Variation Diminishing Transformations: A Direct Approach to Total Positivity and Its Statistical Applications

LAWRENCE D. BROWN, IAIN M. JOHNSTONE, and K. BRENDA MacGIBBON*

Karlin has shown that the variation diminishing property possessed by totally positive distributions allows a unified and straightforward presentation of many basic properties of hypothesis tests. His work in fact reveals that total positivity is equivalent to a suitably formulated variation diminishing property. This largely expository article begins with the latter concept and consequently gives a more direct account of the theory. This approach avoids the extensive mathematical preliminaries previously required and isolates the more important statistical property.

KEY WORDS: Variation diminishing transformation; Total positivity; Sign changes; Polya frequency function; Hypothesis tests.

1. INTRODUCTION

Karlin (1956, 1957a,b, 1958) developed a unified approach to a collection of classical problems of testing and decision theory. He showed how these problems could be solved by appealing to variation diminishing properties of special families of one parameter distributions. In Karlin's development these variation diminishing (or variation reducing (VR)) properties were established by applying a second concept, now called total positivity (TP).

In nearly all statistical applications of total positivity, the argument is the same—establish total positivity and then apply the basic theorem that asserts that total positivity implies the variation reducing property. Karlin's Theorem (1968, 5.3.1), however, actually asserts the equivalence of TP and the VR property. Consequently the detour via TP is logically unnecessary for statistical theory.

In this paper we adopt a more direct approach and take the variation reducing property as basic, giving appropriate definitions and criteria for checking directly whether a family of densities is VR. This route eliminates both the need to define and study the concept of total positivity, and the extensive mathematical preliminaries that must precede its use. Our method emphasizes the more fundamental importance of the VR property for statistics and shows that for the applications of common interest, total positivity is an inessential detour. (In certain less common situations, however, an appeal to total positivity may still be more convenient.) Our mathematical tools are simple, mainly those of linear algebra and elementary calculus, although the proof of our central Lemma 3.2. is considerably simplified by an appeal to a separating hyperplane theorem in $\mathbf{R}^k$.

The simplest and most common total positivity property is TP_2, which is equivalent to monotone likelihood ratio (MLR). In this case Karlin's basic variation diminishing theorem states that the expectation $E_\theta g(X)$ of a monotone function g will itself be monotone in θ. This may be seen heuristically by recalling that an MLR family is stochastically increasing, which in turn suggests the monotonicity of $E_\theta g(X)$ for any monotone g. The case of TP_n for $n > 2$ is not as simple to describe as that of TP_2 but is also statistically important (see, e.g., Ex. 4.2). Corresponding to TP_n is the notion of VR_n, which is formally defined in Section 2.

In the two action decision theory framework, Karlin described complete classes of procedures and determined the Bayes and admissible rules for hypothesis testing. He used these results to establish for TP densities the familiar results on existence and nonexistence of uniformly most powerful (UMP) and uniformly most powerful unbiased (UMPU) one-sided and two-sided tests. These results were expounded (largely for exponential families) in Lehmann (1959). Similar discussions were also given for locally most powerful unbiased tests (Rao's (1973, p. 454) definition), for likelihood ratio tests of composite null versus composite alternative, and for envelope power functions. We illustrate in Section 4 how Karlin's theory may be recast in the VR setting with a discussion of UMPU two-sided tests of a simple null hypothesis.

The variation reducing property has been applied in the literature in many different statistical contexts since Karlin's work. Examples include the study of the combination of independent one-sided test statistics (Van Zwet and Oosterhoff 1967); scale families of symmetric

* Lawrence D. Brown is Professor, Department of Mathematics, Cornell University, Ithaca, NY 14853. Iain M. Johnstone is Assistant Professor, Department of Statistics, Stanford University, Stanford, CA 94305. K. Brenda MacGibbon is Adjunct Associate Professor, Department of Mathematics, Concordia University, 1455 de Maisonneuve Blvd. W., Montreal, Quebec, Canada H3G 1M8. Brown's research was supported in part by NSF Grant MCS 7824175, Johnstone's by an Australian National University Postbachelor Travelling Scholarship held at Cornell and MacGibbon's in part by NSERC of Canada, Operating Grant No. A4088.

December 1981, Volume 76, Number 376
Theory and Methods Section

and one-sided stable densities (Kanter 1975); the structure of optional screening and classification procedures (Marshall and Olkin 1968); qualitative properties of the power function of the F test (Farrell 1968); inequalities on the multinormal integral useful in optimal design (Rinott and Santner 1977); comparison of large deviation rates for differing distribution functions (Lynch 1979); and in the theory of hypothesis tests (Cohen 1965; Meeden 1971). In most of these cases it is the SVR_3 property or higher that is used—there are many further instances employing SVR_2 (which is equivalent to strict MLR). In these applications the desired (S)VR_n property is generally established through the (S)TP_n property and Karlin's basic theorem asserting that (S)TP_n implies (S)VR_n. Thus our approach can in principle be easily applied in all this work to directly establish the desired (S)VR_n property.

Section 2 contains the definitions and their more immediate consequences. Following Karlin (1968, Ch. 5), we distinguish between strict sign changes (S^+) and nonstrict sign changes (S^-) and give corresponding VR definitions (VR and SVR). This device often yields more detailed information on, say, the multiplicity of a zero of a function than use of the S^- operator alone would give. In Section 3, VR behavior of $f_\theta(x)$ on finite sets of θ and x values is shown to suffice for the general VR property, and the composition theorem is derived. These important results show that exponential families and the common noncentral densities are SVR. Section 4 contains the example mentioned earlier together with a quick proof of the monotonicity of power functions of one-sided tests from (S)VR families. Appendix 1 contains a detailed proof of our main theorem and a lemma used in Section 4. Appendix 2 defines the (S)TP property and proves the equivalence of (S)TP and (S)VR. The brevity of this proof reinforces our feeling that the approach beginning with the (S)VR definitions is a more natural one.

2. DEFINITIONS AND SIMPLE EXAMPLES

We begin by describing the variation of a function $g:X \to \mathbf{R}$ through two extreme definitions of sign changes and initial sign. Here $X \subset \mathbf{R}$. First assume that X is finite, with $X = \{x_1, \ldots, x_n\}$, $x_1 < \ldots < x_n$. The function g is then completely specified by X and the vector $(g(x_1), \ldots, g(x_n))$. It is therefore convenient to use vectors both in the following definitions and in Section 3.

Definition 2.1. Let $g = (g_1, \ldots, g_n) \in \mathbf{R}^n$. $S^-(g)$ denotes the number of sign changes of sequence $g_1, \ldots, g_n$, *ignoring zeros*. We make the convention that $S^-(\mathbf{0}) = -1$. This is reasonable if one views $S^-(g)$ as one less than the number of "intervals" on which g is strictly positive or strictly negative.

$S^+(g)$ denotes the *maximum* number of sign changes of the sequence $g_1, \ldots, g_n$ that can be obtained by counting zeros as either $+$ or $-$. Clearly $S^+(g) = \overline{\lim}_{h\to g} S^-(h)$.

Example 2.1. If $g' = (0, 2, 0, -1, -2)$ and $g'' = (0, 2, 0, 0, -2)$, then $S^-(g') = S^-(g'') = 1$, but $S^+(g') = 2$ and $S^+(g'') = 4$.

Definition 2.2. Now consider an arbitrary set $X \subset \mathbf{R}$. Let $g:X \to \mathbf{R}$ and $V \subset X$. Let g_V denote the restriction of g to V. If V is a finite set, we regard g_V as a vector. Define also

$$X_p = \{V \subset X: \text{ cardinality of } V = p\},\ p \geq 1.$$

$$X_f = \bigcup_{p=1}^{\infty} X_p, \quad \text{the family of all finite subsets of } X.$$

We shall always list the elements of a set of X_f in increasing order. We now define $S^\pm(g) = \sup_{V\in X_f} S^\pm(g_V)$. Figure 1 shows two examples.

