Discrete Random Variables
(Devore Chapter Three)

1016-351-03: Probability*

Winter 2009-2010

Contents

Tuesday 15 December 2009

Guest lecture by Dr. Chulmin Kim; see Dr. Kim's notes and worksheet

0 Bayes's Theorem

1 Random Variables

1.1 Probability Mass Function

1.2 Cumulative Distribution Function

Practice Problems

3.7, 3.9, 3.13, 3.21, 3.23, 3.25

Thursday 17 December 2009

Guest lecture by Dr. Chulmin Kim; see Dr. Kim's notes and worksheet

2 Expected Values

3 Binomial Distribution

Practice Problems

3.29, 3.33, 3.43, 3.47 (use formula first, then verify w/table), 3.55, 3.59

Tuesday 5 January 2010

4 More on the Binomial Distribution

Recall: if we have a fixed number n of trials, each of which has a chance p of success, then if X is the random variable representing the total number of successes, X is called a binomial random variable, and we write

$$X \sim \mathrm{Bin}(n, p) \tag{4.1}$$

and the pmf of X follows a binomial distribution

$$p(x) = b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \ldots, n \tag{4.2}$$

because

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \tag{4.3}$$

("n choose x") is the number of possible sequences of x successes and $n - x$ failures, and $p^x(1-p)^{n-x}$ is the probability of any one such sequence.

Also, the cdf of X is written

$$F(x) = B(x; n, p) = \sum_{y=0}^{x} b(y; n, p) \tag{4.4}$$

Now, $b(x; n, p)$ and $B(x; n, p)$ are explicitly defined, and you could in principle work them out by hand every time you need one. However, for all but the smallest values of x and $n - x$, that becomes tedious rather quickly. So this is the first of unfortunately many examples in probability and statistics where the standard approach is to look up the distribution in a statistical table. This is something of an anachronism, since these things can be easily calculated on a computer these days. But there is some middle ground between relying entirely on brute force and tables, and letting a statistical software package do your entire analysis for you. It's sort of like trigonometry: before the advent of calculators, you had to interpolate sines and cosines from values in a table. Now you can punch them into your calculator or call a function in a computer program. But you don't rely on a trigonometry package to work out all of your triangles for you. (And you hopefully remember the sines and cosines of a few simple angles like 0°, 30°, 45° and 90°.) So you shouldn't be ashamed to leave the calculating of something like $B(15; 35, .7)$ to a computer, but you should remember where it came from. (And also be able to write a little script in the language of your choice to calculate it.)

4.1 Hypergeometric Distribution

The binomial distribution is appropriate when the probability of each success or failure doesn't depend on the number of successes, like when calculating the probability of getting a total of three sixes when rolling ten six-sided dice. However, it doesn't give the right probabilities when you're taking elements out of a finite population. For example, consider the probability of getting exactly three spades in a five-card poker hand. Although 13 of the 52 cards in the deck are spades, we know from our example of calculating the probability of a flush that it's not just $b(3;5,1/4)$. Instead we need to calculate the number of different hands which have three spades and two non-spades, and divide it by the total number of possible poker hands.

- There are $\binom{13}{3} = \frac{13!}{3!10!}$ combinations of 3 out of the 13 spades.
- There are $\binom{39}{2} = \frac{39!}{2!37!}$ combinations of $5-3=2$ out of the $52-13=39$ non-spades.
- There are $\binom{52}{5} = \frac{52!}{5!47!}$ combinations of 5 out of the 52 cards.

So there are $\binom{13}{3}\binom{39}{2}$ different hands with exactly 3 spades, out of $\binom{52}{5}$ total hands, and the probability is

$$p(3\spadesuit) = \frac{\binom{13}{3}\binom{52-13}{5-3}}{\binom{52}{5}} \tag{4.5}$$

This is a special case of something called the *Hypergeometric Distribution.* If we randomly draw n items out of a set of N, M of which are successes, the number of successes we'll draw is a random variable X whose pmf is

$$h(x;n,M,N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \tag{4.6}$$

We say that the n items are drawn "without replacement". If we drew cards one at a time and shuffled them back into the deck, we would be back in the situation where the probability of success for each card was the same no matter how many spades had already been drawn, and in that case we'd be back to a binomial distrubution. It must also be the case that the hypergeometric distribution reduces to a binomial one when the number of draws is much less than the number of total objects, since in that case the probability of success on each draw shouldn't be effected much by the removal of a few successes or failures from the pool.

