Math/Stat 251 Fall 2015
Summary of Continuous Random Variables

A random variable is called *continuous* if there exists a function $f : \mathbb{R} \to [0, \infty)$ such that

$$\mathbf{P}\left(X \in B\right) = \int_B f(x)\,\mathrm{d}x$$

for every set $B \subseteq \mathbb{R}$. We call f the *density function* of X. Sometimes we might write $f(x) = f_X(x)$ to stress that this is the density function of the random variable X.

One particularly important choice of B is the following. If $B = (-\infty, x]$, then we call

$$\mathbf{P}\left(X \in (-\infty, x]\right) = \mathbf{P}\left(X \le x\right) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u$$

the *distribution function* of X and write

$$F(x) = \mathbf{P}\left(X \le x\right).$$

Sometimes we might write $F(x) = F_X(x)$ to stress that this is the distribution function of the random variable X. The distribution function and the density function are related by the fundamental theorem of calculus, namely

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x} F(x).$$

Remark. The point-of-view of probability is that we want to analyze the situation before the experiment takes place. This means that we will never actually observe the value of X. Instead, the best we can do is describe the probabilities associated with the likelihoods of the various possible values of X. This is what the density function tells us. Areas under the density curve correspond to probabilities.

Six particularly important families of random variables that we study have names. They are the uniform, exponential, normal, gamma, beta, and Cauchy families.

Instead of just computing probabilities for the likelihoods of various values of X, we might be interested in some sort of "average" value of X. The *median* of X is defined to be that value of a such that

$$\mathbf{P}\left(X \le a\right) = \frac{1}{2} \quad \text{or equivalently} \quad \int_{-\infty}^{a} f(x)\,\mathrm{d}x = \frac{1}{2}.$$

Visually, the median is the value a for which half the area under the density curve is to the left of a and half the area is to the right of a. Beyond computing the median, however, it is a hard object to analyze theoretically. Instead, we might opt for the *mean*, or *expected value*, of X as our measure of "average." The mean of X is defined as

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x.$$

This is a "natural" definition if we view the average of a continuous random variable as some sort of continuum limit of the weighted average of a discrete random variable. In addition, for any positive integer k we can define the kth moment of X to be

$$\mathbb{E}(X^k) = \int_{-\infty}^{\infty} x^k f(x)\,\mathrm{d}x.$$

The moments that concern us in Math/Stat 251 are the first moment, i.e., the mean $\mathbb{E}(X)$, and the second moment, i.e., $\mathbb{E}(X^2)$. We define the *variance* of X to be

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,\mathrm{d}x$$

where $\mu = \mathbb{E}(X)$. It is sometimes easier to compute the variance by the equivalent formula

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \mathbb{E}(X^2) - \mu^2.$$

The square root of the variance is called the *standard deviation* and is denoted by σ so that $\sigma = \sqrt{\mathrm{Var}(X)}$. If we think of the mean as the "centre of mass" of the density, then the standard deviation measures the "spread" of the density, or equivalently, how tightly the density is clustered around the mean. Chebychev's inequality makes this precise, namely if $k > 0$, then

$$\mathbf{P}\left(|X-\mu| \le k\sigma\right) \ge 1 - \frac{1}{k^2}.$$

(Usually we think of k as a positive integer, but it does not necessarily have to be so.) Pictorially, if we draw the density curve, mark the numbers μ, $\mu + k\sigma$, and $\mu - k\sigma$ on the horizontal axis, then the area under the curve between $\mu - k\sigma$ and $\mu + k\sigma$ is at least $1 - k^{-2}$.

The *moment generating function* is a tool for computing the moments of X. We define the moment generating function of X by

$$m(t) = \mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,\mathrm{d}x.$$

Sometimes we might write $m(t) = m_X(t)$ to stress that this is the moment generating function of the random variable X. Taking successive derivatives of m and evaluating at $t = 0$ yields the following result, namely

$$m^{(k)}(0) = \frac{\mathrm{d}^k}{\mathrm{d}t^k} m(t)\bigg|_{t=0} = \mathbb{E}(X^k).$$

Very often we might care about a function of a random variable, say $Y = g(X)$. Using just the density function of X, namely $f_X(x)$, we can compute the expected value of $g(X)$ using the law of the unconscious statistician, namely

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x. \qquad (*)$$

However, we might also want to know the density function of Y. Here is the standard technique for doing such. This is one place where the distribution function is useful. We want the density of Y. We know that $f_Y(y) = F_Y'(y)$. We also know that by definition

$$F_Y(y) = \mathbf{P}\left(Y \le y\right).$$

Substitute in $Y = g(X)$, solve for X, and use the fact that you now have an expression for a probability involving X that can be given in terms of an integral of the density function of X. This integral expression can then be differentiated using the fundamental theorem of calculus to produce $f_Y(y)$. For instance, if g is an increasing function, then

$$F_Y(y) = \mathbf{P}\left(Y \le y\right) = \mathbf{P}\left(g(X) \le y\right) = \mathbf{P}\left(X \le g^{-1}(y)\right) = \int_{-\infty}^{g^{-1}(y)} f_X(x)\,\mathrm{d}x.$$

