Homework 1

AMS 207   Spring 2018

Wyara Vanesa Moura e Silva   April 19

Question 3

The failure time of a pump follows a two-parameter exponential distribution,

$$f(y|b,m) = \frac{1}{b}\exp\left\{-\frac{(y-m)}{b}\right\}, \quad \text{when} \quad y \geq m.$$

For the two-parameter exponential distribution this function is defined when $f(y|b,m) > 0$, and $b > 0$. In which, m is the location parameter.

3.1

Obtain the likelihood for b and m based on an i.i.d. sample of size n.
Solution:

The likehood function for b and m is given by:

$$
\begin{aligned}
f(\mathbf{y}|b,m) &= \prod_{i=1}^{n}\frac{1}{b}\exp\left\{-\frac{(y_i - m)}{b}\right\}\mathbb{I}\left(y_i \geq m\right) \\
&= \left(\frac{1}{b}\right)^n\exp\left\{-\frac{1}{b}\sum_{i=1}^{n}(y_i - m)\right\}\mathbb{1}\left(\min\left\{y_i\right\}_{i=1}^{n} \geq m\right) \\
&= \left(\frac{1}{b}\right)^n\exp\left\{-\frac{1}{b}\left(\sum_{i=1}^{n}y_i - nm\right)\right\}\mathbb{1}\left(\min\left\{y_i\right\}_{i=1}^{n} \geq m\right)
\end{aligned}
$$

3.2

Consider a suitable transformation that maps the parameters b and m to the plane.
Solution:

As $b > 0$, a suitable transformation would be to use the logarithm transformation $\log(b)$, which will maps the parameter space to the real line $\mathbb{R}$.

Regarding the location parameter m, we have that the constraint for this parameter is $y \geq m$. As $y > 0$ being the time of failure. Then, using the logarithm transformation of the $\log(y_{(1)} - m)$ the parameter space will be now the real line $\mathbb{R}$.

3.3

Consider a sample of 8 pumps, where all pumps failed at some time. The smallest failure time was 23721 minutes and the total testing time for all pumps was 15962989 minutes. Assume a sensible prior for the transformed parameters and explore the contours of the posterior distribution.
Solution:

As informations given by the question, we have: the smallest failure time which is $y_{(1)} = 23721$ minutes. In addition, the total testing time for $n = 8$ pumps (where all pumps failed at

some time) which is $\sum_{i=1}^{n} y_i = 15962989$ minutes. The $1/b$ represents the mean time to failure. We can conclude by looking the likelihood function that the $y_{(1)}$ and $\sum_{i=1}^{n} y_i$ happens to be the sufficient statistics for the parameters m and b, respectively.

It was chose to use a noninformative prior, a flat prior with uniform distribution for the parameters (b, m), then it was got as posterior distribution:

$$
\begin{aligned}
\pi(b, m | \mathbf{y}) &= f(\mathbf{y} | b, m) \pi(b, m) \\
&= \left(\frac{1}{b}\right)^n \exp\left\{-\frac{1}{b}\left(\sum_{i=1}^{n} y_i - nm\right)\right\} \mathbb{1}\left(\min\{y_i\}_{i=1}^{n} \geq m\right)
\end{aligned}
$$

Now, in order to do the transformation for this posterior function, for the transformed parameters given below, and using the Jacobian matrix result:

$$
\begin{aligned}
\theta_1 &= \log(b) \\
\theta_2 &= \log(y_{(1)} - m) \\
J(\theta_1, \theta_2) &= \begin{pmatrix} \exp(\theta_1) & 0 \\ 0 & -\exp(\theta_2) \end{pmatrix}
\end{aligned}
$$

