Math 191 Notes, 2003 October 2

Distribution Functions

Properties:

1. $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to\infty} F(x) = 1$.

   - Survivor scenario, choose random angle, X, uniformly in $(-\frac{\pi}{2}, \frac{\pi}{2})$. What is $F(y)$ for the y-coordinate when reaching the river?

     $$\mathbb{P}(X \le x) = \frac{x + \pi/2}{\pi}$$

     Convert this into a probability for the other random variable which we have chosen to call Y. Since $Y = \tan X$

     $$\mathbb{P}(Y \le y) = \mathbb{P}(X \le \arctan y) = \frac{\arctan y + \pi/2}{\pi} = \frac{1}{2} + \frac{1}{\pi}\arctan y$$

     "Cauchy" (what happens with the square is that it diverges, so in that case there would be no answer).

   - Increasing Sequence

     $$A = \{\omega \in \Omega | X(\omega) \le n\}$$

     $$A_1 \subset A_2 \subset A_3 \cdots$$

     $$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} = \Omega$$

     $$\lim_{n\to\infty} \mathbb{P}(A_n) = 1 = \mathbb{P}(A)$$

     But everything here must be done using countable sequences of disjoint sets, and this works with limits only if we know (as we do) that the distribution function is non-decreasing.

2. If $x < y$, then $F(x) \le F(y)$.

   $$F(y) \quad = \quad \mathbb{P}(X \le y) = \mathbb{P}(X \le x) + \mathbb{P}(x < X \le y)$$

3. $F(x)$ is right-continuous. That is,

   $$\lim_{x\downarrow 0} F(x+h) = F(x)$$

   Proof. Set up a decreasing sequence of sets $B_n = \{\omega \in \Omega | X(\omega) \le x + \frac{1}{n}\}$. $\lim B_n = B = \{\omega \in \Omega | X(\omega) \le x\}$. By general principles proved earlier, $\lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}(B)$.

- Let's see what happens if we try left-continuous (what goes wrong?). Increasing sequence:
$$A_n = \{\omega \in \Omega | X(\omega) \leq x - \frac{1}{n}\}$$
But x is not included so we will have jumps and bad stuff.

Discrete and Continuous Random Variables

Misnomer: there are lots of random variables that do not fall into either class, but these are two that are easy to handle with regular calculus. (This appears to have nothing to do with probability).

Discrete X takes value in some countable set (say, rolling two dice). Then the "mass function" $p(x) = \mathbb{P}(X = x)$. Once we have the mass function, we can rebuild the distribution function by saying that

$$F(x) = \sum_{x_i \leq x} p(x)$$

In the single variable case, going back and forth is easy. When you have at least two, it's much trickier.

Continuous There exists a density function $f(u)$ such that I can reconstruct the distribution function by integrating the density function.

$$F(x) = \int_{-\infty}^{x} f(u) du$$

Note that this defines $f(u)$ on all points except a finite number of discontinuities.

Singular Distribution for the Schwarzkopf attack. See figure.

Countable positive probability points

There are only countably many values of x for which $p(X = x) > 0$ is possible.
Proof: Look at the points for which $1/2 < \mathbb{P}(x) \leq 1$. What is the maximum number of points for which a random variable can have that value with probability greater than 1/2? 1. More generally, the maximum number of points you can have for the interval $\frac{1}{n+1} < \mathbb{P}(x) \leq \frac{1}{n}$ is n. Thus we have a countable union of countable sets, so the big set is countable.

Examples with darts

Most dart boards have radius 3 and area 9π in some arbitrary units. Assume dart hits board with all loations equally likely. Define random variable $R(\omega)$ = distance from the center (bullseye). What's the distribution function $F(R) = \mathbb{P}(R \leq r)$? It depends not on the radius, but the square of the radius, which means that the distribution function between 0 and 3 is a parabola. If we were to differentiate this to get the density function, which is linear. When this sort of thing happens in physics, it is called a "Phase space effect." But when you do this, you have to be very careful with what your primary assumption is.

Now, what would be a fair scoring scheme? Well, that's a random variable! Well, the distribution function for that random variable is a multi-step function.

"Random vectors"

A random vector is a sequence of random variables, for example (X_1, X_2), say these are the horizontal and vertical components of where the dart hits the dart board. More generally, consider $(X_1, X_2, \ldots, X_n)$.

The book has introduced convenient but perilous notation. In general, if $\vec{x} \leq \vec{y}$ means $(x_1 \leq y_1, x_2 \leq y_2, \ldots)$. This is perilous because the set is not well-ordered (not every pair x, y can be written that way). The convenience is that we can say that the (Joint) distribution for a random vector:

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y)$$

This is sufficient information to deduce the probability of hitting any specific region on the dart board.

$$F(\vec{x}) = \mathbb{P}(\vec{X} \leq \vec{x})$$

Discrete case, 2-component $\mathbb{Z}$-vector

Grid starting in bottom left corner $= (1, 1)$. $F(x, y) = \mathbb{P}($ 1st roll $\leq x$, 2nd roll $\leq y)$. What is the mass function? That is, what is $p(x, y) = \mathbb{P}(X = x, Y = y)$? The continuous case, we have the answer for the unbounded rectangles, and we want the answer for small discrete, bounded rectangles. Simply use inclusion-exclusion.

$$A : \{X < x, Y \leq y\}$$

$$B : \{X \leq x, Y < y\}$$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

$$\mathbb{P}(X \leq x, Y \leq y) - \mathbb{P}(X = x, Y = y) =$$

$$\mathbb{P}(X < x, Y \leq y) + \mathbb{P}(X \leq x, Y = y) - \mathbb{P}(X < x, Y < y)$$