

ON LEAST FAVORABLE SET FUNCTIONS

by

Olaf Krafft

University of North Carolina  
and

Technische Hochschule Karlsruhe

Institute of Statistics Mimeo Series No. 528

This research was supported by the  
Air Force Office of Scientific  
Research Contract No. AF-AFOSR-760-65

Department of Statistics  
University of North Carolina  
Chapel Hill, N. C.

1. Introduction and Summary.

The concept of a least favorable distribution was introduced by Wald [14]. Lehmann and Stein [12] showed its use for testing a composite hypothesis against a simple alternative. In order to study tests of level  $\alpha$  which minimize the maximum power Lehmann [10] introduced the notion of a pair of least favorable distributions at level  $\alpha$ . In [8] and [9] it is shown that least favorable distributions can be considered as solutions of certain infinite linear programs.

Whereas the existence of an optimal test can be proved under mild conditions which guarantee the weak compactness of the class of all test functions, the problem of the existence of a least favorable distribution at level  $\alpha$  requires quite a few additional considerations. The reason for that consists essentially in the fact that the action space of the statistician contains only two points whilst the parameter space in general contains infinitely many points. The main aim of this paper is to discuss some aspects of this problem. We will restrict ourselves to the case of testing a composite hypothesis  $H$  at level  $\alpha$  against a simple alternative  $K$ . The results can be generalized to the case that  $K$  is also composite and they are similar for the symmetrical version of the test problem described in [8], but those extensions would give no better insight into the problem. Since least favorable distributions can be useful also for non-parametric problems, see for instance Hoeffding [6], we will make as few restrictions on the parameter space as possible.

The basic lemmata for our discussion will be the representation theorems for linear functionals on certain linear topological spaces and a convergence theorem for measures which can be regarded as a generalization

---

1) This research was supported in part by the Air Force Office of Scientific Research Contract No. AF-AFOSR-760-65

of Helly's theorem. All of these lemmata guarantee only the existence of least favorable distributions so that our theorems will give no new methods to construct least favorable distributions.

In section 2 we will introduce the definition of a least favorable set function which is more general than the traditional one in the sense that we do not require the  $\sigma$ -additivity of a least favorable measure. This generalization does not permit the usual interpretation of a least favorable distribution as an optimal strategy of Nature. However, this approach seems to be justified by the fact that the existence of a least favorable set function can be shown under rather mild conditions. (The advantage of such a generalization for stochastic processes is remarked and used also by Dubins and Savage [3]). The principal tool in this section will be the theorem on a separating hyperplane for convex sets. In a corollary we will give conditions under which the least favorable set functions are also  $\sigma$ -additive so that the common conclusions can be derived.

In section 3 we introduce the concept of a least favorable support in order to weaken the condition of the compactness of the parameter space used by Wald [14]. A least favorable support is essentially the set in the parameter space on which a least favorable distribution is concentrated. Under the assumption that there exists a compact least favorable support, a least favorable measure can be shown to be the weak limit of a sequence of measures each of which is concentrated on finitely many points of the parameter space.

The topological and measure theoretic notions used in this paper may be found in the book by Dunford and Schwartz [4].

## 2. Least favorable set functions.

Let  $W = \{w_\theta : \theta \in \Theta\}$  be a class of probability measures on a measurable space  $(X, \mathfrak{B})$ ,  $w_\theta$ , a probability measure on  $(X, \mathfrak{B})$  not belonging to  $W$ . We assume that  $W + \{w_\theta\}$  is dominated by a  $\sigma$ -finite measure  $\mu$ . The corresponding densities with respect to  $\mu$  are denoted by  $p_\theta(x)$  and  $p_{\theta'}(x)$ ; throughout the paper  $p_\theta(x)$  is assumed to be continuous in  $\theta$  (for  $\mu$ -almost all  $x$ ) in the topology which is indicated in the conditions of the respective theorems. By  $\Phi$  we denote the class of all test functions  $\varphi$ :

$$\Phi = \{\varphi(x) : \varphi(x) \text{ } \mathfrak{B} \text{-measurable, } 0 \leq \varphi(x) \leq 1 \text{ for all } x \in X\}.$$

The problem of testing the hypothesis  $H: \theta \in \Theta$  at level  $\alpha$ ,  $\alpha \in (0, 1)$ , against the alternative  $K: \theta = \theta'$  consists in the determination of a function

$$\varphi^* \in \Phi_{\alpha, H}$$

$$\Phi_{\alpha, H} = \{\varphi(x) : \varphi(x) \in \Phi, E_\theta \varphi(X) \leq \alpha \text{ for all } \theta \in \Theta\},$$

such that  $E_{\theta'} \varphi^*(X) \geq E_\theta \varphi(X)$  for all  $\varphi \in \Phi_{\alpha, H}$ .

