Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MSA101 / MVE187 Computational methods for Bayesian statistics

Exam 27 October 2018, 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. (6 points) Assume $X \mid \theta \sim \text{Uniform}(0, \theta)$ for some parameter $\theta > 0$.
   (a) Let $\pi(\theta)$ be a prior for the upper bound of the density for X. Assume $\pi(\theta)$ is non-zero on an interval $[M, \infty)$ and nowhere else. Find the interval on which the posteior $\pi(\theta \mid X)$ is non-zero.
   (b) Prove that the Pareto family of densities is a conjugate prior for θ.
   (c) Compute the prior predictive distribution for this conjugacy pair.
2. (3 points) Give a description of how to do rejection sampling.
3. (8 points) Consider the following model, with parameters $\theta = (\mu_1, \mu_2, \mu_3, \tau_1, \tau_2, \tau_3, \beta)$ and observed variables $y = (y_{11}, \ldots, y_{33})$:

$$\begin{aligned} \mu_i &\sim \text{Normal}(0, 1) \ \text{ for } i = 1, 2, 3. \\ \beta &\sim \text{Gamma}(2, 2) \\ \tau_i \mid \beta &\sim \text{Gamma}(2, \beta) \ \text{ for } i = 1, 2, 3. \\ y_{ij} \mid \mu_i, \tau_i &\sim \text{Normal}(\mu_i, \tau_i^{-1}) \ \text{ for } i = 1, 2, 3, \ j = 1, 2, 3. \end{aligned}$$

   (a) Draw the Bayesian Network corresponding to the model.
   (b) Write down a function $f(\theta)$ such that $f(\theta) = C + \log(\pi(\theta \mid y))$, where C is a constant not depending on θ. Make sure to remove from $f(\theta)$ all additive constants not depending on θ.
   (c) One way to simulate from the posterior $\pi(\theta \mid y)$ is to use MCMC with a proposal function where θ^{new} with new values for the 7 variables in θ are proposed as follows:

$$\begin{aligned} \mu_i^{new} \mid \mu_i &\sim \text{Normal}(\mu_i, 0.1), \ \text{ for } i = 1, 2, 3. \\ \tau_i^{new} &\sim \text{Exponential}(2), \ \text{ for } i = 1, 2, 3. \\ \beta^{new} &\sim \text{Exponential}(2) \end{aligned}$$

   Write down the formula for the acceptance probability in this case; you may write the formula using the function f from (b).

(d) Assume you instead would like to use Gibbs sampling to simulate from the posterior described above. Specify the formulas for all the conditional distributions one would need to simulate from in this case.

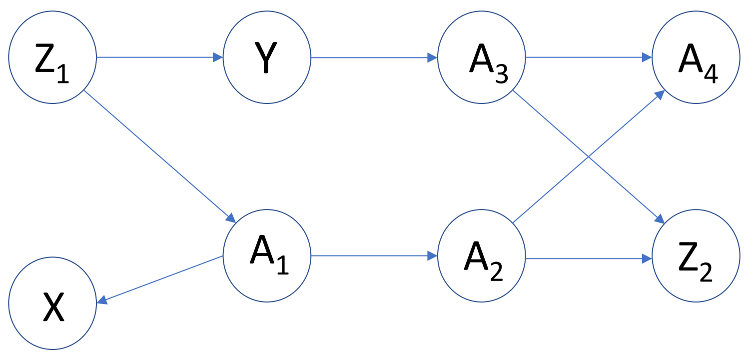

Figure 1: The Bayesian Network used in the question below. NOTE: In the original exam, there was an error; instead of one Z_1 and one Z_2, the figure above contained Z_1 both places.

4. (3 points) A joint probability distribution involving random variables X, Y, Z_1, and Z_2 can be represented by the Bayesian network of Figure 1.

   (a) Is necessarily $X \coprod Y$, i.e., are X and Y necessarily independent (i.e., can you prove that $X \coprod Y$ from the graph given in Figure 1)? Answer yes or no, and give an argument for your answer.

   (b) Is necessarily $X \coprod Y \mid Z_1$? Answer yes or no, and give an argument for your answer.

   (c) Is necessarily $X \coprod Y \mid \{Z_1, Z_2\}$? Answer yes or no, and give an argument for your answer.

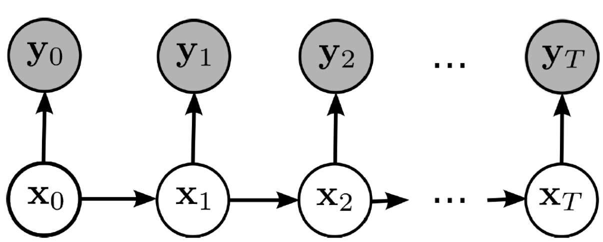

Figure 2: The hidden Markov model used in the question below.

5. (7 points) Consider the Hidden Markov Model of Figure 2. We assume the X_i variables have possible values $u_1, \dots, u_n$, with a fixed distribution $\pi(X_0)$ and a fixed transition matrix specifying $\pi(X_i \mid X_{i-1})$ for all $i = 1, \dots, T$. We also assume that, for $i = 0, \dots, T$,

$$Y_i \mid X_i \sim \text{Normal}(X_i, \tau^{-1})$$

where τ is an unknown precision. We use the prior $\pi(\tau) = \text{Gamma}(\tau; 1, 1)$. We would like to estimate the value of τ maximizing the marginal posterior $\pi(\tau \mid Y_0, \dots, Y_T)$, using the EM-algorithm, with the X_i variables as the augmented data.

   (a) Write down the log-likelihood $\log(\pi(X_0, \dots, X_T, Y_0, \dots, Y_T \mid \tau))$, removing all additive terms not depending on τ.

   (b) Fixing $\tau = \tau^{old}$, describe the algorithm to compute the expectation of the log-likelihood above under the distribution $\pi(X_0, \dots, X_T \mid Y_0, \dots, Y_T, \tau^{old})$. Describe the idea and the steps, but not necessarily all details.

   (c) Find a formula for the τ maximizing

$$\mathrm{E}\left[\log\left(\pi\left(X_0, \dots, X_T, Y_0, \dots, Y_T \mid \tau\right)\right)\right] + \log(\pi(\tau)).$$

   (d) In what circumstances would repeated runs of the complete algorithm above with the same τ_0 as starting points give different results? In what circumstances would repeated runs with different τ_0 as starting points give different results?

6. (3 points) Assume you have defined a Gaussian Markov Random Field over the variables $\{X_1, \dots, X_n\}$. Prove that there is a line between X_i and X_j for $i \neq j$ in the Markov network for the random field if and only if in the precision matrix $P = [p_{ij}]$, the term p_{ij} is different from zero.

Appendix: Some probability distributions

The Bernoulli distribution

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

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

The Beta distribution

If $x \ge 0$ 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 $0 \le p \le 1$, then the probability mass function is

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

The Cauchy distribution

If $x \ge 0$ 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 $x \ge 0$ 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 expectation and variance are α/β and α/β^2, respectively, while the mode is $(\alpha - 1)/\beta$ (when $\alpha \geq 1$).

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 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 Pareto distribution

If the real number $x \in [M, \infty)$ has a Pareto(M, α) distribution with parameters $M > 0$ and $\alpha > 0$, its density on this interval is given by

$$\pi(x \mid M, \alpha) = \alpha M^\alpha x^{-(\alpha+1)}$$

The Uniform distribution

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

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