Some Common Non-Normal Distributions

"The trouble with normal is that it always gets worse"

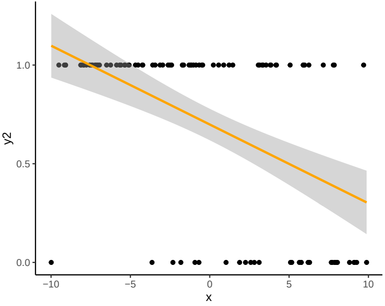
Samuel Robinson, Ph.D.

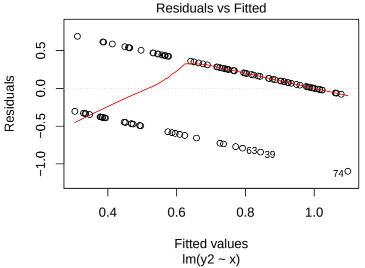
November 19, 2020

Outline

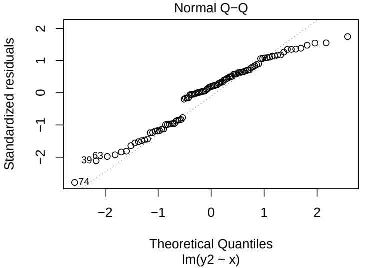
- Meet (some of) the exponential family!
  - Normal
  - Binomial
  - Poisson
  - Beta-Binomial
  - Negative Binomial
- "Play time"

Problem: not everything is normal

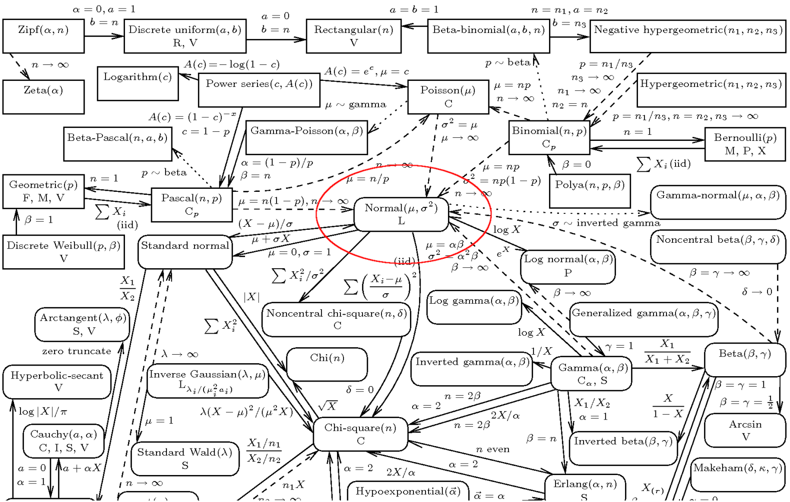




- Some types of data can never be transformed to make the residuals normal
- Solution: **use the distribution that generates the data!**

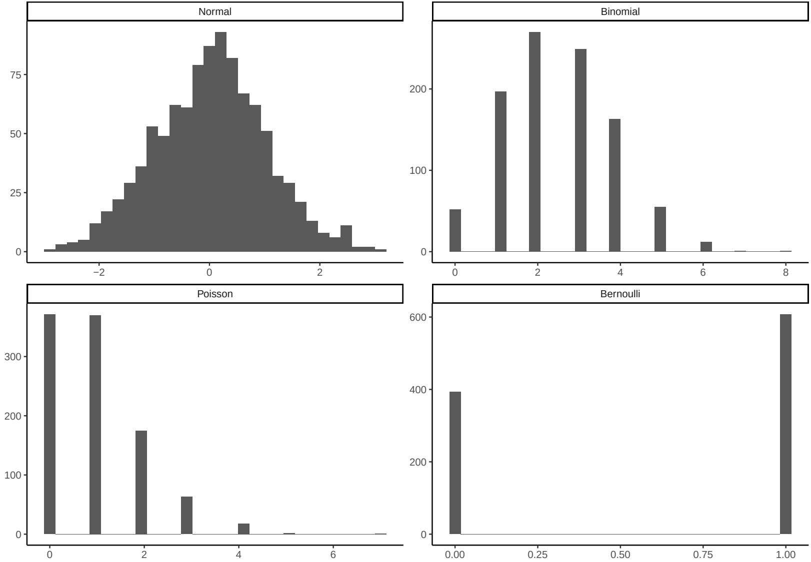


But how do I know which distribution to use?



