Regression Analysis and Analysis of Variance

Exercises 5

Exercise 5.1
Calculate the Fisher information for the following distributions:
(a) Binomial $b(n, \lambda)$,  
(b) Poisson $p(\lambda)$,  
(c) Exponential $e(\lambda)$,  
(d) Negative binomial $nb(\lambda)$,

Exercise 5.2
Calculate the variance of the maximum likelihood estimators for the relevant parameter for the distributions of the previous exercise.

Exercise 5.3
Consider the mixture distribution $(1 - \epsilon)N(0, \sigma^2) + \epsilon N(0, 9\sigma^2)$. Determine the value of $\sigma = \sigma(\epsilon)$ so that the distribution has variance 1.

Exercise 5.4
For $\epsilon = 0(0.05)0.2$ generate samples of size $n = 20$ of the distribution of the previous exercise with $\sigma = \sigma(\epsilon)$. Use the mean as an estimate of the location parameter and calculate the usual 95% confidence interval based on the $t$-distribution. By simulating such samples 10000 times derive the mean length of the confidence interval and its coverage probability and plot these against the values of $\epsilon$.

Exercise 5.5
Let $q_1(0.95, n)$ and $q_2(0.95, n)$ be the 0.95–quantiles of the statistics $OR_1$ and $OR_2$ of Exercise 3.1 For a sample $x_n = (x_1, \ldots, x_n)$ define possible outliers $x_j$ by

$$|x_j - \text{med}(x_n)| \geq q_1(0.95, n) \text{MAD}(x_n)$$
$$|x_j - \text{shorth}(x_n)| \geq q_2(0.95, n) \text{lshorth}(x_n).$$

By generating normal samples with outliers try and determine which is the better outlier identifier.