Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MSA100 / MVE186 Computer Intensive Statistical Methods

Re-exam 5 June 2017, 8:30 - 12:30
Examiner: Petter Mostad, jour: Sebastian Jobjörnsson
Allowed aids: None.
The appendix of this exam contains informaion about probability distributions.
Total number of points: 30. To pass, at least 12 points are needed

1. (2 points) Consider the following situation: Data $x_1, \dots, x_n$ is assumed to be a sample from a normal distribution with expectation θ and variance 1. Consider also the following two statements: "A 95% confidence interval for θ is the interval $[2.3, 2.5]$." and "A 95% credibility interval for θ is the interval $[2.3, 2.5]$." Describe the context in which each statement is used, and write down a correct and precise interpretation of the statement in that context.

2. (7 points) Assume x has a Negative Binomial distribution with fixed known r, and with parameter p (see Appendix).

   (a) Prove that the Beta family of probability distributions (see Appendix) is a conjugate family for the p parameter.

   (b) Compute the formula for the prior predictive distribution for x.

   (c) Find the posterior predictive distribution for an additional observation x_{NEW}, if the prior for p is Beta(2, 1), the value of r is 2, and the observed value is $x = 1$.

   (d) Assume there is a choice between two prior models for p: In model 1, $p \sim \text{Beta}(\alpha_1, \beta_1)$ and in model 2, $p \sim \text{Beta}(\alpha_2, \beta_2)$. Compute the formula for the Bayes factor B which can be used to make a choice between the models.

   (e) Assume the prior for p is a weighted mean of the two Beta distributions with parameters (α_1, β_1) and (α_2, β_2), respectively. Assume they have equal weights in the prior. Compute the posterior density for p given an observation x. You may express the posterior in terms of the B computed in (d) above.

3. (4 points) Assume the only (pseudo) random numbers your computer tool can generate are those that have a Uniform distribution on the interval $[0, 1]$, but otherwise the tool has the capabilities of, say, R. Describe in detail how you would simulate a random number from the following distributions:

   (a) An exponential distribution with parameter 2.7.

   (b) An Inverse Gamma distribution with parameters 2.7 and 9.1.

(c) The distribution on x whose density on the whole real line is proportional to

$$f(x) = \sum_{i=1}^{7} w_i \exp\left(-\frac{1}{2}(x - u_i)^2\right)$$

where $w_1, w_2, \ldots, w_7$ are given positive numbers summing to 1 and $u_1, \ldots, u_k$ are given real numbers.

4. (4 points) Answer each of the following questions in a couple of sentences:

   (a) What is a Bayesian Network?

   (b) What is a Markov Network?

   (c) What is the difference between a causal network and a Bayesian Network with the same structure?

   (d) What is the connection between the precision matrix of a Gaussian Markov random field and its Markov graph?

5. (2 points) What is Monte Carlo Integration? Explain briefly how you can obtain estimates of the accuracy of the result from a Monte Carlo integration.

6. (5 points) Consider the following model:The data consists of k groups of counts, with s counts in each group; we denote the data with c_{ij}, where $i = 1, \ldots, k$, $j = 1, \ldots, s$, and where each c_{ij} is a nonnegative integer. We assume the counts $c_{i1}, c_{i2}, \ldots, c_{is}$ represent a sample from a Poisson distribution with parameter λ_i. We also assume $\lambda_1, \lambda_2, \ldots, \lambda_k$ represent a sample from a Gamma distribution with parameters α and β. We use a flat prior for α and a Gamma(5, 2) prior for β.

