Gaussian processes and Bayesian moment estimation*

Jean-Pierre Florens[†]
Toulouse School of Economics
Université de Toulouse 1 - Capitole

Anna Simoni[‡]
CNRS - THEMA
Université de Cergy-Pontoise

September 4, 2012

– PRELIMINARY –

Abstract

When a large number of moment restrictions is available there may be restrictions that are more important or credible than others. In these situations it might be desirable to weight each restriction based on our beliefs. This is automatically implemented by a Bayesian procedure. We study, in this paper, how to impose moment restrictions on the data distribution through a semiparametric prior distribution for the data generating process F and the structural parameter θ. We show that a Gaussian process prior for the density function associated with F is particularly convenient in order to impose over-identifying restrictions and allows to have a posterior distribution in closed-form. The posterior distribution resulting from our prior specification is shown to be consistent and asymptotically normal.

Key words: Moment conditions, Gaussian processes, overidentification, posterior consistency.

JEL code: C11, C14, C13

1 Introduction

In many practical applications and empirical economic studies a large set of moment restrictions characterizing the parameter of interest is available. Examples are provided for instance in Cazals *et al.* (2004) and Fève *et al.* (2006). Such a situation is complicated to manage since it requires cumbersome computations due to the high number of moment restrictions. It is often the case that the researcher does not equally believe in all the restrictions. Therefore, it is desirable to have a procedure that assigns a specific weight to each moment restriction based on our beliefs and that updates it (and eventually decides

*The authors are grateful to participants to seminars and conferences in: Bristol and NASM 2012 at Northwestern.
†Toulouse School of Economics - 21, allée de Brienne - 31000 Toulouse (France). Email: jean-pierre.florens@tse-fr.eu
‡Corresponding author. CNRS - THEMA, Université de Cergy-Pontoise - 33, boulevard du Port, 95011 Cergy-Pontoise (France). Email: simoni.anna@gmail.com

whether to include or not a restriction) based on the information contained in the data. This may be easily done by a Bayesian procedure where each moment restriction can have a different weight based on the prior and posterior distribution.

The purpose of this paper is to develop a Bayesian approach to the generalized method of moments (GMM). In a Bayesian framework, the computation of a posterior distribution requires the specification of a likelihood function – or sampling distribution – and a prior distribution. In many cases of econometric practice, however, the researcher has only limited information on the data generating process (DGP). This is typical in the GMM framework where the structural information on the DGP is limited to a set of moment conditions. Any parametric specification of the likelihood function is, therefore, completely arbitrary. In this paper we study how to formulate a sampling distribution based only on such a set of moment restrictions.

Let x be a random element in $\mathbb{R}^m$ with distribution F and $x_1, \ldots, x_n$ be an *i.i.d.* sample of x. We are interested in a vectorial parameter $\theta \in \Theta \subset \mathbb{R}^k$ which is linked to F through the relation (moment restrictions)

$$A(\theta, F) = \mathbf{E}^F\left(h(\theta, x)\right) = 0$$

where h is a known function with values in $\mathbb{R}^r$. This model is semiparametric since it includes a finite dimensional structural parameter θ and a functional parameter F which, apart from the moment restrictions, is not at all constraint.

We impose the moment restrictions in the prior for (θ, F) so that the random parameter generated from the prior satisfies the moment restrictions by construction. Specification of semiparametric priors which incorporate moment restrictions may encounter difficulties depending on the relationship existing between θ and F. More precisely, when the model is just-identified, that is $k = r$, the relation $A(\theta, F) = 0$ characterizes θ as an explicit function of F: $\theta = B(F)$, where B is a function defined on the space of probability distributions. For particular functional forms of B, the prior of θ may be recovered from the prior of F and automatically satisfies the constraints.

On the contrary, in an overidentified model where $k < r$, a solution to $A(\theta, F) = 0$ exists only for some particular F so that the distribution F must be constraint to guarantee the existence of a solution to the moment equation. In a Bayesian approach this entails that if we endow F with a prior distribution then this one can not be determined independently of θ and vice-versa. In an overidentified model, the restrictions on F are in general complicated to incorporate in its prior distribution. The approach proposed in Florens and Rolin (1994), for instance, which is based on a Dirichlet process prior distribution, presents several difficulties to deal with overidentified models. Our proposal improves the treatment of overidentified models and allows to deal with just-identified as well as over-identified models by imposing easily the moment restrictions in the prior for the data probability density function.

The purpose of developing Bayesian estimation under moment restrictions has already been undertaken by several papers. Kitamura and Otsu (2011) use a Dirichlet process prior (see Ferguson (1973, 1974)) and then construct the restricted prior on F by minimizing the Kullback-Leibler divergence with respect to the Dirichlet process prior under the moment constraint. A Dirichlet process prior has nice properties due to the fact that it is a natural conjugate of the *i.i.d.* model, however the treatment of the overidentified case is much more complicated. Empirical Likelihood methods have been proposed in the literature as an alternative to the Dirichlet process prior. Kim (2002) proposes a limited information likelihood approach which allows to derive a posterior distribution for θ even when the true likelihood is not available. Schennach (2005) proposes a maximum entropy nonparametric Bayesian procedure which, instead of employing the Dirichlet process prior, rely on a non-informative prior on the space of distributions.

This paper proposes a new Bayesian approach to GMM based on Gaussian process ($\mathcal{GP}$) priors. At the best of our knowledge this prior has not been used yet in the GMM framework. We do not restrict the DGP F except for the fact that we assume it admits a density function f with respect to some positive measure Π and satisfies the moment restrictions. Then, we specify a $\mathcal{GP}$ prior for f conditional on θ. The essential reason for the appropriateness of a $\mathcal{GP}$ prior in a GMM framework is due to the fact that $A(\theta, F) = 0$ is a linear constraint in f. The linearity of the model matches extremely well with a $\mathcal{GP}$ prior since it allows to incorporate the (over-identifying) moment conditions in an easy way by constraining the prior mean and prior covariance of f.

An advantage of our method is that, in both the just-identified and overidentified cases, the moment restrictions are imposed directly through the (conditional) prior of f (given θ) without requiring a second step projection as in Kitamura and Otsu (2011). In the overidentified case we first specify a prior on θ and then we specify a $\mathcal{GP}$ prior on f conditional on θ. In the just-identified case we may either proceed as in the overidentified case or specify an unrestricted $\mathcal{GP}$ prior on f and deduce from it the prior for θ through the (linear) transformation $\theta = B(f)$. After observing the data we compute the posteriors – both marginal and conditional – for θ and f. For estimation purposes we are interested in the marginal posterior distribution of θ. This is usually not available in closed-form but it is possible to simulate easily from it by using MCMC methods.

The second main novelty of our approach is the way in which we construct the sampling distribution. Instead of using directly F, we construct a functional transformation of the data set that weakly converges towards a $\mathcal{GP}$. In this way our analysis benefits of the advantages of a conjugate model without assuming any functional form for the sampling distribution. The motivation for this choice is that if we used F as the sampling distribution, then we would have neither a conjugate model nor a closed form for the posterior distribution of f given θ. On the contrary our approach allows for conjugacy and makes computations quite easy.

In the next section we present our approach. In section 3 we analyze asymptotic properties of the posterior distribution of θ and of f. In section 4 we detail how to implement our method for both the just identified case and the overidentified case.

2 The General Semiparametric Model

Throughout the paper we denote the true data generating process by F_* and its density with respect to some positive measure Π by f_*. Therefore, $x_1, \ldots, x_n$ are *i.i.d.* observations each distributed according to F_*. The data generating process for x could be more general than an *i.i.d.* sampling process but we focus on this case for simplicity. We denote by θ_* the true value of θ which satisfies $\mathbf{E}^{F_*}(h(\theta_*, x)) = 0$.

The general model is based on the relation $\mathbf{E}^F(h(\theta, x)) = 0$ where $h : \Theta \times \mathbb{R}^m \to \mathbb{R}^r$ is a known function and F is absolutely continuous with respect to some positive measure Π with density function f. The parameters of the model are (θ, f). While θ is the parameter of interest and has finite dimension, f is a functional nuisance parameter. Let $\Theta \subseteq \mathbb{R}^k$ and $\mathcal{E}_M \subseteq M$ where M denotes the set of probability density functions on $\mathbb{R}^m$. The parameter space is

$$\Lambda = \left\{ (\theta, f) \in \Theta \times \mathcal{E}_M; \int h(\theta, x) f(x) d\Pi = 0 \right\}$$

so that a prior distribution on (θ, f) must incorporate the moment restriction. The model is made up of three elements that we detail in the following: a prior on θ, a conditional prior on f, given θ, and the sampling model.

2.1 Prior distribution

We put a prior probability measure μ on the pair (θ, f) of the form $\mu = \mu_\theta \otimes \mu_f^\theta$, where μ_θ denotes a marginal distribution on θ and μ_f^θ denotes a conditional probability distribution on f given θ.

Prior on θ

The parameter of interest $\theta \in \Theta \subset \mathbb{R}^k$ is endowed with a prior distribution, denoted by μ_θ. If it admits a density with respect to the Lebesgue measure we denote this density by $\mu_\theta(\theta)$ as well, by abuse of notation. We can specify any prior distribution which incorporates any information available to the econometrician about the parameter θ of interest.

Conditional prior on f given θ

Let S be a subset of $\mathbb{R}^m$ endowed with the trace of the Borelian σ-field $\mathfrak{B}_S$ and Π be a measure on this subset. We denote by $\mathcal{E} = L^2(S, \mathfrak{B}_S, \Pi)$ the Hilbert space of square integrable functions on S and by $\mathfrak{B}_{\mathcal{E}}$ the Borel σ-field generated by the open sets of $\mathcal{E}$. We assume that the true probability density function (*pdf*) f_* belongs to the space $\mathcal{E}_M := \mathcal{E} \cap M$. The function f is the functional parameter of our model and since it is the density of F with respect to Π it must satisfy the restriction $\int f d\Pi = 1$. Further, we make the assumption of square integrability of f with respect to Π, that is, $\int f^2 d\Pi < \infty$. This restriction reduces the parameter space to a subset of M and is verified for instance if f is bounded and Π is a bounded measure.

The conditional prior distribution of f, conditional on θ, is specified as a Gaussian distribution on the Borel σ-field generated by the open sets of $\mathcal{E}$ with mean function $f_{0\theta} \in \mathcal{E}_M$ and covariance operator $\Omega_{0\theta} : \mathcal{E} \to \mathcal{E}$. We denote this prior distribution by μ_f^θ. The covariance operator $\Omega_{0\theta}$ is one-to-one, linear, positive semidefinite, self-adjoint and trace-class. A trace-class operator is a compact operator with eigenvalues that are summable. Remark that this guarantee that the trajectories f generated by μ_f^θ satisfy $\int f^2 d\Pi < \infty$.

This prior distribution has to be "compatible" with the moment conditions. This means that, for any given θ, μ_f^θ must generate *pdf*s f that satisfy the moment conditions with probability 1. We implement this by imposing the following restrictions on $f_{0\theta}$ and $\Omega_{0\theta}$.

Restriction 1 (Restrictions on $f_{0\theta}$)**.** The prior mean function $f_{0\theta}$ has to be a *pdf* on S with respect to Π and has to verify the condition

$$\int h(\theta, x) f_{0\theta}(x) \Pi(dx) = 0. \tag{2.1}$$

Restriction 2 (Restrictions on $\Omega_{0\theta}$)**.** The operator $\Omega_{0\theta}$ must be specified such that

$$\begin{cases} \Omega_{0\theta}^{1/2} h(\theta, x) & = & 0 \\ \Omega_{0\theta}^{1/2} 1 & = & 0. \end{cases} \tag{2.2}$$

The conditions in (2.2) imply that the operator $\Omega_{0\theta}$ is not injective. In fact, the null space of $\Omega_{0\theta}$, denoted by $\mathcal{N}(\Omega_{0\theta})$, contains effectively the constant 1 – which implies that the trajectory f generated by the prior integrates to 1 almost surely – and the function $h(\theta, x)$ – which implies that the trajectory f satisfies almost surely the moment condition. In practice, this means that $\Omega_{0\theta}$ is degenerate in the directions along which we want that the corresponding projections of f and $f_{0\theta}$ are equal. This is the meaning of the next lemma

Lemma 2.1. *The conditional Gaussian prior distribution μ_f^θ, with mean function $f_{0\theta}$ and covariance operator $\Omega_{0\theta}$ satisfying the restrictions 1 and 2, generates trajectories f which satisfy μ_f^θ-a.s. the conditions*

$$\int f(x)\Pi(dx) = 1 \quad \textit{and} \quad \int h(\theta, x) f(x)\Pi(dx) = 0.$$

Proof. Let $\mathcal{H}(\Omega_{0\theta})$ denote the reproducing kernel Hilbert space associated with $\Omega_{0\theta}$ and embedded in $\mathcal{E}$ and $\overline{\mathcal{H}(\Omega_{0\theta})}$ denote its closure. If $f|\theta \sim \mathcal{N}(f_{0\theta}, \Omega_{0\theta})$ then $(f - f_{0\theta}) \in \overline{\mathcal{H}(\Omega_{0\theta})}$, μ_F^θ-almost surely. Moreover, $\mathcal{H}(\Omega_{0\theta}) = \mathcal{D}(\Omega_{0\theta}^{-1/2}) = \mathcal{R}(\Omega_{0\theta}^{1/2})$ where $\mathcal{D}$ and $\mathcal{R}$ denote the domain and the range of an operator, respectively. This means that $\forall \varphi \in \mathcal{H}(\Omega_{0\theta})$ there exists $\psi \in \mathcal{E}$ such that $\varphi = \Omega_{0\theta}^{\frac{1}{2}}\psi$. Moreover, for any $\varphi \in \mathcal{H}(\Omega_{0\theta})$ we have $< \varphi, h(\theta, \cdot) > = \int \varphi(x) h(\theta, x)\Pi(dx) = < \Omega_{0\theta}^{\frac{1}{2}}\psi, h(\theta, \cdot) > = < \psi, \Omega_{0\theta}^{\frac{1}{2}} h(\theta, \cdot) > = 0$ and $< \varphi, 1 > = 0$ by a similar argument. Hence,

$$\mathcal{H}(\Omega_{0\theta}) \subset \left\{ \varphi \in \mathcal{E} \; ; \; \int \varphi(x) h(\theta, x)\Pi(dx) = 0 \text{ and } \int \varphi(x)\Pi(dx) = 0 \right\}. \tag{2.3}$$

Since the set on the right of this inclusion is closed we have

$$\overline{\mathcal{H}(\Omega_{0\theta})} \subset \left\{ \varphi \in \mathcal{E} \; ; \; \int \varphi(x) h(\theta, x)\Pi(dx) = 0 \text{ and } \int \varphi(x)\Pi(dx) = 0 \right\}.$$

We deduce that μ_F^θ-almost surely

$$\int (f - f_{0\theta})(x)\Pi(dx) = 0 \quad \text{and} \quad \int (f - f_{0\theta})(x) h(\theta, x)\Pi(dx) = 0.$$

Condition (2.1) and the fact that $f_{0\theta}$ is a *pdf* imply the results of the lemma.

