Classroom Notes

Math 394B Summer 2005

Week 7

1 Continuous Random Variables

The discussion in class is meant to underline how the shift from discrete to continuous RVs entails a number of traps that usually go unnoticed, but that may occasionally snap...

1.1 Continuous and Absolutely Continuous RVs

In any elementary treatment you will have that (absolutely) continuous RVs are defined by the fact that

$$P\left[a < X \leq b\right] = F_X\left(b\right) - F_X\left(a\right) = \int_a^b f_X\left(x\right) dx$$

for a suitable function f_X (the "density" of X). Note that this means that the cdf F_X has to be *differentiable* (except, possibly, at a finite number of points). Thus, in general,

$$f_X\left(x\right) = \frac{dF_X\left(x\right)}{dx}$$

Note that it is possible to write (or, better said, define) monotone increasing functions that are not differentiable at a countable dense set of values - such functions do not really allow for a density function. A famous (advanced) example is a function that is increasing only at the points that form the standard Cantor set - not easy to visualize, but easily defined in advanced real analysis textbooks.

The restriction to absolutely continuous (ac) RVs means that we expect

$$P\left[X \in \left(a - \varepsilon, a + \varepsilon\right)\right] \simeq C\left(a\right) \varepsilon$$

i.e., that the probability of falling in a small interval around a value is, more or less, proportional to the length of the interval - the constant could be different from value to value. It is also almost implied (we won't go into the fine technical point) that the density function should be continuous, except, possibly, at a finite number of points. The issue here is that f_X is supposed to be integrable, and even though it is not strictly necessary for a function to be continuous, in order to be Riemann integrable, it sure helps!

1.2 Discrete Approximations For Continuous RVs

Suppose we have a RV X that "takes all possible real values", and wish to study it with the tools we already have.

A natural easy step would be to take the range of X, and cut it up in small intervals, say $\ldots, (x_{n-1}, x_n], (x_n, x_{n+1}], \ldots$ (perhaps, beginning with a big semi-infinite interval $(-\infty, x_1]$, and close with another one, $(x_N, \infty])$. Then, assuming each interval is short enough, we can think of X having approximately a constant value in it (since the differences $x_k - x_{k-1}$ are supposed to be very small, we could think that for all ωsuch that $x_{k-1} < X(\omega) \leq x_k$, taking, say, $X(\omega) = \frac{x_{k-1}+x_k}{2}$ would cause a negligible error), like X_k in $(x_k, x_{k+1}]$, and use, instead of X, the discrete RV X_N, defined by

$$X_N(\omega) = X_k \quad \text{when } X(\omega) \in (x_k, x_{k+1}]$$

All usual calculations we have done on discrete RVs could then be transferred to continuous RVs, by (hopefully convergent) limit operations.

The "glitch" in this program is that we would be working with

$$P[X_N = X_k] = P[x_k < X \leq X_{k+1}]$$

and, for sufficiently elaborate RVs, the right hand side can be very tricky to evaluate. In particular, it is not obvious that we could, for instance, compute

$$EX = \lim_{N \to \infty} EX_N = \lim \sum_k X_k P[x_k < X \leq x_{k+1}]$$

(and in fact, in general, this limit would not correspond to a Riemann integral!).

Our restriction to ac RVs allow us to consider only the case when there is a nice function f_X, such that

$$P[x_k < X \leq x_{k+1}] = \int_{x_k}^{x_{k+1}} f_X(x)\, dx$$

in which case

$$\lim \sum_k X_k P[x_k < X \leq x_{k+1}] = \lim \sum_k X_k \int_{x_k}^{x_{k+1}} f_X(x)\, dx$$

which, after observing that $x_k < X_k < x_{k+1}$, and that f_X is hopefully continuous, will be in fact equal to

$$\int x f_X(x)\, dx$$

and this will be a rightful definition for EX.

Remark The caveat "except, possibly, at a finite number of points" is there simply to allow for densities that are discontinuous, but such that we can easily compute their anti-derivative anyway, by breaking up the domain of integration in subintervals where f is continuous.

Remark There is no special difficulty in considering also "mixed" RVs, whose distribution is a mix of discrete and continuous. A typical example is the lifetime of a manufactured product. It has a finite probability, say p_0, of being DOA, i.e. to come put of the factory in a non working condition. Otherwise, it will have a certain probability $f_X(x)\,dx$ of failing between times x and $x + dx$. Its distribution would then look like

$$P[X = 0] = p_0$$

$$P[0 < a < t \le b] = \int_a^b f_X(t)\,dt$$

with

$$p_0 + \int_0^\infty f_X(t)\,dt = 1$$

2 Vector RVs

One comment about "jointly distributed RVs" is in order. Suppose we are interested in two RVs, X, Y, and that they are not independent.

The best way to proceed is to consider the pair as a *vector RV*, taking values in $\mathbb{R}^2$. Hence, we will be considering events of the form

$$\mathbb{Z} = (X, Y) \in A$$

where A is a (reasonable) subset of $\mathbb{R}^2$. As in the case of a single RV, we would limit ourselves to the absolutely continuous case, i.e., when there is a function $f_{X,Y}(x, y)$ such that

$$P[(X, Y) \in A] = \int\int_A f_{X,Y}(x, y)\,dxdy \tag{1}$$

Note that for double (and, with obvious extensions, triple, quadruple,...) integrals we need, not only nice fs, but also nice As! In fact, we could probably only deal with domains A that are "normal with respect one of the axes" (as you learned in multi-variable calculus). Of course, this is true only because we are restricted to Riemann integration. After a course in advanced real analysis, and Lebesgue integration, most of the restrictions (and all the unnatural ones) will fall by the wayside.

