Minimax and Bayes Rules

In order to find minimax estimators, one general result is based on finding Bayes estimators with good properties.

The idea: Assume we want to find

$$\max_s h(t,s)$$

where h is a continuous function. Assume that for some measure we know that for

$$\max_s h(t_0,s) \le \int h(t_0,u) f(u)\,du$$

for some density function $f(u)$ and t_0 is a minimizer of

$$t \to \int h(t,u) f(u)\,du$$

then t_0 is also a minimizer of $\max_s h(t,s)$.

Indeed, we have

$$\int h(t,u) f(u)\,du \le \max_u h(t,u)$$

and thus for $t = t_0$ we get that

$$\max h(t_0,s) \le \int h(t_0,u) f(u)\,du \le \int h(t,u) f(u)\,du \le$$
$$\le \max_u h(t,u)$$

therefore t_0 is a minimizer of $\max_s h(t,s)$.

In fact we also have from this that t_0 is a minimax for h.

Theorem: Assume

$$r(f, \hat{\theta}^f) = \inf_{\hat{\theta}} r(f, \hat{\theta})$$

where $r(f, \hat{\theta}) = \int R(\hat{\theta}, \theta) f(\theta) d\theta$

If $R(\theta, \hat{\theta}^f) \le r(f, \hat{\theta}^f)$ for all θ, then $\hat{\theta}^f$ is the minimax.

Pf As before we have that if $\hat{\theta}^f$ is not, then there exists another $\hat{\theta}$ such that

$$r(f, \hat{\theta}) \le \max_{\theta} R(\theta, \hat{\theta}) < \max_{\theta} R(\theta, \hat{\theta}^f) \le r(f, \hat{\theta}^f)$$

which leads to a contradiction.

Remark Notice here that $R(\theta, \hat{\theta}^f) \le r(f, \hat{\theta}^f)$ is actually equivalent to $R(\theta, \hat{\theta}^f) = r(f, \hat{\theta}^f)$ because we always have the reverse inequality

Corollary Assume $\hat{\theta}$ is the Bayes rule with respect to some prior f. If $R(\hat{\theta}, \theta)$ is constant w.r.t θ, then $\hat{\theta}$ is a minimax.

Pf Since $R(\hat{\theta}, \theta)$ is constant, $R(\hat{\theta}, \theta) = r(f, \hat{\theta})$ and we can apply the Theorem

Ex For the Bernoulli model,

$$\hat{p}(x^n) = \frac{\sum X_i + \alpha}{\alpha + \beta + n}$$ if we take the prior $B(\alpha, \beta)$. Thus we have that

$$R(p; \hat{p}) = \mathbb{E}[(p - \hat{p})]^2 = \frac{\alpha^2 + p(n - 2\alpha(\alpha+\beta)) + p^2((\alpha+\beta)^2 - n)}{(\alpha+\beta+n)^2}$$

If we make this constant we actually get minimax estimators.

Ex If we change the loss function

$$L(p,\hat{p}) = \frac{(p-\hat{p})^2}{p(1-p)}$$

and let $\hat{p} = \bar{x}$

The risk in this case is $R(p,\hat{p}) = E\left[\frac{(p-\hat{p})^2}{p(1-p)}\right] = \frac{1}{n}$

thus the risk is constant. Now for this loss function and uniform prior,

$$r(\hat{p}|x) = \int_0^1 \frac{(p-\hat{p})^2}{p(1-p)} f(p|x)\,dp$$

$$= c\int_0^1 \frac{(p-\hat{p})^2}{p(1-p)} p^s (1-p)^{n-s} dp$$

$$= c\int_0^1 (p-\hat{p})^2 p^{s-1} (1-p)^{n-s-1} dp$$

The minimizer is actually given by

$$\hat{p} \int p^{s-1} (1-p)^{n-s-1} dp = \int p^s (1-p)^{n-s-1} dp$$

$$\hat{p} \frac{\cancel{\Gamma(s)\Gamma(n-s)}}{\Gamma(\cancel{n})} = \frac{\cancel{\Gamma(s+1)\Gamma(n-s)}}{\cancel{\Gamma(n+1)}\, n}$$

so $\hat{p} = \frac{s}{n} = \bar{x}$. Thus $\hat{p}$ is still a Bayes estimator and

Th Let $X_1, \ldots, X_n \sim N(\theta, 1)$ and $\hat{\theta} = \bar{X}$. Then $\hat{\theta}$ is the minimax with respect to convex loss functions.

