

# Introduction to Generalised Linear Models for Ecologists

Dr Niamh Mimmagh

[niamh@prstats.org](mailto:niamh@prstats.org)

<https://github.com/niamhmimmagh/glmm01>

# Poisson Regression

- Poisson regression assumes the variance equals the mean ( $Var(Y) = \lambda$ ). However, real-world count data often show more variability than this model expects. This extra variability is called overdispersion.
- When overdispersion is present, standard errors may be underestimated, leading to overly optimistic p-values and confidence intervals.
- Similarly, binomial regression assumes the variance is  $n\pi(1-\pi)$  for each observation, where  $n$  is the number of trials and  $\pi$  is the success probability.
- But in practice, binary or proportion data can also exhibit extra-binomial variation, meaning the observed variance is larger than  $n\pi(1-\pi)$ .

# Overdispersion

- Overdispersion occurs when the observed variance exceeds the theoretical variance expected under the model.
- This violates model assumptions and leads to underestimated standard errors and misleading p-values.
- Overdispersion can be caused by:
  - Unobserved heterogeneity between observations,
  - Missing important predictors,
  - Clustering or correlation among data points,
  - Excess zeros in the data.

# What if We Ignore Overdispersion?

- If we ignore overdispersion, we will underestimate the parameter variance.
- Standard errors come directly from variance
- The standard error (SE) of a coefficient estimate  $\hat{\beta}_j$  is:  $SE(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)}$
- A typical 95% confidence interval for  $\hat{\beta}_j$  is:  $\hat{\beta} \pm 1.96SE(\hat{\beta}_j)$
- If  $SE(\hat{\beta}_j)$  is too small, then the CI becomes narrower than it should be. This makes the CI look more “certain” than reality justifies.

# What if We Ignore Overdispersion?

- If  $SE(\hat{\beta}_j)$  is underestimated, the denominator is too small, so the z-score becomes too large. Large z-scores leads to tiny p-values, which mean you wrongly conclude “significant effect” more often. This type of result is called a false positive or a Type I error.
- This will lead to model that is too optimistic – the model will be overconfident, suggesting it knows the parameter more precisely than it really does. We are more likely to declare an effect significant, even when its just noise, and the reported confidence intervals won’t actually contain the true values as often as they claim.
- So underestimating variance means we systematically underestimate uncertainty, leading to too-small SEs, too-narrow CIs, and too-small p-values.



# Diagnosing Overdispersion in Poisson Models

- To diagnose overdispersion in a Poisson model, fit the model and compute the dispersion statistic:

$$\Phi = \frac{\textit{Residual Deviance}}{\textit{Residual degrees of freedom}}$$

- If  $\Phi \approx 1$ , the model fits the assumed mean–variance relationship well.
- If the  $\Phi \gg 1$ , there is overdispersion present (too much variability for the assumed distribution).
- If the  $\Phi \ll 1$ , there is underdispersion present (too little variability, which is rare).

# Diagnosing Overdispersion in Binomial Models

- For binomial models, you can also use the residual deviance divided by the residual degrees of freedom or the Pearson chi-squared statistic divided by the degrees of freedom.

$$\Phi_{dev} = \frac{\textit{Residual Deviance}}{\textit{Residual degrees of freedom}}$$

