Note Title

Stat 543 1-14-05

1/14/2005

Comment regarding Bayes paradigm: In general,

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

doesn't come out "cleanly" (i.e. the posterior isn't immediately recognizeable as being of some standard form) so that numerical analysis and/or simulation is needed to discern the properties of the posterior ... especially when θ is multidimensional ... MCMC

Statistical Decision Theory –

if statistical theory is the mathematics of "rational" data analysis, this material for bringing notions of optimality to bear

The basic elements of this theory are:

θ "parameter" belonging to Θ parameter space

a "action" belonging to $\mathcal{A}$ action space

$L(\theta, a) \geq 0$ "loss/cost" suffered if the parameter is θ and I take action a

$X \sim f(x|\theta)$ a random observable whose dsn depends on the parameter

$\delta(x)$ a decision rule (function from observation space to $\mathcal{A}$) – $\delta(x)$ is the decision/action taken on the basis of observing $X = x$

$L(\theta, \delta(X))$ (random) loss suffered using the decision rule δ

$E_\theta L(\theta, \delta(X)) = R(\theta, \delta)$ the "risk" associated with decision rule δ . . . (a function of θ)

For G some prior dsn on Θ

$$R(G, \delta) = \int R(\theta, \delta)\, g(\theta)\, d\theta \quad \left(\text{cont}^{\text{s}}\ G \text{ case}\right)$$

is the so called "Bayes risk" of δ against the prior dsn G

<u>Example</u> $\theta \in \Theta = \mathbb{R}$

$a \in \mathcal{A} = \mathbb{R}$

$L(\theta, a) = (\theta - a)^2$ "squared error loss"

suppose, e.g., $X \sim N(\theta, 1)$

$\delta(x)$ is just a function from $\mathbb{R}$ to $\mathbb{R}$

$$E_\theta L(\theta, \delta(X)) = \underbrace{E_\theta(\theta - \delta(X))^2}_{\text{The mean squared error of } \delta(X) \text{ as an approximator of } \theta}$$

$$= \underbrace{(\theta - E_\theta \delta(X))^2}_{\text{bias}^2} + \mathrm{Var}_\theta\, \delta(X)$$

e.g. $\delta_1(X) = X$ has risk function

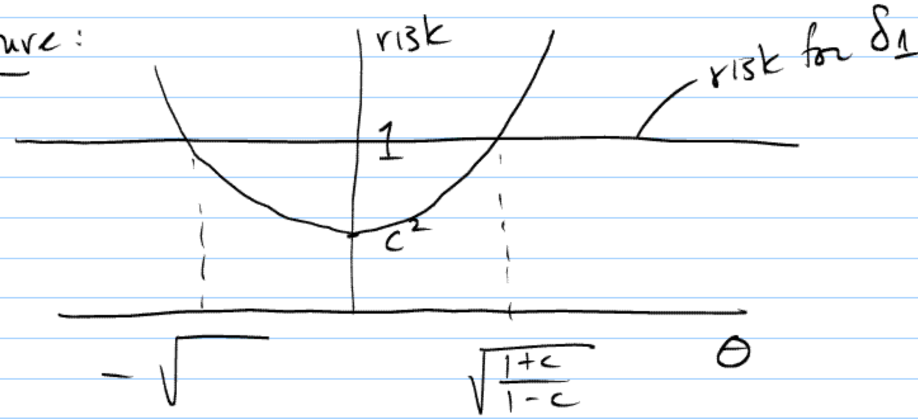
$$E_\theta(\theta - X)^2 = \mathrm{Var}_\theta X = 1 \quad \text{(a constant)}$$

For $c \in (0,1)$ define $\delta_c(X) = cX$

This has risk function

$$E_\theta(\theta - cX)^2 = (1-c)^2\theta^2 + c^2$$

Picture:



Usually comparison of decision rules amounts to comparison of functions (non-constant) of Θ ... ???? There are various principles that have been put forth for doing such comparison

minimax

Bayes > this averaging reduces comparison of 2 functions to comparison of 2 numbers

How to find decision rules that are "best"/"optimal" according to one of theses is the subject of whole courses

A second simple example is

<u>Example</u> $\Theta = \mathbb{R}$

$\mathcal{A} = \{0,1\}$

$$L(\theta,a) = \underbrace{I[a=0]\,I[\theta>17] + I[a=1]\,I[\theta\le 17]}$$

this is a "0-1" loss function appropriate if I'm trying to decide whether or not $\theta \le 17$

if, e.g., $X \sim N(\theta,1)$

δ maps $\mathbb{R} \to \{0,1\}$

I might consider decision criteria

$\delta_c(x) = I[x > c]$

$R(\theta, \delta_c) = E_\theta L(\theta, \delta_c(X))$

$\vdots$

$= I[\theta > 17]\,\Phi(c-\theta)$

$\quad + I[\theta \le 17]\,(1 - \Phi(c-\theta))$

Cartoon