The Annals of Statistics
1979, Vol. 7, No. 5, 1066–1077

COMPARISON OF EXPERIMENTS AND INFORMATION MEASURES[1]

BY PREM K. GOEL AND MORRIS H. DEGROOT

Purdue University and Carnegie-Mellon University

Let $\mathcal{E}_X = \{X, S_X; P_\theta, \theta \in \Theta\}$ and $\mathcal{E}_Y = \{Y, S_Y; Q_\theta, \theta \in \Theta\}$ be two statistical experiments with the same parameter space Θ. Some implications of the sufficiency of $\mathcal{E}_X$ for $\mathcal{E}_Y$, according to Blackwell's definition, are given in terms of Kullback-Leibler information and Fisher information matrices. For a scale parameter θ, and $k_1 > k_2 > 0$, the experiment with parameter θ^{k_1} is proved to be sufficient for the experiment with parameter θ^{k_2} for a class of distributions including the gamma distribution and the normal distribution with known mean. Some results of Stone are generalized to the class of experiments with both location and scale parameters. A concept of sufficiency is proposed in which $\mathcal{E}_X$ is more informative than $\mathcal{E}_Y$ for a fixed prior distribution of θ if the expected Bayes risk from $\mathcal{E}_X$ is not greater than that from $\mathcal{E}_Y$ for every decision problem involving θ. This concept is then used to develop a definition of marginal Bayesian sufficiency in the presence of nuisance parameters.

1. Introduction and summary. Let $\mathcal{E}_X = \{X, S_X; P_\theta, \theta \in \Theta\}$ denote a statistical experiment in which a random variable or random vector X defined on some sample space S_X is to be observed, and the distribution P_θ of X depends on a parameter θ whose value is unknown and lies in some parameter space Θ. Also, let $\mathcal{E}_Y = \{Y, S_Y; Q_\theta, \theta \in \Theta\}$ denote another statistical experiment with the same parameter space Θ. Blackwell's (1951) method for comparing two experiments states that the experiment $\mathcal{E}_X$ is sufficient for the experiment $\mathcal{E}_Y$ (denoted $\mathcal{E}_X \succcurlyeq \mathcal{E}_Y$) if there exists a stochastic transformation of X to a random variable $Z(X)$ such that, for each $\theta \in \Theta$, the random variables $Z(X)$ and Y have identical distributions. It was proved by Blackwell (1953) and Boll (1955) that this method of comparison is equivalent to Bohnenblust, Shapley and Sherman's method for comparing two experiments [see Blackwell (1951)] which states that $\mathcal{E}_X$ is more informative than $\mathcal{E}_Y$ if for every decision problem involving θ and every prior distribution on Θ, the expected Bayes risk from $\mathcal{E}_X$ is not greater than that from $\mathcal{E}_Y$. LeCam (1964) generalized this notion to a concept of approximate sufficiency or ε-deficiency of $\mathcal{E}_X$ relative to $\mathcal{E}_Y$. Torgersen (1970, 1972, 1976) and Hansen and Torgersen (1974) have extended these results and applied them to interesting special classes of experiments. Some other papers on this topic are DeGroot (1962,

Received February 1977, revised June 1978.

[1]This research was supported in part by the National Science Foundation under Grant SOC 77-07548 at Carnegie-Mellon University.

AMS 1970 *subject classifications.* Primary 62B15; secondary 62B10.

Key words and phrases. Comparison of experiments, Kullback-Leibler information, Fisher information matrix, Bayes risk, marginal Bayesian sufficiency, informative experiment, self-decomposable characteristic functions.

1966), Torgersen (1977), Feldman (1972), and Gerber (1977). We shall now give a summary of the results presented in this paper.

In Section 2, we summarize various implications of the relation $\mathcal{E}_X \geqslant \mathcal{E}_Y$ in terms of Kullback-Leibler (K-L) information and Fisher information.

In Section 3, it is assumed that θ is a scale parameter in the distribution of Y and that X has this same distribution except that θ is replaced by $\theta^k (k > 1)$. Let W denote a random variable with distribution identical to that of X with $\theta = 1$, and let $\varphi(t)$ denote the characteristic function of $\log W$. As stated in Lemma 2, if $\varphi(t)/\varphi(t/k)$ is a characteristic function, then $\mathcal{E}_X \geqslant \mathcal{E}_Y$. It is noted that $\varphi(t)$ satisfies this condition if and only if $\varphi(t)$ is a self-decomposable characteristic function [see Lukacs (1970), page 161]. Let $G_n(a, b)$ denote an experiment in which a random sample of size n is taken from a gamma distribution $G(a, b)$ with parameters $a > 0$ and $b > 0$, for which the density function is

$$g(w) = \frac{w^{a-1}}{b^a \Gamma(a)} \exp(-w/b) \text{ for } w > 0. \tag{1.1}$$

Also, let $N_n(\mu, \sigma^2)$ denote an experiment in which a random sample of size n is taken from a normal distribution with mean μ and variance σ^2. The result in Lemma 2 is used to prove that for any known numbers $a > 0$ and $b > 0$, $G_n(a, b\theta^k) \geqslant G_n(a, b\theta)$ for all $k > 1$, and that $N_n(0, \sigma^{2k}) \geqslant N_n(0, \sigma^2)$ for all $k > 1$.

In Section 4, we extend some of the results obtained by Stone (1961) for a class of experiments with location parameter θ to the class of experiments with both a location parameter μ and a scale parameter σ. Various results for normal distributions are summarized. We consider an experiment $\mathcal{E}^*(c)$ involving a random variable having a pdf of the form $(c/\sigma)f[c(x - \mu)/\sigma]$. In Theorem 5 we study the relation between the Fisher information matrices for two experiments $\mathcal{E}^*(c_1)$ and $\mathcal{E}^*(c_2)$ when f is a symmetric function.