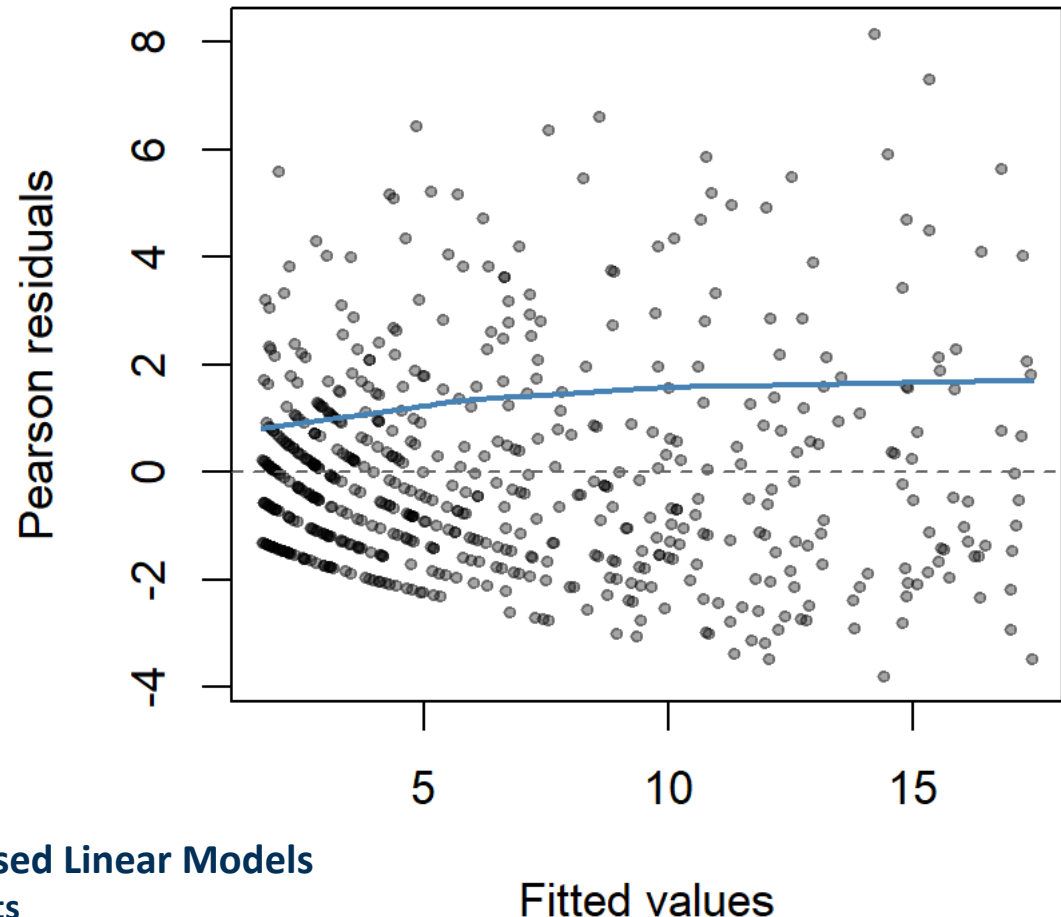
- Under a well-specified model with no overdispersion,  $\Phi_{dev} \approx 1$ .
- If  $\Phi_{dev} \gg 1$ , there is overdispersion and  $\Phi_{dev} \ll 1$  indicates underdispersion.

$$\Phi_{\chi} = \frac{\textit{Pearson Chi-squared statistic}}{\textit{Residual degrees of freedom}}$$

- Again,  $\Phi_{\chi} \approx 1$  indicate a good fit. Larger values suggest overdispersion.

# Visualising Overdispersion

- Residual plots can be useful for visualising overdispersion.
- Under a mis-specified model, residuals may show a fan shape, with increasing spread for larger fitted values. This indicates the variance is underestimated.





# Quasi-Models

- Quasi models keep the same link function and mean structure (so fitted means don't change). They estimate a dispersion parameter  $\Phi$  from the residual deviance:

$$\hat{\Phi} = \frac{\textit{Residual Deviance}}{\textit{Residual degrees of freedom}}$$

- This quantifies how much larger the variance is compared to the nominal distribution, and is used to rescale the variance. This affects standard errors, confidence intervals, and p-values.
- For quasi-Poisson:

$$\textit{Var}(Y) = \Phi \lambda$$

- For quasi-binomial:

$$\textit{Var}(Y) = \Phi n \pi (1 - \pi)$$

# Fitting Quasi-Models

- In R, you can fit a quasi-Poisson or quasi-binomial model using:

```
glm(y ~ x, family = quasipoisson)
glm(y ~ x, family = quasibinomial)
```

- The coefficients will be the same as the Poisson/binomial model but the standard errors will be adjusted for overdispersion.

