Introduction to Applied Bayesian Analysis in Wildlife Ecology

Jeffrey W. Doser

May 11, 2024

Bayesian generalized linear models

Generalized linear models

- Recall the general linear model framework that unified many common statistical approaches (e.g., ANOVA, t-test, linear regression)
- Genera**lized** Linear Models (GLMs) are even broader

Generalized linear models

- Recall the general linear model framework that unified many common statistical approaches (e.g., ANOVA, t-test, linear regression)
- Genera**lized** Linear Models (GLMs) are even broader

Generalized Linear Models

Multinomial regression

Chi-square tests

Poisson regression

General linear models

Logistic regression

Log-linear models

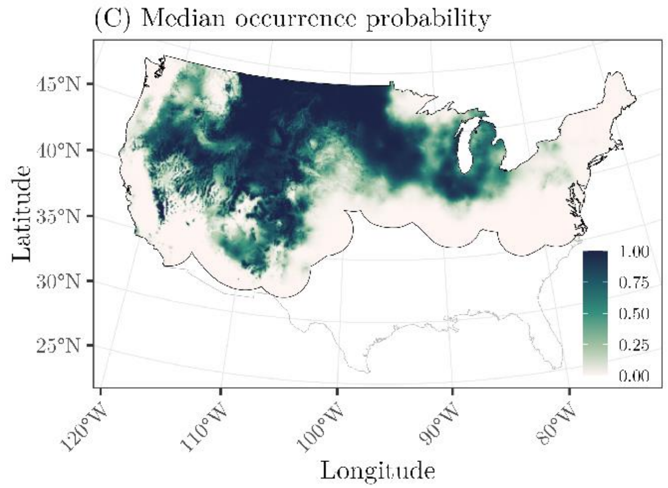
Why are GLMs so useful?

Why are GLMs so useful?

Allow us to model data that do not work well with the Normal distribution (e.g., counts, binary variables, proportions, rates, etc.)

Why are GLMs so useful?

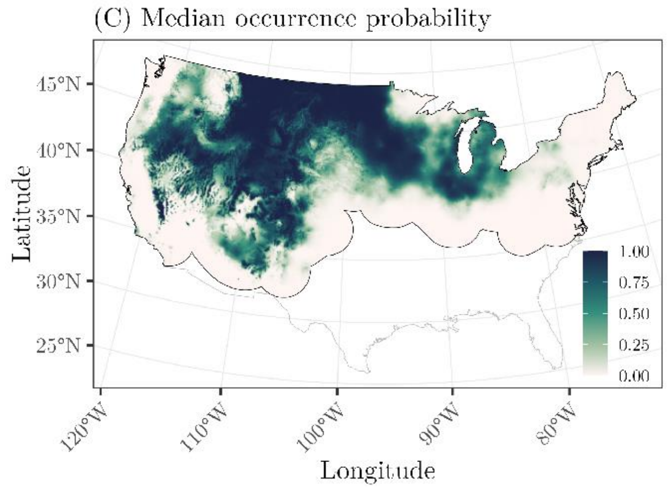
Allow us to model data that do not work well with the Normal distribution (e.g., counts, binary variables, proportions, rates, etc.)



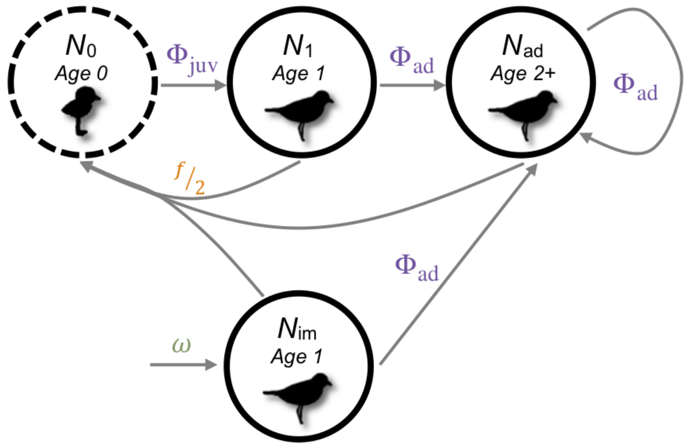
Species distribution models

Why are GLMs so useful?

Allow us to model data that do not work well with the Normal distribution (e.g., counts, binary variables, proportions, rates, etc.)



Species distribution models



Estimate demographic rates

GLM Recipe

Statistical Distribution

Describes random variation in the response y. This is the stochastic (random) part of the model.

GLM Recipe

Statistical Distribution

Describes random variation in the response *y*. This is the stochastic (random) part of the model.

Linear Predictor

Relates the response variable to the predictor variables (parameters are linear)

GLM Recipe

Statistical Distribution

Describes random variation in the response *y*. This is the stochastic (random) part of the model.

Linear Predictor

Relates the response variable to the predictor variables (parameters are linear)

Link Function

Applied to the expectation of the response variable. Connects the mean to its linear predictor.

GLM Recipe

Same recipe for Bayesian and frequentist. A model is a model!

Statistical Distribution

Describes random variation in the response *y*. This is the stochastic (random) part of the model.

Linear Predictor

Relates the response variable to the predictor variables (parameters are linear)

Link Function

Applied to the expectation of the response variable. Connects the mean to its linear predictor.

Some common GLMs

Normal

Random part	$y_i \sim \mathrm{Normal}(\mu_i, \sigma^2)$
Deterministic part	$\eta_i = \beta_0 + \beta_1 \cdot x_i$
Link function	$\eta_i = \mu_i \;\; \text{(identity)}$

Continuous data

Some common GLMs

Normal

Random part $y_i \sim \text{Normal}(\mu_i, \sigma^2)$

Deterministic part $\eta_i = \beta_0 + \beta_1 \cdot x_i$

Link function $\eta_i = \mu_i \;\; \text{(identity)}$

Poisson

Random part $y_i \sim \text{Poisson}(\lambda_i)$

Deterministic part $\eta_i = \beta_0 + \beta_1 \cdot x_i$

Link function $\eta_i = \log(\lambda_i)$

Continuous data

Count data

Some common GLMs

Normal

Random part $y_i \sim \text{Normal}(\mu_i, \sigma^2)$

Deterministic part $\eta_i = \beta_0 + \beta_1 \cdot x_i$

Link function $\eta_i = \mu_i \;\; \text{(identity)}$

Continuous data

Poisson

Random part $y_i \sim \text{Poisson}(\lambda_i)$

Deterministic part $\eta_i = \beta_0 + \beta_1 \cdot x_i$

Link function $\eta_i = \log(\lambda_i)$

Count data

Binomial

Random part $y_i \sim \text{Binomial}(p_i, N)$

Deterministic part $\eta_i = \beta_0 + \beta_1 \cdot x_i$

Link function $\eta_i = \text{logit}(p_i) = \log\left(\frac{p}{1-p}\right)$

Binary data/proportions

GLM Recipe

Same recipe for Bayesian and frequentist. A model is a model!

Statistical Distribution

Describes random variation in the response y. This is the stochastic (random) part of the model.