Since two experiments $\mathcal{E}_X$ and $\mathcal{E}_Y$ may not be comparable in Blackwell's sense, Feldman (1972) introduced another definition in which $\mathcal{E}_X$ is more informative than $\mathcal{E}_Y$ for a fixed decision problem involving θ if, for every prior distribution on Θ, the expected Bayes risk from $\mathcal{E}_X$ is not greater than that from $\mathcal{E}_Y$. In Section 5, we propose an alternative definition in which $\mathcal{E}_X$ is more informative than $\mathcal{E}_Y$ with respect to a fixed prior distribution on Θ if, for every decision problem involving θ, the expected Bayes risk from $\mathcal{E}_X$ is not greater than that from $\mathcal{E}_Y$. We then apply this concept to problems in which $\boldsymbol{\theta}$ is a vector with a given prior distribution, and we are interested in decision problems involving only some of the components of $\boldsymbol{\theta}$. We present a definition of the marginal Bayesian sufficiency of $\mathcal{E}_X$ for $\mathcal{E}_Y$ in this context. Some examples are given to illustrate the usefulness of this concept.

2. Relationships between sufficiency and information. Consider again two arbitrary experiments $\mathcal{E}_X$ and $\mathcal{E}_Y$ with the same parameter space Θ as defined at the beginning of Section 1. We shall assume that there exist generalized probability density functions (gpdf's) $p(x|\theta)$ and $q(y|\theta)$ for the distributions P_θ and Q_θ, with

respect to some σ-finite measures μ and ν respectively. We shall now investigate the implications of the relation $\mathcal{E}_X \succcurlyeq \mathcal{E}_Y$ in terms of some well-known information measures. Let Ξ denote the class of all prior distributions on the parameter space Θ. Given two prior distributions $\xi_1, \xi_2 \in \Xi$, let $p_i(x)$ denote the marginal gpdf $\int_\Theta p(x|\theta)d\xi_i(\theta)$, for $i = 1, 2$, and let $I_X(\xi_1, \xi_2)$ denote the K-L information contained in $\mathcal{E}_X$ for discriminating between $p_1(x)$ and $p_2(x)$, defined by

$$I_X(\xi_1, \xi_2) = \int_{S_X} p_1(x) \log \frac{p_1(x)}{p_2(x)} d\mu(x). \tag{2.1}$$

If ξ_1 assigns probability 1 to a point $\theta = \theta_0$, we shall denote $I_X(\xi_1, \xi_2)$ by $I_X(\theta_0, \xi_2)$. The K-L information $I_Y(\xi_1, \xi_2)$ contained in $\mathcal{E}_Y$ is defined analogously.

Lindley (1956) has shown that if $\mathcal{E}_X \succcurlyeq \mathcal{E}_Y$, then the Shannon information contained in $\mathcal{E}_X$ is at least as large as that contained in $\mathcal{E}_Y$. That is, if $\mathcal{E}_X \succcurlyeq \mathcal{E}_Y$ then

$$\int_\Theta I_X(\theta, \xi)d\xi(\theta) \geqslant \int_\Theta I_Y(\theta, \xi)d\xi(\theta) \quad \text{for all } \xi \in \Xi. \tag{2.2}$$

If (2.2) holds for $\mathcal{E}_X$ and $\mathcal{E}_Y$, we shall denote it by $\mathcal{E}_X \succcurlyeq_L \mathcal{E}_Y$. The following stronger version of Lindley's result was proved by Sakaguchi (1964).

LEMMA 1. *If* $\mathcal{E}_X \succcurlyeq \mathcal{E}_Y$, *then*

$$I_X(\theta, \xi) \geqslant I_Y(\theta, \xi) \quad \textit{for all } \theta \in \Theta \quad \textit{and } \xi \in \Xi. \tag{2.3}$$

In fact, this result can be easily extended as follows:

THEOREM 1. *If* $\mathcal{E}_X \succcurlyeq \mathcal{E}_Y$, *then*

$$I_X(\xi_1, \xi_2) \geqslant I_Y(\xi_1, \xi_2) \quad \text{for all } \xi_1, \xi_2 \in \Xi. \tag{2.4}$$

PROOF. The relation $\mathcal{E}_X \succcurlyeq \mathcal{E}_Y$ is preserved when the parameter space Θ is enlarged to include all convex combinations of the basic distributions P_θ of X and convex combinations of the distributions Q_θ of Y. Therefore, the distribution ξ in (2.4) can be assumed, without loss of generality, to be degenerate. Hence, the theorem follows from Lemma 1. □

The following example shows that Lemma 1 is stronger than Lindley's result.

EXAMPLE 1. Let $\mathcal{E}(\theta_1, \theta_2, \theta_3)$ denote an experiment in which a coin with unknown probability of heads θ is flipped n times and the parameter space Θ contains only three points $0 \leqslant \theta_1 \leqslant \theta_2 \leqslant \theta_3 \leqslant 1$. Blackwell (1951) remarks that the experiment $\mathcal{E}_X \equiv \mathcal{E}(0, \frac{1}{2}, 1)$ is not sufficient for the experiment $\mathcal{E}_Y \equiv \mathcal{E}(0, \frac{1}{2}, \frac{1}{2})$ even though our intuition suggests the contrary. In other words, suppose that there are only three possible states of nature and we have a choice of either (i) observing n flips of a coin $\mathcal{E}_X$ for which the probability of heads is 0, $\frac{1}{2}$, or 1 according as θ_1, θ_2 or θ_3 is correct, or (ii) observing n flips of a coin $\mathcal{E}_Y$ for which the probability of heads is 0 if θ_1 is correct, but is $\frac{1}{2}$ if either θ_2 or θ_3 is correct. It would seem at first glance that $\mathcal{E}_X$ must always be at least as useful as $\mathcal{E}_Y$, but Blackwell pointed out that this conclusion is not correct.

Lindley (1956) showed that $\mathcal{E}_X \succcurlyeq_L \mathcal{E}_Y$ for these experiments. However, it can be shown that $I_X(\theta_2, \xi) \geq I_Y(\theta_2, \xi)$ if and only if $3\xi_1 + \xi_2 \leq 1$, where $\xi_i = \xi(\theta_i)$. Hence (2.3) does not hold for $\theta = \theta_2$ and a prior distribution ξ for which $3\xi_1 + \xi_2 > 1$.