Then, the posterior distribution of the transformed parameters will be

$$
\begin{aligned}
\pi(\theta_1, \theta_2 | \mathbf{y}) &= \left(\frac{1}{e^{\theta_1}}\right)^n \exp\left\{-\frac{1}{e^{\theta_1}}\left(\sum_{i=1}^{n} y_i - n(y_{(1)} - e^{\theta_2})\right)\right\} |J(\theta_1, \theta_2)| \mathbb{1}(\theta_1 \in \mathbb{R}) \mathbb{1}(\theta_2 \in \mathbb{R}) \\
&= \left(\frac{1}{e^{\theta_1}}\right)^n \exp\left\{-\frac{1}{e^{\theta_1}}\left(\sum_{i=1}^{n} y_i - n(y_{(1)} - e^{\theta_2})\right)\right\} \exp(\theta_1 + \theta_2) \mathbb{1}(\theta_1 \in \mathbb{R}) \mathbb{1}(\theta_2 \in \mathbb{R})
\end{aligned}
$$

And the log-posterior distribution it is given by:

$$
\log(\pi(\theta_1, \theta_2 | \mathbf{y})) = -n\theta_1 - \frac{1}{e^{\theta_1}}\left(\sum_{i=1}^{n} y_i - n(y_{(1)} - e^{\theta_2})\right) + (\theta_1 + \theta_2)
$$

In Figure 1 is shown the contour plot for the joint posterior distribution of the transformed parameters (θ_1, θ_2), we can notice that the joint distribution exhibit an asymmetric shape.

Figure 1: Contour plot of the joint posterior distribution of the transformed parameters θ_1 and θ_2.
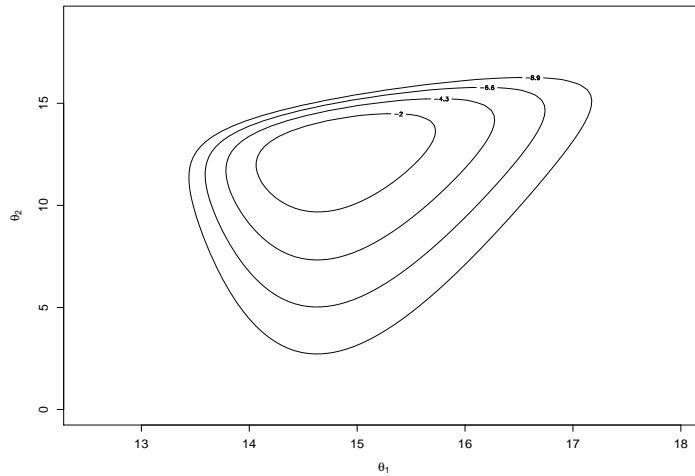
3.4

Find a normal approximation to the posterior distribution of the transformed parameters.
Solution:

Using a large n and given that the $\pi(\theta_1, \theta_2)|\mathbf{y})$ satisfy some regularity conditions, we can do the normal approximation for this posterior by: First, using the second order Taylor series expansion evaluated under $\hat{\theta}_1$ and $\hat{\theta}_2$, in which these are the maximum likelihood estimators. Then, setting $\boldsymbol{\theta} = [\theta_1, \theta_2]^T$ and $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \hat{\theta}_1]^T$, we get the approximation:

$$\log(\pi(\boldsymbol{\theta}|\mathbf{y})) \approx \log(\pi(\hat{\boldsymbol{\theta}}|\mathbf{y}) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathcal{H}_{\log(\pi(\hat{\boldsymbol{\theta}}|\mathbf{y}))}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \tag{1}$$

The results values that compose the hessian matrix $\mathcal{H}_{\log(\pi(\hat{\boldsymbol{\theta}}|\mathbf{y}))}$ are:

$$\mathcal{H}_{\log(\pi(\hat{\boldsymbol{\theta}}|\mathbf{y}))} = \begin{pmatrix} -\left[e^{-\theta_1} \left(\sum_{i=1}^{n} y_i - n(y_{(1)} - e^{\theta_2}) \right) \right] & ne^{-\theta_1}e^{\theta_2} \\ ne^{-\theta_1}e^{\theta_2} & n(-e^{-\theta_1})e^{\theta_2} \end{pmatrix}$$

Then, using that expansion in (1), the approximation for the posterior density will have as distribution: Normal $\left(\hat{\boldsymbol{\theta}}, [\mathrm{I}_f(\hat{\boldsymbol{\theta}})]^{-1} \right)$. Which we can find the Fisher information by using the hessian matrix. It was used the `optim` function in `R` to find the values for the parameters, and the results were:

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= (14.7821, 12.7027) \\ \left[\mathrm{I}_f(\hat{\boldsymbol{\theta}}) \right]^{-1} &= \begin{pmatrix} 0.1666696 & 0.1666696 \\ 0.1666696 & 1.1666411 \end{pmatrix} \end{aligned} \tag{2}$$

Figure 2 displays the contours of the normal approximation of the transformed parameters θ_1 and θ_2.

