8.3 EXERCISES

BMED6420
Brani Vidakovic; Fall 2018
Consult the class slides, hints, and cited literature for the solution of exercise problems.

1. Dukes' C Colorectal Cancer and Diet Treatment. Colorectal cancer is a common cause of death. In the advanced stage of disease, when the disease is first diagnosed in many patients, surgery is the only treatment. Cytotoxic drugs, when given as an adjunct to surgery, do not prevent relapse and do not increase the survival in patients with advanced disease.

Interest has been shown, at least by patients, in a nutritional approach to treatment, where diet plays a critical role in the disease management program.

In a controlled clinical trial, McIllmurray and Turkie (1987) evaluated the diet treatment in patients with Dukes' C colorectal cancer, because the residual tumour mass is small after operation, the relapse rate is high, and no other effective treatment is available. The diet treatment consisted of linolenic acid, an oil extract of the seed from the evening primrose plant *Onagraceae Oenothera biennis* and vitamin E.

The data for the treatment and control patients are given below:

Treatment	Survival time (months)
Linoleic acid ($n_1 = 25$)	1+, 5+, 6, 6, 9+, 10, 10, 10+, 12, 12, 12, 12, 12+, 13+, 15+, 16+, 20+, 24, 24+, 27+, 32, 34+, 36+, 36+, 44+
Control ($n_2 = 24$)	3+, 6, 6, 6, 6, 8, 8, 12, 12, 12+, 15+, 16+, 18+, 18+, 20, 22+, 24, 28+, 28+, 28+, 30, 30+, 33+, 42

Fit the data with Weibull distribution, taking the treatment/control (1/0) as a covariate. Place noninformative priors on all parameters. Consult the file `Leukemia.odc` from the repository. Is the linoleic acid treatment beneficial? Comment.

2. Censored Rayleigh.

The lifetime (in hours) of a certain sensor has Rayleigh distribution, with survival function

$$S(t) = \exp\left\{-\frac{1}{2}\lambda t^2\right\}, \ \lambda > 0.$$

Twelve sensors are placed under test for 100 hours, and the following failure times are recorded 23, 40, 41, 67, 69, 72, 84, 84, 88, 100+, 100+. Here + denotes a censored time.

(a) If failure times $t_1, \dots, t_r$ are observed, and $t_{r+1}^+, \dots, t_n^+$ are censored, find the Bayes estimator of λ . Use noninformative gamma prior on λ.

(b) Evaluate $S(t)$ for $t = 60$ and find 95% Credible Set.

(c) The MLE for λ is

$$\hat{\lambda} = \frac{2r}{\sum_{i=1}^{r} t_i + \sum_{i=r+1}^{n} t_i^{+2}}.$$

Evaluate the MLE for the given data and comment on closeness to the Bayes estimator in (a).

Hint: Rayleigh distribution is not implemented as default in WinBUGS/OpenBUGS and has to be dealt using zero-tricks. Additional complication is censoring. Here is an example how to do censoring on distributions defined by zero-trick. In the set of variables $y_1, \ldots, y_9$ from normal distribution with mean μ and variance 1, the last three values are censored at $t = 8$.

The standard representation with censoring

```
model {
  for (i in 1:6) {y[i] ~ dnorm(mu, 1)}        # uncensored data
  for (i in 7:9) {y[i] ~ dnorm(mu, 1)I(8,)}   # censored data
  mu                   ~ dunif(0, 100)
}

Data:
list(y = c(6,6,6,7,7,7,NA,NA,NA))

 node  mean  sd  MC error 2.5% median 97.5% start sample
mu 7.193 0.3478 0.003604 6.515 7.19 7.875 1001 10000
```

is equivalent to

```
model {
  for (i in 1:6) {y[i] ~ dnorm(mu, 1)}
  for (i in 1:3) {
    zeros[i]          <- 0
    zeros[i]           ~ dpois(p[i])
    p[i]              <- -log(phi(mu-8))
    #each censored observation provides term P(Y-mu>8-mu) to
    #the likelihood of mu, which is equal to 1-phi(8-mu)=phi(mu-8),
    # for phi being the cdf of standard normal
  }
  mu                   ~ dunif(0, 100)
}

Data:
```

```
list(y = c(6,6,6,7,7,7,NA,NA,NA))

 node  mean  sd  MC error 2.5% median 97.5% start sample
mu 7.192 0.348 0.003687 6.519 7.185 7.882 1001 10000
```

3. Stagnant Water with MAR Data.

Carlin et al (1992)[1] analyse data on stagnation of water by piecing together linear parametric forms.

