705 09/06/2020

Youtube 03/10/16

Lecture Notes 8 (Minimax Theory) cont.

$R(\theta, \hat{\theta}) = \mathbb{E}_\theta[L(\theta, \hat{\theta})]$ - "risk as exp. of loss"

$R_n = \inf_{\hat{\theta}} \sup_{\theta} R(\theta, \hat{\theta})$ - "minimax risk"

$B_\pi(\hat{\theta}) = \int R(\theta, \hat{\theta}) \pi(\theta) d\theta$ - "Bayes risk"

(*) Bayes risk - integral of risk wrt prior; minimiser of Bayes risk is Bayes estimator

(*) maximum risk; minimiser of maximum risk is minimax estimator; alternatively, minimax estimator achieves minimax risk.

LW: Finding minimiser of Bayes risk is 'easy'; minimise of maximum risk is more difficult.

(*) Can use minimisation of Bayes risk as a route to getting minimax estimator

LW: Derive a different expression for Bayes risk

1.3 Bayes estimator

- π prior distri.

Recall θ is an r.v. under this paradigm

- After observing $X^n = (X_1, \dots, X_n)$

- Posterior distn: $L(\theta) = p(x^n; \theta)$ (likelihood fn.)

$$P(\theta \in A | X^n) = P(\theta | X_1, \dots, X_n) = \frac{\int_A p(X_1, \dots, X_n | \theta) \pi(\theta) d\theta}{\int_\Theta p(X_1, \dots, X_n | \theta) \pi(\theta) d\theta}$$

$$= \frac{\int_A L(\theta) \pi(\theta) d\theta}{\int_\Theta L(\theta) \pi(\theta) d\theta}$$

Define the posterior risk of an estimator $\hat{\theta}(x^n)$ by:-

$$r(\hat{\theta}|x_1,\dots,x_n) = \int L(\theta, \hat{\theta}(x_1,\dots,x_n))\, p(\theta|x_1,\dots,x_n)\, d\theta$$

(*) Alternative way of treating loss function in the context of expectation
(*) conventionally, risk is defined as expectation of loss with respect to the sampling distribution $E_\theta[L(\theta,\hat{\theta})]$
(*) Instead, here; we use Bayes theorem to get a posterior distribution; and also density for θ
(*) Take loss function $L(\theta,\hat{\theta})$ which we don't know due to presence of unknown θ.
(*) But we have a posterior density/distri on $\theta \longrightarrow$ integrate out θ.
(*) Posterior risk is integral of loss with respect to a posterior density/distri?
(R) - review.

(*) Marginal distri of X^n: (recall Bayesian para. now)

- $m(x^n) = m(x_1,\dots,x_n) = \int p(x^n|\theta)\pi(\theta)\, d\theta$

Theorem 6

(as an useful alternative to original defin.)

- The Bayes risk $B_\pi(\hat{\theta})$ satisfies/can be written as:-

$$B_\pi(\hat{\theta}) = \int \underbrace{r(\hat{\theta}|x^n)}_{(i)} m(x^n)\, dx^n = \int\cdots\int r(\hat{\theta}|x_1,\dots,x_n)\, m(x_1,\dots,x_n)\, dx_1 \dots dx_n$$

- Let $\hat{\theta}(x^n)$ be the value of θ that minimises $r(\hat{\theta}|x^n)$ (posterior risk)

(*) Then $\hat{\theta}$ is the Bayes estimator
$\hat{\theta}(x_1,\dots,x_n)$

IW:
- minimise (i) for each x^i, then entire integral is minimised.
- Note that all terms are non-negative
- minimising each element of an integral (aka summation) $\longrightarrow$ minimise the integral.

(*) In many cases, posterior risk is easy to minimise

(*) Definition of Bayes risk (original) is used for theoretical reasons; but for purposes of calculation/computation, we use the specification of Bayes risk in terms of the posterior i.e. Theorem 6.