Linear Predictor

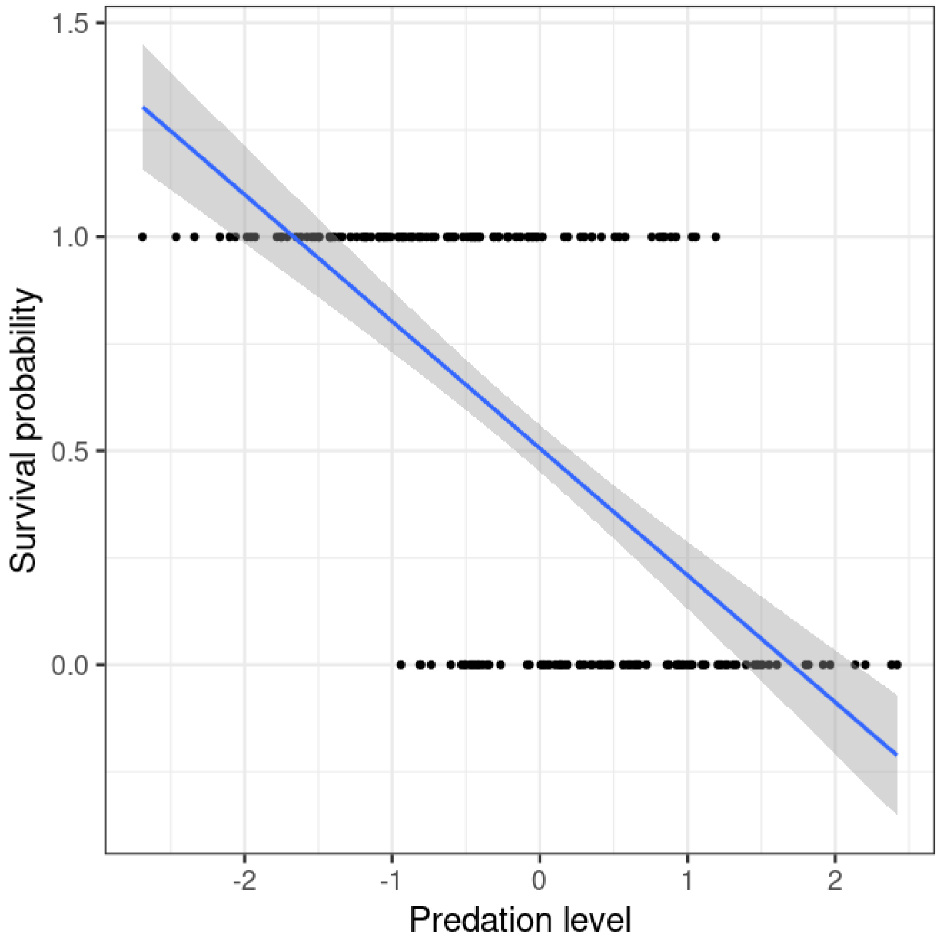
Relates the response variable to the predictor variables (parameters are linear)

Link Function

Applied to the expectation of the response variable. Connects the mean to its linear predictor.

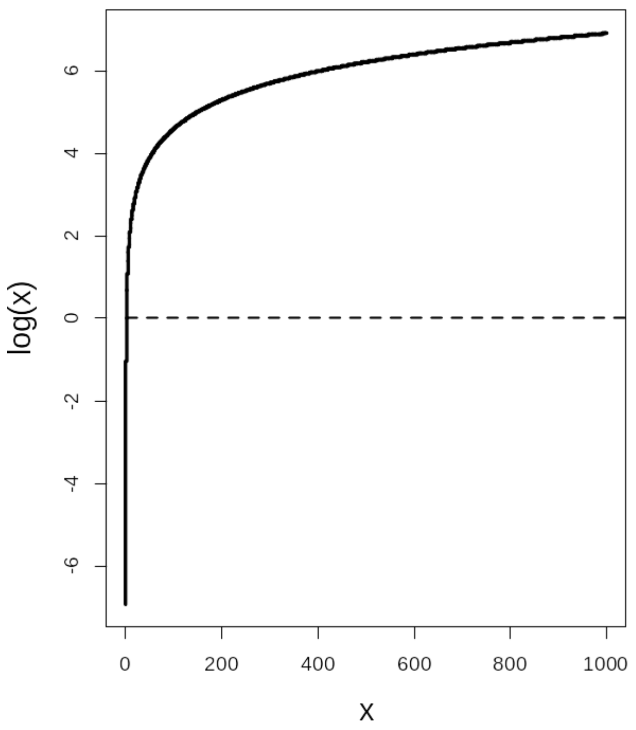
Why do we need link functions?

- Without link functions, estimated values would fall outside of the possible values allowed by the data
  - Predictions are often not helpful
  - Inference is unreliable (i.e., you can't trust the estimated parameter values)
- Remember: we want our model to accurately characterize the data we have



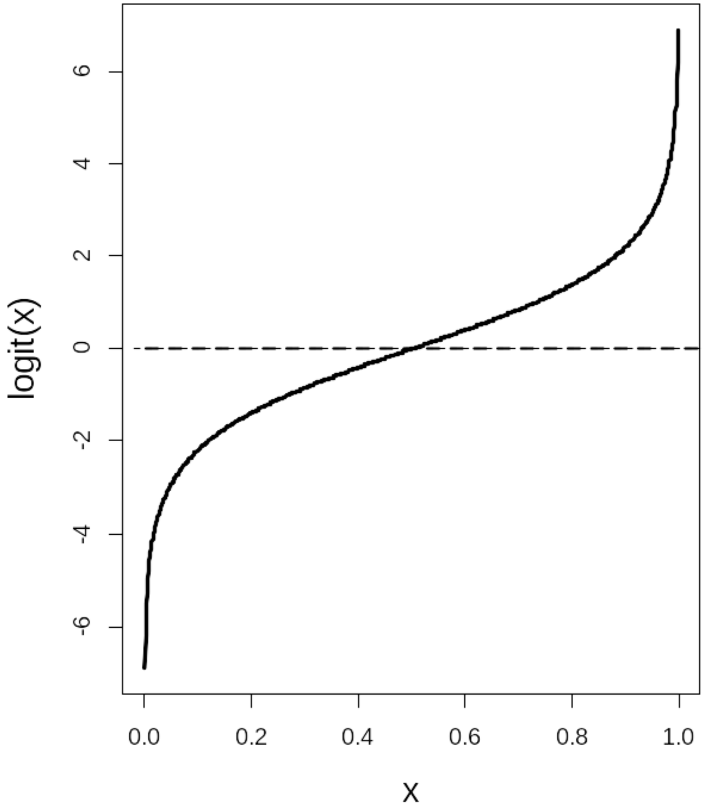
More on link functions

- Suppose we are modeling count data (whole numbers)
- Expected count must be greater than or equal to zero
- The log function (base e) takes positive values and transforms them to continuous positive and negative values



