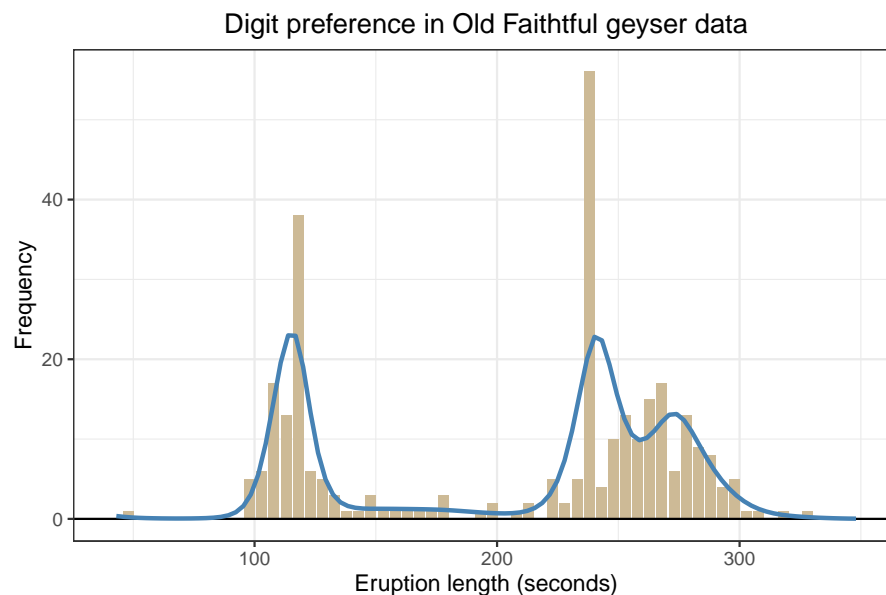


## Density estimation with digit preference (Old Faithful geyser data)

---



Digit preference in Old Faithful eruption times. Uncertain nightly observations have been rounded to exactly 2 and 4 minutes. Optimal  $\log_{10}(\lambda) = -1.5$  determined by AIC. Cubic B-spline basis with 20 segments and third order penalty. R code in `f-geyser-with-DP.R`

---

```
# Show digit preference (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(MASS)
library(JOPS)

u = round(geyser$duration * 60, 1)
h = hist(u, breaks = seq(40, 350, by = 5) + 0.5, plot = F)
x = h$mids
y = h$counts
Data = data.frame(x, y)

llas = seq(-3, 1, by = 0.1)
aics = NULL
nseg = 20
for (lla in llas) {
  fit = psPoisson(x, y, nseg = nseg, lambda = 10^lla, show = F)
  aics = c(aics, fit$aic)
  cat(fit$lambda, fit$aic, "\n")
}
k = which.min(aics)
lambda = 10^llas[k]
fit = psPoisson(x, y, nseg = nseg, lambda = lambda, show = F)
Fit = data.frame(x = fit$xgrid, y = fit$mugrid)

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

```
JOPS_theme()  
# Save graph  
print(plt)
```

---