Bayesian Statistics, Final Exam, MIMUW UW, January 2019

1. Let $X_1, \ldots, X_n$ be an i.i.d. sample from the normal distribution $\mathrm{N}(\theta, \sigma^2)$ with known σ. Consider the standard conjugate prior $\theta \sim \mathrm{N}(\mu, v^2)$.

   (a) Compute the shortest credible interval for θ at level $1 - \alpha = 0.95$, i.e. $[a, b]$ such that $\mathbb{P}(a \le \theta \le b | x_1, \ldots, x_n) = 1 - \alpha$.

   (b) What is the mninimum sample size n such that the length of this interval is bounded by a given d, i.e. $b - a \le d$?

   (c) Give the answer to the previous question in the case when $v = \infty$ (pass to the limit as $v \to \infty$).

   Solution: We know that the posterior is

$$\mathrm{N}\left(\mu_x, \sigma_x^2\right), \text{ where } \mu_x = \frac{nv^2\bar{x} + \sigma^2\mu}{nv^2 + \sigma^2}, \;\; \sigma_x^2 = \frac{\sigma^2 v^2}{nv^2 + \sigma^2}.$$

   (a) The shortest credible interval is a HPD region, i.e. the set where the posterior density is above a level. Since the normal density is unimodal and symmetric, the HPD is of the form $[\mu_x - d/2, \mu_x + d/2]$. The desired probability $1 - \alpha$ obtains if we choose $d/2 = z\sigma_x$, where $\Phi(z) = 1 - \alpha/2$. In our case $z \approx 2$, so the answer is

$$[\mu_x - 2\sigma - x, \mu_x + 2\sigma - x]$$

   (b) $2\sigma_x \le d$ iff $n \ge 16\sigma^2/d^2 - \sigma^2/v^2$, from the formula for σ_x^2.

   (c) $\sigma_x^2 \sim \sigma^2/n$ and we get $n \ge 16\sigma^2/d^2$.

---

2. As in the previous problem, $X_1, \ldots, X_n$ is an i.i.d. sample from $\mathrm{N}(\theta, \sigma^2)$ with known σ and $\theta \sim \mathrm{N}(\mu, v^2)$.

   (a) Give the (marginal) distribution of $\bar{X}$ in this model.

   (b) Give the (marginal) distribution of $X_1 - X_2$.

   (c) Compute $\mathrm{Cov}(\theta, \bar{X} - \theta)$.

   (d) Compute $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n X_i^2$.

   Solution: The conditional distribution of $Y_i = X_i - \theta | \theta$ is $\mathrm{N}(0, \sigma^2)$. It does not involve θ, so Y_is are independent of θ. This fact, verified at the tutorials, facilitates computations:

   (a) $\bar{X} \sim \mathrm{N}(\mu, v^2 + \sigma^2/n)$, because $\bar{X} = \theta + \bar{Y}$ and $\bar{Y} \sim \mathrm{N}(0, \sigma^2/n)$.

   (b) $X_1 - X_2 \sim \mathrm{N}(0, 2\sigma^2)$, obviously.

   (c) $\mathrm{Cov}(\theta, \bar{X} - \theta) = 0$ by independence.

   (d) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n X_i^2 = \mathbb{E}(X_1^2 | \theta)$ by the SLLN and conditional independence. Since $\mathbb{E}(X_1^2 | \theta) = \theta^2 + \sigma^2$, we get

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n X_i^2 = \theta^2 + \sigma^2.$$

   Remark: This is a.s. convergence to a *random variable*, not to a number. We cannot apply SLLN to X_i^2s unconditionally, because they are not unconditionally independent.

---

3. We observe a (1-dim)random variable X, which belongs to one of two classes: $C = 1$ or $C = 2$. Class-conditional probability distributions are normal,

$$(X|C = 1) \sim \mathrm{N}(-\mu, 1) \text{ and } (X|C = 2) \sim \mathrm{N}(\mu, 1).$$

Prior probabilities of both the classes are equal, $\mathbb{P}(C = 1) = \mathbb{P}(C = 2) = 1/2$. We are to compute a "discriminant function" $\delta : \mathbb{R} \to \mathbb{R}$ which minimizes the Bayes risk $r(\delta) = \mathbb{E}\ell(C, \delta(X))$ for the following exponential loss function:

$$\ell(c, \delta(x)) = \begin{cases} \mathrm{e}^{-\delta(x)} & \text{for } c = 2; \\ \mathrm{e}^{\delta(x)} & \text{for } c = 1. \end{cases}$$

(a) Compute the optimal function δ^*.

(b) Compute the posterior risk $r_x(\delta^*)$ for this function.

(c) Compute the Bayes risk $r(\delta^*)$ for this function in terms of μ.