Figure 2: Contour plot of the normal approximation of the transformed parameters θ_1 and θ_2.
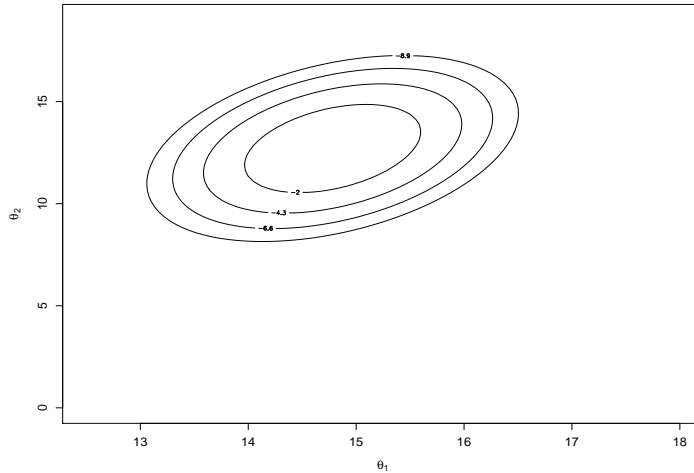
Comparing Figure 1 and 2 we can see a difference between the exact and the normal approximation for the posterior. As in Figure 1, we can see the asymmetrical shape of the distribution in the contour plot in 2 using the normal approximation.

Figure 3 shows the contour plot of the exact density overlaid on the normal approximation of the transformed parameters θ_1 and θ_2. We can notice that for value of θ_2 less than 8, the normal approximation does not make a good fit. In addition, the tails of the exact density seems to be heavier than the normal approximation for the posterior.

Figure 3: Contour plot of the exact overlaid on the normal approximation of the transformed parameters θ_1 and θ_2.
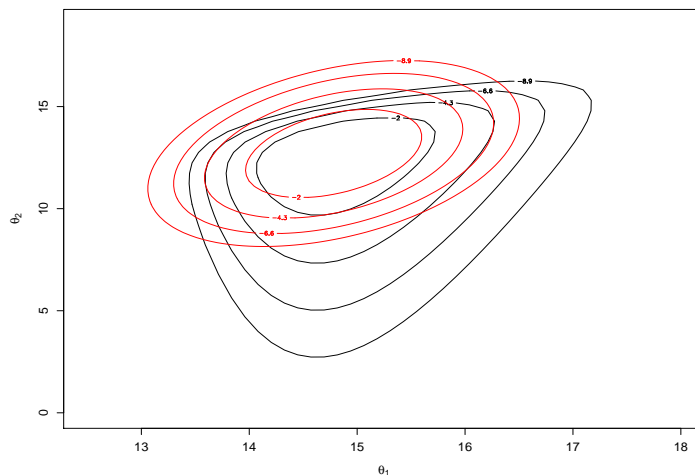
3.5

Use rejection sampling and SIR to approximate the posterior distribution. Compare.
Solution:

For the rejection sampling method; first it was chose a bivariate t distribution as proposal distribution $p(\boldsymbol{\theta})$ that resembles the posterior density, and another reason is that we need a function that has thicker tails.

The parameters for the proposal distribution are given by: $p(\boldsymbol{\theta}) \sim BVT_3(\hat{\boldsymbol{\theta}}, \hat{\Sigma})$. In which, the degree of freedom is $\nu = 3$, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ found in Question **3.3** was used as the location parameter, and covariance matrix $\hat{\Sigma}$ found in Question **3.3** was used as scale matrix. These values are in (2).

