

1.5 Measuring the goodness of fit:

Special case:

Assume in the full model $\theta_i, i = 1, \dots, n$, are **n** parameters, one per observation.

Thus, $\mu_i = b'(\theta_i)$, are also **n** parameters, one per observation. Then, the log-likelihood function is

$$l = \sum_{i=1}^n \left\{ \frac{[y_i \theta_i - b(\theta_i)]}{a(\phi)} + c(y_i, \phi) \right\}.$$

Then,

$$\begin{aligned} \frac{\partial l}{\partial \theta_i} &= \frac{[y_i - b'(\theta_i)]}{a(\phi)} = \frac{(y_i - \mu_i)}{a(\phi)} = 0 \\ \Rightarrow \hat{\mu}_i &= y_i. \end{aligned}$$

Suppose we want to use a simplest (null) model to fit the data in which

$$\mu_1 = \mu_2 = \dots = \mu_n = \mu \Leftrightarrow \theta_1 = \theta_2 = \dots = \theta_n = \theta$$

Then, in the null model,

$$\begin{aligned} \frac{\partial l}{\partial \theta} &= \sum_{i=1}^n \frac{[y_i - b'(\theta)]}{a(\phi)} = \sum_{i=1}^n \frac{(y_i - \mu)}{a(\phi)} = 0 \\ \Rightarrow \hat{\mu} &= \frac{\sum_{i=1}^n y_i}{n} = \bar{y}. \end{aligned}$$

Let

$$\hat{\theta}_i = (b')^{-1}(\hat{\mu}_i)$$

and

$$\hat{\theta} = (b')^{-1}(\hat{\mu}) = (b')^{-1}(\bar{y}).$$

Denote

$$l(\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_n) = l(y_1, y_2, \dots, y_n) = \sum_{i=1}^n \left\{ \frac{[y_i \hat{\theta}_i - b(\hat{\theta}_i)]}{a(\phi)} + c(y_i, \phi) \right\}$$

and

$$l(\hat{\mu}) = \sum_{i=1}^n \left\{ \frac{[y_i \hat{\theta} - b(\hat{\theta})]}{a(\phi)} + c(y_i, \phi) \right\}.$$

Then,

$$\begin{aligned} 2[\log(\text{likelihood in full model}) - \log(\text{likelihood in null model})] \\ = 2[l(y_1, y_2, \dots, y_n) - l(\hat{\mu})] \end{aligned}$$

$$= \frac{2}{a(\phi)} \cdot \sum_{i=1}^n [y_i(\hat{\theta}_i - \hat{\theta}) + b(\hat{\theta}) - b(\hat{\theta}_i)] \\ = \frac{D(y_1, y_2, \dots, y_n | \hat{\mu})}{a(\phi)},$$

where

$$D(y_1, y_2, \dots, y_n | \hat{\mu}) = 2 \sum_{i=1}^n [y_i(\hat{\theta}_i - \hat{\theta}) + b(\hat{\theta}) - b(\hat{\theta}_i)]$$

is called the **deviance** and can be used to measure the goodness of fit with the null model.

Example 1 (normal distribution, continue):

$Y \sim N(\mu, \sigma^2)$. Then

$$\theta = \mu, b(\theta) = \frac{\theta^2}{2} = \frac{\mu^2}{2}.$$

$$D(y_1, y_2, \dots, y_n | \hat{\mu}) = 2 \sum_{i=1}^n [y_i(\hat{\theta}_i - \hat{\theta}) + b(\hat{\theta}) - b(\hat{\theta}_i)] \\ = 2 \sum_{i=1}^n [y_i(\hat{\mu}_i - \hat{\mu}) + b(\hat{\mu}) - b(\hat{\mu}_i)] \\ = 2 \sum_{i=1}^n \left[y_i(y_i - \bar{y}) + \frac{\bar{y}^2}{2} - \frac{y_i^2}{2} \right] \\ = \sum_{i=1}^n (y_i - \bar{y})^2$$

Note:

In the above example, $a(\phi) = \sigma^2 = \phi$ and

$$\frac{D(y_1, y_2, \dots, y_n | \hat{\mu})}{a(\phi)} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Also,

$$2[\log(\text{likelihood in full model}) - \log(\text{likelihood in null model})] \\ = 2[l(y_1, y_2, \dots, y_n) - l(\hat{\mu})] \\ = -2 \log \Lambda \sim \chi^2_{n-1}$$

where

$$\Lambda = \frac{\text{likelihood in null model}}{\text{likelihood in full model}} \equiv \text{likelihood ratio.}$$

Example 2 (Poisson distribution, continue):

