Generalized Linear Models:
The Big Picture

STA303/STA1002: Methods of Data Analysis II, Summer 2016

Michael Guerzhoy

GLMs

- $g(\mu) = X\beta$
- $Y \sim dist_{\mu}, E(Y) = \mu$

g: link function

$dist_{\mu}$: some prob. distribution ("family" in R)

GLMs

	Link function	Distribution	In R
Logistic Regression	logit(π_i)=$X_i\beta$	$Y_i \sim Bernoulli(\pi_i)$	binomial(link=logit)
Linear Regression	$1 \times \mu_i = X_i\beta$	$Y_i \sim N(\mu_i, 1)$	gaussian(link=identity)
Poisson Regression	log(λ_i)=$X_i\beta$	$Y_i \sim Poisson(\lambda_i)$	poisson(link=log)
Poisson Regression	$1 \times \lambda_i = X_i\beta$	$Y_i \sim Poisson(\lambda_i)$	poisson(link=identity)

- Each distribution has a default ("canonical") link function, but other link functions can be used
  - E.g., identity link for Poisson
  - The link functions in the table above are $logit, identity, log, identity$

GLMs and Maximum Likelihood: Gaussian(link=identity)

- Data: $(\mathrm{x}_1, y_1), (\mathrm{x}_2, y_2), \dots, (\mathrm{x}_n, y_n)$,
- Likelihood: $\Pi_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_i\beta)^2}{2\sigma^2}\right), \sigma = 1$
- Log-likelihood: $-\Sigma_i \left(\frac{(y_i - x_i\beta)^2}{2\sigma^2}\right) - \mathrm{N}\sqrt{2\pi\sigma^2}$
  - Maximized when $\Sigma_i (y_i - x_i\beta)^2$ is minimized (i.e., $\beta = (x'x)^{-1}x'y$)

  Formally:
  - $\mathrm{argmax}_\beta \left(-\Sigma_i \left(\frac{(y_i - x_i\beta)^2}{2\sigma^2}\right) - \mathrm{N}\sqrt{2\pi\sigma^2}\right) = argmin_\beta \Sigma_i (y_i - \beta x_i)^2$
  $= argmin_\beta (y - x\beta)'(y - x\beta)$
  - $\frac{\partial}{\partial\beta}(y - x\beta)'(y - x\beta) = -2x'y + 2x'x\beta = 0$
  - $x'x\beta = 2x'y$
  - $\beta = (x'x)^{-1}x'y$

Cars, in R

Overdispersion

- Deviance: $const - 2\log \mathrm{P}(\mathrm{y}|\beta)$
- For the Gaussian distribution with $\sigma = 1$:
  - $-2\log \mathrm{P}(\mathrm{y}|\beta) = -2\Sigma_i\left(-\frac{(y_i - x_i\beta)^2}{2\sigma^2}\right) - \mathrm{N}\sqrt{2\pi\sigma^2} = \Sigma_i((y_i - x_i\beta)^2) + const$

- Estimate: $\hat{\psi} = \frac{Deviance}{Degrees\ of\ Freedom}$
- For the Gaussian distribution with assumed $\sigma = 1$:
  - $\hat{\psi} \approx \sigma^2_{actual}$
  - The average squared residual

- $SE_{\hat{\psi}}(\beta) = \sqrt{\hat{\psi}}SE_{est}(\beta)$
  - More uncertainty in the estimate the larger the average squared residual

(Overdispersion in Cars in R)

Overdispersion in general

- Estimate: $\hat{\psi} = \frac{Deviance}{Degrees\ of\ Freedom}$
- Multiply all uncertainty by $\sqrt{\hat{\psi}}$
  - Analogous to first estimating a linear regression assuming $\sigma^2 = 1$, and then scaling all uncertainties by $\hat{\sigma}$

Goodness of fit (Gaussian)

