LECTURE 15

§18. Minimax Estimates

In general, we have a loss function $L(\theta, \hat{\theta}(x_1, x_2, ..., x_n))$, which reflects how much we lose, if the real value of the parameter θ, and our estimate is $\hat{\theta}(x_1, x_2, ..., x_n)$. We will substitute some estimates $\hat{\theta}(x_1, x_2, ..., x_n)$, where $x_1, ..., x_n$ is the sample. Thus, our losses will be a random variable, which we, in some way, would like to minimize.

Example 18.1. Let the owner of the bakery estimate the number of people θ, who will come to him today. If his estimate $\hat{\theta}$ turns out to be too low, he will lose $c_1 \cdot (\theta - \hat{\theta})$ money, where c_1 is the benefit received by the owner from each loaf. If overpriced, it will lose $c_2 \cdot (\hat{\theta} - \theta)$, where c_2 is the cost price (price of production) of one loaf. At the same time, he will use some estimate $\hat{\theta}$ related to the statistics of previous days, so his losses will be random, dependent on θ and on the sample. For the baker, his loss is a key characteristic of the estimation method, prevailing over the good properties of the estimates.

From the loss function, we will require that the function $L(\theta, \hat{\theta})$ be nonnegative and be 0 at $\hat{\theta} = \theta$. It is not very convenient to work with a random loss and we will deal with minimizing its average, i.e. minimizing the risk function:

$$R(\theta, \hat{\theta}) = E_\theta L(\theta, \hat{\theta}(x_1, ..., x_n)) = \int_{\mathbf{R}^n} L(\theta, \hat{\theta}(x_1, ..., x_n))\, p(x_1, ..., x_n)/\theta)\, dx_1\, dx_2 ... dx_n.$$

Example 18.2. In Example 18.1, we assume that x_i have a normal distribution with mean θ and variance 1 and the loss function is absolute error loss function, that is

$$L(\theta, \hat{\theta}) = |\hat{\theta} - \theta|.$$

. Then, in the estimate $\hat{\theta} = x_1$, the risk function will be equal to

$$c_2 E(x_1 - \theta) I_{x_1 > \theta} + c_1 E(\theta - x_1) I_{x_1 < \theta} =$$

$$= \frac{c_2}{\sqrt{2\pi}} \int_\theta^{+\infty} (x - \theta) \exp\left\{-\frac{(x-\theta)^2}{2}\right\} dx + \frac{c_1}{\sqrt{2\pi}} \int_{-\infty}^{\theta} (\theta - x) \exp\left\{-\frac{(x-\theta)^2}{2}\right\} dx =$$

Let us make change variable $x-\theta=y$, $dx=dy$, we obtain

$$= \frac{c_2}{\sqrt{2\pi}} \int_0^{+\infty} y \exp\left\{-\frac{y^2}{2}\right\} dy - \frac{c_1}{\sqrt{2\pi}} \int_{-\infty}^{0} y \exp\left\{-\frac{y^2}{2}\right\} dy =$$

$$= \frac{c_1+c_2}{\sqrt{2\pi}} \int_0^{\infty} x\, e^{-x^2/2}\, dx = \frac{c_1+c_2}{\sqrt{2\pi}}.$$

Often, loss functions are considered that depend only on the difference θ and $\hat\theta$. In this case, the quadratic loss function

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

plays a special role. Its importance follows from the following consideration: any smooth two-fold differentiable loss function L, depending only on $\theta-\hat\theta$, decomposes into a Taylor series

$$L(\theta-\hat\theta) = a + b(\theta-\hat\theta) + c(\theta-\hat\theta)^2 + \varepsilon,$$

where $a=0$, $b=0$ from the conditions of nonnegativity and equality 0 at point 0. In turn, it is in order, ε is $o((\theta-\hat\theta)^2)$, with closely evaluating θ evaluations (for example, for consistent $\hat\theta$ and large n), the risk function will behave as $c(\theta-\hat\theta)^2$.

