Solutions for the exam in Statistical Reasoning

1. **Bayes theorem** $\boxed{10}$

   (a) $\boxed{5}$ <u>EVENTS</u>: 1st urn (U1), 2nd urn (U2), and red ball is drawn (R).

   $$p(U1|R) = \frac{P(R|U1) \cdot P(U1)}{P(R)} = \frac{0.5 \cdot 0.5}{0.5 \cdot 0.5 + 0.3 \cdot 0.5} = \frac{5}{8}$$

   (b) $\boxed{5}$ <u>EVENTS</u>: Fair coin (F), two-headed coin (T), and head (H).

   $$p(F|H) = \frac{P(H|F) \cdot P(F)}{P(H)} = \frac{0.5 \cdot 0.5}{0.5 \cdot 0.5 + 1 \cdot 0.5} = \frac{1}{3}$$

2. **Density of Gaussian distribution** $\boxed{10}$

   (a) $\boxed{5}$ Transform the PDF of the Gaussian distribution:

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

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

   $$= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot \exp\{-0.5 \cdot x^2 \cdot \frac{1}{\sigma^2} + x \cdot \frac{\mu}{\sigma^2}\} \cdot \exp\{-0.5 \cdot \frac{\mu^2}{\sigma^2}\}$$

   Thus it can be seen:

   $$c = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot \exp\{-0.5 \cdot \frac{\mu^2}{\sigma^2}\}$$

   $$a = \frac{1}{\sigma^2}$$

   $$b = \frac{\mu}{\sigma^2}$$

   (b) $\boxed{5}$ It follows from (a) that the posterior must be a Gaussian distribution with mean μ and variance σ^2 where:

   $$a = \frac{1}{\sigma^2} = 0.5 \Rightarrow \sigma^2 = 2$$

   $$b = \frac{\mu}{\sigma^2} = \frac{\mu}{2} = 0.5 \Rightarrow \mu = 1$$

3. **Exponential-Gamma model** $\boxed{25}$
Let Gamma(a,b) denote the Gamma distribution with parameters a and b.

(a) $\boxed{2}$ The Exponential distribution is a special case of the Gamma distribution, because with $a = 1$ and $b = \lambda$: $\text{Exp}(\lambda) = \text{Gamma}(1,\lambda)$.

(b) $\boxed{3}$ The joint PDF is:

$$p(y_1, \ldots, y_n|\lambda) = \prod_{i=1}^{n} p(y_i|\lambda) = \prod_{i=1}^{n} \lambda \cdot e^{-\lambda \cdot y_i} = \lambda^n \cdot e^{-\lambda \cdot \sum_{i=1}^{n} y_i}$$

(c) $\boxed{10}$ Compute the posterior distribution of λ:

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

$$= \lambda^n \cdot e^{-\lambda \cdot \sum_{i=1}^{n} y_i} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \lambda^{\alpha-1} \cdot e^{-\beta \cdot \lambda}$$

$$\propto \lambda^n \cdot e^{-\lambda \cdot \sum_{i=1}^{n} y_i} \cdot \lambda^{\alpha-1} \cdot e^{-\beta \cdot \lambda}$$

$$= \lambda^{n+\alpha-1} \cdot e^{-\lambda \cdot (\sum_{i=1}^{n} y_i + \beta)}$$

Hence, the PDF of the posterior is proportional to the PDF of a Gamma distribution with parameters $\tilde{\alpha} = n + \alpha$ and $\tilde{\beta} = \sum_{i=1}^{n} y_i + \beta$. Thus:

$$\lambda|(Y_1 = y_1, \ldots, Y_n = y_n) \sim \text{Gamma}(n + \alpha, \sum_{i=1}^{n} y_i + \beta)$$

(d) $\boxed{5}$ Interpretation of the hyperparameters: There are α pseudo observations whose values sum up to β.

4. **Predictive distribution of the Exponential-Gamma model** $\boxed{25}$

(a) $\boxed{15}$ For predictive distributions we have:

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

Here:

$$p(\tilde{y}|y_1, \ldots, y_n) = \int \lambda \cdot e^{-\lambda \cdot \tilde{y}} \cdot \frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})} \cdot \lambda^{\tilde{\alpha}-1} \cdot e^{-\tilde{\beta} \cdot \lambda} d\lambda = \frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})} \cdot \int \lambda^{\tilde{\alpha}} \cdot e^{-\lambda \cdot (\tilde{\beta} + \tilde{y})} d\lambda$$

$$= \frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})} \cdot \frac{\Gamma(\tilde{\alpha} + 1)}{(\tilde{\beta} + \tilde{y})^{\tilde{\alpha}+1}} \cdot \int \frac{(\tilde{\beta} + \tilde{y})^{\tilde{\alpha}+1}}{\Gamma(\tilde{\alpha} + 1)} \cdot \lambda^{(\tilde{\alpha}+1)-1} \cdot e^{-\lambda \cdot (\tilde{\beta} + \tilde{y})} d\lambda$$

$$= \frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})} \cdot \frac{\Gamma(\tilde{\alpha} + 1)}{(\tilde{\beta} + \tilde{y})^{\tilde{\alpha}+1}} \cdot 1 = (\frac{\tilde{\beta} + \tilde{y}}{\tilde{\beta}})^{-(\tilde{\alpha}+1)} \cdot \frac{1}{\tilde{\beta}} \cdot \tilde{\alpha} = \frac{\tilde{\alpha}}{\tilde{\beta}}(1 + \frac{\tilde{y}}{\tilde{\beta}})^{-(\tilde{\alpha}+1)}$$

And this is the PDF of a Lomax distribution with parameters $\tilde{\alpha}$ and $\tilde{\beta}$.