Then, in order to find the identifiable constant M, it is necessary to maximize the function $\log(\pi(\boldsymbol{\theta}|\mathbf{y})) - \log(p(\boldsymbol{\theta}))$. And it was found that the maximum value $\log(M)$ occurs at the value of $\boldsymbol{\theta} = (15.6444, 13.5649)$. In which we can notice that this values of $\boldsymbol{\theta}$ are not at the extreme part of the parametric space. This gives an indication that the value found for M is in fact an approximate maximum. The value found was $M = -96.2857$ for the identifiable constant.

The original sample size was 15,000, and the acceptance rate was 0.58, being a reasonable value. Figure 4 shows the contour plot with the simulated draws from the rejection method.

We can see in Figure 4 that most of the draws fall within the inner contour of the posterior density. However, at the two upper ends there are no points.

Figure 4: Contour plot of transformed parameters θ_1 and θ_2 with simulated draws from the rejection algorithm.
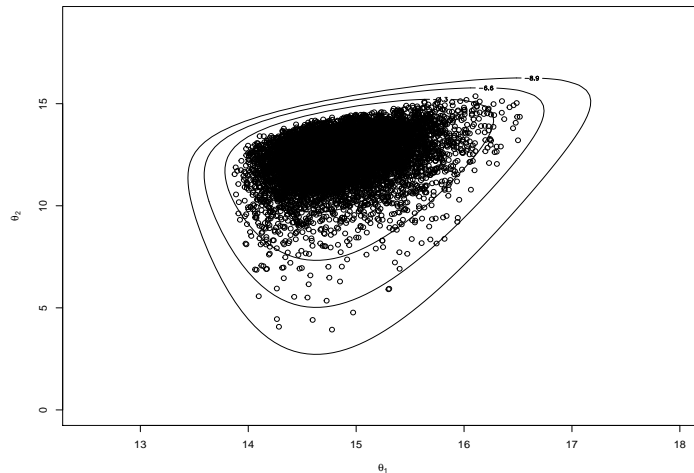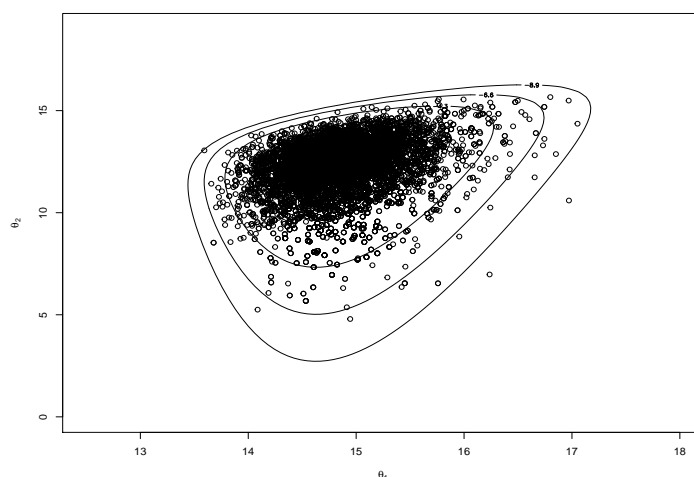

For the sampling importance resampling (SIR), as in the rejection method it was chose a bivariate t distribution as proposal distribution $p(\boldsymbol{\theta})$, where $p(\boldsymbol{\theta}) \sim BVT_3(\hat{\boldsymbol{\theta}}, \hat{\Sigma})$. Then, it was computed the weights.

The sample size was chose as the size of the accepted sample in the rejection method $n = 8794$. Figure 5 shows the contour plot of the transformed parameters $(\boldsymbol{\theta})$ with the simulated draws from the SIR method.

Looking at Figure 5 it is noticed that even two points being out of the contour plot, the most of the draws fall within the inner contour of the posterior density, and now the two upper ends are more covered by the simulated points.

Figure 5: Contour plot of transformed parameters θ_1 and θ_2 with simulated draws from the SIR algorithm.

Thus, by looking the scatted plots in Figure 4 and 5 of the simulated points using the two methods, seems that the SIR method gets more accurate results for the simulated points because

5

the simulated draws had covered better the exact density of the posterior distribution.

