STA2101: Likelihood Cheatsheet

$Y = Y_1, \ldots, Y_n$ independently distributed with densities $f(y_i \mid x_i; \theta), \theta \in \Theta \subset \mathbb{R}^p; y_i \in \mathbb{R}$. The observations are independent, but not identically distributed, due to the dependence on the $p \times 1$ vector x_i. Independence is critical, but i.d. can usually be handled, so the dependence on x_i below is often suppressed.

Likelihood function is the joint probability of the observations, considered as a function of the parameter

$$L(\theta; y) \propto \prod_{i=1}^{n} f(y_i \mid x_i; \theta)$$

Log-likelihood function

$$\ell(\theta; y) = \log L(\theta; y) = \sum_{i=1}^{n} \log f(y_i \mid x_i; \theta)$$

maximum likelihood estimator (MLE)

$$\hat{\theta} = \hat{\theta}(\boldsymbol{x}) = \arg\sup_{\theta} \ell(\theta)$$

score function and score equation

$$\ell'(\theta; y) = \frac{\partial}{\partial \theta} \ell(\theta; y) \qquad \ell'(\hat{\theta}; y) = 0$$

observed Fisher information

$$j(\hat{\theta}) = -\left. \frac{\partial^2}{\partial \theta \partial \theta^T} \ell(\theta; y) \right|_{\theta = \hat{\theta}}$$

Wald statistic (standardized MLE)

$$\hat{\theta} \,\dot{\sim}\, N_d[\theta_0, \{j(\hat{\theta})\}^{-1}], \qquad \text{under sampling from } f(y; \theta_0)$$

$$\hat{\theta}_k \,\dot{\sim}\, N[\theta_{0k}, \{j(\hat{\theta})\}^{-1}_{kk}], \qquad \text{under sampling from } f(y; \theta_0)$$

(log)-likelihood ratio statistic

$$w(\theta) = 2\{\ell(\hat{\theta}; y) - \ell(\theta_0; y)\} \,\dot{\sim}\, \chi^2_p, \qquad \text{under sampling from } f(y; \theta_0)$$

These approximations come from asymptotic theory of inference, $n \to \infty, p$ fixed.

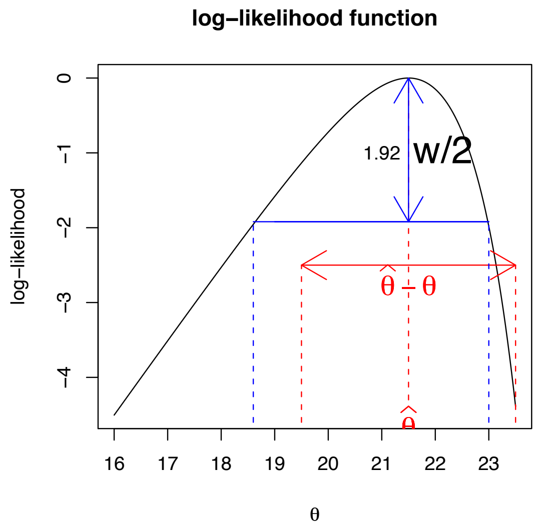

Figure 1: This is the log-likelihood function for θ with a single observation $y = 21.5$ from the density $f(y;\theta) = \exp\{-(y-\theta) - e^{(y-\theta)}\}$.

Profile log-likelihood function $\ell_{\rm p}(\psi)$: Assume $\theta = (\psi, \lambda), \psi \in \mathbb{R}^q, \lambda \in \mathbb{R}^{p-q}$.

$$\ell_{\rm p}(\psi) = \ell(\psi, \hat\lambda_\psi), \quad \hat\lambda_\psi = \arg\sup_\lambda \ell(\psi, \lambda)$$

Wald statistic

$$\hat\psi \mathrel{\dot\sim} N_q\{\psi, j_{\rm p}(\hat\psi)^{-1}\}, \quad j_{\rm p}(\psi) = -\ell''_{\rm p}(\psi), \qquad \text{under sampling from } f(y;\psi,\lambda)$$

(profile) (log)-likelihood ratio statistic

$$w(\psi) = 2\{\ell_{\rm p}(\hat\psi) - \ell_{\rm p}(\psi)\} = 2\{\ell(\hat\psi, \hat\lambda) - \ell(\psi, \hat\lambda_\psi)\} \mathrel{\dot\sim} \chi^2_q, \qquad \text{under sampling from } f(y;\psi,\lambda)$$

This means we can use the same approximations with the profile log-likelihood function as with the log-likelihood function. On the next page are two plots of a bivariate log-likelihood function. Can you create a plot of the profile log-likelihood function for the log-odds ratio parameter $\psi = \log[p_1(1-p_2)/\{(1-p_1)p_2\}]$?

Figure 2: log-likelihood function for (p_1, p_2) in a model for two independent binomial observations, with $n_1 = n_2 = 200, y_1 = 160, y_2 = 148$.