Probability distributions

A Field Guide to Probability theory & distributions for biologists

Part 2

September 24th 2015

Goals for today

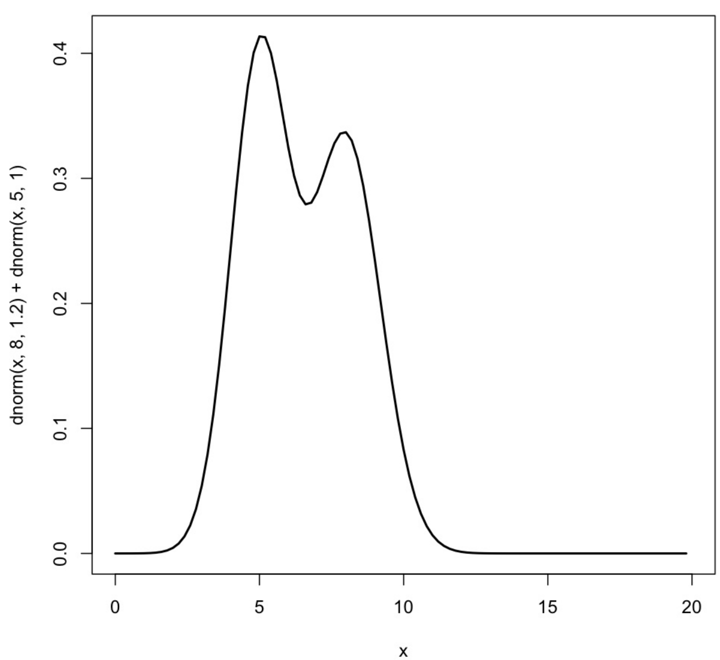
- Discuss Method of moments estimators for the parameters for these distributions.

- Review the relationships between a few sampling distributions.

- Do some example calculations in R, and use knitr and git (github) to make a repository of it.

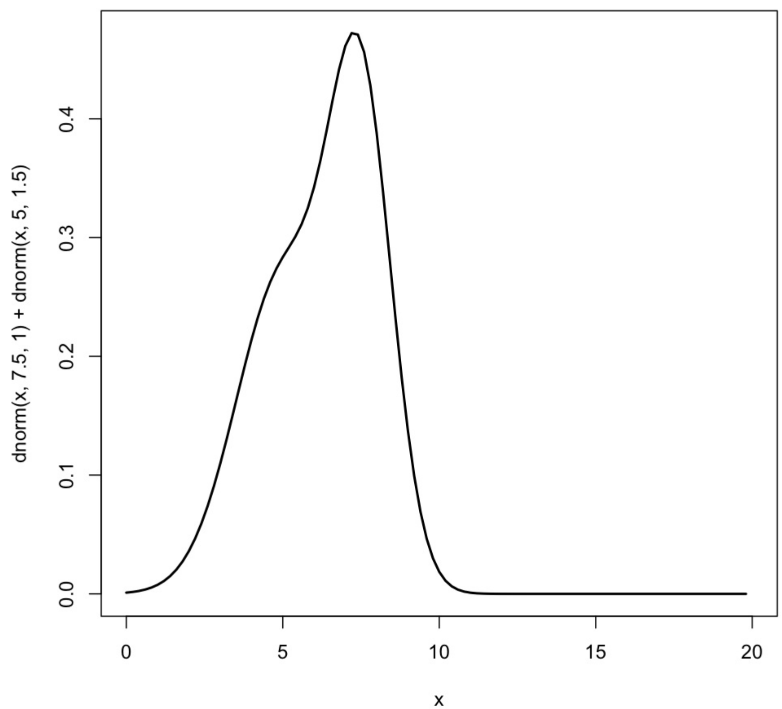
In statistical models we are interested both in modeling the data as a whole, and the distribution of the error in particular.

- If your data looked like this, what kind of distribution might you use?



In statistical models we are interested both in modeling the data as a whole, and the distribution of the error in particular.

- If your data looked like this, what kind of distribution might you use?



Both examples were the result of a mixture of two normal distributions

You could use a mixture of two normals with 4 parameters (mean and variance for each peak). However Occam’s Razor suggests this may not be required as only the location/mean differs.

An important point is that you may not need to fit your whole data set to a particular probability distribution, per se. Instead remember that there may be other model parameters (I.e. a mean for each peak) to consider, with a single underlying error distribution.

Keep in my mind..

- It is not the distribution of the whole data set that will necessarily determine what distribution to use to model it. Instead we are often more concerned with the distribution of the residual variation once we have accounted for all the parameters that we are estimating.

Parameterizing a probability model

- Generally if you “know” the parameter values for a distribution, you can visualize it, which is quite useful.

- For example the Normal distribution has a scale parameter (sd) and a location parameter (mean). If you can estimate these you can visualize the distribution.

Parameterizing a probability model

- Each distribution has moments (1st moment being the mean, second moment the variance…).
- Moments are the expected values of the powers of a random variable.

- $E[x^1]$ first moment
- $E[x^2]$ second moment

If we can estimate these parameters, even roughly we can get an idea of what the distributions may look like.

Expected values

- What is the "expected value" from a normal six sided die?

$$\frac{1+2+3+4+5+6}{6} = 3.5$$

Not the most probable outcome, but the expected outcome (or mean). It is "expected" from the standpoint of repeating the experiment over and over.

