The Annals of Mathematical Statistics
1970, Vol. 41, No. 5, 1760–1762

A COMMENT ON THE COMPOUND DECISION THEORY

BY ELSE SANDVED

University of Oslo

1. An example by Robbins. The Compound Decision Theory was introduced by Robbins [1] and has been developed particularly by him and Ester Samuel in several papers (see for instance [3]). To get an illustration of the concept, we shall consider the following simple example, given by Robbins in [1].

Let $x_1, \cdots, x_n$ be independent random variables, each normally distributed with variance 1 and with means $\theta_1, \cdots, \theta_n$, respectively, where $\theta_i = +1$ or -1. On the basis of $x_1, \cdots, x_n$ we are to decide, for every i, whether the true value of θ_i is 1 or -1. Let Ω denote the set of all 2^n possible parameter points $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)$ and let $w(\boldsymbol{\theta}', \boldsymbol{\theta}) = n^{-1}$ (no. of i for which $\theta_i' \neq \theta_i$) be the loss involved when the true parameter point is $\boldsymbol{\theta}$ and the decision $(\boldsymbol{\theta} = \boldsymbol{\theta}')$ is taken, $\boldsymbol{\theta} \in \Omega$, $\boldsymbol{\theta}' \in \Omega$.

A simple and reasonable decision rule, when the loss function is as above, seems to be the rule

$$\tilde{R}\text{: estimate } \theta_i \text{ by } \operatorname{sgn}(x_i); \qquad i = 1, \cdots, n.$$

The corresponding risk function $L(\tilde{R}, \boldsymbol{\theta}) = Ew(\boldsymbol{\theta}', \boldsymbol{\theta})$ equals $F(-1) = 0.1587$ for all $\boldsymbol{\theta}$, where F is the cumulative standard normal distribution function. $\tilde{R}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$, and Robbins shows that $\tilde{R}$ is the unique minimax decision rule.

2. The Bayes Case. Suppose that in the example above the θ_i's are independent random variables taking the values 1 and -1 with probabilities p and $1-p$, respectively, where p is known. Let $u(x_i)$ be the conditional probability of estimating θ_i to be 1, given x_i. The corresponding risk

$$p \int f(x-1)(1-u(x))\,dx + (1-p)\int f(x+1)u(x)\,dx,$$

where f is the standard normal density function, is minimized by the rule

$$R_p\text{: estimate } \theta_i \text{ by } \operatorname{sgn}(x_i - \tfrac{1}{2}\ln[(1-p)p^{-1}]); \qquad i = 1, \cdots, n$$

which has the risk

$$h(p) = pF(-1+\tfrac{1}{2}\ln[(1-p)p^{-1}]) + (1-p)F(-1-\tfrac{1}{2}\ln[(1-p)p^{-1}]).$$

$h(p)$ is less than $F(-1)$ for $p \neq 0.5$ and equal to $F(-1)$ for $p = 0.5$, and R_p will therefore be preferable to $\tilde{R}$ in this case, unless $p = 0.5$.

3. The Empirical Bayes Case. If in the Bayes Case above p is unknown, and the n x_i's are used to estimate p, then a decision rule corresponding to R_p, with p substituted by an estimate of p, could be used. This would be an example of an Empirical Bayes Case. See Robbins [2].

Received June 30, 1969; revised March 23, 1970.

4. The Compound Decision Case. Let us denote the problem in Section 1 of the present paper as a Compound Decision Problem if it satisfies the description of Robbins in [1]: *No relation whatever is assumed to hold amongst the unknown parameters* θ_i. Then the frequency p of θ_i's equal to one is completely unknown, but may be estimated by means of $x_1, \cdots, x_n$. The estimator $v = \frac{1}{2}(1+\bar{x})$, where $\bar{x} = n^{-1}\sum x_i$, is unbiased for p. As v can take on values outside [0, 1], it is truncated at 0 and 1, and the resulting estimator

$$\begin{aligned} v' &= 0 && \text{if} \quad v \leqq 0 \\ &= v && \text{if} \quad 0 < v < 1 \\ &= 1 && \text{if} \quad v \geqq 1 \end{aligned}$$

is substituted for p in R_p. Hence one gets the decision rule

$$\begin{aligned} R^*: \quad \text{estimate} \quad \theta_i \text{ by } & -1 && \text{if} \quad \bar{x} \leqq -1; \\ \text{by } & \operatorname{sgn}(x_i - \tfrac{1}{2}\ln[(1-\bar{x})(1+\bar{x})^{-1}]) && \text{if} \ -1 < \bar{x} < 1; \\ \text{by } & \ 1 && \text{if} \quad \bar{x} \geqq 1. \end{aligned}$$

The risk function $h(p, n)$ for R^*, the risk function $h(p) = \lim_{n\to\infty} h(p, n)$ for R_p in Section 2 and the risk function $F(-1) \equiv 0.1587$ for $\tilde{R}$ in Section 1 are compared in Table 1 of [1]. For $p = 0.5$, $h(p, n)$ is always greater than $F(-1)$, though the difference is very small for large n. For any $p \neq 0.5$, $h(p, n)$ is less than $F(-1)$ for large enough n. For p near 0 or 1, $h(p, n)$ is much less than $F(-1)$, at least for n as large as 100. Considering this as an argument for preferring R^* to $\tilde{R}$ will have certain consequences, as will be shown in the next section.