$Y \sim P(\mu)$. Then,

$$\theta = \log(\mu), b(\theta) = \exp(\theta) = \mu.$$

Therefore

$$\begin{aligned} D(y_1, y_2, \dots, y_n | \hat{\mu}) &= 2 \sum_{i=1}^n [y_i(\hat{\theta}_i - \hat{\theta}) + b(\hat{\theta}) - b(\hat{\theta}_i)] \\ &= 2 \sum_{i=1}^n \{y_i[\log(\hat{\mu}_i) - \log(\hat{\mu})] + \hat{\mu} - \hat{\mu}_i\} \\ &= 2 \sum_{i=1}^n \{y_i[\log(y_i) - \log(\bar{y})] + \bar{y} - y_i\} \\ &= 2 \sum_{i=1}^n \left[y_i \log\left(\frac{y_i}{\bar{y}}\right) - (y_i - \bar{y}) \right] \end{aligned}$$

Example 3 (binomial distribution, continue):

$Y \sim \frac{B(m, p)}{m}, 0 \leq Y \leq 1 \Rightarrow \text{Binomial distribution in frequency.}$

Then,

$$\theta = \log\left(\frac{p}{1-p}\right) = \log\left(\frac{\mu}{1-\mu}\right),$$

$$b(\theta) = \log[1 + \exp(\theta)] = \log\left(\frac{1}{1-p}\right) = \log\left(\frac{1}{1-\mu}\right).$$

Therefore

$$\begin{aligned} D(y_1, y_2, \dots, y_n | \hat{\mu}) &= 2 \sum_{i=1}^n [y_i(\hat{\theta}_i - \hat{\theta}) + b(\hat{\theta}) - b(\hat{\theta}_i)] \\ &= 2 \sum_{i=1}^n \left\{ y_i \left[\log\left(\frac{\hat{\mu}_i}{1-\hat{\mu}_i}\right) - \log\left(\frac{\hat{\mu}}{1-\hat{\mu}}\right) \right] + \log\left(\frac{1}{1-\hat{\mu}}\right) - \log\left(\frac{1}{1-\hat{\mu}_i}\right) \right\} \\ &= 2 \sum_{i=1}^n \left\{ y_i \left[\log\left(\frac{y_i}{1-y_i}\right) - \log\left(\frac{\bar{y}}{1-\bar{y}}\right) \right] + \log\left(\frac{1}{1-\bar{y}}\right) - \log\left(\frac{1}{1-y_i}\right) \right\} \\ &= 2 \sum_{i=1}^n \left[y_i \log\left(\frac{y_i}{\bar{y}}\right) + (1-y_i) \log\left(\frac{1-y_i}{1-\bar{y}}\right) \right] \end{aligned}$$

Example 4 (gamma distribution, continue):

$Y \sim G(\mu, \nu)$. Then,

$$\theta = \frac{-1}{\mu}, b(\theta) = -\log(-\theta) = -\log\left(\frac{1}{\mu}\right).$$

Therefore

$$\begin{aligned} D(y_1, y_2, \dots, y_n | \hat{\mu}) &= 2 \sum_{i=1}^n [y_i(\hat{\theta}_i - \hat{\theta}) + b(\hat{\theta}) - b(\hat{\theta}_i)] \\ &= 2 \sum_{i=1}^n \left[y_i \left(\frac{-1}{\hat{\mu}_i} + \frac{1}{\hat{\mu}} \right) - \log\left(\frac{1}{\hat{\mu}}\right) + \log\left(\frac{1}{\hat{\mu}_i}\right) \right] \\ &= 2 \sum_{i=1}^n \left[y_i \left(\frac{-1}{y_i} + \frac{1}{\bar{y}} \right) - \log\left(\frac{1}{\bar{y}}\right) + \log\left(\frac{1}{y_i}\right) \right] \\ &= 2 \sum_{i=1}^n \left[\left(\frac{y_i - \bar{y}}{\bar{y}} \right) - \log\left(\frac{y_i}{\bar{y}}\right) \right] \end{aligned}$$

Example 5 (inverse Gaussian distribution, continue):

$Y \sim IG(\mu, \sigma^2)$. Then

$$\theta = \frac{-1}{2\mu^2}, b(\theta) = -\sqrt{-2\theta} = \frac{-1}{\mu},$$

Therefore

$$\begin{aligned} D(y_1, y_2, \dots, y_n | \hat{\mu}) &= 2 \sum_{i=1}^n [y_i(\hat{\theta}_i - \hat{\theta}) + b(\hat{\theta}) - b(\hat{\theta}_i)] \\ &= 2 \sum_{i=1}^n \left[y_i \left(\frac{-1}{2\hat{\mu}_i^2} + \frac{1}{2\hat{\mu}^2} \right) - \frac{1}{\hat{\mu}} + \frac{1}{\hat{\mu}_i} \right] \\ &= 2 \sum_{i=1}^n \left[y_i \left(\frac{-1}{2y_i} + \frac{1}{2\bar{y}^2} \right) - \frac{1}{\bar{y}} + \frac{1}{y_i} \right] \\ &= 2 \sum_{i=1}^n \left[\frac{y_i}{2\bar{y}^2} + \frac{1}{2y_i} - \frac{1}{\bar{y}} \right] \\ &= \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{y_i \bar{y}^2} \end{aligned}$$

