Petter Mostad
Applied Mathematics and Statistics
Chalmers

Suggested solutions for
MSA100 / MVE186 Computer Intensive Statistical Methods
Exam 22 October 2016

1. (a)
$$\pi(\theta \mid x) \propto \pi(x \mid \theta)\pi(\theta) = \frac{1}{\theta} I[0 \le x \le \theta] \frac{1}{\theta^2} I[1 \le \theta] = \frac{1}{\theta^3} I[\max(1, x) \le \theta]$$
We can read off this equation that the posterior is a Pareto$(\max(1, x), 2)$ distribution.

   (b) If the prior is Pareto(M, α), we get
$$\pi(\theta \mid x) \propto \pi(x \mid \theta)\pi(\theta) = \frac{1}{\theta} I[0 \le x \le \theta] \frac{1}{\theta^{\alpha+1}} I[M \le \theta] = \frac{1}{\theta^{\alpha+2}} I[\max(x, M) \le \theta]$$
so the posterior is Pareto$(\max(x, M), \alpha + 1)$. As the posterior is in the same family as the prior, this family is conjugate.

   (c)
$$\pi(x) = \frac{\pi(x \mid \theta)\pi(\theta)}{\pi(\theta \mid x)} = \frac{\frac{1}{\theta} I[0 \le x \le \theta] \frac{1}{\theta^2} I[1 \le \theta]}{2 \max(1, x)^2 \frac{1}{\theta^3} I[\max(1, x) \le \theta]} = \begin{cases} 1/2 & \text{if } 0 \le x \le 1 \\ 1/(2x^2) & \text{if} x > 1 \end{cases}$$

2. (a) The Cauchy distribution has a simple expression for the cumulative distribution function F given in the appendix. Writing $u = F(x)$ and solving for x gives
$$x = \mu + \gamma \tan\left(\pi\left(u - \frac{1}{2}\right)\right) = 3 + \tan\left(\pi\left(U - \frac{1}{2}\right)\right)$$
So the algorithm simulates u from Uniform[0, 1] and transforms the output with the equation above.

   (b) The probability of observing values above, say, 20, from a Poisson distribution with intensity 2.9 will be completely ignorable. Thus, an algorithm computes the values
$$p_k = \exp(-2.9)\frac{2.9^k}{k!}$$
for $k = 0, \ldots, 20$ and then simulates from the discrete distribution with possible values $0, \ldots, 20$ and probabilities $p_0, \ldots, p_{20}$.

   (c) As the exponential distribution has a cumulative distribution function F given in the appendix, we may write $u = F(x)$ and solve for x to get
$$x = -\frac{1}{3}\log(1 - u).$$

Using the relation $u = F(x)$ we get that $x > 10$ corresponds to $1 - u < \exp(-30)$. So writing u_0 for another Uniform[0, 1] variable and $(1 - u) = e^{-30}u_0$, an algorithm consists of simulating u_0, and then computing

$$x = -\frac{1}{3}\log(e^{-30}u_0) = 10 - \frac{1}{3}\log(u_0)$$

Indeed, an alternative approach is to use the properties of the Exponential distribuiton to see directly that simulating from Exponential(3) and then adding 10 is a valid algorithm.

3. (a)

$$\begin{aligned}
&\pi(y_1, \ldots, y_n, c_1, \ldots, c_n, \lambda_1, \ldots, \lambda_n, \alpha, \beta) \\
&\propto \prod_{i=1}^{n} \left[\left(I[y_i = c_i] I[c_i < 3] + I[y_i = \text{"many"}] I[c_i \geq 3] \right) e^{-\lambda_i} \frac{\lambda_i^{c_i}}{c_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} \exp(-\beta\lambda_i) \right]
\end{aligned}$$

(b) We go through all the groups of variables below:

- When all variables except the y_i are fixed, these can be computed from the c_i, setting $y_i = c_i$ when $c_i < 3$ and $y_i =$ "many" otherwise.
- When all variables except the c_i are fixed, there are two cases: If $y_i \neq$ "many", we compute $c_i = y_i$. If $y_i =$ "many", we get from (a) that

  $$\pi(c_i \mid \ldots) \propto I[c_i \geq 3] e^{-\lambda_i} \frac{\lambda_i^{c_i}}{c_i!}$$

