EXERCISE 8: REPEATED COUNT MODEL (ROYLE)

In collaboration with Heather McKenney

University of Vermont, Rubenstein School of Environment and Natural Resources

Please cite this work as: Donovan, T. M. and J. Hines. 2007. Exercises in occupancy modeling and estimation.
<http://www.uvm.edu/envnr/vtcfwru/spreadsheets/occupancy.htm

TABLE OF CONTENTS

ROYLE REPEATED COUNT SPREADSHEET MODEL

OBJECTIVES

- To understand the basics of the Poisson and Binomial distributions.

- To learn and understand the basic mixture model for estimating abundance, and how it fits into a multinomial maximum likelihood analysis.

- To use Solver to find the maximum likelihood estimates for the probability of detection and lambda, the average site abundance.

- To assess deviance of the saturated model.

- To introduce concepts of model fit.

- To learn how to simulate basic mixture data.

INTRODUCTION

Suppose that you want to estimate the size of a breeding songbird population across a large area. You decide that you can't survey the entire area, and instead decide to estimate abundance at a number of study sites within your study area. On each visit, you conduct a standardized survey of one kind or another. Point counts are a popular choice, in which the observer visits a site and records all birds heard or seen within a specified time period. These counts are repeated over the course of the breeding season (say, 5 times) under the assumption that the population is closed to any changes in animals over the study period (i.e., no births, no deaths, no immigrants, no emigrants). A sample of your data may look like this:

	A	B	C	D	E	F
7				Visit		
8	Site	1	2	3	4	5
9	1	2	1	1	2	2
10	2	1	2	2	1	2
11	3	0	4	1	1	1
12	4	2	0	0	0	1
13	5	1	0	2	1	2
14	6	0	0	1	0	0
15	7	0	1	0	0	0
16	8	1	0	0	2	1
17	9	2	3	1	1	0
18	10	0	0	0	1	0
19	11	2	2	1	0	0
20	12	1	0	1	1	1
21	13	1	0	0	1	0
22	14	2	0	0	2	2
23	15	1	1	1	1	1
24	16	1	2	0	1	1
25	17	0	0	1	0	0
26	18	1	1	2	0	1
27	19	1	0	0	1	0
28	20	2	1	0	0	0

The data shown above are hypothetical data collected at 20 different sites (or 20 different point count locations in the study area). Let's let R denote the total number of sites; R = 20. Let's let T denote the total number of surveys for each site; T = 5. In the first visit to site 1, you detected 2 individuals of the species of interest. The second visit to that same site yielded a count of 1 individual; the third visit yielded a count of 1 individual, and the fourth and fifth visits yielded counts of 2 individuals. This is a very common scenario in bird-survey work. During any given survey, not all individuals will be detected. Because we assume that the population is closed to changes in abundance over the course of the 5 surveys, study site number 1 MUST have had AT LEAST two individuals because we counted two individuals in surveys 1, 4, and 5. This also means that we missed at least 1 individual on surveys 2 and 3 because only one individual was detected during those surveys.

Now, one way to estimate abundance from these data is to simply take the mean of the counts, and carry on. (I've done this myself, before learning better methods of analysis!). As we've seen, though, the mean will be biased low because of imperfect detection. For site 1, the mean abundance of the raw count data is 1.6 individuals, and given our assumptions of population closure, we know that there are AT LEAST 2 individuals present. Another common method is to use the maximum count from the 5 surveys. In this case, the answer is 2. That's all well and good, but suppose our species is very elusive . . . . it sings sporadically and is very cryptic. How do we know whether the maximum count is representative of the true number of animals that occur at the site? Well, we don't know. And this is the major reason why researchers can no longer analyze the mean or maximum of count data and expect to have their results published.

So, how can these data be analyzed in a rigorous fashion to account for imperfect detection? Andy Royle figured one method out, and in this exercise we will learn about the Royle Repeated Count model for analyzing data such as those shown above. This model is described in Chapter 5 of the book, Occupancy Estimation and Modeling. The original paper describing the model is Royle, J.A. 2004. N-Mixture Models for Estimating Population Size from Spatially Replicated Counts. Biometrics 60, 108-115.

ASSUMPTIONS OF THE ROYLE COUNT MODEL

There are several assumptions of this model. We already mentioned a key one: that the population is assumed to be demographically closed over the course of the T surveys. There are two more critical assumptions, namely: (1) the spatial distribution of the animals across the R survey sites follows some kind of prior distribution, such as the Poisson distribution, and (2) the probability of detecting n

animals at a site represents a binomial trial (Bernoulli trial) of how many animals are actually at that site. In a nutshell, the Royle Count Model is a mixture of the Poisson and Binomial distributions, and the goal is to find the parameters that shape the Poisson and Binomial distributions in such as way that the results of the model will yield data that "match" our observed field data. Let's get started.

POISSON DISTRIBUTION

The Royle count model starts with the assumption that the spatial distribution of animals is governed by some statistical distribution, such as the Poisson. The spatial distribution of animals is simply how many animals occur at each site within the study area. Each of the survey sites will contain some number of animals (some sites may contain 0 animals, some may contain 1 animal, some may contain 2 animals, etc.). That number, the site abundance, is a function of the mechanisms governing the distribution.

A prior distribution is specified, or chosen, based on how you think the animal species is really distributed. If you were in the planning stages of your survey and had not yet collected any data, you would ask yourself, "How are these animals distributed in space?" _Prior_ to collecting any data, we specify the Poisson distribution—we consider the Poisson to accurately represent the true spatial distribution of our target species. You've probably been exposed to the Poisson distribution in your studies already, but let's review the Poisson Distribution in some detail in case you've forgotten.

The Poisson distribution is used to model the number of certain randomly occurring events, like the number of car accidents in your home town, or the number of

individuals of a species within each of your survey sites. In the car accident example, each accident is independent of every other accident and the number of accidents in any time period is random and independent of any other time period. The spatial distribution of animals can also meet these Poisson assumptions when the number of animals inhabiting one site is random and independent of the number of animals at other sites. This means that if you conduct a point count survey at site 1, the number of animals in site 2 will be independent of the number of animals in site 1. (Clearly, there are many ways in which this assumption can be violated. For example, sites that are located near each other may not be truly independent if there is spatial autocorrelation among the sites. We'll discuss some options that account for non-independence later in the exercise).

The Royle-count model assumes that each of the R sites in your occupancy survey is home to some number of animals that can be modeled by a specified prior distribution like the Poisson. We also assume this number does not change over the course of your study. This additional assumption means repeated sampling visits must be completed within a relatively short period of time.

The Poisson distribution has a single parameter, λ ("lambda"), the mean. In this case, lambda is the mean or average abundance across the R sites. The Poisson distribution returns the probability of any level of abundance x from 0 to ∞ given some lambda. We often don't know what lambda is, but we can make some guesses. For example, if you think $\lambda = 3$, this means that the average abundance of animals across all sites is 3 animals. Given this information, you can find the probability that a specific number of animals will occur at a given site. For example, when lambda = 3, the probability of a single site having an abundance of 5 is 0.10. Where

does 0.10 come from? It is calculated with the probability density formula for the Poisson:

$$f_x = \frac{e^{-\lambda}\lambda^x}{x!}$$

where lambda is the mean of the Poisson distribution, and x is the "event" of interest, which in this case is the number of animals at a given site: x = 5. (Note: "f_x" is a generic term for any probability distribution. The term to the right of the equals sign is unique to the Poisson.) If you calculate this function for lambda = 3 and x = 5 animals, the result is a probability of 0.10.

$$f_5 = \frac{\exp(-3)3^5}{5*4*3*2*1} = 0.10$$

The distribution of these Poisson probabilities over a range of values of x when $\lambda = 3$ looks like this:
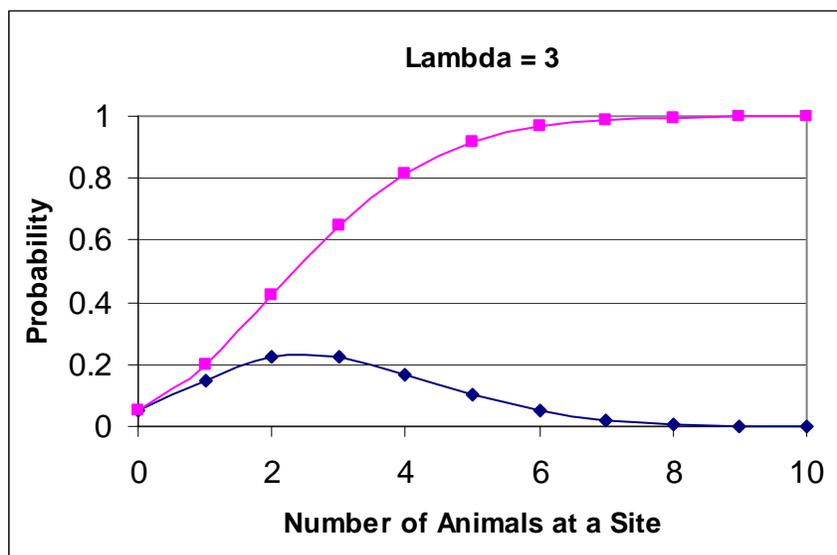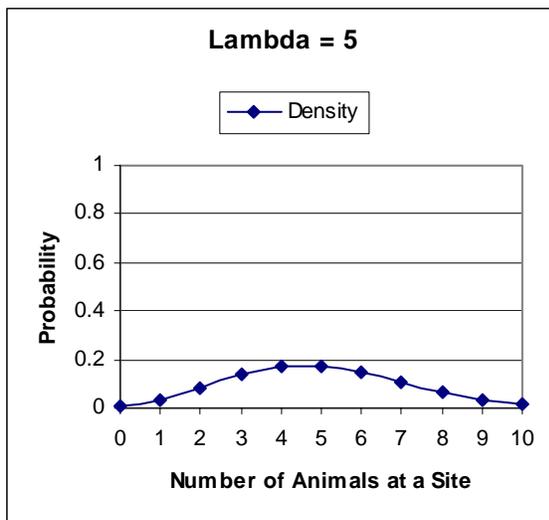
Lambda = 3

The blue points (diamonds) show the probabilities of a given site being inhabited by x individuals when the lambda is 3, where x is 0 to 10 animals and is shown on the x-axis. The graph would take a different shape if lambda were different. Notice that the peak of this blue curve is around 3. When lambda = 3, x = 3 has the highest probability of occurrence. There will be quite a few sites with 0, 1, 2, 4, and 5 animals, and fewer sites with more than 5 animals. While it is possible to have x = 8 animals at a site when lambda = 3 animals, it isn't nearly as probable as having x = 3 animals.

