CIRCUITS SYSTEMS SIGNAL PROCESSING
VOL. 16, NO. 1, 1997, PP. 83–89

ON IDENTIFIABILITY AND INFORMATION-REGULARITY IN PARAMETRIZED NORMAL DISTRIBUTIONS*

Bertrand Hochwald[1] *and Arye Nehorai*[2]

Abstract. We describe methods to establish identifiability and information-regularity of parameters in normal distributions. Parameters are considered identifiable when they are determined uniquely by the probability distribution and they are information-regular when their Fisher information matrix is full rank. In normal distributions, information-regularity implies local identifiability, but the converse is not always true. Using the theory of holomorphic mappings, we show when the converse is true, allowing information-regularity to be established without having to explicitly compute the information matrix. Some examples are given.

1. Introduction

Two criteria are often used to determine that the parametrization of a random variable's distribution is well behaved. Suppose an unknown parameter θ is an element of a set $\Theta \subset \mathbb{R}^p$, and $y \in \mathbb{R}^m$ is an observed random variable with distribution $F(y, \theta)$. The first criterion is an identifiability condition. For any $\theta' \in \Theta$ different from θ, $F(y, \theta') \neq F(y, \theta)$ for at least one value of y. The second is an information-regularity condition dependent on the existence of a differentiable density, $f(y, \theta)$: The Fisher information matrix of θ,

$$\mathrm{FIM}(\theta) \triangleq \mathrm{E}_\theta\{(\partial/\partial\theta)\log f(y,\theta)[(\partial/\partial\theta)\log f(y,\theta)]^T\} \in \mathbb{R}^{p\times p},$$

is full rank.

Satisfying these conditions is generally important. Assuming a collection of infinitely many independent identically distributed observations is available, Wald,

* Received August 22, 1995; accepted January 29, 1996. This work was supported by the Air Force Office of Scientific Research under Grant no. F49620-93-1-0096, the National Science Foundation under Grant no. MIP-9122753, and the Office of Naval Research under Grant no. N00014-91-J-1298.
[1] Mathematics of Communications Research Department, Lucent Technologies Bell Laboratories, Murray Hill, NJ, 07947.
[2] Department of Electrical Engineering and Computer Science, University of Illinois at Chicago, Chicago, Illinois.

in [7], uses identifiability to guarantee almost sure convergence of a class of θ estimates that includes the maximum likelihood estimate. In [8], Wald uses information-regularity to prove asymptotic normality of the maximum likelihood estimate.

Verifying information-regularity is troublesome, especially when p is large. We show, however, that there is an intimate connection between regularity and identifiability when the observations have a normal distribution. The connection is exploited to give a useful method, based on the powerful theorems of holomorphic functions in several complex variables, to check regularity without computing the Fisher information matrix.

When the identifiability condition holds for all $\theta \in \Theta$, we call Θ *identifiable.* Elements of Θ for which FIM(θ) is full rank are called *information-regular* or simply *regular*, and the remaining elements are called *singular.*

In the next section we characterize identifiability and regularity for observations with a normal distribution. Section 3 presents the main result, Theorem 1, which connects the two conditions when the distribution's mean and covariance are holomorphic mappings. We conclude with some examples and a summary.

2. Characterization of identifiability and regularity

Identifiability has a simple characterization when $F(y, \theta)$ is normal. Suppose $f(y, \theta)$ is a normal density with mean $\mu(\theta)$ and covariance $R(\theta)$. Define $\varphi(\theta)$ to be a vector comprising the distinct elements of $\mu(\theta)$ and $R(\theta)$ that are explicit functions of θ. Then it is clear that φ maps $\mathbb{R}^p$ to $\mathbb{R}^q$ where $q \leq m + m(m+1)/2$, and Θ is identifiable if and only if the map $\theta \mapsto \varphi(\theta)$ is injective on Θ.

