Random Variables and Distribution Functions

September 9, 2009

Definition 1. *A* **random variable** *is a real valued function from the sample space.*

$$X : S \to \mathbb{R}.$$

Generally speaking, we shall use capital letters near the end of the alphabet, e.g. X, Y, Z for random variables. The range of a random variable is sometimes called the **state space**.

For any event A, an easy to define random variable is the **indicator function**

$$I_A(s) = \begin{cases} 1 & \text{if } s \in A, \text{ and} \\ 0 & \text{if } s \notin A. \end{cases}$$

Exercise 2. *Roll a die three times and consider the sample space* $S = \{(x,y,z); x,y,z = 1,2,3,4,5,6\}$ *and give some random variables on* S.

Exercise 3. *Flip a coin 10 times and consider the sample space* $S = \{H,T\}^{10}$ *and give some random variables on* S.

We often create new random variable via composition of functions. Thus, if X is a random variable, then so is

$$X^2, \quad \exp \alpha X, \quad \sqrt{X^2+1}$$

and so on.

1 Distribution Functions

Definition 4. *A* **(cumulative) distribution function** *of a random variable* X *is defined by*

$$F_X(x) = P\{s \in S; X(s) \le x\} = P\{X \le x\}.$$

For random variables we often interchangeably write

$$\{X \in B\} = \{s \in S; X(s) \in B\} = X^{-1}(B).$$

For the complement

$$P\{X > x\} = 1 - P\{X \le x\} = 1 - F_X(s).$$

Choose $a < b$, then the event $\{X \le a\} \subset \{X \le b\}$. Their set theoretic difference $\{X \le b\} \backslash \{X \le a\} = \{a < X \le b\}$ and consequently

$$P\{a < X \le b\} = P(\{X \le b\} \backslash \{X \le a\}) = P\{X \le b\} - P\{X \le a\} = F_X(b) - F_X(a).$$

Exercise 5. *If $P(A) = p$, give the cumulative distribution function for I_A.*

Exercise 6. *Let X be the sum of the values for two rolls of a die. Give the cumulative distribution function for X.*

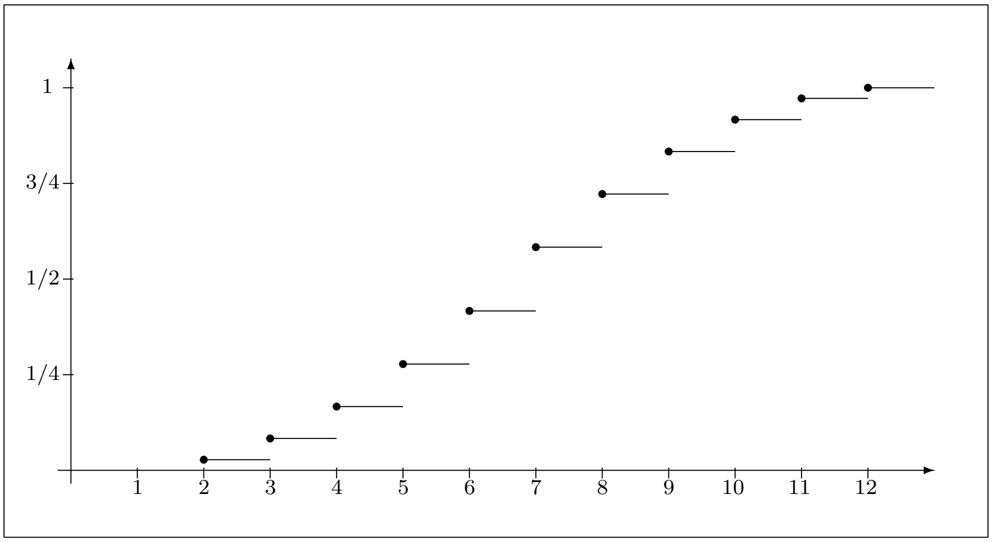

Figure 1: Graph of F_X, the cumulative distributtion function for the sum of the values for two rolls of a die.

If we look at the graph of the cumulative distribution function, we see that it is constant in between the possible values for X and that the jump size at x is equal to $P\{X = x\}$.

In the example,

$$P\{X \leq 5\} = P\{X = 2\} + P\{X = 3\} + P\{X = 4\} + P\{X = 5\} = \sum_{x \leq 5} P\{X = x\}.$$

We shall call a random variable **discrete** if it has a finite or countably infinite state space. Thus, we have in general that:

$$P\{a < X \leq b\} = \sum_{a < x \leq b} P\{X = x\}.$$

Exercise 7. *Let X be the number of heads on three independent flips of a biased coin that turns ups heads with probability p. Give the cumulative distribution function for X.*

Exercise 8. *Let X be the number of spades in a collection of three cards. Give the cumulative distribution function for X.*

2 Properties of the Distribution Function

We shall use the following variant on the countably additivity of P.

