Problem class 2: Statistical decision theory, Bayesian point estimation, Credible sets

Lecturer: Georgios Karagiannis georgios.karagiannis@durham.ac.uk

1 Bayesian point estimation

Exercise 1. ($\star\star$)Consider observables $x = (x_1, ..., x_n)$. Consider the Bayesian model

$$\begin{cases} x_i|\theta & \overset{\text{IID}}{\sim} \text{N}(\theta, 1), \quad i = 1, ..., n \\ \theta & \sim \Pi(\theta) \end{cases}$$

where $\pi(\theta) \propto 1$ and that we have only one observable. Consider the LINEX loss function

$$\ell(\theta, \delta) = \exp(c(\theta - \delta)) - c(\theta - \delta) - 1$$

1. Show that $\ell(\theta, \delta) \geq 0$
2. Find the Bayes estimator $\hat{\delta}$ under LINEX loss function and under the given Bayesian model.

Hint-1: Random variable B follows a log-normal distribution $B \sim \text{LN}(\mu_A, \sigma_A^2)$ with parameters μ_A, σ_A^2 if $B = \exp(A)$ where $A \sim \text{N}(\mu_A, \sigma_A^2)$.

Hint-2: If $B \sim \text{LN}(\mu_A, \sigma_A^2)$ then $\text{E}_{\text{LN}(\mu_A, \sigma_A^2)}(B) = \exp(\mu_A + \frac{\sigma_A^2}{2})$.

Hint-3: It is

$$-\frac{1}{2}\frac{(\mu - \mu_1)^2}{v_1^2} - \frac{1}{2}\frac{(\mu - \mu_2)^2}{v_2^2} ... - \frac{1}{2}\frac{(\mu - \mu_n)^2}{v_n^2} = -\frac{1}{2}\frac{(\mu - \hat{\mu})^2}{\hat{v}^2} + C$$

where

$$\hat{v}^2 = \left(\sum_{i=1}^{n} \frac{1}{v_i^2}\right)^{-1}; \quad \hat{\mu} = \hat{v}^2 \left(\sum_{i=1}^{n} \frac{\mu_i}{v_i^2}\right); \quad C = \frac{1}{2}\frac{\hat{\mu}^2}{\hat{v}^2} - \frac{1}{2}\sum_{i=1}^{n} \frac{\mu_i^2}{v_i^2}$$

Exercise 2. ($\star\star$)Suppose we wish to estimate the values of a collection of discrete random variables $\vec{X} = X_1, \ldots, X_n$. We have a posterior joint probability mass function for these variables, $p(\vec{x}|y) = p(x_1, \ldots, x_n|y)$ based on some data y. We decide to use the following loss function:

$$\ell(\hat{\vec{x}}, \vec{x}) = \sum_{i=1}^{n} (1 - \delta(\hat{x}_i, x_i)) \tag{1}$$

where $\delta(a, b) = 1$ if $a = b$ and zero otherwise.

1. Derive an expression for the estimated values, found by minimizing the expectation of the loss function. [Hint: use linearity of expectation.]
2. When the probability distribution is a posterior distribution in some problem, this type of estimate is sometimes called 'maximum posterior marginal' (MPM) estimate. Explain why this name is appropriate.
3. Explain in words what the loss function is measuring. Compare with the loss function for MAP estimation.

2 Credible sets

Exercise 3. ($\star\star$) (Example from the Lecture's handout) Consider a Bayesian model

$$\begin{cases} y_i|\mu & \overset{\text{iid}}{\sim} \text{N}_d(\mu,\Sigma), \qquad i=1,...,n \\ \mu & \sim \text{N}_d(\mu_0,\Sigma_0) \end{cases}$$

where uncertain $\mu \in \mathbb{R}^d$, $d \geq 1$, and known Σ, μ_0, Σ_0. Find the C_a parametric HPD credible set for μ.

