LECTURE 13

If we have $(x_1, x_2, \ldots, x_n)$ a sample, then the proof have the following form

We show that

$$\arg\min_x \sum_{i=1}^{n} |x_i - x|^2 = mean$$

$$S(x) = \sum_{i=1}^{n} |x_i - x|^2 = \sum_{i=1}^{n} (x_i^2 - 2x_i x + x^2) = \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} 2x_i x + \sum_{i=1}^{n} x^2 = \sum_{i=1}^{n} x_i^2 - 2x \sum_{i=1}^{n} x_i + nx^2.$$

We have to solve the equation

$$S'(x) = 0$$

that is

$$-2 \sum_{i=1}^{n} x_i + 2nx = 0$$

Therefore, we obtain

$$x = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} = mean$$

Since $S''(x) = n > 0$, therefore $x = \bar{x}$ is minimum of $S(x)$.

Absolute loss function $E|\hat{\theta} - \theta|$.

Theorem 17.3.

$$\hat{\theta} = \arg\min_{\hat{\theta}} \int |\theta - \hat{\theta}| \, p(\theta | x_1, x_2, ..., x_n) \, d\theta$$

then $\hat{\theta}$ is a posterior distribution median.

Let's prove for continuous probability distributions.

Proof:

$$E|\hat{\theta} - \theta| = \int |\theta - \hat{\theta}| p(\theta | x_1, x_2, ..., x_n) \, d\theta =$$

$$= \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) \, p(\theta | x_1, x_2, ..., x_n) \, d\theta + \int_{\hat{\theta}}^{+\infty} (\theta - \hat{\theta}) p(\theta | x_1, x_2, ..., x_n) \, d\theta$$

Calculate the first derivative in $\hat{\theta}$ and equating to 0, we get

$$\int_{-\infty}^{\hat{\theta}} p(\theta | x_1, x_2, ..., x_n) \, d\theta = \int_{\hat{\theta}}^{+\infty} p(\theta | x_1, x_2, ..., x_n) \, d\theta$$

Therefore, we get

$$2\int_{-\infty}^{\hat{\theta}} p(\theta|x_1, x_2, ..., x_n)\, d\theta =$$
$$= \int_{-\infty}^{\infty} p(\theta|x_1, x_2, ..., x_n)\, d\theta = 1$$

That is

$$\int_{-\infty}^{\hat{\theta}} p(\theta|x_1, x_2, ..., x_n)\, d\theta = \frac{1}{2}$$

implying that $\hat{\theta}$ is a posterior $p(\theta|x_1, x_2, ..., x_n)$ density median.

Another proof: Let's consider $E|\hat{\theta} - \theta|$

$$E|\hat{\theta} - \theta| = \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta)\, dF + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta})\, dF =$$

$$= \int_{-\infty}^{M} (\hat{\theta} - M + M - \theta) dF + \int_{M}^{\hat{\theta}} (\hat{\theta} - \theta) dF + \int_{\hat{\theta}}^{M} (\theta - \hat{\theta})\, dF + \int_{M}^{\infty} (\theta - M + M - \hat{\theta})\, dF =$$

$$= \int_{-\infty}^{M} (M-\theta)\, dF + (\hat{\theta}-M) \int_{-\infty}^{M} dF + \int_{M}^{\hat{\theta}} (\hat{\theta}-\theta)\, dF + \int_{\hat{\theta}}^{M} (\theta-\hat{\theta})\, dF + \int_{M}^{\infty} (\theta-M)\, dF + (M-\hat{\theta}) \int_{M}^{\infty} dF =$$

$$= E|\theta - M| + (\hat{\theta} - M) \int_{-\infty}^{M} dF + \int_{M}^{\hat{\theta}} (\hat{\theta} - \theta)\, dF + \int_{\hat{\theta}}^{M} (\theta - \hat{\theta})\, dF + (M - \hat{\theta}) \int_{M}^{\infty} dF =$$

$$= E|\theta - M| + \int_{M}^{\hat{\theta}} (\hat{\theta} - \theta)\, dF + \int_{\hat{\theta}}^{M} (\theta - \hat{\theta})\, dF.$$

Therefore, we obtain

$$\min_{\hat{\theta}} E|\hat{\theta} - \theta| \qquad \text{when} \qquad \hat{\theta} = M.$$

The proof bellow is convenient, if we consider sample median.

