Math 541: Statistical Theory II

Statistical Inference and Method of Moment

Instructor: Songfeng Zheng

1 Statistical Inference Problems

In probability problems, we are given a probability distribution, and the purpose is to to analyze the property (Mean, variable, etc.) of the random variable coming from this distribution. Statistics is the converse problem: we are given a set of random variables coming from an unknown distribution, and the task is to make inference about the property of the underlying probability distribution.

In real-world applications, the true distribution is rarely known to us, but the data is not hard to obtain. Therefore, in applications, we deal with statistics problems to discover the underlying law behind the data. In this sense, statistics helps us in scientific discovering. For example, Kepler discovered the orbit of planet by analyzing the observation data.

1.1 Parameter and Parameter Space

In many statistical problems, the probability distribution which generates the observed data is completely known except for the values of one or more parameters. For example, it might be known that the life of a lamp bulb has an exponential distribution with parameter β, but the exact value of β might be unknown. As another example, suppose that the distribution of the heights of the individuals in a certain population is known to be a normal distribution with mean μ and variance σ^2, but the exact values of μ and σ^2 are unknown.

In a statistical problem, any characteristic of the distribution generating the experimental data which has an unknown value, such as the mean μ and variance σ^2 in the example just presented, is called a *parameter* of the distribution. Parameter or parameter vector is usually denoted as θ in this note, and we denote Θ as the set of all the possible values of parameter θ, and it is called *parameter space*.

Thus, in the first example we presented, the parameter β of the exponential distribution must be positive. Therefore, unless certain positive values of β can be explicitly ruled out as possible values of β, the parameter space Θ will be the set of all positive numbers. In the second example, the mean μ and variance σ^2 of the normal distribution can be regarded as a pair of parameters. Here the value of μ can be any real number and σ^2 must be positive. Therefore, the parameter space Θ can be taken to be the set of all pairs (μ, σ^2) such that $-\infty < \mu < \infty$ and $\sigma^2 > 0$. More specifically, if the normal distribution in this example represents the distribution of the heights in inches of the individuals in some given population, we might be certain that $30 < \mu < 100$ and $\sigma^2 < 50$. In this case, the parameter space Θ can be taken to be the smaller set of all pairs (μ, σ^2) such that $30 < \mu < 100$ and $0 < \sigma^2 < 50$.

1.2 Settings in Statistical Inference Problem

In a statistics problem, we usually assume we have a random sample $X_1, \cdots, X_n$ which is taken from a distribution $f(x|\theta)$[1] with the unknown parameter(s) θ. We further suppose $X_1, \cdots, X_n$ are independent identically distributed (i.i.d.). Then the statistical inference problem is to make some statements about the unknown parameter(s) θ.

In this course, we will have three types of problem:

- Parameter estimation: to estimate the value of the unknown parameters, including evaluating the quality of the estimation;
- Interval estimation: with a certain confidence, to get an interval which contains the parameter;
- Hypothesis testing: to test whether or not a given statement about the parameter is true.

2 Estimator and Estimate

Given an i.i.d. random sample $X_1, \cdots, X_n$ which is taken from a distribution $f(x|\theta)$ with the unknown parameter(s) θ which lies in the parameter space Θ. Clearly, the the value of θ must be estimated from the observed values in the sample.

An *estimator* of the parameter θ, based on the random variables $X_1, \cdots, X_n$, is a real-valued function $\delta(X_1, \cdots, X_n)$ which specifies the estimated value of θ for each possible set of values of $X_1, \cdots, X_n$. In other words, if the observed values of $X_1, \cdots, X_n$ turn out to be $x_1, \cdots, x_n$, then the estimated value of θ is $\delta(x_1, \cdots, x_n)$. Since the value of θ must belong to the space Θ, it is reasonable to require that every possible value of an estimator $\delta(X_1, \cdots, X_n)$ must also belong to Θ.

