THE USE OF LEAST FAVORABLE DISTRIBUTIONS IN TESTING COMPOSITE HYPOTHESES

By H. E. Reinhardt

Montana State University[1]

0. Introduction and summary. The usual method of finding a most powerful size α test of a composite hypothesis against a simple alternative is the guessing of a Least Favorable Distribution (LFD)—introduced at various levels of generality by Neyman and Pearson [6], Wald [7], Lehmann [4], and Lehmann and Stein [5]—and testing the mixture of the distributions of the hypothesis over this LFD against the alternate using the Neyman-Pearson Fundamental Lemma. In guessing LFD's statisticians have looked for a mixture which is "like" the alternate.

In this paper, the notion of Uniformly Least Favorable Mixture (ULFM) is introduced. In Section 2, we show that a ULFM is a point in the convex set of mixtures of the hypothesis which is closest (in the sense of the $\mathfrak{L}^1$ norm) to the alternate. The condition is not sufficient. More generally, any LFM corresponds to a point which is closest to the alternate in some expansion or contraction of this set of mixtures. A sufficient condition for ULFM's is, essentially, that the nuisance parameter can take on the same values in the alternate as in the hypothesis. In Section 3, we consider the case where no ULFM exists. We show, *inter alia*, that any distribution is least favorable for a closed set of α's. (A pathological example shows that this closed set need not be the union of a finite number of closed intervals.)

1. Notation and definitions. We consider a family f_θ, $\theta \varepsilon \Omega$, of densities and a density g with respect to a σ-finite measure μ over a measurable space $(\mathfrak{X}, \mathfrak{A})$. For tests φ, i.e., measurable functions φ such that $0 \leqq \varphi(x) \leqq 1$ for all x, we shall use the inner product notation

$$(\varphi, f_\theta) = \int_{\mathfrak{X}} \varphi(x) f_\theta(x)\, d\mu(x). \tag{1}$$

For the problem of testing a composite hypothesis $H: f_\theta$, $\theta \varepsilon \Omega$, against the simple alternative g a *most powerful level* α *test* is a test φ which maximizes the power (φ, g) among all tests satisfying $(\varphi, f_\theta) \leqq \alpha$ for all $\theta \varepsilon \Omega$.

We assume there is a σ-algebra $\mathfrak{B}$ on the indexing set Ω such that $f_\theta(x)$ is measurable on $\mathfrak{A} \times \mathfrak{B}$. If λ is a probability measure over $(\Omega, \mathfrak{B})$ we define the mixture f_λ by

$$f_\lambda(x) = \int_\Omega f_\theta(x)\, d\lambda(\theta). \tag{2}$$

Received June 6, 1960; revised July 24, 1961.

[1] Address during the academic year 1961–62: Army Mathematics Research Center, Madison, Misconsin.

Then λ is called a *least favorable distribution* and f_λ a *least favorable mixture* (LFM) at level α for testing $H: f_\theta$, $\theta \varepsilon \Omega$, against g if a most powerful level α test of f_λ against g is also most powerful for testing H against g. If f_λ is least favorable at every level α, $0 \leqq \alpha \leqq 1$, f is called a *uniformly least favorable mixture* (ULFM) for testing H against g.

To obtain a most powerful test for testing $H: f_\theta$, $\theta \varepsilon \Omega$, against g and a least favorable distribution λ or mixture f_λ we shall frequently use the following generalization of the fundamental lemma of Neyman and Pearson by Lehmann and Stein ([5], Theorem 1, or Corollary 5, p. 92, of [3]):

THEOREM 1.1. *Suppose that λ is a probability distribution over Ω and that Ω' is a subset of Ω with $\lambda(\Omega') = 1$. Let φ be a test such that*

$$\varphi(x) = 1 \quad \text{if} \quad g(x) > kf_\lambda(x)$$

$$\varphi(x) = 0 \quad \text{if} \quad g(x) < kf_\lambda(x).$$

Then φ is a most powerful level α test for testing H against g provided

$$(3) \qquad (\varphi, f_{\theta'}) = \sup_{\theta \varepsilon \Omega}(\varphi, f_\theta) = \alpha \qquad \text{for all } \theta' \varepsilon \Omega'.$$

2. Uniformly least favorable distributions. In [3] and [5] Lehmann and Lehmann and Stein, respectively, give examples of problems where the least favorable distribution depends on the level of significance. The case where the LFD is independent of the level of significance is more tractable and we consider it first. The following theorem shows the relation between ULFM's and the $\mathfrak{L}^1$ norm:

THEOREM 2.1. *If f_λ is a* ULFM *for testing $H: f_\theta$, $\theta \varepsilon \Omega$, against g, then $f_\lambda - g$ is a point of smallest norm in the convex set $\{f_\nu - g\}$*. (Here $\{f_\nu\}$ is the convex set of mixtures of f_θ's formed by averaging with respect to a probability measure on the space Ω. Thus $\{f_\nu\}$ is a convex set in the positive part of the unit sphere in an $\mathfrak{L}^1$ space.)

We omit the direct proof of Theorem 2.1 since it is a special case (for $k = 1$) of Theorem 3.2.

For Bernoulli distributions with a single trial the situation is particularly simple. The positive part of the unit sphere is the line segment in the Euclidean plane consisting of points $(x, 1 - x)$ with $0 \leqq x \leqq 1$. Thus the convex set spanned by the hypothesis is a line segment which we can take to be closed. The closest point to the alternate is the one whose first co-ordinate minimizes $|x - p|$ (where $(p, 1 - p)$ is the distribution under the alternate). This closest point is, in fact, a ULFM. (If min $|x - p| = 0$, the alternate is in the hypothesis and $\varphi(x) \equiv \alpha$ is most powerful. If min $|x - p| \neq 0$ there is a non-trivial test.)

If we were able to find the LFM for a sample of n independent observations knowing it for a sample of one observation, we would not have to do the problem over for each sample size. Such an inductive procedure is possible in a fairly special case.

LEMMA 2.2. *If* $\mu\{x \mid f(x) = kg(x)\} = 0$ *for every* k, *then*

$$(\mu \times \mu)\{(x, y) \mid f(x)f(y) = kg(x)g(y)\} = 0$$

for every k.

PROOF. Let $E = \{(x, y) \mid f(x)f(y) = kg(x)g(y)\}$. Consider a section E^{y_0} of E by y_0, i.e., the set of all x such that $(x, y_0) \varepsilon E$. Hence $x \varepsilon E^{y_0}$ if and only if

$$f(x) = kg(x)g(y_0)/f(y_0).$$

Then by the condition of the lemma $\mu(E^{y_0}) = 0$, and this is true for all y_0 except possibly the set where $f(y_0) = 0$. But this set has measure 0 by condition (taking $k = 0$). Thus all y sections have measure 0 and hence E has product measure 0 ([2], Theorem 36A).

THEOREM 2.3. *Let* f_{θ_0} *be a* ULFM *for testing* $H_1 : f_\theta$, $\theta \varepsilon \Omega$, *against* g; *if* $\theta_0 \varepsilon \Omega$ *and if* $\mu\{x \mid f_{\theta_0}(x) = kg(x)\} = 0$ *for each* k, *then for* n *independent observations of* X, *the density* $\prod_{i=1}^n f_{\theta_0}(x_i)$ *is uniformly least favorable for testing* H_n: $\prod_{i=1}^n f_{\theta_0}(x_i)$, $\theta \varepsilon \Omega$, *against* $\prod_{i=1}^n g(x_i)$.

PROOF. We prove the statement only for $n = 2$. The proof can easily be generalized by an induction.

