A ROBUST VERSION OF THE PROBABILITY RATIO TEST

By Peter J. Huber

Swiss Federal Institute of Technology, Zurich

1. Introduction and summary. A statistical procedure is called robust, if its performance is insensitive to small deviations of the actual situation from the idealized theoretical model. In particular, a robust procedure should be insensitive to the presence of a few "bad" observations; that is, a small minority of the observations should never be able to override the evidence of the majority. (But at the same time the discordant minority might be a prime source of information for improving the theoretical model!)

The classical probability ratio test is not robust in this sense: a single factor $p_1(x_j)/p_0(x_j)$ equal (or almost equal) to 0 or ∞ may upset the test statistic $T(x) = \prod_1^n p_1(x_j)/p_0(x_j)$. This leads to the conjecture that appropriate robust substitutes to both fixed sample size and sequential probability ratio tests might be obtained by censoring the single factors at some fixed numbers $c' < c''$. Thus, one would replace the test statistic by $T'(x) = \prod_1^n \pi(x_j)$, where $\pi(x_j) = \max(c', \min(c'', p_1(x_j)/p_0(x_j)))$.

The problem of robustly testing a simple hypothesis P_0 against a simple alternative P_1 may be formalized by assuming that the true underlying distribution lies in some neighborhood of either of the idealized model distributions P_0 or P_1.

The present paper exhibits two different types of such neighborhoods for which the above mentioned test, to be called *censored probability ratio test*, is most robust in a well defined minimax sense.

The problem solved here originated through the earlier paper Huber (1964), over the question how to test hypotheses about the mean of contaminated normal distributions.

2. Setup of the problem. Let $(\mathfrak{X}, \mathfrak{A})$ be a measurable space, and let P_0, P_1 be two distinct probability measures on it, having densities p_0, p_1 with respect to some measure μ, e.g. $\mu = P_0 + P_1$. In order to formalize the possibility of unknown small deviations from the idealized models P_i we blow them up to composite hypotheses

$$\mathcal{P}_i = \{Q \mid Q = (1 - \epsilon_i)P_i + \epsilon_i H_i, H_i \in \mathcal{H}\}, \qquad (i = 0, 1),$$

where $0 \leqq \epsilon_i < 1$ are fixed numbers, and $\mathcal{H}$ denotes the class of all probability measures on $(\mathfrak{X}, \mathfrak{A})$. We shall always assume that $\mathcal{P}_0$ and $\mathcal{P}_1$ do not overlap (cf. the remark after Lemma 2 below).

Let ϕ be any test between $\mathcal{P}_0$ and $\mathcal{P}_1$, rejecting $\mathcal{P}_i$ with conditional probability $\phi_i(x)$, given that $x = (x_1, \cdots, x_n)$ has been observed. Assume that a loss $L_i > 0$ is incurred if $\mathcal{P}_i$ is falsely rejected, then the expected loss, or risk, is $R(Q_i', \phi)$

Received 6 April 1965.

$= L_i E_{Q_i'}(\phi_i)$, if Q_i' is the true underlying distribution. Here and in the following Q_i' is used as a generic notation for elements of $\mathcal{P}_i$. We are interested in the following minimax testing problems:

(i) to minimize $\max_{i=0,1} \sup_{Q_i'} R(Q_i', \phi)$;

(ii) to minimize $\sup_{Q_1'} R(Q_1', \phi)$, subject to $\sup_{Q_0'} R(Q_0', \phi) \leqq \alpha$; ["maximin test", see e.g. Lehmann (1959), p. 327];

(iii) the mixed Bayes-minimax problem, to minimize the maximum risk if $\mathcal{P}_i$ is true with known prior probability λ_i, i.e. to minimize

$$\sup_{Q_0', Q_1'} (\lambda_0 R(Q_0', \phi) + \lambda_1 R(Q_1', \phi)).$$

We shall see that there is a pair of distributions $Q_i \in \mathcal{P}_i$ such that for any probability ratio test between Q_0 and Q_1,

$$R(Q_i', \phi) \leqq R(Q_i, \phi), \qquad (i = 0, 1). \tag{1}$$

Now consider the auxiliary problems

(i′) to minimize $\max_{i=0,1} R(Q_i, \phi)$;

(ii′) to minimize $R(Q_1, \phi)$ subject to $R(Q_0, \phi) \leqq \alpha$;

(iii′) to minimize $\lambda_0 R(Q_0, \phi) + \lambda_1 R(Q_1, \phi)$.

It is well known that these auxiliary problems can be solved by probability ratio tests; let ϕ be a probability ratio test solving a particular one of them. In view of (1), the same test then is a solution of the corresponding unprimed problem.

In other words, we have proved that (1) implies the following theorem (compare Lehmann (1959), Chapter 8, for the notion of "least favorable"):

THEOREM 1. *The pair* (Q_0, Q_1) *is least favorable for any of the testing problems* (i), (ii) *or* (iii).

3. The least favorable pair of distributions. Intuitively speaking, if there exist two elements $Q_i \in \mathcal{P}_i$ $(i = 0, 1)$ having the above mentioned properties, then Q_0 should be "as close as possible" to P_1, and similarly with reversed indices. Furthermore, according to the conjecture mentioned in the introduction, the probability ratio of Q_1 and Q_0 should correspond to a censored version of that of P_1 and P_0. This led to the following trial version (which turned out to be successful). Define the Q_i by their densities with respect to μ as follows:

$$\begin{aligned} q_0(x) &= (1 - \epsilon_0) p_0(x) && \text{for } p_1(x)/p_0(x) < c'' \\ &= (1/c'')(1 - \epsilon_0) p_1(x) && \text{for } p_1(x)/p_0(x) \geqq c'', \\ q_1(x) &= (1 - \epsilon_1) p_1(x) && \text{for } p_1(x)/p_0(x) > c' \\ &= c'(1 - \epsilon_1) p_0(x) && \text{for } p_1(x)/p_0(x) \leqq c'. \end{aligned} \tag{2}$$

The numbers $0 \leqq c' < c'' \leqq \infty$ have to be determined such that q_0, q_1 are probability densities, i.e.,

$$\begin{aligned} (1 - \epsilon_0)\{P_0[p_1/p_0 < c''] + (c'')^{-1} P_1[p_1/p_0 \geqq c'']\} &= 1, \\ (1 - \epsilon_1)\{P_1[p_1/p_0 > c'] + c' P_0[p_1/p_0 \leqq c']\} &= 1, \end{aligned} \tag{3}$$

and then one will have $Q_i \varepsilon \mathcal{P}_i$, as one checks easily. By interchanging the roles of P_0 and P_1 and of c' and $(c'')^{-1}$ the two relations (3) are interchanged, so it suffices to discuss the existence of a solution for the second one of them, as follows: Let $c_0 = \text{ess inf}_{[\mu]}\, p_1(x)/p_0(x)$, and let $f(c) = P_1[p_1/p_0 > c] + cP_0[p_1/p_0 \leqq c]$.

LEMMA 1. *The function f is continuous;* $f(c) = 1$ *for* $0 \leqq c \leqq c_0$; $f(c)$ *is strictly increasing for* $c > c_0$, *and tends to* ∞ *as* $c \uparrow \infty$.

PROOF. We have

$$f(c) = 1 + cP_0[p_1/p_0 \leqq c] - P_1[p_1/p_0 \leqq c]$$
$$= 1 + \int_{[p_1/p_0 \leqq c]} (c - p_1/p_0)p_0\, d\mu.$$

Thus

$$f(c + \Delta) - f(c) = \int_{[c<p_1/p_0 \leqq c+\Delta]} (c + \Delta - p_1/p_0)p_0\, d\mu + \Delta \int_{[p_1/p_0 \leqq c]} p_0\, d\mu,$$

hence $0 \leqq f(c + \Delta) - f(c) \leqq \Delta$ for any positive Δ, and continuity and monotonicity follow. If $c \geqq c_1 > c_0$, then $f(c + \Delta) - f(c) \geqq \Delta P_0[p_1/p_0 \leqq c_1] > 0$, thus $f(c)$ is strictly increasing and tends to ∞ as $c \uparrow \infty$.

Lemma 1 implies in particular that the Equations (3) have solutions for all $0 \leqq \epsilon_i < 1$, and that these solutions are unique if $0 < \epsilon_i < 1$. Moreover, if the ϵ_i are sufficiently small, then we have also $c' < c''$, provided $P_0 \neq P_1$.

It follows from (2), with $b = (1 - \epsilon_1)/(1 - \epsilon_0)$, and if $c' < c''$, that

$$
\begin{aligned}
q_1(x)/q_0(x) &= bc' &&\text{for } p_1/p_0 \leqq c' \\
&= bp_1(x)/p_0(x) &&\text{for } c' < p_1/p_0 < c'' \qquad (4)\\
&= bc'' &&\text{for } p_1/p_0 \geqq c''.
\end{aligned}
$$

In other words, the probability ratio $q_1(x)/q_0(x)$ equals b times the probability ratio $p_1(x)/p_0(x)$ censored at the points c' and c''.

LEMMA 2. *For any* $Q_i' \varepsilon \mathcal{P}_i$ $(i = 0, 1)$, *and any real number t, we have*

$$Q_0'[q_1/q_0 < t] \geqq Q_0[q_1/q_0 < t] \geqq Q_1[q_1/q_0 < t] \geqq Q_1'[q_1/q_0 < t],$$

provided $c' < c''$.

PROOF. The lemma is trivially true for $t \leqq bc'$ and for $t > bc''$. Assume $bc' < t \leqq bc''$, and let E be the event $[q_1/q_0 < t]$. Then

$$Q_0'(E) = (1 - \epsilon_0)P_0(E) + \epsilon_0 H_0'(E) \geqq (1 - \epsilon_0)P_0(E) = Q_0(E),$$

and similarly

$$Q_1'(E) = (1 - \epsilon_1)P_1(E) + \epsilon_1 H_1'(E) \leqq (1 - \epsilon_1)P_1(E) + \epsilon_1 = Q_1(E).$$

The middle inequality expresses the well known fact that the power of a probability ratio test never falls below its size [Lehmann (1959) p. 67].

REMARK. If the ϵ_i are sufficiently small, then $Q_0 \neq Q_1$, and the middle inequality in Lemma 2 must be strict for some t. This implies that the $\mathcal{P}_i$ are disjoint for sufficiently small ϵ_i.

Put $\gamma(x) = \log q_1(x)/q_0(x)$. Lemma 2 implies that there exist a random variable v and nondecreasing functions $f_0' \leqq f_0 \leqq f_1 \leqq f_1'$ of v such that the probability distributions of $f_i(v)$ and $f_i'(v)$ coincide with the distributions of $\gamma(x)$ under Q_i and Q_i' respectively [see, e.g., Lehmann (1959), p. 73].

4. Fixed sample size problems. This and the following section are based exclusively on the validity of Lemma 2, and do not use particular properties of the $\mathcal{P}_i$.

It is well known that the testing problems (i'), (ii') and (iii') defined in Section 2 can be solved by probability ratio tests (for (ii'), this is the conclusion of the Neyman-Pearson lemma, (i') is an easy consequence of it, and (iii') is immediate). Now let ϕ be any probability ratio test between Q_0 and Q_1, i.e. a test which rejects $\mathcal{P}_0$, given that $x = (x_1, \cdots, x_n)$ has been observed, with a conditional probability $\phi(x)$ satisfying

$$\begin{aligned} \phi(x) &= 1 \quad \text{if} \quad \gamma_n(x) > K, \\ &= \kappa \quad \text{if} \quad \gamma_n(x) = K, \\ &= 0 \quad \text{if} \quad \gamma_n(x) < K, \end{aligned} \tag{5}$$

where $\gamma_n(x) = \sum_{j=1}^n \gamma(x_j)$; K and $0 \leqq \kappa \leqq 1$ are some numbers.

Then for any losses L_0, L_1, and any $Q_i' \in \mathcal{P}_i$, the risk is

$$\begin{aligned} R(Q_0', \phi) &= L_0\{Q_0'[\gamma_n > K] + \kappa Q_0'[\gamma_n = K]\} \\ &= L_0\{\kappa Q_0'[\gamma_n \geqq K + (1 - \kappa)Q_0'[\gamma_n > K]\} \\ &= L_0\{\kappa P[\textstyle\sum_j f_0'(v_j) \geqq K] + (1 - \kappa)P[\sum_j f_0'(v_j) > K]\} \\ &\leqq L_0\{\kappa P[\textstyle\sum_j f_0(v_j) \geqq K] + (1 - \kappa)P[\sum_j f_0(v_j) > K]\} \\ &= R(Q_0, \phi). \end{aligned}$$

Here, the v_j are independent replicas of the random variable v introduced at the end of the preceding section. Similarly, one shows $R(Q_1', \phi) \leqq R(Q_1, \phi)$, which establishes (1). Thus, the proof of Theorem 1 is complete.

REMARK. The limiting case $c' = c''$ is of particular interest. Assume for simplicity $\epsilon_0 = \epsilon_1$, then (3) yields for the common value $c' = c'' = 1$. Assume furthermore that $P_0(p_1/p_0 = 1) = P_1(p_1/p_0 = 1) = 0$. It is convenient to normalize γ by putting $\gamma'(x) = (\gamma(x) - \log c')/(\log c'' - \log c')$. Then, as $c' \uparrow 1$, $c'' \downarrow 1$, the test statistic $\gamma_n'(x) = \sum_j \gamma'(x_j)$ tends to the number of times the inequality $p_1(x_j) > p_0(x_j)$ holds. So the limiting test is a *sign test*.

5. Sequential tests. Let δ be a sequential probability ratio test of Q_0 against Q_1, terminating as soon as $K' < \sum_{j=1}^n \gamma(x_j) < K''$ is violated for the first time $n = N(x)$. If one replaces the stochastic process $\gamma(x_j)$ by $f_i(v_j)$ and $f_i'(v_j)$ respectively ($i = 0, 1$), one sees that if $\sum f_0'(v_j)$ leaves the interval (K', K'') first at K'', then $\sum f_0(v_j)$ does so even earlier. Therefore, the probabilities of error satisfy the inequality $Q_0'[\gamma_N \geqq K''] \leqq Q_0[\gamma_N \geqq K'']$, and similarly $Q_1'[\gamma_N \leqq K'] \leqq Q_1[\gamma_N \leqq K']$.

Thus, as far as the probabilities of error are concerned, the pair (Q_0, Q_1) is also least favorable in the sequential case.

The behavior of the expected sample sizes is more difficult to determine, and I have only the following partial result: If the probabilities of error tend to zero, then the pair (Q_0, Q_1) is asymptotically least favorable with respect to expected sample size.

PROOF. If the probabilities of error tend to zero, which means that $K' \to -\infty$, $K'' \to +\infty$, then the expected sample sizes behave asymptotically as $E_{Q_0'}(N) \sim K'/E(f_0')$ and $E_{Q_0}(N) \sim K'/E(f_0)$, respectively. We have $f_0' \leqq f_0$; if $E(f_0') = E(f_0)$, the stochastic behavior of our test is the same under both Q_0 and Q_0', and the assertion is trivially true. Thus we may assume $E(f_0') < E(f_0) < 0$. But then $E_{Q_0}(N) \geqq E_{Q_0'}(N)$ for sufficiently large negative values of K', and similarly $E_{Q_1}(N) \geqq E_{Q_1'}(N)$ for sufficiently large positive values of K''.

Notice that δ is not quite a censored probability ratio test, unless $\epsilon_0 = \epsilon_1$, since it has a built-in drift (the factor b in (4)).

6. Monotone likelihood ratio. Assume that $p_\theta(x)$ is a family of densities having monotone likelihood ratio in x [cf. Lehmann (1959), p. 68]. We cannot test the hypothesis $H(\theta \leqq \theta_0)$ against the alternative $K(\theta > \theta_0)$ if we allow arbitrary contamination, because hypothesis and alternative would in general overlap. We can however test $H(\theta \leqq \theta_0)$ against $K(\theta \geqq \theta_1)$, $\theta_0 < \theta_1$. Then it turns out that the least favorable pair of distributions for testing $H(\theta = \theta_0)$ against $K(\theta = \theta_1)$ under contamination is also least favorable for the broader problem of testing $H(\theta \leqq \theta_0)$ against $K(\theta \geqq \theta_1)$ under contamination. The proof generalizes that of Lemma 2. Let $Q_\theta' = (1 - \epsilon_0)P_\theta + \epsilon_0 H'$. For $t \leqq bc'$ and for $t > bc''$, one has $Q_\theta'[q_1/q_0 < t] = 0$ and 1 respectively for all Q_θ'. For $bc' < t \leqq bc''$ and $\theta \leqq \theta_0$, one has (with the aid of Lemma 2, p. 74 of Lehmann (1959))

$$\begin{aligned} Q_\theta'[q_1/q_0 < t] &= (1 - \epsilon_0)P_\theta[q_1/q_0 < t] + \epsilon_0 H'[q_1/q_0 < t] \\ &\geqq (1 - \epsilon_0)P_{\theta_0}[q_1/q_0 < t] \geqq Q_{\theta_0}[q_1/q_0 < t], \end{aligned}$$

and similarly for $\theta \geqq \theta_1$, $Q_\theta'[q_1/q_0 < t] \leqq Q_{\theta_1}[q_1/q_0 < t]$, where Q_{θ_0}, Q_{θ_1} is the least favorable pair for the narrower testing problem. Thus, a variant of Lemma 2 holds, and the conclusions of Sections 4 and 5 generalize.

7. Uncertainty in terms of total variation. Let $\mathcal{P}_i = \{Q \mid \|Q - P_i\| \leqq \epsilon\}$, where $\|\ \|$ denotes total variation.

Now define two probability measures Q_0, Q_1 by their densities as follows:

$$q_0(x) = (1 + c')^{-1}(p_0(x) + p_1(x)), \quad q_1(x) = [c'/(1 + c')](p_0(x) + p_1(x)) \qquad \text{for } p_1(x)/p_0(x) \leqq c',$$

$$(6) \quad q_0(x) = p_0(x), \qquad q_1(x) = p_1(x) \qquad \text{for } c' < p_1(x)/p_0(x) < c'',$$

$$q_0(x) = (1 + c'')^{-1}(p_0(x) + p_1(x)), \quad q_1(x) = [c''/(1 + c'')](p_0(x) + p_1(x)) \qquad \text{for } p_1(x)/p_0(x) \geqq c''.$$

Obviously, the probability ratio $q_1(x)/q_0(x)$ equals the probability ratio $p_1(x)/p_0(x)$ censored at the points c', c''.

The numbers c', c'' have to be determined such that Q_0, Q_1 are probability measures and $\|Q_0 - P_0\| = \|Q_1 - P_1\| = \epsilon$, which is the case if

$$(7) \qquad \int_{[p_1/p_0 \leqq c']} (q_1 - p_1)\, d\mu = \tfrac{1}{2}\epsilon, \quad \int_{[p_1/p_0 \geqq c'']} (q_0 - p_0)\, d\mu = \tfrac{1}{2}\epsilon.$$

Since both conditions (7) are analogous, we shall only investigate the first one. It is convenient to put $k' = c'/(1 + c')$, and to write the first of the conditions (7) as

$$\int_{[p_1 \leqq (p_0+p_1)k']} ((p_0 + p_1)k' - p_1)\, d\mu = \tfrac{1}{2}\epsilon.$$

Define the function

$$g(k) = \int_{[p_1 \leqq (p_0+p_1)k]} ((p_0 + p_1)k - p_1)\, d\mu.$$

Let $\Delta \geqq 0$, then

$$g(k + \Delta) - g(k) = \int_{L_1} ((p_0 + p_1)(k + \Delta) - p_1)\, d\mu + \Delta \int_{L_2} (p_0 + p_1)\, d\mu,$$

where $L_1 = [(p_0 + p_1)k < p_1 \leqq (p_0 + p_1)(k + \Delta)]$ and $L_2 = [p_1 \leqq (p_0 + p_1)k]$. Thus $0 \leqq g(k + \Delta) - g(k) \leqq 2\Delta$, therefore g is continuous and monotone increasing. If we put $k_0 = \text{ess inf}_{[\mu]}\, p_1/(p_0 + p_1)$, the following lemma is immediate.

LEMMA 3. *The function g is continuous*; $g(k) = 0$ *for* $0 \leqq k \leqq k_0$, $g(1) = 1$, *and $g(k)$ is strictly increasing for* $k_0 < k < 1$.

Lemma 3 implies in particular that the Equations (7) have solutions for all $0 \leqq \epsilon < 2$, that these solutions are unique for $\epsilon > 0$, and that $c' < c''$, provided $P_0 \neq P_1$ and ϵ is sufficiently small.

Now one establishes the validity of Lemma 2 as in Section 3. Hence all consequences of Lemma 2 remain valid in the present context, in particular all the results of Sections 4 and 5; also the result of Section 6 holds. The proofs have to be modified as follows, e.g. in the case of Section 6:

For $t \leqq c'$ and for $t > c''$, one has $Q_\theta'[q_1/q_0 < t] = 0$ and 1 respectively for all Q_θ'. For $c' < t \leqq c''$ and $\theta \leqq \theta_0$, one has $Q_\theta'[q_1/q_0 < t] \geqq P_\theta[q_1/q_0 < t] - \tfrac{1}{2}\epsilon \geqq P_{\theta_0}[q_1/q_0 < t] - \tfrac{1}{2}\epsilon = Q_{\theta_0}[q_1/q_0 < t]$, etc.

8. Open problems.

(i) Is it true that the pair (Q_0, Q_1) is least favorable for the expected sample size, if the probabilities of error are smaller than $\tfrac{1}{2}$ (Section 5)?

(ii) What are the minimax tests, if both hypothesis P_0 and alternative P_1 are contaminated by the *same* unknown distribution?

(iii) Instead of the two-decision problem, one might consider the k-decision problem of deciding between $\mathcal{P}_1, \cdots, \mathcal{P}_k$, $k > 2$. Presumably, one would first attack the mixed Bayes-minimax problem, assuming that $\mathcal{P}_i$ is true with *a priori* probability λ_i.

REFERENCES

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

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.