□

Remark 2.1. Our assumption implies that $\int f d\Pi = 1$ but it does not ensure that $f \geq 0$. This condition is incompatible with the choice of a Gaussian prior. The alternative would be to write $f = g^2$, $g \in \mathcal{E}$, and to specify a conditional prior distribution, given θ, for g instead of for f. We do not pursue this approach here since it would lead to a non-linear inverse problem that is beyond the scope of this paper.

From a practical implementation point of view, the construction of a covariance operator $\Omega_{0\theta}$ which satisfies (2.2) may appear complicated. In reality, such a construction may be realized quite easily by using the following procedure based on the eigensystem $(\lambda_{\theta j}, \varphi_{\theta j})_{j \in \mathbb{N}}$ of $\Omega_{0\theta}$, where $\lambda_{\theta j}$ and $\varphi_{\theta j}$ denote the eigenvalues and eigenfunctions of $\Omega_{0\theta}$, respectively. Let us consider the null space $\mathcal{N}(\Omega_{0\theta}) \subset \mathcal{E}$ which is generated by 1 and the elements of $h(\theta, \cdot)$. Suppose that this subspace has dimension $r + 1$. We can always construct an orthonormal basis $\{\varphi_{\theta j}\}_{j \geq 0}$ of $\mathcal{E}$ where the $r + 1$ first elements $(\varphi_{\theta 0}, \varphi_{\theta 1}, \ldots, \varphi_{\theta r})$ are the elements that generate $\mathcal{N}(\Omega_{0\theta})$, that is, $\varphi_{\theta 0} = 1$ and $(\varphi_{\theta 1}, \ldots, \varphi_{\theta r})' = h$. Thus, we can construct $\Omega_{0\theta}$ as

$$\Omega_{0\theta}g = \sum_{j=0}^{\infty} \lambda_{\theta j} < g, \varphi_{\theta j} > \varphi_{\theta j}, \qquad g \in \mathcal{E}.$$

If we assume $\lambda_{\theta j} = 0$, $\forall j = 0, 1, \dots, r$, then condition (2.2) is fulfilled since $< \varphi_{\theta j}, \varphi_{\theta j'} >= \delta_{jj'}$, where $\delta_{jj'}$ denotes the Kronecker delta. In order to completely specify $\Omega_{0\theta}$ we have to choose the remaining components $\{\varphi_{\theta j}\}_{j>r}$ such that $\{\varphi_{\theta j}\}_{j\geq 0}$ forms a basis of $\mathcal{E}$ and $\{\lambda_{\theta j}\}_{j>r}$ such that $\sum_{j>r} \lambda_{\theta j} < \infty$. In section 4 we provide some examples that explain in a detailed way the construction of $\Omega_{0\theta}$.

Remark 2.2. In the just-identified case where $r = k$ and θ is a linear transformation of f we may adopt an alternative scheme for constructing the prior on (θ, f). Since the moment restrictions $\mathbf{E}^F(h(\theta, x)) = 0$ rewrite in an explicit form as $\theta = B(f)$, where B is a linear functional, then we may recover the prior of θ through a transformation of the prior for f. In this case we specify a Gaussian process prior μ_f for f with a mean function f_0 restricted to be a *pdf* and a covariance operator Ω_0 restricted to satisfy $\Omega_0^{1/2}1 = 0$. If, for instance, $\theta = \mathbf{E}^f(x)$ then $B(f) =< f, \iota >$ where $\iota \in \mathcal{E}$ denotes the identity function $\iota(x) = x$. The prior for θ recovered from μ_f would be $\mathcal{N}(< f_0, \iota >, < \Omega_0 \iota, \iota >)$.

For clarity reasons, we summarize in the table 1 below the notation used for the prior distributions in the overidentified and in the just-identified case.

Table 1: Prior distribution

Case:	over-identified	just-identified: 1st possibility	just-identified: 2nd possibility
Marginal of θ	$\mu_\theta(\theta)$	$\mu_\theta(\theta)$	$\mu_\theta(\theta)$ through $\theta = B(f)$
Conditional of $f\|\theta$	$\mu_f^\theta(f\|\theta)$	$\mu_f^\theta(f\|\theta)$	–
Marginal of f	–	–	$\mu_f(f)$

2.2 The sampling model

Conditional on f, the sample likelihood is $\prod_{i=1}^n f(x_i)$. While this is the natural choice for the sampling distribution it has the disadvantage to make the posterior distribution of f given θ not available in closed-form. Indeed, a Gaussian prior distribution is usually used in Bayesian modeling with the purpose of making the analysis of the posterior distribution mathematical tractable. For these reasons and in order to exploit the advantage of a conjugate model we propose a different and new way for the construction of the sampling model.

We construct the sampling distribution by considering a functional transformation $\hat{r}$ of the sample $x_1, \dots, x_n$. This transformation $\hat{r}$ is chosen by the researcher such that the following characteristics are satisfied. *I.* $\hat{r}$ converges weakly towards a Gaussian process; *II.*

it is an observable element of an infinite-dimensional Hilbert space, for instance a L^2-space; *III.* it is linked to the nuisance parameter f according to the following linear scheme

$$\hat{r} = Kf + U \tag{2.4}$$

where $K : \mathcal{E} \to \mathcal{F}$ is a linear operator, $\mathcal{F}$ is an infinite-dimensional separable Hilbert space and U is a Hilbert space-valued random variable (H-r.v.). We recall that, for a complete probability space $(Z, \mathcal{Z}, \mathbb{P})$, U is a H-r.v. if it defines a measurable map $U : (Z, \mathcal{Z}, \mathbb{P}) \to (\mathcal{F}, \mathfrak{B}_{\mathcal{F}})$, where $\mathfrak{B}_{\mathcal{F}}$ denotes the Borel σ-fields generated by the open sets of $\mathcal{F}$.

More precisely, let $T \subset \mathbb{R}^p$, we first select a function $k(t,x) : T \times S \to \mathbb{R}_+$ that is a measurable function of one observation $\forall t \in T$. We then represent the data through the expectation of $k(t,\cdot)$ under the empirical measure:

$$\hat{r} = \frac{1}{n}\sum_{i=1}^{n} k(t, x_i).$$

Thus, by denoting with $Kf := \int k(t,x)f(x)\Pi(dx)$ the expectation of $k(t,\cdot)$ under F, model (2.4) rewrites:

$$\hat{r} = \frac{1}{n}\sum_{i=1}^{n} k(t, x_i) = \int k(t,x)f(x)\Pi(dx) + U(t). \tag{2.5}$$

Moreover, the function k must be such that $r := Kf$ and $\hat{r}$ are elements of $\mathcal{F} = L^2(T, \mathfrak{B}_T, \rho)$ with ρ a measure on T. Here $\mathfrak{B}_T$ denotes the Borel σ-field generated by the open sets of T. Conditionally on f, the expectation of $\hat{r}$ is equal to Kf and the error term U has zero mean and covariance kernel

$$\sigma^F(t,s) = \mathbf{E}^F U(t)U(s) = \frac{1}{n}\left[\mathbf{E}^F\left(k(t,x)k(s,x)\right) - \mathbf{E}^F(k(t,x))\mathbf{E}^F(k(s,x))\right].$$

We denote by $P_{n,*}^{f_*}$ the true distribution function of $\hat{r}$ satisfying $\hat{r} = Kf_* + U_*$ where U_* is an $H - r.v.$ with zero mean and covariance kernel $\sigma^{F_*}(t,s)$ by construction. Similarly, we denote by $P_{n,*}^{f}$ the conditional distribution of $\hat{r}$ given f satisfying $\hat{r} = Kf + U$ and based on the true $P_{n,*}^{f_*}$. In general, $P_{n,*}^{f}$ is either unknown or not suitable in order to construct the posterior distribution. For this reason we consider as the sampling distribution an approximation of $P_{n,*}^{f}$ that we denote by P_n^f and that is the weak limit of $P_{n,*}^{f}$ as $n \to \infty$. Therefore, the sampling model that we consider is *misspecified* in finite samples. In practice, it is sufficient to choose $k(t,\cdot)$ to be Donsker so that the weak limit of $P_{n,*}^{f}$ is a Gaussian distribution with mean Kf and covariance kernel $\frac{1}{n}\left[\mathbf{E}^F\left(k(t,x)k(s,x)\right) - \mathbf{E}^F(k(t,x))\mathbf{E}^F(k(s,x))\right]$. Therefore, the sampling distribution P_n^f that we use in the following is

$$
\begin{aligned}
P_n^f &= \mathcal{N}(Kf, \Sigma_n), \qquad \Sigma_n = \frac{1}{n}\Sigma : \mathcal{F} \to \mathcal{F} \qquad (2.6)\\
\Sigma\varphi &= \int \left[\mathbf{E}^F\left(k(t,x)k(s,x)\right) - \mathbf{E}^F(k(t,x))\mathbf{E}^F(k(s,x))\right]\varphi(s)ds, \quad \varphi \in \mathcal{F}.
\end{aligned}
$$

Due to the Gaussianity of the prior μ_f^θ of f, a Gaussian distribution is a convenient choice for P_n^f. Under P_n^f, U is a zero-mean Gaussian H-r.v. with covariance operator Σ_n which is one-to-one, linear, positive definite, self-adjoint and trace-class. In several examples the covariance operator Σ_n is unknown and therefore estimated. We estimate it in a frequentist way by replacing F with the empirical *cdf*. We have shown in Florens and Simoni (2012a) that this does not affect any asymptotic properties of our procedure. We clarify our construction of the sampling model (2.4) in the next example.

Example 2.1. Let us suppose that we dispose of an *i.i.d.* sample of x: $(x_1, \ldots, x_n)$, where $x_i \in \mathbb{R}$, $i = 1, \ldots, n$. By using this sample we can construct a functional transformation $\hat{r}$. For example, $\hat{r}$ may be the empirical cumulative distribution function (*cdf*) $\hat{F}(t) = \frac{1}{n}\sum_{i=1}^n 1\{x_i \leq t\}$ or the empirical characteristic function $\hat{\Phi}(t) = \frac{1}{n}\sum_{i=1}^n e^{itx_i}$ for $t \in \mathbb{R}$. In these two cases we can write:

$$
\begin{aligned}
\hat{F}(t) &= \int 1\{s \leq t\} f(s)\Pi(ds) + U(t),\\
\widehat{\Phi}(t) &= \int e^{its} f(s)\Pi(ds) + U(t),
\end{aligned}
$$

respectively. In the first case $\hat{r} = \hat{F}$ and $\forall \varphi \in \mathcal{E}$, $K\varphi = \int 1\{s \leq t\}\varphi(s)\Pi(ds) = F(t)$, while $\hat{r} = \widehat{\Phi}$ and $\forall \varphi \in \mathcal{E}$, $K\varphi = \int e^{its}\varphi(s)\Pi(ds) = \Phi(t)$ in the second case. In these two cases, by the Donsker's theorem, U is asymptotically Gaussian with zero mean and covariance operator characterized by the kernel $\frac{1}{n}(F(s \wedge t) - F(s)F(t))$ in the first case and $\frac{1}{n}(\Phi(s+t) - \Phi(s)\Phi(t))$ in the second case. These variances are clearly unknown when f is unknown but we can estimate them consistently by replacing F and Φ by $\hat{F}$ and $\hat{\Phi}$, respectively.

The following lemma gives an useful characterization of the operator Σ_n in terms of K and its adjoint K^*. We recall that the adjoint K^* is such that $< K\varphi, \psi >=< \varphi, K^*\psi >$, $\forall \varphi \in \mathcal{E}$ and $\psi \in \mathcal{F}$. In our case $K\varphi = \int_S k(t,x)\varphi(x)\Pi(dx)$ and $\mathcal{F} = L^2(T, \mathfrak{B}_T, \rho)$, then an elementary computation shows that $K^*\psi = \int_T k(t,x)\psi(t)\rho(dt)$.

Lemma 2.2. *Let $K : \mathcal{E} \to \mathcal{F}$ be the operator: $\forall \varphi \in \mathcal{E}$, $K\varphi = \int_S k(t,x)\varphi(x)\Pi(dx)$ and $K^* : \mathcal{F} \to \mathcal{E}$ be its adjoint, that is, $\forall \psi \in \mathcal{F}$, $K^*\psi = \int_T k(t,x)\psi(t)\rho(dt)$. Moreover, denote with f_* the true value of f that characterizes the DGP. Thus, the operator $\Sigma_n = \frac{1}{n}\Sigma$ takes the form*

$$\forall \psi \in \mathcal{F}, \quad \Sigma\psi = KM_f K^*\psi - (KM_f 1) < M_f, K^*\psi > \tag{2.7}$$

where $\Sigma : \mathcal{F} \to \mathcal{F}$ *and* $M_f : \mathcal{E} \to \mathcal{E}$ *is the multiplication operator* $\forall \varphi \in \mathcal{E}$, $M_f\varphi = f_*(x)\varphi(x)$.

Proof. The result follows trivially from the definition of the covariance operator $\Sigma_n : \mathcal{F} \to \mathcal{F}$: $\forall \psi \in \mathcal{F}$,

$$\begin{aligned}
\Sigma_n\psi &= \frac{1}{n}\left[\int_T \int_S (k(t,x)k(s,x))\, f_*(x)\Pi(dx)\psi(t)\rho(dt) - \int_T\int_S k(t,x)f_*(x)\Pi(dx)\left(\int_S k(s,x)f_*(x)\Pi(dx)\right)\psi(t)\rho(dt)\right]\\
&= \frac{1}{n}\left[\int_S k(s,x)f_*(x)\int_T k(t,x)\psi(t)\rho(dt)\Pi(dx) - \int_S k(s,x)f_*(x)\Pi(dx)\left(\int_S\int_T k(t,x)\psi(t)\rho(dt)f_*(x)\Pi(dx)\right)\right]\\
&= \frac{1}{n}\left[KM_fK^*\psi - (KM_f1) < M_f, K^*\psi >\right]
\end{aligned}$$

where the second equality has been obtained by using the Fubini's theorem.