By Lemma 2.2 and the Neyman-Pearson Lemma there is for each α a most powerful level α test φ for testing $f_{\theta_0}(x)f_{\theta_0}(y)$ against $g(x)g(y)$ such that

$$\varphi(x, y) = 1 \quad \text{if} \quad g(x)g(y) > kf_{\theta_0}(x)f_{\theta_0}(y), \qquad \text{and}$$

$$\varphi(x, y) = 0 \quad \text{if} \quad g(x)g(y) \leqq kf_{\theta_0}(x)f_{\theta_0}(y).$$

Thus φ is the indicator function I_S of a set S. Since f_{θ_0} is a ULFM, the section S^y of S by y is a most powerful test for testing H_1 against g for $g(y) > 0$ and $S^y = \phi$ for $g(y) = 0$. Hence

$$\int_{S^y} [f_{\theta_0}(x) - f_\theta(x)]\, d\mu(x) \geqq 0 \tag{4}$$

and similarly

$$\int_{S^x} [f_{\theta_0}(y) - f_\theta(y)]\, d\mu(y) \geqq 0. \tag{5}$$

Applying Fubini's theorem we obtain

$$\int_S [f_{\theta_0}(x)f_{\theta_0}(y) - f_\theta(x)f_\theta(y)]\, d(\mu \times \mu) \geqq 0. \tag{6}$$

Thus $\varphi = I_S$ is uniformly most powerful for testing H_2 against $g(x)g(y)$ ([3], Theorem 3.7). Hence $f_{\theta_0}(x)f_{\theta_0}(y)$ is a ULFM.

The LFM has to be in H_1, for otherwise we would be looking at

$$\prod_{i=1}^n \int f_\theta(x_i)d\lambda(\theta);$$

and this, in general, is not of the form $\int \prod_{i=1}^{n} f_\theta(x_i)d\lambda(\theta)$. It is mixtures of the latter sort that are available as potential LFM's.

In many densities of practical importance, the most natural parametrization is given by an indexing set which is a product space. (For normal distributions the pair (μ, σ^2) where μ is the mean and σ^2 the variance is a natural choice.) In many hypothesis testing problems, the statistican is only interested in one co-ordinate of the parameter. However, the nature of the observations he can make forces him to consider the other co-ordinates (traditionally called "nuisance parameters").

For problems of a certain form we can eliminate the nuisance parameters from consideration. Toward this end, let X, Y be independent random variables whose joint probability measure is absolutely continuous with respect to the product measure $\mu \times v$; thus,

$$P((X, Y) \varepsilon A) = \iint_A f_\theta(x)g_\eta(y)\, d\mu \times v. \tag{7}$$

THEOREM 2.4. *If there exists a probability measure* λ *on* H *such that* $h(y) = \int_{\mathrm{H}} g_\eta(y)d\lambda(\eta)$, *then* $f_{\theta_0}(x)h(y)$ *is a* ULFM *for testing* H: $f_{\theta_0}(x)g_\eta(y)$, $\eta \varepsilon \mathrm{H}$, *against* $f_{\theta_0}(x)h(y)$.

PROOF. A most powerful level α test for testing $f_{\theta_0}(x)h(y)$ against $f_{\theta_1}(x)h(y)$ is given by

$$\varphi(x) = 1 \quad \text{if} \quad f_{\theta_1}(x)h(y) > kf_{\theta_0}(x)h(y)$$
$$\varphi(x) = 0 \quad \text{if} \quad f_{\theta_1}(x)h(y) < kf_{\theta_0}(x)h(y),$$

where k is chosen so $\int \varphi(x)f_{\theta_0}(x)d\mu = \alpha$. Since φ is independent of y it follows that φ is most powerful for testing H against $f_{\theta_1}(x)h(y)$. Since α is arbitrary $f_{\theta_0}(x)h(y)$ is uniformly least favorable. (Theorem 2.1 suggests that this is a natural candidate for ULFM.)

COROLLARY 2.5. (See [1], p. 86.) *For testing* H: $f_{\theta_0}(x)g_\eta(y)$, $\eta \varepsilon \mathrm{H}$, *against* $f_{\theta_1}(x)g_\eta(y)$, $\eta \varepsilon \mathrm{H}$, *there is a uniformly most powerful test given by*

$$\varphi(x, y) = 1 \quad \text{if} \quad f_{\theta_1}(x) > kf_{\theta_0}(x)$$
$$\varphi(x, y) = 0 \quad \text{if} \quad f_{\theta_1}(x) < kf_{\theta_0}(x).$$

PROOF. Apply the theorem to the problem of testing $f_{\theta_0}(x)g_\eta(y)$, $\eta \varepsilon \mathrm{H}$, against $f_{\theta_1}(x)g_{\eta_1}(y)$ for each η_1. A ULFM is $f_{\theta_0}(x)g_{\eta_1}(y)$.

