On Tukey's Polyefficiency

The assumed underlying distribution of the Data does not usually represent the reality which is often "somewhere in between" of what experts think about. Thus, statistical inference based on one model is usually not accurate with real data. In finite samples, the need of robustness of conclusions due to deviations from the assumed model motivated John Tukey to introduce the notion of **Polyefficiency**.

Note: First draft of this section. Please let me know if there are typos or something is not clear. Thank you!

Set-up of a statistical estimation problem

Observe data $\mathbf{X}_n = (X_1, \ldots, X_n)$ from a model/density $f(x, \theta)$ with the parameter θ unknown.

- $S_n = S(\mathbf{X_n})$: estimate of interest for θ.
- $T_n = T(\mathbf{X_n})$: a generic estimate of θ.
- $\mathcal{R}(T_n, \theta)$: the cost (called "Risk") in estimating θ with T_n, calculated under f.

$\mathcal{R}(x, y)$ is a distance-measure with properties: *i)* $\mathcal{R}(x, x) = 0$,
ii) $\mathcal{R}(x, y) = \mathcal{R}(y, x)$, for every x, y in the domain of $\mathcal{R}$.
What is missing to make $\mathcal{R}$ a distance? (More to make you realize is a distance-measure.)

Example: $\mathcal{R}(T_n, \theta) = E(T_n - \theta)^2$ with the expected value taken under f.

Definition: The efficiency of estimate S_n for θ and f, within a class $\mathcal{C}$ of estimates is

$$\frac{\inf\{\mathcal{R}(T_n, \theta); T_n \in \mathcal{C}\}}{\mathcal{R}(S_n, \theta)} = \inf\{\frac{\mathcal{R}(T_n, \theta)}{\mathcal{R}(S_n, \theta)}; T_n \in \mathcal{C}\} \tag{1}$$

It is desired that the efficiency of S_n is near 1 (from below).

The class $\mathcal{C}$: cannot iclude all possible estimates of θ because then estimates which are each a constant number will be in $\mathcal{C}$ and therefore $\inf_n\{\mathcal{R}(T_n, \theta); T_n \in \mathcal{C}\} = 0$.

Example: When θ is a location parameter and S_n is unbiased estimate, then $\mathcal{C}$ can be the class of all unbiased estimates of θ, and under mild conditions on f the Cramer-Rao bound provides

$$\inf\{E(T_n - \theta)^2; T_n \in \mathcal{C}\}$$

to determine efficiency of S_n.

The efficiency of an estimate S_n is usually computed at the assumed underlying distribution f. Unfortunately estimates that have high efficiency for the assumed underlying distribution do not necessarily have high efficiency for the true distribution. This fact motivated several notions of robustness of an estimate, most relying on the asymptotic properties of the estimators (Huber, 1964; Hampel, 1971).

A different notion of robustness, based on the finite sample behavior of estimates, was introduced by Tukey (1987) and Mosteller & Tukey (1977, p. 206). Because it is rare that a real situation can be represented by a single assumed distribution, Tukey proposed to calculate efficiencies at a reasonable collection of distributions called **"corners"**. This has led to a new notion of efficiency.

We are interested in having **high efficiency in a variety of situations rather than in one situation**. Thus, a finite number of alternative densities/distributions is considered, each representing relatively diverse circumstances.

Definition: The **Polyefficiency of** S_n is the infimum of the efficiencies of S_n at a "reasonable" collection of densities/distributions called **"corners"**.

Interest is restricted to estimates with high polyefficiency.

Definition: An estimate that achieves the supremum of the polyefficiencies is called **efficient-robust or polyefficient optimal with respect to the corners**.

Tukey usually considers three corners: the **normal** $N(0, 1)$ distribution; the **slash distribution** obtained by dividing a standard normal random variable with an independent uniform $U(0, 1)$ random variable; **the one-wild distribution** obtained sampling 95% N(0, 1) random variables and 5% N(0, 100) random variables.

Estimates with high polyefficiency have been constructed, under different circumstances, in unpublished Princeton theses by R. Guarino and G. S. Easton among others.