The pink curve (squares) shows the <u>cumulative</u> probabilities for each value of x. The pink point corresponding to x = 5 shows the probability of <u>5 or fewer</u> individuals inhabiting the site (the sum of the individual probabilities for x = 0 through 5). Since the mean is 3, most sites probably have abundances around 3 so the cumulative probability for x = 5 is quite high (0.92, in fact). This means that there is a 92% chance that a site will have 5 or fewer animals on it. The cumulative probability for x = 8 is 0.996, meaning that virtually ALL sites will have 8 or fewer animals on them when λ = 3. We can compute the probability that a site will have 100 animals on it when λ = 3, but we already know that this number will be very, very, very small, and that the cumulative probability will be 1.00 minus this extremely small number.

It's easy to generate such probabilities in Excel with the POISSON function. In this function, you enter x and lambda, and then tell Excel whether you want the cumulative probability ("true") or individual mass probability ("false"). For example, we used "=Poisson(5,3,false)" to obtain the probability of observing 5 events when

λ = 3, and we used "=Poisson(5,3,true)" to obtain the cumulative probability of observing <u>at least</u> 5 events when λ = 3.

Here are some more examples of interpretation of the Poisson distribution and its single parameter, lambda.

Lambda = 5

If lambda = 5 (as shown to the left), probabilities are highest for site abundance between, say 3 and 7 animals. Relatively many sites may have 3, 4, 5, 6, or 7 animals. The probability of x = 1 or x = 9 is still above 0, but this probability is very small. There probably will be few sites with 1, 2, 8, or 9 animals, and very little chance of a site having 0 or 10+ animals. We could carry out the function for all values of x up to ∞, and the probabilities would just get smaller and smaller as we moved away from lambda. It wouldn't take long for them to be essentially 0. Consider that, for lambda = 5 as in the graph above, the Poisson probability of x = 10 is 0.018. For x = 20 it is 0.00000027. It's very unlikely that a site would have 20 animals when lambda = 5.

The intent behind all of this is to "define" the function we will use to calculate the probability of a given level of abundance at any site. Why do we care? Because the Royle count model assumes that whether an animal is detected at a site is a function of site abundance (more on this in a bit). We need to know how likely one abundance is relative to another. Without specifying a prior distribution, we would

be saying, in effect, "We think any number of animals at this site is as likely as any other number. A site abundance of 2 is as likely as 20." This is not only uninformative, it's totally unrealistic. Spatial distributions of organism _do_ follow mechanisms that can be represented by probability distributions. By specifying a prior distribution, we can quantify the probabilities of site abundance being 2 or 20. For the Poisson, we can do this provided we know the average abundance across all sites. In practice, we won't know true abundances, so lambda is one of the key parameters estimated by the Royle count model.

THE BINOMIAL DISTRIBUTION

So, we now know that there is some true, but unknown number of animals at each site, and our goal is to estimate this number. But let's assume that we DO know how many animals occur at a site, and we'll call this number capital N. Given N number of animals occur at a site, the Royle count model then calculates the probability of observing lowercase n animals at that site with a Binomial probability. For example, if N = 10, the binomial function can be used to calculate the probability of detecting 0, 1, 2, 3, ... 10 animals at the site. You studied the binomial function in Exercise 1, so this should be a review.

The binomial distribution is widely used for problems where there are a fixed number of tests or trials (N) and when each trial can have only one of two outcomes (e.g., success or failure, detect or don't detect, heads or tails). The formula is written in the orange box below:

$$\text{BINOMIAL:} \quad f(n \mid N, p) = \binom{N}{n} p^n (1-p)^{N-n}$$

In this formulation, the number of successes is denoted as n, and the probability of success is usually denoted as p. A typical example considers the probability of getting 3 heads, given 10 coin flips and given that the coin is fair (p = 0.5). The binomial probability function is written $f(3|10, 0.5)$, where the vertical bar | means "given" and is read, "the probability of observing 3 heads, *given* 10 coin flips and the probability of a head (success) is 0.5."

Let's break the right hand side of the binomial probability function into pieces. The portion p^n and $(1-p)^{N-n}$ gives p (the probability of success, or heads) raised to the number of times the success occurred (n) and 1-p (the probability of a failure, or tails) raised to the number of times the failures occurred (N – n). But if you flip a fair coin 10 times, there are many ways you could end up with three heads. For instance, the first three tosses could be heads and the rest could be tails (HHHTTTTTTT). Or the first seven could be tails and the last three could be heads (TTTTTTTHHH). Or you could alternate getting heads and tails (e.g., THTHTHTTTT). The portion of the binomial probability function in brackets is called the binomial coefficient, and accounts for ALL the possible ways in which three heads and seven tails could be obtained. Now we can calculate the probability of observing 3 heads, given 10 coin flips and a fair coin as:

$$\text{BINOMIAL:} \quad f(3|10,0.5) = \binom{10}{3} 0.5^3 (1-0.5)^{10-3} = 0.117$$

A graph of the binomial distribution when N = 10 and p = 0.5 is shown below.

Binomial Distribution, N = 10, p = 0.5
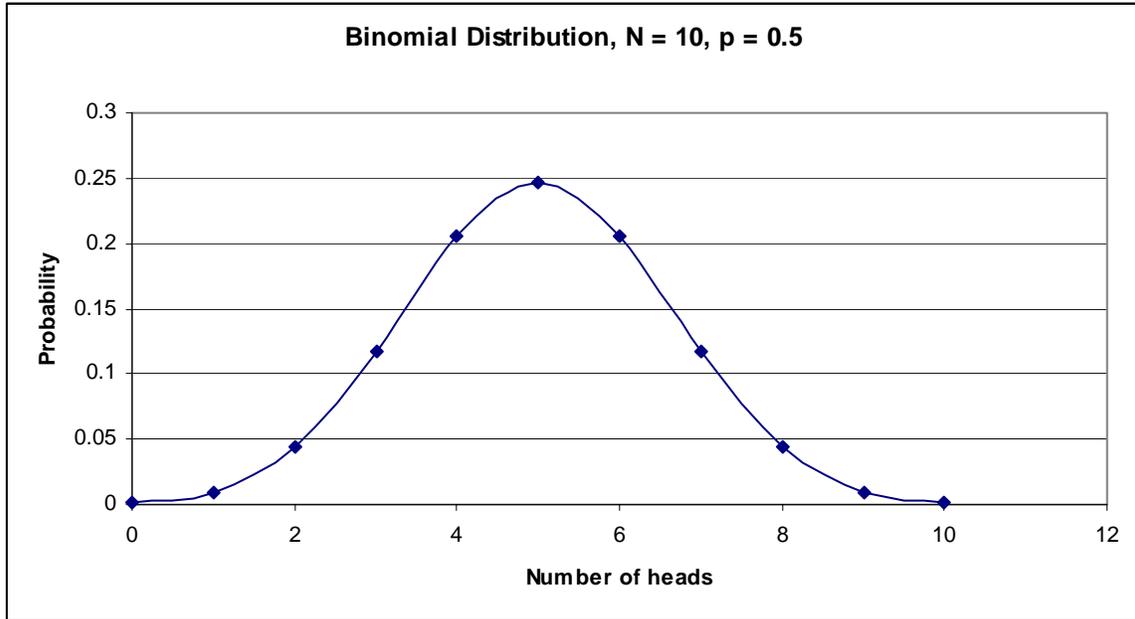
This graph shows the probability of observing lowercase n successes (heads) out of 10 coin flips when the coin is fair. Note that the x axis ranges from 0 to N (10). When the coin is fair and is flipped 10 times, it's most likely that you will end up with 5 heads (probability = 0.246). However, 4 or 6 heads is also likely (probability = 0.205). It's less likely that you will end up with 3 heads (probability = 0.117), although it is certainly possible.

The graph above shows the binomial distribution for N = 10 and p = 0.5. This distribution would change if either N or p changes. For example, below is the binomial distribution for N = 10 and p = 0.2.