Let  $\mathfrak{H}$  be a  $\sigma$ -field over  $\Theta$ ,  $\lambda$  a probability measure on  $(\Theta, \mathfrak{H})$  and assume  $p_\theta(x)$  to be  $\mathfrak{H} \times \mathfrak{B}$ -measurable. If  $\varphi_\lambda^*$  denotes an optimal test for testing the simple hypothesis  $H_\lambda$  that the true density is  $p_\lambda = \int_\Theta p_\theta(x) d\lambda(\theta)$  against  $K$  at level  $\alpha$ , then  $\lambda^*$  is called a least favorable distribution at level  $\alpha$  iff

$$(1) \quad E_{\theta'} \varphi_\lambda^*(X) \leq E_\theta \varphi_\lambda^*(X) \text{ for all } \lambda.$$

A measure  $\lambda^*$  which is defined in this way has a direct statistical interpretation. For mathematical purposes, however, it seems to be adequate to use a more general definition. This will be done by departing from the requirement that  $\lambda$  is  $\sigma$ -additive:

Definition 1. Let  $\Theta$  be a topological space and  $\mathfrak{H}$  be a field on  $\Theta$  such that  $p_\theta(x)$  is  $\mathfrak{H} \times \mathfrak{B}$ -measurable. By  $ba(\mathfrak{H})$  we denote the class of all bounded additive set functions on  $\mathfrak{H}$ . Then an  $l^* \in ba(\mathfrak{H})$ ,  $l^* \geq 0$ , is called a least favorable set function (l. f. s. f.) for testing  $H$  at level  $\alpha$  against  $K$

iff for the optimal test  $\varphi^*$  and  $l^*$  the following conditions are satisfied:

$$(i) \quad E_{\theta} \varphi^* = \alpha \quad l^* \text{-a.e.},$$

$$(ii) \quad E_{\theta'} \varphi^* - \int_{\Theta} E_{\theta} \varphi^* d l^* \geq E_{\theta'} \varphi - \int_{\Theta} E_{\theta} \varphi d l^* \text{ for all } \varphi \in \mathcal{F}.$$

To show that when  $\mathcal{M}$  is a  $\sigma$ -field and  $l^*$  is  $\sigma$ -additive definition 1 is equivalent with that given in connection with formula (1) we will use the following

Lemma 1. Necessary and sufficient for that  $\varphi^* \in \mathcal{F}_{\alpha, H}$  is optimal and  $\lambda^*$  is a least favorable distribution is the existence of a number  $c > 0$  such that

$$(2) \quad E_{\theta} \varphi^* = \alpha \quad \lambda^* \text{- a.e.},$$

$$(3) \quad \varphi^*(x) = \begin{cases} 1, & p_{\theta'}(x) > c \int_{\Theta} p_{\theta}(x) d \lambda^* \\ \mu\text{-a.e.}, & \\ 0, & p_{\theta'}(x) < c \int_{\Theta} p_{\theta}(x) d \lambda^* . \end{cases}$$

The proof of this lemma which is an extension of corollary 5 by Lehmann [10], p. 92, is based on the fact that a least favorable distribution can be shown to be equal to  $\tilde{\lambda} = \lambda^* / \lambda^*(\Theta)$ , where  $\lambda^*$  is that finite measure  $\lambda$  for which the function

$$(4) \quad f(\lambda) = \alpha \lambda(\Theta) + \int_{\mathcal{F}} (p_{\theta'}(x) - \int_{\Theta} p_{\theta}(x) d \lambda)^+ d \mu$$

achieves its minimum. The statement of lemma 1 then follows from the equality

$$(5) \quad E_{\theta'} \varphi^*(x) = \inf f(\lambda)$$

where the infimum is to be taken with respect to all finite measures  $\lambda$  which are concentrated on finitely many points of  $\Theta$ . A proof of lemma 1 and of formula (5) which will be used also in the proof of theorem 3 may be found in [9].

Theorem 1. Let  $\mathcal{M}$  be a  $\sigma$ -field and the l.f.s.f  $l^*$  be non-degenerate and  $\sigma$ -additive. Then  $\lambda^* = l^* / l^*(\Theta)$  is a least favorable distribution as defined in connection with (1). If, conversely,  $\lambda^*$  is a least favorable distribution and  $c$  is the constant of lemma 1 then  $l^* = c\lambda^*$  will satisfy definition 1.

