BCB 568 Homework Assignment 3

Schuyler Smith

February 22, 2018

Asymptotic results for multinomial probabilities.

a. Since $\sum_{i=1}^m p_i = 1$ and $g(\boldsymbol{p}) = 0$, we have $g(\boldsymbol{p}) = 1 - \sum_{i=1}^m p_i$. Since $P(\boldsymbol{x}) = \frac{n!}{x_1!x_2!\dots x_m!} p_1^{x_1} p_2^{x_2} \dots p_m^{x_m} = n! \prod_{i=1}^m \frac{p_i^{x_i}}{x_i}$, the log-likelihood $l(\boldsymbol{p}|\boldsymbol{x})$ is given by

$$\ln l(\boldsymbol{p}|\boldsymbol{x}) = \ln \left(n! \prod_{i=1}^m \frac{p_i^{x_i}}{x_i} \right) = \ln n! + \sum_{i=1}^m x_i \ln p_i - \sum_{i=1}^m \ln x_i!.$$

Taking the partial derivative of $l(\boldsymbol{p}|\boldsymbol{x}) - \lambda g(\boldsymbol{p}) = 0$, we have

$$\begin{aligned} 0 &= \frac{\partial}{\partial p_i} \Big[l(\boldsymbol{p}|\boldsymbol{x}) - \lambda g(\boldsymbol{p}) \Big] = \frac{\partial}{\partial p_i} \left[\ln n! + \sum_{i=1}^m x_i \ln p_i - \sum_{i=1}^m \ln x_i! - \lambda + \lambda \sum_{i=1}^m p_i \right] \\ &= \frac{x_i}{p_i} - \lambda. \end{aligned}$$

So $p_i = \frac{x_i}{\lambda}$.

$$1 = \sum_{i=1}^m p_i = \sum_{i=1}^m \frac{x_i}{\lambda} = \frac{n}{\lambda},$$

showing that $\lambda = n$. Thus, our MLE $\hat{p}_i = \frac{x_i}{n}$.

b. See included R-script and Figures.

c. Testing significant difference with $\alpha = 0.05$ we rejected the hypotheses that they are significantly different.

Bootstrap results for multinomial probabilities.

a. Computed in included R-code.

b. Computed in included R-code.

Parametric bootstrap tests of a Poisson clustering model.

a. Let $\boldsymbol{z} = (z_1, z_2, \dots, z_m)$ and $\boldsymbol{x} = (x_1, x_2, \dots, x_m)$. Then the likelihood L is given by

$$L(\lambda_1, \lambda_2, \dots, \lambda_k, \boldsymbol{z}|\boldsymbol{x}) = P(\boldsymbol{x}|\lambda_1, \lambda_2, \dots, \lambda_k, \boldsymbol{z}) = \prod_{i=1}^{m} P(x_i|\lambda_1, \lambda_2, \dots, \lambda_k, \boldsymbol{z}) = \prod_{i=1}^{m} \frac{e^{-\lambda_{z_i}} \lambda_{z_i}^{x_i}}{x_i!}. \tag{1}$$

The log-likelihood LL is then given by

$$\begin{aligned} LL(\lambda_1, \lambda_2, \dots, \lambda_k, \boldsymbol{z}|\boldsymbol{x}) &= \ln L(\lambda_1, \lambda_2, \dots, \lambda_k, \boldsymbol{z}|\boldsymbol{x}) = \ln\left(\prod_{i=1}^{m} \frac{e^{-\lambda_{z_i}} \lambda_{z_i}^{x_i}}{x_i!}\right) = \sum_{i=1}^{m} \ln\left(\frac{e^{-\lambda_{z_i}} \lambda_{z_i}^{x_i}}{x_i!}\right) \\ &= \sum_{i=1}^{m} \Big[\ln e^{-\lambda_{z_i}} + \ln \lambda_{z_i}^{x_i} - \ln x_i!\Big] = \sum_{i=1}^{m} \Big[-\lambda_{z_i} + x_i \ln \lambda_{z_i} - \ln x_i!\Big]. \end{aligned} \tag{2}$$

b. R-code.

c. We talked about boundaries. Our value of K may lie on the boundary of Θ, since it is possible that our number of genomes is equaly to our number of groups, so we cannot meet the necesary conditions. If they were able to be met we could have the asymptotic distribution of the likelihood ratio test statistic $-2\ln(\Lambda)$ will be distributed according to χ_1^2.

d.