# How Quasi-Model Adjust Inference

- Quasi models do not change the estimated means but they do inflate the standard errors.

$$SE(\hat{\beta})_{quasi} = \sqrt{\Phi} SE(\hat{\beta})_{Poisson/Binomial}$$

- This leads to wider confidence intervals and more reliable p-values.
- Why multiply by the square root of the overdispersion parameter?
- Because it's the variance we're inflating by  $\Phi$ , and the standard error is the square root of the variance.

# Coding Demo

# Beyond Quasi-Likelihood

- Quasi models are a quick fix for overdispersion. They improve uncertainty estimates without changing the predicted means. They are simple but limited when overdispersion is strong.
- While quasi-likelihood models adjust standard errors, they do not explain the source of extra variability.
- Models intended specifically for dealing with overdispersion introduce additional parameters that explicitly model the variance.

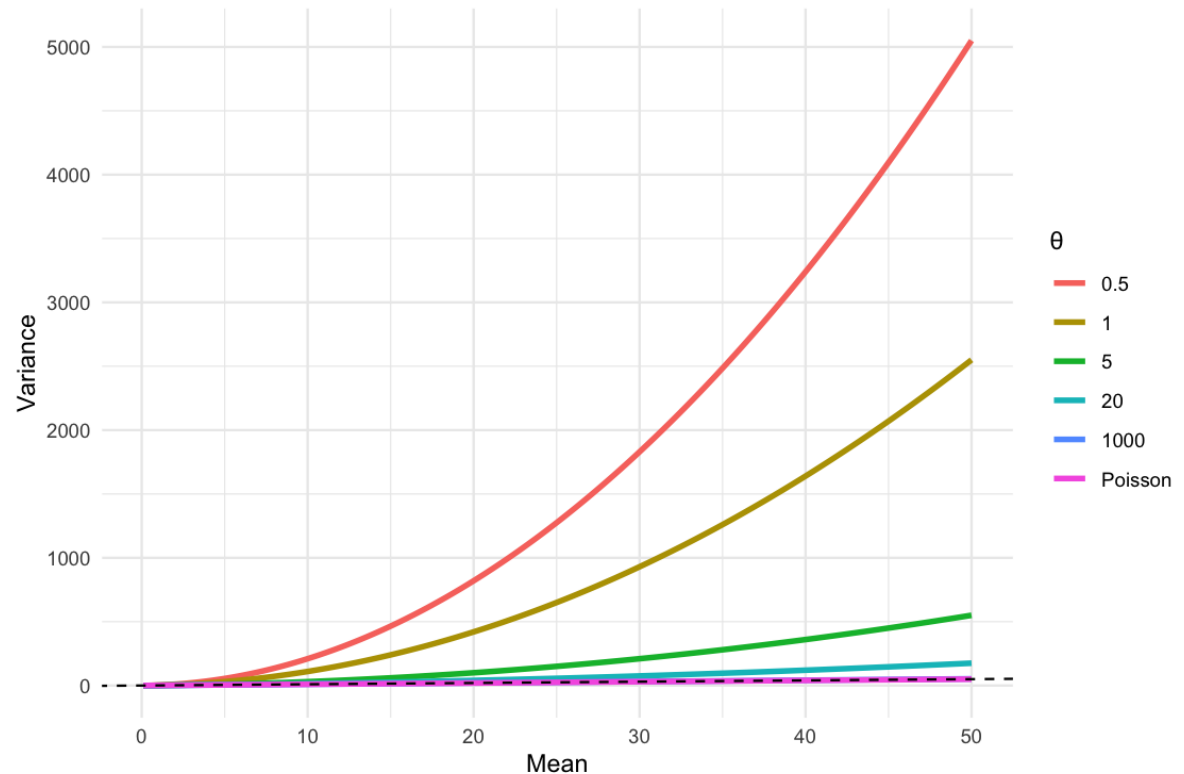
# Negative Binomial Distribution

$$Y_i \sim \text{Negative Binomial}(\lambda_i, \theta)$$
$$\log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

- The negative binomial distribution has mean  $\lambda_i$  and variance  $\lambda_i + \frac{\lambda_i^2}{\theta}$
- This additional term allows the variance to grow faster than the mean – it is no longer assumed that the variance is equal to the mean (as it is in the Poisson distribution).
- The dispersion parameter  $\theta$  controls the amount of overdispersion, with smaller  $\theta$  indicating greater dispersion.

