Count Response

STAT 526
Professor Olga Vitek

March 31, 2011
Basic Poisson Regression
(Faraway Ch. 3)
Poisson Distribution

• \( Y \): Counts of events that occur randomly over fixed interval of time or space
  
  – No upper limit for the number of events
  
  – Approximates \( Binomial(n, \pi) \) distribution when \( n \) is large and \( \pi \) is small. Then \( \lambda = n\pi \).
  
  – Approaches Normality as \( \lambda \) increases

• \( P\{Y = y\} = \frac{e^{-\lambda}\lambda^y}{y!}, \ y = 1, \ldots \)

• \( E\{Y\} = Var\{Y\} = \lambda \)

• If \( Y_i \sim Poisson(\lambda_i), \ i = 1, \ldots, \)
  
  then \( \sum_i Y_i \sim Poisson(\sum \lambda_i) \)
Poisson($\lambda$) for various $\lambda$

- $\lambda = 0.5$
- $\lambda = 1$
- $\lambda = 3$
- $\lambda = 6$
- $\lambda = 9$
- $\lambda = 12$
Poisson Regression

• Model

\[ Y_i \overset{ind}{\sim} \text{Poisson}(\lambda_i), \text{ where } \]
\[ \log \lambda_i = X_i' \beta \quad (= \eta, \text{ link}) \]

\[ - \beta_j = \text{difference in } \log \text{E}\{Y\} \text{ following a unit change in predictor } X_j, \text{ while the others are constant} \]

• Log-likelihood

\[ l(\beta) = \log \prod_{i=1}^{n} \left[ \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \right] = \sum_{i=1}^{n} \left( y_i \log \lambda_i - \lambda_i - \log y_i! \right) \]

\[ = \sum_{i=1}^{n} \left( y_i X_i' \beta - \exp X_i' \beta \right) + \text{constant} \]

• Pearson \( X^2 \)

\[ X^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i} \]

• Deviance

\[ D = 2 \sum_{i=1}^{n} \left( y_i \log \frac{y_i}{\hat{\lambda}_i} - (y_i - \hat{\lambda}_i) \right) \]

- Different numeric values of deviance for grouped and ungrouped data
Inference About Individual \( \beta_j \): Wald Test

- Same as for Binomial response

- Test \( H_0 : \beta_j = 0 \) versus \( H_a : \beta_j \neq 0 \).

- Test statistic \( z^* = \frac{b_j - 0}{s\{b_j\}} \)

- **Approximate** variance \( s^2\{b\} \)

\[
s^2\{b\} = \left( \left[ - \frac{\partial^2 \log_e L(\beta)}{\partial \beta_j \partial \beta_j'} \right]_{\beta=b} \right)^{-1}
\]

- **Approximate** distribution of \( z \)
  
  - \( z^* \sim N(0, 1) \). Alternatively, \( (z^*)^2 \sim \chi^2_1 \)
  
  - reject \( H_0 \) if \( |z^*| > z^{1-\alpha/2} \)
  
  - CI for \( \beta_j \): \( b_j \pm z^{1-\alpha/2} s\{b_j\} \)
Example: Galapagos Plants
(Faraway Ch. 3)

- Study the number of plant species on 30 Galapagos islands as function of geographic variables

```r
library(faraway)
data(gala)
?gala

# faraway does not use the 2nd predictor,
# so we’ll remove it too.
gala <- gala[, -2]
head(gala)
```

<table>
<thead>
<tr>
<th>Species</th>
<th>Species</th>
<th>Area</th>
<th>Elevation</th>
<th>Nearest</th>
<th>Scruz</th>
<th>Adjacent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltra</td>
<td>58</td>
<td>25.09</td>
<td>346</td>
<td>0.6</td>
<td>0.6</td>
<td>1.84</td>
</tr>
<tr>
<td>Bartolome</td>
<td>31</td>
<td>1.24</td>
<td>109</td>
<td>0.6</td>
<td>26.3</td>
<td>572.33</td>
</tr>
<tr>
<td>Caldwell</td>
<td>3</td>
<td>0.21</td>
<td>114</td>
<td>2.8</td>
<td>58.7</td>
<td>0.78</td>
</tr>
<tr>
<td>Champion</td>
<td>25</td>
<td>0.10</td>
<td>46</td>
<td>1.9</td>
<td>47.4</td>
<td>0.18</td>
</tr>
<tr>
<td>Coamano</td>
<td>2</td>
<td>0.05</td>
<td>77</td>
<td>1.9</td>
<td>1.9</td>
<td>903.82</td>
</tr>
<tr>
<td>Daphne.Major</td>
<td>18</td>
<td>0.34</td>
<td>119</td>
<td>8.0</td>
<td>8.0</td>
<td>1.84</td>
</tr>
</tbody>
</table>
> fit1 <- glm(Species~., family=poisson, data=gala)
> summary(fit1)