- If the model is correct (and there is no overdispersion),

$$Deviance \sim \chi^2(Npoints - Nparameters)$$

- Test:
  - 1-pchisq(Deviance, df= $Npoints - Nparameters$) < thr means the residuals are too large and there is lack of fit

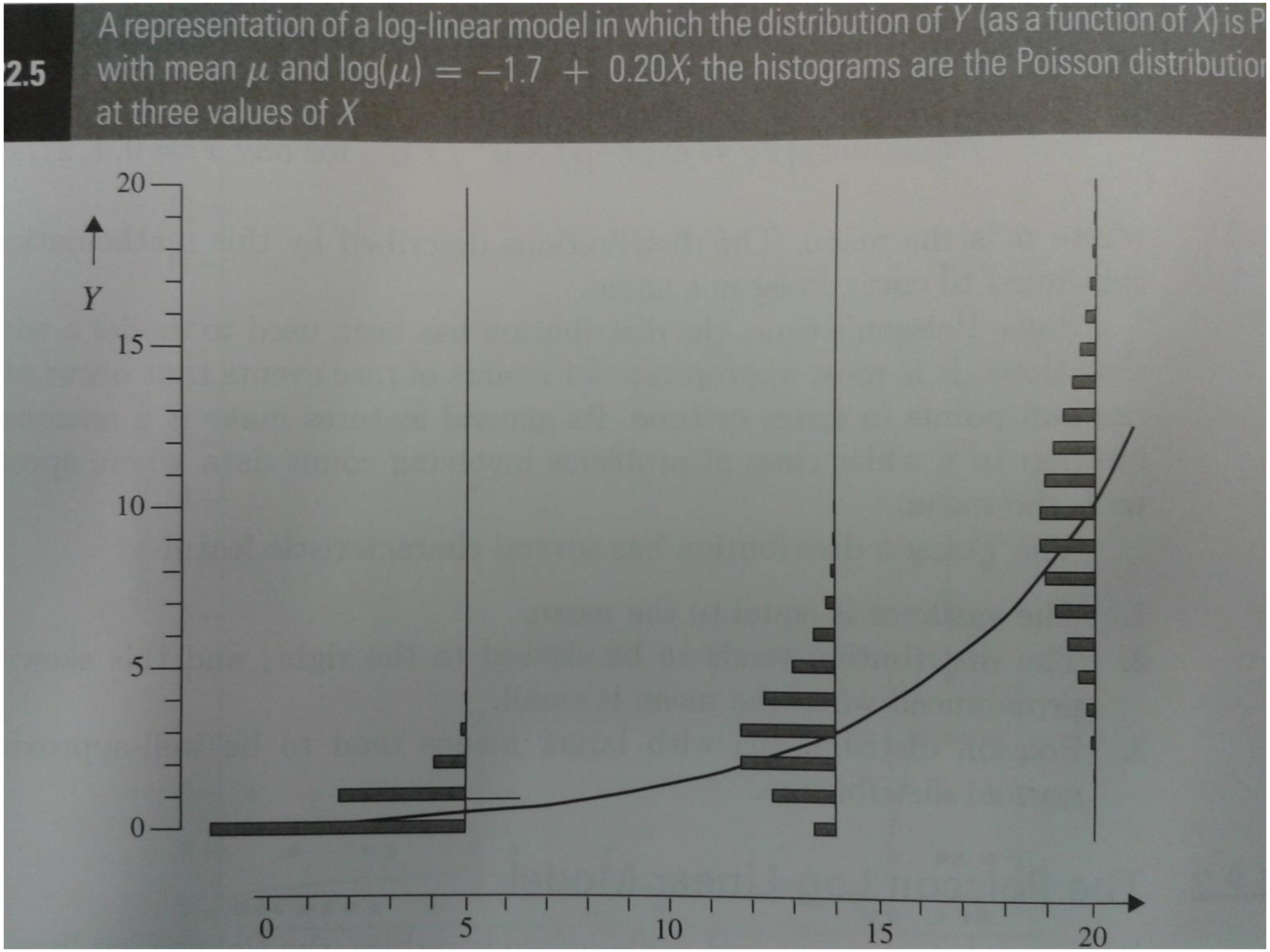
Likelihood Ratio Test

- $(Residual\ deviance\ A) - (Residual\ Deviance\ B) \sim \chi^2(df)$ if the additional parameters are all 0, larger than expected if not
  - One sided test 1-pchisq($diff\ in\ deviance$)
- Partial F-test, if the additional parameters are all 0:
  - $1/\sigma^2 SSE_{full} \sim \chi^2(df_1), 1/\sigma^2 SSE_{reduced} \sim \chi^2(df_2)$
  - $\frac{1}{\sigma^2}\left(SSE_{reduced} - SSE_{full}\right) \sim \chi^2(df_2 - df_1)$
- If we happen to know that $\sigma^2 = 1$, can perform a chi-squared test.
- If not:
  - Estimate σ^2 using MSE_{full}, and perform the partial F-test
  - $F = \frac{\frac{SSE_{reduced} - SSE_{full}}{df_2 - df_1}}{\frac{SSE_{full}}{df_1}} \sim \frac{\frac{\chi^2(df_2 - df_1)}{df_2 - df_1}}{\frac{X^2(df_1)}{df_1}} = F(df_2 - df_1, df_1)$
  - IF F is large, the reduction in the SSE is significant
    - One sided test

Residuals

- Pearson residuals (for logistic)
  - $P_{res,i} = \frac{y_i - m_i \hat{\pi}_{M,i}}{\sqrt{m_i \hat{\pi}_{M,i}(1-\hat{\pi}_{M,i})}}$
  - Approximately N(0, 1) if the model is correct
- Deviance residuals:
  - $sign(y_i - \pi_i)\sqrt{2\{y_i \log\left(\frac{y_i}{\pi_i}\right) + (1 - y_i)\log(\frac{1-y_i}{1-\pi_i})\}}$
  - Squares add up to the Deviance and are chi-square distributed
- Residuals (for Gaussian family)
