Stat 543 2-14-05

Bayes Estimators – There is a "standard" Bayes way of cooking up estimators – This has both a "philosophical" justification and decision-theoretic justification

"Philosophical" motivation:

likelihood $L(\theta)$
prior $g(\theta)$

and "posterior" (conditional dsn of θ given x)

$$\propto L(\theta)\, g(\theta)$$

and only this is relevant to inference

a Bayesian takes some measure of "central tendency" of the posterior as an estimate of θ

<u>Example</u> $X \sim N(\theta, 1)$ prior $\theta \sim N(0, \gamma^2)$

$$L(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}} \qquad g(\theta) = \frac{1}{\sqrt{2\pi\gamma^2}} e^{-\frac{\theta^2}{2\gamma^2}}$$

$$L(\theta) g(\theta) \propto \underbrace{e^{-\frac{1}{2}\left[\left(1+\frac{1}{\gamma^2}\right)\theta^2 - 2x\theta\right]}}_{e^{-\frac{1}{\frac{1}{1+\frac{1}{\gamma^2}}}\frac{1}{2}\left[\theta^2 - 2\frac{x}{1+\frac{1}{\gamma^2}}\theta\right]}}$$

i.e. the posterior dsn of $\theta \mid X = x$ is

$$N\left(\frac{\gamma^2}{1+\gamma^2}x, \frac{\gamma^2}{1+\gamma^2}\right)$$

The mean (median, mode) of this dsn is

$$\frac{\gamma^2}{1+\gamma^2}x$$

<u>so a</u> Bayes estimator of θ is $\delta(X) = \frac{\gamma^2}{1+\gamma^2}X$

<u>Example</u> $X \sim$ Poisson θ, prior $\theta \sim Exp(\beta)$

$$L(\theta) = \frac{e^{-\theta}\theta^x}{x!} \qquad g(\theta) = \frac{1}{\beta}e^{-\theta/\beta}$$

$L(\theta)\, g(\theta) \propto$ posterior

$$L(\theta)\, g(\theta) \propto \underbrace{\theta^{x} e^{-\theta(1+\frac{1}{\beta})}}$$

shape

B+D notation (Γ dsn with parameters $p = x+1$ and $\lambda = (1+\frac{1}{\beta})$

scale

These two examples have used "special" priors that fit together nicely with the likelihood to give convenient forms for the posterior — the jargon is "conjugate priors" —

Notice that the prescription posterior $\propto L(\theta)\, g(\theta)$

doesn't even require that $g(\theta)$ actually specify a <u>probability</u> dsn – e.g. if I'm willing to think of θ as "uniformly distributed on $\mathbb{R}^1$" I might also be willing to write

$$L(\theta)\, \underset{\substack{\uparrow \\ 1}}{g(\theta)} = L(\theta)$$

<u>Example</u> $X \sim N(\theta, 1)$ $\qquad$ "$\theta \sim U$ on $\mathbb{R}^1$"

$$L(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}} \qquad g(\theta) = c$$

$$L(\theta)\, g(\theta) \propto \underline{\underline{e^{-\frac{1}{2}(x-\theta)^2}}}$$

formally this specifies a posterior for θ that is $N(x, 1)$ and thus a "Bayes" estimator for θ that is

$$\delta(X) = X$$

Bayesians would call this formalism use of an "<u>improper prior</u>"

dsn with infinite total mass, not with total mass = 1

A decision-theoretic argument for the posterior mean as an estimator of a real-valued $\gamma(\theta)$

$$X \sim f(x|\theta) \qquad \theta \sim g(\theta) \quad \Longrightarrow \quad \text{joint dsn} \quad f(x,\theta) = f(x|\theta)\, g(\theta)$$

$$\Downarrow$$

$$g(\theta|x) \qquad \text{conditional pmf or pdf of } \theta \mid X = x$$

$$f(x)$$

$$L(\theta, a) \quad , \quad \delta(x)$$

$$R(\theta, \delta) = \int L(\theta, \delta(x))\, f(x|\theta)\, dx$$

$$R(G,\delta) = \int \left[\int L(\theta,\delta(x))\, f(x|\theta)\, dx \right] g(\theta)\, d\theta$$

$$= \iint L(\theta,\delta(x))\, \underline{f(x,\theta)}\, dx\, d\theta$$

write the "other" way

$$= \int \left[\int L(\theta,\delta(x))\, g(\theta|x)\, d\theta \right] f(x)\, dx$$

To minimize $R(G,\delta)$ it suffices to choose $\delta(x)$ to minimize $[\qquad\qquad]$ one x at a time — i.e. a Bayes decision rule for each x seeks an action, a,

that minimizes conditional expected loss given $X = x$

So for $L(\theta, a) = (\gamma(\theta) - a)^2$ this is

$$\int \gamma(\theta)\, g(\theta \mid x)\, d\theta$$