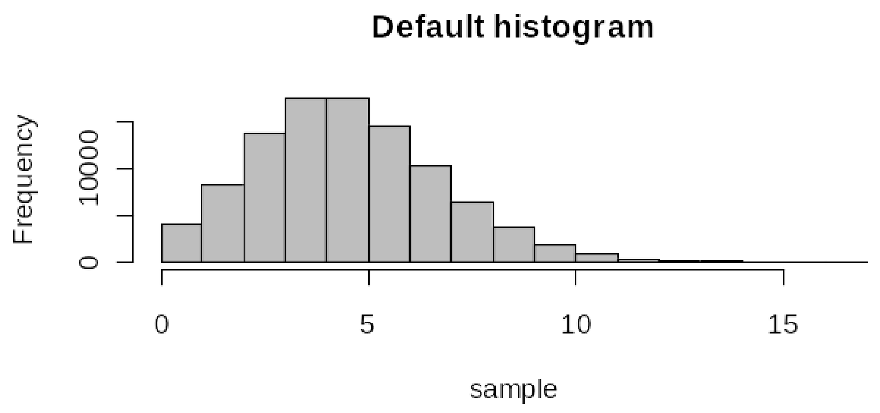
More on link functions

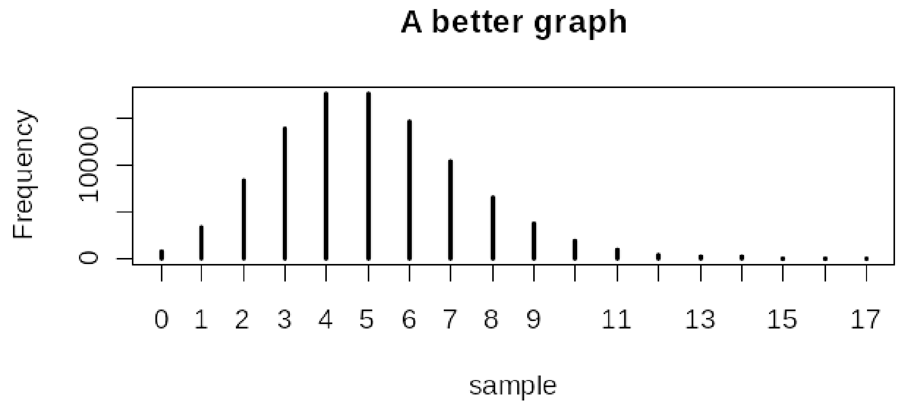
- Suppose we are modeling the probability of an event occurring
- Probability must be between 0 and 1
- The logit function takes values between 0 and 1 and transforms them to continuous positive and negative values



Poisson GLMs

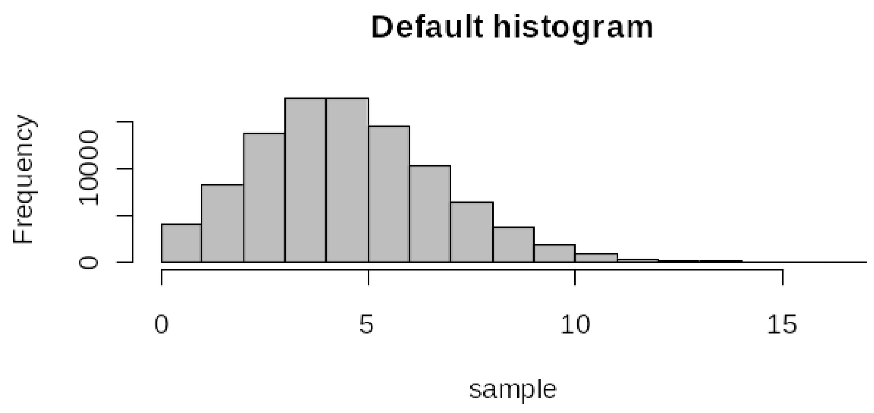
- Poisson distribution is a discrete distribution that is useful for modeling counts

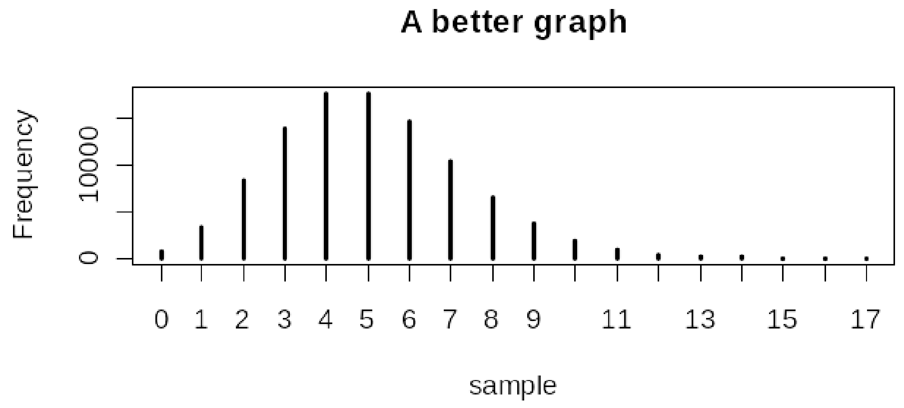




Poisson GLMs

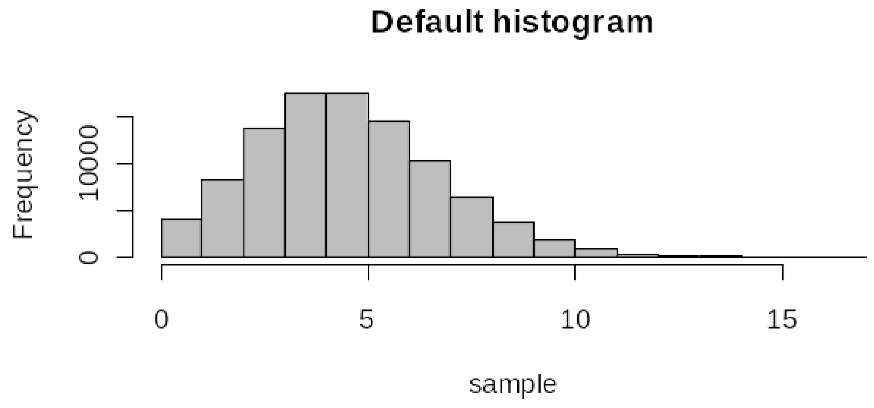
- Poisson distribution is a discrete distribution that is useful for modeling counts
- Notation: y ~ Poisson(λ), where λ > 0
- Mean: λ
- Variance: λ

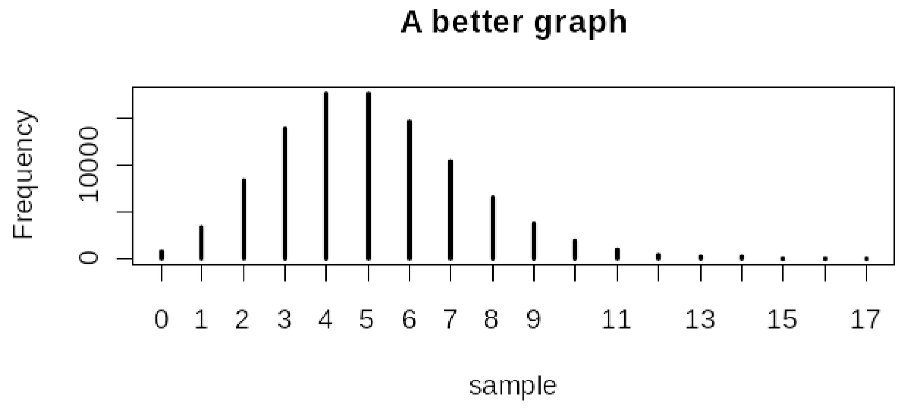




Poisson GLMs

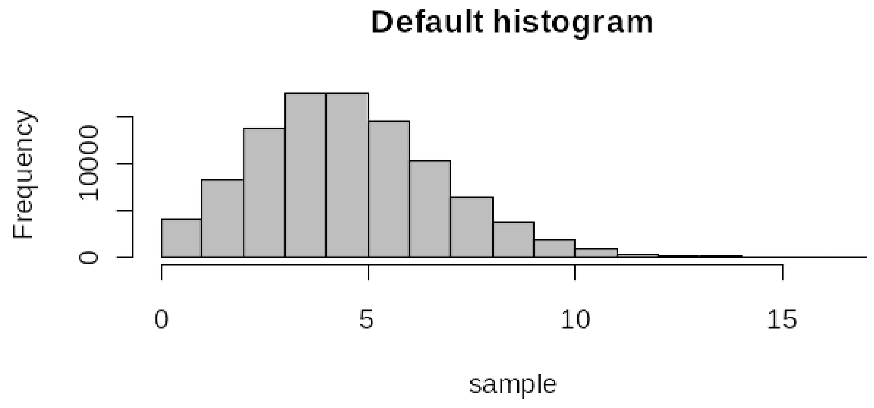
- Poisson distribution is a discrete distribution that is useful for modeling counts
- Notation: y ~ Poisson(λ), where λ > 0
- Mean: λ
- Variance: λ
- Possible values for y: positive integer numbers
    - Also called the **support** of a distribution

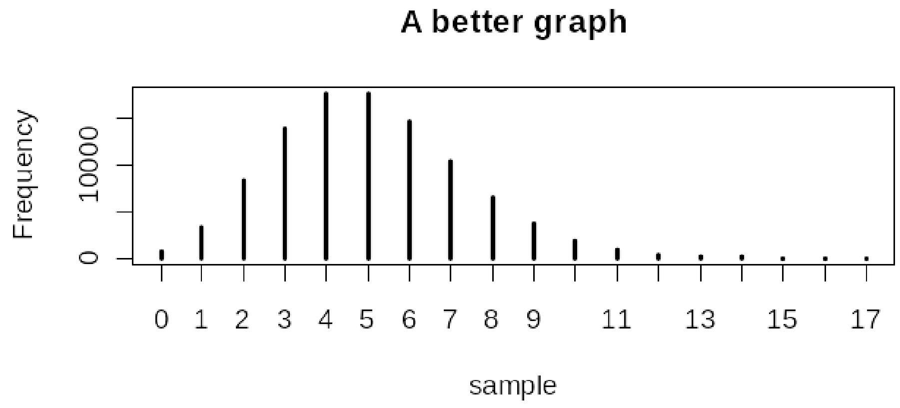




Poisson GLMs

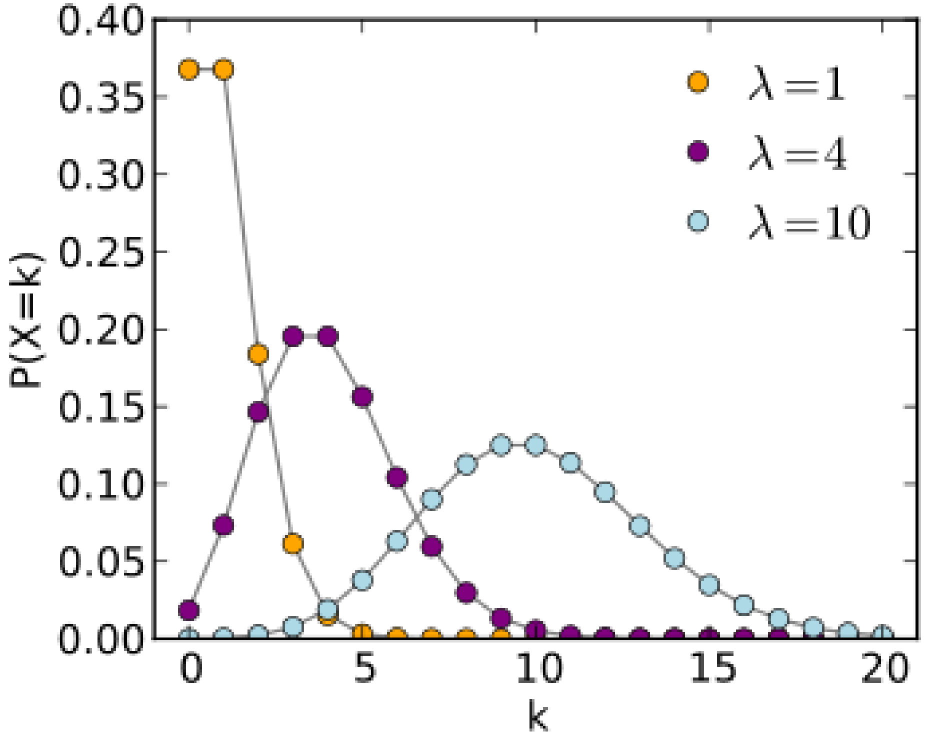
- Poisson distribution is a discrete distribution that is useful for modeling counts
- Notation: y ~ Poisson(λ), where λ > 0
- Mean: λ
- Variance: λ
- Possible values for y: positive integer numbers
  - Also called the **support** of a distribution
- Examples:
  - Number of birds that fly over a migration site within 10 minutes
  - Number of plants per sample quadrat





Poisson distribution

What do you notice about the shape of the curve as λ gets larger?



Special things to consider for Bayesian Poisson GLM

Special things to consider for Bayesian Poisson GLM

$$y_i \sim \text{Poisson}(\lambda_i)$$
$$\eta_i = \beta_0 + \beta_1 \cdot x_i$$
$$\eta_i = \log(\lambda_i)$$

Special things to consider for Bayesian Poisson GLM

- Not much!

$$
\begin{aligned}
y_i &\sim \text{Poisson}(\lambda_i) \\
\eta_i &= \beta_0 + \beta_1 \cdot x_i \\
\eta_i &= \log(\lambda_i)
\end{aligned}
$$

Special things to consider for Bayesian Poisson GLM

- Not much!
- What do we need a prior for?
- If we assume λ is constant, we need a prior for λ
  - Gamma distribution is good
- If we include any covariates (i.e., predictor variables), we specify priors on the intercept and regression coefficients just like with a linear model

$$y_i \sim \text{Poisson}(\lambda_i)$$
$$\eta_i = \beta_0 + \beta_1 \cdot x_i$$
$$\eta_i = \log(\lambda_i)$$

Example: Bayesian Poisson GLM

Hooded warbler relative abundance

4a-poisson-glm-brms.R

Binomial GLM

Binomial GLM

- Overall idea: model a binary response against a linear predictor via a link function
- Also called **Logistic Regression** for a binary response

Binomial GLM

- Overall idea: model a binary response against a linear predictor via a link function
- Also called **Logistic Regression** for a binary response

Binomial

Random part	$y_i \sim \text{Binomial}(p_i, N)$
Deterministic part	$\eta_i = \beta_0 + \beta_1 \cdot x_i$
Link function	$\eta_i = \text{logit}(p_i) = \log\left(\frac{p}{1-p}\right)$

Binomial GLMs/Logistic Regression

- Used for situations with binary outcomes: Bernoulli trial
  - Heads or tails of a coin
  - Presence/absence of a species
  - Dead/alive
  - Success/failure of breeding
  - Occurrence (proportion) of a color morph

Binomial GLMs/Logistic Regression

- Suppose we flip a coin N times (independent trials). Our data y is the number of successes (heads) out of the N trials

