MATH1280 Chapter 5-6 Quick Notes

Robert Hoot

November 6, 2015

Contents

Chapter 1

Binomial Distribution

1.1 Quick Facts

Usage	Use the binomial distribution (random variable) when an event is either a success or failure (e.g., coin flip).
Equation	$X = \text{Binomial}(n, p)$ (Yakir, 2011, p. 66). n = number of successes (MUST BE A NON-NEGATIVE INTEGER) p = probability of success, such as .5 for a coin flip
Expectation	$E(X) = np$ (Yakir, 2011, p. 69).
Population Variance Pop. Std. Deviation	$np(1-p)$ $\sqrt{np(1-p)}$ (Yakir, 2011, p. 69)
`pbinom(X, size=n, prob=p)`	For an integer value X, show the probability getting X or fewer successes when repeating an experiment n times if the probability of success for each experiment is p . example: `pbinom(4, size=10, prob=.3)` tells you the probability getting 4 or fewer success when repeating an experiment 10 times if each experiment had a .3 probability of success.
`qbinom(Q, size=n, prob=p)`	For a probability value Q (between 0 and 1), show the value from a binomial distribution that is of percentile Q, where the binomial distribution is based on samples of size n and probability of success p. example: `qbinom(.975, size=10, prob=.3)` tells you the integer value from the sample space of the distribution that is greater than or equal to 97.5% of the values in that distribution.

`dbinom(X, size=n, prob=p)`	Show the probability of randomly obtaining EXACTLY X successes when repeating an experiment n times with each experiment having a probability of success of p. example: `dbinom(3, size=10, prob=.3)` shows the probability of ramdomly getting exactly 3 successes when you repeat an experiment 10 times and each experiment has a .3 probability of succes.

1.2 R Probability Density

`dbinom(nbr_of_successes, sample_size, prob)`
Example: what is the probability of rolling 3 heads in 5 coin flips:

```
> dbinom(3, size=5, prob=.5)
[1] 0.3125
```

1.3 R Cumulative Probability Density

`pbinom(nbr_of_successes, sample_size, prob)`

Example: the probability of randomly getting from 0 to 10 ones when rolling a single die 10 times (a 6-sided cube with each side having probability of success 1/6).

```
> success.count <- 0:10
> binom.prob <- round(dbinom(success.count, 10, 1/6), 3)
> cum.prob <- round(pbinom(success.count, 10, 1/6),3)
> data.frame(success.count, binom.prob, cum.prob)
   success.count binom.prob cum.prob
1              0      0.162    0.162
```

2	1	0.323	0.485
3	2	0.291	0.775
4	3	0.155	0.930
5	4	0.054	0.985
6	5	0.013	0.998
7	6	0.002	1.000
8	7	0.000	1.000
9	8	0.000	1.000
10	9	0.000	1.000
11	10	0.000	1.000

Chapter 2

Poisson Distribution

2.1 Quick Facts

Usage	Use the Poisson distribution (random variable) when you know the expected count of successes in a fixed period of time, such as number of incoming phone calls per minute, and the count is subjectively large. If you were using binomial, and n is large and p is small, you can use Poisson as a close estimate of the binomial (Yakir, 2011, p. 71).
Equation	$X = \text{Poisson}(\lambda)$ (Yakir, 2011, p. 72). $\lambda =$ the expectation (number of successes in a given period of time; it can be a decimal value)
Expectation	$E(X) = \lambda$ (Yakir, 2011, p. 72).
Population Variance Pop. Std. Deviation	λ (Yakir, 2011, p. 74) $\sqrt{\lambda}$ (Yakir, 2011, p. 74)

`ppois(X, lambda=L)`	For an integer value X, show the probability getting X or fewer successes when the expectation of the Poisson distribution is L. example: `ppois(4, lambda=7.3)` tells you the probability getting 4 or fewer success when from a Poisson experiment that has an expectation of 7.3.
`qpois(Q, lambda=L)`	For a probability value Q (between 0 and 1), show the value from a Poisson distribution that is of percentile Q, where the Poisson distribution has an expectation of L. example: `qpois(.975, lambda=7.3)` tells you the integer value from the sample space of the distribution that is greater than or equal to 97.5% of the values in that distribution.
`dpois(X, lambda=L)`	Show the probability of randomly obtaining EXACTLY X successes from a Poisson distribution that has an expectation of L. example: `dbinom(3, lambda=7.3)` shows the probability of randomly getting exactly 3 successes from a Poisson distribution that has an expectation of 7.3.

