

**Handout 12: Credible sets<sup>a</sup>**

Lecturer: Georgios P. Karagiannis

georgios.karagiannis@durham.ac.uk

**Aim:** To explain and produce credible regions in the Bayesian framework.**References:**

- Berger, J. O. (2013; Section 4.3.2). Statistical decision theory and Bayesian analysis. Springer Science & Business Media.
- Robert, C. (2007; Section 5.5). The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer Science & Business Media.
- The Matrix Cookbook (Section 7) <http://matrixcookbook.com>

**Web applets:**

- [https://georgios-stats-1.shinyapps.io/demo\\_CredibleSets/](https://georgios-stats-1.shinyapps.io/demo_CredibleSets/)

<sup>a</sup>Author: Georgios P. Karagiannis.**1 Set-up and aim***Notation 1.* Consider a Bayesian model

$$\begin{cases} y|\theta & \sim F(y|\theta) \\ \theta & \sim \Pi(\cdot) \end{cases}$$

where  $y := (y_1, \dots, y_n) \in \mathcal{Y}$  is a sequence of observables, assumed to be generated from the parametric sampling distribution  $F(y|\theta)$  with pdf/pmf  $f(y|\theta)$  and labeled by an unknown parameter  $\theta \in \Theta$  with a prior distribution  $\Pi(\theta)$  with pdf/pmf  $\pi(\theta)$ . Also assume a sequence of  $m$  future outcomes  $z = (y_{n+1}, \dots, y_{n+m})$ .

**AIM:** Instead of just reporting a point value for  $\theta$  (or  $z$ ) and the associated standard error, it is often desirable and clearer to report sets of values  $C_a \subseteq \Theta$  (or  $C_a \subseteq \mathcal{Z}$ ) with a specified probability  $a$  reflecting Your believe that  $\theta \in C_a$  (or  $z \in C_a$ ).

*Note 2.* Recall that

- Posterior degree of believe about uncertain parameter  $\theta \in \Theta \subseteq \mathbb{R}^d$  is quantified via the posterior distribution  $\Pi(\theta|y)$ ;

$$d\Pi(\theta|y) = \pi(\theta|y)d\theta$$

with cdf  $\Pi(\theta|y)$  and pdf/pmf  $\pi(\theta|y)$ .

- Degree of believe about a future sequence of outcomes  $z = (y_{n+1}, \dots, y_{n+m}) \in \mathcal{Z}$  is quantified via the predictive distribution  $G(z|y)$ ;

$$dG(z|y) = g(z|y)dz$$

with cdf  $G(z|y)$  and pdf/pmf  $g(z|y)$ .

*Notation 3.* We present the parametric and predictive credible intervals in a unified framework. Consider unknown random quantity  $x \in \mathcal{X} \subseteq \mathbb{R}^k$  following a distribution  $Q(x|y)$ ;

$$dQ(x|y) = q(x|y)dx$$

with cdf  $Q(x|y)$  and pdf/pmf  $q(x|y)$ . These are dummies for the following:

- In parametric inference, we have  $x \equiv \theta$ ,  $Q \equiv \Pi$ ,  $q \equiv \pi$ , and  $k = d$ .
- In predictive inference, we have  $x \equiv z$ ,  $Q \equiv G$ ,  $q \equiv g$ , and  $k = m$ .
- Note that  $x$  can also be any function of  $\theta$  or  $z$ .

## 2 Credible Sets

**Definition 4.** A set  $C_a \subseteq \mathcal{X}$  is called ‘ $100(1 - a)\%$ ’ posterior credible set for  $x$ , with respect to the posterior distribution  $Q(x|y)$  if

$$1 - a \leq P_Q(x \in C_a|y) = \int 1(x \in C_a) dQ(x|y)$$

*Note 5.* In Bayesian stats (unlike frequentist stats) we can correctly say that the  $(1 - a)100\%$  credible set  $C_a$  of unknown parameter  $\theta$  means that the probability that  $\theta$  is in  $C_a$  is  $(1 - a)100\%$ . This is theoretically correct as everything unknown/uncertain is a random quantity following a distribution reflecting Your degree of believe.