□

The following lemma states the relationship between the range of K and the range of $\Sigma^{\frac{1}{2}}$. We denote by $\mathfrak{D}$ the subset of $\mathcal{E}$ whose elements integrate to 0 with respect to Π:

$$\mathfrak{D} := \left\{ g \in \mathcal{E};\ \int g(x)\Pi(dx) = 0 \right\}.$$

We remark that $\mathfrak{D}$ contains the subset of functions in $\mathcal{E}$ that are the difference of *pdf* of F with respect to Π. Moreover, $\mathcal{R}(\Omega_{0\theta}^{\frac{1}{2}}) \subset \mathfrak{D}$ where $\mathcal{R}(\Omega_{0\theta}^{\frac{1}{2}}) \equiv \mathcal{H}(\Omega_{0\theta})$ has been defined in (2.3).

Lemma 2.3. *Let* $K : \mathcal{E} \to \mathcal{F}$ *be the operator:* $\forall \varphi \in \mathcal{E}$, $K\varphi = \int_S k(t,x)\varphi(x)\Pi(dx)$ *and denote by* $K|_{\mathfrak{D}}$ *the operator* K *restricted to* $\mathfrak{D} \subset \mathcal{E}$. *Then, if* $K|_{\mathfrak{D}}$ *is injective we have*

$$\mathcal{R}(K|_{\mathfrak{D}}) = \mathcal{D}(\Sigma^{-\frac{1}{2}}).$$

Proof. We can rewrite Σ as

$$\begin{aligned}
\forall \psi \in \mathcal{F}, \quad \Sigma\psi &= \int_T \mathbf{E}\left(v(x,t)v(x,s)\right)\psi(t)\rho(dt)\\
&= \int_T\int_S \left(v(x,t)v(x,s)\right) f_*(x)\Pi(dx)\psi(t)\rho(dt)
\end{aligned}$$

where $v(x,t) = [k(x,t) - \mathbf{E}(k(x,t))]$. Then, $\forall \psi \in \mathcal{F}$ we can write $\Sigma\psi = RM_fR^*\psi$ where $R : \mathcal{E} \to \mathcal{F}$, $M_f : \mathcal{E} \to \mathcal{E}$ and $R^* : \mathcal{F} \to \mathcal{E}$ are the operators defined as

$$\begin{aligned}
\forall \psi \in \mathcal{F}, \quad R^*\psi &= \int_T v(x,t)\psi(t)\rho(dt)\\
\forall \varphi \in \mathcal{E}, \quad M_f\varphi &= f_*(x)\varphi(x)\\
\forall \varphi \in \mathcal{E}, \quad R\varphi &= \int_S v(x,t)\varphi(x)\Pi(dx).
\end{aligned}$$

Moreover, we have $\mathcal{D}(\Sigma^{-\frac{1}{2}}) = \mathcal{R}(\Sigma^{\frac{1}{2}}) = \mathcal{R}((RM_f R^*)^{\frac{1}{2}}) = \mathcal{R}(RM_f^{1/2})$.

Let $h \in \mathcal{R}(K)$, that is, there exists a $g \in \mathcal{E}$ such that $h(t) = \int_S k(t,x)g(x)\Pi(dx)$. Then $h \in \mathcal{D}(\Sigma^{-\frac{1}{2}})$ if there exists an element $\nu \in \mathcal{E}$ such that $h(t) = \int_S v(x,t) f_*^{\frac{1}{2}}(x)\nu(x)\Pi(dx)$. By developing this equality, the element ν has to satisfy

$$\begin{aligned} & \int_S k(t,x)g(x)\Pi(dx) = \int_S v(x,t) f_*^{\frac{1}{2}}(x)\nu(x)\Pi(dx) \\ \Leftrightarrow & \int_S k(t,x)g(x)\Pi(dx) = \int_S \left[k(x,t) - \left(\int_S k(x,t) f_*(x)\Pi(dx)\right)\right] f_*^{\frac{1}{2}}(x)\nu(x)\Pi(dx) \\ \Leftrightarrow & \int_S k(t,x)g(x)\Pi(dx) = \int_S k(x,t)\left[f_*^{\frac{1}{2}}(x)\nu(x) - f_*(x)\left(\int_S f_*^{\frac{1}{2}}(x)\nu(x)\Pi(dx)\right)\right]\Pi(dx). \end{aligned}$$

If K is injective it follows that such an element ν must satisfy

$$g(x) = f_*^{\frac{1}{2}}\nu(x) - f_*(x)\left(\int_S f_*^{\frac{1}{2}}(x)\nu(x)\Pi(dx)\right)$$

which in turn implies that $\int_S g(x)\Pi(dx) = 0$, *i.e.* that $h \in \mathcal{R}(K|_{\mathfrak{D}})$. Therefore, one solution is $\nu(x) = f_*^{-\frac{1}{2}} g(x)$ which proves that the range of the truncated operator $K|_{\mathfrak{D}}$ in contained in $\mathcal{D}(\Sigma^{-\frac{1}{2}})$. On the other side, let $h \in \mathcal{D}(\Sigma^{-\frac{1}{2}})$, then there exists a $\nu \in \mathcal{E}$ such that $h = \int_S v(x,t) f_*^{\frac{1}{2}}(x)\nu(x)\Pi(dx)$. By the previous argument and under the assumption that $K|_{\mathfrak{D}}$ is injective, this implies that $h \in \mathcal{R}(K|_{\mathfrak{D}})$ since there exists $g \in \mathfrak{D}$ such that $g(x) = f_*^{\frac{1}{2}}\nu(x) - f_*(x)\left(\int_S f_*^{\frac{1}{2}}(x)\nu(x)\Pi(dx)\right)$. This shows the inclusion of $\mathcal{D}(\Sigma^{-\frac{1}{2}})$ in $\mathcal{R}(K|_{\mathfrak{D}})$ and concludes the proof.

□

2.3 Posterior distribution

The posterior distribution is constructed by using the approximated (or misspecified) sampling distribution P_n^f. The Bayesian model can be summarized in the following way:

$$\begin{cases} \theta & \sim \quad \mu_\theta \\ f|\theta & \sim \quad \mu_f^\theta \sim \mathcal{N}(f_{0\theta}, \Omega_{0\theta}), \qquad \int h(\theta,x) f_{0\theta}(x)\Pi(dx) = 0 \quad \text{and} \quad \Omega_{0\theta}^{\frac{1}{2}}(1, h(\theta,\cdot)')' = 0 \\ \hat{r}|f,\theta & \sim \quad \hat{r}|f \sim P_n^f \sim \mathcal{N}(Kf, \Sigma_n) \end{cases}$$

which defines a joint distribution on $\Lambda \times \mathcal{F}$. This joint probability distribution may be examined under different aspects. First, let us consider the joint conditional distribution of $(f, \hat{r})$ conditional on θ. Following Theorem 1 in Florens and Simoni (2012a) we can show that

$$\left(\begin{array}{c} f \\ \hat{r} \end{array}\right)\Bigg|\theta \sim \mathcal{N}\left(\left(\begin{array}{c} f_{0\theta} \\ Kf_{0\theta} \end{array}\right), \left(\begin{array}{cc} \Omega_{0\theta} & \Omega_{0\theta}K^* \\ K\Omega_{0\theta} & \Sigma_n + K\Omega_{0\theta}K^* \end{array}\right)\right) \tag{2.8}$$

where the operator $(\Sigma_n + K\Omega_{0\theta}K^*)$ is an operator from $\mathcal{F}$ to $\mathcal{F}$, while $\Omega_{0\theta}K^* : \mathcal{F} \to \mathcal{E}$ and $K\Omega_{0\theta} : \mathcal{E} \to \mathcal{F}$.

From (2.8) we deduce the sampling distribution of $\hat{r}$ conditional on θ by integrating out f:

$$\hat{r}|\theta \sim \mathcal{N}(Kf_{0\theta}, \Sigma_n + K\Omega_{0\theta}K^*). \tag{2.9}$$

We denote by P_n^θ this distribution. The marginal posterior for $\theta \in \Theta$ depends on the nuisance parameter f only through the integrated sampling distribution P_n^θ.

2.3.1 Conditional posterior distribution of f, given θ

The conditional distribution of f given $(\hat{r}, \theta)$, that is, the posterior distribution of f, is a Gaussian distribution. This has been proven for instance in Florens and Simoni (2012a). This distribution is fully characterized by its mean and variance and, in general, the computation of these moments rises problems of regularization when the dimension of the problem is infinite. While this point has been broadly discussed in (Florens and Simoni , 2012a,b) and references therein, in this section we analyze it in the particular case considered in the paper where the operators take a specific form.
We recall briefly the problem encountered in the computation of the moments of the Gaussian posterior distribution of f given θ is the following. It is well known that in finite dimensional problems the conditional moments of joint Gaussian distributions require the inversion of the covariance matrix of the conditioning variable. So that in our case we should inverse $(\Sigma_n + K\Omega_{0\theta}K^*)$ in order to construct the posterior mean and covariance of f given $(\hat{r}, \theta)$. The problem arises because the inverse operator $(\Sigma_n + K\Omega_{0\theta}K^*)^{-1}$ is in general defined only on a subset of $\mathcal{F}$ of $P_{n,*}^\theta$-measure 0. Therefore, in general there is no closed-form available for the posterior mean and variance of $\mu_f^{\hat{r},\theta}$.

However, in the framework under consideration we determine mild conditions that allows to solve this problem so that the inversion of $(\Sigma_n + K\Omega_{0\theta}K^*)$, necessary for constructing the posterior mean and variance of f, does not rise any continuity problem. Now, we are going to illustrate these conditions in the lemmas below.

Lemma 2.4. *Consider the Gaussian distribution (2.8) on $\mathfrak{B}_\mathcal{E} \times \mathfrak{B}_\mathcal{F}$ and assume that $f_*^{-1/2} \in \mathcal{R}(K^*)$. Then, the conditional distribution on $\mathfrak{B}_\mathcal{E}$ conditional on $\mathfrak{B}_\mathcal{F} \times \mathfrak{B}$, denoted by $\mu_f^{\hat{r},\theta}$, exists, is regular and almost surely unique. It is Gaussian with mean*

$$\mathbf{E}[f|\hat{r}] = f_{0\theta} + A(\hat{r} - Kf_{0\theta}) \tag{2.10}$$

and trace class covariance operator

$$Var[f|\hat{r}] = \Omega_{0\theta} - AK\Omega_{0\theta} : \mathcal{E} \to \mathcal{E} \tag{2.11}$$

where

$$A := \Omega_{0\theta} M_f^{-1/2} \left(\frac{1}{n} I - \frac{1}{n} M_f^{1/2} < M_f^{1/2}, \cdot > + M_f^{-1/2} \Omega_{0\theta} M_f^{-1/2} \right)^{-1} ((K^*)^{-1} M_f^{-1/2})^*$$

is a continuous and linear operator from $\mathcal{F}$ to $\mathcal{E}$.

Proof. The first part of the theorem follows from theorem 1 (ii) in Florens and Simoni (2012b). From this result, since $\Sigma_n = \frac{1}{n}\Sigma$, where $\Sigma : \mathcal{F} \to \mathcal{F}$ is defined in lemma 2.2, we know that $\mathbf{E}[f|\hat{r}] = f_{0\theta} + \Omega_{0\theta} K^* (\frac{1}{n}\Sigma + K\Omega_{0\theta}K^*)^{-1}(\hat{r} - Kf_{0\theta})$ and $Var[f|\hat{r}] = \Omega_{0\theta} - \Omega_{0\theta}K^*(\frac{1}{n}\Sigma + K\Omega_{0\theta}K^*)^{-1}K\Omega_{0\theta}$. Hence, we have to show that $\Omega_{0\theta}K^*(\frac{1}{n}\Sigma + K\Omega_{0\theta}K^*)^{-1} = A$ and that A is continuous and linear. Denote $\tilde{M} = \left(\frac{1}{n} I - \frac{1}{n} M_f^{\frac{1}{2}} < M_f^{\frac{1}{2}}, \cdot > + M_f^{-\frac{1}{2}} \Omega_{0\theta} M_f^{-\frac{1}{2}} \right)^{-1}$ and

$$\breve{M} = \left(\frac{1}{n} K M_f K^* - \frac{1}{n} (K M_f 1) < M_f, K^* \cdot > + K\Omega_{0\theta} K^* \right)^{-1}.$$

By using the result of lemma 2.2, we can rewrite the operator $\Omega_{0\theta}K^*(\frac{1}{n}\Sigma + K\Omega_{0\theta}K^*)^{-1}$ as

$$\Omega_{0\theta} M_f^{-\frac{1}{2}} \tilde{M} ((K^*)^{-1} M_f^{-\frac{1}{2}})^* + \Omega_{0\theta} \left[K^* \breve{M} - M_f^{-\frac{1}{2}} \tilde{M} ((K^*)^{-1} M_f^{-\frac{1}{2}})^* \right].$$

This is equal to $\Omega_{0\theta} M_f^{-\frac{1}{2}} \tilde{M} ((K^*)^{-1} M_f^{-\frac{1}{2}})^*$ since $\left[K^* \breve{M} - M_f^{-\frac{1}{2}} \tilde{M} ((K^*)^{-1} M_f^{-\frac{1}{2}})^* \right]$ is equal to

$$\begin{aligned} & \left[K^* - M_f^{-\frac{1}{2}} \tilde{M} ((K^*)^{-1} M_f^{-\frac{1}{2}})^* \breve{M}^{-1} \right] \breve{M} \\ = & M_f^{-\frac{1}{2}} \tilde{M} \left[\left(\frac{1}{n} M_f^{\frac{1}{2}} - \frac{1}{n} M_f^{\frac{1}{2}} < M_f, \cdot > + M_f^{-\frac{1}{2}} \Omega_{0\theta} \right) K^* - ((K^*)^{-1} M_f^{-\frac{1}{2}})^* \breve{M}^{-1} \right] \breve{M} \end{aligned}$$

which is zero.

We now show that the operator A is continuous and linear on $\mathcal{F}$. First, remark that the assumption $f_*^{-\frac{1}{2}} \in \mathcal{R}(K^*)$ ensures that $(K^*)^{-1} M_f^{-\frac{1}{2}}$ exists and is bounded. Since $\Omega_{0\theta}$ is the covariance operator of a Gaussian measure on a Hilbert space then, it is trace class. This means that $\Omega_{0\theta}^{\frac{1}{2}}$ is Hilbert-Schmidt, which is a compact operator. Therefore, since the product of two bounded and compact operators is compact, it follows that $\Omega_{0\theta}$, $\Omega_{0\theta} M_f^{-\frac{1}{2}}$ and $M_f^{-\frac{1}{2}} \Omega_{0\theta} M_f^{-\frac{1}{2}}$ are compact.

It is also easy to show that the operator $\frac{1}{n} M_f^{\frac{1}{2}} < M_f^{\frac{1}{2}}, \cdot >: \mathcal{E} \to \mathcal{E}$ is compact since its Hilbert-Schmidt norm is equal to 1. In particular this operator has rank equal to 1 since it has only one eignevalue different from 0 and which is equal to 1. This eigenvalue corresponds to the eigenfunction $f_*^{\frac{1}{2}}$. Therefore, the operator $(\frac{1}{n} M_f^{\frac{1}{2}} < M_f^{\frac{1}{2}}, \cdot > - M_f^{-\frac{1}{2}} \Omega_{0\theta} M_f^{-\frac{1}{2}})$ is

compact.
By the Cauchy-Schwartz inequality we have

$$\begin{aligned}\forall\phi\in\mathcal{E},\quad <\tilde{M}^{-1}\phi,\phi> &= \frac{1}{n}||\phi||^2-\frac{1}{n}<f_*^{\frac{1}{2}},\phi>^2+<\Omega_{0\theta}^{\frac{1}{2}}f_*^{-\frac{1}{2}}\phi,\Omega_{0\theta}^{\frac{1}{2}}f_*^{-\frac{1}{2}}\phi>\\ &\geq \frac{1}{n}||\phi||^2-\frac{1}{n}||f_*^{\frac{1}{2}}||^2||\phi||^2+||\Omega_{0\theta}^{\frac{1}{2}}f_*^{-\frac{1}{2}}\phi||^2\\ &\geq ||\Omega_{0\theta}^{\frac{1}{2}}f_*^{-\frac{1}{2}}\phi||^2\geq 0\end{aligned}$$

since $||f_*^{\frac{1}{2}}||^2=1$. Therefore, we conclude that $\tilde{M}$ is injective. Then, from the Riesz Theorem 3.4 in Kress (1999) it follows that the operator $\tilde{M}:\mathcal{E}\to\mathcal{E}$ is bounded.

Finally, the operator A is bounded and linear since it is the product of bounded linear operators. We conclude that A is a continuous operator from $\mathcal{F}$ to $\mathcal{E}$.

□

Remark 2.3. If $f_*^{-1}\in\mathcal{R}(K^*)$ then the operator $A:\mathcal{F}\to\mathcal{E}$ of the theorem may be written in an equivalent way as: $\forall\varphi\in\mathcal{F}$

$$A\varphi=\Omega_{0\theta}\left(\frac{1}{n}I+\frac{1}{n}<f_*,\cdot>+f_*^{-1}\Omega_{0\theta}\right)^{-1}((K^*)^{-1}f_*^{-1})^*. \tag{2.12}$$

Remark 2.4. If f_* is assumed to be bounded away from 0 and ∞ on its support, then the condition $f_*^{-1}\in\mathcal{R}(K^*)$, as well as the condition $f_*^{-1/2}\in\mathcal{R}(K^*)$, can not be satisfied if $k(t,x)$ is such that $\forall\psi\in\mathcal{F}$, $K^*\psi=\int_T k(t,x)\psi(t)\rho(dt)$ vanishes at some x in the support of f_*. This excludes the kernel $k(t,x)=1\{x\leq t\}$ when T is equal to a compact set, say $T=[a,b]$. This remark suggests that some care must be taken by the researcher when he/she chooses the operator K according to its prior information about f_*.

The next lemma provides a condition alternative to the one given in lemma 2.4 which also guarantees continuity of the inverse of $(\Sigma_n+K\Omega_{0\theta}K^*)$.

Lemma 2.5. *Consider the Gaussian distribution (2.8) on $\mathfrak{B}_{\mathcal{E}}\times\mathfrak{B}_{\mathcal{F}}$ and assume that $K|_{\mathcal{H}(\Omega_{0\theta})}$ is injective and that $\Omega_{0\theta}$ is such that $\mathcal{R}(K\Omega_{0\theta}^{\frac{1}{2}})\subseteq\mathcal{R}(\Sigma)$. Then, the result of lemma 2.4 holds with A equal to*

$$A:=\Omega_{0\theta}^{1/2}\left(\frac{1}{n}I+\Omega_{0\theta}^{1/2}K^*\Sigma^{-1}K\Omega_{0\theta}^{1/2}\right)^{-1}(\Sigma^{-1}K\Omega_{0\theta}^{1/2})^*.$$

Proof. Since $K\Omega_{0\theta}^{\frac{1}{2}}=K|_{\mathcal{H}(\Omega_{0\theta})}\Omega_{0\theta}^{\frac{1}{2}}$ and $K|_{\mathcal{H}(\Omega_{0\theta})}$ is injective by assumption then $\Sigma^{-\frac{1}{2}}K|_{\mathcal{H}(\Omega_{0\theta})}$ is well defined by lemma 2.3. By applying theorem 1 *(iii)* in Florens and Simoni (2012b) we conclude.

□

The trajectories of f generated by the conditional posterior distribution $\mu_f^{\hat{r},\theta}$ verify almost surely the moment conditions and integrate to 1. This can be proved by an argument similar to the one used to prove Lemma 2.1. First, remark that the posterior covariance operator satisfies the moment restrictions:

$$[\Omega_{0\theta} - AK\Omega_{0\theta}]^{1/2}(1, h'(\theta, \cdot))' = [I - AK]^{1/2}\Omega_{0\theta}^{1/2}(1, h'(\theta, \cdot))' = 0$$

where we have factorized $\Omega_{0\theta}^{\frac{1}{2}}$ on the left and used assumption (2.2). Moreover, a trajectory f drawn from the posterior $\mu_f^{\theta,\hat{r}}$ is such that $(f - f_{0\theta}) \in \overline{\mathcal{H}(\Omega_{0\theta} - AK\Omega_{0\theta})}$, $\mu_f^{\theta,\hat{r}}$-a.s. Now, for any $\varphi \in \mathcal{H}(\Omega_{0\theta} - AK\Omega_{0\theta})$ we have $< \varphi, h(\theta, \cdot) >=< [\Omega_{0\theta}^{\frac{1}{2}} - AK\Omega_{0\theta}]\psi, \Omega_{0\theta}^{\frac{1}{2}} h(\theta, \cdot) >= 0$, for some $\psi \in \mathcal{E}$, and $< \varphi, 1 >= 0$ by a similar argument. This shows that

$$\mathcal{H}(\Omega_{0\theta} - AK\Omega_{0\theta}) \subset \left\{\varphi \in \mathcal{E}\ ;\ \int \varphi(x)h(\theta, x)\Pi(dx) = 0 \text{ and } \int \varphi(x)\Pi(dx) = 0\right\}$$

and since the set on the right of this inclusion is closed we have

$$\overline{\mathcal{H}(\Omega_{0\theta} - AK\Omega_{0\theta})} \subset \left\{\varphi \in \mathcal{E}\ ;\ \int \varphi(x)h(\theta, x)\Pi(dx) = 0 \text{ and } \int \varphi(x)\Pi(dx) = 0\right\}.$$

Therefore, $\mu_f^{\theta,\hat{r}}$-a.s. a trajectory f drawn from $\mu_f^{\theta,\hat{r}}$ is such $\int(f - f_{0\theta})(x)\Pi(dx) = 0$ and $\int(f - f_{0\theta})(x)h(\theta, x)\Pi(dx) = 0$ which implies: $\int f(x)\Pi(dx) = 1$ and $\int f(x)h(\theta, x)\Pi(dx) = 0$.

Remark 2.5. The posterior distribution of f conditional on θ gives the revision of the prior on f except in the direction of the constant and of the moment conditions that remain unchanged. A possible strategy would be to estimate also θ by maximum likelihood by using the density given by $\mathbf{E}(f|\hat{r}, \theta)$ as the probability density of the data. We could also take an Empirical Bayes approach which consists in obtaining the posterior on θ by starting from the marginal likelihood. We do not develop this approach but we use a completely Bayesian approach by trying to recover a conditional distribution of θ conditional on $\hat{r}$.

Remark 2.6. When neither the conditions of lemma 2.4 nor the conditions of lemma 2.5 are satisfied then we can not use the exact posterior distribution $\mu_f^{\theta,\hat{r}}$. Instead, we use the *regularized posterior distribution* denoted by $\mu_{f,\tau}^{\theta,\hat{r}}$, where $\tau > 0$ is a regularization parameter that must be suitable chosen and that converges to 0 with n. This distribution has been proposed by Florens and Simoni (2012a) and we refer to this paper for a complete description of it. Here, we only give its expression: $\mu_{f,\tau}^{\theta,\hat{r}}$ is a Gaussian distribution with mean function

$$\mathbf{E}[f|\hat{r}, \tau] = f_{0\theta} + A_\tau(\hat{r} - Kf_{0\theta}) \tag{2.13}$$

and covariance operator

$$Var[f|\hat{r},\tau] = \Omega_{0\theta} - A_\tau K\Omega_{0\theta} : \mathcal{E} \to \mathcal{E} \tag{2.14}$$

where

$$A_\tau := \Omega_{0\theta}K^*\left(\tau I + \frac{1}{n}I + K\Omega_{0\theta}K^*\right)^{-1} : \mathcal{E} \to \mathcal{E}. \tag{2.15}$$

2.3.2 Posterior distribution of θ

We have stressed that the marginal posterior for θ, denoted by $\mu_\theta^{\hat{r}}$, can be obtained by using the marginal sampling distribution P_n^θ given in (2.9). In order to obtain a closed-form expression for the marginal posterior $\mu_\theta^{\hat{r}}$ or at least to simulate through an MCMC procedure it is suitable to find a dominating measure, say P_n^0, for P_n^θ and to characterize the likelihood of P_n^θ with respect to P_n^0. The following theorem, which is a slight modification of Theorem 3.4 in Kuo (1975, page 125), characterizes a probability measure P_n^0 which is equivalent to P_n^θ and the corresponding likelihood of P_n^θ with respect to P_n^0.

Theorem 2.1. *Let $\tilde{f} \in \mathcal{E}$ denote a probability density function (with respect to Π)and P_n^0 be a Gaussian measure with mean $K\tilde{f}$ and covariance operator $n^{-1}\Sigma$, i.e. $P_n^0 = \mathcal{N}(K\tilde{f}, n^{-1}\Sigma)$. If $K|_{\mathfrak{D}}$ is injective then P_n^0 and P_n^θ are equivalent. Moreover, assume that one of the following conditions is satisfied*

(i) $\mathcal{R}(K\Omega_{0\theta}^{\frac{1}{2}}) \subset \mathcal{D}(\Sigma^{-1})$;

(ii) the operators Σ and $\Sigma^{-1/2}K\Omega_{0\theta}K^*\Sigma^{-1/2}$ have the same eigenfunctions.

Then the Radon-Nikodym derivative is given by

$$\frac{dP_n^\theta}{dP_n^0} = \prod_{j=1}^\infty \frac{1}{\sqrt{nl_j^2+1}} e^{\frac{1}{2(l_j^2+n^{-1})}\left(nl_j^2z_j^2 - A_j^2 + 2z_jA_j\right)}, \tag{2.16}$$

with $z_j = <\hat{r} - K\tilde{f}, \Sigma^{-1/2}\varphi_j>$, l_j^2 and φ_j the eigenvalues and eigenfunctions of $\Sigma^{-1/2}K\Omega_{0\theta}K^*\Sigma^{-1/2}$ and A_j the expectation of z_j under P_n^θ.

The random variable $\sqrt{n}z_j$ has a standard Gaussian distribution under P_n^0. If condition *(i)* holds then z_j is well defined since $l_j^2\varphi_j = \Sigma^{-1/2}K\Omega_{0\theta}K^*\Sigma^{-1/2}\varphi_j$ and $\Sigma^{-1/2}\varphi_j = l_j^{-1}\Sigma^{-1}K\Omega_{0\theta}K^*\Sigma^{-1/2}\varphi_j$ which is well-defined under the assumption $\mathcal{R}(K\Omega_{0\theta}^{\frac{1}{2}}) \subset \mathcal{D}(\Sigma^{-1})$. If condition *(ii)* holds then z_j is well defined since φ_j is an eigenfunction of Σ as well as of $\Sigma^{-1/2}$ so that $\forall j \in \mathbb{N}$, there exists $\lambda_{j\Sigma}$ such that $\Sigma^{-1/2}\varphi_j = \lambda_{j\Sigma}^{-1/2}\varphi_j$ and, in this case, $z_j = \frac{<\hat{r}-K\tilde{f},\varphi_j>}{\sqrt{\lambda_{j\Sigma}}}$. We also remark that we can use any density function for the mean function $\tilde{f}$ as long as it does not depend on θ. For instance, it could be $\tilde{f} = f_*$ even if it is unknown in practice.

Proof of Theorem 2.1 In this proof we denote $B = \Sigma^{-1/2}K\Omega_{0\theta}^{1/2}$. To prove that P_n^θ and P_n^0 are equivalent we first rewrite the covariance operator of P_n^θ as

$$\left(n^{-1}\Sigma + K\Omega_{0\theta}K^*\right) = \sqrt{n^{-1}}\Sigma^{\frac{1}{2}}\left[I + n\Sigma^{-\frac{1}{2}}K\Omega_{0\theta}K^*\Sigma^{-\frac{1}{2}}\right]\Sigma^{\frac{1}{2}}\sqrt{n^{-1}}.$$

Then according to theorem 3.3 p.125 in Kuo (1975) we have to verify that $K(\tilde{f} - f_{0\theta}) \in \mathcal{R}(\Sigma^{1/2})$ and that $\left[I + n\Sigma^{-\frac{1}{2}}K\Omega_{0\theta}K^*\Sigma^{-\frac{1}{2}}\right]$ is positive definite, bounded, invertible with $n\Sigma^{-\frac{1}{2}}K\Omega_{0\theta}K^*\Sigma^{-\frac{1}{2}}$ Hilbert Schmidt.

