

## Single prior Bayes

Lecture 15

Mean that the distribution of the parameter

( $\Theta$ )

depends on a (new) parameter, hyperparameter  
 $\gamma$ ,  $\sim \Gamma$

$$\text{④ } x | \gamma \sim \pi(\theta | \gamma),$$

$$\bar{x} | \theta \sim f(x | \theta).$$

Here both  $\pi$  and  $\gamma$  are known.

We want to find the Bayes estimator

$$\hat{\delta}(x) = \underset{\theta}{\operatorname{argmax}} \int L(\theta, \delta(x)) \bar{\pi}(\theta | x, \gamma) d\theta$$

where

$$\bar{\pi}(\theta | x, \gamma) = \frac{f(x | \theta) \pi(\theta | \gamma)}{\int f(x | u) \pi(u | \gamma) du}$$

is the posterior distribution.

①

If we have quadratic loss

$$L(\theta, d) = (\hat{\theta}(\theta) - d)^2$$

then we know that the Bayes estimator is given as a conditional expectation

$$\begin{aligned}\hat{\theta}_{\text{Bayes}}(x) &= E(\hat{\theta} | \bar{x} = x) \\ &= \int \theta \pi(\theta | \bar{x}, x) d\theta \\ &= \frac{\int \theta f(x|\theta) \pi(\theta) d\theta}{\int f(x|\theta) \pi(\theta) d\theta}\end{aligned}$$

Two questions that are interesting :

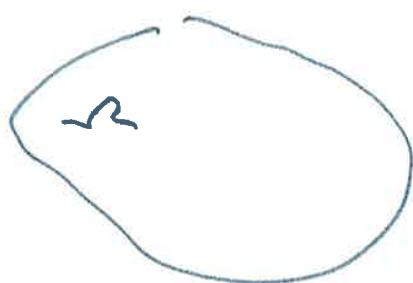
1) How does one calculate

$$E(\hat{\theta} | \bar{x} = x) ?$$

2) How good is the estimator?

→ (to be completed.)

## Minimax estimation



$$\Lambda(\theta)$$

$$r_\Lambda(\delta) = \int R(\theta, \delta) d\Lambda(\theta)$$

Bayes risk.

$$\sup_{\theta \in \Theta} R(\theta, \delta)$$

maximum risk

Bayes estimator

$$\hat{\delta}_\Lambda = \operatorname{argmin}_\delta \int R(\theta, \delta) d\Lambda(\theta)$$

minimax

$$\hat{\delta} = \operatorname{argmin}_\delta \sup_{\theta \in \Theta} R(\theta, \delta) \quad (1)$$

Define, if it exists, the minimax estimator  $\hat{\delta}$  by (1).

Def. A prior distribution  $\Lambda$  for  $\Theta$  is called least favorable for estimating  $g(\theta)$  if

$$r_\Lambda \geq r_{\Lambda'}$$

for any other prior  $\Lambda'$ .

(3)

where

$$r_\lambda = r_\lambda(\delta_\lambda) = r(A, \delta_\lambda)$$

$$r_{\lambda'} = r_{\lambda'}(\delta_{\lambda'}) = r(A', \delta_{\lambda'})$$

$$r_\lambda(\delta) = \int R(G, \delta) d\lambda(\alpha)$$

### Theorem

Assume  $\lambda$  is a prior distribution and suppose that the Bayes estimator  $\delta_\lambda$  has a risk

$$(2) \quad \int R(\theta, \delta_\lambda) d\lambda(\theta) = \sup_{\delta} R(\theta, \delta_\lambda)$$

Then

- (i)  $\delta_\lambda$  is minimax
- (ii) If  $\delta_\lambda$  is the \* unique Bayes estimator, then it is also the unique minimax estimator.
- (iii)  $\lambda$  is least favorable.

### Proof.

Let  $\delta \neq \delta_\lambda$  be an arbitrary estimator.

Then

$$\sup_{\theta} R(\theta, \delta) \geq \int R(\theta, \delta) d\lambda(\theta)$$

$$\leq \int R(\theta, \delta) dR(\theta)$$

Since

$$\delta_\lambda \text{ is the } \rightarrow \geq \int R(\theta, \delta_\lambda) d\lambda(\theta)$$

Bayes estimator

$$\xrightarrow{\text{by (2)}} = \sup_{\theta} R(\theta, \delta_\lambda)$$

and there  $\delta_\lambda$  has a smaller maximum  
risk than  $\delta$ . Since  $\delta$  is arbitrary,  
 $\delta_\lambda$  is a minimax estimator.

(ii) If the Bayes estimator, the inequality  
is strict, then we have strict  
inequality in the LHD and RHS  
of (i) and therefore  $\delta_\lambda$  is  
the unique minimax estimator

⑤

(iii) Let  $\lambda' \neq \lambda$  be an arbitrary prior on  $\theta$ . Let  $\delta_\lambda, \delta_{\lambda'}$  be the corresponding Bayes estimators.

$$\underline{r_{\lambda'}(\delta_{\lambda'})} = \int R(\theta, \delta_{\lambda'}) d\lambda'(\theta)$$

$$\text{since } \delta_{\lambda'} \rightarrow \underbrace{\int R(\theta, \delta_\lambda) d\lambda'(\theta)}_{r_{\lambda'}(\delta_\lambda)} \leq \sup_{\theta} R(\theta, \delta_\lambda)$$

minimizes the Bayes risk  $r_{\lambda'}$

$$\rightarrow = \underline{r_\lambda(\delta_\lambda)}$$

by (2)

Therefore  $\lambda$  is least favorable.

