Generalized Linear Models

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

Samuel Robinson, Ph.D.

Sept 29, 2023

Part 1: The exponential family

Outline

- Meet (some of) the exponential family!

Christmas gifts for the nerds in your life

Outline

- Meet (some of) the exponential family!
    - Normal

Christmas gifts for the nerds in your life

Outline

- Meet (some of) the exponential family!
  - Normal
  - Binomial

Christmas gifts for the nerds in your life

Outline

- Meet (some of) the exponential family!
  - Normal
  - Binomial
  - Poisson

Christmas gifts for the nerds in your life

Outline

- Meet (some of) the exponential family!
  - Normal
  - Binomial
  - Poisson
  - Beta-Binomial

Christmas gifts for the nerds in your life

Outline

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

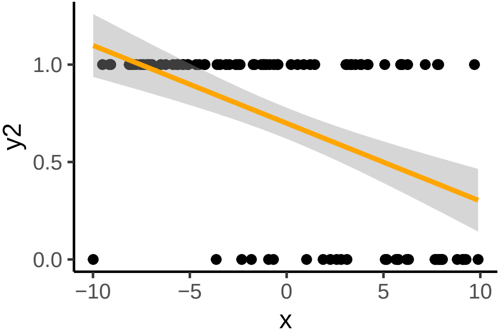
Christmas gifts for the nerds in your life

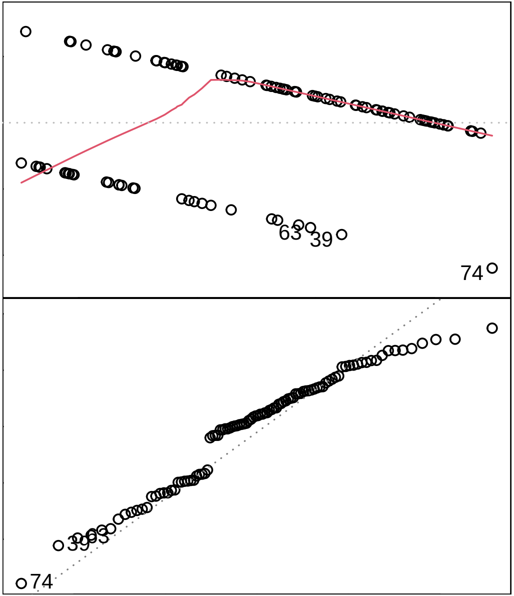
Outline

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

Christmas gifts for the nerds in your life

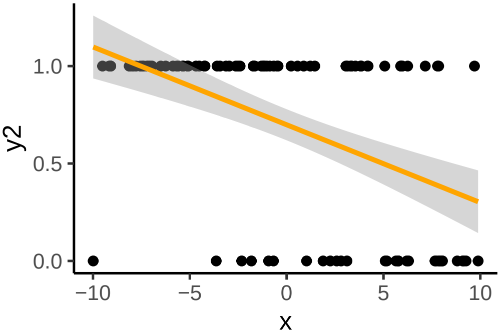
Problem: not everything is normal

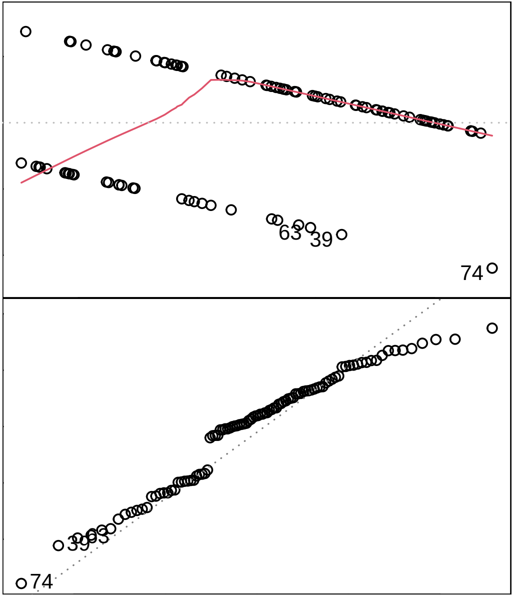




- Some types of data can never be transformed to make the residuals normal

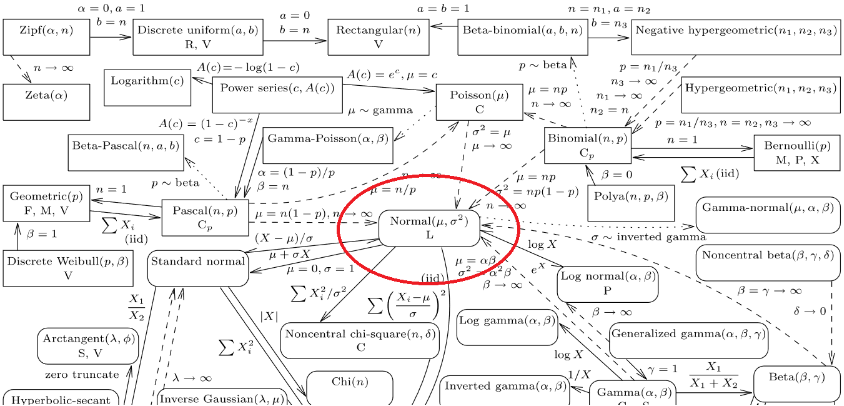
Problem: not everything is normal





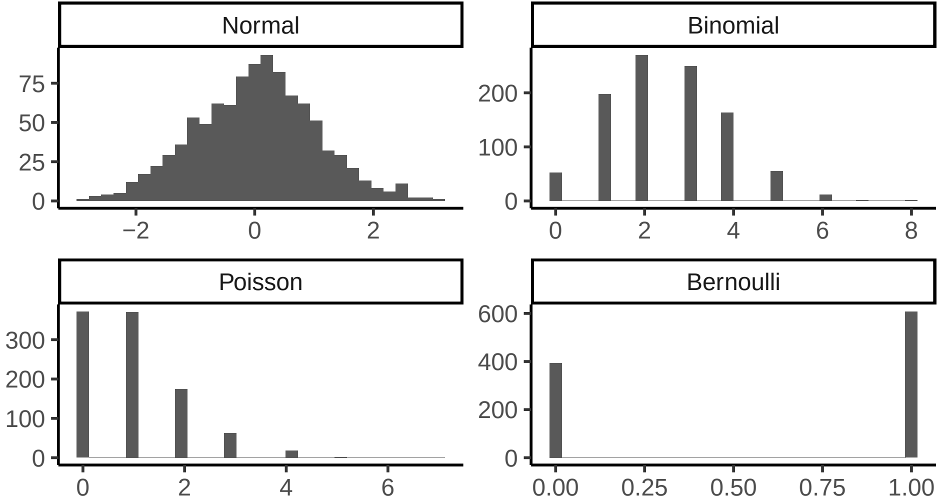
- Some types of data can never be transformed to make the residuals normal
- Solution: **use the distribution that generates the data!**

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

A Galton Board in action:

The Normal Distribution (aka *Gaussian*)

- Imagine many random + and - numbers added together
- If you do this *many* times:

A Galton Board in action:

The Normal Distribution (aka *Gaussian*)

- Imagine many random + and - numbers added together
- If you do this *many* times:
  - Most cancel out (somewhere around 0)

A Galton Board in action:

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)

A Galton Board in action:

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

A Galton Board in action:

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)

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

The Normal Distribution - scary math!

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

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

- Probability distribution function (PDF) for the Normal distribution

The Normal Distribution - scary math!

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

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

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

The Normal Distribution - scary math!

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

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

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

The Normal Distribution - scary math!

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

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

- 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?

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

$$= \sim 0.24$$

In R, this is easy:

The Normal Distribution - scary math!

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

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

- 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?

