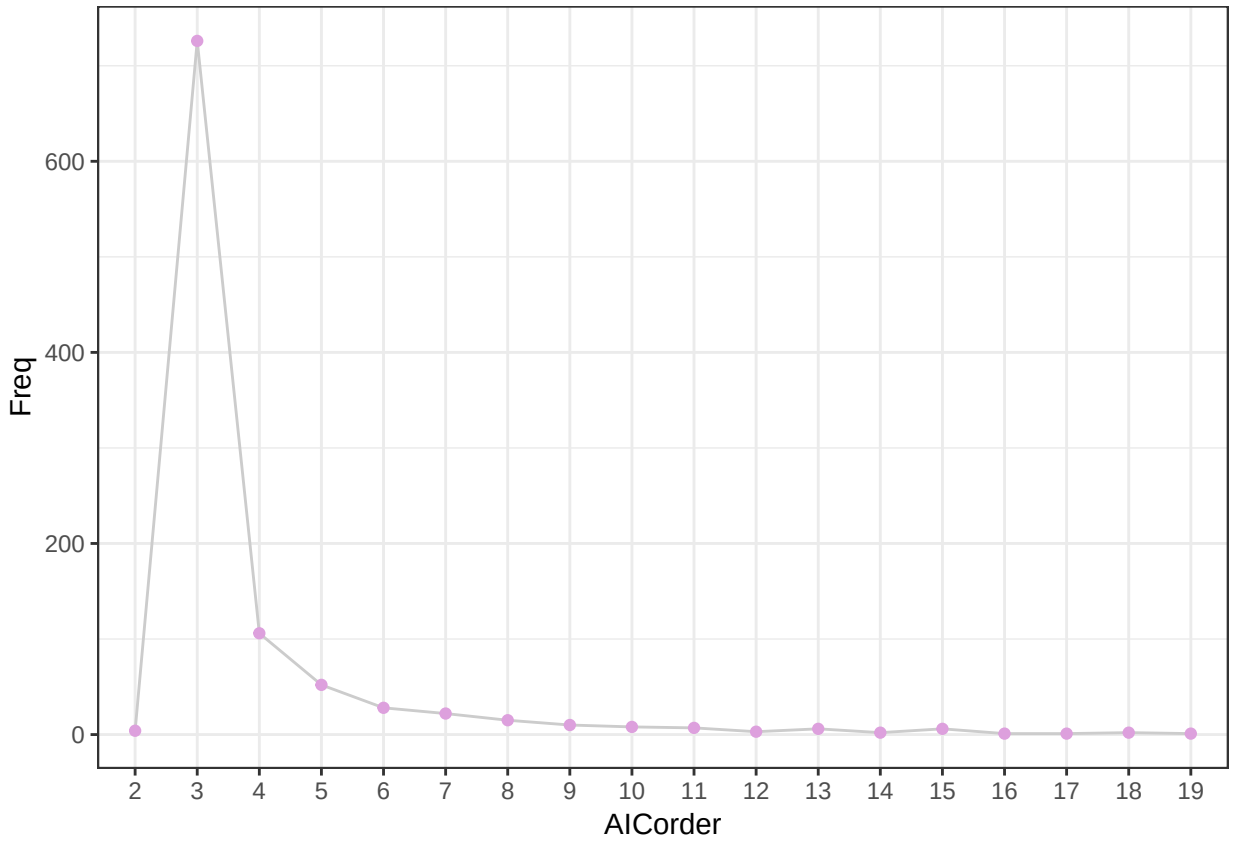
Exercises

Bruna Wundervald

April, 2019

```
set.seed(2019)
B <- 1000
n <- 25
M <- 3
pmax <- 20

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

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

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

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)

  formula <- "y ~ x"
  mod <- glm(formula, family = poisson)
  aics[b, 2 ] <- AIC(mod)

  for(j in 3:pmax){
    formula <- paste(formula, " + I(x^", j - 1, ")", sep = "")
    mod <- glm(formula, family = poisson)
    aics[b, j] <- AIC(mod)
  }
}

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

```
tAIC %>%
  as.data.frame() %>%
  ggplot(aes(AICorder, Freq, group = 1)) +
  geom_line(colour = "gray80") +
  geom_point(colour = "plum") +
  theme_bw()
```



1. Investigate the performance of AIC as a model selection tool for $n = 25, 50, 100, 1000$.