From the definition, an estimator $\delta(X_1, \cdots, X_n)$ is a function of the random variables $X_1, \cdots, X_n$, therefore, in general, the estimator itself is a random variable, and its probability distribution

[1] We write the distribution model as $f(x|\theta)$ to emphasize the dependence of f on θ.

can be derived from the joint distribution of $X_1, \cdots, X_n$. On the other hand, an *estimate* is a specific value $\delta(x_1, \cdots, x_n)$ of the estimator that is determined by using specific observed values $x_1, \cdots, x_n$. It will often be convenient to use vector notation and to let $\mathbf{X} = (X_1, \cdots, X_n)$ and $\mathbf{x} = (x_1, \cdots, x_n)$. With this notation, an estimator is a function $\delta(\mathbf{X})$ of the random vector $\mathbf{X}$, and an estimate is a specific value $\delta(\mathbf{x})$.

3 Method of Moment

The method of moment is probably the oldest method for constructing an estimator, dating back at least to Karl Pearson, an English mathematical statistician, in the late 1800's. The advantage of method of moment is that it is quite easy to use; however, the quality of the result from method of moment is not very good. We will define the quality and see some examples of the quality of method of moment later in this course.

Suppose a random variable X has density $f(x|\theta)$, and this should be understood as point mass function when the random variable is discrete.

The k-th *theoretical moment* of this random variable is defined as

$$\mu_k = E(X^k) = \int x^k f(x|\theta) dx$$

or

$$\mu_k = E(X^k) = \sum_x x^k f(x|\theta).$$

If $X_1, \cdots, X_n$ are i.i.d. random variables from that distribution, the k-th *sample moment* is defined as

$$m_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k,$$

thus m_k can be viewed as an estimator for μ_k. From the law of large number, we have $m_k \to \mu_k$ in probability as $n \to \infty$.

If we equate μ_k to m_k, usually we will get an equation about the unknown parameter. Solving this equation will help us get the estimator of the unknown parameter. If the probability distribution has p unknown parameters, the method of moment estimators are found by equating the first p sample moments to corresponding p theoretical moments (which will probably depend on other parameters), and solving the resulting system of simultaneous equations.

To illustrate the procedure of method of moment, we consider several examples.

4 Examples

Example 1: Suppose that X is a discrete random variable with the following probability mass function: where $0 \leq \theta \leq 1$ is a parameter. The following 10 independent observations

X	0	1	2	3
$P(X)$	$2\theta/3$	$\theta/3$	$2(1-\theta)/3$	$(1-\theta)/3$

were taken from such a distribution: (3,0,2,1,3,2,1,0,2,1). Please use the method of moment to find the estimate of θ.

Solution: The theoretical mean value is

$$E(X) = \sum_{x=0}^{3} xP(x) = 0 \cdot \frac{2\theta}{3} + 1 \cdot \frac{\theta}{3} + 2 \cdot \frac{2(1-\theta)}{3} + 3 \cdot \frac{(1-\theta)}{3} = \frac{7}{3} - 2\theta$$

The sample mean value is:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{3+0+2+1+3+2+1+0+2+1}{10} = 1.5$$

We need to solve the equation $\frac{7}{3} - 2\theta = 1.5$, and we finally get the method of moment estimation $\hat{\theta} = \frac{5}{12}$.

In this example, we choose $k = 1$. The reason is that when k is small, it will be convenient to calculate the k-th theoretical moment and k-th sample moment. Another reason for using small k is that if k is too big, the k-th theoretical moment might not exist. Although in this example, it may not matter when we try to use large k's, but in applications, we prefer to use small k. This is our first rule for selecting k.

Example 2: Suppose $X_1, X_2, \cdots, X_n$ are i.i.d. random variables with probability density function

$$f(x|\sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right),$$

please use the method of moment to estimate σ.

Solution: If we calculate the first order theoretical moment, we would have:

$$E(X) = \int_{-\infty}^{\infty} xf(x|\sigma)dx = \int_{-\infty}^{\infty} x\frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right) dx = 0.$$

Thus, if we try to solve equation $E(X) = \overline{X}$, we will not get the estimator, because $E(X)$ does not contain the unknown parameter σ.

