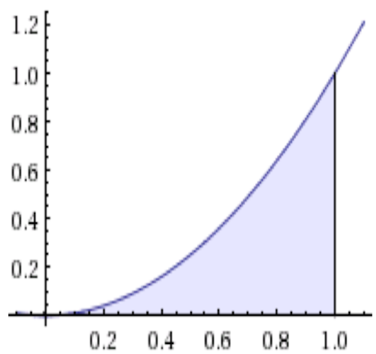
Continuous distributions
Math 217 Probability and Statistics
Prof. D. Joyce, Fall 2014

Back when we first looked at axioms for probability distributions, we looked as one example of a continuous distribution, namely a uniform distribution on an interval. We also looked at its cumulative distribution function and its probability density function. We'll review those concepts, but first let's look at sources of other continuous probabilities.

Monte Carlo estimates. Monte Carlo estimation is a method that uses random processes to make estimates when direct methods are too complicated or too time consuming to work. The examples we'll look at aren't complicated, but they will illustrate what the method is.

Our first example will use the Monte Carlo method to estimate a certain region inside the unit square in the (x, y)-plane. The *unit square* consists of points (x, y) where both x and y have values between 0 and 1. Consider the region R inside that unit square under the curve $y = x^2$.



With calculus, we can compute the area of that region exactly.

$$\int_0^1 x^2\,dx = \left.\frac{x^3}{3}\right|_0^1 = \frac{1}{3}.$$

For the experiment, choose a random point (X, Y) in the unit square. Here, we assume that both X and Y are uniform continuous distributions on the unit interval. Then we can ask, what is the probability that the random point (X, Y) lies in the region R?

We'll look at the Java applet that estimates this value by repeated simulation of the experiment. The greater the number of simulations, the closer the estimate is. One run of this computer experiment with 10,000 trials gave an estimate of 0.325. Different runs give different estimates.

Buffon's needle experiment. In 1836, long before computers, Buffon used a experiment something like the last example to estimate π. He suggested drawing parallel lines a unit distance apart on the the plane, then randomly dropping a needle of length 1 on the plane. He proposed that the probability that the needle intersects a line was $2/\pi$.

We'll look at the Java applet that simulates this experiment. The number of trials has to be very large to get a good estimate of π.

Here's an argument for why this Monte Carlo experiment should estimate π. Make the parallel lines be the horizontal lines $y = n$ for all integers n. Let d be the distance from the center of the pin to the nearest line. This d is actually a random variable uniformly distributed between 0 and $\frac{1}{2}$. Let θ be the angle of the needle; it's a uniform random variable between 0 and $\pi/2$ (if we measure angles in radians). Thus, the position of the needle is a point (θ, d) in the rectangle $[0, \frac{\pi}{2}] \times [0, \frac{1}{2}]$. Now, the needle will intersects one of the horizontal lines if $d/\sin\theta \leq \frac{1}{2}$, that is, if $d \leq \frac{1}{2}\sin\theta$.

So, the fraction of the rectangle $[0, \frac{\pi}{2}] \times [0, \frac{1}{2}]$ that lies below the curve $d/\sin\theta \leq \frac{1}{2}$ is the probability that the needle intersects one of the horizontal lines. The area of that region can be computed using integrals.

$$\int_0^{\pi/2} \tfrac{1}{2}\sin\theta\, d\theta = -\tfrac{1}{2}\cos\theta\Big|_0^{\pi/2} = -\tfrac{1}{2}\cos\tfrac{\pi}{2} + \tfrac{1}{2}\cos 0 = \tfrac{1}{2}$$

The area of the whole rectangle $[0, \frac{\pi}{2}] \times [0, \frac{1}{2}]$ is $\pi/4$, so the fraction of the rectangle that lies below the curve is $\dfrac{1/2}{\pi/4} = \dfrac{2}{\pi}$.

Of course, this experiment has been performed and the resulting experimental values are near $2/\pi$.

The Poisson process. Another source of continuous probabilities is what's called the Poisson process. It's the continuous version of the Bernoulli process. You can think of a Bernoulli process as giving events that occur at times t that are integers, where an event occurs when there is a success on the t^{th} Bernoulli trial. With that interpretation, we can interpret p as the expected number of events per unit time. We saw how expected time to the next event is $1/p$ when we evaluated the mean of the geometric distribution.

Now imagine t as a real variable. Events could occur at any time t, not just times that are whole numbers. We can still have an expected number of events per unit time, usually denoted λ, and it will turn out that the expected time to the next event will be $1/\lambda$.