  - $(y_i - \hat{y}_i)$
  - Approximately N(0, 1) if the model is correct
  - Sum of Squares (SSE) is chi-square distributed if the model is correct

Residuals – Elephant Example



Residuals – Elephant example

- Compute $(y_i - \hat{y}_i)$
- If the Poisson model uses the log-link, we have
  - $\hat{y}_i = \exp(\beta_0 + \beta_1 x_i)$
  - $y_i \sim Poisson(\mu_i)$
  - Residual distribution should be like the Poisson distribution around each of the means

Choice of Link Function (one covariate)

- Log link function :
  - $\log(\mu_i) = \beta_0 + \beta_1 x_i$ so $\mu_i = \exp(\beta_0)\exp(\beta_1)^{x_i}$
  - An increase of 1 in x_i means μ_i gets multiplied by $\exp(\beta_1)$
- Identity link function:
  - $\mu_i = \beta_0 + \beta_1 x_i$
  - An increase of 1 in x_i means μ_i increases by β_1

Example

- Predicting salary from age
  - Possibility 1: the salary grows by x% every year
    - Log-link is appropriate
  - Possibility 2: the salary grows by $z every year
    - Identity link is appropriate

Choice of link function

- When the data is distributed using $Bernoulli(\pi_i)$, cannot generally use identity or log link
  - Why?
- Logit link function:
  - $logit(\pi_i) = \beta_0 + \beta_1 x_i$
  - An increase by 1 in x_i means the log-odds grow by β_1
  - How does π_i change?
    - Depends on what it started out being
    - $logistic(\beta_0 + \beta_1 x_i + \beta_1)/logistic(\beta_0 + \beta_1 x_i)$

```
> c(plogis(3)/plogis(2),  plogis(4)/plogis(3))
[1] 1.081491 1.030905
> c(plogis(3)-plogis(2),  plogis(4)-plogis(3))
[1] 0.07177705 0.02943966
```