

# Computation II

Statistics 220

Spring 2005



# Acceptance-Rejection and Log-concave Densities

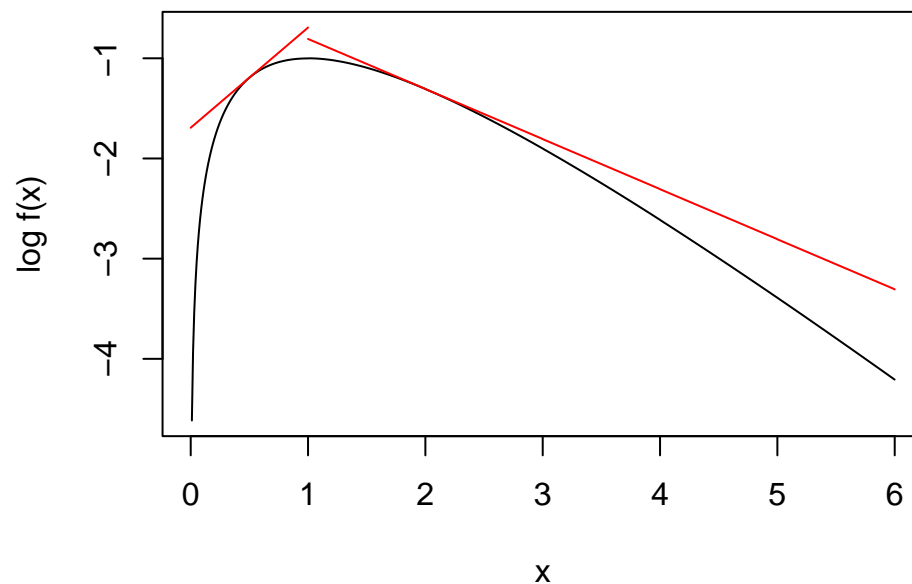
There is one class of distributions where it is easy to set up an acceptance-rejection scheme.

It is the case when the log of the density function is concave on the support of the distribution.

If  $f(x)$  is log concave, any tangent line to  $\log f(x)$  will lie above  $\log f(x)$  (call it  $l(x) = a + bx$ ).

Thus  $h(x) = e^{l(x)} = e^a e^{bx}$  lies above  $f(x)$ . Note that this looks like a scaled exponential density.

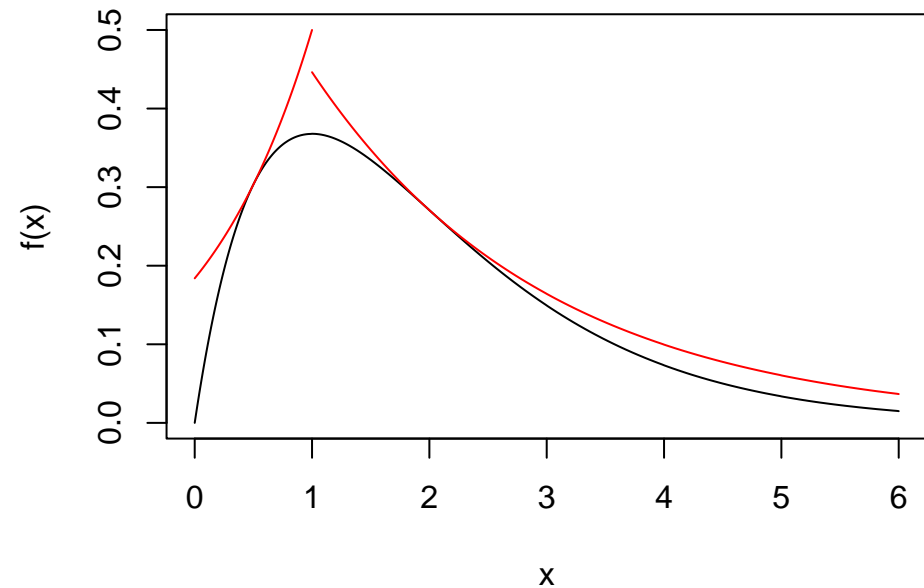
This suggests that exponential distributions can be used as the majorizing distribution.



A strictly log concave density is unimodal. The mode may occur at an endpoint or in the middle.

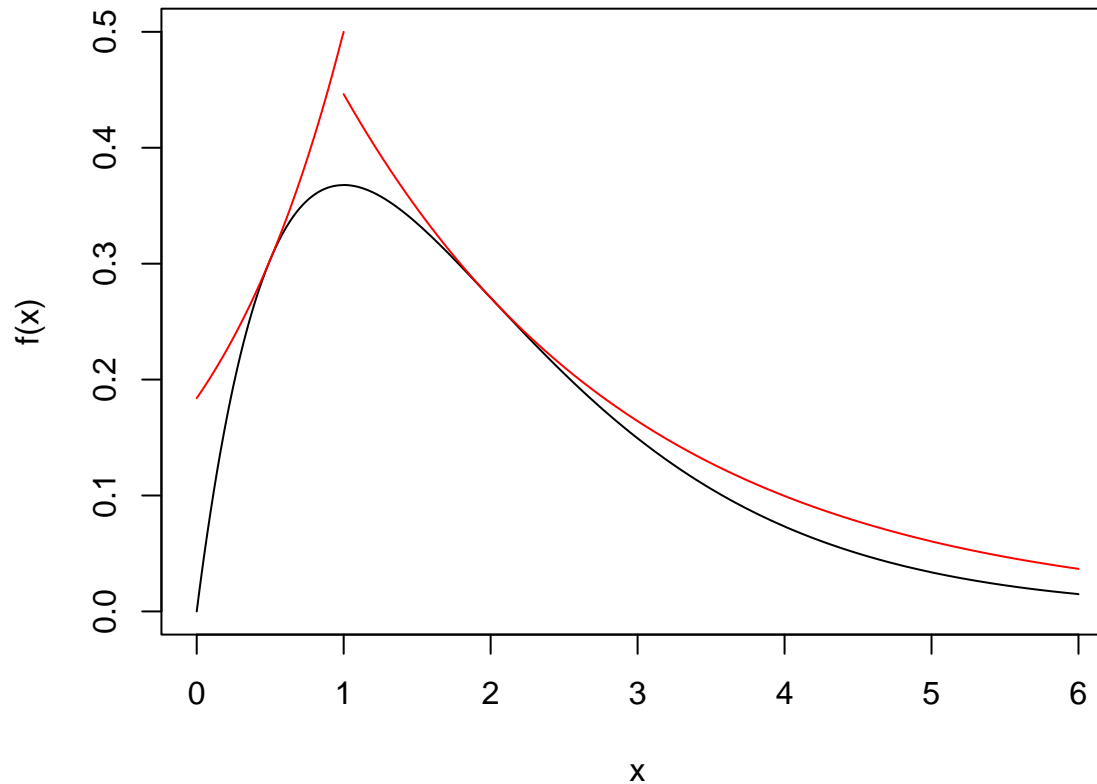
If the mode occurs at an endpoint, a single exponential can be used (as with the half normal example).

If the mode occurs in the middle of the range, two exponential envelopes are needed, one for left of the mode, the other for right of the mode).



Example: Gamma(2,1)

The mode for a Gamma is  $\frac{\alpha-1}{\beta}$  (so 1 for this example). So lets use shifted exponentials starting at the mode of 1.



- Left side:

$$g_l(x) = \frac{1}{\mu_l} \exp((x - 1)/\mu_l) I(x < 1)$$

- Right side:

$$g_r(x) = \frac{1}{\mu_r} \exp(-(x - 1)/\mu_r) I(x \geq 1)$$

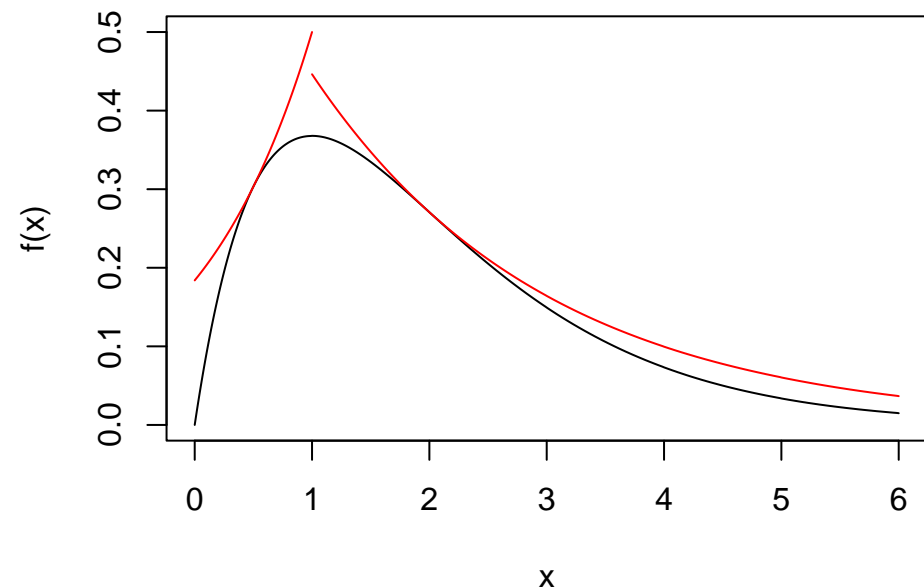
The choice of  $\mu_l$  and  $\mu_r$  depend on where you want the majorized densities  $g_l(x)$  and  $g_r(x)$  to be tangent to  $f(x)$ .

Lets set  $x_l = 0.5$  and  $x_r = 2$  as the tangent points.

The total area under

$$h(x) = c_l g_l(x) + c_r g_r(x)$$

is  $c = c_l + c_r$ .



For the example,  $c_l = 0.5$  and  $c_r = 0.8925$  so the rejection rate for this sampler is just under 30%.

To determine which exponentials to use, solve the system (for given  $x_l$  and  $x_r$ )

$$f(x_l) = c_l g_l(x_l) \quad f(x_r) = c_l g_r(x_r)$$

$$f'(x_l) = c_l g'_l(x_l) \quad f'(x_r) = c_l g'_r(x_r)$$

Solving this gives

$$\lambda_l = \frac{f'(x_l)}{f(x_l)} \quad \lambda_r = -\frac{f'(x_r)}{f(x_r)}$$
$$c_l = \frac{(f(x_l))^2}{f'(x_l)} e^{-\lambda_l(x_l-m)} \quad c_r = -\frac{(f(x_r))^2}{f'(x_r)} e^{\lambda_l(x_r-m)}$$

where  $m$  is the mode of  $f(x)$  and  $\lambda_i = \frac{1}{\mu_i}$

The optimal choices for  $x_l$  and  $x_r$  can be found by minimizing  $c_l$  and  $c_r$  separately.

## Implementing the proposal distribution

The majorizing distribution  $g(x)$  is a mixture distribution

$$g(x) = \frac{c_l}{c_l + c_r} g_l(x) + \frac{c_r}{c_l + c_r} g_r(x)$$

So for the proposal  $x$ , you want to sample from  $g_l(x)$  with probability  $\frac{c_l}{c_l + c_r}$  and from  $g_r(x)$  with probability  $\frac{c_r}{c_l + c_r}$ .

This can be implemented by

1. Draw  $v \sim U(0, 1)$
2. If  $v < \frac{c_l}{c_l + c_r}$   
draw  $y \sim \text{Exp}(\mu_l)$  and set  $x = m - y$ .  
else  
draw  $y \sim \text{Exp}(\mu_r)$  and set  $x = m + y$

## Discrete Log-concave Distributions

Let  $X$  be a random variable defined on the non-negative integers. In this case, log-concave is defined as

$$\log f(x) \geq \frac{1}{2} [\log f(x-1) + \log f(x+1)]$$

which is equivalent to

$$(f(x))^2 \geq f(x-1)f(x+1)$$

for all integers  $x$ ,

A possible majorizing distribution in the discrete case is the geometric distribution with

$$g(x) = p(1-p)^x \quad x = 0, 1, 2, \dots$$

(See Lange for further details, particularly on  $p_l, p_r, x_l, x_r$ )



# Markov Chain Monte Carlo (MCMC)

Instead of generating independent samples, generate dependent samples via a Markov Chain

$$\theta^0 \rightarrow \theta^1 \rightarrow \theta^2 \rightarrow \theta^3 \rightarrow \dots$$

where the stationary distribution of the chain is the desired distribution  $p(\theta)$ .