We can show this from the formula, that $h(x;n,M,N) \approx b(x;n,M/N)$ when $n \ll N$, $x \ll M$ and $n-x \ll N-M$. We first note that

$$\begin{aligned}\binom{N}{n} &= \frac{N!}{(N-n)!n!} = \frac{N(N-1)\cdots(N-n+1)}{n!} \\ &= \frac{N^n}{n!}\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)\end{aligned} \tag{4.7}$$

In the limit of large N, each of the factors in parentheses goes to 1, so we can replace $\binom{N}{n}$ with $\frac{N^n}{n!}$. If we do the same thing with the other combinations, we get

$$\begin{aligned} h(x; n, M, N) &\approx \frac{\frac{M^x}{x!}\frac{(N-M)^{n-x}}{(n-x)!}}{\frac{N^n}{n!}} = \frac{n!}{x!(n-x)!}\frac{M^x(N-M)^{n-x}}{N^x N^{n-x}} \\ &= \binom{n}{x}\left(\frac{M}{N}\right)^x\left(\frac{N-M}{N}\right)^{n-x} = b\left(x; N, \frac{M}{N}\right) \end{aligned} . \tag{4.8}$$

4.2 Negative Binomial Distribution

Return now to a scenario where the per-trial chance of success is fixed. The binomial distribution is appropriate when the number of trials, n, is fixed ahead of time. If instead we decide that we'll keep doing trials until we have a specified number of successes, r, the pmf for the number of failures, X, that we have in the meantime is $nb(x; r, p)$. Note that this is not n times $b(\cdots)$; the "nb" is taken as a unit. The total number of trials is $X + r$ (which is also a random variable).

This probability is a little more involved to estimate. To get $X = x$, we have to have $r - 1$ successes (and x failures) in the first $x + r - 1$ trials, and then a success in the last trial, so

$$\begin{aligned} nb(x; r, p) = b(r-1; x+r-1, p) \cdot p &= \binom{x+r-1}{r-1} p^{r-1}(1-p)^x p \\ &= \binom{x+r-1}{r-1} p^r (1-p)^x \end{aligned} \tag{4.9}$$

We can have any (non-negative integer) number of failures, so the pmf is defined for $x = 0, 1, 2, \ldots$.

4.3 Summary of Binomial and Related Distributions

The binomial, hypergeometric and negative binomial distributions are summarized in this table:

Distribution	Trials	Successes	Failures	Prob/Success
Binomial	n, fixed	X, rv	$n - X$, rv	p, fixed
Hypergeometric	n, fixed	X, rv	$n - X$, rv	drawn from M succ in pop of N
Negative Binomial	$X + r$, rv	r, fixed	X, rv	p, fixed

Distribution	pmf	Domain
Binomial	$b(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x}$	$0 \le x \le n$
Hypergeometric	$h(x;n,M,N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$\max(0, n-N+M) \le x \le \min(n,M)$
Negative Binomial	$nb(x;r,p) = \binom{x+r-1}{r-1} p^r (1-p)^x$	$0 \le x$

Practice Problems

3.65, 3.69, 3.71, 3.73, 3.75, 3.77

Thursday 7 January 2010

5 Poisson Distribution

Devore introduces the Poisson distribution starting with its pmf and then delving into the situations where it's relevant. Let's come at it from the other side. Consider the numerous statistical statements[1] you get like:

1. On average 118.3 people per day are killed in traffic accidents in the US
2. On average there are 367.2 gamma-ray bursts detected per year by orbiting satellites
3. During a rainstorm, an average of 929.4 raindrops falls on a square foot of ground each minute

In each of these cases, the number of events, X, occurring in one representative interval is a discrete random variable with a probability mass function. The average number of events occurring is a parameter of this distribution, which we traditionally write as λ. From the information given above, we only know the mean value of the distribution, $E(X) = \lambda$.

The key bit of extra information that makes the pmf a Poisson distribution is what happens if you break the interval up into smaller pieces. If each of the pieces can be treated as the same and independent of the others, the Poisson distribution is appropriate. So if we divide the day into 100 equal pieces of 14 minutes 24 seconds each, is the number of traffic deaths in each

[1] I fabricated the last few significant figures in each of these numbers to produce a concrete example.

an independent random variable with a mean value of 11.83? Now, in practice this assumption will often not quite be true: the rate of traffic deaths is higher at some times of the day than others, some patches of ground may be more prone to be rained on because of wind patterns, etc, but we can imagine an idealized situation in which this subdivision works.

Okay, so how do we get the pmf? Take the interval in question (time interval, area of ground, or whatever) and divide it into n pieces. Each one of them will have an average of λ/n events. If we make n really big, so that $\lambda/n \ll 1$, the probability of getting one event in that little piece will be small, and the probability that two or more of them happen to occur in the same piece is even smaller, and we can ignore it to a good approximation. (We can always make the approximation better by making n bigger.) That means that the number of events in that little piece, call it Y, has a pmf of

$$p(Y=0) \approx 1-p \tag{5.1a}$$
$$p(Y=1) = p \tag{5.1b}$$
$$p(Y>1) \approx 0 \tag{5.1c}$$

In order to have $E(Y) = \lambda/n$, the probability of an event occurring in that little piece has to be $p = \lambda/n$.

