Exam Statistical Reasoning

Date: Friday, November 11, 2016
Time: 09.00-12.00
Place: 5419.0013 (Landleven 12 (Kapteynborg))
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 write your name and student number on each paper sheet.
- There are 6 exercises and the number of points per exercise are indicated within boxes. You can reach 90 points. The exam grade g will be computed as follows: $g = \frac{10+p}{100}$, where p is the number of points you have reached.
- If you have to derive/show/compute something, then include the relevant equations and/or a short description so as to show how you obtained the result.
- **We wish you success with the completion of the exam!**

START OF EXAM

1. **Posterior Distribution of Binomial-Beta Model.** **[15]**
Consider the random variables $Y_1, \ldots, Y_n$ which are independently Binomial-distributed with an unknown probability parameter $\theta \in [0,1]$ and a known sample size N:

$$Y_1, \ldots, Y_n | \theta \sim \text{BIN}(N, \theta)$$

The prior of θ is a Beta distribution with fix hyperparameters $a > 0$ and $b > 0$:

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

The realisations $Y_1 = y_1, \ldots, Y_n = y_n$ have been observed.

EXERCISE: Compute the posterior distribution of θ.

HINTS:
(1) Recall the pdf of a Binomial distribution with parameters $\theta \in [0,1]$ and $N \in \mathbb{N}$. For $x = 0, \ldots, N$:

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

(2) Recall the pdf of a Beta distribution with parameters $a > 0$ and $b > 0$. For $x \in [0,1]$:

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

2. **Marginal Likelihood of Poisson-Exponential Model.** **15**

Consider one single random variable Y whose distribution is a Poisson distribution with parameter $\theta > 0$. Conditional on θ the density of Y is given by:

$$p(y|\theta) = \frac{\theta^y \cdot e^{-\theta}}{y!}$$

for all $y \in \mathbb{N}_0$.

Assume that the unknown parameter θ is exponentially distributed with a fix hyperparameter $\lambda > 0$. The density of θ is then:

$$p(\theta) = \lambda \cdot e^{-\lambda\theta}$$

for all $\theta > 0$.

EXERCISE: Show that the marginal distribution of Y has the density (pdf):

$$p(y) = \left(\frac{1}{1+\lambda}\right)^y \cdot \frac{\lambda}{1+\lambda} \qquad (y \in \mathbb{N}_0)$$

HINTS:

(1) Recall the pdf of a Gamma distribution with parameters α and β. For $x \in \mathbb{R}^+$:

$$p(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\beta \cdot x}$$

3. **Predictive Distribution of Multinomial-Dirichlet Model.** **15**

Let the random vector $(N_1, \ldots, N_K)^T$ be multinomial distributed with parameters $(\theta_1, \ldots, \theta_K)$ and n so that the density (pdf) is given by:

$$p(n_1, \ldots, n_K|\theta_1, \ldots, \theta_K) = \frac{n!}{n_1! \cdot \ldots \cdot n_K!} \cdot \prod_{k=1}^{K} \theta_k^{n_k}$$

where $\theta_k > 0$ for all k, $\theta_1 + \ldots + \theta_K = 1$, $n_k \in \{0, \ldots, n\}$ for all k and $n_1 + \ldots + n_K = n$. Assume that the parameter vector $(\theta_1, \ldots, \theta_K)^T$ is Dirichlet distributed with pdf:

$$p(\theta_1, ..., \theta_K) = \frac{\Gamma(\sum_{k=1}^{K} \alpha_k)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \cdot \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

where $\alpha_1, \ldots, \alpha_K > 0$ are fixed hyperparameters. In the lecture we have shown (Exercise 1 on Assignment no. 2) that the posterior distribution of $(\theta_1, \ldots, \theta_K)^T$, given the realisations $N_1 = n_1, \ldots, N_K = n_K$, is a Dirichlet distribution with parameters $(n_1 + \alpha_1, \ldots, n_K + \alpha_K)$.

EXERCISES:

(a) Just give the pdf $p(\theta_1, \ldots, \theta_K|n_1, \ldots, n_K)$ of the posterior distribution. 5

(b) Show that the predictive probability for a new realisation of the form

$$(\tilde{n}_1, \ldots, \tilde{n}_{j-1}, \tilde{n}_j, \tilde{n}_{j+1}, \ldots, \tilde{n}_K) = (0, \ldots, 0, 1, 0 \ldots, 0)$$

is given by:

$$p(\tilde{n}_1, \ldots, \tilde{n}_K|n_1, \ldots, n_K) = \frac{\alpha_j + n_j}{\alpha + n}$$

where $\alpha = \alpha_1 + \ldots + \alpha_K$. 10

4. **Gibbs Sampling - Pseudo Code.** **10**
Consider a Bayesian model with three unknown parameters $\theta_1 \in \Theta_1$, $\theta_2 \in \Theta_2$ and $\theta_3 \in \Theta_3$, and assume that the full conditional distributions can be computed in closed-form. Let $\text{FCD}(\theta_1|\theta_2, \theta_3, \text{D})$, $\text{FCD}(\theta_2|\theta_1, \theta_3, \text{D})$ and $\text{FCD}(\theta_3|\theta_1, \theta_2, \text{D})$ denote the full conditional distributions, where 'D' stands for the observed data.