The corollary says, in effect, that we can not get information about θ by observing a random variable whose distribution is independent of θ!

Theorem 2.4 is a generalization of the method to obtain a uniformly most powerful test for the following example, used by Lehmann and Stein [5] (or see [3], p. 96).

EXAMPLE. Let $X_1, \cdots, X_n$ be independently normally distributed with mean η and variance θ. For given $\theta_1 > \theta_0$, we wish to test the composite hypothesis $H: \theta = \theta_0$, $-\infty < \eta < +\infty$, against the simple alternative $\theta = \theta_1$, $\eta = \eta_1$.

The normal distribution of the parameter η with mean η_1 and variance $(\theta_1 - \theta_0)/n$ is uniformly least favorable.

3. The non-uniform case. In the case where no ULFM exists, a result similar to Theorem 2.1 holds.

LEMMA 3.1. *Let f and g be probability densities with respect to μ; if $k > 0$, then*

$$\|kf - g\| = k - 1 + 2 \sup_{A \in \mathfrak{A}} \int_A (g(x) - kf(x))\, d\mu. \tag{8}$$

PROOF.

$$\begin{aligned} \|kf - g\| &= \int_{\mathfrak{X}} |(kf(x) - g(x))|\, d\mu \\ &= \int_{\mathfrak{X}} (kf(x) - g(x))\, d\mu + 2 \int_{g>kf} (g(x) - kf(x))\, d\mu \\ &= k - 1 + 2 \sup_{A \in \mathfrak{A}} \int_A (g(x) - kf(x))\, d\mu. \end{aligned}$$

THEOREM 3.2. *Suppose $f_{\lambda\alpha}(x)$ is an* LFM *at the level α for testing $H: f_\theta$, $\theta \varepsilon \Omega$, against g and the most powerful test is given by*

$$\varphi(x) = 1 \quad \text{if} \quad g(x) > kf_{\lambda\alpha}(x),$$
$$\varphi(x) = 0 \quad \text{if} \quad g(x) < kf_{\lambda\alpha}(x),$$

then $kf_{\lambda\alpha} - g$ is a point of smallest norm in the convex set $\{kf_\nu - g\}$.

PROOF. Consider any mixture f_ν of f_θ's. Since $\varphi(x)$ is a most powerful level α test we have

$$\int \varphi(x) f_\nu(x)\, d\mu \leqq \alpha. \tag{9}$$

Hence by Lemma 3.1 and by definition of φ

$$\begin{aligned} \|kf_\nu - g\| &= k - 1 + 2 \sup_{A \in \mathfrak{A}} \int_A (g(x) - kf_\nu(x))\, d\mu \\ &\geqq k - 1 + 2 \int \varphi(x)(g(x) - kf_\nu(x))\, d\mu \geqq \|kf_{\lambda\alpha} - g\|. \end{aligned}$$

In solving a problem where there is no ULFM, we would like to know how to proceed from an LFM for α_1 to one for α_2. We prove three useful theorems.

THEOREM 3.3. *The set of* LFM's *for a particular α is convex.*

PROOF. Suppose f_1 and f_2 are both LFM's at level α and suppose $0 \leqq a \leqq 1$. Let a most powerful test be φ. Then

$$\varphi(x) = 1 \quad \text{if} \quad g(x) > k_i f_i(x)$$
$$\varphi(x) = 0 \quad \text{if} \quad g(x) < k_i f_i(x)$$

for $i = 1, 2$. Let k_0 and b be determined by

$$k_0 a = k_1 b \quad \text{and} \quad k_0(1 - a) = k_2(1 - b). \tag{10}$$

Then

$$\varphi(x) = 1 \quad \text{if} \quad g(x) > k_0[af_1(x) + (1 - a)f_2(x)]$$
$$\varphi(x) = 0 \quad \text{if} \quad g(x) < k_0[af_1(x) + (1 - a)f_2(x)].$$

Thus φ is a most powerful test for $af_1 + (1 - a)f_2$ vs. g and consequently $af_1 + (1 - a)f_2$ is least favorable.

COROLLARY 3.4. *Suppose f_1 and f_2 are both least favorable at level α, and the k's of the fundamental lemma are k_1 and k_2 with $k_1 \geqq k_2 > 0$. Then for any k_0, $k_1 \geqq k_0 \geqq k_2$, there is an a, $0 \leqq a \leqq 1$, such that the k of the fundamental lemma for testing $af_1 + (1 - a)f_2$ against g is k_0.*

PROOF. Determine a and b from (10). The corollary then follows from the proof of Theorem 3.3.

THEOREM 3.5. *For testing $H: f_\theta$, $\theta \varepsilon \Omega$, against g, any f_λ is least favorable for a closed set of α's (which may be empty).*

PROOF. Suppose $\lim_n \alpha_n = \alpha$ and f_λ is least favorable at level α_n for each positive integer n. We show f_λ is least favorable at level α. Let the most powerful level α_n test be φ_n. Then, because of the weak* compactness of the set of tests, ([3], Appendix 4) there is a $\varphi(x)$, $0 \leqq \varphi(x) \leqq 1$ and a subsequence of $\{\varphi_n\}$—which we take to be $\{\varphi_n\}$ itself such that $\lim_n (\varphi_n, f) = (\varphi, f)$ for every f. In particular we have (since φ_n is level α_n)

$$(\varphi_n, f_\theta) \leqq \alpha_n \tag{11}$$

so that

$$(\varphi, f_\theta) \leqq \alpha. \tag{12}$$

Thus φ is level α for H.

Similarly, since f_λ is least favorable at level α_n,

$$(\varphi, f_\lambda) = \alpha \tag{13}$$

and φ is level α for f_λ vs. g.

Let the power of φ_n be $1 - \beta_n$, so

$$(\varphi_n, g) = 1 - \beta_n. \tag{14}$$

Now power is a concave and hence continuous function of size so the most powerful level α test of f_λ vs. g has power $1 - \beta = \lim_n(1 - \beta_n)$. But $(\varphi, g) = 1 - \beta$, so φ is a most powerful level α test of f_λ vs. g, and satisfies the size requirements of the original problem. Hence f_λ is least favorable.

Before proceeding to the final theorem of this section, we remark that for testing f vs. g (both simple) and for any $k > 0$, there is an α such that the most powerful level α test of f vs. g is

$$\varphi(x) = 1 \quad \text{if} \quad g(x) > kf(x)$$
$$\varphi(x) = 0 \quad \text{if} \quad g(x) < kf(x).$$

Clearly we can take any α satisfying

$$\int_{kf<g} fd\,\mu \leqq \alpha \leqq \int_{kf\leqq g} fd\,\mu. \tag{15}$$

Further, if we let k_α be the (not necessarily unique) k of the Fundamental Lemma such that φ is a most powerful level α test of f vs. g, then $\alpha_2 < \alpha_1$ implies $k_{\alpha_1} \leqq k_{\alpha_2}$.

THEOREM 3.6. *Suppose for testing $H: f_\theta$, $\theta \varepsilon \Omega$, against g there are a finite number of mixtures of the hypothesis, $f_1, \cdots, f_m$, (not necessarily distinct) each least favorable for an interval I_i of α's such that*

$$\bigcup_{i=1}^{m} I_i = [0, 1]. \tag{16}$$

Then there is an LFM *for some α which is a point closest to g (in $\mathfrak{L}^1$ norm) among the points of the convex set $\{f_\nu\}$.*

PROOF. Let the ends of the intervals be $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_m = 1$. We can assume, without loss of generality, that f_i is least favorable for $\alpha_{i-1} \leqq \alpha \leqq \alpha_i$. Let $k(\alpha, i)$ be the k of the Fundamental Lemma for testing f_i against g at level α. We can take $k(\alpha_0, 1) = \infty$ and $k(\alpha_m, m) = 0$. Hence there is an i such that $k(\alpha_{i-1}, i) \geqq 1 \geqq k(\alpha_i, i)$ or $k(\alpha_i, i) \geqq 1 \geqq k(\alpha_i, i+1)$. In the first case, by the remark above there is an α, $\alpha_{i-1} \leqq \alpha \leqq \alpha_i$, such that $k(\alpha, i) = 1$ and hence by Theorem 3.1 f_i is closest to g. In the second case by Theorem 3.3, $f_\lambda = af_i + (1-a)f_{i+1}$ is least favorable at level α_i, and for a suitable choice of a the k of the Fundamental Lemma is 1 for f_λ by Corollary 3.4. Then f_λ is a point closest to the alternate.