# Negative Binomial Distribution

- The Poisson model is a special case of the Negative Binomial where  $\theta$  approaches infinity.



# Interpreting Negative Binomial Coefficients

- Because we are specifying the link in the same way (using the log link) coefficients in a Negative Binomial model have the same interpretation as in a Poisson model.
- Exponentiating the coefficient gives the rate ratio, or the multiplicative effect on the expected count.
- The  $\theta$  parameter is estimated to capture overdispersion.
- You can compare Poisson and Negative Binomial models using AIC.
- If overdispersion is present, the Negative Binomial model will typically have a lower AIC value – indicating a better fit to the overdispersed data.



# Example

- We have 240 plot-surveys from 80 meadow patches (3 surveys per patch), with equal effort per survey. During the survey, the number of bee/butterfly visits to a plot were recorded for 30 minutes. Predictors are:
  - FlowerDensity (flowers/m<sup>2</sup>),
  - PatchSize (ha),
  - DistHedge (m to nearest hedgerow),
- We expect that counts are overdispersed (variance > mean) due to unmeasured patch quality, temporal bursts, and site-level heterogeneity. We will start by implementing a Poisson model, and then comparing with a negative binomial model.

# Coding Demo

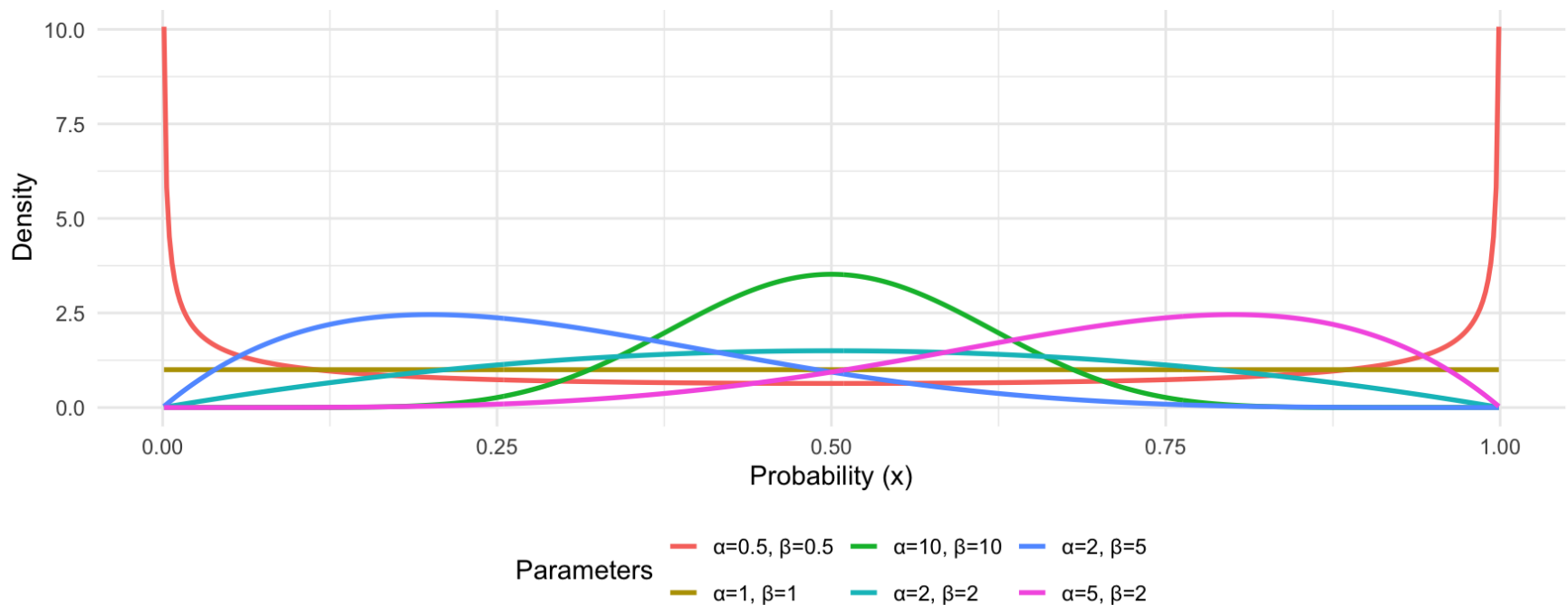
# Beta Distribution

- For proportion data with overdispersion, the beta-binomial model assumes the binomial probability parameter follows a beta distribution, introducing extra variance.
- The Beta distribution is a continuous distribution defined on (0,1), ideal for modeling probabilities. It is defined by two shape parameters:  $\alpha$  and  $\beta$ . It is very flexible: it can be uniform, U-shaped, or peaked.

$$Mean = \frac{\alpha}{\alpha + \beta}$$

$$Variance = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

# Beta Distribution



# Beta Binomial Model

- The Beta-Binomial model assumes the true group probability:

$$\pi_i \sim \text{Beta}(\alpha, \beta)$$
$$Y_i \sim \text{Binomial}(n_i, \pi_i).$$

- The mean of the beta binomial is the same as the binomial:

$$E[Y] = n\mu.$$

- The variance is larger:

$$\text{Var}[Y] = n\mu(1 - \mu) \frac{1 + (n - 1)}{\alpha + \beta + 1}$$

- Extra term accounts for variability in  $\pi$  across groups. This captures overdispersion that the Binomial distribution cannot.

# When to Use the Beta Binomial

- You may encounter situations when using proportional data that would suggest you should use a Beta binomial model rather than a binomial model. These include:
  - Proportion or binary data grouped into clusters.
  - Residual deviance  $\gg$  degrees of freedom under Binomial distribution.
  - Group-specific probabilities vary (unobserved heterogeneity).

# Beta Binomial vs Quasi-Binomial

- A quasi-binomial model rescales variance using the additional overdispersion parameter, but it doesn't explain the source of the variance.
- Whereas the beta-binomial is generative: it models heterogeneity in  $\pi$  explicitly.
- The beta-binomial model also provides interpretable parameters ( $\alpha$ ,  $\beta$ , overdispersion,  $\pi$ ).

# Fitting a Beta Binomial in R

- Use the VGAM package in R:

```
library(VGAM)
model <- vglm(cbind(success, failure) ~ x,
beta.binomialff, data = df)
```



# Interpreting Beta Binomial Coefficients

- Coefficients are interpreted like logistic regression log-odds, while the extra dispersion parameter accounts for additional variability in the proportion.
- $\beta_0$ : the log-odds of success at the reference levels of all predictors.
- $\beta_1$ : the change in log-odds for a 1-unit increase in  $x$ .
- On the odds scale, each unit of  $x$  multiplies the odds by  $e^{\beta_1}$  (the odds ratio)
- The dispersion parameter does not change these interpretations; it explains extra variability around  $\mu_i$  and typically widens the confidence intervals.



# Example

- We have 150 plots across 30 meadow sites (5 plots/site). In each site we sow 50 seeds of a native forb and revisit after 6 weeks.
- The response is the number of seeds that germinated in plot  $i$  (0,...,50)  
Predictors include:
  - CanopyCover (%) ,
  - SoilMoisture (%) ,
  - Management (Mown vs Unmown).
- Expectation: true germination probability  $\pi_i$  varies among plots AND sites (unmeasured microhabitat), producing extra-binomial variation (overdispersion).

