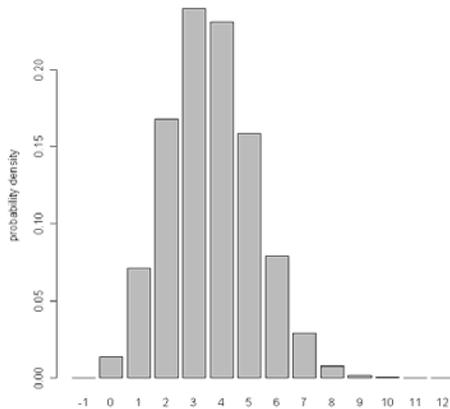


WILD 502

The Binomial Distribution

The binomial distribution is a finite discrete distribution. The binomial distribution arises in situations where one is observing a sequence of what are known as Bernoulli trials. A Bernoulli trial is an experiment which has exactly two possible outcomes: success and failure. Further the probability of success is a fixed number p which does not change no matter how many times you conduct the experiment. A binomial distributed variable counts the number of successes in a sequence of N independent Bernoulli trials. The probability of a success (head) is denoted by p . For N trials you can obtain between 0 and N successes.

A representative example of a binomial density function is plotted below for the case of $p = 0.3$, $N=12$ trials, and for values of k heads = -1, 0, ..., 12. Note, as expected, there is 0 probability of obtaining fewer than 0 heads or more than 12 heads! And, the probability density is at a peak for the case where we observe 3 or 4 heads in 12 trials.



The density function for a quantity having a binomial distribution is given by:

$$f(y|N,p) = \frac{N!}{y! \cdot (N-y)!} \cdot p^y \cdot (1-p)^{N-y} \text{ if } 0 \leq y \leq N, 0 \text{ otherwise}$$

Here y and N are positive integers and $0 < p < 1$ is a probability. The numbers N and p determine the distribution, N representing the number of trials and p the probability of success in each trial. The number $f(y,N,p)$ represents the probability of exactly y successes out of N trials. It makes sense then that this probability is only nonzero for $y=0,1,2,\dots,N$. Here, I've used y for the number of successes but you might see books or lecture notes use k and sometimes you'll see n also.

The Binomial Coefficient calculates how many ways a sample size of y heads can be taken from a population of N coins without replacement. The remainder of the equation calculates how probable any such an outcome is.

So, if we know that adult female red foxes in the Northern Range of Yellowstone National Park have a true underlying survival rate of 0.65, we can calculate the probability that different numbers of females will survive. Imagine that there are 25 adult females in the population at the start of the year. What's the probability that 0, 1, 2, ..., 25 will survive? In **R**, we can calculate this easily with the following command 'dbinom(seq(0,25,1),25,0.65)'. Or, to be more organized, we can type in the following:

```
> x=seq(0,25)
> y=dbinom(x,25,0.65)
> out=cbind(x,y)
> round(out,3)
      x      y
[1,]  0 0.000
[2,]  1 0.000
[3,]  2 0.000
[4,]  3 0.000
[5,]  4 0.000
[6,]  5 0.000
[7,]  6 0.000
[8,]  7 0.000
[9,]  8 0.001
[10,] 9 0.002
[11,] 10 0.006
[12,] 11 0.016
[13,] 12 0.035
[14,] 13 0.065
[15,] 14 0.103
[16,] 15 0.141
[17,] 16 0.163
[18,] 17 0.161
[19,] 18 0.133
[20,] 19 0.091
[21,] 20 0.051
[22,] 21 0.022
[23,] 22 0.008
[24,] 23 0.002
[25,] 24 0.000
[26,] 25 0.000
> sum(y)
[1] 1
```

So, we can see that the most likely outcome is for 16 or 17 to survive but that other numbers are also fairly probable. We also see that the probabilities sum to 1.

If we want to simulate outcomes for this population, we can do so easily using the command 'rbinom' and providing it numbers, e.g, type 'rbinom(1,25,.65)', which yields the outcome of one simulated dataset with 25 individuals and a probability of success of 0.65. When I executed the code, I obtained 14 survivors. Repeating the command will provide additional simulations: when I did 4 more simulations, I obtained 12, 14, 17, & 18.

We can do many simulations by issuing a command like 'rbinom(100, 25, .65)', which is 100 simulations with 25 trials in each & where every trial has a probability of success of 0.65. When I did this, I obtained:

```
[1] 10 18 14 15 17 17 13 14 18 18 12 14 17 14 21 18 13 20 14 15 18 22 20 15 20
[26] 14 17 14 20 17 18 12 16 17 14 19 18 17 17 19 16 19 12 16 19 16 20 18 20 17
[51] 15 17 16 17 14 12 17 18 15 15 16 17 16 19 18 18 13 16 17 18 15 16 16 15 12
[76] 15 15 19 17 13 16 21 17 21 21 14 12 13 14 19 18 15 15 19 16 12 19 16 13 17
```

Where the minimum was 10, the maximum was 22, the mean was 16.34, and the variance was 6.34.

