

AMS 207 HW Assignment 1

Mary Silva

April 14 2018

3 Failure time

The failure time of a pump follows a two-parameter exponential distribution, $f(y|b,m) = 1/b \exp(-(y-m)/b)$, when $y \geq m$.

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

$$f(y_1 \cdots y_n | b, m) = \left(\frac{1}{b}\right)^n \exp\left\{-\sum_{i=1}^n \frac{y_i - m}{b}\right\} I_{(m, \infty)}(y_{(1)})$$

where $y_{(1)}$ is the minimal of y_1, \dots, y_n .

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

Notice that the support of posterior distribution of m is $(-\infty, y_{(1)})$. To map m to the entire real line, let $t = \log(y_{(1)} - m)$, $t \in R$. Let $v = \log(b)$, $v \in R$.

3. Assume a sensible prior for the transformed parameters and explore the contours of the posterior distribution.

Assume uniform prior on m and b .

$$\pi(m, b) \propto 1$$

Then the posterior is defined up to a constant by

$$f(m, b | y) \propto \left(\frac{1}{b}\right)^n \exp\left\{-\sum_{i=1}^n \frac{y_i - m}{b}\right\} I_{(m, \infty)}(y_{(1)})$$

Given $v = \log(b)$, $t = \log(y_{(1)} - e^t)$, $J = e^v e^t$

$$\begin{aligned} f(m, v | y) &\propto e^{-vn} \exp\left\{-\left(\sum_{i=1}^n y_i - ny_{(1)} + ne^t\right)e^{-v}\right\} e^v e^t \\ &= e^{-v(n-1)} \exp\left\{-\left(\sum_{i=1}^n y_i - ny_{(1)} + ne^t\right)e^{-v}\right\} e^t \end{aligned}$$

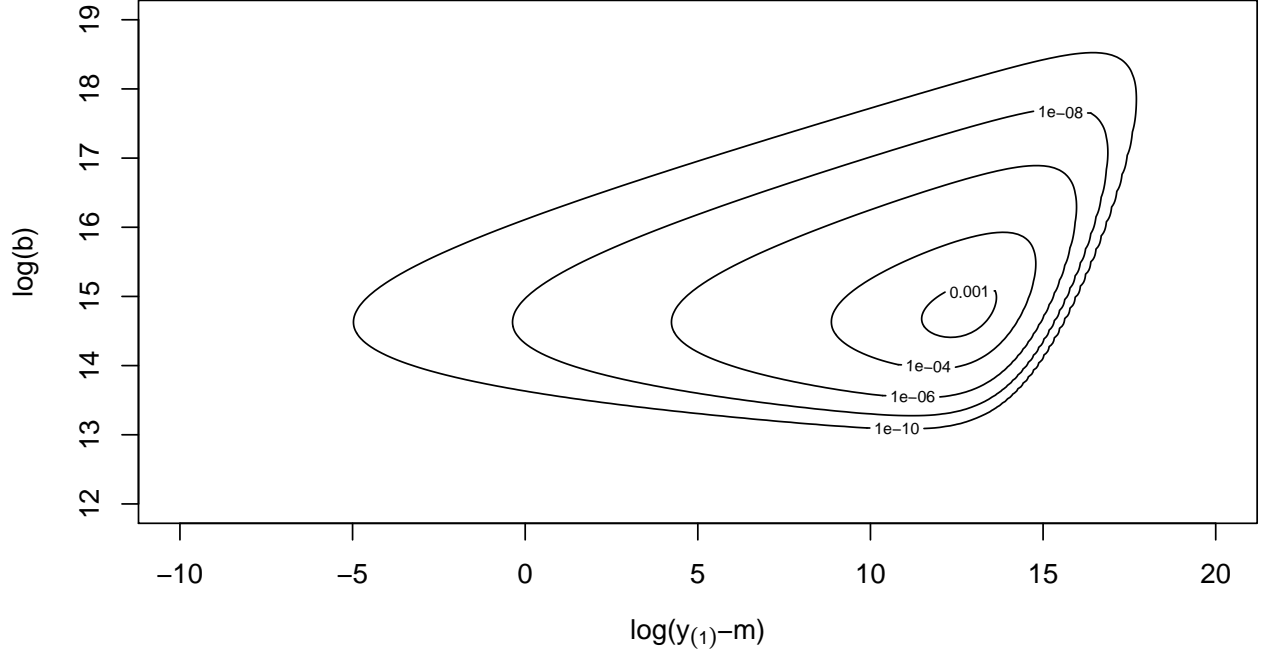


Figure 1: Contour plot of the joint posterior

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

$$\frac{\partial}{\partial v} \log(f, t|y) = -(n-1) + \left(\sum_{i=1}^n y_i - ny_{(1)} + ne^t \right) e^{-v} = 0$$

$$\frac{\partial}{\partial t} \log(f, t|y) = -ne^{-v} e^t + 1 = 0$$

Solving these two equations together we have

$$\hat{v} = \log\left(\frac{\sum_{i=1}^n y_i - ny_{(1)}}{n-2}\right)$$

$$\hat{t} = \log\left(\frac{\sum_{i=1}^n y_i - ny_{(1)}}{n(n-2)}\right)$$

Now take the second derivative to get the information matrix.

$$\begin{aligned} I(t, v) &= \begin{bmatrix} \frac{\partial^2 \log(f(v, t|y))}{\partial t^2} & \frac{\partial^2 \log(f(v, t|y))}{\partial v \partial t} \\ \frac{\partial^2 \log(f(v, t|y))}{\partial v \partial t} & \frac{\partial^2 \log(f(v, t|y))}{\partial v^2} \end{bmatrix} \\ &= \begin{bmatrix} -ne^t e^{-v} & ne^t e^{-v} \\ ne^t e^{-v} & -(\sum_{i=1}^n y_i - ny_{(1)} + ne^t) e^{-v} \end{bmatrix} \end{aligned}$$

Notice that if we write $A = \sum_{i=1}^n y_i - ny_{(1)}$, then lots of the terms cancel out when $I(v, t)$ is evaluated at (\hat{v}, \hat{t}) . We end up with a information matrix at the posterior mode

$$I(\hat{t}, \hat{v}) = \begin{bmatrix} -1 & 1 \\ 1 & -(n-1) \end{bmatrix}$$

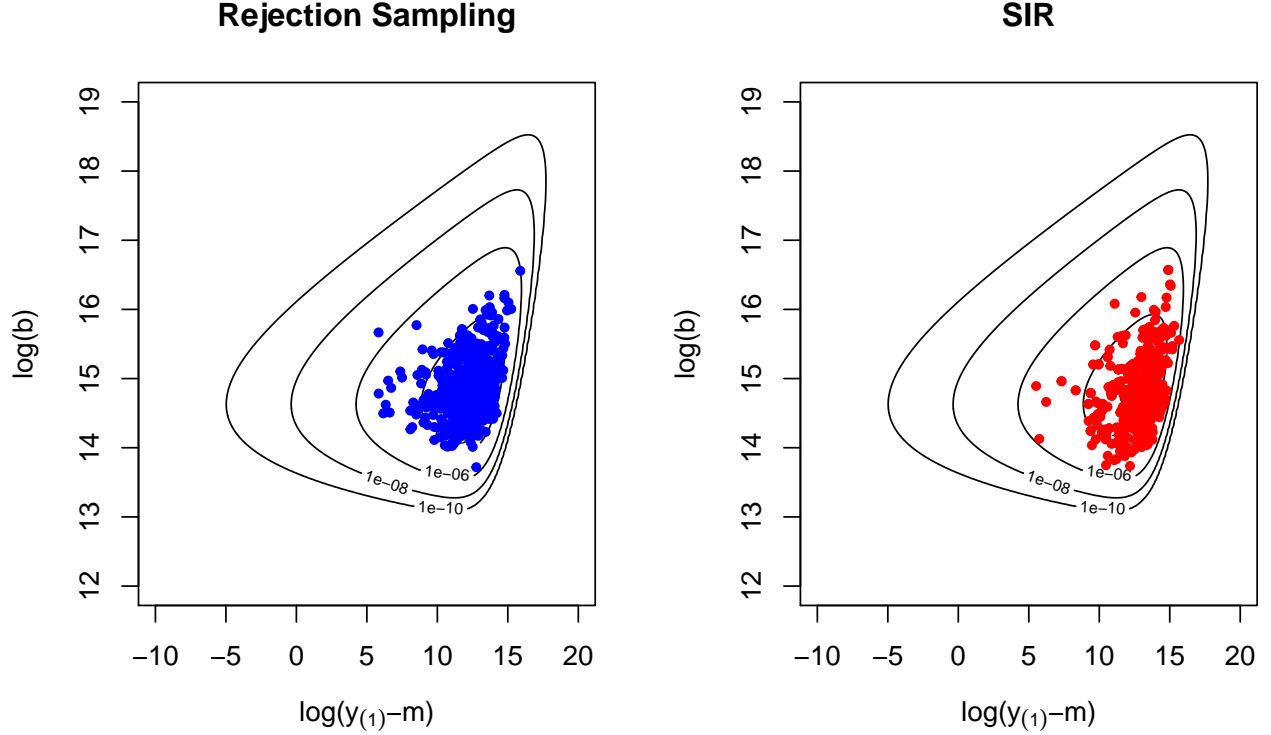


Figure 2: Comparing Rejection Sampling and SIR, $n=200$

The variance covariance matrix for normal approximation is given by $\Sigma = (-I(\hat{v}, \hat{t}))^{-1}$. I will leave the matrix inversion to R.