```
x_sim <- c(25, 50, 100, 1000) %>%
  purrr::map(~{
    rep(seq(from = -10, to = 10, length = .x), each = M)
  })

#-------
for(b in 1:B){

  y <- rpois(n = M * n, lambda = mu(x_sim[[1]]))

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

  formula <- "y ~ x"
  mod <- glm(formula, family = poisson)
  aics[b, 2 ] <- AIC(mod)

  for(j in 3:pmax){
    formula <- paste(formula, " + I(x^", j - 1, ")", sep = "")
    mod <- glm(formula, family = poisson)
    aics[b, j] <- AIC(mod)
  }
}
```

```r
AICorder_25 <- apply(aics, 1, which.min) - 1
tAIC_25 <- table(AICorder_25)
tAIC_25
```

```
## AICorder_25
##   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19
##   2 694 112  55  43  24  16  16   8   6   7   4   3   2   4   1   2   1
```

```r
#-------
for(b in 1:B){

  y <- rpois(n = M * n, lambda = mu(x_sim[[2]]))

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

  formula <- "y ~ x"
  mod <- glm(formula, family = poisson)
  aics[b, 2 ] <- AIC(mod)

  for(j in 3:pmax){
    formula <- paste(formula, " + I(x^", j - 1, ")", sep = "")
    mod <- glm(formula, family = poisson)
    aics[b, j] <- AIC(mod)
  }
}
AICorder_50 <- apply(aics, 1, which.min) - 1
tAIC_50 <- table(AICorder_50)
tAIC_50
```

```
## AICorder_50
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  17  18  19
## 110 579 132  57  27  23  13  21   8   6   5   5   4   3   2   2   1   2
```

```r
#-------
for(b in 1:B){

  y <- rpois(n = M * n, lambda = mu(x_sim[[3]]))

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

  formula <- "y ~ x"
  mod <- glm(formula, family = poisson)
  aics[b, 2 ] <- AIC(mod)

  for(j in 3:pmax){
    formula <- paste(formula, " + I(x^", j - 1, ")", sep = "")
    mod <- glm(formula, family = poisson)
    aics[b, j] <- AIC(mod)
  }
}
```