Proof. The proof follows immediately from lemma 1 observing that because of the  $\sigma$ -additivity of  $l^*$  condition (ii) in definition 1 becomes

$$(6) \quad \int_{\mathbb{X}} (p_{\theta}(x) - \int_{\Theta} p_{\theta}(x) dl^*) \varphi^*(x) d\mu = \max_{\varphi \in \Phi} ,$$

so that  $\varphi^*(x)$  satisfies (3) with  $c = l^*(\Theta)$ .

Conditions (i) and (ii) of our definition of a l.f.s.f. are closely related to formulae used in the proof of the generalized fundamental lemma by Dantzig and Wald [2], the numbers  $k_i$  given there correspond to the values of  $l^*$ . We will prove the existence of a l.f.s.f. along the lines of that proof and of the proof of theorem 1 by Hurwicz and Uzawa [7].

Theorem 2. Let  $\Theta$  be a closed subspace of a normal topological space  $T$ ,  $\mathcal{M}$  be the field generated by the closed sets of  $\Theta$  and  $p_{\theta}(x)$  be  $\mathcal{M} \times \Theta$ -measurable. Then there exists a regular l.f.s.f.

Proof. Since a closed subspace of a normal space is itself normal  $\Theta$  can be assumed to be a normal topological space. Let  $C(\Theta)$  be the space of all bounded continuous real functions on  $\Theta$  with norm  $\|f\| = \sup_{\theta \in \Theta} |f(\theta)|$ . Let  $\varphi^*$  be an optimal test for testing  $H$  at level  $\alpha$  against  $K$  and the subsets  $P$ ,  $Q$  and  $R$  of the Cartesian product  $R^1 \times C(\Theta)$  of  $R^1$  and  $C(\Theta)$  be defined as

$$P = \{ (a, h(\theta)) : a \leq E_{\theta} \varphi, h(\theta) \leq \alpha - E_{\theta} \varphi \text{ for some } \varphi \in \Phi \},$$

$$Q = \{ (a, h(\theta)) : a > E_{\theta} \varphi^*, h(\theta) \geq 0 \text{ for all } \theta \in \Theta \},$$

$$R = \{ (a, h(\theta)) : a \geq E_{\theta} \varphi^*, h(\theta) \geq 0 \text{ for all } \theta \in \Theta \}.$$

If  $(a_1, h_1)$  and  $(a_2, h_2)$  are two points in  $P$  with

$$a_1 \leq E_{\theta, \varphi_1}, \quad h_1(\theta) \leq \alpha - E_{\theta} \varphi_1,$$

$$a_2 \leq E_{\theta, \varphi_2}, \quad h_2(\theta) \leq \alpha - E_{\theta} \varphi_2,$$

the point  $(\rho a_1 + (1-\rho) a_2, \rho h_1 + (1-\rho) h_2)$ ,  $0 \leq \rho \leq 1$ , will be in  $P$  since  $\rho \varphi_1 + (1-\rho) \varphi_2 \in \Phi$ . Hence  $P$  - and similarly  $Q$  and  $R$  - is a convex set in  $R^1 \times C(\Theta)$ . Moreover  $P$  has a non-empty interior: It contains for example the point  $(a = -1, h(\theta) \equiv -1)$  as an interior point. Let  $(a'', h'')$  be a point of  $Q$  and assume  $(a'', h'') \in P$ . Then there will be a  $\varphi'' \in \Phi$  such that

$$a'' \leq E_{\theta, \varphi''} \text{ and } h''(\theta) \leq \alpha - E_{\theta} \varphi''.$$

Since  $h''(\theta) \geq 0$  for all  $\theta \in \Theta$  it follows that  $\varphi''$  is of level  $\alpha$  and hence its power does not exceed the power of  $\varphi^*$ :  $E_{\theta, \varphi''} \leq E_{\theta} \varphi^*$ . The last inequality implies  $a'' \leq E_{\theta, \varphi^*}$  so that  $(a'', h'')$  would not be an element of  $Q$ . This proves that  $P$  and  $Q$  are disjoint sets. Therefore  $P$  and  $Q$  satisfy the conditions of theorem V. 2.8 in [4] which states that  $P$  and  $Q$  can be separated by a non-zero continuous linear functional. According to Theorem IV.6.2 in [4] the dual space of  $C(\Theta)$  can be represented by the space  $rba(\Theta)$ , i. e. by the space of all regular bounded additive set functions on  $\mathcal{U}$ . Hence there exist numbers  $b$  and  $\gamma$  and  $anl^* \in rba(\Theta)$  such that

$$(7) \quad ba' + \int_{\Theta} h' dl^* \leq \gamma \leq ba'' + \int_{\Theta} h'' dl^*$$

holds for all  $(a', h') \in P$  and  $(a'', h'') \in Q$ . The continuity of  $(b, l^*)$  implies that

$$(8) \quad \gamma \leq ba'' + \int_{\Theta} h'' dl^* \quad \text{for all } (a'', h'') \in \bar{Q},$$

$\bar{Q}$  denoting the closed hull of  $Q$ . So (8) will hold for all  $(a'', h'') \in R$ .

