APTS Statistical Modelling: Practical 1

Helen Ogden

Suppose

$$y_{im} \sim \text{Poisson}\left(\mu(x_{im})\right),$$

independently, for $i = 1, \ldots, n$ and $m = 1, \ldots, M$, where

$$\mu(x_{im}) = 8 \exp\left(w(x_{im})\right),$$

for some function $w(.)$.

Suppose $M = 3$,

$$x_{im} = x_i = -10 + 20\,\frac{i-1}{n-1},$$

and

$$w(x) = 0.001\left(100 + x + x^2 + x^3\right).$$

Consider the following simulation study. For $b = 1, \ldots, B$:

- For $i = 1, \ldots, n$ and $m = 1, \ldots, M$, generate

  $$y_{im} \sim \text{Poisson}(\mu(x_{im})).$$

- Record the AIC for models

  $$y_{im} \sim \text{Poisson}(\mu(x_{im})), \qquad \mu(x_{im}) = \exp\left(\sum_{j=1}^{p} \beta_j x_{im}^{j-1}\right),$$

  for $p = 1, \ldots, p_{\max}$, where $p_{\max} = 20$.

You can run this simulation study with the following code:

```
B <- 1000
n <- 1000
M <- 3
pmax <- 20


w <- function(x) {
    0.001 * (100 + x + x^2 + x^3)
}
```

```r
mu <- function(x) {
    8 * exp(w(x))
}

x <- rep(seq(from = -10, to = 10, length = n), each = M)

aics <- matrix(0, nrow = B, ncol = pmax)

for(b in 1:B){
    y <- rpois(n = M * n, lambda = mu(x))

    mod <- glm(y ~ 1, family = poisson)
    aics[b, 1] <- AIC(mod)

    for(p in 2:pmax) {
        modp <- glm(y ~ poly(x, p - 1), family = poisson)
        aics[b,p] <- AIC(modp)
    }
}

AICorder <- apply(aics, 1, which.min) - 1
tAIC <- table(AICorder)
tAIC
```

Tasks

1. Modify the code above to investigate the performance of AIC as a model selection tool for $n = 25, 50, 100, 1000$. If your simulation study is taking too long to run, try reducing B to 100.

2. Vary the simulation model, using

$$w(x) = \frac{1.2}{1 + \exp(-x)},$$

   to see how AIC performs when the fitted models do not include the simulation model.

3. Modify the code to compute the values of BIC. Repeat the simulation studies from parts 1 and 2, using BIC to compare models. How do the results with AIC and BIC compare?