```r
AICorder_100 <- apply(aics, 1, which.min) - 1
tAIC_100 <- table(AICorder_100)
tAIC_100
```

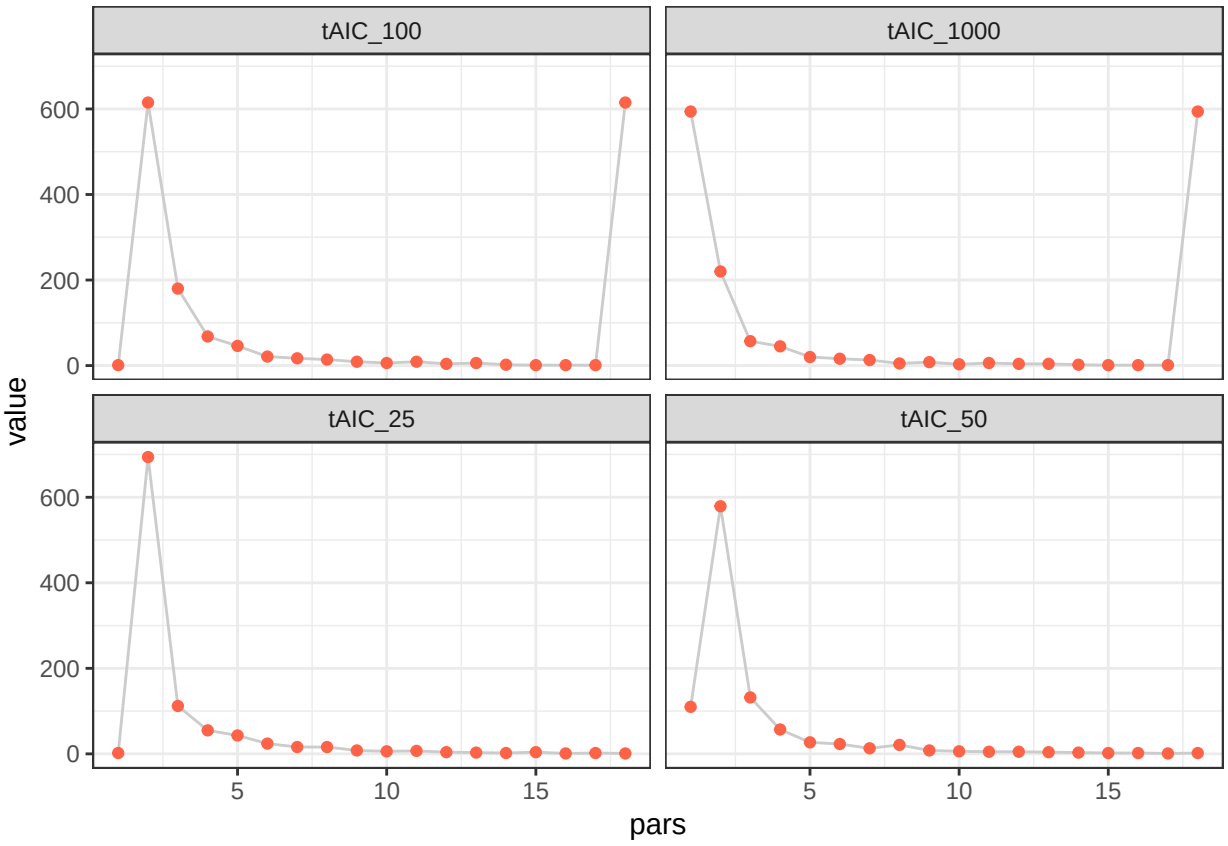
```
## AICorder_100
##   0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  16
##   1 615 180  68  46  21  17  14   9   6   9   4   6   2   1   1
```

```r
#-------
for(b in 1:B){

  y <- rpois(n = M * n, lambda = mu(x_sim[[3]]))

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

  formula <- "y ~ x"
  mod <- glm(formula, family = poisson)
  aics[b, 2 ] <- AIC(mod)

  for(j in 3:pmax){
    formula <- paste(formula, " + I(x^", j - 1, ")", sep = "")
    mod <- glm(formula, family = poisson)
    aics[b, j] <- AIC(mod)
  }
}

AICorder_1000 <- apply(aics, 1, which.min) - 1
tAIC_1000 <- table(AICorder_1000)
tAIC_1000
```

```
## AICorder_1000
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  17  19
## 594 220  57  45  20  16  13   5   8   3   6   4   4   2   1   1   1
```

```r
#--------

cbind(tAIC_25, tAIC_50, tAIC_100, tAIC_1000) %>%
  as.data.frame() %>%
  gather() %>%
   mutate(pars = rep(1:18,  4)) %>%
  ggplot(aes(pars, value, group = 1)) +
  geom_line(colour = "gray80") +
  geom_point(colour = "tomato") +
  facet_wrap(~key) +
  theme_bw()
```

