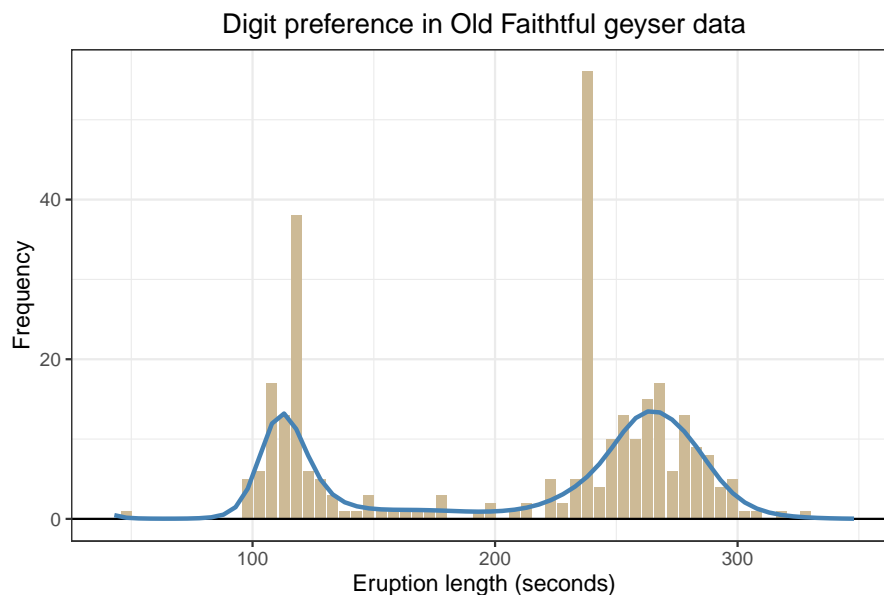


## Using weights to handle digit preference (Old Faithful data)

---



Using weights to remedy digit preference in Old Faithful eruption times. Uncertain nightly observations have been rounded to exactly 2 and 4 minutes. These bins have been given zero weights. Optimal  $\log_{10}(\lambda) = -0.8$  determined by AIC. Cubic B-spline basis with 20 segments and third order penalty. R code in `f-geyser-weights.R`

---

```
# Use weights to handle 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)

# Get the data
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)

# Set weights to right digit preference
w0 = rep(1, length(y))
w0[16] = w0[40] = 0

B = bbase(x, nseg = 20)
n = ncol(B)
D = diff(diag(n), diff = 3)
P = t(D) %*% D

# Fit for series of lambdas
llas = seq(-3, 2, by = 0.1)
aics = NULL
nseg = 20

eta = log(y + 1)
```

```

a = 0
Mus = NULL
for (lla in llas) {
  lambda = 10^lla
  for (it in 1:20) {
    mu = exp(eta)
    w = c(w0 * mu)
    W = diag(c(w0 * mu))
    r = w0 * (y - mu + mu * eta)
    G = t(B) %>% W %>% B
    anew = solve(G + lambda * P, t(B) %>% r)
    da = max(abs(anew - a))
    # cat(it, lla, da, '\n')
    if (da < 1e-04)
      break
    a = anew
    eta = B %>% a
  }
  Mus = cbind(Mus, mu)
  K = solve(G + lambda * P, G)
  ed = sum(diag(K))
  ok = y > 0
  dev = 2 * sum(w0[ok] * y[ok] * log(y[ok]/mu[ok]))
  aic = dev + 2 * ed
  aics = c(aics, aic)
}

# Pick best lambda
k = which.min(aics)
lambda = 10^llas[k]
mu = Mus[, k]
cat("Optimum: log_10(lambda) = ", llas[k], "\n")

# Make graph
Fit = data.frame(x = x, y = mu)
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 geysir data") +
  JOPS_theme()

# Make and save graph
print(plt)

```

---