Since the point  $(a^* = E_{\theta, \varphi^*}, h^*(\theta) = \alpha - E_{\theta, \varphi^*})$  lies in the intersection of  $P$  and  $R$  the number  $\gamma$  will satisfy

$$(9) \quad \gamma = b E_{\theta, \varphi^*} + \int_{\Theta} (\alpha - E_{\theta, \varphi^*}) d\mathbb{1}^* .$$

The left side of (7) shows that  $\mathbb{1}^*$  is non-negative since otherwise for a sufficiently small  $h'$  we would have  $\int_{\Theta} h' d\mathbb{1}^* > \gamma$ . Similarly  $b$  can be shown to be non-negative. The same inequality (7) can be used to show that  $\mathbb{1}^*$  is non-degenerate if  $E_{\theta, \varphi^*} < 1$ . (If  $E_{\theta, \varphi^*} = 1$  the conditions of definition 1 are trivially satisfied with  $\mathbb{1}^* \equiv 0$ .) For in the case that  $\mathbb{1}^*$  would be degenerated  $b$  would be positive and by choosing  $a' = 1$  it would follow that  $1 \leq E_{\theta, \varphi^*}$ . Assume now  $b$  would be equal to zero. It would follow that  $\mathbb{1}^*$  is non-degenerate and by choosing  $h'(\theta) \equiv \alpha$ ,  $h''(\theta) \equiv 0$  formula(7) would imply

$$\alpha \mathbb{1}^*(\Theta) \leq 0 ,$$

which contradicts  $\alpha \in (0, 1)$ .

Observing that for all  $\varphi \in \Phi$  the points  $(a = E_{\theta, \varphi}, h(\theta) = \alpha - E_{\theta, \varphi})$  are elements of  $P$  and dividing by  $b$  we get from (7) and (9)

$$(10) \quad E_{\theta, \varphi} + \int_{\Theta} (\alpha - E_{\theta, \varphi}) d\tilde{\mathbb{1}}^* \leq E_{\theta, \varphi^*} + \int_{\Theta} (\alpha - E_{\theta, \varphi^*}) d\tilde{\mathbb{1}}^* \text{ for all } \varphi \in \Phi,$$

where  $\tilde{\mathbb{1}}^*$  is defined by  $\tilde{\mathbb{1}}^* = \mathbb{1}^*/b$ . Hence  $\tilde{\mathbb{1}}^*$  satisfies condition (ii) of definition 1. It also satisfies condition (i) since for  $(a'', h'') = (E_{\theta, \varphi^*}, 0)$  we get from (8) and (9)

$$b E_{\theta, \varphi^*} = \int_{\Theta} (\alpha - E_{\theta, \varphi^*}) d\mathbb{1}^* \leq b E_{\theta, \varphi^*}$$

so that

$$\int_{\Theta} (\alpha - E_{\theta, \varphi^*}) d\tilde{\mathbb{1}}^* \leq 0 .$$

(i) follows from the fact that  $\varphi^*$  is of level  $\alpha$ .



In the case that there exists a l.f.s.f. which is  $\sigma$ -additive, lemma 1 shows that the structure of an optimal test  $\phi^*$  can be described by means of a l.f.s.f. A test of the form (3) will be called a likelihood ratio test with respect to  $l^*$  where  $l^* = c\lambda^*$ . In general the knowledge of a l.f.s.f.  $l^*$  does not permit to decide whether an optimal test is a likelihood ratio test w.r.t.  $l^*$  or not. The reason for that is that Fubini's theorem can not be generalized to the case when one of the involved set functions is not  $\sigma$ -additive, see for example Yosida and Hewitt [16]. However, under certain additional conditions on  $\Theta$  and  $p_\theta(x)$  it is possible to represent the dual space of  $C(\Theta)$  by the space of the  $\sigma$ -additive set functions. Some of these conditions are given in the following corollary. There the subsets  $\mathcal{N}(\Theta)$  and  $\mathcal{E}(\Theta)$  of  $C(\Theta)$  are defined respectively as the set of all  $h(\theta) \in C(\Theta)$  which vanish outside compact subsets of  $\Theta$  and as the set of all  $h(\theta) \in C(\Theta)$  which tend to zero at infinity.