Table 1: Variance covariance matrix of normal approximation

	t	v
t	1.1666667	0.1666667
v	0.1666667	0.1666667

$(\hat{t}, \hat{v}) = (12.4149, 14.2431)$. The variance covariance matrix is positive definite (yay!).

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

For rejection sampling I used a bivariate t with 4 degree of freedom, mean (\hat{t}, \hat{v}) , and variance 2Σ from the normal approximation as the proposal density. The log of density ratio upper bound is optimized to be -95.73.

Acceptance rate is 0.3646775

The acceptance rate for rejection sampling is roughly 37%. From figure 2, both sample seem to good representations of the distribution since majority of the points centered around the mode.

To ensure better discretization of the density, I forced the proposal sample for SIR to be 5 times the desired sample size. When not taking such measure, the sample resulted from SIR is less representative of the real posterior distribution. Computationally, both algorithm need to sample more than the desired sample size to get a good representation of the distribution. However, using rejection sampling is potentially time consuming if the acceptance rate is low whereas SIR can render a sample fairly quickly.

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

Use importance sampling

Define $h((t, v)) = (t, v)$, define $w((t, v)) = \frac{p(t, v)}{q(t, v)}$, where p is the posterior density and q is the density of multivariate t with 3 degree of freedom, location at the posterior mode and 2Σ .

Then the expectation of (t, v) is given by

$$E((t, v)) = \frac{\sum_{i=1}^m (t_i, v_i) w((t_i, v_i))}{\sum_{i=1}^m w((t_i, v_i))}$$

The variance is given by

$$Var(t) = \frac{\sum_{i=1}^m (t - \bar{t})^2 w((t_i, v_i))}{\sum_{i=1}^m w((t_i, v_i))}$$

$$Var(v) = \frac{\sum_{i=1}^m (v - \bar{v})^2 w((t_i, v_i))}{\sum_{i=1}^m w((t_i, v_i))}$$

where $(\bar{t}, \bar{v}) = E((t, v))$ from previous step.

Table 2: Posterior mean and variance using Importance Sampling

	$\log(y_{(1)} - m)$	$\log(b)$
Mean	12.24991	14.8818783
Var	1.86855	0.1914875

Use Laplace approximation

Implementation Notes on using Laplace approximation to find $E(g(\theta))$

1. Define $h_1 = \log(g(\theta) * p(\theta|y))$, find its maximum \hat{h}_1 and hessian at mode H_1 .
2. Calculate the determinant of $\Sigma_1 = [-H_1]^{-1}$ to be d_1
3. Define $h = \log(p(\theta|y))$, find its maximum \hat{h} and determinant of inverse of hessian at mode to be d .
4. $E(g(\theta)) = \frac{\exp(h_1)d_1}{\exp(\hat{h})d}$

Since both $t \in R, v \in R$, apply the hack of adding a large number (in my case 100) to both in order to define $g((t, v))$, i.e. $g((t, v)) = (t + 100, v + 100)$ so that $\log(g((t, v)))$ is effectively greater than 0 for optimization purpose. Empirically we are somewhat confident that the mean of t and v are greater than 0, which provides some support for such approach.

However, one draw back of this method is that the approximation linearly depends on the choice of this constant. If a larger constant is chosen instead, the approximation for expectation of t and v both increases. Therefore it is questionable whether this “hack” is reliable.

Table 3: Posterior mean and variance using Laplace approximation

	$\log(y_{(1)} - m)$	$\log(b)$
Mean	12.174751	14.7805855
Var	1.127309	0.1363237

$\log(y_{(1)} - m)$	$\log(b)$
---------------------	-----------

Notice that the results are highly dependent on the starting points of optimization process. So maybe analytical solution to the mode and hessian should be implemented.

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

Acquire a sample of 2000 points with SIR since it is faster. Transform t, v back to m, b by $m = y_{(1)} - e^t, b = e^v$. Then get a sample of $R(10^6)$ by letting $R^{(i)} = \exp\{-(-m^{(i)} + 10^6)/b\}$. Find the sample moments of $\{R^{(i)}\}$ to be approximations of the posterior moments according to WLLN.

Table 4: Summary of posterior distribution of $R(10^6)$

	Mean	Sd	2.5%	25%	50%	75%	97.5%
$R(10^6) Y$	0.5751771	0.1279857	0.3183528	0.4831401	0.580247	0.6635922	0.8101091