The integral in (1) is often also written

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_A(x, y)\, f_{X,Y}(x, y)\,dxdy$$

where $1_A$is the indicator function of the set A. Suppose now that A is a "rectangle", i.e. $A = [a, b] \times [c, d]$, so that (1) can be written as

$$\int_a^b \int_c^d f_{X,Y}(x, y)\,dxdy$$

(the case when A is an "infinite rectangle", i.e. one or more of its vertexes is $\pm\infty$ is treated in the same way). Such an integral is an essentially double integral and, with some restrictions, would be computed in two steps: first, compute, say,

$$g(y) = \int_a^b f_{X,Y}(x, y)\, dx$$

and then

$$\int_c^d g(y)\, dy$$

In the extremely special case when $f_X(x, y) = f_1(x) f_2(y)$, however, the integral would break up as

$$\int_a^b f_1(x)\, dx \cdot \int_c^d f_2(y)\, dy = P[a \le X \le b]\, P[c \le Y \le d]$$

If this is the case for all rectangles, the two variables X, Y (the components of our vector variable $\mathbb{Z}$) are called (as you would expect) *independent.*

Notice that it is not enough for the *analytic form* of the "joint density" $f_{X,Y}$ to be in product form. The product must hold over rectangles. A famous counterexample (bane of students of probability everywhere) is the joint density

$$f(x, y) = \begin{cases} 2 & 0 \le x \le y \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

Now,

$$P[0 \le X \le 1, 0 \le Y \le 1] = \int_0^1 \int_0^1 f(x, y)\, dxdy =$$

$$= \int_0^1 \int_0^1 1_{0 \le x \le y \le 1}(x, y)\, dxdy = \int_0^1 dy \int_0^y 2dx = \left[\frac{2y^2}{2}\right]_0^1 = 1$$

as it should, but

$$\int_0^1 \int_0^1 2dxdy = \int_0^1 dy \int_0^1 2dy = 2$$

I.e., the two variables X, Y are ***not*** independent at all. This can be understood by some reflection: given the structure of the density, we know that certainly $X \le Y$, so knowledge of the value of, say, Y, gives plenty of information about X.

3 What Happened to Conditional Probabilities?

Conditioning is just an essential tool in treating continuous RVs as it is for discrete ones. However, things become really technically challenging if we try to keep the maximum possible generality. In fact, proper definition of conditioning requires some serious real analysis.

As usual, we will restrict to a very special case: that's when the joint density $f_{X,Y}$ is nice, and jointly continuous (you'll remember that "continuity" for functions of several variables has to mean "joint continuity": separate continuity in each variable is almost completely useless).

So we will try to define what it should mean to say

$$P\left[X \in I \,|Y\right]$$

First, we argue as follows: we suppose we have "observed" Y. I.e., $Y = y$. Now, we know that, as written, this statement means little - we really mean $Y \simeq y$, or $y - \varepsilon < Y \leq y + \varepsilon$, or something like that.

Thus we can write

$$P\left[X \in I \,|y - \varepsilon < Y < y + \varepsilon\right]$$

and, more or less thinking in terms of approximating discrete **RV**s, write that this is equal to

$$\frac{P\left[\{X \in I\} \cap \{y - \varepsilon < Y < y + \varepsilon\}\right]}{P\left[\{y - \varepsilon < Y < y + \varepsilon\}\right]}$$

The numerator is, by definition

$$\int_I dx \int_{y-\varepsilon}^{y+\varepsilon} f_{X,Y}\left(x,y\right) dy \simeq \int_I f_{X,Y}\left(x,y\right) 2\varepsilon dx$$

(using one of the mean value theorems), while the denominator is the probability that Y is near y, while X is anywhere, and can be considered in terms of *the density of* Y, as a single **RV**:

$$\int_{-\infty}^{\infty} dx \int_{y-\varepsilon}^{y+\varepsilon} f_{X,Y}\left(x,y\right) dy = \int_{y-\varepsilon}^{y+\varepsilon} \int_{-\infty}^{\infty} f_{X,Y}\left(x,y\right) dx dy =$$

$$= \int_{y-\varepsilon}^{y+\varepsilon} f_Y\left(y\right) dy \simeq f_Y\left(y\right) 2\varepsilon$$

where

$$f_Y\left(y\right) = \int_{-\infty}^{\infty} f_{X,Y}\left(x,y\right)$$

Combining, we arrive at

$$\frac{\int_I f_{X,Y}\left(x,y\right)}{f_Y\left(y\right)} dx$$

which is usually rewritten symbolically as

$$\int_I f_{X|Y}\left(x|y\right) dx$$

with

$$f_{X|Y}\left(x|y\right) = \frac{f_{X,Y}\left(x,y\right)}{f_Y\left(y\right)}$$

defined as the “conditional density of X, given $Y = y$.

Note that in the special case of independent RVs,

$$f_{X,Y}(x, y) = f_X(x)\, f_Y(y)$$

and

$$f_{X|Y}(x|y) = f_X(x)$$

which is very much like the similar statement for discrete RVs.