But we have now described the conditions for a binomial distribution! Each of the n tiny sub-pieces of the interval is like a trial, a piece with an event is a success, and a piece with no event is a failure. So the pmf for this Poisson random variable must be the limit of a binomial distribution as the number of trials gets large:

$$\begin{aligned} p(x;\lambda) &= \lim_{n\to\infty} b(x;n,\lambda/n) = \lim_{n\to\infty} \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \\ &= \frac{\lambda^x}{x!} \lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n \frac{n!}{(n-x)!}\left(\frac{1/n}{1-\lambda/n}\right)^x \end{aligned} \tag{5.2}$$

Now, the ratio of factorials is a product of x things:

$$\frac{n!}{(n-x)!} = n(n-1)(n-2)\cdots(n-x+1) \tag{5.3}$$

The last factor is of course also the product of x things, i.e., x identical copies of

$$\frac{1/n}{1-\lambda/n} = \frac{1}{n-\lambda}\ . \tag{5.4}$$

But that means the two of them together give you

$$\frac{n!}{(n-x)!}\left(\frac{1/n}{1-\lambda/n}\right)^x = \frac{n}{n-\lambda}\frac{n-1}{n-\lambda}\frac{n-2}{n-\lambda}\cdots\frac{n-x+1}{n-\lambda} \tag{5.5}$$

which is the product of x fractions, each of which goes to 1 as n goes to infinity, so we can lose that factor in the limit and get

$$p(x;\lambda) = \frac{\lambda^x}{x!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = \frac{\lambda^x}{x!}e^{-\lambda} \ . \tag{5.6}$$

We've used the exponential function

$$e^\alpha = \lim_{n\to\infty}\left(1+\frac{\alpha}{n}\right)^n \ ; \tag{5.7}$$

This is the form one usually sees in the context of compound interest. [2] Here e is Euler's number, $e = 2.718\ldots$.

Note that in the pmf for a Poisson random variable X,

$$p(x;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda} \tag{5.8}$$

the most daunting part, the exponential, doesn't actually depend on x. It's a normalization constant which depends on the parameter λ. If we only care about the relative probabilities, we could write

$$p(x;\lambda) = \mathcal{N}(\lambda)\,\frac{\lambda^x}{x!} \ ; \tag{5.9}$$

the fact that the constant $\mathcal{N}(\lambda)$ is $e^{-\lambda}$ is required by the normalization

$$\sum_{x=0}^{\infty} p(x;\lambda) = \mathcal{N}(\lambda)\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = \mathcal{N}(\lambda)\,e^{\lambda} \ . \tag{5.10}$$

This uses the Taylor series

$$e^\lambda = \sum_{x=0}^{\infty}\frac{\lambda^x}{x!} \tag{5.11}$$

[2] If the rate times the term is α and we compound the interest at n equally spaced intervlas during the term, the principal grows by a factor of $(1+\alpha/n)^n$; in the limit of continuously compounded interest, the principal doesn't grow infinitely, but only by a factor of e^α.

5.1 Example: Poisson distribution for $\lambda = 2$

We can work out the pmf in this case. The normalization constant is $e^{-2} = .135335$. The pmf values are

$$p(0; 2.) = \frac{2^0}{0!} e^{-2} = .135 \tag{5.12a}$$

$$p(1; 2.) = \frac{2^1}{1!} e^{-2} = .271 \tag{5.12b}$$

$$p(2; 2.) = \frac{2^2}{2!} e^{-2} = .271 \tag{5.12c}$$

$$p(3; 2.) = \frac{2^3}{3!} e^{-2} = .180 \tag{5.12d}$$

$$p(4; 2.) = \frac{2^4}{4!} e^{-2} = .0902 \tag{5.12e}$$

$$p(5; 2.) = \frac{2^5}{5!} e^{-2} = .0361 \tag{5.12f}$$

$$p(6; 2.) = \frac{2^6}{6!} e^{-2} = .0120 \tag{5.12g}$$

$$p(7; 2.) = \frac{2^7}{7!} e^{-2} = .00344 \tag{5.12h}$$

$$p(8; 2.) = \frac{2^8}{8!} e^{-2} = .000859 \tag{5.12i}$$

Note that $p(x; \lambda)$ never goes to zero, but it gets very small. We can also say what the total probability of being above a certain point is, e.g.,

$$\begin{aligned} P(X > 4) = 1 - P(X \leq 4) = 1 - e^{-2} \left(1 + 2 + 2 + \frac{4}{3} + \frac{2}{3}\right) = 1 - 7e^{-2} \\ = 1 - .947 = .053 \end{aligned} \tag{5.13}$$

Practice Problems

3.79, 3.81, 3.83, 3.85, 3.91, 3.93