Topic 15

Maximum Likelihood Estimation

Multidimensional Estimation

Outline

Fisher Information

For a multidimensional parameter space $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$, the Fisher information $I(\theta)$ is a matrix. As with one-dimensional case, the ij-th entry has two alternative expressions, namely,

$$I(\theta)_{ij} = E_\theta \left[\frac{\partial}{\partial \theta_i} \ln L(\theta|X) \frac{\partial}{\partial \theta_j} \ln L(\theta|X) \right] = -E_\theta \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln L(\theta|X) \right].$$

Rather than taking reciprocals to obtain an estimate of the variance, we find the matrix inverse $I(\theta)^{-1}$.

- The diagonal entries of $I(\theta)^{-1}$ gives estimates of variances.
- The off-diagonal entries of $I(\theta)^{-1}$ give estimates of covariances.

Fisher Information

To be precise, for n observations, let $\hat{\theta}_{i,n}(X)$ be the maximum likelihood estimator of the i-th parameter. Then

$$\mathrm{Var}_\theta(\hat{\theta}_{i,n}(X)) \approx \frac{1}{n} I(\theta)^{-1}_{ii} \qquad \mathrm{Cov}_\theta(\hat{\theta}_{i,n}(X), \hat{\theta}_{j,n}(X)) \approx \frac{1}{n} I(\theta)^{-1}_{ij}.$$

When the i-th parameter is θ_i, the asymptotic normality and efficiency can be expressed by noting that the z-score

$$Z_{i,n} = \frac{\hat{\theta}_i(X) - \theta_i}{\sqrt{I(\theta)^{-1}_{ii}/n}}.$$

is approximately a standard normal. As we saw in one dimension, we can replace the information matrix with the observed information matrix,

$$J(\hat{\theta})_{ij} = -\frac{\partial^2}{\partial\theta_i \partial\theta_j} \ln L(\hat{\theta}(X)|X).$$

Distribution of Fitness Effects

We return to the model of the gamma distribution for the distribution of fitness effects of deleterious mutations. To obtain the maximum likelihood estimate for the gamma family of random variables, write the likelihood

$$\begin{aligned} \mathbf{L}(\alpha, \beta|\mathbf{x}) &= \left(\frac{\beta^\alpha}{\Gamma(\alpha)} x_1^{\alpha-1} e^{-\beta x_1}\right) \cdots \left(\frac{\beta^\alpha}{\Gamma(\alpha)} x_n^{\alpha-1} e^{-\beta x_n}\right) \\ &= \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^n (x_1 x_2 \cdots x_n)^{\alpha-1} e^{-\beta(x_1+x_2+\cdots+x_n)} . \end{aligned}$$

and its logarithm

$$\ln \mathbf{L}(\alpha, \beta|\mathbf{x}) = n(\alpha \ln \beta - \ln \Gamma(\alpha)) + (\alpha - 1) \sum_{i=1}^n \ln x_i - \beta \sum_{i=1}^n x_i .$$

The score function is a vector $\left(\frac{\partial}{\partial \alpha} \ln \mathbf{L}(\alpha, \beta|\mathbf{x}), \frac{\partial}{\partial \beta} \ln \mathbf{L}(\alpha, \beta|\mathbf{x})\right)$.

Gamma Distribution

$$\ln \mathbf{L}(\alpha, \beta|\mathbf{x}) = n(\alpha \ln \beta - \ln \Gamma(\alpha)) + (\alpha - 1)\sum_{i=1}^{n} \ln x_i - \beta \sum_{i=1}^{n} x_i.$$

The zeros of the components of the score function determine the maximum likelihood estimators. Thus, to determine these parameters, we solve the equations

$$\frac{\partial}{\partial \alpha} \ln \mathbf{L}(\hat{\alpha}, \hat{\beta}|\mathbf{x}) = n(\ln \hat{\beta} - \frac{d}{d\alpha} \ln \Gamma(\hat{\alpha})) + \sum_{i=1}^{n} \ln x_i = 0$$

and $$\frac{\partial}{\partial \beta} \ln \mathbf{L}(\hat{\alpha}, \hat{\beta}|\mathbf{x}) = n\frac{\hat{\alpha}}{\hat{\beta}} - \sum_{i=1}^{n} x_i = 0, \quad \text{or} \quad \bar{x} = \frac{\hat{\alpha}}{\hat{\beta}}.$$

Substituting $\hat{\beta} = \hat{\alpha}/\bar{x}$ into the first equation results the following relationship for $\hat{\alpha}$.

$$n(\ln \hat{\alpha} - \ln \bar{x} - \frac{d}{d\alpha} \ln \Gamma(\hat{\alpha}) + \overline{\ln x}) = 0$$

Gamma Distribution

This can be solved numerically. The derivative of the logarithm of the gamma function

$$\psi(\alpha) = \frac{d}{d\alpha} \ln \Gamma(\alpha)$$

is know as the digamma function and is called in R with `digamma`.

