Multinomial Distribution

Table 1. The multinomial distribution as a statistical model for the vector of observed proportions $\mathbf{y}$, given the predicted proportions $\mathbf{p}$ and the sample size n.

$$\mathbf{v} \sim \mathcal{M}(\mathbf{p}, n) \tag{T1.1}$$

$$\mathbf{y} = \frac{\mathbf{v}}{\sum_{j=1}^{g} v_j} = \frac{\mathbf{v}}{n} \tag{T1.2}$$

$$P(\mathbf{y} \mid \mathbf{p}, n) = \frac{n!}{(ny_1)!(ny_2)!\ldots(ny_g)!} \prod_{i=1}^{g} p_i^{ny_i} \tag{T1.3}$$

$$\ell(\mathbf{p} \mid \mathbf{y}, n) = -n \sum_{i=1}^{g} y_i \log p_i \tag{T1.4}$$

$$\mathrm{E}[y_i] = p_i \tag{T1.5}$$

$$\mathrm{Var}[y_i] = \frac{p_i(1-p_i)}{n} \tag{T1.6}$$

$$\mathrm{Cov}[y_i, y_j] = -\frac{p_i p_j}{n} \quad \text{for } i \neq j \tag{T1.7}$$

Dirichlet Distribution

Table 2. The Dirichlet distribution as a statistical model for the vector of observed proportions $\mathbf{y}$, given the predicted proportions $\mathbf{p}$ and an effective sample size n. The model applies when $\mathbf{y}$ comes from the composition (T2.2) of independent variates (T2.1) drawn from gamma distributions with a common scale parameter $n > 0$. The approximate modal estimate $\hat{n}$ in (T2.5) depends on Stirling's approximation to the gamma function.

$$v_i \sim \mathcal{G}(\text{shape} = np_i, \text{scale} = n) \tag{T2.1}$$

$$y_i = \frac{v_i}{\sum_{j=1}^{g} v_j} \tag{T2.2}$$

$$P(\mathbf{y} \,|\, \mathbf{p}, n) = \frac{\Gamma(n)}{\Gamma(np_1)\Gamma(np_2)\ldots\Gamma(np_g)} \prod_{i=1}^{g} y_i^{np_i - 1} \tag{T2.3}$$

$$\ell(\mathbf{p}, n \,|\, \mathbf{y}) = \sum_{i=1}^{g} \left[\log\Gamma(np_i) - np_i \log y_i \right] - \log\Gamma(n) \tag{T2.4}$$

$$\hat{n} \approx \frac{g-1}{2} \left(\sum_{i=1}^{g} p_i \log \frac{p_i}{y_i} \right)^{-1} \tag{T2.5}$$

$$\mathrm{E}[y_i] = p_i \tag{T2.6}$$

$$\mathrm{Var}[y_i] = \frac{p_i(1-p_i)}{n+1} \tag{T2.7}$$

$$\mathrm{Cov}[y_i, y_j] = -\frac{p_i p_j}{n+1} \quad \text{for } i \neq j \tag{T2.8}$$

Logistic-Normal Distribution

Table 3. The logistic-normal distribution as a statistical model for the vector of observed proportions $\mathbf{y}$, given the predicted proportions $\mathbf{p}$ and a standard deviation σ. The model applies when $\mathbf{y}$ comes from the logistic transformation (T3.2) of independent variates (T3.1) drawn from normal distributions with a common standard deviation σ. Calculations involve the geometric means $\tilde{y}$ and $\tilde{p}$ of $\mathbf{y}$ and $\mathbf{p}$, respectively. The modal estimate $\hat{\sigma}$ in (T3.5) is exact. The approximations (T3.9)–(T3.11) apply when σ is small.

$$u_i = \log p_i + \sigma\varepsilon_i \quad \text{where } \varepsilon_i \sim \mathcal{N}(0,1) \tag{T3.1}$$

$$y_i = \frac{e^{u_i}}{\sum_{j=1}^{g} e^{u_j}} \tag{T3.2}$$

$$P(\mathbf{y}\,|\,\mathbf{p},\sigma) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{g-1}\left(\sqrt{g}\prod_{i=1}^{g} y_i\right)^{-1}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{g}\left(\log\frac{y_i}{\tilde{y}}-\log\frac{p_i}{\tilde{p}}\right)^2\right] \tag{T3.3}$$

$$\ell(\mathbf{p},\sigma\,|\,\mathbf{y}) = (g-1)\log\sigma + \frac{1}{2\sigma^2}\sum_{i=1}^{g}\left(\log\frac{y_i}{\tilde{y}}-\log\frac{p_i}{\tilde{p}}\right)^2 \tag{T3.4}$$

$$\hat{\sigma}^2 = \frac{1}{g}\sum_{i=1}^{g}\left(\log\frac{y_i}{\tilde{y}}-\log\frac{p_i}{\tilde{p}}\right)^2 \tag{T3.5}$$

$$\mathrm{E}[\log(y_i/y_j)] = \log(p_i/p_j) \tag{T3.6}$$

$$\mathrm{Var}[\log(y_i/y_j)] = 2\sigma^2 \quad \text{for } i \neq j \tag{T3.7}$$

$$\mathrm{Cov}[\log(y_i/y_k), \log(y_j/y_k)] = \sigma^2 \quad \text{for } i \neq j \neq k \neq i \tag{T3.8}$$

$$\mathrm{E}[y_i] \approx p_i \tag{T3.9}$$

$$\mathrm{Var}[y_i] \approx \sigma^2 p_i^2\left(1 - 2p_i + \sum_{i=1}^{g} p_i^2\right) \tag{T3.10}$$

$$\mathrm{Cov}[y_i, y_j] \approx -\sigma^2 p_i p_j\left(p_i + p_j - \sum_{i=1}^{g} p_i^2\right) \quad \text{for } i \neq j \tag{T3.11}$$
