Likelihood

1. Likelihood $L(\theta) = \prod_{i=1}^{n} f(y_i;\theta)$

   Log likelihood $l(\theta; y) = \log L(\theta) = \sum_{i=1}^{n} \log f(y_i;\theta)$

   Score $U(\theta) = l'(\theta)$, MLE $U\left(\hat{\theta}\right) = 0$

   Observed information $J(\theta) = -\frac{\mathrm{d}U}{\mathrm{d}\theta} = -\frac{\mathrm{d}^2 l}{\mathrm{d}\theta^2}$

   Fisher's information $I(\theta) = E[J(\theta)] = -E\left[\frac{\mathrm{d}^2 l}{\mathrm{d}\theta^2}\right]$

   $E[U] = 0$, $Var(U) = E[U^2] - E[U]^2 = E[U^2] = -E[U'] = I$

   $Var\left(\hat{\theta}\right) \approx \frac{1}{I\left(\hat{\theta}\right)}$, $ESE\left(\hat{\theta}\right) = \sqrt{Var\left(\hat{\theta}\right)}$

2. When θ is a p-dimensional parameter,
   $\mathbf{U}(\theta)$ is a vector, $\mathbf{I}(\theta)$ and $\mathbf{J}(\theta)$ are matrices.
   Let $p = 2$, $Var(\hat{\theta}) = I^{-1}(\theta_1, \theta_2) = \begin{pmatrix} Var\left(\hat{\theta}_1\right) & cov\left(\hat{\theta}_1, \hat{\theta}_2\right) \\ cov\left(\hat{\theta}_1, \hat{\theta}_2\right) & Var\left(\hat{\theta}_2\right) \end{pmatrix}$

3. Score test: $\frac{U(\theta)}{\sqrt{I(\theta)}} \sim N(0,1)$ and $\frac{U^2(\theta)}{I(\theta)} \sim \chi_1^2$

   Wald test: $\frac{\hat{\theta} - \theta_0}{\sqrt{I^{-1}(\theta_0)}} \sim N(0,1)$ and $\frac{\left(\hat{\theta} - \theta_0\right)^2}{I^{-1}(\theta_0)} \sim \chi_1^2$

   Likelihood Ratio test: $LR = \frac{L(\theta_0)}{L(\hat{\theta})}$, $-2\log(LR) = -2(l(\theta_0) - l(\hat{\theta}))$, $2\left(l\left(\hat{\theta}\right) - l(\theta_0)\right) \sim \chi_1^2$

   `qnorm(0.975,0,1)=1.96`, `qchisq(0.95,1)=3.84`

4. Fisher's method of scoring $\theta_{r+1} = \theta_r + \frac{U(\theta_r)}{I(\theta_r)}$

   Multiparameter case: $\theta_{r+1} = \theta_r + I^{-1}(\theta_r)U(\theta_r)$

5. Exponential family of distributions $f(y;\theta) = \exp\{a(y)b(\theta) + c(\theta) + d(y)\}$
   $E\{a(Y)\} = -\frac{c'(\theta)}{b'(\theta)}$, $Var\{a(Y)\} = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{\{b'(\theta)\}^3}$

6. Generalised Linear Models
   Link function $\mu_i = E(Y_i)$, $g(\mu_i) = \mathbf{x}_i^T \beta$
   Exponential family: $f(y;\theta) = \exp\{yb(\theta) + c(\theta) + d(y)\}$
   Canonical link: $g(\mu_i) = b(\theta_i)$
   Deviance: $D = -2\left\{l(\hat{\beta}) - l(\text{maximal model}, \Omega)\right\}$, where $\hat{\mu}_i = y_i$ in model Ω
   Pearson residuals: $r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$, where $V(\mu_i) = Var(Y_i)$