Vincent's CS70 Discussion 14A Notes

April 27, 2016

1 Pick a number, any number

Consider the following problem: If we pick a real number uniformly at random from the interval $[0, 1]$, what is the probability that we pick $1/2$?

Almost immediately, we start seeing that we currently lack the tools necessary to answer this question. In fact, the question in its current state makes absolutely no sense. Because there are uncountably many real numbers in the interval from 0 to 1, it must be the case that the probability of selecting any one of them must be zero. However, if this is the case, how can we possibly have that the sum of all possibilities in our sample space be equal to 1?

Eventually, we come to the conclusion that it is more or less senseless to talk about the probability of choosing a single number in our interval. Instead, it seems more natural to ask about the probability of our randomly selected number being in a specified *interval*. For example, given that we want it to be "equally likely" for any single number to be chosen, few would complain if we said that the probability that the number we choose is in the interval $[0, 0.1]$ should be 0.1, or that the probability that our number is in the interval $[0.25, 0.75]$ is 0.5. This intuition brings us to the definitions of continuous random variables, cumulative distribution functions (cdfs) and probability density functions (pdfs).

2 Random variables, cdfs, and pdfs

Above, we saw why it no longer makes sense to speak of individual "sample points" of a continuous probability space $\Omega \subseteq \mathbb{R}$. Instead, we work with calculating the probabilities of events that are *intervals* (or unions of intervals), and it turns out that the analysis of such events is aided greatly by the use of random variables (while random variables are of course extremely useful in the analysis of discrete probability spaces, note that we were able to get away with not defining them in the discrete case for quite some time). As before, the definition of a random variable X is a mapping from $\Omega \to \mathbb{R}$, so we now proceed to the following definition.

Definition 1. *A function $F : \mathbb{R} \to [0, 1]$ is called a cumulative distribution function (cdf) for the continuous random variable X if it satisfies the following axioms.*

1. *F is nondecreasing. That is, $x \le y \Longrightarrow F(x) \le F(y)$.*
2. *$\mathbb{P}(X \le x) = F(x),\ \forall x \in \mathbb{R}$.*

It follows that if we want to compute the probability of our random variable X taking a value in the interval $[a, b]$, that is, we want to compute $\mathbb{P}(a \le X \le b)$, we have that

$$\mathbb{P}(a \le X \le b) = \mathbb{P}(X \le b) - \mathbb{P}(X \le a) = F(b) - F(a).$$

Let's return to our motivating example of picking a number uniformly from the interval $[0, 1]$. Our assumptions give us that if X is a (continuous) random variable that denotes the number picked, then the cdf corresponding to X is given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x \le 1 \\ 1 & \text{otherwise} \end{cases}$$

People who have taken calculus (ideally everyone taking this course) may think that the fact that $\mathbb{P}(a \le X \le b) = F(b) - F(a)$ seems familiar. That sense of familiarity turns out to be justified as the concept of a cdf and the fundamental theorem of calculus are intimately related. Before we see exactly how the two are related, we introduce the notion of a probability density function.

Definition 2. *A function $f : \mathbb{R} \to [0, \infty)$ is called a probability density function (pdf) of the random variable X if it satisfies the following.*

1. *f is nonnegative: $\forall x \in \mathbb{R},\ f(x) \ge 0$.*
2. *$\int_{-\infty}^{\infty} f(x)dx = 1$*
3. *$\mathbb{P}(a \le X \le b) = \int_a^b f(x)dx$*

By applying the fundamental theorem of calculus, the relationship between pdfs and cdfs is immediate. Given a (piecewise differentiable) cdf F corresponding to random variable X, the pdf f of X can be found by differentiating F with respect to x. Similarly, given an integrable pdf f of X, we can compute the cdf to be $F(x) = \int_{-\infty}^{x} f(s)ds$.

If we (once again) return to our example of picking a number uniformly in the interval $[0, 1]$, we have that if X is the random variable that denotes the number picked, the pdf corresponding to X is given by

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

3 Some definitions

We now proceed to define the continuous versions of expectation and variance.

Definition 3. *The expectation of a continuous random variable X with pdf f is defined to be*

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

Note that this definition is completely analogous to the definition of discrete expectation, except here we have integrals replacing sums. Also, it is important to mention that linearity of expectation still holds in the continuous case.

The definition of variance essentially remains completely the same:

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

and once again we have the result that

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

To calculate the expectation and variance of picking a number uniformly in the range $[0, 1]$, we have that since the pdf that we are working with is

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise,} \end{cases},$$

we can compute

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x dx = \left[\frac{x^2}{2} \right]_0^1 = \frac{1}{2}$$

and

$$Var(X) = E[X^2] - (E[X])^2 = \left(\int_{-\infty}^{\infty} x^2 f(x) dx \right) - (E[X])^2$$
$$= \left(\int_0^1 x^2 dx \right) - (\frac{1}{2})^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

Finally, the definition of two independent continuous random variables is again completely analogous to the familiar discrete case.

Definition 4. *Two continuous random variables X, Y are independent if*

$$\mathbb{P}(a \le X \le b \cap c \le Y \le d) = \mathbb{P}(a \le X \le b)\mathbb{P}(c \le Y \le d)$$

We'll cover more about the implications of two random variables being independent next section when we cover joint distributions.

4 The exponential distribution

Finally, we state the definitions and a few elementary results concerning an important continuous distribution (those of you waiting excitedly for the Gaussian distribution and the Central Limit Theorem will have to wait until next time). We won't prove the stated results here, but the curious reader can find the proofs in the lecture slides and lecture note 20.

Definition 5. *We say that a continuous random variable X follows an exponential distribution with parameter $\lambda > 0$ if it has pdf f defined by*

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

Intuitively, we want to think of the exponential distribution as the continuous time analog of the geometric distribution. That is, the exponential distribution models the amount of (continuous) time we have to wait until the occurrence of a specified event. Given our analogy with the geometric distribution, a few of the results stated below should seem familiar.

Theorem 1. *Let X be an exponential random variable with parameter $\lambda > 0$. We have that*

- *The cdf of X is given by* $F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$
- $E[X] = \frac{1}{\lambda}$
- $Var(X) = \frac{1}{\lambda^2}$
- *The exponential distribution is memoryless. That is, if X is an exponential random variable, then*

$$\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t).$$