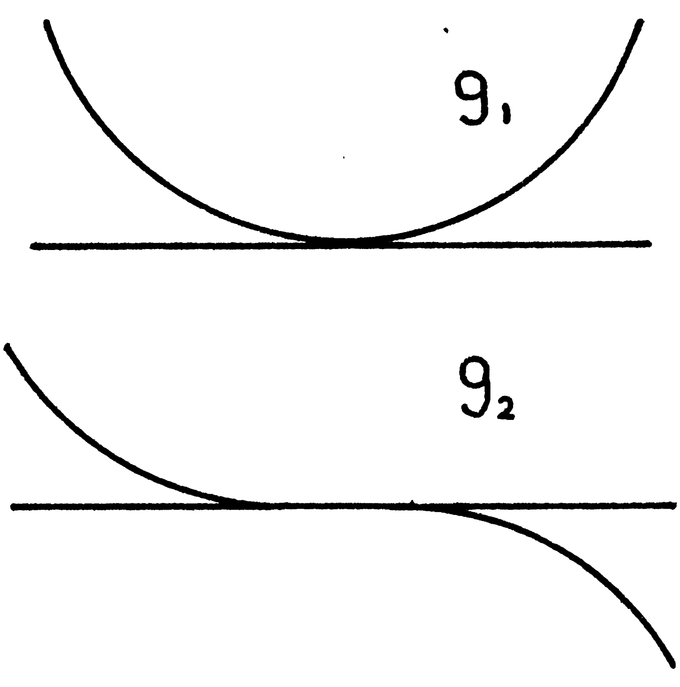
If $g = (g_1, \ldots, g_n) \in \mathbf{R}^n$ and $g \neq \mathbf{0}$, then let g_j denote the first nonzero term. The *initial sign* of g is defined by

$$\begin{aligned} IS^-(g) &= +, \quad \text{if } g_j > 0 \\ &= -, \quad \text{if } g_j < 0, \\ IS^+(g) &= +, \quad \text{if } (-1)^{j+1} g_j > 0 \\ &= -, \quad \text{if } (-1)^{j+1} g_j < 0. \end{aligned}$$

If $g = \mathbf{0}$, then set $IS^-(g) = IS^+(g) = 0$. In Example 2.1, $IS^-(g') = +$, but $IS^+(g') = -$.

If X is infinite and $S^-(g) < \infty$, then $IS^-(g) = IS^-(g_V)$, where V is any finite subset of X for which $S^-(g) = S^-(g_V)$. (The value of IS^- is independent of the choice of V: if V' were another such set with $IS^-(g_{V'}) = -IS^-(g_V)$, then it would follow that $S^-(g) \geq S^-(g_{V\cup V'}) > S^-(g_V) = S^-(g)$, an absurdity.) $IS^+(g)$ is defined in

Figure 1. Graphs of Two Examples, $S^-(g_1) = 0$, $S^+(g_1) = 2$, and $S^-(g_2) = 1$, $S^+(g_2) = \infty$.



an analogous manner. Alternatively, if $h_n \to g$, if $g \not\equiv 0$, and $S^+(g) = \lim_n S^-(h_n)$, then $IS^+(g) = \lim_n IS^-(h_n)$.

It will often be convenient to write $\mathcal{S}^{\pm}(g) = (S^{\pm}(g), IS^{\pm}(g))$. Thus, in Example 2.1, write $\mathcal{S}^-(g') = (1, +)$, $\mathcal{S}^+(g'') = (4, -)$, and so on.

The quantities $S^-(g)$ and $S^+(g)$ can be thought of as a quantitative measure of the amount of fluctuation or variation of the function g. (A more precise, but more complex measure is provided by examining $S^{\pm}(g-c)$ for all constant functions (or vectors) c.) Thus a monotone function has $S^-(g) \leqq 1$ and a unimodal function has $S^-(g) \leqq 2$. Strictly monotone and strictly unimodal functions have $S^+(g) \leqq 1$ and $S^+(g) \leqq 2$, respectively. (In fact, a function is monotone if and only if $S^-(g-c) \leqq 1$ for all c; and analogous remarks hold for unimodality and for strict versions of these properties.)

Suppose $\{f_\theta(x)\}$ is a family of probability densities relative to a measure ν. Consider $\gamma(\theta) = E_\theta g(X)$. We say that the kernel $\{f_\theta(x)\}$ reduces the variation of g if $S^-(\gamma) \le S^-(g)$. (Here the word "reduces" means more precisely "does not increase.") The VR and SVR properties defined below reflect this notion of $E_\theta(g(X))$ reducing the variation of g, and include an additional statement relating the order of sign changes of g and $E_\theta(g(X))$.

Definition 2.3: Variation Reducing Kernels. Suppose $f_\theta(x)$: $\Theta \times X \to [0, \infty)$, $\Theta \subset \mathbf{R}$, $X \subset \mathbf{R}$.

Let ν be a nonnegative measure on X and $g: X \to \mathbf{R}$ a function such that $\int |g| \, d\nu > 0$. Write $\gamma(\theta) = \int f_\theta(x) g(x) \nu(dx)$. We shall say that f is VR_{n+1} on X with parameter in Θ (written $f \in \mathrm{VR}_{n+1}(X, \Theta)$) if the following property holds for all such ν and g.

Property 2.1. $S^-(g) \le n$ implies $S^-(\gamma) \le S^-(g)$. If also $S^-(\gamma) = S^-(g)$, then $IS^-(\gamma) = IS^-(g)$.

In other words f is $\mathrm{VR}_{n+1}(X, \Theta)$ if for any ν and g with $S^-(g) \le n$, the number of sign changes of γ is bounded by the number of changes of g, and if the two are equal, then the changes occur in the same order.

We say that f is $\mathrm{SVR}_{n+1}(X, \Theta)$ if $S^-(\gamma)$ may be replaced by $S^+(\gamma)$ in (2.1). That is, $S^-(g) \le n$ implies $S^+(\gamma) \le S^-(g)$, and if also $S^+(\gamma) = S^-(g)$, then $IS^+(\gamma) = IS^-(g)$.

If γ is only well defined and finite on a subset $\tilde{\Theta}$ of Θ, we will regard $\tilde{\Theta}$ as the domain of γ, and think of the symbol γ as being replaced by $\gamma_{\tilde{\Theta}}$ in the above definitions and throughout the paper.

Remark 2.1. If $V = \{x_1, \ldots, x_p\} \in X_p$ and $\Sigma = \{\theta_1, \ldots, \theta_q\} \in \Theta_q$, then $f_{V \times \Sigma}$ may be regarded as a $q \times p$ matrix F: $F = (f_{ij}) = (f_{\theta_i}(x_j))$. Furthermore, the vector β with coordinates $\beta_i = \gamma(\theta_i) = \Sigma f_{\theta_i}(x_j) g(x_j) \nu(\{x_j\})$ may be written as $\beta = Fa$, where $a_j = g(x_j)\nu(\{x_j\})$. Note that $\mathcal{S}^{\pm}(a) = \mathcal{S}^{\pm}(g)$, and so forth, since we may assume that $\nu(\{x_i\}) > 0$ for each i. This convenient notational device emphasizes how the (S)VR properties relate to systems of linear equations, and will be frequently used.

If $p = q$ and $f \in \mathrm{SVR}_p(X, \Theta)$, then F is nonsingular. Indeed, if $Fa = 0$ and $a \neq 0$, then $S^-(a) \le p - 1 = S^+(0)$. Consequently, from the SVR property, equality must hold, so $IS^-(a) = IS^+(\mathbf{0}) = 0$, which is impossible, since $a \neq \mathbf{0}$.