And if thou gaze long into an abyss, the abyss will also gaze into thee - F. Nietzsche

Let's take a look at some *common* ones!



Time to meet the Exponential family!

The Normal Distribution (aka *Gaussian*)

- Imagine many random + and - numbers added together
- If you do this *many* times:
  - Most cancel out (somewhere around 0)
  - Few are far away from 0 (tails of distribution)
- Common in nature, because of many small + and - factors adding together
  - e.g. Height is driven by many sets of genes

A Galton Board in action:

The Normal Distribution - scary math!

- 2 parameters: mean (μ) and standard deviation (σ)

$$p(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

- Probability distribution function (PDF) for the Normal distribution
- Tells you about the probability of getting some number *given* μ and σ

Example: what is the probability of getting a 4, if the mean is 5 and SD is 1?

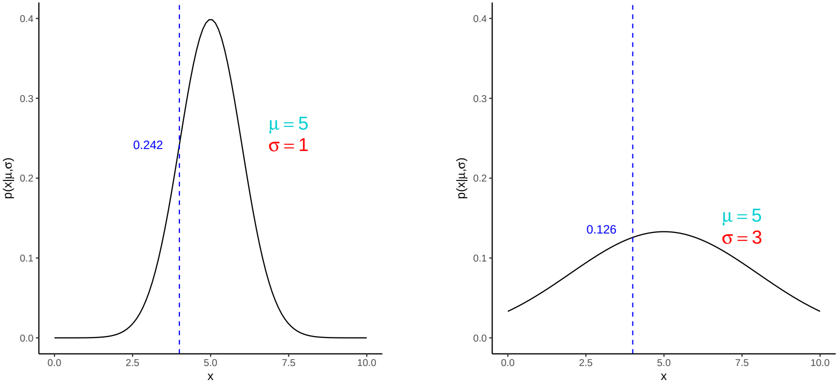
$$p(4|5,1) = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{4-5}{1})^2}$$
$$= \sim 0.24$$

In R, this is easy:

```
#d stands for "density"
dnorm(x=4,mean=5,sd=1)
```

```
## [1] 0.2419707
```

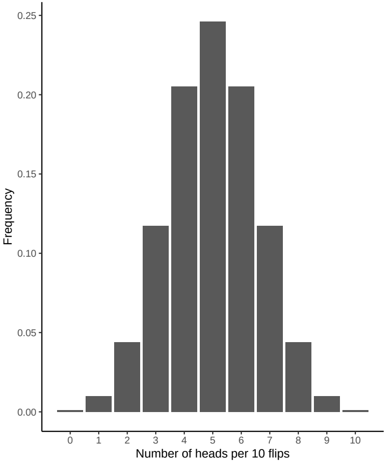
The Normal Distribution



- Probability of x changes with μ and σ
- Left: $\sigma = 1$, Right: $\sigma = 3$

The Binomial Distribution

- Imagine you have 10 coins, and you flip them all
- If you do this *many* times:
  - Most will be about 5 heads/tails
  - Few will be 1 head, 9 tails (or reverse)
- Common in nature where outcomes are binary
  - e.g. 10 seeds from a plant, how many will germinate?
- If $N = 1$, this is called a *Bernoulli trial*



The Binomial Distribution - scary math!

- 1 parameter: probability of success (ϕ), plus...
- Number of "coin flips" (N)

$$p(x|\phi, N) = \binom{N}{x} \phi^x (1-\phi)^{N-x}$$

- Probability mass function (PMF); density = continuous
- Tells you about the probability of getting x "successes" *given* ϕ and N

Example: what is the probability of getting 4 successes, if ϕ is 0.25 and N is 15?

