STA 623 – Fall 2011 – Dr. Charnigo

Section 1.5: Distribution Functions

Cumulative distribution function. Let X be a random variable. The cumulative distribution function of X is denoted by $F(x)$ and defined as $P(X \leq x)$. Note that (lower case) x is a placeholder for 2 or 8 or any other real number. For instance, $F(2) = P(X \leq 2)$ and $F(8) = P(X \leq 8)$.

Example (cumulative distribution function). Suppose that there are two children in a family, not twins, and each child has a 50% chance of inheriting a disease. Since the children are not twins, we may assume that whether the first child inherits the disease is independent of whether the second child inherits the disease. In this case, the probability that neither child inherits the disease is $0.5 \times 0.5 = 0.25$, the probability that both children inherit the disease is also $0.5 \times 0.5 = 0.25$, and the probability that exactly one child inherits the disease is $0.5 = 1 - 0.25 - 0.25$. Let X denote the number of children who inherit the disease. What is $F(x)$?

Properties of cumulative distribution functions. A function $F(x)$ is a cumulative distribution function if and only if the following three conditions hold.

1. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.
2. $F(x)$ is nondecreasing in x.
3. $F(x)$ is right continuous in x.

By right continuous we mean that, for any x, $\lim_{\delta \searrow 0} F(x + \delta) = F(x)$.

Example (properties of cumulative distribution functions). Let us verify that condition 2 holds for any cumulative distribution function $F(x)$. Suppose that $x_1 < x_2$. We must show that $F(x_1) \leq F(x_2)$, which is to say that $P(X \leq x_1) \leq P(X \leq x_2)$. Put $A := \{\omega \in S : X(\omega) \leq x_1\}$ and $B := \{\omega \in S : X(\omega) \leq x_2\}$. Then $A \subset B$ and

$$P(X \leq x_1) = P(A) \leq P(B) = P(X \leq x_2).$$

Let us verify that the second part of condition 1 holds for any cumulative distribution function $F(x)$. Condition 2 implies that $\lim_{x\to\infty} F(x)$ exists.

• If the limit is infinite, then we can find x^* such that $F(x^*) > 1$. But $P(X \leq x^*) > 1$ is nonsensical, so the limit cannot be infinite.

• If the limit is finite and greater than 1, call it G, then we can find x^* such that $F(x^*) > (1+G)/2 > 1$. Again, $P(X \leq x^*) > 1$ is nonsensical, so the limit cannot be greater than 1.

• If the limit is finite and less than 1, call it L, then there does not exist x^* such that $F(x^*) > (1+L)/2$. Yet

$$1 = P(X \in \mathbb{R}) = \sum_{j=-\infty}^{\infty} P(X \in (j, j+1]) = \lim_{k\to\infty} \sum_{j=-k}^{k} P(X \in (j, j+1]),$$

so there exists k^* such that $\sum_{j=-k^*}^{k^*} P(X \in (j, j+1]) > (1+L)/2$. But $F(k^*+1) \geq F(k^*+1) - F(-k^*) = \sum_{j=-k^*}^{k^*} P(X \in (j, j+1])$. Putting $x^* := k^* + 1$ generates a contradiction, so the limit cannot be less than 1.

• The only possibility remaining is that the limit equals 1.

Does there exist C such that $1_{\{0\leq x\leq 10\}}Cx + 1_{\{x>10\}}$ is a cumulative distribution function?

Does there exist C such that $1_{\{0\leq x\leq 10\}}C(x^2 - x) + 1_{\{x>10\}}$ is a cumulative distribution function?

Types of random variables. If $F(x)$ is a step function (i.e., piecewise constant except at values x assumed by X with positive probability), then we say that X is discrete. If $F(x)$ is continuous (i.e., both right continuous and left continuous), then we say that X is continuous. If neither description applies, then X is neither discrete nor continuous.

Example (types of random variables). Some people define a discrete random variable as one that can assume only a finite or countably infinite number of values. To understand why this definition is used, let us show that a random variable cannot assume positive probability at an uncountably infinite number of values:

Suppose that $\mathcal{X}$ is an uncountably infinite set such that $P(X = x) > 0$ for every $x \in \mathcal{X}$. For $j \in \{2, 3, \ldots\}$, let $\mathcal{X}_j$ denote the subset of $\mathcal{X}$ such that $1/j < P(X = x) \leq 1/(j-1)$ for $x \in \mathcal{X}_j$. Since $\mathcal{X}$ is uncountably infinite, there exists $j \in \{2, 3, \ldots\}$ such that $\mathcal{X}_j$ is uncountably infinite. Let $\mathcal{X}_j^*$ be a countably infinite subset of $\mathcal{X}_j$. Then $P(X \in \mathcal{X}_j^*) = \sum_{x \in \mathcal{X}_j^*} P(X = x) \geq \sum_{x \in \mathcal{X}_j^*} 1/j = \infty$, which is nonsense. So, no such $\mathcal{X}$ exists.

Some people define a continuous random variable as one that can assume any value in a continuum. This definition is not correct, as seen next.

A random variable that can assume any value in the continuum $[0, 2]$ but that is neither discrete nor continuous is one whose cumulative distribution function is $1_{\{0 \leq x < 2\}} 0.25x + 1_{\{x \geq 2\}}$. Such a random variable may describe the length of time that a student spends on a two-hour in class examination. There is a 50% chance that the student will spend the full two hours.

A continuous random variable X has $P(X = x) = 0$ for any $x \in \mathbb{R}$. (Is this clear?) So we have $0 = \sum_{x \in \mathbb{R}} P(X = x) \neq P(X \in \mathbb{R}) = 1$. Does that contradict any property we have claimed for random variables?

Identically distributed. We say that two random variables X and Y are identically distributed if they share the same cumulative distribution function.

Example (identically distributed). Suppose that X has cumulative distribution function $\pi^{-1}\arctan(x) + 1/2$. Let $Y := -X$. Then

$$\begin{aligned} P(Y \leq x) &= P(X \geq -x) \\ &= 1 - P(X < -x) \\ &= 1 - P(X \leq -x) \\ &= 1/2 - \pi^{-1}\arctan(-x) \\ &= 1/2 + \pi^{-1}\arctan(x). \end{aligned}$$

The third equality uses the fact that $P(X = -x) = 0$. The fifth equality uses the fact that arctangent is an odd function. Hence, X and Y are identically distributed.

On the other hand, X and Y are clearly not the same. In fact, $P(X = Y) = P(X = 0) = 0$.