STAT 801 Lecture 25

Reading for Today's Lecture:
Goals of Today's Lecture:

- Do Bayesian estimation examples.
- Look at hypothesis testing as decision theory.

Today's notes

Statistical Decision Theory: examples

Example: In estimation theory to estimate a real parameter θ we used $D = \Theta$,

$$L(d, \theta) = (d - \theta)^2$$

and find that the risk of an estimator $\hat{\theta}(X)$ is

$$R_{\hat{\theta}}(\theta) = E[(\hat{\theta} - \theta)^2]$$

which is just the Mean Squared Error of $\hat{\theta}$.

The Bayes estimate of θ is $E(\theta|X)$, the posterior mean of θ.
Example: In $N(\mu, \sigma^2)$ model with σ known a common prior is $\mu \sim N(\nu, \tau^2)$. The resulting posterior distribution is Normal with posterior mean

$$E(\mu|X) = \frac{\frac{n}{\sigma^2}\bar{X} + \frac{1}{\tau^2}\nu}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$$

and posterior variance $1/(n/\sigma^2 + 1/\tau^2)$.
Improper priors: When the density does not integrate to 1 we can still follow the machinery of Bayes' formula to derive a posterior. For instance in the $N(\mu, \sigma^2)$ example consider the prior density $\pi(\theta) \equiv 1$. This "density" integrates to ∞ but using Bayes' theorem to compute the posterior would give

$$\pi(\theta|X) = \frac{(2\pi)^{-n/2}\sigma^{-n}\exp\{-\sum(X_i - \mu)^2/(2\sigma^2)\}}{\int(2\pi)^{-n/2}\sigma^{-n}\exp\{-\sum(X_i - \nu)^2/(2\sigma^2)\}d\nu}$$

It is easy to see that this cancels to the limit of the case previously done when $\tau \to \infty$ giving a $N(\bar{X}, \sigma^2/n)$ density. That is, the Bayes estimate of μ for this improper prior is $\bar{X}$.

Admissibility: Bayes procedures with finite Bayes risk and continuous risk functions are admissible. It follows that for each $w \in (0,1)$ and each real ν the estimate

$$w\bar{X} + (1-w)\nu$$

is admissible. That this is also true for $w = 1$, that is, that $\bar{X}$ is admissible, is much harder to prove.

Minimax estimation: The risk function of $\bar{X}$ is simply σ^2/n. That is, the risk function is constant since it does not depend on μ. Were $\bar{X}$ Bayes for a proper prior this would prove that $\bar{X}$ is minimax. In fact this is also true but hard to prove.

Example: Suppose that given p X has a Binomial(n,p) distribution. We will give p a Beta(α,β) prior density

$$\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

The joint "density" of X and p is

$$\binom{n}{X} p^X (1-p)^{n-X} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

so that the posterior density of p given X is of the form

$$cp^{X+\alpha-1}(1-p)^{n-X+\beta-1}$$

for a suitable normalizing constant c. But this is a Beta$(X+\alpha, n-X+\beta)$ density. The mean of a Beta(α,β) distribution is $\alpha/(\alpha+\beta)$. Thus the Bayes estimate of p is

$$\frac{X+\alpha}{n+\alpha+\beta} = w\hat{p} + (1-w)\frac{\alpha}{\alpha+\beta}$$

where $\hat{p} = X/n$ is the usual mle. Notice that this is again a weighted average of the prior mean and the mle. Notice also that the prior is proper for $\alpha > 0$ and $\beta > 0$. To get $w = 1$ we take $\alpha = \beta = 0$ and use the improper prior

$$\frac{1}{p(1-p)}$$

Again we learn that each $w\hat{p} + (1-w)p_o$ is admissible for $w \in (0,1)$. Again it is true that $\hat{p}$ is admissible but that our theorem is not adequate to prove this fact.