*Note 6.* Note that different sets may satisfy Definition 4 and hence we are interested in using the most useful credible set for our application. This is addressed by imposing additional restrictions.

## 3 Highest probability density Credible sets<sup>1</sup>

*Note 7.* Often it is useful to consider credible sets  $C_a$  which contain values of  $x$  that correspond to the highest pdf/pmf  $q(x|y)$  (aka the most likely values of  $x$ ). Then Definition 4 can be restricted by imposing an extra restriction that:

$$q(x|y) \geq q(x'|y) \text{ for all } x \in C_a, x' \in C_a^c. \quad (1)$$

This leads to the definition of the highest probability density (HPD) set.

**Definition 8.** The  $100(1 - a)\%$  highest probability density (HPD) set for  $x \in \mathcal{X}$  with respect to the posterior distribution  $Q(x|y)$  is the subset  $C_a$  of  $\Theta$  of the form

$$C_a = \{x \in \mathcal{X} : q(x|y) \geq k_a\}$$

where  $k_a$  is the largest constant<sup>2</sup> such that

$$1 - a \leq P_Q(x \in C_a|y)$$

*Note 9.* Credible sets are considered as ‘set estimators’, and hence, they can be produced as Bayes decision rules under a specified loss function.

**Fact 10.** [Minimal size region property] Let random quantity  $x$  follows  $Q(x|y)$ , let  $\mathcal{D} = \{C; P_Q(x \in C|y) \geq 1 - a\}$  be the decision space containing all possible  $(1 - a)$  credible sets of  $x$ , and let the loss function be

$$\ell(x, C) = \kappa \|C\| - I(x \in C), \quad \forall C \in \mathcal{D}, \forall x \in \mathcal{X}, \forall \kappa > 0, \quad (2)$$

where  $\|\cdot\|$  denotes a size of an area. Then:

<sup>1</sup>Web applet: [https://georgios-stats-1.shinyapps.io/demo\\_CredibleSets/](https://georgios-stats-1.shinyapps.io/demo_CredibleSets/)

<sup>2</sup>... this is equivalent to constrain in (1)

1. The Bayes rule (estimator)  $\hat{C}$  has the minimum size among credible sets in  $\mathcal{D}$ .
2.  $\hat{C}$  is the Bayes rule if and only if it is the  $100(1 - a)\%$  highest probability density (HPD) set as defined in Definition 8.

*Proof.* The proof is omitted. (1.) is straightforward; while (2.) is just tricky calculus.  $\square$

*Note.* HPD credible sets are credible sets with the minimum size (by Fact 10). Clearly, loss (2) considers a trade off between two components:  $\|C\|$  measuring the size of the credible set (the smaller the better), and  $1(x \in C)$  indicating coverage of the credible set.

## 4 General discussions

*Remark 11.* HPD credible sets are not, in general, invariant to transformations. If one has computed the HPD set for  $x \sim Q(x|y)$ , the HPD set for  $\varphi = g(x)$  does not necessarily result by converting HPD set for  $x$ . To compute the HPD set for  $\varphi$ , one has to compute the posterior distribution

$$dQ(\varphi|y) = \underbrace{q(g^{-1}(\varphi)|y) \left| \frac{d}{d\varphi} g^{-1}(\varphi) \right|}_{=\pi(\varphi|y)} d\varphi,$$

and then compute the HPD set by implementing Definition 8.

*Note 12.* <sup>3</sup>A (not-that-efficient) algorithm to compute HPD credible sets with a computer is as follows:

1. Create a routine which computes all the solutions  $x^*$  to the equation  $q(x|y) = k_a$ , for a given  $k_a$ . Typically,  $C_a = \{x \in \mathcal{X} : q(x|y) \geq k_a\}$  can be constructed from those solutions.
2. Create a routine which computes

$$P_Q(x \in C_a|y) = \int 1(x \in C_a) dQ(x|y) \quad (3)$$

3. Sequentially solve the equation

$$P_Q(x \in C_a|y) = 1 - a$$

by increasing incrementally  $k_a$  from zero to larger, and stop just before (3) drops below  $1 - a$ .

*Note 13.* For the simple 1D case,  $x \in \mathcal{X}$  with  $\dim(\mathcal{X}) = 1$ , the following theorem can be used to compute HPD credible sets.

