Online Re-Exam Statistical Reasoning 2020/2021

Date: Thursday, June 24, 2021
Time: 15:00-18:00
Place: Nestor Online Exam
Progress code: WBMA038-05

Rules to follow:

- This is an open book online exam. You are allowed to consult: Textbooks, written and printed notes, all files on your computer disc and all files on Nestor.
  You are **neither** allowed to communicate with anyone about the exercises **nor** to use the internet to search for possible solutions.
- **We wish you success with the completion of the exam!**

START OF RE-EXAM

1. **Joint, marginal and conditional probabilities.** $\boxed{\mathbf{15}}$
   There are three urns. The first urn contains 80 red and 20 blue balls. The second urn contains 100 red balls and no blue balls. The third urn contains 10 red and 90 blue balls. One urn is randomly chosen (each urn has probability $p = 1/3$ of being chosen), and a ball is randomly drawn from the selected urn.
   The variable U describes which urn was chosen, $U = i$ for urn i ($i = 1, 2, 3$), and the variable R indicates if a red ($R = 1$) or blue ($R = 0$) ball was drawn.

   (a) $\boxed{5}$ Give the joint distribution of U and R.

   (b) $\boxed{5}$ Give the marginal distribution of U and the marginal distribution of R.

   (c) $\boxed{5}$ Give the conditional distribution of U given $R = 1$.

2. **Marginal Likelihood.** $\boxed{\mathbf{10}}$
   The random variable Y has the density

$$p(y|\theta) = \frac{\theta^y \cdot e^{-\theta}}{\Gamma(y+1)} \quad (y \in \mathbb{N}_0)$$

   On the unknown parameter $\theta > 0$ we impose a prior with density:

$$p(\theta) = \lambda \cdot e^{-\lambda\theta} \quad (\theta > 0)$$

   where $\lambda > 0$ is a known hyperparameter.
   <u>**EXERCISE**</u>: Show that the marginal distribution of Y then has the density:

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

   <u>RECALL</u>:
   The Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ has density

$$p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\beta \cdot x} \quad (x > 0)$$

3. **Power posterior distribution.** $\boxed{10}$
Consider a random (iid) sample from a Gamma distribution with parameters $a_1 > 0$ and $a_2 > 0$, symbolically:

$$Y_1, \dots, Y_n | b_1 \sim \text{GAM}(a_1, b_1)$$

where a_1 is known and b_1 is an unknown parameter. On b_1 we impose a Gamma prior distribution with parameters $a_2 > 0$ and $b_2 > 0$

$$b_1 \sim \text{GAM}(a_2, b_2)$$

EXERCISE: $\boxed{10}$ Compute the power posterior distribution of b_1, for the given inverse temperature $\tau \in [0, 1]$.

RECALL:
The Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ has density:

$$p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\beta \cdot x} \qquad (x > 0)$$

4. **Wishart sample.** $\boxed{30}$
Consider a random (iid) sample from an S-dimensional Wishart distribution:

$$\mathbf{W}_1, \dots, \mathbf{W}_n | \mathbf{V} \sim \mathcal{W}(\alpha, \mathbf{V})$$

where $\alpha > S + 1$, and $\mathbf{V}$ is a positive definite S-by-S matrix, symbolically $\mathbf{V} > 0$. We assume that α is known, and on $\mathbf{V}$ we impose a Wishart prior distribution

$$\mathbf{V} \sim \mathcal{W}(\alpha_2, \mathbf{V}_0)$$

where $\alpha_2 > S + 1$ and $\mathbf{V}_0$ is a positive definite S-by-S matrix.

(a) $\boxed{20}$ Compute the posterior distribution of $\mathbf{V}$.

(b) $\boxed{10}$ Show that the Wishart distribution for $S = 1$ becomes a Gamma distribution (see HINT of Exercise 2 for the density of the Gamma distribution).