Binomial Distribution, N = 10, p = 0.2
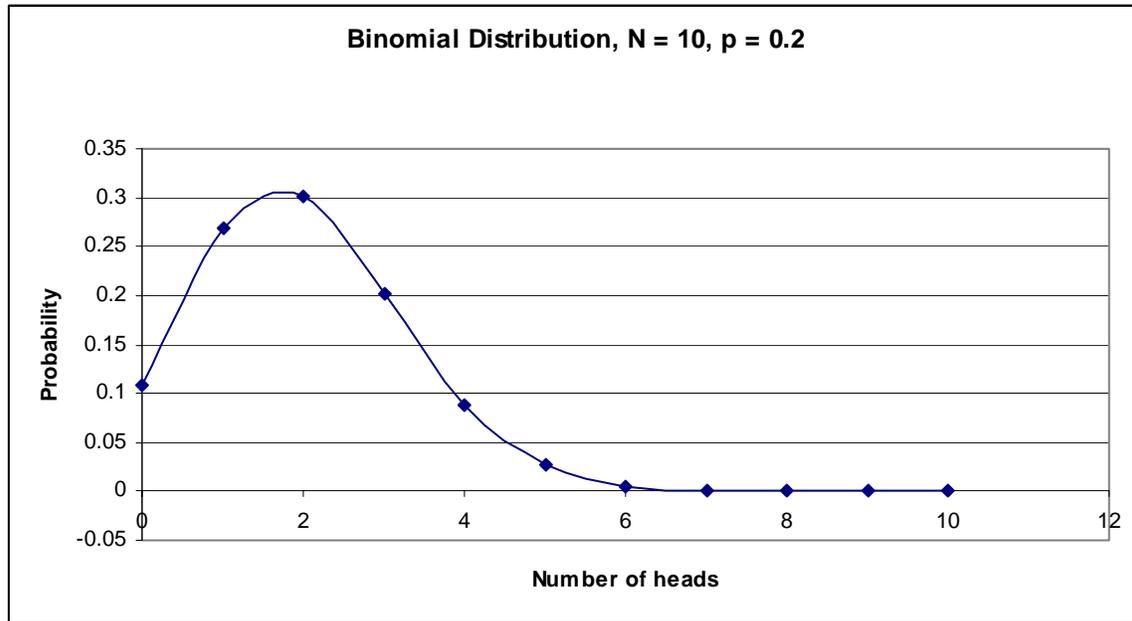
In this example, it's far more probable that you will end up with 0, 1, 2, 3, or 4 heads out of 10 coin flips, and less likely that you will end up with 5 or more heads.

Now let's stop thinking about coin flips and apply the binomial function to surveys of animals. In the Royle count model, the binomial probability $f(3|10, 0.5)$ would consider the probability of observing n = 3 animals, given that N = 10 animals occur at the site and the probability of detecting an animal (p) is 0.5. The number of successes, n, is the number of animals that were detected in a survey. The total number of trials, N, in this exercise represents the total number of animals that actually occur on the site (governed by the Poisson distribution). The probability of success, p, is the probability of detecting an individual animal, given it is present on the site. Thus, if our site has 10 animals on it, and we conduct a survey, we can compute the probability of observing n successes (detections) over 10 trials, given p. So

$f(3|10, 0.5)$ is the probability of observing 3 animals when 10 animals occur, given the probability of detecting an individual is 0.5. If there are 10 animals on the

site, and we are in the field conducting a survey, our data might "look" like this: 1110000000, or 0000000111, or 0101010000. All of these options describe the observation of 3 individuals when 10 individuals actually occur. Of course, we only record the 3 animals we actually observe and do not really know how many we missed and so can't record the 0's.

It's easy to generate binomial probabilities in Excel with the BINOMDIST function. This particular function has four arguments:
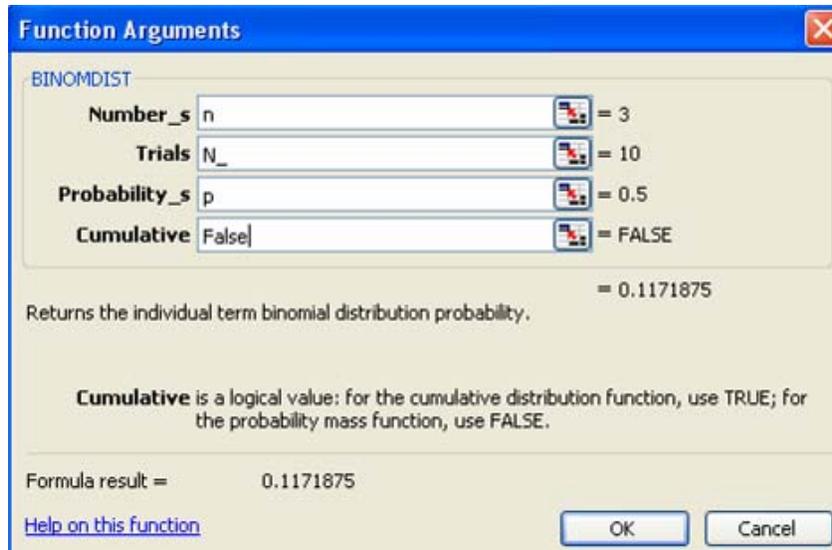
Number_s is the number of successes in trials, n (such as 3 animals detected), Trials is the number of trails, N (such as 10 animals truly on a site), Probability_s is the probability of a success, p (such as the probability of detecting an animal), and Cumulative is where you specify whether you want the cumulative binomial probability (true) or whether you want the probability mass function (false). As you can see, the result for $f(3|10, 0.5) = 0.117$ as we've seen before.

ASSUMPTIONS OF THE BINOMIAL DISTRIBUTION

Two major assumptions of the binomial distribution are that the trials are independent, and probability of success is constant throughout the experiment. In the Royle count model, a single "experiment" is a single survey. In the real world, both of these assumptions can be violated. If you flip a penny, the outcome of the next flip will be completely independent of the outcome of the first flip. But animals are not pennies. Pair bonds and family associations are examples of how the detection of one individual during a given survey can be linked to the detection of a second individual in that same survey period, resulting in extra binomial variation. Additionally, p may not be constant over the course of the experiment….during your survey, any given animal may be highly detectable during one minute and elusive the next. How to deal with these problems is covered later.)

THE ROYLE COUNT MODEL DATA INPUT

OK, now that you've had a refresher course on the Poisson and Binomial distributions, we can forge ahead with our main objective. Our goal, ultimately, is to estimate the abundance of animals across R total sites, each surveyed a total of T times. Our raw field data for R = 20 and T = 5 might look like this:

	A	B	C	D	E	F
7				Visit		
8	Site	1	2	3	4	5
9	1	2	1	1	2	2
10	2	1	2	2	1	2
11	3	0	4	1	1	1
12	4	2	0	0	0	1
13	5	1	0	2	1	2
14	6	0	0	1	0	0
15	7	0	1	0	0	0
16	8	1	0	0	2	1
17	9	2	3	1	1	0
18	10	0	0	0	1	0
19	11	2	2	1	0	0
20	12	1	0	1	1	1
21	13	1	0	0	1	0
22	14	2	0	0	2	2
23	15	1	1	1	1	1
24	16	1	2	0	1	1
25	17	0	0	1	0	0
26	18	1	1	2	0	1
27	19	1	0	0	1	0
28	20	2	1	0	0	0

The total detections for site i at time t is denoted as n_{it}. Thus, for site 1 and survey 1, n_{11} = 2 animals. For site 1 and survey 3, n_{13} = 1. Across the 5 surveys, the maximum number of detections for site 1 was 2. For site 6 and survey 2, n_{62} = 0. The maximum number of detections for site 6 was 1. Each site therefore has an "encounter history" which is made up of the total counts on a survey by survey basis. Site 1 has the history 2 1 1 2 2. Site 3 has the history 0 4 1 1 1, so the maximum count across all surveys for site 3 was 4.

To estimate abundance from this data, remember our two key assumptions: we assume that there is some number (it could be 0) of individuals actually inhabiting each site (N_i), which is governed by the Poisson distribution. We also assume that whether or not you detect the target at that site is going to be a function of the species-specific detection probability (p).

THE ROYLE COUNT MODEL

OK, given these two major assumptions, the Royle count model can be expressed as:

$$L(p, \theta \mid \{n_{it}\})$$
$$= \prod_{i=1}^{R} \left\{ \sum_{N_i = \max_t n_{it}}^{\infty} \left(\prod_{t=1}^{T} \text{Bin}(n_{it}; N_i, p) \right) f(N_i; \theta) \right\},$$

which looks incredibly complicated until you break it down into pieces. Let's say it in words first, and then we'll show you how to set this model in the spreadsheet.

The left side of the equation, $L(p, \theta \mid \{n_{it}\})$ says, "The likelihood of p (the probability of a success, or the probability of detecting an individual that is present at the site) and θ (the mean of the Poisson distribution), given the observed field data $\{n_{it}\}$." Note that Andy Royle calls the mean of the Poisson θ instead of the usual λ, so we'll stick with his notation from this point on. Thus, the Royle count model estimates two critical parameters, p and θ, with maximum likelihood analysis.

