Lectures 10 - Discrete Probability Distributions

Text: WMM, Chapter 5. Sections 5.1-5.6

In the previous section, we learned how to compute means and variances of functions, given a probability distribution function, f(x). In this chapter, we introduce some of the common probability distribution functions (PDFs) for discrete sample spaces. The goal of this section is to become familiar with these probability distributions and, when given a word problem, know which PDF is appropriate.

Discrete Uniform Distribution (p. 114)

If the random variable X assumes the values of x_1, x_2, x_3… x_k with equal probability, then the discrete uniform distribution is given by f(x;k) (The semicolon is used to separate variables from parameters.)

$$f(x;k) = \frac{1}{k} \tag{10.1}$$

Mean and Variance of the Uniform Distribution (p. 115)

Note, although we are giving these formulae for the mean and variance. They are derived from equation (7.2) in the previous section. If you would prefer not to memorize these formulae, that's fine, so long as you know how to derive them when needed.

$$\mu = \frac{1}{k}\sum_{i=1}^{k} x_i \tag{10.2}$$

$$\sigma^2 = \frac{1}{k}\sum_{i=1}^{k} (x_i - \mu)^2 \tag{10.3}$$

Example 10.1: You select a card randomly from a mixed deck of cards. What is the probability you draw a king with an axe or a one-eyed jack? (Note, there is only one king with an axe but there are two jacks shown in profile.) If you assign a numeric value of 1 to the ace, 11 to the jack, 12 to the queen, and 13 to the king, what is the mean value of the card drawn. What is the variance?

The probability of drawing a king with an axe or a one-eyed jack is 3/52 by equation (10.1). The mean is 7 by equation (10.2). The variance is 14 by equation (10.3).

Relationship between binomial, multinomial, hypergeometric, and multivariate hypergeometric PDFs

The next four PDFs we are going to discuss are the binomial, multinomial, hypergeometric, and multivariate hypergeometric PDFs. Which of the four PDFs you need to employ for a given problem depends upon two criteria: (1) how many outcomes an experiment can yield, and (2) whether the probability of a particular outcome changes from one trial to the next. Frequently, the change in probability is due to not replacing some element of the experiment. Therefore, this second factor is noted as replacement vs. no replacement. The following table describes when each of the PDFs should be used.

	replacement	no replacement
2 outcomes	binomial	hypergeometric
n>2 outcomes	multinomial	multivariate hypergeometric

Binomial Distribution (p. 116)

The binomial and multinomial distributions arise from a process called the Bernoulli Process. The Bernoulli process is

1. An experiment that consists of n repeated, **independent** trials.
2. Each trial can have one of two outcomes, success or failure.
3. The probability of success, p, is the same for each trial.

Examples of Bernoulli processes:

10.2. Flipping a coin n times. Success = landing heads up. (Each toss is a trial. Each toss is independent. Each toss has one of two outcomes: heads or tails. The probability for heads is the same for each toss.)

10.3. Grabbing a handful of marbles from a bag or red and black marbles, and **replacing** the marbles between grabs. Success = more than m red marbles in hand. (Each grab is a trial. Each grab is independent, so long as there is replacement. Each grab has one of two outcomes: more than m red marbles or less than or equal to m red marbles; success or failure. Sure the number of red marbles varies, but that's not our criterion for success, only more or less than m. The probability for success is the same for each grab.)

Now the random variable, X, in a binomial distribution, b(x; n, p) , is the number of successes from n Bernoulli trials. So for our first example, flipping a coin n times, the probability of a getting a head in one independent trial is p. For n trials, the Binomial random variable can assume values between 0 (never getting a head) up to n (getting a head every time). The distribution gives the probability for getting a particular value of successes in n trials.

The binomial distribution is (where q the probability of a failure is q = 1 - p)

$$P(X = x) = b(x; n, p) = \binom{n}{x} p^x q^{n-x} \tag{10.4}$$

The mean is

$$\mu = np \tag{10.5}$$

The variance is

$$\sigma^2 = npq \tag{10.6}$$

Frequently what we want is the cumulative PDF,

How to Compute probabilities with the Binomial Distribution with a computer code

The easiest way to compute probabilities with the binomial distribution is to write a short code. For example, we can obtain the

$$P(X = x) = b(x; n, p) = \binom{n}{x} p^x q^{n-x}$$

with the following MATLAB code, binomial.m:

```
function f = binomial(x,n,p)
f = comb(n,x)*p^x*(1-p)^(n-x);
```

This two line program accespts as inputs x, n, and p, and returns $f = b(x; n, p)$. This code accesses the program, comb.m, to obtain the combinations. The code for comb.m is given in lecture packet 1.