Proof: We show that

$$\arg\min_{a} \sum_{i=1}^{n} |x_i - a| = median$$

First we consider the case of two summands

$$S(a) = |x_1 - a| + |x_2 - a|$$

Depending on a and x_1 and x_2 we consider the following three cases

1. $x_1 \le a \le x_2$

$$S(a) = (a - x_1) + (x_2 - a) = x_2 - x_1$$

2. $a < x_1 \le x_2$

$$S(a) = (x_1 - a) + (x_2 - a) = x_1 + x_2 - 2a > x_1 + x_2 - 2x_1 = x_2 - x_1$$

3. $x_1 \le x_2 < a$

$$S(a) = (a - x_1) + (a - x_2) = 2a - x_1 - x_2 > 2x_2 - x_1 - x_2 = x_2 - x_1$$

Thus, $S \ge x_2 - x_1$, and S has a minimal value $S = x_2 - x_1$ if and only if $a \in [x_1, x_2]$.

Now we consider the following intervals

$$[x_{(1)}, x_{(n)}], \qquad [x_{(2)}, x_{(n-1)}], \quad \ldots, [x_{(i)}, x_{(n+1-i)}],$$

where $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ are in the increasing order and $i = 1, \ldots, c$, where

$$c = \begin{cases} \frac{n}{2}, \text{ if } n \text{ is even}, \\ \frac{n+1}{2}, \text{ if } n \text{ is odd} \end{cases}$$

The cases of even or odd of n we consider separately.

Lets n is odd. In this case the central interval is $[x_{\frac{n+1}{2}}, x_{\frac{n+1}{2}}] = x_{\frac{n+1}{2}}$

Consider

$$a \in \bigcap_{i=1}^{c} [x_{(i)}, x_{(n+1-i)}]$$

$$S(a) = \sum_{i=1}^{n} |x_i - a| = (|x_{(1)} - a| + |x_{(n)} - a|) + (|x_{(2)} - a| + |x_{(n-1)} - a|) + \ldots + \frac{1}{2}(|x_{\frac{n+1}{2}} - a| + |x_{\frac{n+1}{2}} - a|)$$

$$a \in [x_{(i)}, x_{(n+1-i)}], \qquad i = 1, \ldots, \frac{n+1}{2}$$

Therefore, by 2 point case a is attained the minimum for any $|x_{(i)} - a| + |x_{(n+1-i)} - a|$, and therefore for whole sum.

Note that $a = x_{\frac{n+1}{2}}$ is just a median.

Let n is even. In this case the central interval is $[x_{\frac{n}{2}}, x_{\frac{n}{2}+1}]$. Therefore, any $a \in [x_{\frac{n}{2}}, x_{\frac{n}{2}+1}]$ by previous case will minimize $S(a)$. We get for a median $a = \frac{x_{\frac{n}{2}+1} + x_{\frac{n}{2}}}{2}$ the same arguments.

Appendix. Median of a probability distribution.

Median is one of the important characteristics of probability distributions. For a random variable $\eta(\omega)$ with $F(x)$ distribution function median is called a number M which satisfy the following two conditions:

$$F(M) \le 1/2 \qquad \text{and} \qquad F(M+0) \ge 1/2. \tag{A1}$$

Note that this definition true in the case if we define distribution function as

$$F(x) = P(\omega : \eta(\omega) < x),$$

then $F(x)$ continuous from the left, $F(x-0) = F(x)$ for any $x \in \mathbf{R}$.

But if we give the following definition for distribution function:

$$F(x) = P(\omega : \eta(\omega) \le x),$$

then definition of median will be the following:

$$F(M-0) \le 1/2 \qquad \text{and} \qquad F(M) \ge 1/2. \tag{A2}$$

What number is a median of $F(x)$ does not depend on the definition of $F(x)$. Any random variable has at least one median M. If $F(x) = 1/2$ for any x from the closed interval, then every point of this interval is a median. If $F(x)$ is strictly monotone function, then median is unique. In a symmetrical case, if median is unique, then it coincides with the expectation, if the mean exists. It is very important that median exists for any probability distribution.

Note that median M we can define also by the following formulae:

$$P(\eta \le M) \ge 1/2 \qquad \text{and} \qquad P(\eta \ge M) \ge 1/2. \tag{A3}$$

Show that this definition of median is coincides with definition (A1) for $F(x) = P(\eta < x)$ and with (A2) for $F(x) = P(\eta \le x)$.

(A1) If $F(x) = P(\eta < x)$, then

$$P(\eta \leq M) = F(M+0) \geq 1/2$$

and

$$P(\eta \geq M) = 1 - P(\eta < M) = 1 - F(M) \geq 1/2,$$

and therefore $F(M) \leq 1/2$. Thus for $F(x) = P(\eta < x)$ we obtain definition (A1).