RECALL:
The density of an N-dimensional Wishart distribution with parameters $\alpha > N + 1$ and positive definite matrix $\mathbf{T}$ is given by:

$$p(\mathbf{W}|\mathbf{T}, \alpha) = \frac{\det(\mathbf{W})^{(\alpha-N-1)/2}}{Z(N, \mathbf{T}, \alpha)} \cdot \exp(-\frac{1}{2} tr(\mathbf{TW})) \qquad (\mathbf{W} \in \mathbb{R}^{N,N} \text{ with } \mathbf{W} > 0)$$

where $tr(.)$ is the trace operator, and $Z(N, \mathbf{T}, \alpha)$ is defined as:

$$Z(N, \mathbf{T}, \alpha) = 2^{N \cdot \alpha/2} \cdot \pi^{N(N-1)/4} \cdot \det(\mathbf{T})^{-\alpha/2} \cdot \prod_{j=1}^{N} \Gamma\left(\frac{\alpha - j + 1}{2}\right)$$

Moreover, recall that we have for the trace operator:

$$tr(\mathbf{AB}) = tr(\mathbf{BA}) \quad \text{and} \quad tr(\mathbf{A}) + tr(\mathbf{B}) = tr(\mathbf{A} + \mathbf{B})$$

5. **Gaussian distributions.** $\boxed{25}$
Consider an (iid) random sample from a Gaussian distribution

$$Y_1, \ldots, Y_n | \mu \sim \mathcal{N}(\mu, \sigma^2)$$

where $\sigma^2 > 0$ is known and μ is an unknown parameter with prior distribution

$$\mu | \theta \sim \mathcal{N}(\theta, \sigma^2)$$

We assume that θ is also unknown and has the hyperprior distribution:

$$\theta \sim \mathcal{N}(1, \sigma^2)$$

We define the random vector

$$\mathbf{Y} := (Y_1, \ldots, Y_n)^\mathsf{T}$$

Use the rules below (see RULES) to compute

(a) $\boxed{5}$ $\mathbf{Y}|\theta$

(b) $\boxed{5}$ $\mathbf{Y}$

(c) $\boxed{5}$ $\mathrm{VAR}(Y_1)$ and $\mathrm{COV}(Y_1, Y_2)$

(d) $\boxed{5}$ $\mu|(\mathbf{Y}, \theta)$

(e) $\boxed{5}$ $\theta|\mu$

and simplify your solutions whenever possible.

RULES:
Let $\mathbf{\Sigma}_1 \in \mathbb{R}^{S,S}$ and $\mathbf{\Sigma}_2 \in \mathbb{R}^{k,k}$ be (positive definite) covariance matrices, let $\boldsymbol{\mu} \in \mathbb{R}^k$ denote a vector, and let $\mathbf{X} \in \mathbb{R}^{S,k}$ be a design matrix. For a Gaussian (regression) likelihood with a Gaussian prior on the regression parameter vector $\boldsymbol{\beta} \in \mathbb{R}^k$:

$$\begin{aligned} \mathbf{y}|\boldsymbol{\beta} &\sim \mathcal{N}_S(\mathbf{X}\boldsymbol{\beta}, \mathbf{\Sigma}_1) \\ \boldsymbol{\beta} &\sim \mathcal{N}_k(\boldsymbol{\mu}, \mathbf{\Sigma}_2) \end{aligned}$$

we have the posterior distribution:

$$\boldsymbol{\beta}|\mathbf{y} \sim \mathcal{N}(\mathbf{\Sigma}_\dagger(\mathbf{X}^\mathsf{T}\mathbf{\Sigma}_1^{-1}\mathbf{y} + \mathbf{\Sigma}_2^{-1}\boldsymbol{\mu}) \ , \ \mathbf{\Sigma}_\dagger) \ \text{ where } \ \mathbf{\Sigma}_\dagger = (\mathbf{\Sigma}_2^{-1} + \mathbf{X}^\mathsf{T}\mathbf{\Sigma}_1^{-1}\mathbf{X})^{-1}$$

and the marginal distribution

$$\mathbf{y} \sim \mathcal{N}_S(\mathbf{X}\boldsymbol{\mu}, \ \mathbf{\Sigma}_1 + \mathbf{X}\mathbf{\Sigma}_2\mathbf{X}^\mathsf{T})$$