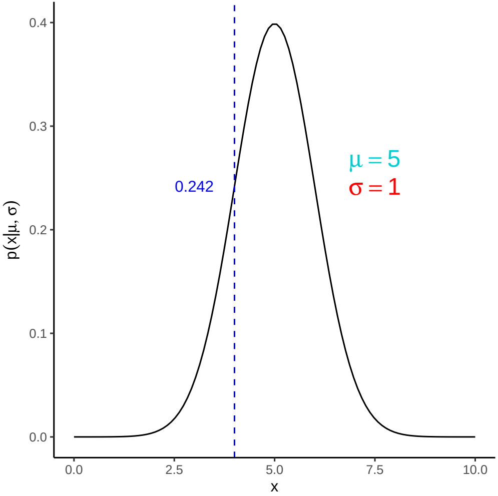
$$p(4|5,1) = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{4-5}{1}\right)^2}$$
$$= \sim 0.24$$

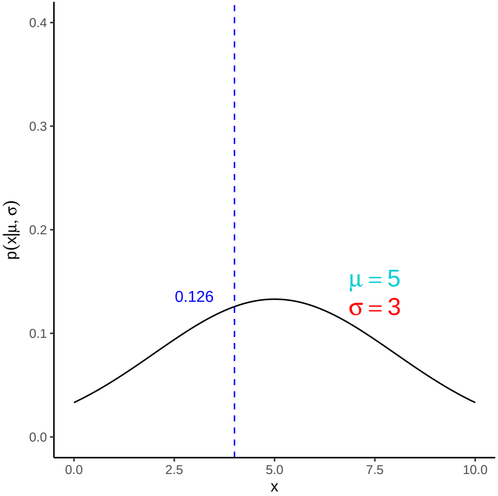
In R, this is easy:

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

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

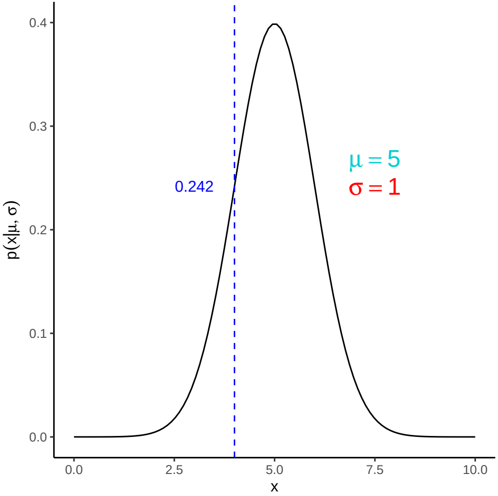
The Normal Distribution

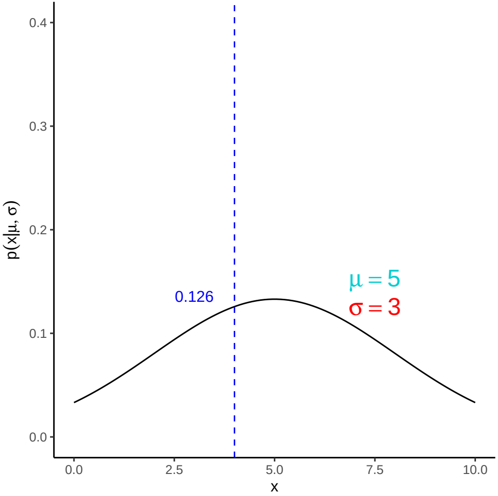




- Probability of x changes with μ and σ

The Normal Distribution

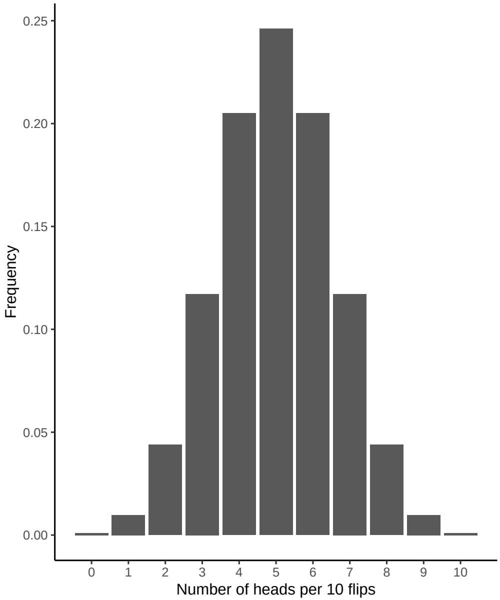




- Probability of x changes with μ and σ
- Left: $\sigma = 1$, Right: $\sigma = 3$

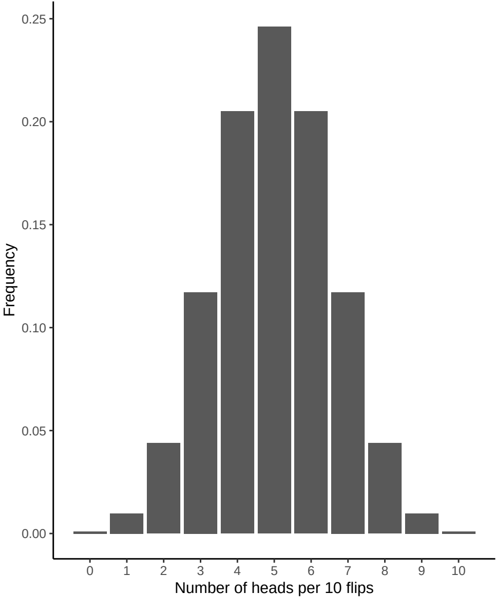
The Binomial Distribution

- Imagine you have 10 coins, and you flip them all



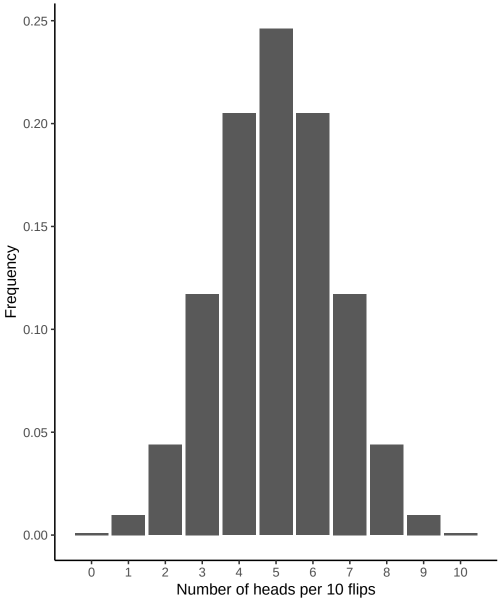
The Binomial Distribution

- Imagine you have 10 coins, and you flip them all
- If you do this *many* times:



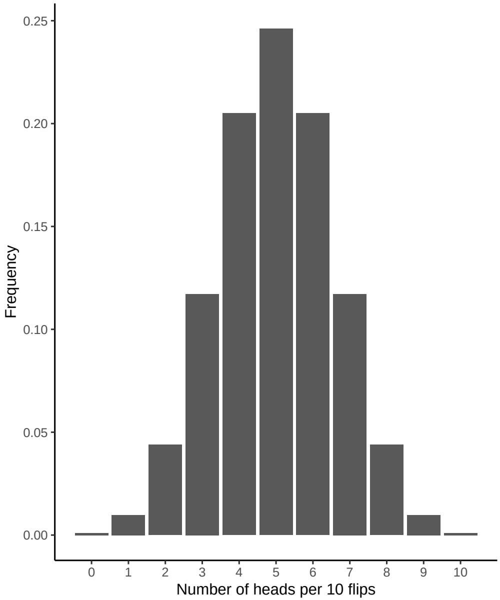
The Binomial Distribution

- Imagine you have 10 coins, and you flip them all
- If you do this *many* times:
    - Most will be about 5 heads/tails



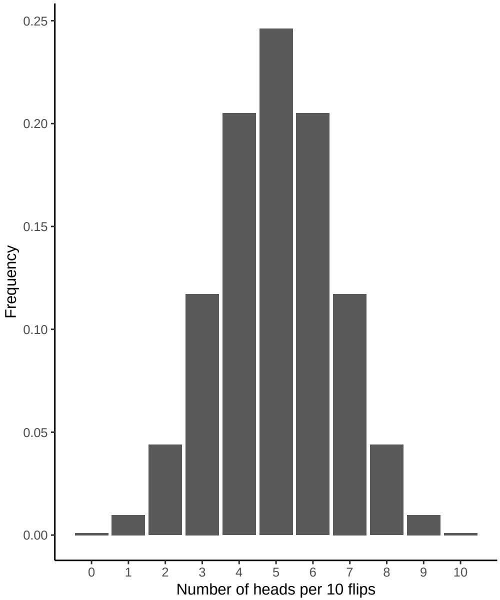
The Binomial Distribution

- 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)



