Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MSA101 / MVE187 Computational methods for Bayesian statistics

Exam 21 October 2017, 14:00 - 18:00
Allowed aids: None.
The appendix of this exam contains informaion about some probability distributions.
Total number of points: 30. To pass, at least 12 points are needed

1. (5 points) Assume x has a Geometric distribution with parameter p.
   (a) If p has a prior that is uniform on $[0, 1]$, find and name the posterior distribution for p when $x = 3$.
   (b) Make a guess for a family of distributions that is conjugate to the Geometric distribution. Prove that this family is a conjugate family.
   (c) Assume p has a prior that is uniform on $[0, 1]$, and that the observation $x = 3$ has been made. Before a new observation y from the Geometric distribution with parameter p is made, one may compute $\pi(y \mid x)$, its predictive distribution. Find a formula for this probability mass function (not depending on p).

2. (2 points) It is possible to show that if $X \sim \text{Gamma}(\alpha_1,\beta)$ and $Y \sim \text{Gamma}(\alpha_2,\beta)$ are independent, then $X+Y \sim \text{Gamma}(\alpha_1+\alpha_2,\beta)$. Use this property to derive an algorithm for simulating from a Gamma(α,β) distribution when α is an integer, using only simulation from the uniform distribution on $[0, 1]$ as a basis.

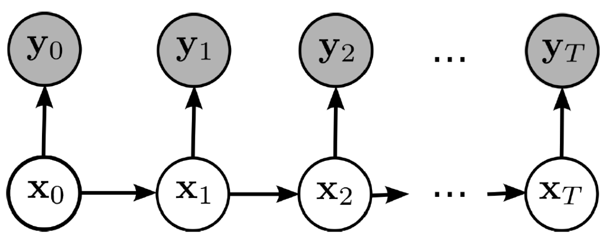

Figure 1: The hidden Markov model used in question 3.

3. (5 points) Consider the hidden Markov model depicted in Figure **??**. We assume that all variables are binary, with values 0 or 1, and that

$$\begin{aligned}
\Pr(x_i = 1 \mid x_{i-1} = 1) &= 0.4 \text{ for } i = 1, \dots, T \\
\Pr(x_i = 1 \mid x_{i-1} = 0) &= 0.2 \text{ for } i = 1, \dots, T \\
\Pr(y_i = 1 \mid x_i = 1) &= 0.8 \text{ for } i = 0, \dots, T \\
\Pr(y_i = 1 \mid x_i = 0) &= 0.3 \text{ for } i = 0, \dots, T \\
\Pr(x_0 = 1) &= 0.1
\end{aligned}$$

We assume data values for $y_0, y_1, \ldots, y_T$ are given.

(a) Define, for $i = 0, \ldots, T$,

$$a_i = \Pr(x_i = 1 \mid y_0, \ldots, y_i)$$

Compute a_0 and a_1 assuming that $y_0 = 1$ and $y_1 = 0$.

(b) Define, for $i = 0, \ldots, T-1$,

$$b_i = \Pr(y_{i+1}, \ldots, y_T \mid x_i = 1)$$

Compute b_{T-1} assuming that $y_T = 1$.

(c) Assume you have computed a_i for $i = 0, \ldots, T$ and b_i for $i = 0, \ldots, T-1$. Describe how you can compute

$$\Pr(x_i = 1 \mid y_0, \ldots, y_T)$$

for $i = 0, \ldots, T-1$.

4. (7 points) Consider the following hierarchical model: We have n groups of observations, each with m observations; we denote the observations with x_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$. For each i, the observations $x_{i1}, \ldots, x_{im}$ are independently exponentially distributed with parameter λ_i. The parameters $\lambda_1, \ldots, \lambda_n$ are Gamma$(4, \beta)$ distributed. We use a Gamma$(3, 4)$ prior for β.

(a) Write down and simplify as much as you can the logarithm of the joint posterior density for the model. You may disregard any additive constants not depending on the parameters β and $\lambda_1, \ldots, \lambda_n$.

(b) Using the result from (a), describe in detail how you can use Gibbs sampling to obtain an (approximate) sample from the joint posterior distribution for $(\beta, \lambda_1, \ldots, \lambda_n)$. Include which distributions you sample from in each step.

(c) Now, assume that the observations x_{ij} are censored, in the sense that for any x_{ij} that is greater than 10, you only know that it is greater than 10, you do not know its value. Describe an extension of the simulation algorithm above which may be used to obtain a sample from the posterior distribution for $(\beta, \lambda_1, \ldots, \lambda_n)$ given the censored data.