Call:
glm(formula = Species ~ ., family = poisson, data = gala)

Deviance Residuals:
     Min       1Q   Median       3Q      Max
-8.2752  -4.4966  -0.9443   1.9168  10.1849

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.155e+00  5.175e-02  60.963  < 2e-16 ***
  Area     -5.799e-04  2.627e-05  -22.074  < 2e-16 ***
Elevation  3.541e-03  8.741e-05   40.507  < 2e-16 ***
Nearest    8.826e-03  1.821e-03    4.846  1.26e-06 ***
Scruz      -5.709e-03  6.256e-04   -9.126  < 2e-16 ***
Adjacent  -6.630e-04  2.933e-05  -22.608  < 2e-16 ***

---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 3510.73  on 29 degrees of freedom
Residual deviance:  716.85  on 24 degrees of freedom
AIC: 889.68
Model Diagnostics
(Faraway Sec. 6.4)
Goodness of Fit Tests

- Residual Deviance
  - Test $H_0$: our model vs $H_a$: saturated model
  - Distribution poorly approximated by $\chi^2$

  ```r
  > pchisq(fit1$deviance,
        fit1$df.residual,lower.tail=FALSE)
  [1] 7.073157e-136  # reject H0
  ```

- Pearson $X^2$
  - Test $H_0$: our model vs $H_a$: saturated model
  - Distribution better approximated by $\chi^2$

  ```r
  > pchisq( sum(residuals(fit1, type="pearson")^2 ),
            fit1$df.residual, lower.tail=FALSE)
  [1] 2.18719e-145  # reject H0
  ```

- Better approximation of $\chi^2$ when Poisson approaches Normal
- Can look more into model diagnostics to find possible reasons for poor fit
Model Diagnostics: Residual Plots

- Response residuals $r_i = y_i - \hat{\lambda}_i$
  - Of limited use: variance is function of the expected value

- Pearson residuals $r_{i,P} = \frac{y_i - \hat{\lambda}_i}{\sqrt{\hat{\lambda}_i}}$
  - $\sum_i r_{i,P}^2 = X^2$
  - Account for differences in variance

- Deviance residuals
  
  $r_{i,D} = \text{sign}(y_i - \hat{\lambda}_i) \left[ 2(y_i \log \frac{y_i}{\hat{\lambda}_i} - (y_i - \hat{\lambda}_i)) \right]^{1/2}$
  
  - Defined by analogy with Pearson residuals
  - Define $d_i = 2(y_i \log \frac{y_i}{\hat{\lambda}_i} - (y_i - \hat{\lambda}_i))$, such that
  
  $\sum_i r_{i,D}^2 = \sum_i d_i = \text{Deviance}$
  
  - Account for differences in variance
Residual Plots (Next Slide)

(a) As in linear regression, can plot response residuals $r_i$ vs $E\{Y_i\} = \hat{\lambda}_i$

\[
\text{plot(residuals(fit1, type="response") ~ predict(fit1, type="response"),}
\]
\[
\quad \text{xlab=expression(hat(lambda)),}
\]
\[
\quad \text{ylab="Response residuals", main="(a)"}
\]
\[
\text{abline(h=0)}
\]

– Of limited use:
  Variance is function of the expected value;
  Most predicted expected values are small