If we wanted the cumulative binomial PDF, $P(X \le r) \equiv B(r; n, p) = \sum_{x=0}^{r} b(x; n, p)$, then we could write a short code and call it, binocumu.m, which would contain

```
function f = binocumu(r,n,p)
f = 0.0
for x = 0:1:r
     f = f + binomial(x,n,p);
end
```

If we wanted the most general code to calculate the probability from the binomial PDF in some arbitrary interval, then we could write in the file binoprob.m

```
function f = binoprob(a,c,n,p)
f = 0.0
for x = a:1:c
     f = f + binomial(x,n,p);
end
```

This file returns the value of $P(a \le X \le c) = \sum_{x=a}^{c} b(x;n,p)$. In the table below we see how the program, binoprob.m can calculate the probability for any arbitrary interval, given the correct values of a and c. The table does not present a complete set of all the possible combinations but does give the general idea.

probability	command line argument
$P(X = a) = P(a \le X \le a)$	`binoprob(a,a,n,p)`
$P(X \le a) = P(0 \le X \le a)$	`binoprob(0,a,n,p)`
$P(X < a) = P(0 \le X \le a-1)$	`binoprob(0,a-1,n,p)`
$P(X \ge a) = P(a \le X \le n)$	`binoprob(a,n,n,p)`
$P(X > a) = P(a+1 \le X \le n)$	`binoprob(a+1,n,n,p)`
$P(a \le X \le c) = P(a \le X \le c)$	`binoprob(a,c,n,p)`
$P(a < X < c) = P(a+1 \le X \le c-1)$	`binoprob(a+1,c-l,n,p)`
$P(a \ge X \ge c) = P(c \le X \le a)$	`binoprob(c,a,n,p)`
$P(a > X > c) = P(c < X < a)$	`binoprob(c+1,a-l,n,p)`

How to Compute probabilities with the Binomial Distribution with Tables

The probability:

$$P(X \le r) \equiv B(r;n,p) = \sum_{x=0}^{r} b(x;n,p) \qquad (10.7)$$

requires the evaluation of equation (10.4) r times. If n and r are large, this can be a time-consuming calculation, although a simple computer code could be used to this rapidly, as we showed above. In the absence of a computer code, values of B(r;n,p) are given in Table A.1 of the Appendix for a range of values of r, n, and p. There are several disadvantages to relying on tabulated values of the binomial PDF. First, the tables only go up to n=20. For higher values, you have to use a computer code or an approximation. Second, the tables only provide values for p = 0.1, 0.2, 0.3,…0.8,0.9. Of course, p, the probability of success can take on any value between 0 and 1. If p = 0.15, you cannot use the table. Of course, you can linearly interpolate, but that interpolation is an approximation because the binomial PDF is not linear in p. Third, the tables only provide the cumulative PDF, $P(X \le r)$. If you have something else, like $P(a \le X \le c)$ or $P(X > a)$, you have to first rearrange the probability into some combination of cumulative PDFs to read the information from the table.

Example 10.2: Here is an example where we use the table to calculate some probabilities based on the binomial PDF. We flip a coin 20 times. A success is heads. The probability of a success in a single trial is 0.5.

What is the probability that 5 of the 20 tosses are heads?
P(X=5) = 0.014786, using equation (10.4).
What are the average number of heads?
The average number of heads (successes) is 10 from equation (10.5)
What is the variance?
The variance is 5 from equation (10.6)
What is the probability of getting between 8 and 12 heads, inclusive?

$$P(8 \le X \le 12) = P(X \le 12) - P(X < 8)$$

$$P(8 \le X \le 12) = P(X \le 12) - P(X \le 7)$$

$$P(8 \le X \le 12) = B(r = 12; n = 20, p = 0.5) - B(r = 7; n = 20, p = 0.5)$$

$$P(8 \le X \le 12) = 0.8684 - 0.1316 = 0.7368$$

So when you give a coin 20 flips, roughly 70 percent of the time, you will wind up with between 8 and 12 heads, inclusive. Does this mean 30% of the time you will wind up with 8 to 12 tails, inclusive? Why or why not?