3.6

Use importance sampling as well as a Laplace approximation to estimate the posterior mean and variance of the transformed parameters.
Solution:

Using the importance sampling, it was chose a bivariate t distribution as proposal distribution $p(\boldsymbol{\theta})$, where $p(\boldsymbol{\theta}) \sim BVT_3(\hat{\boldsymbol{\theta}}, \hat{\Sigma})$, the same parameters as in previous methods.

It was found for the posterior mean $\mathbb{E}(\boldsymbol{\theta}) = (14.8658, 12.2319)$, and for the variance $\mathbb{V}\mathrm{ar} = (0.1787, 1.8085)$

The values found in (2) are the estimates for the posterior moments of the approximating normal density, in summary, we have $\mathbb{E}(\boldsymbol{\theta}) = (14.7821, 12.7027)$, and for the variance $\mathbb{V}\mathrm{ar} = (0.1666, 1.1666)$. Which this approximation is called first order approximation. However, these are not a good estimates given that our exact density differs a little from the normal distribution.

However, as said by Carlin & Louis (2008) "the second-order accuracy provided by Laplace's method is comforting, and is surely better than the first-order accuracy provided by the normal approximation for any given dataset we will still lack a numerical measurement of how far our approximate posterior expectations are from their exact values".

Reference
Carlin, B. P., & Louis, T. A. (2008). Bayesian methods for data analysis. CRC Press.
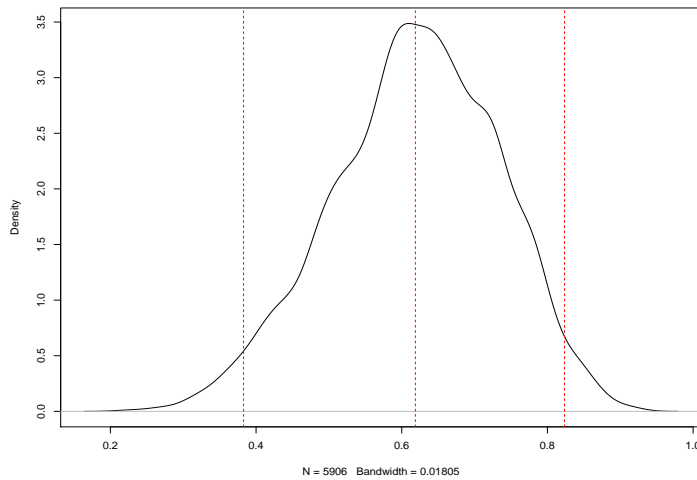
3.7

Define the reliability at time t_0 as $\mathrm{R}(t_0) = \exp(-(t_0-m)/b)$. Describe the posterior moments and the posterior distribution of $\mathrm{R}(10^6)$.
Solution:

The reliability in this case is the probability that a pump does not fail before time t_0. In Figure 6 is shown the posterior density for the reliability at time $t_0 = 10^6$.

Figure 6: Posterior density for the reliability at time $t_0 = 10^6$.

The moments found for this distribution were: $\mathbb{E}(R(10^6)) = 0.6197$ which represents the probability mean of the pump does not fail before 10^6 minutes. For the variance, $\mathbb{V}ar(R(10^6)) = 0.0128$, and the credible interval $(0.3854099, 0.8216955)$.

