Solutions for the exam in Statistical Reasoning

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

1. **Negative binomial distribution with Gamma prior** 20

   (a) 5 Compute the joint density

$$p(y_1,\dots,y_n|\theta,r) = \prod_{i=1}^{n} p(y_i|\theta,r) = \prod_{i=1}^{n} \binom{r+y_i-1}{y_i} \cdot \theta^r \cdot (1-\theta)^{y_i}$$

$$= \left(\prod_{i=1}^{n} \binom{r+y_i-1}{y_i}\right) \cdot \left(\prod_{i=1}^{n} \theta^r \cdot (1-\theta)^{y_i}\right)$$

$$= \left(\prod_{i=1}^{n} \binom{r+y_i-1}{y_i}\right) \cdot \theta^{n\cdot r} \cdot (1-\theta)^{\sum_{i=1}^{n} y_i}$$

   (b) 10 Compute the posterior distribution of θ

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

$$= \left(\prod_{i=1}^{n} \binom{r+y_i-1}{y_i}\right) \cdot \theta^{n\cdot r} \cdot (1-\theta)^{\sum_{i=1}^{n} y_i} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \theta^{a-1} \cdot (1-\theta)^{b-1}$$

$$\propto \theta^{n\cdot r} \cdot (1-\theta)^{\sum_{i=1}^{n} y_i} \cdot \theta^{a-1} \cdot (1-\theta)^{b-1}$$

$$= \theta^{n\cdot r+a-1} \cdot (1-\theta)^{\sum_{i=1}^{n} y_i+b-1}$$

   From the last line it can be seen that the PDF of the posterior is proportional to the PDF of a Beta distribution with parameters $\tilde{a} = n\cdot r + a$ and $\tilde{b} = \sum_{i=1}^{n} y_i + b$. For the posterior distribution we thus have:

$$\theta|(Y_1 = y_1,\dots,Y_n = y_n) \sim \text{Beta}(n\cdot r + a, \sum_{i=1}^{n} y_i + b)$$

   (c) 5 Selecting the parameter $r = 1$ in the PDF of the negative binomial yields:

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

   which is the PDF of the geometric distribution with parameter θ.

   By applying the result from part (a) with $a = b = r = 1$ we immediately obtain the posterior distribution for this Bayesian model:

$$\theta|(Y_1 = y_1,\dots,Y_n = y_n) \sim \text{Beta}(n\cdot r + a, \sum_{i=1}^{n} y_i + b) = \text{Beta}(n+1, \sum_{i=1}^{n} y_i + 1)$$

2. **Uniform distribution with Pareto prior** $\boxed{15}$

(a) $\boxed{3}$ Compute the joint PDF. For $y_i \in \mathbb{R}_0^+$ $(i = 1, \ldots, n)$ we have:

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

$$= \begin{cases} \left(\frac{1}{b}\right)^n, & y_1 \leq b, \ldots, y_n \leq b \\ 0, & \text{else} \end{cases}$$

$$= \begin{cases} \left(\frac{1}{b}\right)^n, & \max\{y_1, \ldots, y_n\} \leq b \\ 0, & \text{else} \end{cases}$$

(b) $\boxed{10}$ Compute the posterior. For $b \in \mathbb{R}^+$ we have:

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

$$= \begin{cases} \left(\frac{1}{b}\right)^n \cdot \left(\frac{m}{b}\right)^k, & \max\{y_1, \ldots, y_n\} \leq b \wedge b \geq m \\ 0, & \text{else} \end{cases}$$

$$\propto \begin{cases} \left(\frac{1}{b}\right)^{n+k}, & \max\{y_1, \ldots, y_n, m\} \leq b \\ 0, & \text{else} \end{cases}$$

$$= \begin{cases} \left(\frac{1}{b}\right)^{n+k}, & b \geq \max\{y_1, \ldots, y_n, m\} \\ 0, & \text{else} \end{cases}$$

From the last line it can be seen that the PDF of the posterior is proportional to the PDF of a Pareto distribution with parameters $\tilde{a} := n + k$ and $\tilde{b} := \max\{y_1, \ldots, y_n, m\}$.