(b) Plot response residuals $r_i$ vs link $\hat{\eta} = X\hat{\beta}$

\[
\text{plot(residuals(fit1, type="response") ~ predict(fit1, type="link"),}
\]
\[
\quad \text{xlab=expression(paste(hat(eta), " = X", hat(beta))),}
\]
\[
\quad \text{ylab="Response residuals", main="(b)"}
\]
\[
\text{abline(h=0)}
\]

– The points on x axis are better separated
  Variance still function of the expected value
Residual Plots (Next Slide)

(c) Plot deviance residuals $r_{i,D}$ vs link $\hat{\eta} = X\hat{\beta}$

```r
plot(residuals(fit1) ~ predict(fit1, type="link"), 
     xlab=expression(paste(hat(eta), " = X", hat(betta))), 
     ylab="Deviance residuals", main="(c)")
abline(h=0)
```

(d) Plot Pearson residuals $r_{i,P}$ vs link $\hat{\eta} = X\hat{\beta}$

```r
plot(residuals(fit1, type="pearson") ~ predict(fit1, type="link"), 
     xlab=expression(paste(hat(eta), " = X", hat(betta))), 
     ylab="Pearson residuals", main="(d)")
abline(h=0)
```

• Interpretation of (c) and (d):
  – Y axis accounts for differences in variance
  – X axis better separates the predicted values
  – Systematic nonlinear relationship → lack of fit
    * No apparent problems for this dataset
  – Large/nonconstant variance → violations of the assumed Poisson distribution
    * Potential overdispersion
Residual Plots

- No obvious problematic trends in (c) or (d)
Model Diagnostics: Functional Form of Predictors

- As in linear regression, can plot $Y$ against each predictor $X_j$

  scatter.smooth(gala$Area, gala$Species)
  scatter.smooth(gala$Elevation, gala$Species)
  scatter.smooth(gala$Nearest, gala$Species)
  scatter.smooth(gala$Scruz, gala$Species)

- The plots are of limited use, and may not show clear trends:
  - Lower response conts are more frequent
  - $\text{Var}\{Y\} = E\{Y\} \rightarrow$ large $Y$ are unreliable

- Transformations can be helpful
  - Look for transformations for predictors even if there is no obvious problems in residual plots
Plotting $Y$ vs $X_j$

- Most predictors indicate some non-linearity
Consider $log(Area)$

(a) $Area$ vs $Species$
plot(gala$Area, gala$Species, main="(a)"

(b) Use $log(Area)$
scatter.smooth(log(gala$Area), gala$Species, main="(b)"

(c) Account for the log link:
$log(Area)$ vs $Species$
scatter.smooth(log(gala$Area), log(gala$Species), main="(c)"

(d) Faraway suggests plotting linearized response:
$Z_i = X_i\hat{\beta} + (Y_i - \hat{\lambda})/\hat{\lambda}$

lambda<- predict(fit1, type="response"
Z <- predict(fit1) + (gala$Species-lambda)/lambda
scatter.smooth(log(gala$Area), z, ylab="Linearlized response", main="(d)"

• $log$ and the linearized response are similar
Consider $\log(Area)$

- Log transformation appears appropriate
Residuals of Original and Log-Transformed Predictors

- Apply log transform to all predictors

\[
\text{fit2} \leftarrow \text{glm(Species} \sim \text{log(Area)+log(Elevation)+log(Nearest)+log(Scruz+0.1)+log(Adjacent),}
\]
\[
\text{family=poisson, data=gala)}
\]

\[
\text{anova(fit1, fit2)}
\]

Analysis of Deviance Table

Model 1: Species \sim Area + Elevation + Nearest + Scruz + Adjacent
Model 2: Species \sim log(Area) + log(Elevation) + log(Nearest) + log(Scruz + 0.1) + log(Adjacent)

<table>
<thead>
<tr>
<th>Resid. Df</th>
<th>Resid. Dev</th>
<th>Df</th>
<th>Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>716.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>359.12</td>
<td>0</td>
<td>357.72</td>
</tr>
</tbody>
</table>

- obtain a substantial reduction in deviance
- the model with all the transformed predictors yields a better overall fit
- Next: visualize the change in plots of response and deviance residuals
Residuals of Original and Log-Transformed Predictors

- More compact residuals after the transform
Model Diagnostics: Partial Residuals and Link

