Exam Statistical Reasoning

Date: Thursday, October 30, 2014
Time: 09.00-12.00
Place: Kapteyenborg, Landleven 12, 5419.0119
Progress code: WISR-11

Rules to follow:

- This is a closed book exam. Consultation of books and notes is not permitted.
- Do not forget to fill in your name and student number.
- The number of points per question are indicated within a box. Ten points are free.
- We wish you success with the completion of the exam!

START OF EXAM

1. **Negative binomial distribution with Gamma prior** $\boxed{20}$

Consider the following sampling model. The random variables $Y_1, \ldots, Y_n$ are negative binomial distributed, symbolically $Y_1, \ldots, Y_n|\theta \sim \text{NBIN}(\theta, r)$, and i.i.d. conditional on the parameters $r \in \mathbb{N}$ and $\theta \in [0, 1]$. Assume that r is known and fixed, while θ is unknown. Impose a Beta prior with the hyperparameters a and b on θ, symbolically $\theta \sim \text{Beta}(a, b)$.

Recall the following: The density (PDF) of a Beta distribution with parameters $a > 0$ and $b > 0$ is given by:

$$p(x|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \theta^{a-1} \cdot (1-\theta)^{b-1}$$

for $x \in [0, 1]$. The density (PDF) of the negative binomial distribution with two parameters $r \in \mathbb{N}$ and $\theta \in [0, 1]$ is given by

$$p(x|\theta, r) = \binom{r+x-1}{x} \cdot \theta^r \cdot (1-\theta)^x$$

for $x \in \mathbb{N}_0$.

(a) $\boxed{5}$ Compute the joint density $p(y_1, \ldots, y_n|\theta, r)$ of the sampling model.

(b) $\boxed{10}$ Compute the posterior distribution of θ, i.e. compute the following conditional distribution $\theta|(Y_1 = y_1, \cdots, Y_n = y_n)$.

(c) $\boxed{5}$ A geometric distribution with parameter $\theta \in [0, 1]$ has the density (PDF):

$$p(x|\theta) = \theta \cdot (1-\theta)^x$$

for $x \in \mathbb{N}_0$.
First, show that the geometric distribution is a special case of the negative binomial distribution. Then determine the posterior distribution of θ when the random variables $Y_1, \ldots, Y_n$ are geometric distributed, $Y_1, \ldots, Y_n|\theta \sim \text{Geo}(\theta)$, and the prior distribution of the parameter θ is the continuous uniform distribution on the interval $[0, 1]$, symbolically: $\theta \sim \text{Uniform}([0, 1])$.
<u>HINT</u>: Recall that: $\text{Uniform}([0, 1]) = \text{Beta}(1, 1)$.

to be continued below

2. **Uniform distribution with Pareto prior** $\boxed{15}$
Consider the following sampling model. The random variables $Y_1, \ldots, Y_n$ are i.i.d. and continuous uniformly distributed on the interval $[0, b]$, symbolically:

$$Y_1, \ldots, Y_n | b \sim \text{Uni}([0, b])$$

where the upper bound of the interval $b \in \mathbb{R}^+$ is unknown. Impose a Pareto prior with parameters $k > 0$ and $m > 0$ on b, symbolically $b \sim \text{Pareto}(k, m)$.

   (a) $\boxed{3}$ Give the joint density $p(y_1, \ldots, y_n|b)$ of the sampling model for the sample spaces $y_i \in \mathbb{R}_0^+$ $(i = 1, \ldots, n)$. That is, give $p(y_1, \ldots, y_n|b)$ for *all* $y_i \in \mathbb{R}_0^+$ $(i = 1, \ldots, n)$ by using a case-by-case definition (see HINT).

   (b) $\boxed{10}$ Show that the posterior distribution of b is a Pareto distribution with parameters $\tilde{k} := k + n$ and $\tilde{m} := \max\{y_1, \ldots, y_n, m\}$, symbolically:

$$b|(Y_1 = y_1, \ldots, Y_n = y_n) \sim \text{Pareto}(\tilde{k}, \tilde{m})$$

   Perform a case-by-case analysis to actually show that $\tilde{m} := \max\{y_1, \ldots, y_n, m\}$.

   (c) $\boxed{2}$ Give an interpretation of the two hyperparameters k and m in terms of pseudo observations.