Since $\mathcal{E}_X$ is not sufficient for $\mathcal{E}_Y$ in this example, there must be a decision problem in which the expected Bayes risk from $\mathcal{E}_Y$ is less than that from $\mathcal{E}_X$. The following simple decision problem has this property. Suppose that the hypothesis $H_0 : \theta = \theta_2$ is to be tested against the alternative $H_1 : \theta \neq \theta_2$ with the usual $0 - 1$ loss function and the prior distribution ξ defined by $\xi(\theta_2) = \lambda$ and $\xi(\theta_1) = \xi(\theta_3) = (1 - \lambda)/2$. If λ satisfies $2^{n-1} < \lambda/(1 - \lambda) < 2^{n-1} + \frac{1}{2}$, then it can be shown that the Bayes rule for the experiment $\mathcal{E}_Y$ is to reject H_0 if the number of heads is 0 and the Bayes rule for the experiment $\mathcal{E}_X$ is to accept H_0 regardless of the outcome. Since the outcome of $\mathcal{E}_X$ is of no value to the experimenter in this decision problem, it can be shown that the expected Bayes risk from $\mathcal{E}_X$ is larger than that from $\mathcal{E}_Y$.

The converse of Theorem 1 does not necessarily hold. In fact, it follows from Torgersen (1970, Proposition 18) that if Θ is finite and, in his terminology, $\mathcal{E}_X$ is more informative than $\mathcal{E}_Y$ for testing problems, then (2.4) holds, even if $\mathcal{E}_X$ is not sufficient for $\mathcal{E}_Y$.

The relation $\mathcal{E}_X \succcurlyeq \mathcal{E}_Y$ implies a similar ordering in terms of Fisher information. Let $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$ and suppose that Θ is an open subset of R^k. Let $\mathbf{i}_X(\boldsymbol{\theta})$ and $\mathbf{i}_Y(\boldsymbol{\theta})$ denote the $k \times k$ Fisher information matrices for the experiments $\mathcal{E}_X$ and $\mathcal{E}_Y$ respectively, under the standard regularity conditions such as those given in Kullback ((1968), pages 26–27). We shall use the notation $\mathcal{E}_X \succcurlyeq_F \mathcal{E}_Y$ whenever $\mathbf{i}_X(\boldsymbol{\theta}) - \mathbf{i}_Y(\boldsymbol{\theta})$ is nonnegative definite. The next theorem essentially follows from the development given in Kullback ((1968), pages 26–28). A sketch of the proof of this result is also given by Barndorff-Nielsen in his discussion of Torgersen (1976). For $k = 1$, the result had been proved by Stone (1961).

THEOREM 2. *If* $\mathcal{E}_X \succcurlyeq \mathcal{E}_Y$, *then* $\mathcal{E}_X \succcurlyeq_F \mathcal{E}_Y$.

REMARK 1. It should be noted that if $\mathbf{i}_X(\boldsymbol{\theta}) - \mathbf{i}_Y(\boldsymbol{\theta})$ is nonnegative definite and $\mathbf{i}_Y(\boldsymbol{\theta})$ is nonnegative definite, then $|\mathbf{i}_X(\boldsymbol{\theta})| \geq |\mathbf{i}_Y(\boldsymbol{\theta})|$ (see Rao (1973), page 70). In other words, the generalized Fisher information in $\mathcal{E}_X$ is at least as large as that in $\mathcal{E}_Y$.

REMARK 2. A counterexample which shows that the converse of Theorem 2 does not necessarily hold is given by Hansen and Torgersen (1974).

3. Comparison of normal experiments with known mean and unknown variance. In this section we shall consider experiments in which a random sample can be taken from a normal distribution for which the mean is known and the variance is unknown. Without loss of generality we shall assume that the known value of the mean is 0.

To begin we note that $N_n(0, \sigma^2) \succcurlyeq N_n(0, \sigma^2 + k^2)$ where k is a given constant. To see this, suppose that the random variable X is distributed as $N(0, \sigma^2)$, the random variable Y is distributed as $N(0, \sigma^2 + k^2)$, and the random variable W is independent of X and has the distribution $N(0, k^2)$. Then $X + W$ has the same distribution as Y for every possible value of σ^2. Hence, $N_1(0, \sigma^2) \succcurlyeq N_1(0, \sigma^2 + k^2)$. However, it is well known that if $\mathcal{E}_X \succcurlyeq \mathcal{E}_Y$ when only one observation is taken in each experiment, then this same relation holds when a random sample of n observations is taken from each experiment [See Blackwell (1951)]. It now follows that $N_n(0, \sigma^2) \succcurlyeq N_n(0, \sigma^2 + k^2)$.

Next, we note that for any given constant $k \neq 0$, the experiments $N_n(0, \sigma^2)$ and $N_n(0, k^2\sigma^2)$ are equivalent in the sense that each is sufficient for the other. This follows from the fact that multiplying each observation in the first experiment by k or multiplying each observation in the second experiment by $1/k$ maps each experiment into the other one.

We turn now to the much more difficult problem of determining whether either of the experiments $N_n(0, \sigma^2)$ and $N_n(0, \sigma^{2k})$ is sufficient for the other, where k is a given positive constant. We shall prove that for $k_1 > k_2 > 0$, $N_n(0, \sigma^{2k_1}) \succcurlyeq N_n(0, \sigma^{2k_2})$. First we present some related results.

LEMMA 2. *Let W be a nonnegative random variable with* pdf $g(w)$ *and let $\varphi(t)$ denote the characteristic function of* $\log W$. *Let $\theta > 0$ be an unknown parameter, let $k > 0$ be a given constant, and let $G_n(\theta^k)$ denote the experiment in which a random sample of n observations is taken from the distribution with* pdf $(1/\theta^k)g(w/\theta^k)$. *For any given constant $c > 0$, define*

$$\psi_c(t) = \frac{\varphi(t)}{\varphi(ct)}, -\infty < t < \infty. \tag{3.1}$$

If $\psi_{k_2/k_1}(t)$ is a characteristic function, then $G_n(\theta^{k_1}) \succcurlyeq G_n(\theta^{k_2})$. Moreover, $G_1(\theta^{k_1}) \succcurlyeq G_1(\theta^{k_2})$ if and only if $\psi_{k_2/k_1}(t)$ is a characteristic function.

The proof of this lemma follows from results in Boll (1955) or Torgersen (1972) by regarding $\log \theta$ as a translation parameter in the distribution of $\log W$.

It should be noted that $\psi_c(t)$, defined in (3.1), is a characteristic function for all $c \in (0, 1)$ if and only if $\varphi(t)$ belongs to the class of self-decomposable characteristic functions, introduced by P. Lévy and A. Ya. Khinchine (See Lukacs (1970), subsection 5.11). Some interesting properties of this class (also called L-class by Gnedenko and Kolmogorov (1954)) are as follows:

(i) All self-decomposable characteristic functions are infinitely divisible.

(ii) If $\varphi(t)$ is a self-decomposable characteristic function, then $\psi_c(t)$ is infinitely divisible.

(iii) All stable characteristic functions are self-decomposable.

(iv) The necessary and sufficient conditions for $\varphi(t)$ to be self-decomposable in terms of Lévy's and Kolmogorov's canonical representations of an infinitely divisible characteristic functions are given in Theorems 1 and 2 of Chapter 6 in Gnedenko and Kolmogorov (1954).

We shall now assume that g is the density function of a gamma distribution $G(a, b)$, defined in (1.1), with known values of a and b, and prove that the assumptions in Lemma 2 hold for this pdf.

THEOREM 3. *Let $G_n(a, b)$ denote the experiment in which a random sample of n observations is taken from the gamma distribution $G(a, b)$ with* pdf (1.1). *Then $G_n(a, b\theta^{k_1}) \geqslant G_n(a, b\theta^{k_2})$, where $\theta > 0$ is an unknown parameter, and a, b, k_1 and k_2 are given positive constants with $k_1 > k_2$.*

PROOF. Let $\varphi(t)$ denote the characteristic function of $\log W$, where W is a random variable with pdf (1.1), and let $\alpha = k_2/k_1$. Then

$$\varphi(t) = \int_0^\infty \frac{1}{b^a\Gamma(a)} w^{it} w^{a-1} e^{-w/b} dw, \tag{3.2}$$

$$= b^{it}\Gamma(a + it)/\Gamma(a).$$

For $\alpha < 1$, consider

$$\psi_\alpha(t) = \frac{\varphi(t)}{\varphi(\alpha t)} = \frac{b^{it(1-\alpha)}\Gamma(a + it)}{\Gamma(a + i\alpha t)}. \tag{3.3}$$

The Weerstrass expansion of $1/\Gamma(z)$ (see Whittaker and Watson (1935), page 236) is

$$\frac{1}{\Gamma(z)} = ze^{\gamma z}\prod_{j=1}^\infty \left\{\left(1 + \frac{z}{j}\right)\exp(-z/j)\right\} \tag{3.4}$$

where γ is the Euler's constant. Therefore, after some algebraic manipulation, $\psi_\alpha(t)$ can be written as

$$\psi_\alpha(t) = b^{it(1-\alpha)}\exp\{-\gamma it(1 - \alpha)\}\left(\frac{a + it\alpha}{a + it}\right)\prod_{j=1}^\infty \left(\frac{j + a + it\alpha}{j + a + it}\right)e^{it(1-\alpha)/j} \tag{3.5}$$

or, equivalently, as

$$\psi_\alpha(t) = \exp\{-it(1 - \alpha)(\gamma - \log b)\}\left[\alpha + (1 - \alpha)\left(1 + \frac{it}{a}\right)^{-1}\right] \tag{3.6}$$

$$\times \prod_{j=1}^\infty \left[\alpha + (1 - \alpha)\left(1 + \frac{it}{j + a}\right)^{-1}\right]e^{it(1-\alpha)/j}.$$

The first factor in (3.6) is the characteristic function of a degenerate random variable T_0 with probability one at the point $[-(1 - \alpha)(\gamma - \log b)]$, and the factor $e^{it(1-\alpha)/j}$ is the characteristic function of a degenerate random variable T_j with probability one at the point $(1 - \alpha)/j, j = 1, 2, \cdots$. Furthermore, $[\alpha + (1 - \alpha)(1 + (it/(j + a))^{-1}]$ is the characteristic function of a random variable $Z_j, j = 0, 1, 2, \cdots$, which takes the value 0 with probability α and, with probability $(1 - \alpha)$, has the pdf

$$f_j(z) = (j + a)\exp[(j + a)z] \quad \text{for} \quad z \leqslant 0 \tag{3.7}$$

$$= 0 \quad \text{for} \quad z > 0.$$

Let $\{T_i, i = 0, 1, 2, \cdots\}$ and $\{Z_i, i = 0, 1, 2, \cdots\}$ be independent sequences of independent random variables with the distributions defined above. Define $S_j = \Sigma_{i=0}^{j}(T_i + Z_i)$, and let $\psi_j(t)$ denote the characteristic function of S_j. It follows from (3.6) and the above discussion that

$$\psi_\alpha(t) = \lim_{j\to\infty}\psi_j(t). \tag{3.8}$$

It is obvious that $\psi_\alpha(0) = 0$. Furthermore, since the gamma function $\Gamma(z)$ is analytic for complex arguments except at the points $z = 0, -1, -2, \cdots$, where it has simple poles, $\psi_\alpha(t)$ is continuous at $t = 0$. Hence by the continuity theorem (Theorem 3.6.1, Lukacs (1970)), $\psi_\alpha(t)$ is a characteristic function. It now follows from Lemma 2 that $G_n(a, b\theta^{k_1}) \geqslant G_n(a, b\theta^{k_2})$ for all $k_1 > k_2 > 0$. □

REMARK. (i) An alternative way to prove that $\Sigma_{i=0}^{j}(Z_i + T_i)$ converges in distribution to a random variable Z, is to use Theorem 3.7.3, Lukacs (1970). Since $\Sigma_{i=0}^{\infty}\operatorname{Var}(Z_i + T_i) = \alpha(1-\alpha)\Sigma_{i=0}^{\infty}1/(i+a)^2\} < \infty$, $\psi_j(t)$ converges to a characteristic function $\psi_\alpha(t)$ as $j \to \infty$.

(ii) Another proof of the fact that $\psi_\alpha(t)$ is a characteristic function could be given by proving that $\varphi(t)$ is self-decomposable by using either Theorem 1 or Theorem 2 in Chapter 6 of Gnedenko and Kolmogorov (1954). However, we prefer the proof given above because it gives the specific form of the random variable Z in the proof of Lemma 2.