(The third equality relies on the fact that g is increasing.) Next,

$$f_Y(y) = F_Y'(y) = \frac{\mathrm{d}}{\mathrm{d}y}\int_{-\infty}^{g^{-1}(y)} f_X(x)\,\mathrm{d}x = f_X(g^{-1}(y))\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y).$$

(This formula relies on the fact that g is increasing. In any given exercise, don't memorize this formula, just solve for X in the distribution function and take derivatives carefully using the fundamental theorem of calculus.) Having the density of Y, we can actually compute $\mathbb{E}(Y)$ directly, namely

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,\mathrm{d}y. \qquad (**)$$

Since $Y = g(X)$, it will always be the case that the numbers given by $(*)$ and $(**)$ are the same. In fact, the equality of these expressions is just the first-year calculus substitution method, or change-of-variables method, in disguise.

Applications of Continuous Random Variables

In industrial applications, we are often concerned with the "first failure time" or the "last failure time" of independent components. Suppose that $X_1, X_2, \ldots, X_n$ are independent, continuous random variables. If $Y = \max\{X_1, \ldots, X_n\}$ and $Z = \min\{X_1, \ldots, X_n\}$, then we are interested in computing probabilities for Y and Z. To do so, we need their density functions. To find their density functions, we use the standard technique of first finding the distribution function, and then differentiating to find the density function.

For the maximum Y, we find

$$\begin{aligned} F_Y(y) = \mathbf{P}\left(Y \le y\right) = \mathbf{P}\left(\max\{X_1, \ldots, X_n\} \le y\right) &= \mathbf{P}\left(X_1 \le y, \ldots, X_n \le y\right) \\ &= \mathbf{P}\left(X_1 \le y\right) \cdots \mathbf{P}\left(X_n \le y\right). \end{aligned}$$

Since we know the density function of each X_i, we can compute $\mathbf{P}\left(X_i \le y\right)$ for each i. This then gives us the required expression for the distribution function of Y which can be differentiated to yield the density function of Y.

For the minimum Z, we start with

$$F_Z(z) = \mathbf{P}(Z \le z)$$

as always. However, this expression is not easy to manipulate for the minimum. Instead, we need this "trick" (which is no longer a trick for you since it is the same thing every time you have a minimum). Write

$$\begin{aligned} F_Z(z) = \mathbf{P}(Z \le z) = 1 - \mathbf{P}(Z > z) = &= 1 - \mathbf{P}(\min\{X_1, \dots, X_n\} > z) \\ &= 1 - \mathbf{P}(X_1 > z, \dots, X_n > z) \\ &= 1 - \mathbf{P}(X_1 > z) \cdots \mathbf{P}(X_n > z). \end{aligned}$$

Since we know the density function of each X_i, we can compute $\mathbf{P}(X_i > z)$ for each i. This then gives us the required expression for the distribution function of Z which can be differentiated to yield the density function of Z.

As a final, though related, application of continuous random variable, we might be interested in computing $\mathbf{P}(X > Y)$ when X and Y are independent continuous random variables. The general technique is to use the continuous version of the law of total probability by conditioning on the value of either X or Y. For instance, conditioning on the value of Y gives

$$\mathbf{P}(X > Y) = \int_{-\infty}^{\infty} \mathbf{P}(X > Y \mid Y = y) \, f_Y(y) \, \mathrm{d}y = \int_{-\infty}^{\infty} \mathbf{P}(X > y) \, f_Y(y) \, \mathrm{d}y$$

where

$$\mathbf{P}(X > y) = \int_y^{\infty} f_X(x) \, \mathrm{d}x.$$

In other words,

$$\mathbf{P}(X > Y) = \int_{-\infty}^{\infty} \int_y^{\infty} f_X(x) f_Y(y) \, \mathrm{d}x \, \mathrm{d}y.$$

Equivalently, we can condition on the value of X in which case we find

$$\begin{aligned} \mathbf{P}(X > Y) = \int_{-\infty}^{\infty} \mathbf{P}(X > Y \mid X = x) \, f_X(x) \, \mathrm{d}x &= \int_{-\infty}^{\infty} \mathbf{P}(Y < x) \, f_X(x) \, \mathrm{d}x \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{x} f_Y(y) \, \mathrm{d}y \right] f_X(x) \, \mathrm{d}x \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{x} f_Y(y) f_X(x) \, \mathrm{d}y \, \mathrm{d}x. \end{aligned}$$