Maximum Likelihood Estimation – the Binomial Distribution

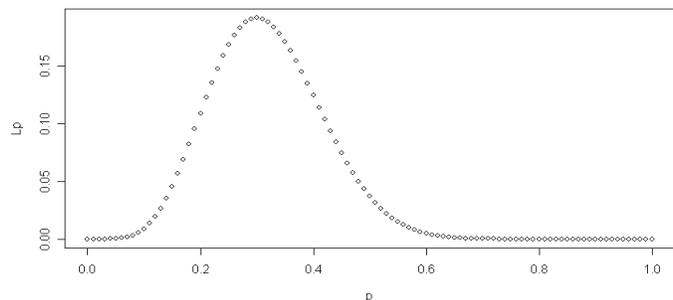
This is all very good if you are working in a situation where you know the parameter value for p , e.g., the fox survival rate. And, it's useful when simulating population dynamics, too. But, in this course, we'll be *estimating* the parameters from data rather than generating data from known parameters. So, ... Estimating binomial probability from data with Maximum Likelihood

$$L(p|y, N) = \frac{N!}{y! \cdot (N - y)!} \cdot p^y \cdot (1 - p)^{N-y} \text{ if } 0 \leq y \leq N, 0 \text{ otherwise}$$

Now, we're estimating how *likely* various values of p are given the *observed data*. Imagine that you radio-collared 20 of the female foxes, studied their survival for 1 year, and found that 6 of 20 survived. What would your guess be for survival rate? I think that it would be 6/20 or 0.3? Let's see if the likelihood equation supports that guess.

We can do this in **R** easily enough with the following commands:

```
> p=seq(0,1,.01)
> N=20
> y=6
>
bc=factorial(N)/(factorial(y)*factorial(N-
y))
> Lp=bc*p^y*(1-p)^(N-y)
> plot(p,Lp)
```



Our guess is supported as the most likely *estimate* of p (\hat{p}) is 0.30. And ... it turns out that the peakedness of that likelihood surface also gives us information on how likely other values of \hat{p} are as well. As you read through the Cooch and White reading for Maximum Likelihood with binomials you'll see that we can use derivatives to get the variance of \hat{p} . And, once we have that we can build confidence intervals, etc.

It turns out that we typically work with log-Likelihoods (or $\ln L$) largely because of analytical advantages that come about when we take logs of these equations.

For the binomial likelihood, we obtain:

$\ln L(p|y, N) = \ln(C) + y \cdot \ln(p) + (N - y) \cdot \ln(1 - p)$, where C is a constant representing the binomial coefficient's value (it's often removed for simplicity as it's truly just a constant).

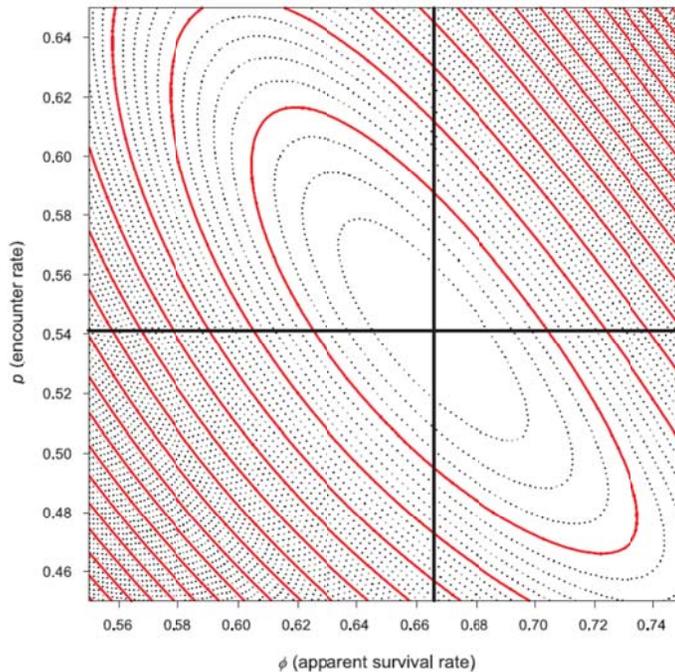
We can actually obtain closed-form estimators for \hat{p} from the binomial likelihood equation in its log form. The steps are provided and explained in *Chapter 1 (Section 1.3) of Cooch and White*. **Do study that section of the chapter carefully.** The most likely value of $\hat{p} = y/N$ and the MLE for the variance of \hat{p} is $\hat{p} \cdot (1 - \hat{p})/N$. From this pair of equations you can see that (1) the variance decreases as N increases and (2) the variance for a given N is largest when $\hat{p} = 0.5$.