For the example for the distribution of fitness effects in humans, a simulated data set (`rgamma(500,0.19,5.18)`) yields $\hat{\alpha} = 0.2006$ and $\hat{\beta} = 5.806$ for maximum likelihood estimates.

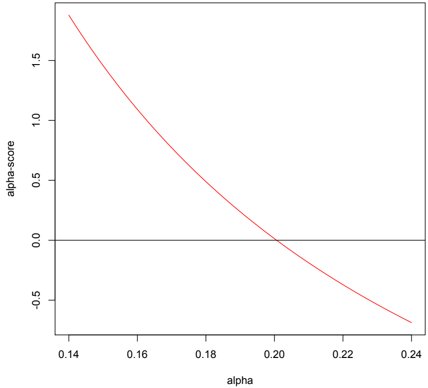

Figure: $\ln \hat{\alpha} - \ln \bar{x} - \frac{d}{d\alpha} \ln \Gamma(\hat{\alpha}) + \overline{\ln x_i}$ crosses the horizontal axis at $\hat{\alpha} = 0.2006$.

Gamma Distribution

Exercise. To determine the variance of these estimators, compute the appropriate second derivatives.

$$I(\alpha,\beta)_{11} = -\frac{\partial^2}{\partial\alpha^2}\ln \mathbf{L}(\alpha,\beta|\mathbf{x}) = n\frac{d^2}{d\alpha^2}\ln\Gamma(\alpha), \quad I(\alpha,\beta)_{22} = -\frac{\partial^2}{\partial\beta^2}\ln \mathbf{L}(\alpha,\beta|\mathbf{x}) = n\frac{\alpha}{\beta^2},$$

$$I(\alpha,\beta)_{12} = -\frac{\partial^2}{\partial\alpha\partial\beta}\ln \mathbf{L}(\alpha,\beta|\mathbf{x}) = -n\frac{1}{\beta}.$$

This give a Fisher information matrix

$$I(\alpha,\beta) = n\begin{pmatrix} \frac{d^2}{d\alpha^2}\ln\Gamma(\alpha) & -\frac{1}{\beta} \\ -\frac{1}{\beta} & \frac{\alpha}{\beta^2} \end{pmatrix} \quad I(0.19, 5.18) = 500\begin{pmatrix} 28.983 & -0.193 \\ -0.193 & 0.007 \end{pmatrix}.$$

NB. $\psi_1(\alpha) = d^2 \ln\Gamma(\alpha)/d\alpha^2$ is known as the trigamma function and is called in R with `trigamma`.

Gamma Distribution

The inverse matrix

$$I(\alpha, \beta)^{-1} = \frac{1}{500} \begin{pmatrix} 0.0422 & 1.1494 \\ 1.1494 & 172.5587 \end{pmatrix}.$$

Thus,

$$\text{Var}(\hat{\alpha}) \approx 8.432 \times 10^{-5} \qquad \sigma_{\hat{\alpha}} \approx 0.00918$$

$$\text{Var}(\hat{\beta}) \approx 0.3451 \qquad \sigma_{\hat{\beta}} \approx 0.5875$$

Compare this with the method of moments estimators

$$\sigma_{\hat{\alpha}} \approx 0.02838 \qquad \sigma_{\hat{\beta}} \approx 0.9769$$

Exercise. Estimate the correlation $\rho(\hat{\alpha}, \hat{\beta})$.

Gamma Distribution

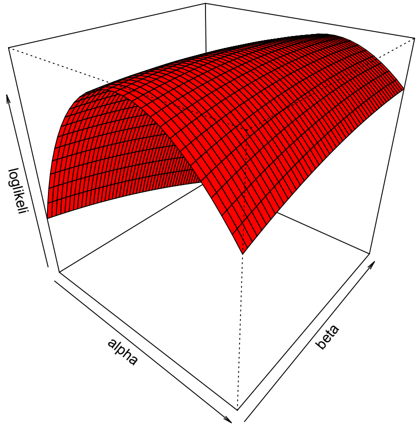

Figure: The log-likelihood surface. The domain is $0.14 \leq \alpha \leq 0.24$ and $5 \leq \beta \leq 7$

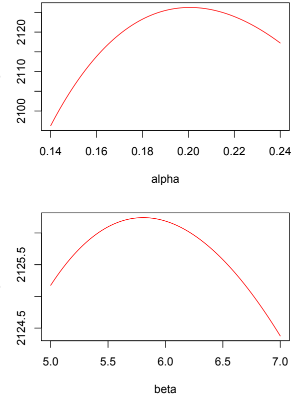

Figure: Graphs of vertical slices through the log-likelihood function surface through the MLE. (top) $\hat{\beta} = 5.806$ (bottom) $\hat{\alpha} = 0.20066$.