planetmath.org

Math for the people, by the people.

generalized linear model

Canonical name	GeneralizedLinearModel
Date of creation	2013-03-22 14:30:11
Last modified on	2013-03-22 14:30:11
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	17
Author	CWoo (3771)
Entry type	Definition
Classification	msc 62J12
Synonym	GLM
Defines	link function
Defines	canonical parameter
Defines	cumulant function
Defines	variance function

Given a random vector, or the response variable, $\mathbf{Y}$, a *generalized linear model*, or GLM for short, is a statistical model $\{f_{\mathbf{Y}}(\boldsymbol{y} \mid \boldsymbol{\theta})\}$ such that

1. the components of $\mathbf{Y}$ are mutually independent of each other,
2. $f_{Y_i}(y_i \mid \theta_i)$ belongs to the exponential family of distributions and has the following canonical form:

   $$f_{Y_i}(y_i \mid \theta_i) = \exp[y\theta_i - b(\theta_i) + c(y)],$$

   where the parameter θ_i is called the *canonical parameter* and $b(\theta_i)$ is called the *cumulant function*.
3. for each component or variate Y_i, with a corresponding set of p covariates X_{ij}, there exists a monotone differentiable function g, called the *link function*, such that

   $$g(\mathrm{E}[Y_i]) = \mathbf{X}_i{}^{\mathrm{T}}\boldsymbol{\beta},$$

   where $\mathbf{X}_i{}^{\mathrm{T}} = (X_{i1}, \dots, X_{ip})$, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^{\mathrm{T}}$ is a parameter vector.

In practice, an extra parameter called the dispersion parameter, ϕ, is introducted to the model to lower a phenonmenon known as overdispersion. The GLM now looks like:

$$f_{Y_i}(y_i \mid \theta_i) = \exp[\frac{y\theta_i - b(\theta_i)}{a(\phi)} + c(y, \phi)]$$

Remarks

- Below is a table of canonical parameters and cumulant functions for some well-known distributions from the exponential family:

distribution	notation	canonical p
http://planetmath.org/NormalRandomVariableNormal	$N(\mu, \sigma^2)$	μ
http://planetmath.org/PoissonRandomVariablePoisson	$Poisson(\mu)$	ln
http://planetmath.org/BernoulliDistribution2Binomial	$Bin(m, \pi)$	logi
http://planetmath.org/GammaRandomVariableGamma	$Gamma(\alpha, \lambda)$	−

- GLM is a direct generalization of the general linear model, which includes linear regression models, ANOVA and ANCOVA. The link function for the general linear model is the identity function $g(\mu) = \mu$.

- For a GLM, $\mathrm{E}[Y] = b'(\theta)$ and $\mathrm{Var}[Y] = b''(\theta)$. $b''(\theta)$, when expressed in terms of $\mu = \mathrm{E}[Y]$, is known as the *variance function* $V(\mu)$. Below are some examples of variance functions:

distribution	notation	variance fu
Normal	$N(\mu, \sigma^2)$	1
Poisson	$Poisson(\mu)$	μ
`http://planetmath.org/BernoulliDistribution2`Binomial	$Bin(m, \pi)$	$\pi(1 -$
Gamma	$Gamma(\alpha, \lambda)$	$\frac{1}{\lambda^2}$

- The logistic regression model, where the response variable Y is categorial in nature, is a special case of GLM, with possible link functions the logit function, $\mathrm{logit}(\pi) = \ln(\mathrm{odds}(\pi))$, the inverse cumulative normal distribution function, or probit function $\Phi^{-1}(\pi)$, or the complementary-log-log function, $\ln(-\ln(1-\pi))$, where the parameter π is between 0 and 1, usually measured as the frequency of occurrences of certain events.

- The log-linear model, where the response variable Y has a Poisson distribution, is also a special case of GLM, with link function the natural logarithm of the parameter μ in question. Poisson distribution is typically used to model count or frequency data.

References

[1] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman & Hall/CRC, 2nd ed., London (1989).

[2] A. J. Dobson, *An Introduction to Generalized Linear Models*, Chapman & Hall, 2nd ed. (2001).