Brazilian Journal of Probability and Statistics (2002), 16, pp. 99–106.

CONSISTENCY AND IDENTIFIABILITY REVISITED

Ernesto San Martín and **Fernando Quintana**

Departamento de Estadística, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile. Emails: esanmart@mat.puc.cl and quintana@mat.puc.cl

Summary
We provide a general framework to review the well-known concept of identifiability and give a formal proof that this is implied by the existence of a consistent estimator. We apply these ideas to the predictive recursion algorithm for finite mixtures to conclude that identifiability is actually equivalent to consistency.

Key words: Convergence of medians; identification of mixtures; predictive recursion.

1 Introduction

The concept of identification is closely related to that of model specification. Following Koopmans and Reiersøl (1950) we can distinguish two components of any statistical model: (i) a *structural model* that formalizes what the contextual theory implies on the process generating the observed and latent variables and (ii) a *measurement model* that connects these variables. This view creates a new problem which logically precedes all inference questions: is the distribution of observables generated by only one structure contained in the set of structures that constitute a model? This is the so-called *identification problem* which becomes a necessary part of the specification problem (Koopmans and Reiersøl, 1950). Note that identifiability is related to knowledge of the probability distribution of observables rather than to a finite sample of observations. See additional discussion in Aldrich (1999), Hurwicz (1950), and Qin (1993). Bayesian approaches to this problem can be found in Kadane (1974), Dawid (1979), and Florens et al. (1990).

In the econometric tradition the link between identifiability and statistical inference is given by the relationship between identifiability and the existence of a consistent estimator. Gabrielsen (1978) suggested a proof of such a relationship, but this proof seems inadequate with respect to

Koopmans and Reiersøl (1950) identification concept. This note proposes an alternative simple proof to Gabrielsen's (1978) claim. The relevance of this proof consists in showing that identifiability is a necessary condition for the convergence in law. The relationship between identifiability and the existence of an asymptotically unbiased estimate is also explored. We apply the results to the predictive recursion algorithm for finite mixtures discussed in Newton (2000) to conclude that identifiability is actually *equivalent* to consistency of the method.

2 Definitions and fundamental concepts

A statistical model is defined as a family of sampling probability distributions indexed by a parameter, that is

$$\mathcal{E} \;=\; \{(X,\mathcal{X}),\; P^\theta \;:\; \theta\in\Theta\}, \tag{2.1}$$

where $(X,\mathcal{X})$ is the *sample space*, P^θ is a *sampling probability* on $(X,\mathcal{X})$ indexed by a parameter θ, and Θ is the *parameter space*; see, *e.g.*, Cox and Hinkley (1974), Raoult (1975), and Barra (1981).

Considering structure (2.1), the identification of any statistical model deals with the *identification of a parametrization.* A parametrization θ is said to be identified if the mapping $\theta \longmapsto P^\theta$ is injective (Koopmans and Reiersøl, 1950), and in this case all *injective* reparametrizations $h(\theta)$ of θ are also identified. When a parametrization θ is unidentified, an identified model can be obtained through reparametrizations (Shao, 1999). As a trivial example, consider the sampling probabilities $\mathcal{N}(\mu_1+\mu_2,\sigma^2)$; if we take $\theta = (\mu_1,\mu_2,\sigma^2) \longmapsto P^\theta \equiv \mathcal{N}(\mu_1+\mu_2,\sigma^2)$, the parametrization θ is unidentified. Nevertheless, the mapping $\lambda = (\mu_1+\mu_2,\sigma^2) \longmapsto P^\lambda \equiv \mathcal{N}(\mu_1+\mu_2,\sigma^2)$ is injective (hence, the parametrization λ is identified), but the function $\theta \longmapsto \lambda$ is non injective.

The consistency criterion for estimation problems is heuristically stated as follows: the statistic applied to the whole population should be equal to the parameter (Fisher, 1922). Formally, this criterion is defined by using convergence in probability; in practice, it is important to make explicit with respect to which probability measure such convergence is taken. In the context of (2.1), a real-valued parameter b is a real function of θ. Thus, a sequence of random variables $\{s_n \,:\, n\in\mathbb{N}\}$ is called *strongly* (resp. *weakly*) consistent for the real-valued parameter b if $s_n \longrightarrow b(\theta)$ almost surely (resp. in probability) with respect to P^θ, for all $\theta\in\Theta$.

3 Interface between identifiability and statistical inference

We establish now links between identifiability, asymptotic unbiasedness and consistency, when the parameter space Θ is a subset of a finite dimensional space.