Even when the exact ML estimator doesn't exist as a simple formula, we can use iterative numerical routines to find solutions and ... it turns out that excellent software exists for doing such estimation. So, Maximum Likelihood Estimation is very useful for obtaining estimates of parameters along with measures of precision. The actual likelihood values are useful for evaluating the strength of evidence for competing models of the process that generated the observed data.

We'll get into details later but one method that we'll use to choose among models involves calculating **AIC** for each competing model, where $AIC = -2(\ln L) + 2(k)$ and k is the number of parameters in the model. Section 4.3 of Cooch & White gives a quick primer on AIC.

It's also worth noting that the method of maximum likelihood provides estimates that are asymptotically unbiased, normally distributed, and of minimum variance among such estimators.

The method of maximum likelihood also extends to problems where we're trying to estimate >1 parameter at a time. For example, when estimating survival from resightings of marked animals in a situation where some marked animals might go undetected on some surveys, we have to estimate both detection probability and survival rate. For a model that constrains detection rate and survival rate to be constant through time, we have to estimate 2 parameters. In such a case, we estimate the likelihood for 2 parameters simultaneously, i.e., we find the pair of values for detection probability and survival rate that are most likely. Visually, this looks like a contour map (likelihood is represented by the contour lines) and various values of detection probability and survival rate appear along the x-y axes.



It turns out that we'll work with some models that have many parameters to be estimated and we'll be glad that we have software that solves for the estimates and their associated variances and covariances. For many situations we'll work with, the actual estimators won't fall out in nice simple forms and produce what are called closed-form solutions. In such cases, there are fortunately good ways to proceed but that's a story for another day!

Appendix to lecture for those interested in deriving the estimators

$$f(y|N,p) = \begin{cases} \frac{N!}{y!(N-y)!} p^y \cdot (1-p)^{(N-y)} & \text{if } 0 < y < N \\ 0 & \text{otherwise} \end{cases}$$

$$L(p|y,N) = \frac{N!}{y!(N-y)!} p^y \cdot (1-p)^{(N-y)}$$

$$\ln L(p|y,N) = y \cdot \ln(p) + (N-y) \cdot \ln(1-p)$$

$$\frac{\partial [\ln L(p|y,N)]}{\partial p} = \frac{y}{p} - \frac{(N-y)}{(1-p)} = 0$$

You can use R to obtain the derivatives if you want:

```
> D(expression(y*log(p)), 'p')
y * (1/p)
> D(expression((N-y)*log(1-p)), 'p')
-((N - y) * (1/(1 - p)))
```

$$\frac{y}{p} = \frac{(N-y)}{(1-p)} \quad \dots \quad y(1-p) = p(N-y)$$

$$y - yp = pN - yp$$

$$y - yp + yp = pN \quad \dots \quad y = pN$$

$$p = \frac{y}{N}$$

Next, take the 2nd derivatives as 1st step in obtaining equation for variance

You can again use R to obtain the derivatives (take the derivative of the 1st derivative)

```
> D(expression(y/p - (N-y)/(1-p)), 'p')
-(y/p^2 + (N - y)/(1 - p)^2)
```

$$\frac{\partial^2 \ln L}{\partial p^2} = \frac{\partial}{\partial p} \left(\frac{y}{p} - \frac{(N-y)}{(1-p)} \right) = -\frac{y}{p^2} - \frac{(N-y)}{(1-p)^2}$$

Now substitute y/N for one of the p 's each time it appears

$$\frac{\partial^2 \ln L}{\partial p^2} = -\frac{y}{p \cdot \frac{y}{N}} - \frac{(N-y)}{\left(1 - \frac{y}{N}\right) \cdot (1-p)}$$

Then rewrite $(N-y)$ as $(N)(1-y/N)$ to obtain:

$$-\frac{N}{p} - \frac{N \cdot (1 - \frac{y}{N})}{\left(1 - \frac{y}{N}\right) \cdot (1-p)} = -\frac{N}{p} - \frac{N}{(1-p)} = -\left[\frac{N \cdot (1-p) + (N \cdot p)}{p \cdot (1-p)} \right] = -\left[\frac{N - Np + N \cdot p}{p \cdot (1-p)} \right] = -\frac{N}{p \cdot (1-p)}$$

The 2nd derivative of the log-likelihood evaluated at \hat{p} provides us with the Hessian. That's useful because the negative inverse of the Hessian provides us with a maximum likelihood estimator for the variance of \hat{p} . Here, that estimator is $\frac{p(1-p)}{N}$.

By the way, when the likelihood involves more parameters, then you work with the relevant partial derivatives of the likelihood or $\log(L)$ to get the MLE's for each of the parameters, the variances of the estimated parameters, and the estimated covariances of the estimated parameters.