For a quadratic loss function, the risk function will simply be equal to $E_\theta(\hat\theta-\theta)^2$. In particular, for an unbiased estimate it will be $\mathrm{Var}_\theta(\hat\theta)$. For biased estimates due to the following equalities

$$E(X-a)^2 = E(X-EX+EX-a)^2 = E(X-EX)^2 + (EX-a)^2 + 2(EX-a)E(X-EX)$$

$$= \mathrm{Var}X + (E(X)-a)^2,$$

the quadratic risk is $\mathrm{Var}_\theta(\hat\theta) + [E(\hat\theta)-\theta]^2$.

We would like to choose an estimator so as to minimize the risk function. However, the risk function at each estimator will depend on θ and therefore it is necessary to specify how are we going to choose the lesser among the functions. There are two main approaches here.

1) Limitations of the considered estimates. If we restrict the set of estimates under consideration to a certain class, then it may turn out that in this class one of the risk

functions lies below all others, that is, for all $\theta \in \Theta$, inequality $R(\theta, \hat{\theta}) \leq R(\theta, \hat{\theta}_1)$ holds. Such an estimator is called uniformly the most powerful estimator in its class.

2) Bayes approach. Consider the following example.

Example 18.3. Suppose one observation was taken of a random variable X which yielded the value 2. The density function for X is

$$p(x/\theta) = \begin{cases} \dfrac{1}{\theta} & \text{if} \quad 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}.$$

and prior distribution for parameter θ is

$$p(\theta) = \begin{cases} \dfrac{3}{\theta^4} & \text{if} \quad 1 < \theta < \infty \\ 0 & \text{otherwise} \end{cases}.$$

If the loss function is $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$, then what is the Bayes" estimate for θ?

Solution. Thus, the joint density function of the sample and the parameter θ is given by

$$f(x, \theta) = p(\theta) \cdot p(x/\theta) = \frac{3}{\theta^4} \cdot \frac{1}{\theta} = \begin{cases} \dfrac{3}{\theta^5} & \text{if} \quad 0 < x < \theta \quad \text{and} \quad 1 < \theta < \infty \\ 0 & \text{otherwise} \end{cases}.$$

The marginal density of the sample is given by

$$f(x) = \int_x^\infty f(x, \theta)\, d\theta = \int_x^\infty 3\, \theta^{-5}\, d\theta = \frac{3}{4x^4}.$$

Thus, if $x = 2$, then $f(2) = \dfrac{3}{64}$. The posterior density of θ when $x = 2$ is given by

$$p(\theta/x = 2) = \frac{f(2, \theta)}{f(2)} = \begin{cases} \dfrac{64}{\theta^5} & \text{if} \quad 2 < \theta < \infty \\ 0 & \text{otherwise} \end{cases}.$$

Now, we find the Bayes estimator by minimizing the expression

$$E[L(\theta, \hat{\theta})/x = 2].$$

That is,

$$\Psi(\hat{\theta}) = \int_\Omega L(\theta, \hat{\theta})\, p(\theta/x = 2)\, d\theta = \int_2^\infty (\theta - \hat{\theta})^2\, p(\theta/x = 2)\, d\theta =$$

$$= 64 \int_2^\infty (\theta - \hat{\theta})^2 \, \theta^{-5} \, d\theta.$$

We want to find the value of $\hat{\theta}$ which yields a minimum of $\Psi(\hat{\theta})$. This can be done by taking the derivative of $\Psi(\hat{\theta})$ and evaluating where the derivative is zero.

$$\frac{d}{d\hat{\theta}} \Psi(\hat{\theta}) = \frac{d}{d\hat{\theta}} \left[64 \int_2^\infty (\theta - \hat{\theta})^2 \, \theta^{-5} \, d\theta \right] =$$

$$= 128 \int_2^\infty (\hat{\theta} - \theta) \, \theta^{-5} \, d\theta = 128 \int_2^\infty \hat{\theta} \, \theta^{-5} \, d\theta - 128 \int_2^\infty \theta^{-4} \, d\theta = 2\hat{\theta} - \frac{16}{3}.$$

Setting this derivative of $\Psi(\hat{\theta})$ to zero and solving or $\hat{\theta}$, we get

$$\hat{\theta} = \frac{8}{3}.$$

Since

$$\frac{d^2 \Psi(\hat{\theta})}{d\hat{\theta}^2} = 2$$

the function $\Psi(\hat{\theta})$ has a minimum at $\hat{\theta} = 8/3$. Hence, the Bayes' estimator of θ is $8/3$.