Expected values from a distribution

If X takes on discrete values, and is a random variable with probability ***mass*** function p(x).

$$E(X) = \sum_i x_i p(x_i)$$

If X is a continuous random variable with probability ***density*** function f(x).

$$E(X) = \int x f(x)\,dx$$

Empirical moments

$$E[x^1] = \frac{1}{n}\sum_{i=1}^{n} x_i^1$$

$$E[(x - \bar{x})^2] = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

Look in Bolker for the proof

We can use these empirical moments with those of the theoretical distribution as a first guess on the parameter estimates.

Using the method of moments

Some things to consider

- There are three types of parameters for probability distributions; shape, scale and location.
- Some distributions have only 2 (normal or 1 (Poisson) parameters.

Parameters for probability distributions

- Location = mean for a normal. Scale = Standard Deviation. (NO shape, that is why it is symmetrical).

- The fact that location=mean and scale = Sd is NOT TRUE for all probability distributions.

Discrete distributions

Binomial

Poisson

Negative-binomial

Random variables

- This is what we want to know the probability distribution of.

- I.e. $P(x|\text{some distribution})$

I will use “x” to be the random variable in each case.

Binomial

Let's say you set up a series of enclosures. Within each enclosure you place 25 flies, and a pre-determined set of predators.
You want to know what the distribution (across enclosures) of flies getting eaten is, based on a pre-determined probability of success for a given predator species.

You can set this up as a binomial problem.

N (R calls this size) = 25 (the total # of individuals or "trials" for predation) in the enclosure
p = probability of a successful predation "trial" (the coin toss)
x = # trials of successful predation. This is what we usually want for the probability distribution.

Binomial

$$\binom{N}{x} p^{x}(1-p)^{N-x}$$

p= probability of success
p^x= number of times you are successful
(1-p)^N-x = times you aren't successful

$$\binom{N}{x} = \frac{N!}{x!(N-x)!}$$

You can think of this in two ways.
A) A normalizing constant so that probabilities sum to 1.
B) # of different combinations to allow for x "successful" predation events out of N total.

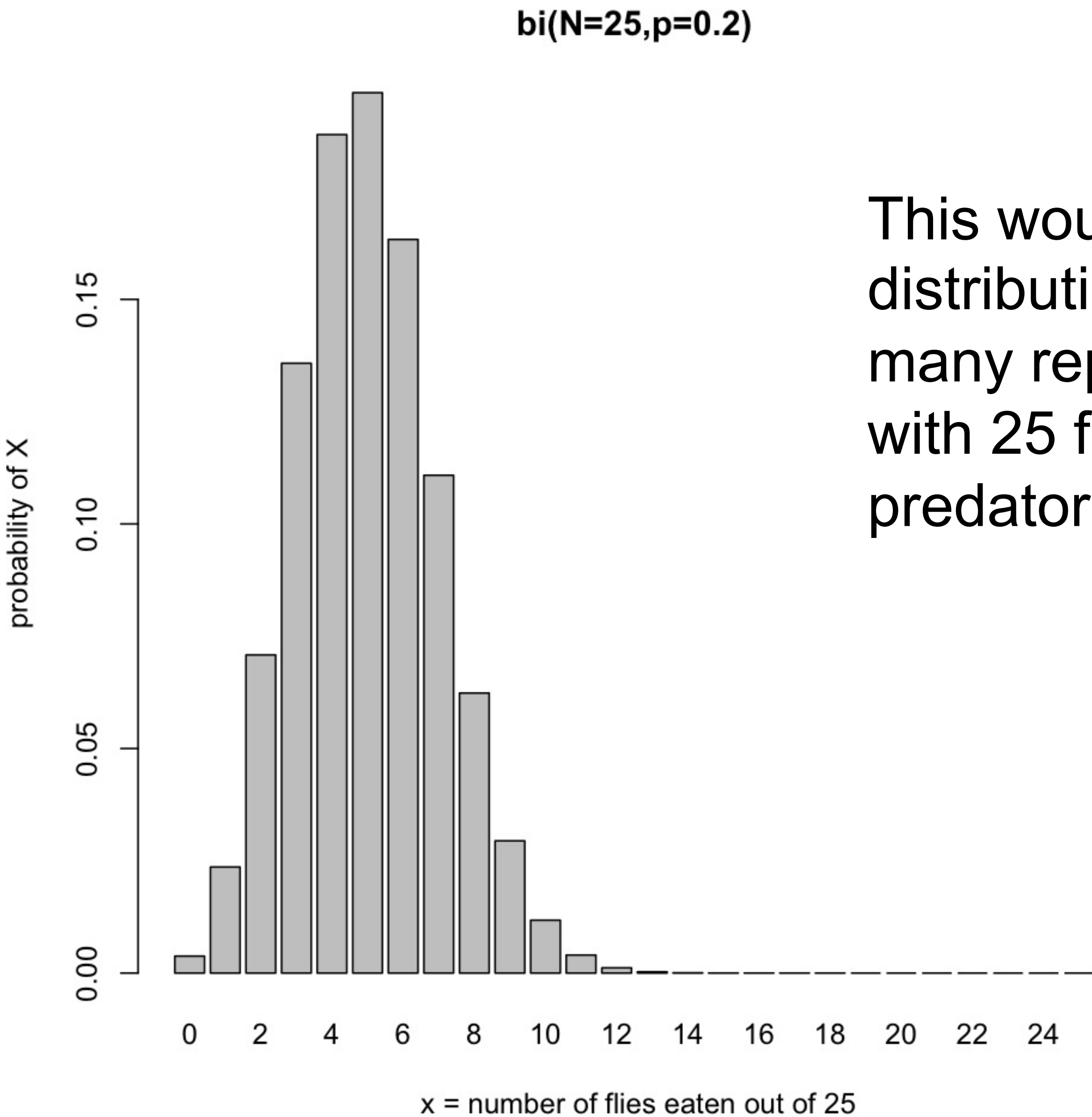
You will often see x=k and hear " N choose k"

