CHAPTER 5

Inference

5.8 Exercises

5.1 Consider the single response variable Y with $Y \sim \text{Bin}(n, \pi)$.

(a) Find the Wald statistic $(\widehat{\pi} - \pi)^T \mathfrak{I}(\widehat{\pi} - \pi)$, where $\widehat{\pi}$ is the maximum likelihood estimator of π and $\mathfrak{I}$ is the information.

(b) Verify that the Wald statistic is the same as the score statistic $U^T \mathfrak{I}^{-1} U$ in this case (see Example 5.2.2).

(c) Find the deviance

$$2[l(\widehat{\pi}; y) - l(\pi; y)].$$

(d) For large samples, both the Wald/score statistic and the deviance approximately have the $\chi^2(1)$ distribution. For $n = 10$ and $y = 3$, use both statistics to assess the adequacy of the models:

(i) $\pi = 0.1$; (ii) $\pi = 0.3$; (iii) $\pi = 0.5$.

Do the two statistics lead to the same conclusions?

5.2 Consider a random sample $Y_1, \ldots, Y_N$ with the exponential distribution

$$f(y_i; \theta_i) = \theta_i \exp(-y_i \theta_i).$$

Derive the deviance by comparing the maximal model with different values of θ_i for each Y_i and the model with $\theta_i = \theta$ for all i.

5.3 Suppose $Y_1, \ldots, Y_N$ are independent identically distributed random variables with the Pareto distribution with parameter θ.

(a) Find the maximum likelihood estimator $\widehat{\theta}$ of θ.

(b) Find the Wald statistic for making inferences about θ (Hint: Use the results from Exercise 3.10).

(c) Use the Wald statistic to obtain an expression for an approximate 95% confidence interval for θ.

(d) Random variables Y with the Pareto distribution with the parameter θ can be generated from random numbers U, which are uniformly distributed between 0 and 1 using the relationship $Y = (1/U)^{1/\theta}$ (Evans et al. 2000). Use this relationship to generate a sample of 100 values of Y with $\theta = 2$. From these data calculate an estimate $\widehat{\theta}$. Repeat this process 20 times and also calculate 95% confidence intervals for θ. Compare the average of the estimates $\widehat{\theta}$ with $\theta = 2$. How many of the confidence intervals contain θ?

5.4 For the leukemia survival data in Exercise 4.2:

(a) Use the Wald statistic to obtain an approximate 95% confidence interval for the parameter β_1.

(b) By comparing the deviances for two appropriate models, test the null hypothesis $\beta_2 = 0$ against the alternative hypothesis $\beta_2 \neq 0$. What can you conclude about the use of the initial white blood cell count as a predictor of survival time?