Remark 2.2. If $\{f_\theta(x)\}$ is a parametric family of probability densities with respect to ν, and X is a random variable with density f_θ, then $\gamma(\theta) = E_\theta g(X)$.

Remark 2.3. The (S)VR property is *hereditary*: if $X' \subset X$ and $\Theta' \subset \Theta$, then $(\mathrm{S})\mathrm{VR}_{n+1}(X, \Theta)$ implies $(\mathrm{S})\mathrm{VR}_{n+1}(X', \Theta')$. This follows from two observations: first, that $S^{\pm}(\gamma_{\Theta'}) \le S^{\pm}(\gamma)$ and that $S^{\pm}(\gamma_{\Theta'}) = S^{\pm}(\gamma)$ implies $IS^{\pm}(\gamma_{\Theta'}) = IS^{\pm}(\gamma)$. Second, if g and ν are defined on X', then they may be extended to X by setting them equal to zero on $X \setminus X'$.

Remark 2.4. The (S)VR property is *cumulative*: $(\mathrm{S})\mathrm{VR}_{n+1}(X, \Theta)$ implies $(\mathrm{S})\mathrm{VR}_m(X, \Theta)$ for all $m \le n + 1$.

Remark 2.5. It may be checked that $f \in \mathrm{VR}_1$ (resp SVR_1) iff $f_\theta(x) \ge 0$ (resp > 0) for all θ, x. Interest focuses therefore on $(\mathrm{S})\mathrm{VR}_p$ for $p \ge 2$. In fact, $(\mathrm{S})\mathrm{VR}_2$ is equivalent to (S)MLR. See Proposition 3.3.

Definition 2.4. Suppose that for all $V \in X_p$, $\Sigma \in \Theta_q$, the function $f_\theta(x) \in (\mathrm{S})\mathrm{VR}_{n+1}(V, \Sigma)$. With slight abuse of notation, we say that $f \in \mathrm{SVR}_{n+1}(X_p, \Theta_q)$.

We can now state a simple consequence of the heredity property. Its converse will be established in Section 3.

Corollary 2.1. $\mathrm{SVR}_{n+1}(X, \Theta) \Rightarrow \mathrm{SVR}_{n+1}(X_{n+1}, \Theta_{n+1})$.

To illustrate the variety of statistical distributions possessing the (S)VR property, we shall need to develop some basic properties of (S)VR families (see Sec. 3). At this stage two simple examples of VR_∞ (that is, VR_n for all n) kernels may be given. The first is nonstatistical, but has an important statistical application (Proposition 3.4). Both families show that SVR is a strictly stronger condition than VR—neither is even SVR_3!

Example 2.2. Let

$$e(x, \theta) = \begin{cases} 1, & \text{if } x \le \theta \\ 0, & \text{if } x > \theta \end{cases} \qquad x, \theta \in \mathbf{R}.$$

We show that $e \in \mathrm{VR}_\infty(\mathbf{R}, \mathbf{R})$. Suppose $\mathcal{S}^-(g) = (n, +)$. $\mathbf{R}$ may then be partitioned into $n + 1$ intervals L_i (with L_i to the left of L_j for $1 \le i < j \le n + 1$) so that $(-1)^{i+1} g \ge 0$ on L_i. Now

$$\gamma(\theta) = \int_{-\infty}^{\infty} e(x, \theta) g(x) \nu(dx) = \int_{-\infty}^{\theta} g(x) \nu(dx),$$

so γ alternately increases or decreases as θ increases through successive L_i. Putting $\tilde{\gamma}_i = \int_{L_i} g(x)\nu(dx)$, $i = 1, \ldots, n + 1$ and $\tilde{\gamma} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_{n+1})$, it is clear that $S^-(\gamma) \le S^-(\tilde{\gamma}) \le n$. If $S^-(\gamma) = n$, then $S^-(\tilde{\gamma}) = n$ also, so that $\tilde{\gamma}_1 > 0$ and $IS^-(\gamma) = IS^-(g)$. Since n is arbitrary, e is VR_∞, while Remark 2.5 shows that it cannot be SVR_1.

Example 2.3 A VR_∞ *family that is only* SVR_2. Let $f_\theta(x) = 1 + \theta(x - \frac{1}{2})$, $0 \le x \le 1$, $-2 \le \theta \le 2$. Then $\gamma(\theta) = \int_0^1 f_\theta(x)g(x)\nu(dx) = (1 - \theta/2)\int_0^1 g\, d\nu + \theta\int_0^1 xg\, d\nu$. γ is linear in θ, so $S^-(\gamma) \le 1$ and $S^-(\gamma) \le S^-(g)$, whatever the value of $S^-(g)$. Now suppose that $S^-(g) = (1, +)$ with $\int |g(x)|\, \nu(dx) > 0$. We show that $S^-(\gamma) = S^+(\gamma) = 0$ or $\mathcal{S}^-(\gamma) = (1, +)$. Let the sign change of g occur at $x = c$. Then $(x - c)g(x) \le 0$ for all $x \in [0, 1]$. If $\nu\{x: (x - c)g(x) < 0\} = 0$, the result is easy; while if $\nu\{x: (x - c)g(x) < 0\} > 0$, we get $\gamma(2) = 2\int_0^1 xg\, d\nu < 2c\int_0^1 g\, d\nu = c[\gamma(2) + \gamma(-2)]$. Hence $(1 - c)\gamma(2) < c\gamma(-2)$, and the conclusion follows.

Therefore $\{f_\theta(x)\}$ is both VR_∞ and SVR_2. However, SVR_3 is impossible because if g is a suitably chosen second degree polynomial, one can have $\gamma(\theta) \equiv 0$.

3. WHICH FAMILIES ARE (S)VR?

We now derive important properties of (S)VR transformations that enable one to identify (S)VR families. Theorem 3.1 is the basic theorem and has the only nontrivial proof needed in the theory. It is later used to show that exponential families are SVR and to establish the symmetry of (S)VR and its equivalence to total positivity.

Theorem 3.1. $f \in (S)VR_{n+1}(X, \Theta)$ iff $f \in (S)VR_{n+1}(X_{n+1}, \Theta_{n+1})$.

Outline of Proof. Necessity follows from the heredity property (Remark 2.2, Corollary 2.1). The sufficiency proof has two major steps.

Lemma 3.1. $(S)VR_{n+1}(X_{n+m}, \Theta_{n+1}) \Rightarrow (S)VR_{n+1}(X_{n+m+1}, \Theta_{n+1})$ for all $m \ge 1$.

Lemma 3.2. $(S)VR_{n+1}(X_f, \Theta_{n+1}) \Rightarrow (S)VR_{n+1}(X, \Theta_{n+1})$.

The conclusion of Lemma 3.1 implies (by induction on m) the hypothesis of Lemma 3.2. An easy contradiction argument shows that $(S)VR_{n+1}(X, \Theta_{n+1}) \Rightarrow (S)VR_{n+1}(X, \Theta)$. In fact, if $(S)VR_{n+1}(X, \Theta)$ fails for some pair of functions $g(x)$ and $\gamma(\theta)$, then there will exist a subset Σ of Θ containing at most $n + 1$ points for which $S^\pm(\gamma_\Sigma) = S^-(g)$, but $IS^\pm(\gamma_\Sigma) \ne IS^-(g)$. Thus for each value of n the theorem follows from Lemmas 3.1 and 3.2.

