3.2 Likelihood functions:

1. Poisson distribution

The distribution of a Poisson random variable Y is

$$P(Y = y) = \frac{exp(-\mu)\mu^y}{y!}, y = 0, 1, 2, \cdots$$

The cumulant generating function is

$$\kappa_Y = \mu[exp(t) - 1].$$

Thus,

$$E(Y) = \kappa_T'(0) = [\mu \cdot exp(t)]_{t=0} = \mu$$

and

$$E(Y) = \kappa_T''(0) = [\mu \cdot exp(t)]_{t=0} = \mu.$$

2. The Poisson log-likelihood function

The Poisson log-likelihood function for independent $Y_i \sim P(\mu_i), i = 1, \cdots, n,$ is

$$l(\mu_1, \mu_2, \cdots, \mu_n) \propto \sum_{i=1}^{n} [y_i \cdot log(\mu_i) - \mu_i],$$

where $E(Y_i) = \mu_i$. The deviance function is

$$\begin{aligned} &D(y_1, y_2, \cdots, y_n | \tilde{\mu}_1, \tilde{\mu}_2, \cdots, \tilde{\mu}_n) \\ &= 2\sum_{i=1}^{n}\left[y_i log\left(\frac{y_i}{\tilde{\mu}_i}\right) - (y_i - \tilde{\mu}_i)\right] \\ &= 2\sum_{i=1}^{n}\left[y_i log\left(\frac{y_i}{\tilde{\mu}_i}\right)\right] - 2\sum_{i=1}^{n}(y_i - \tilde{\mu}_i) \end{aligned}$$

where $\tilde{\mu}_i = g^{-1}(x_i\hat{\beta})$ is the estimate of $E(Y_i) = \mu_i$.

Note:

If a constant term (the intercept) is included in the model, it can be shown that

$$\sum_{i=1}^{n}(y_i - \tilde{\mu}_i) = 0.$$

Thus, the deviance function can be reduced to

$$D(y_1, y_2, \cdots, y_n | \tilde{\mu}_1, \tilde{\mu}_2, \cdots, \tilde{\mu}_n) = 2\sum_{i=1}^{n}\left[y_i log\left(\frac{y_i}{\tilde{\mu}_i}\right)\right].$$

Note:

The deviance function is closely related to Pearson's statistic. Since

$$\frac{y_i - \tilde{\mu}_i}{\tilde{\mu}_i} = \varepsilon_i \Leftrightarrow \frac{y_i}{\tilde{\mu}_i} = 1 + \varepsilon_i \Leftrightarrow y_i = \tilde{\mu}_i(1 + \varepsilon_i),$$

then

$$\begin{aligned}
y_i \cdot log\left(\frac{y_i}{\tilde{\mu}_i}\right) - (y_i - \tilde{\mu}_i) &= \tilde{\mu}_i(1 + \varepsilon_i) \cdot log\left(\frac{y_i}{\tilde{\mu}_i}\right) - \tilde{\mu}_i \varepsilon_i \\
&= \tilde{\mu}_i[(1 + \varepsilon_i) \cdot log(1 + \varepsilon_i) - \varepsilon_i] \\
&= \tilde{\mu}_i\left[(1 + \varepsilon_i)\left(\varepsilon_i - \frac{\varepsilon_i^2}{2} + \cdots\right) - \varepsilon_i\right] \\
&= \tilde{\mu}_i\left[\left(\varepsilon_i - \frac{\varepsilon_i^2}{2} + \varepsilon_i^2 - \frac{\varepsilon_i^3}{2} + \cdots\right) - \varepsilon_i\right] \\
&= \tilde{\mu}_i\left(\frac{\varepsilon_i^2}{2} + \cdots\right) \\
&\approx \frac{\tilde{\mu}_i \varepsilon_i^2}{2},
\end{aligned}$$

where

$$log(1 + \varepsilon_i) = \varepsilon_i - \frac{\varepsilon_i^2}{2} + \cdots.$$

Thus, as the model includes the intercept,

$$\begin{aligned}
D(y_1, y_2, \cdots, y_n | \tilde{\mu}_1, \tilde{\mu}_2, \cdots, \tilde{\mu}_n) &= 2\sum_{i=1}^{n}\left[y_i log\left(\frac{y_i}{\tilde{\mu}_i}\right)\right] \\
&\approx \sum_{i=1}^{n}\frac{(y_i - \tilde{\mu}_i)^2}{\tilde{\mu}_i} \\
&\equiv \text{Pearson's statistic}
\end{aligned}$$

Note:

σ^2 can be estimated by

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

X^2 is the sum of Pearson's residuals, i.e., Pearson's statistic.