# Coding Demo

# Zero-Inflated Models vs Overdispersion

- Sometimes overdispersion is driven by more zeros than the Poisson/NB can generate. Zero-inflated Poisson/NegBin add a separate zero process (a Bernoulli “always zero” state) alongside the count model, so you estimate both the zero-inflation probability and the usual mean.
- Quasi models (mild overdispersion): Quasi-Poisson / quasi-binomial inflate the variance with a single dispersion factor but keep the same mean model. They’re quick and robust, but they’re quasi-likelihood (no AIC/LRT), so use mainly when extra variance is modest.
- Negative binomial / beta-binomial (explicit variance): These are full likelihood models that add a dispersion parameter and specify how variance grows with the mean. They’re generative, support AIC/LRT, and are a good default when overdispersion is more than mild.

# Comparing Models

- Different models may all fit overdispersed data. They encode different mechanisms (extra variance, structural zeros, heterogeneity), so they can imply different conclusions and predictions.
- Comparing models helps us choose the one that balances fit, complexity, and predictive accuracy.
- Aim for the simplest plausible model that captures the signal; use ecological reasoning to judge whether the model's assumptions make sense.
- We will use tools like AIC, BIC, and cross-validation to compare models. Residual diagnostics and visual checks are also important to understand fit



# Information Criteria for Model Comparison

- AIC (Akaike Information Criterion) balances model fit and complexity. A lower AIC indicates a better trade-off between fit and simplicity.

$$AIC = -2 \log L + 2k$$

where  $L$  is the likelihood and  $k$  is the number of parameters

- BIC (Bayesian Information Criterion) penalises complexity more heavily than AIC. BIC tends to favour simpler models, especially for larger datasets.

$$BIC = -2 \log L + k \log(n)$$

where  $n$  is the sample size

# Quasi Models Cannot Use AIC

- In a standard GLM (e.g. Poisson, Binomial, Negative Binomial), the likelihood is the probability of the observed data given the model parameters.

$$L(\theta) = P(Y | \theta) L(\theta) = P(Y | \theta)$$

- This provides a full probability model.
- Quasi-models do not assume a full probability distribution. Instead, they only specify the mean–variance relationship (e.g.  $Var(Y) = \phi\lambda$  for quasi-Poisson). Because they are not based on a full distribution, they lack a “true” likelihood function.
- Consequence: You cannot compute AIC or BIC (which depend on likelihood) for quasi-models. Quasi models are useful for valid standard errors under overdispersion, but model comparison must use other methods.

# Model Diagnostics

## Residual plots

- Residuals vs fitted values: check for curvature, non-linearity, or signs of over/under-dispersion (fan shapes). These are useful for spotting lack of fit in the mean–variance relationship, but: Poisson and Negative Binomial often give similar residual plots (same mean structure).

## QQ plots

- These are used to compare the distribution of residuals to theoretical expectation. They can identify skewness, heavy tails, or other deviations.



# Model Diagnostics

## **Influence and leverage plots**

- Leverage plots: show points with unusual predictor values (high leverage).
- Cook's distance (or influence plots): highlight observations that are both high leverage and have large residuals (i.e. influential points).

## **Numerical checks for dispersion**

- Dispersion parameters much greater than 1 indicate overdispersion.

# Coefficient Precision Across Models

- Overdispersion inflates uncertainty. Models that account for it typically produce wider confidence intervals than a Poisson/Binomial fit.
- Quasi-Poisson / Quasi-binomial: point estimates ( $\beta$ ) are the same as the Poisson/Binomial with the same link; standard errors are inflated by  $\sqrt{\phi}$ , so CIs widen. (Quasi-likelihood  $\Rightarrow$  no AIC/LRT.)
- Negative Binomial / Beta-binomial: change the likelihood and variance model; both estimates and standard errors can shift. Report the dispersion.
- Diagnostic cue: if CIs widen but estimates are stable, the main problem was underestimated variance; if estimates also move, the mean–variance structure (or excess zeros) was misspecified.