5. (4 points)

(a) Give a description of slice sampling: What it is, and how it works.

(b) Assume we have defined a density

$$\pi(x) \propto \frac{\exp\left(-(x+1)^2\right)}{3 + x^4}$$

for positive real x. Describe in detail how a slice sampler would work in this case.

6. (7 points) Consider the following model: For $i = 1, \ldots, n$, we have unobserved indicators $X_i \sim \text{Bernoulli}(\theta)$, with a uniform prior for θ. For data $y_1, \ldots, y_n$ we have

$$y_i \mid X_i = 0 \sim \text{Normal}\,(0, 1)$$

and

$$y_i \mid X_i = 1 \sim \text{Cauchy}(0, 1)$$

(a) Write down the formula for $\log(\pi(y_1, \ldots, y_n, X_1, \ldots, X_n \mid \theta))$, the logarithm of the full data likelihood.

(b) Compute the formula for $w_i = \Pr[X_i = 1 \mid y_1, \ldots, y_n, \theta']$ (with $i = 1, \ldots, n$).

(c) We would like to find the maximum likelihood estimate for θ using the EM algorithm: Write down the function $Q(\theta \mid \theta')$ of that algorithm.

(d) Find a formula for the θ maximizing $Q(\theta \mid \theta')$.

(e) Explain how you would implement the EM algorithm for this model.

Appendix: Some probability distributions

The Bernoulli distribution

If $x \in \{0, 1\}$ has a Bernoulli(p) distribution, with $0 \leq p \leq 1$, then the probability mass function is

$$\pi(x) = p^x(1-p)^{1-x}.$$

The Beta distribution

If $x \geq 0$ has a Beta(α, β) distribution with $\alpha > 0$ and $\beta > 0$ then the density is

$$\pi(x \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}.$$

The Beta-Binomial distribution

If $x \in \{0, 1, 2, \ldots, n\}$ has a Beta-Binomial(n, α, β) distribution, with n a positive integer and parameters $\alpha > 0$ and $\beta > 0$, then the probability mass function is

$$\pi(x \mid n, \alpha, \beta) = \binom{n}{x} \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n+\alpha+\beta)}.$$

The Binomial distribution

If $x \in \{0, 1, 2, \ldots, n\}$ has a Binomial(n, p) distribution, with n a positive integer and $0 \leq p \leq 1$, then the probability mass function is

$$\pi(x \mid n, p) = \binom{n}{x} p^x(1-p)^{n-x}.$$

The Cauchy distribution

If $x \geq 0$ has a Cauchy(μ, γ) distribution, with $\gamma > 0$, then the probability density is

$$\pi(x \mid \mu, \gamma) = \frac{1}{\pi\gamma\left(1 + \left(\frac{x-\mu}{\gamma}\right)^2\right)}.$$

The Exponential distribution

If $x \geq 0$ has an Exponential(λ) distribution with $\lambda > 0$ as parameter, then the density is

$$\pi(x \mid \lambda) = \lambda \exp(-\lambda x)$$

and the cumulative distribution function is

$$F(x) = 1 - \exp(-\lambda x).$$

The Gamma distribution

If $x > 0$ has a Gamma(α,β) distribution, with $\alpha > 0$ and $\beta > 0$, then the density is

$$\pi(x \mid \alpha\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x).$$

The Geometric distribution

If the non-negative integer x has a Geometric distribution with parameter $p \in [0, 1]$, its probability mass function is given by

$$\pi(x \mid p) = (1 - p)^x p.$$

The Logistic distribution

If x has a Logistic(μ, s) distribution, with $s > 0$, then the density is

$$\pi(x \mid \mu, s) = \frac{\exp\left(-\frac{x-\mu}{s}\right)}{s\left(1 + \exp\left(-\frac{x-\mu}{s}\right)\right)^2}$$

and the cumulative density function is given by

$$F(x) = \frac{1}{1 + \exp\left(-\frac{x-\mu}{s}\right)}.$$

The Normal distribution

If the real x has a Normal distribution with parameters μ and σ^2, its density is given by

$$\pi(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{\sigma^2}(x - \mu)^2\right).$$

The Uniform distribution

If $x \in [a, b]$ has a Uniform(a, b) distribution with $b > a$, then the density is given by

$$\pi(x \mid a, b) = \frac{1}{b - a}.$$