(a) Partial contribution of a predictor:
Plot \( Z_i - X_i'\hat{\beta} + X_{ij}\hat{\beta}_j \) vs \( X_j \)

```r
lambda <- predict(fit2, type="response")
u <- (gala$Species-lambda)/lambda +
    coef(fit2)[2]*log(gala$Area)
scatter.smooth(log(gala$Area), u,
    ylab="Partial residual", main="Partial residuals")
```

(b) Choice of link function:
Plot \( Z_i \) vs \( X_i'\hat{\beta} \)

```r
z <- predict(fit2) + (gala$Species-lambda)/lambda
scatter.smooth(predict(fit2), z,
    ylab="Linearlized response",
    main="Diagnostics for link")
```

- No major problems in this dataset
  - Points roughly forming straight lines indicate no gross departures from the assumptions
  - Possible two outliers
Model Diagnostics: Partial Residuals and Link

- No major problems in this dataset
  - Points roughly forming straight lines indicate no gross departures from the assumptions
  - Two potential outliers
Model Diagnostics: Outliers

• Same concepts as in linear and logistic regression

• Automated plots in R (next slide)

  for (i in 1:4)
    plot(fit2, which=i)

• Manual plots in R (following slide)

  halfnorm( rstudent(fit2),
            main="Half-normal qqplot of Jacknife residuals")

  halfnorm( influence(fit2)$hat,
            main="Half-normal qqplot of leverages")

  halfnorm( cooks.distance(fit2),
            main="Half-normal qqplot of Cook’s distance")

  plot( influence(fit2)$coef[,5], xlab="Obs.No",
       ylab="dBeta",
       main="Change in Scruz coef after deletion")
Diagnostics: Outliers

- 'Residuals' = deviance residuals;
- 'Predicted values' = \( X_i \hat{\beta} \);
- 'Std. deviance resid' = \( r_{SD} = \frac{r_D}{\sqrt{\hat{\phi}(1-h_{ii})}} \) (p. 124);
- 'Cook's distance' (p. 125).
• #25 = 'Scruz'. The influence can be due to adding 0.1 at the log transform.
Overdispersion
(Faraway Ch. 3)
Overdispersion

• Suppose $Y \sim Poisson(\lambda)$
  – as for Binomial response, overdispersion arises
    in case of hierarchical or clustered observations

• Suppose $\lambda$ itself is a r.v. with $E\{\lambda\} = \theta$.
  – $E\{Y\} = E\{E\{Y|\lambda\}\} = \theta$
  – $Var\{Y\} = E\{Var\{Y|\lambda\}\} + Var\{E\{Y|\lambda\}\}$
    $= \theta + Var\{\lambda\} > \theta$

• Quasipoisson distribution
  – $E\{Y\} = \lambda$
  – $Var\{Y\} = \phi \lambda$
  – Can estimate $\hat{\phi}$: $\hat{\phi} = \frac{1}{n-p} X^2 = \frac{1}{n-p} \sum_{i=1}^{n} \frac{(y_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i}$

• Comparing nested models in presence of
  overdispersion:
  – Test $H_0$ : reduced model vs $H_a$ : full model
  – $F = \frac{(D_0 - D_1)/(df_0 - df_1)}{\hat{\phi}_1} \overset{approx.}{\sim} F(df_0 - df_1, df_1)$
Example: Galapagos Plants

- Fit a simplified model with 2 predictors

```r
> fit3 <- glm(Species~ log(Area)+log(Adjacent),
   family=poisson, data=gala)
```

Coefficients:

|                | Estimate | Std. Error | z value | Pr(>|z|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | 3.276683 | 0.044132   | 74.25   | <2e-16 *** |
| log(Area)      | 0.375029 | 0.008023   | 46.74   | <2e-16 *** |
| log(Adjacent)  | -0.095747| 0.006118   | -15.65  | <2e-16 *** |

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 3510.73 on 29 degrees of freedom
Residual deviance: 395.54 on 27 degrees of freedom
AIC: 562.37