General case:

The null model we commonly use to fit the data involves p parameters, $g(\mu_i) = x_i\beta, i = 1, 2, \dots, n$. Let $\tilde{\mu}_i = g^{-1}(x_i\hat{\beta})$, where $\hat{\beta}$ is the iterated reweighted least square estimate. Then, the deviance function

$$\begin{aligned} D(y_1, y_2, \dots, y_n | \tilde{\mu}_1, \tilde{\mu}_2, \dots, \tilde{\mu}_n) \\ = 2a(\phi)[l(y_1, y_2, \dots, y_n) - l(\tilde{\mu}_1, \tilde{\mu}_2, \dots, \tilde{\mu}_n)] \\ = 2 \sum_{i=1}^n [y_i(\hat{\theta}_i - \tilde{\theta}_i) + b(\tilde{\theta}_i) - b(\hat{\theta}_i)] \end{aligned}$$

where

$$\hat{\theta}_i = (\mathbf{b}')^{-1}(\hat{\mu}_i) = (\mathbf{b}')^{-1}(y_i)$$

and

$$\tilde{\theta}_i = (\mathbf{b}')^{-1}(\tilde{\mu}_i).$$

The deviance function can be used to assess the discrepancy of a fit with $g(\mu_i) = x_i\beta$. Intuitively, large deviance implies that the “distance” between the data y_i and the fitted value $\tilde{\mu}_i = g^{-1}(x_i\hat{\beta})$ is large. That is, the fit with $g(\mu_i) = x_i\beta$ might not be sensible.

The forms of the deviances for different distribution are given below:

Normal distribution:

$$\sum_{i=1}^n (y_i - \tilde{\mu}_i)^2$$

Poisson distribution:

$$2 \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{\tilde{\mu}_i} \right) - (y_i - \tilde{\mu}_i) \right]$$

Binomial distribution (in frequency):

$$2 \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{\tilde{\mu}_i} \right) + (1 - y_i) \log \left(\frac{1 - y_i}{1 - \tilde{\mu}_i} \right) \right]$$

Gamma distribution:

$$2 \sum_{i=1}^n \left[\left(\frac{y_i - \tilde{\mu}_i}{\tilde{\mu}_i} \right) - \log \left(\frac{y_i}{\tilde{\mu}_i} \right) \right]$$

Inverse Gaussian distribution:

$$\sum_{i=1}^n \frac{(y_i - \tilde{\mu}_i)^2}{y_i \tilde{\mu}_i^2}$$

Supplement: Deviance function

Let k be the difference between the number of parameters in the full model and the one in the null model. Then,

$$-2 \log \Lambda = \frac{\text{deviance function}}{a(\phi)} \approx \chi^2_k,$$

where

$$\Lambda = \frac{\text{likelihood in null model}}{\text{likelihood in full model}} \equiv \text{likelihood ratio}$$

and

$$\begin{aligned} \text{deviance function} &= 2a(\phi)[\log(\text{likelihood in full model}) \\ &\quad - \log(\text{likelihood in null model})] \end{aligned}$$

Case 1:

Full model: $\mu_i, i = 1, 2, \dots, n.$

Null model: $\mu \Rightarrow H_0: \mu_1 = \mu_2 = \dots = \mu_n = \mu.$

Then,

$$D(y_1, y_2, \dots, y_n | \hat{\mu}) \approx a(\phi) \chi^2_{n-1}.$$

Case 2: β is a $p \times 1$ vector of parameters.

Full model: $\mu_i, i = 1, 2, \dots, n.$

Null model: $g(\mu_i) = x_i \beta, i = 1, 2, \dots, n \Rightarrow H_0: g(\mu) = x \beta.$

Then,

$$D(y_1, y_2, \dots, y_n | \tilde{\mu}_1, \tilde{\mu}_2, \dots, \tilde{\mu}_n) \approx a(\phi) \chi^2_{n-p},$$

where $\tilde{\mu}_i = g^{-1}(x_i \hat{\beta}).$

Case 3 (most commonly used): β is a $p \times 1$ vector of parameters

$\tilde{\beta}$ is a $q \times 1$ vector of parameters.

Full model: $g(\mu_i) = x_i \beta, i = 1, 2, \dots, n.$

Null model: $g(\mu_i) = z_i \tilde{\beta}, i = 1, 2, \dots, n \Rightarrow H_0: g(\mu) = z \tilde{\beta}.$

Then,

$$D(\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_n | \tilde{\mu}_1, \tilde{\mu}_2, \dots, \tilde{\mu}_n) \approx a(\phi) \chi^2_{p-q},$$

where $\hat{\mu}_i = g^{-1}(x_i \hat{\beta})$ and $\tilde{\mu}_i = g^{-1}(z_i \tilde{\beta}).$