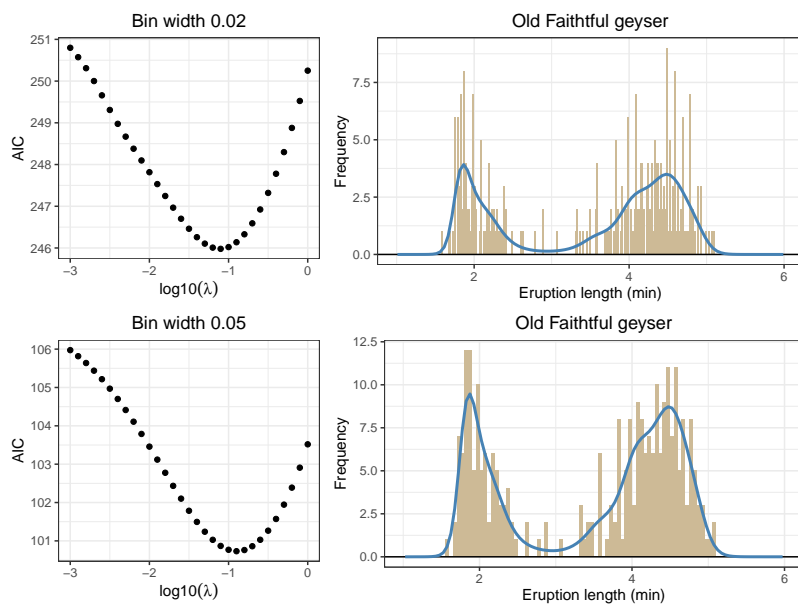


Density estimation, optimal on AIC (Old Faithful geyser data)



P-spline density estimation for Old Faithful eruption length using bin widths 0.02 (top row) and 0.05 (bottom row). B-splines with 20 segments and second order penalty. The left column shows the AIC profiles, the right column the histograms and fits. R code in `f-geyseropt.R`

```
# Density estimation, optimal on AIC (Old Faithful geyser data)
# A graph in the book 'Practical Smoothing. The Joys of P-splines'
# Paul Eilers and Brian Marx, 2019

library(ggplot2)
library(gridExtra)
library(JOPS)

# Get the data
data(faithful)
v = faithful[, 1] # Eruption length

# Histogram with narrow bin width
bw = 0.02
hst = hist(v, breaks = seq(1, 6, by = bw), plot = F)
x = hst$mids
y = hst$counts
Data = data.frame(x = x, y = y)

# Poisson smoothing
lambdas = 10^seq(-3, -0, by = 0.1)
aics = NULL
for (lambda in lambdas) {
  fit = psPoisson(x, y, nseg = 20, pord = 2, lambda = lambda, show = F)
  aics = c(aics, fit$aic)
}

# Finding min AIC
ka = which.min(aics)
lambda = lambdas[ka]

# Gridded data for plotting using opt tuning
fit = psPoisson(x, y, nseg = 20, pord = 2, lambda = lambda, show = F)
```

```

F1 = data.frame(x = log10(lambdas), y = aics)
Fit = data.frame(x = fit$xgrid, y = fit$mugrid)
plt1 = ggplot(aes(x = x, y = y), data = F1) +
  geom_point() + ggtitle(paste("Bin width", bw)) +
  xlab(expression(log10(lambda))) + ylab('AIC') +
  JOPS_theme()

plt2 = ggplot(aes(x = x, y = y, fill = I("wheat3")), data = Data) +
  geom_bar(stat = "identity") +
  geom_hline(yintercept = 0) +
  xlab("Eruption length (min)") + ylab("Frequency") +
  ggtitle("Old Faithful geyser") +
  geom_line(data = Fit, col = I("steelblue"), size = 1.0) +
  JOPS_theme()

# Histogram using wider bin width
bw = 0.05
hst = hist(v, breaks = seq(1, 6, by = bw), plot = F)
x = hst$mids
y = hst$counts
Data = data.frame(x = x, y = y)

# lambdas = 10 ^ seq(-4, -0, by = 0.1)
aics = NULL
for (lambda in lambdas) {
  fit = psPoisson(x, y, nseg = 20, pord = 2, lambda = lambda, show = F)
  aics = c(aics, fit$aic)
}
ka = which.min(aics)
lambda = lambdas[ka]

fit = psPoisson(x, y, nseg = 20, pord = 2, lambda = lambda, show = F)
F1 = data.frame(x = log10(lambdas), y = aics)
Fit = data.frame(x = fit$xgrid, y = fit$mugrid)
plt3 = ggplot(aes(x = x, y = y), data = F1) +
  geom_point() + ggtitle(paste("Bin width", bw)) +
  xlab(expression(log10(lambda))) + ylab('AIC') +
  JOPS_theme()

plt4 = ggplot(aes(x = x, y = y, fill = I("wheat3")), data = Data) +
  geom_bar(stat = "identity") +
  geom_hline(yintercept = 0) +
  xlab("Eruption length (min)") + ylab("Frequency") +
  ggtitle("Old Faithful geyser") +
  geom_line(data = Fit, col = I("steelblue"), size = 1) +
  JOPS_theme()

# Make and save graph
grid.arrange(plt1, plt2, plt3, plt4, ncol = 2, nrow = 2, widths = c(4,
6))

```