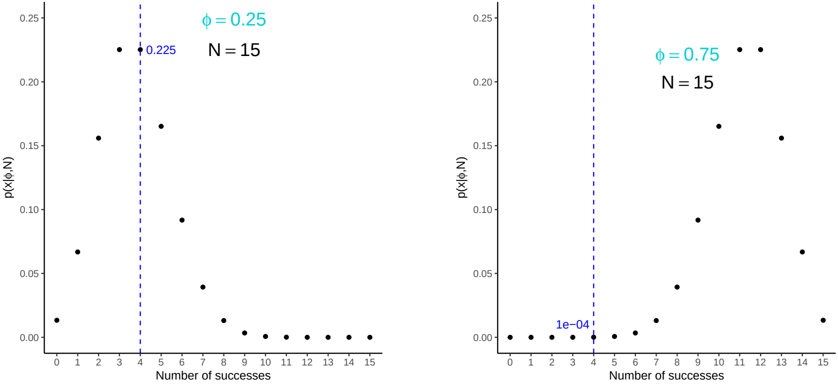
$$p(4|0.25, 15) = \binom{15}{4} 0.25^4 (1-0.25)^{15-4}$$
$$= \sim 0.23$$

In R, this is easy:

```
dbinom(x=4,size=15,prob=0.25)
```

```
## [1] 0.2251991
```

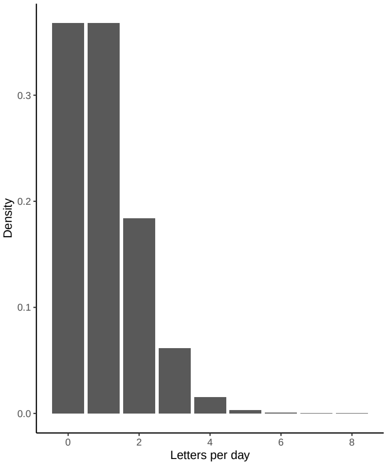
The Binomial Distribution



- Probability of x "successes" changes with ϕ and N

The Poisson Distribution

- Imagine a rare event (e.g. getting a non-junk mail letter)
- If you record the number of events every day:
  - Most days, you'll get 0 or maybe 1 letter
  - On some rare days, you'll get 3 or 4 letters
- Common in nature where rare events are measured over time/space:
  - e.g. Number of bats caught in a net (per night)



- Equivalent to Binomial distribution, where N is unknown

The Poisson Distribution - scary math!

- 1 parameter: rate parameter (λ)

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

- Probability mass function (PMF)
- Tells you about the probability of getting x counts *given* λ

Example: what is the probability of getting 2 counts, if λ is 1?

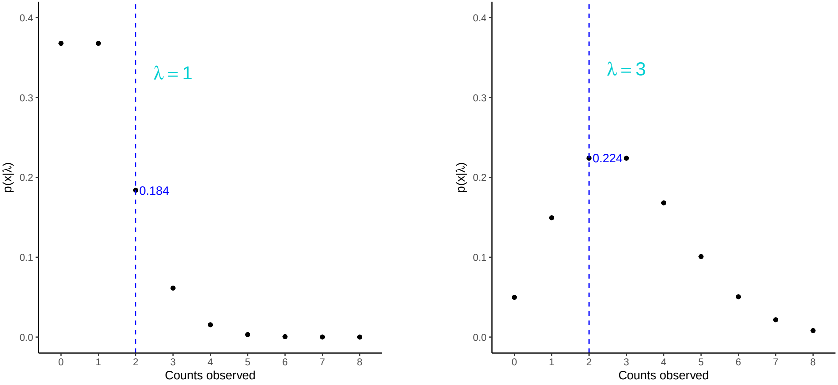
$$\begin{aligned} p(2|1) &= \frac{1^2 e^{-1}}{2!} \\ &= \sim 0.18 \end{aligned}$$

In R, this is easy:

```
dpois(x=2,lambda=1)
```

```
## [1] 0.1839397
```

The Poisson Distribution



- Probability of x counts changes with λ

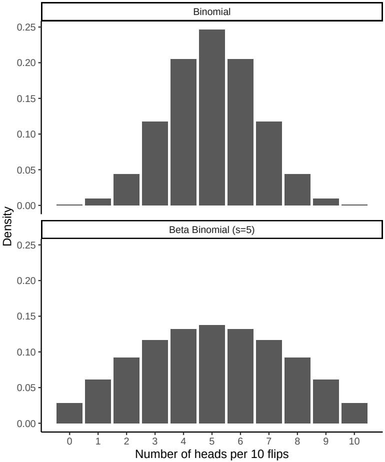
More complications:

- The Normal distribution has a parameter for the mean and SD, but. . .
- What about the Binomial and Poisson distributions?
  - Binomial: mean $= Np$, SD $= \sqrt{Np(1-p)}$
  - Poisson: mean $= \lambda$, SD $= \sqrt{\lambda}$
- What if our data have additional variance?
  - *Beta Binomial* and *Negative Binomial* distributions

The Beta Binomial Distibution

- Many "coin-flip" processes have longer tails than standard Binomial
  - e.g. numbers of males/females in families
- Beta-binomial adds additional dispersion to coin flip process
- 2 parameters: ϕ and s (if s is large, similar to Binomial)
  - Also requires: N

```
#Extra distributions
library(rmutil)
dbetabinom(x,m=phi,size=N,s=5)
```

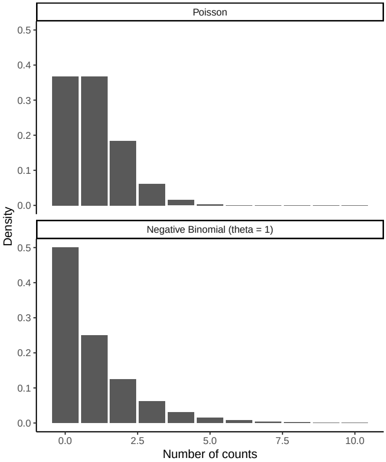


The Negative Binomial Distribution

Unfortunately, *almost nothing* in ecology actually follows a Poisson distribution

- Negative Binomial is similar to a Poisson, but can have longer tails
- Also called: *Polya* distibution (`nbinom2` in many GLM commands)
- Parameters: μ and θ (if θ is large, close to Poisson)

```r
#size = theta parameter
dnbinom(x,mu,size=1)
```



Summary of Common "Starter" Distributions

- Continuous data, spanning - or + numbers:
  - Normal (transformed or regular)
- Count data
  - Poisson, Negative Binomial
- Count data of successes *and* failures
  - Binomial + Beta Binomial

These are by *no means* the only useful distributions, but are fairly common

A Challenger Approaches! (Part 1)

Let's say that you've collected data at 2 different sites. Which distributions would you start with for the following data?

- Bat weights
- Total bats per box (or small nest)
- Number of male and female bats
- Number of pups per female bat
- Record of occupied/unoccupied nests
- Size of trees (DBH or height)

A Challenger Approaches! (Part 2)

Now that you've figured out which distribution, try simulating some data from each site, and plot it!

- Bat weights: *Normal*
  - `rnorm(n,mean,sd)`
- Total bats per box: *Poisson or NB*
  - `rpois(n,lambda)` or `rnbinom(n,mu,size)`
- Number of male/female bats: *Binomial or Beta Binomial*
  - `rbinom(n, size, prob)` or `rbetabinom(n,size,m,s)`
- Number of pups per female bat: *Poisson or NB*
  - See above
- Record of occupied/unoccupied nests: *Binomial*
  - `rbinom(n, 1, prob)` aka. *Bernoulli* distribution
- Size of trees (DBH or height): *log-Normal*
  - `exp(rnorm(n,mean,sd))`

Distributional pillows

Christmas gifts for the huge nerds in your life