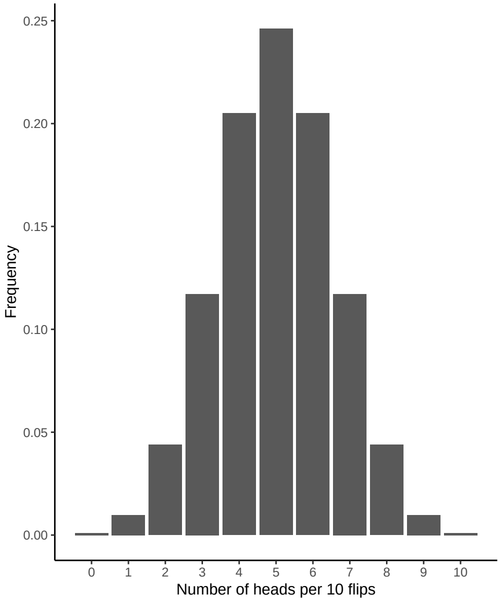
The Binomial Distribution

- 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



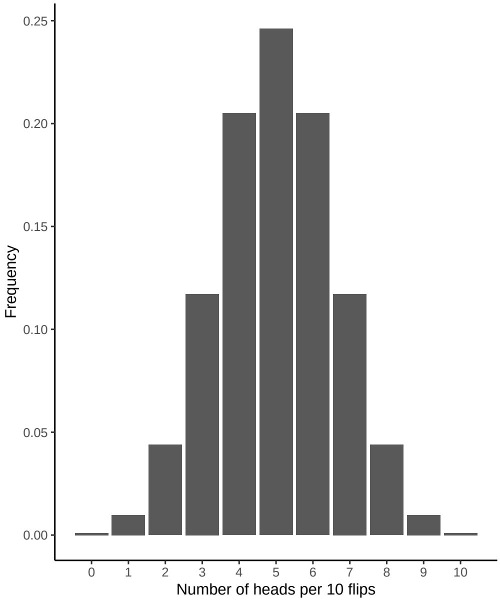
The Binomial Distribution

- 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?



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. . .

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

The Binomial Distribution - scary math!

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

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

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

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

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

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?

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

$$= \sim 0.23$$

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?

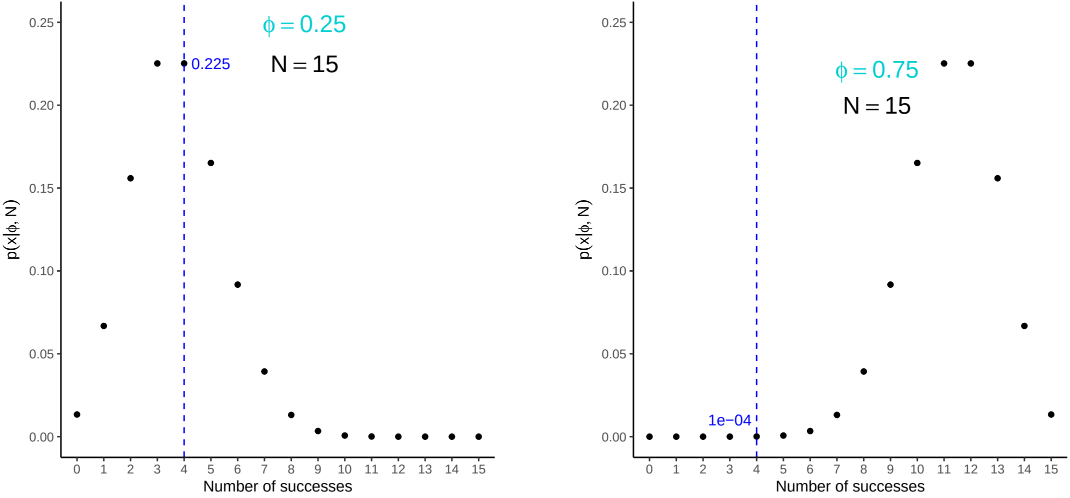
$$p(4|0.25, 15) = \binom{15}{4} 0.25^4 (1-0.25)^{15-4}$$
$$= \sim 0.23$$

In R, this is easy:

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

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

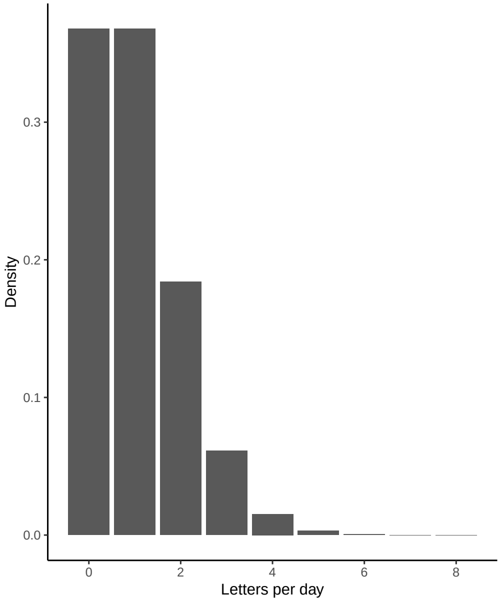
The Binomial Distribution



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

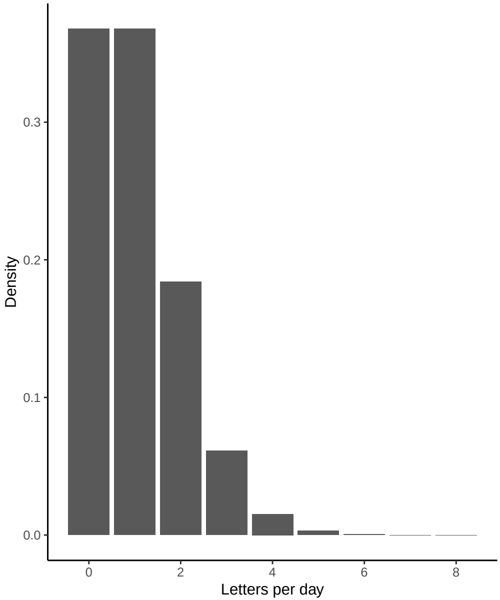
The Poisson Distribution

- Imagine a rare event (e.g. getting a non-junk mail letter)



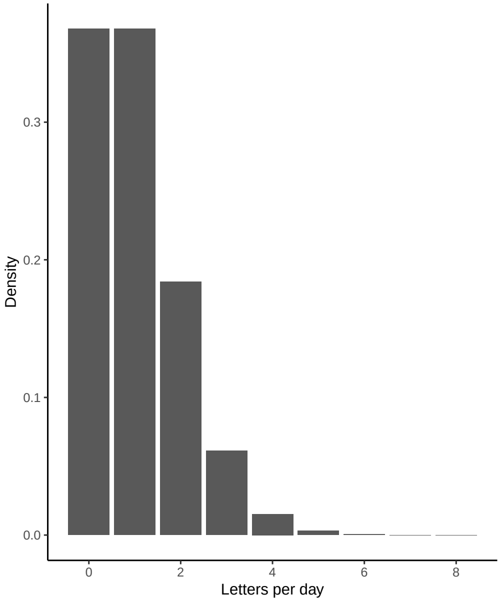
The Poisson Distribution

- Imagine a rare event (e.g. getting a non-junk mail letter)
- If you record the number of events every day:



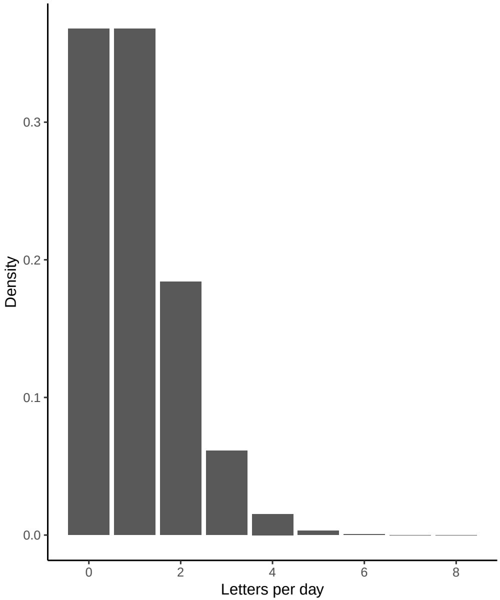
The Poisson Distribution

- 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



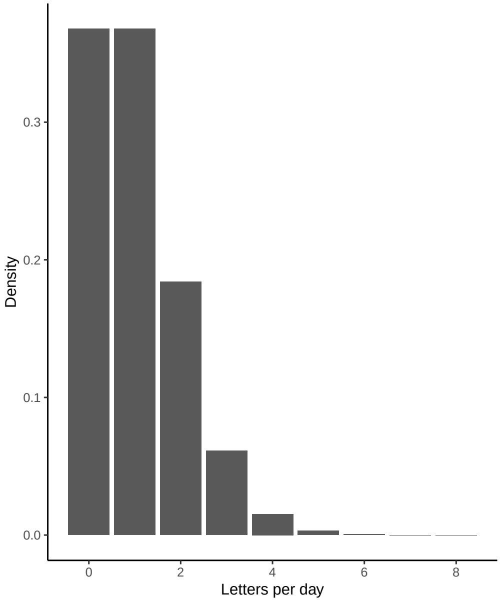
The Poisson Distribution

- 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



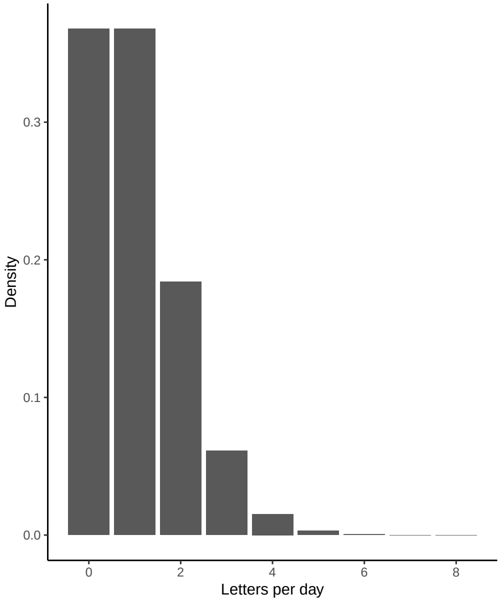
The Poisson Distribution

- 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:



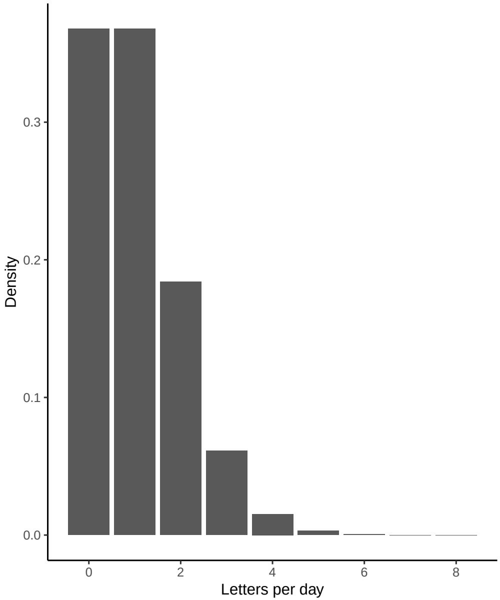
The Poisson Distribution

- 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 bugs caught in a net (per sweep)



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 bugs caught in a net (per sweep)
- Equivalent to Binomial distribution, where N is unknown



The Poisson Distribution - scary math!

- 1 parameter: rate parameter (λ)

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

The Poisson Distribution - scary math!

- 1 parameter: rate parameter (λ)

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

- Probability mass function (PMF)

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

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

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?

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

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?

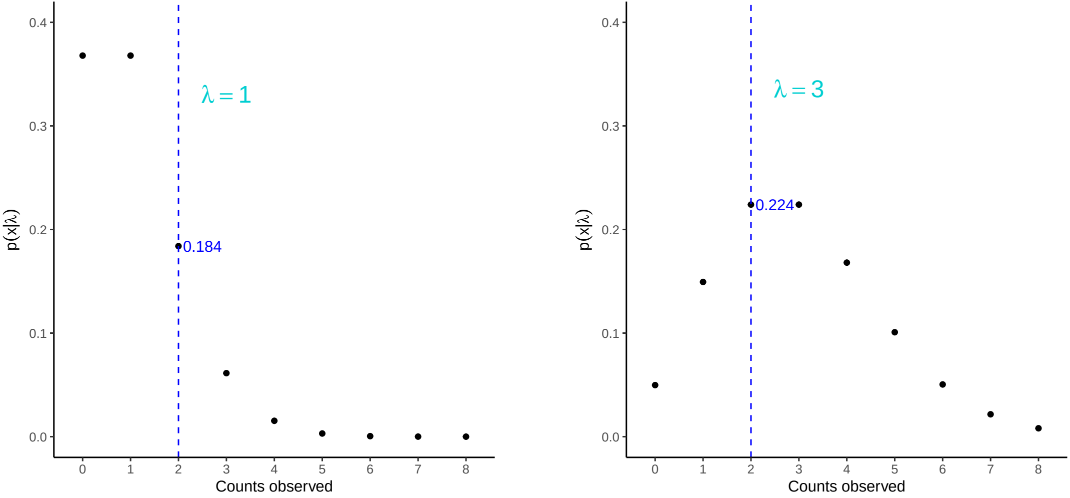
$$\begin{aligned} p(2|1) &= \frac{1^2 e^{-1}}{2!} \\ &= \sim 0.18 \end{aligned}$$

In R, this is easy:

```r
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...

More complications:

- The Normal distribution has a parameter for the mean and SD, but. . .
- What about the Binomial and Poisson distributions?

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 $= \sqrt{Np(1-p)}$

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 $= \sqrt{\lambda}$

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?

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

The Beta Binomial Distibution

- Many "coin-flip" processes have longer tails than standard Binomial
  - e.g. numbers of males/females in families

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

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)

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

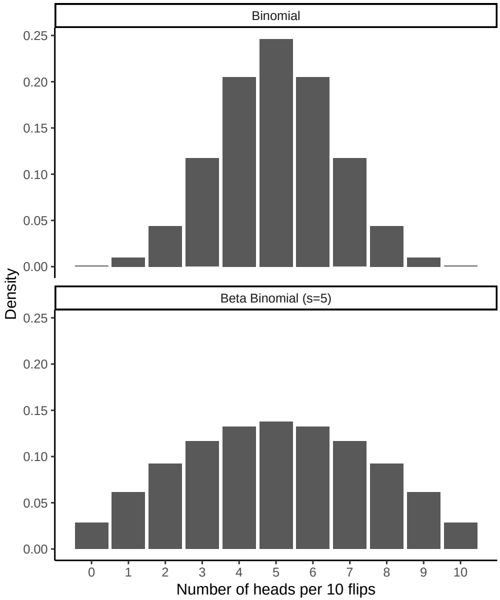
The Beta Binomial Distibution

- 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

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

```r
#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

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

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)

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)

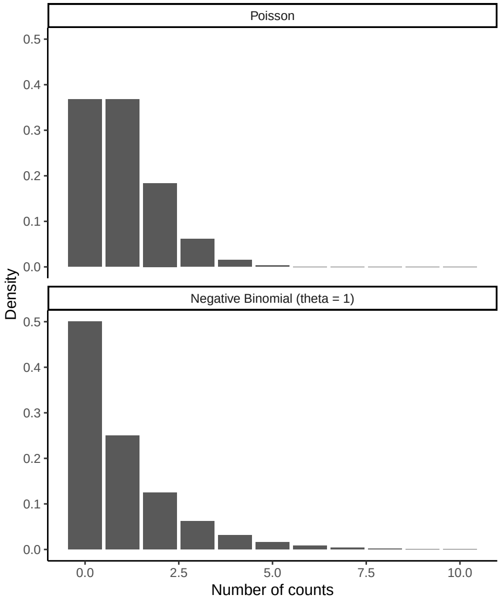
The Negative Binomial Distribution

- 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)

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)

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



Summary of Common "Starter" Distributions

- Continuous data, spanning - or + numbers:

Summary of Common "Starter" Distributions

- Continuous data, spanning - or + numbers:
  - Normal (transformed or regular)

Summary of Common "Starter" Distributions

- Continuous data, spanning - or + numbers:
  - Normal (transformed or regular)
- Count data

Summary of Common “Starter” Distributions

- Continuous data, spanning - or + numbers:
  - Normal (transformed or regular)
- Count data
  - Poisson, Negative Binomial

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

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

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

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

First challenge (Part 1)

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

- Insects caught in a trap (per day)

First challenge (Part 1)

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

- Insects caught in a trap (per day)
- Weight of seeds from a plant

First challenge (Part 1)

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

- Insects caught in a trap (per day)
- Weight of seeds from a plant
- Occupied/unoccupied nest sites

First challenge (Part 1)

