Used on count data
- Discrete
- Non-negative

Issues.

1. Non-linear / no negatives ⇒ must use exponential = $y = e^{b_0 + b_1x_1 + b_2x_2 \dots}$

2. Varying variance: with count data, variance is higher with higher count (y)
⇒ distribution needs to increase variance when expected count increases

- Normal distribution also allows negative numbers in error

⇓

distribution changes according to expected count + skewed to not include <0

⇓ distribution becomes poisson, not normal

Assumptions

1. Variance = mean ← y-axis

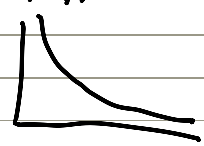
$y = e^{b_0 + b_1x_1 + b_2x_2 \dots}$

$\ln(y) = b_0 + b_1x_1 + b_2x_2 \dots + b_nx_n$

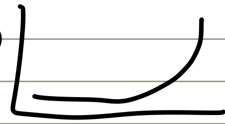
one variable:

$\ln(y) = b_0 + b_1x_1$

$b_1 < 0$



$b_1 > 0$



$y = e^{b_0 + b_1x_1}$

e^{b_0} ⇒ when $x=0$

e^{b_1} ⇒ % change per x ⇒ incidence rate ratio

Fitting Model

using maximum likelihood estimation

ex. X_i = # of textbooks, 50 students $(X_1 \dots X_{50})$

$\sum x_i = 150$

We want the mean # of books a person buys

$X_i \sim$ Poisson Distribution(y)

$$f(x_i \mid y) = \frac{e^{-y} y^{x_i}}{x_i!}$$

Likelihood:

$$L(y \mid x_1 \ldots x_{50}) = \prod_{i=1}^{50} f(x_i \mid y)$$ → log of products = sum
$\log(ab) = \log a + \log b$

Log both sides:

$$\log\left(L(y \mid x_1 \ldots x_{50})\right) = \sum_{i=1}^{50} \log\left(\frac{e^{-y} y^{x_i}}{x_i!}\right) = \sum_{i=1}^{50}\left[-y + x_i \log y - \log(x_i!)\right]$$

constant with respect to y ⇒ can be dropped

$$\log\left(L(y \mid x_1 \ldots x_{50})\right) \propto \sum_{i=1}^{50}(-y + x_i \log y) = -50y + \sum_{i=1}^{50} x_i \log y$$

proportion

Minimizing function: $\frac{d\, f(\ldots)}{dy} = 0$

$$\frac{d}{dy} \log\left(L(y \mid x_1 \ldots x_{50})\right) = -50 + \frac{\sum_{i=1}^{50} x_i}{y} = 0$$

$$50 = \frac{\sum_{i=1}^{50} x_i}{y}$$

$$y = \frac{\sum_{i=1}^{50} x_i}{50} = \frac{150}{50} = 3$$ ⇒ each student, on average, buys 3 books