Petter Mostad
Applied Mathematics and Statistics
Chalmers

Suggested solutions for
MSA101 / MVE187 Computational methods for Bayesian statistics
8 January 2019

1. A $(1-\alpha)100\%$ credibility interval for a real stochastic variable θ is an interval such that the probability that θ is in the interval is α. A $(1-\alpha)100\%$ confidence interval for θ assumes that θ is an unknown parameter, not a random variable. One also assumes defined a random distribution for data determined by θ. Then a $(1-\alpha)100\%$ confidence interval for θ is an interval that is a function of the data in such a way that, with probability α, it contains θ. Even though credibility intervals and confidence intervals are generally incompatible concepts, a confidence interval may in many practical situations correspond to a credibility interval with a specific choice of prior.

2. (a) The cumulative distribution function is given by

$$F(t) = \int_0^t \frac{\pi}{2}\cos\left(\frac{\pi}{2}x\right)dx = \left[\sin\left(\frac{\pi}{2}x\right)\right]_0^t = \sin\left(\frac{\pi}{2}t\right)$$

Thus, we get $F^{-1}(x) = \frac{2}{\pi}\arcsin(x)$. To simulate from the density, simulate $U \sim \text{Uniform}(0,1)$ and compute $F^{-1}(U) = \frac{2}{\pi}\arcsin(U)$.

(b) As $0 \le \cos\left(\frac{\pi}{2}x\right) \le 1$ and $0 \le e^{-x} \le 1$ for all $x \in [0,1]$, we get that $\pi(x) \le 1$ for all $x \in [0,1]$, and we can for example use the uniform distribution on $(0,1)$ as a proposal density. Thus, a rejection sampling algorithm for the density $\pi(x)$ can proceed as follows:

i. Simulate, independently, $U_1, U_2 \sim \text{Uniform}(0,1)$.
ii. If $U_2 \le \cos\left(\frac{\pi}{2}U_1\right)\exp(-U_1)$ report U_1 as the sampled value. Otherwise, return to step i.

(c) For slice sampling, note that the density consists of two factors, both of which are decreasing for $x \in (0,1)$. Note also that $t = \cos\left(\frac{\pi}{2}x\right)$ can be solved to $x = \frac{2}{\pi}\arccos(t)$ and that $t = e^{-x}$ can be solved to $x = -\log(t)$. Thus, a slice sampling algorithm for the density $\pi(x)$ can proceed as follows: First, fix an initial value $x^{(0)} \in (0,1)$. Then, for $i = 1, 2, \ldots,$

i. Simulate $u_1^{(i)}$ uniformly on $\left(0, \cos\left(\frac{\pi}{2}x^{(i-1)}\right)\right)$.
ii. Simulate $u_2^{(i)}$ uniformly on $\left(0, e^{-x^{(i-1)}}\right)$.
iii. Simulate $x^{(i)}$ uniformly on $\left(0, \min\left(\frac{2}{\pi}\arcsin(u_1^{(i)}), -\log(u_2^{(i)})\right)\right)$.

After a sufficient number of simulations, the distribution of $x^{(i)}$ will be approximately that of the density $\pi(x)$.

3. (a) The entropy is the expected information:

$$H[X] = \int -\log(\pi(x))\pi(x)\,dx$$

where the integral is taken over all possible values for x.

(b) We get

$$\begin{aligned} H[X] &= \mathrm{E}\left[-\log(\lambda\exp(-\lambda x))\right] \\ &= \mathrm{E}\left[-\log\lambda + \lambda x\right] \\ &= -\log\lambda + \lambda\,\mathrm{E}\left[x\right] \\ &= -\log\lambda + \lambda\frac{1}{\lambda} \\ &= 1 - \log\lambda \end{aligned}$$

4. (a) If $\tau \sim \text{Gamma}(\alpha,\beta)$ we get

$$\begin{aligned} \pi(\tau \mid x) &\propto_\tau \pi(x \mid \tau)\pi(\tau) \\ &\propto_\tau \frac{1}{\sqrt{2\pi\tau^{-1}}}\exp\left(-\frac{\tau}{2}x^2\right)\tau^{\alpha-1}\exp(-\beta\tau) \\ &\propto_\tau \tau^{1/2}\exp\left(-\frac{\tau}{2}x^2\right)\tau^{\alpha-1}\exp(-\beta\tau) \\ &\propto_\tau \tau^{\alpha+1/2-1}\exp\left(-\left(\beta+\frac{1}{2}x^2\right)\tau\right). \end{aligned}$$

Thus

$$\tau \mid x \sim \text{Gamma}\left(\alpha + 1/2, \beta + \frac{1}{2}x^2\right).$$

As the posterior is a Gamma distribution for any Gamma distribution as prior, we get that the Gamma family is a conjugate prior family.