<u>HINT</u>: The densities (PDFs) of the uniform distribution on $[0, b]$ and the Pareto distribution with parameters k and m are as follows. For $x \in \mathbb{R}_0^+$:

$$p(x|b) = \begin{cases} \frac{1}{b}, & x \leq b \\ 0, & \text{else} \end{cases}$$

$$p(x|k, m) = \begin{cases} \left(\frac{m}{x}\right)^k, & x \geq m \\ 0, & \text{else} \end{cases}$$

3. **Predictive distribution** $\boxed{20}$
Consider the following sampling model. The random variables $Y_1, \ldots, Y_n$ are i.i.d. and Gaussian distributed, symbolically: $Y_1, \ldots, Y_n|\mu, \sigma^2 \sim N(\mu, \sigma^2)$, where the variance parameter is known, $\sigma^2 = 1$, and the expectation parameter μ is unknown. Given the observed data $(Y_1 = y_1, \ldots, Y_n = y_n)$ assume that a Gaussian prior, $\mu \sim N(\mu_0, \tau_0^2)$, has been imposed on μ and that it is already known that the posterior distribution of μ is the standard Gaussian distribution:

$$\mu|(Y_1 = y_1, \ldots, Y_n = y_n) \sim N(0, 1)$$

The goal is to compute the predictive distribution for a new random variable $\tilde{Y}$.

<u>HINTS:</u>

(1) As usual, we have $\tilde{Y}|\mu, \sigma^2 \sim N(\mu, \sigma^2)$ and $\tilde{Y}, Y_1, \ldots, Y_n$ are i.i.d..

(2) The density (PDF) of a Gaussian distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}^+$ is given by:

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot e^{-\frac{1}{2} \cdot \frac{1}{\sigma^2} \cdot (x-\mu)^2}$$

for $x \in \mathbb{R}$.

to be continued below

(a) $\boxed{5}$ Given a Bayesian model with a one-dimensional parameter $\mu \in \mathbb{R}$, e.g. the Gaussian model defined above. Show that the density (PDF) of the predictive distribution $\tilde{Y}|(Y_1 = y_1, \ldots, Y_n = y_n)$ is given by:

$$p(\tilde{y}|y_1, \ldots, y_n) = \int p(\tilde{y}|\mu) \cdot p(\mu|y_1, \ldots, y_n) d\mu$$

(b) $\boxed{5}$ Show that the following relationship is true. For $\tilde{y}, \mu \in \mathbb{R}$:

$$e^{-0.5 \cdot (\tilde{y}-\mu)^2} \cdot e^{-0.5 \cdot \mu^2} = \frac{1}{\sqrt{2}} \cdot e^{-0.5 \cdot \frac{\tilde{y}^2}{2}} \cdot \frac{1}{\sqrt{0.5}} \cdot e^{-0.5 \cdot \frac{(\mu - 0.5 \cdot \tilde{y})^2}{0.5}}$$

(c) $\boxed{10}$ Compute the predictive distribution $\tilde{Y}|(Y_1 = y_1, \ldots, Y_n = y_n)$ for the Gaussian Bayesian model, defined above.

4. **Discrete Markov chains** $\boxed{20}$
Consider a simple discrete random variable X with sample space $\Theta = \{1, 2, 3, 4\}$ and probability density function (PDF): $p_X(1) = 0.1$, $p_X(2) = 0.4$, $p_X(3) = 0.4$, and $p_X(4) = 0.1$. The goal is to construct a Metropolis-Hastings MCMC sampling scheme to generate a sample from the distribution of X. Assume that proposal moves are designed such that the sixteen proposal probabilities are given by: $Q(1, 1) = 0$, $Q(1, 2) = 0.8$, $Q(1, 3) = 0.2$, $Q(1, 4) = 0$, $Q(2, 1) = 0.2$, $Q(2, 2) = 0$, $Q(2, 3) = 0$, $Q(2, 4) = 0.8$, $Q(3, 1) = 1$, $Q(3, 2) = 0$, $Q(3, 3) = 0$, $Q(3, 4) = 0$, $Q(4, 1) = 0$, $Q(4, 2) = 1$, $Q(4, 3) = 0$, and $Q(4, 4) = 0$, where $Q(i, j)$ is the probability to propose to move from state $i \in \Theta$ to state $j \in \Theta$.

(a) $\boxed{6}$ Compute the following six Metropolis-Hastings acceptance probabilities $A(1, 2)$, $A(2, 1)$, $A(1, 3)$, $A(3, 1)$, $A(2, 4)$ and $A(4, 2)$, where $A(i, j)$ is the probability to accept a move from state $i \in \Theta$ to state $j \in \Theta$.

(b) $\boxed{3}$ Compute the following six (1-step) transition probabilities $T(1, 2)$, $T(2, 1)$, $T(1, 3)$, $T(3, 1)$, $T(2, 4)$ and $T(4, 2)$, where $T(i, j)$ is the probability for a move (transition) from state $i \in \Theta$ to state $j \in \Theta$. HINT: $10/32 = 5/16 = 0.3125$.

(c) $\boxed{4}$ Compute the following four (1-step) transition probabilities $T(1, 1)$, $T(2, 2)$ and $T(3, 3)$ and $T(4, 4)$, where $T(i, i)$ is the probability for a move from $i \in \Theta$ to $i \in \Theta$, i.e. the probability for staying at state i.

(d) $\boxed{3}$ Give the 4-by-4 (1-step) transition probability matrix T of the resulting Markov chain.