R Code

See below all the R Code used to perform the analysis.

```
rm(list=ls(all=TRUE))
library(mvtnorm)

# pump failured
n=8

# sufficient statistic
y1=23721
sumy=15962989

# part 3.3
logpost = function(theta,postpar) {
  theta1 = theta[1]
  theta2 = theta[2]
  sumy = postpar$sumy
  y1 = postpar$y1
  n = postpar$n
  logpostr = -n*theta1 - ((1/exp(theta1)*(sumy - n*(y1 - exp(theta2)))))
  + theta1 + theta2
  return(logpostr)}

postpar=list(sumy=sumy,y1=y1,n=n)

logf = function(theta, postpar) {
  if (is.matrix(theta) == TRUE) {
    val = matrix(0, c(dim(theta)[1], 1))
    for (j in 1:dim(theta)[1]) val[j] = logpost(theta[j,], postpar)
  }
  else val = logpost(theta, postpar)
  return(val)
}

limits=c(12,18,0,25)
ng=50
x0 = seq(limits[1], limits[2], len = ng)
y0 = seq(limits[3], limits[4], len = ng)
X = outer(x0, rep(1, ng))
Y = outer(rep(1, ng), y0)
n2 = ng^2
den = logf(cbind(X[1:n2], Y[1:n2]), postpar)
den = den - max(den)
den = matrix(den, c(ng, ng))
contour(x0, y0, den,levels = seq(-8.9, 0, by = 2.3), xlim=c(12.5,18),
ylim=c(0,19),lwd = 1, xlab=expression(theta[1]),
        ylab=expression(theta[2]),cex.axis=1.2, cex.lab=1.2)
```

```
# part 3.4
start=c(25,25)
theta_mle = optim(par=start, logpost, postpar=postpar, method="BFGS",
hessian = TRUE, control = list(fnscale = -1))
theta_mle$par
theta_mle$hessian

fisher = solve(-theta_mle$hessian)
fisher

mean1 = c(theta_mle$par[1],theta_mle$par[2])
Var1  = fisher
tpar=list(m=mean1,var=Var1,df=1)

logf_n = function(theta, tpar) {
  if (is.matrix(theta) == TRUE) {
    val = matrix(0, c(dim(theta)[1], 1))
    for (j in 1:dim(theta)[1]) val[j] = dmvnorm(theta[j,], mean = tpar$m,
    sigma = tpar$var,log=TRUE)
  } else val = dmvnorm(theta, mean = tpar$m, sigma = tpar$var,log=TRUE)
  return(val)
}

limits=c(12,18,0,25)
ng=100
x0 = seq(limits[1], limits[2], len = ng)
y0 = seq(limits[3], limits[4], len = ng)
X = outer(x0, rep(1, ng))
Y = outer(rep(1, ng), y0)
n2 = ng^2
den1 = logf_n(cbind(X[1:n2], Y[1:n2]), tpar)
den1 = den1 - max(den1)
den1 = matrix(den1, c(ng, ng))
contour(x0, y0, den1, levels = seq(-8.9, 0, by = 2.3), xlim=c(12.5,18),
ylim=c(0,19),lwd = 1, xlab=expression(theta[1]),
        ylab=expression(theta[2]),cex.axis=1.2, cex.lab=1.2)

# part 3.5
# reject sampling
log_rej = function(theta,postpar,tpar){
  l_rej = logpost(theta,postpar) - dmt(theta, mean = c(tpar$m), S = tpar$var,
   df = tpar$df,log=TRUE)
  return(l_rej)
}

tpar=list(m=mean1,var=Var1,df=3)

start=c(15,15)
```

```
theta_rej=optim(par=start,log_rej,postpar=postpar,tpar=tpar,method="BFGS",
    hessian = TRUE, control = list(fnscale = -1))
theta_rej$par

dmax=log_rej(theta_rej$par,postpar=postpar,tpar=tpar)
dmax


# accept/reject method
set.seed(7)
M = 15000
d = length(tpar$m)
theta = rmt(M,mean = c(tpar$m), S = tpar$var, df = tpar$df)
lf = matrix(0, c(dim(theta)[1], 1))

for (j in 1:dim(theta)[1]) lf[j] = logpost(theta[j,],postpar=postpar)

lg = dmt(theta,  mean = c(tpar$m), S = tpar$var, df = tpar$df, log = TRUE)
if (d == 1) {
  prob = exp(c(lf) - lg - dmax)
  draws1=(theta[runif(n) < prob])
} else {
  prob = exp(lf - lg - dmax)
  draws1=(theta[runif(n) < prob, ])
}

nd = nrow(draws1)
accept_rate = nd/M