2.2 R Probability Density

This is an example from Yakir (2011), pp. 71-72. If we expect two phone calls to arrive at the switchboard every minute then we might be able to model the distribution of calls using a Poisson distribution with an lambda of 2, where the lambda represents the expected number of calls per minute.

We will use the dpois() function in R. To get help on the function, run this in interactive R: ?dpois here are some values from this discrete distribution:

```
> dpois(c(0:10), lambda=2)
 [1] 1.353353e-01 2.706706e-01 2.706706e-01 1.804470e-01 9.022352e-02
 [6] 3.608941e-02 1.202980e-02 3.437087e-03 8.592716e-04 1.909493e-04
[11] 3.818985e-05
```

That looks ugly, so let's round to 3 decimal places:

```
> round(dpois(c(0:10), lambda=2), 3)
 [1] 0.135 0.271 0.271 0.180 0.090 0.036 0.012 0.003 0.001 0.000 0.000
```

The results mean that there is a .135 probability that there will be zero calls in the next minute. There is a .271 probability that there will be 1 call, and .271 for 2 calls, and .180 for three calls...

I can add explanations by gluing the values together using the paste() function (you do not need to do this for the class)

```
> nbr.of.calls <- 0:10
> prob <- round(dpois(c(0:10), lambda=2), 3)
> explanation <- paste('The probability of randomly getting exactly',
  nbr.of.calls, 'calls in one minute is:', prob)
> explanation
 [1] "The probability of randomly getting exactly 0 calls in one minute is: 0.135"
 [2] "The probability of randomly getting exactly 1 calls in one minute is: 0.271"
 [3] "The probability of randomly getting exactly 2 calls in one minute is: 0.271"
 [4] "The probability of randomly getting exactly 3 calls in one minute is: 0.18"
 [5] "The probability of randomly getting exactly 4 calls in one minute is: 0.09"
 [6] "The probability of randomly getting exactly 5 calls in one minute is: 0.036"
 [7] "The probability of randomly getting exactly 6 calls in one minute is: 0.012"
 [8] "The probability of randomly getting exactly 7 calls in one minute is: 0.003"
 [9] "The probability of randomly getting exactly 8 calls in one minute is: 0.001"
[10] "The probability of randomly getting exactly 9 calls in one minute is: 0"
[11] "The probability of randomly getting exactly 10 calls in one minute is: 0"
```

2.3 Cumulative Probability Density

```
> cum.prob <- round(ppois(c(0:10), lambda=2), 3)
> explanation <- paste('The cumulative probability of randomly getting',
```

```
  nbr.of.calls, 'calls in one minute is:', cum.prob)
> explanation
 [1] ``The cumulative probability of randomly getting 0 calls in one minute is: 0.135''
 [2] ``The cumulative probability of randomly getting 1 calls in one minute is: 0.406''
 [3] ``The cumulative probability of randomly getting 2 calls in one minute is: 0.677''
 [4] ``The cumulative probability of randomly getting 3 calls in one minute is: 0.857''
 [5] ``The cumulative probability of randomly getting 4 calls in one minute is: 0.947''
 [6] ``The cumulative probability of randomly getting 5 calls in one minute is: 0.983''
 [7] ``The cumulative probability of randomly getting 6 calls in one minute is: 0.995''
 [8] ``The cumulative probability of randomly getting 7 calls in one minute is: 0.999''
 [9] ``The cumulative probability of randomly getting 8 calls in one minute is: 1''
[10] ``The cumulative probability of randomly getting 9 calls in one minute is: 1''
[11] ``The cumulative probability of randomly getting 10 calls in one minute is: 1''
```

2.4 Using Poisson to Estimate Binomial

If you have a binomial experiment (each event is either a success or failure), you might be able to model it using a Poisson distribution. You can use the Poisson distribution to estimate binomial only if your binomial experiment has a large n and a small p—such as the chance of getting a hole in one in golf or the chance of have a part failure on a machine. Do not use this trick to estimate a coin flip experiment because the p value is too big.