**Theorem 14.** Let  $x \in \mathbb{R}$  be a continuous random variable following distribution  $Q(x|y)$  with unimodal density  $q(x|y)$ . If the interval  $C_a = [L, U]$  satisfies

1.  $\int_L^U q(x|y) dx = 1 - a$ ,
2.  $q(U) = q(L) > 0$ , and
3.  $x_{\text{mode}} \in (L, U)$ , where  $x_{\text{mode}}$  is the mode of  $q(x|y)$ ,

then it is the HPD interval of  $x$  with respect to  $Q(x|y)$ .

---

<sup>3</sup>Web-applet [https://georgios-stats-1.shinyapps.io/demo\\_CredibleSets/](https://georgios-stats-1.shinyapps.io/demo_CredibleSets/)



Figure 1: Schematic of Theorem 14 (in Fig. 1(1a)) and Note 12 (in Fig. 1(1a) & Fig. 1(1b))

*Proof.* Out of scope. Use of the mean values theorem to prove. See, Casella, G., & Berger, R. L. (2002; pp. 441-443). Statistical inference (Vol. 2). Pacific Grove, CA: Duxbury.  $\square$

*Remark 15.* Theorem 14 suggests a procedure to find the boundaries of  $C_a$  in 1D cases. As is Figure 1a, we can imagine a horizontal bar which moves from the maximum of the density to zero, and intersects the density at locations which are the potential boundaries of  $C_a$ . The limits of the credible set are where the density above the two points the intersection take place (shaded area) is equal to  $1 - a$ . This mechanism is also described in the algorithm in suggested in Note 12 and hence can also be used in multimodal densities (Figure 1b) or multivariate ones.

## 5 Examples

**Example 16.** Consider a Bayesian model

$$\begin{cases} y_i | \mu & \stackrel{\text{iid}}{\sim} N_d(\mu, \Sigma), & i = 1, \dots, n \\ \mu & \sim N_d(\mu_0, \Sigma_0) \end{cases}$$

where uncertain  $\mu \in \mathbb{R}^d$ ,  $d \geq 1$ , and known  $\Sigma, \mu_0, \Sigma_0$ . Find the  $C_a$  parametric HPD credible set for  $\mu$ .

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

**Hint-2:** It is

$$\begin{aligned}
-\frac{1}{2} \sum_{i=1}^n (x - \mu_i)^\top \Sigma_i^{-1} (x - \mu_i) &= -\frac{1}{2} (x - \hat{\mu})^\top \hat{\Sigma}^{-1} (x - \hat{\mu}) + C(\hat{\mu}, \hat{\Sigma}) \quad ; \\
\hat{\Sigma} &= \left( \sum_{i=1}^n \Sigma_i^{-1} \right)^{-1}; \quad \hat{\mu} = \hat{\Sigma} \left( \sum_{i=1}^n \Sigma_i^{-1} \mu_i \right); \\
C(\hat{\mu}, \hat{\Sigma}) &= \frac{1}{2} \underbrace{\left( \sum_{i=1}^n \Sigma_i^{-1} \mu_i \right)^\top \left( \sum_{i=1}^n \Sigma_i^{-1} \right)^{-1} \left( \sum_{i=1}^n \Sigma_i^{-1} \mu_i \right) - \sum_{i=1}^n \mu_i^\top \Sigma_i^{-1} \mu_i}_{=\text{independent of } x}
\end{aligned}$$

**Solution.** I will use the Definition 8.

- First, I compute the posterior of  $\mu$ . It is

$$\begin{aligned}
\pi(\mu|y) &\propto f(y|\mu) \pi(\mu) = \prod_{i=1}^n \mathcal{N}_d(y_i|\mu, \Sigma) \mathcal{N}_d(\mu|\mu_0, \Sigma_0) \\
&\propto \exp \left( -\frac{1}{2} \sum_{i=1}^n (y_i - \mu)^\top \Sigma^{-1} (y_i - \mu) - \frac{1}{2} (\mu - \mu_0)^\top \Sigma_0^{-1} (\mu - \mu_0) \right) \\
&\propto \exp \left( -\frac{1}{2} (\mu - \hat{\mu}_n)^\top \hat{\Sigma}_n^{-1} (\mu - \hat{\mu}_n) \right)
\end{aligned}$$

where

$$\hat{\Sigma}_n = (n\Sigma^{-1} + \Sigma_0^{-1})^{-1}; \quad \hat{\mu}_n = \hat{\Sigma}_n (n\Sigma^{-1}\bar{y} + \Sigma_0^{-1}\mu_0)$$