Example

- If predator species 1 had a per "trial" probability of successfully eating a prey item of 0.2, what would be the probability of exactly 10 flies (out of the 25) being eaten in a single enclosure.

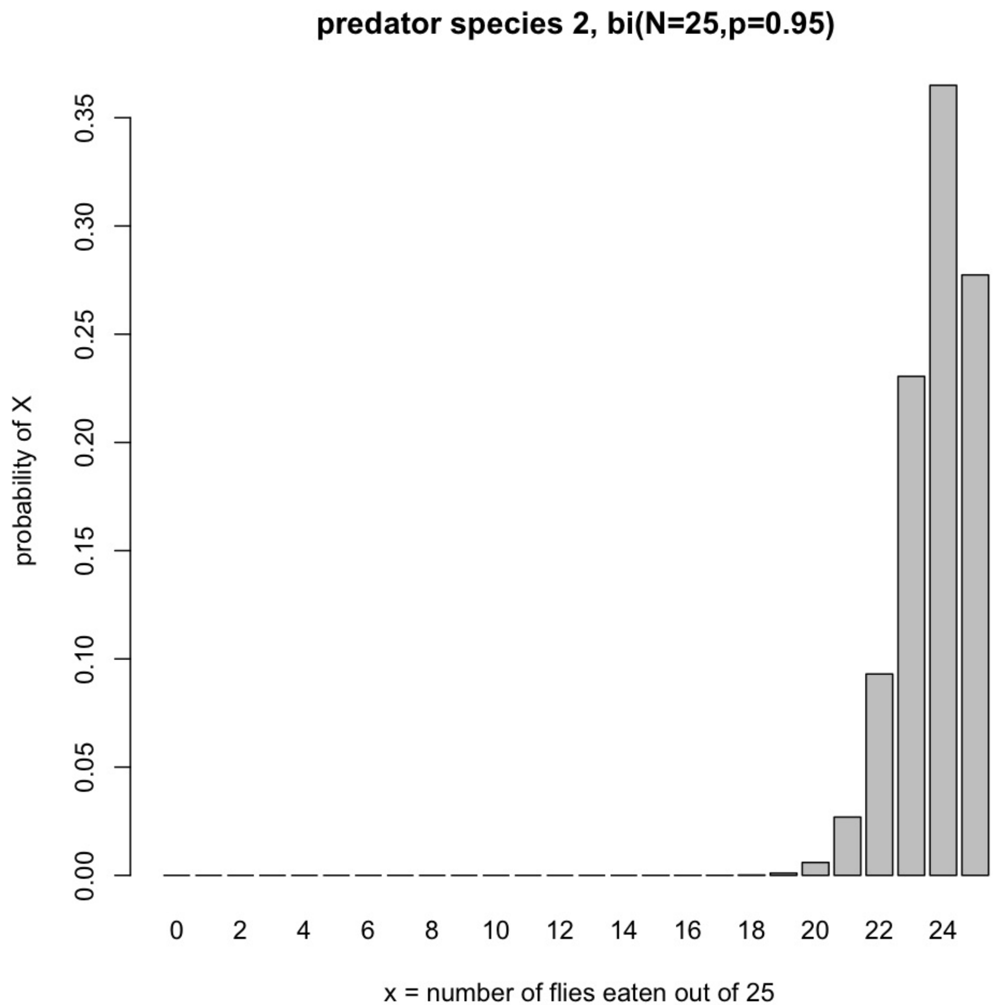
$P(x=10 \mid bi(N=25,p=0.2)) = 0.0118$

Not so high. We can look at the expected probability distribution for different values of x.



This would be the expected distribution if we set up many replicate enclosures with 25 flies and this predator.

Predator species 2 is much hungrier….



If you are modeling using a binomial distribution, which parameter are you estimating?

- N, the number of trials
- p, the proportion of successes
- x, the number of successful trials

The rub…

- Usually we are not interested in the probability of a given number of “successful” trials, but in estimating the parameter, p itself.

