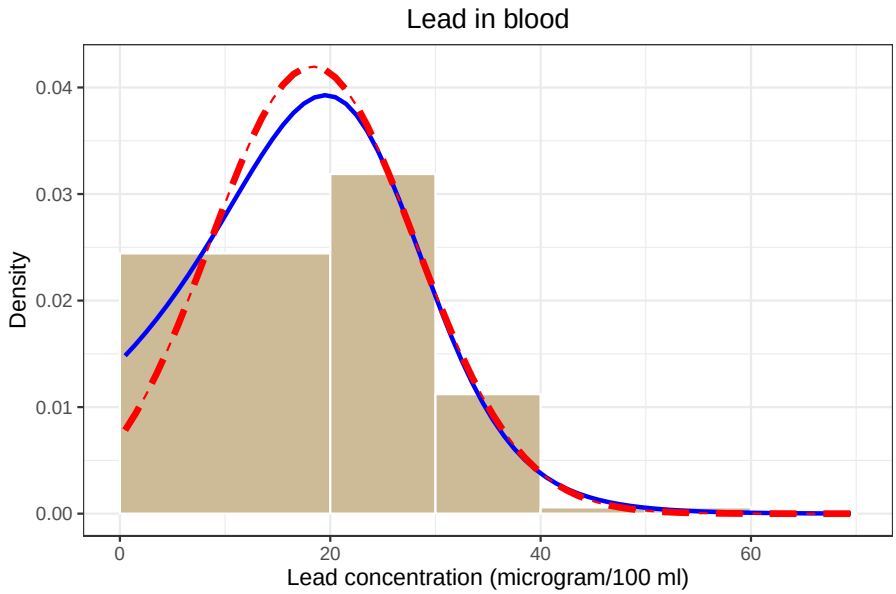
Composite link function density estimation (Lead in blood data)



Raw density of lead concentrations in the blood of New York children and the estimated smooth density, based on optimal AIC. Solid line: second order penalty; broken line: third order penalty. R code in `f-lead2.R`

```
# Composite link function density estimation (Lead in blood data)
# A graph in the book 'Practical Smoothing. The Joys of P-splines'
# Paul Eilers and Brian Marx, 2019

library(ggplot2)
library(JOPS)

# Input data
cb <- c(0, 20, 30, 40, 50, 60)
ce <- c(20, 30, 40, 50, 60, 70)
y <- c(79, 54, 19, 1, 1, 0)
m <- length(y)
n <- ce[m]

C <- matrix(0, m, n)
for (i in 1:m) C[i, cb[i]:ce[i]] <- 1

mids = (cb + ce)/2 - 0.5
widths = ce - cb + 1
dens = y/widths/sum(y)

# Prepare B-spline matrix and penalty matrix
x <- 1:n
B <- bbase(x)
lambda2 = 3.12
fit2 = pclm(y, C, B, lambda = lambda2, pord = 2, show = T)
gam2 = fit2$gamma
lambda3 = 50.1
fit3 = pclm(y, C, B, lambda = lambda3, pord = 3, show = T)
gam3 = fit3$gamma

# Plot it with ggplot2
Fit2 = data.frame(x = x - 0.5, y = gam2/sum(gam2))
Fit3 = data.frame(x = x - 0.5, y = gam3/sum(gam3))
```

```
Dat = data.frame(cb, ce, dens)
plt2 = ggplot(Dat, aes(ymin = 0)) +
  geom_rect(aes(xmin = cb, xmax = ce, ymax = dens, fill = I("wheat3"),color=I('white'))) +
  geom_line(aes(x = x, y = y), size = 1, linetype = 1,data = Fit2, col = I("blue")) +
  geom_line(aes(x = x, y = y), size = 0.5, linetype = 2,data = Fit3, col = I("red")) +
  geom_line(aes(x = x, y = y), size = 1.5, linetype = 2,data = Fit3, col = I("red")) +
  xlab("Lead concentration (microgram/100 ml)") + ylab("Density") +
  ggtitle("Lead in blood") +
  JOPS_theme()

# Save graph
print(plt2)
```