If $A_1 \subset A_2 \subset \cdots$ and $A = \cup_{n=1}^{\infty} A_n = A$, then

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

Similarly, if If $B_1 \supset B_2 \supset \cdots$ and $B = \cap_{n=1}^{\infty} B_n = B$, then

$$\lim_{n\to\infty} P(B_n) = P(B).$$

Also, we can determine limits by sequences:

$$\lim_{x\to x_0} g(x) = L \quad \text{if and only if} \quad \lim_{n\to\infty} g(x_n) = L$$

for all sequences $\{x_n; \geq 1\}$ that converge to x_0.

A distribution function F_X is characterized by:

1. F_x is nondecreasing.
   Let $x_1 < x_2$, then
   $$\{X \leq x_1\} \subset \{X \leq x_2\}$$
   $$P\{X \leq x_1\} \leq P\{X \leq x_2\}, \quad F_X(x_1) \leq F_X(x_2)$$
2. $\lim_{x\to\infty} F_X(x) = 1$.
   Let $x_n \to \infty$ be an increasing sequence. Then
   $$x_1 < x_2 < \cdots$$
   $$\{X \leq x_1\} \subset \{X \leq x_2\} \subset \cdots, \quad \bigcup_{n=1}^{\infty} \{X \leq x_n\} = S.$$
   $$P\{X \leq x_1\} \leq P\{X \leq x_2\} \leq \cdots.$$
   Now use the continuity identity (1).
3. $\lim_{x\to-\infty} F_X(x) = 0$.
   Let $x_n \to -\infty$ be a decreasing sequence. Then
   $$x_1 > x_2 > \cdots$$
   $$\{X \leq x_1\} \supset \{X \leq x_2\} \supset \cdots, \quad \bigcap_{n=1}^{\infty} \{X \leq x_n\} = \emptyset.$$
   $$P\{X \leq x_1\} \geq P\{X \leq x_2\} \geq \cdots.$$
   Now use the continuity identity (2).
4. F_X is right continuous, i.e., $\lim_{x\to x_0+} F_X(x) = F_X(x_0)$. Let $x_n \to x_0$ be a decreasing sequence. Then
   $$x_1 > x_2 > \cdots,$$
   $$\{X \leq x_1\} \supset \{X \leq x_2\} \supset \cdots, \quad \bigcap_{n=1}^{\infty} \{X \leq x_n\} = \{X \leq x_0\}.$$
   $$P\{X \leq x_1\} \geq P\{X \leq x_2\} \geq \cdots.$$
   Again use the continuity identity (2).

Although we will not show it here. If a function F has properties 1-4, then we can give a random variable with dstribution function F

In addition, $F_X(x_0-) = \lim_{x \to x_0-} F_X(x)$ exists. To see this, let $x_n \to x_0$ be an increasing sequence. Then

$$x_1 < x_2 < \cdot$$

$$\{X \le x_1\} \subset \{X \le x_2\} \subset \cdots, \quad \bigcup_{n=1}^{\infty} \{X \le x_n\} = \{X < x_0\}.$$

and use continuity identity (1). Notice that the jump in F at x_0,

$$F_X(x_0) - F_X(x_0-) = P\{X \le x_0\} - P\{X < x_0\} = P(\{X \le x_0\} \backslash \{X < x_0\}) = P\{X = x\}.$$

3 Continuous Random Variables

Consider a dartboard having unit radius.

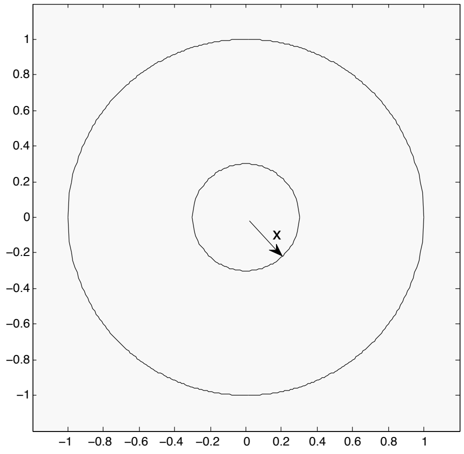

Figure 2: A dart board

Let X be the distance from the center and set, for $x \in [0, 1]$,

$$F_X(x) = P\{X \le x\} = \frac{\text{area inside circle of radius } x}{\text{area or circle}} = \frac{\pi x^2}{\pi 1^2} = x^2.$$

Thus, we have

$$F_X(x) = \begin{cases} 0 & \text{if } x \le 0, \\ x^2 & \text{if } 0 < x \le 1, \\ 1 & \text{if } x > 1. \end{cases}$$

Exercise 9. *An* **exponential random variable** X *has cumulative distribution function*

$$F_X(x) = P\{X \leq x\} = 1 - \exp(-x/\beta)$$

for some $\beta > 0$. *Show that* F_X *has the properties of a distribution function.*

Exercise 10. *The time until the next bus arrives is a exponential random variable with* $\beta = 10$ *minutes. A person waits for a bus at the bus stop until the bus arrives, giving up if when the wait reaches 20 minutes. Give the cumulative distribution function for* T *the time that the peson remains at the bus station and sketch a graph.*