

Bayesian inference and minimax (ctd) Lecture 16

Sometimes one would like to have an improper prior.
For instance, suppose for $X/\theta \sim$

$$N(\mu, \sigma^2)$$

with σ^2 known, so that $\Theta = \mu$ and then we would like to have $\Omega = \mathbb{R}$ (① is should be a r.v. drawn from a d.f. λ).

One would maybe like to have (Lebesgue) measure as prior, but
 $\lambda(\mathbb{R}) = \infty$.
ordinary length measure

We don't know, so far, how to do this.

We let $\{\lambda_n\}_{n \geq 1}$ be a sequence of prior d.f.'s, that in some sense approximates the desired λ .

Def. Let $\{\lambda_n\}_{n \geq 1}$ be a sequence of priors, and let δ_n be Bayes estimator corresponding to λ_n .

①

Let

$$r_n = \int R(\theta, \delta_n) d\Lambda_n(\theta)$$

be Bayes risk for δ_n . Define

$$r = \lim_{n \rightarrow \infty} r_n$$

If for every proper prior distribution λ' we have

$$r_{\lambda'} \leq r$$

then the sequence $\{\Lambda_n\}_{n \geq 1}$ is called least favorable

#

Recall:

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

is the Bayes risk of the Bayes estimator $\delta_{\lambda'}$.

(2)

Theorem

Let $\{\Lambda_n\}$ be a sequence of priors,
such

$$r = \lim_{n \rightarrow \infty} r_n \quad (1)$$

is the limit of their Bayes risks

$$r_n = \int R(\theta, \delta_{\Lambda_n}) d\Lambda_n(\theta).$$

Assume the estimator δ satisfies

$$\sup_{\theta} R(\theta, \delta) = r \quad (2)$$

Then δ is minimax and $\{\Lambda_n\}$
is least favorable.

Proof.

(minimax)
estimator. Then

Take δ' any other

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

$$\rightarrow \geq \int R(\theta, \delta_{\Lambda_n}) d\Lambda_n(\theta) = r_n$$

③

Since δ_{Λ_n} is the Bayes est. corresp. to Λ_n

Then (1) and (2) imply

$$\sup_{\delta} R(\theta, \delta') \geq \sup_{\delta} R(\theta, \delta)$$

and therefore δ is minimax.

(ν_n) least favorable

* Λ any proper prior distribution,
and let δ_λ be corresponding Bayes estimator.

Then

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

$$\leq \int R(\theta, \delta) d\Lambda(\theta)$$

$$\leq \sup_{\delta} R(\theta, \delta) = \lim_{n \rightarrow \infty} r_n$$

#

①

Note that if δ_n are unique Bayes estimators (for each μ , δ_n is the unique Bayes estimator corresponding to Λ_n) then we have inequality

$$\sup_{\delta} R(\theta, \delta) \geq \int R(\theta, \delta') d\Lambda_n(\theta) > \underbrace{\int R(\theta, \delta_{\Lambda_n}) d\Lambda_n(\theta)}_{r_n}$$

$$\sup_{\delta} R(\theta, \delta) > r_n$$

which could be transformed to

$$\sup_{\delta} R(\theta, \delta) \geq r$$

under limit operation, therefore

uniqueness of the Bayes estimator does not imply uniqueness of the minimax estimator.

(5)

In order to calculate the Bayes risk we may use:

Lemma

Let δ_λ be the Bayes estimator of $f(\theta)$ corresponding to λ , and assume we have quadratic loss. The Bayes risk is given by

$$r_\lambda = \int \text{Var}(g(\theta) | \bar{X} = x) dF_{\bar{X}}(x)$$

Proof.

We know that the Bayes estimator for quadratic risk (loss)

$$\delta_\lambda(x) = E(g(\theta) | \bar{X} = x). \quad (4)$$

Its Bayes risk is

$$\begin{aligned} r_\lambda &= \int R(\theta, \delta_\lambda) d\lambda(\theta) \\ &= \int E(L(\theta, \delta_\lambda(\bar{X})) | \theta = \theta) d\lambda(\theta) \\ &= E(L(\theta, \delta_\lambda(\bar{X}))) \\ &= E(E(L(\theta, \delta_\lambda(\bar{X})) | \theta)) \\ &= E(E(L(\theta, \delta_\lambda(\bar{X})) | \bar{X})) \end{aligned}$$

6

$$= \int E(L(\mathbb{M}), \delta_\lambda(\bar{X}) | \bar{X}=x) dF_{\bar{X}}(x)$$

$$= \int E((g(\mathbb{M}) - \delta_\lambda(\bar{X}))^2 | \bar{X}=x) dF_{\bar{X}}(x)$$

$$\rightarrow = \int E\left(\left[g(\mathbb{M}) - E(g(\mathbb{M}) | \bar{X}=x)\right]^2 | \bar{X}=x\right) dF_{\bar{X}}(x)$$

by def

$$\sigma_\lambda^2(x) = \int \text{Var}(g(\mathbb{M}) | \bar{X}=x) dF_{\bar{X}}(x)$$