Hint-1: If $z=(z_1,...,z_d)^\top$ such as $z_j \overset{\text{iid}}{\sim} \text{N}(0,1)$ for $j=1,...,d$, and $\xi = z^\top z = \sum_{j=1}^d z_j^2$, then $\xi \sim \chi^2_d$

Hint-2: It is

$$-\frac{1}{2}\sum_{i=1}^n (x-\mu_i)^\top \Sigma_i^{-1}(x-\mu_i)) = -\frac{1}{2}(x-\hat{\mu})^\top \hat{\Sigma}^{-1}(x-\hat{\mu})) + C(\hat{\mu},\hat{\Sigma}) \quad ;$$

$$\hat{\Sigma} = (\sum_{i=1}^n \Sigma_i^{-1})^{-1}; \quad \hat{\mu} = \hat{\Sigma}(\sum_{i=1}^n \Sigma_i^{-1}\mu_i);$$

$$C(\hat{\mu},\hat{\Sigma}) = \underbrace{\frac{1}{2}(\sum_{i=1}^n \Sigma_i^{-1}\mu_i)^\top (\sum_{i=1}^n \Sigma_i^{-1})^{-1}(\sum_{i=1}^n \Sigma_i^{-1}\mu_i) - \frac{1}{2}\sum_{i=1}^n \mu_i^\top \Sigma_i^{-1}\mu_i}_{=\text{independent of } x}$$

Example 4. ($\star\star$) (Example from the Lecture's handout) Assume a 1-dimensional random quantity $x \sim Q(x|y)$, with unimodal density $q(x|y)$. Show that the $(1-a)$-credible interval $C_a=[L,U]$ for x as a Bayesian rule C_a under the loss function

$$\ell(x,C_a;L,U) = k(U-L) - 1(x\in[L,U]), \quad \text{with} \quad k \in (0, \max_{\forall x \in \mathbb{R}}(q(x|y)))$$

is given by $q(L)=q(U)=k$, and $\mathsf{P}_Q(x\in[L,U]|y) = 1-a$.

Discuss known properties of the derived credible interval.

Solution. The decision space is $\mathcal{D} = \{C_a = [L,U] \,:\, \Pr_Q(x \in C_a|y) = 1-a\}$. It is

$$\begin{aligned}\text{E}_Q\left(\ell(x,C_a;L)|y\right) &= \int \left(k(U-L) - 1(x\in[L,U])\right)\text{d}Q(x|y) \\ &= \int k(U-L)q(x|y)\text{d}x - \int_L^U q(x|y)\text{d}x = k(U-L) - \int_{-\infty}^U q(x|y)\text{d}x + \int_{-\infty}^L q(x|y)\text{d}x\end{aligned}$$

To find the critical values $\hat{L}$, and $\hat{U}$ for L and U, it is

$$\begin{aligned}0 = &\left.\frac{\text{d}}{\text{d}L}\text{E}_Q\left(\ell(x,C_a;L)|y\right)\right|_{C_a=[\hat{L},\hat{U}]} = \frac{\text{d}}{\text{d}L}\left.\left(k(U-L) - \int_{-\infty}^U q(x|y)\text{d}x + \int_{-\infty}^L q(x|y)\text{d}x\right)\right|_{C_a=[\hat{L},\hat{U}]} \\ &= -k + q(\hat{L}|y) \implies q(\hat{L}|y) = k \\ 0 = &\left.\frac{\text{d}}{\text{d}U}\text{E}_Q\left(\ell(x,C_a;U)|y\right)\right|_{C_a=[\hat{L},\hat{U}]} = ... = k - q(\hat{U}|y) \implies q(\hat{U}|y) = k\end{aligned}$$

which are minimizers because

$$\frac{\mathrm{d}^2}{\mathrm{d}L^2}\mathrm{E}_Q\left(\ell(x,C_a;L)|y\right)\Big|_{C_a=[\hat{L},\hat{U}]} = \frac{\mathrm{d}}{\mathrm{d}L}q(L|y)\Big|_{\hat{L}} > 0\ ; \qquad \frac{\mathrm{d}^2}{\mathrm{d}L\mathrm{d}U}\mathrm{E}_Q\left(\ell(x,C_a;L)|y\right)\Big|_{C_a=[\hat{L},\hat{U}]} = 0$$
$$\frac{\mathrm{d}^2}{\mathrm{d}U^2}\mathrm{E}_Q\left(\ell(x,C_a;U)|y\right)\Big|_{C_a=[\hat{L},\hat{U}]} = -\frac{\mathrm{d}}{\mathrm{d}U}q(U|y)\Big|_{\hat{U}} > 0$$

So it is $C_a = [\hat{L},\hat{U}]$ such that $q(\hat{L}|y) = q(\hat{U}|y) = k$, and $\Pr_Q(x \in [\hat{L},\hat{U}]|y) = 1-a$.

Based on Theorem **??**, it is the HPD credible interval and in fact the shorter length credible interval.