The Markov Chain is defined by a transition distribution  $T_t(\theta^t|\theta^{t-1})$ , which describes the possible moves when you are in state  $\theta^{t-1}$

Useful for a wide range of problems.

Popular for Bayesian analyses, but it is a general sampling procedure. For example, it has been used for calculating likelihoods in genetic linkage analysis.

# Gibbs Sampling

One example of MCMC

Idea: Break the random variable  $\theta$  in  $k$  pieces ( $\theta = \{\theta_1, \theta_2, \dots, \theta_k\}$ ) and sample the pieces sequentially. (The pieces  $\theta_i$  could be univariate or multivariate.)

1. Initialize chain:  $\theta^0 = \{\theta_1^0, \theta_2^0, \dots, \theta_k^0\}$  by some mechanism.
2. At time  $t$ , sample  $\theta^t = \{\theta_1^t, \theta_2^t, \dots, \theta_k^t\}$  by
  - Step 1: sample  $\theta_1^t \sim p(\theta_1 | \theta_2^{t-1}, \dots, \theta_k^{t-1})$
  - Step 2: sample  $\theta_2^t \sim p(\theta_2 | \theta_1^t, \theta_3^{t-1}, \dots, \theta_k^{t-1})$
  - Step  $j$ : sample  $\theta_j^t \sim p(\theta_j | \theta_1^t, \dots, \theta_{j-1}^t, \theta_{j+1}^{t-1}, \dots, \theta_k^{t-1})$
  - Step  $k$ : sample  $\theta_k^t \sim p(\theta_k | \theta_1^t, \dots, \theta_{k-1}^t)$

Under certain regularity conditions, the realizations  $\theta^1, \theta^2, \theta^3, \dots$  form a Markov chain with stationary distribution  $p(\theta)$ . Thus the realizations can be treated as dependent samples from the desired distribution.

Example: Nuclear Pump Failure

Gaver & O'Muircheartaigh (Technometrics, 1987)

Gelfand & Smith (JASA, 1990)

Observed 10 nuclear reactor pumps and counted the number of failures for each pump.

Pump	Failures ( $s_i$ )	Observation Time ( $t_i$ )	Observed Rate ( $l_i$ )
1	5	94.320	0.053
2	1	15.720	0.064
3	5	62.880	0.080
4	14	125.760	0.111
5	3	5.240	0.573
6	19	31.440	0.604
7	1	1.048	0.954
8	1	1.048	0.954
9	4	2.096	1.910
10	22	10.480	2.099

Observation time in 1000's of hours

Observed Rate = # Failure / 1000 hours

Want to determine the true failure rate for each pump with the following hierarchical model

$$\begin{aligned} s_i | \lambda_i &\stackrel{ind}{\sim} \text{Poisson}(\lambda_i t_i) \\ \lambda_i | \beta &\stackrel{iid}{\sim} \text{Gamma}(\alpha, \beta) \\ \beta &\sim \text{Gamma}(\gamma, \delta) \end{aligned}$$

Note that this is a slightly different parameterization but the same model that Gelfand and Smith used.

In this example,  $\alpha$  will be assumed to be a fixed parameter. We could be a prior on it, or as Gelfand and Smith do, estimate it from the data and take an empirical Bayes solution.