(e) $\boxed{4}$ Give a graphical representation of the transition matrix. That is, represent each of the four states $\{1, 2, 3, 4\}$ as a node. Then use directed edges to indicate all those state space transitions that have a positive probability. As usual, label each edge with its transition probability.

to be continued below

5. **Full conditional distributions and MCMC Sampling** 15

Consider the following hierarchical Bayesian model: The sampling model is given by n variables $Y_1, \dots, Y_n$ which are exponentially distributed, $Y_1, \dots, Y_n|\theta \sim \text{EXP}(\theta)$, where $\theta > 0$ is the parameter of the Exponential distribution and $Y_1, \dots, Y_n$ are i.i.d. conditional on the unknown parameter θ. A Gamma distribution with hyperparameters $a > 0$ and $b > 0$ is imposed on θ, where a is known and b is unknown. Therefore a Gamma hyperprior with the fixed/known hyperhyperparameters $\alpha > 0$ and $\beta > 0$ is imposed on b. Mathematically, we thus have

$$Y_1, \dots, Y_n|\theta \sim \text{EXP}(\theta)$$

$$\theta|b \sim \text{Gamma}(a, b)$$

$$b \sim \text{Gamma}(\alpha, \beta)$$

where $a > 0$, $\alpha > 0$, and $\beta > 0$ are fixed and known.

(a) 5 Give a graphical model representation of this hierarchical model.

(b) 5 Show that the densitities of the two full conditional distributions fullfil:

$$p(\theta|b, y_1, \dots, y_n) \propto p(y_1, \dots, y_n|\theta) \cdot p(\theta|b)$$

$$p(b|\theta, y_1, \dots, y_n) \propto p(\theta|b) \cdot p(b)$$

(c) 5 Do **not** compute the two full conditional distributions. Just assume that both full conditional distributions $p(\theta|b, y_1, \dots, y_n)$ and $p(b|\theta, y_1, \dots, y_n)$ can be computed in closed form. Describe a MCMC sampling scheme based on two Gibbs sampling steps to generate a sample $(b^{(1)}, \theta^{(1)}), \dots, (b^{(T)}, \theta^{(T)})$ from the joint posterior distribution: $(\theta, b)|(Y_1 = y_1, \dots, Y_n = y_n)$. You can provide either some pseudo-code (see below) or a verbal description of the algorithm.
HINT: Proposed structure of your pseudo code:

Initialisation: Set $b^{(1)} =$ ..., and set $\theta^{(1)} =$ ...

Iterations: For $t = 2, \dots, T$:
Sample ... from ...
Sample ... from ...

Output: ...

to be continued below

6. **Monte Carlo approximation** $\boxed{10}$

Consider a general Bayesian model. The sampling model is given by n variables $Y_1, \ldots, Y_n$ whose distribution depends on a one-dimensional parameter $\theta \in \mathbb{R}$. The variables $Y_1, \ldots, Y_n$ are i.i.d. conditional on the unknown parameter θ having the joint density (PDF) $p(y_1, \ldots, y_n|\theta) = \prod_{i=1}^n p(y_i|\theta)$, where $y_1, \ldots, y_n$ are the observed data. A prior distribution with some fixed hyperparameters is imposed on θ, and the density of the prior distribution is given by $p(\theta)$.

(a) $\boxed{2}$ Give an equation for the density (PDF) of the posterior distribution of θ, i.e. give the standard definition of $p(\theta|y_1, \ldots, y_n)$.

(b) $\boxed{2}$ Give the definition of the normalisation constant: $p(y_1, \ldots, y_n)$.

(c) $\boxed{3}$ Assume that the predictive distribution cannot be computed in closed form. Describe how a Monte Carlo approximation can be employed to obtain a sample of size T from the predictive distribution $\tilde{Y}|(Y_1 = y_1, \ldots, Y_n = y_n)$, where, as usual, $\tilde{Y}$ and $Y_1, \ldots, Y_n$ are i.i.d. conditional on the parameter θ.

(d) $\boxed{3}$ Describe how the normalisation constant $p(y_1, \ldots, y_n)$ from part (b) can be approximated with the Monte Carlo method.

<u>HINTS:</u> For parts (c) and (d) of this exercise: You can provide some pseudo code, e.g. as follows:
To take n samples $x_1, \ldots, x_n$ from the distribution of a random variable X, whose distribution depends on a parameter θ, you can for example write:
Sample: $x_1 \sim p(x|\theta), \ldots, x_n \sim p(x|\theta)$, where $p(x|\theta)$ is the density (PDF) of X given θ.
To indicate that you take the mean of $x_1, \ldots, x_n$, you can for example write:
Compute: $\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$.
To indicate that you compute the joint density of $x_1, \ldots, x_n$ you can for example write:
Compute: $p(x_1, \ldots, x_n|\theta) = \prod_{i=1}^n p(x_i|\theta)$.
etc.

END OF EXAM