```
## Warning in cbind(tAIC_25, tAIC_50, tAIC_100, tAIC_1000): number of rows of
## result is not a multiple of vector length (arg 3)
```



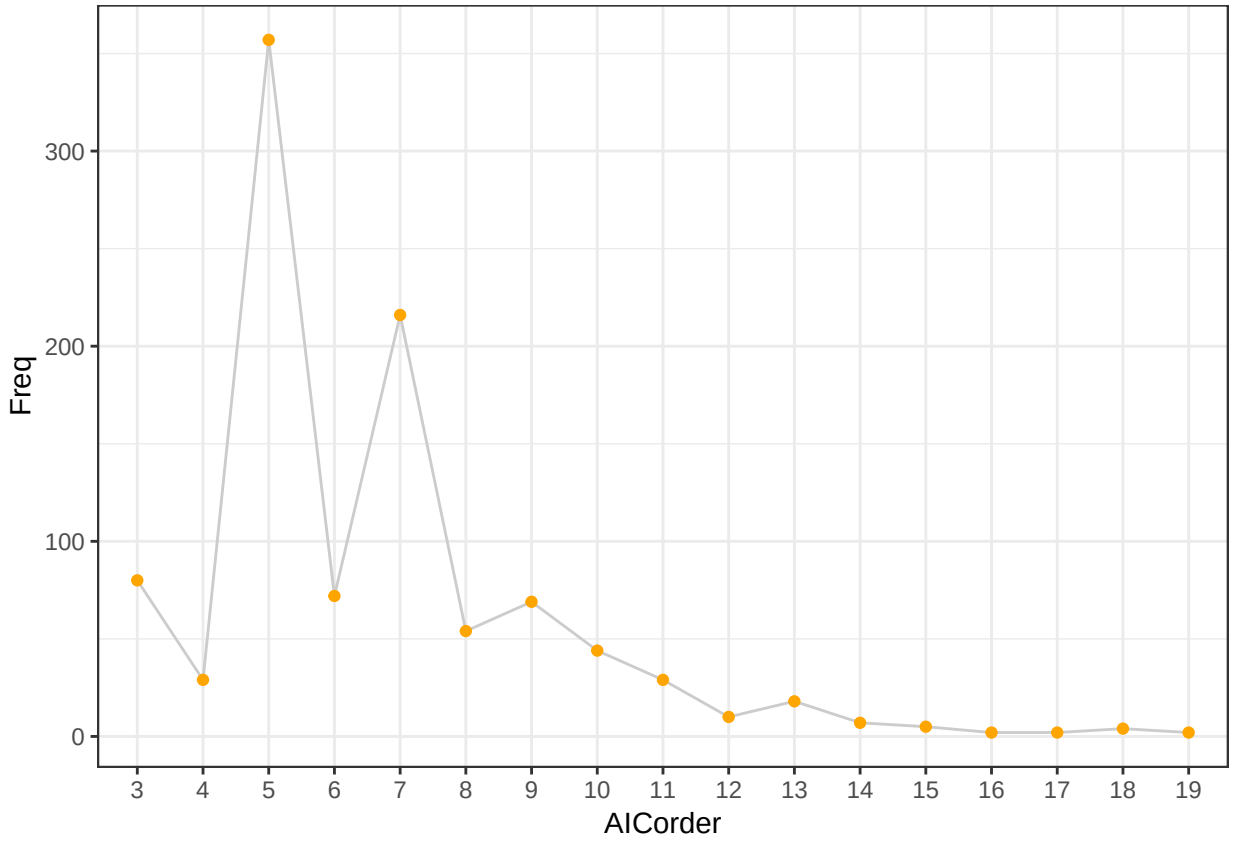
2. Vary the simulation, using

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

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

```
w <- function(x){
  1.2 /(1 + exp(-x))
}

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

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)