Major criticisms of polyefficiency are the arbitrariness in the choice of the corners, and that nothing is known about interpolation among these corners (Tukey, 1987, p. 5). The idea behind this notion of robustness is that **if an estimate performs well at each of the corners** chosen to represent the extreme kinds of data **it should also perform well at distributions that lie in between them, in a sense not yet specified.**

By enlarging the set of possible corners to be used **in a given situation** the first criticism is relaxed. Thus, a finite number k of corners is adopted, with k as large as desired. Usually, this idea is interpreted by condidering in R convex combinations of the corners densities, $g_1, \ldots, g_k$, defined on R, i.e.

$$a_1 g_1(x) + \ldots + a_k g_k(x), \qquad x \in R, \tag{2}$$

with the conditions

$$\sum_{i=1}^{k} a_i = 1, \qquad a_i \geq 0,\ i = 1, \ldots, k. \tag{3}$$

With a sample of size n the question is whether it is more informative to look at the sample as vector of observations or simply as n observations. In the latter case the joint density of the sample is product of densities (2). Looking at vector $\mathbf{X}_n$ the corners in (2) are each a density in R^n.

Seeing corners as **distributions of the** n **dimensional sample** $\mathbf{X}_n$, it is shown that if an estimate S_n has high polyefficiency it will also have **at least equally high efficiency** over all mixtures of these corners in (2), thus relaxing the second criticism.

The following elementary lemma is used; it can be proved by induction.

Lemma: Let a_i, u_i, v_i be all positive, $i = 1, \ldots, k$.Then,

$$\frac{a_1 u_1 + \ldots + a_k u_k}{a_1 v_1 + \ldots + a_k v_k} \geq \min\{\frac{u_1}{v_1}, \ldots, \frac{u_k}{v_k}\} \tag{4}$$

Assume now that the n**-dimensional random vector** $\mathbf{X}(\in R^n)$ is observed with density

$$f_{\theta,\mathbf{a}}(\mathbf{x}) = a_1 g_{\theta,1}(\mathbf{x}) + ... + a_k g_{\theta,k}(\mathbf{x}), \tag{5}$$

$\underline{\mathbf{x} \in R^n}$, $\mathbf{a} = (a_1, \ldots, a_k)$, with i-th corner density $g_{\theta,i},\ a_i > 0, i = 1, \ldots, k.$

Let $S_n = S(X)$ be the estimate of θ in class $\mathcal{C}$ we compute its polyefficiency. T_n is a generic estimate in class $\mathcal{C}$. $R_{\mathbf{a}}(\theta, T_n)$ is the risk function of T_n with respect to $f_{\theta,\mathbf{a}}$. Observe that if $\mathbf{e}_i$ is the vector with all coordinates $(\mathbf{e}_i)_j = 0, \jmath \neq i$, and the i-th coordinate $(\mathbf{e}_i)_i = 1$, then $R_{\mathbf{e}_i}$ is the Risk with respect to $g_{\theta,i}$ in (5), $i = 1, \ldots, k$. If in (4) we identify u_i with the risk $R_{\mathbf{e}_i}, (\theta, T_n)$ and v_i with the risk $R_{\mathbf{e}_i}(\theta, S_n)$, where $(\mathbf{e}_i)_j = 1 \; if \; j = i$ and 0 otherwise, we immediately have the following.

Proposition. For each $\mathbf{a} \in R^k$ satisfying (3),

$$\inf_{T_n \in \mathcal{C}} \{\frac{R_{\mathbf{a}}(\theta, T_n)}{R_{\mathbf{a}}(\theta, S_n)}\} \geq \min_{1 \leq i \leq k} [\inf_{T_n \in \mathcal{C}} \frac{R_{\mathbf{e}_i}(\theta, T_n)}{R_{\mathbf{e}_i}(\theta, S_n)}]. \tag{6}$$

The proposition can be rephrased: high polyefficiency implies at least as high an efficiency for any element of the smallest convex set containing the corners.

An example can be found showing that for a family of distributions with an infinite number of extreme points, one cannot necessarily find a lower bound on the efficiency of an estimate based on its polyefficiency at a finite number of corners.

REFERENCES

Hampel, F. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42**, 1887-96.

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73-101.

Mosteller, F. and Tukey, J. W. (1977). *Data Analysis and Regression.* Reading, Mass.: Addison-Wesley. Tukey, J. W. (1987). Configural polysampling. *SIAM Rev.* **29**, 1-20.