- P(D|H)
- P(x|bi(N=25,p=?)

binomial

- $0 \le x \le N$
- Mean = Np (how do you estimate p)
- Var = Np(1-p)

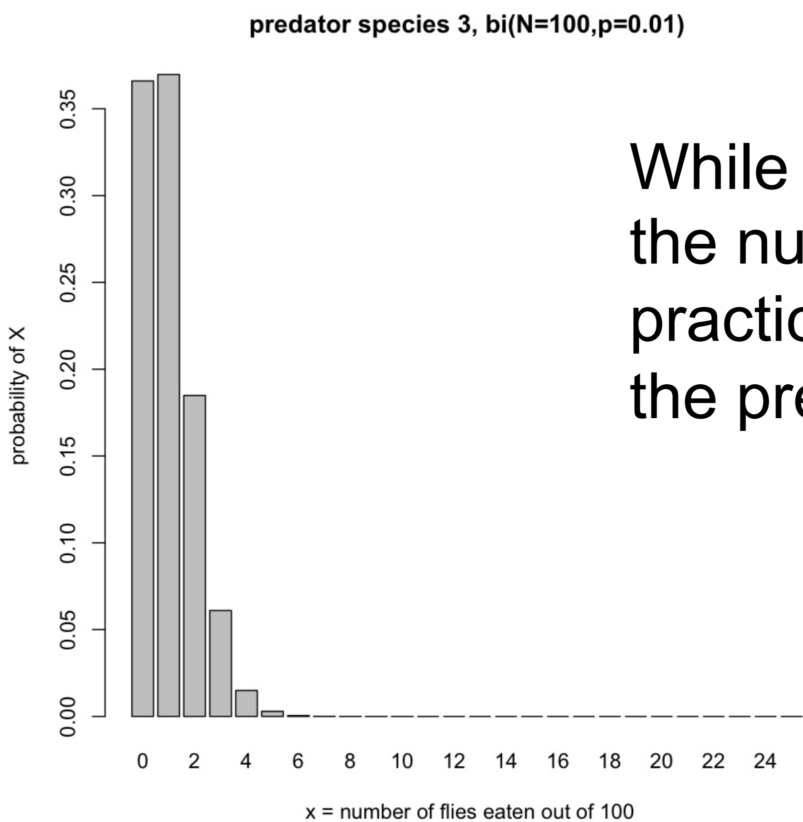
- Let’s try some examples in R, simulating the Drosophila predation experiment.

Let’s play.

```
x <- rbinom(n = 50, size = 25, prob = 0.5)

plot(density(x, bw=1.2))
```

Let’s say we had 100 flies per enclosure, and predator species 3 was really ineffective, p=0.01



While there may be a theoretical limit to the number of flies that can be eaten, practically speaking it is unlimited since the predation probability is so low.

Binomial Examples

- Number of surviving individuals out of an initial sample
- Number of infested / affected animals in a sample
- Number of a particular class (e.g. haplotype) in a larger population

Poisson

binomial distributions have a fixed number of trials, whereas Poisson have ~infinite number of trials

- When you have a discrete random variable where the probability of a “successful” trial is very small, but the theoretical (or practical) range is effectively infinite, you can use a poisson distribution.

- Useful for counting # of “rare” events, like new migrants to a population/year.
- # of new mutations/offspring..

Poisson

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

x is our random variable (# events/unit sampling effort)
λ Is the "rate" parameter. Expected number per sample
λ is the mean and the variance!!!!

For its relation to a binomial when N is large and p is small
 $\lambda = N*p$

Poisson Examples

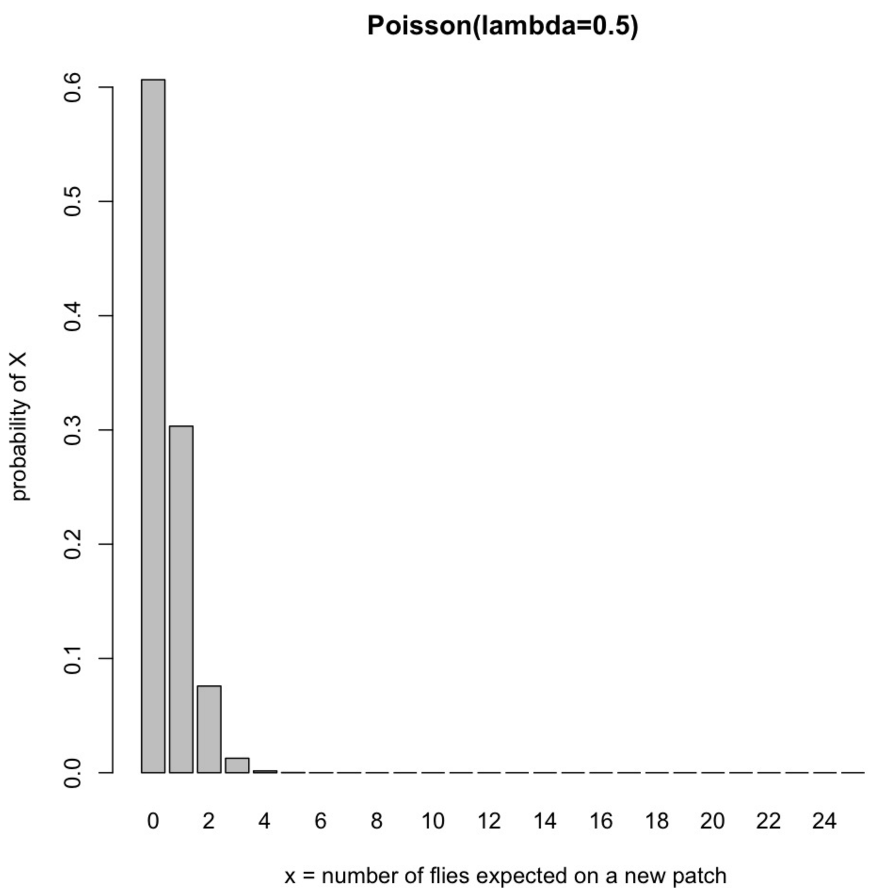
- Number of seeds in a gap
- Number of offspring in a season
- Number of prey caught per unit time
- Fluctuation test for directed mutations
- Use when binomial doesn’t work because there’s always the rare possibility of large counts