The right side of the equal sign is best described by describing the individual pieces. Let's start with the term

$$\left(\prod_{t=1}^{T} Bin(n_{it}; N_i, p) \right)$$

which focuses on the survey results for a <u>single</u> site. The Π symbol means "multiply" (like the Σ symbol means "sum"). So we are going to multiply some numbers together -- specifically 5 numbers, from survey 1 (t = 1) to the final survey (T = 5 in our example). Now, exactly what numbers will we multiply? We will multiply the binomial probability of detecting n_{it} animals (successes) out of N total

animals at the site, given the probability of detection is p, which is computed for each of the 5 surveys. For example, if there are 5 animals at the site 1 and p = 0.4, and we detect 2 of them in survey 1, 1 in survey 2, 0 in survey 3, 3 in survey 4, and 3 in survey 5, we simply multiply the binomial probabilities:

	G	H	I	J	K	L
19	p = 0.4	Survey Results for Site 1				
20	Site	1	2	3	4	5
21	No. Animals Detected	2	1	0	3	3
22	Binomial Probability	0.3456	0.2592	0.07776	0.2304	0.2304

For site 1, if Ni = 5 and p = 0.4, the term

$$(\prod_{t=1}^{T} Bin(n_{it}; N_i, p))$$

is equal to 0.3456 * 0.2592 * 0.0776 * 0.2304 * 0.2304, which is 0.000369769. There! That wasn't so bad, was it? This particular result is for probability of getting a 2 1 0 3 3 history for site 1, given that N_i = 5 and p = 0.4. Of course, if we changed either N_i or p, our results would be different. Find this term in the Royle count model below:

$$L(p, \theta \mid \{n_{it}\}) = \prod_{i=1}^{R} \left\{ \sum_{N_i = \max_t n_{it}}^{\infty} \left(\prod_{t=1}^{T} Bin(n_{it}; N_i, p) \right) f(N_i; \theta) \right\},$$

OK, in the Royle count model, this result is multiplied by a term, $f(N_i; \theta)$. Hopefully, you remember that this is the Poisson probability that there are actually N_i individuals at site i, given that the mean abundance across all sites is θ.

So, to carry our previous example through to the end, the probability of observing a 2 1 0 3 3 history for site 1, given N_i is 5 and p is 0.4 is

0.3456 * 0.2592 * 0.0776 * 0.2304 * 0.2304 * the Poisson probability that the site contains N_i individuals, given θ. If θ = 3, the Poisson probability that a site contains 5 animals is 0.100818813, and the term

$$(\prod_t^T Bin(n_{it}; N_i, p)) f(N_i; \theta)$$

would be 0.3456 * 0.2592 * 0.0776 * 0.2304 * 0.2304 * 0.100818813.

If θ = 9, the Poisson probability that a site contains 5 animals is 0.0607. In this case, the term

$$(\prod_t^T Bin(n_{it}; N_i, p)) f(N_i; \theta)$$

would be 0.3456 * 0.2592 * 0.0776 * 0.2304 * 0.2304 * 0.0607.

Make sense? In other words, we essentially weigh the product of our 5 binomial probabilities by the Poisson probability that the site actually contains N_i individuals, given θ.

You might be wondering, "how will we know what Ni is for a given site when all we record is the number we actually observe?" Well, we don't know what N_i is. All we really know is the maximum number we observed across the 5 surveys. So we play a little "what if" game. We essentially say, "what if N_i is 1? What if N_i is 2? What if N_i is 3? What if Ni is 100?" And then we compute the following portion of the Royle count model likelihood function

$$\left(\prod_t^T Bin(n_{it}; N_i, p)\right) f(N_i; \theta)$$

for ALL possible values of N_i.

That's what the term, $\sum\limits_{Ni=\max; n_{it}}^{\infty}$ essentially indicates: sum across all possible values of

N_i (from the maximum count at the site to infinity), and add them together.

$$\sum_{Ni=\max n_{it}}^{\infty} \left(\prod_{t=1}^T Bin(n_{it}; N_i, p) f(Ni; \theta)\right)$$

$$\sum_{Ni=\max n_{it}}^{\infty} \left(\prod_{t=1}^T Bin(n_{it}; N_i, p) f(Ni; \theta)\right)$$

Realistically, you can let N_i go from 0 to about 100 (that's what we have in the spreadsheet). You may want to let N_i extend to 200 if you are dealing with a very common animal. You start with the maximum value observed in survey for the site because this is the minimum number of animals that are KNOWN to occur on the site. (Remember, the surveys were conducted under the assumption of closure, so if you detect a maximum of 4 animals at a site during any given survey, you know that 4 is the minimum number of animals that can occur at that site. Then you let N_i go from 4 to 100.)

Let's try it with our previous example. Site 1 had a history of 2 1 0 3 3. Let's assume p = 0.4 and θ = 3, and we'll calculate the binomial product times the Poisson probability for possible values of Ni, starting at 3 and ending at 100. Then, the Royle count model

$$\sum_{Ni=\max n_{it}}^{\infty} (\prod_{t=1}^{T} Bin(n_{it}; N_i, p) f(Ni; \theta)$$

becomes

$$\sum_{3}^{100} (\prod_{t=1}^{T} Bin(n_{it}; N_i, 0.4) f(N_i; 3)$$

for site 1. So, for each value of Ni, we will estimate

$$(\prod_{t}^{T} Bin(n_{it}; N_i, p)) f(N_i; \theta)$$

for site 1, given p = 0.4 and θ = 3, and then will add the results across all possible values of Ni.

OK, to this point our focus has been on a single site. Naturally, we will do this for all R sites in the study. The final step is to multiply the site-by-site results together, as indicated by the Π symbol, starting with site i = 1 and ending with site R. Thus, you now have all the parts of the Royle Count Model formulation:

$$L(p, \theta \mid \{n_{it}\})$$
$$= \prod_{i=1}^{R} \left\{ \sum_{N_i = \max_t n_{it}}^{\infty} \left(\prod_{t=1}^{T} \mathrm{Bin}(n_{it}; N_i, p) \right) f(N_i; \theta) \right\},$$

THE ROYLE COUNT MODEL SPREADSHEET INPUTS

Whew! We've covered a lot of ground. Some of the concepts may seem a bit blurry to you, but hopefully the concepts we discussed previously will become more clear as we work through the spreadsheet. Open up your spreadsheet and click on the sheet named Royle Count Model.

In this example, we will be estimating mean abundance across R = 20 sites that were surveyed a total of T = 5 times. The data are given in cells B9:F28.

|	A	B	C	D	E	F
7				Visit		
8	Site	1	2	3	4	5
9	1	2	1	1	2	2
10	2	1	2	2	1	2
11	3	0	4	1	1	1
12	4	2	0	0	0	1
13	5	1	0	2	1	2
14	6	0	0	1	0	0
15	7	0	1	0	0	0
16	8	1	0	0	2	1
17	9	2	3	1	1	0
18	10	0	0	0	1	0
19	11	2	2	1	0	0
20	12	1	0	1	1	1
21	13	1	0	0	1	0
22	14	2	0	0	2	2
23	15	1	1	1	1	1
24	16	1	2	0	1	1
25	17	0	0	1	0	0
26	18	1	1	2	0	1
27	19	1	0	0	1	0
28	20	2	1	0	0	0

That's it in terms of data entry, and these are the same data that you will paste into PRESENCE later on.

ROYLE COUNT MODEL PARAMETERS AND OUTPUTS

At the top of the spreadsheet, you'll see a section labeled Inputs, Parameters, and Outputs.

	B	C	D	E	F
1	Inputs, Parameters, Outputs				
2	p	p beta	θ	θ beta	N hat
3	0.5		1.0000		0.00000
4	Log L	-2Log L	K	AIC	AICc
5	-130.838	261.677	2	265.67684	261.677

The key parameters that will be estimated in this model are p (cell B3) and theta
(θ), in cell E3. The goal is to find estimates of p (detection probability) and θ (the
average abundance of animals across the R sites) that will generate results that
closely match the field data that we collected. We don't know what these values
are....Solver will find them for us. But Solver won't find these directly...instead it
will work on the betas that are linked to these estimates. As a very quick
refresher, p is a probability that is bounded between 0 and 1, while θ is a positive
integer. If we plan to do some linear modeling (that is, constrain p or θ to be a
function of predictor variables, such as habitat, time of year, etc. within the model
itself), we need to unbound these parameters so that they range from plus infinity
to minus infinity. To achieve this, we use a logit transformation for p (which has
the form exp(beta)/(1+exp(beta)) and we use the log transformation for θ (which
has the form exp(beta)). Thus, Solver will find a beta for p (cell C3) and a beta for
θ (cell E3), and then will back-transform these betas into a probability (p; cell B3)
or into a positive integer (θ; cell D3). The picture below shows how this works. On
the x-axis are possible beta values. The transformed θ values associated with each
beta are shown in the diamonds (center axis), while the transformed p values
associated with each beta are shown in the squares (right axis).
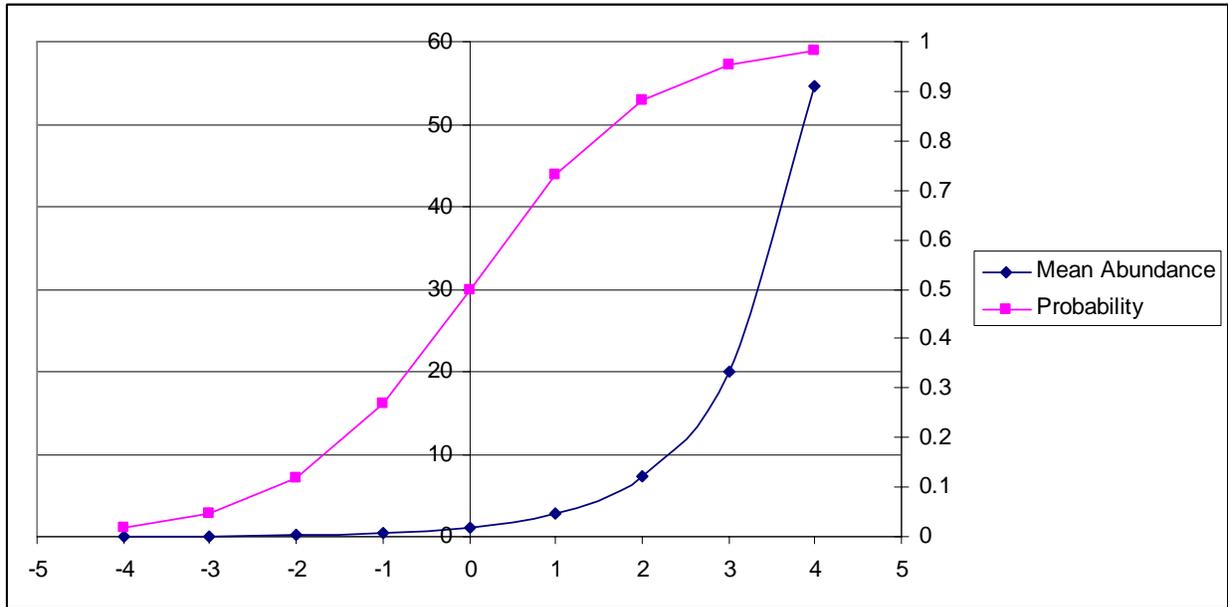
Note that beta values < 0 correspond to θ values close to 0. If beta = 0, average site abundance is 1, and if beta = 2, average site abundance = 7.4. Note that, for p, beta values < -4 correspond to p = 0 while beta values > 4 correspond to p = 1. So if you run Solver and find values outside this range, it should be a flag that something is amiss.

