

- 3.1 We consider three nested regression models for modelling the number of car accidents according to region (region). The variables risk class (risk) has 3 categories and the number of years of driving experience (exp) is broken down into 4 categories.

Model	variables	$p + 1$	$\ell(\hat{\beta})$	AIC	BIC
M ₁	risk	3	-244.566	495.132	510.362
M ₂	risk + region	★	-151.620	★	★
M ₃	risk + region + exp	10	-139.734	299.468	350.235

Table 1: Goodness-of-fit measures for three nested regression models with the number of parameters in each model ($p + 1$), the value of the log-likelihood function evaluated at the maximum likelihood estimate ($\ell(\hat{\beta})$) and information criteria.

What is the difference between AIC and BIC of Model M₂ (in absolute value)?

Solution

The difference is approximately 35.54. The only information required is the number of parameters of M₂ and the sample size. We can use the first line to compute the latter, since

$$\text{BIC} = -2\ell(\mathbf{M}_1) + 3\ln(n)$$

and we find upon solving the equation $n = 1184$. It only remains to compute the number of levels for the region variable. This is $\text{DF}(\mathbf{M}_3) - \text{DF}(\mathbf{M}_1) - 3 = 7$; there are $K = 4$ categories for years of experience, but only three parameters. The difference $\text{BIC} - \text{AIC} = 7 \cdot \{\ln(1184) - 2\}$.

- 3.2 A random variable X follows a geometric distribution with parameter p if its probability mass function is

$$P(X = x) = (1 - p)^{x-1} p, \quad x = 1, 2, \dots$$

- Write the likelihood and the log-likelihood of the n sample.
- Derive the maximum likelihood estimator for the parameter p .
- Compute the observed information matrix.
- Suppose we have a sample of 15 observations, $\{5, 6, 3, 7, 1, 2, 11, 8, 7, 34, 1, 7, 10, 1, 0\}$, whose sum is 103. Compute the maximum likelihood estimate and its approximate standard error.
- Compute the likelihood ratio and the Wald test statistics. Perform a test at level 5% for $\mathcal{H}_0 : p_0 = 0.1$ against the two-sided alternative $\mathcal{H}_a : p_0 \neq 0.1$.

Solution

(a)

$$L(p; \mathbf{x}) = \prod_{i=1}^n (1 - p)^{x_i - 1} p = (1 - p)^{\sum_{i=1}^n (x_i - 1)} p^n$$

$$\ell(p; \mathbf{x}) = \ln(1 - p) \sum_{i=1}^n (x_i - 1) + n \ln(p)$$

(b)

$$\frac{d}{dp} \ell(p; \mathbf{x}) = -\frac{1}{(1 - p)} \sum_{i=1}^n (x_i - 1) + \frac{n}{p}$$

Setting the score to zero and re-arranging the expression, we find

$$\frac{1}{p} = \frac{1}{n} \sum_{i=1}^n x_i$$

Since the second derivative is negative, the maximum likelihood estimator is the reciprocal mean, $\hat{p} = \bar{X}^{-1}$.

(c) The observed information function is the negative of the hessian matrix. We have

$$\frac{d^2}{dp^2} \ell(p; \mathbf{x}) = -\frac{n(\bar{x} - 1)}{(1 - p)^2} - \frac{n}{p^2}$$

and the observed information evaluated at the maximum likelihood estimate is $j(\hat{p}) = n\bar{x}^3/(\bar{x} - 1)$.

(d) The maximum likelihood estimate is 0.1456 and its standard error is 0.0347.

(e) The Wald statistic is 1.72 and the likelihood ratio statistic equals 0.558. Both can be compared to χ_1^2 : the respective p -values are 0.19 and 0.45, so we fail to reject the null $\mathcal{H}_0 : p = 0.1$.

3.3 We consider the failure time of an engine based on its level of corrosion w . Failure time T is modelled with an exponential distribution with density $f(t) = \lambda \exp(-\lambda t)$, but whose rate parameter $\lambda = aw^b$; if $b = 0$, the failure time is a^{-1} . Assume we have an n sample of independent observations with w_i assumed known. [Coles (2001)]

- (a) Write down the log-likelihood of the model.
- (b) Derive the observed and the Fisher information matrices.
- (c) Show that the profile log-likelihood for b is

$$\ell_p(b) = n \ln(\hat{a}_b) + b \sum_{i=1}^n \ln(w_i) - \hat{a}_b \sum_{i=1}^n w_i^b t_i,$$

and give an explicit formula for the partial maximum likelihood estimator \hat{a}_b .

Solution

(a) The log-likelihood is

$$\ell(a, b; \mathbf{w}, \mathbf{t}) = n \ln(a) + b \sum_{i=1}^n \ln(w_i) - a \sum_{i=1}^n w_i^b t_i$$

(b) The Fisher information is the expected value of the negative of the hessian matrix. Setting

$$j(a, b) = - \begin{pmatrix} \partial^2 \ell / \partial a^2 & \partial^2 \ell / \partial a \partial b \\ \partial^2 \ell / \partial b \partial a & \partial^2 \ell / \partial b^2 \end{pmatrix} = \begin{pmatrix} na^{-2} & \sum_{i=1}^n w_i^b t_i \ln(w_i) \\ \sum_{i=1}^n w_i^b t_i \ln(w_i) & a \sum_{i=1}^n w_i^b t_i \ln^2(w_i) \end{pmatrix}.$$

The Fisher information is obtained by computing the expectation of each entry of the matrix. Only the T_i terms are constant, and since the formulae only involve terms that are linear in T_i , the Fisher information is

$$I(a, b) = \begin{pmatrix} na^{-2} & a^{-1} \sum_{i=1}^n \ln(w_i) \\ a^{-1} \sum_{i=1}^n \ln(w_i) & \sum_{i=1}^n \ln^2(w_i) \end{pmatrix}$$

given that $E(T_i) = a^{-1} w_i^{-b}$.

(c) The partial maximum \hat{a}_b is obtained by differentiating the log-likelihood as a function of a while treating b as fixed. We get

$$\frac{\partial \ell}{\partial a} = \frac{n}{a} - \sum_{i=1}^n w_i^b t_i = 0$$

and isolating a shows that $\hat{a}_b = n / \sum_{i=1}^n w_i^b t_i$ is the maximum, since the second derivative calculated previously is negative.