Paulino and Pereira (1994) establish that identifiability is a necessary condition for the existence of an unbiased estimator. We extend this result to asymptotically unbiased estimators:

Proposition 1 *The identifiability of the parameter θ is a necessary condition for the existence of an asymptotically unbiased estimate.*

Proof: Let $\{s_n : n \in \mathbb{N}\}$ be an asymptotically unbiased estimate of θ, that is, $\lim_{n\to\infty} \mathbb{E}^\theta s_n = \theta$ for all $\theta \in \Theta$, where $\mathbb{E}^\theta(\cdot)$ denotes the sampling expectation with respect to P^θ. Let $\theta_1, \theta_2 \in \Theta$ such that $P^{\theta_1} = P^{\theta_2}$. Because $\theta_1 = \lim_{n\to\infty} \mathbb{E}^{\theta_1} s_n = \lim_{n\to\infty} \mathbb{E}^{\theta_2} s_n = \theta_2$, it follows that the parametrization θ is identified. ■

A similar argument establishes that the existence of an unbiased estimator of $g(\theta)$ implies the identifiability of θ provided that g is an injective function. We note that Proposition 1 is valid regardless of the dimension of Θ.

The result stating that identifiability is a necessary condition for the consistency belongs to the econometrics "oral tradition":

Proposition 2 *The identifiability of the parameter θ is a necessary condition for the existence of a consistent estimate.*

Gabrielsen (1978) gives a proof, which runs as follows:

> Assume θ is not identifiable. In this case we can find at least two different values θ_1 and θ_2 yielding exactly the same distribution of the observations. Let s_n be an estimator. If s_n is consistent, it should in principle converge to these two values. As the convergence to two values is contradictory to the definition of convergence in probability and thus consistency, it follows that there cannot exist consistent estimators for parameters that are not identified.

A similar argument can be found in Rao (1992), page 134. Gabrielsen's (1978) proof is essentially based on the unicity of the limit. However, this *almost sure unicity is with respect to the probability used for establishing the convergence in probability.* In the context of (2.1), the only involved probability is P^θ, a probability measure defined on the sampling space, not on the parameter space. Gabrielsen's statement implicitly considers an expression of the form $P^\theta(|\theta_1 - \theta_2| > \epsilon)$, where $P^{\theta_1} = P^{\theta_2} \equiv P^\theta$, and the event $\{|\theta_1 - \theta_2| > \epsilon\}$ does not belong to the domain of these probability measures (i.e., the sampling space). In the sequel, we give an alternative proof, which only makes use of the unicity of limits in a non-stochastic setting.

Proof of Proposition 2: Let s_n be a sequence of random variables, assumed weakly consistent for θ. Let $\theta_1, \theta_2 \in \Theta$ such that $P^{\theta_1} = P^{\theta_2} \equiv P^\theta$. It follows that $s_n \longrightarrow \theta_i$ in probability with respect to P^{θ_i} for $i = 1, 2$. This implies that $(s_n - \theta_i) \longrightarrow \delta_0$ in law, where δ_0 is a pointmass at 0. Therefore, every accumulation point of a sequence of medians of $s_n - \theta_i$ is a median of δ_0 (see Theorem 2.2.3 in Lukacs, 1968). But δ_0 has a unique median, namely 0 and it follows that any sequence of medians of s_n converge to both θ_1 and θ_2. By the unicity of the limits (in the real line), we conclude that $\theta_1 = \theta_2$. ■

Proposition 2 and its proof deserve the following comments. Firstly, the argument given is valid for $\Theta \subset \mathbb{R}$. The result applies to $\Theta \subset \mathbb{R}^m$ by repeating the proof in each coordinate. Secondly, if there exists a consistent estimate of $g(\theta)$, then the parameter θ is identified provided that g is an injective function. And thirdly, Proposition 2 holds no matter how the consistent estimates are derived.

4 Consistency of the predictive recursion algorithm for finite mixtures

We explore now the link between identifiability and consistency in a finite mixture model. We do this by showing that a nonparametric estimator of the mixing probabilities $(\pi_1, \ldots, \pi_n)$ –the so-called *predictive recursion* algorithm– is consistent provided that the parametrization $(\pi_1, \ldots, \pi_n)$ is identifiable.

Consider a sample $x_1, \ldots, x_n$ from the following finite mixture of Binomial distributions:

$$g(x) = \pi_1 p_T(x|\theta_1) + \cdots + \pi_m p_T(x|\theta_m), \tag{4.1}$$

where $p_T(x|\theta) = \binom{T}{x}\theta^x(1-\theta)^{T-x}$, $\pi = \{\pi_i : 1 \leq i \leq m\}$ is a probability distribution and $\theta_1, \ldots, \theta_m$ are m distinct and known elements of $(0, 1)$. Letting $f^\star(\theta_i) = \pi_i$ for $i = 1, \ldots, m$, and $\Theta = \{\theta_1, \ldots, \theta_m\}$, (4.1) becomes $g_{f^*}(x) = \sum_{i=1}^{m} p_T(x|\theta_i) f^\star(\theta_i)$ and the statistical experiment is given by

$$\mathcal{E} \ = \ \{(\{1, \ldots, T\}, 2^{\{1,\ldots,T\}}), \, g_{f^*}(\cdot) \ : \ f^\star \in \mathcal{P}(\Theta, 2^\Theta)\}, \tag{4.2}$$

where 2^A denotes the power set of A and $\mathcal{P}(\Theta, 2^\Theta)$ is the set of probability distributions on Θ. Condition $m \leq T+1$ ensures identifiability of the mixing distribution $f^\star$ (Lindsay, 1995).

