EE 178/278A Probabilistic Systems Analysis

Spring 2014   Tse/Hussami   Lecture 11

A Brief Introduction to Continuous Probability

Up to now we have focused exclusively on *discrete* probability spaces Ω, where the number of outcomes $\omega \in \Omega$ is either finite or countably infinite (such as the integers). As a consequence we have only been able to talk about *discrete* random variables, which take on only a finite (or countably infinite) number of values.

But in real life many quantities that we wish to model probabilistically are *continuous-valued*; examples include the position of a particle in a box, the time at which a certain incident happens, or the direction of travel of a meteorite. In this lecture, we discuss how to extend the concepts we've seen in the discrete setting to this continuous setting. As we shall see, everything translates in a natural way once we have set up the right framework. The framework involves some elementary calculus.

Continuous uniform probability spaces

Suppose we spin a "wheel of fortune" and record the position of the pointer on the outer circumference of the wheel. Assuming that the circumference is of length ℓ and that the wheel is unbiased, the position is presumably equally likely to take on any value in the real interval $[0, \ell]$. How do we model this experiment using a probability space?

Consider for a moment the (almost) analogous discrete setting, where the pointer can stop only at a finite number m of positions distributed evenly around the wheel. (If m is very large, then presumably this is in some sense similar to the continuous setting.) Then we would model this situation using the discrete sample space $\Omega = \{0, \frac{\ell}{m}, \frac{2\ell}{m}, \ldots, \frac{(m-1)\ell}{m}\}$, with uniform probabilities $\mathbf{P}(\omega) = \frac{1}{m}$ for each $\omega \in \Omega$. In the continuous world, however, we get into trouble if we try the same approach. If we let ω range over all real numbers in $[0, \ell]$, what value should we assign to each $\mathbf{P}(\omega)$? By uniformity this probability should be the same for all ω, but then if we assign to it any positive value, the sum of all probabilities $\mathbf{P}(\omega)$ for $\omega \in \Omega$ will be ∞! Thus $\mathbf{P}(\omega)$ must be zero for all $\omega \in \Omega$. But if all of our outcomes have probability zero, then we are unable to assign meaningful probabilities to any events!

To rescue this situation, consider instead any non-empty *interval* $[a, b] \subseteq [0, \ell]$. Can we assign a non-zero probability value to this interval? Since the total probability assigned to $[0, \ell]$ must be 1, and since we want our probability to be uniform, the natural assignment of probability to the interval $[a, b]$ is

$$\mathbf{P}([a, b]) = \frac{\text{length of } [a, b]}{\text{length of } [0, \ell]} = \frac{b - a}{\ell}. \tag{1}$$

In other words, the probability of an interval is proportional to its length.

Note that intervals are subsets of the sample space Ω and are therefore *events*. So in continuous probability, we are assigning probabilities to certain basic events, in contrast to discrete probability, where we assigned probability to *points* in the sample space. But what about probabilities of other events? Actually, by specifying the probability of intervals we have also specified the probability of any event E which can be written
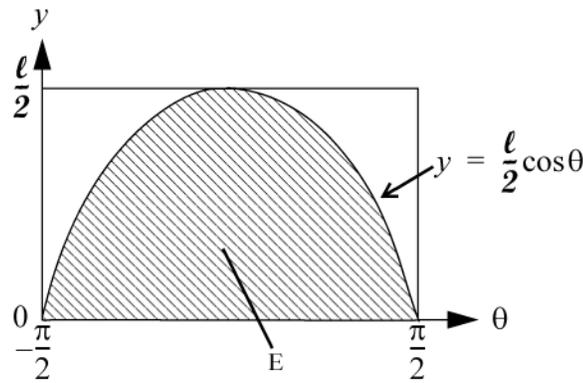
Figure 1: The event E as a subset of the sample space Ω.

as the disjoint union of (a finite or countably infinite number of) intervals, $E = \cup_i E_i$. For then we can write $\mathbf{P}(E) = \sum_i \mathbf{P}(E_i)$, in analogous fashion to the discrete case. Thus for example the probability that the pointer ends up in the first or third quadrants of the wheel is $\frac{\ell/4}{\ell} + \frac{\ell/4}{\ell} = \frac{1}{2}$. For all practical purposes, such events are all we really need.[1]

An example: Buffon's needle

Here is a simple application of continuous probability to the analysis of a classical procedure for estimating the value of π known as *Buffon's needle*, after its 18th century inventor Georges-Louis Leclerc, Comte de Buffon.

Here we are given a needle of length ℓ, and a board ruled with horizontal lines at distance ℓ apart. The experiment consists of throwing the needle randomly onto the board and observing whether or not it crosses one of the lines. We shall see below that (assuming a perfectly random throw) the probability of this event is exactly $2/\pi$. This means that, if we perform the experiment many times and record the *proportion* of throws on which the needle crosses a line, then the Law of Large Numbers (Lecture Note 10) tells us that we will get a good estimate of the quantity $2/\pi$, and therefore also of π; and we can use Chebyshev's inequality as in the other estimation problems we considered in that same Lecture Note to determine how many throws we need in order to achieve specified accuracy and confidence.

To analyze the experiment, we first need to specify the probability space. Note that the position where the needle lands is completely specified by two numbers: the vertical distance y between the midpoint of the needle and the closest horizontal line, and the angle θ between the needle and the vertical. The vertical distance y ranges between 0 and $\ell/2$, while θ ranges between $-\pi/2$ and $\pi/2$. Thus, the sample space is the rectangle $\Omega = [-\pi/2, \pi/2] \times [0, \ell/2]$. Note that, compared to the wheel-of-fortune example, the sample space is two-dimensional rather than one-dimensional. But like the wheel-of-fortune example, the sample space is also continuous.

Now let E denote the event that the needle crosses a line. It is a subset of the sample space Ω. We need to identify this subset explicitly. By elementary geometry the vertical distance of the endpoint of the needle from its midpoint is $\frac{\ell}{2}\cos\theta$, so the needle will cross the line if and only if $y \leq \frac{\ell}{2}\cos\theta$. The event E is sketched in Figure 1.

---

[1]A formal treatment of which events can be assigned a well-defined probability requires a discussion of *measure theory*, which is beyond the scope of this course.

Since we are assuming a completely random throw, probability of the event E is:

$$\mathbf{P}(E) = \frac{\text{area of } E}{\text{area of } \Omega}.$$

This is the two-dimensional generalization of equation (1) in the wheel-of-fortune example, where the probability of landing in an interval is proportional to the length of the interval.

The area of the whole sample space is $\pi\ell/2$. The area of E is:

$$\int_{-\pi/2}^{\pi/2} \frac{\ell}{2} \cos\theta \, d\theta = \left[\frac{\ell}{2} \sin\theta\right]_{-\pi/2}^{\pi/2} = \ell.$$

Hence,

$$\mathbf{P}(E) = \frac{\ell}{\pi\ell/2} = \frac{2}{\pi}.$$

This is exactly what we claimed at the beginning of the section!

Continuous random variables

Recall that in the discrete setting we typically work with *random variables* and their distributions, rather than directly with probability spaces and events. This is even more so in continuous probability, since numerical quantities are almost always involved. In the wheel-of-fortune example, the position X of the pointer is a random variable. In the Buffon needle example, the vertical distance Y and the angle Θ are random variables. In fact, they are all *continuous* random variables. These random variables are all relatively simple, in the sense that they are all *uniformly* distributed on the range of values they can take on. (Because of this simplicity, we didn't even need to worry about the random variables explicitly when calculating probabilities in these examples, and instead reason directly with the sample space.) But more complicated random variables do not have a uniform distribution. How, precisely, should we define the distribution of a general continuous random variable? In the discrete case the distribution of a r.v. X is described by specifying, for each possible value a, the probability $\mathbf{P}(X = a)$. But for the r.v. X corresponding to the position of the pointer, we have $\mathbf{P}(X = a) = 0$ for every a, so we run into the same problem as we encountered above in defining the probability space.

The resolution is essentially the same: instead of specifying $\mathbf{P}(X = a)$, we instead specify $\mathbf{P}(a < X \le b)$ for all intervals $[a, b]$.[2] To do this formally, we need to introduce the concept of a *probability density function* (sometimes referred to just as a "density", or a "pdf").

Definition 11.1 (Density): A *probability density function* for a random variable X is a function $f : \mathbb{R} \to \mathbb{R}$ satisfying

$$\mathbf{P}(a < X \le b) = \int_a^b f(x)dx \qquad \text{for all } a \le b. \tag{2}$$

Let's examine this definition. Note that the definite integral is just the area under the curve f between the values a and b (Figure 2(a)). Thus f plays a similar role to the "histogram" we sometimes draw to picture the distribution of a discrete random variable.