COMPUTING THE LOG LIKELIHOOD FOR A SINGLE SITE

OK, now we get to the heart of the spreadsheet, which is where we let Ni for each site range from 0 to 100. Then, for each value of Ni, we compute the product of the binomial probabilities, given p in cell B3, and then multiply the result by the Poisson probability that Ni animals occur at the site, given θ in cell D3. Let's focus on site 1 only. Here are the survey results from site 1:

	A	B	C	D	E	F
7		Visit				
8	Site	1	2	3	4	5
9	1	2	1	1	2	2

So, we know that at least 2 animals occur on site 1. Now scroll down to row 32. In this row, we list the potential N_i that can occur at a site. We start with 0 in cell B32, and end with 100 all the way over in cell CX32. Thus, column B returns a result when the number of individuals at the site is hypothetically 0 animals, column C returns a result when the number of individuals at the site is hypothetically 1 animal, column F is the returns a result when the number of individuals at a site is hypothetically 5 animals, and so on.

	A	B	C	D	E	F	G
31		Potential N Max =>>>					
32	Site	0	1	2	3	4	5
33	1	#NUM!	#NUM!	0.000718515	0.000454685	5.05206E-05	2.2841E-06

Now, click on cell B33, and you'll see the formula
=BINOMDIST($B9,B$32,$B$3,FALSE)*BINOMDIST($C9,B$32,$B$3,FALSE)*BINOMDIST($D9,B$32,$B$3,FALSE)*BINOMDIST($E9,B$32,$B$3,FALSE)*BINOMDIST($F9,B$32,$B$3,FALSE)*POISSON(B32,$D$3,FALSE). Wow! This is

simply $(\prod\limits_{t}^{T} Bin(n_{it}; N_i, p)) f(N_i; \theta)$ for site 1. You can see that the formula returns

#NUM!, indicating that there is a problem with this particular equation. Let's work through the formula so we can understand why this happened. First, notice that the formula
=BINOMDIST($B9,B$32,$B$3,FALSE)*BINOMDIST($C9,B$32,$B$3,FALSE)*BINOMDIST($D9,B$32,$B$3,FALSE)*BINOMDIST($E9,B$32,$B$3,FALSE)*BINOMDIST($F9,B$32,$B$3,FALSE)*POISSON(B32,$D$3,FALSE)
is really 5 binomial probabilities and 1 Poisson probability multiplied together. We've color-coded them above so you can clearly see this. This is exactly what we

need to compute $(\prod_{t}^{T} Bin(n_{it}; N_i, p)) f(N_i; \theta)$ in the Royle count model; we multiply

the t → T binomial probabilities by a Poisson probability.

The first term, BINOMDIST($B9,B$32,$B$3,FALSE), returns the binomial

probability of observing 2 animals in survey 1 (cell B9), given 0 animals occur at the

site (cell B32) and given p in cell B3. Remember that the argument FALSE

indicates that we want the individual, mass probability, and not the cumulative

probability. What's wrong with this? Well, if 0 animals occur on a site (N_i = 0),

there is no way you can observe 2 animals; hence the #NUM! result. Don't let this

throw you…it won't affect our calculations. The next term,

BINOMDIST($C9,B$32,$B$3,FALSE), returns the binomial probability of

observing 1 animal in survey 2 (cell C9), given 0 animals occur at the site (cell B32)

and given p in cell B3. Again, this is a problem because you can't observe 1 animal if

none occur on the site. The three other BINOMDIST functions do essentially the

same thing, but target the number of animals observed in surveys 3, 4, and 5. The

final term, POISSON(B32,$D$3,FALSE), returns the binomial probability that Ni

= 0 (cell B32), given the θ estimate in cell D3.

Now click on cell D33, and you'll see the same basic formula, but this time Excel

returns a number. The equation is

=BINOMDIST($B9,D$32,$B$3,FALSE)*BINOMDIST($C9,D$32,$B$3,FALSE)*B

INOMDIST($D9,D$32,$B$3,FALSE)*BINOMDIST($E9,D$32,$B$3,FALSE)*BI

NOMDIST($F9,D$32,$B$3,FALSE)*POISSON(D32,$D$3,FALSE), and the result

is 0.000718515.

	A	B	C	D
31		Potential N Max =>>>		
32	Site	0	1	2
33	1	#NUM!	#NUM!	0.000718515

In this case, Excel was able to estimate the BINOMDIST function when N_i = 2 because we observed 2 animals at the site. The first term, BINOMDIST($B9,D$32,$B$3,FALSE) returns the binomial probability of observing 2 animals at site 1 in survey 1, given that N_i is 2 (cell D32) and the p estimate in cell B3. The second term BINOMDIST($C9,D$32,$B$3,FALSE) returns the binomial probability of observing 1 animal at site 1 in survey 2, given that N_i is 2 (cell D32) and the p estimate in cell B3. Make sense? We hope so!

	CY	CZ
32	Sum	Ln (sum)
33	0.001226063	-6.703947094
34	0.001226063	-6.703947094
35	9.94236E-07	-13.82129146
36	0.000376453	-7.884717619
37	0.000878729	-7.037034473
38	0.011861162	-4.434485903
39	0.011861162	-4.434485903
40	0.000770462	-7.168520826
41	5.63676E-05	-9.783616492
42	0.011861162	-4.434485903
43	0.000411929	-7.794659131
44	0.014525652	-4.231839067
45	0.012231823	-4.403714244
46	0.0002334	-8.362758077
47	0.017714292	-4.033383503
48	0.00159432	-6.441308039
49	0.011861162	-4.434485903
50	0.00159432	-6.441308039
51	0.012231823	-4.403714244
52	0.000376453	-7.884717619
53	-130.8384206	-130.8384206

Now, scroll across row 33, which contains the result of $(\prod_t^T Bin(n_{it}; N_i, p)) f(N_i; \theta)$ for each and every value of Ni, from 0 to 100. Click on cell CY33 and you'll see the formula =SUMIF(B33:CX33,">0"). This function accomplishes this portion of the Royle count model: $\sum_{Ni=\max n_{it}}^{\infty} (\prod_{t=1}^{T} Bin(n_{it}; N_i, p) f(Ni; \theta)$ for site 1.

**COMPUTING THE LOG LIKELIHOOD ACROSS ALL SITES**

Now, we do the exact same thing for each of the remaining sites. The sum of the binomial and Poisson products for each site is given in cells

CY33:CY52. The natural log of the result is computed in cells CZ33:CZ52 with a LN function.

The final step is to compute the final likelihood. This can be done in two ways. First, in cell CY53, the formula computes the likelihood of the data, given p and θ, as =LN(PRODUCT(CY33:CY52)). First, the product of cells CY33:CY52 is calculated, and the natural log of this result is the likelihood. Second, in cell CZ53, the sum of the natural logs is calculated as =SUM(CZ33:CZ52). Either way the result is the same. Now, the only thing left to do is to maximize this result to obtain maximum likelihood estimates for and θ.