The proof of Lemmas 3.1 and 3.2 involves two induction arguments. Both lemmas, and the theorem, are clearly valid when $n = 0$ (Remark 2.5). Now, suppose the lemmas and the theorem are valid for $n \leqq N$ and that $f \in (S)VR_{N+1}(X_{N+1}, \Theta_{N+1})$. Remarks 2.3 and 2.4 show that $f \in (S)VR_N(X_N, \Theta_N)$ and, consequently, $f \in (S)VR_N(X, \Theta)$, by the induction hypothesis. Lemma 3.1 can now be established for $n = N + 1$ by induction on m. Let g be a function on X_{N+m+1} and γ the corresponding function on Θ_{N+1}. If $S^-(g) < N$, (2.1) follows from the induction hypothesis on N. In the other case, $S^-(g) = N$, we show by algebraic manipulation how to construct a proper subset X_{N+m} of X_{N+m+1} and a $\hat{g}$ on X_{N+m} such that the corresponding $\hat{\gamma}$ equals γ. The induction hypothesis on m may be applied to $\hat{g}$ and $\hat{\gamma}$, and it follows that γ has the desired property (2.1). The details of this argument are given in Appendix 1.

When g in the definition of (S)VR is restricted to be a continuous function then Lemma 3.2 is in principle an easy consequence of Lemma 3.1. One need only choose nested subsets $X_n \subset X_f$ whose union is dense in X, define corresponding measures $\nu_n \to \nu$ and pass to the limit. Presumably this type of argument can be modified and extended to treat the general situation involving an arbitrary g. We give a different proof, however, in Appendix 1 that involves the separating hyperplane theorem on Euclidean space, and avoids any such limiting arguments.

Theorem 3.1 may be immediately applied to prove that the (S)VR property is symmetric.

Theorem 3.2. $(S)VR_{N+1}(X, \Theta) \Leftrightarrow (S)VR_{n+1}(\Theta, X)$.

Proof. The proof is by induction, the case $n = 0$ being trivial. By Theorem 1, it suffices to prove the result for $(n + 1) \times (n + 1)$ matrices F. That is, we must show $F \in (S)VR \Rightarrow F^t \in (S)VR$, where F^t denotes the transpose of F.

VR *case*. A simple reduction argument like that which begins the proof of Lemma 3.1 (Appendix 1) shows we may suppose that $b = F^t\alpha$ with $\mathcal{S}^-(\alpha) = (n, +)$, $S^-(b) = n$ and rank $F^t = n + 1$. The goal is to show that $IS^-(b) = +$.

Solve $\alpha = Fg$ for g: $VR_{n+1}(X_{n+1}, \Theta_{n+1})$ ensures that $\mathcal{S}^-(g) = (n, +)$. Thus $b^tg = \alpha^tFg = \alpha^t\alpha > 0$, which implies that $IS^-(b) = IS^-(g) = +$, since b and g have length $n + 1$ with $S^-(b) = S^-(g) = n$.

SVR *case*. Now use $S^+(b)$ and $IS^+(b)$ and note that $\mathcal{S}^-(\alpha) = (n, +)$ implies $\mathcal{S}^+(\alpha) = (n, +)$ and the result follows as before. Alternatively, one can use the device given in the Proof of Lemma 3.1 for the SVR case.

Both multiplication by (strictly) positive functions and strictly monotone transformations of Θ alone, or of X alone preserve $(S)VR_{n+1}$. The proofs of the statements below are simple applications of the definitions.

Proposition 3.1. (a) If $a(\theta) > 0$, $b(x) > 0$ on Θ, X, then $f_\theta(x) \in SVR_{n+1} \Rightarrow a(\theta)b(x)f_\theta(x) \in SVR_{n+1}$. (b) If $a(\theta) \ge 0$, $b(x) \ge 0$ on Θ, X, then $f_\theta(x) \in VR_{n+1} \Rightarrow a(\theta)b(x)f_\theta(x) \in VR_{n+1}$.

Proposition 3.2. If a: $\Theta \to \Theta'$, b: $X \to X'$ are strictly monotone and onto and $f_\theta(x)$ is $(S)VR_{n+1}(X, \Theta)$, then $f'_{\theta'}(x') = f_{a^{-1}\theta}(b^{-1}x)$ is $(S)VR_{n+1}(X', \Theta')$.

Example 3.1: Exponential families are SVR_∞. If ν is a σ-finite measure on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ then the corresponding exponential family has density $f_\theta(x) = e^{\theta x - \psi(\theta)}$ with respect to ν, where $e^{\psi(\theta)} = \int e^{\theta x}\nu(dx)$. Here Θ is the interval of θ values for which the Laplace transform is finite, and X may be taken as $(-\infty, \infty)$. To demonstrate that $f_\theta(x)$ is $SVR_\infty(\Theta, x)$ (and so also $SVR_\infty(X, \Theta)$) we show, by induction, that $e^{\theta x} \in \mathrm{SVR}_n(\mathbf{R}, \mathbf{R})$, and then invoke Proposition 3.1.

Choose $\theta_1 < \ldots < \theta_{n+1}$ and $a \in \mathbf{R}^{n+1}$ satisfying $\mathcal{S}^-(a) = (n, +)$. Let

$$\begin{aligned} f(x) &= \sum_{i=1}^{n+1} a_i e^{\theta_i x} \\ &= e^{\theta_1 x} \sum_{i=1}^{n+1} a_i e^{(\theta_i - \theta_1)x} \\ &= e^{\theta_1 x} g(x), \quad \text{say.} \end{aligned}$$

By Theorem 3.1, it remains to check that $\mathcal{S}^+(f) = (n, +)$. Now

$$g'(x) = \sum_{i=2}^{n+1} a_i(\theta_i - \theta_1)e^{(\theta_i - \theta_1)x},$$

so by induction and Proposition 3.1, $S^+(g') \le n - 1$, yielding in turn by Rolle's theorem $S^+(g) \le n$. Thus $S^+(f) \le n$, and $IS^+(f) = +$ follows from the representation

$$\begin{aligned} f(x) &= a_1 e^{\theta_1 x}\left(1 + \frac{1}{a_1}\sum_{i=2}^{n+1} a_i e^{(\theta_i - \theta_1)x}\right) \\ &\sim a_1 e^{\theta_1 x} \quad \text{as } x \to -\infty. \end{aligned}$$

Example 3.2. The central $\chi^2_{(j)}$ distribution is SVR_∞ with parameter j in $\mathbf{Z}^+$. Indeed the density is $f_j(x) = x^{j/2-1} e^{-x/2}/2^{j/2}\Gamma(j/2)$, $j \in \mathbf{Z}^+$, $x \in \mathbf{R}^+$, so the only term of interest is $x^{j/2}$. But this may be rewritten as $e^{j/2 \log x}$, so the result follows from Example 3.1 and Proposition 3.2.

Proposition 3.3. The (strict) monotone likelihood ratio [(S)MLR] property for $f_\theta(x)$ on X, Θ is equivalent to $(\mathrm{S})\mathrm{VR}_2(X, \Theta)$.

Proof: $\mathrm{VR}_2 \Rightarrow$ MLR: For any $x_1 < x_2$ in X and $a \in \mathbf{R}$,

$$af_\theta(x_1) + f_\theta(x_2) \quad \text{and hence} \quad a + \frac{f_\theta(x_2)}{f_\theta(x_1)}$$

have at most one S^- sign change as functions of θ. This is true for all real a, so $f_\theta(x_2)/f_\theta(x_1)$ is monotone in θ. It then follows from the initial sign property that the ratio is actually increasing in θ.

$\mathrm{SVR}_2 \Rightarrow$ SMLR. If there were $\theta_1 < \theta_2$ for which $f_{\theta_2}(x_2)/f_{\theta_2}(x_1) = f_{\theta_1}(x_2)/f_{\theta_1}(x_1)$, then the matrix $(f_{\theta_i}(x_j))_{i,j=1,2}$ would be singular, in contravention of Remark 2.1.

(S)MLR $\Rightarrow$ $(\mathrm{S})\mathrm{VR}_2$. This may be found in Lehmann (1959, p. 74).

The following composition theorem allows one to identify many important (S)VR families from the stock already available.