(b) We get

$$\begin{aligned} \pi(x) &= \frac{\pi(x \mid \tau)\pi(\tau)}{\pi(\tau \mid x)} \\ &\propto_x \frac{\text{Normal}(x; 0, \tau^{-1})}{\text{Gamma}\left(\tau; \alpha + \frac{1}{2}, \beta + \frac{1}{2}x^2\right)} \\ &\propto_x \frac{\exp\left(-\frac{\tau}{2}x^2\right)}{\frac{\left(\beta+\frac{1}{2}x^2\right)^{\alpha+1/2}}{\Gamma(\alpha+1/2)}\tau^{\alpha+1/2}\exp\left(-\left(\beta+\frac{1}{2}x^2\right)\tau\right)} \\ &\propto_x \frac{1}{(\beta + \frac{1}{2}x^2)^{\alpha+1/2}} \end{aligned}$$

(c) Comparing with the density of the non-standard t distribution, we make the following stepwise changes to bring the density on the right form:

$$\begin{aligned}
\pi(x) &\propto_x \frac{1}{(\beta + \frac{1}{2}x^2)^{\alpha+1/2}} \\
&\propto_x \left(\beta + \frac{1}{2}x^2\right)^{-\frac{2\alpha+1}{2}} \\
&\propto_x \left(1 + \frac{x^2}{2\beta}\right)^{-\frac{2\alpha+1}{2}} \\
&\propto_x \left(1 + \frac{1}{2\alpha} \cdot \frac{x^2}{\beta/\alpha}\right)^{-\frac{2\alpha+1}{2}}.
\end{aligned}$$

We can now see that the marginal density for x is a non-standard t distribution with the following parameters: Degrees of freedom $\nu = 2\alpha$, expectation $\mu = 0$, and $\sigma^2 = \beta/\alpha$.

(d) We see that the prior is now a mixture of distributions from the Gamma family, and thus also conjugate. As the prior weights are both 1/2, the weights in the posterior can be found by comparing the marginal density of the data x given the prior Gamma(3, 7) with the marginal density given the prior Gamma(1, 1). According to (b) and (c) these marginal densities are $t(x; 2 \cdot 3, 0, 7/3)$ and $t(x; 2 \cdot 1, 0, 1/1)$, respectively. Bringing this together, we get the posterior

$$\begin{aligned}
\pi(\tau \mid x) &= \frac{t(x; 2 \cdot 3, 0, 7/3)}{t(x; 2 \cdot 3, 0, 7/3) + t(x; 2 \cdot 1, 0, 1/1)} \text{Gamma}\left(\tau; 3 + \frac{1}{2}, 7 + \frac{1}{2}x^2\right) \\
&+ \frac{t(x; 2 \cdot 1, 0, 1/1)}{t(x; 2 \cdot 3, 0, 7/3) + t(x; 2 \cdot 1, 0, 1/1)} \text{Gamma}\left(\tau; 1 + \frac{1}{2}, 1 + \frac{1}{2}x^2\right) \\
&= \frac{t(x; 6, 0, 7/3)}{t(x; 6, 0, 7/3) + t(x; 2, 0, 1)} \text{Gamma}\left(\tau; \frac{7}{2}, \frac{15}{2}x^2\right) \\
&+ \frac{t(x; 2, 0, 1)}{t(x; 6, 0, 7/3) + t(x; 2, 0, 1)} \text{Gamma}\left(\tau; \frac{3}{2}, \frac{3}{2}x^2\right).
\end{aligned}$$

5. (a) One can use the Forward Backward algorithm.

(b) One can use the Viterbi algorithm.