Corollary 1. Let  $\mathcal{H}$  be the  $\sigma$ -field generated by the closed sets of  $\Theta$  and  $p_\theta(x)$  be  $\mathcal{H} \times \mathcal{B}$ -measurable. If  $\Theta$  is a closed subspace of either

- (a) a compact Hausdorff space and  $p_\theta(x) \in C(\Theta)$  for  $\mu$ -almost all  $x \in \mathcal{X}$ , or
- (b) a locally compact space and  $p_\theta(x) \in \mathcal{N}(\Theta)$  for  $\mu$ -almost all  $x \in \mathcal{X}$ , or
- (c) a locally compact space and  $p_\theta(x) \in \mathcal{E}(\Theta)$  for  $\mu$ -almost all  $x \in \mathcal{X}$ ,

then there exists a regular  $\sigma$ -additive l.f.s.f.

Proof. The statement follows from the proof of theorem 2 and

- in case (a) from theorem IV.6.3 in [4],
- in case (b) from theorem D in [5], sec. 56,
- in case (c) from exercise 9 in [1], p. 67.

Remark: Part (c) of corollary 1 is related to assumption B used by Lehmann [11]. Concerning our condition that  $\Theta$  should be a closed set the author was informed by Prof. Lehmann that such a condition is necessary in his paper too.

Under any of the conditions of corollary 1 the optimal test will be a likelihood ratio test w.r.t. the corresponding l.f.s.f. If none of the conditions holds an optimal test is at least approximable by a likelihood ratio test w.r.t. a least favorable measure which is concentrated on finitely many points of  $\Theta$  as it is shown in the following

Theorem 3. Let  $\varphi^*$  be an optimal test for testing H at level  $\alpha$  against K. Then for every  $\epsilon > 0$  there exists a measure  $\lambda'$  on  $(\Theta, \mathcal{H})$  which is concentrated on finitely many points of  $\Theta$  and a test  $\varphi'$ ,

$$(11) \quad \varphi'(x) = \begin{cases} 1, & p_{\theta_1}(x) > \int_{\Theta} p_{\theta}(x) d\lambda' \\ 0, & p_{\theta_1}(x) < \int_{\Theta} p_{\theta}(x) d\lambda' \end{cases},$$

$$(12) \quad \int_{\Theta} E_{\theta} \varphi' d\lambda' = \alpha \lambda'(\Theta),$$

such that

$$(13) \quad E_{\theta_1} \varphi' - E_{\theta_1} \varphi^* \leq \epsilon.$$

Proof. Formula (5) shows that for every  $\epsilon > 0$  we can find a  $\lambda''$  which is concentrated on finitely many points of  $\Theta$ , say on  $\{\theta_1, \theta_2, \dots, \theta_n\}$ , such that

$$(14) \quad f(\lambda'') - E_{\theta_1} \varphi^* \leq \epsilon.$$

Since  $\{\theta_1, \theta_2, \dots, \theta_n\}$  is a compact set in  $\Theta$  we can apply corollary 1 for testing  $H': \theta \in \{\theta_1, \theta_2, \dots, \theta_n\}$  at level  $\alpha$  against K. Let  $\varphi'$  be an optimal test and  $\lambda'$  a l.f.s.f. for  $H'$  against K. From (5) and (14) we get

$$E_{\theta_1} \varphi' - E_{\theta_1} \varphi^* \leq f(\lambda'') - E_{\theta_1} \varphi^* \leq \epsilon$$

so that (13) is satisfied. The equalities (11) and (12) follow from lemma 1.

### 3. Least favorable supports.

In this section we will discuss one important case for which a least favorable distribution exists but where in general the conditions of the corollary in section 2 are not satisfied. It is the case in which a uniformly most powerful test turns out to be a test for two simple hypotheses. That will be true for example if the class  $W$  is such that the corresponding densities have monotone likelihood ratio and one-sided hypotheses are considered. As a simple example take the following:

Let  $W$  be the class of normal distributions with known mean  $a_0$  and unknown variance  $\sigma^2$ ,  $\mu$  the Lebesgue measure,  $H: \sigma^2 \leq \sigma_0^2$ ,  $K: \sigma^2 = \sigma_1^2, \sigma_0^2 < \sigma_1^2$ .

Here a least favorable distribution exists (it is concentrated on the point  $\sigma_0^2$ ) but none of the conditions of the corollary in section 2 is satisfied. To study problems of that kind the following definition may be useful:

Definition 2. Let  $\mathcal{H}$  be a  $\sigma$ -field over  $\Theta$ . An  $G \in \mathcal{H}$  is called a least favorable support iff for every  $\theta \notin G$  there is a  $\tilde{\theta} \in G$  such that

$$(15) \quad E_{\tilde{\theta}} \varphi_G^*(X) \geq E_{\theta} \varphi_G^*(X),$$

where  $\varphi_G^*(x)$  denotes an optimal test for testing  $H_G: \theta \in G$  at level  $\alpha$  against  $K$ .

Definition 2 is suggested by the following consideration: It can be shown that there exists always an optimal test  $\varphi^*$  such that

$$(16) \quad \sup_{\theta \in \Theta} E_{\theta} \varphi^*(X) = \alpha$$

otherwise the test

$$\tilde{\varphi}^*(x) = \{ (1 - \alpha) \varphi^*(x) + \alpha - \sup_{\theta \in \Theta} E_{\theta} \varphi^*(x) \} / (1 - \sup_{\theta \in \Theta} E_{\theta} \varphi^*(x))$$

would satisfy (16) and would have at least as much power as  $\varphi^*$ . Because of (2) no least favorable distribution would exist if

$$(17) \quad E_{\theta} \varphi^* (X) < \alpha \text{ for all } \theta \in \Theta.$$

By definition 2 we have characterized a set  $G$  for which the equality

$$E_{\theta} \varphi^* (X) = \alpha \text{ holds as it is seen by}$$

Lemma 2. If  $G$  is a least favorable support then the test  $\varphi_G^*$  is also optimal for  $H: \theta \in \Theta$  at level  $\alpha$  against  $K$ .

Proof. Since  $\varphi_G^*$  is at level  $\alpha$  for  $H_G: \theta \in G$  we have

$$(18) \quad E_{\theta} \varphi_G^* (X) \leq \alpha \text{ for all } \theta \in G.$$

Condition (15) implies that (18) holds for all  $\theta \in \Theta$ . Since  $\Phi_{\alpha, H}$  is a

subset of  $\Phi_{\alpha, H_G}$  and  $\varphi_G^*$  is optimal in  $\Phi_{\alpha, H_G}$ ,  $\varphi_G^*$  will be optimal also in  $\Phi_{\alpha, H}$ .

Examples of least favorable supports are for instance the sets on which the least favorable distributions derived in [10], 3.8, 3.9 and 8.2 or in [6] are concentrated. The following simple example shows that a least favorable support may not be determined uniquely.

Let  $W$  be the class of distributions which have densities  $p_{\theta}(x) = \theta^x (1-\theta)^{1-x}$  with respect to the counting measure  $\mu$ ,  $\Theta = [0, 1]$ ,  $H: \theta \in \Theta = (0, 1/2) \cup (1/2, 1)$ ,  $K: \theta = 1/2$ . Here every  $G(\eta) = \{\eta, 1-\eta\}$ ,  $\eta \in \Theta$ , will be a least favorable support since  $\varphi_{G(\eta)}^*$  will be identically equal to  $\alpha$  for every  $\eta \in \Theta$ .

An interesting example for that (17) holds and hence no least favorable distribution exists is given by Lehmann [11], p. 413; another typical case is discussed in [15], example 2.23.

In the proof of the existence theorem of this section we will use the following

Lemma 3. Let  $\{f_n(x)\}$  be a sequence of  $\mu$ -integrable functions which converges pointwise  $\mu$ -a.e to the  $\mu$ -integrable function  $f_0(x)$ . Then it follows that

$$\lim_{n \rightarrow \infty} \int_{S_n} f_o(x) d\mu = \int_{S_o} f_o(x) d\mu,$$

where the sets  $S_n$  and  $S_o$  are defined by

$$S_n = \{x: f_n(x) > 0\} \text{ and } S_o = \{x: f_o(x) > 0\}.$$

Proof. Let  $T$  be defined by  $T = \{x: f_o(x) = 0\}$  and let  $I_S(x)$  be the characteristic function of a set  $S$ . If  $x_o \in \limsup S_n$ , i.e. if  $f_n(x_o) > 0$  for infinitely many  $n$ , it follows that  $\lim f_n(x_o) = f_o(x_o) \geq 0$ .

Hence we have

$$(19) \quad \limsup S_n \subset S_o \cup T.$$

By similar arguments we can show that

$$(20) \quad \liminf S_n \supset S_o$$

and that for  $x_o \notin S_o \cup T$

$$(21) \quad \liminf I_{S_n}(x_o) = \limsup I_{S_n}(x_o) = 0.$$

From formulae (19)-(21) it follows that

$$\limsup f_o(x) I_{S_n}(x) \leq f_o(x) \limsup I_{S_n}(x) \leq f_o(x) I_{S_o \cup T}(x), \text{ and}$$

$$f_o(x) I_{S_o}(x) \leq f_o(x) \liminf I_{S_n}(x) \leq \liminf f_o(x) I_{S_n}(x).$$

Since  $f_o(x) I_{S_n}(x)$  is bounded by the  $\mu$ -integrable function  $|f_o(x)|$  we can apply the lemma of Fatou and get

$$\begin{aligned} \limsup \int f_o(x) I_{S_n}(x) d\mu &\leq \int \limsup f_o(x) I_{S_n}(x) d\mu \leq \\ \int f_o(x) \limsup I_{S_n}(x) d\mu &\leq \int f_o(x) I_{S_o \cup T}(x) d\mu = \int f_o(x) I_{S_o}(x) d\mu \leq \\ \int f_o(x) \liminf I_{S_n}(x) d\mu &\leq \int \liminf f_o(x) I_{S_n}(x) d\mu \leq \\ \liminf \int f_o(x) I_{S_n}(x) d\mu, \end{aligned}$$

which shows that  $\lim_{n \rightarrow \infty} \int f_o(x) I_{S_n}(x) d\mu = \int f_o(x) I_{S_o}(x) d\mu$ .

Theorem 4. Let  $\Theta$  be a complete, separable metric space and  $\mathcal{M}$  be the  $\sigma$ -field generated by the closed sets of  $\Theta$ .  $p_\theta(x)$  is assumed to be  $\mu \times \mathcal{M}$ -measurable and to be bounded by a  $\mu$ -integrable function  $g(x)$  such that

$p_\theta(x) \leq g(x)$  for all  $\theta \in \Theta$ . If under these conditions there exists a compact least favorable support  $G \in \mathcal{H}$  then there exists a regular  $\sigma$ -additive l.f.s.f.  $\lambda^*$  on  $\mathcal{H}$ . Moreover  $\lambda^*(\Theta)$  will be less than or equal to  $1/\alpha$ .

Proof. According to the remarks made in connection with formulae (4) and (5)  $\lambda^*$  will be a l.f.s.f. if

$$E_{\theta, p^*} = f(\lambda^*).$$

Because of lemma 2 it will be sufficient to show that a  $\lambda^*$  exists which satisfies

$$\delta = E_{\theta, p^*} = f(\lambda^*).$$

From (5) it follows that there exists a sequence  $\{\lambda_n\}$  of finite measures  $\lambda_n$  such that

$$\delta = \lim_{n \rightarrow \infty} (\alpha \lambda_n(G) + \int_{\mathcal{X}} (p_{\theta'}(x) - \int p_{\theta}(x) d\lambda_n)^+ d\mu).$$

Since we can choose  $\{\lambda_n\}$  in such a way that  $\lambda_n$  is concentrated on finitely many points of  $G$ , i.e.  $\lambda_n$  for fixed  $n$  is concentrated on a compact subset of  $G$ ,  $\{\lambda_n\}$  can be assumed to be a sequence of regular measures. For such measures a weak convergence is defined in the following way:  $\{\lambda_n\}$  converges weakly to  $\lambda_0$  iff  $\int_{\Theta} f(\theta) d\lambda_n$  converges to  $\int_{\Theta} f(\theta) d\lambda_0$  for all bounded continuous functions  $f$ . By Prohorov [13], theorem 1.12, it is shown that  $\{\lambda_n\}$  will have a weak limit  $\lambda_0$  iff there exists a constant  $M$  such that

$$(22) \quad \lambda_n(G) \leq M \quad \text{for all } n,$$

and for every  $\epsilon > 0$  there exists a compact set  $K_\epsilon$  such that for all  $n$

$$(23) \quad \lambda_n(G - K_\epsilon) < \epsilon.$$

Since  $f(\lambda_n) \geq \alpha \lambda_n(G)$  holds for all  $n$  we can choose  $\{\lambda_n\}$  so that

$$(24) \quad 1 \geq \delta \geq \alpha \lambda_n(G),$$

and hence (22) will be satisfied. From the compactness of  $G$  it follows that (23) is also satisfied so that  $\{\lambda_n\}$  has a weak limit  $\lambda_0$ .

Formula (24) shows that  $\lambda_0(G)$  is bounded by  $1/\alpha$ .