(iii) Using Theorem 3.7.6 of Lukacs (1970), it can be shown that the distribution function of the random variable Z is continuous.

We shall now prove the main result of this section.

THEOREM 4. *Let k_1 and k_2 be given constants satisfying $k_1 > k_2 > 0$. Then $N_n(0, \sigma^{2k_1}) \geqslant N_n(0, \sigma^{2k_2})$.*

PROOF. As explained earlier, we can assume, without loss of generality, that $k_2 = 1$ and $k_1 = k$. Let X and Y denote the observations in the experiments $N_1(0, \sigma^{2k})$ and $N_1(0, \sigma^2)$, respectively. Since X^2/σ^{2k} and Y^2/σ^2 have the same χ^2 distribution, it follows from Theorem 3 that the experiments in which X^2 is observed is sufficient for that in which Y^2 is observed. Furthermore, since X^2 is a sufficient statistic for the experiment $N_1(0, \sigma^{2k})$ and Y^2 is a sufficient statistic for the experiment $N_1(0, \sigma^2)$, it follows that $N_1(0, \sigma^{2k}) \geqslant N_1(0, \sigma^2)$. Hence, $N_n(0, \sigma^{2k}) \geqslant N_n(0, \sigma^2)$. □

If X and Y have the distributions specified in the proof of Theorem 4, we now know how to generate a random variable equivalent to an observation on Y from an observation on X.

Let the random variable Z be as defined in the proof of Theorem 3, independently of X, with $a = \frac{1}{2}$, $b = 2$, and $\alpha = k_2/k_1$, and let Y' be defined as follows

$$\begin{aligned} Y' &= |X|^\alpha e^{Z/2} && \text{with probability } \tfrac{1}{2}, \\ &= -|X|^\alpha e^{Z/2} && \text{with probability } \tfrac{1}{2}. \end{aligned}$$

Then Y' has the same distribution as Y for every possible value of σ^2.

4. Comparison of experiments with location and scale parameters. Stone (1961) considers the class of experiments $\{\mathcal{E}(c);\ c > 0\}$ where $\mathcal{E}(c)$ is the experiment in which an observation is taken from the pdf $cf[c(x-\theta)]$, for a fixed pdf f and $\Theta = R$. For given values of c_1 and c_2, he obtains conditions under which $\mathcal{E}(c_1) \geqslant \mathcal{E}(c_2)$, $\mathcal{E}(c_1) \geqslant_L \mathcal{E}(c_2)$, or $\mathcal{E}(c_1) \geqslant_F \mathcal{E}(c_2)$. Let $\varphi(t)$ denote the characteristic function of the pdf f. Stone shows that if $f(\cdot)$ is bounded and $c_1 > c_2 > 0$, then a sufficient condition for $\mathcal{E}(c_1) \geqslant \mathcal{E}(c_2)$ is that

$$\psi(t) = \frac{\varphi(t/c_2)}{\varphi(t/c_1)} \tag{4.1}$$

be a characteristic function. However, it follows from the references mentioned after Lemma 2 that the boundedness of $f(\cdot)$ is not needed in this result. Furthermore, it follows that if $\varphi(t)$ is a self-decomposable characteristic function, then $\mathcal{E}(c_1) \geqslant \mathcal{E}(c_2)$ for all $c_1 > c_2 > 0$.

Stone also established that if $f(\cdot)$ is bounded and the family of pdf's $\{f(u-\theta);\ \theta \in R\}$ is boundedly complete, then a necessary condition that $\mathcal{E}(c_1) \geqslant \mathcal{E}(c_2)$ whenever $c_1 > c_2 > 0$ is that $\psi(t)$ be a characteristic function. In addition, if all the cumulants of $f(\cdot)$ exist, Stone proves that $\psi(t)$ is a characteristic function, only if (i) $f(\cdot)$ is a normal density or (ii) the even-order cumulants of $f(\cdot)$ are positive. After proving this result, he states that "it is possible that condition (ii) is inconsistent with $\mathcal{E}(c_1) \geqslant \mathcal{E}(c_2)$ whenever $c_1 > c_2$, in which event, yet another characterization of the normal distribution would be provided". However, for $f(u) = \exp(-u)$, $u > 0$, and $c_1 > c_2 > 0$, it can be shown that $\psi(t)$ is a characteristic function and therefore $\mathcal{E}(c_1) \geqslant \mathcal{E}(c_2)$. Furthermore, all the cumulants of $f(\cdot)$ exist, all the even order cumulants of $f(\cdot)$ are positive, $f(\cdot)$ is bounded, and the family of distributions $\{f(u-\theta);\ \theta \in R\}$ is boundedly complete. Hence, this result does not provide yet another characterization of the normal distribution, as suggested by Stone.

A natural extension of the above results is to consider the class of experiments $\{\mathcal{E}^*(c);\ c > 0\}$ such that $\mathcal{E}^*(c)$ is the experiment in which an observation is taken from the pdf $(c/\sigma)f[c(x-\mu)/\sigma]$, where f is a given pdf and the parameter space is $\Theta = \{(\mu, \sigma) : \mu \in R, \sigma > 0\}$. One may ask whether $\mathcal{E}^*(c_1) \geqslant \mathcal{E}^*(c_2)$ for $c_1 > c_2 > 0$. In particular, one may ask whether $N_n(\mu, \sigma^2)$ is sufficient for $N_n(\mu, \sigma^2/c^2)$, where $c < 1$ is a known constant. The answer is negative, as shown by Boll (1955). In fact, it follows from Theorem 3.1 of Hansen and Torgersen (1974) that $N_{n_1}(\mu, \sigma^2/c_1^2) \geqslant N_{n_2}(\mu, \sigma^2/c_2^2)$ if and only if either $n_1c_1^2 = n_2c_2^2$ and $n_1 \geqslant n_2$ or $n_1c_1^2 > n_2c_2^2$ and $n_1 \geqslant n_2 + 1$.

Stone (1961) also proved that in the location parameter case, $\mathcal{E}(c_1) \geqslant_F \mathcal{E}(c_2)$ for $c_1 > c_2 > 0$, whenever the Fisher information exists. We shall now extend this result to the family of experiments $\mathcal{E}^*(c)$ with both location and scale parameters, defined at the beginning of this section.