Theorem 3.3. (a) If $f_\xi(x)$ and $g_\theta(\xi)$ are VR_{n+1}, then $h_\theta(x) = \int g_\theta(\xi)f_\xi(x)\mathrm{m}(d\xi)$ is VR_{n+1}. (b) If f is VR_{n+1} and g is SVR_{n+1} (or vice-versa), then h is SVR_{n+1} if m satisfies

$$\begin{aligned} S^-(\ell) \le n \quad &\text{and} \quad \int |\,\ell(x)|\, d\nu(x) > 0 \\ &\Rightarrow \int |\,\lambda(\xi)|\, dm(\xi) > 0, \quad \text{where} \qquad (3.1) \\ \lambda(\xi) &= \int f_\xi(x)\ell(x)\, d\nu(x). \end{aligned}$$

Remark. (3.1) certainly holds if f is SVR_{n+1} and the support of the measure $m(\cdot)$ contains more than n points.

Proof. We consider only (b), since (a) is similar. Suppose that $S^-(\ell) \le n$ and $\int |\,\ell(x)|\, d\nu(x) > 0$. Let $\gamma(\theta) = \int h_\theta(x)\ell(x)\, d\nu$. By Fubini's theorem,

$$\gamma(\theta) = \int g_\theta(\xi)\lambda(\xi)m(d\xi).$$

Using (3.2), $S^+(\gamma) \le S^-(\lambda) \le S^-(\ell)$, while $S^+(\gamma) = S^-(\ell)$ implies equality throughout, hence $IS^+(\gamma) = IS^-(\lambda) = IS^-(\ell)$.

Example 3.3. The noncentral χ^2 and F distributions are SVR_∞ in their noncentrality parameter λ. Recall that a $\chi^2{}_n(\lambda)$ variable may be regarded as a compound central χ^2 variable, having $n + 2\nu$ degrees of freedom, where ν is Poisson with parameter λ. Writing $g_{n,\lambda}(x)$, $f_j(x)$, and $p_\lambda(j)$ for the densities of $\chi^2{}_{(n)}(\lambda)$, $\chi^2{}_{(j)}$, and Poisson (λ) respectively, we have

$$g_{n,\lambda}(x) = \sum_{j=0}^{\infty} p_\lambda(j) f_{n+2j}(x).$$

Since the Poisson distribution belongs to the exponential family, it follows from Examples 3.1 and 3.2 and the previous theorem (m is counting measure) that $\chi^2{}_{(n)}(\lambda)$ is SVR_∞.

For the noncentral F density we note that the density h_λ of the ratio X/Y of two independent positive random variables having densities g_λ (for X) and f (for Y) may be written as

$$h_\lambda(t) = \int_0^\infty (x/t^2) f\left(\frac{x}{t}\right) g_\lambda(x)\, dx.$$

Now take X to be $\chi^2{}_n(\lambda)$, Y to be $\chi^2{}_m$, so that $f(x/t) = c_m(x/t)^{(m/2)-1}e^{-x/2t}$, which is $\mathrm{SVR}_\infty(\mathbf{R}^+, \mathbf{R}^+)$ in x and t because $t \to -\frac{1}{2}t$ is strictly monotone on $\mathbf{R}^+$. The composition theorem now applies to show that $h_\lambda(t)$ is SVR_∞. We were shown this proof by R.H. Farrell.

Other families have integral representations for their densities that permit use of the composition theorem to establish SVR_∞. These include the noncentral t density in the noncentrality parameter, the sample and multiple correlation coefficients and generalized variance, and other densities arising in multivariate analysis. For details, see Karlin (1968, 3.4).

Definition 3.1. $f(t) \ge 0$ is called a $(\mathrm{S})\mathrm{PF}_{n+1}$ (for Polya Frequency) function if $f(t - \theta)$ is in $(\mathrm{S})\mathrm{VR}_{n+1}(\mathbf{R}, \mathbf{R})$, and a $(\mathrm{S})\mathrm{PF}_{n+1}$ sequence if $f(t - \theta) \in (\mathrm{S})\mathrm{VR}_{n+1}(\mathbf{Z}, \mathbf{Z})$.

Examples of PF_∞ functions of statistical importance are the normal density $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$ and the Heaviside step function $e(x) = I_{0,\infty}(x)$ (see Example 2.2).

Proposition 3.4. (a) If X_1, X_2 are independent rv's with (S)PF$_{n+1}$ densities, then $X_1 + X_2$ has an (S)PF$_{n+1}$ density.
(b) If $f(x)$ is (S)Pf$_{n+1}$, so is $F(x) = \int_{-\infty}^{x} f(y)\,dy$.

Proof. (a) The density of $X_1 + X_2$ is the convolution of the densities of X_1 and X_2 so Theorem 3.3 can be applied. (b) $F(x - \theta) = \int_{-\infty}^{\infty} f(x - \xi)e(\theta, \xi)\,d\xi$, where e is the kernel of Example 2.2. Notice that although e is VR$_\infty$ but not SVR, Condition (3.1) holds for Lebesgue measure, so that the composition theorem applies.

4. SOME ILLUSTRATIVE SVR ARGUMENTS

We present two examples showing how basic statistical results can be derived using the SVR property as a starting point. (Refer to the Introduction for further discussion and references to many more examples.)

Example 4.1: Power Functions of One-Sided Tests. Suppose that $f(\theta, x)$ is VR$_2$ and that φ is the critical function of a one-sided test of H_0: $\theta = \theta_0$ versus H_1: $\theta > \theta_0$. That is,

$$\begin{aligned}\varphi(x) = \Pr(\text{reject } H_0 \mid X = x) &= 1, \quad \text{if } x > x_0,\\ &= \lambda, \quad \text{if } x = x_0,\\ &= 0, \quad \text{if } x < x_0,\end{aligned}$$

where $0 \le \lambda \le 1$. For any constant γ, $S^-(\varphi(X) - \gamma) \le 1$; so the VR property implies that $S^-(E_\theta\varphi(X) - \gamma) \le 1$. Thus the power function of any one-sided test must be monotone. If f is SVR$_2$, and S^+ is used in place of S^-, we find that the power function is *strictly* monotone.

Example 4.2: UMP Unbiased Two-Sided Tests on a Single Parameter. Consider the problem of testing H_0: $\theta = \theta_0$ versus H_1: $\theta \ne \theta_0$ where the family $\{f_\theta(x)\}$ of probability densities wrt some σ-finite measure μ is SVR$_3$. Karlin's approach to this problem was decision theoretic (see also Ferguson 1967, 5.3 for a special case) and use of the SVR property changes only the mathematical tools needed for the argument rather than its general outline. Lehmann used an optimization theory approach based on the generalized Neyman Pearson (GNP) Lemma (Lehmann, 1959, 3.5), but restricted consideration to one-parameter exponential families. We use this example both to illustrate the recasting of Karlin's argument and to indicate the extension of Lehmann's approach to SVR$_3$ families, which accommodates distributions such as those of noncentral t and the correlation coefficient (cf. also Lehmann 1959, Problems 3.24, 3.25). We shall assume that (a) all power functions $\beta_\varphi(\theta)$ are continuous and may be differentiated under the integral, and (b) $\mu\{x: (\partial/\partial\theta)\log f_{\theta_0}(x) \ne 0\} > 0$. In other words, the Fisher information (e.g., Rao 1973, p. 379) about θ contained in X is positive at θ_0. We shall establish the existence of a unique UMPU test.