- Since $(\tilde{f} - f_{0\theta}) \in \mathfrak{D}$ and since $K|_{\mathfrak{D}}$ is injective then, by lemma 2.3, $K(\tilde{f} - f_{0\theta}) \in \mathcal{R}(\Sigma^{1/2})$.
- *Positive definiteness.* It is trivial to show that the operator $(I + nBB^*)$ is self-adjoint, *i.e.* $(I + nBB^*)^* = (I + nBB^*)$. Moreover, $\forall\varphi \in \mathcal{F}$, $\varphi \neq 0$
$$< (I + nBB^*)\varphi, \varphi >=< \varphi, \varphi > +n < B^*\varphi, B^*\varphi >= ||\varphi||^2 + n||B^*\varphi|| > 0.$$
- *Boundedness.* By lemma 2.3, if $K|_{\mathfrak{D}}$ is injective, the operators B and B^* are bounded ; the operator I is bounded by definition and a linear combination of bounded operators is bounded, see Remark 2.7 in Kress (1999).
- *Continuously invertible.* The operator $(I + nBB^*)$ is continuously invertible if its inverse is bounded, *i.e.* there exists a positive number C such that $||(I+nBB^*)^{-1}\varphi|| \leq C||\varphi||$, $\forall\varphi \in \mathcal{F}$. We have $||(I + nBB^*)^{-1}\varphi|| \leq (\sup_j \frac{n^{-1}}{n^{-1}+l_j^2})||\varphi|| = ||\varphi||$, $\forall\varphi \in \mathcal{F}$.
- *Hilbert-Schmidt.* We consider the Hilbert-Schmidt norm $||nBB^*||_{HS} = \frac{1}{\alpha}\sqrt{tr((BB^*)^2)}$. Now, $tr((BB^*)^2) = tr(\Omega_0\tilde{B}^*\tilde{B}\Omega_{0\theta}\tilde{B}^*\tilde{B}) \leq tr(\Omega_{0\theta})||\tilde{B}^*\tilde{B}\Omega_{0\theta}\tilde{B}^*\tilde{B}|| < \infty$ since $\tilde{B} := \Sigma^{-\frac{1}{2}}K|_{\mathcal{H}(\Omega_{0\theta})}$ has a bounded norm by lemma 2.3.

This shows that P_n^θ and P_n^0 are equivalent.

Next we derive (2.16). Let $z_j =< \hat{r} - K\tilde{f}, \Sigma^{-1/2}\varphi_j >$. This variable is defined for every $j \in \mathbb{N}$ if either *(i)* or *(ii)* is satisfied. By theorem 2.1 in Kuo (1975, page 116):

$$\frac{dP_n^\theta}{dP_n^0} = \prod_{j=1}^{\infty}\frac{d\nu_j}{d\mu_j}$$

where ν_j denotes the distribution of $\sqrt{n}z_j$ under P_n^θ and μ_j denotes the distribution of $\sqrt{n}z_j$ under P_n^0. By writing down the likelihoods of ν_j and μ_j with respect to the Lebesgue measure we obtain

$$\frac{dP_n^\theta}{dP_n^0} = \prod_{j=1}^{\infty}\frac{\left(1 + l_j^2 n\right)^{-1/2}\exp\{-\frac{1}{2}(z_j - < K(f_{0,\theta} - \tilde{f}), \Sigma^{-1/2}\varphi_j >)^2 n\left(1 + l_j^2 n\right)^{-1}\}}{\exp\{-\frac{1}{2}z_j^2 n\}}$$

which, after simplifications, gives the result.

$\square$

Theorem 2.1 is stated for a fixed n. In section 3, where the asymptotic behavior of the posterior distribution is analyzed, we need to replace the fixed prior for f with a scaled one. This will be made by replacing, when necessary, $\Omega_{0\theta}$ by $\frac{1}{\alpha n}\Omega_{0\theta}$ where $\alpha > 0$ and $\alpha \to 0$.

The marginal posterior distribution of θ can be used to compute a point estimator of θ. The maximum a posterior (MAP) estimator is particularly suitable and plays an important role in the study of the asymptotic properties of $\mu_\theta^{\hat{r}}$. The MAP θ_n is defined as

$$\begin{aligned}\theta_n &= \arg\max_{\theta\in\Theta} d\mu_\theta^{\hat{r}} \qquad (2.17)\\ &= \arg\max_{\theta\in\Theta} \frac{dP_n^\theta(\hat{r})\mu_\theta(d\theta)}{\int_\Theta dP_n^\theta(\hat{r})\mu_\theta(d\theta)} = \arg\max_{\theta\in\Theta} \frac{\frac{dP_n^\theta}{dP_n^0}(\hat{r})\mu_\theta(d\theta)}{\int_\Theta \frac{dP_n^\theta}{dP_n^0}(\hat{r})\mu_\theta(d\theta)}.\end{aligned}$$

Since the denominator of the posterior distribution does not depend on θ it plays no role in the optimization.

In general, when the conditional prior distribution on f, given θ, is very precise the MAP will essentially be equivalent to the maximum likelihood estimator (MLE) that we would obtain if we use the prior mean function $f_{0\theta}$ as the likelihood. On the contrary, with a prior μ_f^θ almost uninformative the MAP will be close to the GMM estimator (up to a prior on θ). The next example shows this argument in a rigorous way.

Example 2.2. Consider a function $h(\theta, x)$ that after normalization is of the form: $h(\theta, x) = a(x) - b(\theta)$ with $a, b \in \mathbb{R}^r$ and $\theta \in \mathbb{R}^k$, $k \leq r$ so that the model is in general over-identified and $Var(h(\theta, x)) = I_r$, where I_r denotes the r-dimensional identity matrix. This implies that the classical GMM estimator is solution of

$$\min_\theta \sum_{j=1}^r \left(\frac{1}{n}\sum_{i=1}^n \varphi_j(x_i) - b_j(\theta)\right)^2$$

with $a(x) = (a_1(x), \ldots, a_r(x))'$and $b(\theta) = (b_1(\theta), \ldots, b_r(\theta))'$.

Assume in this example that Π is the true distribution F_* which implies that $f_* = 1$. Denote $\varphi_j(x) \equiv \varphi_j(x; \theta) = (a_j(x) - b_j(\theta))$ for $j = 1, \ldots, r$ and $\varphi_0 = 1$. Under these assumptions the functions $(1, \varphi_1(x), \ldots, \varphi_r(x))$ form an orthonormal system in $\mathcal{E}$ and we can complete this system to form an orthonormal basis $\{\varphi_j\}_{j\geq 0}$. Since the $span\{1, \varphi_1(x), \ldots, \varphi_r(x)\}$ does not depend on θ then the same holds for its orthogonal and $\{\varphi_j\}_{j>r}$ are independent of θ. As described in section 2.1, the prior distribution μ_f^θ on f is $\mathcal{N}(f_{0\theta}, \Omega_0)$ where $f_{0\theta}$ verifies $\int a(x) f_{0\theta}(x)\Pi(x)dx = b(\theta)$ and Ω_0 verifies

$$\Omega_0 u = \lambda_1 < u, 1 > + \sum_{j=1}^r \lambda_j < u, \varphi_j > \varphi_j + \sum_{j=r+1}^\infty \lambda_j < u, \varphi_j > \varphi_j, \quad \forall u \in \mathcal{E}$$

where $\sum_j \lambda_j < \infty$ and $\lambda_j = 0$, $\forall j = 0, \ldots, r$. Therefore, Ω_0 is independent of θ.

In order to construct the sampling model we choose an operator K (that is, a function $k(x,t)$) with range in $\mathcal{F}$, singular functions $\{\varphi_j\}_{j\geq 0}$ and singular values $\{\lambda_{jK}\}_{j\geq 0}$, where $\{\lambda_{jK}\}_{j\geq 0}$ must be a non-increasing sequence of positive elements. Therefore, we have

$$K^*K\varphi_j = \lambda_{jK}^2 \varphi_j$$

and if we define $\psi_j \in \mathcal{F}$ as $K\varphi_j = \lambda_{jK}\psi_j$, $\lambda_{jK} \neq 0$, for every $j \geq 0$, we also have

$$K^*\psi_j = \lambda_{jK}\varphi_j \quad \text{and} \quad KK^*\psi_j = \lambda_{jK}^2\psi_j.$$

In practice, the operator K takes the form: $\forall \phi \in \mathcal{E}$, $K\phi = \sum_{j=0}^{\infty} \lambda_{jK} < \phi, \varphi_j > \psi_j$, where $\{\psi_j\}_{j\geq 0}$ is an orthonormal basis in $\mathcal{F}$. The first $r+1$ basis functions $\{\psi_j\}_{j=0}^{r}$ might also depend on θ. This construction of K will allow us to have a suitable spectrum of Σ. In fact, under our assumptions we can verify that $\Sigma\psi_j = \lambda_{jK}^2\psi_j$ for $j \geq 1$ and $\Sigma\psi_0 = 0$. To see this we write Σ in the form given in lemma 2.2: $\Sigma = KM_fK^* - KM_f < M_f, K^*\cdot >$ and if $f_* = 1$ we have

$$\begin{aligned}
\Sigma\psi_j &= KK^*\psi_j - K1 < 1, \lambda_{jK}\varphi_j >, \quad \text{for } j \neq 0 \\
\Sigma\psi_0 &= \lambda_{0K}^2\psi_0 - (\lambda_{0K}\psi_0)\lambda_{0K} < 1, \varphi_0 > .
\end{aligned}$$

Since $< 1, \varphi_j >= 0$ for $j \geq 1$ and $< 1, \varphi_0 >= 1$ we get the result.

From the result of theorem 2.1 the marginal likelihood is proportional to

$$\exp\left\{-\frac{1}{2}||\hat{r} - Kf_{0\theta}||^2_{\Sigma_n + K\Omega_0K^*}\right\}$$

where $||\cdot||^2_{\Sigma_n + K\Omega_0K^*}$ denotes the square of the norm in the reproducing kernel Hilbert space associated with the operator $(\Sigma_n + K\Omega_0K^*)$. The eigenvalues of this operator are the functions $\{\psi_j\}_{j\geq 0}$ previously constructed and the eigenvalues are denoted $\{\mu_{nj}^2\}_{j\geq 0}$ and given by

$$\begin{aligned}
\mu_{n0}^2 &= 0 \\
\mu_{nj}^2 &= \frac{1}{n}\lambda_{jK}^2, \quad \text{for } j = 1, \ldots, r \\
\mu_{nj}^2 &= \lambda_{jK}^2\left(\frac{1}{n} + \lambda_j\right).
\end{aligned}$$

Therefore, we can rewrite:

$$
\begin{aligned}
||\hat{r} - Kf_{0\theta}||^2_{\Sigma_n + K\Omega_0 K^*} &= \sum_{j;\, \mu_{nj}\neq 0} \frac{< \hat{r} - Kf_{0\theta}, \psi_j >^2}{\mu_{nj}^2} \\
&= \sum_{j;\, \mu_{nj}\neq 0} \mu_{nj}^{-2} \left(\frac{1}{n}\sum_{i=1}^{n} \int k(t,x_i)\psi_j(t)\rho(t)dt - \int\int k(t,x)\psi_j(t) f_{0\theta}(x)\Pi(dx)\rho(t)dt \right)^2 \\
&= \sum_{j;\, \mu_{nj}\neq 0} \mu_{nj}^{-2}\lambda_{jK}^2 \left(\frac{1}{n}\sum_{i=1}^{n} \varphi_j(x_i) - \int \varphi_j(x) f_{0\theta}(x)\Pi(dx) \right)^2 \\
&= \sum_{j=1}^{r} n \left(\frac{1}{n}\sum_{i=1}^{n} a_j(x_i) - b_j(\theta) \right)^2 + \sum_{j>r} \frac{1}{n^{-1} + \lambda_j} \left(\frac{1}{n}\sum_{i=1}^{n} \varphi_j(x_i) - \mathbf{E}_{0\theta}(\varphi_j) \right)^2
\end{aligned}
$$

for every $f_{0\theta}$ which satisfies $\int h(\theta,x) f_{0\theta}(x)\Pi(dx) = 0$. We have used $\mathbf{E}_{0\theta}$ to denote the expectation taken with respect to $f_{0\theta}$. Hence, the MAP verifies

$$
\begin{aligned}
\theta_n &= \arg\min_{\theta\in\Theta} \sum_{j=1}^{r} n \left(\frac{1}{n}\sum_{i=1}^{n} a_j(x_i) - b_j(\theta) \right)^2 + \sum_{j>r} \frac{\left(n^{-1}\sum_{i=1}^{n} \varphi_j(x_i) - \mathbf{E}_{0\theta}(\varphi_j)\right)^2}{n^{-1} + \lambda_j} \\
&= \arg\min_{\theta\in\Theta} \sum_{j=1}^{r} \left(\frac{1}{n}\sum_{i=1}^{n} a_j(x_i) - b_j(\theta) \right)^2 + \frac{1}{n}\sum_{j>r} \frac{\left(n^{-1}\sum_{i=1}^{n} \varphi_j(x_i) - \mathbf{E}_{0\theta}(\varphi_j)\right)^2}{n^{-1} + \lambda_j}. \qquad (2.18)
\end{aligned}
$$

These formulas clearly show that the prior distribution μ_f^θ completes the moment conditions and extends them to a continuum of moment conditions. In the case of an almost noninformative prior we have: $\lambda_j \to \infty$, $\forall j > r$ so that (2.18) is exactly the expression of the GMM. In the case of a perfectly informative prior (that is, $f = f_{0\theta}$ a.s. and $\lambda_j = 0$ for every j) the expression (2.18) becomes

$$
\theta_n = \arg\min ||\hat{r} - Kf_{0\theta}||^2_{\Sigma_n}.
$$

In this case the MAP is equivalent to the MLE obtained by using $f_{0\theta}$ as the likelihood in the sense that it possesses the same asymptotic distribution under very general conditions on K, see Carrasco and Florens (2012). A sufficient condition for this is that the closure of the vector space generated by the family $\{k(t,x)\}$ in $\mathcal{E}$ be equal to $\mathcal{E}$. Remark that this is the case for $k(t,x) = 1(x \leq t)$ and $k(t,x) = e^{itx}$ with $t, x \in \mathbb{R}$.

Remark 2.7. We have already discussed (see Remark 2.2) the possibility of using a different prior scheme when we are in the just-identified case and θ can be written as a linear functional of f. In that case, given a Gaussian process prior on f, the prior of θ is recovered through the transformation $\theta = B(f)$. The posterior distribution for θ is recovered from the posterior distribution of f (which is obviously unconditional on θ) through the transformation $B(f)$.

For clarity reasons, we summarize in tables 2 and 3 below the notation used for the sampling distribution (the true and the approximated one) and for the posterior distributions for both the overidentified and the just-identified cases.

Table 2: Sampling distribution

Sampling distribution:	Conditional on f_*	Conditional on f	Marginal
True	$P_{n,*}^{f_*}$	$P_{n,*}^{f}$	–
Approximated	$P_n^{f_*}$	P_n^f	P_n^θ

Table 3: Posterior distribution

Case:	over-identified	just-identified: 1st possibility	just-identified: 2nd possibility
Marginal of θ	$\mu_\theta^{\hat{r}}(\theta\|\hat{r})$	$\mu_\theta^{\hat{r}}(\theta\|\hat{r})$	$\mu_\theta^{\hat{r}}(\theta\|\hat{r})$ through $\theta = B(f)$
Conditional of $f\|\theta$	$\mu_f^{\hat{r},\theta}(f\|\hat{r},\theta)$	$\mu_f^{\hat{r},\theta}(f\|\hat{r},\theta)$	–
Regularized Conditional of $f\|\theta$	$\mu_{f,\tau}^{\hat{r},\theta}(f\|\hat{r},\theta,\tau)$	$\mu_{f,\tau}^{\hat{r},\theta}(f\|\hat{r},\theta,\tau)$	–
Marginal of f	–	–	$\mu_f^{\hat{r}}(f\|\hat{r})$
Regularized of f	–	–	$\mu_{f,\tau}^{\hat{r}}(f\|\hat{r},\tau)$

3 Asymptotic Analysis

In this section we focus on the asymptotic properties of our approach. Along all this section we replace $\Omega_{0\theta}$ by $\frac{1}{\alpha n}\Omega_{0\theta}$ where $\alpha > 0$ and $\alpha \to 0$. This expression is very general since depending on the choice of α the prior μ_f^θ is: *(i)* shrinking (when $\alpha n \to \infty$), *(ii)* spreading out (when $\alpha = o(n^{-1})$) and *(iii)* fixed (when $\alpha = n^{-1}$). In some cases a scaling prior is necessary in order to obtain the minimax rates of convergence for the posterior distribution $\mu_f^{\theta,\hat{r}}$ (see Florens and Simoni (2012b)).

We analyze three issues: *(i)* convergence of the "misspecified" posterior $\mu_\theta^{\hat{r}}$ towards the true marginal posterior of θ for the just-identified case where θ is a linear functional of f (section 3.1); *(ii)* posterior consistency of $\mu_\theta^{\hat{r}}$ and *(iii)* convergence of $\mu_\theta^{\hat{r}}$ towards a normal distribution (section 3.2).

Along all this section we use the posterior distribution $\mu_f^{\theta,\hat{r}}$ given either in lemma 2.4 or in lemma 2.5. Therefore, we implicitly assume that the conditions of these lemmas are satisfied. When this is not the case, then our asymptotic results can be easily extended to the case where the exact posterior $\mu_f^{\theta,\hat{r}}$ is replaced by the *regularized posterior distribution* $\mu_{f,\tau}^{\theta,\hat{r}}$ discussed in Remark 2.6.

3.1 Asymptotic for linear functionals: the just-identified case

We first analyze asymptotic for the just-identified case where $k = r$ and where $\theta = B(f)$ writes as an explicit linear functional of f. Without loss of generality we can write $\theta =< f, g >$ for any $g \in \mathcal{E}^r$. In fact, by the Riesz theorem there exists a unique $g \in \mathcal{E}$ such that $B(f) =< f, g >$. We consider the situation described in Remark 2.2 where the prior distribution of θ is specified through the prior distribution of f. The analysis for this case is quite simple since we have a closed-form for the posterior distribution of θ.

The results of section 2.3.1 and Remark 2.7 imply that the posterior distribution of θ is Gaussian with mean $< \mathbf{E}[f|\hat{r}], g >$ and variance $< Var[f|\hat{r}]g, g >$ where $\mathbf{E}[f|\hat{r}]$ and $Var[f|\hat{r}]$ have been defined in lemma 2.4. We denote by $\mu_\theta^{\hat{r}}$ this distribution and we stress that this is the approximated posterior distribution of θ where the approximation is due to the fact that the Gaussian sampling distribution P_n^f we have used is not the true one (but only the weak limit in distribution of the true sampling distribution $P_{n,*}^f$). Denote $\hat{\theta} := n^{-1}\sum_{i=1}^n g(x_i)$ the method of moment estimator and $\sigma^2 =< f_* g, g > - < f_*, g >^2$ the true variance of g. The next theorem states that the total variation distance – denoted by $||\cdot||_{TV}$ – between $\mu_\theta^{\hat{r}}$ and $\mathcal{N}(\hat{\theta}, \frac{\sigma^2}{n})$ converges to 0 in probability. For this result we need the following assumptions:

TV-1. There exists a kernel function C such that $\forall h > 0$ small and $\forall u$:

$$\left(f_0 - \frac{1}{h}C\left(\frac{x-u}{h}\right)\frac{1}{\pi(x)}\right) \in \mathcal{R}\left(\Omega_0^{1/2}(T^*T)^{\beta/2}\right)$$

for some $\beta > 0$.

TV-2. There exists a kernel function C such that $\forall h > 0$ small and $\forall u$:

$$\left|\left|g(x_i) - \int g(x)\frac{1}{h}C\left(\frac{x-u}{h}\right)dx\right|\right| = \mathcal{O}(h^2).$$

TV-3. There exists a kernel function C such that $\forall h > 0$ small and $\forall u$:

$$\left|\left|k(u,t) - \int k(u,t)\frac{1}{h}C\left(\frac{x-u}{h}\right)dx\right|\right| = \mathcal{O}(h^2.$$

Theorem 3.1. *Let $\hat{\theta} = n^{-1}\sum_{i=1}^n g(x_i)$ and consider the Gaussian model 2.8 independent of θ with the prior covariance operator Ω_0 replaced by $\frac{1}{\alpha n}\Omega_0$ where $\alpha > 0$, $\alpha \to 0$ and $n\alpha^{\beta\wedge 2} \to 0$. Let the assumptions of lemma 2.5 and assumptions TV-1 - TV-3 hold true. Define $\xi_i := h^{-1}C\left(\frac{x-x_i}{h}\right)\frac{1}{\pi(x)}$, $\forall i = 1, \ldots, n$, and assume that there exists a random element $\exists \zeta_i \in \mathcal{E}$ such that: (i) $n^{-1}\sum_{i=1}^n ||\zeta_i|| = \mathcal{O}_p(1)$ and (ii) $(\xi_i - f_0) = \Omega_0^{1/2}\left(\Sigma^{-1/2}K\Omega_0K^*\Sigma^{-1/2}\right)^{\beta/2}\zeta_i$ for some $\beta > 0$. Hence, :*

$$\left|\left|\mu_\theta^{\hat r} - \mathcal{N}(\hat\theta, \frac{\sigma^2}{n})\right|\right|_{TV} \to 0$$

in $P_n^{f_*}$*-probability.*

Proof. Let f denote the density function of the posterior $\mu_\theta^{\hat r}$ and

$$\left|\left|\mu_\theta^{\hat r} - \mathcal{N}(\hat\theta, \frac{\sigma^2}{n})\right|\right|_{TV} \leq \left|\left|\mu_\theta^{\hat r} - \mathcal{N}\left(< \mathbf{E}[f|\hat r], g >, \frac{\sigma^2}{n}\right)\right|\right|_{TV} + \left|\left|\mathcal{N}\left(< \mathbf{E}[f|\hat r], g >, \frac{\sigma^2}{n}\right) - \mathcal{N}\left(\hat\theta, \frac{\sigma^2}{n}\right)\right|\right|_{TV}. \quad (3.1)$$

By trivial algebra it is possible to show that

$$\left|\left|\mathcal{N}(< \mathbf{E}[f|\hat r], g >, \sigma^2/n) - \mathcal{N}(\hat\theta, \sigma^2/n)\right|\right|_{TV} = 4\left[\Phi\left(\frac{\sqrt{n}| < \mathbf{E}[f|\hat r], g > - \hat\theta|}{2\sigma}\right) - \frac{1}{2}\right]$$

and

$$\left|\left|\mu_\theta^{\hat r} - \mathcal{N}\left(< \mathbf{E}[f|\hat r], g >, \frac{\sigma^2}{n}\right)\right|\right|_{TV} = 4\left[\Phi\left(\sqrt{\log\left(\frac{n\tau^2}{\sigma^2}\right)}\frac{\sqrt{n}\tau}{\sqrt{|\sigma^2 - n\tau^2|}}\right) - \Phi\left(\sqrt{\log\left(\frac{n\tau^2}{\sigma^2}\right)}\frac{\sigma}{\sqrt{|\sigma^2 - n\tau^2|}}\right)\right]$$

where $\Phi(\cdot)$ denotes the *cdf* of a $\mathcal{N}(0,1)$-distribution and $\tau^2 =< Var[f|\hat r]g, g >$. We start by computing the rate for $\left|\frac{\sigma^2}{n} - \tau^2\right|$. Remark that under the conditions of lemma 2.5 we can write the posterior variance either in the form given in the lemma or in the form given in lemma 2.4. We use this second expression:

$$\begin{aligned}
\left|\frac{\sigma^2}{n} - \tau^2\right| &= \left|< Var[f|\hat r]g, g > - \frac{\sigma^2}{n}\right| = \left|< Var[f|\hat r]g, g > - \frac{< f_*(g - \mathbf{E}_*g), g >}{n}\right| \\
&= \frac{1}{n}\left|< \Omega_0(\alpha f_* - \alpha f_* < f_*, \cdot > + \Omega_0)^{-1} f_*(g - \mathbf{E}_*g), g > - < f_*(g - \mathbf{E}_*g), g >\right| \\
&= \frac{1}{n}\left|< \left[\Omega_0(\alpha f_* - \alpha f_* < f_*, \cdot > + \Omega_0)^{-1} - I\right] f_*(g - \mathbf{E}_*g), g >\right| \\
&= \frac{1}{n}\left|< \left[\alpha f_* - \alpha f_* < f_*, \cdot >\right](\alpha f_* - \alpha f_* < f_*, \cdot > + \Omega_0)^{-1} f_*(g - \mathbf{E}_*g), g >\right| \\
&= \frac{\alpha}{n}\Big| < f_*(\alpha f_* - \alpha f_* < f_*, \cdot > + \Omega_0)^{-1} f_*(g - \mathbf{E}_*g), g > \\
&\quad - < f_* < f_*, \cdot > (\alpha f_* - \alpha f_* < f_*, \cdot > + \Omega_0)^{-1} f_*(g - \mathbf{E}_*g), g > \Big| \\
&= \frac{\alpha}{n}\Big| < (\alpha f_* - \alpha f_* < f_*, \cdot > + \Omega_0)^{-1} f_*(g - \mathbf{E}_*g), f_*g > \\
&\quad - < f_*, (\alpha f_* - \alpha f_* < f_*, \cdot > + \Omega_0)^{-1} f_*(g - \mathbf{E}_*g) >< f_*, g > \Big|.
\end{aligned}$$

Remark that $f_*(g - \mathbf{E}_*g) \in \mathcal{R}(\Omega_0^{1/2})$ since $\mathcal{R}(\Omega_0^{1/2}) = \overline{\mathcal{N}(\Omega_0^{1/2})}^{\perp}$ and $\int f_*(g - \mathbf{E}_*g) d\Pi = 0$. Thus, there exists $\nu \in \mathcal{E}$ such that $f_*(g - \mathbf{E}_*g) = \Omega_0^{1/2}\nu$ and

$$
\begin{aligned}
\left|\frac{\sigma^2}{n}-\tau^2\right| &= \frac{\alpha}{n}\Big| < (\alpha f_* - \alpha f_* < f_*, \cdot > +\Omega_0)^{-1}\Omega_0^{1/2}\nu, f_* g > \\
&\quad - < f_*, (\alpha f_* - \alpha f_* < f_*, \cdot > +\Omega_0)^{-1}\Omega_0^{1/2}\nu >< f_*, g > \Big| \\
&= \frac{\alpha}{n}\Big| < (\alpha I - \alpha f_*^{1/2} < f_*^{1/2}, \cdot > + f_*^{-1/2}\Omega_0 f_*^{-1/2})^{-1} f_*^{-1/2}\Omega_0^{1/2}\nu, f_*^{1/2} g > \\
&\quad - < f_*, (\alpha f_* - \alpha f_* < f_*, \cdot > +\Omega_0)^{-1}\Omega_0^{1/2}\nu >< f_*, g > \Big| \\
&= \frac{\alpha}{n}\Big| < (\alpha I + f_*^{-1/2}\Omega_0 f_*^{-1/2})^{-1} f_*^{-1/2}\Omega_0^{1/2}\nu, f_* g > \\
&\quad - < (\alpha I - \alpha f_*^{1/2} < f_*^{1/2}, \cdot > + f_*^{-1/2}\Omega_0 f_*^{-1/2})^{-1}(\alpha f_*^{1/2} < f_*^{1/2}, \cdot >) \times \\
&\quad (\alpha I + f_*^{-1/2}\Omega_0 f_*^{-1/2})^{-1} f_*^{-1/2}\Omega_0^{1/2}\nu, f_*^{1/2} g > \\
&\quad - < f_*^{1/2}, (\alpha I + f_*^{-1/2}\Omega_0 f_*^{-1/2})^{-1} f_*^{-1/2}\Omega_0^{1/2}\nu > \\
&\quad + < f_*^{1/2}, (\alpha I - \alpha f_*^{1/2} < f_*^{1/2}, \cdot > + f_*^{-1/2}\Omega_0 f_*^{-1/2})^{-1}(\alpha f_*^{1/2} < f_*^{1/2}, \cdot >) \times \\
&\quad (\alpha I + f_*^{-1/2}\Omega_0 f_*^{-1/2})^{-1} f_*^{-1/2}\Omega_0^{1/2}\nu >< f_*, g > \Big| \\
&= \mathcal{O}\left(2\frac{\sqrt{\alpha}}{n}||\nu||\sqrt{\mathbf{E}_* g^2} + 2\frac{\sqrt{\alpha}}{n}||\nu||\right) = \mathcal{O}\left(\frac{\sqrt{\alpha}}{n}\right)
\end{aligned}
$$