An example of random occurrences over time is radioactive decay. Suppose we have a mass of some radioactive element. A Geiger counter can listen to decays of this substance, and it sounds like random clicks. Each occurrence is the decay of a one atom that can be detected by a Geiger counter. Different radioactive elements have different rates λ of decay.

It's surprising how many different phenomena are modeled by Poisson processes. Failures of light bulbs and failures of computer hard drives aren't exactly modeled, but they're pretty close. Arrivals of customers at various queues (like bank queues and supermarket queues) are closely modeled by Poisson processes.

Poisson studied events occurring in continuous time like this. He came up with some axioms to describe the situation and used analysis to determine the probability density functions for random variables associated to it. We'll look at this process later in detail, but for now, let's look at one particular random variable, T, the time to the next event. Its distribution is called the *exponential distribution*.

The exponential density function of a Poisson process with parameter λ is

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{for } t > 0 \\ 0 & \text{otherwise.} \end{cases}$$

We can find the corresponding cumulative distribution function by integration. For $t \le 0$, of course $F(t) = 0$, but for $t > 0$, we find

$$F(t) = \int_{-\infty}^{t} f(x)\,dx = \int_{0}^{t} \lambda e^{-\lambda x}\,dx = -\,e^{-\lambda x}\Big|_{0}^{t} = -\,e^{-\lambda t} + e^{-\lambda 0} = 1 - e^{-\lambda t}$$

With this c.d.f. we can answer various questions involving events in a Poisson process. For instance, the probability that the first event occurs before time 1 is

$$P(T \le 1) = F(1) = 1 - e^{-\lambda},$$

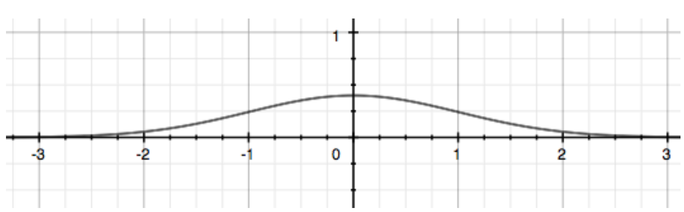
and the probability that the first event occurs between time 1 and time 2 is

$$P(T \in [1,2]) = F(2) - F(1) = (1 - e^{-2\lambda}) - (1 - e^{-\lambda}) = e^{-\lambda} - e^{-2\lambda}.$$

The normal distribution. By far, the most important continuous distribution is the normal distribution (also called a Gaussian distribution). Actually, normal random variables form a whole family that includes the standard normal distribution Z. It's traditional to use the independent variable z for a standard normal distribution. This Z has the probability density function

$$f_Z(z) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}.$$

Note that the integral of this function is not an elementary function, that is, can't be expressed in terms of algebraic functions, trig functions, exponents and logs. Instead, its values are found from tables (which you can find in every statistics textbook) or approximated as needed by calculators or computers. The term "Bell curve" refers to this density function. Here's a graph of it where the axes have the same scale. Usually you'll see it with an exaggerated vertical scale.



The rest of the normal distributions come from scaling and translating this the standard normal Z. If $X = \sigma Z + \mu$, where σ and μ are constants, then X is called a *normal random variable with mean μ and standard deviation σ.*

The density function for a normal distribution $X = \sigma Z + \mu$ is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/(2\sigma^2)} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Note that $\exp x$ is another notation used for e^x when the exponent is complicated. The expectation $E(X)$ of this distribution is μ, and its standard deviation is σ.

Normal distributions are very commonly found. They arise from "natural" variation. For instance if you measure the same thing many times you will get different measurements that tend to be normally distributed around the actual value. For an example from nature, different plants (or animals) growing in the same environment vary in their size and other measurable attributes, and these attributes tend to be normally distributed.

The Central Limit Theorem. This natural variation can be explained by summing together many small variations due to different factors. The mathematical explanation is found in the Central Limit Theorem or its many generalizations. It says that if you average many independent random variables $X_1 + X_2 + \cdots + X_n$ all having the same mean μ and standard deviation σ, then the average

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

is very close to a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$. More precisely, the limit as $n \to \infty$ of $(\overline{X} - \mu)\sqrt{n}$ is exactly a standard normal distribution.

There are generalizations of the Central Limit Theorem that weaken the requirements of independence and having the same means and standard deviations. That explains why when there are many factors that influence an attribute, the result is close to a normal distribution.

Many of the techniques of statistics rely on this Central Limit Theorem to estimate from statistical measurements both the means and the standard deviations. We'll prove it later in the course.

Math 217 Home Page at `http://math.clarku.edu/~djoyce/ma217/`