  In other words, c_i has a Poisson distribution with parameter λ_i, truncated with $I[c_i \geq 3]$.
- When all variables except the λ_i are fixed, we get from (a) that

  $$\pi(\lambda_i \mid \ldots) \propto \lambda_i^{\alpha + c_i - 1} \exp(-(\beta + 1)\lambda_i)$$

  so λ_i has a Gamma$(\alpha + c_i, \beta + 1)$ distribution.
- When all variables except α are fixed, we get from (a) that

  $$\pi(\alpha \mid \ldots) \propto \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right)^n \left(\prod_{i=1}^{n} \lambda_i \right)^{\alpha - 1}$$

- When all variables except β are fixed, we get from (a) that

  $$\pi(\beta \mid \ldots) \propto \beta^{n\alpha} \exp\left(-\beta \sum_{i=1}^{n} \lambda_i \right)$$

  Thus β has a Gamma $\left(n\alpha + 1, \sum_{i=1}^{n} \lambda_i\right)$ distribution.

(c)
- For the c_i, which have a Poisson distribution with parameter λ_i truncated with $I[c_i \geq 3]$: When λ_i is high, it may be simplest to simulate from this distribution using rejection sampling, throwing away samples that are less than 3. When λ_i is low, it may be easiest to compute explicitly the probabilities for observing each $c_i \geq 3$ up to some cutoff, and then simulate from the resulting discrete distribution.
- For the α, the conditional distribution given in (b) can be re-written as

$$\pi(\alpha \mid \dots) \propto \frac{1}{\Gamma(\alpha)^n} \left(\beta \prod_{i=1}^{n} \lambda_i \right)^{\alpha}$$

Many alternative simulation methods are possible, and it is easier to select between them if one can first explore the function above numerically. However, even without that, one may notice that the presence of the factor $1/\Gamma(\alpha)^n$ means the function will approach zero very quickly when α increases, so it can easily be dominated by instrumental distributions with thicker tails. One alternative is to numerically find the α maximizing the function, and use this to select an instrumental density, for example a normal or a Gamma density, in either rejection sampling or possibly sampling importance resampling. As the function goes so fast to zero when α increases, an alternative is to find a reasonable cutoff and then simulate with a numerical approximation, for example a grid, on the interval up to this cutoff.

(d) A short outline of the EM algorithm is an acceptable answer, but here is a more detailed description of the EM algorithm for this situation: First, we notice that the distributions $c_i \mid \lambda_i$ and $\lambda_i \mid \alpha, \beta$ are conjugate. This means that the distribution $c_i \mid \alpha, \beta$ has an analytic form, which we can find as follows:

$$\begin{aligned}
\pi(c_i \mid \alpha, \beta) &= \frac{\pi(c_i \mid \lambda_i)\pi(\lambda_i \mid \alpha, \beta)}{\pi(\lambda_i \mid c_i, \alpha, \beta)} \\
&= \frac{\exp(-\lambda_i)\frac{\lambda_i^{c_i}}{c_i!}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda_i^{\alpha-1}\exp(-\lambda_i\beta)}{\frac{(\beta+1)^{\alpha+c_i}}{\Gamma(\alpha+c_i)}\lambda_i^{\alpha+c_i-1}\exp(-\lambda_i(\beta+1))} \\
&= \frac{\Gamma(\alpha+c_i)}{\Gamma(\alpha)c_i!} \cdot \frac{\beta^{\alpha}}{(\beta+1)^{\alpha+c_i}} \\
&= \binom{c_i+\alpha-1}{c_i} \left(\frac{\beta}{\beta+1}\right)^{\alpha} \left(\frac{1}{\beta+1}\right)^{c_i}
\end{aligned}$$

This can be recognized as a Negative Binomial distribution for c_i with parameters $r = \alpha$ and $p = 1/(\beta+1)$. Thus we can represent our model as having parameters $\theta = (\alpha, \beta)$, with a flat prior, and that the complete data $c_1, \dots, c_n$ is a set of independent observations from a Negative Binomial distribution with parameters $r = \alpha$ and $p = 1/(\beta+1)$. The reduced data is $y_1, \dots, y_n$.

In the EM algorithm we consider the logarithm of the complete data likelihood:

$$\begin{aligned}
& \log\left(\pi(c_1,\ldots,c_n \mid \alpha,\beta)\right) \\
= \;& \log\left[\prod_{i=1}^{n}\binom{c_i+\alpha-1}{c_i}\left(\frac{\beta}{\beta+1}\right)^{\alpha}\left(\frac{1}{\beta+1}\right)^{c_i}\right] \\
= \;& \sum_{i=1}^{n}\left[\ell\Gamma(c_i+\alpha)-\ell\Gamma(\alpha)-\ell\Gamma(c_i-1)-c_i\log(\beta+1)+\alpha\log(\beta/(\beta+1))\right] \\
= \;& n\left(\alpha\log(\beta/(\beta+1))-\ell\Gamma(\alpha)\right)+\sum_{i=1}^{n}\left[\ell\Gamma(c_i+\alpha)-\ell\Gamma(c_i-1)-c_i\log(\beta+1)\right]
\end{aligned}$$

where $\ell\Gamma$ is the logarithm of the Γ function. The $Q(\theta \mid \theta')$ of the E-step of the EM algorith consists of computing

$$\begin{aligned}
& Q(\theta \mid \theta') \\
= \;& E\left[\log(\pi(c_1,\ldots,c_n \mid \alpha,\beta)) \mid y_1,\ldots,y_n\right] \\
= \;& n\left(\alpha\log(\beta/(\beta+1))-\ell\Gamma(\alpha)\right)+\sum_{i=1}^{n}E\left[\ell\Gamma(c_i+\alpha)-\ell\Gamma(c_i-1)-c_i\log(\beta+1) \mid y_i\right]
\end{aligned}$$

where we take the expectation for each c_i over its distribution given parameters $\theta' = (\alpha',\beta')$, conditional on y_i. Let $p_{ij} = \Pr(c_i = j)$ under this distribution: For those i such that $y_i \neq$ "many" we get that $p_{ij} = 1$ for $j = y_i$ and $p_{ij} = 0$ otherwise. For those i where $y_i =$ "many", p_{ij} is the probability of j in the Negative Binomial distribution with parameters α' and $1/(\beta'+1)$ truncated to have 3 or more observations. We then get

$$\begin{aligned}
& Q(\theta \mid \theta') \\
= \;& n\left(\alpha\log(\beta/(\beta+1))-\ell\Gamma(\alpha)\right) \\
& +\sum_{i=1}^{n}\sum_{j=1}^{\infty}p_{ij}\left[\ell\Gamma(j+\alpha)-\ell\Gamma(j-1)-j\log(\beta+1)\right] \\
\approx \;& n\left(\alpha\log(\beta/(\beta+1))-\ell\Gamma(\alpha)\right)+\sum_{j=1}^{J}p_j\left[\ell\Gamma(j+\alpha)-\ell\Gamma(j-1)-j\log(\beta+1)\right]
\end{aligned}$$

where $p_j = \sum_{i=1}^{n} p_{ij}$ and where we in the last step disregard the terms for which the p_j are very small.

We may now move on to the M-step of the algorithm, where the expression above is maximized over $\theta = (\alpha,\beta)$, which in our case can be done numerically.

4. (a) We are given a function $f(x)$ which is proportional to the density we would like to simulate from. We also use a proposal density $q(y \mid x)$ which it must be possible to simulate from given an x. The Metropolis Hastings algorithm starts with generating some starting value x_0. Then, for $i = 1, 2, \ldots,$

- Simulate a new proposed value y using the density $q(y \mid x_{i-1})$.
- Compute the acceptance probability

$$p(y, x_{i-1}) = \min\left(1, \frac{f(y)q(x_{i-1} \mid y)}{f(x_{i-1})q(y \mid x_{i-1})}\right)$$

- Set x_i equal to y with probability $p(y, x_{i-1})$, otherwise equal to x_{i-1}.

(b) The Kernel function $K(y \mid x)$ of a Markov chain $x_0, \ldots, x_i, x_{i+1}, \ldots$, is the conditional probability of $y = x_{i+1}$ given $x = x_i$. The Markov chain satisfies the detailed balance condition relative to a density f if it satisfies

$$f(x)K(y \mid x) = f(y)K(x \mid y)$$

for all x and y.

(c) Taking the integral over x of the equation above, we get

$$\int f(x)K(y \mid x)\, dx = \int f(y)K(x \mid y)\, dx = f(y)$$

This means that, if x_i is distributed according to f, then x_{i+1} will also be distributed according to f, so f is a stationary distribution.

(d) Assume x is the value of the Metropolis Hastings chain at step i, let $K(y \mid y)$ be the Kernel function defined by the MH algorithm, and let $q(y \mid x)$ be the proposal density. Assume y is the proposed value, and assume first the acceptance probability $p(y, x) < 1$. Then

$$\begin{aligned} f(x)K(y \mid x) &= f(x)q(y \mid x)p(y, x) = f(x)q(y \mid x)\frac{f(y)q(x \mid y)}{f(x)q(y \mid x)} = f(y)q(x \mid y) \\ &= f(y)K(x \mid y) \end{aligned}$$

The last step is true because $p(x, y) = 1/p(y, x) > 1$. Clearly, by symmetry, we also get the detalied balance condition when $p(y, x) > 1$.

5. (a) Defining

$$\begin{aligned} y_1 \mid x &\sim \text{Uniform}(0, \exp(-x^{2.3})) \\ y_2 \mid x &\sim \text{Uniform}(0, \exp(-x^{1.3})) \\ y_3 \mid x &\sim \text{Uniform}(0, 1/(1 + x^3)) \end{aligned}$$

leads to the joint density

$$\begin{aligned} &\pi(x, y_1, y_2, y_3) = \pi(x)\pi(y_1 \mid x)\pi(y_2 \mid x)\pi(y_3 \mid x) \\ = \;& \exp(-x^{2.3}) \cdot \exp(-x^{1.3}) \cdot \frac{1}{1 + x^3} \\ & \cdot \frac{I[0 \le y_1 \le \exp(-x^{2.3})]}{\exp(-x^{2.3})} \cdot \frac{I[0 \le y_2 \le \exp(-x^{2.3})]}{\exp(-x^{2.3})} \cdot \frac{I[0 \le y_3 \le 1/(1 + x^3)]}{1/(1 + x^3)} \\ = \;& I[0 \le y_1 \le \exp(-x^{2.3})] \cdot I[0 \le y_2 \le \exp(-x^{2.3})] \cdot I[0 \le y_3 \le 1/(1 + x^3)] \end{aligned}$$

(b) The slice sampler is a Gibbs sampler simulating from the conditional distributions of the four variables. The conditional distributions of y_1, y_2, and y_3 are uniform and given directly above. For x we get

$$\begin{aligned} & \pi(x \mid y_1, y_2, y_3) \\ \propto \; & I[0 \leq x \leq (-\log(y_1))^{1/2.3}] \cdot I[0 \leq x \leq (-\log(y_2))^{1/1.3}] \cdot I[0 \leq x \leq (1/y_3 - 1)^{1/3}] \\ \propto \; & I\left[0 \leq x \leq \min\left((-\log(y_1))^{1/2.3}, (-\log(y_2))^{1/1.3}, (1/y_3 - 1)^{1/3}\right)\right] \end{aligned}$$

which indicates the uniform distribution to simulate from.

6. (a) As $|\sin(14x)| \leq 1$ and $|\cos(19x)| \leq 1$ for any x, we get that $|\sin(14x) + \cos(19x)| \leq 2$ and

$$f_0(x) = (\sin(14x) + \cos(19x))^2 \leq 4$$

for any x. Thus if g denotes the uniform density on the interval $[0, 1]$, we have

$$4g(x) \geq f_0(x)$$

for all x, and g should be a reasonable instrumental function.

(b) The algorithm is

- Simulate $x \sim \text{Uniform}(0, 1)$.
- Simulate $u \sim \text{Uniform}(0, 1)$.
- If $u < \frac{1}{4} f_0(x)$ then accept x, otherwise reject x and start over.

(c) If we store the acceptance rate R from the algorithm above, we get

$$R \approx \Pr\left(u < \frac{1}{4} f_0(x)\right) = E\left(\frac{1}{4} f_0(x)\right) = \frac{1}{4} \int_0^1 f_0(x)\, dx = \frac{1}{4C} \int_0^1 \pi(x)\, dx = \frac{1}{4C}$$

Thus we estimate

$$C \approx \frac{1}{4R}.$$

One may also take the average of the computed values $f_0(x)$ for all the simulated x above, to obtain a Monte Carlo estimate for $\int_0^1 f_0(x)\, dx$, and from that an estimate for C.

7. There is a problem with the simulation, as it is often "stuck" at the same value in large periods. This is a problem as the simulated sample then approximates badly the target density, in particular in the regions where it seems to become stuck.

The simulation seems to get stuck for low values, around the interval $[-1, 0]$, while it is not stuck in the interval around $[2, 4]$. If we denote by $f(x)$ the target density and $g(x)$ the proposal density used in the independent proposals, the plot indicates that $f(x)/g(x)$ is much higher than 1 in the region around $[-1, 0]$, and much lower than 1 in the region around $[2, 4]$. To improve the simulation, the proposal density g should be adjusted so that the quotient becomes closer to 1 in both regions.