The expectation for a binomial experiment is $\mathrm{E}(x) = np$. If your sample size is $n = 500$ and the probability of success for each event is $p = .01$ then your expectation is 5. For Poisson, the expectation (which is equal to lambda for Poisson distribution) will be of the entire set of 500 events, and the expectation would be 5 successes ($\lambda = 5$). The standard deviation would be $\sqrt{\lambda}$ (Yakir, 2011, p. 74). The following code plots the probability of getting various numbers of successes using the binomial distribution (random variable) versus the Poisson distribution (random variable). The two lines are nearly on top of each other. The result of the following code is in Figure 2.1.

```
n <- 500
p <- .01
mu.binomial <- n * p
```

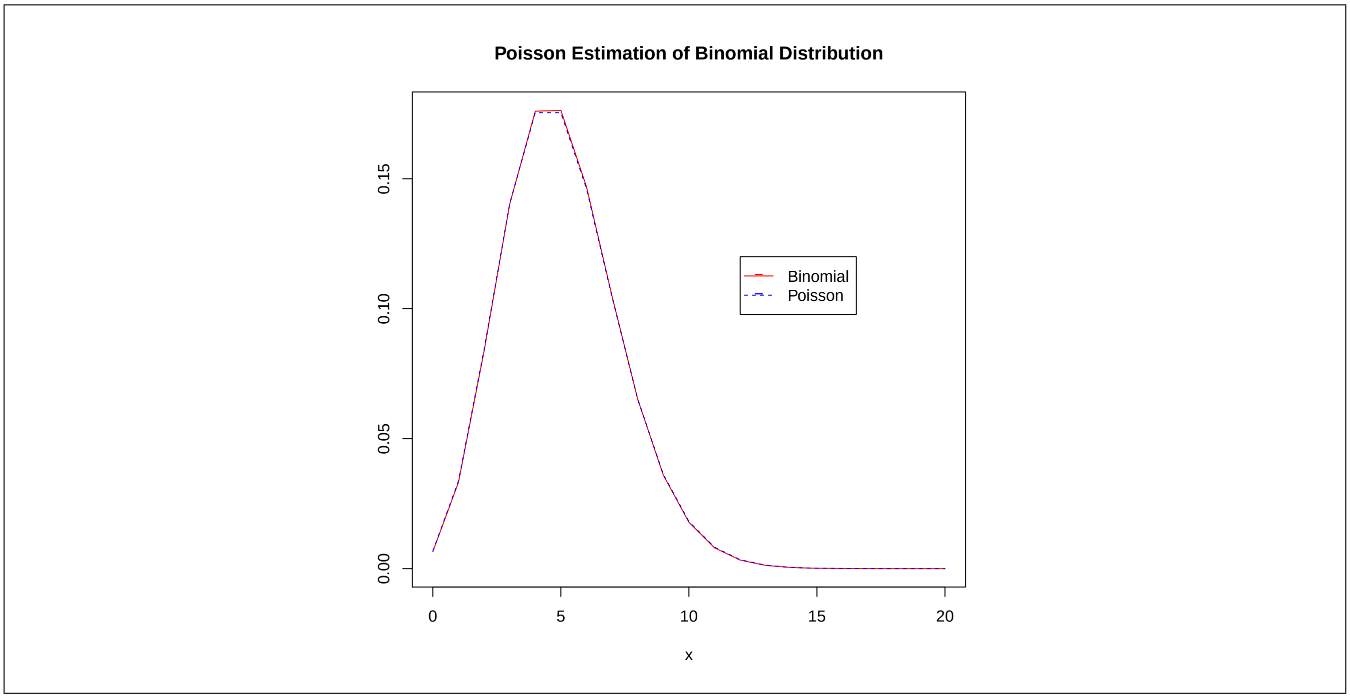

Figure 2.1: Poisson Estimate of Binomial Distribution ($n = 500$, $p = .01$)

```
x <- 0:20
plot(x, dbinom(x, size=n, prob=p), type='l',col='red',
  ylab='',
  main='Poisson Estimation of Binomial Distribution')
# lty=2 makes it a dashed line
lines(x, dpois(x, lambda=n*p), type='l',col='blue', lty=2)
legend(x=12, y=.12, col=c(2, 4), lty=c(1,2),
        legend=c('Binomial', 'Poisson'), pch='--')
```

Chapter 3

Uniform Distribution

3.1 Quick Facts

Usage	Use the uniform distribution (random variable) when each possible value in the sample space is equally likely. You would use this when modeling a lottery.
Equation	$X = \text{Uniform}(a, b)$ (Yakir, 2011, p. 75). $a =$ the lowest value in the range of possible values. $b =$ the highest value in the range of possible values
Expectation	$E(X) = \frac{(a+b)}{2}$ (Yakir, 2011, p. 78).
Population Variance Pop. Std. Deviation	$\frac{(b-a)^2}{12}$ (Yakir, 2011, p. 78) $\sqrt{\frac{(b-a)^2}{12}}$