- y_i is log flow rate down the channel.
- x_i is log height of stagnant surface levels for different surfactants i.

Proposed model is

$$\begin{aligned} y_i &\sim \mathcal{N}(\mu_i, \sigma^2) \\ \mu_i &= \alpha + \beta_1 \cdot x_i + \beta_2 \cdot (x_i - \theta)_+ \end{aligned}$$

Here a_+ is a if $a \geq 0$ and 0 if $a < 0$.

According to this model, regression slope is β_1 for $x < \theta$ and $\beta_1 + \beta_2$ for $x \geq \theta$.

The original exercise is modified to have two y's and two x's missing at random.

```
model {
  for (i in 1:N) {
    y[i]     ~ dnorm(mu[i], tau)
    mu[i]  <- alpha + beta[1]*x[i] + beta[2]*(x[i] - theta)
             * step(x[i] - theta)
  }
  for( i in 1:N) {
      x[i] ~ dunif(-5,5)}
  tau       ~ dgamma(0.001, 0.001)
  alpha     ~ dnorm(0.0, 1.0E-6)
  for (j in 1:2) {
    beta[j] ~ dnorm(0.0, 1.0E-6)
  }
  sigma    <- 1/sqrt(tau)
  theta     ~ dunif(-1.3, 1.1)
}

Data
list(y = c(1.12, 1.12, 0.99, 1.03, 0.92, NA, 0.81, 0.83, 0.65, 0.67, 0.60,
0.59, 0.51, 0.44, 0.43, 0.43, 0.33, 0.30, 0.25, NA, 0.13, -0.01, -0.13,
```

[1]Hierarchical Bayesian Analysis of Changepoint Problems Bradley P. Carlin, Alan E. Gelfand and Adrian F. M. Smith Journal of the Royal Statistical Society. Series C (Applied Statistics) Vol. 41, No. 2 (1992), pp. 389-405.

```
 -0.14, -0.30, -0.33, -0.46,  -0.43, -0.65),
     x = c(-1.39, -1.39, -1.08, -1.08, -0.94, -0.80, -0.63, -0.63, -0.25, -0.25,
     -0.12, NA, 0.01, 0.11, 0.11, 0.11,  0.25, 0.25, 0.34, 0.34, 0.44, 0.59,
     0.70, 0.70, 0.85, NA,  0.99, 0.99, 1.19),
     N = 29)


Inits
list(alpha = 0.2, beta = c(-0.45, 0), tau = 5, theta = 0)

+ Gen Inits

Complete data results:
 node  mean  sd  MC error 2.5% median 97.5% start sample
alpha 0.5482 0.01337 4.934E-4 0.5228 0.5475 0.5756 501 10000
beta[1] -0.4187 0.01555 5.097E-4 -0.4485 -0.4193 -0.3869 501 10000
beta[2] -0.5944 0.0212 2.248E-4 -0.6362 -0.5942 -0.5528 501 10000
sigma 0.02218 0.003352 5.138E-5 0.01673 0.02178 0.02995 501 10000
theta 0.02563 0.03378 0.001399 -0.04016 0.02783 0.08707 501 10000

Omitted from the original data  y_6=0.90,  y_20 = 0.24, x_12=-0.12, x_26=0.85

mean sd MC_error val2.5pc median val97.5pc start sample
alpha 0.5498 0.01519 2.87E-4 0.5204 0.5496 0.5794 1001 100000
beta[1] -0.4153 0.01676 3.064E-4 -0.4485 -0.4153 -0.3826 1001 100000
beta[2] -0.5892 0.02205 2.957E-4 -0.6323 -0.5895 -0.5453 1001 100000
sigma 0.02197 0.00356 2.742E-5 0.01632 0.02151 0.03022 1001 100000
theta 0.01631 0.03713 7.755E-4 -0.05168 0.01715 0.08537 1001 100000
x[12] -0.1016 0.05593 5.299E-4 -0.2168 -0.09895 -0.002997 1001 100000
x[26] 0.8855 0.0236 1.313E-4 0.839 0.8854 0.9327 1001 100000
y[6] 0.8821 0.02345 1.156E-4 0.8356 0.8821 0.9281 1001 100000
y[20] 0.2179 0.02347 1.358E-4 0.1715 0.2179 0.2644 1001 100000
```