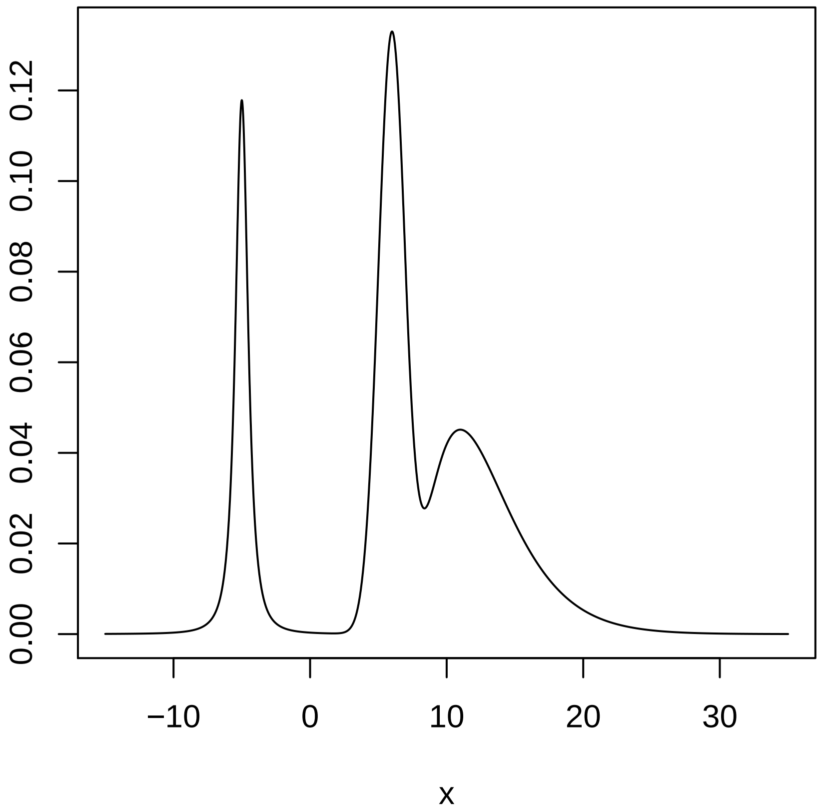
   (a) Write down and simplify the log posterior density for the model[1]. In other words write down a function of the parameters that (up to an additive constant) is equal to the logarithm of the density of the parameters given the observed data $c_{11}, \ldots, c_{ks}$.

   (b) Give a general description of Gibbs sampling. Include the general idea, and mention how one can prove that it produces an (approximate) sample from the posterior of the parameters given the data in the model above.

   (c) Describe how you would implement Gibbs sampling for the model above. In particular, at each point where a number should be simulated and that simulation can be done from a standard parametric distribution, identify that distribution.

7. (6 points) The figure below shows a plot of a probability density function from which you would like to get a sample:

[1] In the original exam, one asks for the log-likelihood here



(a) Suppose you would like to sample from the distribution above using rejection sampling (e.g., the Accept-Reject method). Give a description of exactly how the algorithm would work, including an explicit candidate density, with parameter values. Base your proposal on looking at the figure above, and motivate your choices.

(b) Suppose you would like to sample from the distribution above using a Metropolis Hastings algorithm with an independent proposal function. Give an example of a proposal function that you would use, including its parametric form and reasonable values for its parameter or parameters. Base your proposal on looking at the figure above, and motivate your choice. Describe the differences, if any, between the type of samples generated and those generated according to the method in (a).

(c) Suppose you would like to sample from the distribution using a Metropolis Hastings algorithm with a symmetric proposal function (a random walk Metropolis Hastings). Give an example of a proposal function that you would use, indlucing its parametric form and a reasonable values for its parameter or parameters. Base your proposal on looking at the figure above, and motivate your choice.

(d) Using the proposal function from (c), describe what would happen in the algorithm if you rescale the proposal function to have a much smaller variance. Why would this be a problem? Also, answer the same questions assuming you rescale to a much larger variance.

Appendix: Some probability distributions

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 Inverse Gamma distribution

We say that $x > 0$ has an Inverse Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$,

$$x \sim \text{Inv-Gamma}(\alpha, \beta),$$

if the density is given by

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

Note that if $x \sim \text{Inv-Gamma}(\alpha, \beta)$ then $1/x \sim \text{Gamma}(\alpha, \beta)$.

The Poisson distribution

If k has a Poisson(λ) distribution with intensity $\lambda > 0$, then the probabilities are given by

$$\pi(k \mid \lambda) = \exp(-\lambda)\frac{\lambda^k}{k!}$$

The Normal distribution

We say that a real number x has a Normal distribution with parameters μ and $\sigma^2 > 0$,

$$x \sim \text{Normal}(\mu, \sigma^2)$$

if it has density

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

The Beta distribution

We say that $x \geq 0$ has a Beta distribution with parameters $\alpha > 0$ and $\beta > 0$,

$$x \sim \text{Beta}(\alpha, \beta),$$

if it has density

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

The Exponential distribution

If $x > 0$ has an Exponential(λ) distribution, with $\lambda > 0$, the density is given by

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

and the cumulative density function is given by

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

The Negative Binomial distribution

We say that x has a Negative Binomial distribution with parameters r and p, where x and r are nonnegative integers and $p > 0$, writing

$$x \sim \text{Neg-Bin}(r, p),$$

if it has probability mass function

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

If the parameter r represents a number of "failures" and p the probability of "success", x represents the number of successes observed until r failures are observed.