Poisson, Negative binomial & Exponential Regression

Poisson Regression → It is used to model response variable = (y-value) that are "counts". It tells you which explanatory variable have a statistically significant effect on the response variable.

In other words.

→ It tells you which x-value works on the y-value. It is best used for "Rare events." as these tend to follow a "poisson distribution"

Ex No of Computer infrastructure crashes in MNC
No of Flights Cancelled without weather reasons.

For large means. The normal distribution is a good approximation of poisson distribution. Therefore poisson Regression is more suited to cases where response variable is a small integer.

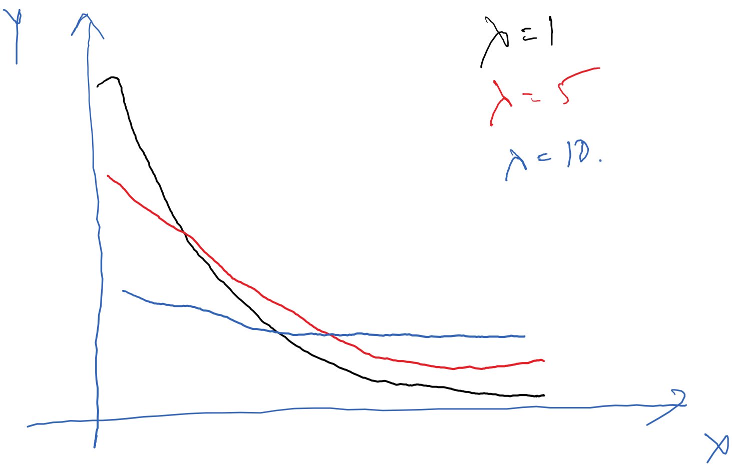
Poisson Regression is used only for Neumerical Continuous data. The same technique can be used for modeling categorical explanatory variables or counts in the cells of the contengency table. When used this way model is called loglinear model

→ Assumptions for Poisson Regression

- → Y-values are counts
- → Counts must be positive integers
  - → No factors / Negative values because poisson is discrete
- → Counts must follow poisson distribution
- → Explanatory variables must be continuous / Dichatomous

→ Explanatory variables must be Continuous / Dichotomous

→ Observations must be independent.



$\lambda = \mu = n \cdot p$ for large values.

Poisson distribution

→ family of curves.

Negative binomial Regression

It is used for Over-dispersed Count data, when the Conditional Variance exceeds the conditional mean.

⇒ It is Considered as generalised form of poisson Regression since it has same mean structure as poisson & an extra parameter to model Over dispersion

→ Negative binomial is same as binomial experiment but binomial has fixed no of trials.

→ Conditions for negative binomial dist

1) Fixed no of n trials.
2) Each trial is independent
3) Only two outcomes are possible → Dichatomous
4) probability of success of each trial (p) is a constant
5) A Random Variable Y = the no of success.

Ex. Deck of Cards → select a card. → Replace & Repeat 20 times

Y is the no of Aces we draw

→ Binomial dist with difference

→ No of trials are not fixed

→ A random Var Y = the no of Trials needed to make 'r' success

→ Same experiment → Repeat till you get 10 Aces → Negative binomial.

Regular binomial → Success Counts

Negative binomial → Failure Counts → So Called Negative binomial

Formula

$x \rightarrow$ no of trials

$r \rightarrow$ rate of success.

$$b(x ; r, p) = {}_{x-1}C_{r-1} * p^r * (1-p)^{x-r}$$

$$\mu = \frac{r}{p} = \frac{\text{no of trials}}{\text{prob of success for any trial}}$$

Solving negative binomial problem

$$nb(x ; r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x \qquad x = 0, 1, 2, \ldots$$

r = no of success & p is the probability of success

Ex. Ask 15 ppl & I get 5 votes , $P = 20\%$

$\Rightarrow P = 20\%$

$r = 5$

failures $x = 15 - 5 = 10$

$$nb(10 : 5, 0.2) = ({}_{14}C_4)(0.2)^5(0.8)^{10} = 0.34$$

or 34%

binomial dist

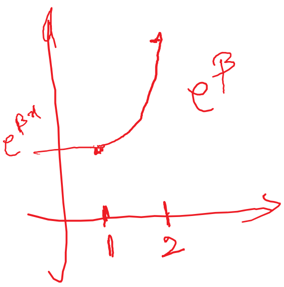
$\Rightarrow$ geometric distribution is a special case of Negative Binomi...

Exponential Regression

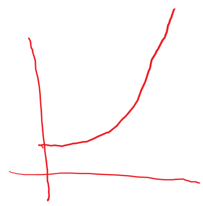
Some times Non-linear models are made linear to fit models using transformation.

Assume exponential model $y = \alpha e^{\beta x}$

take log on b.s $\Rightarrow$ $\underbrace{\ln(y) = \ln(\alpha) + \beta x}_{\text{linear form}}$



$\Rightarrow$ $y' = \alpha' + \beta x + \varepsilon$ $\quad$ ε is error term



Observation : Since

$$\alpha e^{\beta(x+1)} = \alpha e^{\beta x} \cdot e^{\beta}$$

An increase in x by one unit will result y multiplied by e^{β}

e^{β} is exponential

$\Rightarrow$ $\ln(y) = \beta x + \delta$ is reffered as log-level regression

Clearly they can be expressed as $y = \alpha e^{\beta x}$ by setting $\alpha = e^{\delta}$

Fitting model is given in the code.