if $\mathbf{E}_* g^2 < \infty$. Note that to obtain the big $\mathcal{O}$ in the last line we have used the Cauchy-Schwarz inequality. Next, we study the term $| < \mathbf{E}[f|\hat{r}], g > -\hat{\theta}|$. Define $\xi_i := h^{-1}C\left(\frac{x-x_i}{h}\right)\frac{1}{\pi(x)}$. By using the expression of the posterior mean given in lemma 2.5 and denoting $B = \Sigma^{-1/2}K\Omega_0^{1/2}$ we obtain:

$$
\begin{aligned}
| < \mathbf{E}[f|\hat{r}], g > -\hat{\theta}| &= \left| < \mathbf{E}[f|\hat{r}], g > -n^{-1}\sum_{i=1}^n g(x_i)\right| \\
&= \left| < f_0 + \Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}B^*\Sigma^{-1/2}\left[n^{-1}\sum_{i=1}^n k(x_i,t) - Kf_0\right], g > -n^{-1}\sum_{i=1}^n g(x_i)\right| \\
&= \Big| < f_0 + \Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}B^*\Sigma^{-1/2}\left[n^{-1}\sum_{i=1}^n K\xi_i + \mathcal{O}_p(h^2) - Kf_0\right], g > \\
&\quad -n^{-1}\sum_{i=1}^n < \xi_i, g > + \mathcal{O}_p(h^2)\Big|
\end{aligned}
$$

since $n^{-1}\sum_{i=1}^n g(x_i) = n^{-1}\sum_{i=1}^n < \xi_i, g > +\mathcal{O}_p(h^2)$ and $n^{-1}\sum_{i=1}^n k(x_i,t) = n^{-1}\sum_{i=1}^n K\xi_i + \mathcal{O}_p(h^2)$ under assumption 2 and 3. Therefore,

$$
\begin{aligned}
| < \mathbf{E}[f|\hat{r}], g > -\hat{\theta}| &= | < n^{-1}\sum_{i=1}^n\left[\Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}B^*\Sigma^{-1/2}K(\xi_i - f_0) - (\xi_i - f_0)\right], g > \\
&\quad + < \Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}B^*\Sigma^{-1/2}\mathcal{O}_p(h^2), g > + \mathcal{O}_p(h^2)| =: |\mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3|.
\end{aligned}
$$

Since $(\xi_i - f_0) \in \mathcal{R}(\Omega_0^{1/2})$ then there exists $\eta_i \in \mathcal{E}$ such that $(\xi_i - f_0) = \Omega_0^{1/2}\eta_i$; moreover, there exists $\zeta_i \in \mathcal{E}$ (function of the data x_i) such that $\eta_i = (T^*T)^{\beta/2}\zeta_i$ for some $\beta > 0$. Hence,

$$
\begin{aligned}
|\mathcal{A}_1| &= \left| < n^{-1}\sum_{i=1}^{n}\left[\Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}B^*\Sigma^{-1/2}K(\xi_i - f_0) - (\xi_i - f_0)\right], g > \right| \\
&= \alpha\left| < \Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}n^{-1}\sum_{i=1}^{n}\eta_i, g > \right| = \alpha\left| < \Omega_0^{\frac{1}{2}}(\alpha I + B^*B)^{-1}(B^*B)^{\beta/2}n^{-1}\sum_{i=1}^{n}\zeta_i, g > \right| \\
&= \mathcal{O}_p\left(\alpha^{(\beta\wedge 2)/2}\right)
\end{aligned}
$$

if $n^{-1}\sum_i^n ||\zeta_i|| = \mathcal{O}_p(1)$. Since $\mathcal{R}(K\Omega_0^{1/2}) \subseteq \mathcal{R}(\Sigma)$ term $|\mathcal{A}_2|$ is well-defined and $|\mathcal{A}_2| = \mathcal{O}_p(\alpha^{-1}h^2)$. Finally, we choose h that converges to 0 sufficiently fast to guarantee that $\alpha^{-1}h^2 \to 0$. Under the condition that $n\alpha^{\beta\wedge 2} \to 0$ the first term of (3.1) converges to 0.

3.2 Posterior Consistency

In this section we study the consistency of the posterior distribution of θ. Posterior consistency for $\mu_f^{\hat{r},\theta}$ and $\mu_{f,\tau}^{\hat{r},\theta}$ has been shown respectively in Florens and Simoni (2012a) and Florens and Simoni (2012b).

Let $\Theta_n = \{\theta \in \Theta;\ \sqrt{n}||\theta - \theta_*|| \leq M_n\}$ for every sequence $M_n \to \infty$ and

$$
\mathcal{F}(\theta) = \left\{ f \in \mathcal{E}_M; \int h(x,\theta)f(x)d\Pi(x) = 0 \right\}.
$$

We want to show that the posterior measure $\mu_\theta^{\hat{r}}(\Theta_n)$ converges to 1 in $P_{n,*}^{f_*}$-probability. Define $P_n^{f_*} = \mathcal{N}(Kf_*, \frac{1}{n}\Sigma)$. By theorem 2.1, $P_n^{f_*}$ dominates P_n^θ so that we define $p_{n\theta} = dP_n^\theta/dP_n^{f_*}$. For a covariance operator $C : \mathcal{F} \to \mathcal{F}$ and $\varphi \in \mathcal{R}(C^{1/2})$ denote $||\varphi||_C$ the norm in the reproducing kernel Hilbert space associated with C defined as $||\varphi||_C^2 =< C^{-1/2}\varphi, C^{-1/2}\varphi >$. We introduce the following assumptions:

A-1. There exists a constant $\underline{c} > 0$ such that for every $\theta \in \Theta_n^c$

$$
\underline{c}||\theta - \theta_*|| \leq \inf_{f_{0\theta}\in\mathcal{F}(\theta)} ||\Sigma^{-1/2}K(f_{0\theta} - f_*)||_{\alpha I + BB^*}.
$$

A-2. For the constant $\underline{c} > 0$ defined in A-1, the set

$$
\tilde{\Theta}_n := \left\{ \theta;\ \inf_{f_{0\theta}\in\mathcal{F}(\theta)} ||\Sigma^{-1/2}K(f_{0\theta} - f_*)||_{\alpha I + BB^*} \leq \frac{\underline{c}M_n}{\sqrt{n}2} \right\} \subset \Theta_n
$$

is non empty.

A-3. The prior distribution μ_θ is continuous in θ and $0 < \mu_\theta(\theta) < \infty$ for every $\theta \in \Theta$.

For a probability measure P and an integrable function g we use the notation Pg to abbreviate $\int gdP$.

Theorem 3.2. *Under A-1, A-2 and A-3:*

$$P_{n,*}^{f_*}\mu_\theta^{\hat{r}}(\theta \in \Theta_n^c|\hat{r}) \to 0.$$

Proof. Define the event $A_n := \left\{\sum_{j=1}^\infty \frac{\alpha n}{(l_j^2+\alpha)}\frac{1}{\alpha}l_j^2 z_j^2 < \underline{c}\alpha M_n^2/2\right\}$ where $z_j =< \hat{r}-Kf_*, \Sigma^{-1/2}\varphi_j >$, $\forall j$. By the Markov's inequality the probability of this event, under $P_{n,*}^{f_*}$, converges to 1 if $\alpha M_n^2 \to \infty$. In fact, we have

$$P_{n,*}^{f_*}A_n^c \quad \leq \quad \frac{2}{\underline{c}\alpha M_n^2}\mathbf{E}_*\left|\sum_{j=1}^\infty \frac{\alpha n}{(l_j^2+\alpha)}\frac{1}{\alpha}l_j^2 z_j^2\right| = \frac{2}{\underline{c}\alpha M_n^2}\sum_{j=1}^\infty \frac{l_j^2}{(l_j^2+\alpha)}$$

which converges to 0 if $\alpha M_n^2 \to \infty$ and $\sum_{j=1}^\infty \frac{l_j^2}{(l_j^2+\alpha)} < \infty$.

The quantity of interest $P_{n,*}^{f_*}\mu_\theta^{\hat{r}}(\theta \in \Theta_n^c|\hat{r})$ may be rewritten as

$$\begin{aligned} P_{n,*}^{f_*}\mu_\theta^{\hat{r}}(\theta \in \Theta_n^c|\hat{r}) &= P_{n,*}^{f_*}\frac{\int_{\Theta_n^c} p_{n\theta}(\hat{r})\mu_\theta(d\theta)}{\int_\Theta p_{n\theta}(\hat{r})\mu_\theta(d\theta)}I_{A_n} + P_{n,*}^{f_*}\frac{\int_{\Theta_n^c} p_{n\theta}(\hat{r})\mu_\theta(d\theta)}{\int_\Theta p_{n\theta}(\hat{r})\mu_\theta(d\theta)}I_{A_n^c} && (3.2) \\ &= P_{n,*}^{f_*}\frac{\int_{\Theta_n^c} p_{n\theta}(\hat{r})\mu_\theta(d\theta)}{\int_\Theta p_{n\theta}(\hat{r})\mu_\theta(d\theta)}I_{A_n} + o(1) && (3.3) \end{aligned}$$

where I_A denotes the indicator function of an event A. Now, in order to upper bound the numerator and lower bound the denominator we use the explicit form for $p_{n\theta}$ given in (2.16) with $\tilde{f}$ replaced by f_* and $\Omega_{0\theta}$ replaced by $(\alpha n)^{-1}\Omega_{0\theta}$:

$$p_{n\theta} \quad = \quad \prod_{j=1}^\infty \sqrt{\frac{\alpha}{l_j^2+\alpha}}e^{\frac{\alpha n}{2(l_j^2+\alpha)}\left(\alpha^{-1}l_j^2 z_j^2 - A_j^2 + 2z_j A_j\right)}$$

where $A_j =< K(f_{0\theta}-f_*), \Sigma^{-1/2}\varphi_j >$ and $z_j =< \hat{r}-Kf_*, \Sigma^{-1/2}\varphi_j >$, $\forall j$. Therefore, since $\frac{\alpha}{l_j^2+\alpha} \leq 1$ and $\frac{\alpha}{l_j^2+\alpha} \geq \frac{\alpha}{l_1^2+\alpha}$:

$$\begin{aligned} &P_{n,*}^{f_*}\frac{\int_{\Theta_n^c} p_{n\theta}(\hat{r})\mu_\theta(d\theta)}{\int_\Theta p_{n\theta}(\hat{r})\mu_\theta(d\theta)}I_{A_n} \leq P_{n,*}^{f_*}\frac{\int_{\Theta_n^c}\exp\left\{\frac{1}{2}\sum_{j=1}^\infty \frac{\alpha n}{(l_j^2+\alpha)}\left(\alpha^{-1}l_j^2 z_j^2 - A_j^2 + 2z_j A_j\right)\right\}\mu_\theta(d\theta)}{\sqrt{\frac{\alpha}{(l_1^2+\alpha)}}\int_\Theta \exp\{-\frac{1}{2}\sum_{j=1}^\infty z_j^2\frac{\alpha n}{\alpha+l_j^2} + \frac{1}{2}\sum_{j=1}^\infty z_j^2 n + \sum_{j=1}^\infty z_j A_j\frac{\alpha n}{\alpha+l_j^2} - \frac{1}{2}\sum_{j=1}^\infty A_j^2\frac{\alpha n}{\alpha+l_j^2}\}\mu_\theta(d\theta)}I_{A_n} \\ &\leq P_{n,*}^{f_*}\frac{\int_{\Theta_n^c}\exp\left\{\frac{1}{2}\sum_{j=1}^\infty \frac{\alpha n}{(l_j^2+\alpha)}\alpha^{-1}l_j^2 z_j^2 - \frac{1}{2}||\Sigma^{-1/2}K(f_{0\theta}-f_*)||^2_{\alpha I+BB^*} + \mathcal{O}_p\left(\sqrt{\alpha n}||\Sigma^{-1/2}K(f_{0\theta}-f_*)||_{\alpha I+BB^*}\right)\right\}\mu_\theta(d\theta)}{\sqrt{\frac{\alpha}{(l_1^2+\alpha)}}\int_\Theta \exp\{\sum_{j=1}^\infty z_j A_j\frac{\alpha n}{\alpha+l_j^2} - \frac{1}{2}\sum_{j=1}^\infty A_j^2\frac{\alpha n}{\alpha+l_j^2}\}\mu_\theta(d\theta)}I_{A_n} \\ &= P_{n,*}^{f_*}\frac{\int_{\Theta_n^c}\exp\left\{\frac{1}{2}\sum_{j=1}^\infty \frac{n}{(l_j^2+\alpha)}l_j^2 z_j^2 - \frac{1}{2}||\Sigma^{-1/2}K(f_{0\theta}-f_*)||^2_{\alpha I+BB^*} + \mathcal{O}_p\left(\sqrt{\alpha n}||\Sigma^{-1/2}K(f_{0\theta}-f_*)||_{\alpha I+BB^*}\right)\right\}\mu_\theta(d\theta)}{\sqrt{\frac{\alpha}{(l_1^2+\alpha)}}\int_\Theta \exp\left\{-\frac{1}{2}\alpha n||\Sigma^{-1/2}K(f_{0\theta}-f_*)||^2_{\alpha I+BB^*}(1+\mathcal{O}_p((\alpha n)^{-1/2}||\Sigma^{-1/2}K(f_{0\theta}-f_*)||^{-1}_{\alpha I+BB^*}))\right\}\mu_\theta(d\theta)}I_{A_n} \end{aligned}$$

since $\sum_{j=1}^{\infty} A_j^2 \frac{1}{\alpha + l_j^2} = ||\Sigma^{-1/2} K(f_{0\theta} - f_*)||^2_{\alpha I + BB^*}$ and

$$\sum_j z_j A_j \alpha n / (\alpha + l_j^2) = \mathcal{O}_p\left(\sqrt{\alpha n} ||\Sigma^{-1/2} K(f_{0\theta} - f_*)||_{\alpha I + BB^*}\right).$$