The predictive recursion algorithm (Newton and Zhang, 1999, Newton et al., 1998) applied to this case provides a sequence of probability distributions on Θ adopting the form

$$f_n(\theta_i) = (1-w_n)f_{n-1}(\theta_i) + w_n \times \frac{p_T(x_n|\theta_i)f_{n-1}(\theta_i)}{c(x_n, f_{n-1})}, \qquad (4.3)$$

where $c(x,f) = \sum_{j=1}^{m} p_T(x|\theta_j)f(\theta_j)$. Here, the user defines a decreasing sequence of weights $\{w_n\}$ (a default choice is $w_n = (1+n)^{-1}$) and an initial density $f_0(\cdot)$ on $\mathcal{P}(\Theta, 2^\Theta)$. Note that $g_{f^*}(x) = c(x, f^\star)$. Under the assumption that $\sum_{n=1}^{\infty} w_n$ diverges, Newton (2000) shows that $f_n(\theta_i)$ converges *surely* to $f_\infty(\theta_i)$ for all $i = 1, \ldots, n$, where f_∞ is a probability distribution on Θ. Under some extra technical conditions on the sequence of weights it follows that

$$f_\infty(\theta_i) = \sum_{x=0}^{T} \frac{p_T(x|\theta_i)f_\infty(\theta_i)}{c(x, f_\infty)}\, c(x, f^\star) \qquad (4.4)$$

i.e., f_n converges to a solution of a *self consistency equation*. It is immediately seen that $f_\infty = f^\star$ solves this equation. But identifiability implies that the mapping $f \longmapsto c(\cdot, f)$ is injective, so that (4.4) has $f_\infty = f^\star$ as its only solution. In other words, under the technical conditions discussed above, identifiability implies consistency of $\{f_n\}$. By Proposition 2, we conclude that for the binomial mixture with $m \leq T+1$, identifiability is actually a necessary and sufficient condition for the consistency of $\{f_n\}$.

As a simple illustration, Figure 1 depicts the pointwise convergence of $\{f_n\}$ when the mixing distribution is $f^\star = (2/5, 1/5, 2/5)$ on $\Theta = (1/4, 1/2, 3/4)$, under two circumstances. In both cases we chose f_0 to be uniform over Θ. The left column of plots is obtained using $T = 5$, and horizontal lines represent the values of $f^*(\theta_i)$, which agree with the limits of $\{f_n(\theta_i)\}$. Note how this situation changes in the right column, where we chose $T = 1$. The sequences are still convergent, but their limits disagree with those obtained from $f^\star$. Of course, the reason for that behavior in the latter is the unidentifiability that comes from the fact that $m > T+1$.

5 Concluding remarks

The relationship between identifiability and existence of a consistent estimate can already be found in Reiersøl (1950). It seems that Gabrielsen

