David Brenner/Donald A.S. Fraser/Günter Menges/Erika Rost

Model Analysis with Structural and Stochastic Partial Information

1. Introduction

We consider the information processing that arises when a structural model involving a shape parameter is combined with background information in the form of stochastic partial information (SPI). More particularly, we find that the standard factorization of the structural model is preserved provided the stochastic partial information is symmetric with respect to the group parameter. Otherwise, we note that the structural factorization is lost and the stochastic partial information relates in general to the full model.

Suppose that an investigation of a process or system yields a structural statistical model (Fraser, 1979) with a symmetry (group) parameter θ for the presentation and a shape parameter λ for the underlying (objectively-identified) error distribution (Brenner and Fraser, 1980). Typically, interest focuses on the symmetry parameter θ while the shape parameter λ enters as an additional although essential parameter that bears information concerning the primary parameter θ. In addition, suppose that the background information has been organized and developed as stochastic partial information (Kofler, Menges, 1976; Kofler, Menges, Fahrion, Huschens, Kuss, 1980) concerning θ and λ. We examine the case where this SPI is symmetric with respect to the group parameter θ.

2. Basic Information Types

The main ideas underlying structural models and partial information theory are informational and rely on the following definitions and principles:

(i) An initial given consists of three components

(a) a statistical model M
(b) additional background information A
(c) observed data D.

(ii) Statistical inference is the transformation of the given (M,A,D) into organized information I directly describing the unknowns; thus (M,A,D) → I.

(iii) All three components of the given must be fully utilized in the transformation process: all the given and nothing further. Any remaining gap may need to be filled by decision theoretical or other means, with utility or loss functions and with decision criteria, that is, in inference problems, with the additives discussed in a recent article (Brenner, Fraser, Monette 1981).

3. Additional Knowledge

Background knowledge refers to all kinds of knowledge that are relevant to the investigation other than the data themselves. We recognize a natural division of the knowledge:

(a) Model-determining knowledge
(b) Additional background knowledge.

The first is the background knowledge about the variables, the unknowns and the structure of the system as investigated. As formalized, this becomes the information that forms the statistical model.

The second, the additional background knowledge is the information that cannot be incorporated into the model and includes information concerning basic unknowns of the system and information relevant to the inference process itself. It may, for example, refer to quality aspects of the data like "inference supporting", there being a major difference between the results of controlled experiments and results that are non-experimental. Or it may relate to the purpose of the investigation, like the additives in (Brenner, Fraser, Monette 1981). Or it may provide assessments of the possible values of the parameters, as, for example, that certain subregions are more likely than others, or extensions of this leading to formulation as stochastic partial information.

The ideal case of additional background knowledge is that with the highest degree of acuity - the case where the investigation is a controlled experiment concerning a physical system. This ideal is of course an extreme with several opposing extremes. One extreme is that involving the absence of additional background knowledge - where we have data but no model. Some authors recommend exploratory data analysis for this. Another extreme is that involving an abundance of additional background knowledge, as for example in the social sciences and humanities. This can involve frequencies, weights, evaluations, evaluations of evaluations, and so on. It is here that rules of compilation, of aggregation, and of evaluation are needed towards which SPI is a first step.

4. The Concept of Stochastic Partial Information (SPI)

As we want to link the SPI-concept (Kofler, Menges, Fahrion, Huschens, Kuss, 1981) with that of structural analysis in the sense of Fraser, we must define the state space in a special manner.

The state space will be denoted by Φ, where Φ is the parameter space $\{\lambda\}$ of the additional parameter λ alone or the parameter space $\{\theta,\lambda\}$ embracing the group (symmetry) parameter θ and the additional parameter λ:; thus

$$\Phi = \begin{cases} \{\lambda\} & (1) \\ \{\theta,\lambda\} & (2) \end{cases}$$

depending on whether a model-data factorization remains valid or not.