The first step is to note that the family $\mathcal{M}_2$ of procedures with critical functions of the form

$$\begin{aligned}\varphi(x) &= 1, \quad \text{if } x < x_1 \text{ or } x > x_2\\ &= \lambda_i, \quad \text{if } x = x_i, \quad i = 1, 2\\ &= 0, \quad \text{if } x_1 < x < x_2\end{aligned} \tag{4.1}$$

form a complete class for this testing problem. Here $-\infty \le x_1 \le x_2 \le +\infty$ and $\lambda_i \in [0, 1]$. For the proof, see Karlin (1956). A modified form of Karlin's argument, using SVR$_3$ methods, will now be sketched. Let φ be an arbitrary critical function. By the complete class theorem for Bayes procedures for decision problems with finite parameter set (Ferguson 1967, p. 87), there is a test $\tilde\varphi_\epsilon$ dominating φ on the set $\Theta_\epsilon = \{\theta_0 - \epsilon, \theta_0, \theta_0 + \epsilon\}$ that is Bayes with respect to some prior $\{\pi_{-1}, \pi_0, \pi_{+1}\}$ on Θ_ϵ. A little calculation shows that $\tilde\varphi_\epsilon$ must have the form

$$\begin{aligned}\tilde\varphi_\epsilon(x) &= 1\\ &\qquad \text{when } \pi_{-1}f_{\theta_0-\epsilon} - \pi_0 f_{\theta_0} + \pi_1 f_{\theta_0+\epsilon} > 0\\ &= 0\\ &\qquad \text{when } \pi_{-1}f_{\theta_0-\epsilon} - \pi_0 f_{\theta_0} + \pi_1 f_{\theta_0+\epsilon} < 0.\end{aligned} \tag{4.2}$$

The SVR$_3$ property now implies that $\tilde\varphi_\epsilon(x) \in \mathcal{M}_2$. Hence $S^-(\varphi - \tilde\varphi_\epsilon) \le 2$ and so by SVR$_3$ again, $\beta_{\varphi_\epsilon}(\theta) \ge \beta_\varphi(\theta)$ for all $\theta \notin (\theta_0 - \epsilon, \theta_0 + \epsilon)$. Tests in $\mathcal{M}_2$ are determined by four parameters, hence we can extract a subsequence φ_{ϵ_n} and limit $\varphi_0 \in \mathcal{M}_2$ such that $\varphi_{\epsilon_n}(x) \to \varphi_0(x)$ for all $x \in \mathbf{R}$. It follows that φ_0 dominates φ, and if $\mu\{\varphi_0 \ne \varphi\} > 0$, then $S^+(\beta_\varphi - \beta_{\varphi_0}) \le 2$, so that φ_0 is strictly better than φ except at θ_0. Hence $\mathcal{M}_2$ is a complete class.

Remark 1. If we seek a test that minimizes $\beta_{\varphi_\epsilon}(\theta_0)$ subject to $\beta_{\varphi_\epsilon}(\theta_0 \pm \epsilon) = \beta_\varphi(\theta_0 \pm \epsilon)$, then Lehmann's GNP lemma implies that a test of the form (4.2) results. Thus, an approach independent of decision theory is also possible.

Let θ_1 be a fixed alternative and φ_1 a test that maximizes $\beta_\varphi(\theta_1)$ among unbiased tests of size α. By the complete class theorem, we may assume that $\varphi_1 \in \mathcal{M}_2$. Since φ_1 is of size α and unbiased, it must be genuinely two-sided, that is, x_1 and x_2 are both finite.

There is in fact only one unbiased size α test $\in \mathcal{M}_2$. To show this we note that any unbiased size α test satisfies $\beta'_\varphi(\theta_0) = 0$. Suppose that φ_2 is another size α test $\in \mathcal{M}_2$ (with x_1, x_2 finite) satisfying $\beta'_{\varphi_2}(\theta_0) = 0$. Since both tests are of the form (4.1), with x_i finite, and since both have size α, either $\varphi_1 = \varphi_2$ ae (μ) or $S^-(\varphi_1 - \varphi_2) = 1$. On the other hand, $\beta'_{\varphi_1}(\theta_0) - \beta'_{\varphi_2}(\theta_0) = E_{\theta_0}h(\varphi_1 - \varphi_2)$ where $h(x) = (\partial/\partial\theta)\ln f_{\theta_0}(x)$. Let us accept for the moment the truth of the following lemma, which will be proved in Appendix 1.

Lemma (4.1). If $f_\theta(x)$ is SVR_3, Θ contains a neighborhood of θ_0, and $h(x) = (\partial/\partial\theta) \ln f_\theta(x)|_{\theta=\theta_0}$ exists, then either h is identically constant or strictly increasing.

Since $\beta'_{\varphi_1}(\theta_0) = E_{\theta_0} h\varphi_1 = 0$, the only possible constant value for h is zero, and this is excluded by the assumption of positive Fisher information. Thus h is strictly increasing. If $\mu\{\varphi_1 \neq \varphi_2\} > 0$, then $A = \{\varphi_1 < \varphi_2\}$ and $B = \{\varphi_1 > \varphi_2\}$ are disjoint intervals with $E_{\theta_0}(\varphi_2 - \varphi_1)\chi_A = E_{\theta_0}(\varphi_1 - \varphi_2)\chi_B > 0$. But h is strictly increasing, so that $E_{\theta_0} h(\varphi_1 - \varphi_2) \neq 0$, which is a contradiction. Hence $\varphi_1 = \varphi_2$ a.e. Thus φ_1 is the unique UMPU test for this problem.

Remark 2. The condition (b) is needed only for our (and Karlin's) argument but it is not clear that it is required for the result. If $X \sim N(\theta^3, 1)$, then all power functions have zero derivative at zero, and yet a unique UMPU test exists (as may be seen from the obvious reparametrization).

Remark 3. Similar (indeed, substantially simpler) SVR_3 arguments may be given to establish the existence, form and unicity of UMP or UMPU tests for the other two sided hypotheses considered in Lehmann (1959, 3.7 and 4.2).

APPENDIX 1: COMPLETION OF PROOF OF THEOREM 3.1

Proof of Lemma 3.1: VR *case.* Choose $n + m + 1$ points $x_1 < \ldots < x_{n+m+1}$ and $n + 1$ points $\theta_1 < \ldots < \theta_{n+1}$. Let $F = (f_{\theta_i}(x_j))$ and label its columns $f_1, \ldots, f_{n+m+1}$. Suppose $\beta = Fa$. β has $n + 1$ components, a has $n + m + 1$. We are concerned only with vectors a for which $S^-(a) \leq n$. If $S^-(a) < n$, β must satisfy (2.1), since $f \in VR_n(X, \Theta)$ by induction. So it remains only to consider the case $\mathcal{S}^-(a) = (n, +)$, for if $IS^-(a) = -$, just replace a by $-a$.

If a_{j_0} were 0 for some j_0, then forming $\hat{F}$ by deleting the j_0th column of F and $\hat{a}$ by deleting a_{j_0} from A, one could write $\beta = \hat{F}\hat{a}$ and deduce from $f \in VR_{n+1}(X_{n+m}, \Theta_{n+1})$ that β satisfied (2.1). Hence we may assume that $a_j \neq 0$ for all j. Finally, since β has only $n + 1$ components and $S^-(a) = n$, we may assume that $S^-(\beta) = n$. Our goal is now to show that $IS^-(\beta) = +$.

The rank of F must be $n + 1$. Otherwise β would be a linear combination of n columns of F and one could write $\beta = Fe$, where e has at most n nonzero coordinates. Hence $S^-(e) \leq n - 1 < n = S^-(\beta)$, which contradicts $VR_n(X_{n+1}, \Theta_{n+1})$.