`punif(X, min=A, max=B)`	For a value X (decimals are OK), show the probability of randomly selecting a value of X or lower from a uniform distribution that has values between A and B. example: `punif(4, min=1, max=9)` tells you the probability of randomly selecting a value of 4 or lower from a uniform distribution that has values between 1 and 9.
`qunif(Q, min=A, max=B)`	For a probability value Q (between 0 and 1), show the value from a uniform distribution that is of percentile Q, where the uniform distribution has values between A and B. example: `qunif(.975, min=1, max=9)` tells you the value from the sample space of a uniform distribution that is greater than or equal to 97.5% of the values in that distribution.
`dunif(X, min=A, max=B)`	Avoid using this function in this class. The result is NOT the probability of getting exactly X (you could multiply the result by a small $\pm\delta x$ value to estimate the probability of selecting something in that range, but it is easier to use two `punif` functions).

3.2 Using the Uniform Distribution

The uniform distribution can be used to model random selection of values from a segment of the number line. Any real, decimal value withing the upper and lower boundaries of the distribution is allowed (is in the sample space). If I ask you to pick a number between 1 and 10 and I say that you can pick any decimal value, then the answer could be modeled with a uniform distribution (if you pick randomly). A value of 1.3345384783497 is a legitimate choice as is a similar value with a million additional decimal places.

Chapter 4

Exponential Distribution

4.1 Quick Facts

Usage	Use the exponential distribution (random variable) when you want to model the time between events, such as the amount of time between phone calls at a call center (the actual time intervals need to vary as predicted by the model, so it is not a good idea to predict the time between ticks of a clock using this distribution).
Equation	$X = \text{Exponential}(\lambda)$ (Yakir, 2011, p. 79). $\lambda =$ the rate variable—if you get 20 calls per minute, the rate is 20; if you get 1 every two minutes then the rate is 0.50
Expectation	$E(X) = \frac{1}{\lambda}$ (Yakir, 2011, p. 80).
Population Variance Pop. Std. Deviation	$\frac{1}{\lambda^2}$ (Yakir, 2011, p. 80) $\sqrt{\frac{1}{\lambda^2}}$

`pexp(X, rate=R)`	For a value X (decimals are OK), show the probability of randomly selecting a value of X or lower from an exponential distribution that has rate R. example: `pexp(4, rate=4)` tells you the probability of randomly selecting a value of 4 or lower from an exponential distribution that has a rate of 4.
`qexp(Q, rate=R)`	For a probability value Q (between 0 and 1), show the value from an exponential distribution that is of percentile Q, where the exponential distribution has rate R. example: `qexp(.975, rate=4)` tells you the value from the sample space of an exponential distribution that is greater than or equal to 97.5% of the values in that distribution.
`dexp(X, rate=R)`	Avoid using this function in this class. The result is NOT the probability of getting exactly X (you could multiply the result by a small $\pm\delta x$ value to estimate the probability of selecting something in that range, but it is easier to use two `pexp` functions).

4.2 R Probability

What is the probability that the time between calls will be between 1 minute and 2 minutes when the expectation is 2 calls per minute. Note that `lower.tail=TRUE` is the default, but I show it as a reminder:

```
> pexp(2, rate=2, lower.tail=TRUE) - pexp(1, rate=2, lower.tail=TRUE)
[1] 0.1170196
```

What is the probability that the time between calls will be less than 1 minute when the expectation is 2 calls per minute. Note that `lower.tail=TRUE` is the default, but I show it as a reminder:

```
> pexp(1, rate=2, lower.tail=TRUE)
```

```
[1] 0.8646647
```

What is the probability that the time between calls be MORE than 1 minute when the expectation is 2 calls per minute. (note that `lower.tail` is set to `FALSE`, meaning that I get the probability of being higher than 1).

```
> pexp(1, rate=2, lower.tail=FALSE)
[1] 0.1353353
> # the above is the same as
> 1 - pexp(1, rate=2, lower.tail=TRUE)
[1] 0.1353353
```

