Discrete Random Variables

October 15, 2009

Let S be a sample space. Recall that if $A \subseteq S$ is an event, $P(A)$ denotes the probability of A. (Probability is a function on *events*.)

0.1 Random Varaibles, Probablity Mass Function, Expectation and Variance.

In this discussion, we will assume that S is discrete (i.e., finite or countable). We shall see later (i.e., in Chapter 6) that this assumption is not essential—that is to say, we can develop a perfectly sound theory of random variables without it. However, the assumption is often satisfied in practice. When this is true, it is possible to compute the probabilities of all events from the probabilities of the outcomes they contain (Fact 1, below). This simplifes some proofs (e.g., Fact 3, below.)

Fact 1. If A is an event, $P(A) = \sum_{s \in A} P(\{s\})$

Definition. A function from S to $\mathbb{R}$ is called a *random variable.*

Remark. We really should have mentioned the requirement that X be *measurable*. We will see the need for this assumption later. For the time being, it is not essential to include this stipulation, since all real-valued functions on a discrete set are measurable.

Remark. Note that given *any* sample space (discrete or not), and any random variable, we may make a coarser sample space from S by treating the events of the form $\{\, s \in S \mid X(s) = x \,\}$, $x \in \mathbb{R}$, as outcomes. As long as we ask no questions that concern events that are any finer (i.e., smaller) than these, we lose nothing.

Remark. We say that X is discrete if its set of values $\{\, X(s) \mid s \in S \,\}$ is discrete. Certainly, if S is discrete, then so is X. In case only X is discrete, then by the maneuver in the last remark, we may replace the sample space on which X is defined with a discrete one. This shows that confining attention to discrete sample spaces in not a serious limitation.

Definition. The *probability mass function of X* is

$$p(x) := P(X = x) = P(\{\, s \in S \mid X(s) = x \,\}).$$

Fact 2. $\sum_{x\in\mathbb{R}} p\,(x) = 1$.

Definition. The *expected value of* X is $\mathrm{E}(X) := \sum_{x\in\mathbb{R}} x\, p\,(x)$.

Remark. When S is discrete, $\mathrm{E}(X) = \sum_{s\in S} X(s)\, P(\{s\})$.

Fact 3. If $g : \mathbb{R} \to \mathbb{R}$, then $g\,(X) = g \circ X$ is a random variable, and

$$\mathrm{E}(g\,(X)) = \sum_{s\in S} g\,(X(s))\, P(s) = \sum_{x\in\mathbb{R}} g\,(x)\, p\,(x).$$

Fact 4. If a and b are constants, $\mathrm{E}(a\,X + b) = a\,\mathrm{E}(X) + b$.

Definition. The *variance of* X is $\mathrm{Var}(X) := \mathrm{E}\Big(\big(X - \mathrm{E}(X)\big)^2\Big)$.

Fact 5. $\mathrm{Var}(X) = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2$.

Fact 6. $\mathrm{Var}(a\,X + b) = a^2\,\mathrm{Var}(X)$.

0.2 Binomial Random Variable

Motivating example. Consider the experiment of flipping a coin n times. The sample space is $\{H,T\}^n$; it has 2^n outcomes. Let $p \in [0,1]$. If the coin is biased and lands on heads with probability p, then we would assign probabilities to outcomes by the rule $P(s) = p^i(1-p)^{n-i}$, where i is the number of Hs in s. The probability of an outcome depends only on the number of heads in the outcome, and not the pattern in which they occur. The number of outcomes with i heads is $\binom{n}{i}$, and so the probability of getting i heads is $\binom{n}{i}p^i(1-p)^{n-i}$.

Definition. Let X be a random variable. X is said to be *a binomial random variable with parameters* n *and* p if its values are $0, 1, 2, \ldots, n$ and its probability mass function is

$$p\,(i) = P(X = i) = \binom{n}{i} p^i(1-p)^{n-i}, \quad i \in \{0, 1, 2, \ldots, n\},$$

and $p\,(i) = 0$ for $i \notin \{0, 1, 2, \ldots, n\}$.

Fact. Let X binomial random variable with parameters n and p. Then:

- $\mathrm{E}(X) = n\,p$;
- $\mathrm{Var}(X) = n\,p\,(1-p)$.

0.3 Poisson Random Variable

Motivating example. Consider the experiment of taking a container full of rice from a huge supply that contains a modest number of red grains thoroughly mixed among white ones. We can view each red grain as being on trial, with success meaning inclusion among the taken grains. The the probability of success for any given grain is vanishingly small, but there are a vast number of grains on trial. If on average there are a red grains in every container, then the number of grains taken will vary around a is a way that can be estimated by the binomial distribution (as demonstrated in class).

