3.3 Comparison of two or more Poisson means:

Suppose that $Y_i \sim P(\mu_i), i = 1, \cdots, k,$ are independent Poisson random variables with $log(\mu_i) = \beta_0 + \beta_1 x_i$ and that we require to test the composite null hypothesis

$$\begin{aligned} &H_0: \mu_1 = \mu_2 = \cdots = \mu_k = exp(\beta_0) \\ \Leftrightarrow\ &H_0: log(\mu_1) = log(\mu_2) = \cdots = log(\mu_k) = \beta_0 \\ \Leftrightarrow\ &H_0: \beta_1 = 0 \end{aligned}$$

The alternative hypotheses under consideration are $H_1: \beta_1 > 0$, or $H_1: \beta_1 < 0$ or $H_1: \beta_1 \neq 0$. Denote $Y = (Y_1, Y_2, \cdots, Y_k)$. Standard theory of significance testing leads to consideration of the test statistic

$$T(Y) = \ T(Y_1, Y_2, \cdots, Y_k) = \sum_{i=1}^{k} x_i Y_i$$

conditionally on the observed value of

$$m(Y) = m(Y_1, Y_2, \cdots, Y_k) = \sum_{i=1}^{k} Y_i$$

which is the sufficient statistic for β_0. For example, as $H_a: \beta_1 > 0$, the test is

$$\begin{aligned} &\text{reject } H_0 \text{ as } T(y) > c_0(m) \\ &\text{reject } H_0 \text{ with probability } w(m) \text{ as } T(y) = c_0(m) \\ &\text{not reject } H_0 \text{ as } T(y) < c_0(m) \end{aligned}$$

given the data

$$\mathrm{y} = (y_1, y_2, \cdots, y_k), m = \sum_{i=1}^{k} y_i,$$

where $c_0(m)$ and $w(m)$ can be obtained by solving

$$P(T(Y) > c_0(m) | m(Y) = m) + w(m) P(T(Y) = c_0(m) | m(Y) = m) = \alpha$$

under $H_0, i.e., \beta_1 = 0$ and $w(m)$ is some constant depending on m.

Note that under the null hypothesis, we regard the data as having multinomial distribution with index m and parameter vector $\left({}^{1}/_{k}, {}^{1}/_{k}, \cdots, {}^{1}/_{k}\right)$ independent of β_0, i.e.,

$$P\left(Y_1 = y_1, Y_2 = y_2, \cdots, Y_k = y_k | m(Y) = \sum_{i=1}^{k} Y_i = m\right)$$

$$= \frac{m!}{\prod_{i=1}^{k} y_i!} \prod_{i=1}^{k} \left(\frac{1}{k}\right)^{y_i} = \frac{m!}{\prod_{i=1}^{k} y_i!} \left(\frac{1}{k}\right)^{\sum_{i=1}^{k} y_i}$$

$$= \frac{m!}{\prod_{i=1}^{k} y_i!} \frac{1}{k^m}$$

Note:

Let $Y = (Y_1, Y_2, \cdots, Y_k)$ has the probability density function or probability distribution function

$$f(y) = f(y_1, y_2, \cdots, y_k) = c(\theta, \eta) exp\left[\theta T(y) + \sum_{j=1}^{r} \eta_j m_j(y)\right] h(y).$$

Let

$$m(y) = \big(m_1(y), m_2(y), \cdots, m_r(y)\big).$$

Then, for testing $H_0: \theta \le \theta_0 \;\; vs \;\; H_1: \theta > \theta_0$, an UMP (uniformly most powerful) unbiased level-test given $m(Y) = m$ is

$$\begin{aligned} &\text{reject } H_0 \text{ as } T(y) > c_0(m) \\ &\text{reject } H_0 \text{ with probability } w(m) \text{ as } T(y) = c_0(m) \\ &\text{not reject } H_0 \text{ as } T(y) < c_0(m) \end{aligned}$$

where $c_0(m)$ and $w(m)$ can be obtained by solving

$$P(T(Y) > c_0(m) | m(Y) = m) + w(m) P(T(Y) = c_0(m) | m(Y) = m) = \alpha$$

under the null hypothesis.