- Log-log plot of mean vs variance (next slide)

```r
lambda <- predict(fit3, type="response")
plot(log(lambda),
    log((gala$Species -lambda)^2),
    xlab=expression(hat(lambda)),
    ylab=expression((y - hat(lambda))^2))
abline(0,1)
```
Diagnostics: Overdispersion

- X axis: \( \log(E\{Y\}) \) of each observation
- Y axis: \( \log(\text{Var}\{Y\}) \) of each observation
- The variance is increasing with the mean
- The variance is larger than the mean
  \( \rightarrow \) indicates overdispersion

\[
\log((y - \hat{\lambda})^2) = \log(\lambda^2) + \log(\hat{\lambda})
\]

6-28
Account for Overdispersion

- Estimate the dispersion parameter

  ```
  > sum(residuals(fit3, type="pearson")^2) / fit3$df.res
  [1] 16.52745  # >> 1, i.e. overdispersion present
  ```

- Quasipoisson model

  ```
  > summary(glm(Species~ log(Area)+log(Adjacent),
  family=quasipoisson, data=gala))
  ```

  Coefficients:

  | Estimate | Std. Error | t value | Pr(>|t|)   |
  |----------|------------|---------|-----------|
  | (Intercept) | 3.27668    | 0.17942 | 18.26     | < 2e-16 *** |
  | log(Area)   | 0.37503    | 0.03262 | 11.50     | 6.56e-12 *** |
  | log(Adjacent) | -0.09575  | 0.02487 | -3.85     | 0.000658 *** |
  ...

  (Dispersion parameter for quasipoisson family taken to be 16.52745)

  Null deviance: 3510.73 on 29 degrees of freedom
  Residual deviance: 395.54 on 27 degrees of freedom
  AIC: NA

- Same parameters, larger SE, weaker significance
Rate Models
(Faraway Ch. 3)
Motivation

• Each $Y_i$ can represent a different interval in space or time
  – # of crimes in cities of different size
  – # of customers served by workers who work different hours
  – # cars running red light in different intersections

• Modeling $Y_i$ as Poisson response is often more appropriate than Binomial
  – counts can be small as compared to the total (crimes)
  – the total may not be a count (workers, intersections)

• Goal: express a common effect of covariates on all $Y_i$, while accounting for differences in 'exposure'
  – 'exposure' needs to be a defined variable
Formulation

• Model $Y_i$ as

\[ Y_i \sim \text{Poisson}(\lambda_i), \quad \lambda_i = \text{exposure}_i e^{x_i'\beta}, \]

\[ \text{i.e. } \log(\lambda_i) = \log(\text{exposure}_i) + x_i'\beta \]

• The model is equivalent to using \textit{exposure} as a predictor with the coefficient=1

  – \textit{exposure} is often called an \textit{offset} in this context
Example: Gamma Radiation (Faraway Sec. 3.2)

- Effect of gamma radiation on the numbers of chromosomal abnormalities (ca) in cells

- Different number of cells (cells, in hundreds) exposed to a dose and a rate of radiation

- cells can be viewed as 'exposure' variable

library(faraway)
data(dicentric)
?dicentricic

> head(dicentric)
cells ca doseamt doserate
1 478 25 1 0.10
2 1907 102 1 0.25
3 2258 149 1 0.50
4 2329 160 1 1.00
5 1238 75 1 1.50
6 1491 100 1 2.00
Visualize the data

- Interaction plot: proportion of cells with abnormality
  - The effect of dose rate is multiplicative

```R
> with(dicentric, 
  interaction.plot(doseamt, doserate, ca/cells))
```
Model Without Offset

> # indicate that 'dose' is categorical:
> dicentric$doseF <- factor(dicentric$doseamt)

> # the total number of cells as an independent predictor
> fit4 <- glm(ca ~ log(cells) + log(doserate)*doseF,
>             family=poisson, data=dicentric)

...  
Coefficients:  
  Estimate  Std. Error    z value  Pr(>|z|)  
(Intercept)  -2.76534   0.38116   -7.255    4.02e-13 ***
log(cells)    1.00252   0.05137   19.517      < 2e-16 ***
log(doserate) 0.07200   0.03547    2.030     0.042403 *
doseF2.5    1.62984   0.10273   15.866      < 2e-16 ***
doseF5       2.76673   0.12287   22.517      < 2e-16 ***
log(doserate):doseF2.5  0.16111  0.04837    3.331    7.03e-06 ***
log(doserate):doseF5    0.19316  0.04299    4.493   7.03e-06 ***
...

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 916.127 on 26 degrees of freedom
  Residual deviance: 21.748 on 20 degrees of freedom
  AIC: 211.15

• The coefficient of log(cells) is close to 1, and the offset is appropriate
Model With Offset

> fit5 <- glm(ca ~ offset(log(cells)) + log(doserate)*doseF, 
  family=poisson, data=dicentric)

...

Coefficients:

|            | Estimate | Std. Error | z value | Pr(>|z|) |
|------------|----------|------------|---------|----------|
| (Intercept)| -2.74671 | 0.03426    | -80.165 | < 2e-16  *** |
| log(doserate) | 0.07178  | 0.03518    | 2.041   | 0.041299 * |
| doseF2.5   | 1.62542  | 0.04946    | 32.863  | < 2e-16  *** |
| doseF5     | 2.76109  | 0.04349    | 63.491  | < 2e-16  *** |
| log(doserate):doseF2.5 | 0.16122  | 0.04830    | 3.338   | 0.000844 *** |
| log(doserate):doseF5  | 0.19350  | 0.04243    | 4.561   | 5.1e-06  *** |

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 4753.00 on 26 degrees of freedom
Residual deviance: 21.75 on 21 degrees of freedom
AIC: 209.16

- The coefficients did not change much from the previous model.
- Good residual deviance, no evidence of overdispersion.
Grouped vs Ungrouped Data

- Since $Y_i^{\text{ind}} \sim \text{Poisson}(\lambda_i), \ i = 1, \ldots,$
  
  $\rightarrow \sum_i Y_i \sim \text{Poisson}(\sum_i \lambda_i),$

  we can model aggregate data

  - as in Binomial or Multinomial response, log-likelihood only involves sums of $y_i$ with same covariate patterns

  - e.g. add total cells and counts of abnormalities for entries with same doseamt and doserate

- Rate models: individual vs grouped
  
  - different deviances

  - same parameter estimates

  - same comparison of nested models

- Models with no offset: indiv. vs grouped
  
  - different deviances

  - same parameters except intercept

  - same comparison of nested models
Allowing for Extra Variation:

Negative Binomial

(Faraway Sec. 3.3)
Negative Binomial: Distribution

- Independent trials with \( P\{\text{success}\} = p \)
  - \( Z = \) the number of trials until \( k \)th success
  - \( Z \sim \text{NB}(p, k), \; Z = k, k+1, \ldots \)

- Probability distribution
  - \( P\{Z = z\} = \binom{z-1}{k-1} p^k (1-p)^z \)
  - \( E\{Z\} = \frac{k}{p} \)
  - \( Var\{Z\} = \frac{k(1-p)}{p^2} \)

- Obtained as the marginal distribution of \( Y|\lambda \sim \text{Poisson}(\lambda) \) where \( \lambda \sim \text{G}(k, \alpha) \)
  - \( P\{Y = y\} = \frac{1}{y!} \lambda^y e^{-\lambda}, \; Y = 0, 1, \ldots \)
  - \( f(\lambda) = \frac{\alpha^k}{\Gamma(k)} \lambda^{k-1} e^{-\alpha\lambda}, \; \lambda > 0 \)
Negative Binomial: Genesis

• Suppose \( Y|\lambda \sim \text{Poisson}(\lambda), \lambda \sim G(k, \alpha) \)

  – Joint distribution of \( Y \) and \( \lambda \):
    
    \[
p(Y = y, \lambda) = \frac{\alpha^k}{\Gamma(k)} \frac{\lambda^{y+k-1} e^{-(\alpha+1)\lambda}}{y!}
    \]

  – Marginal distribution of \( Y \) (by integrating \( \lambda \)):
    
    \[
p(Y = y) = \frac{\alpha^k}{\Gamma(k)} \Gamma(k) \int_0^\infty \frac{\lambda^{y+k-1} e^{-(\alpha+1)\lambda}}{y!} d\lambda
    \]
    
    \[
    = \frac{\alpha^k}{\Gamma(k)} \frac{\Gamma(y + k)}{\Gamma(y + 1)(\alpha + 1)^{y+k}} \Gamma(y + 1)
    \]
    
    \[
    = \frac{(y + k - 1)!}{(k - 1)!y!} \left( \frac{\alpha}{\alpha + 1} \right)^k \left( \frac{1}{\alpha + 1} \right)^y
    \]
    
    \[
    = \binom{y + k - 1}{k - 1} \left( \frac{\alpha}{\alpha + 1} \right)^k \left( \frac{1}{\alpha + 1} \right)^y
    \]

• \( \Rightarrow \) \( Z = Y + k \sim NB\left(\frac{\alpha}{\alpha + 1}, k\right), \ Y = 0, 1, \ldots \)
Negative Binomial: Model

- \( Y + k \sim NB(\frac{\alpha}{\alpha+1}, k) \), \( Y = 0, 1, \ldots \)
  
  \[
  - E\{Y\} = E\{Y + k\} - k = \frac{k}{\alpha/(\alpha+1)} - k = \frac{k}{\alpha} \quad (\text{denote } = \mu)
  \]
  
  \[
  - Var\{Y\} = Var\{Y + k\} = \frac{k(1-\alpha/(\alpha+1))}{\alpha^2/(\alpha+1)^2} = \frac{k}{\alpha} + \frac{k}{\alpha^2}
  = \mu + \mu^2/k \quad (= \text{Poisson variance + extra component})
  \]

- Log-likelihood of \( y_1, \ldots, y_n \)
  
  \[
  \sum_{i=1}^{n} \left( y_i \log \frac{1}{\alpha_i + 1} + k \log \frac{\alpha_i}{\alpha_i + 1} \right) + \text{constant}
  \]
  
  \[
  \overset{k/\alpha = \mu}{=} \sum_{i=1}^{n} \left( y_i \log \frac{\mu_i}{k + \mu_i} + k \log \frac{1}{k + \mu_i} \right) + \text{constant}
  \]
  
  - Can model \( \log \frac{\mu_i}{k + \mu_i} = x_i' \beta \)

- \( k \) is usually assumed same for all \( y_i \)
  
  - fixed \text{ a priori}, or estimated from the data
  
  - assumes constant coefficient of variation of \( \lambda \):
  
  \[
  \sqrt{Var\{\lambda\}}/E\{\lambda\} = \frac{1}{k}
  \]
Example: Galapagos Plants

- Negative Binomial in an instance of the Exponential Family of distributions when \( k \) is fixed
- Fix \( k \), use glm.

```r
> fit.nb <- glm(Species~ log(Area)+log(Adjacent),
               family=negative.binomial(1), data=gala)
```

Coefficients:

|               | Estimate | Std. Error | t value | Pr(>|t|)  |
|---------------|----------|------------|---------|-----------|
| (Intercept)   | 3.27257  | 0.15304    | 21.384  | < 2e-16 *** |
| log(Area)     | 0.35100  | 0.03773    | 9.304   | 6.52e-10 *** |
| log(Adjacent) | -0.03204 | 0.04015    | -0.798  | 0.432     |

(Dispersion parameter for Negative Binomial(1) family taken to be 0.4650222)

Null deviance: 54.069  on 29  degrees of freedom
Residual deviance: 13.965 on 27  degrees of freedom
AIC: 292.97
Example: Galapagos Plants

• Use glm.nb, and estimate $k$ from data

```r
> fit.nb1 <- glm.nb(Species~ log(Area)+log(Adjacent),
                 data=gala)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.27777    0.14495  22.613  <2e-16 ***
 log(Area)   0.34973    0.03541   9.875  <2e-16 ***
 log(Adjacent) -0.03316   0.03737  -0.887    0.375
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for Negative Binomial(2.6196)
 family taken to be 1)

Null deviance: 134.240  on 29 degrees of freedom
Residual deviance: 32.741  on 27 degrees of freedom
AIC: 284.99

Theta:  2.620
Std. Err.:  0.753
```

• Not a big difference in model fit in this example