(c) $\boxed{2}$ Interpretation: There are k additional pseudo observations, and the maximum of those pseudo observations is equal to m.

3. **Predictive distribution** $\boxed{20}$

(a) $\boxed{5}$ For the PDF of the predictive distribution we have:

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

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

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

(b) $\boxed{5}$ On the left we have:

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

$$= e^{-0.5 \cdot \tilde{y}^2 + \tilde{y} \cdot \mu - \mu^2} =: (\star)$$

On the right we have:

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

$$= e^{-0.25\cdot\tilde{y}^2} \cdot e^{-(\mu-0.5\cdot\tilde{y})^2} = e^{-0.25\cdot\tilde{y}^2} \cdot e^{-(\mu^2-\mu\cdot\tilde{y}+0.25\cdot\tilde{y}^2)} = e^{-0.25\cdot\tilde{y}^2} \cdot e^{-\mu^2+\mu\cdot\tilde{y}-0.25\cdot\tilde{y}^2}$$

$$= e^{-0.25\cdot\tilde{y}^2-\mu^2+\mu\cdot\tilde{y}-0.25\cdot\tilde{y}^2} = e^{-0.5\cdot\tilde{y}^2+\mu\cdot\tilde{y}-\mu^2} = (\star)$$

(c) $\boxed{10}$ Here we have for the posterior: $\mu|(Y_1 = y_1, \ldots, Y_n = y_n) \sim N(0, 1)$, so that

$$p(\mu|y_1, \ldots, y_n) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot\mu^2}$$

and

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

Compute the PDF of the predictive distribution:

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

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

Using the result from part (b) it follows:

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

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

The integral is equal to 1, as the integrand is the PDF of the normal distribution with expectation $0.5 \cdot \tilde{y}$ and variance 0.5. Hence, we obtain:

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

and this is the PDF of the normal distribution with expectation 0 and variance 2. Consequently, we have the predictive distribution:

$$\tilde{Y}|(Y_1 = y_1, \ldots, Y_n = y_n) \sim N(0, 2)$$

4. **Discrete Markov chains** $\boxed{20}$

(a) $\boxed{6}$ For $i \neq j$ the acceptance probabilities are given by:

$$A(i, j) = \min\left\{1, \frac{p(j)}{p(i)} \cdot \frac{Q(j, i)}{Q(i, j)}\right\}$$

So we obtain:

$$A(1,2) = \min\left\{1, \frac{0.4}{0.1} \cdot \frac{0.2}{0.8}\right\} = 1$$

$$A(2,1) = \min\left\{1, \frac{0.1}{0.4} \cdot \frac{0.8}{0.2}\right\} = 1$$

$$A(1,3) = \min\left\{1, \frac{0.4}{0.1} \cdot \frac{1}{0.2}\right\} = 1$$

$$A(3,1) = \min\left\{1, \frac{0.1}{0.4} \cdot \frac{0.2}{1}\right\} = \frac{1}{20} = 0.05$$

$$A(2,4) = \min\left\{1, \frac{0.1}{0.4} \cdot \frac{1}{0.8}\right\} = \frac{1}{4} \cdot \frac{10}{8} = \frac{5}{16} = 0.3125$$

$$A(4,2) = \min\left\{1, \frac{0.4}{0.1} \cdot \frac{0.8}{1}\right\} = 1$$

(b) $\boxed{3}$ For $i \neq j$ we have: $T(i,j) = Q(i,j) \cdot A(i,j)$:

$$T(1,2) = Q(1,2) \cdot A(1,2) = 0.8 \cdot 1 = 0.8$$
$$T(2,1) = Q(2,1) \cdot A(2,1) = 0.2 \cdot 1 = 0.2$$
$$T(1,3) = Q(1,3) \cdot A(1,3) = 0.2 \cdot 1 = 0.2$$
$$T(3,1) = Q(3,1) \cdot A(3,1) = 1 \cdot 0.05 = 0.05$$
$$T(2,4) = Q(2,4) \cdot A(2,4) = 0.8 \cdot 0.3125 = \frac{8}{10} \cdot \frac{5}{16} = 0.25$$
$$T(4,2) = Q(4,2) \cdot A(4,2) = 1 \cdot 1 = 1$$