THEOREM 5. *Suppose that the* pdf $f(x)$ *is symmetric around the point* $x = 0$. *For* $i = 1, 2$, *let* $\mathcal{E}_i$ *denote the experiment in which* n_i *independent replications of the experiment* $\mathcal{E}^*(c_i)$ *are performed, where* c_1 *and* c_2 *are given positive constants. Assume*

that the Fisher information matrices $\mathbf{i}_1(\mu, \sigma)$ *and* $\mathbf{i}_2(\mu, \sigma)$ *exist for these two experiments. Then* $\mathcal{E}_1 \succcurlyeq_F \mathcal{E}_2$ *if and only if* $n_1 \geq n_2$ *and* $n_1 c_1^2 \geq n_2 c_2^2$.

PROOF. It can be shown that if $f(\cdot)$ is symmetric, then the matrix $\mathbf{i}_1(\mu, \sigma) - \mathbf{i}_2(\mu, \sigma)$ is diagonal with diagonal elements $(n_1 c_1^2 - n_2 c_2^2)A$ and $(n_1^2 - n_2^2)B$, where A and B are the positive diagonal elements of the information matrix for the experiment $\mathcal{E}^*(1)$. Hence, $\mathbf{i}_1(\mu, \sigma) - \mathbf{i}_2(\mu, \sigma)$ will be nonnegative definite if and only if $n_1 c_1^2 \geq n_2 c_2^2$ and $n_1^2 \geq n_2^2$. □

It has been pointed out by a referee that for any given pdf f and any positive constants $c_1 \neq c_2$, the experiments $\mathcal{E}^*(c_1)$ and $\mathcal{E}^*(c_2)$ are never comparable with respect to the sufficiency ordering $\succcurlyeq$.

5. Marginally sufficient experiments. In general, the relation $\mathcal{E}_X \succcurlyeq \mathcal{E}_Y$ is equivalent to the requirement that $\mathcal{E}_X$ is at least as preferred as $\mathcal{E}_Y$ for *every decision problem* involving the parameter θ and *every prior distribution* on Θ. Therefore, it is a very restrictive relation and induces only a partial ordering on the class $E(\Theta)$ of all possible experiments with parameter space Θ. Feldman (1972) studied certain properties of orderings of $E(\Theta)$ induced by the weakened requirement that in a *fixed decision problem*, the expected Bayes risk from $\mathcal{E}_X$ be not greater than that from $\mathcal{E}_Y$ for *every prior distribution* $\xi \in \Xi$. Following DeGroot (1962), he identified the decision problem with an uncertainty function $\mathcal{U}(\xi)$ defined on Ξ and considered the experiment $\mathcal{E}_X$ to be at least as informative as the experiment $\mathcal{E}_Y$ with respect to $\mathcal{U}$ if $\mathcal{U}(\xi|X) \leq \mathcal{U}(\xi|Y)$ for all $\xi \in \Xi$, where $\mathcal{U}(\xi|X)$ is the expected posterior uncertainty if X is observed and the prior distribution is ξ and $\mathcal{U}(\xi|Y)$ is the corresponding value for the observation Y.

An alternative possibility for comparing experiments is to consider a *fixed prior distribution* ξ and study the ordering on $E(\Theta)$ induced by the requirement that the expected Bayes risk from $\mathcal{E}_X$ be not greater than that from $\mathcal{E}_Y$ for *every decision problem* involving θ. In this case, we will say that $\mathcal{E}_X$ is at least as informative as $\mathcal{E}_Y$ with respect to the prior distribution ξ.

If $\mathcal{E}_X$ is at least as informative as $\mathcal{E}_Y$ for a fixed decision problem, then every experimenter interested in that decision problem will prefer $\mathcal{E}_X$ to $\mathcal{E}_Y$ since any risk function that can be attained from $\mathcal{E}_Y$ can be matched or dominated by one from $\mathcal{E}_X$. On the other hand, if $\mathcal{E}_X$ is at least as informative as $\mathcal{E}_Y$ with respect to a prior distribution ξ, then an experimenter with prior distribution ξ on Θ will prefer $\mathcal{E}_X$ to $\mathcal{E}_Y$ regardless of his decision problem. We shall now give an example to illustrate this concept.

EXAMPLE 2. Let $c_1 > c_2 > 0$ be given constants and for $i = 1, 2$, let X_i denote a random variable with the normal distribution $N(\mu, \sigma^2/c_i^2)$. Suppose that the prior distribution of (μ, σ) is concentrated on just two points such that $\Pr[(\mu, \sigma) = (0, 1)] = \xi$ and $\Pr[(\mu, \sigma) = (\mu_0, \sigma_0)] = 1 - \xi$ where $0 < \xi < 1$, σ_0 and μ_0 are known and arbitrary. It follows from Bradt and Karlin (1956) that when the parameter space is regarded as containing just these two points, $N_1(\mu, \sigma^2/c_1^2)$ is sufficient for

$N_1(\mu, \sigma^2/c_2^2)$, and therefore $N_n(\mu, \sigma^2/c_1^2) \geq N_n(\mu, \sigma^2/c_2^2)$. Hence $N_n(\mu, \sigma^2/c_1^2)$ is more informative than $N_n(\mu, \sigma^2/c_2^2)$ with respect to this prior distribution.

Torgersen (1976, Theorem 1) shows that if Θ is countable, then the partial ordering of experiments with respect to a given prior distribution ξ such that $\xi(\theta) > 0$ for each $\theta \in \Theta$ is equivalent to the usual sufficiency partial ordering $\geq$. However, the concept of relative informativeness with respect to a prior distribution ξ is especially useful when the parameter $\boldsymbol{\theta}$ is vector valued, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, and the experimenter is interested only in $\boldsymbol{\theta}_1$; i.e., $\boldsymbol{\theta}_2$ is a nuisance parameter. For example, in the experiment $\mathcal{E}^*(c)$ defined in Section 4, corresponding to the pdf $(c/\sigma)f[(c/\sigma)(x - \mu)]$, the decision problems of interest may involve only μ or only σ. A detailed discussion on the elimination of nuisance parameters in the framework of classical statistical inference is given by Basu (1977). For a given prior distribution, $\xi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \xi_1(\boldsymbol{\theta}_1)\xi_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1)$, a Bayesian statistician who is interested only in $\boldsymbol{\theta}_1$, because the loss depends only on $\boldsymbol{\theta}_1$, will eliminate $\boldsymbol{\theta}_2$ from the analysis and use the prior pdf $\xi_1(\boldsymbol{\theta}_1)$ together with the conditional pdf

$$g(x|\boldsymbol{\theta}_1) = \int_{\Theta_2} p(x|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\xi_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1). \tag{5.1}$$

Consider a particular decision problem with $\Theta = \{\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) | \boldsymbol{\theta}_1 \in \Theta_1, \boldsymbol{\theta}_2 \in \Theta_2\}$ and a given class D of all possible decisions d, and let $l(\boldsymbol{\theta}, d)$ denote the loss incurred from any decision $d \in D$ when $\boldsymbol{\theta} \in \Theta$ is true. We shall say that the decision problem involves only $\boldsymbol{\theta}_1$ if, for every pair $(\boldsymbol{\theta}_1, d)$,

$$l[(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2), d] = l[(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*), d] \quad \text{for all} \quad \boldsymbol{\theta}_2^* \in \Theta_2, \tag{5.2}$$

i.e., l depends only on the value of $\boldsymbol{\theta}_1$ and the value of d, and not on the value of $\boldsymbol{\theta}_2$. For such decision problems, we now present a natural and useful concept of *marginal Bayesian sufficiency* with respect to a given prior distribution $\xi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$.

DEFINITION. The experiment $\mathcal{E}_X$ is marginally sufficient for $\mathcal{E}_Y$, denoted by $\mathcal{E}_X \geq \mathcal{E}_Y(\boldsymbol{\theta}_1)$, with respect to the prior distribution $\xi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ if the expected Bayes risk from $\mathcal{E}_X$ is not greater than that from $\mathcal{E}_Y$ for every decision problem involving only $\boldsymbol{\theta}_1$, when the prior distribution is $\xi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$.

For each $\theta_1 \in \Theta_1$, let G_{θ_1} denote the distribution on S_X represented by the conditional pdf given by (5.1), and let H_{θ_1} denote the analogous distribution on S_Y. Also, let $\mathcal{E}_X' = \{X, S_X; G_{\theta_1}, \theta_1 \in \Theta_1\}$ and $\mathcal{E}_Y' = \{Y, S_Y; H_{\theta_1}, \theta_1 \in \Theta_1\}$. If $\mathcal{E}_X' \geq \mathcal{E}_Y'$ then it will be true that $\mathcal{E}_X \geq \mathcal{E}_Y(\boldsymbol{\theta}_1)$ with respect to any prior distribution ξ^* that yields the same conditional distribution $\xi_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1)$ as ξ. In this case, we shall say that $\mathcal{E}_X \geq \mathcal{E}_Y(\boldsymbol{\theta}_1)$ with respect to the conditional prior distribution $\xi_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1)$.

We shall now give some examples of marginal Bayesian sufficiency.

EXAMPLE 3. For a given pdf f, let $\mathcal{E}^*(c)$ denote the experiment defined in Section 4, and let $\varphi(t) = \int_R e^{itu} f(u)du$. For any joint prior distribution of μ and σ, let $\xi_2(\sigma)$ denote the marginal prior distribution of σ and let $\varphi_1(t) = \int_0^\infty \varphi(t\sigma) d\xi_2(\sigma)$. It follows from (5.1) that if μ and σ are independent under their joint prior distribution, then $g(x|\mu)$ is of the form $cg^*[c(x - \mu)]$. Therefore, if $\varphi_1(t)$ is a self-decomposable characteristic function, then it follows from the result of Stone,

presented at the beginning of Section 4, that $\mathcal{E}^*(c_1) \geq \mathcal{E}^*(c_2)(\mu)$ with respect to the conditional prior distribution $\xi_2(\sigma)$ for $c_1 > c_2 > 0$.

In particular, let $f(u)$ be the standard normal pdf and let either (i) the prior density of σ^2 be a gamma distribution of the form $G(\alpha, \beta)$, or (ii) the prior density of $(1/\sigma^2)$ be a gamma distribution of the form $G(\frac{1}{2}, \beta)$. By carrying out the analysis indicated in this example, it can be shown that $N_1(\mu, \sigma^2/c_1) \geq N_1(\mu, \sigma^2/c_2)(\mu)$ with respect to both of these conditional prior distributions of σ^2 given μ.

EXAMPLE 4. Let $c_1 > c_2 > 0$ be given constants and, for $i = 1, 2$, let X_i denote a random variable with the normal distribution $N[\mu, \sigma^2/c_i]$. Suppose that our interest lies in decision problems involving only σ^2. If μ and σ are independent under their joint prior distribution, and if the marginal distribution of μ is a normal distribution $N(m, \tau^2)$, then it follows that, given σ^2, X_i is distributed as $N[m, \sigma^2/c_i^2 + \tau^2]$. Let W be distributed as $N[(1 - c_2/c_1)m, (1 - (c_2^2/c_1^2))\tau^2]$ independently of X_2. Then it can be verified that $(c_2/c_1)X_2 + W$ has the same distribution as X_1 for every possible value of σ^2. Hence, $N_1(\mu, \sigma^2/c_2^2) \geq N_1(\mu, \sigma^2/c_1^2)(\sigma^2)$ with respect to this conditional prior distribution of μ. In fact, using the joint distribution of $\bar{X}$ and S^2 from a random sample of n observations given σ^2, it can be shown that $N_n(\mu, \sigma^2/c_2^2) \geq N_n(\mu, \sigma^2/c_1^2)(\sigma^2)$ with respect to this conditional prior distribution of μ.

However, if the conditional prior distribution of μ given σ^2 is $N(m, \sigma^2/\tau^2)$, then it follows that X_i is distributed as $N[m, \sigma^2(1/c_i^2 + 1/\tau^2)]$, given σ^2. Therefore, the experiments $N_1(\mu, \sigma^2/c_1^2)$ and $N_1(\mu, \sigma^2/c_2^2)$ are sufficient for each other with respect to this conditional prior distribution of μ. Furthermore, it can be shown that the experiments $N_n(\mu, \sigma^2/c_1^2)$ and $N_n(\mu, \sigma^2/c_2^2)$ are sufficient for each other with respect to this conditional prior distribution of μ.

EXAMPLE 5. For given constants $c_1 > c_2 > 0$, consider again the normal experiments $N_n(\mu, \sigma^2/c_2^2)$, $i = 1, 2$, and suppose that the joint prior distribution of μ and σ^2 is a conjugate normal-gamma distribution such that the conditional distribution of μ given σ^2 is $N[m, \sigma^2/\tau^2]$ and the distribution of $(1/\sigma^2)$ is $G(\alpha, \beta)$. It follows from Example 3 that for decision problems involving only σ^2, the experiments $N_n(\mu, \sigma^2/c_1^2)$ and $N_n(\mu, \sigma^2/c_2^2)$ are marginally equivalent with respect to this joint prior distribution. However, for decision problems involving only μ, it is not known whether one of these experiments is marginally sufficient for the other with respect to this conjugate joint prior distribution. We can prove, however, that for estimating any of the functions μ, μ/σ, μ/σ^2, $\mu\sigma$ and $\mu\sigma^2$ with squared-error loss, the experiment $N_n(\mu, \sigma^2/c_1^2)$ has a smaller expected Bayes risk than the experiment $N_n(\mu, \sigma^2/c_2^2)$ for this conjugate prior distribution.

EXAMPLE 6. If a statistic $T(\mathbf{Y})$ is partially sufficient for the parameter $\boldsymbol{\theta}_1$, according to Fraser's definition (1956), then it can be proved that the experiment $\mathcal{E}_T$, in which only T is observed, satisfies $\mathcal{E}_T \geq \mathcal{E}_{\mathbf{Y}}(\boldsymbol{\theta}_1)$ with respect to any prior distribution $\xi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ under which $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are independent. For example, if

$Y_1, \cdots, Y_n$ are independent and identically distributed with a gamma distribution $G(\alpha, \beta)$, then $T = \Sigma_1^n Y_i$ is partially sufficient for β in Fraser's sense and, therefore, $\mathcal{E}_T \geqslant \mathcal{E}_Y(\beta)$ with respect to any prior distribution for which α and β are independent.

Acknowledgment. We are deeply indebted to the referees for extremely valuable comments and for pointing out some key references that led to a thorough revision of this paper.

REFERENCES

[1] BASU, D. (1977). On the elimination of nuisance parameters. *J. Amer. Statist. Assoc.* **72** 355–366.
[2] BLACKWELL, D. (1951). Comparison of experiments. *Proc. Second Berkeley Symp. Math. Statist. Probability*. Univ. California Press. 93–102.
[3] BLACKWELL, D. (1953). Equivalent comparison of experiments. *Ann. Math. Statist.* **24** 265–272.
[4] BOLL, CHARLES H. (1955). Comparison of experiments in the infinite case and the use of invariance in establishing sufficiency. Unpublished Ph.D. thesis, Depart. Statist., Stanford Univ.
[5] BRADT, R. N. and KARLIN, S. (1956). On the design and comparison of certain dichotomous experiments. *Ann. Math. Statist.* **27** 390–409.
[6] DEGROOT, M. H. (1962). Uncertainty, information and sequential experiments. *Ann. Math. Statist.* **33** 404–419.
[7] DEGROOT, M. H. (1966). Optimal allocation of observations. *Ann. Inst. Statist. Math.* **18** 13–28.
[8] DEGROOT, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.
[9] FELDMAN, D. (1972). Some properties of Bayesian ordering of experiments. *Ann. Math. Statist.* **43** 1428–1440.
[10] FRASER, D. A. S. (1956). Sufficient statistics with nuisance parameters. *Ann. Math. Statist.* **27** 838–842.
[11] GERBER, HANS U. (1977). Uncertainty functions with a constant rate of reduction and comparison of experiments. *J. Amer. Statist. Assoc.* **72** 899–900.
[12] GNEDENKO, B. V. and KOLMOGOROV, A. N. (1954). *Limit Theorems for Sums of Independent Random Variables*. Addison-Wesley, Reading, Mass.
[13] HANSEN, O. H. and TORGERSEN, E. N. (1974). Comparison of linear normal experiments. *Ann. Statist.* **2** 367–373.
[14] KULLBACK, S. (1968). *Information Theory and Statistics*. Dover, New York.
[15] LECAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35** 1419–1455.
[16] LINDLEY, D. V. (1956). On a measure of the information provided by an experiment. *Ann. Math. Statist.* **27** 986–1005.
[17] LUKACS, E. (1970). *Characteristic Functions, 2nd ed.* Hafner, New York.
[18] RAO, C. R. (1973). *Linear Statistical Inference and its Applications, 2nd ed.* Wiley, New York.
[19] SAKAGUCHI, M. (1964). Information theory and decision making. Unpublished lecture notes, Statist. Depart., The George Washington Univ., Washington, D.C.
[20] STONE, M. (1961). Non-equivalent comparisons of experiments and their use for experiments involving location parameters. *Ann. Math. Statist.* **32** 326–332.
[21] TORGERSEN, E. N. (1970). Comparison of experiments when the parameter space is finite. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **16** 219–249.
[22] TORGERSEN, E. N. (1972). Comparison of translation experiments. *Ann. Math. Statist.* **43** 1383–1399.
[23] TORGERSEN, E. N. (1976). Comparison of statistical experiments. *Scand. J. Statist.* **3** 186–208.
[24] TORGERSEN, E. N. (1977). Prediction sufficiency when the loss function does not depend on the unknown parameter. *Ann. Statist.* **5** 155–163.
[25] WHITTAKER, E. T. and WATSON, G. N. (1935). *A Course on Modern Analysis, 4th ed.* Cambridge Univ. Press, Cambridge.

DEPARTMENT OF STATISTICS
MATHEMATICAL SCIENCES BUILDING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907

DEPARTMENT OF STATISTICS
CARNEGIE-MELLON UNIVERSITY
SCHENLEY PARK
PITTSBURGH, PENNSYLVANIA 15213