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

- Insects caught in a trap (per day)
- Weight of seeds from a plant
- Occupied/unoccupied nest sites
- Chemical concentrations

First challenge (Part 1)

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

- Insects caught in a trap (per day)
- Weight of seeds from a plant
- Occupied/unoccupied nest sites
- Chemical concentrations
- Size of trees (DBH or height)

First challenge (Part 1)

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

- Insects caught in a trap (per day)
- Weight of seeds from a plant
- Occupied/unoccupied nest sites
- Chemical concentrations
- Size of trees (DBH or height)
- Number of male and female bats

Second challenge (Part 2)

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

- Insects caught in a trap (per day): *Poisson or NB*

Second challenge (Part 2)

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

- Insects caught in a trap (per day): *Poisson or NB*
  - `rpois(n,lambda)` or `rnbinom(n,mu,size)`

Second challenge (Part 2)

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

- Insects caught in a trap (per day): *Poisson or NB*
  - `rpois(n,lambda)` or `rnbinom(n,mu,size)`
- Weight of seeds: *Normal*

Second challenge (Part 2)

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

- Insects caught in a trap (per day): *Poisson or NB*
  - `rpois(n,lambda)` or `rnbinom(n,mu,size)`
- Weight of seeds: *Normal*
  - `rnorm(n,mean,sd)`

Second challenge (Part 2)

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

- Insects caught in a trap (per day): *Poisson or NB*
  - `rpois(n,lambda)` or `rnbinom(n,mu,size)`
- Weight of seeds: *Normal*
  - `rnorm(n,mean,sd)`
- Occupied/unoccupied nest sites: *Binomial*

Second challenge (Part 2)

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

- Insects caught in a trap (per day): *Poisson or NB*
  - `rpois(n,lambda)` or `rnbinom(n,mu,size)`
- Weight of seeds: *Normal*
  - `rnorm(n,mean,sd)`
- Occupied/unoccupied nest sites: *Binomial*
  - `rbinom(n, 1, prob)` aka. *Bernoulli* distribution

Second challenge (Part 2)

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

- Insects caught in a trap (per day): *Poisson or NB*
  - `rpois(n,lambda)` or `rnbinom(n,mu,size)`
- Weight of seeds: *Normal*
  - `rnorm(n,mean,sd)`
- Occupied/unoccupied nest sites: *Binomial*
  - `rbinom(n, 1, prob)` aka. *Bernoulli* distribution
- Chemical concentrations in a pond: *Normal*

Second challenge (Part 2)

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

- Insects caught in a trap (per day): *Poisson or NB*
  - `rpois(n,lambda)` or `rnbinom(n,mu,size)`
- Weight of seeds: *Normal*
  - `rnorm(n,mean,sd)`
- Occupied/unoccupied nest sites: *Binomial*
  - `rbinom(n, 1, prob)` aka. *Bernoulli* distribution
- Chemical concentrations in a pond: *Normal*
  - `rnorm(n,mean,sd)`

Second challenge (Part 2)

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

- Insects caught in a trap (per day): *Poisson or NB*
  - `rpois(n,lambda)` or `rnbinom(n,mu,size)`
- Weight of seeds: *Normal*
  - `rnorm(n,mean,sd)`
- Occupied/unoccupied nest sites: *Binomial*
  - `rbinom(n, 1, prob)` aka. *Bernoulli* distribution
- Chemical concentrations in a pond: *Normal*
  - `rnorm(n,mean,sd)`
- Size of trees (DBH or height):*log-Normal*

Second challenge (Part 2)

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

- Insects caught in a trap (per day): *Poisson or NB*
  - `rpois(n,lambda)` or `rnbinom(n,mu,size)`
- Weight of seeds: *Normal*
  - `rnorm(n,mean,sd)`
- Occupied/unoccupied nest sites: *Binomial*
  - `rbinom(n, 1, prob)` aka. *Bernoulli* distribution
- Chemical concentrations in a pond: *Normal*
  - `rnorm(n,mean,sd)`
- Size of trees (DBH or height):*log-Normal*
  - `exp(rnorm(n,mean,sd))`

Second challenge (Part 2)

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

- Insects caught in a trap (per day): *Poisson or NB*
  - `rpois(n,lambda)` or `rnbinom(n,mu,size)`
- Weight of seeds: *Normal*
  - `rnorm(n,mean,sd)`
- Occupied/unoccupied nest sites: *Binomial*
  - `rbinom(n, 1, prob)` aka. *Bernoulli* distribution
- Chemical concentrations in a pond: *Normal*
  - `rnorm(n,mean,sd)`
- Size of trees (DBH or height):*log-Normal*
  - `exp(rnorm(n,mean,sd))`
- Number of male and female bats: *Binomial or Beta Binomial*

Second challenge (Part 2)

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

- Insects caught in a trap (per day): *Poisson or NB*
    - `rpois(n,lambda)` or `rnbinom(n,mu,size)`
- Weight of seeds: *Normal*
    - `rnorm(n,mean,sd)`
- Occupied/unoccupied nest sites: *Binomial*
    - `rbinom(n, 1, prob)` aka. *Bernoulli* distribution
- Chemical concentrations in a pond: *Normal*
    - `rnorm(n,mean,sd)`
- Size of trees (DBH or height):*log-Normal*
    - `exp(rnorm(n,mean,sd))`
- Number of male and female bats: *Binomial or Beta Binomial*
    - `rbinom(n, size, prob)` or `rbetabinom(n,size,m,s)`

Part 2: Maximum likelihood and GLMs

Outline

- Maximum likelihood

Outline

- Maximum likelihood
    - A way to think about data

Outline

- Maximum likelihood
  - A way to think about data
  - Likelihood vs Probability

Outline

- Maximum likelihood
  - A way to think about data
  - Likelihood vs Probability
- Generalized linear models

Outline

- Maximum likelihood
    - A way to think about data
    - Likelihood vs Probability
- Generalized linear models
    - Link functions

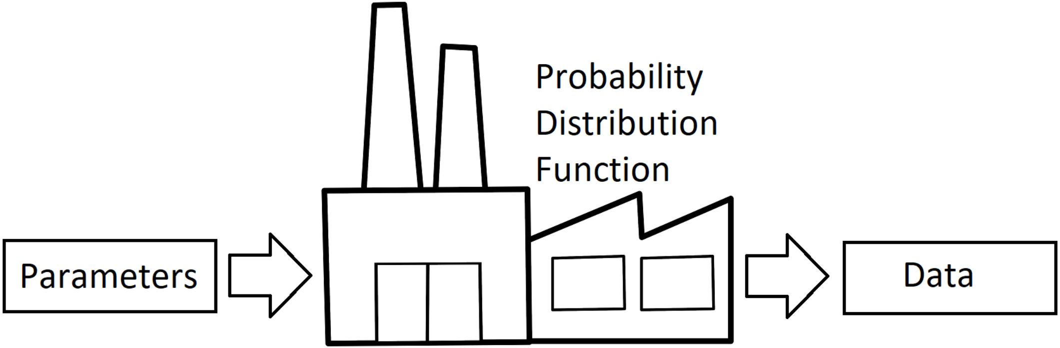
Outline

- Maximum likelihood
    - A way to think about data
    - Likelihood vs Probability
- Generalized linear models
    - Link functions
    - Predictors -> Linear model

How is our data made?

Making data can be thought of as a *factory*

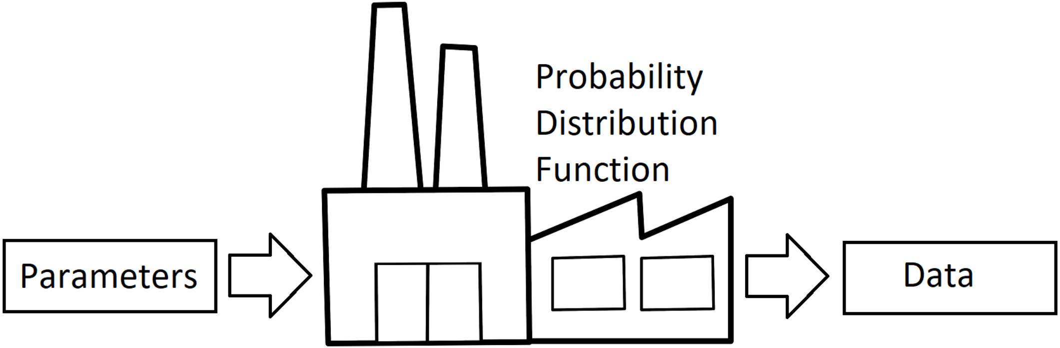
- Input: **parameters** (things that guide the process)



How is our data made?

Making data can be thought of as a *factory*

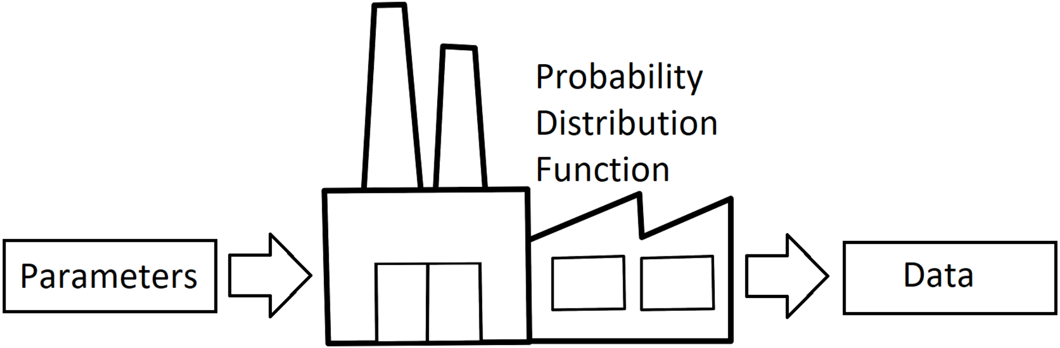
- Input: **parameters** (things that guide the process)
- Process: **probability function**



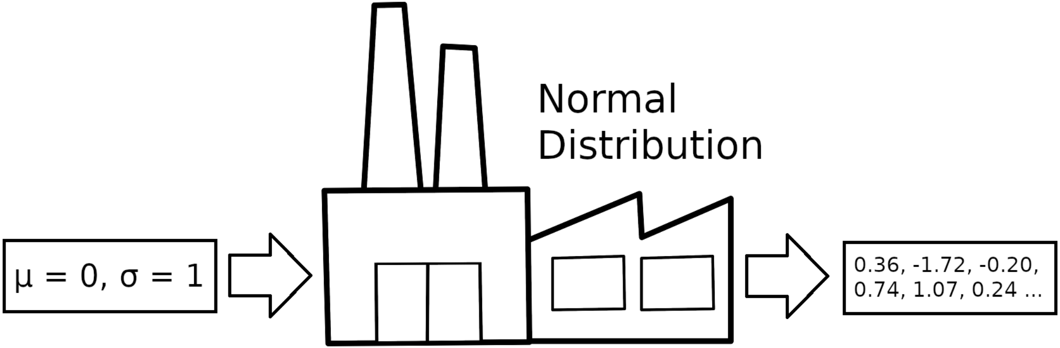
How is our data made?

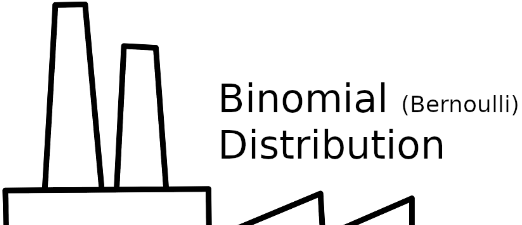
Making data can be thought of as a *factory*

- Input: **parameters** (things that guide the process)
- Process: **probability function**
- Output: **data** (things made by the process)



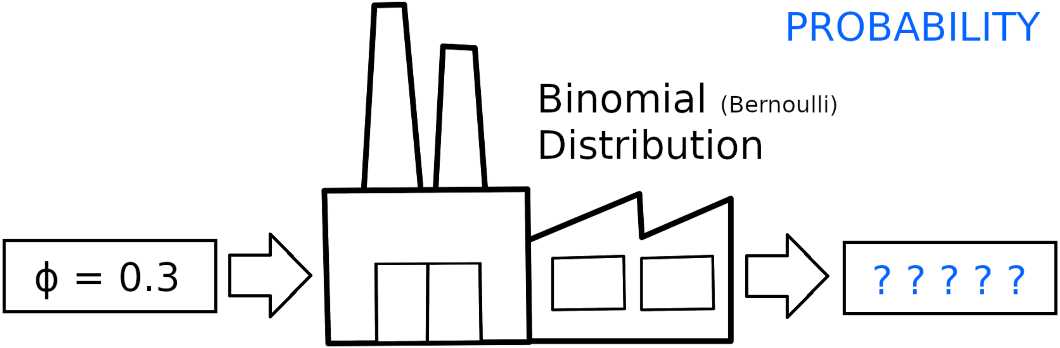
Examples


μ = 0, σ = 1
Normal
Distribution
0.36, -1.72, -0.20,
0.74, 1.07, 0.24 ...

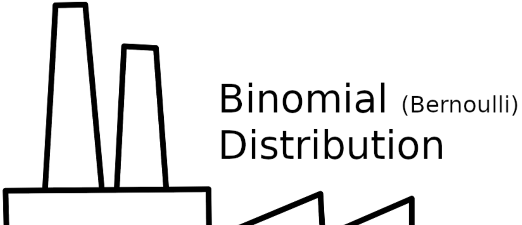

Binomial (Bernoulli)
Distribution

Likelihood vs Probability

PROBABILITY



LIKELIHOOD



Likelihood vs Probability (cont.)

Probability and likelihood both use the same PDF

- "I know that $\phi = 0.3$. What is the chance of getting 2 heads and a tail?"

```r
dbinom(1,1,0.3)*dbinom(1,1,0.3)*dbinom(0,1,0.3)
```

```
## [1] 0.063
```


```r
dbinom(1,1,0.3)*dbinom(1,1,0.3)*dbinom(0,1,0.3)
```

```
## [1] 0.063
```

Likelihood vs Probability (cont.)

Probability and likelihood both use the same PDF

- "I know that $\phi = 0.3$. What is the chance of getting 2 heads and a tail?"

```
dbinom(1,1,0.3)*dbinom(1,1,0.3)*dbinom(0,1,0.3)
```

```
## [1] 0.063
```

- "I got 2 heads and a tail. What is the likelihood that $\phi = 0.3$?"

```
dbinom(1,1,0.3)*dbinom(1,1,0.3)*dbinom(0,1,0.3)
```

```
## [1] 0.063
```

Likelihood vs Probability (cont.)

Let's see how *likelihood* changes with different values of ϕ:

```
#phi = 0.3
dbinom(1,1,0.3)*dbinom(1,1,0.3)*dbinom(0,1,0.3)
```

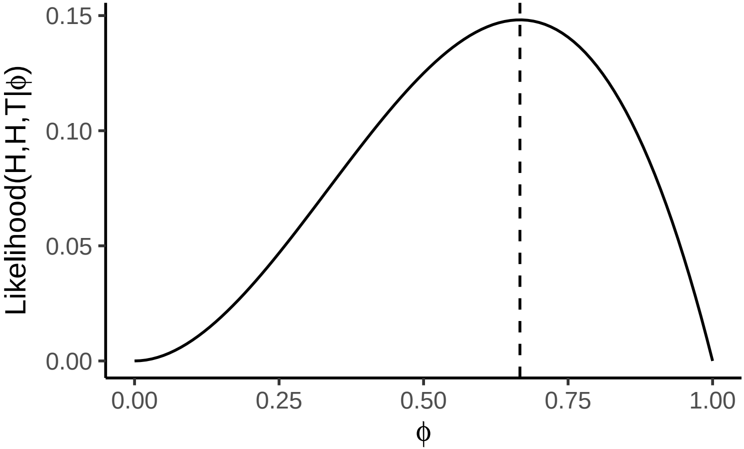
```
## [1] 0.063
```

```
#phi = 0.7
dbinom(1,1,0.7)*dbinom(1,1,0.7)*dbinom(0,1,0.7)
```

```
## [1] 0.147
```

Likelihood of $\phi = 0.7$ is higher, i.e. $\phi = 0.7$ matches our data *better*

Likelihood



The best match (maximum likelihood value) is at $\phi = 0.666$ (2 heads out of 3 flips)

Generalized Linear Models

`glm()` will fit a model like this, and find the ML solution

```r
dat <- data.frame(flips=c(1,1,0)) #Data (2 heads, 1 tail)
mod1 <- glm(flips~1,data=dat,family='binomial') #Note family specification
summary(mod1)
```

```
##
## Call:
## glm(formula = flips ~ 1, family = "binomial", data = dat)
##
## Deviance Residuals:
##       1        2        3
##  0.9005   0.9005  -1.4823
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.6931     1.2247   0.566    0.571
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3.8191  on 2  degrees of freedom
## Residual deviance: 3.8191  on 2  degrees of freedom
## AIC: 5.8191
##
## Number of Fisher Scoring iterations: 4
```

Wait... our estimate should be 0.666 (2/3), not 0.693!

Link functions

- Some parameters of PDFs have *limits*

Link functions

- Some parameters of PDFs have *limits*
    - Normal: $-\infty < \mu < \infty$, $0 < \sigma$

Link functions

- Some parameters of PDFs have *limits*
  - Normal: $-\infty < \mu < \infty$, $0 < \sigma$
  - Binomial: $0 < \phi < 1$

Link functions

- Some parameters of PDFs have *limits*
  - Normal: $-\infty < \mu < \infty$, $0 < \sigma$
  - Binomial: $0 < \phi < 1$
  - Poisson: $0 < \lambda$

Link functions

- Some parameters of PDFs have *limits*
  - Normal: $-\infty < \mu < \infty$, $0 < \sigma$
  - Binomial: $0 < \phi < 1$
  - Poisson: $0 < \lambda$
- GLMs use *link functions* to map values onto the appropriate parameter range

Link functions

- Some parameters of PDFs have *limits*
  - Normal: $-\infty < \mu < \infty$, $0 < \sigma$
  - Binomial: $0 < \phi < 1$
  - Poisson: $0 < \lambda$
- GLMs use *link functions* to map values onto the appropriate parameter range
  - Normal: Identity (i.e. $\times 1$)

Link functions

- Some parameters of PDFs have *limits*
  - Normal: $-\infty < \mu < \infty$, $0 < \sigma$
  - Binomial: $0 < \phi < 1$
  - Poisson: $0 < \lambda$
- GLMs use *link functions* to map values onto the appropriate parameter range
  - Normal: Identity (i.e. $\times 1$)
  - Binomial: Logit

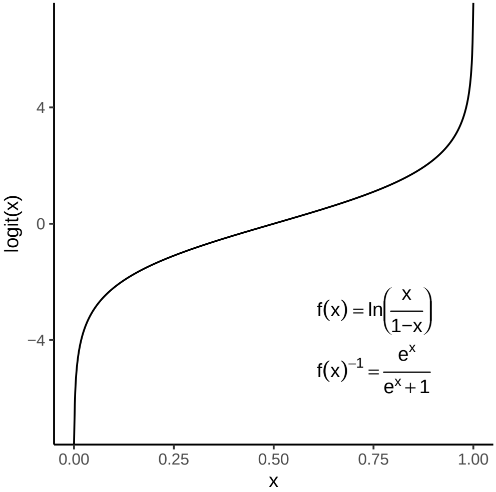
Link functions

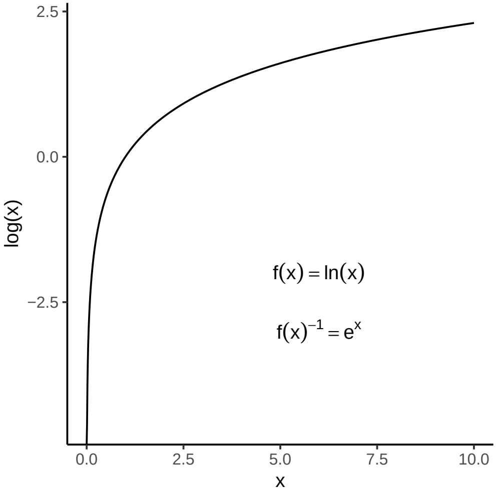
- Some parameters of PDFs have *limits*
  - Normal: $-\infty < \mu < \infty$, $0 < \sigma$
  - Binomial: $0 < \phi < 1$
  - Poisson: $0 < \lambda$
- GLMs use *link functions* to map values onto the appropriate parameter range
  - Normal: Identity (i.e. $\times 1$)
  - Binomial: Logit
  - Poisson/NB: Log

Link functions

- Some parameters of PDFs have *limits*
  - Normal: $-\infty < \mu < \infty$, $0 < \sigma$
  - Binomial: $0 < \phi < 1$
  - Poisson: $0 < \lambda$
- GLMs use *link functions* to map values onto the appropriate parameter range
  - Normal: Identity (i.e. $\times 1$)
  - Binomial: Logit
  - Poisson/NB: Log
- $\mathit{logit}(0.693) = 0.666$, so the GLM actually got it right!

What do these functions look like?

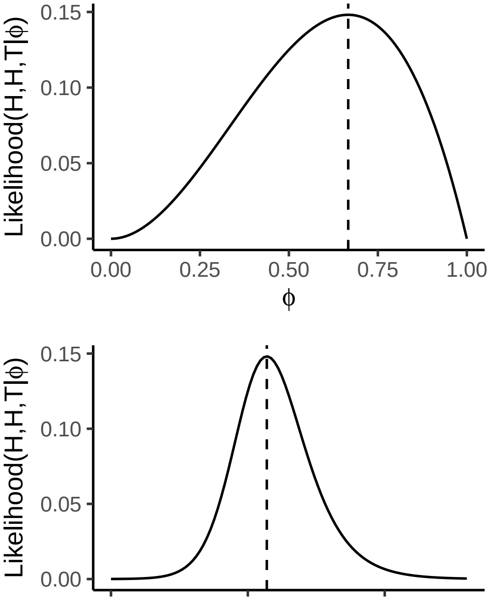




- These functions map parameter values from the appropriate range (0-1 or 0-∞) onto $-\infty$ to $+\infty$

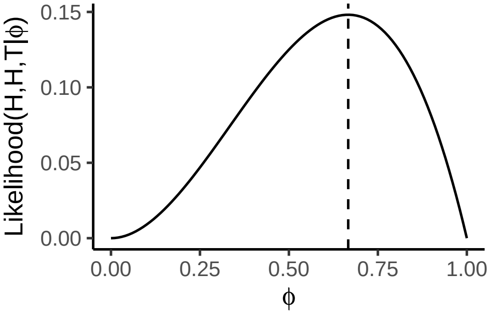
Why do we bother with these link function?

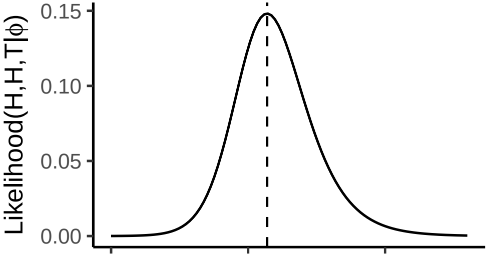
- Likelihood functions are not symmetrical on the regular scale



Why do we bother with these link function?

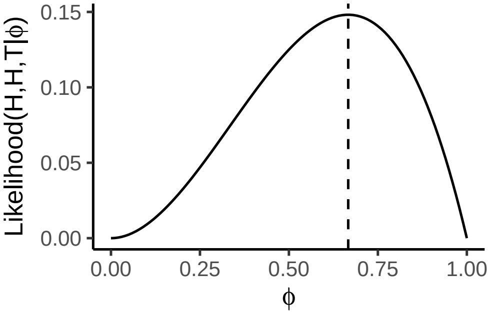
- Likelihood functions are not symmetrical on the regular scale
- On the link-scale, they are closer to a normal distribution

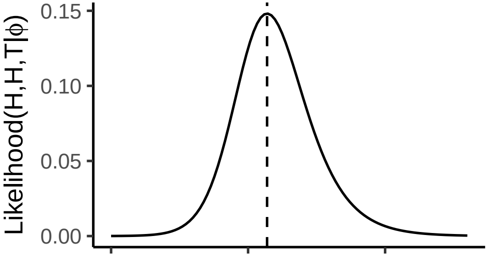




Why do we bother with these link function?

- Likelihood functions are not symmetrical on the regular scale
- On the link-scale, they are closer to a normal distribution
- Makes it easier for R to find the ML estimate (and confidence intervals)





How do linear models fit into this?

- Usually we aren't interested in finding only a single parameter ϕ.

$$logit(\hat{\phi}) = b_0 + b_1 x_1 ... + b_i x_i$$
$$flips \sim Binomial(\hat{\phi})$$

Instead of finding ϕ, **R finds the coefficients (**b_0**,** b_1 **. . .** b_i**) that create** ϕ

How do linear models fit into this?