Poisson

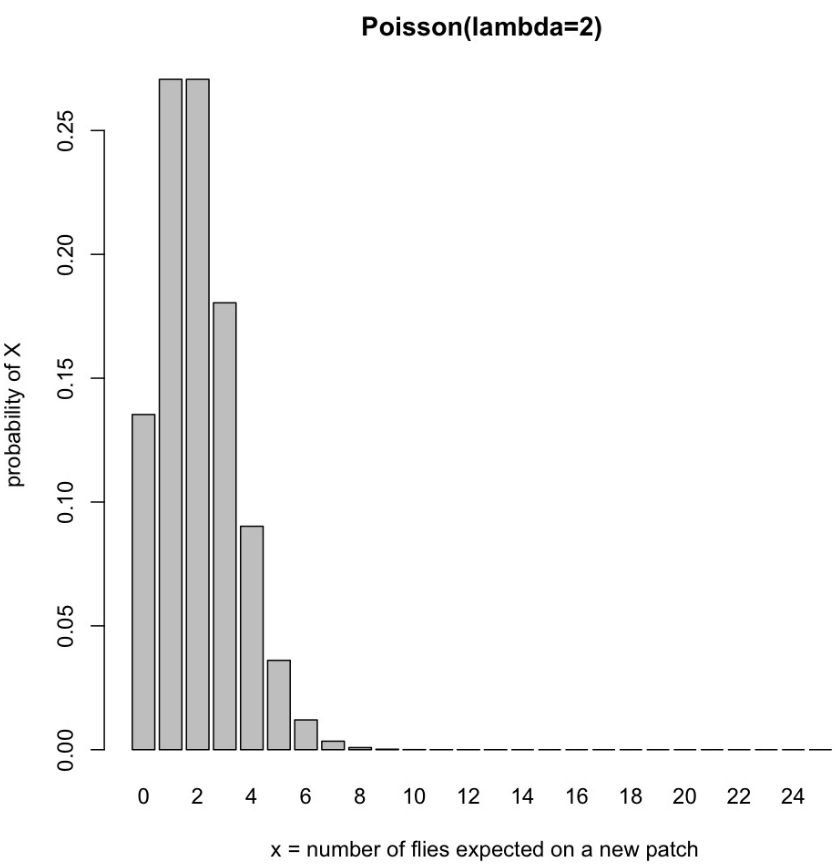
- Let’s say flies disperse to colonize a new patch at a very low rate (previous estimates suggest we will observe one fly for every two new patches we examine, $\lambda=0.5$).
- What is the probability of observing 2 flies on a new patch of land?

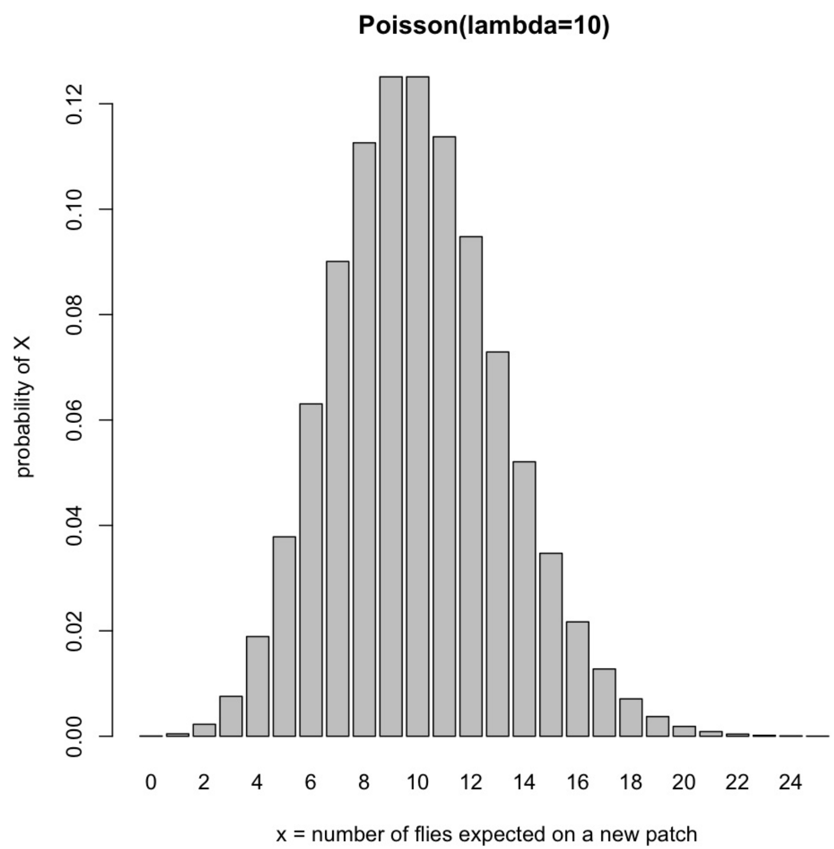
$P(x=2|\ \text{poisson}(\lambda=0.5)) = 0.076$

Probability of observing x number of flies on a patch given lambda=0.5



What happens as lambda increases?