Remark: A "*discriminant function*" $\delta : \mathbb{R} \to \mathbb{R}$ should not be confused with a *classification rule* which is a function $\mathbb{R} \to \{1, 2\}$. The problem of finding a "*discriminant function*" makes sense, because we want "the bigger $\delta(x)$ – the higher posterior probability $\mathbb{P}(C = 2|x)$". In fact, the answer to question (a) is very intuitive:

Solution:

(a) $\delta^*(x) = \frac{1}{2} \log \frac{\pi_x(2)}{\pi_x(1)}$, where $\pi_x(c) = \mathbb{P}(C = c|x)$. Indeed, the standard way is to minimize the posterior risk $r_x(a) = \pi_x(2)\mathrm{e}^{-a} + \pi_x(1)\mathrm{e}^{a}$ w.r.t. $a = \delta(x)$ with fixed x. Differentiating $r_x(a)$ w.r.t. $a = \delta(x)$ we get $\mathrm{e}^{2a}\pi_x(1) = \pi_x(2)$. If the class-conditional distributions are normal with equal variances, we have $\frac{\pi_x(2)}{\pi_x(1)} = 2\mu x$ (this is a special case of a formula derived at our tutorials and can easily be directly computed). Hence the answer is:

$$\text{In general } \delta^*(x) = \frac{1}{2} \log \frac{\pi_x(2)}{\pi_x(1)}. \text{ In our normal model } = \mu x.$$

(b) The posterior risk $r_x(\delta^*)$ is obtained by substituting the formula for $a = \delta^*(x)$ into the expression for $r_x(a)$. The answer is:

$$\text{In general } r_x(\delta^*) = \sqrt{\pi_x(2)\pi_x(1)}. \text{ In our model } = \frac{1}{\mathrm{e}^{\mu x/2} + \mathrm{e}^{-\mu x/2}}$$

(c) $r(\delta^*) = \frac{1}{2}\mathbb{E}(\mathrm{e}^{\mu X}|C = 1) + \frac{1}{2}\mathbb{E}(\mathrm{e}^{-\mu X}|C = 2)$. Using the formula for the m.g.f. of a normal distribution we get $\mathbb{E}(\mathrm{e}^{\mu X}|C = 1) = \mathrm{e}^{-\mu^2+\mu^2/2}$ so by symmetry,

$$r(\delta^*) = \mathrm{e}^{-\mu^2/2}.$$

---

4. Let $X|\theta \sim \text{Poiss}(\theta)$. We observe $X = x$ and consider the problem of testing

$H_0 : \theta = \lambda$, where λ is a given positive number;

against

$H_1 : \theta \sim \text{Ex}(1/\lambda)$ (the exponential distribution with expectation λ).

(a) Compute the Bayes Factor $B_{10}(x) = \mathbb{P}(x|H_1)/\mathbb{P}(x|H_0)$ (my error, sorry!).

(b) For what value of x the Bayes Factor $B_{10}(x)$ is minimal?

(c) Consider the symmetric 0-1 loss and equal prior probabilities of the two hypotheses, $\mathbb{P}(H_0) = \mathbb{P}(H_1) = 1/2$. Let δ^* be the optimal test, i.e. the Bayes decision rule $\delta^* : \{0, 1, 2, \ldots\} \to \{0, 1\}$. What is the decision $\delta^*(x)$ in the special case $\lambda = 1$ and $x = 2$?

Solution:

(a) Of course, $\mathbb{P}(x|H_0) = \mathrm{e}^{-\lambda}\frac{\lambda^x}{x!}$. To compute $\mathbb{P}(x|H_1)$ we need to integrate out θ: $\int_0^\infty \mathrm{e}^{-\theta}\frac{\theta^x}{x!}\frac{1}{\lambda}\mathrm{e}^{-\theta/\lambda}\mathrm{d}\theta = \frac{1}{\lambda x!}\int_0^\infty \theta^x \mathrm{e}^{-\theta(1+1/\lambda)}\mathrm{d}\theta = \frac{1}{\lambda x!}\frac{\Gamma(x+1)}{(1+1/\lambda)^{x+1}} = \frac{1}{\lambda}\left(\frac{\lambda}{\lambda+1}\right)^{x+1}$. Hence

$$B_{10}(x) = \frac{\mathbb{P}(x|H_1)}{\mathbb{P}(x|H_0)} = \mathrm{e}^{\lambda}\frac{x!}{(\lambda+1)^{x+1}}.$$

(b) $B_{10}(x)/B_{10}(x-1) = \frac{x}{\lambda+1}$, so the minimum of $B_{10}(x)$ is attained for $x = \lfloor \lambda + 1 \rfloor$.

(c) The decision $\delta^*(x)$ is 1 iff $B_{10}(x) > 1$. For $\lambda = 1$ and $x = 2$ we obtain $B_{10}(x) = \mathrm{e}/4 < 1$, so we do not reject H_0.

