planetmath.org

Math for the people, by the people.

deviance

Canonical name	Deviance
Date of creation	2013-03-22 14:34:04
Last modified on	2013-03-22 14:34:04
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	8
Author	CWoo (3771)
Entry type	Definition
Classification	msc 62J12
Defines	null model
Defines	saturated model

Background

In testing the fit of a generalized linear model $\mathcal{P}$ of some data (with response variable $\mathbf{Y}$ and explanatory variable(s) $\mathbf{X}$), one way is to compare $\mathcal{P}$ with a similar model $\mathcal{P}_0$. By similarity we mean: given $\mathcal{P}$ with the response variable $Y_i \sim f_{Y_i}$ and link function g such that $g(\mathrm{E}[Y_i]) = \mathbf{X}_i{}^{\mathrm{T}}\boldsymbol{\beta}$, the model $\mathcal{P}_0$

1. is a generalized linear model of the same data,
2. has the response variable Y distributed as f_Y, same as found in $\mathcal{P}$
3. has the same link function g as found in $\mathcal{P}$, such that $g(\mathrm{E}[Y_i]) = \mathbf{X}_i{}^{\mathrm{T}}\boldsymbol{\beta_0}$

Notice that the only possible difference is found in the parameters $\boldsymbol{\beta}$.

It is desirable for this $\mathcal{P}_0$ to be served as a base model in case when more than one models are being assessed. Two possible candidates for $\mathcal{P}_0$ are the *null model* and the *saturated model*. The null model $\mathcal{P}_{null}$ is one in which only one parameter μ is used so that $g(\mathrm{E}[Y_i]) = \mu$, all responses have the same predicted outcome. The saturated model $\mathcal{P}_{max}$ is the other extreme where the maximum number of parameters are used in the model so that the observed response values equal to the predicted response values exactly, $g(\mathrm{E}[Y_i]) = \mathbf{X}_i{}^{\mathrm{T}}\boldsymbol{\beta}_{max} = y_i$

Definition The *deviance* of a model $\mathcal{P}$ (generalized linear model) is given by

$$\mathrm{dev}(\mathcal{P}) = 2\big[\ell(\hat{\boldsymbol{\beta}}_{max} \mid \mathbf{y}) - \ell(\hat{\boldsymbol{\beta}} \mid \mathbf{y})\big],$$

where ℓ is the log-likelihood function, $\hat{\boldsymbol{\beta}}$ is the MLE of the parameter vector $\boldsymbol{\beta}$ from $\mathcal{P}$ and $\hat{\boldsymbol{\beta}}_{max}$ is the MLE of parameter vector $\boldsymbol{\beta}_{max}$ from the saturated model $\mathcal{P}_{max}$.

Example For a normal or general linear model, where the link function is the identity:

$$\mathrm{E}[Y_i] = \mathbf{x}_i{}^{\mathrm{T}}\boldsymbol{\beta},$$

where the Y_i's are mutually independent and normally distributed as $N(\mu_i, \sigma^2)$. The log-likelihood function is given by

$$\ell(\boldsymbol{\beta} \mid \mathbf{y}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu_i)^2 - \frac{n\ln(2\pi\sigma^2)}{2},$$

where $\mu_i = \mathbf{x}_i{}^{\mathrm{T}}\boldsymbol{\beta}$ is the predicted response values, and n is the number of observations.

For the model in question, suppose $\hat{\mu}_i = {\mathbf{X}_i}^{\mathrm{T}}\hat{\boldsymbol{\beta}}$ is the expected mean calculated from the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ of the parameter vector $\boldsymbol{\beta}$. So,

$$\ell(\hat{\boldsymbol{\beta}} \mid \mathbf{y}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2 - \frac{n\ln(2\pi\sigma^2)}{2},$$

For the saturated model $\mathcal{P}_{max}$, the predicted value $(\hat{\mu}_{max})_i =$ the observed response value y_i. Therefore,

$$\ell(\hat{\boldsymbol{\beta}}_{max} \mid \mathbf{y}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - (\hat{\mu}_{max})_i)^2 - \frac{n\ln(2\pi\sigma^2)}{2} = -\frac{n\ln(2\pi\sigma^2)}{2}.$$

So the deviance is

$$\mathrm{dev}(\mathcal{P}) = 2\big[\ell(\hat{\boldsymbol{\beta}}_{max} \mid \mathbf{y}) - \ell(\hat{\boldsymbol{\beta}} \mid \mathbf{y})\big] = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2,$$

which is exactly the residual sum of squares, or RSS, used in regression models.

Remarks

- The deviance is necessarily non-negative.
- The distribution of the deviance is asymptotically a `http://planetmath.org/ChiSquaredRan` square distribution with $n-p$ degress of freedom, where n is the number of observations and p is the number of parameters in the model $\mathcal{P}$.
- If two generalized linear models $\mathcal{P}_1$ and $\mathcal{P}_2$ are nested, say $\mathcal{P}_1$ is nested within $\mathcal{P}_2$, we can perform hypothesis testing H_0: the model for the data is $\mathcal{P}_1$ with p_1 parameters, against H_1: the model for the data is the more general $\mathcal{P}_2$ with p_2 parameters, where $p_1 < p_2$. The deviance difference Δ(dev)= dev($\mathcal{P}_2$) − dev($\mathcal{P}_1$) can be used as a test statistic and it is approximately a chi square distribution with $p_2 - p_1$ degrees of freedom.

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).