We will see this result w.r.t the square loss function later on.

Ex If we restrict our attention to $\theta \in [-1, 1]$, then the unique minimax estimator under the square loss is given by

$$\hat{\theta}(x) = \tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}.$$

To do this we look for a Bayes rule with a given prior such that the risk function is constant. Let $f(\theta)$ be the prior with this property. Taking $f(\theta) = \frac{1}{2}(\delta_a + \delta_{-a})$, then

$$\hat{\theta}(x) = E[\theta \mid X = x] = \int \theta f(\theta \mid x)\, d\theta$$

$$= \int \theta \frac{f(x \mid \theta) f(\theta)}{m(x)}\, d\theta$$

where $m(x) = \int f(x \mid \theta) f(\theta)\, d\theta = \frac{1}{2}(f(x \mid 1) f(1)) + \frac{1}{2} f(x \mid -1) f(-1)$

$$= \frac{1}{2}\left(\frac{1}{\sqrt{2\pi}} e^{-(x-a)^2/2} + \frac{1}{\sqrt{2\pi}} e^{-(x+a)^2/2}\right)$$

and then

$$\hat{\theta}(x) = \frac{a e^{-(x-a)^2/2} - a e^{-(x+a)^2/2}}{e^{-(x-a)^2/2} + e^{-(x+a)^2/2}}$$

$$= \frac{a\left(e^{ax} - e^{-ax}\right)}{e^{ax} + e^{-ax}}.$$

Now $r(f,\hat{\theta}) = \int (\hat{\theta}(x)-\theta)^2 f(x|\theta) f(\theta)\, dx\, d\theta$

$$= \int (\hat{\theta}(x)-a)^2 f(x|a) f(a)\, dx + \int (\hat{\theta}(x)+a)^2 f(x|-a) f(-a)\, dx$$

$$= \frac{a^2}{2}\int (\tanh(ax)+1)^2 e^{-(x-a)^2/2}\, dx + \frac{a^2}{2}\int (\tanh(ax)-1)^2 e^{-(x+a)^2/2}\, dx$$

$x \to -x$

$$= a^2 \int (\tanh(ax)-1)^2 e^{-(x+a)^2/2}\, dx$$

$x \to -x$

$$= a^2 \int (\tanh(a(y-a))-1)^2 e^{-y^2/2}\, dy$$

Now $R(\theta,\hat{\theta}) = \int (\hat{\theta}(x)-\theta)^2 f(x|\theta)\, d\theta$

$$= \int (a\tanh(ax)-\theta)^2 e^{-(x-\theta)^2/2}\, dx$$

$$= \int (a\tanh(a(y-\theta))-\theta)^2 e^{-y^2/2}\, dy$$

Now $\theta \to (a\tanh(a(y-\theta))-\theta)^2$ is a convex function of θ, thus it attains its maximum on the boundary of Θ. Therefore

$$(\tanh(a(y-\theta))-\theta)^2 \le (\tanh(a(y-1))-1)^2$$

which is the risk $r(f,\hat{\theta})$ for $\boxed{a=1}$.

Therefore what we obtained is that $R(\theta,\hat{\theta}) \le r(f,\hat{\theta})$ and this shows that $\hat{\theta}$ is minimax

We return to the case of the normal distribution. To deal with it we use the following result.