Therefore choose columns $(f_{j_1}, \ldots, f_{j_{n+1}}) = F_1$ so that F_1 is invertible. Let $T = \{j_1, \ldots, j_{n+1}\}$. Solve the equation $\beta = F_1 b^0$ for b^0, noting that $\mathcal{S}^-(b^0) = (n, IS^-(\beta))$ from $VR_{n+1}(X_{n+1}, \Theta_{n+1})$. Put $b_{j_i} = b_i^0$ and $b_j = 0$ for $j \notin T$. For the extended b also, $\mathcal{S}^-(b) = (n, IS^-(\beta))$ and

$$(1 + c)\beta = F(a + cb), \quad c \in \mathbf{R}.$$

Let $T' = \{j: a_j b_j < 0\}$. If T' is empty, $a_j b_j > 0$ for all $b_j \neq 0$, so $IS^-(b) = IS^-(a) = +$, and hence $IS^-(\beta) = +$. If T' is nonempty, let $\hat{c} = \min_{j \in T'} |a_j/b_j| > 0$. Then $d = a + \hat{c}b$ has a zero component, but still $d_i a_i \geq 0$ for all i, so that $S^-(d) \leq S^-(a) = n$. Omit this component and the corresponding column from F obtaining $\hat{d}$ and $\hat{F}$ and the equation $(\hat{c} + 1)\beta = \hat{F}\hat{d}$. By $VR_{n+1}(X_{m+n}, \Theta_{n+1})$ it follows that $\mathcal{S}^-(\hat{d}) = (n, IS^-(\beta))$. Thus $S^-(\hat{d}) = S^-(a)$ and so $d_i a_i \geq 0$ implies that $IS^-(\hat{d}) = IS^-(a) = +$. Hence $IS^-(\beta) = +$ in this case also.

SVR *case.* This time we may suppose that $\mathcal{S}^-(a) = (n, +)$, all $a_j \neq 0$ and $S^+(\beta) = n$, and shall show that $IS^+(\beta) = +$. Choose $\beta_m \to \beta$ so that $S^-(\beta_m) = n$. Since rank $F = n + 1$ (by Remark 2.1), there exist $a_m \to a$ such that $Fa_m = \beta_m$ (take $a_m = F'(FF')^{-1}\beta_m + (I - F'(FF')^{-1}F)a$). For large m, $\mathcal{S}^-(a_m) = (n, +)$ so the argument in the VR case shows that $IS^-(\beta_m) = +$. Hence $IS^+(\beta) = +$ and this completes the proof of Lemma 1.

Proof of Lemma 3.2: VR *case.* Let ν and g be given, and $\gamma(\theta) = \int f_\theta(x)g(x)\nu(dx)$. Since $f \in VR_n(X, \Theta)$ by the induction on n, we may assume, as in Lema 3.1, that $\mathcal{S}^-(g) = (n, +)$. Our goal is to show that $IS^-(\gamma) = +$. To this end let $\Sigma = \{\theta_1 < \ldots < \theta_{n+1}\} \subset \Theta$ be chosen so that $S^-(\gamma_\Sigma) = n$. It will suffice to show that $IS^-(\gamma_\Sigma) = +$.

Let

$$C = \{\eta \in R^{n+1}: \eta_j = \sum_{i=1}^{N} a_i f_{\theta_j}(x_i) g(x_i), \quad a_i \geq 0,$$

$$\sum_{i=1}^{N} a_i = 1, \quad N \geq 1, \quad x_1 < \ldots < x_N \in X\}$$

and

$$D = \{\eta \in R^{n+1}: \mathcal{S}^-(\eta) = (n, -)\}.$$

C and D are convex and $VR_{n+1}(X_f, \Theta_{n+1})$ implies $C \cap D = \emptyset$. By the separating hyperplane theorem in $\mathbf{R}^{n+1}$, there exists $\beta \neq 0$ such that

$$\beta \cdot \eta \geq \beta \cdot \zeta \quad \text{for all} \quad \eta \in C, \zeta \in D.$$

In particular $\beta \cdot \eta \geq 0$ for all $\eta \in C$, $(0 \in \bar{D}!)$, so

$$g(x) \sum_{j=1}^{n+1} \beta_j f_{\theta_j}(x) \geq 0 \quad \text{for all } x \in X.$$

Now if $IS^-(\gamma_\Sigma) = -$, then $\gamma_\Sigma \in \text{int } D$. Moreover, the normal vector $-\beta$ to the hyperplane lies in $\bar{D}$, because D is an orthant of $\mathbf{R}^{n+1}$. So, putting $\gamma_j = \gamma(\theta_j)$

$$0 > \sum_j \beta_j \gamma_j = \sum_j \beta_j \int f_{\theta_j}(x) g(x) \nu(dx) \geq 0,$$

which is an impossibility. Hence $IS^-(\gamma_\Sigma) = +$, as required.

SVR *case.* This time we assume $S^+(\gamma) = n$ and aim to show that $IS^+(\gamma) = +$. The function $m(x) = \sum_{j=1}^{n+1} \beta_j f_{\theta_j}(x)$ can have at most n zeros in X, since Remark 2.1 guarantees that any $(n + 1) \times (n + 1)$ matrix $(f_{\theta_j}(x_i))$ is nonsingular. We may further assume that $|g(z)|$

$\nu(dz)$ is not a discrete measure on a finite number of points, for we would then conclude that $IS^+(\gamma) = +$ from Lemma 3.1. Thus, if $IS^+(\gamma) = -$ or 0 (implying $\gamma \in \bar{D}$), we would obtain the contradiction

$$0 < \int_X m(x)g(x)\, d\nu(x) = \sum_j \beta_j \gamma_j \leq 0.$$

Hence $IS^+(\gamma) = +$. This completes the proof of Lemma 3.2 and the theorem.

Proof of Lemma 4.1. First note that $h(x) = c$ iff $\tilde{h}(x) = 0$, where $\tilde{h}(x) = (\partial/\partial\theta) \ln e^{-c\theta} f_\theta(x)|_{\theta=\theta_0}$ and $e^{-c\theta} f_\theta(x)$ is SVR_3 by Proposition 3.1. It therefore suffices to prove that if h is not identically zero, then h can have at most one zero. By invoking Proposition 3.1, we may assume without loss of generality that $f_{\theta_0}(x) = 1$ for all x, so that $h(x) = (\partial/\partial\theta) f_{\theta_0}(x)$. By virtue of SVR_2, $f_{\theta_0+h}(x) - f_{\theta_0}(x)$ is strictly increasing (for $h > 0$) and hence $h(x)$ is nondecreasing. If h has at least two zeros, then there exist $x_{-1} < x_0 < x_1$ such that either $h(x_{-1}) < h(x_0) = 0 = h(x_1)$ or $h(x_{-1}) = h(x_0) = 0 < h(x_1)$. Suppose the former (a symmetric argument will apply to the latter case). By a further appeal to Proposition 3.1, we may suppose that $f_\theta(x_0) = 1$ for all θ. Indeed, let $f^*_\theta(x) = f^*_\theta(x)/f_\theta(x_0)$. Then f^*_θ is SVR_3, $f^*_{\theta_0} \equiv 1$, $f^*_\theta(x_0) \equiv 1$, and $h(x) = (\partial/\partial\theta) f^*_{\theta_0}(x)$ satisfies $h^*(x) = h(x)$ because $h(x_0) = 0$.

Choose now $\theta_{-1} < \theta_0$. Since $f_{\theta_0}(x_0) = f_{\theta_{-1}}(x_0)$, we may choose $b > 0$ small enough so that $h(x_{-1}) < b(f_{\theta_0}(x_{-1}) - f_{\theta_{-1}}(x_{-1})) < 0$ using SVR_2. Note also from SVR_2 that $f_{\theta_0}(x_1) - f_{\theta_{-1}}(x_1) > 0$. Then

$$\lim_{h \downarrow 0} [f_{\theta_0+h}(x_i) - f_{\theta_0}(x_i)]/h - b(f_{\theta_0}(x_i) - f_{\theta_{-1}}(x_i))$$
$$= h(x_i) - b(f_{\theta_0}(x_i) - f_{\theta_{-1}}(x_i)),$$

which is < 0, $= 0$, < 0 according as $i = -1, 0, 1$. Hence for $h = h_0 > 0$ sufficiently small, $g(x) = bf_{\theta_{-1}}(x) - (b + h_0^{-1}) f_{\theta_0}(x) + h_0^{-1} f_{\theta_0+h_0}(x)$ satisfies $S^+(g) \geq 3$ or $S^+(g) = (2, -)$. This contradicts the SVR_3 property of f and completes the proof of the lemma.