Def. 1 Let $(\Phi,\mathcal{F})$ be the corresponding measurable space. Then, P, the set of all probability measures on $(\Phi,\mathcal{F})$ is the "generalized distribution simplex" (Verallgemeinertes Verteilungssimplex) in the sense of Kofler, Menges et al. (1980, p.160).

Def. 2 A subset $Q \subseteq P$ is a convex subset of the distribution simplex if

$$p_1, p_2 \in Q \wedge \alpha \in (0,1) \Rightarrow \alpha p_1 + (1-\alpha)p_2 \in Q.$$

The definitions 3 and 4 in Kofler, Menges et al. (1980, p. 161f) apply correspondingly:

Def. 3 For $Q \subseteq P$, $\mathrm{Conv}(Q) := \cap\{C \subseteq P\colon Q \subseteq C,\ C \text{ convex}\}$ is the convex closure of Q.

Def. 4 The convex closure of a finite subset of P is called a <u>convex polyhedron</u>; the union of a finite number of convex polyhedra is called a <u>polyhedron</u>.

Note that every polyhedron is a subset of the (generalized) distribution simplex P.

As we noted in section 2, a statistical model M expresses a certain background knowledge concerning the relations between the states $\mathcal{S} \in \Phi$ and possible response observations. According to the principle of etiality (Hartwig, 1956) a specific model "information" can only refer to a distribution of the $\mathcal{S}$'s. Let any such distribution be represented by an element $p \in P$; then an "information" called <u>stochastic information</u> (SI) is given by a subset $S \subseteq P$, S being the solution set of the boundary conditions generating the "information"*).

In order to be able to operate on P (as well as for some other reasons that are made evident in the original paper) we have to introduce a measurable space $(P,\mathcal{P})$ of the distributions $p \in P$:

Def. 5 The σ-algebra generated by the set of convex polyhedra in $\mathcal{P}$,

$$\mathcal{P} := \sigma\{Q \subseteq P : Q \text{ is convex polyhedron}\},$$

---

*) It should be noted that a polyhedron can be considered as the solution set based on logical connections of a finite number of linear inequalities. This is one reason why the concept of polyhedron is important for SPI.

is called the σ-algebra of the stochastic information on $(\Phi, \mathcal{F})$. An element S of $\mathcal{S}$ is called stochastic information (SI).

In general we note that the σ-algebra generated by a class M of subsets of P, is given by

$$\sigma(M) := \cap\{\mathfrak{M} : M \subseteq \mathfrak{M} \text{ and } \mathfrak{M} \text{ is a } \sigma\text{-algebra on } P\}.$$

Let us now consider different types of stochastic information.

Def. 6 A stochastic information $S \subset P$ is called linear: $\longleftrightarrow$ S is a polyhedron.

The linear case is the only important one from the practical point of view, as in practice, systems of linear inequalities or logical combinations of these are used to express restrictions describing p in S.

Def. 7 A stochastic information $S \subset P$ is called complete: $\longleftrightarrow$ the number of elements of S is 1.

Def. 8 A stochastic information is called partial and denoted by SPI: $\longleftrightarrow$ $S \subset P$ and S contains at least two elements.

In the remainder of this section we restrict our attention to the case of a finite parameter-space which has particular significance for applications. For this the versatility and generality of the concept is exemplified by means of three features: simplicity, multi-stage modelling and partitioning.

Def. 9 We call a stochastic information $L \subseteq P$ a linear partial information (LPI) on the distribution p : $\longleftrightarrow$ L can be expressed by a finite number of inequalities. By $\mathcal{L}(\Phi)$ we denote the set of all LPI,

$$\mathcal{L}(\Phi) := \{L \subseteq P : L \text{ is LPI}\}.$$

$\mathcal{L}(.)$ may be also called the $\mathcal{L}$-operator.

Def. 10 An LPI is called simple, an SLPI: $\longleftrightarrow$ it can be expressed in the form

$$\text{SLPI} := \{p \in P : Ap \geqslant b,\ A \in \mathbb{R}^{m \times n},\ b \in \mathbb{R}^{m}\}.$$

SLPI's are of great practical use for a number of reasons:

(a) An SLPI is compact. This is important if the shape parameter problem is formulated as a problem of optimization. Under very general conditions, the SLPI-optimization problem yields a minimum or maximum.

(b) SLPI is a restricted convex polyhedron; this is of importance because the SLPI can then be completely characterized by its vertices.

(c) SLPI provides the possibility to formulate, formalize, and at the same time utilize partial knowledge of the distribution of the additional shape parameter.

(d) SLPI allows the softening of classical stochastic concepts without suspending the notion of probability.

In order to make the SPI- or SLPI-concept still more versatile for practical application, we allow for stochastic information of stochastic information as now outlined.

Def. 11 For a given Φ we call

$$U^{(1)} := \mathcal{L}(\Phi)$$

the first level of uncertainty.

From $U^{(1)}$ we select a finite subset

$$\Phi^{(1)} = \{\mathcal{J}_1^{(1)}, \mathcal{J}_2^{(1)}, \ldots, \mathcal{J}_m^{(1)}\},$$

apply to it the $\mathcal{L}$-operator, and obtain the set $\mathcal{L}(\Phi^{(1)})$ of all LPI's for distributions on $\Phi^{(1)}$. The union of all $\mathcal{L}(\Phi^{(1)})$ for all finite subsets of $U^{(1)}$ forms then the second level of uncertainty

$$U^{(2)} = \cup\{\mathcal{L}(\Phi^{(1)}) : \Phi^{(1)} \subset U^{(1)}, \Phi^{(1)} \text{ finite}\}.$$

In general, for k=2,3,... we define the k'th level of uncertainty.

Def. 12 If for a given Φ, the (k-1)th level of uncertainty is $U^{(k-1)}$, then

$$U^{(k)} := \cup\{\mathcal{L}(\Phi^{(k-1)}) : \Phi^{(k-1)} \subset U^{(k-1)}, \Phi^{(k-1)} \text{ finite}\}$$

is called k'th level of uncertainty.

By means of the concepts defined in the definitions 11 and 12, we obtain any structuring of indeterminateness regarding Φ.

Several important means towards this involve the partitioning, evaluation, and reduction of indeterminateness regarding Φ.

A general reduction theorem establishes that any general multi-stage structured LPI can be reduced and expressed equivalently as a single LPI, a subset of the original first level LPI.

5. Structural Model

We consider a process or system with response Y having possible distributions that exhibit the symmetries underlying the structural model. For details on the structural model see Fraser (1979); the objective grounds for the structural model in terms of objective distribution form have been discussed in Brenner and Fraser (1980) and in terms of symmetries in the ordinary model in Brenner and Fraser (1981).

Let $\mathcal{Y}$ be the response space for the response variable Y. In accord with Brenner and Fraser (1980) we assume that distribution form is objective. Accordingly, let Z be a variable for the error distribution representing the objective distribution form, and let θ, in an exact transformation group G on $\mathcal{Y}$, give the presentation of $Y = \theta Z$ in terms of the error variable Z. Also let λ in Λ be the parameter for the distribution P_λ for the error variable Z. As part of the validity of the model we assume that θ and λ are nontrivial (identifiable) and that no further symmetries are available for description in terms of distribution form (Brenner and Fraser, 1981). We summarize the model as

$$\begin{aligned} Y &= \theta Z \qquad && \theta \varepsilon G \\ Z &\sim P_\lambda && \lambda \varepsilon \Lambda . \end{aligned} \tag{1}$$

We now consider the necessary reductions that occur with the structural model together with a value Y_o for the response variable Y. Let Z_o denote the corresponding realized value for the error variable Z.

The observable portion of the realized value Z_o is the orbit $GZ_o = GY_o$. Let $P_{\lambda,GZ}$ be the marginal distribution for the variable GZ. We thus have the model $P_{\lambda,GZ}$, λ in Λ, together with the observed value GY_o for the variable GZ.