- Usually we aren't interested in finding only a single parameter ϕ.
- Solution: ϕ becomes a *linear* function of the predictors

$$logit(\hat{\phi}) = b_0 + b_1 x_1 ... + b_i x_i$$

$$flips \sim Binomial(\hat{\phi})$$

Instead of finding ϕ, **R finds the coefficients (**b_0**,** b_1 **. . .** b_i**) that create** ϕ

How do linear models fit into this?

- Usually we aren't interested in finding only a single parameter ϕ.
- Solution: ϕ becomes a *linear* function of the predictors
- Simple linear models take the form:

$$\hat{y} = b_0 + b_1 x_1 ... + b_i x_i$$
$$y \sim Normal(\hat{y}, \sigma)$$

$$logit(\hat{\phi}) = b_0 + b_1 x_1 ... + b_i x_i$$
$$flips \sim Binomial(\hat{\phi})$$

Instead of finding ϕ, **R finds the coefficients (**b_0**,** b_1 **. . .** b_i**) that create** ϕ

How do linear models fit into this?

- Usually we aren't interested in finding only a single parameter ϕ.
- Solution: ϕ becomes a *linear* function of the predictors
- Simple linear models take the form:

$$\hat{y} = b_0 + b_1 x_1 ... + b_i x_i$$
$$y \sim Normal(\hat{y}, \sigma)$$

- Generalized linear models are similar, except that:

$$logit(\hat{\phi}) = b_0 + b_1 x_1 ... + b_i x_i$$
$$flips \sim Binomial(\hat{\phi})$$

Instead of finding ϕ, **R finds the coefficients (**b_0**,** b_1 **. . .** b_i**) that create** ϕ

How do linear models fit into this?

- Usually we aren't interested in finding only a single parameter ϕ.
- Solution: ϕ becomes a *linear* function of the predictors
- Simple linear models take the form:

$$\hat{y} = b_0 + b_1 x_1 ... + b_i x_i$$
$$y \sim Normal(\hat{y}, \sigma)$$

- Generalized linear models are similar, except that:

1. Expected value (ϕ) fed through a link function

$$logit(\hat{\phi}) = b_0 + b_1 x_1 ... + b_i x_i$$
$$flips \sim Binomial(\hat{\phi})$$

Instead of finding ϕ, **R finds the coefficients (b_0, b_1 . . . b_i) that create** ϕ

How do linear models fit into this?

- Usually we aren't interested in finding only a single parameter ϕ.
- Solution: ϕ becomes a *linear* function of the predictors
- Simple linear models take the form:

$$\hat{y} = b_0 + b_1 x_1 ... + b_i x_i$$
$$y \sim Normal(\hat{y}, \sigma)$$

- Generalized linear models are similar, except that:

1. Expected value (ϕ) fed through a link function
2. Data is fit to a non-normal probability function

$$logit(\hat{\phi}) = b_0 + b_1 x_1 ... + b_i x_i$$
$$flips \sim Binomial(\hat{\phi})$$

Instead of finding ϕ, **R finds the coefficients (b_0, b_1 . . . b_i) that create** ϕ

How do I fit GLMs in R?

Syntax and model output is very similar to `lm`

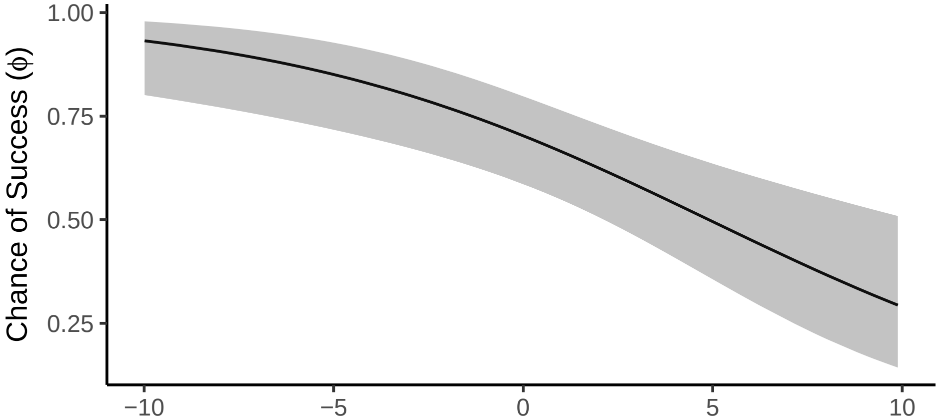
```
# y ~ x, where x is the predictor of y (~1 for just intercept)
mod_binomial <- glm(y2 ~ x1 + x2 , data = d1, family = 'binomial') #Fit a binomial GLM
summary(mod_binomial)
```

```
##
## Call:
## glm(formula = y2 ~ x1 + x2, family = "binomial", data = d1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0050  -0.9493   0.3924   0.8336   1.6806
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.81748    0.25851   3.162 0.001565 **
## x1          -0.17576    0.04871  -3.608 0.000309 ***
## x2           0.30193    0.09950   3.034 0.002410 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 129.49  on 99  degrees of freedom
## Residual deviance: 102.98  on 97  degrees of freedom
## AIC: 108.98
##
## Number of Fisher Scoring iterations: 4
```

How do I get partial effects plots?

crPlot (from car) and ggpredict (ggeffects) work with fitted glm models

```
ggpredict(mod_binomial, terms='x1 [all]') %>% #Partial effect of x1 term
  ggplot(aes(x, predicted)) + geom_line() +
  geom_ribbon(aes(ymin = conf.low, ymax = conf.high), alpha = 0.3) +
  labs(x = 'x1', y = expression(paste('Chance of Success (',phi,')')))
```



A challenger approaches!

- Dr. Roberto Darkley (Robert Barkley's evil nemesis) sent 2 people out to check out some bat roosts in Edmonton and Calgary. One of them dutifully counted bats at each roost, but the other one was really lazy, and just recorded "bats or no bats" (1 or 0).

A challenger approaches!

- Dr. Roberto Darkley (Robert Barkley's evil nemesis) sent 2 people out to check out some bat roosts in Edmonton and Calgary. One of them dutifully counted bats at each roost, but the other one was really lazy, and just recorded "bats or no bats" (1 or 0).
- Fit a model to each of their data (found in `batDatGLM.csv`) using a GLM

A challenger approaches!

- Dr. Roberto Darkley (Robert Barkley's evil nemesis) sent 2 people out to check out some bat roosts in Edmonton and Calgary. One of them dutifully counted bats at each roost, but the other one was really lazy, and just recorded "bats or no bats" (1 or 0).
- Fit a model to each of their data (found in `batDatGLM.csv`) using a GLM
    - `batCounts` should be modeled using a Poisson GLM, and `batPres` should use a Binomial GLM

A challenger approaches!

- Dr. Roberto Darkley (Robert Barkley's evil nemesis) sent 2 people out to check out some bat roosts in Edmonton and Calgary. One of them dutifully counted bats at each roost, but the other one was really lazy, and just recorded "bats or no bats" (1 or 0).
- Fit a model to each of their data (found in `batDatGLM.csv`) using a GLM
  - `batCounts` should be modeled using a Poisson GLM, and `batPres` should use a Binomial GLM
  - Terms to include: `city` and `size` (no interaction)

A challenger approaches!

- Dr. Roberto Darkley (Robert Barkley's evil nemesis) sent 2 people out to check out some bat roosts in Edmonton and Calgary. One of them dutifully counted bats at each roost, but the other one was really lazy, and just recorded "bats or no bats" (1 or 0).
- Fit a model to each of their data (found in `batDatGLM.csv`) using a GLM
  - `batCounts` should be modeled using a Poisson GLM, and `batPres` should use a Binomial GLM
  - Terms to include: `city` and `size` (no interaction)
- How do the models look? Compare the coefficients and see if they are different

A challenger approaches!

- Dr. Roberto Darkley (Robert Barkley's evil nemesis) sent 2 people out to check out some bat roosts in Edmonton and Calgary. One of them dutifully counted bats at each roost, but the other one was really lazy, and just recorded "bats or no bats" (1 or 0).
- Fit a model to each of their data (found in `batDatGLM.csv`) using a GLM
    - `batCounts` should be modeled using a Poisson GLM, and `batPres` should use a Binomial GLM
    - Terms to include: `city` and `size` (no interaction)
- How do the models look? Compare the coefficients and see if they are different
    - Bonus: make a partial regression plot of terms in the Poisson GLM

Model results