Session 6

Bayesian Methods

Yuri Balasanov

University of Chicago, MScA

For any inquiries contact the author, Yuri Balasanov, at
ybalasan@uchicago.edu or yuri.balasanov@iLykei.com .

Outline of the Session

- Scales of data
- Review of generalized linear model
- Links and their properties
- Types of models parameterizations and distributions

Possible Scales of Data

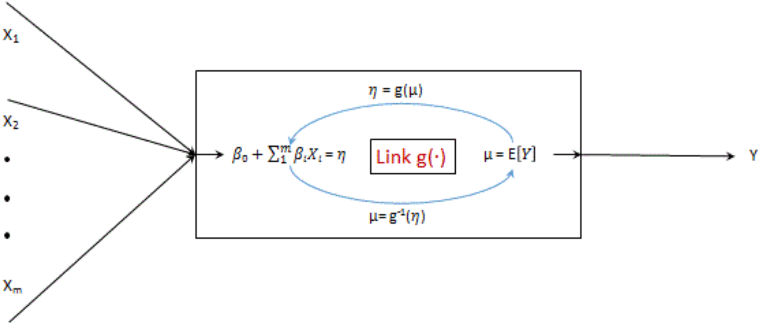
As we now, data may come in different formats (scales):

- **Metric**, i.e. real numbers. E.g. temperature in one of the scales, Celsius, Fahrenheit, Kelvin.
  Metric scales can be of several types:
  - **Ratio:** there is natural zero, e.g. temperature in Kelvin: $0K$ means "no temperature" or no heat can be extracted from gas cooled down to absolute zero. The coldest spot $(1K)$ in the universe is Boomerang Nebula in the constellation of Centaurus, 5000 light-years away
  - **Interval** type: no natural zero, e.g. temperature measurement in Celsius or Fahreinheit do not have natural zero, zero is a relative point
  - **Count** type: a non-negative integer. Differences are meaningful
- **Ordinal**, i.e. ranks: these are like counts, but differences are meaningless, e.g. the University of Chicago is numer 11 in the Times Higher Education World Reputation Rankings 2016. We know UofC is behind Harvard ranked number 1, but how much?
- **Nominal** type: categorical variables, names; we cannot tell which category is better

Generalized Linear Model

- When either predictors or output of the data are not of metric type or not consistent with Gaussian assumption application of linear model based on least squares method, strictly speaking, is not correct
- Generalization of linear model to a broader class than Gaussian linear model was developed in 1980s by John Nelder and Robert Wedderburn
- GLM extends linear model by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value
- Different models of the GLM family result in different shapes of likelihood functions which is important for Bayesian analysis

Links



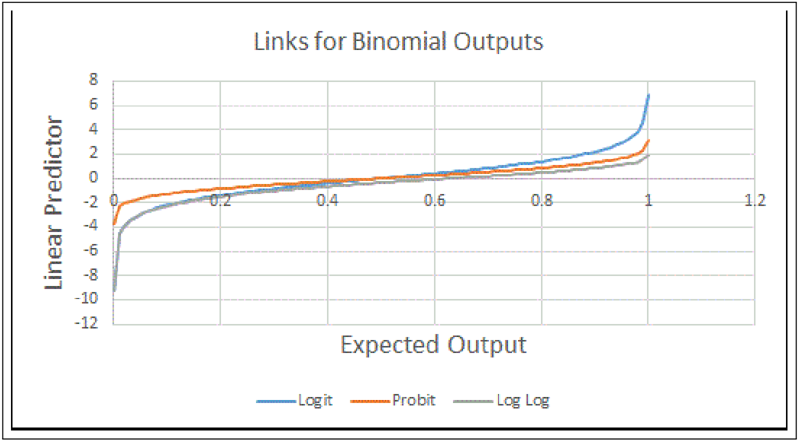
Links

- Link is a function g connecting linear predictor $\eta = \beta_0 + \sum_1^m \beta_i \mathbf{X}_i$ and the expectation of the output $\mu = \mathbb{E}\left[\mathbf{Y}\right]$: $g\left(\mu\right) = \eta$.
- In classical linear model the linear predictor and the expected output are identical: $g\left(\mu\right) \equiv \mu = \eta$. Both μ and η are real numbers from $(-\infty, \infty)$.
- When the output originates from counts and the distribution is Poisson $\mu > 0, \eta \in (-\infty, \infty)$. In order to match the domains of the expected output and the linear predictor we use link $\eta = \log\left(\mu\right)$ or $\mu = e^{\eta}$.
- When the output is binomial, $\mu \in (0, 1)$ and the link should map that interval onto the whole real line. This is usually done by one of the following three link functions:
    - Logit: $\eta = \log\left(\frac{\mu}{1-\mu}\right)$;
    - Probit: $\eta = \Phi^{-1}\left(\mu\right)$, where $\Phi\left(x\right)$ is cumulative normal distribution function;
    - Complementary log-log: $\eta = \ln\left\{-\ln\left(1-\mu\right)\right\}$