The unobservable portion of Z_o is its location on the observed orbit $GZ_o = GY_o$ and this is described by the conditional probability of Z given $GZ = GY_o$. For notation, let $[Z]$ in G be such that $Z = [Z]D(Z)$; this implicitly defines a reference or cross-section point $D(Z)$ on the orbit through Z. We then have $D(Y) = D(Z)$ expressing $GY = GZ$; and we have $[Y] = \theta[Z]$ giving the presentation of the observed $[Y_o]$ in terms of the unobserved $[Z_o]$ based on group coordinates on the observed orbit. Let $P_\lambda^{GY_o}$ be the conditional distribution for $[Z]$ given the orbit $GZ = GY_o$. We thus obtain the model $[Y] = \theta[Z]$, $[Z] \sim P_\lambda^{GY_o}$, θ in G, λ in Λ together with the observed value $[Y_o]$ for the response.

With the preceding notation we have that the initial model-data combination

$$\begin{array}{llll} \mathcal{M}: & Y = \theta Z & \theta\varepsilon G & \\ & Z \sim P_{\lambda} & \lambda\varepsilon\Lambda & \quad (2) \\ \mathcal{D}: & Y_{o} & & \end{array}$$

factors into a first model-data combination of classical form

$$\begin{array}{llll} \mathcal{M}_1: & GZ \sim P_{\lambda,GZ} & \lambda\varepsilon\Lambda & \quad (3) \\ \mathcal{D}_1: & GY_{o} & & \end{array}$$

and a second model-data combination of structural form

$$\begin{array}{llll} \mathcal{M}_2: & [Y] = \theta[Z] & \theta\varepsilon G & \\ & \quad GY_{o} & & \\ & [Z] \sim P_{\lambda} & \lambda\varepsilon\Lambda & \quad (4) \\ \mathcal{D}_2: & [Y_{o}] & & \end{array}$$

Details concerning this factorization of the structural model with data have been discussed in Fraser (1979, Chapter 6) and concerning the necessary reduction implicit in the factorization in Fraser (1979, Chapter 3).

6. Structural Model, Symmetry, and SPI

We consider the analysis of the structural model (1) in the presence of symmetric SPI concerning the parameter (θ,λ) in $G\times\Lambda$. In this we assume that the SPI is in the reduced form of a set S of probability measures p on the product space $G\times\Lambda$ and suppose that each measure p can be presented in terms of a marginal probability or probability density $p_*(\lambda)$ for λ and a conditional probability or probability density function $p^{\lambda}(\theta)$ for θ given λ, that is $p(\theta,\lambda) = p^{\lambda}(\theta)\, p_*(\lambda)$; the support measure for θ is taken to be the right invariant measure (Fraser, 1972a, Section 7), which represents invariance concerning origin in the use of the group G as parameter space.

We define two levels of symmetry for SPI concerning (θ,λ).

Def. 13 The set S is weakly symmetric for θ in G if

$$S = \{S^{\lambda}(\theta)p_*(\lambda) : p_*\varepsilon S_*\}$$

and $S^{\lambda}(\theta)$ is right symmetric with respect to G where

$$S^{\lambda}(\theta) = \{p^{\lambda}(\theta)\} , \; S_* = \{p_*\}.$$

Def. 14 The set S is strongly symmetric for θ in G if

$$S = S^* \times S_*$$

where $S^* = S^\lambda$ is right symmetric with respect to G and independent of λ in Λ.

The factorization of a structural model is based on the availability of valid conditional distributions given observed data. This question has been examined in Fraser (1972b) and the validity of conditional probability was derived from no differential information given the condition other than that provided by the probabilities of the model. In the present context we then obtain validity for given λ on the basis of weak symmetry and obtain general validity on the basis of strong symmetry.

We summarize in the following theorem.

Theorem The structural model (1) with strongly symmetric SPI on $G \times \Lambda$ given by $S = S^* \times S_*$ factors into the marginal model as in (3) with S_* and the conditional model as in (4) with S^*.

References

Brenner, D. and D.A.S. Fraser: The identification of distribution form. Statistische Hefte, 21 (1980), pp. 296-304.

Brenner, D. and D.A.S. Fraser: A simplification of the traditional statistical model in the presence of symmetry. Statistische Hefte, 22 (1981), pp. 195-206.

Brenner, D., D.A.S. Fraser, G. Monette: Theories of inference or simple additives. Statistische Hefte, 22 (1981), pp. 232-234.

Fraser, D.A.S.: Bayes, likelihood, or structural. Annals Mathematical Statistics, 43 (1972a), pp. 777-790.

Fraser, D.A.S.: Events, information processing and the structured model. In: Foundations of Statistical Inference. Ed.: V.P. Godambe and D.A. Sprott. Holt, Rinehart and Winston. Toronto 1972b.

Fraser, D.A.S.: Inference and Linear Models. McGraw Hill. New York 1979.

Kofler, E. and G. Menges: Entscheidungen bei unvollständiger Information. Springer. Berlin, Heidelberg, New York 1976.

Kofler, E., G. Menges, R. Fahrion, S. Huschens and U. Kuss: Stochastische Partielle Information (SPI). Statistische Hefte, 21 (1980), pp. 160-167.

Kuss, U.: C-optimale Entscheidungen. Statistische Hefte, 21 (1980), pp. 261-279.

Menges, G.: Adaptive Statistik. Statistische Hefte, 21 (1980), pp. 182-208.

Hartmut Sangmeister

Aktuelle Probleme der internationalen Wirtschafts- und Sozialstatistik

- Ein Überblick über neuere Literatur -

Auf der 51. Jahreshauptversammlung der Deutschen Statistischen Gesellschaft 1980 in Hamburg wurde die Feststellung getroffen, daß die internationale Statistik ein höchst unbefriedigendes Bild böte; konstatiert wurden ein generelles Theoriedefizit, ein weitgehendes methodologisches Defizit und die fast völlige Ignoranz der Fehler, die den internationalen Statistiken anhaften (Menges 1981). Diesem entmutigenden Befund wird jeder, der sich mit internationaler Statistik inhaltlich beschäftigt, nur zustimmen können.

Angesichts dieser desolaten Situation greift man voller Erwartung auf Besserung zu der fast 400 Seiten umfassenden Neuerscheinung:

R. Zwer: Internationale Wirtschafts- und Sozialstatistik. Lehrbuch über die Methoden und Probleme ihrer wichtigsten Teilgebiete. (Internationale Standardlehrbücher der Wirtschafts- und Sozialwissenschaften) München-Wien 1981.

Wer sich von dem Untertitel des Buches eine Methodenlehre (oder gar eine Theorie) der internationalen Statistik erhofft, wird möglicherweise enttäuscht sein; denn was der Autor liefert, ist im wesentlichen eine detaillierte und kritische Darstellung wichtiger Bereiche aus den Arbeitsgebieten der Statistischen Ämter der Vereinten Nationen und der Europäischen Gemeinschaften.

Etwa die Hälfte des Buches nehmen die beiden Kapitel über internationale Volkswirtschaftliche Gesamtrechnungen und Input-Output-Analysen ein. Dabei werden die Systeme der Vereinten Nationen und der Europäischen Gemeinschaften (sowie nationale Besonderheiten, insbesondere in der Bundesrepublik Deutschland) in einer Ausführlichkeit und Gründlichkeit dargestellt, wie man es wohl kaum an anderer Stelle findet, es sei denn, man ziehe die originalen Richtlinien, Empfehlungen und Erläuterungen der verschiedenen Institutionen zu Rate.

Neben einem Überblick über die historische Entstehung der verschiedenen Systeme inter- und supranationaler Volkswirtschaftlicher Gesamtrechnungen sowie Einzelheiten über deren rechtliche Grundlagen und organisatorische Durchführung erhält der Leser umfassende Informationen über die inhaltliche Ausgestaltung der Systeme. Besondere Berücksichtigung findet dabei die Frage der Vergleichbarkeit von statistischen Einheiten, Merkmalen und Aggregaten. Denn die Vergleichbar-