Def We say that the sequence of priors (f_n) is <u>least favorable</u> if for any prior f,

$$r(f,\hat{\theta}^{f}) \le r = \lim_{n\to\infty} r(f_n,\hat{\theta}^{f_n})$$

<u>Theorem</u> If $(f_n)_{n\ge1}$ is a sequence such that

$$r = \lim_{n\to\infty} r(f_n,\hat{\theta}^{f_n})$$ and

$\hat{\theta}$ is such that

$$r = \sup_{\theta} R(\hat{\theta},\theta)$$

then $\hat{\theta}$ is minimax & (f_n) is least favorable.

Pf Assume that $\hat{\theta}'$ is another estimator. Then

$$\sup_{\theta} R(\hat{\theta}',\theta) \ge \int R(\hat{\theta}',\theta) f_n(\theta)\,d\theta \ge$$
$$\ge r(f_n,\hat{\theta}^{f_n}) \xrightarrow[n\to\infty]{} r$$

and this holds for every n. Thus taking the limit we get that

$$\sup_{\theta} R(\hat{\theta}',\theta) \ge \sup_{\theta} R(\hat{\theta},\theta) = r$$

thus $\hat{\theta}$ is minimax.

In addition we get that (f_n) is least favorable because if f is another prior, then

$$r(f,\hat{\theta}^{f}) = \int R(\hat{\theta}^{f},\theta) f(\theta)\,d\theta \le \int R(\hat{\theta},\theta) f(\theta)\,d\theta$$
$$\le \int \sup_{\theta} R(\hat{\theta},\theta) f(\theta)\,d\theta = r$$

Now we apply this to the normal case. (7)

Take $X_1, \dots, X_n \sim N(\theta, \sigma^2)$ and take $\theta \sim N(a, b^2)$ and observe that the Bayesian estimator for the square loss is

$$\theta_{a,b}(\vec{x}) = \frac{n\bar{x}/\sigma^2 + a/b^2}{n/\sigma^2 + 1/b^2}$$

The posterior variance is $\left(\theta|\vec{X} \sim N\left(\frac{n\bar{x}/\sigma^2 + a/b^2}{n/\sigma^2 + 1/b^2}, \frac{1}{n/\sigma^2 + 1/b^2}\right)\right)$

$$r(f, \hat{\theta}^f) = \int R(\hat{\theta}^f, \theta) f(\theta)\, d\theta$$
$$= \int r(\hat{\theta}^f | \vec{x})\, m(x)\, dx$$

and $r(\hat{\theta}^f | \vec{x}) = \mathrm{Var}(\theta | \vec{x}) = \dfrac{1}{n/\sigma^2 + 1/b^2}$

Therefore we obtain that

$$r(f, \hat{\theta}^f) = \frac{1}{n/\sigma^2 + 1/b^2}$$

Letting $b \uparrow \infty$ we see that

$$r(f_b, \hat{\theta}^{f_b}) \longrightarrow \sigma^2/n$$

which is the risk of $\bar{X}$. (Indeed $R(\bar{X}, \theta) = \sigma^2/n$ for all θ, therefore $\bar{X}$ is minimax.

In fact $R(\hat{\theta}_{a,b}, \theta) = E\left[\left(\dfrac{n\bar{X}/\sigma^2 + a/b^2}{n/\sigma^2 + 1/b^2} - \theta\right)^2\right]$

$$= \frac{1}{(n/\sigma^2 + 1/b^2)^2} E\left[\left(n(\bar{X} - \theta)/\sigma^2 + (a-\theta)/b^2\right)^2\right]$$

$$= \frac{n/\sigma^2 + (a-\theta)^2/b^4}{(n/\sigma^2 + 1/b^2)^2}$$

further

$$r(f, \hat{\theta}^f) = \frac{1}{\sqrt{2\pi b^2}} \int \frac{n/\sigma^2 + (a-\theta)^2/b^4}{(n/\sigma^2 + 1/b^2)^2} e^{-(a-\theta)^2/2b^2}\, d\theta$$

$$= \frac{n/\sigma^2 + 1/b^2}{(n/\sigma^2 + 1/b^2)^2} = \frac{1}{n/\sigma^2 + 1/b^2} \xrightarrow[b \to \infty]{} \frac{\sigma^2}{n}.$$