APPENDIX 2: EQUIVALENCE OF (S)VR AND (STRICT) TOTAL POSITIVITY

Definition. (Karlin 1968, p. 11). A function $f(x, \theta)$, defined for $x \in X \subset \mathbf{R}$, $\theta \in \Theta \subset \mathbf{R}$ is called (strictly) totally positive of order r (STP_r) if whenever $x_1 < \ldots < x_m$; $\theta_1 < \ldots < \theta_m$,

$$f\begin{pmatrix} x_1, \ldots, x_m \\ \theta_1, \ldots, \theta_m \end{pmatrix} = \det(f(x_i, \theta_j))_{i,j=1,\ldots,m}$$
$$\geq 0 \text{ (resp. } > 0), \ 1 \leq m \leq r.$$

An $n \times m$ matrix U is therefore $(S)TP_q$ ($q \leq m \wedge n$) if and only if all its $q \times q$ minors are nonnegative (positive).

Remark A.1. Notice that from the very definition, $(S)TP_{n+1}(X, \Theta) \Leftrightarrow (S)TP_{n+1}(X_{n+1}\Theta_{n+1})$.

Theorem A.1. f is $(S)VR_{n+1}(X, \Theta)$ iff it is $(S)TP_{n+1}(X, \Theta)$.

Proof. The outline is as follows:

$SVR_{n+1}(X, \Theta)$

(by Theorem 3.1)

$\Leftrightarrow SVR_{n+1}(X_{n+1}, \Theta_{n+1})$

(by Lemma A.1)

$\Leftrightarrow TP(X_{n+1}, \Theta_{n+1})$

(by Remark A.1)

$\Leftrightarrow TP(X, \Theta)$

It therefore remains only to establish the following lemma.

Lemma A.1. Let U be a $q \times q$ matrix. U is $(S)VR_q$ iff U is $(S)TP_q$.

Proof if U is SVR_q. The proof is by induction, the case $p = 1$ being trivial. Suppose that U is STP_p and let $U_0 = (u_{i_\nu j_\mu})$, $\nu, \mu = 1, \ldots p + 1$ be an arbitrarily chosen $(p + 1) \times (p + 1)$ submatrix of U.

Choose $\beta \in \mathbf{R}^{p+1}$ with $\mathcal{S}^+(\beta) = \mathcal{S}^-(\beta) = (p + 1, +)$ and, recalling that U_0 is nonsingular, solve $\beta = U_0 a$ for a, with $\mathcal{S}^-(a) = (p + 1, +)$. By Cramer's rule,

$$a_1 = \frac{1}{|U_0|} \begin{vmatrix} \beta_1 & u_{i_1 j_2} & \cdots & u_{i_1 j_{p+1}} \\ \vdots & & & \\ \beta_{p+1} & u_{i_{p+1} j_2} & \cdots & u_{i_{p+1} j_{p+1}} \end{vmatrix}.$$

Hence,

$$|U_0| = \frac{1}{a_1} \sum_{k=1}^{p+1} (-1)^{k+1} \times \beta_k U_0 \begin{pmatrix} i_1 \ldots \hat{i}_k \ldots i_{p+1} \\ j_2 \ldots\ldots\ldots j_{p+1} \end{pmatrix}, \quad \text{(A.1)}$$

where $\hat{i}_k$ means that i_k is omitted. Now $a_1 > 0$; $(-1)^{k+1}\beta_k > 0$; so by the induction hypothesis, $|U_0| > 0$, as required.

If U is VR_q. The proof is similarly by induction. Suppose U is TP_p but not TP_{p+1}. Then there is a $(p + 1) \times (p + 1)$ submatrix U_0 with $|U_0| < 0$. But we may obtain (A.1) almost exactly as before, which would imply $|U_0| \geq 0$ by TP_p; a contradiction.

If U is STP_q. By induction we may suppose that U is SVR_{q-1}. Let $\beta = Ua$. The only new case is $\mathcal{S}^-(a) = (q - 1, +)$, $S^+(\beta) = q - 1$. Since U is nonsingular, $IS^+(\beta) \neq 0$, and furthermore (A.1) holds with U_0 replaced by U. In (A.1) all the minors are positive, as are a_1 and $|U|$, while $(-1)^{k+1}\beta_k$ has a constant sign, which must therefore be positive. Hence $IS^+(\beta) = +$, as required.

If U is TP_q. The argument is analogous to the STP_q case, except that the nonsingularity of U follows from the assumption that $S^-(\beta) = q$, exactly as in the third paragraph of the proof of Lemma 3.1 (VR case).

Remark. The proof of Lemma A.1 is essentially contained in Karlin (1968, 5.1 and 5.2) but is enormously simplified by both the restriction to total positivity (Karlin discussed the more general concept of "sign regularity") and the emphasis on the initial sign condition in the definition of (S)VR.

[*Received September 1980. Revised May 1981.*]

REFERENCES

COHEN, A. (1965), "A Hybrid Problem on the Exponential Family," *Annals of Mathematical Statistics*, 36, 1185.
FARRELL, R.H. (1968), "On the Admissibility at ∞, Within the Class of Randomized Designs, of Balanced Designs," *Annals of Mathematical Statistics*, 39, 1978–94.
FERGUSON, T.S. (1967), *Mathematical Statistics*, New York: Academic Press.
KANTER, MAREK (1975), "Stable Densities under Change of Scale and Total Variation Inequalities," *Annals of Probability*, 3, 697–707.
KARLIN, SAMUEL (1956), "Decision Theory for Pólya Distributions, Case of Two Actions, I," *Proceedings of the Third Berkeley Symposium on Probability and Statistics* (vol. 1), Berkeley, Calif.: University of California Press, 115–129.
——— (1957a), "Polya-Type Distributions, II," *Annals of Mathematical Statistics*, 28, 281–308.
——— (1957b), "Polya-Type Distributions, III: Admissibility for Multiaction Problems," *Annals of Mathematical Statistics*, 28, 839–60.
——— (1958), "Polya-Type Distributions, IV: Some Principles of Selecting a Single Procedure from a Complete Class," *Annals of Mathematical Statistics*, 29, 1–21.
——— (1968), *Total Positivity*, Stanford, Calif.: Stanford University Press.
KIEFER, J.C. (1958), "On the Non-Randomized Optimality and Randomized Non-Optimality of Symmetrical Designs," *Annals of Mathematical Statistics*, 29, 675–99.
LEHMANN, E.L. (1959), *Testing Statistical Hypotheses*, New York: John Wiley.
LYNCH, JAMES (1979), "Sign Changes of the Difference of Convex Functions and Their Application to Large Deviation Rates," *Annals of Probability*, 7, 96–108.
MARSHALL, A.W., and OLKIN, I. (1968), "A General Approach to Some Screening and Classification Problems," *Journal of The Royal Statistical Society*, Ser. B, 30, 407.
MEEDEN, G. (1971), "Finding Best Tests Approximately for Testing Hypotheses About a Random Parameter," *Annals of Mathematical Statistics*, 42, 1452.
RAO, C.R. (1973), *Linear Statistical Inference*, New York: John Wiley.
RINOTT, Y., and SANTNER, T.J. (1977), "An Inequality for Mulivariate Normal Probabilities With Applications to a Design Problem," *Annals of Statistics*, 5, 1228–1234.
VAN ZWET, W.R., and OOSTERHOFF, J. (1967), "On the Combination of Independent Test Statistics," *Annals of Mathematical Statistics*, 38, 659–680.