III R. A. FISHER'S THEORY OF ESTIMATION[5]

The problem of estimation of the unknown parameter θ is the problem of finding a function $t(x_1,\ldots,x_n)$ of the observations such that t can be considered in a certain sense as a "good" or "best" estimate of θ. Since the estimate $t(x_1,\ldots,x_n)$ is a random variable, we cannot expect that its value should coincide with that of the unknown parameter, but we will try to choose $t(x_1,\ldots,x_n)$ in such a way as to make as great as possible the probability of the value of t lying as near as possible to the value of the unknown parameter θ.

This is a somewhat vague formulation of the requirement for a "good" or "best" statistical estimate. It can be made precise in different ways. Markoff[6], for instance, defines the notion of <u>a "best" estimate</u> as follows: A statistic t (we shall call any function of the observations a <u>statistic</u>) is a best estimate of θ if

(1) t is an unbiased estimate of θ, i.e., $E_\theta(t) = \theta$ identically in θ where $E_\theta(t)$ denotes the expected value of t under the assumption that θ is the true value of the parameter.

(2) $E_\theta(t-\theta)^2 \leq E_\theta(t'-\theta)^2$ identically in θ for all t' which satisfy (1).

This definition of a "best estimate" seems to be a reasonable and acceptable one since, in general, the smaller the variance of t the greater is the probability that t will lie in a small

---

5) See references 3 - 6

6) See reference 15, p.344

neighborhood of θ. It should be remarked that although (by virtue of Tshebisheff's inequality) smallness of the variance implies that the probability of t lying in a small neighborhood of θ is small, the converse is not necessarily true. It may happen that a statistic t has a large variance and, nevertheless, the probability of t lying in a small neighborhood of θ is high. This circumstance constitutes some argument against Markoff's definition. A more serious difficulty is, however, the fact that a best estimate in Markoff's sense seldom exists.

R. A. Fisher's theory of estimation is based on the <u>principle of the maximum likelihood.</u> It is assumed that a probability density

$$p(x_1,\ldots,x_n,\ \theta)$$

exists in the sample space, i.e., for any measurable subset W of the sample space

$$P(W|\theta) = \int_W p(x_1,\ldots,x_n,\ \theta)\,dx.$$

In particular, the cumulative distribution function is given by

$$F(x_1,\ldots,x_n\,,\ \theta) = \int_{-\infty}^{x_n}\int_{-\infty}^{x_{n-1}}\cdots\int_{-\infty}^{x_1} P(v_1,\ldots,v_n,\ \theta)dv_1,\ldots dv_n.$$

The maximum likelihood estimate $\theta_n(x_1,\ldots,x_n)$ is defined as that value of θ for which $p(x_1,\ldots,x_n,\theta)$ becomes a maximum. Now assume that $X_1,\ldots,X_n$ are n independently distributed random variables each having the same distribution. This can also be expressed by saying that $x_1,\ldots,x_n$ are n independent observations on the same random variable X. The main result of Fisher's theory of estimation can be stated as follows: If $x_1,\ldots,x_n$ are n independent observations (n = 1,..., ad inf.) on the same random variable X and if the distribution of X satisfies certain conditions (which are not too restrictive and in practical application are frequently fulfilled), then $\hat{\theta}_n$ is <u>an efficient estimate</u>. The definition of an efficient estimate is given as follows:

A sequence $\{t_n\}$ (n = 1,..., ad inf.) of statistics is called an efficient estimate of θ (the subscript n indicates the number of observations of which t_n is a function) if

(1) the limit distribution of $\sqrt{n}\,(t_n - \theta)$ is a normal distribution with zero mean and finite variance, and

(2) for any sequence $\{t_n'\}$ of statistics which satisfies (1)

$$\sigma^2/\sigma'^2 \leq 1$$

where $\sigma^2 = \lim E_\theta \left[\sqrt{n}\,(t_n - \theta)\right]^2$

and $\sigma'^2 = \lim E_\theta \left[\sqrt{n}\,(t_n' - \theta)\right]^2$

The ratio σ^2/σ'^2 is called the efficiency of $\{t_n'\}$ which is always ≤ 1.

Vaguely speaking, in large samples the maximum likelihood estimate has the smallest variance compared with any other statistic which is in the limit normally distributed. The restriction of the comparison to statistics which are in the limit normally distributed seems to be a serious one. However, as recent results show, the maximum likelihood estimate has a much stronger property than efficiency, and it can be considered as a "best" large sample estimate of θ compared even with statistics which are not normally distributed in the limit.[7)]

---

7) See reference 20

The question of consistency and limit distribution of the maximum likelihood estimate has been treated by H. Hotelling, 7. A complete proof has been given by J. L. Doob, 1.

As an example, let $x_1, \ldots, x_n$ be n independent observations on a normally distributed variate X with unknown mean and unit variance. It can easily be verified that the maximum likelihood estimate of θ is given by

$$\hat{\theta}_n(x_1, \ldots, x_n) = \frac{x_1 + \ldots + x_n}{n}$$

Let $t_n(x_1, \ldots, x_n)$ be the median of the observations $x_1, \ldots, x_n$. It can be shown that the limit distribution of $\sqrt{n}\,(t_n - \theta)$ is normal with zero mean and variance $\frac{\pi}{2}$. Hence, the efficiency of the median for estimating θ is equal to $\frac{2}{\pi} = 0.6366\ldots$