Stat120B: Introduction to Prabability and Statistics **Lecture Notes Archive**

Method of Moments

Lecturer: Gang Liang

1 Introduction to Parameter Estimation

The difference between proabability and statistics is that (1) in probability, we assume the distributions (pmf or pdf) of random variables are known, then we want to infer some properties of samples drawn from the distribution; (2) in statistics, our view point is more practical. The assumption is that we don't exactly know the distribution of the random variable, but would like to infer some properties of the distribution throught the observed samples.

The setup we will adopt is that (1) we know the distribution family of the random variable, for instance, exponential or Bernoulli, (2) but we do not know the parameter of the distribution. Then our goal is to infer the parameter(s) from an independent identically distributed (IID) sample.

Example: There is a biased coin with $P(head) = p$, but p is unknown to us. We conduct n independent experiments, and the results of these experiments are

$$X_1, X_2, \ldots, X_n.$$

The question is how to infer p from the sample data points.

Two of the most widely used principles in parameter estimation are the method of moment (MOM) and maximum likelihood (ML). Next we'll talk about the method of moments.

2 Method of Moment (MOM)

The main idea of parameter estimation is that *a sample is a subset of the population independently drawn from it*, hence a miniature of the population. One consequence of this observation is that the sample and the population should share many similarities, in other words, many properties derived from the sample can be generalized to the population (of course with cautions).

The method of moment (MOM) claims that the moments (recall what's moment?) calculated based on the sample should be similar to those of population. Such relations are used to produce the estimate of the unknown parameters.

Setup: Suppose the pdf/pmf of a distribution is

$$f_X(x; \theta), \text{ where } \theta \text{ is a vector of unknown parameters.}$$

Suppose the dimension of θ is k. Examples of θ are λ in the exponential distribution, and (μ, σ^2) in the normal distribution. We observe an sample independently drawn from the distribution:

$$X_1, X_2, X_3, \ldots, X_n.$$

Definition: an **estimator** of an unknown parameter θ is a function of the observed samples.

Sample moments: for the given sample $X_1, \dots, X_n$, the sample moment is defined as

$$\hat{m}_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r \text{ for } r \text{ to be positive integers.}$$

Population moments: the population moments are just the moments we defined before.

$$m_r = E(X^r)$$

Note that these quantities are functions of unknown parameter θ.

2.1 Procedure (MOM)

1. Compute the population moments (i.e., the moments based on the pdf/pmf):

$$\begin{cases} m_1 = E(X^1) = f_1(\theta_1, \dots, \theta_k) \\ m_2 = E(X^2) = f_2(\theta_1, \dots, \theta_k) \\ \dots \\ m_k = E(X^k) = f_k(\theta_1, \dots, \theta_k) \end{cases}$$

2. There are k equations and k parameters. Solve these equations,

$$\begin{cases} \theta_1 = g_1(m_1, \dots, m_k) \\ \theta_2 = g_2(m_1, \dots, m_k) \\ \dots \\ \theta_k = g_k(m_1, \dots, m_k) \end{cases}$$

3. Replacing the population moments with the sample moments yields the MOM estimate of θ:

$$\begin{cases} \hat{\theta}_1 = g_1(\hat{m}_1, \dots, \hat{m}_k) \\ \hat{\theta}_2 = g_2(\hat{m}_1, \dots, \hat{m}_k) \\ \dots \\ \hat{\theta}_k = g_k(\hat{m}_1, \dots, \hat{m}_k) \end{cases}$$

2.2 Example: (Bernoulli distribution)

There is only one parameter p for the Bernoulli random variable, so one equation will be sufficient to estimate p. Given $X_1, \dots, X_n$, the first sample moment is

$$\hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.$$

For Bernoulli distribution, we have

$$m_1 = E(X) = p.$$

Solving this equation (of course, it is trivia here, but it is important to realize we are solving an equation), we have

$$p = m_1.$$

Replacing the population moment m_1 with its sample counter part,

$$\hat{p}_{MOM} = \hat{m}_1 = \bar{X}.$$