The Poisson Distribution

A Distribution of Counts

- How many people walk into your store at any given hour interval?
- Can approximate with binomial experiments:
  - 60 1-minute intervals
  - Each 1-minute interval:
    - Success: a person walks in
    - Not Success: nobody walks in
- BUT: we want finer grained time intervals than 1-minute
  - Becomes more realistic as the time interval approaches 0
  - As time interval $\rightarrow$ 0: Poisson Process

Poisson Process

- Counts the number of (successes) events in continuous time
- Events are independent
- Each point in time (or space) is equally likely for an event to occur
  - The probability to come in at any time is uniform

Poisson Distribution

- Random variable Y: the number of events that occur in some time interval
- Examples:
  - # of births per hour in a hospital
  - # of cars passing by 79 Madison in some time interval
  - # calls to a switchboard in an hour
  - # disease cases within a given town (one disease, not contagious)
  - # goals in a hockey game
  - # soldiers in the Prussian cavalry killed by horse kicks, by corp membership and year

Poisson Distribution

- Discrete
- Probability Mass Function:

$$P(Y \mid \lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \;;\; y = 0, 1, 2, \ldots$$

$E(Y) = \lambda$

$Var(Y) = \lambda$

Poisson Distribution Example

- On average, 5 people walk into my store at any given hour.
- During this specific hour, what is the probability that 0 people walk into my store?

Poisson Distribution Example

- On average, 5 people walk into my store at any given hour.
- During this specific hour, what is the probability that 0 people walk into my store?

$$P(Y \mid \lambda) = \frac{\lambda^{y} e^{-\lambda}}{y!}$$

$$P(Y=0 \mid \lambda = 5) = \frac{5^{0} e^{-5}}{0!}$$

$$= e^{-5}$$

Note: the distribution of inter-arrival times between events follows an *exponential distribution*.

GLM for Count Data

- Response variable: counts, Y (= 0 1, 2, ...)
- Traditional to assume Poisson distribution
  - $Y \sim Poisson(\lambda)$

$$log(Y) \quad = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

$$log(E(Y_X)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

log link function

systematic component

GLM for Count Data

- Response variable: counts, Y (= 0 1, 2, ...)
- Traditional to assume Poisson distribution
  - $Y \sim Poisson(\lambda)$

$$log(Y) \quad = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

$$\underbrace{log}_{\text{log link function}}(\underbrace{E(Y_X)}_{\lambda_X}) = \underbrace{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p}_{\text{systematic component}}$$

GLM for Count Data

- Response variable: counts, Y (= 0 1, 2, …)
- Traditional to assume Poisson distribution
  - $Y \sim Poisson(\lambda)$

$$log(Y) \quad = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

$$log(E(Y_X)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

log link function

systematic component

λ_X

$$\lambda_X = E(Y_X)$$

$$\lambda_X = exp\,(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)$$

1 unit change in x → multiplicative change in conditional mean $E(Y_X)$

OLS versus GLM

OLS:

- $log(Y) \sim N(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p, \sigma^2)$
- If $Y = 0$ → ad hoc solutions, i.e.: $log(0+1)$

GLM:

- $E(Y_X) = \lambda_X = \sigma_X{}^2$
- If $Y = 0$ → works
- Overdispersion: $Var(Y) >> E(Y)$

Measuring Goodness of Fit for GLMs

Goodness of fit for each parameter β_i

Wald test:

- Analogous to t-tests in Linear Regression
- Looks at each parameter β_i separately
- H_0 (null): $\beta_i = 0$
- H_a (alternative): $\beta_i \neq 0$

Goodness of fit for entire model

Pearson's Chi-Squared statistic

- Analogous to whole model F-test in Linear Regression
- Looks at each parameter β_i separately
- H_0 (null): $g(Y) = \beta_0$
- H_a (alternative): *otherwise*

Deviance

Unlike in OLS, there is no R^2

- Instead: *Deviance*, *D*
- $D = 2$ *[log lik(saturated model) – log lik(your proposed model)]*
- $D = -2[\log P(y|\beta) - \log P(y|\beta_{saturated})]$
- Looks at how closely the fitted values from your model match actual Y values from the raw data.
- In a saturated model: perfectly overfit model.
- In a good model, want small deviance
- D ~ $X_{n\text{-}p}{}^2$

Deviance is useful

…because you can compare models.

Fuller model: $g(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p,$

Reduced model: $g(Y) = \beta_0 + \beta_1 x_1$

Question: do all the other covariates ($\beta_2 x_2 + \ldots + \beta_p x_p$) contribute to your model?

- H_0 (null): $\beta_2 = \beta_3 = \ldots = \beta_p = 0$
- H_a (alternative): Not all $\beta_i \neq 0$; $i = 2, 3, \ldots, p$

$$D_{red} - D_{full} \sim X_{full\ -\ red}^2$$

If $p < 0.05$ → fuller model is better