Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MSA101 / MVE187 Computational methods for Bayesian statistics

Exam 29 August 2019, 8:30 - 12:30
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 has a Negative Binomial distribution with parameters p and r, and assume the prior for p is a Beta distribution with parameters α and β.
   (a) Find the posterior distribution for p given x.
   (b) Find the probability mass function for the prior predictive distribution for x (i.e., the marginal distribution for x).
   (c) Assume the prior for p instead of the above has a density
   $$\pi(p) = ap^{\alpha-1}(1-p)^{\beta-1} + bp^{\gamma-1}(1-p)^{\delta-1}$$
   What form does the posterior density have? Compute the posterior density.

2. (4 points) Define a density on the positive real line with
   $$p(x) = Ce^{-3x}|sinx|$$
   where C is some constant.
   (a) Describe an efficient way to obtain an exact sample from this density. Describe the algorithm in detail and name it. When describing the algorithm, take as the starting point that you can only simulate from the Uniform(0, 1) density.
   (b) Describe how to obtain an estimate for C from the algorithm.

3. (3 points)
   (a) Assume X has a Normal distribution with parameters μ and σ^2. Compute the entropy $H[X]$.
   (b) Assume $p(x)$ and $q(x)$ are two densities on real numbers. Define the Kullback-Leibler divergence, or Kullback-Leibler "distance" $KL[p||q]$ from p to q.

4. (3 points) A joint probability distribution involving random variables X, Y and $Z_1, \ldots, Z_6$ can be represented by the Bayesian network of Figure 1.

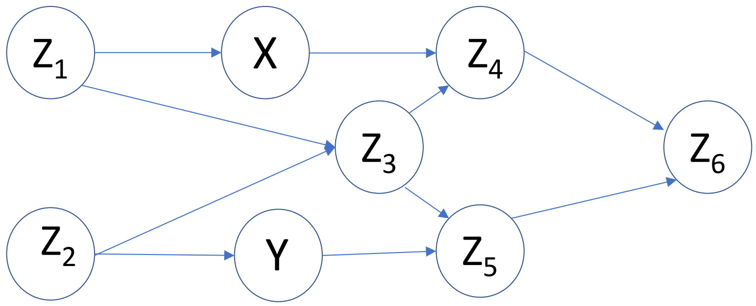

Figure 1: The Bayesian Network used in the question below.

(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 $X \coprod Y \mid \{Z_1, Z_2\}$, i.e., are X and Y necessarily independent given Z_1 and Z_2? Answer yes or no, and give an argument for your answer.

(c) Is necessarily $X \coprod Y \mid \{Z_4, Z_5\}$, i.e., are X and Y necessarily independent given Z_4 and Z_5? Answer yes or no, and give an argument for your answer.

5. (7 points) Assume the model

$$
\begin{aligned}
\alpha & \quad && \text{Uniform on the finite set } \{1, 2, 3, \ldots, 9, 10\}. \\
\beta & \sim && \text{Gamma}(3, 7) \\
\lambda \mid \alpha, \beta & \sim && \text{Gamma}(\alpha, \beta) \\
y \mid \lambda & \sim && \text{Poisson}(4\lambda)
\end{aligned}
$$

where y is observed data.

(a) Write down the logarithm of the posterior distribution for (α, β, λ) up to a constant not depending on α, β, or λ.

(b) Write down the precise steps for generating an approximate sample from the posterior $\pi(\alpha, \beta, \lambda \mid y)$ using Gibbs sampling.

(c) Write down the precise steps for generating an approximate sample from the posterior $\pi(\alpha, \beta, \lambda \mid y)$ using Metropolis-Hastings random walk with a symmetric proposal function that you specify.

6. (3 points). Assume a model with real variables y, θ, and z is specified with a conditional density $\pi(y \mid \theta, z)$ and a prior density $\pi(\theta, z)$, where y is the data. Assuming z is regarded as the "augmented data", describe the EM (Expectation Maximization) algorithm in this context: What would be its goal? What would be the outline of the algorithm? What exactly would be done in the E step? What exactly would be the done in the M step?

7. (2 points) What is Variational Bayes? Explain briefly.

8. (2 points) What is the Viterbi algorithm? Describe what its goal is, and outline briefly how it works.

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 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 Negative Binomial distribution

If the non-negative integer x has a Negative Binomial distribution with parameters p and r, so that p with $0 < p < 1$ is the probability of success in each independent trial, $r > 0$ is a given number of failed trials, and x represents the number of successul trials when r failures have occurred, then the probability mass function of x is given by

$$\pi(x \mid p, r) = \frac{(x + r - 1)!}{x!(r - 1)!} p^x (1 - p)^r$$

We write $x \mid p, r \sim$ Negative-Binomial(p, r) and $\pi(x \mid p, r) =$ Negative-Binomial$(x; p, r)$.

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 Poisson distribution

If the non-negative integer x has a Poisson distribution wiht parameter $\lambda > 0$, its probability mass function is

$$\pi(x \mid \lambda) = e^{-\lambda} \frac{\lambda^x}{x!}.$$

We write $x \mid \lambda \sim \text{Poisson}(\lambda)$ and $\pi(x \mid \lambda) = \text{Poisson}(x; \lambda)$.

The Uniform distribution

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

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