MAXIMIZING THE LOG LIKELIHOOD

To maximize the likelihood, all we need to do is open Solver, and set cell CY53 to a maximum by changing cells C3, E3.
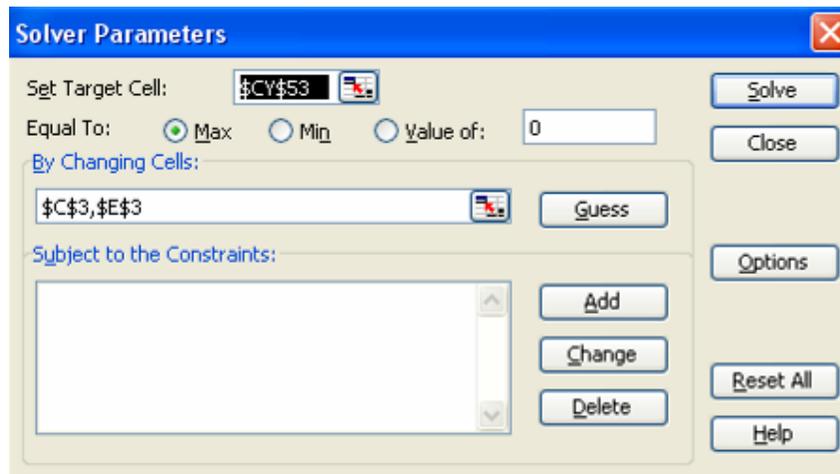
Press Solve, and Solver will work through many combinations of betas until it finds a solution that provides the maximum likelihood. Here are our results:

	B	C	D	E	F
1	Inputs, Parameters, Outputs				
2	p	p beta	θ	θ beta	N hat
3	0.164643094	-1.624078998	5.1019	1.629621246	102.03884
4	Log L	-2Log L	K	AIC	AICc
5	-115.108	230.217	2	234.21651	233.137

From our field data, Solver found p = 0.1646 (cell B3), indicating that if an individual occurs at a site, it has a 16.46% chance of being detected by the observer on a survey. That's pretty low. The average abundance of animals across the 20 sites is 5.1019 (cell D3). N hat (cell F3) is estimated for the 20 study sites as θ * R, or 5.1019 * 20 = 102.04. The maximized $Log_e L$ for these data is -115.108 (cell B5), the $-2Log_e L$ is 230.217 (cell C5). K is 2 for this model because we estimated two parameters, p and θ. AIC is computed in cell E5 as $-2Log_e L + 2K$, and AICc is the second order correction.

Really, that's all there is to it. In most cases, you'll want to add covariates to the model, and you can easily do so with this model. (In fact, the sheet labeled "Covariate Model" provides an example of how to include covariates. We won't cover it in detail here, but the same principles that we covered in the basic occupancy model (Chapters 4 and 5) apply to Royle count model as well.

SIMULATING REPEATED COUNT DATA

The last thing we need to do before running the analysis in PRESENCE is to briefly describe how data can be simulated for the Royle count model. Click on the sheet labeled "Simulate Data." The key inputs for this model are in cells B2 (where you enter lambda, or the average abundance of animals across sites) and B3 (where you enter p, or the probability of detecting an animal at a site, given it is present).

The spreadsheet is set up for situations when the total number of animals at a site does not exceed 10 (so keep lambda around 5 or less). Also, this model performs well when p is not very high or very low, so enter a p value that is somewhat in the middle-ground.

|	A	B
1	**Data Simulation Inputs**	
2	Lambda =	3
3	p	0.2

Once you have settled on some values, the next step is to "populate" the study sites with animals. For example, the grid in cells F8:J27 shows the abundance of animals in 20 sites for all 5 time periods. Note that the abundance is constant across periods….per the assumption that the population is demographically closed during the survey period. This grid represents the true abundance of animals at each site…a field observer will detect only a portion of these animals. So, how do we "populate" the sites?

	E	F	G	H	I	J
4				Simulate Data		
5						
6				True Abundance (ni)		
7	Site	1	2	3	4	5
8	1	1	1	1	1	1
9	2	3	3	3	3	3
10	3	5	5	5	5	5
11	4	3	3	3	3	3
12	5	1	1	1	1	1
13	6	2	2	2	2	2
14	7	2	2	2	2	2
15	8	3	3	3	3	3
16	9	2	2	2	2	2
17	10	0	0	0	0	0
18	11	2	2	2	2	2
19	12	7	7	7	7	7
20	13	0	0	0	0	0
21	14	2	2	2	2	2
22	15	2	2	2	2	2
23	16	2	2	2	2	2
24	17	2	2	2	2	2
25	18	5	5	5	5	5
26	19	3	3	3	3	3
27	20	2	2	2	2	2
28	max =	7				

First, take a look at cells A5:C17.

	A	B	C
5		Density	Cumulative
6	Number		0
7	0	0.049787068	0.049787068
8	1	0.149361205	0.199148273
9	2	0.224041808	0.423190081
10	3	0.224041808	0.647231889
11	4	0.168031356	0.815263245
12	5	0.100818813	0.916082058
13	6	0.050409407	0.966491465
14	7	0.021604031	0.988095496
15	8	0.008101512	0.996197008
16	9	0.002700504	0.998897512
17	10	0.000810151	0.999707663

These cells give the POISSON probabilities that a site will have 0, 1, 2, …10 animals at the site, given the lambda value you entered in cell B2. Cells B7:B17 give the mass probability function, while cells C7:C17 give the cumulative Poisson probability. Click on cell B7 and you'll see the formula =POISSON(A7,$B$2,FALSE), which returns the probability that a site will contain 0 animals for the lambda specified in cell B2. Click on cell C7 and you'll see the formula =POISSON(A7,$B$2,TRUE), which returns the probability that a site will contain AT LEAST 0 animals for the lambda specified in cell B2. These formulae are copied down for 10 animals. (If you want to simulate data where sites can have more animals, just copy these formula down as far as you need to. A graph showing the mass and cumulative POISSON probabilities is included on the spreadsheet to help you visualize what how animals might be distributed across the 20 sites. Below is the graph for lambda = 3.
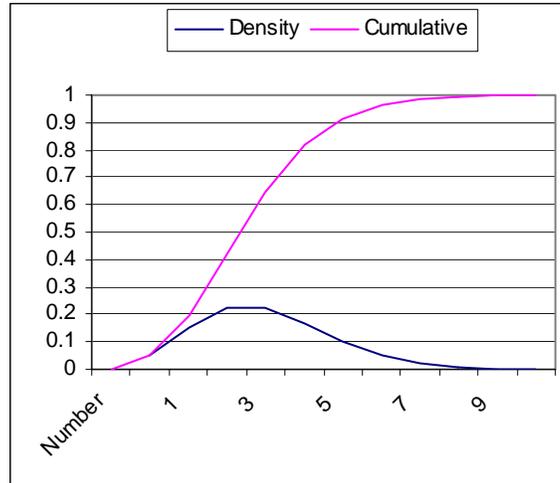
Now, how do we use this information to "populate" our 20 sites? Easy. Click on cell F8 and you'll see the formula =LOOKUP(RAND(),$C$6:$C$17,$A$7:$A$17). This populates site 1 with a certain number of animals. The function uses a LOOKUP function. The function looks up a random number, RAND(), in series of cells C6:C17, which is the cumulative Poisson probabilities. Because the cumulative probabilities are ordered, LOOKUP doesn't need to find an exact match...it finds the closest match and then returns the associated abundance listed in cells A7:A17. This formula is copied down for all 20 sites. Now, when you press F9, the calculate key, Excel will draw new random numbers and you'll see the site abundances change; the mean should stay roughly the same however. We only need to "populate" cells for survey 1. After survey 1, we know that the number of animals at each site stays the same across all 5 surveys, so we just enter equations so that the abundance of animals on a site during surveys 2-5 is the same as the abundance of animals at the site on survey 1. Get it? By using the cumulative Poisson probabilities, we end up with 20 sites with different numbers of animals, but the mean abundance across sites should be similar to the lambda value entered in cell B2.

OK, now that we know how many animals actually occur on each site, we need to figure out how many animals a field technician will count during a survey, given p (the probability of detection) in cell B3. But before we do that, scroll down to cells L35:W47.

	L	M	N	O	P	Q	R	S	T	U	V	W
34	BINOMIAL						N Trials					
35		0	1	2	3	4	5	6	7	8	9	10
36	n = successes	0	0	0	0	0	0	0	0	0	0	0
37	0	1	0.8	0.64	0.512	0.4096	0.32768	0.262144	0.2097152	0.16777216	0.134217728	0.107374182
38	1	#NUM!	1	0.96	0.896	0.8192	0.73728	0.65536	0.5767168	0.50331648	0.436207616	0.375809638
39	2	#NUM!	#NUM!	1	0.992	0.9728	0.94208	0.90112	0.851968	0.79691776	0.738197504	0.677799526
40	3	#NUM!	#NUM!	#NUM!	1	0.9984	0.99328	0.98304	0.966656	0.9437184	0.914358272	0.879126118
41	4	#NUM!	#NUM!	#NUM!	#NUM!	1	0.99968	0.9984	0.995328	0.9895936	0.98041856	0.967206502
42	5	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	1	0.999936	0.9996288	0.99876864	0.996933632	0.993630618
43	6	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	1	0.9999872	0.99991552	0.999686144	0.999135642
44	7	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	1	0.99999744	0.999981056	0.999922074
45	8	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	1	0.999999488	0.999995802
46	9	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	1	0.999999898
47	10	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	1

This symmetrical grid provides the cumulative probability of observing n successes in N total trials, given p provided in cell B3. The trials are listed in cells M35:W35, and go from 0 to 10. As we've said previously, this corresponds to 0 to 10 animals that may occur on any given site. If you want to simulate data where sites can have more than 10 animals, feel free to extend this grid. The number of successes are listed in cells L37:L47, and range from 0 to 10. Now, click on cell M37 and you'll see the function =BINOMDIST($L$37,M35,$B$3,TRUE). For this first grid cell, we compute the binomial probability of observing 0 animals at a site (n = 0), given N = 0. The argument TRUE indicates that we want the cumulative Binomial probability. The answer should be 1. If N is 0, then the probability that n = 0 is 1. Now click on cell N37 and you'll see the function =BINOMDIST($L$37,N35,$B$3,TRUE). This function returns the cumulative binomial probability of 0 successes (n) given 1 trial (N = 1) and given p in cell B3. The result is 0.8. Cell N38 has the formula =BINOMDIST($L$38,N35,$B$3,TRUE), and returns the cumulative binomial probability for at LEAST 1 success (0 + 1), given 1 trial. That answer is also 1. Notice that the diagonal on the grid contains the result, 1. This is because the

cumulative probability of getting at least n successes in N trials is 1 whenever n = N. Work your way through this table. You'll notice that, for any given N the cumulative probabilities of detecting n animals are ordered – we'll use these cumulative probabilities to derive field data.