The risk function of $w\hat{p} + (1-w)p_0$ is

$$R(p) = E[(w\hat{p} + (1-w)p_0 - p)^2]$$

which is

$$w^2 Var(\hat{p}) + (wp + (1-w)p - p)^2 = w^2 p(1-p)/n + (1-w)^2(p-p_0)^2$$

This risk function will be constant if the coefficients of both p^2 and of p in the risk are 0. The coefficient of p is

$$-w^2/n + (1-w)^2$$

so $w = n^{1/2}/(1+n^{1/2})$. The coefficient of p is then

$$w^2/n - 2p_0(1-w)^2$$

which will vanish if $2p_0 = 1$ or $p_0 = 1/2$. Working backwards we find that to get these values for w and p_0 we require $\alpha = \beta$. Moreover the equation

$$w^2/(1-w)^2 = n$$

gives

$$n/(\alpha+\beta) = \sqrt{n}$$

or $\alpha = \beta = \sqrt{n}/2$. The minimax estimate of p is

$$\frac{\sqrt{n}}{1+\sqrt{n}}\hat{p} + \frac{1}{1+\sqrt{n}}\frac{1}{2}$$

Example: Now suppose that $X_1, \ldots, X_n$ are iid $MVN(\mu, \Sigma)$ with Σ known. Consider as the improper prior for μ which is constant. The posterior density of μ given X is then $MVN(\bar{X}, \Sigma/n)$.

For multivariate estimation it is common to extend the notion of squared error loss by defining

$$L(\hat{\theta}, \theta) = \sum(\hat{\theta}_i - \theta_i)^2 = (\hat{\theta} - \theta)^t(\hat{\theta} - \theta)\,.$$

For this loss function the risk is the sum of the MSEs of the individual components and the Bayes estimate is the posterior mean again. Thus $\bar{X}$ is Bayes for an improper prior in this problem. It turns out that $\bar{X}$ is minimax; its risk function is the constant $trace(\Sigma)/n$. If the dimension p of θ is 1 or 2 then $\bar{X}$ is also admissible but if $p \geq 3$ then it is inadmissible. This fact was first demonstrated by James and Stein who produced an estimate which is better, in terms of this risk function, for every μ. The "improved" estimator, called the James Stein estimator, is essentially never used.

Hypothesis Testing and Decision Theory

One common decision analysis of hypothesis testing takes $D = \{0, 1\}$ and $L(d, \theta) = 1(\text{make an error})$ or more generally $L(0, \theta) = \ell_1 1(\theta \in \Theta_1)$ and $L(1, \theta) = \ell_2 1(\theta \in \Theta_0)$ for two positive constants ℓ_1 and ℓ_2. We make the decision space convex by allowing a decision to be a probability measure on D. Any such measure can be specified by $\delta = P(\text{reject})$ so $\mathcal{D} = [0, 1]$. The loss function of $\delta \in [0, 1]$ is

$$L(\delta, \theta) = (1 - \delta)\ell_1 1(\theta \in \Theta_1) + \delta \ell_0 1(\theta \in \Theta_0) .$$

Simple hypotheses: A prior is just two numbers π_0 and π_1 which are non-negative and sum to 1. A procedure is a map from the data space to $\mathcal{D}$ which is exactly what a test function was. The risk function of a procedure $\phi(X)$ is a pair of numbers:

$$R_\phi(\theta_0) = E_0(L(\delta, \theta_0))$$

and

$$R_\phi(\theta_1) = E_1(L(\delta, \theta_1))$$

We find

$$R_\phi(\theta_0) = \ell_0 E_0(\phi(X)) = \ell_0 \alpha$$

and

$$R_\phi(\theta_1) = \ell_1 E_1(1 - \phi(X)) = \ell_1 \beta$$

The Bayes risk of ϕ is

$$\pi_0 \ell_0 \alpha + \pi_1 \ell_1 \beta$$

We saw in the hypothesis testing section that this is minimized by

$$\phi(X) = 1(f_1(X)/f_0(X) > \pi_1 \ell_1 / (\pi_0 \ell_0))$$

which is a likelihood ratio test. These tests are Bayes and admissible. The risk is constant if $\beta \ell_1 = \alpha \ell_0$; you can use this to find the minimax test in this context.

Composite hypotheses: