研究ノート

Robust Sequential Analysis for Special Capacities.

by

Hajime Takahashi

Abstract.

Huber type robustness will be considered for some extensions of Wald's Sequential Probability Ratio Test, including Wald's three decision problem and the Kiefer–Weiss formulation. The results of Huber (1965, 1968), Huber and Strassen (1973), Rieder (1977) and Österreicher (1978) will be extended to derive a least favorable tuple in the multiple decision problem. And then the asymptotically least favorable Kiefer–Weiss procedure together with its asymptotic relative efficiency for the ε–contamination and the total variation models will be discussed.

Introduction. It is known that the classical likelihood ratio test is not robust (cf. Huber, 1965). On the other hand, many sequential procedures are highly dependent on the likelihood ratio. There has been some previous work on sequential rank tests (cf. Bradley, Merchant and Wilcoxon, 1966, Berk and Savage, 1968). We shall, however, consider Huber type robustness, so that we may robustify some extensions of Wald's Sequential Probability

Ratio Test (SPRT for short), including Wald's three decision problem (Sobel and Wald, 1949) and the Kiefer–Weiss formulation (Kiefer and Weiss, 1957).

Robust Testing Problem. We briefly review the robust testing problem. Let $(\Omega, \boldsymbol{B})$ be a measurable space and $\boldsymbol{M}$ be the set of all measures on it, endowed with the vague topology. For given constants ε_i, δ_i and $P_i \in \boldsymbol{M}$ such that $0 \leqq \varepsilon_i,\ \delta_i < 1,\ 0 \leqq \varepsilon_i + \delta_i < 1$, and $P_i \neq P_j (i \neq j) i, j = 1, 2, \ldots, k, k \geqq 2$, we define neighborhood $\boldsymbol{P}_i$ of P_i by

(1) $\boldsymbol{P}_i = \{Q(B) \leqq v_i(B),\ \text{for all } B \in \boldsymbol{B}\}$

where

$$v_i(B) = \begin{cases} [(1-\varepsilon_i)P_i(B) + \varepsilon_i + \delta_i] \wedge 1 & B \neq \phi \\ 0 & B = \phi \end{cases}$$

$i = 1, 2, \ldots, k$. Then v_i has all properties of 2–alternating Choquet capacity except for the continuity property (4) in Huber and Strassen (1973), which would require a compact Ω (cf. Rieder, 1977). Note that the neighborhood $\boldsymbol{P}_i$ exhausts many of the interesting models such as the ε–contamination model and the total variation model (Huber, 1964, Huber and Strassen, 1973). Let $r : \boldsymbol{M}^k \to [-\infty, \infty]$, $k \geqq 2$, be a lower semicontinuous function with respect to the canonical topology on $\boldsymbol{M}^k$.

Definition 1. (Österreicher and Kusolitsh, 1975). A tuple $(Q_1^{(r)}, Q_2^{(r)}, \ldots, Q_k^{(r)}) \in \boldsymbol{P}_1 \times \boldsymbol{P}_2 \times \cdots \times \boldsymbol{P}_k$ is called a least favorable tuple (LFT) with respect to r and $\boldsymbol{P}_i (i = 1, 2, \ldots, k)$ *if*

$$r(Q_1^{(r)}, Q_2^{(r)}, \ldots, Q_k^{(r)}) = \sup \{r(Q'_1, Q'_2, \ldots, Q'_k)\}$$

where the supreme is taken for all $Q'_i \in \boldsymbol{P}_i (i = 1, 2, \ldots, k)$.

Österreicher and Kusolitsh (1975) showed the existence of a LFT when $\boldsymbol{P}_i$'s are weakly closed. It is not difficult to extend their result to our $\boldsymbol{P}_i$'s. Moreover lower semicontinuity of r is satisfied by the problem we discuss below. On the other hand the construction of a LFT is not easy to carry out.

In the simplest case when $k=2$, we have several results by Huber (1965, 1968), Huber and Strassen (1973), Rieder (1977), and Österreicher (1978). In this case, in view of Neyman–Pearson lemma, it suffices to consider the following modefication of LFT.

Definition 2. (Huber, 1965, Huber and Strassen, 1973). A pair (Q_1^*, Q_2^*) ε $\boldsymbol{P}_1 \times \boldsymbol{P}_2$ is called a least favorable pair (LFP) if

$$Q_1^*(\pi>t) = \sup \{Q_1(\pi>t) : Q_1 \in \boldsymbol{P}_1\} \quad \text{and}$$

$$Q_2^*(\pi>t) = \inf \{Q_2(\pi>t) : Q_2 \in \boldsymbol{P}_2\} \quad \text{for all } t>0,$$

where π is a version of Generalized Radon–Nikodym derivetive dQ_2^*/dQ_1^*. Existence and construction of LFP is discussed by Huber (1965, 1968), Huber and Strassen (1973), and Rieder (1977). The last two papers observed that π is also a version of dv_2/dv_1, and then considered a Bayes test between v_1 and v_2 with a priori distribution $t/(1+t)$ and $1/(1+t)$, $t \geqq 0$ to construct π. By restricting to a special capacity of the form (1), Rieder (1977) gave a constructive proof in which he derived Huber's LFP. The derivation of (Q_1^*, Q_2) from π is a routine work. Lately Österreicher (1978) gave a completly different method. He first considers a problem of testing P_1 against $\boldsymbol{P}_2'$ where $\boldsymbol{P}_i'$ is is defined by (1) with v_i replaced by $v_i' = (1-\varepsilon_i)P_i + \varepsilon_i + \delta_i$. He constructed a least favorable probability measure Q_2^* ε $\boldsymbol{P}_2'$ by utilizing the convexity of the risk set and that the risk set contains points (0, 1) and(1, 0). Repeating the same argument for $\boldsymbol{P}_1'$ vs Q_2^* to derive Q_1^*, he then showed the pair thus obtained is LFP for $\boldsymbol{P}_1' \times \boldsymbol{P}_2'$.