Poisson mean and variance

- When lambda is small for your random variable, you will often find that your data is “over-dispersed”.
- That is there is more variation that expected under Poisson (lambda)
- Similarly when lambda gets large, you will often find that there is less variation than expected under Poisson(lambda).

Negative binomial

- In ecology the Neg. Binomial is mostly used like a Poisson, but when you need more dispersion of x (it needs to be spread out more).

Negative binomial

$$\text{Negative Binomial Distribution} = \frac{\Gamma(k+x)}{\Gamma(k)x!}\left(\frac{k}{k+\mu}\right)^{k}\left(\frac{\mu}{k+\mu}\right)^{x}$$

Expected number of counts = μ (mu in R)
Over-dispersion parameter = k (size in R)

Mean = μ
Variance = $\mu + \mu^2/k$

Negative Binomial Examples

- Essentially the same as Poisson, but allows for heterogeneity
- Number of individuals per patch
- Distributions of parasites with individual hosts
- Number of seedlings in a gap

Continuous distribution

- Normal/Gaussian
- Beta
- Gamma Family (Gamma, Exponential, Chi Squared)

Normal/Gaussian Distribution

Symmetric distribution with two parameters mean (location) and scale (sd).

Central limit theorem as poor approximation sometimes, binomial(N,p is small)

Gamma Family of Distributions

- Continuous distributions
- Bounded by zero (no negative values)
- Includes Gamma, exponential and chi-squared (the latter two are special cases of the Gamma distribution).

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$$

important parameters are Alpha and Beta

Gamma Distribution

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$$

- We just have a two parameter model (α,β).
- We don’t care how the functional form was derived, what we care about are its properties.

Gamma distribution

- Often described as the distribution of waiting times until a fixed number of events take place.

- More generally it is used because it is an extremely flexible in shape and scale.
- Useful when data is over-dispersed on a normal distribution

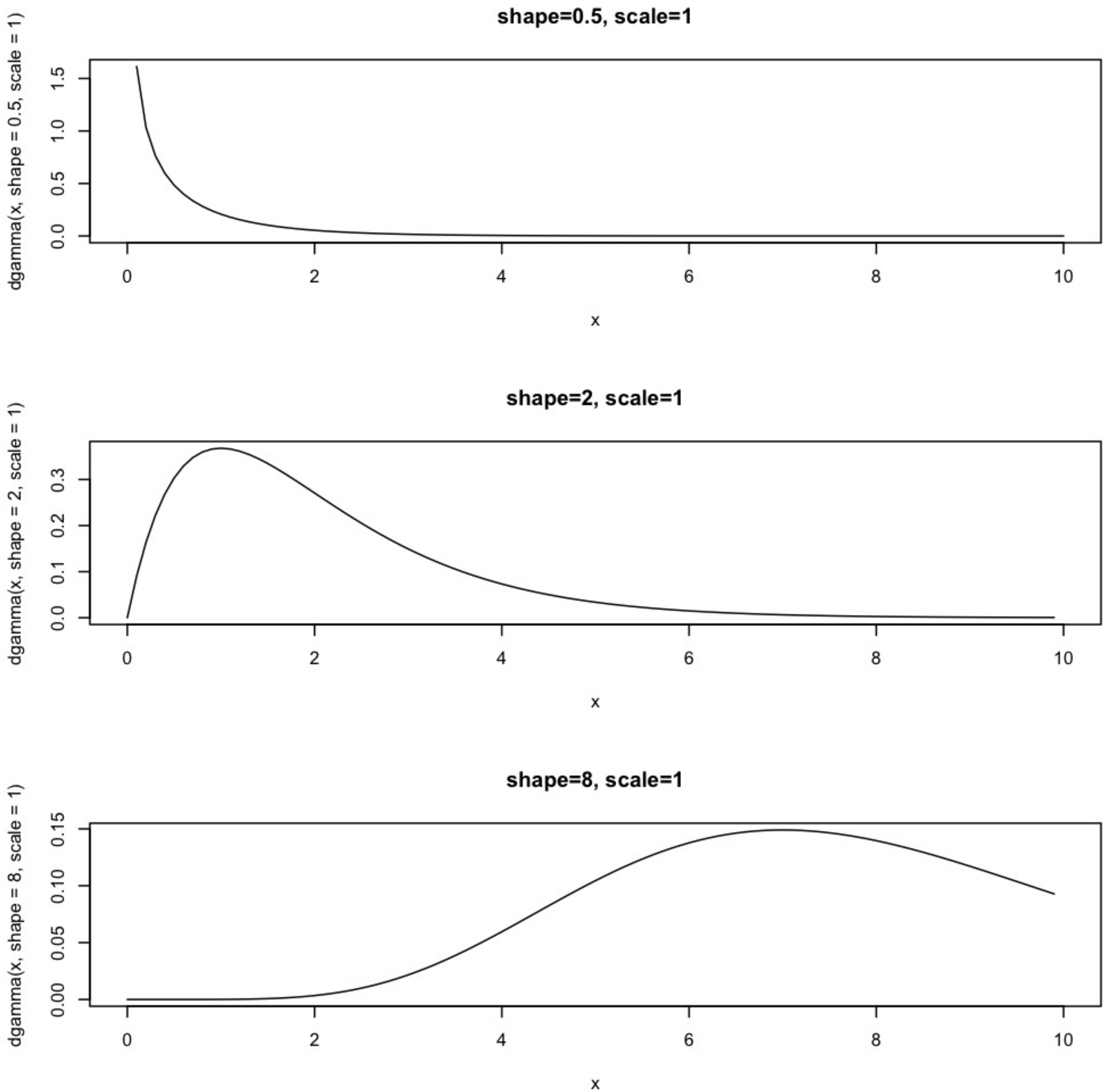
Gamma

- α= shape / # of events
- β= scale / length per event
- Sometimes (r) rate is used instead of scale. Rate = 1/ scale.