In Example 18.3, we have found the Bayes' estimate of θ by directly minimizing the

$$\int_\Omega L(\theta, \hat{\theta}) \, p(\theta/X_1, X_2, ..., X_n) \, d\theta$$

with respect to $\hat{\theta}$.

The next result is very useful while finding the Bayes' estimate using a quadratic loss function. Note that if $L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$, then

$$\int_\Omega L(\theta, \hat{\theta}) \, p(\theta/X_1, X_2, ..., X_n) \, d\theta$$

is $E\left[(\theta - \hat{\theta})^2 / X_1, X_2, ..., X_n \right]$. The following theorem is based on the fact that the function ϕ defined by $\phi(c) = E[(X - c)^2]$ attains minimum if $c = EX$.

Theorem. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with density $f(x/\theta)$, where θ is the unknown parameter to be estimated. If the loss function is squared error, then Bayes' estimator $\hat{\theta}$ of parameter θ is given by

$$\hat{\theta} = E\left[\theta / X_1, X_2, ..., X_n\right], \tag{4}$$

where the expectation is taken with respect to posterior density $p(\theta/X_1, X_2, ..., X_n)$.

Therefore in Example 18.3 we can obtain $\hat{\theta}$ using Theorem (see (4)):

$$\hat{\theta} = \int_2^{\infty} \theta \frac{64}{\theta^5}\, d\theta = 64 \int_2^{\infty} \theta^{-4}\, d\theta = -\frac{64}{3} \cdot \frac{1}{\theta^3}\bigg|_2^{\infty} = \frac{8}{3}.$$

Now we give several examples to illustrate the use of this theorem.

Example 20. Suppose the prior distribution of θ is uniform over the interval $(0, 1)$. Given θ, the population X is uniform over the interval $(0, \theta)$. If the squared error loss function is used, find the Bayes' estimator of θ based on a sample of size one.

Solution. The prior density of θ is given by

$$p(\theta) = \begin{cases} 1 & \text{if} \quad 0 < \theta < 1 \\ 0 & \text{otherwise} \end{cases}.$$

The density of population X is given by

$$p(x/\theta) = \begin{cases} \frac{1}{\theta} & \text{if} \quad 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}.$$

The joint density of the sample and the parameter θ is given by

$$f(x, \theta) = p(\theta) \cdot p(x/\theta) = \begin{cases} \frac{1}{\theta} & \text{if} \quad 0 < x < \theta < 1 \\ 0 & \text{otherwise} \end{cases}.$$

The marginal density of the sample is given by

$$f(x) = \int_x^1 f(x, \theta)\, d\theta = \int_x^1 \frac{1}{\theta}\, d\theta = \begin{cases} -\ln x & \text{if} \quad 0 < x < \theta < 1 \\ 0 & \text{otherwise} \end{cases}.$$

The posterior density of θ when the sample is x given by

$$p(\theta/x) = \frac{f(x, \theta)}{f(x)} = \begin{cases} -\frac{1}{\theta \cdot \ln x} & \text{if} \quad 0 < x < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}.$$

Since the loss function is quadratic error, therefore the Bayes' estimator of θ is

$$\hat{\theta} = E[\theta/x] = \int_x^1 \theta p(\theta/x)\, d\theta = -\int_x^1 \theta \frac{1}{\theta \cdot \ln x}\, d\theta = -\frac{1}{\ln x} \int_x^1 d\theta = \frac{x - 1}{\ln x}.$$

Thus the Bayes' estimator of θ based on one observation X is

$$\hat{\theta} = \frac{X - 1}{\ln X}.$$