The Annals of Statistics
1974, Vol. 2, No. 1, 199–204

ON SYMMETRIC COMPOUND DECISION RULES FOR DICHOTOMIES

BY J. B. COPAS

University of Essex

When an admissible symmetric compound decision rule is applied to a sequence of simple hypothesis testing problems, the decisions are shown to exactly reflect the ordering of the component likelihood ratios. This leads to a characterization of admissible procedures which is closely related to the method ordinarily used in constructing compound decision rules. The extension to estimation problems is indicated.

1. Introduction. Consider a simple dichotomy hypothesis testing problem of testing $\theta = \pm 1$ on the basis of an observation x whose probability density (or probability function) is $\phi_\theta(x)$. If $t = \pm 1$ is the decision reached, the loss is taken to be zero if $t = \theta$, a if $\theta = +1$ and $t = -1$, and b if $\theta = -1$ and $t = +1$ $(a, b > 0)$. Define the likelihood ratio $z = \phi_{+1}(x)/\phi_{-1}(x)$. This paper discusses the non-sequential compound decision problem formed by n problems of this type. We index the components by the suffix i, $i = 1, 2, \cdots, n$, and denote the vectors of observations, likelihood ratios, parameters and decisions by $\mathbf{x}$, $\mathbf{z}$, $\boldsymbol{\theta}$, and $\mathbf{t}$, respectively. As usual the x_i's are assumed independent and the compound loss is taken to be the arithmetic average of the component losses. General discussions of the compound decision problem can be found in [1], [4] and [6].

A (randomized) compound decision rule δ is characterised by a vector $\boldsymbol{\delta}(\mathbf{x})$ of functions $\delta_i(\mathbf{x})$, where $\delta_i(\mathbf{x})$ is the probability (conditional on $\mathbf{x}$) of reaching the decision $t_i = +1$ in the ith component, $i = 1, 2, \cdots, n$. The rule δ is said to be *symmetric* if for almost all $\mathbf{x}$

$$\boldsymbol{\delta}(q\mathbf{x}) = q(\boldsymbol{\delta}(\mathbf{x})) , \qquad \text{all } q \in Q , \tag{1}$$

where Q is the set of permutations of the integers $1, 2, \cdots, n$, and where $q\mathbf{y}$, for any $q \in Q$ and any n-vector $\mathbf{y}$, is defined by $[q\mathbf{y}]_j = y_k$, $k = q^{-1}j$, $j = 1, 2, \cdots, n$. Additionally it is said to be *simple symmetric* if $\delta_i(\mathbf{x}) = \delta^0(x_i)$ for some function $\delta^0(x)$.

Now define f to be the proportion of $+1$'s in $\boldsymbol{\theta}$, namely $f = f(\boldsymbol{\theta}) = (2n)^{-1} \sum (1 + \theta_i)$. It is clear that the risk $r(\delta, \boldsymbol{\theta})$ of a symmetric rule δ can depend on $\boldsymbol{\theta}$ only through the value of f, and it is shown in [2] and [5] that when this parameter f is given an a priori probability distribution the corresponding symmetric Bayes procedure gives decisions of the form

$$t_i = \operatorname{sgn}(z_i - \lambda(\mathbf{x}_{(i)})) , \qquad i = 1, 2, \cdots, n , \tag{2}$$

Received September 1970; revised February 1973.
AMS 1970 *subject classification.* Primary 62C25.
Key words and phrases. Compound decision problem, admissibility, tests of simple hypotheses, likelihood ratios, monotone likelihood ratio.

(sgn (0) may be arbitrarily randomized between ± 1) where $\mathbf{x}_{(i)} = (x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$ and $\lambda(\mathbf{x}_{(i)})$ is a symmetric function of the $(n-1)$ values x_j, $j \neq i$. An explicit expression for $\lambda(\mathbf{x}_{(i)})$ in terms of the a priori probability distribution is given in those two papers but is not repeated here.

If it happened that the value of f were known, the optimum *simple symmetric* rule would give decisions

$$t_i = \operatorname{sgn}\left(z_i - \frac{b(1-f)}{af}\right), \qquad i = 1, 2, \cdots, n,$$

which suggests the approach adopted in the literature of estimating f by some function $\hat{f}(\mathbf{x})$ and then substituting to give the (non-simple) rule

$$t_i = \operatorname{sgn}\left(z_i - \frac{b(1-\hat{f}(\mathbf{x}))}{a\hat{f}(\mathbf{x})}\right), \qquad i = 1, 2, \cdots, n. \tag{3}$$

For example, if $h(x)$ is an unbiased estimate of θ in the component problem and $\bar{h}(\mathbf{x}) = n^{-1} \sum h(x_i)$, one possible estimate of f is

$$\begin{aligned} \hat{f}(\mathbf{x}) &= 0 & &\text{if } \bar{h}(\mathbf{x}) \leqq -1 \\ &= \tfrac{1}{2}(1 + \bar{h}(\mathbf{x})) & &\text{if } |\bar{h}(\mathbf{x})| < 1 \\ &= 1 & &\text{if } \bar{h}(\mathbf{x}) \geqq 1 \end{aligned} \tag{4}$$

The rule (3) with $\hat{f}(\mathbf{x})$ given by (4) is considered in [2], [3] and [5], but despite the favourable asymptotic properties proved in these papers the authors of [2] and [5] suggest that this rule is inadmissible. We show that every admissible compound decision rule is of the form (3), although the appropriate estimates of f cannot in general be identified with (4). This result was quoted without proof in [1].

2. An ordering property of admissible rules and the possible admissibility of (3). A compound decision rule is said to be *ordered* if, for almost all $\mathbf{x}$,

$$z_i < z_j \quad \text{and} \quad \delta_i(\mathbf{x}) > 0 \quad \text{imply} \quad \delta_j(\mathbf{x}) = 1, \qquad 1 \leqq i, j \leqq n. \tag{5}$$

We prove

THEOREM 1. *If a symmetric compound decision rule is admissible, then it is ordered.*

PROOF. Let δ be a symmetric compound decision rule which is not ordered. Then there exists a set S of values of $\mathbf{x}$ within which the implication in (5) fails, where $P_{\boldsymbol{\theta}}(S) > 0$ for some $\boldsymbol{\theta}$ and where $\mathbf{x} \in S$ implies $q\mathbf{x} \in S$ for every $q \in Q$. We shall construct a symmetric compound decision rule δ' which differs from δ only on S, and which satisfies

$$r(\delta', \boldsymbol{\theta}) \leqq r(\delta, \boldsymbol{\theta}) \tag{6}$$

for all $\boldsymbol{\theta}$, but with strict inequality for at least one value of $\boldsymbol{\theta}$. Let, for any n-vector $\boldsymbol{\xi}$, $A(\boldsymbol{\xi}) = \{\mathbf{x} : \mathbf{x} = q\boldsymbol{\xi}, q \in Q\}$, and denote by $L(\boldsymbol{\theta}, \delta(\mathbf{x}))$ the expected loss of δ conditional on $\mathbf{x}$. Then (6) will follow if we can show that, given any $\boldsymbol{\xi} \in S$,

$$E_{\boldsymbol{\theta}}\{L(\boldsymbol{\theta}, \delta'(\mathbf{x})) - L(\boldsymbol{\theta}, \delta(\mathbf{x})) \mid A(\boldsymbol{\xi})\} \leqq 0 \tag{7}$$

for all $\boldsymbol{\theta}$ satisfying

$$\prod_1^n \phi_{\theta_i}(x_i) > 0 \qquad \text{for some} \quad \mathbf{x} \in A(\boldsymbol{\xi}) . \tag{8}$$

For each $\boldsymbol{\xi} \in S$, the values of $\boldsymbol{\delta}'$ within $A(\boldsymbol{\xi})$ are constructed from those of $\boldsymbol{\delta}$ as follows.

Firstly note that as we are considering symmetric procedures, we may effectively replace $\boldsymbol{\xi}$ by any member of $A(\boldsymbol{\xi})$ in view of (1), and so no generality is lost by supposing that

$$\zeta_1 = \phi_{+1}(\xi_1)/\phi_{-1}(\xi_1) < \zeta_2 = \phi_{+1}(\xi_2)/\phi_{-1}(\xi_2) ,$$

and that for $\mathbf{x} \in A(\boldsymbol{\xi})$,

$$\begin{aligned} \delta_i(\mathbf{x}) &= \alpha && \text{if} \quad x_i = \xi_1 \\ &= 1 - \beta && \text{if} \quad x_i = \xi_2 , \qquad i = 1, 2, \cdots, n , \end{aligned}$$

with $\alpha, \beta > 0$. Allowing for the possibility of ties in the components of $\boldsymbol{\xi}$, suppose that the value of ξ_j occurs $n_j - 1$ times in the remaining components $\xi_3, \cdots, \xi_n$, where $n_j \geqq 1, j = 1, 2$. Choose ε with $0 < \varepsilon \leqq \min(\alpha, \beta)$ and define for $\mathbf{x} \in A(\boldsymbol{\xi})$

$$\begin{aligned} \delta_i'(\mathbf{x}) &= \alpha - n_1^{-1}\varepsilon && \text{if} \quad x_i = \xi_1 \\ &= 1 - \beta + n_2^{-1}\varepsilon && \text{if} \quad x_i = \xi_2 \\ &= \delta_i(\mathbf{x}) && \text{otherwise} , \qquad i = 1, 2, \cdots, n . \end{aligned}$$

Now take any fixed $\boldsymbol{\theta}$ satisfying (8) and define, for $\mathbf{x} \in A(\boldsymbol{\xi})$, $s_j(\mathbf{x})$ to be the number of components i of $\mathbf{x}$ with both $x_i = \xi_j$ and $\theta_i = +1$, where $0 \leqq s_j(\mathbf{x}) \leqq n_j$, $j = 1, 2$. Then when $\mathbf{x} \in A(\boldsymbol{\xi})$

$$\begin{aligned} L(\boldsymbol{\theta}, \delta'(\mathbf{x})) - L(\boldsymbol{\theta}, \delta(\mathbf{x})) &= \frac{\varepsilon(a + b)}{n}\left(\frac{s_1(\mathbf{x})}{n_1} - \frac{s_2(\mathbf{x})}{n_2}\right) \\ &= \frac{\varepsilon(a + b)n'}{nn_1 n_2}\left(s_1(\mathbf{x}) - \frac{n_1 s'(\mathbf{x})}{n'}\right) , \end{aligned} \tag{9}$$

where $n' = n_1 + n_2$ and $s'(\mathbf{x}) = s_1(\mathbf{x}) + s_2(\mathbf{x})$. If, for a given integer s',

$$P_{\boldsymbol{\theta}}(s'(\mathbf{x}) = s' \mid A(\boldsymbol{\xi})) > 0 \tag{10}$$

and $0 < \zeta_1, \zeta_2 < \infty$, it is easy to show that the conditional expectation (under $\boldsymbol{\theta}$) of $s_1(\mathbf{x})$ given that $\mathbf{x} \in A(\boldsymbol{\xi})$ and $s'(\mathbf{x}) = s'$ is equal to

$$\sum lB(l)/\sum B(l) , \tag{11}$$

where

$$B(l) = \binom{s'}{l}\binom{n' - s'}{n_1 - l}\left(\frac{\zeta_1}{\zeta_2}\right)^l , \tag{12}$$

with the summation in (11) taken over all values of l for which (12) is defined. In this case

$$E_{\boldsymbol{\theta}}(s_1(\mathbf{x}) \mid A(\boldsymbol{\xi}), s'(\mathbf{x}) = s') \leqq \frac{n_1 s'}{n'} , \tag{13}$$

since $(\zeta_1/\zeta_2) < 1$ and the right-hand side of (13) is the value (11) would take had ζ_1 and ζ_2 been equal. If (10) holds but $\zeta_1 = 0$ or $\zeta_2 = \infty$, then $s_1(\mathbf{x})$ can only take the values 0 or $s' - n_2$ respectively, in which cases (13) is also true. Thus, as (8) implies that (10) holds for at least one value of s', we have

$$E_{\boldsymbol{\theta}}\left(s_1(\mathbf{x}) - \frac{n_1 s'(\mathbf{x})}{n'} \,\middle|\, A(\boldsymbol{\xi})\right) \leqq 0\,,$$

which, together with (9), implies (7).

Now if s' satisfies (10) and

$$(14) \qquad 0 < s' < n'\,,$$

the inequality in (13) is strict. But to each $\boldsymbol{\xi} \in S$ there corresponds at least one pair $\boldsymbol{\theta}$, s' for which (8), (10) and (14) hold, since the possibility that the likelihood ratios of $\boldsymbol{\xi}$ are either all zero or all infinite is ruled out by the definition of S. Hence, as there are only a finite number of possible values of $\boldsymbol{\theta}$, and S is non-null, there must exist $\boldsymbol{\theta}$ at which the inequality in (6) is strict.

A referee has pointed out that Theorem 1 can also be proved from the expressions for $\lambda(\mathbf{x}_{(i)})$ contained in [2] and [5], but the proof we have given above is constructive and admits generalizations such as that of Theorem 3 below.

We now point out a consequence of the ordering property, and hence a corollary of Theorem 1.

Let δ be an (arbitrary) ordered symmetric compound decision rule, and define

$$C = \{\mathbf{x} : z_1 \geqq z_2 \geqq \cdots \geqq z_n\}\,.$$

Then, by (5), there exists a unique pair of integer-valued functions $0 \leqq c(\mathbf{x}) < d(\mathbf{x}) \leqq n + 1$ such that $\delta_i(\mathbf{x}) = 1$ precisely when $i \leqq c(\mathbf{x})$ and $\delta_i(\mathbf{x}) = 0$ precisely when $i \geqq d(\mathbf{x})$. Define, for $\mathbf{x} \in C$,

$$(15) \qquad \begin{aligned} \lambda^*(\mathbf{x}) &= z_{c(\mathbf{x})+1} && \text{if} \quad c(\mathbf{x}) < d(\mathbf{x}) - 1 \\ &= \tfrac{1}{2}(z_{c(\mathbf{x})} + z_{d(\mathbf{x})}) && \text{if} \quad c(\mathbf{x}) = d(\mathbf{x}) - 1\,, \quad 0 < c(\mathbf{x}) < n \\ &= \infty && \text{if} \quad c(\mathbf{x}) = 0\,, \quad d(\mathbf{x}) = 1 \\ &= 0 && \text{if} \quad c(\mathbf{x}) = n\,, \quad d(\mathbf{x}) = n + 1\,. \end{aligned}$$

Then, for almost all $\mathbf{x} \in C$, the decisions of δ can be written in the form

$$(16) \qquad t_i(\mathbf{x}) = \operatorname{sgn}(z_i - \lambda^*(\mathbf{x}))\,, \qquad i = 1, 2, \cdots, n\,,$$

where sgn (0) indicates that the decision may be randomized.

It is obvious that if $\mathbf{x}', \mathbf{x}'' \in C$ with $q\mathbf{x}' = \mathbf{x}''$ for some $q \in Q$, then $\lambda^*(\mathbf{x}') = \lambda^*(\mathbf{x}'')$. Hence we may extend the definition of $\lambda^*(\mathbf{x})$ to all observations $\mathbf{x}$ by defining $\lambda^*(\mathbf{x}) = \lambda^*(\boldsymbol{\xi})$ where $\boldsymbol{\xi}$ satisfies $\boldsymbol{\xi} = q\mathbf{x}$, $\boldsymbol{\xi} \in C$, $q \in Q$. Thus, using (1) and (16), we have for almost all $\mathbf{x}$

$$(17) \qquad \begin{aligned} t_i(\mathbf{x}) = t_{qi}(q\mathbf{x}) &= \operatorname{sgn}([q\mathbf{z}]_{qi} - \lambda^*(q\mathbf{x})) \\ &= \operatorname{sgn}(z_i - \lambda^*(\mathbf{x}))\,, \qquad i = 1, 2, \cdots, n\,. \end{aligned}$$

Of course, a complete specification of δ requires the particular convention to be adopted when sgn (0) is obtained in (17).

We see that the cut-off $\lambda(\mathbf{x}_{(i)})$ in (2), which in general varies from one component to another, has been replaced in (17) by the single function $\lambda^*(\mathbf{x})$. By writing $\hat{f}(\mathbf{x}) = b(a\lambda^*(\mathbf{x}) + b)^{-1}$, Theorem 1 gives

THEOREM 2. *If δ is an admissible symmetric compound decision rule, it can be expressed in the form* (3) *for some symmetric function $\hat{f}(\mathbf{x})$.*

Note that, for a given ordered symmetric rule, there will in general be other functions besides (15) for which (16) and (17) hold.

EXAMPLE. Suppose that δ decides $t_1 = t_2 = \cdots = t_n = +1$ if $z_1 z_2 \cdots z_n \geqq K$ and $t_1 = t_2 = \cdots = t_n = -1$ if $z_1 z_2 \cdots z_n < K$, where K is a positive constant. This symmetric rule is Bayes against the a priori distribution $P(f = 1) = 1 - P(f = 0) = b(aK + b)^{-1}$. For (17) to be valid in this case, it is necessary that $\lambda^*(\mathbf{x})$ satisfy

$$\lambda^*(\mathbf{x}) \leqq \min(z_1, z_2, \cdots, z_n) \qquad \text{if} \quad z_1 z_2 \cdots z_n \geqq K$$

and

$$\lambda^*(\mathbf{x}) \geqq \max(z_1, z_2, \cdots, z_n) \qquad \text{if} \quad z_1 z_2 \cdots z_n < K\,,$$

and so $\lambda^*(\mathbf{x})$, and hence any choice of $\hat{f}(\mathbf{x})$ in (3), must have a discontinuity at (almost) all points $\mathbf{x}$ satisfying $z_1 z_2 \cdots z_n = K$. It follows from this requirement that, at least when the distribution of z in the component problem is continuous, $\hat{f}(\mathbf{x})$ cannot be of the form (4). Note that one possible choice of $\hat{f}(\mathbf{x})$ is

$$\begin{aligned} \hat{f}(\mathbf{x}) &= 1 \qquad \text{if} \quad z_1 z_2 \cdots z_n \geqq K \\ &= 0 \qquad \text{if} \quad z_1 z_2 \cdots z_n < K \end{aligned}$$

which is simply a Bayes solution to the associated simple hypothesis testing problem of testing $f = 1$ against $f = 0$.

3. Extensions. If a *non-randomized* compound decision rule is ordered, then it operates by maximizing the likelihood conditional on the marginal frequencies with which the possible decisions are made. As such, this concept extends to other component decision problems in which the decision and parameter spaces coincide. To avoid needless redefinition of terms we use the same notation as above but extend it to more general decision problems in the obvious way.

Supposing the decision and parameter spaces to be identical, we say a non-randomized compound decision rule giving decisions $\mathbf{t}(\mathbf{x})$ is *ordered* if, for almost all $\mathbf{x}$, and for $1 \leqq i, j \leqq n$,

$$\phi_u(x_i)\phi_v(x_j) \geqq \phi_v(x_i)\phi_u(x_j) \tag{18}$$

where

$$u = t_i(\mathbf{x})\,, \qquad v = t_j(\mathbf{x})\,.$$

The desirability of this property depends on the particular loss function and family $\phi_\theta(x)$. However, in many problems of interest, the analogue of Theorem 1 remains true, and we illustrate this by the following.

THEOREM 3. *Suppose that the components are estimation problems with continuous observations and squared-error loss, and that the family $\phi_\theta(\mathbf{x})$ has the monotone likelihood ratio property. Then if a symmetric non-randomized compound decision rule is admissable, it is ordered.*

The proof of Theorem 3 is very similar to that of Theorem 1 and is omitted.

Acknowledgment. I am indebted to a referee for helpful comments on an earlier draft of this paper.

REFERENCES

[1] COPAS, J. B. (1969). Compound decisions and empirical Bayes (with discussion). *J. Roy. Statist. Soc. Ser. B* **31** 397–425.

[2] HANNAN, J. F. and ROBBINS, H. (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **26** 37–51.

[3] HANNAN, J. F. and VAN RYZIN, J. R. (1965). Rate of convergence in the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **36** 1743–1752.

[4] MARITZ, J. S. (1970). *Empirical Bayes Methods*. Methuen, London.

[5] ROBBINS, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. Math. Statist. Prob.* Univ. of California Press.

[6] SAMUEL, ESTER (1967). The compound statistical decision problem. *Sankhyā Ser. A* **29** 123–140.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SALFORD
SALFORD MS 4WT
ENGLAND