(4)

#

Note that if $\text{Var}(g(\mathbb{M}) | \bar{X}=x)$ is not a function of x , then

$$r_\lambda = \text{Var}(g(\mathbb{M}) | \bar{X}=x)$$

Ex (pre-example)

x_1, \dots, x_n i.i.d. $N(\theta, \sigma^2)$, σ^2 known,
 θ unknown. (a) r.v. prior distribution

$$(a) \sim N(\mu, b^2)$$

hyper-parameters

(7)

joint distribution of x_1, \dots, x_n and θ

$$f(x, \theta) \propto \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2 - \frac{1}{2b^2} (\theta - \mu)^2 \right]$$

The marginal density for x is

$$f(x) = \int_{-\infty}^{\infty} f(x, \theta) d\theta$$

so that

$$f(\theta|x) = C(x) \cdot$$

$$\cdot \exp \left(-\frac{1}{2} \left(\theta^2 \left\{ \frac{n}{\sigma^2} + \frac{1}{b^2} \right\} + \theta \left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu}{b^2} \right) \right) \right)$$

(note that $C(x)$ has absorbed the factor $e^{-\frac{1}{2\sigma^2} \sum x_i^2}$)

$$= C(x) \exp \left[-\frac{1}{2} \left(\frac{n}{\sigma^2} + \frac{1}{b^2} \right) (\theta^2 - 2\theta \cdot \left(\frac{n\bar{x}/\sigma^2 + \mu/b^2}{n/\sigma^2 + 1/b^2} \right)) \right]$$

(8)

This is a Gaussian distribution with

$$E(\hat{\theta}|x) = \frac{n\bar{x}/\sigma^2 + \mu/b^2}{n/\sigma^2 + 1/b^2}$$

$$\text{Var}(\hat{\theta}|x) = \frac{1}{n/\sigma^2 + 1/b^2}$$

Assumed we have squared error loss,
and $g(\theta) = \theta$ estimator then

(i) Before any data, the estimator from the prior distribution is

$$E(\hat{\theta}) = \mu.$$

(ii) Model, frequentist approach (no prior)

$$\hat{\theta}(x) = \bar{x}$$

(based on x_1, \dots, x_n i.i.d. of $N(\theta, \sigma^2)$)

(iii) Bayes estimator

$$\hat{\theta}_1(x) = E(\hat{\theta} | \mathcal{X} = x)$$

(9)

$$= \frac{n\bar{x}/\sigma^2 + \mu/b^2}{n/\sigma^2 + 1/b^2}$$

$$= \bar{x} \left(\frac{n/\sigma^2}{n/\sigma^2 + 1/b^2} \right) + \mu \left(\frac{1/b^2}{n/\sigma^2 + 1/b^2} \right)$$

i.e. a convex combination of the estimator based on ~~only~~ prior belief and estimator based on ~~only~~ data

We see

$$n \rightarrow \infty \Rightarrow \delta_n(x) \rightarrow \bar{x}$$

$$b \rightarrow 0 \quad (\text{Var}(\theta) \rightarrow 0) \Rightarrow \delta_n \rightarrow \mu$$

$$b \rightarrow \infty \quad (\text{Var}(\theta) \rightarrow \infty)$$

$$\Rightarrow \delta_n(x) \rightarrow \bar{x}$$

#

Ex.

Let (\bar{X}_1, \bar{X}_n) be an i.i.d. sequence with $X_i \in N(\theta, \sigma^2)$. Want to estimate $g(\theta) \equiv \theta$, assume quadratic loss. Assume σ^2 known. Want to show that \bar{X} is minimax!

The var of \bar{X} is σ^2/n . We want to find a sequence of Bayes estimators s.t. $r_m \rightarrow \sigma^2/n$; $m \rightarrow \infty$. and noting that

$$\sup_{\theta} R(\theta, \bar{X}) = \sup_{\theta} \frac{\sigma^2}{n} = \frac{\sigma^2}{n}$$

this will be enough.

Take prior distribution $N(\mu, b^2) = \Lambda$ for θ . We know that the Bayes estimator of θ is

$$\hat{\theta}_{\Lambda}(x) = \frac{n\bar{x}/\sigma^2 + \mu/b^2}{n/\sigma^2 + 1/b^2}$$

and

(11)

and posterior variance

$$\begin{aligned} & \text{Var}(\theta | \bar{X} = x) \\ &= \frac{1}{n/\sigma^2 + 1/b^2} \end{aligned}$$

and this does not depend on x , and is therefore the Bayes risk for δ_{λ}

$$r_{\lambda} = \frac{1}{n/\sigma^2 + 1/b^2}$$

Now let $b \rightarrow \infty$, then

$$r_{\lambda_b} \rightarrow \sigma^2/n = \sup_{\theta} R(\theta, \bar{X})$$

Therefore $\delta(\bar{X}) = \bar{X}$ is minimax.

□