Thus one might be able to proceed stepwise—finding an LFM as a mixture closest to the alternate, then finding the set of α's for which this is least favorable. For each point which is a boundary point of this set of α's there may be another LFM. (This is true if there are only a finite number of LFM's.) For this LFM, we could proceed as with the first one. This procedure works perfectly for some problems which will be considered elsewhere. The theorems suggest, however, that LFM's are going to be difficult to find in general.

We conclude with a pathological example. We take as sample space the positive integers; the distributions can be represented as sequences $\{a_n\}$ with $a_n \geqq 0$ and $\sum a_n = 1$. We consider three sequences defined inductively.

$$\begin{array}{lll} a_1 = 1/3 & b_1 = 1/3 & c_n = (1/2)^n \\ a_2 = 2/9 + \epsilon & b_2 = 2/9 - \epsilon & \\ a_3 = 4/27 - \epsilon & b_3 = 4/27 + \epsilon & \\ a_4 = 2^3/3^4 & b_4 = 2^3/3^4 & \end{array}$$

$$
\begin{aligned}
a_5 &= (2/3)^3 b_2 & b_5 &= (2/3)^3 a_2 \\
a_6 &= (2/3)^3 b_3 & b_6 &= (2/3)^3 a_3 \\
&\vdots \\
a_{n+6} &= (2/3)^6 a_n & b_{n+6} &= (2/3)^6 b_n
\end{aligned}
$$

where $0 < \epsilon < 2/81$. Then the sequences $\{a_n/c_n\}$ and $\{b_n/c_n\}$ are monotone increasing and

$$\sum_{n=1}^{6k+2} a_n > \sum_{n=1}^{6k+2} b_n \quad \text{and} \quad \sum_{n=1}^{6k+5} b_n > \sum_{n=1}^{6k+5} a_n \qquad \text{for } k = 0, 1, 2, \cdots .$$

Hence for testing H: $\{a_n\}$, $\{b_n\}$ against $\{c_n\}$ any most powerful level α test is of the form

$$
\begin{aligned}
\varphi(n) &= 1 \quad \text{if} \quad n < n_0 \\
\varphi(n) &= 0 \quad \text{if} \quad n > n_0 \, ;
\end{aligned}
$$

however no uniformly least favorable distribution exists. Further $\{a_n\}$ is least favorable for a closed set of α's which is not the union of a finite number of closed intervals.

The example is a fairly natural one to construct from the point of view of hypothesis testing. What it says about the $\mathcal{L}^1$ norm as a byproduct is something of a surprise (at least to the originator).

4. Acknowledgments. Professor James G. Wendel offered many helpful suggestions and much encouragement during the development of this paper. The referee offered many suggestions which added clarity and conciseness to the exposition.

REFERENCES

[1] D. A. S. FRASER, *Nonparametric Methods in Statistics*, John Wiley and Sons, New York, 1957.

[2] PAUL R. HALMOS, *Measure Theory*, D. Van Nostrand, New York, 1950.

[3] E. L. LEHMANN, *Testing Statistical Hypotheses*, John Wiley and Sons, New York, 1959.

[4] E. L. LEHMANN, "On the existence of least favorable distributions," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 408–416.

[5] E. L. LEHMANN AND C. STEIN, "Most powerful tests of composite hypotheses I. Normal distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 495–516.

[6] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses I. Unbiased critical regions of type A and type A_1," *Statistical Research Memoirs*, Vol. 1 (1936), pp. 1–37.

[7] ABRAHAM WALD, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.