Suppose φ_i, $i = 1, \ldots, q$, are the coordinate functions of φ and there exists a set of indices, $i_1, \ldots, i_p \in \{1, \ldots, q\}$, making the map $\theta \mapsto \varphi^*(\theta) \triangleq (\varphi_{i_1}(\theta), \ldots, \varphi_{i_p}(\theta))$ injective on Θ. Then we call Θ *strongly identifiable* and φ^* a *representative mapping* of Θ. Strongly identifiable sets are identifiable, but the converse is not necessarily true.

Regularity has the following characterization when $\varphi(\theta)$ is continuously differentiable. FIM(θ) > 0, or θ is a regular point, if and only if $\partial\varphi(\theta)/\partial\theta$ has (full) rank p. This result is a direct consequence of Lemma 6.1 of Chapter 9 in [1].

When p is large, both $\partial\varphi(\theta)/\partial\theta$ and FIM(θ) are often cumbersome and have ranks that are difficult to compute. There is, however, a simple way to find singular points of θ without computing either matrix. If $\partial\varphi(\theta)/\partial\theta$ has rank p, then there exists a set of p indices, $j_1, \ldots, j_p \in \{1, \ldots, q\}$, and a mapping, $\varphi^{**} \triangleq (\varphi_{j_1}, \ldots, \varphi_{j_p})$ of Θ into $\mathbb{R}^p$, such that $\partial\varphi^{**}(\theta)/\partial\theta$ is full rank. The inverse function theorem then implies that there exist open neighborhoods U of θ and V of $\varphi^{**}(\theta)$ on which $\varphi^{**} : U \to V$ is a bijection. Hence, φ^{**} is a representative mapping of U, and U is strongly identifiable. We restate the result in a more useful form.

Proposition 1. *If $\theta \in \Theta$ is a point having no strongly identifiable open neighborhood in $\mathbb{R}^p$, then θ is a singular point.*

We illustrate this result with a simple example, to which we will also refer later, that shows that the condition in Proposition 1 is not necessary to the proposition's conclusion.

Example 1. Consider $y \sim \mathcal{N}(\theta^k, 1)$, where k is a positive integer. Then $\varphi(\theta) = \theta^k$; if k is even, $\Theta = \mathbb{R}^+$ is identifiable, and if k is odd; $\Theta = \mathbb{R}$ is identifiable. For any $\theta \neq 0$ there exists an open subset of $\mathbb{R}$ containing θ such that $\varphi(\theta)$ is injective. This is not true, however, when $\theta = 0$ and k is even. Hence, by Proposition 1, $\theta = 0$ is a singular point whenever k is even. But note that $\mathrm{FIM}(\theta) = k^2\theta^{2k-2}$, and therefore $\theta = 0$ is a singular point for all $k > 1$, which Proposition 1 fails to notice when k is odd. A remedy is proposed in the next section.

3. Main result

Proposition 1 does not necessarily find all the singular points of Θ. In this section we present a method for finding all the singular points by examining φ as a function of complex arguments. In Example 1 with $k > 1$, as a holomorphic function of $z \in \mathbb{C}$, $\varphi(z)$ does not map open sets containing $z = 0$ injectively. As we will show, this will allow us to conclude that $z = 0$ is a singular point.

The reward for viewing the parametrization in complex variables is a near converse to Proposition 1. We assume $\Omega \subset \mathbb{C}^p$ is a set chosen so that $\Theta \subset \Omega$, and we carry over from the previous section the notions of identifiability and regularity to points of Ω. That is, Ω is identifiable if the map $z \mapsto \varphi(z)$ is injective on Ω, and $z \in \Omega$ is regular if $\partial\varphi(z)/\partial z$ has full rank. Theorem 1 is the main result.

Theorem 1. *Let $\varphi(z) \in \mathbb{C}^q$ be a holomorphic mapping of $z \in \Omega \subset \bigcup_{\alpha\in\mathcal{A}} \Omega_\alpha$, where for each α in an index set $\mathcal{A}$, $\Omega_\alpha \subset \mathbb{C}^p$ is open.* (a) *Suppose $z \in \Omega$ is a point having no strongly identifiable open neighborhood in $\mathbb{C}^p$. Then z is a singular point.* (b) *Conversely, suppose that for each α there exists a representative mapping $\varphi_\alpha^* : \Omega_\alpha \to \mathbb{C}^p$, making Ω_α strongly identificable. Then every point of Ω is regular.*

Proof. Assertion (a) follows exactly as in Proposition 1, from the inverse mapping theorem for holomorphic mappings [3, Theorem C6]. Assertion (b) comes from the following powerful lemma, which has no counterpart for real mappings.

Lemma 1 [4, Corollary E10]. *If $\pi : U \to V$ is a bijective holomorphic mapping from an open set $U \subset \mathbb{C}^p$ onto a subset $V \subset \mathbb{C}^p$, then V is necessarily an open subset of $\mathbb{C}^p$ and π is a biholomorphic mapping.*

A biholomorphic mapping is, by definition, a holomorphic mapping admitting a holomorphic inverse. It follows from $z = (\pi \circ \pi^{-1})(z)$ that $(\partial\pi(z)/\partial z)^{-1} = \partial\pi^{-1}(z)/\partial z$, so $\partial\pi(z)/\partial z$ is nonsingular; a biholomorphic mapping has a full-rank Jacobian matrix.

Let $z \in \Omega$. Then $z \in \Omega_\alpha$ for some $\alpha \in \mathcal{A}$. By assumption, there exists a representative mapping $\varphi_\alpha^*(z) : \Omega_\alpha \to V \subset \mathbb{C}^p$, where Ω_α is open. As a consequence of Lemma 1, $\partial\varphi_\alpha^*(z)/\partial z$ is full rank, and because $\partial\varphi_\alpha^*(z)/\partial z$ is a submatrix of $\partial\varphi(z)/\partial z$, the latter also has full rank. Therefore, z is regular. □

In the special case when $q = p$ and Ω is open, we may take $\Omega_\alpha = \Omega$. We then have the corollary that identifiability and regularity coincide on Ω.

Corollary 1. *Let $\varphi(z) : \Omega \to \mathbb{C}^p$ be a holomorphic mapping where $\Omega \subset \mathbb{C}^p$ is open. If all points of Ω are regular, then Ω is (strongly) identifiable. Conversely, if Ω is (strongly) identifiable, then all its points are also regular.*

To apply Theorem 1, one must show that $\varphi(z)$ is a holomorphic mapping; that is, the coordinate functions $\varphi_i(z)$, $i = 1, \ldots, q$, are holomorphic on Ω. This task is made simple by a theorem due to Hartogs.

Hartogs' Theorem [3, Theorem B6]. *If a complex-valued function is holomorphic in each variable separately in an open domain $U \subset \mathbb{C}^p$, then it is holomorphic in U.*

Thus we may use standard tools (power series expansions, etc.) to show that each $\varphi_i(z)$ is holomorphic in $z_1, \ldots, z_p$ separately and, by Hartogs' theorem, conclude that $\varphi(z)$ is a holomorphic mapping. As with Lemma 1, this theorem is peculiar to holomorphic functions and has no counterpart for real functions.

$\varphi(z)$ is often holomorphic because, in many applications, a parametrization is the result of mathematical analysis of a physical problem, and many functions encountered in physics are holomorphic. We present some examples to illustrate the results.

Example 2. In Example 1, $\varphi(z) = z^k$, $k > 1$, injectively maps open sets in $\mathbb{C}$ containing a if and only if $a \neq 0$. Thus, by Theorem 1, $z = 0$ is the one and only singular point whenever $k > 1$.

The next example demonstrates that strong identifiability, and not just identifiability, of some open Ω is needed for regularity.

Example 3. Let

$$y = \begin{bmatrix} \theta^2 \\ \theta^3 \end{bmatrix} + e$$

where $e \in \mathbb{R}^2$ is a zero-mean normal random variable with known covariance. In this example, $\varphi(z) = (z^2, z^3)$ has an inverse for all $z \in \mathbb{C}$, implying that $\mathbb{C}$ is identifiable. Observe, however, that $z = 0$ has no strongly identifiable open neighborhood because neither z^2 nor z^3 can be taken as a representative mapping of open sets containing the origin. It follows from Theorem 1(a) that the origin is a singular point.

The final example shows that regularity can be much easier to prove using Theorem 1, rather than explicitly computing the rank of $\partial\varphi(\theta)/\partial\theta$ or FIM(θ).

Example 4 (adapted from [2]). Consider the problem of fitting n exponentials to m time-samples. The model for the jth sample is $y_j = \sum_{i=1}^{n} x_i e^{\alpha_i t_j} + e_j$, $j = 1, \ldots, m$, where $\alpha_i \in \mathbb{R}$ is the ith signal's decay or growth rate, and e_j is the jth component of $e \in \mathbb{R}^m$, a normal random variable with zero mean and arbitrary known covariance. With $t_1 = 0$ and integer $t_1 < t_2 < \cdots < t_m$, $y = A(\alpha)x + e$, where $y = [y_1, \ldots, y_m]^T$, $x = [x_1, \ldots, x_n]^T$, and

$$A(\alpha) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{\alpha_1 t_2} & e^{\alpha_2 t_2} & \cdots & e^{\alpha_n t_2} \\ \vdots & \vdots & \ddots & \vdots \\ e^{\alpha_1 t_m} & e^{\alpha_2 t_m} & \cdots & e^{\alpha_n t_m} \end{bmatrix}.$$

We assume the α_i's are distinct and consider two possibilities for the x_i's: (i) they are unknown deterministic quantities; (ii) they are normal random variables with unknown power, uncorrelated with each other and e. In (i), $\Theta = \mathbb{R}^{2n}$, and in (ii), $\Theta = \mathbb{R}^n \times \mathbb{R}^{+n}$. We are interested in the maximum n for which every point of Θ is regular.

(i) In this case $\theta = (\alpha_1, \ldots, \alpha_n, x_1, \ldots, x_n)$ and

$$\varphi(\theta) = \left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i e^{\alpha_i t_2}, \ldots, \sum_{i=1}^{n} x_i e^{\alpha_i t_m} \right).$$

It follows from the nonnegativity of the exponential function and a determinantal result given in [6, p. 126] that $A(\alpha)$ is full rank as long as $n \le m$. We may therefore employ standard results on identifiability [5], [9] to conclude that Θ is identifiable if and only if $2n \le m$. As we now show, Theorem 1 can be used to prove that points of Θ are also regular for all such n, avoiding the cumbersome process of computing the rank of $\partial\varphi(\theta)/\partial\theta \in \mathbb{R}^{m\times 2n}$ or of FIM(θ) $\in \mathbb{R}^{2n\times 2n}$.

Considering φ as a function of complex arguments $z_1, \ldots, z_{2n}$, we see that $\varphi(z)$ is holomorphic in each z_i separately (each φ_i is a sum of products of holomorphic functions of the z_i's). Therefore, by Hartogs' theorem, $\varphi(z)$ is holomorphic in z.

For every $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ (α_i distinct), $A(\alpha)$ is full rank. Consequently, each α has an open neighborhood $N_\alpha \subset \mathbb{C}^n$, for which $A(\alpha')$ remains full rank when $\alpha' \in N_\alpha$. By shrinking N_α, if necessary, we may employ the complex-version results of [5], [9] to conclude that there exists an open identifiable $\Omega_\alpha \subset \mathbb{C}^{2n}$ as long as $2n \le m$; Θ can be covered by such Ω_α's.

Because $\varphi_1, \ldots, \varphi_{2n}$ is a representative mapping of each Ω_α when $2n \leq m$ (this too follows immediately from standard identifiability results), regularity for points of Θ follows from Theorem 1. Conversely, Theorem 1 also implies that every point of Θ is singular if $2n > m$.

Remark. When $2n \leq m$, regularity holds even if the covariance of e is unknown, because $\varphi(z)$ can be augmented to include the unknown entries of R, yielding a trivially holomorphic and injective φ. Verifying this by computing $\text{FIM}(\theta) \in \mathbb{R}^{[2n+m(m+1)/2]\times[2n+m(m+1)/2]}$ would indeed be tedious.

(ii) In this case $\theta = (\alpha_1, \ldots, \alpha_n, \sigma_1^2, \ldots, \sigma_n^2)$, where $\sigma_i^2 = Ex_i^2$, and

$$\varphi(\theta) = \left(\sum_{i=1}^{n} \sigma_i^2, \ldots, \sum_{i=1}^{n} \sigma_i^2 e^{\alpha_i(t_j+t_k)}, \ldots, \sum_{i=1}^{n} \sigma_i^2 e^{2\alpha_i t_m}\right), \qquad j, k = 1, \ldots, m$$

excluding terms with nondistinct $t_j + t_k$. It follows from the results in [5], [9] that Θ is identifiable for all possible α_i and σ_i^2 if and only if $n \leq m - 1$. As in (i), $\varphi(z)$ is holomorphic, and there is a collection of open identifiable Ω_α covering Θ. Because there are at least $2m - 1$ distinct $t_j + t_k$, we may choose $2n < 2m - 1$ distinct φ_i as representative mappings of each Ω_α. Hence the elements of Θ are regular if and only if $n \leq m - 1$.

3.1. *Summary*

1. In a normal distribution, to show identifiability of $\Theta \subset \mathbb{R}^p$, establish that the mapping $\varphi : \Theta \to \mathbb{R}^q$ is injective.
2. To find the singular and regular points of the distribution's parametrization:
   a. Establish that $\varphi(z)$ is holomorphic for z in some domain $\Omega \subset \mathbb{C}^p$, where $\Theta \subset \Omega$. This can be done, employing Hartogs' theorem, by verifying that the coordinate functions φ_i, $i = 1, \ldots, q$, are holomorphic in each variable $z_1, \ldots, z_p$ separately.
   b. *To show singularity at points*: Locate points $z \in \Theta$ having no strongly identifiable open neighborhood in $\mathbb{C}^p$.
   c. *To show regularity on a set*: Select a representative mapping on an open subset of $\mathbb{C}^p$; every point of the subset is regular. If Ω can be covered by open sets and corresponding representative mappings, then every point of Ω, and hence Θ, is regular.

4. Conclusion

We have presented a way to unify the notions of identifiability and rgularity using the theory of holomorphic functions of several variables. Example 4 showed that regularity in some models can be more easily determined by examining φ as a holomorphic mapping than by computing the rank of $\partial\varphi(\theta)/\partial\theta$ or $\mathrm{FIM}(\theta)$. This is often true when p is large, for large models are often designed by bringing together the pieces of many smaller, well-parametrized models, whereas matrix ranks are usually cumbersome to compute.

References

[1] P. E. Caines, *Linear Stochastic Systems*, New York: Wiley, 1988.

[2] G. H. Golub andV. Pereyra, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate, *SIAM J. Numer. Anal.*, vol. 10, no. 2, pp. 413–432, April 1973.

[3] R. C. Gunning, *Introduction to Holomorphic Functions of Several Variables, Volume I: Function Theory*, Pacific Grove: Brooks/Cole, 1990.

[4] R. C. Gunning, *Introduction to Holomorphic Functions of Several Variables, Volume II: Local Theory*, Pacific Grove: Brooks/Cole, 1990.

[5] B. Hochwald and A. Nehorai, Identifiability in array processing models with vector-sensor applications, *IEEE Trans. Signal Processing*, vol. 44, no. 1, pp. 83–95, Jan. 1996.

[6] J. Pila, Geometric and arithmetic postulation of the exponential function, *J. Australian Math. Soc.*, vol. 54A, pp. 111–127, 1993.

[7] A. Wald, Note on the consistency of the maximum likelihood estimate, *Ann. Math. Statist.*, vol. 20, pp. 595–601, 1949.

[8] A. Wald, Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Trans. Amer. Math. Soc.*, vol. 54, pp. 426–482, 1943.

[9] M. Wax and I. Zizkind, On unique localization of multiple sources by passive sensor arrays, *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, no. 7, pp. 996–1000, July 1989.