It remains to show that

$$(25) \quad \lim_{n \rightarrow \infty} \int_x (p_{\theta_n}(x) - \int p_{\theta}(x) d\lambda_n)^+ d\mu = \int (p_{\theta}(x) - \int p_{\theta}(x) d\lambda_0)^+ d\mu.$$

The weak convergence of  $\{\lambda_n\}$  to  $\lambda_0$  implies that

$$\lim_{n \rightarrow \infty} \int p_{\theta}(x) d\lambda_n = \int p_{\theta}(x) d\lambda_0 \quad \mu \text{-a.e.}$$

Let  $f_n(x)$  and  $f_0(x)$  be defined by

$$f_n(x) = p_{\theta_n}(x) - \int p_{\theta}(x) d\lambda_n, \quad f_0(x) = p_{\theta}(x) - \int p_{\theta}(x) d\lambda_0$$

and  $S_n$  and  $S_0$  be defined as in lemma 3. Since by assumption

$$|f_n(x)| \leq p_{\theta_n}(x) + g(x)/\alpha,$$

we can for  $f_n(x)$  apply the theorem of the dominated convergence. Hence

$\int_{S_n} f_n(x) d\mu$  will tend to  $\int_{S_0} f_0(x) d\mu$  uniformly in  $S_n$  as  $n$  tends to infinity so that for every  $\epsilon > 0$  the inequality

$$(26) \quad \left| \int_{S_n} f_n(x) d\mu - \int_{S_n} f_0(x) d\mu \right| \leq \epsilon/2$$

holds for sufficiently large  $n$ .

Since lemma 3 implies that for sufficiently large  $n$

$$(27) \quad \left| \int_{S_n} f_0(x) d\mu - \int_{S_0} f_0(x) d\mu \right| \leq \epsilon/2,$$

(25) can be shown to be true with the help of the triangle inequality.

By defining  $\lambda^*(C) = \lambda_0(CG)$ ,  $C \in \mathcal{U}$ , we get a measure which has the desired properties.

Remark 1. It is likely that the compactness condition on the least favorable support  $G$  can be weakened by a similar condition as it is given in (23).

Remark 2. Clearly the methods of section 2 can be applied also in connection with a least favorable support. On the other hand if  $\Theta$  itself is compact the method used in the proof of theorem 3 is also applicable to prove the existence of a least favorable measure.



## References

- [1] Bourbaki, N. (1952). Éléments de Mathématique, VI, Intégration.  
Act. Sci. et Ind. 1175. Hermann et Cie., Paris.
- [2] Dantzig, G. B. and Wald, A. (1951). On the fundamental lemma of  
Neyman and Pearson. Ann. Math. Statist. 22 87-93.
- [3] Dubins, L. E. and Savage, L. J. (1965). How to gamble if you  
must. McGraw-Hill, New York.
- [4] Dunford, N. and Schwartz, J. T. (1966). Linear Operators, 1.  
Interscience Publishers, New York.
- [5] Halmos, P. R. (1964). Measure Theory. Van Nostrand, Princeton.
- [6] Hoeffding, W. (1951). Optimum non-parametric tests. Proc. Sec.  
Berkeley Symp. Math. Stat. Prob. 83-92.
- [7] Hurwicz, L. and Uzawa, H. (1958). A note on the Lagrangian saddle-  
points 103-113. Studies in Linear and Non-Linear Programming.  
Stanford University Press, Stanford.
- [8] Krafft, O. (196 ). Eine symmetrische Behandlung des Testproblems.  
Will be published in Arch. Math.
- [9] Krafft, O. and Witting, H. (1967) Optimale Tests and ungünstigste  
Verteilungen. Will be published in Z. Wahrscheinlichkeitstheorie  
and Verw. Gebiete.
- [10] Lehmann, E. L. (1959). Testing Statistical Hypotheses.  
Wiley, New York.
- [11] Lehmann, E. L. (1952). On the existence of least favorable  
distributions. Ann. Math. Statist. 23 408-416.
- [12] Lehmann, E. L. and Stein, C. (1948). Most powerful tests of composite  
hypotheses. Ann. Math. Statist. 19 495-516.

- [13] Prohorov, Yu. V. (1956). Convergence of random processes and limit theorems in probability theory. (Russian) Teor. Verojatnost. i Primenen. 1 177-238 (English translation in Theor. Probab. and Appl. 1 157-214).
- [14] Wald, A. (1950). Statistical Decision Functions. Wiley, New York.
- [15] Witting H. (1966). Mathematische Statistik. Teubner, Stuttgart.
- [16] Yosida, K. and Hewitt, E. (1952) Finitely additive measures. Trans. Amer. Math. Soc. 72 46-66.