I recognize that  $\pi(\mu|y) = \mathcal{N}_d(\mu|\hat{\mu}_n, \hat{\Sigma}_n)$ , and hence  $\mu|y \sim \mathcal{N}_d(\hat{\mu}_n, \hat{\Sigma}_n)$

- Now let's implement Definition 8. So,

$$\begin{aligned}
C_a &= \{ \mu \in \mathbb{R}^d : \pi(\mu|y) \geq k_a \} \\
&= \{ \mu \in \mathbb{R}^d : \mathcal{N}_d(\mu|\hat{\mu}_n, \hat{\Sigma}_n) \geq k_a \} \\
&= \left\{ \mu \in \mathbb{R}^d : (\mu - \hat{\mu}_n)^\top \hat{\Sigma}_n^{-1} (\mu - \hat{\mu}_n) \leq \underbrace{-\log(2\pi \det(\hat{\Sigma}_n))}_{=\tilde{k}_a} k_a \right\} \tag{4}
\end{aligned}$$

and I want the smallest constant  $\tilde{k}_a$  (aka the largest constant  $k_a$ ) such that

$$\begin{aligned}
\mathbb{P}_\Pi(\mu \in C_a|y) &\geq 1 - a \iff \\
\mathbb{P}_\Pi \left( \underbrace{(\mu - \hat{\mu}_n)^\top \hat{\Sigma}_n^{-1} (\mu - \hat{\mu}_n)}_{=\xi} \leq \tilde{k}_a \right) &\geq 1 - a \tag{5}
\end{aligned}$$

- I need to find quantile  $\tilde{k}_a$ . This requires to find the distribution of  $\xi$ . I know that

$$\xi = (\mu - \hat{\mu}_n)^\top \hat{\Sigma}_n^{-1} (\mu - \hat{\mu}_n) \sim \chi_d^2 \tag{6}$$

because  $\xi = z^\top z = \sum_{j=1}^n z_j$  with  $z = L^{-1}(\mu - \hat{\mu}_n) \sim \mathcal{N}_d(0, I_d)$  where  $L$  is the lower matrix of the Cholesky decomposition of  $\hat{\Sigma}_n = L^\top L$ .

Hence Eq. 5, (due to Eqs. 4, 6) becomes

$$P_{\chi_d^2}((\mu - \hat{\mu}_n)^\top \hat{\Sigma}_n^{-1}(\mu - \hat{\mu}_n) \leq \tilde{k}_a) = 1 - a \quad (7)$$

which means that,  $\tilde{k}_a$  is the  $1 - a$  quantile of the  $\chi_d^2$  distribution, aka  $\tilde{k}_a = \chi_{d,1-a}^2$

- Hence, the  $C_a$  parametric HPD credible set for  $\mu$  is

$$C_a = \{\mu \in \mathbb{R}^d : (\mu - \hat{\mu}_n)^\top \hat{\Sigma}_n^{-1}(\mu - \hat{\mu}_n) \leq \chi_{d,1-a}^2\}$$

**Example 17.** Consider an exchangeable sequence of observables  $y := (y_1, \dots, y_n) \in \mathbb{R}^n$  from model

$$\begin{cases} y_i | \theta & \stackrel{\text{iid}}{\sim} \text{Br}(\theta), & i = 1, \dots, n \\ \theta & \sim \text{Be}(a, b) \end{cases}$$

where  $a = b = 2$ ,  $n = 30$ , and  $\sum_{i=1}^{30} y_i = 15$ . Find the 2-sides  $C_a$  parametric HPD credible interval for  $\theta$ . Consider  $a = 0.95$ .