Now, let us calculate the second order theoretical moment, we have

$$\mu_2 = E(X^2) \quad = \quad \int_{-\infty}^{\infty} x^2 f(x|\sigma)dx = \int_{-\infty}^{\infty} x^2 \frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right) dx$$

$$\begin{aligned}
&= \int_0^\infty x^2 \frac{1}{\sigma} \exp\left(-\frac{|x|}{\sigma}\right) dx = \sigma^2 \int_0^\infty y^2 e^{-y} dy \quad (\text{Let } x = \sigma y) \\
&= -\sigma^2 \int_0^\infty y^2 de^{-y} = -\sigma^2 y^2 e^{-y} \mid_0^\infty + \sigma^2 \int_0^\infty 2ye^{-y} dy \\
&= 0 - 2\sigma^2 \int_0^\infty y de^{-y} = -2\sigma^2 y e^{-y} \mid_0^\infty + 2\sigma^2 \int_0^\infty e^{-y} dy \\
&= 0 + 2\sigma^2 e^{-y} \mid_0^\infty = 2\sigma^2
\end{aligned}$$

The second order sample moment is:

$$m_2 = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

Solving the equation $\mu_2 = m_2$, i.e. $2\sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$, we can obtain the estimate of σ:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n X_i^2}{2n}}$$

From this example, we can see that we also want to choose a k, such that $E(X^k)$ is a function of the unknown parameters. Only when it contains the unknown parameters, can we solve the equation. This is our second rule for selecting k.

Example 3: Use the method of moment to estimate the parameter θ for the uniform density $f(x|\theta) = \frac{1}{\theta}$, with $0 \le x \le \theta$, based on a random sample $X_1, \cdots, X_n$.

Solution: The first theoretical moment for this distribution is

$$E(X) = \int_0^\theta x \cdot \frac{1}{\theta} dx = \frac{x^2}{2\theta} \mid_0^\theta = \frac{\theta^2}{2\theta} - 0 = \frac{\theta}{2}$$

Equate the first theoretical moment to the first sample moment, we have

$$E(X) = \bar{X} \Rightarrow \frac{\theta}{2} = \overline{X} \Rightarrow \hat{\theta} = 2\overline{X} = \frac{2}{n} \sum_{i=1}^n X_i$$

as the method of moment estimate.

Example 4: Use the method of moment to estimate the parameters μ and σ^2 for the normal density

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},$$

based on a random sample $X_1, \cdots, X_n$.

Solution: The first and second theoretical moments for the normal distribution are

$$\mu_1 = E(X) = \mu \quad \text{and} \quad \mu_2 = E(X^2) = \mu^2 + \sigma^2$$

The first and second sample moments are

$$m_1 = \overline{X} \quad \text{and} \quad m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Solving the equations

$$\mu = \overline{X} \quad \text{and} \quad \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

We have the method of moment estimate

$$\hat{\mu} = \overline{X} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

5 Exercises

Exercise 1: Let $X_1, \cdots, X_n$ be an i.i.d. sample from an exponential distribution with the density function

$$f(x|\beta) = \frac{1}{\beta} e^{-\frac{x}{\beta}}, \text{ with } \quad 0 \le x < \infty.$$

Please use method of moment to estimate the parameter β.

Exercise 2: Suppose that $X_1, X_2, \cdots, X_n$ are i.i.d. random variables on the interval $[0, 1]$ with the density function

$$f(x|\alpha) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}[x(1-x)]^{\alpha-1}$$

where $\alpha > 0$ is a parameter to be estimated from the sample. It can be shown that

$$E(X) = \frac{1}{2}$$

$$Var(X) = \frac{1}{4(2\alpha+1)}$$

Please use method of moment to estimate the parameter α.

Exercise 3: The Pareto distribution has been used in economics as a model for a density function with a slowly decaying tail:

$$f(x|x_0, \theta) = \theta x_0^{\theta} x^{-\theta-1}, \quad x \ge x_0, \quad \theta > 1$$

Assume that $x_0 > 0$ is given and that $X_1, X_2, \cdots, X_n$ is an i.i.d. sample. Find the method of moments estimate of θ.