Binomial GLMs/Logistic Regression

- Suppose we flip a coin N times (independent trials). Our data y is the number of successes (heads) out of the N trials
- Pr(y heads given N trials) → depends on the probability (p) of success of each trial

Binomial GLMs/Logistic Regression

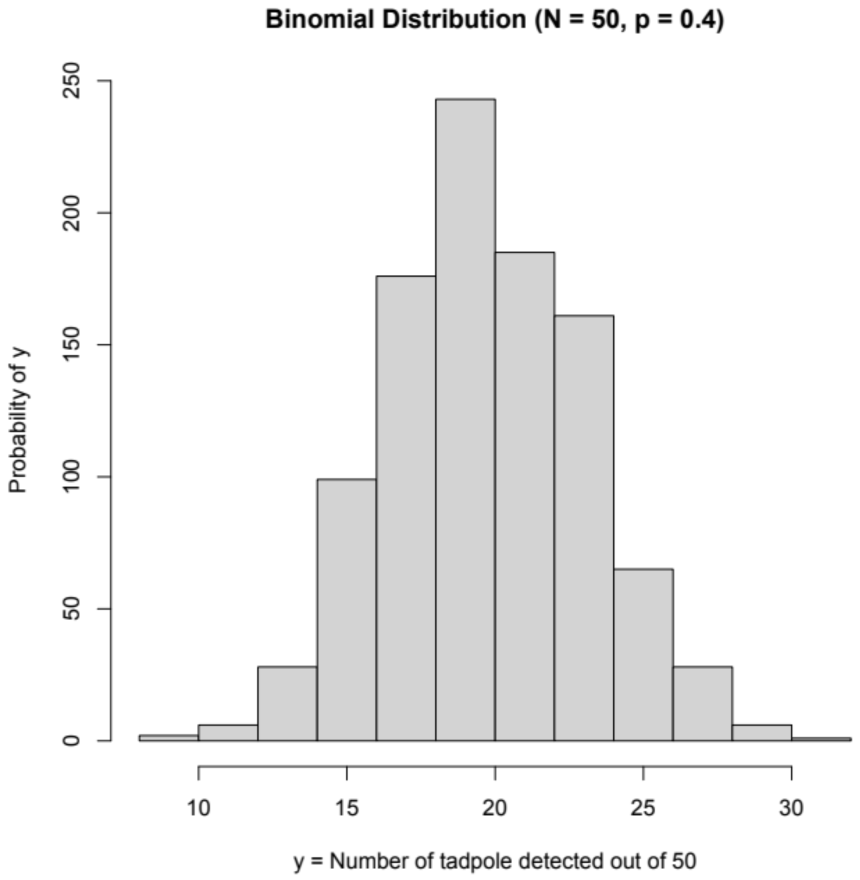
- Suppose we flip a coin N times (independent trials). Our data y is the number of successes (heads) out of the N trials
- Pr(y heads given N trials) → depends on the probability (p) of success of each trial
- Possible outcomes of y → (0, 1, 2, …, N)

Binomial GLMs/Logistic Regression

- Suppose we flip a coin N times (independent trials). Our data y is the number of successes (heads) out of the N trials
- Pr(y heads given N trials) → depends on the probability (p) of success of each trial
- Possible outcomes of y → (0, 1, 2, …, N)
- The Bernoulli distribution is a special case of Binomial when N = 1
- May also see proportion data described as Binomial data
- Ex: Proportion of leaves with caterpillars
- This proportion is y / N (the number of leaves with caterpillars divided by the total number of leaves)

Binomial distribution

- Two parameters (p and N), but generally N is observed and not treated as a parameter in analysis
- Bernoulli distribution is a special case when N = 1
- Note that when N is big, the binomial distribution closely resembles the shape of a normal distribution



Interpreting the binomial GLM

$$y_i \sim \text{Binomial}(p_i, N_i)$$
$$\eta_i = \beta_0 + \beta_1 \cdot x_i$$
$$\eta_i = \text{logit}(p_i) = \log\left(\frac{p_i}{1 - p_i}\right)$$

- If $\beta_1 > 0$, p increases as x increases
- If $\beta_1 < 0$, p decreases as x increases
- If $\beta_1 = 0$, p is constant across all values of x
- Together with plots of your results, this provides a good way of interpreting Binomial GLM results

Interpreting the binomial GLM: Odds

- Odds are an alternative way to represent chance and assess probability
- Odds of a "success"

$$\text{odds}(y_i = 1) = \frac{p_i}{1-p_i}$$

Interpreting the binomial GLM: Odds

- Odds are an alternative way to represent chance and assess probability
- Odds of a "success"

$$\text{odds}(y_i = 1) = \frac{p_i}{1-p_i}$$

- In words: the probability an event will occur divided by the probability the event will not occur

Interpreting the binomial GLM: Odds

- Odds are an alternative way to represent chance and assess probability
- Odds of a "success"

$$\text{odds}(y_i = 1) = \frac{p_i}{1-p_i}$$

- In words: the probability an event will occur divided by the probability the event will not occur
- Odds = 1 means the event is equally likely to occur as not occur

Interpreting the binomial GLM: Odds

- Odds are an alternative way to represent chance and assess probability
- Odds of a "success"

$$\text{odds}(y_i = 1) = \frac{p_i}{1 - p_i}$$

- In words: the probability an event will occur divided by the probability the event will not occur
- Odds = 1 means the event is equally likely to occur as not occur
- Example: odds of rolling a 6 on a six-sided dice:
  - (1 / 6) / (1 - (1 / 6)) = 1 / 5 = 0.2

A little bit of math

- The logit transformation is the log odds, which we can manipulate below

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \log(\text{odds}(y_i))$$

A little bit of math

- The logit transformation is the log odds, which we can manipulate below

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \log(\text{odds}(y_i))$$

$$\text{odds}(y_i) = \frac{p_i}{1-p_i} = \exp(\text{logit}(p_i)) = e^{\beta_0+\beta_1 \cdot x_i} = e^{\beta_0} e^{\beta_1 \cdot x}$$

Interpreting binomial GLMs: Odds ratios

- How do the odds change when x increases by 1 unit?

$$e^{\beta_0}\, e^{\beta_1 \cdot (x+1)} = e^{\beta_0}\, e^{\beta_1 \cdot x}\, e^{\beta_1}$$

Interpreting binomial GLMs: Odds ratios

- How do the odds change when x increases by 1 unit?

$$e^{\beta_0}\, e^{\beta_1 \cdot (x+1)} = e^{\beta_0}\, e^{\beta_1 \cdot x}\, e^{\beta_1}$$

- How does this compare to our value with x? We form an odds ratio:

$$\frac{\text{odds}_{x+1}}{\text{odds}_x} = \frac{e^{\beta 0}\, e^{\beta 1 \cdot (x+1)}}{e^{\beta 0}\, e^{\beta 1 \cdot x}} = \frac{e^{\beta 0}\, e^{\beta 1 \cdot x}\, e^{\beta 1}}{e^{\beta 0}\, e^{\beta 1 \cdot x}} = e^{\beta_1}$$

Interpreting binomial GLMs: Odds ratios

- How do the odds change when x increases by 1 unit?

$$e^{\beta_0}\, e^{\beta_1 \cdot (x+1)} = e^{\beta_0}\, e^{\beta_1 \cdot x}\, e^{\beta_1}$$