Now let's take a look at how we use the cumulative BINOMDIST function to derive "field data". Let's take a look at how to do that. In cells M8:Q27, we derive field data for sites where the true abundance is 0 to 5.

	L	M	N	O	P	Q
6		Field Data (Abundance = 0 to 5)				
7	Site	1	2	3	4	5
8	1	0	1	0	1	0
9	2	1	0	1	0	0
10	3	FALSE	FALSE	FALSE	FALSE	FALSE
11	4	FALSE	FALSE	FALSE	FALSE	FALSE
12	5	1	0	1	0	0

In cells S8:W27, we derive field data for sites where the true abundance is 6-10.

	R	S	T	U	V	W
6		Field Data (Abundance = 6 to 10)				
7	Site	1	2	3	4	5
8	1	FALSE	FALSE	FALSE	FALSE	FALSE
9	2	FALSE	FALSE	FALSE	FALSE	FALSE
10	3	1	2	2	1	3
11	4	1	2	0	0	1
12	5	FALSE	FALSE	FALSE	FALSE	FALSE

Only the first five sites are shown above. We had to split this part into the two sections because Excel can handle only so many nested IF functions, which we'll cover in a second.

Click on cell M8 and you'll see the formula
=IF(F8=0,LOOKUP(RAND(),$M$36:$M$47,$L$37:$L$47),IF(F8=1,LOOKUP(RAND
(),$N$36:$N$47,$L$37:$L$47),IF(F8=2,LOOKUP(RAND(),$O$36:$O$47,$L$37:

$L$47),IF(F8=3,LOOKUP(RAND(),$P$36:$P$47,$L$37:$L$47),IF(F8=4,LOOKUP(RAND(),$Q$36:$Q$47,$L$37:$L$47),IF(F8=5,LOOKUP(RAND(),$R$36:$R$47,$L$37:$L$47)))))))). This formula is a series of 6 nested IF functions. Remember, an IF function in Excel has three arguments: IF(logical test, value if true, value if false). Let's look at the very first IF function. The logical test is =IF(F8=0). If site 1's true abundance is 0, then lookup a random number in the cells M36:M47, and return the number of successes (detections) in corresponding cells L37:L47. If site 1's abundance is NOT 0, the formula moves to the next IF function: IF(F8=1,LOOKUP(RAND(),$N$36:$N$47,$L$37:$L$47),LOOKUP(RAND(),$M$36:$M$47,$L$37:$L$47). This next function asks if site 1's abundance is 1. If it is (true), then the spreadsheet looks up a random number in cells N36:N47 (the ordered, cumulative binomial probabilities for N = 1) and returns the corresponding detections in cells L37:L47. If site 1's abundance is NOT 1, then the formula moves onto the next IF function. In this way, the complete formula returns the observed number of individuals for the site as long as the abundance is 5 or less. If the abundance is greater than 5, then none of the IF functions evaluate to "true", and so the word "FALSE" appears on the spreadsheet. It's a bit confusing because we had to break this portion up into two sections.

Now, we need to bring our field data back into one piece, and this is done in cells F32:J51. Click on any of the cells and you'll see them tagged back to the results from the nested IF functions.

	E	F	G	H	I	J
30		Actual Survey Abundance (ni) Paste into Model				
31	Site	1	2	3	4	5
32	1	0	1	1	0	1
33	2	0	2	0	0	0
34	3	0	0	1	0	0
35	4	0	0	0	1	2
36	5	3	1	2	2	2
37	6	0	0	0	0	0
38	7	1	1	1	0	0
39	8	1	0	0	0	0
40	9	0	2	1	0	2
41	10	0	0	0	0	0
42	11	1	1	2	1	2
43	12	0	0	0	2	2
44	13	2	1	2	2	3
45	14	0	0	0	0	1
46	15	0	1	2	1	1
47	16	1	0	0	1	0
48	17	1	0	0	1	0
49	18	0	0	1	1	0
50	19	0	0	0	0	0
51	20	0	0	1	0	0

These are the data that you can copy and paste into the sheet labeled Royle Count Model (don't do that until you've analyzed the current data in PRESENCE!).

Thus, given lambda and p, the simulated data consists of 1) the true site abundances, as determined by the cumulative Poisson function, and 2) the observed "field data" as determined by the cumulative Binomial Distribution function. Press F9, and you'll see your data change. Press F9 100 times, and you've created 100 simulated datasets for the lambda and p that you specify in cells B2:B3.

	E	F	G	H	I	J
4		Simulate Data				
5						
6		True Abundance (ni)				
7	Site	1	2	3	4	5
8	1	9	9	9	9	9
9	2	3	3	3	3	3
10	3	2	2	2	2	2
11	4	1	1	1	1	1
12	5	3	3	3	3	3
13	6	2	2	2	2	2
14	7	5	5	5	5	5
15	8	8	8	8	8	8
16	9	2	2	2	2	2
17	10	7	7	7	7	7
18	11	2	2	2	2	2
19	12	3	3	3	3	3
20	13	1	1	1	1	1
21	14	3	3	3	3	3
22	15	3	3	3	3	3
23	16	4	4	4	4	4
24	17	4	4	4	4	4
25	18	3	3	3	3	3
26	19	1	1	1	1	1
27	20	3	3	3	3	3
28	max =	9				
29						
30		Actual Survey Abundance (ni) Paste into Model				
31	Site	1	2	3	4	5
32	1	3	2	0	5	4
33	2	2	1	0	1	1
34	3	0	0	0	0	0
35	4	0	1	0	0	1
36	5	1	1	1	0	2
37	6	0	1	0	0	0
38	7	1	1	1	1	0
39	8	0	1	4	3	2
40	9	1	2	0	1	1
41	10	1	2	1	0	3
42	11	0	0	2	0	1
43	12	1	0	0	0	1
44	13	0	0	0	1	0
45	14	2	1	1	1	0
46	15	1	1	0	1	1
47	16	2	2	2	1	2
48	17	2	0	0	1	1
49	18	2	0	1	0	1
50	19	1	0	0	0	1
51	20	2	1	1	1	2

REPEATED COUNT MODEL (ROYLE) ANALYSIS IN PROGRAM PRESENCE

GETTING STARTED

OK, all that's left is to run the analysis in PRESENCE. By now you should know the drill. Open PRESENCE, and then go to File | New Project. In the "Enter Specifications" form, enter a title for this project (e.g., Royle Count). We'll be analyzing the data we analyzed in the spreadsheet, which contained data for 20 sites and 5 occasions. Enter the number 5 in the text box labeled "No. Occasions/Season."
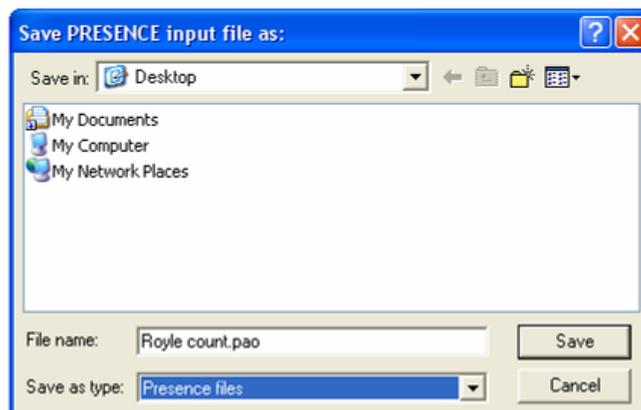

Now we are ready to input our data. Click on the button labeled "Input Data Form," and you'll be presented with a new form. Return to your spreadsheet, and copy cells B9:F28, and then click again on the Input Data Form and select Edit | Paste | Paste Values.

Data Input Form

File   Edit   Simulate

Presence/Absence data

data	1-1	2-1	3-1	4-1	5-1
site 1	2	1	1	2	2
site 2	1	2	2	1	2
site 3	0	4	1	1	1
site 4	2	0	0	0	1
site 5	1	0	2	1	2
site 6	0	0	1	0	0
site 7	0	1	0	0	0
site 8	1	0	0	2	1
site 9	2	3	1	1	0
site 10	0	0	0	1	0
site 11	2	2	1	0	0
site 12	1	0	1	1	1
site 13	1	0	0	1	0
site 14	2	0	0	2	2
site 15	1	1	1	1	1
site 16	1	2	0	1	1
site 17	0	0	1	0	0
site 18	1	1	2	0	1
site 19	1	0	0	1	0
site 20	2	1	0	0	0

This is our full dataset. Notice that PRESENCE labels the surveys 1-1, 2-1, 3-1, 4-1, 5-1, where the first number indicates the survey number, and the second number indicates the season or period. In this example, all five surveys were completed within a single season.