Definition. Let X be a random variable. X is said to be *a Poisson random variable with parameter* a if its probability mass function is

$$p\,(i) = P(X = i) = \frac{e^{-a}\,a^i}{i\,!}, \quad i \in \{0, 1, 2, \ldots\},$$

and $p\,(i) = 0$ for $i \notin \{0, 1, 2, \ldots\}$.

Fact. Let X Poisson random variable with parameter a. Then:

- $\mathrm{E}(X) = a$;
- $\mathrm{Var}(X) = a$.

Fact. The Poisson random variable with parameter a and the binomial random variable with parameters n (number of flips) and a/n (probability of success) are good approximations of one another if n is very large. Indeed, as $n \to \infty$, the probability mass function associated with this binomial random variable tends to the probability mass function of the Poisson with paramter a.

0.4 Geometric Random Variable

Motivating example. Roll a die until you get a one. More generally, repeat a trial with probablity p of success over and over until a success is achieved. Let X be the number of trials required.

Definition. Let X be a random variable. X is said to be *a geometric random variable with parameter* p if its probability mass function is

$$p\,(i) = P(X = i) = p\,(1 - p)^{i-1}, \quad i \in \{1, 2, \ldots\},$$

and $p\,(i) = 0$ for $i \notin \{1, 2, \ldots\}$.

Fact. Let X be a geometric random variable with parameter p. Then:

- $\mathrm{E}(X) = 1/p$;
- $\mathrm{Var}(X) = \frac{1-p}{p^2}$.

0.5 Negative Binomial Random Variable

Motivating example. Roll a die until you get four ones. More generally, repeat a trial with probablity p of success over and over until r successes are achieved. Let X be the number of trials required.

Definition. Let X be a random variable. X is said to be *a negative binomial random variable with parameters r and p* if its probability mass function is

$$p\,(i) = P(X = i) = \binom{i-1}{r-1} p^{r-1}(1-p)^{i-r}, \quad i \in \{r, r+1, r+2, \ldots\},$$

and $p\,(i) = 0$ for $i \notin \{r, r+1, r+2, \ldots, n\}$.

Fact. Let X be a geometric random variable with parameter p. Then:

- $\mathrm{E}(X) = r/p$;
- $\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$.

0.6 Hypergeometric Random Variable

Motivating example. From an urn containing N beads of which D have a distinguishing property, take n without replacement, where $n \leq \min\{D, N-D\}$. Let X be the number of distinguished beads.

Definition. Let X be a random variable. X is said to be *a hypergeometric binomial random variable with parameters N, D and $n \leq \min\{D, N-D\}$* if its probability mass function is

$$p\,(i) = P(X = i) = \binom{D}{i}\binom{N-D}{n-i}, \quad i \in \{0, 1, 2, \ldots, n\},$$

and $p\,(i) = 0$ for $i \notin \{0, 1, 2, \ldots, n\}$.

Fact. Let X be a geometric random variable with parameter p. Then:

- $\mathrm{E}(X) = \frac{nD}{N}$;
- $\mathrm{Var}(X) = n\,p\,(1-p)\Big(1 - \frac{n-1}{N-1}\Big)$, where $p = D/N$.

0.7 The Socks-in-the-Dryer Random Variable

Remark. This is optional. I put it here because I am obsessed with the problem.

Motivating example. Put n pairs of socks, each a different color, in the dryer. Remove them one at a time until a match is made. Let X_n be the number of socks taken when the match occurs.

Elaboration. The probability of getting a match on the first draw is 0 and on the second draw it is $\frac{1}{2n-1}$. If $k > 1$ socks have been taken without a match, then there are $2n-k$ socks remaining in the dryer, $(2n-k)-k = 2(n-k)$ ways to take *another* sock without making a match, and k ways to make a match. The probability of getting a match on the third draw, therefore, is $\frac{2(n-1)}{2n-1} \cdot \frac{2}{2n-2}$. In general, the probability of getting the first match on the i^{th} draw is

$$\frac{2(n-1)}{2n-1} \cdot \frac{2(n-2)}{2n-2} \cdots \frac{2(n-i+2)}{2n-i+2} \cdot \frac{i-1}{2n-i+1}.$$

The probability mass function of this variable, therefore, is given by

$$p\,(i) = P(X_n = i) = \frac{(i-1)\,2^{i-2}}{2n-i+1} \prod_{j=1}^{i-2} \frac{n-j}{2n-j}, \quad i \in \{2, 3, \ldots, n\},$$

and $p\,(i) = 0$ for $i \notin \{2, 3, \ldots, n\}$. (If the upper index of an indexed product is less than the lower, we take it to be 1.) Remarkably, there is a very simple formula for the expectation of this random variable:

$$\mathrm{E}(X_n) = 4^n \Big/ \binom{2n}{n} = \prod_{i=1}^{n} \frac{2i}{2i-1}.$$

I have not found a nice formula for $\mathrm{Var}(E_n)$, and I do not have a simple demonstration for the formula for $\mathrm{E}(X_n)$.