Ⅲ

Bayesian Estimation

1. The Bayesian approach to statistics

Suppose parameter θ has a prior distribution, $\tau(\theta)$,

and we know the family of distribution $p(\underline{x}|\theta)$.
Sampling

Then r.v. $(\underline{X}, \Theta)$ is jointly distributed as:

$$(\underline{X}, \Theta) \sim p(\underline{x}|\theta) \cdot \tau(\theta)$$

And the posterior distribution of θ given the sample $\underline{x}$ is: $p(\theta|\underline{x}) = \dfrac{p(\underline{x}|\theta)\tau(\theta)}{m(\underline{x})}$

with $m(\underline{x}) = \int p(\underline{x}|\theta) \cdot \tau(\theta) d$

2. Bayes estimators

Design a MMSE estimator $T(\underline{X})$ of θ.

$$\because MSE \equiv \mathbb{E}[T(\underline{X}) - \theta]^2$$

$$= \iint (T(\underline{x}) - \theta)^2 p(\underline{x}|\theta)\tau(\theta)\, d\underline{x}\, d\theta$$

$$= \iint (T(\underline{x}) - \theta)^2 p(\theta|\underline{x})\tau(\underline{x})\, d\underline{x}\, d\theta$$

where $\tau(\underline{x})$ is

$$= \int \left[\int (T(\underline{x}) - \theta)^2 p(\theta|\underline{x})\, d\theta \right] \tau(\underline{x})\, d\underline{x}$$

let $t = T(\underline{x})$, then $\arg\min_t \int (t-\theta)^2 p(\theta|\underline{x})\, d\theta = \arg\min_t \mathbb{E}[(t-\Theta)^2|\underline{x}] = \arg\min_t \{Var(\theta|\underline{x}) + [\mathbb{E}(\Theta|x) - t]^2\}$

we get $t = \mathbb{E}(\theta|\underline{x})$, is the posterior mean.

$(T(\underline{X}) = \mathbb{E}(\theta|\underline{X}))$

Def: $\mathcal{F}$ is the class of sampling distributions $f(x|\theta)$,

Π is a class of prior distributions.

Π is called a <u>conjugate family</u> for $\mathcal{F}$, if the posterior distribution $f(\theta|x) \in \Pi$, $\forall f(x|\theta) \in \mathcal{F}$, $\pi(\theta) \in \Pi$, $x \in \mathcal{U}$.

E.g.: 1° The beta family ($Beta(\alpha,\beta)$) is conjugate for the binomial family ($Binomial(n,p)$). (for p).

2° The normal family ($N(\mu,\sigma^2)$) is its own conjugate family. (for μ)

Note: 1° Without any observation, our best estimate of θ is the prior mean, $E\theta$.

2° Ignoring the prior information, we might use the sample mean, $\overline{X}$ as the estimator for θ. (Say when $EX = \theta$).

3° The Bayes estimator, i.e. the posterior mean $E(\theta|\mathbf{X})$, is in some cases a linear combination of the <u>prior mean</u> and the <u>sample mean</u>. | As the prior information becomes more vague, the Bayesian estimator tends to give more weight to the sample information; if the prior information is good, then more weight is given to the prior mean.