  formula <- "y ~ x"
```

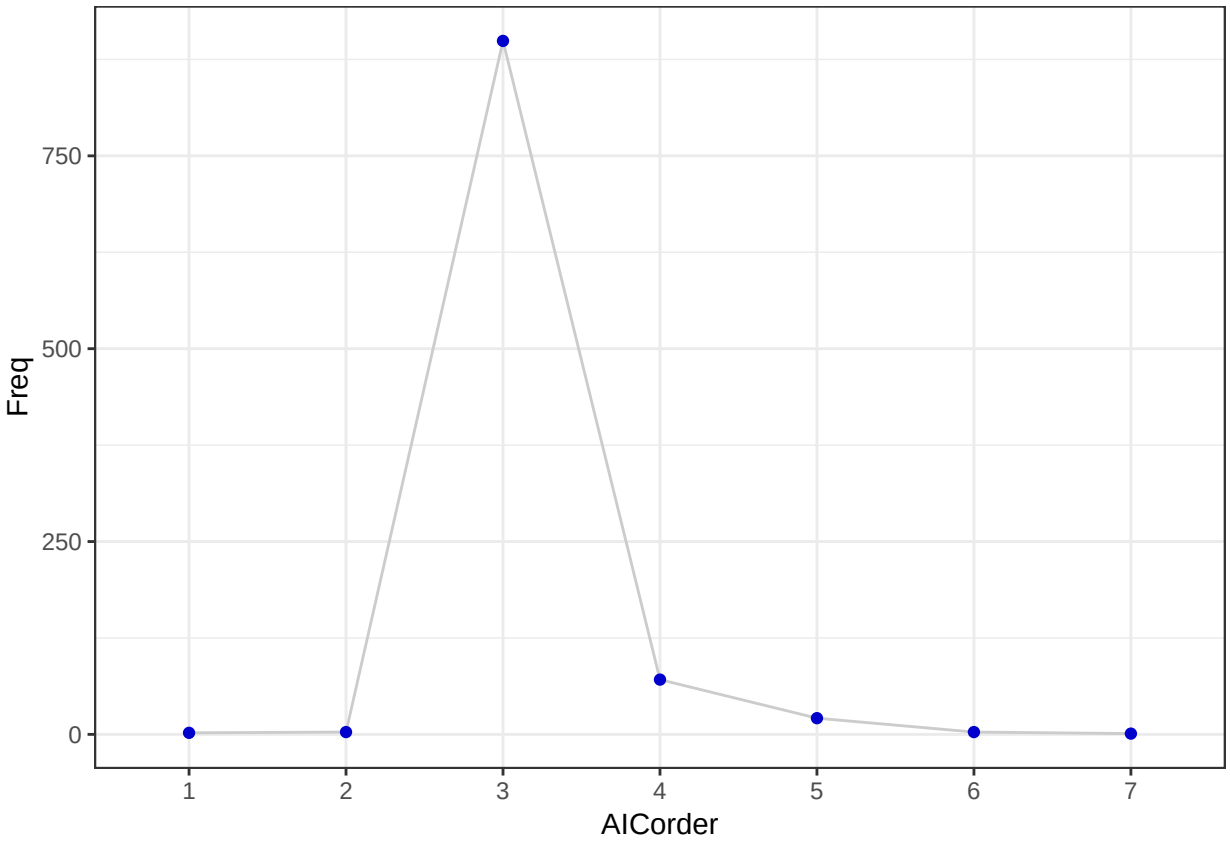
```r
  mod <- glm(formula, family = poisson)
  aics[b, 2 ] <- AIC(mod)

  for(j in 3:pmax){
    formula <- paste(formula, " + I(x^", j - 1, ")", sep = "")
    mod <- glm(formula, family = poisson)
    aics[b, j] <- AIC(mod)
  }
}

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

tAIC %>%
  as.data.frame() %>%
  ggplot(aes(AICorder, Freq, group = 1)) +
  geom_line(colour = "gray80") +
  geom_point(colour = "orange") +
  theme_bw()
```



3. Modify the code above to compute the values of BIC and AIC_c, where

$$AIC_c = AIC + \frac{2p^2 + 2p}{n - p - 1}$$

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

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

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)
  p <- length(mod$coefficients)
  aics[b, 1] <- AIC(mod) + (2*p^2 + 2*p)/(n - p - 1)

  formula <- "y ~ x"
  mod <- glm(formula, family = poisson)
  p <- length(mod$coefficients)
  aics[b, 2 ] <- AIC(mod) + (2*p^2 + 2*p)/(n - p - 1)

  for(j in 3:pmax){
    formula <- paste(formula, " + I(x^", j - 1, ")", sep = "")
    mod <- glm(formula, family = poisson)
    p <- length(mod$coefficients)
    aics[b, j] <- AIC(mod) + (2*p^2 + 2*p)/(n - p - 1)
  }
}


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

tAIC %>%
  as.data.frame() %>%
  ggplot(aes(AICorder, Freq, group = 1)) +
  geom_line(colour = "gray80") +
  geom_point(colour = "blue3") +
  theme_bw()
```


750
500
250
0
Freq
1
2
3
4
5
6
7
AICorder