Therefore, for independent Poisson random variables $Y_i \sim P(\mu_i), i = 1, \cdots, k,$ with $log(\mu_i) = \beta_0 + \beta_1 x_i$, the probability distribution function is

$$f(y) = \prod_{i=1}^{k} \frac{\mu_i^{y_i} \cdot exp(-\mu_i)}{y_i!} = exp\left(-\sum_{i=1}^{k} \mu_i\right) exp\left[\sum_{i=1}^{k} log(\mu_i^{y_i})\right] \cdot \frac{1}{\prod_{i=1}^{k} y_i!}$$

$$= exp\left(-\sum_{i=1}^{k} \mu_i\right) exp\left[\sum_{i=1}^{k} y_i \cdot log(\mu_i)\right] \cdot \frac{1}{\prod_{i=1}^{k} y_i!}$$

$$= exp\left(-\sum_{i=1}^{k} \mu_i\right) exp\left[\sum_{i=1}^{k} y_i(\beta_0 + \beta_1 x_i)\right] \cdot \frac{1}{\prod_{i=1}^{k} y_i!}$$

$$= exp\left(-\sum_{i=1}^{k} \mu_i\right) exp\left[\beta_1 \sum_{i=1}^{k} x_i y_i + \beta_0 \sum_{i=1}^{k} y_i\right] \cdot \frac{1}{\prod_{i=1}^{k} y_i!}$$

$$= c(\beta_0, \beta_1) exp[\beta_1 T(y) + \beta_0 m(y)] h(y)$$

where

$$c(\beta_0, \beta_1) = exp\left(-\sum_{i=1}^{k} \mu_i\right), T(y) = \sum_{i=1}^{k} x_i y_i\,, m(y) = \sum_{i=1}^{k} y_i\,, h(y) = \frac{1}{\prod_{i=1}^{k} y_i!}.$$

Note:

The Poisson log-likelihood function for (β_0, β_1) in this problem is

$$l(\beta_0, \beta_1) \propto \sum_{i=1}^{k} [y_i \cdot log(\mu_i) - \mu_i]$$

$$= \sum_{i=1}^{k} [y_i(\beta_0 + \beta_1 x_i) - exp(\beta_0 + \beta_1 x_i)]$$

$$= \beta_0 \sum_{i=1}^{k} y_i + \beta_1 \sum_{i=1}^{k} x_i y_i - \sum_{i=1}^{k} exp(\beta_0 + \beta_1 x_i)$$

Denote

$$\tau = \sum_{i=1}^{k} exp(\beta_0 + \beta_1 x_i) = \sum_{i=1}^{k} \mu_i.$$

Then, the log-likelihood for (τ, β_1) becomes

$$l(\beta_0, \beta_1) \propto \beta_0 \sum_{i=1}^{k} y_i + \beta_1 \sum_{i=1}^{k} x_i y_i - \sum_{i=1}^{k} exp(\beta_0 + \beta_1 x_i)$$

$$= \beta_0 m + \beta_1 \sum_{i=1}^{k} x_i y_i - \tau$$

$$= m \cdot log(\tau) - \tau + \beta_1 \sum_{i=1}^{k} x_i y_i - [m \cdot log(\tau) - \beta_0 m]$$

$$= m \cdot log(\tau) - \tau + \beta_1 \sum_{i=1}^{k} x_i y_i - m \cdot log\left[\frac{\tau}{exp(\beta_0)}\right]$$

$$= m \cdot log(\tau) - \tau + \beta_1 \sum_{i=1}^{k} x_i y_i - m \cdot log\left[\sum_{i=1}^{k} exp(\beta_1 x_i)\right]$$

$$= l_{m(Y)}(\tau) + l_{Y|m(Y)}(\beta_1)$$

where

$$l_{m(Y)}(\tau) \propto \; m \cdot log(\tau) - \tau$$

is the Poisson log-likelihood based on

$$m = \sum_{i=1}^{k} y_i = m(Y) \sim P(\tau)$$

and

$$l_{Y|m(Y)}(\beta_1) \propto \beta_1 \sum_{i=1}^{k} x_i y_i - m \cdot log\left[\sum_{i=1}^{k} exp(\beta_1 x_i)\right]$$

is the multinomial log-likelihood for β_1 **based on conditional distribution,**

$$Y|m(Y) = m \equiv Y_1, Y_2, \cdots, Y_k|m(Y) = m \sim M(m, \pi_1, \cdots, \pi_k),$$

where $M(m, \pi_1, \cdots, \pi_k)$ **is a multinomial distribution with index** m **and parameters**

$$\pi_j = \frac{exp(\beta_1 x_j)}{\sum_{i=1}^{k} exp(\beta_1 x_i)}, j = 1, \cdots, k.$$