The next step is to save these data, so go to File | Save As, and enter a name for this input file and store it in a place where you can readily retrieve it.

Save PRESENCE input file as:

Save in: Desktop

My Documents
My Computer
My Network Places

File name: Royle count.pao   Save

Save as type: Presence files   Cancel

Now, close the data file, and return to the "Enter Specifications" form. Click on the button labeled "Click to select file" and navigate to your freshly created input file.
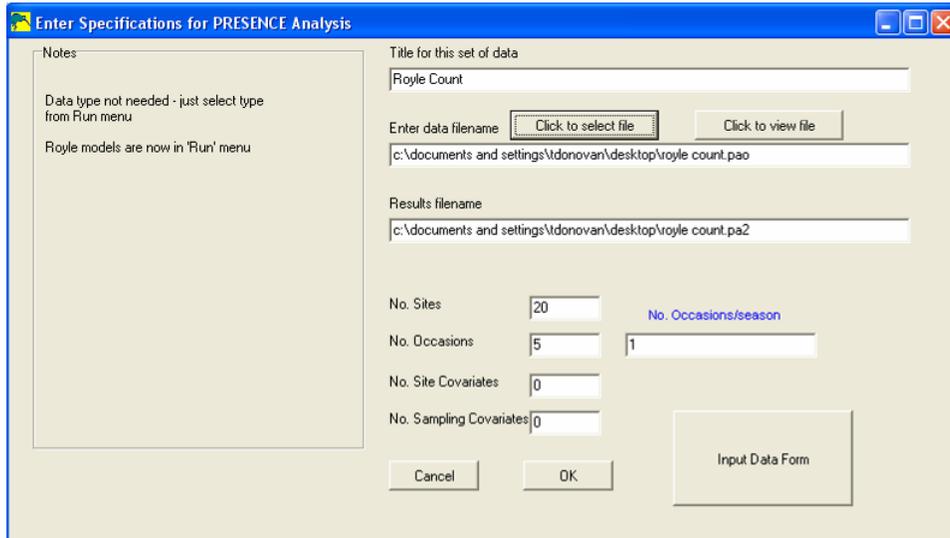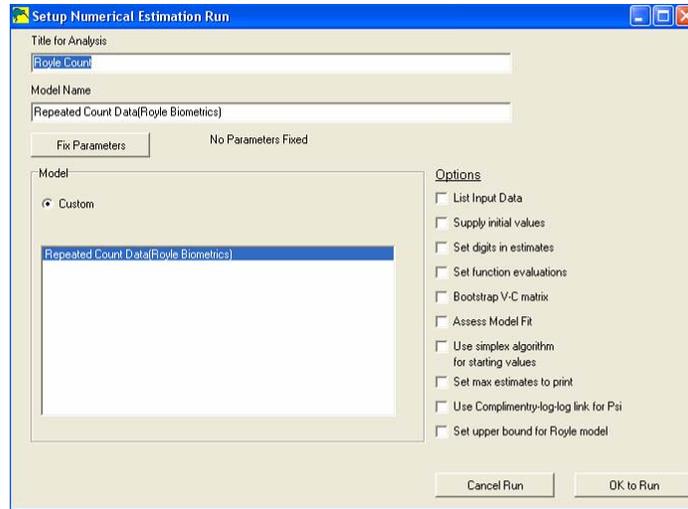

When you are finished, press OK, and the Results Browser will now appear.


RUNNING THE ROYLE COUNT MODEL

Now we're ready to run an analysis, and it's extremely simple. On the PRESENCE main page (the one showing the book), go to Run | Analysis: Repeated Count Data (Royle Biometrics):
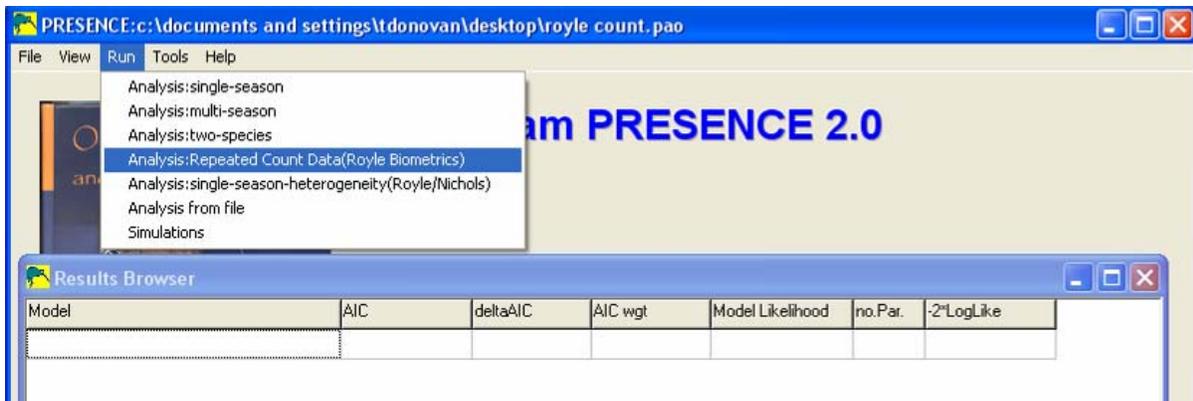
You'll then be presented with a new form called "Setup Numerical Estimation Run."

There aren't a lot of bells and whistles on this page. You can check off a variety of options for the model output, but we really don't need them for our purpose in this exercise. So all that's left to do is to press the "OK to Run" button. PRESENCE will run its optimization routine, and once it has found a solution, will ask if you want to append the results to the Results Browser:
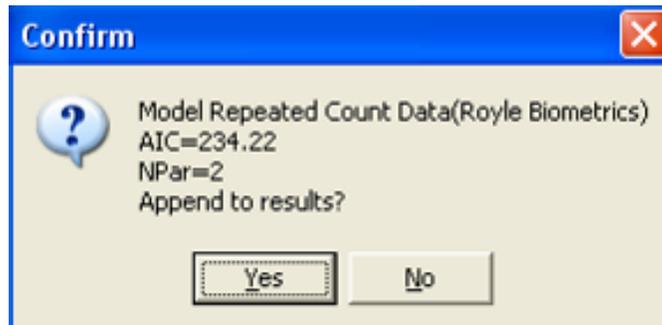
Click Yes, and the Results Browser should now look like this:

Results Browser

Model	AIC	deltaAIC	AIC wgt	Model Likelihood	no.Par.	-2*LogLike
Repeated Count Data(Royle Biometrics)	234.22	0.00	1.0000	1.0000	2	230.22

Because we won't be running multiple models in this exercise, let's dive straight into the model results. Right-click on the model title in the Results Browser to bring up the full model results:

```
Royle Model w/ species counts (k=200)
Number of groups              = 1
Number of sites               = 20
Number of sampling occasions  = 5
Number of missing observations = 0


Matrix 1: rows=3, cols=3
-,a1,a2,
p  1 0
lambda  0 1
=========================
Number of parameters          = 2
Number of function calls       = 57
Final function value           = 115.108253
-2log(likelihood)              = 230.216506
AIC                            = 234.216506
Naive occupancy estimate       = 1.000000


Untransformed (beta) parameters:
Estimated parameter                     estimate  std.err
---------------------------             --------  -------
beta0                                 = -1.6241   0.6806
beta1                                 = 1.6296    0.5684

beta var-cov matrix:
   0.4632    -0.3750
  -0.3750     0.3231

                    MODEL PARAMETERS:
Estimated parameter                     estimate  std.err  95% confidence interval
---------------------------             --------  -------  -------------------------
Detection probability (c)             = 0.1646   0.0936   -0.0188 -  0.3481
Avg. abundance/sample unit(lambda) =    5.10     2.90     -0.58 -   10.79

Derrived  parameter                     estimate  std.err  95% confidence interval
---------------------------             --------  -------  -------------------------
Occupancy (psi)                       = 0.9939   0.0176    0.9593 -  1.0285
Total Abundance (N)                   = 102.04   58.00    -11.64 -  215.71
-----------------------------------------------------------
CPU time: 1.0 seconds
```

For comparison, here are the spreadsheet results, and they are a match.

	B	C	D	E	F
1	Inputs, Parameters, Outputs				
2	p	p beta	θ	θ beta	N hat
3	0.164643094	-1.624078998	5.1019	1.629621246	102.03884
4	Log L	-2Log L	K	AIC	AICc
5	-115.108	230.217	2	234.21651	233.137

You should always do your analysis in PRESENCE because it provides the standard errors and 95% confidence limits for both the betas and the parameters themselves, and knowing this information is essential before you draw conclusions from your data. In this example, the estimate of p was 0.1646, but the 95% confidence intervals ranged from 0 to 0.34. Lambda (which Royle calls θ) is estimated by PRESENCE as 5.10 with 95% confidence intervals of 0.58 to 10.79. That's quite a spread…these estimates are not precise. And total abundance across the 20 sites was estimated between 11.64 and 215.71 (Jim, are the negative signs in the PRESENCE output correct?). Depending on the objectives of your study, you might not be very happy with this conclusion. These large standard errors and confidence intervals are undoubtedly due to the very small sample size (R = 20) explored in the spreadsheet. Nevertheless, when planning a study, it is useful to explore various options for increasing R or increasing T so that you can obtain estimates that are both precise and unbiased.