

Continuous probability distributions, Part I
 Math 121 Calculus II
 Spring 2015

One of the major applications of calculus is to probability and statistics, in particular, to continuous distributions. Not all of probability needs calculus; in the study of finite probability there are only finitely many possible outcomes, and for that, arithmetic and algebra are enough. But when there are infinitely many possible cases, the methods of calculus—derivatives, integrals, and infinite series—are essential.

A quick background look at the discrete case. In discrete probability, you only consider a finite number of possible outcomes. For instance, if you toss a coin there are two outcomes, heads or tails. If you roll a die, there are six. If you ask someone if they prefer candidate A, candidate B, or candidate C, then there are three outcomes.

Each outcome has a particular probability. Sometimes that probability is known, but sometimes the problem is to determine the probabilities. For instance, if you know that your coin is fair, then heads and tails have the same probability. The numerical probability for each outcome is a number between 0 and 1. That number is intended to indicate the relative frequency of that outcome. For a fair coin, heads and tails each come up about half the time, so probabilities of $\frac{1}{2}$ are assigned to each outcome. Notationally, $P(\text{heads}) = \frac{1}{2}$ and $P(\text{tails}) = \frac{1}{2}$.

Note that the sum of the probabilities of all the outcomes is 1. If you have a fair die, there are six outcomes each with the same probability, so each probability is $\frac{1}{6}$.

In statistics, the problem is determine these probabilities when they're not known beforehand. For instance, when you want to know the popular preferences for candidates A, B, and C, you can gather statistics by surveying a small sample of the population. If you have a random sample, then the relative frequencies for your sample give you estimates for the probabilities. Say you sample 100 people at random and you find that of those 23 preferred candidate A, 52 preferred candidate B, and 25 preferred candidate C. That suggests that the probability that a random person prefers A is 0.23, $P(B) = 0.52$, and $P(C) = 0.25$. Statistics doesn't stop there, though, since you know these suggestions for probabilities aren't perfect; you also want to estimate the errors in these estimates.

We'll assume that we know the probabilities so that we don't have to get into the difficulties of statistics.

Continuous probability. In many applications there aren't only finitely many outcomes but infinitely many; the outcome may be any real number in some interval (perhaps an infinite interval). An outcome X that can take a real number as a value is called a *random variable*. We'll look at three examples of continuous probabilities. The first will be a uniform random variable on an interval $[a, b]$.

Suppose that a number X is chosen at random on the interval $[a, b]$ where a and b are real constants. There are infinitely many possible outcomes for X . We'll take the case where

every outcome is equally likely. Since there are infinitely many of them, the probability of any one outcome is 0. Probabilities aren't enough, by themselves, to help like they were in the finite case. For continuous probability, we'll assume that the probability that X equals a particular value x is 0, for each number x . That's denoted as $P(X=x) = 0$.

The cumulative distribution function, c.d.f., $F(x)$. Since probabilities of outcomes aren't enough (since they're all 0), we'll need something else. We'll use two related functions, the cumulative distribution function $F(x)$, often called the c.d.f., and the density function $f(x)$.

Let's take the case where the interval is $[0, 3]$. Our random variable X takes some value between 0 and 3. We want to capture the notion of uniformity. We'll do this by assuming that the probability that X takes a value in any subinterval $[c, d]$ only depends on the length of the interval. So for instance, $P(X \in [0, 1.5]) = P(X \in [1.5, 3])$ so each equals $\frac{1}{2}$. Likewise, $P(X \in [0, 1]) = P(X \in [1, 2]) = P(X \in [2, 3]) = \frac{1}{3}$. In fact, the probability that X lies in $[c, d]$ is proportional to the length $d - c$ of the subinterval. Thus, $P(X \in [c, d]) = \frac{1}{3}(d - c)$.

For any continuous random variable the cumulative distribution function $F(x)$ is enough to determine probabilities of intervals. The c.d.f. F gives the probability that X lies in the interval $(-\infty, x]$

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

We can use F to determine probabilities of intervals as follows

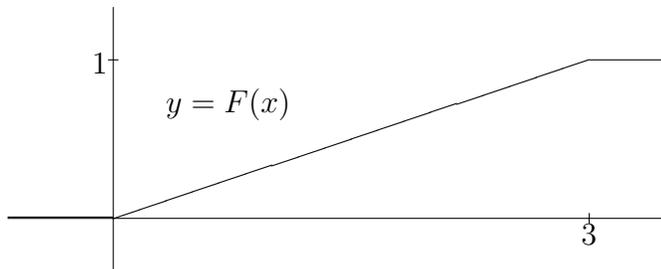
$$P(X \in [a, b]) = P(X \leq b) - P(X \leq a) = F(b) - F(a).$$

(Note that $P(X=a) = 0$.)

For a uniformly continuous random variable on $[0, 3]$ the c.d.f. is defined by

$$F(X) = \begin{cases} 0 & \text{if } x \leq 0 \\ x/3 & \text{if } 0 \leq x \leq 3 \\ 1 & \text{if } 3 \leq x \end{cases}$$

Its graph is



More generally, a uniformly continuous random variable on $[a, b]$ has a c.d.f. which is 0 for $x \leq a$, 1 for $x \geq b$, and increases linearly from $x = a$ to $x = b$ with derivative $\frac{1}{b - a}$ on $[a, b]$.

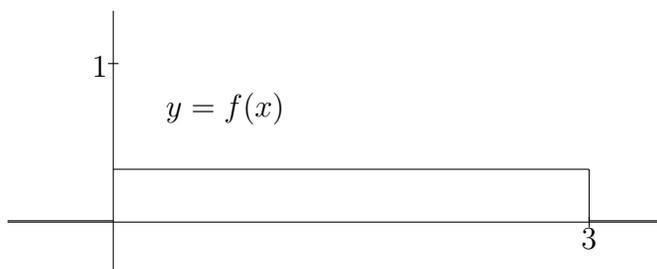
The probability density function $f(x)$. Although the c.d.f. is enough to understand continuous random variables (even when they're not uniform), a related function, the density function, carries just as much information, and its graph visually displays the situation better. In general, densities are derivatives, and in this case, the probability density function $f(x)$ is the derivative of the c.d.f. $F(x)$. Conversely, the c.d.f. is the integral of the density function.

$$f(x) = F'(x) \quad \text{and} \quad F(x) = \int_{-\infty}^x f(t) dt$$

For our example of the uniformly continuous random variable on $[0, 3]$ the probability density function is defined by

$$f(X) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1/3 & \text{if } 0 \leq x \leq 3 \\ 0 & \text{if } 3 \leq x \end{cases}$$

Its graph is



More generally, a uniformly continuous random variable on $[a, b]$ has a density function which is 0 except on the interval $[a, b]$ where it is the reciprocal of the length of the interval, $\frac{1}{b-a}$.

A nonuniform example, the exponential distribution. In a Poisson process, events occur uniformly randomly over time. Events aren't regularly spaced but are random with a constant rate of events per unit time. That rate is denoted λ .

For a Poisson process, the time T to the next occurrence is a random variable with a particular distribution, namely, the exponential distribution.

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 and exponential distributions. 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.

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 $t \leq 0$, 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

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

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

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

Continued in Part II.

Math 121 Home Page at <http://math.clarku.edu/~ma121/>