**Solution.**

- The posterior distribution of  $\theta$  is  $\text{Be}(a + n\bar{y}, b + n - n\bar{y})$ , because

$$\pi(\theta|y) \propto \prod_{i=1}^n \text{Br}(y_i|\theta) \text{Be}(\theta|a, b) \propto \prod_{i=1}^n \theta^{y_i} (1 - \theta)^{1-y_i} \theta^{a-1} (1 - \theta)^{b-1} \propto \theta^{n\bar{y}+a-1} (1 - \theta)^{n-n\bar{y}+b-1}$$

After substituting the values of the fixed parameters, I get  $\pi(\theta|y) = \text{Be}(\theta|a_n = 17, b_n = 17)$ .

- To find the 2-sides  $C_a$  parametric HPD credible interval for  $\theta$ , I use Theorem 14.

$$1 - a = \int_L^U \text{Be}(\theta|17, 17) d\theta = P_{\text{Be}(17,17)}(\theta < U) - P_{\text{Be}(17,17)}(\theta < L)$$

I note that the posterior is symmetric around 0.5 because  $a_n = b_n$ . Then,

$$1 - a = P_{\text{Be}(17,17)}(\theta < U) - (1 - P_{\text{Be}(17,17)}(\theta < U)) = 2P_{\text{Be}(17,17)}(\theta < U) - 1$$

so  $P_{\text{Be}(17,17)}(\theta < U) = 1 - a/2$  and  $L = 1 - U$ . For  $a = 0.95$ , the 95% posterior credible interval for  $\theta$  is

$$[L, U] = [0.36, 0.64].$$

- Note that, if we follow the same procedure, the compute the 95% prior credible interval for  $\theta$  is

$$[L, U] = [0.14, 0.85].$$

As expected, the posterior 95 credible interval is narrower than the corresponding posterior one. (Try to check it in R).

```
> install.packages('HDInterval')
> library('HDInterval')
> hdi(qbeta, 0.95, shape1=17, shape2=17)
lower upper
0.3354445 0.6645555
```

**Example 18.** Assume an 1- dimensional random quantity  $x \sim Q(x|y)$ . In the Lecture Handout (Handout 11: Bayesian point estimation), discussed the following Hint:

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

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

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

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

$$\ell(x, C_a; \varpi_L, \varpi_U) = \ell(x, L; \varpi_L) + \ell(x, U; \varpi_U) \quad (8)$$

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 (8) 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 (8) by computing  $L$  and  $U$ .

**Solution.** It is given that

$$\begin{aligned} 0 &= \frac{d}{d\delta} E_Q(\ell(x, \delta; \varpi)|y) \Big|_{\delta=\hat{\delta}} = \frac{d}{d\delta} \int \ell(x, \delta; \varpi) dQ(x|y) \Big|_{\delta=\hat{\delta}} \implies \hat{\delta} = x_{\varpi} \\ &= (1 - \varpi)P_Q(\{x \leq \hat{\delta}\}|y) - \varpi P_Q(\{x \leq \hat{\delta}\}^c|y) \implies \hat{\delta} = x_{\varpi} \end{aligned}$$

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

$$\begin{aligned} 0 &= \frac{d}{dL} E_Q(\ell(x, C_a; \varpi_L, \varpi_U)|y) \Big|_{C_a=[\hat{L}, \hat{U}]} = E_Q(\ell(x, L; \varpi_L)|y) \Big|_{L=\hat{L}} \implies \hat{L} = x_{\varpi_L} \\ 0 &= \frac{d}{dU} E_Q(\ell(x, C_a; \varpi_L, \varpi_U)|y) \Big|_{C_a=[\hat{L}, \hat{U}]} = E_Q(\ell(x, U; \varpi_U)|y) \Big|_{U=\hat{U}} \implies \hat{U} = x_{\varpi_U} \end{aligned}$$

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

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

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

$$\frac{d}{dU} \frac{d}{dL} E_Q(\ell(x, C_a; \varpi_L, \varpi_U)|y) \Big|_{C_a=[\hat{L}, \hat{U}]} = 0$$

and hence the determinant of the Hessian is 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  $\varpi_U = 1 - a$ .

## Practice

**Question 19.** To practice try to work on the Exercises 65, and 74 from the Exercise sheet.