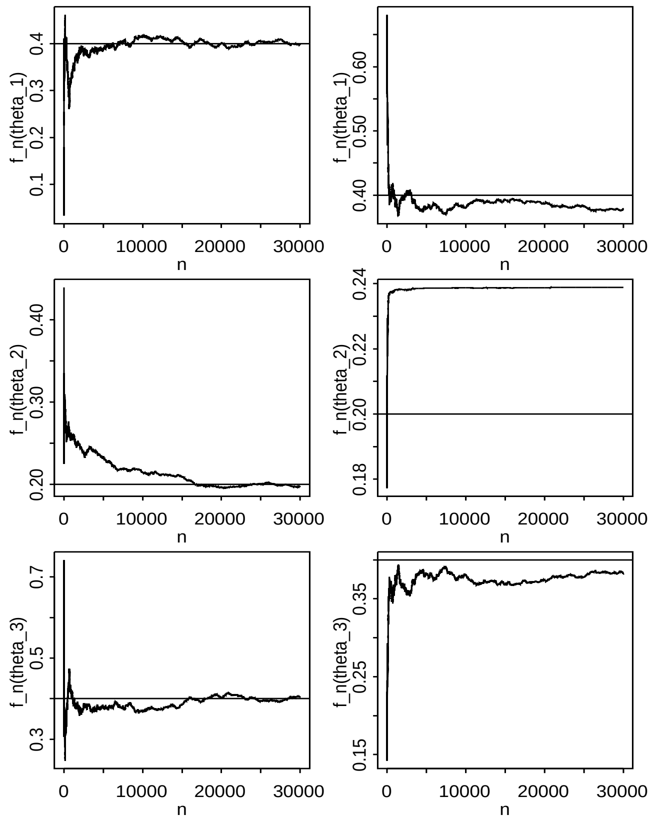

Figure 1 *Convergence of predictive recursion. Figures to the left represent the sequence* $f_n(\theta_i)$ *versus* n *for* $i = 1, 2, 3$ *when* $T = 5$, $m = 3$ *and* $\Theta = \{1/4, 1/2, 3/4\}$. *The right column displays* $f_n(\theta_i)$ *versus* n *for* $i = 1, 2, 3$ *when* $T = 1$, $m = 3$ *and* $\Theta = \{1/4, 1/2, 3/4\}$.

(1978) contains a first attempt to prove it. This paper shows that his approach is inadequate with respect to the identification concept introduced by Koopmans and Reiersøl (1950), suggesting a formal proof. The link identifiability-consistency is relevant in the following terms: if a parametrization θ is unidentified, there does not exist a statistical procedure capable of providing us with a consistent estimate of θ.

The second main result of this paper consists of showing the equivalence between identifiability and consistency in the predictive recursion algorithm for the case of finite mixture of Binomials. The relevant point here is to show that a statistical procedure (as the predictive recursion algorithm) can be interpreted with respect to a statistical experiment only if this experiment is identified. Consequently, if a given model is generalized through a more complex (structural) parametrization, such extension makes statistical sense only if the model is identified.

Acknowledgements

Work of the first author was partially supported by Grant FONDECYT N^o 3010069, whereas the second author was partially supported by Grant FONDECYT N^o 1990430. The authors thank the Editor and two referees for their valuable comments which helped improving the final version of this manuscript.

(Received December, 2001. Revised June, 2002.)

References

Aldrich, J. H. (1999). Determinacy in the linear model: Gauss to Bose and Koopmans. *International Statistical Review*, **67**, 211–219.

Barra, J. R. (1981). *Mathematical Basis of Statistics*. New York: Academic Press.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Dawid, A. P. (1979). Conditional independence in statistical theory (with discussion). *Journal of the Royal Statistical Society* B, **41**, 1–31.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London* A, **222**, 309–368.

Florens, J. P., Mouchart, M. and Rolin, J. M. (1990). *Elements of Bayesian Statistics*. New York: Marcel Dekker.

Gabrielsen, A. (1978). Consistency and identifiability. *Journal of Econometrics*, **8**, 261–263.

Hurwicz, L. (1950). Generalization of the concept of identification, in T. C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models, Cowles Commission Research Monograph 11.* New York: Wiley, 238–257.

Kadane, J. B. (1947). The role of identification in Bayesian theory, in S. E. Fienberg and A. Zellner (eds.), *Studies in Bayesian Econometrics and Statistics.* Amsterdam: North Holland, 175–191.

Koopmans, T. C. and Reiersøl (1950). The identification of structural characteristics. *The Annals of Mathematical Statistics*, **21**, 165–181.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications.* California: Institute of Mathematical Statistics.

Lukacs, E. (1968). *Stochastic Convergence.* Lexington: D. C. Heath and Company.

Newton, M. A. (2000). On a nonparametric recursive estimator of the mixing distribution. *Technical Report*, Department of Statistics, University of Wisconsin-Madison. To appear in Sankhya.

Newton, M. A., Quintana, F. A. and Zhang, Y. (1998). Nonparametric Bayes methods using predictive updating, in D. Dey, P. Müller and D. Sinha (eds.), *Practical Nonparametric and Semiparametric Bayesian Statistics.* New York: Springer, 45–61.

Newton, M. A. and Zand, Y. (1999). A recursive algorithm for nonparametric analysis with missing data. *Biometrika*, **86**, 15–26.

Paulino, C. D. M. and Pereira, C. A. B. (1994). On identifiability of parametric models. *Journal of the Italian Statistical Society*, **3**, 125–151.

Qin, D. (1993). *The Formation of Econometrics. A Historical Perspective.* Oxford: Clarendon Press.

Rao, B. L. S. P. (1992). *Identifiability in Stochastic Models. Characterization of Probability Distributions.* Boston: Academic Press.

Raoult, J. P. (1975). *Structures Statistiques.* Paris: Presses Universitaires de France.

Reiersøl, O. (1950). Idenfiability of a linear relation between variables which are subject to error. *Econometrica*, **18**, 375–389.

Shao, J. (1999). *Mathematical Statistics.* New York: Springer.