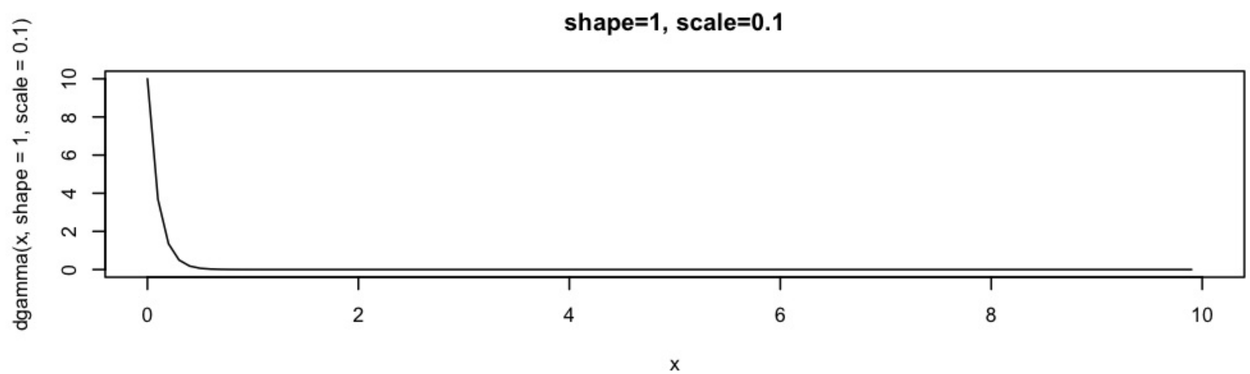
Mean = $\alpha\beta$

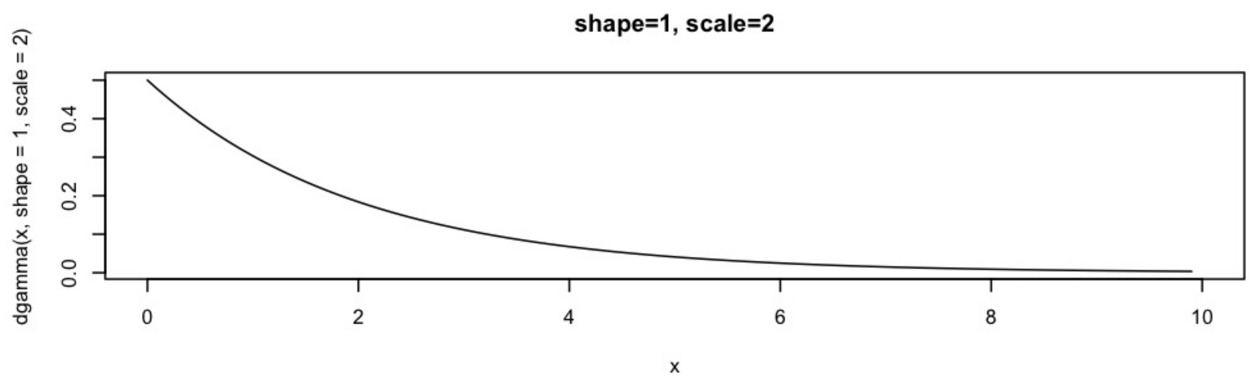
Variance = $\alpha\beta^2$

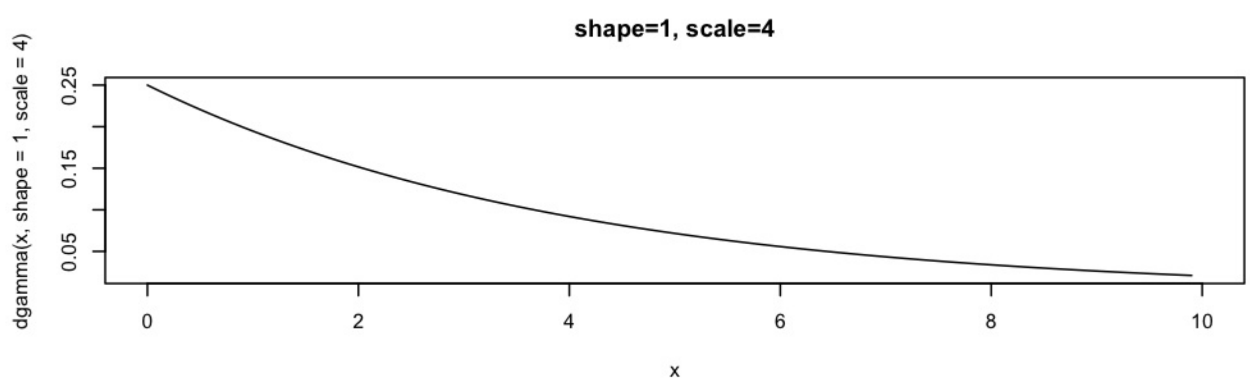
- x must be non-negative

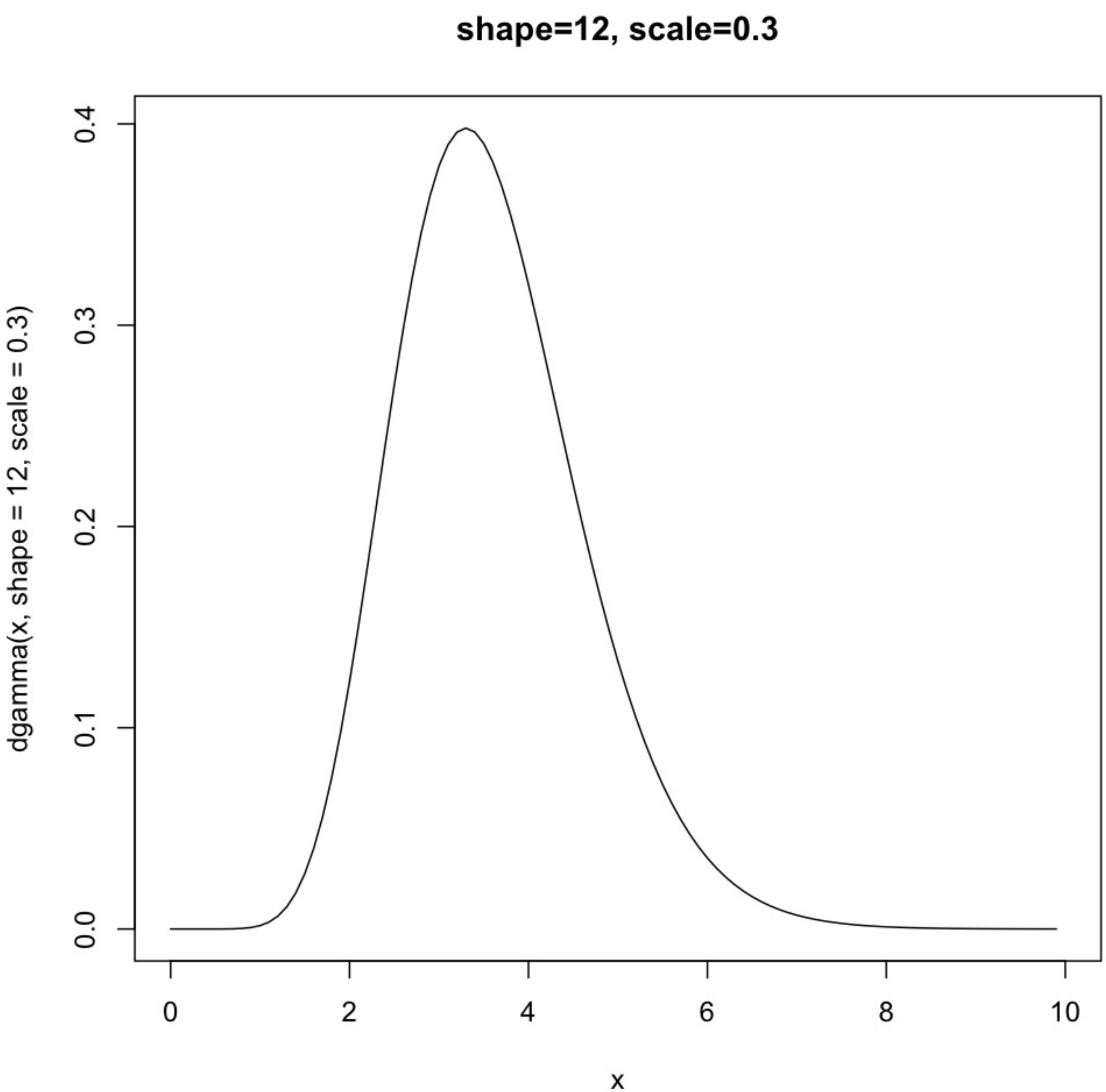


shape=0.5, scale=1
dgamma(x, shape = 0.5, scale = 1)
0.0
0.5
1.0
1.5
0
2
4
6
8
10
x
shape=2, scale=1
dgamma(x, shape = 2, scale = 1)
0.0
0.1
0.2
0.3
0
2
4
6
8
10
x
shape=8, scale=1
dgamma(x, shape = 8, scale = 1)
0.00
0.05
0.10
0.15
0
2
4
6
8
10
x


shape=1, scale=0.1
dgamma(x, shape = 1, scale = 0.1)
0
2
4
6
8
10
0
2
4
6
8
10
x


shape=1, scale=2
dgamma(x, shape = 1, scale = 2)
0.0
0.2
0.4
0
2
4
6
8
10
x


shape=1, scale=4
dgamma(x, shape = 1, scale = 4)
0.05
0.15
0.25
0
2
4
6
8
10
x


shape=12, scale=0.3
dgamma(x, shape = 12, scale = 0.3)
0.0
0.1
0.2
0.3
0.4
0
2
4
6
8
10
x

Let’s play with the gamma.

```
y <- rgamma(50, shape=2, scale = 1)
plot(density(y))
```

Exponential - special case of gamma distribution

Describes the waiting time for a single event to happen (shape =1)

Useful for situations where most of the probability “mass” is near zero.

beta

- Continuous, but constrained on 0,1.
- Useful for modeling probabilities or proportions

- We will be mostly interested in beta as the conjugate prior for the binomial when we talk about Bayesian estimation.