Admissible Solutions of Finite State Sequence Compound Decision Problems*

by

Stephen B. Vardeman
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #472

September 1975

*This research was supported in part by NSF Grants No.s GP-33677X1 and GP-31123X at Michigan State University.

0. Abstract

A general method of constructing procedures which are both admissible and asymptotically optimal in finite state sequence compound decision problems is suggested and applied to the situation of a two state classification component. When used in an empirical Bayes setting, procedures so constructed are seen to be both admissible and asymptotically optimal.

1. Introduction

We consider a situation in which independent structurally identical decision problems are to be faced serially. Numerous authors have produced procedures for various types of component problems satisfying the classical compound optimality criterion but there has been little study of the finite N properties of such asymptotically optimal rules. Indeed it is possible that in some cases procedures exist with better N problem average risk functions for each N.

In this paper we give a natural notion of admissibility for sequence compound rules and in the case of a finite state component problem, suggest a method of producing procedures which satisfy both the admissibility criterion and the classical asymptotic optimality criterion. The method is applied to the situation where the component problem is a two state classification problem. The proof of the asymptotic optimality of the resulting admissible sequence compound rule is carried out under a smoothness condition on the two possible distributions of the component problem likelihood ratio statistic and depends upon an estimation result of Gilliland, Hannan and Huang (1974), developed in their study of Bayes procedures in non-sequential versions of the compound problem. Finally we note that the suggested method of constructing good sequence compound rules can also produce admissible, asymptotically optimal empirical Bayes procedures.

2. Notation and Generalities

We consider a component decision problem with states $\theta \in \Theta$ indexing distributions P_θ on a sample space $(\mathcal{X}, \mathcal{F})$, possible actions $a \in \mathcal{A}$, loss function $L(\cdot,\cdot)$ and decision rules $d(\cdot)$, measurable functions on $\mathcal{X}$ into $\mathcal{A}$. The risk of a rule $d(\cdot)$ when state θ holds will be denoted by $R(\theta,d) = \int L(\theta,d(x))dP_\theta(x)$

and for a signed measure G on Θ, $R(G,d)$ will abbreviate $\int R(\theta,d)dG(\theta)$. $d_G(\cdot)$ will stand for a Bayes rule versus G in the component problem (that is a $d^\circ(\cdot)$ such that $R(G,d^\circ) = \inf_d R(G,d)$) and $R(G)$ will denote the minimum Bayes risk against G, $R(G,d_G)$.

The problem addressed here is "What are good procedures when one is to face a sequence of independent decision problems, all with the above structure?" Decision rules in such a situation are sequences $\underline{\delta} = (\delta_1,\delta_2,\ldots)$ of measurable functions, $\delta_i(\cdot)$ mapping the first i observations $\underline{X}_i = (X_1,\ldots,X_i)$ into an action a_i to be taken in the ith problem. For a sequence of states $\underline{\theta} = (\theta_1,\theta_2,\ldots)$ and a sequence compound decision rule $\underline{\delta}$, we will denote the average risk of $\underline{\delta}$ through the first N problems when $\underline{\theta}$ holds as

$$R_N(\underline{\theta},\underline{\delta}) = \frac{1}{N}\sum_{i=1}^{N} E\, L(\theta_i,\delta_i(\underline{X}_i))$$

$$= \frac{1}{N}\sum_{i=1}^{N} \int L(\theta_i,\delta_i(\underline{x}_i))dP_{\underline{\theta}_i}(\underline{x}_i)$$

where $\underline{\theta}_i$ denotes $(\theta_i,\ldots,\theta_i)$ and $P_{\underline{\theta}_i} = P_{\theta_1}x\ldots xP_{\theta_i}$, the distribution of $\underline{X}_i$. $R_N(\underline{\theta},\underline{\delta})$ clearly depends on $\underline{\theta}$ only through $\underline{\theta}_N$. For $\underline{G}_N$ a signed measure on Θ^N let $\underline{G}_N^i$ denote the marginal of $\underline{G}_N$ on the first i coordinates of Θ^N. In notation similar to that in the component problem, take as the N problem Bayes risk of the rule $\underline{\delta}$ against $\underline{G}_N$

$$R_N(\underline{G}_N,\underline{\delta}) = \int \frac{1}{N}\sum_{i=1}^{N} \int L(\theta_i,\delta_i(\underline{x}_i))dP_{\underline{\theta}_i}(\underline{x}_i)d\underline{G}_N(\underline{\theta}_N)$$

$$= \frac{1}{N}\sum_{i=1}^{N} \iint L(\theta_i,\delta_i(\underline{x}_i))dP_{\underline{\theta}_i}(\underline{x}_i)d\underline{G}_N^i(\underline{\theta}_i).$$

The classical optimality criterion for a sequence compound procedure is that its N problem risk be asymptotically no larger than the minimum that could be obtained if before facing any decisions one was furnished with E_N, the empiric distribution of states θ_1 through θ_N, and determined to choose a fixed $d(\cdot)$ and in the ith problem take action $d(X_i)$. That is,

Definition 2.1 A sequence compound procedure $\underline{\delta}$ is called s.c. optimal provided

$$\overline{\lim_N} \, (R_N(\underline{\theta},\underline{\delta}) - R(E_N)) \leq 0.$$

As indicated before, taken alone such a definition of optimality is open to criticism on the basis that when considered as a function of $\underline{\theta}_N$, the N problem risk function of an s.c. optimal rule $\underline{\delta}$, $R_N(\underline{\theta},\underline{\delta})$ may well be inadmissible for each N. Hence

Definition 2.2 A sequence compound procedure $\underline{\delta}$ will be called s.c. admissible provided when considered as a function of $\underline{\theta}_N$, $R_N(\underline{\theta},\underline{\delta})$ is admissible for each N.

s.c. admissibility does not imply s.c. optimality. The main result of this paper is the demonstration of sequence compound rules for a two state classification problem component which satisfy both definitions 2.1 and 2.2.

3. Considerations for Finite Θ

For $\Theta = \{1,2,\ldots,m\}$ a method of showing the s.c. admissibility of a procedure $\underline{\delta}^\circ$ would be to for each N produce a distribution $\underline{G}_N$ on Θ^N such that $\underline{G}_N(\underline{\theta}_N) > 0$ for each $\underline{\theta}_N \in \Theta^N$ and such that $\underline{\delta}^\circ$ minimizes $R_N(\underline{G}_N,\cdot)$. For $\underline{\delta}^\circ$ to minimize $R_N(\underline{G}_N,\cdot)$ it is necessary and sufficient that δ_i° minimize

$$(1) \qquad \sum_{\underline{\theta}_i \in \Theta^i} \int L(\theta_i,\delta_i(\underline{x}_i))\, dP_{\underline{\theta}_i}(\underline{x}_i)\, \underline{G}_N^i(\underline{\theta}_i)$$

for each i=1,2,...,N over choices of measurable maps δ_i from $\mathcal{X}^i$ to $\mathcal{A}$. (In the terminology of Gilliland and Hannan (1969) such a δ_i° would be Bayes versus $\underline{G}_N^i$ in a Γ^i decision problem.) Notice that for μ a sigma finite measure dominating $P_1,\ldots,P_m$ and $f = \frac{dP_\theta}{d\mu}$, subject to measurability considerations, the choice of δ_i as

$$(2)\qquad \delta_i(\underline{x}_i) = \begin{cases} \text{an a which minimizes} \\ \sum_{\underline{\theta}_i \in \Theta^i} L(\theta_i,a)\underline{G}_N^i(\underline{\theta}_i) \prod_{j=1}^{i} f_{\theta_j}(x_j) \end{cases}$$

will minimize (1). It is informative to rewrite (2) as

$$(3)\qquad \delta_i(\underline{x}_i) = \begin{cases} \text{an a which minimizes} \\ \sum_{k=1}^{m} f_k(x_i)L(k,a)\Big(\sum_{\underline{\theta}_i \in \Theta^i \ni \theta_i=k} \underline{G}_N^i(\underline{\theta}_i) \prod_{j=1}^{i-1} f_{\theta_j}(x_j)\Big) \end{cases}$$

(interpreting the empty product as 1 in the case i=1), because abbreviating $\sum_{\underline{\theta}_i \in \Theta^i \ni \theta_i=k} \underline{G}_N^i(\underline{\theta}_i) \prod_{j=1}^{i=1} f_{\theta_j}(x_j)$ to $w_{k,i}(\underline{G}_N^i)$, it is then apparent that for fixed $\underline{x}_{i-1}, \delta_i(\underline{x}_i)$ is a component problem Bayes rule against a measure giving mass $w_{k,i}(\underline{G}_N)$ to each state k=1,...,m.

Now a standard method of producing s.c. optimal rules in finite state settings is to at problem i estimate E_{i-1} in some consistent fashion, say by $\hat{E}_{i-1}$ and to take action $d_{\hat{E}_{i-1}}(X_i)$ (see for example Hannan (1956),(1957), Van Ryzin (1966) or Vardeman (1975).) This suggests that to produce a procedure satisfying definitions 2.1 and 2.2 one might search for a sequence of distributions $(G_1,\underline{G}_2,\ldots)$ such that

a) $\underline{G}_i$ is a distribution on Θ^i such that $\underline{G}_i(\underline{\theta}_i) > 0$ for all $\underline{\theta}_i \in \Theta^i$,

b) $\underline{G}_{i-1}$ is the marginal of $\underline{G}_i$ on the first i-1 coordinates of Θ^i,

and c) when normalized the weights $w_{1,i}(\underline{G}_i), \ldots w_{m,i}(\underline{G}_i)$ give a consistent estimate of E_{i-1}.

One might then take δ_i to be of form (3) with $\underline{G}_i$ replacing $\underline{G}_N^i$ and attempt to prove s.c. optimality for the resulting rule. We proceed to carry out such a program for a two state classification component problem.

4. Admissible, Asymptotically Optimal Two State Classification Rules

For this section we specialize to the case where $\Theta = \{0,1\}$ and P_0 and P_1 are distinct probability measures on $(\mathcal{X}, \mathcal{F})$. Let $\mu = P_0 + P_1$, $0 \leq f_0 = \frac{dP_0}{d\mu} \leq 1$, $0 \leq f_1 = \frac{dP_1}{d\mu} \leq 1$ and note that the extended real valued $\rho = \frac{f_1}{f_0}$ is well defined (i.e. not $\frac{0}{0}$) a.e. μ. Take $G = \Theta$ and assume the loss structure

$$L(\theta,a) = \begin{cases} 0 & \text{for } \theta=a \\ L_0 & \text{for } \theta=0 \text{ and } a=1 \\ L_1 & \text{for } \theta=1 \text{ and } a=0 \end{cases}$$

where L_0 and L_1 are positive real numbers. L will abbreviate the ratio $\frac{L_0}{L_1}$.

An intuitively appealing sequence of distributions on $\Theta, \Theta^2, \ldots$, for which there is available the kind of consistency result for the weights $w_{k,i}(\underline{G}_i)$ alluded to in §3, was suggested by Robbins (1951) in his original treatment of the non-sequential compound decision problem. Denote by $\underline{H}_i$ the probability on Θ^i which is symmetric and places equal mass on the subsets of Θ^i defined by $\sum_{j=1}^{i} \theta_j = k$, $k = 0,1,\ldots,i$ (that is for which $\underline{H}_i(\{(\theta_1,\ldots,\theta_i)\}) = ((i+1)\binom{i}{\alpha})^{-1}$ where $\alpha = \sum_{j=1}^{i} \theta_i$). It is a simple exercise to show that $\underline{H}_{i-1}$ is the marginal distribution of $\underline{H}_i$ on the first i-1 coordinates of Θ^i and $\underline{H}_i(\{\underline{\theta}_i\}) > 0$ for each $\underline{\theta}_i \in \Theta^i$. Further, with p_i denoting $\frac{1}{i}\sum_{j=1}^{i} \theta_j$, Δ

abbreviating $\int f_1 d(P_1-P_0)$, and $w_{k,i}$ standing for $\sum_{\underline{\theta}_i \in \Theta^i} \ni \theta_i=k \underline{H}_i(\underline{\theta}_i) \prod_{j=1}^{i-1} f_{\theta_j}(X_j)$ for k=0 and 1, Gilliland, Hannan, and Huang (1974) prove the following.

Proposition 4.1. For $\underline{X}_{i-1}$ distributed as $P_{\underline{\theta}_{i-1}}$

$$E\left|\frac{w_{0,i}}{w_{0,i}+w_{1,i}} - (1-p_{i-1})\right| = E\left|\frac{w_{1,i}}{w_{0,i}+w_{1,i}} - p_{i-1}\right| \leq \left(\frac{4\Delta^{-1}\sqrt{2\pi(i-1)}+1}{i+1}\right)$$

for any $\underline{\theta}_{i-1} \in \Theta^{i-1}$.

Thus with $\hat{p}_{i-1} = w_{1,i}(w_{0,i} + w_{1,i})^{-1}$, the sequence compound procedure $\underline{\phi} = (\phi_1,\phi_2,\ldots)$ defined by

$$\phi_i(\underline{X}_i) = I[\rho(X_i) \geq L \frac{1-\hat{p}_{i-1}}{\hat{p}_{i-1}}]$$

is at least s.c. admissible and the estimation result lends hope of proving uniform s.c. optimality at a good rate. Adding a condition on the possible distributions of $\rho(X)$ we can prove,

Theorem 4.2. For k=0,1 let ν_k denote the distribution of $L(L+\rho(X))^{-1}$ for X with distribution P_k. If there exists a $\gamma \in (0,1]$ and real number C such that for any two real numbers $0 \leq a \leq b \leq 1$, $\nu_k([a,b]) \leq C(b-a)^{\gamma}$, then there exists a real number $\mathcal{K}$ depending only on Δ, $\max(L_0,L_1)$, and C such that

$$R_N(\underline{\theta},\underline{\phi}) - R(E_N) \leq \mathcal{K} N^{-\frac{\gamma}{2}}.$$

Proof. It is standard in proofs of s.c. optimality (see for example Vardeman (1975)) to note that $\frac{1}{N}\sum_{i=1}^{N} R(\theta_i,d_{E_i}) \leq R(E_N)$ so that

$$R_N(\underline{\theta},\underline{\phi}) - R(E_N) \leq \frac{1}{N}\sum_{i=1}^{N} E(L(\theta_i,\phi_i(\underline{X}_i)) - L(\theta_i,d_{E_i}(X_i))). \tag{4}$$

The function $d_{E_i}(x) = I[\rho(x) \geq L(1-p_i)p_i^{-1}]$ is component Bayes versus E_i so that the right side of (4) may be bounded by

$$(5)\quad \frac{1}{N}\{L_0 \sum_{i\ge 2,\theta_i=0} EI[L(1-\hat{p}_{i-1})\hat{p}_{i-1}^{-1} \le \rho(X_i) < L(1-p_i)p_i^{-1}]$$

$$+ L_1 \sum_{i\ge 2,\theta_i=1} EI[L(1-p_i)p_i^{-1} \le \rho(X_i) < (1-\hat{p}_{i-1})\hat{p}_{i-1}^{-1}]\}.$$

But now consider a typical summand above.

$$EI[L(1-\hat{p}_{i-1})\hat{p}_{i-1}^{-1} \le \rho(X_i) < L(1-p_i)p_i^{-1}] = EE[I[p_i < L(L+\rho(X_i))^{-1} \le \hat{p}_{i-1}]\,|\,\underline{X}_{i-1}]$$

$$\le E\ C|\hat{p}_{i-1}-p_i|^{\gamma}$$

by the assumption on the distribution of $\rho(X_i)$. Further for $i > 1$

$$E|\hat{p}_{i-1}-p_i|^{\gamma} \le E(|\hat{p}_{i-1}-p_{i-1}|+|p_i-p_{i-1}|)^{\gamma}$$

$$\le (E|\hat{p}_{i-1}-p_{i-1}|+|p_i-p_{i-1}|)^{\gamma}.$$

So applying proposition 4.2 together with the fact that $|p_i-p_{i-1}| \le (i-1)^{-1}$ for $i > 1$, we have $R_N(\underline{\theta},\underline{\phi})-R(E_N)$ is no larger than

$$C \max(L_0,L_1)\ \frac{1}{N}\left(1 + \sum_{i=2}^{N} ((4\Delta^{-1}\sqrt{2\pi(i-1)}+1)(i+1)^{-1} + (i-1)^{-1})^{\gamma}\right)$$

and the result follows. □

Several comments are in order. The first is that the $\gamma=1$ version of the condition on the distribution of $L(L+\rho(X))^{-1}$ is similar to one used by Hannan and Van Ryzin (1965) in an investigation of a non-sequential compound classification problem and can be verified by showing ν_0 and ν_1 have bounded densities with respect to Lebesque measure. Such is the case for example for P_0 the normal (0,1) distribution and P_1 the normal (β,1) distribution. The second is that proof obviously carries over practically verbatim to any other sequence of distributions $G_1,\underline{G}_2,\ldots$ for which the weights $w_{k,i}(\underline{G}_i)$ are consistent for E_{i-1} at an $i^{-\frac{1}{2}}$ rate. Gilliland, Hannan and Huang (1974) have considered distributions $\underline{G}_i$ defined by $\underline{G}_i(\underline{\theta}_i) = \int t^{\alpha}(1-t)^{i-\alpha}\, d\Lambda(t)$ with $\alpha = \sum_{j=1}^{i}\theta_j$ for a wide class of probabilities Λ on (0,1) and proved analogues of

proposition 4.1 for such $\underline{G}_i$. Note that since for $\underline{\theta}_{i-1} \in \Theta^{i-1}$ with

$\alpha = \sum_{j=1}^{i-1} \theta_j$ we have $\underline{G}_i^{i-1}(\underline{\theta}_{i-1}) = \int (t^{\alpha}(1-t)^{i-\alpha} + t^{\alpha+1}(1-t)^{i-\alpha-1})d\Lambda(t) = \underline{G}_{i-1}(\underline{\theta}_{i-1})$

such sequences of distributions can be used to produce whole classes of s.c. admissible, s.c. optimal classification rules.

5. Admissible Finite State Empirical Bayes Procedures

The attractive sequence compound properties of the kind of procedures discussed in the previous sections carry over to empirical Bayes problems. That is, consider a situation where $\theta_1, \theta_2, \ldots, \theta_{N+1}$ are i.i.d. according to some unknown prior G on Θ, available are observations $(X_1, \ldots, \underline{X}_{N+1}) = \underline{X}_{N+1}$ with conditional distribution $P_{\underline{\theta}_{N+1}}$ and an action a is to be taken and loss $L(\theta_{N+1}, a)$ suffered. Many authors have considered procedures of the form $d_{\hat{G}}(X_{N+1})$ where $\hat{G}$ is an estimate of G based on $\underline{X}_N$ and proved

$$\overline{\lim}\ E\ L(\theta_{N+1}, d_{\hat{G}}(X_{N+1})) - R(G) \leq 0.$$

Such an asymptotic optimality property again does not guarantee admissibility for a fixed N. That is there may well be $\delta(\underline{X}_{N+1})$ for which

$$E[L(\theta_{N+1}, d_{\hat{G}}(X_{N+1})) - L(\theta_{N+1}, \delta(\underline{X}_{N+1})) \,|\, \underline{\theta}_{N+1}] \geq 0$$

for each $\underline{\theta}_{N+1} \in \Theta^{N+1}$ with strict inequality for at least one $\underline{\theta}_{N+1}$. But for finite Θ component problems, carrying out the program described in section 3 can produce rules that are not only s.c. admissible and s.c. optimal but also both admissible and asymptotically optimal in the empirical Bayes problem.

That is, if $G_1, \underline{G}_2, \ldots$ are distributions such that

a) of section 3 holds, $\delta_{N+1}(\underline{X}_{N+1})$ of the form

$$\delta_{N+1}(\underline{X}_{N+1}) = \begin{cases} \text{an a which minimizes} \\ \sum_{k=1}^{m} f_k(X_{N+1}) L(k,a) \left(\sum_{\underline{\theta}_{N+1} \in \Theta^{N+1}} \underline{G}_{N+1}(\underline{\theta}_{N+1}) \prod_{i=1}^{N} f_{\theta_i}(X_i) \right) \end{cases}$$

is at least admissible in the empirical Bayes problem. If in addition c) of section 3 holds, subject to measurability considerations the following lemma can be used to prove asymptotic optimality of δ_{N+1}.

Lemma 5.1. Let $\underline{X}_{N+1}$ have conditional distribution $P_{\underline{\theta}_{N+1}}$ given $\underline{\theta}_{N+1}$, and let $\theta_1,\ldots,\theta_{N+1}$ be i.i.d. according to G on $\Theta = \{1,2,\ldots,m\}$. Suppose $0 \le L(\theta,a)$ and $\exists\, B < \infty$ such that for each d, $\int L(\theta,d(x))dP_\theta(x) \le B$. Then for a random vector $V(\underline{X}_N) = (V_1(\underline{X}_N),\ldots,V_m(\underline{X}_N))$,

$$E\, L(\theta_{N+1},d_{V(\underline{X}_N)}(X_{N+1}))-R(G) \le B \sum_{k=1}^{m} E|G(\{k\})-V_k(\underline{X}_N)|.$$

Proof. Iterating expectations and abbreviating $V(\underline{X}_N)$ to V,

$$(6)\quad E\, L(\theta_{N+1},d_V(X_{N+1}))-R(G)=EE[L(\theta_{N+1},d_V(X_{N+1}))-R(G)\,|\,\underline{X}_N].$$

$$=E(R(G,d_V)-R(G,d_G)).$$

But by the minimizing property of a component Bayes rule the right side of (6) is bounded by

$$\text{r.h.s.}(6) \le E(R(G,d_V)-R(G,d_G)-(R(V,d_V)-R(V,d_G))).$$

$$= \sum_{k=1}^{m} E(G(\{k\})-V_k(\underline{X}_N))(R(k,d_V)-R(k,d_G)).$$

$$\le B\, E \sum_{k=1}^{m} |G(\{k\})-V_k(\underline{X}_N)|. \quad \square$$

For example, returning to the two state classification component of section 4, $\phi_{N+1}(\underline{X}_{N+1})$ is admissible in the empirical Bayes classification problem, and since for fixed $\underline{X}_N$, $\phi_{N+1}(\underline{X}_{N+1})$ is component Bayes versus the prior giving weights $1-\hat{p}_N$ and $\hat{p}_N$ to states 0 and 1 respectively, by the lemma

$$E\, L(\theta_{N+1},\phi_{N+1}(\underline{X}_{N+1}))-R(G) \le 2\max(L_0,L_1)E|G(\{1\})-\hat{p}_N|.$$

But triangulating about $p_N = \frac{1}{N}\sum_{i=1}^{N} \theta_i$, applying proposition 4.1, and the moment inequality

$$E|G(\{1\})-\hat{p}_N| \leq \frac{4\Delta^{-1}\sqrt{2\pi N}+1}{N+2} + \left(\frac{G(\{1\})(1-G(\{1\}))}{N}\right)^{\frac{1}{2}}$$

It is typical of results in this area that the empirical Bayes optimality of $\underline{\phi}$ follows under less stringent assumptions than the s.c. optimality, that is, no regularity of the distribution of $\rho(X)$ is needed.

6. Acknowledgement

Lemma 5.1 is a specialization of an unpublished lemma reducing the question of asymptotic optimality of procedures in an extended version of the finite state empirical Bayes problem to the question of L_1 consistency of estimates of distributions on productsof Θ developed by the author while a student at Michigan State University. The author wishes to thank Professor James Hannan for his help in simplifying that lemma.

References

Gilliland, Dennis C. and Hannan, James F. (1969). On the extended compound decision problem. Ann. Math. Statist. 40 1536-1541.

Gilliland, Dennis C., Hannan, James F. and Huang, J. S. (1974). Asymptotic solutions to the two state component compound decision problem, Bayes versus diffuse priors on proportions. RM-320, Statistics and Probability, Michigan State University.

Hannan, James F. (1956). The dynamical statistical decision problem when the component problem involves a finite number, m, of distributions. (Abstract) Ann. Math. Statist. 27 212.

Hannan, James (1957). Approximation to Bayes risk in repeated play. Contributions to the Theory of Games 3 97-139. Princeton University Press.

Hannan, James F. and Van Ryzin, J. R. (1965). Rates of convergence in the compound decision problem for two completely specified distributions. Ann. Math. Statist. 36 1743-1752.

Robbins, Herbert (1951). Asymptotically subminimax solutions of compound statistical decision problems. Proc. Second Berkeley Symp. Math. Statist. Prob., 131-148. University of California Press.

Van Ryzin, J. R. (1966). The sequential compound decision problem with m x n finite loss matrix. Ann. Math. Statist. 37 954-975.

Vardeman, Stephen B. (1975). $O(N^{-\frac{1}{2}})$ convergence in the finite state restricted risk component sequence compound decision problem. RM-349, Statistics and Probability, Michigan State University.