Want to determine the following posterior distributions

1.  $p(\lambda_i|s)$  for each pump
2.  $p(\beta|s)$  and  $p\left(\frac{1}{\beta}|s\right)$

Note the both sets of these distributions are difficult to get analytically. It is possible to show that

$$p(\lambda|s) \propto \frac{1}{(\delta + \sum \lambda_i)^{10\alpha+\gamma}} \prod \frac{t_i^{\alpha+s_i} \lambda^{\alpha+s_i-1} e^{-\lambda t_i}}{\Gamma(\alpha + s_i)}$$

Note that the  $\lambda$ 's are correlated and trying to get the marginal for each looks to be intractable analytically.

Instead lets run a Gibbs sampler to determine  $p(\lambda, \beta|s)$  from which we can get the desired posteriors.

One possible Gibbs scheme is

- Step 1: sample  $\lambda_1 \sim p(\lambda_1 | \lambda_{(-1)}, \beta, s)$
- Step 2: sample  $\lambda_2 \sim p(\lambda_2 | \lambda_{(-2)}, \beta, s)$
- ...
- Step 10: sample  $\lambda_{10} \sim p(\lambda_{10} | \lambda_{(-10)}, \beta, s)$
- Step 11: sample  $\beta \sim p(\beta | \lambda, s)$

where  $\lambda_{(-j)} = \{\lambda_1, \dots, \lambda_{j-1}, \lambda_{j+1}, \dots, \lambda_{10}\}$

Need the following conditional distributions

$$\begin{aligned}\lambda_j &\sim p(\lambda_j | \lambda_{(-j)}, \beta, s) = p(\lambda_j | \beta, s_j) \\ &= \text{Gamma}(\alpha + s_j, \beta + t_j)\end{aligned}$$

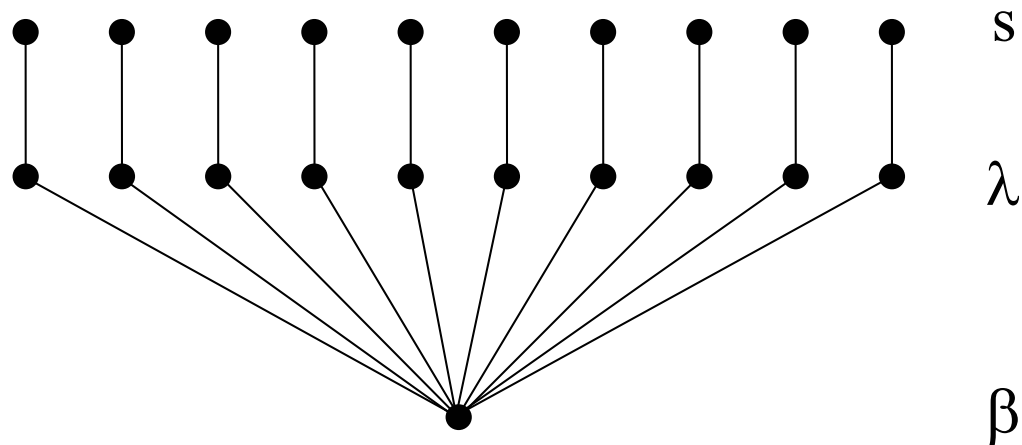
$$\begin{aligned}\beta &\sim p(\beta | \lambda, s) = p(\beta | \lambda) \\ &= \text{Gamma}(\gamma + 10\alpha, \delta + \sum \lambda)\end{aligned}$$

These can be gotten from the joint distribution by including only the terms in the product that contain the random variable of interest

$$p(s, \lambda, \beta) = \left( \prod_{i=1}^{10} \frac{(\lambda_i t_i)^{s_i} e^{-\lambda_i t_i}}{s_i!} \right) \left( \prod_{i=1}^{10} \lambda_i^{\alpha-1} \beta^{\alpha} \frac{e^{-\lambda_i \beta}}{\Gamma(\alpha)} \right) \beta^{\delta-1} \gamma^{\delta} \frac{e^{-\beta \delta}}{\Gamma(\gamma)}$$



Equivalently, you can do this by looking at the graph structure of the model by only including terms that correspond to edges joining to the node of interest. (e.g. for  $\beta$ , which edges connect with the node for  $\beta$ .)