In order for the definition to make sense, f must obey certain properties. Some of these are technical in nature, which basically just ensure that the integral is always well defined; we shall not dwell on this issue

---

[2]Note that it does not matter whether or not we include the endpoints a, b; since $\mathbf{P}(X = a) = \mathbf{P}(X = b) = 0$, we have $\mathbf{P}(a < X < b) = \mathbf{P}(a < X \le b) = \mathbf{P}(a < X < b)$.
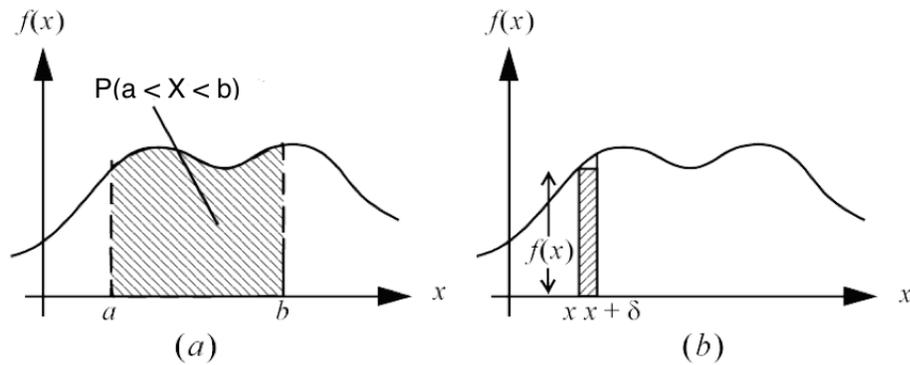
Figure 2: (a) The area under the density curve between a and b is the probability that the random variable lies in that range. (b) For small δ, the area under the curve between x and $x + \delta$ can be well approximated by the area of the rectangle of height $f(x)$ and width δ.

here since all the densities that we will meet will be well behaved. What about some more basic properties of f? First, it must be the case that f is a non-negative function; for if f took on negative values we could find an interval in which the integral is negative, so we would have a negative probability for some event! Second, since the r.v. X must take on some value everywhere in the space, we must have

$$\int_{-\infty}^{\infty} f(x)dx = \mathbf{P}(-\infty < X < \infty) = 1. \tag{3}$$

In other words, the total area under the curve f must be 1.

A caveat is in order here. Following the "histogram" analogy above, it is tempting to think of $f(x)$ as a "probability." However, $f(x)$ doesn't itself correspond to the probability of anything! For one thing, there is no requirement that $f(x)$ be bounded by 1 (and indeed, we shall see examples of densities in which $f(x)$ is greater than 1 for some x). To connect $f(x)$ with probabilities, we need to look at a very small interval $[x, x + \delta]$ close to x. Assuming that the interval $[x, x + \delta]$ is so small that the function f doesn't change much over that interval. we have

$$\mathbf{P}(x < X \le x + \delta) = \int_{x}^{x+\delta} f(z)dz \approx \delta f(x). \tag{4}$$

This approximation is illustrated in Figure 2(b). Equivalently,

$$f(x) \approx \frac{\mathbf{P}(x < X \le x + \delta)}{\delta}. \tag{5}$$

The approximation in (5) becomes more accurate as δ becomes small. Hence, more formally, we can relate density and probability by taking limits:

$$f(x) = \lim_{\delta \to 0} \frac{\mathbf{P}(x < X \le x + \delta)}{\delta}. \tag{6}$$

Thus we can interpret $f(x)$ as the "probability per unit length" in the vicinity of x. Note that while the equation (2) allows us to compute probabilities given the probability density function, the equation (6) allows us to compute the probability density function given probabilities. Both relationships are useful in problems.
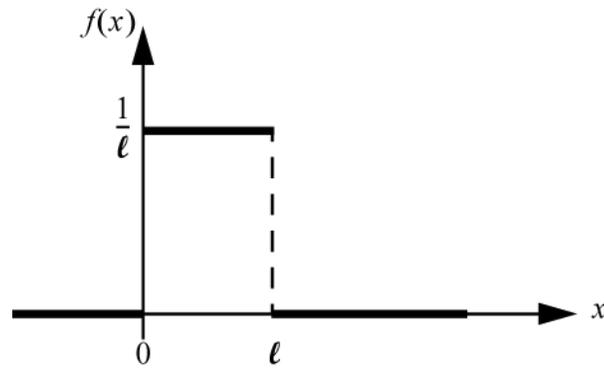
Figure 3: The density function of the wheel-of-fortune r.v. X.