5. A sequence R_1^*, R_2^*, $\cdots$ of rules. Consider first the asymptotic case where we assume that any sequence $\frac{1}{2}n^{-1}(\theta_{i_1}+\theta_{i_2}+\cdots+\theta_{i_n}+n)$, where $i_1 < i_2 < \cdots < i_n$, has a limit as $n \to \infty$. Then the asymptotic risk of R^* is $h(p)$, where $p = \lim_{n\to\infty} \frac{1}{2}n^{-1}(\theta_1+\cdots+\theta_n+n)$. Now it is possible to find a sequence of rules, say R_1^*, $R_2^*, \cdots$, etc., where R_1^* is asymptotically uniformly at least as good as R^*, and where R_{i+1}^* is asymptotically uniformly at least as good as $R_i^*, i = 1, 2, \cdots$, etc. This sequence runs as follows: Denote the original sequence of problems by (θ_1, x_1), $(\theta_2, x_2), \cdots$, etc. Then R_1^* consists in applying R^* separately on the two subsequences of problems $(\theta_1, x_1), (\theta_3, x_3), (\theta_5, x_5), \cdots$ and $(\theta_2, x_2), (\theta_4, x_4), (\theta_6, x_6), \cdots$. Let $p_1 = \frac{1}{2}k^{-1}(\theta_1+\theta_3+\cdots+\theta_{2k-1}+k)$ and $p_2 = \frac{1}{2}k^{-1}(\theta_2+\theta_4+\cdots+\theta_{2k}+k)$. Then $p = \frac{1}{2}(\lim_{k\to\infty} p_1 + \lim_{k\to\infty} p_2)$, and because of the concavity of $h(p)$, the asymptotic risk of R_1^*, namely $\frac{1}{2}(h(\lim p_1)+h(\lim p_2))$, is less than $h(p)$, unless $\lim p_1 = \lim p_2$, in which case $\frac{1}{2}(h(\lim p_1)+h(\lim p_2)) = h(p)$. Hence, if $\lim p_1 \neq \lim p_2$, then R_1^* is asymptotically uniformly better than R^*.

The construction of R_2^*, R_3^*, $\cdots$, etc. is this: The relation between R_{i+1}^* and R_i^* is the same as the relation between R_1^* and R^*, $i = 1, 2, \cdots$, etc. For instance, the construction of $R_2{}^*$ consists in applying the rule R^* separately on the four subsequences of problems

$$\begin{array}{llll}(\theta_1, x_1), & (\theta_5, x_5), & (\theta_9, x_9), \cdots \\ (\theta_2, x_2), & (\theta_6, x_6), & (\theta_{10}, x_{10}), \cdots \\ (\theta_3, x_3), & (\theta_7, x_7), & (\theta_{11}, x_{11}), \cdots \\ (\theta_4, x_4), & (\theta_8, x_8), & (\theta_{12}, x_{12}), \cdots .\end{array}$$

Let us now consider the more interesting case where $n = 2k$ is large but fixed. If the problems are presented to the statistician in *random* order, then this is not a Compound Decision Problem according to the description of Robbins: "No relation whatever is assumed to hold amongst the unknown parameters θ_i", because randomization creates relations between the θ_i's, for instance the relation that $p_1 \approx p_2$ with high probability. Hence this situation, where the θ_i's are presented in random order, should rather be called an Empirical Bayes Case.

Now consider the case where the problems are *not* presented in random order, but according to something else, for instance according to time order. If we do not believe that Nature randomizes the problems for us, then there is no reason why p_1 should be near p_2, and if p_1 and p_2 are not close together, then $R_1{}^*$ is better than R^* (much better if p_1 and p_2 differ considerably), because if $n = 2k$ is large, then k is also large. Continuing this way one gets the conclusion: The same sort of argument for preferring R^* to $\tilde{R}$ applies for preferring R_{i+1}^* to $R_i{}^*$, at least as long as the number of problems in the subsequences is large.

REFERENCES

[1] Robbins, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 131–148.
[2] Robbins, H. (1956). An Empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Prob.* 157–164.
[3] Samuel, Ester (1967). The compound statistical decision problem. *Sankhyā* **29** 123–139.