In these graphs, for every factor of the joint distribution, the nodes for the variables in the factor are joined.

$$p(s, \lambda, \beta) = \left( \prod_{i=1}^{10} \frac{(\lambda_i t_i)^{s_i} e^{-\lambda_i t_i}}{s_i!} \right) \left( \prod_{i=1}^{10} \lambda_i^{\alpha-1} \beta^\alpha \frac{e^{-\lambda_i \beta}}{\Gamma(\alpha)} \right) \beta^{\delta-1} \gamma^\delta \frac{e^{-\beta \delta}}{\Gamma(\gamma)}$$

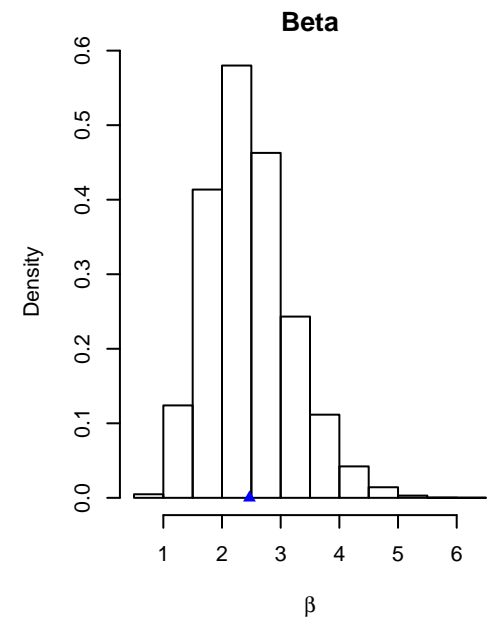
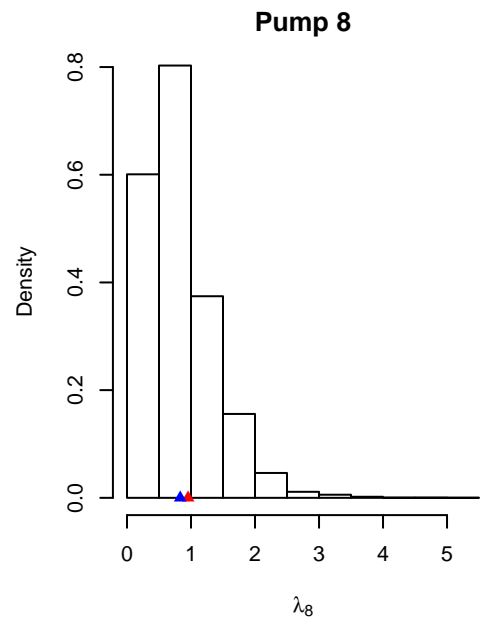
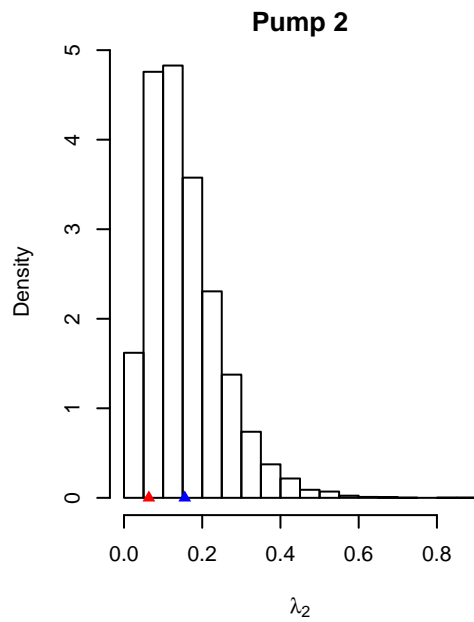
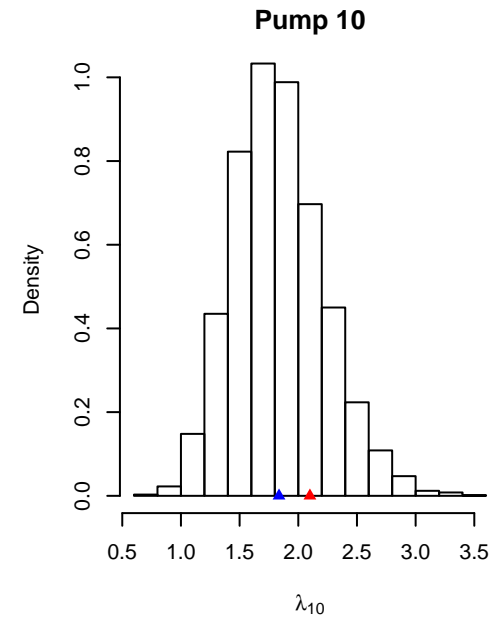
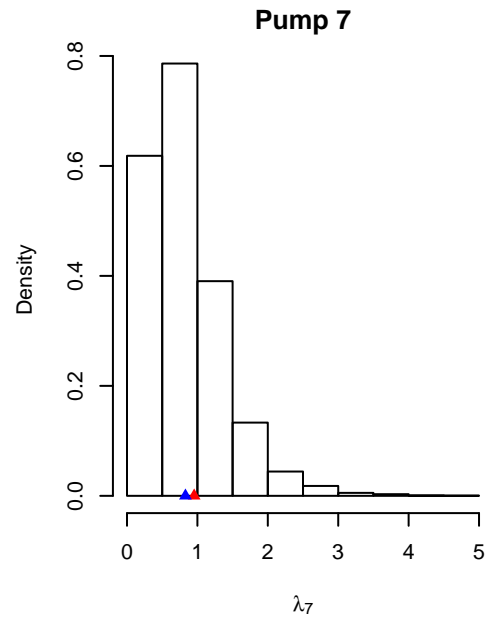
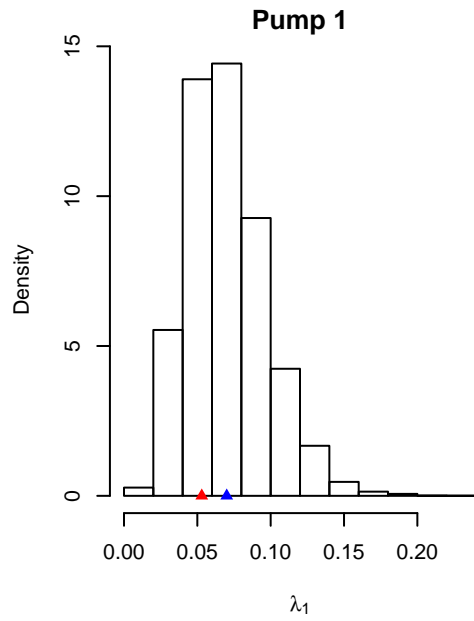
Example Run:

The following values were used for the prior parameters

$$\alpha = 1.8 \quad \delta = 1 \quad \gamma = 0.1$$

$n = 10000$  imputations were generated after a burn in of 1000 imputations.

The starting values for the chain were  $\beta^0 = \bar{l} = 1.33$ ,  $\lambda_i = l_i$ .



Pump	$l_i$	$t_i$	$E[\lambda_i s]$	Med( $\lambda_i s$ )	SD( $\lambda_i s$ )
1	0.0530	94.320	0.0700	0.0667	0.0267
2	0.0636	15.720	0.1553	0.1368	0.0935
3	0.0795	62.880	0.1044	0.0991	0.0401
4	0.1113	125.760	0.1231	0.1204	0.0307
5	0.5725	5.240	0.6283	0.5861	0.2914
6	0.6043	31.440	0.6167	0.6071	0.1343
7	0.9541	1.048	0.8298	0.7188	0.5379
8	0.9541	1.048	0.8316	0.7178	0.5302
9	1.9083	2.096	1.3020	1.2150	0.5744
10	2.0992	10.480	1.8358	1.8087	0.3873

$$E \left[ \frac{\alpha}{\beta} | s \right] = 0.7929$$

$E[\beta s]$	Med( $\beta s$ )	SD( $\beta s$ )
2.4678	2.3970	0.7074