(*) Proof of T.6 → motivation:-

- Note that the term within $\int R(\theta,\hat{\theta})\pi(\theta)\,d\theta$, $R(\theta,\hat{\theta})$ is an expected value under a sampling distribution (i.e. x^n "given" θ).
- Note that there is also a prior $\pi(\theta)$ within the integral. (distri of θ).
- Essence is just switching integral order, manip. probabilities.

Theorem 6 (Proof)

- Let $p(x,\theta) = p(x|\theta)\pi(\theta)$ be the joint density of X and θ. (i)
- The Bayes risk:-

$$B_\pi(\hat{\theta}) = \int \underbrace{R(\theta,\hat{\theta})}\pi(\theta)\,d\theta = \int\left(\int L(\theta,\hat{\theta}(x^n))p(x|\theta)\,dx^n\right)\pi(\theta)\,d\theta$$

-subs $R(\theta,\hat{\theta})$

$$= \iint L(\theta,\hat{\theta}(x^n))\underbrace{p(x,\theta)}_{\pi(\theta|x^n)}\,dx^n\,d\theta$$

-using above (i)

$$= \iint L(\theta,\hat{\theta}(x^n))p(\theta|x^n)m(x^n)\,dx^n\,d\theta$$

-decomposing joint differently

$$= \int\left(\int L(\theta,\hat{\theta}(x^n))\pi(\theta|x^n)\,d\theta\right)m(x^n)\,dx^n$$

$$= \int r(\hat{\theta}|x^n)m(x^n)\,dx^n$$

- Bayes risk can be viewed as expect. of usual risk $R(\theta,\hat{\theta})$ wrt prior distri

OR expect. of posterior risk wrt marginal

(*) To find the Bayes estimator, it is sufficient to minimise posterior risk for each x^n

(*) i.e. selecting $\hat{\theta}(x^n)$ to be value that minimises $r(\hat{\theta}|x^n)$ then we will minimise the integrand at every x and thus minimise

$$\int r(\hat{\theta}|x^n)\, m(x^n)\, dx^n$$

(*) use this to find Bayes est. for specific loss fns.

Theorem 7

- For $L(\theta,\hat{\theta}) = (\theta-\hat{\theta})^2$ — note $\hat{\theta} = \hat{\theta}(x_1,\dots,x_n)$

- posterior risk:

$$r(\hat{\theta}|x_1,\dots,x_n) = \int (\theta-\hat{\theta})^2 p(\theta|x_1,\dots,x_n) = \int (\theta-\hat{\theta}(x^n))^2 p(\theta|x^n)\, d\theta$$

- Bayes estimator $\hat{\theta}(x^n)$ minimises $r(\hat{\theta}|x^n)$

- Hence;

$$\frac{dr(\hat{\theta}|x^n)}{d\hat{\theta}} = 2\int (\hat{\theta}(x^n)-\theta)\pi(\theta|x^n)\, d\theta = 0$$

$$\Rightarrow \int (\hat{\theta}(x^n)-\theta)\pi(\theta|x^n)\, d\theta = 0$$

- split integral

$$\Rightarrow \int \hat{\theta}(x^n)\pi(\theta|x^n)\, d\theta - \int \theta\pi(\theta|x^n)\, d\theta = 0$$

$$\int \pi(\theta|x^n)\, d\theta = 1$$

$$\Rightarrow \underset{=\hat{\theta}(x_1,\dots,x_n)}{\hat{\theta}} = \int \theta\pi(\theta|x^n)\, d\theta = \mathbb{E}(\theta|X^n=x^n)$$

→ mean of θ wrt posterior density

(*) LW: For L_2 loss; find Bayes est. expl. by multiplying likelihood and prior to get post. density/distri

- Find mean of that distri and you have Bayes estimator wrt L_2 loss. (i.e. minimise of posterior risk)

(*) By 1.6 ; this also minimises the original definition of Bayes risk

∴ Bayes est. w.r.t L_2 loss

(*) mean of the posterior is the Bayes estimator for L_2 loss

---

(*) lecture terminates prematurely after 19 mins

- 30+ mins missing
- Go through it yourself