**18.440 Probability and Random Variables      Fall Semester, 2003**

## Supplementary notes on continuous random variables

The text passes over in silence some basic questions related to continuous random variables. The reason seems to be a desire to keep the calculus needed to single variable as much as possible. That's a laudable goal, but of course it's not necessary for MIT students, who have multiple integrals for breakfast. These notes are meant to provide a little bit of additional background for the ideas in the text. I will still be a bit sloppy. None of this can be done in a mathematically consistent and precise way without Lebesgue integrals, whose very pronunciation can cause strong men to weep. But I promise not to be too misleading.

The first question is where continuous random variables come from. A random variable is a function defined on a sample space; but in Chapter 5, sample spaces are almost nowhere to be found. So where do you get sample spaces on which continuous random variables can be defined? Here's one place ...

**Example 1.** Supppose that $a < b$ are real numbers, and that $S = [a, b] \subset \mathbb{R}$ is the interval from $a$ to $b$. Suppose $m$ is a non-negative real-valued function defined on $[a, b]$, and assume that $\int_a^b m(s)\, ds = 1$. We define a probability $P$ on subsets $E \subset [a, b]$ by

$$P(E) = \int_E m(s)\, ds. \qquad \text{SAMPLE SPACE PROBABILITY}$$

Here are some things to notice about this probability. First, the assumption on $m$ was included to make $P(S) = 1$. Second, the probability takes values between 0 and 1 (as it had better). Third, the probability of a single point (a single element of $S$) is always zero, since $\int_c^c m(s)\, ds = 0$.

A *random variable on* $S$ is a real-valued function $X$ on $[a, b]$. The *expected value* of $X$ is

$$E(X) = \int_a^b X(s) m(s)\, ds. \qquad \text{SAMPLE SPACE EXPECTATION}$$

The *second moment* of $X$ is

$$E(X^2) = \int_a^b X^2(s) m(s)\, ds$$

and the *variance* of $X$ is

$$\mathrm{Var}(X) = E((X - E(X))^2) = \int_a^b (X(s) - E(X))^2 m(s)\, ds. \qquad \text{SAMPLE SPACE VARIANCE}$$

In fact you can define the expected value of any function $g(X)$ of the random variable $X$:

$$E(g(X)) = \int_a^b g(X(s)) m(s)\, ds.$$

I hope these definitions look reasonable in light of the definition of "average value" that you learned in calculus. Unfortunately, they look quite different from the definitions in the text. In order to connect the two, we need to perform a change of variables in these integrals. Before doing that, let's look at an example just to keep things concrete.

**Example of Example 1.** A passenger arrives at a bus stop at a time that is uniformly distributed between 7:00 and 7:30. Buses depart at 7:05, 7:20, 7:35, and so on. What is the probability of having to wait at least ten minutes for a bus? How long should the passenger expect to wait on average?

**Solution** For this problem a reasonable sample space is the interval $[0, 30]$: the point $s$ corresponds to arriving exactly $s$ minutes after 7:00. (Being MIT passengers, we keep track of times as real numbers: you can arrive at 10.13579 minutes after 7:00, in which case you missed the 7:10 bus (by more than eight seconds).) To say that the arrival time is "uniformly distributed" over the half hour means that the function $m(s) = 1/30$: the probability of arriving during some specified interval is one thirtieth of its length (in minutes).

Waiting at least ten minutes for a bus means arriving in one of the intervals (5,10] (missing the 7:05 bus, and waiting at least ten minutes for the 7:20) or (20,25]. The probability is therefore

$$P(\text{ten minute wait}) = P((5,10] \cup (20,25]) = 1/30(\text{length}) = 1/30(5+5) = 1/3.$$

The random variable that gives the actual waiting time is

$$X(s) = \begin{cases} 5 - s, & \text{if } s \in [0,5]; \\ 20 - s, & \text{if } s \in (5,20]; \\ 35 - s, & \text{if } s \in (20,30]. \end{cases}$$

The expected value of $X$ is therefore

$$\begin{aligned} E(X) &= \frac{1}{30}\int_0^{30} X(s)\,ds \\ &= \frac{1}{30}\int_0^5 (5-s)\,ds + \frac{1}{30}\int_5^{20} (20-s)\,ds + \frac{1}{30}\int_{20}^{30} (35-s)\,ds \\ &= \frac{1}{30}(5^2/2 + 15^2/2 + 10^2) = \frac{1}{30}(225) = 7.5. \end{aligned}$$

So the expected waiting time (with buses every fifteen minutes) is seven and a half minutes. Not terrifically surprising.