Multinomial Distribution

If a Bernoulli trial can have more than 2 outcomes (success or failure) then it ceases to be a Bernoulli trial and becomes a multinomial experiment. In the multinomial experiment, there are k outcomes and n trials. Each outcome has a result E_i. There are now k random variables X_i, each representing the probability of obtaining result E_i in X_i of the n trials.

The distribution of a multinomial experiment is

$$P(\{X = x\}) = m(\{x\}; n, \{p\}, k) = \binom{n}{x_1, x_2 \ldots x_k} \prod_{i=1}^{k} p_i^{x_i} \qquad (10.8)$$

where $\sum_{i=1}^{k} x_i = n$ and $\sum_{i=1}^{k} p_i = 1$.

Example Problem using multinomial PDF:
A hypothetical statistical analysis of people moving to Knoxville, TN shows that 25% of people who move here do so to attend the University, 55% move here for a professional position, and the remaining 20% for some other reason. If you ask 10 new arrivals in Knoxville, why they moved here, what is the probability that all of them moved here to go to UT?

$$P(\{X\} = [10,0,0]) = m([10,0,0];10,[0.25,0.55,0.20],3) = \frac{10!}{10!0!0!} 0.25^{10} 0.55^{0} 0.20^{0} = 9.54e-007$$

How to Compute probabilities with the Multinomial Distribution with a computer code

You can write a code to evaluate $P(\{X = x\}) = m(\{x\};n,\{p\},k)$, such as the one in multinomial.m

```
function prob = multinomial(x,n,p,k)
prob = factorial(n);
for i = 1:1:k
   prob = prob/factorial(x(i))*p(i)^x(i);
end
```

This makes use of a program called factorial.m, which has the following form.

```
function f = factorial(n)
f = 1.0;
for i = n:-1:2
     f = f*i;
end
```

In the code, multinomial.m, x and p are vectors of length k. This code would be run at the command prompt with something like

```
multinomial([2,4,3],9,[0.5,0.3,0.2],3)
```

The vector arguments in multinomial must be enclosed by brackets.

Hypergeometric Distribution p. 126
The hypergeometric distribution applies when

1. A random sample of size n is selected **without replacement** from a sample space containing N total items.
2. k of the N items may be classified as successes and N-k are classified as failures. (Therefore, there are only 2 outcomes.)

The probability distribution of the hypergeometric random variable X, the number of successes in a random sample of size n selected from a sample space containing a total of N items, in which k of N are will be labeled as a success and N - k will be labeled failure is

$$h(x;N,n,k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad \text{for } x = 0,1,2...n \tag{10.9}$$

The mean of the hypergeometric distribution h(x;N,n,k) is

$$\mu = \frac{nk}{N} \tag{10.10}$$

and the variance is

$$\sigma^2 = \left(\frac{N-n}{N-1}\right)\frac{nk}{N}\left(1-\frac{k}{N}\right) \tag{10.11}$$

Example 10.5: What is the probability of getting dealt 4 of a kind in a hand of five-card-stud poker?

We can use the hypergeometric distribution on this problem because we are going to select n=5 from N=52. We don't care about what value of the cards the four of a kind is in so we can just calculate the result for aces and then multiply that probability by 13 since there are 13 values of cards and a four of a kind of any of them are equally likely. Therefore, a success is an ace and a failure is not an ace. k = 4 aces. For a four of kind x = four aces.

$$h(4;52,5,4) = \frac{\binom{4}{4}\binom{52-4}{5-4}}{\binom{52}{5}} = \frac{\binom{4}{4}\binom{48}{1}}{\binom{52}{5}} = \frac{1}{54145}$$

We multiply that by 13 to get the probability of a four-of-a-kind as being: 0.00024.
Or, one out of every 4165 hands dealt is probably a four of a kind.