- How does this compare to our value with x? We form an odds ratio:

$$\frac{\text{odds}_{x+1}}{\text{odds}_x} = \frac{e^{\beta 0}\, e^{\beta 1 \cdot (x+1)}}{e^{\beta 0}\, e^{\beta 1 \cdot x}} = \frac{e^{\beta 0}\, e^{\beta 1 \cdot x}\, e^{\beta 1}}{e^{\beta 0}\, e^{\beta 1 \cdot x}} = e^{\beta_1}$$

Interpretation: for every one unit increase in x, the odds increase multiplicatively by exp(β_1)

Example: Odds ratios

$$\frac{\text{odds}_{x+1}}{\text{odds}_x} = \frac{e^{\beta_0} e^{\beta_1 \cdot (x+1)}}{e^{\beta_0} e^{\beta_1 \cdot x}} = \frac{e^{\beta_0} e^{\beta_1 \cdot x} e^{\beta_1}}{e^{\beta_0} e^{\beta_1 \cdot x}} = e^{\beta_1}$$

- Suppose we are studying presence/absence of white-tailed deer across multiple forest stands in an intensively managed forest region.
- Suppose x = 1 if the stand was logged in the last 10 years, x = 0 if it was not
- We fit our model using `brms` and found $\beta_1 = 1.5$

Example: Odds ratios

$$\frac{\text{odds}_{x+1}}{\text{odds}_x} = \frac{e^{\beta_0} e^{\beta_1 \cdot (x+1)}}{e^{\beta_0} e^{\beta_1 \cdot x}} = \frac{e^{\beta_0} e^{\beta_1 \cdot x} e^{\beta_1}}{e^{\beta_0} e^{\beta_1 \cdot x}} = e^{\beta_1}$$

- Suppose we are studying presence/absence of white-tailed deer across multiple forest stands in an intensively managed forest region.
- Suppose x = 1 if the stand was logged in the last 10 years, x = 0 if it was not
- We fit our model using `brms` and found $\beta_1 = 1.5$

Interpretation: white-tailed deer are exp(1.5) = 4.48 times more likely to use a recently logged stand than a stand that wasn't recently logged.

Special things to consider with Bayesian binomial GLMs

Special things to consider with Bayesian binomial GLMs

- If p is constant, a good choice for our prior distribution is a beta distribution or a uniform(0, 1) distribution

Special things to consider with Bayesian binomial GLMs

- If p is constant, a good choice for our prior distribution is a beta distribution or a uniform(0, 1) distribution
- If including covariates, we again assign priors for the regression coefficients

Special things to consider with Bayesian binomial GLMs

- If p is constant, a good choice for our prior distribution is a beta distribution or a uniform(0, 1) distribution
- If including covariates, we again assign priors for the regression coefficients

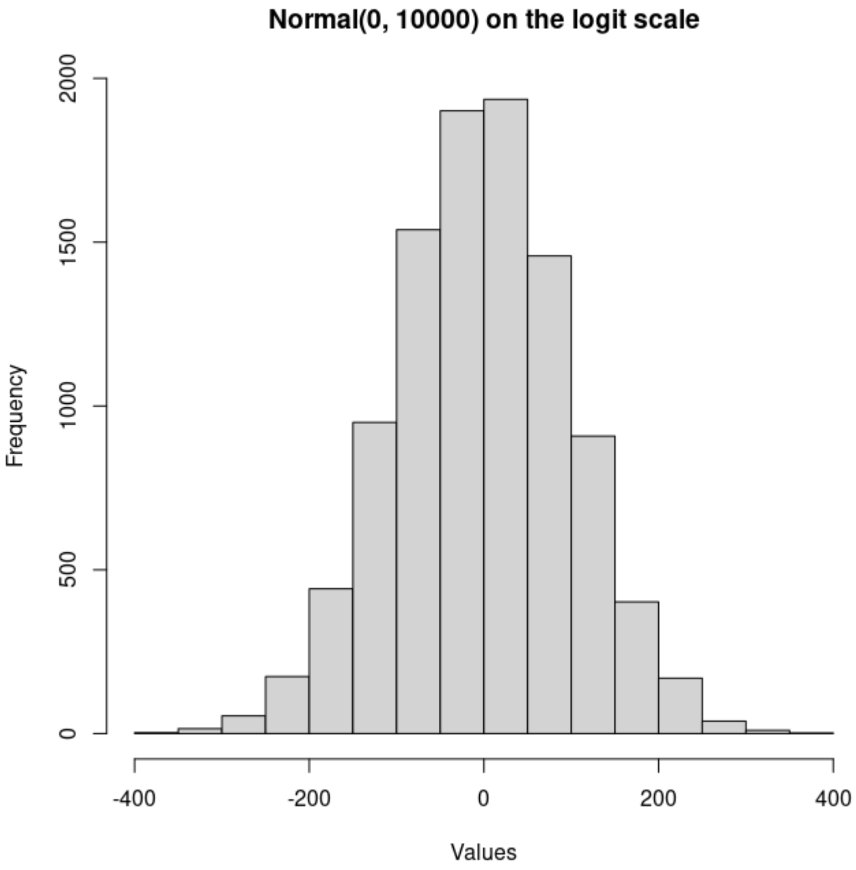
NOTE: need to be very careful with prior distributions for regression coefficients when using link functions

Priors in Bayesian binomial GLMs

- Recall a common prior distribution for regression coefficients is a normal distribution with a mean of 0 centered with a large variance

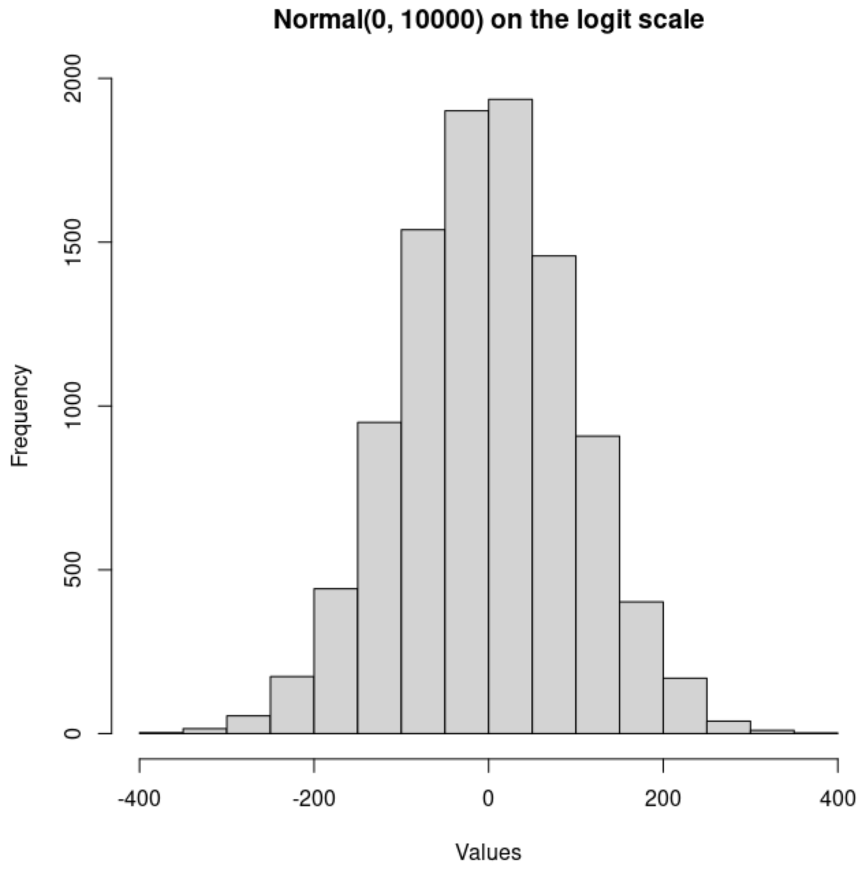
Priors in Bayesian binomial GLMs

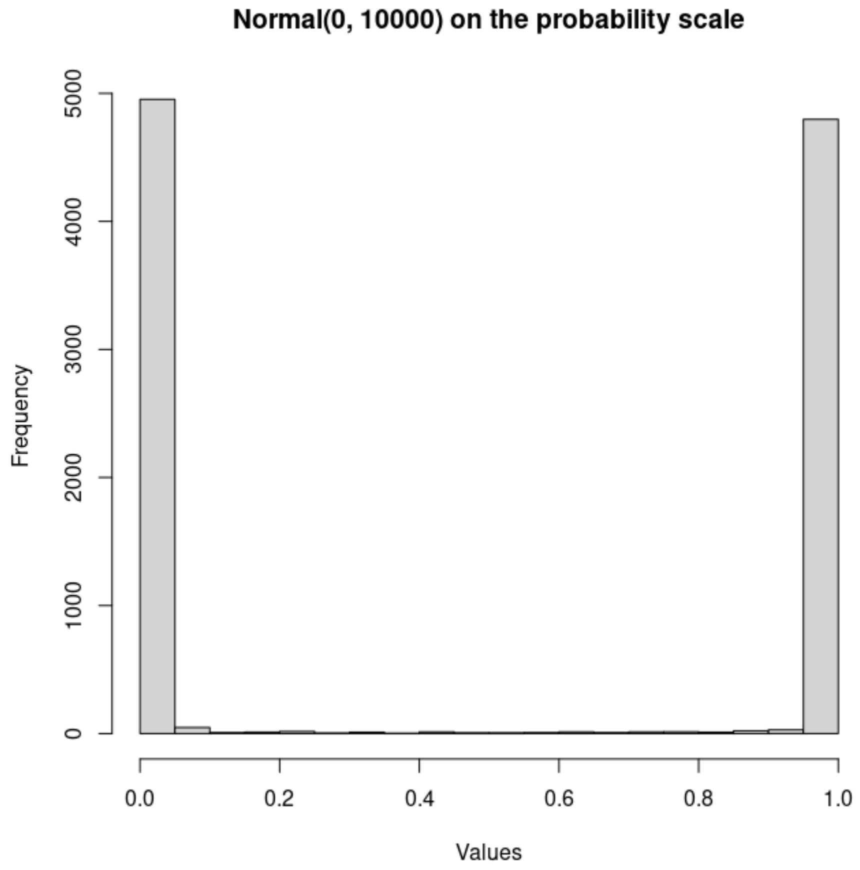
- Recall a common prior distribution for regression coefficients is a normal distribution with a mean of 0 centered with a large variance



Priors in Bayesian binomial GLMs

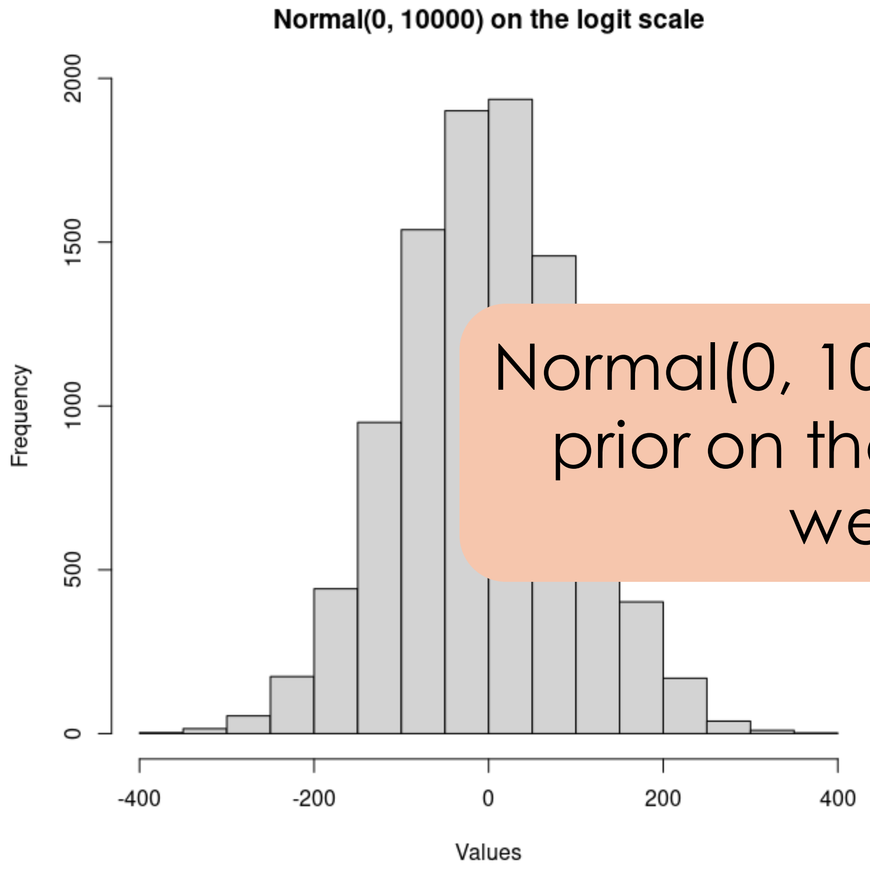
- Recall a common prior distribution for regression coefficients is a normal distribution with a mean of 0 centered with a large variance

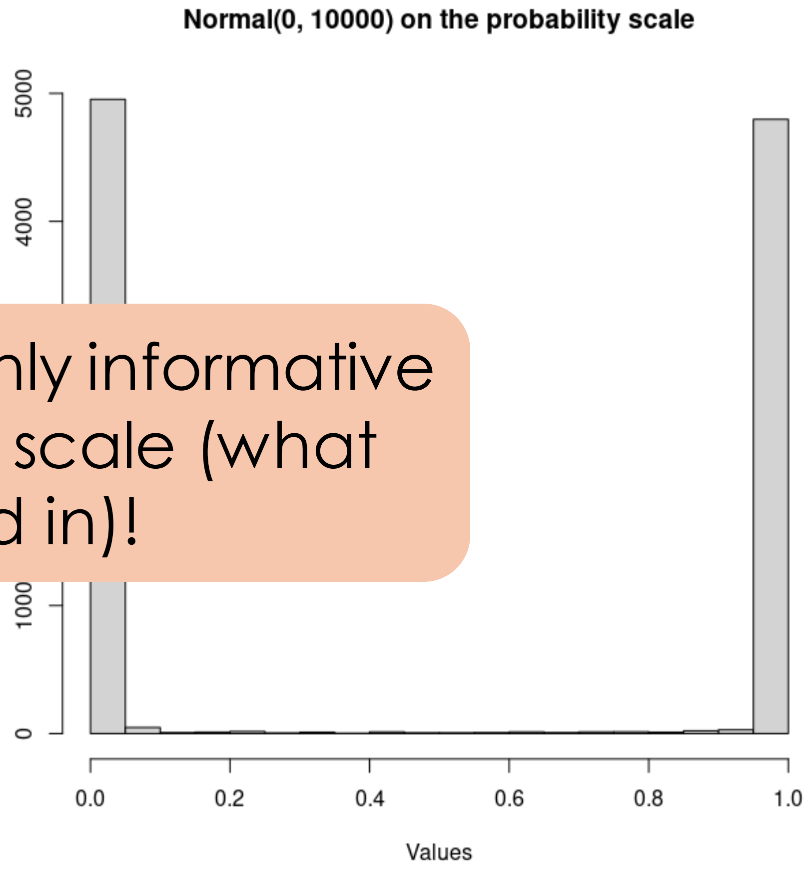




Priors in Bayesian binomial GLMs

- Recall a common prior distribution for regression coefficients is a normal distribution with a mean of 0 centered with a large variance

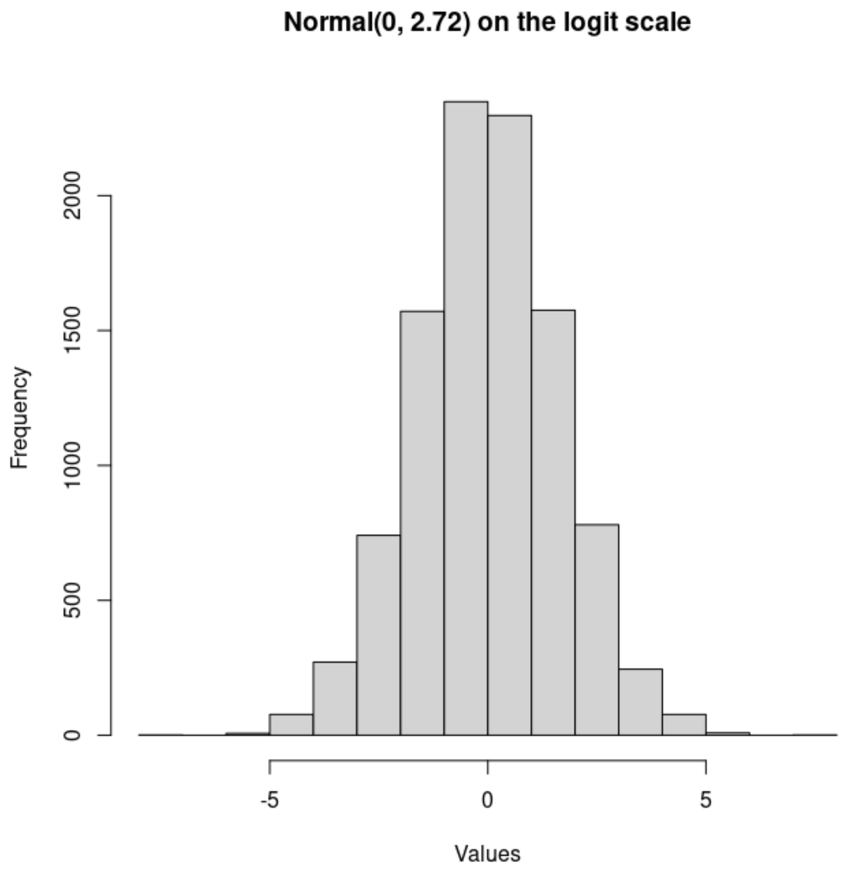


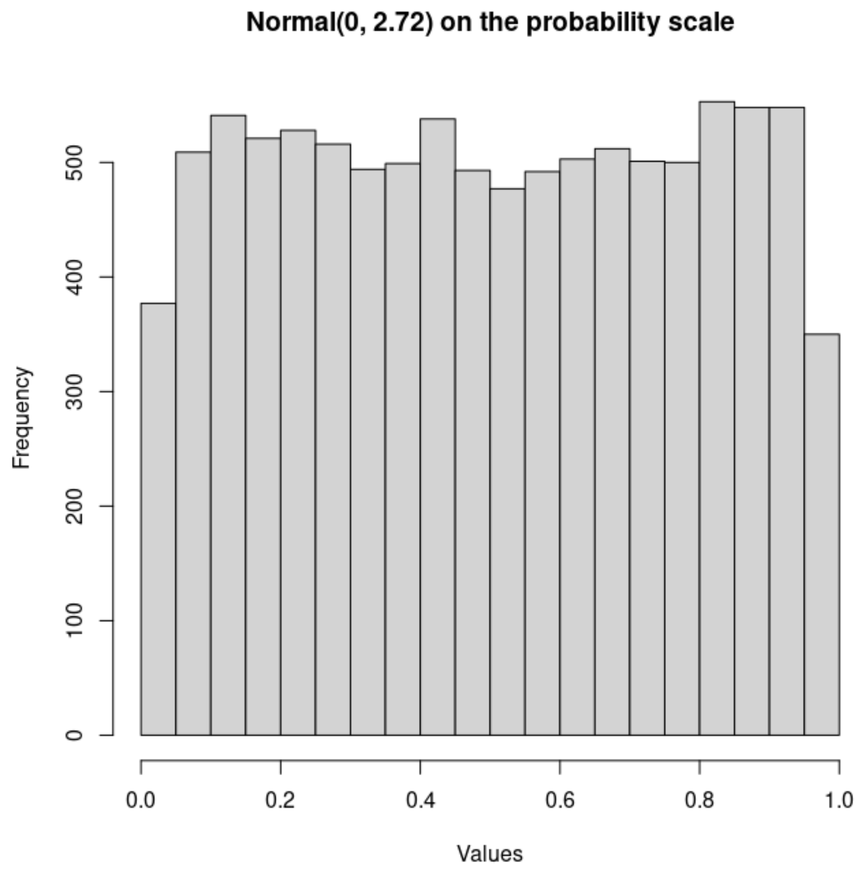


Normal(0, 10000) is a highly informative prior on the probability scale (what we're interested in)!

Priors in Bayesian binomial GLMs

- Instead, a good prior is Normal(0, 2.72)





`brms` can fit a TON of GLMs

`brms` can fit a TON of GLMs

- A historical benefit to using Bayesian programming languages like `NIMBLE`, `Stan`, `JAGS` is their flexibility to fit different models

`brms` can fit a TON of GLMs

- A historical benefit to using Bayesian programming languages like `NIMBLE`, `Stan`, `JAGS` is their flexibility to fit different models
- `brms` now provides a lot of that flexibility, with an easy-to-use formula-based interface

`brms` can fit a TON of GLMs

- A historical benefit to using Bayesian programming languages like `NIMBLE`, `Stan`, `JAGS` is their flexibility to fit different models
- `brms` now provides a lot of that flexibility, with an easy-to-use formula-based interface
- Lots of zero-inflated models, hurdle models, exponential, Weibull, negative binomial, geometric, etc.

`brms` can fit a TON of GLMs

- A historical benefit to using Bayesian programming languages like `NIMBLE`, `Stan`, `JAGS` is their flexibility to fit different models
- `brms` now provides a lot of that flexibility, with an easy-to-use formula-based interface
- Lots of zero-inflated models, hurdle models, exponential, Weibull, negative binomial, geometric, etc.
- See here for more details on possibilities

Example: Bayesian Binomial GLM

Distribution of eastern hemlock across Vermont

`4c-binomial-glm-brms.R`