Now let's go back and put our wheel-of-fortune r.v. X into this framework. What should be the density of X? Well, we want X to have non-zero probability only on the interval $[0, \ell]$, so we should certainly have $f(x) = 0$ for $x < 0$ and for $x > \ell$. Within the interval $[0, \ell]$ we want the distribution of X to be uniform, which means we should take $f(x) = c$ for $0 \leq x \leq \ell$. What should be the value of c? This is determined by the requirement (3) that the total area under f is 1. The area under the above curve is $\int_{-\infty}^{\infty} f(x) dx = \int_0^{\ell} c \, dx = c\ell$, so we must take $c = \frac{1}{\ell}$. Summarizing, then, the density of the uniform distribution on $[0, \ell]$ is given by

$$f(x) = \begin{cases} 0 & \text{for } x < 0; \\ 1/\ell & \text{for } 0 \leq x \leq \ell; \\ 0 & \text{for } x > \ell. \end{cases}$$

This is plotted in Figure 3. Note that $f(x)$ can certainly be greater than 1, depending on the value of ℓ.

Another Example: Suppose you throw a dart and it lands uniformly at random on a target which is a disk of unit radius. What is the probability density function of the distance of the dart from the center of the disk?

Let X be the distance of the dart from the center of the disk. We first calculate the probability that X is between x and $x + \delta$. If x is negative or greater than or equal to 1, this probability is zero, so we focus on the case that x is between 0 and 1. The event in question is that the dart lands in the ring (annulus) shown in Figure 4. Since the dart lands uniformly at random on the disk, the probability of the event is just the ratio of the area of the ring and the area of the disk. Hence,

$$\mathbf{P}(x < X \leq x + \delta) \quad = \quad \frac{\pi[(x + \delta)^2 - x^2]}{\pi(1)^2} \tag{7}$$
$$= \quad x^2 + 2\delta x + \delta^2 - x^2 = 2\delta x - \delta^2.$$

Using equation (6), we can now compute the probability density function of X:

$$f(x) = \lim_{\delta \to 0} \frac{\mathbf{P}(x < X \leq x + \delta)}{\delta} = f(x) = \lim_{\delta \to 0} \frac{2\delta x - \delta^2}{\delta} = 2x.$$

Summarizing, we have

$$f(x) = \begin{cases} 0 & \text{for } x \leq 0; \\ 2x & \text{for } 0 \leq x < 1; \\ 0 & \text{for } x \geq 1. \end{cases}$$
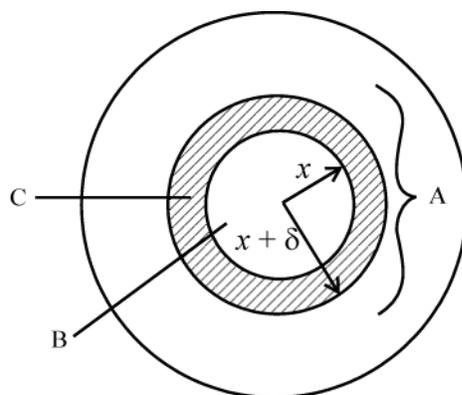
Figure 4: The sample space is the disk of unit radius. The event of interest is the ring.
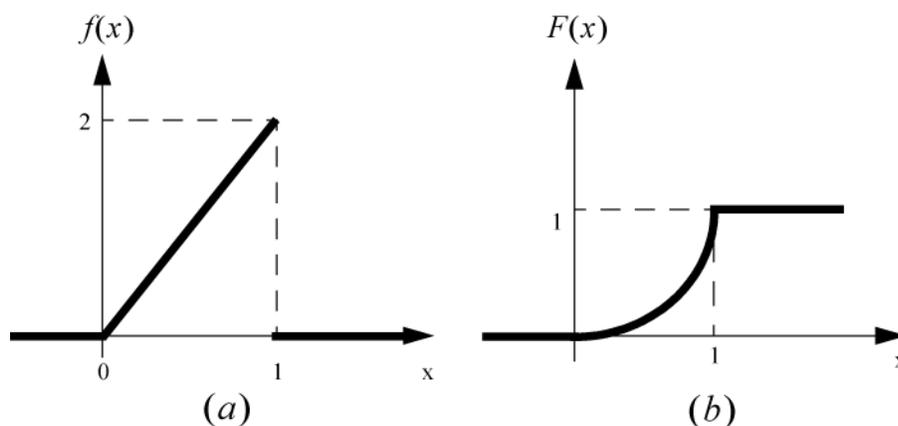

