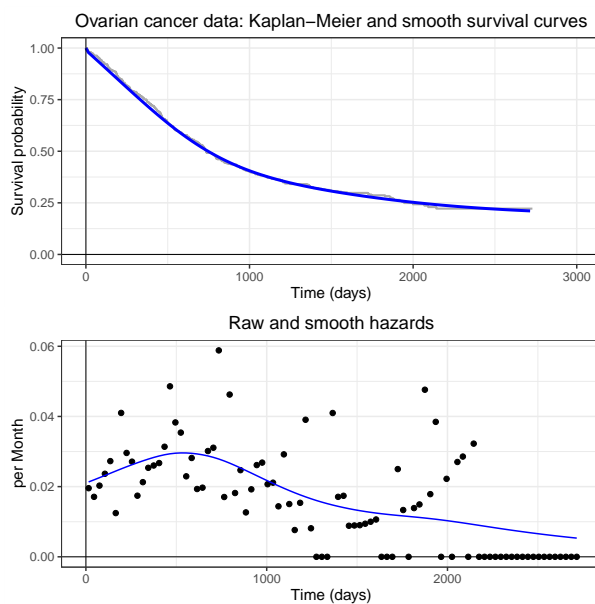


Survival hazard smoothing, mixed model for λ (Ovarian cancer data)



Hazard smoothing with P-splines for the ovarian cancer data. Top: smooth survival curve compared to Kaplan-Meier. Bottom: raw and smooth hazard functions; λ optimized with the Harville-Fellner-Schall algorithm. R code in `f-surv-OVA.R`

```
# Survival analysis hazard smoothing, mixed model for lambda (Ovarian cancer data)
# A graph in the book 'Practical Smoothing. The Joys of P-splines'
# Paul Eilers and Brian Marx, 2019
```

```
library(CoxRidge)
library(survminer)
library(JOPS)
library(gridExtra)
```

```
# Load data
data(ova)
time = ova$time;
dead = ova$death;
m = length(time)
dova = data.frame(time = time, dead = dead)
```

```
# Discretise time
dt = 30;
tm = floor(time / dt) + 1;
tmax = max(tm);
Tm = outer(tm, rep(1, tmax));
e = rep(1, m)
Tau = outer(e, 1:tmax);
R = Tm >= Tau;
Y = (Tm == Tau) * (outer(dead, rep(1, tmax)));
r = apply(R, 2, sum)
y = apply(Y, 2, sum)
n = length(y)
x = ((1:n) - 0.5) * dt
```

```
# Prepare penalty
D = diff(diag(n), diff = 2)
P = t(D) %>% D
```

```

# Initalize
lambda = 10
eta = log((y + 1) / (r + 2))

# Iterate
for (it in 1:40) {

  # Update log-hazard
  mu = r * exp(eta)
  W = diag(c(mu))
  eta_old = eta
  G = W + lambda * P
  eta = solve(G, y - mu + mu * eta)
  delta = max(abs(eta - eta_old))
  if (delta < 1e-3) break

  # Update lambda
  H = solve(G, W)
  ed = sum(diag(H))
  tau2 = sum((D %**% eta) ^ 2) / (ed - 2)
  lambda = 1/ tau2
  cat(it, delta, lambda, ed, '\n')
}

# Data frames for ggplot
haz = c(0, exp((eta)))
DF = data.frame(x = c(0, x), y = cumprod(1 - haz))
DF2 = data.frame(x = x, raw = y / (r + 1e-6), u = y, haz = exp(eta))

# Survival object for Kaplan-Meier plot
surph = survfit(Surv(time, dead) ~ 1, data = dova)

# Plot Kaplan-Meier curve
kmp = ggsvrplot(surph, data= dova, palette = 'darkgrey',
               conf.int = FALSE, censor = FALSE, legend= "none",
               legend.title = '', legend.labs = 'K-M', size = 0.7)

# Kaplan-Meier plot and smooth survival curve
kpl = kmp$plot
plt1 = kpl +
  geom_line(aes(x = x, y = y), col = 'blue', size = 1, data = DF) +
  xlab('Time (days)') +
  ggtitle('Ovarian cancer data: Kaplan-Meier and smooth survival curves') +
  geom_hline(yintercept = 0, size = 0.3) +
  geom_vline(xintercept = 0, size = 0.3) +
  theme(
    plot.title = element_text(size=16),
    axis.title.x = element_text(size=16),
    axis.title.y = element_text(size=16)) +
  JOPS_theme() +
  theme(legend.position = "none")

# Raw and smooth hazards
plt2 = ggplot(data = DF2) +
  geom_point(aes(x = x, y = raw)) +
  xlab('Time (days)') + ylab('per Month') +
  geom_hline(yintercept = 0, size = 0.3) +
  geom_vline(xintercept = 0, size = 0.3) +
  ggtitle('Raw and smooth hazards') +
  geom_line(aes(x = x, y = haz) , col = 'blue') +
  theme(
    plot.title = element_text(size=16),
    axis.title.x = element_text(size=16),
    axis.title.y = element_text(size=16)) +
  JOPS_theme()

plts = grid.arrange(plt1, plt2, nrow = 2, ncol = 1)

```

`plot(plts)`