**Relating Example 1 expected values to book's expected values.** The text says that a random variable $X$ has "probability density function" $f$ if for any set $B$ of real numbers,

$$P(X \in B) = \int_B f(x)\,dx. \qquad \text{BOOK 5.1.1}$$

Here is a slightly less formal way to say the same thing: for any real number $x$, and any small $\Delta x$,

$$P(x \le X \le x + \Delta x) \approx f(x)\Delta x. \qquad \text{BOOK 5.1.1 REPHRASED}$$

So let's look now at the setting of Example 1 (particularly the formula for expectation), and try to change variables in the integral over $S$. We want to go from the old variable $s$ to the new variable $X(s)$. Say that $X$ takes values in the interval $[c, d]$. Think about breaking up the integral into pieces corresponding to little ranges of values of $X$; that is, chopping up $[c, d]$ into little pieces. What's the contribution of values of $X$ between $x$ and $x + \Delta x$? It's

$$\int_{x \le X(s) \le x + \Delta x} X(s) m(s)\, ds.$$

The function $X(s)$ stays close to the constant $x$ here, so this integral is approximately

$$x \int_{x \le X(s) \le x + \Delta x} m(s)\, ds = xP(x \le X \le x + \Delta x).$$

According to the rephrased definition of density functions, this last probability is approximately $f(x)\Delta x$. Putting it all together gives

$$\int_{x \le X(s) \le x + \Delta x} X(s) m(s)\, ds \approx x f(x) \Delta x.$$

Now add up these formulas over a lot of little intervals of values of $X(s)$. On the left the pieces add up to the integral of $X(s)m(s)$ over $S$. On the right, the pieces give a Riemann sum for $\int_c^d x f(x)\, dx$. The conclusion (taking limits as $\Delta x$ goes to zero) is

$$\int_a^b X(s) m(s)\, ds = \int_c^d x f(x)\, dx.$$

The formula on the left is the "sample space" formula for expected values, from Example 1 above. The formula on the right is the book's definition of expected value, from page 191. (I assumed that $X$ took values in the finite range $[c, d]$, in order to reduce the mathematical sloppiness in the argument. That's why the integral here goes from $c$ to $d$, while the one in the book goes from $-\infty$ to $\infty$.

Computationally this discussion has left one serious question hanging: in the setting of Example 1, how can you hope to find a probability density function satisfying the requirement BOOK 5.1.1 REPHRASED?

**Computing probability density functions in Example 1.** Suppose that the function $X(s)$ has a derivative (except for finitely many bad values of $s$), and that the derivative is not zero (at except for another finite bad set). Fix some value of $x$ of $X(s)$. The first problem is to find when $X$ takes this value; that is, to look for solutions of $X(s) = x$. Suppose the solutions are

$$s_1, s_2, \ldots, s_r \in [a, b];$$

that is, that

$$X(s_i) = x, \quad (i = 1, \ldots, r), \qquad X(s) \ne x \quad (s \notin \{s_i\}).$$

If $t$ is small, then the linear approximation to $X$ near $s_i$ says

$$X(s_i + t) \approx x + tX'(s_i).$$

This suggests that

$$\{s \in S \mid X(s) \in [x, x + \Delta x]\} \approx \{s \in S \mid s \text{ between } s_i \text{ and } s_i + \Delta x / X'(s_i)\}.$$

(Of course this only makes sense when all the $X'(s_i)$ are defined and non-zero, but we're assuming that's usually true.)