(b) $\boxed{10}$ Pseudo code:

- Sample $\lambda^{(1)} \sim \text{Gamma}(\tilde{\alpha}, \tilde{\beta}), \ldots, \lambda^{(R)} \sim \text{Gamma}(\tilde{\alpha}, \tilde{\beta})$.
- Then sample $\tilde{y}^{(1)} \sim \text{Exp}(\lambda^{(1)}), \ldots, \tilde{y}^{(R)} \sim \text{Exp}(\lambda^{(R)})$.
- It holds: $\frac{1}{R} \cdot \sum_{r=1}^{R} \tilde{y}^{(r)} \to E[\tilde{Y}|Y_1, \ldots, Y_n]$ for $R \to \infty$.

5. **Discrete Markov chains** $\boxed{15}$

(a) $\boxed{5}$ For $i \neq j$ the acceptance probabilities are given by:

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

Therefore:

$$A(1,2) = \min\left\{1, \frac{0.4}{0.5} \cdot \frac{0.1}{1}\right\} = 0.08$$
$$A(2,1) = \min\left\{1, \frac{0.5}{0.4} \cdot \frac{1}{0.1}\right\} = 1$$
$$A(2,3) = \min\left\{1, \frac{0.1}{0.4} \cdot \frac{1}{0.9}\right\} = \frac{10}{36}$$
$$A(3,2) = \min\left\{1, \frac{0.4}{0.1} \cdot \frac{0.9}{1}\right\} = 1$$

(b) $\boxed{10}$ For $i \neq j$: $T(i,j) = Q(i,j) \cdot A(i,j)$:

$$T(1,2) = Q(1,2) \cdot A(1,2) = 1 \cdot 0.08 = 0.08$$
$$T(2,1) = Q(2,1) \cdot A(2,1) = 0.1 \cdot 1 = 0.1$$
$$T(1,3) = Q(1,3) \cdot A(1,3) = 0 \cdot A(1,3) = 0$$
$$T(3,1) = Q(3,1) \cdot A(3,1) = 0 \cdot A(3,1) = 0$$
$$T(3,2) = Q(3,2) \cdot A(3,2) = 1 \cdot 1 = 1$$
$$T(2,3) = Q(2,3) \cdot A(2,3) = 0.9 \cdot \frac{10}{36} = 0.25$$

And the diagonal elements $T(i,i)$ for $i = 1,2,3$ are:

$$T(1,1) = 1 - T(1,2) - T(1,3) = 1 - 0.08 - 0 = 0.92$$
$$T(2,2) = 1 - T(2,1) - T(2,3) = 1 - 0.1 - 0.25 = 0.65$$
$$T(3,3) = 1 - T(3,1) - T(3,2) = 1 - 0 - 1 = 0$$

Therefore, the transition matrix is given by:

$$T = \begin{pmatrix} 0.92 & 0.08 & 0 \\ 0.1 & 0.65 & 0.25 \\ 0 & 1 & 0 \end{pmatrix}$$

6. **Full conditional distributions** $\boxed{15}$

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

(b) $\boxed{5}$ Right (R) or Wrong (W):

(i) W, (ii) R, (iii) R, (iv) W, and (v) W.

(c) $\boxed{5}$ EXAMPLE: The 1st full conditional distribution:

$$p(\theta|a, b, y_1, \ldots, y_n)$$

'by definition':

$$= \frac{p(\theta, a, b, y_1, \ldots, y_n)}{p(a, b, y_1, \ldots, y_n)}$$

'as the denominator is independent of θ'

$$\propto p(\theta, a, b, y_1, \ldots, y_n)$$

'the joint PDF can be factorized':

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

'as the factor $p(a, b)$ is independent of θ'

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

given θ, $y_1, \ldots, y_n$ are independent of the hyperparameters a and b

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

ANALOGOUSLY IT CAN BE DERIVED:

$$p(b|a, \theta, y_1, \ldots, y_n) = \frac{p(\theta, a, b, y_1, \ldots, y_n)}{p(a, b, y_1, \ldots, y_n)} \propto p(\theta, a, b, y_1, \ldots, y_n)$$

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

$$p(a|b, \theta, y_1, \ldots, y_n) = \frac{p(\theta, a, b, y_1, \ldots, y_n)}{p(a, b, y_1, \ldots, y_n)} \propto p(\theta, a, b, y_1, \ldots, y_n)$$

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

END