A Note on Efficiency

By H. Strasser, Gießen

Summary: A concept of efficiency for estimator sequences is defined the spirit of which is Bayesian. This concept states an optimum property within the class of all estimator sequences. Under weak regularity conditions efficiency in the new sense implies several of the classical efficiency concepts which state optimum properties in a smaller class of estimator sequences. It is shown that under certain regularity conditions a sequence of estimators \((T_n)\) is efficient in the new sense up to an error term of order \(0 ((\log n)^{1/2})\) iff

\[
\sqrt{n} (T_n - \theta_n) \rightarrow c_w(\theta) \text{ in } P_\theta^n \text{-probability if } \theta \in \Theta
\]

with accuracy of the order \(0 ((\log n)^{1/2} n^{-1/2})\). Here \((\theta_n)\) denotes the sequence of maximum likelihood estimates and \(c_w(\theta)\) is a constant depending on \(\theta \in \Theta\) and the loss function \(W\). The particular case of symmetric loss functions \(W\) implies \(c_w \equiv 0\). As a side result it follows that the sequence of Bayes estimates \(S^n_{W,\lambda}\) for a regular prior \(\lambda\) is asymptotically normally distributed with mean \(\theta + c (\theta)n^{-1/2}\) and minimal variance.

1. The General Case

Let \((X, A)\) be a measurable space, \(\Theta \subseteq \mathbb{R}\) an open interval and \(\mathcal{P}_\Theta | A, \theta \in \Theta\), a family of probability measures such that \((\theta, A) \rightarrow P_\theta(A)\) is a Borel measurable kernel on \(\Theta \times A\). A probability measure \(\lambda\) on the Borel \(\sigma\)-algebra \(\mathcal{B}\) of \(\Theta\) is called prior distribution. In the following \(W: \mathbb{R} \rightarrow \mathbb{R}^+, W(0) = 0\), denotes a bounded, continuous loss function.

Definition 1: A denotes the class of all prior distributions \(\lambda\) which have a density with respect to the Lebesgue measure and satisfy the following condition: For every \(\theta \in \Theta\) there exists a neighborhood \(U(\theta)\) of \(\theta\) and \(c_\theta > 0\) such that

\[
\left| \frac{p(\sigma')}{p(\sigma)} - 1 \right| < c_\theta \left| \sigma - \sigma' \right| \quad \text{if } \sigma, \sigma' \in U(\theta).
\]

For every \(\lambda \in \Lambda, R_n^\lambda\) denotes the probability measure on \(\mathcal{B} \otimes \mathcal{A}^n\) which is defined by

\[
R_n^\lambda (\Sigma \times A) = \int_{A} P_{\theta}(A) \lambda (d\theta), \Sigma \in \mathcal{B}, A \in \mathcal{A}^n.
\]

1) Prof. Dr. H. Strasser, Math. Inst. Justus-Liebig Universität, Arndtstr. 2, D-6301 Gießen, Germany.
We assume that for every \( \lambda \in \Lambda \) and every \( n \in \mathbb{N} \) there exists a regular version \( R_{n, x}^\lambda \) of the conditional probability \( \Sigma \rightarrow R_{n, x}^\lambda (\Sigma | A_n) (x), x \in X^n, \Sigma \in \mathcal{B} \).

Any sequence of \( \mathcal{A}^n \)-measurable mappings \( T_n : X^n \rightarrow \Theta \) is called a sequence of estimators. We deal with the comparison of estimator sequences in terms of the loss generated by deviations of order \( n^{-1/2} \) of the estimators from the true parameter value. Thus, for any estimator \( T_n \) we define the risk for a loss function \( W \) by

\[
r_{n, W}(T_n, \theta) = \int W(\sqrt{n}(T_n(x) - \theta)) P^n_\theta (dx), \theta \in \Theta.
\]

As is well known it cannot be expected that there exists an estimator sequence \( (T_n) \) which satisfies

\[
\lim_{n \to \infty} r_{n, W}(T_n, \theta) \leq \lim_{n \to \infty} r_{n, W}(S_n, \theta)
\]

for every estimator sequence \( (S_n) \) and every \( \theta \in \Theta \). In this paper, therefore, another optimum property is proposed. To this end we define the posterior risk given \( x \in X^n \) of an estimate \( T_n \) with respect to \( \mathcal{X} \) by

\[
\rho_{n, W}(T_n(x) \mid \lambda, x) = \int W(\sqrt{n}(T_n(x) - \omega)) R_{n, x}^\lambda (d\omega), x \in X^n.
\]

**Definition 2:** A sequence \( (T_n) \) is called efficient for a class \( \mathcal{W}_1 \) of loss functions \( W \) if for every \( \lambda \in \Lambda \), every \( W \in \mathcal{W}_1 \) and every \( \theta \in \Theta \)

\[
\rho_{n, W}(T_n(x) \mid \lambda, x) - \inf_{\tau \in \Theta} \rho_{n, W}(\tau \mid \lambda, x) \xrightarrow{P^n_\theta} 0.
\]

The following remarks may serve as a motivation for our concept of efficiency.

**Remark 1:** It seems to be natural to call a sequence \( (T_n) \) efficient for \( \mathcal{W}_1 \) if for every \( \lambda \in \Lambda \), every \( W \in \mathcal{W}_1 \) and every sequence \( (S_n) \)

\[
\lim_{n \to \infty} \int \left[ r_{n, W}(T_n, \theta) - r_{n, W}(S_n, \theta) \right] \lambda (d\theta) \leq 0.
\]

This condition is equivalent with

\[
\lim_{n \to \infty} \int \left[ \rho_{n, W}(T_n(x) \mid \lambda, x) - \inf_{\tau \in \Theta} \rho_{n, W}(\tau \mid \lambda, x) \right] P^n_\theta (dx) \lambda (d\theta) = 0
\]

for every \( \lambda \in \Lambda \). It follows that (1) \( \Rightarrow \) (2) and thus (1) is a stronger optimum property of \( (T_n) \) than (2).

**Remark 2:** If condition (2) holds for every \( \lambda \in \Lambda \), \( W \in \mathcal{W}_1 \), and every sequence \( (S_n) \), there are implications being related to other approaches to efficiency (cf. the papers of Bahadur, [1960], Wolfowitz [1965], Schmetterer [1966] and Pfanzagl [1970]):

(A) Denote by \( \mathcal{V} \) the class of all sequences \( (S_n) \) such that the sequence of probability