(A2) If $F(x) = P(\eta \leq x)$, then

$$P(\eta \leq M) = F(M) \geq 1/2$$

and

$$P(\eta \geq M) = 1 - P(\eta < M) = 1 - F(M-0) \geq 1/2,$$

and therefore $F(M-0) \leq 1/2$. Thus for $F(x) = P(\eta \leq x)$ we obtain definition (A2).

Example 17.1. An example of discrete random variable:

$$\begin{array}{lll} \eta: & 0 & 1,000 \\ p: & 0.99 & 0.01 \end{array}$$

Then we can easily see the median is 0.

Example 17.2. Another example from discrete distribution:

$$\begin{array}{lll} \eta: & 0 & 1,000 \\ p: & 0.5 & 0.5 \end{array}$$

Then we see that the median is not unique. In fact, all real values in the interval $[0, 1,000]$ are medians.

Example 17.3. In practice, however, the median may be calculated as follows:
If there are N numerical data points, then by ordering the data values (either non-decreasingly or non-increasingly),
a) the $\dfrac{N+1}{2}$-th data point is the median if N is odd, and
b) the midpoint of the $(N-1)$-th and the $(N+1)$-th data points is the median if N is even.

Example 17.4. The median of a normal distribution $\eta \sim \mathcal{N}(\mu, \sigma^2)$, ($\mu$ is the mean, and σ^2 is the variance of η) is μ.

In fact, for a normal distribution,

mean = median = mode.

Example 17.5. Let η have a density function:

$$f(x) = \begin{cases} \frac{1}{6}(x+1) & \text{if} \quad 1 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the median.

Proof: It is not difficult to calculate distribution function of η:

$$F(x) = \begin{cases} 0 & \text{if} \quad x \le 1 \\ \frac{1}{12}x^2 + \frac{1}{6}x - \frac{1}{4} & \text{if} \quad 1 \le x \le 3 \\ 1 & \text{if} \quad x \ge 3 \end{cases}$$

Therefore we need to solve the equation $F(M) = \frac{1}{2}$, that is

$$\frac{1}{12}x^2 + \frac{1}{6}x - \frac{1}{4} = \frac{1}{2},$$

and the median is $M = 2.16227766$.

Indicator loss function

Theorem 17.4. If

$$\hat{\theta} = \arg\min_{\hat{\theta}} \int L(\theta, \hat{\theta})\, p(\theta | x_1, x_2, ..., x_n)\, d\theta$$

where

$$L(\theta, \hat{\theta}) = \begin{cases} 0, & |\theta - \hat{\theta}| < \delta \\ 1, & |\theta - \hat{\theta}| \ge \delta \end{cases}$$

then $\hat{\theta}$ is a posterior distribution mode.

Proof.

$$\int L(\theta, \hat{\theta})\, p(\theta | x_1, x_2, ..., x_n)\, d\theta = \int_{-\infty}^{\hat{\theta}-\delta} 1 \cdot p(\theta | x_1, x_2, ..., x_n)\, d\theta + \int_{\hat{\theta}+\delta}^{\infty} 1 \cdot p(\theta | x_1, x_2, ..., x_n)\, d\theta$$

Simplifying we get

$$\int L(\theta,\hat{\theta})p(\theta|x_1,x_2,...,x_n)\,d\theta = 1 - \int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} 1\cdot p(\theta|x_1,x_2,...,x_n)\,d\theta$$

we minimize, if maximize the following

$$\int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} p(\theta|x_1,x_2,...,x_n)\,d\theta$$

For δ and smooth $p(\theta|x_1,x_2,...,x_n)$ it allow his maximum when $p(\theta|x_1,x_2,...,x_n)$ gets his maximum value. Therefore the estimate is a mode (the maximal value of posterior density). Thus, it name is maximum a posterior, or MAP estimate. The proof is complete.

Consider MAP estimate

$$\hat{\theta}_{MAP} = \arg\max_\theta p(\theta|x_1,x_2,...,x_n)$$

It follows by Bayes formula

$$\hat{\theta}_{MAP} = \arg\max_\theta \frac{p(x_1,x_2,...,x_n|\theta)\,p(\theta)}{p(x_1,x_2,...,x_n)}$$

Therefore, since $p(x_1,x_2,...,x_n)$ not depends on θ, we have

$$\hat{\theta}_{MAP} = \arg\max_\theta p(x_1,x_2,...,x_n|\theta)\,p(\theta)$$

it is like to maximal likelihood estimate $\hat{\theta}_{ML} = \arg\max_\theta p(x_1,x_2,...,x_n|\theta)$, but it contains prior density.