Example 10.6: A quality control engineer selects a sample of n=3 screws from a box containing N=100 screws. Of these, 100 screws k=10 are defective. What is the probability distribution for X = the number of defective screws that the quality control engineer finds?

First, we see that we can use the hypergeometric distribution because we select n from N, with k of N defined as successes (in this case the detection of a defect). Second, to get the probability distribution we need to find h(x;N,n,k) for all values of x. Since x ranges from 0 to n, we have to solve equation (10.9) for x = 0, 1, 2, and 3.

$$h(x=0;100,3,10) = \frac{\binom{10}{0}\binom{100-10}{3-0}}{\binom{100}{3}} = \frac{\binom{10}{0}\binom{90}{3}}{\binom{100}{3}} = 0.7265$$

$$h(x=1;100,3,10) = \frac{\binom{10}{1}\binom{100-10}{3-1}}{\binom{100}{3}} = \frac{\binom{10}{1}\binom{90}{2}}{\binom{100}{3}} = 0.2477$$

$$h(x=2;100,3,10) = \frac{\binom{10}{2}\binom{100-10}{3-2}}{\binom{100}{3}} = \frac{\binom{10}{2}\binom{90}{1}}{\binom{100}{3}} = 0.0250$$

$$h(x=3;100,3,10) = \frac{\binom{10}{3}\binom{100-10}{3-3}}{\binom{100}{3}} = \frac{\binom{10}{3}\binom{90}{0}}{\binom{100}{3}} = 0.0007$$

The sum of these probabilities = 0.9999, which is close enough to one, seeing as we only used 4 sig figs.

Binomial Approximation to the hypergeometric distribution

If the population size, N, gets too large in the hypergeometric distribution then we will have problems calculating N! However, if the population gets so large, then whether experiment with or without replacement makes less difference. You can see that if the population was infinitely large, replacement would make no difference at all. For large samples, we can approximate the hypergeometric distribution by the binomial distribution. In this case the hypergeometric parameters and variables:

$$h(x;N,n,k)$$

can be approximated by the binomial variables:

$b(x;n,p)$ where $p = \frac{k}{N}$, the probability of finding k in N.

How to Compute probabilities with the Hypergeometric Distribution with a computer code

You can write a code to evaluate $P(X = x) = h(x;N,n,k)$, such as the one in hypergeo.m

```
function prob = hypergeo(x,ntot,nsamp,k)
denom = comb(ntot,nsamp);
numerator = comb(k,x)*comb(ntot-k,nsamp-x);
prob = numerator/denom;
```

This code requires the presence of comb.m.

Multivariate hypergeometric distribution

Just as the binomial distribution can be adjusted to account for multiple random variables (i.e. the multinomial distribution) so too can the hypergeometric distribution account for multiple random variables (multivariate hypergeometric distribution). The multivariate hypergeometirc distribution is defined as

$$h_m(\{x\};N,n,\{a\},k) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2}\binom{a_3}{x_3}\cdots\binom{a_k}{x_k}}{\binom{N}{n}}$$

where x_i is the number of outcomes of the ith result and a_i is the number of objects of the ith type in the total population of N, and k is the number of types of outcomes.

How to Compute probabilities with the Multivariate Hypergeometric PDF with a computer code

You can write a code to evaluate $P(\{X = x\}) = h_m(\{x\};N,n,\{a\},k)$, such as the one in multihypergeo.m

```
function prob = multihypergeo(x,ntot,nsamp,a,k)
denom = comb(ntot,nsamp);
numerator = 1.0;
for i = 1:1:k
   numerator = numerator*comb(a(i),x(i));
end
prob = numerator/denom;
```

This code requires the presence of comb.m. In the code, multihypergeo.m, x and p are vectors of length k. This code would be run at the command prompt with something like

```
multihypergeo([2,4,3],100,9,[50,30,20],3)
```

The vector arguments in multinomial must be enclosed by brackets.

Example Problem using multivariate hypergeometric PDF:
A unethical vendor has some defective computer merchandise that he is trying to unload. He has 24 computers. Of these, 12 are ok, 4 have bad motherboards, 2 have bad video cards, and 8 have bad sound cards. If we go into buy 5 computers from this vendor, what is the probability we get 3 good computers, 1 with a bad sound card and 1 with a bad video card?

$$P(\{X\} = [3,0,1,1]) = h_m([3,0,1,1];24,5,[12,4,2,8],k) = \frac{\binom{12}{3}\binom{4}{0}\binom{2}{1}\binom{8}{1}}{\binom{24}{5}} = 0.08282$$

Negative Binomial Distribution p. 133
The negative binomial distribution applies when

1. An experiment that consists of x repeated, **independent** trials.
2. Each trial can have one of two outcomes, success or failure.
3. The probability of success, p, is the same for each trial.
4. The trials are continued until we achieve the k^{th} success.

This is the Bernoulli process, except that in the binomial distribution, we fixed n trials and allowed x, the number of successes to vary. In the negative binomial distribution, we fix k successes and allow the number of trials, now x, to vary.

The probability distribution of the negative binomial random variable X, the number of trials needed to obtain k successes is

$$b^*(x;k,p) = \binom{x-1}{k-1} p^k q^{x-k} \quad \text{for } x = k, k+1, k+2... \qquad (10.12)$$

Remember q = 1-p

Example 10.6: (example 5.13, WMM 134) What is the probability when flipping three coins of getting all heads or all tails for the second time on the fifth toss?

Here we can use the negative binomial because we know that we want the k=2 success, for a independent trial with p = 1/4 for the specific case where x = 5.

$$b^*(5;2,1/4) = \binom{5-1}{2-1}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3 = (4)\left(\frac{1}{16}\right)\left(\frac{27}{64}\right) = \frac{27}{256} = 0.1055$$

How to Compute probabilities with the negative binomial PDF with a computer code

You can write a code to evaluate $P(X = x) = b^*(x;k,p)$, such as the one in negbinomial.m

```
function prob = negbinomial(x,k,p)
prob = comb(x-1,k-1)*p^k*(1-p)^(x-k);
```

This code requires the presence of comb.m. This code will work for the geometric distribution as well, by setting the input argument, k, equal to one.

Geometric Distribution p. 134

The geometric distribution is a subset of the negative binomial distribution when k=1. That is, the geometric distribution gives the probability that the first success occurs on the random variable X, the number of the trial.

The probability distribution of the geometric random variable X, the number of trials needed to obtain the first success is

$$g(x;p) = pq^{x-1} \quad \text{for } x = 1,2,3... \tag{10.13}$$

The mean of a random variable following the geometric distribution is

$$\mu = \frac{1}{p} \tag{10.14}$$

and the variance is

$$\sigma^2 = \frac{1-p}{p^2} \tag{10.15}$$

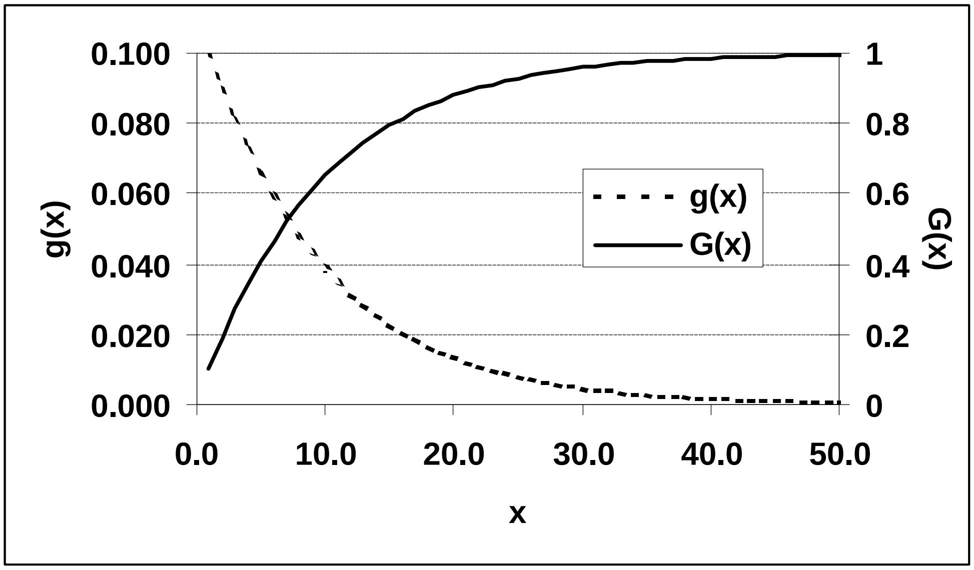
Remember q = 1-p

Example 10.7: (example 5.13, WMM 134) A recalcitrant child is told he cannot leave the dinner table until she eats one pea. Each time the child brings that ominous pea close to her mouth, there is a 90% chance her will crumbles and the spoon shakes, and the pea falls to the floor, where it is gobbled up by the family dog, forcing the child to try again. What is the probability that the child eats the pea on the first through fiftieth try?

For x = 1

$$g(x = 1; p = 0.1) = (0.1)(0.9)^{1-1} = 0.1$$

Similar calculations yield the distribution



where I have also plotted the cumulative geometric distribution,

$$G(x,p) = \sum_{i=1}^{x} g(x' = i;p) \tag{10.16}$$

which gives the probability that the pea has been eaten by the attempt x.

Poisson Distribution p. 135

When we looked at binomial, negative binomial, hypergeometric, geometric distributions, we had two outcomes success and failure. In the Poisson distribution, the random variable X is a number. In fact, it is the number of outcomes (no longer classified as a success or failure) during a given interval (of time or space). For example, the random variable X could be the number of baseball games postponed due to rain in a baseball season, or the number of bacteria in a petri dish.

The Poisson process is a collection of Poisson experiments, with the properties

1. The number of outcomes in one interval is independent of the number that occurs in any disjoint interval.
2. The probability that a single outcome will occur during a very short interval is proportional to the length of the interval and does not depend on the number of outcomes outside the interval.
3. The probability that more than one outcome will occur in an infinitesimally small interval is negligible.

In the Poisson distribution, t is the size of the interval, λ is the rate of the occurrence of the outcome, and X is the number of outcomes occurring in interval t. The probability distribution of the Poisson random variable X is

$$p(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} \quad \text{for } x = 0,1,2.............. \tag{10.17}$$

The cumulative probability distribution, that is the probability for getting anywhere between 0 and r outcomes, inclusive is

$$P(r;\lambda t) = \sum_{x=0}^{r} p(x;\lambda t) \tag{10.18}$$

Values for the cumulative Poisson probability sum are given in table A.2 of WMM. However, you can generate such a table in about one minute using Excel, for any value of r, λ, and t. (Make one column that lists values of x. Make a second column that contains $p(x;\lambda t)$ from equation (10.17), and make a third column that sums the second column from the beginning where x=0 up to the current row, for each row. Voila! A table of the Poisson probability.)

The mean and the variance of the Poisson distribution are

$$\mu = \sigma^2 = \lambda t \tag{10.19}$$

The Poisson Distribution is the asymptotical form of the binomial distribution when n, the number of trials, goes to infinity, p, the probability of a success goes to zero, and the mean (np) remains constant. There is a proof of this in WMM on pp. 138-139.

Example 10.8: Historical quality control studies at a plant indicate that there is a defect rate of 1 in a thousand products. What is the probability that in 10000 products there are exactly 5 defects? Less than or equal to 5 defects?

Using equation (10.17), we have:

$$p(x = 5;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} = \frac{e^{-10}(10)^5}{5!} = 0.0378$$

Using equation 10.18 and table A.2, we have

$$P(r = 5;\lambda t = 10) = 0.0671$$

How to Compute probabilities with the Binomial Distribution with a computer code

The easiest way to compute probabilities with the Poisson distribution is to write a short code. For example, we can obtain the

$$p(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} \quad \text{for } x = 0,1,2............ \tag{10.17}$$

with the following MATLAB code, poisson.m:

```
function f = poisson(x,p)
f= exp(-p)*p^x/factorial(x);
```

This two line program accepts as inputs x and p, and returns $f = p(x;p)$. This code accesses the program, factorial.m, to obtain the combinations.

If we wanted the most general code to calculate the probability from the Poisson PDF in some arbitrary interval, then we could write in the file poisprob.m

```
function f = poisprob(a,c,p)
f = 0.0;
for x = a:1:c
  f = f + poisson(x,p);
end
```

This file returns the value of $P(a \le X \le c) = \sum_{x=a}^{c} p(x;p)$.