Chapter 5

Normal Distribution

5.1 Quick Facts

Usage	Widely used distribution to model natural variation in the physical and social sciences.
Equation	$X = \text{Normal}(\mu, \sigma^2)$ (Yakir, 2011, p. 88). $\mu =$ the population expectation $\sigma =$ the population standard deviation (see Yakir, 2011, pp. 57-59 and remember that the sample standard deviation has n - 1 in the denominator)
Expectation	$E(X) = \mu$ (Yakir, 2011, pp. 63, 88).
Population Variance Pop. Std. Deviation	σ^2 (Yakir, 2011, p. 88) σ

`pnorm(X, mean=`μ`, sd=`σ`)`	For a value X (decimals are OK), show the probability of randomly selecting a value of X or lower from an normal distribution that has mean of μ and a standard deviation of σ. example: `pnorm(4, mean=10, sd=1.3)` tells you the probability of randomly selecting a value of 4 or lower from an normal distribution that has a mean of 10 and a standard deviation of 1.3.
`qnorm(Q, mean=`μ`, sd=`σ`)`	For a probability value Q (between 0 and 1), show the value from an normal distribution that is of percentile Q, where the normal distribution has a mean of μ and a standard deviation of σ. example: `qnorm(.975, mean=10, sd=1.3)` tells you the value from the sample space of an normal distribution that is greater than or equal to 97.5% of the values in that distribution.
`dnorm(X, mean=`μ`, sd=`σ`)`	Avoid using this function in this class. The result is NOT the probability of getting exactly X (you could multiply the result by a small $\pm\delta x$ value to estimate the probability of selecting something in that range, but it is easier to use two `pnorm` functions).

5.2 Find the Probability When You Have x-values

Imagine that the height of 2-year-old children can be modeled using a normal distribution. If the expectation of a population is 50 ($\mu = 50$), and the standard deviation is 3 ($\sigma = 5$),

What is the probability that a 2-year old child will be UP TO 40 cm tall? The `pnorm` function shows the probability of being LESS THAN OR EQUAL TO the first value in the function, so:

```
> pnorm(40, mean=50, sd=3)
```

What is the probability that a 2-year old child will be TALLER than 40 cm? The `pnorm` function shows the probability of being LESS THAN OR EQUAL TO the first value in the function, but we want the probability of being higher, so take 1- minus that `pnorm()` value:

```
> 1 - pnorm(40, mean=50, sd=3)
```

What is the probability that a child is TALLER than 56 cm? You want to find the probability of randomly selecting a value greater than 56 from this distribution:

```
> 1 - pnorm(56, mean=50, sd=3)
```

What is the probability that a child is between 44 cm and 56 cm? Note that if your answer is negative, you made a mistake (probability values are always between 0 and 1).

```
> pnorm(56, mean=50, sd=3) - pnorm(44, mean=50, sd=3)
[1] 0.9544997
```

5.3 Find the Probability When you have μ and σ

What is the probability that an observation will be more than 3 standard deviations above the mean? We can use the standard normal distribution with $\mu = 0$ and $\sigma = 1$, but you can experiment with all other values for μ and σ and see what you find. First note that this will show the probability of getting less than 3:

```
> pnorm(3, mean=0, sd=1)
[1] 0.9986501

> # Now find the answer
> 1 - pnorm(3, mean=0, sd=1)
[1] 0.001349898
```

The probability of finding an observation more than 3 standard deviations above the mean is about .00135.

Try another set of values where the x-value is 3 standard deviations above the mean. In this case, I add "3 times the standard deviation" to the mean of the distribution, and test the probability for that value. In the code below, `mu` is the population mean and `s` is the population standard deviation:

```
> mu = 100
> s = 2
> test.value = mu + 3 * s
> test.value
[1] 106
> 1 - pnorm(test.value, mean=mu, sd=s)
[1] 0.001349898
```

5.4 Find the x-value When You Have a Probability

If you are given a x and asked to find a probability, it might be a question to find a certain percentile of a distribution. You want to identify the tallest 5% of 2-year old children. What is the criterion value (which height value is the cutoff point that identifies the tallest 5% of 2-year olds)? We look for the point such that the value of x is bigger than 95% of the other values:

```
> qnorm(.95, mean=50, sd=3)
[1] 54.93456
```

Children who are 54.93456 or more cm tall are in the top 5% (based on a normal distribution with mean of 50 cm and a standard deviation of 3 cm).

How about the shortest 5%? We look for the value of x that is in the bottom 5%:

```
> qnorm(.05, mean=50, sd=3)
```