Example 5. ($\star\star$) (Example from the Lecture's handout) Assume an 1- dimensional random quantity $x \sim Q(x|y)$. In the Lecture Handout (Handout 11: Bayesian point estimation), discussed the follwoing Hint:

Hint: The Bayes estimate $\hat{\delta}$ of x under the linear loss function

$$\ell(x,\delta;\varpi) = (1-\varpi)(\delta-x)1_{x\le\delta}(\delta) + \varpi(x-\delta)1_{x>\delta}(\delta),$$

where $\varpi \in [0,1]$, is the ϖ-th quantile of distribution Q, let's denote it as x_ϖ.

1. Derive the $(1-a)$-credible interval $C_a = [L,U]$ for x as a Bayesian rule C_a under the loss function

$$\ell(x,C_a;\varpi_L,\varpi_U) = \ell(x,L;\varpi_L) + \ell(x,U;\varpi_U) \tag{2}$$

by computing L and U.

2. Your client is worried the same both for under-estimation and over-estimation; derive a suitable $(1-a)$-credible interval $C_a = [L,U]$ based on (2) by computing L, and U.

3. Your client is worried only for over-estimation; derive a suitable $(1-a)$-credible interval $C_a = [L,U]$ based on (2) by computing L and U.

Solution. It is given that

$$0 = \frac{\mathrm{d}}{\mathrm{d}\delta}\mathrm{E}_Q\left(\ell(x,\delta;\varpi)|y\right)\Big|_{\delta=\hat{\delta}} = \frac{\mathrm{d}}{\mathrm{d}\delta}\int \ell(x,\delta;\varpi)\mathrm{d}Q(x|y)\Big|_{\delta=\hat{\delta}} \implies \hat{\delta} = x_\varpi$$
$$= (1-\varpi)\Pr_Q\left(\{x\le\hat{\delta}\}|y\right) - \varpi\Pr_Q\left(\{x\le\hat{\delta}\}^{\complement}|y\right) \implies \hat{\delta} = x_\varpi$$

1. The decision space is $\mathcal{D} = \{C_a = [L,U]\ :\ \Pr_Q(x\in C_a|y) = 1-a\}$. Therefore, to find the Bayes rule (or Bayes estimate) of $C_a = [L,U]$ I need to minimize the expected posterior loss $\mathrm{E}_Q\left(\ell(x,C_a;\varpi_L,\varpi_U)|y\right)$ with respect to C_a or equivalently L, U, so

$$0 = \frac{\mathrm{d}}{\mathrm{d}L}\mathrm{E}_Q\left(\ell(x,C_a;\varpi_L,\varpi_U)|y\right)\Big|_{C_a=[\hat{L},\hat{U}]} = \mathrm{E}_Q\left(\ell(x,L;\varpi_L)|y\right)\big|_{L=\hat{L}} \implies \hat{L} = x_{\varpi_L}$$
$$0 = \frac{\mathrm{d}}{\mathrm{d}U}\mathrm{E}_Q\left(\ell(x,C_a;\varpi_L,\varpi_U)|y\right)\Big|_{C_a=[\hat{L},\hat{U}]} = \mathrm{E}_Q\left(\ell(x,U;\varpi_U)|y\right)\big|_{U=\hat{U}} \implies \hat{U} = x_{\varpi_U}$$

So $x \in [x_{\varpi_L}, x_{\varpi_U}]$ where $\varpi_U + \varpi_L = 1-a$. It is the minimum because

$$\frac{\mathrm{d}^2}{\mathrm{d}U^2}\mathrm{E}_Q\left(\ell(x,C_a;\varpi_L,\varpi_U)|y\right)\Big|_{C_a=[\hat{L},\hat{U}]} = q(\hat{U}|y) > 0$$

$$\frac{\mathrm{d}^2}{\mathrm{d}L^2}\mathrm{E}_Q\left(\ell(x,C_a;\varpi_L,\varpi_U)|y\right)\Big|_{C_a=[\hat{L},\hat{U}]} = q(\hat{L}|y) > 0$$

$$\frac{\mathrm{d}}{\mathrm{d}U}\frac{\mathrm{d}}{\mathrm{d}L}\mathrm{E}_Q\left(\ell(x, C_a; \varpi_L, \varpi_U)|y\right)\Big|_{C_a=[\hat{L},\hat{U}]} = 0$$

and hence the determinant of the Hessian in positive.

2. Then I can use the equi-tail interval: $x \in [x_{a/2}, x_{1-a/2}]$ with $\varpi_L = a/2$ and $\varpi_U = 1 - a/2$
3. Then I can use the lower-tail interval: $x \in (-\infty, x_{1-a}]$ with $\varpi_L = 0$ and $\varpi_U = 1 - a$.