Chapter 3 – Discrete Probability Models

Sections 3.1-3.3: The tools for describing discrete probability models

3.1 Random variables, X
3.2 Probability Distribution of X
3.3 Expected Value of X

Random variable (p 98 for formal definition) – really just a variable whose values are determined by a "random" process, or in other words, a variable that records the various possible values for the outcomes of an experiment.

e.g. X1=number of heads in 10 tosses of a fair coin
X2=number of visible stars in the sky
Y1=proportion of a hamburger that is fat
Y2=time until the next major earthquake occurs in Vancouver

Discrete random variable: One that has a countable number of values (e.g. X1 or X2)
Continuous random variable: One that has a continuum of values (e.g. Y1, which can be any value in [0,1]; or Y2, which could be any value in {0,1,2,....})

Rest of chapter focuses on models for special discrete random variables.
If the "experiment" we were discussing in Ch 1 & 2 produces values that can be listed (are countable), then we can imagine certain probabilities being attached to each possible value. It is very helpful for analysis when we have a model for the list of probabilities, and it turns out that there are several experiments where this model is known, either exactly or approximately.

Of course, one general model is just to say that the sample space associated with a particular random variable X is given by a table of values of P(X=x):

e.g

X	P(X=x)
0	.25
1	.40
2	.20
3	.15

Any such definition
But there are some random variables where the P(X=x) can be written as a simple function of x. For example, Suppose Y can take values 0,1,2,... with probabilities $P(Y=y)=.1*(.9)^{(y-1)}$ for y=1,2,.... These probabilities do sum to 1 as you can check. What "experiment" would actually produce values of Y with this particular sequence of probabilties? One answer, as you will see (p 133), is that if Y is the number of random selections of the digits {0,1,2,...,9} that it takes until you get "0" for the first time. This is

closely related to a discrete probability model called the geometric distribution – which we will meet again later in this chapter.

The above was to convey what discrete probability models are, and to hint at why they are useful. Now we back up to the simplest models - actually the simplest one is very important: it is the Bernoulli model. This is just a model for the occurrence or non-occurrence of an event: $P(X=x) = p^x(1-p)^{(1-x)}$ where x=0 or 1, and $0 \le p \le 1$. Compare this with the definition on p 99. Note this is equivalent to the tabular definition.

x	P(X=x)
0	(1-p)
1	p

which explains why the apparently complex formula above is actually as general as the definition on p 99.

We need some more general tools to describe discrete distributions – we will use some similar tools for continuous distributions (in Ch 4). The tools are *probability models for random variables,* and *expected values of random variables*. The latter gives a way to describe the location and spread of random variables.

First a note on possible confusion with words to describe random variables.

A random variable X may have a probability distribution P(X=x) which may be described by a table of probabilities (x, P(X=x)) or by a model formula P(X=x)=f(x). Furthermore, if we have several observed values of this random variable, we may describe the distribution of the data. The word "distribution" pops up in many ways. We say X has a certain distribution, or that the distribution of X is given by the table of (x,P(X=x)), or that f(x) is the distribution of X.

Now back to the general tools to describe distributions. Just as in Ch 1 when we were describing data distributions, we can describe a probability distribution by describing its location (or center) and spread. In spite of the discussion of alternative measures of location and spread for descriptive purposes, we usually use mean and standard deviation for characterizing probability distributions. Since we have a formula for the relative probabilities in our models, the mean and standard deviation is usually enough to fix the shape of the whole distribution. For example if we have a model $P(X=x) = p(1-p)^{(1-x)}$ x=1,2,3,..., just knowing the mean is 5 will imply that p = 1/5 (its not obvious but will be shown). The point is that mean and sd can be expressed in terms of the parameters of the model, and the parameters of the model can often be expressed in terms of the mean and sd (for many models).

For any X for which P(X=x) is known (either as numerical probabilities or as known functions of the parameters), you can compute the mean (known as the *expected value* of X and denoted E(X) in this context when the probabilities are stated exactly) from

E(X) = $\sum_x xP(X = x)$ where the summation is intended to be over all values of x that have positive probability. Why is this considered to be a "mean"?

Suppose we have a rv that takes the following values with equal probability {1,1,1,1,2,2,2,2,3,3}. The mean would clearly be the sum/10 = 1.8.
We could have first counted the proportion of the time each value occurs: 1 occurs .4 of the time, 2 also occurs .4 of the time, and 3 occurs .2 of the time. So the average could be computed as [(1 x 4)+(2 x 4)+(3 x 2)]/10 = 1.8, or, bringing the 10 inside, (1 x .4) + (2 x .4) + (3 x .2) = 1.8. This last form is just the formula $\sum_x xP(X = x)$.
So the "expected value of X" or E(X) is the (theoretical) mean of the distribution of X.

Going back to {1,1,1,1,2,2,2,2,3,3}, if I add 3 to each value, how much does this change the mean? Obviously, from 1.8 to 4.8, right? Also, if I multiply each data value by 2, what happens to the mean? 1.8 to 3.6. Now look at the box on the bottom of p 115. So if you know the mean of X in {1,1,1,1,2,2,2,2,3,3}, you don't have to redo the whole calculation if you want to know E(2X +3), the mean of 2X +3. What is it?

Now what if you wanted the mean of X^2, where X is randomly drawn from {1,1,1,1,2,2,2,2,3,3}? Is it 1.8^2? No. (You can check it would be larger than that).
We need to go back to the relative frequency of 1,2,and 3.
$E(X^2) = 1^2 \times .4 + 2^2 \times .4 + 3^2 \times .2 = .4+1.6+1.8 = 3.8$

The general rule for finding means of functions of X is given by the first box on p 115.

Now one very special function of X is called the variance of X. It is defined as the average squared deviation from the mean of X. (p 116).

Going back to {1,1,1,1,2,2,2,2,3,3}, the mean was 1.8, so the deviations are {-.8,-.8,-.8,-.8,.2,.2,.2,.2,1.2,1.2} and the squares of these are {.64,.64,.64,.64,.04,.04,.04,.04,1.44,1.44} and the average of these is 0.56.
(Note "mean" is always the same as "average")

So the variance of X is 0.56.

Of course we could have computed it as $(-.8)^2 \times .4 + (.2)^2 \times .4 + (1.2)^2 \times .2 = 0.56$.
This is the formula in the box p 116.

Now because the variance of X is based on the deviations from the mean, it is reasonable to think of "variance" as a measure of variation. It is in a way, but note that it is not in the same units as the original data. If the original data were in feet, the variance would be in square feet. It would be best to report the variability of the data in the same units as the data. So the measure to do this is the standard deviation which is the square root of the variance. In this example, SD = $(0.56)^{1/2} = 0.75$.

The alternative formula for calculating the variance is in the box on p 117.

In our example, the values of X^2 are {1,1,1,1,4,4,4,4,9,9} and the average of these is 3.8 so by the formula on p 117 we have
variance of $X = 3.8 - (1.8)^2 = 3.80\text{-}3.24 = 0.56$

Now, at last we can get to the meat of this course! A good understanding of a few models will be a powerful tool for predicting the effects of unexplained or uncontrolled variation in measurements. The models we cover in this chapter are the Bernoulli, Binomial, Hypergeometric, Negative Binomial, and the Poisson Distribution – sections 3.4 to 3.6 of this chapter.