We attempt to generalize these methods in the multiple decision problem with three states of nature. Our approach will be to consider a test of the form: Accept P_i if and only if

$t_i q_i^* = \max\ t_j q_j^*,\ j=1,2,3,$

where $t_i \geqq 0$, $t_1+t_2+t_3=1$ and q_i^* is a density function of $Q_i^*(i=1,2,3)$. Although the actual calculation leading to the LFT may be very complicated, we expect that an extension of Rieder's method or Österreicher's will provide a formula in reasonably definitive form.

Kiefer and Weiss Formulation. Kiefer and Weiss (1657) considered a version of Weld's SPRT in which there are three states of nature (P_1, P_2, P_3) and they described a sequential procedure which minimizes the expected sample size at P_3 subject to bounds on the error probabilities at P_1 and P_2. Under some regularity conditions (which are satisfied by the one parameter exponential family), their procedure is a Generalized SPRT of P_1 against P_2. Here GSPRT is meant to be a usual SPRT with boundaries which are a function of the sample size. Although there are several critisisms against their approach (cf. Chernoff, 1972, p. 76), it will be reasonable to take their formulation as a starting point for constructing the theory of robust sequential analysis for composite hypothesis.

We shall formulate the robust version of Kiefer–Weiss problem as follows. Assume that the ideal probability measure P_i belongs to a one parameter exponential family with parameters $\theta_i(1,2,3)$ such that $\theta_1<\theta_3<\theta_2$. We then construct a LFT (Q_1^*, Q_2^*, Q_3^*) for a multiple decision problem with three states of nature (P_1, P_2, P_3) as described in the previous section. Our first goal is to prove that Kiefer–Weiss procedure for (Q_1^*, Q_2^*, Q_3^*) is asymptotically least favorable in the sense of Huber (1965); it maximizes $E_{Q_3'}(T^*)$ for all $Q_3' \varepsilon P_3$ as error probabilities go to zero, where T^* is a stopping time for the Kiefer–Weiss procedure for the LFT. We next consider the comparison of $E_{Q_3^*}(T^*)$ and $E_{P_3}(T)$, where T denotes a stopping time of Kiefer

Weiss procdure for (P_1, P_2, P_3), We will define and calculate an asymptotic relative efficiency $E_{Q_3^*}(T^*)/E_{P_3}(T)$ as error probabilities go to zero (or more conveniently, as the cost per observation goes to zero).

Several technical problems, including the question of whether the LFT (Q_1^*, Q_2^*, Q_3^*) satisfies the assumption of Kiefer and Weiss(1957), will be investigated. We anticipate that these assumptions may not hold in general. However, since Huber(1965, 1968) and Rieder(1977) have shown that the density functions of LFP agree with the ideal ones up to the multiplicative constants $(1-\varepsilon_i)$ in the center part and we may choose a suitable version for the tail part which includes the other density function and the ε_i and δ_i 's, we anticipate that we will be able to extend their result.

Other Problems. The LFT may also be applied to construct the robust version of Wald's three deesion problem. One approach is to apply Sobel and Wald's method directly to the LFT (Q_1^*, Q_2^*, Q_3^*) and show that it is least favorable in the sense of Huber. A more difficult but useful problem is to extend to the robust version of Kiefer–Weiss method to the case where nuisance parameters are present. One approach to this problem will be to construct a robust version of the repeated likelihood ratio test with the help of Huber's M–estimator (cf. Huber, 1964).

References.

Berk, R. H. and Savage, I. R. (1968). The information in rank–order and the stopping time of some associated SPRT's. Ann. Math. Statists. 39 1661–1674.

Bradley, R. A., Merchant, S. D. and Wilcoxon, F. (1964). Sequential rank tests. II. Modified two smple procedures. Technometrics 8 615–623.

Chernoff, H. (1972). Sequential analysis and optimal design. SIAM, Philadelphia.

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

——, (1965). A robust version of the probability ratio test. Ann. Math. Statists. 36 1753-1758.

——, (1968). Robust confidence limits. Z. Wahrscheinlichkeitstheorie verw. 10 269-278.

Huber, P. and Strassen, V. (1973). Minimax tests and the Neyman-Pearson lemma for capacities. Ann. Statists. 1 251-263.

Kiefer, J. and Weiss, L. (1957). Some properties of generalized sequential probability ratio test. Ann. Math. Statists. 28 57-74.

Österreicher, F. (1978). On the construction of least favorable pairs of distributions. Z. Wahrscheinlichkeitstheorie verw. 43 49-55.

Österreicher, F. and Kusolitsch, N. (1975). On least favorable tulles for adecision model. Coll. Math. Soc. János Bólyai, Topics in Information Theory, 479-491.

Rieder, H. (1977). Least favorable pairs for special capacities. Ann. Statists. 5 909-921.

Sobel, M. and Wald, A. (1949). A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of normal distribution. Ann. Math. Statists. 20 502-522.