#

⑥.

### Corollary

If the Bayes estimator has constant risk, i.e.  $R(\theta, \delta_\lambda)$  does not depend on  $\theta$ , then  $\delta_\lambda$  is also minimax.

### Proof

Then

$$\int R(\theta, \delta_\lambda) d\Lambda(\theta) = \sup_{\theta} R(\theta, \delta_\lambda)$$

#

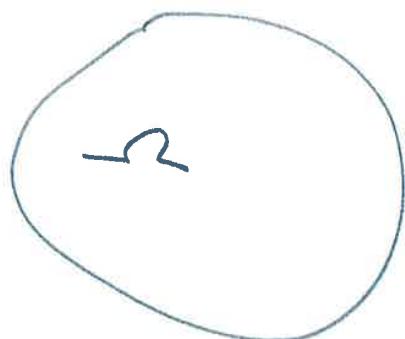
### Corollary

Let  $\delta_\lambda$  be the Bayes estimator for  $\lambda$ , and let

$$\mathcal{R}_\lambda = \left\{ \theta \in \Omega : R(\theta, \delta_\lambda) = \sup_{\theta'} R(\theta', \delta_\lambda) \right\}$$

and assume

$$\Lambda(\mathcal{R}_\lambda) = 1.$$



Then  $\delta_\lambda$  is minimax.

(7)

Proof

A Bayes estimator is only determined modulo  $\lambda$ -null sets.

$$\begin{aligned} & \int R(\theta, \delta) d\lambda(\theta) \\ &= \int_{\Omega_\lambda} R(\theta, \delta) d\lambda(\theta) + \int_{\Omega_\lambda^C} R(\theta, \delta) d\lambda(\theta) \\ &\quad \underbrace{\qquad\qquad}_{\#} = 0 \end{aligned}$$

Ex.

$X \in \text{Bin}(n, p)$ ,  $p$  is distributed according to the Beta distribution as  $\lambda = B(a, b)$ .

We have established that (with quadratic loss) the Bayes estimator of  $p$  is

$$\hat{\delta}_\lambda(x) = \frac{a+x}{a+b+n}$$

(8)

The risk function

$$R(p, \delta_{\lambda}) = \text{Var}(\delta_{\lambda}) + (\mathbb{E}(\delta_{\lambda}) - p)^2$$

$$\begin{aligned} R(p, \delta_{\lambda}) &= \frac{npq}{(a+b+n)^2} + \left( \frac{a+np - p(a+b+n)}{a+b+n} \right)^2 \\ &= \frac{1}{(a+b+n)^2} ( npq + (a+np - pa - pb - np)^2 ) \\ &= \frac{1}{(a+b+n)^2} ( npq + (ap - bp)^2 ) \\ &= \frac{a^2 + (n - 2a(a+b))p + ((a+b)^2 - n)p^2}{(a+b+n)^2} \end{aligned}$$

This is a polynomial in  $p$ , of degree 2,  
constant i.f.f.

$$(a+b)^2 = n$$

$$2a(a+b) = n$$

which has solution  $a=b=\frac{\sqrt{n}}{2}$ .

(9)

Thus, with  $a=b=\sqrt{n}/2$  as the parameter in the prior distribution, the Bayes estimator is

$$\hat{\lambda}(x) = \frac{x + \sqrt{n}/2}{\sqrt{n} + n}$$

and this has constant risk and therefore it is minimax. This was the Bayes estimator for  $\lambda = B(a,b)$  with  $a=b=\sqrt{n}/2$ .

Now let  $\lambda$  be an arbitrary distribution on  $[0,1]$ . The Bayes estimator for this prior is

$$\begin{aligned}\hat{\lambda}(x) &= E(\lambda|x) = \int_0^1 p d\lambda(p|x) \\ &= \int_0^1 p f(x|p) \frac{1}{f(x)} d\lambda(p) \\ &= \frac{\int p p^x (1-p)^{n-x} d\lambda(p)}{\int p^x (1-p)^{n-x} d\lambda(p)} \quad (3)\end{aligned}$$

(10)

The power expansion

$$(1-p)^{n-x} = 1 + a_1 p + \dots + a_{n-x} p^{n-x}$$

and plug-in into (3) to obtain

$$\delta_\lambda(x) = \frac{\int_0^1 p^{x+1} + a_1 p^{x+2} + \dots + a_{n-x} p^{n+1} d\lambda(p)}{\int_0^1 p^x + a_1 p^{x+1} + \dots + a_{n-1} p^n d\lambda(p)}$$

Now for instance

$$\int_0^1 p^{x+1} d\lambda(p) = E_\lambda(p^{x+1})$$

$$\int_0^1 p^{n+1} d\lambda(p) = E_\lambda(p^{n+1})$$

The Bayes estimator depends on the prior distribution  $\lambda$  only via the n+1 first moments in that distribution.

Therefore two prior with the same  
first u+ moments will give the  
same Bayes estimator.

#

(2)