EXERCISE: Give pseudo-code for an MCMC algorithm which generates a sample from the joint posterior distribution of θ_1, θ_2 and θ_3.
HINT: Assume that there is neither need for a burn-in phase nor for thinning out.

5. **Hierarchical Bayesian Model - Coupled Variances.** **20**
Consider a hierarchical Bayesian model where the data stem from two groups. There are the random variables $Y_1, \ldots, Y_{n_1}$ from group 1, and the random variables $Z_1, \ldots, Z_{n_2}$ from group 2. The observed data are the $n := n_1 + n_2$ realisations:

$$Y_1 = y_1, \ldots, Y_{n_1} = y_{n_1}, Z_1 = z_1, \ldots, Z_{n_2} = z_{n_2}$$

The variables within the first group are Gaussian distributed with a known mean μ and an unknown variance σ_1^2:

$$Y_1, \ldots, Y_{n_1}|\sigma_1^2 \sim \text{N}(\mu, \sigma_1^2)$$

The variables within the second group are Gaussian distributed with the same known mean μ but with another unknown variance σ_2^2:

$$Z_1, \ldots, Z_{n_2}|\sigma_2^2 \sim \text{N}(\mu, \sigma_2^2)$$

The variance parameters σ_i^2 $(i = 1, 2)$ are both assumed to be Inverse-Gamma distributed with hyperparameters $a = 1$ and $b > 0$:

$$\sigma_1^{-2} \sim \text{GAM}(1, b)$$

$$\sigma_2^{-2} \sim \text{GAM}(1, b)$$

Since the hyperparameter b is unknown, it obtains a Gamma distribution with hyperparameters $\alpha > 0$ and $\beta > 0$ as hyperprior:

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

EXERCISE: Give a graphical model representation of this model and compute the three full conditional distributions. 5+5+5+5

HINTS:
(1) Recall the pdf of a Gamma distribution with parameters α and β. For $x \in \mathbb{R}^+$:

$$p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\beta \cdot x}$$

(2) Recall the pdf of a Gaussian $\text{N}(\mu, \sigma^2)$ distribution. For $x \in \mathbb{R}$:

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot \exp\{-0.5 \cdot \frac{(x-\mu)^2}{\sigma^2}\}$$

6. **Metropolis-Hastings Sampling and Monte Carlo Approximations.** 15

Consider a hierarchical Bayesian model, where the sampling model consists of n variables $Y_1, \ldots, Y_n$ whose joint distribution depends on a parameter $\theta_1 \in \Theta_1$. Conditional on θ_1 the variables $Y_1, \ldots, Y_n$ are i.i.d. with the likelihood:

$$p(y_1, \ldots, y_n|\theta_1) = \prod_{i=1}^{n} p(y_i|\theta_1)$$

where $y_1, \ldots, y_n$ are the observed realisations.

The prior distribution of θ_1 depends on a hyperparameter $\theta_2 \in \Theta_2$; the density of θ_1 is thus of the form $p(\theta_1|\theta_2)$, and the hyperprior of the hyperparameter θ_2 is a uniform distribution on Θ_2.

Assume that a Metropolis-Hastings MCMC sampling algorithm is used to generate a sample of the joint posterior distribution. The algorithm consists of two substeps:

- The first step proposes to move from θ_1 to $\theta_1^\star$, while θ_2 is left unchanged. The proposal mechanism is such that the new parameter $\theta_1^\star$ is always sampled from the prior distribution. Hence, the proposal probability is

$$Q((\theta_1, \theta_2), (\theta_1^\star, \theta_2)) = p(\theta_1^\star|\theta_2)$$

  for all θ_1,$\theta_1^\star \in \Theta_1$, and $\theta_2 \in \Theta_2$.

- The second step proposes to move from θ_2 to $\theta_2^\star$, while θ_1 is left unchanged. The proposal probability in this second step is always:

$$Q((\theta_1, \theta_2), (\theta_1, \theta_2^\star)) = 0.2$$

  for all $\theta_1 \in \Theta_1$ and $\theta_2, \theta_2^\star \in \Theta_2$.

For the exercises below assume further that:
$p(y_1, \ldots, y_n|\theta_1) = 3$, $p(y_1, \ldots, y_n|\theta_1^\star) = 2, p(\theta_1|\theta_2) = 0.5$, and $p(\theta_1|\theta_2^\star) = 0.4$.

EXERCISES:

(a) Compute the Metropolis-Hastings acceptance probability for the first sub-move from (θ_1, θ_2) to $(\theta_1^\star, \theta_2)$. 5

(b) Compute the Metropolis-Hastings acceptance probability for the second sub-move from (θ_1, θ_2) to $(\theta_1, \theta_2^\star)$. 5

(c) Assume that you have a posterior sample $(\theta_1^{(t)}, \theta_2^{(t)})_{t=1,\ldots,T}$. Give an equation for a Monte Carlo approximation of the predictive probability $p(\tilde{y}|y_1, \ldots, y_n)$, i.e. for an approximation of the density of the predictive distribution at $\tilde{y}$. 5

END OF EXAM