Bayesian Identification of Semi-Parametric Binary Response Models

Michel MOUCHART[1], Jean-Marie ROLIN[1] and Eliana SCHEIHING[2]

July 23, 1997

Abstract

In this paper, minimal conditions under which a semi-parametric binary response model is identified in a Bayesian framework are presented and compared to the conditions usually required in a sampling theory framework.

Running headline: Semi-parametric Binary Response Models.
Keywords: Binary response models, Non parametric Bayesian Statistics, Dirichlet processes, Identification
AMS MOS primary: 62F15
secondary: 62J02, 62G99
Acknowledgments: Research supports from the contract "Projet d'Actions de Recherche Concertées" No. 93/98-164, from the Belgian Program on Interuniversity Poles of Attraction (No. 26), from the Belgian OSTC Transport and Mobility (contract TRB3/007), from the Departmental fellowship of UCL and from DID-UACh grant S-96-30 are gratefully acknowledged. Visits of the first two authors to the Department of Mathematical Engineering of the University of Chile (Santiago) and to the Institute of Computer Science at the University Austral of Chile (Valdivia) were instrumental in provoking discussions with members of those departments.

[1]CORE and Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
[2]Instituto de Informática, Universidad Austral de Chile

1 Introduction

Let us consider a sample of n random vectors $X_i = (Y_i, Z_i)$, $1 \leq i \leq n$, where Y_i is a binary response variable ($Y_i \in \{0, 1\}$) and $Z_i \in \mathbb{R}^p$ is a vector of p explanatory variables. The binary response model specifies only the conditional model generating $(Y_i|Z_i)$, leaving the model generating Z_i virtually unspecified. Because of the binary nature of Y_i, such models are also known as conditional Bernoulli models. The probit and logit models are the most popular purely parametric versions of this model. They are based on a parametrized family of distribution functions, namely normal or logistic, called a link function, and valuated at a parametrized function of the explanatory variables, called an index function.

In this paper we consider the problem of identification of such models under a semi-parametric specification: functional for the link function and euclidean for the index function. In section 2 we review the sampling theory of this issue for the purely parametric, the non-parametric and the semi-parametric case, respectively. In the purely parametric and the semi-parametric cases, we limit our review for the sake of exposition to the index functions that are linear in a euclidean parameter. In section 3, we present a Bayesian analysis of the semi-parametric specification; the main result is that under a quite usual prior specification and a very general condition on the trajectories of the explanatory variables, the model is identified in such a Bayesian set-up whereas identification in sampling theory required further specific restrictions. Section 4 gives some concluding remarks on this seemingly disturbing result. We gather in appendices some mathematical tools necessary for a precise statement of the results.

2 Conditional Bernoulli Models: A Review of Sampling Theory

The data is a sample of n random variables, (Y_i, Z_i); $1 \leq i \leq n$, where Z_i is a random variable defined on a probability space $(\Omega, \mathcal{M})$ with values in $(\mathbb{R}^p, \mathcal{B}^p)$ ($\mathcal{B}^p$ denotes de Borel subsets of $\mathbb{R}^p$) and Y_i is a binary random variable, i.e. a random variable defined on $(\Omega, \mathcal{M})$ with values in $\{0, 1\}$. As usually done in conditional models, we assume that the sequence of Y_i's is sifted by the sequence of Z_i's (see Florens et al. (1990, section 7.6); more explicitely: for any n, conditionally on $Z_1^n = \{Z_i : 1 \leq i \leq n\}$ the random variables Y_i, $1 \leq i \leq n$, are mutually independent and the distribution of

Y_i only depends on Z_i, i.e.

$$
\begin{array}{rcl}
\underset{1\leq i\leq n}{\perp\!\!\!\perp} Y_i & \mid & Z_1^n, \theta \\
 & & \\
Y_i \perp\!\!\!\perp Z_1^n & \mid & Z_i, \theta
\end{array} \tag{1}
$$

where θ is a parameter of the distribution to be made precise later on. *Remark:* As far as suitable, we stick to the notational device of using latin letters for observable random variables, greek letters for unobservable random variables and script letters for σ-fields.

2.1 General Parametric Model

A general parametric model is characterized by the following assumptions:

$$
P[Y_i = 1 \mid Z_i, \theta] \quad = \quad E(Y_i \mid Z_i, \theta) = F_\varphi[g(\gamma, Z_i)]
$$

where F_φ is a distribution function on $I\!\!R$, called a link function, known up to a Euclidean parameter φ and g is a known function, called an index function, parametrized by γ. Thus $\theta = (\varphi, \gamma) \in \Phi \times \Gamma$ is a Euclidean parameter, where Φ and Γ are suitable subsets of $I\!\!R^l$ and $I\!\!R^k$ respectively.

In this general parametric case, the model for fixed n is identified if the following implication holds:

$$
\begin{array}{rrcll}
 & F_{\varphi_1}[g(\gamma_1, Z_i)] & = & F_{\varphi_2}[g(\gamma_2, Z_i)] & \forall\ 1 \leq i < n \\
\Rightarrow & (\varphi_1, \gamma_1) & = & (\varphi_2, \gamma_2) &
\end{array} \tag{2}
$$

When condition (2) is not satisfied, two parameters values $\theta_1 = (\varphi_1, \gamma_1)$ and $\theta_2 = (\varphi_2, \gamma_2)$ are said observationally equivalent once

$$
F_{\varphi_1}[g(\gamma_1, Z_i)] = F_{\varphi_2}[g(\gamma_2, Z_i)] \quad \forall\ 1 \leq i < n
$$

Clearly, condition (2) crucially depends on the form of the F_φ and g functions. Let us now look at parametric models with linear index functions.

2.2 Linear Parametric Models. Probit and Logit Models

Let us consider a linear index function:

$$
g(\gamma, Z_i) \quad = \quad \alpha + \beta' Z_i
$$

i.e. $\gamma = (\alpha, \beta) \in I\!R^{p+1}$ $(k = p + 1)$

In the logit model,

$$F_\varphi(x) \quad = \quad L\left(\frac{x-\mu}{\sigma}\right)$$

where

$$L(y) = \frac{e^y}{1+e^y}$$

and in the probit model,

$$F_\varphi(x) \quad = \quad N\left(\frac{x-\mu}{\sigma}\right)$$

where N is the standard normal distribution function.

In both cases, $\varphi = (\mu, \sigma) \in \Phi = I\!R \times I\!R_0^+$ and since L and N are invertible, the condition of identification becomes:

$$\frac{\alpha_1 - \mu_1}{\sigma_1} + \left(\frac{1}{\sigma_1}\beta_1\right)' Z_i = \frac{\alpha_2 - \mu_2}{\sigma_2} + \left(\frac{1}{\sigma_2}\beta_2\right)' Z_i \qquad \forall\, 1 \le i < n$$

implies $\quad (\alpha_1, \beta_1, \mu_1, \sigma_1) = (\alpha_2, \beta_2, \mu_2, \sigma_2)$

Clearly, if

$$\frac{\alpha_1 - \mu_1}{\sigma_1} = \frac{\alpha_2 - \mu_2}{\sigma_2} \tag{3}$$

$$\text{and} \qquad \frac{1}{\sigma_1}\beta_1 = \frac{1}{\sigma_2}\beta_2, \tag{4}$$

θ_1 and θ_2 are observationally equivalent and if there exist $i_1, i_2, \cdots, i_{p+1}$ such that $\{\tilde{Z}_{i_1}, \tilde{Z}_{i_2}, \cdots, \tilde{Z}_{i_{p+1}}\}$ are linearly independent vectors in $I\!R^{p+1}$ (where $\tilde{Z}'_{i_j} = (1, Z'_{i_j})$), conditions (3) and (4) are necessary for observational equivalence. The identified parameters are therefore

$$\alpha_* = \frac{\alpha - \mu}{\sigma} \qquad \beta_* = \frac{1}{\sigma}\beta$$

and the usual identification restrictions are

$$\mu = 0 \qquad \text{and} \qquad \sigma = 1$$

2.3 Non Parametric Model

In the non parametric model we set

$$P[Y_i = 1 \mid Z_i, \theta] \quad = \quad E[Y_i \mid Z_i, \theta] = \varphi[\gamma(Z_i)]$$

where φ is an unknown distribution function on $I\!R$ and γ is an unknown (measurable) function defined on $I\!R^p$ with values in $I\!R$; in this case $\theta = (\varphi, \gamma)$ is a purely functional parameter.
The identification problem in the purely non parametric case may be viewed as follows. Let $h : I\!R \rightarrow I\!R$ be a strictly increasing continuous function. Note that $\theta_2 = (\varphi_2, \gamma_2)$ where

$$\varphi_2 = \varphi_1 \circ h \qquad \gamma_2 = h^{-1} \circ \gamma_1 \tag{5}$$

produces the same model as $\theta_1 = (\varphi_1, \gamma_1)$. Conversely,

$$\varphi_2(\gamma_2(Z_i)) = \varphi_1(\gamma_1(Z_i)) \qquad \forall\, i \in I\!N \tag{6}$$

implies

$$\varphi_2(\gamma_2(z)) = \varphi_1(\gamma_1(z)) \qquad \forall\, z \in I\!R^p \tag{7}$$

as soon as $\gamma_1, \gamma_2, \varphi_1$ and φ_2 are continuous and $\{Z_i : i \in I\!N\}$ is a dense subset of $I\!R^p$ and, once φ_2 is strictly increasing, we obtain (5) with $h^{-1} = \varphi_2^{-1} \circ \varphi_1$.

Remark: as we shall see later on, a usual prior specification on the space of the distribution functions, the Dirichlet process, generates a.s. trajectories within the space of discrete distributions. It should be mentionned that if $h : I\!R \rightarrow I\!R$ is an increasing right-continuous function, its inverse may be defined as $\overline{h}(u) = inf\{t : h(t) > u\}$; moreover $\overline{h}$ is strictly increasing if and only if h is continuous, in which case $h[\overline{h}(u)] = u$. Thus (5) to (7) are still valid with h^{-1} replaced by $\overline{h}$ once φ_2 is strictly increasing.

It seems difficult in this context to relax the assumptions of continuity of γ and of continuity and strict increase of φ. It seems also difficult to define reasonable identification restrictions for such a general model. Without entering into mathematical technicalities, the difficulty lies in constructing a manageable section of the observational equivalence when defined on spaces so big as the product space of increasing right-continuous functions times the space of continuous functions say. A practical solution, provided by semiparametric models consists in requiring that one of the two factor spaces be finite dimensional. However, the assumption that $\{Z_i : i \in I\!N\}$ is a dense subset of $I\!R^p$ is a natural extension to the non parametric case of the rank condition required in the linear parametric case. Indeed in a non parametric set-up, identification makes sense only in the asymptotic model.

2.4 Semi-Parametric Model

Here, the model is given by

$$P[Y_i = 1 \mid Z_i, \theta] \;=\; E[Y_i \mid Z_i, \theta] = \varphi[g(\gamma, Z_i)]$$

where φ is an unknown distribution function on $I\!R$ and g is a known (measurable) function defined on $\Gamma \times I\!R^p$ with values in $I\!R$. Therefore $\theta = (\varphi, \gamma)$ is a parameter "half functional- half Euclidean". Here also the identification problem relies heavily on the form of the function g and we will look at the semi-parametric linear (in the Euclidean parameter) case, i.e., the case where the index function is linear in γ.

2.5 Semi-Parametric Linear Model

The model becomes

$$E[Y_i \mid Z_i, \theta] \;=\; \varphi(\alpha + \beta' Z_i) \tag{8}$$

Now $\gamma = (\alpha, \beta) \in \Gamma \subset I\!R^{p+1}$ and $\theta = (\alpha, \beta, \varphi) \in \Gamma \times \Phi$, where Φ is the space of all distribution functions on $I\!R$.

The analysis of identification, in a sampling theory framework, runs as follows. On one hand, if there exists $\mu \in I\!R$ and $\sigma \in I\!R_0^+$, such that

$$\begin{aligned} \beta_2 &= \frac{1}{\sigma}\beta_1 \\ \alpha_2 &= \frac{\alpha_1 - \mu}{\sigma} \\ \varphi_2(x) &= \varphi_1(\mu + \sigma x) \quad \forall\, x \in I\!R \end{aligned} \tag{9}$$

then clearly

$$\varphi_2(\alpha_2 + \beta_2' Z_i) = \varphi_1(\alpha_1 + \beta_1' Z_i) \qquad \forall\, i \in I\!N \tag{10}$$

Thus $\theta_1 = (\alpha_1, \beta_1, \varphi_1)$ and $\theta_2 = (\alpha_2, \beta_2, \varphi_2)$ are observationally equivalent in such a case.

On the other hand suppose that (10) holds and that $\{Z_i : i \in I\!N\}$ is a dense subset of $I\!R^p$. We shall show that provided one of the β_i's is non zero, there exists $(\mu, \sigma) \in I\!R \times I\!R_0^+$ such that (9) holds and therefore that $\theta_1 = \theta_2$ once μ and σ are restricted to a particular value.

Define in $I\!R^p$ the following open neighborhoods of xe_j where e_j is the j^{th} column of a unit matrix (of dimension p) and $x \in I\!R$:

$$I_\epsilon(j,x) \quad = \quad \{z \in I\!R^p : x - \epsilon < z_j < x + \epsilon, \;\; -\epsilon < z_k < \epsilon, \;\; \forall k \neq j\}$$

Since φ_1 is increasing and right-continuous, we have

$$\sup_{\{i: Z_i \in I_\epsilon(j,x)\}} \varphi_1(\alpha_1 + \beta_1' Z_i) = \varphi_1 \left[\left(\alpha_1 + \beta_{1j} x + \epsilon \sum_{1 \le k \le p} |\beta_{1k}| \right) - \right]$$

where

$$\varphi(y-) = \sup_{z<y} \varphi(z)$$

Hence,

$$\varphi_1(\alpha_1 + \beta_{1j} x) \quad = \quad \inf_{\epsilon > 0} \sup_{\{i: Z_i \in I_\epsilon(j,x)\}} \varphi_1(\alpha_1 + \beta_1' Z_i) \tag{11}$$

Therefore, given that

$$\sum_{1 \le k \le p} |\beta_{2k}| > 0 \qquad \text{and} \qquad \sum_{1 \le k \le p} |\beta_{1k}| > 0$$

we obtain:

$$\varphi_2(\alpha_2 + \beta_{2j} x) = \varphi_1(\alpha_1 + \beta_{1j} x) \qquad \forall\, x \in I\!R \qquad 1 \le j \le p \tag{12}$$

Since there exists j_0 such that $\beta_{1j_0} > 0$ (resp. < 0), it is easily seen that (12) entails that $\beta_{2j_0} > 0$ (resp. < 0). Hence,

$$\varphi_2(y) = \varphi_1(\alpha_1 - \frac{\beta_{1j_0}}{\beta_{2j_0}} \alpha_2 + \frac{\beta_{1j_0}}{\beta_{2j_0}} y) \tag{13}$$

Therefore defining

$$\mu = \alpha_1 - \frac{\beta_{1j_0}}{\beta_{2j_0}} \alpha_2 \qquad \text{and} \qquad \sigma = \frac{\beta_{1j_0}}{\beta_{2j_0}} \in (0, \infty) \tag{14}$$

(13) may be written as:

$$\varphi_2(y) = \varphi_1(\mu + \sigma y) \qquad \forall\, y \in I\!R \tag{15}$$

furthermore (14) implies:

$$\alpha_2 = \frac{\alpha_1 - \mu}{\sigma}; \tag{16}$$

starting from (12), we successively get

$$\begin{aligned}
\varphi_1(\alpha_1 + \beta_{1j}x) &= \varphi_2(\alpha_2 + \beta_{2j}x) && \text{by (12)} \\
&= \varphi_1(\mu + \sigma(\alpha_2 + \beta_{2j}x)) && \text{by (15)} \\
&= \varphi_1(\alpha_1 + \sigma\beta_{2j}x) \qquad \forall\, x \in I\!R && \text{by (16)}
\end{aligned}$$

implying that

$$\beta_{2j} = \frac{1}{\sigma}\beta_{1j} \qquad 1 \le j \le p$$

as long as φ_1 is not a point mass (at α_1).
Identifying restrictions are typically obtained as follows. If we restrict φ to be a distribution function with expectation zero and variance 1, i.e.

$$\int_{-\infty}^{\infty} x d\varphi(x) = 0, \qquad \int_{-\infty}^{\infty} x^2 d\varphi(x) = 1$$

this implies that $\mu = 0$ and $\sigma = 1$, i.e.

$$\begin{aligned}
\varphi_2(y) &= \varphi_1(y) \qquad \forall\, y \in I\!R \\
\beta_{2j} &= \beta_{1j} \qquad 1 \le j \le p \\
\alpha_2 &= \alpha_1
\end{aligned}$$

Putting all these remarks together entails that the semi-parametric linear model is identified in a sampling theory framework if the following three conditions hold:

(i) $\{Z_i : i \in I\!N\}$ is a.s. a dense subset of $I\!R^p$

(ii) $\exists\ j_0$ such that $\beta_{j_0} \neq 0$

(iii) φ is a distribution function on $I\!R$ such that

$$\int_{-\infty}^{\infty} x d\varphi(x) = 0, \qquad \int_{-\infty}^{\infty} x^2 d\varphi(x) = 1$$

Let us remark that (iii) may be replaced by two conditions on fractiles

$$\varphi(x_{p_1}) = p_1 \qquad \text{and} \qquad \varphi(x_{p_2}) = p_2$$

where $p_1 < p_2 \in (0,1)$. In this case, no assumptions about the existence of the first two moments is needed.
Note that assumption (iii) prevents φ from being a point mass.

3 Bayesian Identification of the Semi-Parametric Linear Model

In the semi-parametric linear model given by (1) and (8), the parameter is $\theta = (\varphi, \alpha, \beta)$. The Bayesian analysis involves specifying a prior distribution on the parameter space $(\Theta, \mathcal{A})$. We will consider the following (proper) prior specification:

$$\begin{aligned} &(i) && \varphi \perp\!\!\!\perp \alpha \perp\!\!\!\perp \beta \perp\!\!\!\perp Z_1^\infty \\ &(ii) && (\alpha, \beta) \sim m \\ &(iii) && \varphi \sim \mathcal{D}i(n_0 F_0) \end{aligned} \tag{17}$$

where m is an arbitrary (provided only that $\alpha \perp\!\!\!\perp \beta$) prior probability measure on $(\mathbb{R}^{p+1}, \mathcal{B}^{p+1})$ and φ is a random distribution function on $(\mathbb{R}, \mathcal{B})$ distributed as a Dirichlet process of parameter F_0, a non degenerate distribution function on $(\mathbb{R}, \mathcal{B})$, and $n_0 \in \mathbb{R}_0^+$. Recall that the Dirichlet process is the natural conjugate prior distribution of the empirical distribution function (see Ferguson (1973)).
As seen in the appendix, the identification of the conditional independent Bayesian experiment, as specified in condition (1), rests on verifying:

$$\mathcal{A} \subset \overline{\mathcal{Z}_1^\infty \vee \bigvee_{i \in \mathbb{N}} (\mathcal{A} \vee \mathcal{Z}_i)\mathcal{Y}_i}$$

where

$$\mathcal{A} = \sigma(\varphi, \alpha, \beta) \qquad \mathcal{Z}_1^\infty = \bigvee_{i \in \mathbb{N}} \mathcal{Z}_i = \bigvee_{i \in \mathbb{N}} \sigma(Z_i), \qquad \mathcal{Y}_i = \sigma(Y_i)$$

In the binary response model we have:

$$(\mathcal{A} \vee \mathcal{Z}_i)\mathcal{Y}_i = \sigma\{E(Y_i \mid Z_i, \theta)\} \supset \sigma\{\varphi(\alpha + \beta' Z_i)\}$$

where ϕ, α ,β are random elements the distribution of which is taken conditionally on Z_i. Recall also that the null subsets of $\mathcal{A} \vee \mathcal{Z}_i$ are included in $(\mathcal{A} \vee \mathcal{Z}_i)\mathcal{Y}_i$.
The main result of this paper is given by the following theorem:

Theorem: The Bayesian semi-parametric linear model defined by (1), (8) and (17) is identified in a Bayesian framework under the following conditions.

$(A1)$ $\{Z_i(\omega) : i \in I\!N\}$ is a dense subset of $I\!R^p$ $\forall \omega \in \Omega - N$ where $N \in \mathcal{Z}_1^\infty$ and $P(N) = 0$ (i.e. $\{Z_i : i \in I\!N\}$ is a.s. dense in $I\!R^p$).

$(A2)$ $\exists\, j_0$ such that $P[\beta_{j_0} \neq 0] = 1$

$(A3)$ F_0 is not a point mass, i.e. $\exists\, x_0 \in I\!R$ such that $0 < F_0(x_0) < 1$

$\blacksquare$

The proof of this theorem, based on three lemmas, rests on the following idea. Consider first a particular value $x \in I\!R$. In a first lemma we show that for any $j = 1, \cdots, p$, the real-valued function of θ, $\varphi(\alpha + \beta_j x)$, is identified i.e. is a.s. $(\mathcal{A} \vee \mathcal{Z}_1^\infty)\mathcal{S}$-measurable, where $\mathcal{S} = \mathcal{X}_1^\infty = \mathcal{Y}_1^\infty \vee \mathcal{Z}_1^\infty = \bigvee_{i \in I\!N}(\mathcal{Y}_i \vee \mathcal{Z}_i)$ (see Appendix) because the trajectories in $\overline{Z_1^\infty}$ are dense in $I\!R^p$ almost surely only. Next consider the function-valued functions of θ , i.e. the σ-field

$$\mathcal{U}_j = \sigma(\{\varphi(\alpha + \beta_j x) : x \in I\!R\})$$

along with

$$\mathcal{U} = \bigvee_{1 \leq j \leq p} \mathcal{U}_j$$

Thus Lemma 3.1 guarantees that $\mathcal{U} \subset \overline{(\mathcal{A} \vee \mathcal{Z}_1^\infty)\mathcal{S}}$.
It will be convenient to reparametrize $(\varphi, \alpha, \beta_j : 1 \leq j \leq p)$ into $(\psi_j, \alpha, \gamma_j, \beta_{j_0})$ where j_0 is such that $P(\beta_{j_0} = 0) = 0$ (A.2), and

$$\begin{aligned} \gamma_j &= \frac{\beta_j}{\beta_{j_0}} \, 1\!\!1_{\{\beta_{j_0} \neq 0\}} \qquad j \neq j_0 \\ \psi_j(x) &= \varphi(\alpha + \beta_j x) \qquad 1 \leq j \leq p \end{aligned}$$

So that $\mathcal{A} = \sigma(\varphi, \alpha, \beta_j : 1 \leq j \leq p) = \sigma(\psi_j, \alpha, \gamma_j, \beta_{j_0} : 1 \leq j \leq p)$. Then the proof that $\mathcal{A} \subset \overline{\mathcal{U}}$ will be decomposed into $\psi_j, \alpha, \gamma_j, \beta_{j_0} \in [\overline{\mathcal{U}}]$, where $[\overline{\mathcal{U}}]$ denotes the set of real-valued $\overline{\mathcal{U}}$-measurable functions. The second lemma proves that for any $j \neq j_0$, γ_j is $\overline{\mathcal{U}}$-measurable. The third lemma shows that α and β_{j_0} are also $\overline{\mathcal{U}}$-measurable (more precisely they

are $\overline{\mathcal{U}}_{j_0}$-measurable). The proof is completed by noting that for any x, $\varphi(x)$ is also $\overline{\mathcal{U}}_{j_0}$-measurable.
Remark: Since $\overline{(\mathcal{A} \vee \mathcal{Z}_1^\infty)\mathcal{S}}$ contains all the null sets of $(\mathcal{A} \vee \mathcal{Z}_1^\infty)$, we only have to prove $\mathcal{A} \subset \overline{\mathcal{U}}(\subset \overline{(\mathcal{A} \vee \mathcal{Z}_1^\infty)\mathcal{S}})$.

Note that for a particular value $x \in I\!R$ and a particular $j(j = 1, \cdots, p)$, $\varphi(\alpha + \beta_j x)$ is a function of the parameters, i.e. a random variable $\mathcal{A}$-measurable. The first lemma aserts that such a function is identified.

Lemma 3.1 *Under* $(A1)$ *and* $(A2)$, *for any* $x \in I\!R$ *and* $j = 1, \cdots, p$*:*

$$\psi_j(x) \in [\overline{(\mathcal{A} \vee \mathcal{Z}_1^\infty)\mathcal{S}}]$$

Proof: First notice that a basic property of the Dirichlet process (17(iii)) ensures that φ is a.s. an increasing right-continuous function. Therefore, under $(A1)$ and $(A2)$, we may repeat the same argument as in Section (2.5) and rewrite (11) as an a.s. equality with respect to the joint probability on $\mathcal{A} \vee \mathcal{Z}_1^\infty$, in the form:

$$\varphi(\alpha + \beta_j x) = \inf_{k \in I\!N_0} \sup_{\{i : Z_i(\omega) \in I_{\frac{1}{k}}(j,x)\}} \varphi(\alpha + \beta' Z_i)$$

Clearly, each $\varphi(\alpha+\beta' Z_i)$ is measurable with respect to $\overline{(\mathcal{A} \vee \mathcal{Z}_i)\mathcal{Y}_i}$, thus the right-hand side is a random variable measurable with respect to $\overline{(\mathcal{A} \vee \mathcal{Z}_1^\infty)\mathcal{S}}$.
∎

Lemma 3.2

$$\gamma_j \in [\overline{\mathcal{U}}] \qquad \forall\, j \neq j_0$$

Proof: Let us first remark that because φ is distributed as a Dirichlet process (17(iii)), we may restrict the attention to the trajectories almost surely satisfying: $\lim_{y \to \infty} \varphi(y) = 1$ and $\lim_{y \to -\infty} \varphi(y) = 0$. Since $P[\beta_{j_0} = 0] = 0$, we have:

$$\begin{aligned} \{\beta_{j_0} > 0\} &= \{\lim_{x \to \infty} \psi_{j_0}(x) = 1\} \\ \{\beta_{j_0} < 0\} &= \{\lim_{x \to \infty} \psi_{j_0}(x) = 0\} \end{aligned}$$

Similarly, $\forall\, j \neq j_0$,

$$\begin{aligned}
\{\beta_j = 0\} &= \{\lim_{x\to\infty} \psi_j(x) = \lim_{x\to-\infty} \psi_j(x)\} \\
\{\beta_j > 0\} &= \{\lim_{x\to\infty} \psi_j(x) = 1, \lim_{x\to-\infty} \psi_j(x) = 0\} \\
\{\beta_j < 0\} &= \{\lim_{x\to\infty} \psi_j(x) = 0, \lim_{x\to-\infty} \psi_j(x) = 1\}
\end{aligned}$$

From Lemma 3.1 each right-hand side event is clearly in the corresponding $\mathcal{U}_j$ and therefore in $\mathcal{U}$. So we will show that

$$\gamma_j \mathbb{1}_{\{\beta_{j_0}>0,\, \beta_j>0\}} \in \overline{\mathcal{U}}$$

The proof is the same (mutatis mutandis) for the other cases, i.e., $\beta_{j_0} > 0$ and $\beta_j < 0$ or $\beta_{j_0} < 0$ and $\beta_j > 0$ or $\beta_{j_0} < 0$ and $\beta_j < 0$.

Let us consider the right median of ψ_{j_0}, i.e.

$$\eta \quad = \quad \inf\left\{x : \psi_{j_0}(x) > \frac{1}{2}\right\}$$

Thus η is a stopping time for the process ψ_{j_0} and therefore

$$\eta, \psi_{j_0}(\eta-), \psi_{j_0}(\eta) \in [\mathcal{U}_{j_0}].$$

Now by definition,

$$0 \leq \psi_{j_0}(\eta-) \leq \frac{1}{2} \leq \psi_{j_0}(\eta) \leq 1$$

Note that $\varphi(\alpha+\beta_{j_0}x)$ cannot be degenerate at η. Indeed, as $\varphi \sim \mathcal{D}i(n_0F_0)$, $\varphi = \epsilon_{\alpha+\beta_{j_0}\eta}$ would imply $F_0 = \epsilon_{x_0}$ (and $x_0 = \alpha + \beta_{j_0}\eta)$) in contradiction with $(A3)$.

Therefore we have almost surely $\psi_{j_0}(\eta-) > 0$ or $\psi_{j_0}(\eta) < 1$ (or both). Now, assuming $\psi_{j_0}(\eta) < 1$, let

$$\begin{aligned}
\kappa_1 &= \frac{1}{2}\psi_{j_0}(\eta) \\
\kappa_2 &= \frac{1}{2} + \frac{1}{2}\psi_{j_0}(\eta) \\
\eta_1 &= \inf\{x : \psi_{j_0}(x) > \kappa_1\} \\
\eta_2 &= \inf\{x : \psi_{j_0}(x) > \kappa_2\}
\end{aligned}$$

Clearly

$$0 < \kappa_1 < \psi_{j_0}(\eta) < \kappa_2 < 1$$

and by definition

$$0 \leq \psi_{j_0}(\eta_1-) \leq \kappa_1 < \psi_{j_0}(\eta) < \kappa_2 \leq \psi_{j_0}(\eta_2) \leq 1$$

therefore

$$-\infty < \eta_1 \leq \eta < \eta_2 < \infty$$

i.e. $\eta_2 - \eta_1 > 0$. Note also that $\kappa_1, \kappa_2, \eta_1$ and $\eta_2 \in [\mathcal{U}_{j_0}]$.
Now, it is easy to see that

$$\psi_j(x) \quad = \quad \psi_{j_0}(\gamma_j x)$$

and if we define

$$\eta_{1j} \quad = \quad \inf\{x : \psi_j(x) > \kappa_1\}$$

$$\eta_{2j} \quad = \quad \inf\{x : \psi_j(x) > \kappa_2\}$$

$\eta_{1j},\, \eta_{2j} \in [\overline{\mathcal{U}}_{j_0} \vee \overline{\mathcal{U}}_j] \subset [\overline{\mathcal{U}}]$ and

$$\eta_{1j} = \frac{1}{\gamma_j}\eta_1 \quad \text{and} \quad \eta_{2j} = \frac{1}{\gamma_j}\eta_2$$

Therefore

$$\eta_{2j} - \eta_{1j} \quad = \quad \frac{\eta_2 - \eta_1}{\gamma_j} \; > \; 0$$

and on $\{\beta_{j_0} > 0, \beta_j > 0\}$:

$$\gamma_j \quad = \quad \frac{\eta_2 - \eta_1}{\eta_{2j} - \eta_{1j}} \in [\overline{\mathcal{U}}]$$

Note that similar arguments may be developed when $\psi_{j_0}(M-) > 0$. ■

Lemma 3.3

$$\alpha \in [\overline{\mathcal{U}}_{j_0}] \qquad \textit{and} \qquad \beta_{j_0} \in [\overline{\mathcal{U}}_{j_0}]$$

Proof: Note first that a Dirichlet process is preserved under functional transformation, i.e. $\varphi \sim \mathcal{D}i(n_0 F_0)$ and $g : I\!R \rightarrow I\!R$ Borel function imply $\varphi \circ g^{-1} \sim \mathcal{D}i(n_0 F_0 \circ g^{-1})$. This entails that

$$\psi_j \mid \alpha, \beta \sim \mathcal{D}i(n_0 \xi_j) \tag{18}$$

where

$$\xi_j(x) = F_0(\alpha + \beta_j x)$$

Let us denote the set of atoms of F_0 as follows:

$$\begin{aligned} D &= \{x \in \mathbb{R} : F_0(x) - F_0(x-) > 0\} \\ &= \{d_j : j \in J\} \end{aligned}$$

where J is at most countable. Let also:

$$f_{0j} = F_0(d_j) - F_0(d_j-)$$

Let us now decompose F_0 into its difuse (F_{0c}) and its discrete (F_{0d}) parts, i.e.

$$F_0 \quad = \quad l_0 F_{0c} + (1 - l_0) F_{0d}$$

where

$$\begin{aligned} 1 - l_0 &= \sum_{j \in J} f_{0j} \\ F_{0c}(x) &= F_0(x \mid D^c) \\ F_{0d}(x) &= F_0(x \mid D) = \frac{1}{1 - l_0} \sum_{j \in J} f_{0j} \mathbb{1}_{\{d_j \leq x\}} \end{aligned}$$

(for definiteness take $F_{0c} = 0$ if $l_0 = 0$ or $F_{0d} = 0$ if $l_0 = 1$). Let us now decompose φ with respect to D as follows:

$$\varphi = \lambda \varphi_c + (1 - \lambda) \varphi_d \tag{19}$$

where

$$\begin{aligned} \varphi_{0j} &= \varphi(d_j) - \varphi(d_j-) \\ \lambda &= 1 - \sum_{j \in J} \varphi_{0j} \\ \varphi_c(x) &= \varphi(x \mid D^c) \\ \varphi_d(x) &= \varphi(x \mid D) = \frac{1}{1 - \lambda} \sum_{j \in J} \varphi_{0j} \mathbb{1}_{\{d_j \leq x\}} \end{aligned}$$

This decomposition has the following property (see, e.g., Rolin (1992) or Florens, Rolin (1994)):

$$
\begin{array}{lll}
(i) & \lambda & \perp\!\!\!\perp \varphi_c \perp\!\!\!\perp \varphi_d \\
(ii) & \lambda & \sim Be(n_0 l_0, n_0(1-l_0)) \\
(iii) & \varphi_c & \sim \mathcal{D}i(n_0 l_0 F_{0c}) \\
(iv) & \varphi_d & \sim \mathcal{D}i(n_0(1-l_0)F_{0d})
\end{array}
$$

Let us look at the transformed process ψ_{j_0} and define the set of atoms of ξ_{j_0}

$$D_0 \quad = \quad \{x : \xi_{j_0}(x) - \xi_{j_0}(x-) > 0\}$$

clearly

$$D_0 \quad = \quad \left\{ \frac{x-\alpha}{\beta_{j_0}} : x \in D \right\} = \left\{ \frac{d_j - \alpha}{\beta_{j_0}} : j \in J \right\} = \{\delta_{0j} : j \in J\} \tag{20}$$

In view of (18), it is known that the trajectories of $(\psi_{j_0} \mid \alpha, \beta)$ are a.s discrete. Let us decompose its jumps into those of fixed locations, i.e. in D_0, and those randomly located in D_0^c. More specifically, for any $x \in I\!R$, we have on $\{\beta_{j_0} > 0\}$:

$$
\begin{aligned}
\frac{1}{ln(r)} \sum_{y \leq x} 1\!\!1_{\{\psi_{j_0}(y) - \psi_{j_0}(y-) > \frac{1}{r}\}} \quad &= \quad \frac{1}{ln(r)} \sum_{\{y \in D_0 | y \leq x\}} 1\!\!1_{\{\psi_{j_0}(y) - \psi_{j_0}(y-) > \frac{1}{r}\}} \\
&\quad + \frac{1}{lnr} \sum_{\{y \in D_0^c | y \leq x\}} 1\!\!1_{\{\psi_{j_0}(y) - \psi_{j_0}(y-) > \frac{1}{r}\}}
\end{aligned} \tag{21}
$$

Now, as $r \to \infty$, conditionally on α, β_j, it may be shown that the first term of the right-hand side of (21) tends to zero in probability (this is trivial if J is finite and this will always be the case in practical situations). The second term may be written as

$$\frac{ln(\lambda r)}{lnr} \frac{1}{ln(\lambda r)} \sum_{y \leq x} 1\!\!1_{\{\psi_{cj_0}(y) - \psi_{cj_0}(y-) > \frac{1}{\lambda r}\}}$$

where $\psi_{cj_0}(y) = \varphi_c(\alpha + \beta_{j_0} y)$.

Note that

$$\frac{ln(\lambda r)}{lnr} \to 1 \qquad \text{as} \qquad r \to \infty$$

and by a result of Florens, Mouchart and Rolin (1992) (Corollary A.3), conditionally on $(\alpha, \beta_j, \lambda)$,

$$\frac{1}{ln(\lambda r)} \sum_{y \leq x} \mathbb{1}_{\{\psi_{cj_0}(y) - \psi_{cj_0}(y-) > \frac{1}{\lambda r}\}}$$

converges in probability to $n_0 l_0 \xi_{cj_0}(x)$, where $\xi_{cj_0}(x) = F_{0c}(\alpha + \beta_{j_0} x)$.
By Egoroff's theorem, there exists an almost surely unconditionally convergent subsequence (see, e.g., Neveu(1964), Proposition II.4.3). Therefore, if $\beta_{j_0} > 0$ and if the continuous part is present, (i.e. $F_{0c} \neq 0$), then

$$\xi_{cj_0}(x) \in [\overline{\mathcal{U}}_{j_0}] \qquad \forall\, x \in I\!R$$

Now let us consider p_1, p_2 such that $0 < p_1 < p_2 < 1$ and define:

$$m_i \;=\; \inf\{x : F_{0c}(x) > p_i\} \qquad i = 1, 2$$

Clearly

$$0 < F_{0c}(m_1) = p_1 < F_{0c}(m_2) = p_2 < 1$$

$$\text{and} \qquad -\infty < m_1 < m_2 < +\infty$$

Let us define

$$\nu_i \;=\; \inf\{x : \xi_{cj_0}(x) > p_i\} = \frac{m_i - \alpha}{\beta_{j_0}} \qquad i = 1, 2$$

then

$$\nu_i \in \overline{\mathcal{U}}_{j_0} \qquad i = 1, 2$$

Hence

$$\nu_2 - \nu_1 = \frac{m_2 - m_1}{\beta_{j_0}} > 0$$

and

$$\beta_{j_0} \;=\; \frac{m_2 - m_1}{\nu_2 - \nu_1} \in [\overline{\mathcal{U}}_{j_0}]$$

$$\alpha \;=\; m_1 - \frac{m_2 - m_1}{\nu_2 - \nu_1}\nu_1 \in [\overline{\mathcal{U}}_{j_0}]$$

Finally, if $F_{0c} = 0$, under assumption (A3) there are at least two fixed points of discontinuity and in view of (19) and (20),

$$\psi_{j_0} \;=\; \sum_{j \in J} \varphi_{0j} \epsilon_{\delta_{0j}}$$

Choosing (in a measurable way) $\delta_{0j} \neq \delta_{0j'}$ two discontinuity points of ψ_{j_0} we obtain

$$\begin{aligned}
\beta_{j_0} &= \frac{d_{j'} - d_j}{\delta_{0j'} - \delta_{0j}} \in [\overline{\mathcal{U}}_{j_0}] \\
\alpha &= d_j - \frac{d_{j'} - d_j}{\delta_{0j'} - \delta_{0j}} \delta_{0j} \in [\overline{\mathcal{U}}_{j_0}]
\end{aligned}$$

∎

To complete the proof of the theorem, it suffices to note that

$$\forall\, x \in I\!R: \qquad \varphi(x) \;=\; \psi_{j_0}\left(\frac{x-\alpha}{\beta_{j_0}}\right) \in [\overline{\mathcal{U}}_{j_0}]$$

and that

$$\beta_j \;=\; \gamma_j \beta_{j_0} \qquad \forall\, 1 \leq j \leq p,\; j \neq j_0$$

Remark: It may be noticed that using the argument of convergence of lemma 3, the proof of lemma 2 would have been greatly simplified. However, we tried to delay as far as possible the use of the prior specification of φ, i.e., the Dirichlet process. The only reference to this prior in Lemma 2 serves to obtain the following property: if $\varphi = \epsilon_X$ where X is a random variable, i.e., φ is a random point mass, then if $\varphi \sim \mathcal{D}i(n_0 F_0)$ and if $n_0 > 0$, then $X = x$ a.s., i.e. φ is a fixed point mass. Let us remind that if $n_0 \rightarrow 0$ (this corresponds to an improper prior), then $\varphi \rightarrow \epsilon_X$ weakly where X is a random variable with distribution F_0 (see, e.g., Tiwari, Sethuraman (1982)). In such a case if $\varphi = \varepsilon_X$, on $\{\beta_{j_0} > 0, \beta_j > 0\}$, $\psi_j = \varepsilon_{X_j}$ and $\psi_{j_0} = \varepsilon_{X_{j_0}}$ where $X_j = \dfrac{X-\alpha}{\beta_j} = \dfrac{1}{\gamma_j} X_{j_0} = \dfrac{1}{\gamma_j}\dfrac{X-\alpha}{\beta_{j_0}}$. Therefore, even in this degenerate case, $\gamma_j \in [\mathcal{U}] \quad \forall\, j \neq j_0$, since $\gamma_j = \dfrac{X_{j_0}}{X_j}$. But clearly α and β_{j_0} are not $\mathcal{U}$-measurable since $\mathcal{U}_{j_0} = \sigma\{\frac{X-\alpha}{\beta_{j_0}}\}$.

4 Comments

The practical implication of section 2 and 3 is the following. In a sampling theory framework, the model $E(Y|Z) = \varphi(\alpha + \beta' Z)$ raises an identification problem typically solved either by constraining φ through moment (zero-mean and unit-variance) or fractile conditions or by constraining the coefficients (typically $\alpha = 0$ and $\beta_1 = 1$). In a Bayesian framework, under a rather natural prior specification, one should not restrict the prior specification to be supported by a restricted set of the link functions φ or a restricted set of the euclidean parameters (α, β). Adding such restrictions acts as restrictions on the sampling process, possibly "rejected" by the data.

This is an interesting example of a discrepancy between sampling theory and Bayesian identification. A deeper understanding of such a result requires further elaboration. This is the object of the next coming comments.

Note first that in a semi-parametric specification, identification needs to be defined in the asymptotic model i.e. the sampling model generating $X_1^\infty = \{X_i = (Y_i, Z_i) : i \in \mathbb{N}_0\}$: there is indeed no hope that an infinity of characteristics of φ could be identified from a finite sample. Next, the very idea of a conditional model is that the specification of the marginal process, generating $Z_1^\infty = \{Z_i, i \in \mathbb{N}_0\}$, should be as general as possible leaving to the conditional model, generating $(Y_1^\infty | Z_1^\infty)$, the burden of the most specific hypotheses.

In the model under consideration, the main hypothesis on the marginal process is that the trajectories Z_1^∞ should conform a dense subset of $\mathbb{R}^p$ with probability 1 (condition (i)). This seemingly natural requirement is identical in the sampling theory framework and in the Bayesian framework; it is satisfied, for instance, when Z_i is (infinitely) generated by an i.i.d. process with a full rank covariance matrix but is also satisfied in most "regular" dynamic processes. For the conditional model, both in the sampling theory and in the Bayesian framework, it is assumed that the sequence Y_1^∞ is sifted, in the sense of Theorem 7.6.10 of Florens, Mouchart, Rolin (1990), by Z_1^∞, i.e., that in the sampling given Z_1^∞, the Y_i's are mutually independent and depends on Z_i only; the Bernoulli character of Y_i is a simple consequence of its binary nature.

The main difference relies in the concept of identification. In sampling theory framework, the traditional concept of identification, in a conditional model, concerns the injectivity of the mapping $\theta \rightarrow P^{\theta, \mathcal{Z}}$, the sampling probability on $(S, \mathcal{S})$ conditional to $\mathcal{Z}$; whether this concept is pointwise on the trajectory of Z or for almost all trajectories of Z, is not at stake in this discussion. The Bayesian concept relates to the measurability of every function of the parameter with respect to the σ-field generated by the sampling expectations. Rather than entering into technicalities about the relationship between these two approaches (for an introduction to this theme see e.g. E.B.S. Section 4.6.2., p.171), let us notice that the standard identification condition in a sampling theory consists in fixing arbitrary values for two (well chosen) characteristics of the unknown φ (such as $\mu = 0$ and $\sigma = 1$). Our main theorem means that in a Bayesian framework, the fact of fixing an (arbitrary) prior expectation for φ, namely F_0, also provides identification; the difference is therefore that F_0 is characterized by a countable infinity of characteristics which are imposed "on the average" only whereas in a sampling theory framework two characteristics are imposed with prior probability one.

A Appendix

A.1 Introduction

In this appendix we briefly survey the main tools we use in this paper. They allow for a precise treatment of the identification of a Bayesian conditional model. In a first part we survey the probabilistic tools, namely the identification among σ-fields by means of the projections among σ-fields. Next we use these tools for the treatment of the identification problem, including in particular a comparison with the sampling theory treatment. The general presentation is sketchy: more details, proofs and motivation are given in Florens et al (1990, sections 4.3 and 4.5 for A.2 and 4.6 for A.3) (For short we refer to Florens et al (1990) as E.B.S.).

A.2 Projection and Identification among σ-Fields

Let us consider an abstract probability space $(M, \mathcal{M}, Q)$. $\mathcal{M}_i$ denotes a sub-σ-field of $\mathcal{M}$ and $\overline{\mathcal{M}_i}$ denotes its measurable completion, i.e. the σ-field generated by $\mathcal{M}_i$ and the measurable $Q-$null sets. More precisely, denoting the trivial σ-field (respectively the completed trivial σ-field) as $\mathcal{M}_0 = \{\varphi, M\}$ (respectively as $\overline{\mathcal{M}_0} = \{A \in \mathcal{M} : Q(A)^2 = Q(A)\}$), we have: $\overline{\mathcal{M}_i} = \mathcal{M}_i \vee \mathcal{M}_0$.

Definition (E.B.S. def. 4.3.1): The *projection of* $\mathcal{M}_1$ *on* $\mathcal{M}_2$ is denoted by $\mathcal{M}_2\mathcal{M}_1$ and defined as follows:

$$\mathcal{M}_2\mathcal{M}_1 = \{\mathcal{M}_2 m_1 : m_1 \in [\mathcal{M}_1]^+\}$$

where $[\mathcal{M}_1]^+$ denotes the set of positive $\mathcal{M}_1$-measurable functions defined on M and $\mathcal{M}_2 m_1$ is a short notation for the expectation of m_1 conditional to $\mathcal{M}_2$: $\mathcal{M}_2 m_1 = E(m_1 \mid \mathcal{M}_2)$. ■

Remark: This definition should be interpreted as the σ-field generated by every version of the conditional expectation of every positive $\mathcal{M}_2-$measurable function. This operation eventually depends crucially on the $Q-$null sets. We basically use the following results:

Theorem (E.B.S. theorem 4.3.3): For any $\mathcal{M}_1, \mathcal{M}_2$ and $\mathcal{M}_3$ sub-σ-fields of $\mathcal{M}$:

(i) $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{M}_2\mathcal{M}_1$

(ii) $\mathcal{M}_3 \subset \mathcal{M}_2$ and $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{M}_3 \Rightarrow \mathcal{M}_2\mathcal{M}_1 \subset \overline{\mathcal{M}_3}$

where $\cdot \perp\!\!\!\perp \cdot \mid \cdot$ denotes conditional independence in $(M, \mathcal{M}, P)$. ∎

This theorem says that $\mathcal{M}_2\mathcal{M}_1$ may also be constructed as the intersection of all sub-σ-fields of $\mathcal{M}_2$ containing the $Q-$null sets of $\mathcal{M}_2$ and conditionally on which $\mathcal{M}_1$ and $\mathcal{M}_2$ are independent, i.e.:

$$\mathcal{M}_2\mathcal{M}_1 = \bigcap_{\mathcal{L} \in \mathcal{F}_{1.2}} (\overline{\mathcal{L}} \cap \mathcal{M}_2)$$

where $\mathcal{F}_{1.2} = \{\mathcal{L} \subset \mathcal{M}_2 : \mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{L}\}$: This is Mac Kean(1963)'s definition of projection with a slight modification: Mac Kean used Lebesgue completion whereas we use the measurable completion.

Definition (E.B.S. def. 4.5.1) $\mathcal{M}_1$ *is identified by* $\mathcal{M}_2$ *conditionally on* $\mathcal{M}_3$, and it is denoted

$$\mathcal{M}_1 < \mathcal{M}_2 \mid \mathcal{M}_3$$

if and only if

$$(\mathcal{M}_1 \vee \mathcal{M}_3)(\mathcal{M}_2 \vee \mathcal{M}_3) = \mathcal{M}_1 \vee \mathcal{M}_3$$

∎

In order to interpret this concept, consider a $\sigma-$field as a structure of information. Supposing first $\mathcal{M}_3 = \mathcal{M}_0$ (we shall write $\mathcal{M}_1 < \mathcal{M}_2$ for $\mathcal{M}_1 < \mathcal{M}_2 \mid \mathcal{M}_0$), we have: $\mathcal{M}_1 < \mathcal{M}_2 \Leftrightarrow \mathcal{M}_1\mathcal{M}_2 = \mathcal{M}_1$.

This may be interpreted as saying that the information contained in $\mathcal{M}_2$ is rich enough (with respect to $\mathcal{M}_1$) for re-constructing the information contained in $\mathcal{M}_1$ only from the conditional expectation of the (positive) $\mathcal{M}_2-$measurable function conditionally on $\mathcal{M}_1$. If $\mathcal{M}_3$ is not trivial, we have

$$\mathcal{M}_1 < \mathcal{M}_2 \mid \mathcal{M}_3 \Leftrightarrow (\mathcal{M}_1 \vee \mathcal{M}_3) < (\mathcal{M}_2 \vee \mathcal{M}_3)$$

in which case $\mathcal{M}_3$ represents an information common to both the identified and the identifying $\sigma-$fields.

A.3 Identification in Bayesian Experiments

We shall now use the tools presented in A.2 for the case of a Bayesian experiment, i.e. a probability space

$$\mathcal{E} = (M, \mathcal{M}, Q) = (\Theta \times S, \mathcal{A} \vee \mathcal{S}, Q)$$

where $(\Theta, \mathcal{A})$ is the parameter space, $(S, \mathcal{S})$ is the sample space and Q is constructed from

$$Q(A \times U) = \int_A P^\theta(U) m(d\theta) \qquad A \in \mathcal{A}, U \in \mathcal{S}$$

where m is the prior probability on $(\Theta, \mathcal{A})$ and $P^\theta(\cdot)$ is the sampling transition defined on $(\Theta \times \mathcal{S})$. The dual desintegration of Q is given by:

$$Q(A \times U) = \int_U m^x(A) P(dx)$$

where P is the predictive probability on $(S, \mathcal{S})$: $P(U) = Q(\Theta \times U)$ and $m^x(\cdot)$ is a posterior probability defined on $(S \times \mathcal{A})$. Note that, by construction, $P^\theta(\cdot)$ is a regular conditional probability whereas the existence of a regular version of $m^x(\cdot)$ is subject to conditions to be explicited. In short we shall write:

$$Q = m \otimes P^{\mathcal{A}} = P \otimes m^{\mathcal{S}}$$

Remark: This synthetic presentation of a Bayesian experiment rests on the convention that a σ-field on a factor space is identified with the corresponding σ-field on the product space. More precisely, $\mathcal{A} \vee \mathcal{S}$, as a wedge of σ-fields, should be written as $[\mathcal{A} \times (S, \phi)] \vee [(\Theta, \phi) \times \mathcal{S}]$ and this will not be distinguished from $\mathcal{A} \otimes \mathcal{S}$, the product σ-field, in as far as the context avoids any ambigüity. More details on this kind of presentation may be found in E.B.S., chapter1.

Definition (E.B.S. def. 4.6.1) A *Bayesian experiment is identified* if and only if the σ-field of the parameter is identified by the σ-field of the observation, i.e.

$$\mathcal{A}\mathcal{S} = \mathcal{A}$$

■

In other words, Bayesian identification is viewed as a problem of minimal sufficiency of the parameter, indeed, Theorem A.2.3 precisely means that $\mathcal{A}\mathcal{S}$ is the minimal sufficient parameter on $\mathcal{E}$.
It also means that every (measurable) function of the parameters may be represented a.s. as a function of (countably many) sampling expectations. Note also that, given $\mathcal{A}\mathcal{S} \subset \mathcal{A}$ is always true, proving identification boils down to prove $\mathcal{A} \subset \mathcal{A}\mathcal{S}$.

A.4 Identification in Conditional Bayesian Experiments

Let us now consider a regular conditional Bayesian experiment

$$\mathcal{E}^{\mathcal{Z}} = (\Theta \times S, \mathcal{A} \vee \mathcal{S}, Q^{\mathcal{Z}})$$

where $\mathcal{Z}$ is a sub-σ-field of $\mathcal{S}$, corresponding to a partial observation of $(S, \mathcal{S})$; typically $\mathcal{Z}$ would represent either a statistic, defined on $(S, \mathcal{S})$, or a sub-vector of the observations, often called exogenous variables. The main regularity requirement is that the conditional probability $Q^{\mathcal{Z}}$ admits a regular enough version that ensures the same decomposition as above, namely:

$$Q^{\mathcal{Z}} = m^{\mathcal{Z}} \otimes P^{\mathcal{A} \vee \mathcal{Z}} = P^{\mathcal{Z}} \otimes m^{\mathcal{S}}$$

For reasons commented in E.B.S. (section 4.4) it is natural to look for parameters sufficient for the conditional experiment $\mathcal{E}^{\mathcal{Z}}$, that are minimal in the class of parameters of $\mathcal{E}^{\mathcal{Z}}$ that contain all the null sets of $\mathcal{A} \vee \mathcal{Z}$ and the conditioning information $\mathcal{Z}$. From theorems 4.3.10 and 4.3.13 in E.B.S., such a minimal sufficient parameter is given by the projection $(\mathcal{A} \vee \mathcal{Z})\mathcal{S}$. Hence the following definition.

Definition (E.B.S. def. 4.6.1)) A *conditional Bayesian experiment* $\mathcal{E}^{\mathcal{Z}}$ is *identified* if and only if the σ-field of the parameter is identified by the σ-field of the observation conditionally on $\mathcal{Z}$, i.e.

$$(\mathcal{A} \vee \mathcal{Z})\mathcal{S} = \mathcal{A} \vee \mathcal{Z}$$

∎

Because in general we have (E.B.S proposition 4.3.2(i))

$$\mathcal{Z} \subset (\mathcal{A} \vee \mathcal{Z})\mathcal{S} \subset (\mathcal{A} \vee \mathcal{Z})$$

Bayesian identification in conditional experiments is equivalently verified through the condition

$$\mathcal{A} \subset (\mathcal{A} \vee \mathcal{Z})\mathcal{S}$$

which heuristically means that every $\mathcal{A}$-measurable function may be a.s. obtained as a Borel function of (countably many) $\mathcal{Z}$-conditional sampling expectation of (real-valued, measurable) functions defined on the sample space.

A.5 Identification in Conditional Independent Bayesian Experiments

In a conditional independent Bayesian experiment, we consider an infinite sequence of random vectors $X_i = (Y_i, Z_i), i \in I\!N$, defined on the sample space $(S, \mathcal{S})$. Thus if X_i is valued on $I\!R^m$, we may take $S = (I\!R^m)^{I\!N}$.
The distinction between Y_i and Z_i is based on the property that the sequence (Y_i) is "sifted" by the sequence (Z_i), considered as "exogenous" or "explanatory" variables, more precisely:

$$
\begin{array}{ll}
(i) & \underset{i \in I\!N}{\perp\!\!\!\perp} Y_i \quad | \quad Z_1^\infty, \theta \\
 & \\
(ii) & Y_i \perp\!\!\!\perp Z_1^\infty \quad | \quad Z_i, \theta
\end{array}
\tag{22}
$$

as mentionned in the introduction of Section 2, where Z_1^∞ stands for $\{Z_i : i \in I\!N\}$. From the presentation above, a σ-algebraic approach is most convenient for dealing with Bayesian identification, as this one is basically a measurability condition. Let us accordingly denote the corresponding σ-fields as follows: $\sigma(X_i) = \mathcal{X}_i = \mathcal{Y}_i \vee \mathcal{Z}_i$ with $\mathcal{Y}_i = \sigma(Y_i)$ and $\mathcal{Z}_i = \sigma(Z_i)$; we shall also use the notation $\mathcal{Z} = \mathcal{Z}_1^\infty = \bigvee_{i \in I\!N} \mathcal{Z}_i$. We shall also consider that $\mathcal{S} = \mathcal{X}_1^\infty = \mathcal{Y}_1^\infty \vee \mathcal{Z}_1^\infty$, i.e. there is no more sample information than that coming from the sequence $X_i = (Y_i, Z_i)$, $i \in I\!N$.

From section A.4, the conditional experiment $\mathcal{E}^{\mathcal{Z}}$, with $\mathcal{Z} = \sigma(Z_1^\infty)$, is identified if:

$$\mathcal{A} \subset \overline{(\mathcal{A} \vee \mathcal{Z}_1^\infty)\mathcal{S}}$$

Now by theorem 4.3.9 (ii) and corollary 7.6.7 (i) of E.B.S., we have successively by (22)(i)

$$\overline{(\mathcal{A} \vee \mathcal{Z}_1^\infty)\mathcal{S}} = \mathcal{Z}_1^\infty \vee \overline{(\mathcal{A} \vee \mathcal{Z}_1^\infty)\mathcal{Y}_1^\infty} = \mathcal{Z}_1^\infty \vee \bigvee_{i \in I\!N} \overline{(\mathcal{A} \vee \mathcal{Z}_1^\infty)\mathcal{Y}_i}$$

But by theorem 4.3.6(iii) of E.B.S. and (22)(ii), we have

$$\overline{(\mathcal{A} \vee \mathcal{Z}_1^\infty)\mathcal{Y}_i} = \overline{(\mathcal{A} \vee \mathcal{Z}_i)\mathcal{Y}_i}$$

Therefore, such a model is identified if

$$\mathcal{A} \subset \mathcal{Z}_1^\infty \vee \bigvee_{i \in I\!N} \overline{(\mathcal{A} \vee \mathcal{Z}_i)\mathcal{Y}_i}$$

References

[1] Ferguson T.S. (1973), A Bayesian analysis of some nonparametric problems, *Annals of Statistics*, 1, 209-230.

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

[3] Florens J.P., Mouchart M. and J.-M. Rolin (1992), Bayesian analysis of mixtures: some results on exact estimability and identification, in *Bayesian Statistics*, 4, edited by J.M. Bernardo and al. Oxford Sciences Publications, 127-145.

[4] Florens J.P. and J.-M. Rolin (1994), Bayes, bootstrap, moments, Discussion paper 9413, Institut de Statistique, Université Catholique de Louvain, Belgium.

[5] Mac Kean H.P. Jr (1963), Brownian motion with a several-dimensional time, *Theory of Probability and its Applications*, 8, 335-354.

[6] Neveu, J. (1964) *Bases Mathématiques du Calcul des Probabilités*, Paris:Masson. (Second edition:1970). English translation: *Mathematical Foundations of the Calculus of probability* (1965). San Francisco: Holden-Day.

[7] Rolin J.-M. (1992), Some useful properties of the Dirichlet process, Discussion paper 9202, Institut de Statistique, Université Catholique de Louvain, Belgium.

[8] Sethuraman J. and R.C. Tiwari (1982), Convergence of Dirichlet measures and the interpretation of their parameter, *Statistical Decision Theory and Related Topics* III 2, 305-315.

M. Mouchart and J.-M. Rolin, Institut de Statistique, Université Catholique de Louvain, 20 voie du Roman Pays, B-1348 Louvain-la-Neuve, Belgium.

E. Scheihing, Instituto de Informática, Universidad Austral de Chile, Casilla 567, Valdivia, Chile.