Now the ith interval in this last formula has length $\Delta x/|X'(s_i)|$. According to the formula SAMPLE SPACE EXPECTATION from Example 1, its probability is approximately $(\Delta x)m(s_i)/|X'(s_i)|$. Putting all these intervals together, we conclude that

$$P(x \leq X(s) \leq x + \Delta x) \approx (\Delta x) \sum_{i=1}^{r} m(s_i)/|X'(s_i)|.$$

Comparing this formula with the definition of density functions says

$$f(x) = \sum_{i=1}^{r} m(s_i)/|X'(s_i)|.$$

Keeping in mind where the $s_i$ came from gives

$$f(x) = \sum_{s \in S, X(s) = x} m(s)/|X'(s)|.$$

DENSITY FUNCTION FORMULA

**Example of computing probability density functions.** Suppose we are in the setting of **Example of Example 1** (waiting for buses), and we want to compute a probability density function for the waiting time random variable $X(s)$. This random variable takes values between 0 and 15, so we need to compute $f(x)$ for $x \in [0, 15]$. The function $X(s)$ has a derivative everywhere except at $s = 5$ and $s = 20$; the derivative is always $-1$. The DENSITY FUNCTION FORMULA above says

$$f(x) = \sum_{s \in [0,30], f(s) = x} m(s)/|X'(s)|.$$

The derivative here is always 1, and $m(s)$ is always 1/30, so

$$f(x) = (1/30)(\text{number of } s \text{ such that } f(s) = x.)$$

So we need (for fixed $x \in [0, 15]$) to count the solutions of $f(s) = x$. For that we refer to the formulas for $X(s)$ in the example. If $x \in [0, 5)$, then the solutions are $5 - x$ and $20 - x$. If $x \in (5, 15)$, then the solutions are $20 - x$ and $35 - x$. If $x = 5$ there are three solutions ($s = 0$, $s = 15$, and $s = 30$) and if $x = 15$ there are none; but these possibilities have probability zero. The conclusion is that the number of solutions is (usually) two, and the probability density function for the waiting time random variable is

$$f(x) = 1/15 \qquad (x \in [0, 15]).$$

The whole point of these notes was supposed to be that you ate multiple integrals for breakfast; but I haven't used any multiple integrals yet. So here they are ...

**Example 2.** . Supppose that $S \subset \mathbb{R}^n$ is some region in $n$-dimensional space. (You can perfectly well take $n = 2$ or $3$ if you prefer.) Suppose $m$ is a non-negative real-valued function defined on $S$, and assume that $\int_S m(s)\,ds = 1$. We define a probability $P$ on subsets $E \subset S$ by

$$P(E) = \int_E m(s)\,ds.$$

SAMPLE SPACE PROBABILITY

Here are some things to notice about this probability. First, the assumption on $m$ was included to make $P(S) = 1$. Second, the probability takes values between 0 and 1 (as it had better). Third, the probability of a "lower dimensional subset" (like a line in $\mathbb{R}^2$ is always zero.

A *random variable on* $S$ is a real-valued function $X$ on $[a, b]$. The *expected value* of $X$ is

$$E(X) = \int_S X(s)m(s)\,ds.$$

SAMPLE SPACE EXPECTATION

The *second moment* of $X$ is

$$E(X^2) = \int_S X^2(s)m(s)\,ds$$

and the *variance* of $X$ is

$$\mathrm{Var}(X) = E((X - E(X))^2) = \int_S (X(s) - E(X))^2 m(s)\,ds.$$

SAMPLE SPACE VARIANCE

A lot of what I said about the sample space $[a, b]$ carries over to this setting without any change. Even the section **Relating Example 1 expected values to book's expected values** can be carried over pretty easily. The only difference is that when we looked at the part of $[a, b]$ where $X$ is close to $x$, we got a nice little collection of short intervals (obtained by fattening up the points $s$ where $X(s) = x$). When we look in $\mathbb{R}^n$ at the set where $X$ is equal to $x$, we get a (possibly very complicated) $(n-1)$-dimensional surface; and fattening it up is even more complicated. But there is still an analogue of the formula we got for a density function:

$$f(x) = \int_{\{s \in S, X(s) = x\}} m(s)/|\nabla X(s)|ds.$$

DENSITY FUNCTION FORMULA

The integral extends over the surface defined by $X(s) = s$. The gradient $\nabla X$ is a vector; what appears here is the length of the gradient. I won't try to write precise conditions making this true: at least we need $X$ to have a non-zero derivative most places, or the formula doesn't even make sense.