6. (a) We get

$$\begin{aligned}
\ell(z,\theta) &= \log\left[\pi(y_1,\ldots,y_{14},z,\theta)\right] \\
&= \log\left[\prod_{i=1}^{14}\text{Poisson}(y_i;z)\cdot\text{Binomial}(z;9,\theta)\cdot\text{Uniform}(0,1)\right] \\
&= \sum_{i=1}^{14}\log\left(e^{-z}\frac{z^{y_i}}{(y_i)!}\right)+\log\left(\frac{9!}{z!(9-z)!}\theta^z(1-\theta)^{9-z}\right) \\
&= -14z+\left(\sum_{i=1}^{14}y_i\right)\log z-\sum_{i=1}^{14}\log((y_i)!) \\
&\quad +\log(9!)-\log(z!)-\log((9-z)!)+z\log\theta+(9-z)\log(1-\theta)
\end{aligned}$$

(b) From (a) we can read off that

$$\pi(\theta\mid z,y_1,\ldots,y_{14})\propto_\theta \theta^z(1-\theta)^{9-z}$$

Thus $\theta\mid z,y_1\ldots,y_{14}$ has a Beta distribution with parameters $z+1$ and $10-z$:

$$\theta\mid z,y_1,\ldots,y_{14}\sim\text{Beta}(z+1,10-z).$$

(c) $z\mid\theta,y_1,\ldots,y_{14}$ does not have a distribution that is easy to describe analytically. However, we may instead compute a value proportional to the posterior for the 10 different possible values for z: $0,1,\ldots,9$ and then normalize this vector. Specifically, define

$$\ell_0(z,\theta)=-14z+\left(\sum_{i=1}^{14}y_i\right)\log z-\log(z!)-\log((9-z)!)+z\log\theta+(9-z)\log(1-\theta)$$

so that ℓ_0 corresponds to ℓ but with additive constants not depending on z or θ removed. Define, for $i=0,1,\ldots,9$,

$$v_i=\frac{\exp\left(\ell_0(i,\theta)\right)}{\sum_{j=0}^{9}\exp\left(\ell_0(j,\theta)\right)}.$$

Then $v=(v_0,\ldots,v_9)$ is the probability vector describing the posterior density for $z\mid\theta,y_1,\ldots,y_{14}$, and we can use this vector to simulate from this posterior density.

(d) First, we would select starting values for z and θ, for example $\theta=0.5$ and $z=5$. Then, we would iterate between simulating from $\pi(\theta\mid z,y_1,\ldots,y_{14})$ as described in (b) and simulating from $\pi(z\mid\theta,y_1,\ldots,y_{14})$ as described in (c).

(e) As with the Gibbs sampling, we first find some starting value for θ, for example $\theta_0=0.5$. We then iteratively perform the E and the M steps: In the E step we compute

$$\text{E}_z\left[\ell(z,\theta)\right]$$

where we take the expectation over the distribution $z \mid y_1, \ldots, y_{14}, \theta^{(old)}$, where $\theta^{(old)}$ is the value of θ in the previous step of the EM algorithm. Using the expression from (a) we can write

$$\mathrm{E}_z\left[\ell(z, \theta)\right] = C + \mathrm{E}_z(z) \log \theta + (9 - \mathrm{E}_z(z)) \log(1 - \theta)$$

where C is a constant not depending on θ. Note that we can compute

$$\mathrm{E}_z(z) = \sum_{n=1}^{9} n v_n$$

using the vector v computed as in (c).
In the M step we find the θ maximizing the expectation above: If

$$f(\theta) = \mathrm{E}_z(z) \log \theta + (9 - \mathrm{E}_z(z)) \log(1 - \theta)$$

then

$$f'(\theta) = \frac{\mathrm{E}_z(z)}{\theta} - \frac{9 - \mathrm{E}_z(z)}{1 - \theta}$$

and setting $f'(\theta) = 0$ results in

$$\theta = \frac{1}{9} \mathrm{E}_z(z) = \frac{1}{9} \sum_{n=1}^{9} n v_n.$$

7. Approximate Bayesian Computing is a way to make approximate Bayesian inference when it is difficult to write down an expression proportional to the likelihood of the data given the parameters of the model. Instead, one takes as the starting point the naive approach of simulating values of θ from the prior, and rejecting all those for which simulated values of the data do not approximately match the data. Through a series of smart tricks, this is transformed into a viable computational method.