overdispersion

Canonical name	Overdispersion
Date of creation	2013-03-22 14:30:34
Last modified on	2013-03-22 14:30:34
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	10
Author	CWoo (3771)
Entry type	Definition
Classification	msc 62J12
Defines	dispersion parameter

When applying the generalized linear model or GLM to the real world, a phenomenon called *overdispersion* occurs when the observed variance of the data is larger than the predicted variance. This is particularly apparent in the case of a Poisson regression model, where

$$\text{predicted variance} = \text{predicted mean},$$

or the binary logistic regression model, where

$$\text{predicted variance} = \text{predicted mean(1- predicted mean)}.$$

A parameter, called the *dispersion parameter*, ϕ, is introducted to the model to lower this overdispersion effect.

The GLM, with the inclusion of this dispersion parameter, has the following density function:

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

Dispersion parameters for some of the well known distributions from the exponential family are listed in the following table:

distribution	notation	dispersion param
http://planetmath.org/NormalRandomVariableNormal	$N(\mu, \sigma^2)$	σ^2
http://planetmath.org/PoissonRandomVariablePoisson	$Poisson(\mu)$	1
http://planetmath.org/BernoulliDistribution2Binomial	$Bin(m, \pi)$	$\frac{1}{m}$
http://planetmath.org/GammaRandomVariableGamma	$Gamma(\alpha, \lambda)$	$\frac{1}{\alpha}$

References

[1] J. M. Hilbe, *Negative Binomial Regression*, Cambridge University Press, Cambridge (2007).

[2] A. Agresti, *An Introduction to Categorical Data Analysis*, Wiley & Sons, New York (1996).

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

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