5. Let $(X_0, X_1, \dots, X_k) = X_{0:k}$ be a Markov chain on the state space $\{1, 2\}$ with the transition matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}.$$

We assume that the chain is strictly stationary (the initial distribution $\mathbb{P}(X_0 = 1) = \mathbb{P}(X_0 = 2) = 1/2$ is stationary). We consider a Hidden Markov Model (HMM): variables $X_0, X_1, \dots, X_k$ are not directly observed. We observe random variables $Y_0, Y_1, \dots, Y_k$ with values in $\{1, 2\}$ such that Y_i depends only on X_i,

$$L(x, y) = \mathbb{P}(Y_i = y | X_i = x) = \begin{cases} 1-\varepsilon & \text{for } y = x; \\ \varepsilon & \text{for } y \neq x, \end{cases} \qquad x, y \in \{1, 2\}.$$

We are interested in the posterior distribution $\pi_{\text{post}}(x_{0:k}) = \mathbb{P}(X_{0:k} = x_{0:k} | y_{0:k})$ on the space $\{1, 2\}^{k+1}$. To construct a Gibbs Sampler, we need full conditional distributions $\pi_{\text{post}}(\cdot)$, that is $\pi_{\text{post}}(x_i | x_{-i})$ where $x_{-i} = (x_0, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$.

(a) Compute $\mathbb{P}(X_i = 2 | X_{i-1} = 2, X_{i+1} = 1, Y_i = 1)$ (for $0 < i < k$).

(b) Consider other configurations (there are not so many essentially different cases).

Solution: By Bayes formula,

$$\begin{aligned} &\mathbb{P}(X_i = x | X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, Y_i = y) \\ &= \frac{\mathbb{P}(X_{i-1} = x_{i-1}, X_i = x, X_{i+1} = x_{i+1}, Y_i = y_i)}{\sum_{x'} \mathbb{P}(X_{i-1} = x_{i-1}, X_i = x', X_{i+1} = x_{i+1}, Y_i = y)} \\ &= \frac{P(x_{i-1}, x) P(x, x_{i+1}) L(x, y)}{\sum_{x'} P(x_{i-1}, x') P(x', x_{i+1}) L(x', y)}. \end{aligned}$$

Thus for our model

(a) $\mathbb{P}(X_i = 2 | X_{i-1} = 2, X_{i+1} = 1, Y_i = 1)$

$$\begin{aligned} &= \frac{P(2,2)P(2,1)L(2,1)}{P(2,2)P(2,1)L(2,1) + P(2,1)P(1,1)L(1,1)} \\ &= \frac{(1-\alpha)\alpha\varepsilon}{(1-\alpha)\alpha\varepsilon + \alpha(1-\alpha)(1-\varepsilon)} = \varepsilon. \end{aligned}$$

(b) For $0 < i < k$, it is enough to consider three essentially different cases:

- $X_{i-1} \neq X_{i+1}$ (as in question (a)),
- $X_{i-1} = X_{i+1} \neq Y_i$,
- $X_{i-1} = X_{i+1} = Y_i$,

due to the symmetry of the model. (Cases $i = 0$ and $i = k$ have to be considered separately; note that the initial distribution is relevant for $i = 0$).

6. Let $X_1, \ldots, X_n$ be an i.i.d. sample from the density

$$p_\theta(x) = \begin{cases} \theta x^{\theta-1} & \text{for } 0 < x < 1; \\ 0 & \text{otherwise.} \end{cases}$$

The prior distribution of the parameter θ is Gamma(α, λ).

(a) Find a 1-dim sufficient statistic in this model.

(b) Compute the posterior distribution of $\theta | X_1, \ldots, X_n$.

(c) Compute the marginal distribution of X_1.

(d) Compute the predictive distribution of $X_{n+1} | X_1, \ldots, X_n$. Of course, we now assume that $X_1, \ldots, X_n, X_{n+1} | \theta$ are conditionally i.i.d.

Solution:

(a) $\sum \log(x_i)$ (or equivalently $\prod x_i$) is a natural sufficient statistic for the exponential family of densities p_θ:

$$p_\theta(x_1, \ldots, x_n) = \theta^n \left(\prod x_i \right)^{\theta-1}.$$

(b) We have a conjugate prior,

$$\pi(\theta | x_1, \ldots, x_n) \propto \theta^n \left(\prod x_i \right)^{\theta-1} \theta^{\alpha-1} \exp\{-\lambda\theta\}$$

$$\propto \theta^{\alpha+n-1} \exp\left\{ -\left(\lambda - \sum \log x_i \right) \theta \right\},$$

so the posterior is Gamma$\left(\alpha + n, \lambda - \sum \log x_i\right)$.

(c) The marginal:

$$p(x) = \int_0^1 \theta x^{\theta-1} \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp\{-\lambda\theta\} \mathrm{d}\theta$$

$$= \frac{\lambda^\alpha}{x\Gamma(\alpha)} \int_0^1 \theta^\alpha \exp\{-(\lambda - \log x)\theta\} \mathrm{d}\theta = \frac{\lambda^\alpha}{x\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{(\lambda - \log x)^{\alpha+1}}$$

$$= \frac{\alpha\lambda^\alpha}{x(\lambda - \log x)^{\alpha+1}}.$$

(d) The predictive distribution obtains if we substitute posterior parameters in place of prior ones in the formula for the marginal:

$$p(x_{n+1} | x_1, \ldots, x_n) = \frac{(\alpha + n) \left(\lambda - \sum \log x_i\right)^{\alpha+n}}{x_{n+1} \left(\lambda - \sum \log x_i - \log x_{n+1}\right)^{\alpha+n+1}}.$$