accept_rate

plot(draws1[,1],draws1[,2],xlab=expression(theta[1]),xlim=c(12.5,18),
ylim=c(0,19),ylab=expression(theta[2]),pch=1)
contour(x0, y0, den, levels = seq(-8.9, 0, by = 2.3), xlim=c(12.5,18),
ylim=c(0,19),lwd = 1, xlab=expression(theta[1]),ylab=expression(theta[2]),
        cex.axis=1.2, cex.lab=1.2, add=TRUE)

# SIR method
set.seed(7)
sir_sexp = function (logf, tpar, n, infor)
{
  k = length(tpar$m)
  theta = rmt(n, mean = c(tpar$m), S = tpar$var, df = tpar$df)
  lf = matrix(0, c(dim(theta)[1], 1))
  for (j in 1:dim(theta)[1]) lf[j] = logf(theta[j, ], infor)
  lp = dmt(theta, mean = c(tpar$m), S = tpar$var, df = tpar$df,
           log = TRUE)
  md = max(lf - lp)
  wt = exp(lf - lp - md)
  probs = wt/sum(wt)
  indices = sample(1:n, size = n, prob = probs, replace = TRUE)
```

```
    if (k > 1)
      theta = theta[indices, ]
    else theta = theta[indices]
    return(theta)
}

theta=sir_sexp(logpost,tpar,nd,postpar)

plot(theta[,1],theta[,2],xlim=c(12.5,18),ylim=c(0,19),
       xlab=expression(theta[1]),ylab=expression(theta[2]),pch=1)
contour(x0, y0, den, levels = seq(-8.9, 0, by = 2.3), xlim=c(13,17),
ylim=c(0,19),lwd = 1,xlab=expression(theta[1]),ylab=expression(theta[2]),
         cex.axis=1.2, cex.lab=1.2, add=TRUE)


# part 3.6
# importance sampling
post = function(theta,postpar) {
  sumy = postpar$sumy
  y1 = postpar$y1
  n = postpar$n
  theta1 = theta[1]
  theta2 = theta[2]
  logpostr = -n*theta1 - ((1/exp(theta1)*(sumy - n*(y1 - exp(theta2))))))
  + theta1 + theta2
  return(logpostr)}

tpar=list(m=mean1,var=Var1,df=3)
postpar=list(sumy=sumy,y1=y1,n=n)

imporsampl = function (logf, tpar, h, n, data)
{
  theta = rmt(n, mean = c(tpar$m), S = tpar$var, df = tpar$df)
  lf = matrix(0, c(dim(theta)[1], 1))
  for (j in 1:dim(theta)[1]) lf[j] = logf(theta[j, ], data)
  H = lf
  for (j in 1:dim(theta)[1]) H[j] = h(theta[j, ])
  lp = dmt(theta, mean = c(tpar$m), S = tpar$var, df = tpar$df,
           log = TRUE)
  md = max(lf - lp)
  wt = exp(lf - lp - md)
  est = sum(wt * H)/sum(wt)
  SEest = sqrt(sum((H - est)^2 * wt^2))/sum(wt)
  return(list(est = est, se = SEest, theta = theta, wt = wt))
}

# mean
myfunc=function(theta) return(theta[2])
```

```
s=imporsampl(post,tpar,myfunc,10000,postpar)
cbind(s$est,s$se)
mu = s$est
mu

hist(s$wt,freq=FALSE)

# second moment
myfunc=function(theta) return(theta[1]^2)

s=imporsampl(post,tpar,myfunc,10000,postpar)
cbind(s$est,s$se)
mu2 = s$est
mu2

var_theta = mu2 - mu^2
var_theta

# variance
myfunc=function(theta) return((theta[2]-mu)^2)

s=imporsampl(post,tpar,myfunc,10000,postpar)
cbind(s$est,s$se)
var2 = s$est
var2


# part 3.7
b = function(theta) return(exp(theta[,1]))
m = function(theta, smin) return(smin - exp(theta[,2]))

t0 = 10^6
R = function(t0,theta,smin) return(exp(-(t0 - m(theta,smin))/b(theta)))
hist(R(t0,theta,postpar$y1))
plot(density(R(t0,theta,postpar$y1)), main="")
abline(v=mean(R(t0,theta,postpar$y1)), col=2, lty=2)
abline(v=quantile(R(t0,theta,postpar$y1), probs = c(0.025, 0.975)), col=2, lty=2)

mean(R(t0,theta,postpar$y1))
var(R(t0,theta,postpar$y1))
sd(R(t0,theta,postpar$y1))
quantile(R(t0,theta,postpar$y1), probs = c(0.025, 0.975))
```