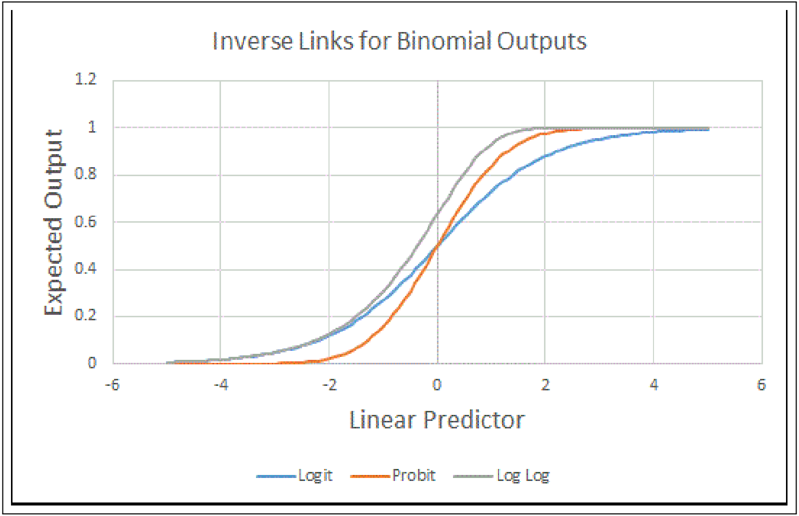
Links Continued

- Another useful type of link functions belongs to the power family:
  - $$\eta = \frac{\mu^\lambda - 1}{\lambda},$$
    which has the smoothness property at $\lambda = 0$
    $$\lim_{\lambda \to 0} \frac{\mu^\lambda - 1}{\lambda} = \ln(\mu);$$
  - Or
    $$\eta = \begin{cases} \mu^\lambda; & \lambda \neq 0, \\ \ln(\mu); & \lambda = 0. \end{cases}$$
    In both cases the point $\lambda = 0$ is special and needs to be taken care of.
- In the case when the output has gamma distribution the canonical link is called reciprocal: $\eta = \frac{1}{\mu}$.
- When the output is distributed as inverse Gaussian, the canonical link is $\eta = \frac{1}{\mu^2}$.

Links for Binomial Outputs



Inverse Links for Binomial Outputs



Inverse Links For Other Distributions of Outputs

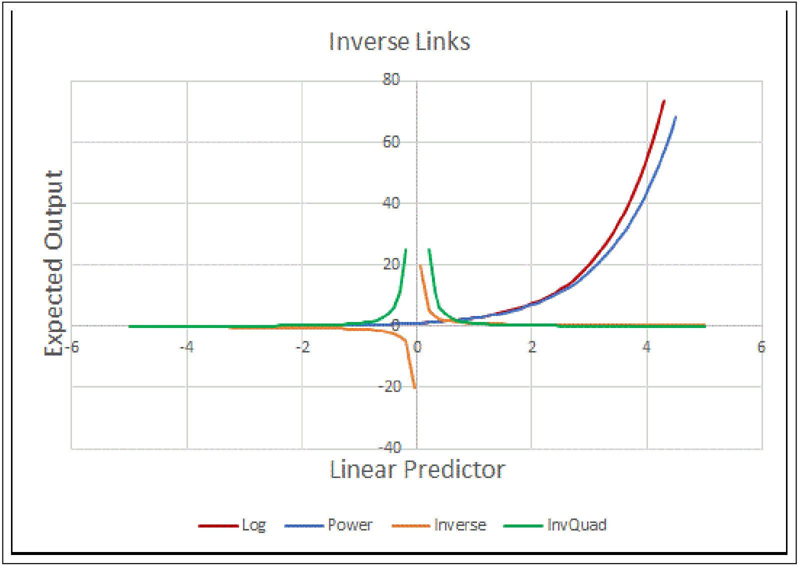


Table of Typical Linear Predictors

Depending on the type of predictors we will look at the following model parameterizations of linear predictors withing the generalized linear model

Scale Type of Predictors					
No Predict		Metric Predict		Nominal Predict	
One Gr.	Two Gr.	Single Pred	Multiple Pred	Single Fact	Multiple Fact
β_0	$\beta_{x=1}$ $\beta_{x=2}$	$\beta_0 + \beta_1 x$	$\beta_0 + \sum_k \beta_k x_k + \sum_{jk} \beta_{j\times k} x_j x_k \ldots$	$\vec{\beta}_0 + \vec{\beta} \cdot \vec{x}$	$\beta_0 + \sum_k \vec{\beta}_k \vec{x}_k + \sum_{jk} \vec{\beta}_{j\times k} \vec{x}_{j\times k} \ldots$

John K. Kruschke, Doing Bayesian Data Analysis, A Tutorial with R, JAGS, and STAN, 2015, Elsevier.

Table of Typical Output Distributions

The table below shows some (but not all) typical distributions of outputs depending on type of output variable

Scale y	Distrib. $y \sim f(\mu; \ldots)$	Inverse link $\mu = g^{-1}(\eta)$
Metric	$y \sim dnorm(\mu, \sigma)$	$\mu = \eta$
Binary	$y \sim dbern(\mu)$	$\mu = \text{logistic}(\eta)$
Nominal	$y \sim dmultinom\left(\mu_1, \ldots, \mu_p\right)$	$\mu_k = \frac{\exp(\eta_k)}{\sum_j \exp(\eta_j)}$
Ordinal	$y \sim dmultinom\left(\mu_1, \ldots, \mu_p\right)$	$\mu = \Phi\left(\frac{\theta_k - \eta}{\sigma}\right) - \Phi\left(\frac{\theta_{k-1} - \eta}{\sigma}\right)$
Count	$y \sim dpois(\mu)$	$\mu = \exp(\eta)$

John K. Kruschke, Doing Bayesian Data Analysis, A Tutorial with R, JAGS, and STAN, 2015, Elsevier.