(c) $\boxed{4}$ $T(i,i)$ for $i = 1, \ldots, 4$:

$$T(1,1) = 1 - T(1,2) - T(1,3) - T(1,4) = 1 - 0.8 - 0.2 - 0 = 0$$

$$T(2,2) = 1 - T(2,1) - T(2,3) - T(2,4) = 1 - 0.2 - 0 - 0.25 = 0.55$$
$$T(3,3) = 1 - T(3,1) - T(3,2) - T(3,4) = 1 - 0.05 - 0 - 0 = 0.95$$
$$T(4,4) = 1 - T(4,1) - T(4,2) - T(4,3) = 1 - 0 - 1 - 0 = 0$$

(d) $\boxed{3}$ The transition matrix is given by:

$$T = \begin{pmatrix} 0 & 0.8 & 0.2 & 0 \\ 0.2 & 0.55 & 0 & 0.25 \\ 0.05 & 0 & 0.95 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

(e) $\boxed{4}$ There is no digital solution available

5. **Full conditional distributions and MCMC Sampling** $\boxed{15}$

(a) $\boxed{5}$ There is no digital solution available

(b) $\boxed{5}$ Full conditional of θ:

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

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

Full conditional of b:

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

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

(c) $\boxed{5}$ **Initialisation**: Set $b^{(1)} = b \in \mathbb{R}^+$, and set $\theta^{(1)} = \theta \in \mathbb{R}^+$.

Iterations: For $t = 2, \ldots, T$:
Sample $b^{(t)}$ from $p(b|\theta^{(t-1)}, y_1, \ldots, y_n)$.
Sample $\theta^{(t)}$ from $p(\theta|b^{(t)}, y_1, \ldots, y_n)$.

Output: $(b^{(1)}, \theta^{(1)}), \ldots, (b^{(T)}, \theta^{(T)})$

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

(a) $\boxed{2}$ The PDF of the posterior is given by:

$$p(\theta|y_1, \ldots, y_n) = \frac{p(y_1, \ldots, y_n|\theta) \cdot p(\theta)}{p(y_1, \ldots, y_n)}$$

(b) $\boxed{2}$ The normalisation constant is given by:

$$p(y_1, \ldots, y_n) = \int p(\theta, y_1, \ldots, y_n) d\theta = \int p(y_1, \ldots, y_n|\theta) \cdot p(\theta) d\theta$$

(c) $\boxed{3}$ For $t = 1, \ldots, T$:
Sample $\theta^{(1)} \sim p(\theta|y_1, \ldots, y_n), \ldots, \theta^{(T)} \sim p(\theta|y_1, \ldots, y_n)$, and then sample $\tilde{y}^{(1)} \sim p(\tilde{y}|\theta^{(1)}), \ldots, \tilde{y}^{(T)} \sim p(\tilde{y}|\theta^{(T)})$.
$\tilde{y}^{(1)}, \ldots, \tilde{y}^{(T)}$ is a sample from $\tilde{Y}|(Y_1 = y_1, \ldots, Y_n = y_n)$.

(d) $\boxed{3}$ Sample: $\theta^{(1)} \sim p(\theta), \ldots, \theta^{(T)} \sim p(\theta)$.
Then compute: $p^{(1)} := p(y_1, \ldots, y_n|\theta^{(1)}), \ldots, p^{(T)} := p(y_1, \ldots, y_n|\theta^{(T)})$

For $T \to \infty$ it follows: $\frac{1}{T} \sum_{t=1}^{T} p^{(t)} \to p(y_1, \ldots, y_n)$

END