Figure 5: (a) The probability density function and (b) the cumulative distribution function of the distance X from the target center.

It is plotted in Figure 5(a). Note that although the dart lands uniformly inside the target, the distance X from the center is *not* uniformly distributed in the range from 0 to 1. This is because an ring farther away from the center has a larger area than an ring closer to the center with the same width δ. Hence the probability of landing in the ring farther away from the center is larger.

Cumulative Distribution Function

Let us re-interpret equation (7) in the dart throwing example above. In words, we are saying:

$$
\begin{aligned}
\mathbf{P}(x < X \leq x + \delta) &= \frac{\text{area of ring}}{\text{area of target}} \\
&= \frac{(\text{area of disk of radius } x + \delta) - (\text{area of disk of radius } x)}{\text{area of target}} \\
&= \frac{\text{area of disk of radius } x + \delta}{\text{area of target}} - \frac{\text{area of disk of radius } x}{\text{area of target}} \\
&= \mathbf{P}(X \leq x + \delta) - \mathbf{P}(X \leq x).
\end{aligned}
$$

This last equality can be understood directly as follows. The event A that $X \le x + \delta$ (dart lands inside disk of radius $x + \delta$) can be decomposed as a union of two events: 1) the event B that $X \le x$ (dart lands inside disk of radius x), and 2) the event C that $x < X \le x + \delta$ (dart lands inside ring). The two events are *disjoint*. (See Figure 4.) Hence,

$$\mathbf{P}(A) = \mathbf{P}(B) + \mathbf{P}(C)$$

or

$$\mathbf{P}(x < X \le x + \delta) = \mathbf{P}(X \le x + \delta) - \mathbf{P}(X \le x), \tag{8}$$

which is exactly the same as above.

Clearly, the reasoning leading to (8) has nothing much to do the particulars of this example but in fact (8) holds true for *any* random variable X. All we needed are the facts that $A = B \cup C$ and B and C are disjoint events, and the facts are true in general.

Substituting (8) into (6), we obtain:

$$f(x) = \lim_{\delta \to 0} \frac{\mathbf{P}(X \le x + \delta) - \mathbf{P}(X \le x)}{\delta}.$$

What does this equation remind you of? To make things even more explicit, let us define the cumulative distribution function

$$F(x) = \mathbf{P}(X \le x). \tag{9}$$

Then we have:

$$f(x) = \lim_{\delta \to 0} \frac{F(x + \delta) - F(x)}{\delta} = \frac{d}{dx} F(x). \tag{10}$$

The function F has a name: it is called the *cumulative distribution function* of the random variable X (sometimes abbreviated as cdf). It is called "cumulative" because at each value x, $F(x)$ is the cumulative probability up to x. Note that the cumulative distribution function and the probability density function of a random variable contains *exactly* the same information. Given the cumulative distribution function F, one can differentiate to get the probability density function f. Given the probability density function f, one can integrate to get the cumulative distribution function:

$$F(x) = \int_{-\infty}^{x} f(a)da.$$

So strictly speaking, one does not need to introduce the concept of cumulative distribution function at all. However, for many problems, the cumulative distribution function is easier to compute first and from that one can then compute the probability density function.

Summarizing, we have:

Definition 11.2 (Cumulative Distribution Function): The *cumulative distribution function* for a random variable X is a function $F : \mathbb{R} \to \mathbb{R}$ defined to be:

$$F(x) = \mathbf{P}(X \le x). \tag{11}$$

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

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

The cumulative distribution function satisfies the following properties:

1. $0 \leq F(x) \leq 1$

2. $\lim_{x \to -\infty} F(x) = 0$

3. $\lim_{x \to \infty} F(x) = 1$