Moreover, since on A_n $\sum_{j=1}^{\infty} \frac{\alpha n}{(l_j^2 + \alpha)} \frac{1}{\alpha} l_j^2 z_j^2 < \underline{c} \alpha M_n^2 / 2$ we have:

$$\begin{aligned}
&P_{n,*}^{f_*} \frac{\int_{\Theta_n^c} p_{n\theta}(\hat{r}) \mu_\theta(d\theta)}{\int_\Theta p_{n\theta}(\hat{r}) \mu_\theta(d\theta)} I_{A_n} \\
&\leq P_{n,*}^{f_*} \frac{\int_{\Theta_n^c} \exp\left\{ \frac{\underline{c}\alpha M_n^2}{4} - \frac{1}{2} ||\Sigma^{-1/2} K(f_{0\theta} - f_*)||^2_{\alpha I + BB^*} + \mathcal{O}_p\left(\sqrt{\alpha n} ||\Sigma^{-1/2} K(f_{0\theta} - f_*)||_{\alpha I + BB^*}\right) \right\} \mu_\theta(d\theta)}{\sqrt{\frac{\alpha}{(l_1^2 + \alpha)}} \int_{\tilde{\Theta}_n} \exp\left\{ -\frac{1}{2} \alpha n ||\Sigma^{-1/2} K(f_{0\theta} - f_*)||^2_{\alpha I + BB^*} (1 + \mathcal{O}_p((\alpha n)^{-1/2} ||\Sigma^{-1/2} K(f_{0\theta} - f_*)||^{-1}_{\alpha I + BB^*})) \right\} \mu_\theta(d\theta)} I_{A_n} \\
&\leq P_{n,*}^{f_*} \frac{\int_{\Theta_n^c} \exp\left\{ \frac{\underline{c}\alpha M_n^2}{4} - \frac{1}{2} ||\Sigma^{-1/2} K(f_{0\theta} - f_*)||^2_{\alpha I + BB^*} + \mathcal{O}_p\left(\sqrt{\alpha n} ||\Sigma^{-1/2} K(f_{0\theta} - f_*)||_{\alpha I + BB^*}\right) \right\} \mu_\theta(d\theta)}{\sqrt{\frac{\alpha}{(l_1^2 + \alpha)}} \exp\left\{ -\alpha n \frac{\underline{c} M_n^2}{n 8} (1 + \mathcal{O}_p((\alpha)^{-1/2} M_n^{-1})) \right\} \int_{\tilde{\Theta}_n} \mu_\theta(d\theta)} I_{A_n} \\
&\leq P_{n,*}^{f_*} \frac{\int_{\Theta_n^c} \exp\left\{ \frac{\underline{c}\alpha M_n^2}{4} - \frac{\alpha n}{2} \underline{c}^2 ||\theta - \theta_*||^2 (1 + \mathcal{O}_p\left(\frac{1}{\sqrt{\alpha n}} \underline{c}^{-1} ||\theta - \theta_*||^{-1}\right)) \right\} \mu_\theta(d\theta)}{\sqrt{\frac{\alpha}{(l_1^2 + \alpha)}} \exp\left\{ -\alpha n \frac{\underline{c} M_n^2}{n 8} (1 + \mathcal{O}_p((\alpha)^{-1/2} M_n^{-1})) \right\} \int_{\tilde{\Theta}_n} \mu_\theta(d\theta)} I_{A_n} \\
&\leq P_{n,*}^{f_*} \frac{\int_{\Theta_n^c} \exp\left\{ \frac{\underline{c}\alpha M_n^2}{4} - \frac{\alpha n}{2} \underline{c} \frac{M_n^2}{n} (1 + \mathcal{O}_p\left(\frac{1}{\sqrt{\alpha n}} \underline{c}^{-1} \frac{\sqrt{n}}{M_n}\right)) \right\} \mu_\theta(d\theta)}{\sqrt{\frac{\alpha}{(l_1^2 + \alpha)}} \exp\left\{ -\alpha n \frac{\underline{c} M_n^2}{n 8} (1 + \mathcal{O}_p((\alpha)^{-1/2} M_n^{-1})) \right\} \int_{\tilde{\Theta}_n} \mu_\theta(d\theta)} I_{A_n} \\
&\leq P_{n,*}^{f_*} \frac{\int_{\Theta_n^c} \exp\left\{ -\frac{\underline{c}\alpha M_n^2}{4} (1 + \mathcal{O}_p\left(\frac{1}{\sqrt{\alpha n}} \underline{c}^{-1} \frac{\sqrt{n}}{M_n}\right)) \right\} \mu_\theta(d\theta)}{\sqrt{\frac{\alpha}{(l_1^2 + \alpha)}} \exp\left\{ -\alpha n \frac{\underline{c} M_n^2}{n 8} (1 + \mathcal{O}_p((\alpha)^{-1/2} M_n^{-1})) \right\} \int_{\tilde{\Theta}_n} \mu_\theta(d\theta)} I_{A_n}
\end{aligned}$$

where we have used assumptions A-1 and A-2 and $0 < \int_{\tilde{\Theta}_n} \mu_\theta(d\theta) < \infty$ under assumptions A-2 - A-3. We conclude that

$$P_{n,*}^{f_*} \mu_\theta^{\hat{r}}(\theta \in \Theta_n^c | \hat{r}) \quad \leq \quad exp\left\{ -\frac{\underline{c}\alpha M_n^2}{4} + \frac{\underline{c}\alpha M_n^2}{8} \right\} P_{n,*}^{f_*}(A_n) const. + P_{n,*}^{f_*}(A_n^c) = o(1).$$

$\square$

3.3 Asymptotic Normality

We show now that asymptotically the posterior distribution of θ behaves like a Normal distribution centered around the MAP estimator θ_n defined in (2.17) as $\theta_n = \arg\max_{\theta \in \Theta} \mu_\theta^{\hat{r}}(d\theta)$. We can equivalently define θ_n as $\theta_n = \arg\max_{\theta \in \Theta} \frac{dP_n^\theta}{dP_n^{f_*}}(\hat{r}) \mu_\theta(d\theta) = \arg\max_{\theta \in \Theta} p_{n\theta}(\hat{r}) \mu_\theta(d\theta)$. In the following we abbreviate $p_{n\theta} \mu_\theta(d\theta) = p_{n\theta}(\hat{r}) \mu_\theta(d\theta)$.

Let $H(\theta_n, \delta) := \{\theta \in \Theta; ||\theta - \theta_n|| < \delta\}$. Remark that under the assumptions of theorem 3.2, for $\delta_n > 0$ such that $\delta_n \to 0$ at a suitable rate the posterior distribution of $H(\theta_n, \delta_n)$ converges to 1. Denote $L_n(\theta) = \log \frac{dP_n^\theta}{dP_n^{f_*}}(\hat{r}) \mu_\theta(d\theta)$. We make the following assumptions

B-1. θ_n is a strict local maximum of $d\mu_\theta^{\hat{r}}$ and $L_n'(\theta_n) := \left.\frac{\partial L_n(\theta)}{\partial \theta}\right|_{\theta = \theta_n} = 0$.

B-2. $\Delta_n = \{-L_n''(\theta_n)\}^{-1} := \left\{ - \frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_n} \right\}^{-1}$ exists and is positive definite.

B-3. $d^2 \to 0$ as $n \to \infty$ where d^2 is the largest eigenvalue of Δ_n.

B-4. For any $\epsilon > 0$, there exists an integer N and $\delta > 0$ such that, for any $n > N$ and $\theta \in H(\theta_n, \delta)$, $L_n''(\theta)$ exists and satisfies

$$I - G(\epsilon) \le L_n''(\theta)\{L_n''(\theta_n)\}^{-1} \le I + G(\epsilon),$$

where I is a $(k \times k)$ identity matrix and $G(\epsilon)$ is a $(k \times k)$ positive semidefinite symmetric matrix whose largest eigenvalue $g(\epsilon)$ converges to 0 as $\epsilon \to 0$.

We provide later sufficient conditions for this assumptions. In particular, assumptions B-3 and B-4 are satisfied if the conditional prior μ_f^θ of f is shrinking, that is, $\Omega_{0\theta}$ is replaced by $\tau\Omega_{0\theta}$ and $\tau = o(n^{-1})$.

Lemma 3.1 and theorem 3.3 below are slightly modifications of results in Kim (2002) and Chen (1985).

Lemma 3.1. *Under A-1, A-2, A-3, B-1, B-2, B-3 and B-4*

$$\lim_{n\to\infty} \mu_\theta^{\hat{r}}(\theta_n)|\Delta_n|^{\frac{1}{2}} \le (2\pi)^{-k/2}. \tag{3.4}$$

Moreover, $\lim_{n\to\infty} \mu_\theta^{\hat{r}}(\theta_n)|\Delta_n|^{\frac{1}{2}} = (2\pi)^{-k/2}$ *in* $P_{n,*}^{f_*}$*-probability if and only if for some* $\delta_n > 0$, $\delta_n \to 0$, $\mu_\theta^{\hat{r}}(H(\theta_n, \delta_n)|\hat{r}) \to 1$.

Proof. Denote $D_{\hat{r}}$ the denominator of $\mu_\theta^{\hat{r}}$. For any $\epsilon > 0$ let n and δ_n be such that B-4 is verified. Under B-1 and B-2, for every $\theta \in H(\theta_n, \delta_n)$ a second order Taylor expansion of $L_n(\theta)$ around θ_n gives

$$\begin{aligned}
p_{n\theta}\mu_\theta(\theta) &= p_{n\theta}\mu_\theta(\theta_n)\exp(L_n(\theta) - L_n(\theta_n)) \\
&= p_{n\theta}\mu_\theta(\theta_n)\exp\left(-\frac{1}{2}(\theta-\theta_n)'L_n''(\tilde{\theta})(\theta-\theta_n)\right) \\
&= p_{n\theta}\mu_\theta(\theta_n)\exp\left(-\frac{1}{2}(\theta-\theta_n)'[I + \Psi_n(\tilde{\theta})]\Delta_n^{-1}(\theta-\theta_n)\right)
\end{aligned}$$

with $\Psi_n(\tilde{\theta}) = \frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\tilde{\theta}} \left[\frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_n} \right] - I$, I a $k \times k$ identity matrix and $\tilde{\theta}$ lies on the segment joining θ and θ_n. Therefore, under B-4, the probability $\mu_\theta^{\hat{r}}(H(\theta_n, \delta_n)|\hat{r})$ defined as

$$\mu_\theta^{\hat{r}}(H(\theta_n, \delta_n)|\hat{r}) = \int_{H(\theta_n,\delta_n)} \mu_\theta^{\hat{r}}(\theta|\hat{r})d\theta$$

is bounded above by

$$
\begin{aligned}
& |I-G(\epsilon)|^{-1/2}|\Delta_n|^{1/2}D_{\hat{r}}^{-1}p_{n\theta}\mu_\theta(\theta_n)\int_{H(\theta_n,\delta_n)}\exp\left(-\frac{1}{2}(\theta-\theta_n)'[I+\Psi_n(\tilde{\theta})]\Delta_n^{-1}(\theta-\theta_n)\right)d\theta \\
\leq\ & |I-G(\epsilon)|^{-1/2}|\Delta_n|^{1/2}D_{\hat{r}}^{-1}p_{n\theta}\mu_\theta(\theta_n)\int_{H(0,u_n)}e^{-z'z/2}dz
\end{aligned}
$$

where $u_n = \delta_n(1+g^{\max}(\epsilon))^{\frac{1}{2}}/d_{\min}$, $g^{\max}(\epsilon)$ is the largest eigenvalue of $G(\epsilon)$ and $d_{\min}$ is the smallest eigenvalue of Δ_n. The second inequality follows from the fact that, after a change of variable, $\delta_n > (\theta-\theta_n)'(\theta-\theta_n) > ||z||\left(\inf d/(1+eigenvalue(\Psi_n))^{1/2}\right) = ||z||u_n/\delta_n$ so that $H(\theta_n,\delta_n)\subset H(0,u_n)$. In a similar way, under B-4, we can bound $\mu_\theta^{\hat{r}}(H(\theta_n,\delta_n)|\hat{r})$ from below by

$$
\begin{aligned}
& |I+G(\epsilon)|^{-1/2}|\Delta_n|^{1/2}D_{\hat{r}}^{-1}p_{n\theta}\mu_\theta(\theta_n)\int_{H(\theta_n,\delta_n)}\exp\left(-\frac{1}{2}(\theta-\theta_n)'[I+\Psi_n(\tilde{\theta})]\Delta_n^{-1}(\theta-\theta_n)\right)d\theta \\
\geq\ & |I+G(\epsilon)|^{-1/2}|\Delta_n|^{1/2}D_{\hat{r}}^{-1}p_{n\theta}\mu_\theta(\theta_n)\int_{H(0,l_n)}e^{-z'z/2}dz
\end{aligned}
$$

where $l_n = \delta_n(1-g^{\max}(\epsilon))^{\frac{1}{2}}/d_{\max}$, $d_{\max}$ is the largest eigenvalue of Δ_n and $H(\theta_n,\delta_n)\supset H(0,l_n)$. Under B-3, $u_n, l_n\to\infty$ as $n\to\infty$. Therefore,

$$
\begin{aligned}
|I-G(\epsilon)|^{1/2}\lim_{n\to\infty}\mu_\theta^{\hat{r}}(H(\theta_n,\delta_n)|\hat{r}) &\leq |2\pi|^{k/2}|\Delta_n|^{1/2}D_{\hat{r}}^{-1}\lim_{n\to\infty}p_{n\theta}\mu_\theta(\theta_n) \\
&\leq |I+G(\epsilon)|^{1/2}\lim_{n\to\infty}\mu_\theta^{\hat{r}}(H(\theta_n,\delta_n)|\hat{r})
\end{aligned}
$$

and (3.4) is implied by the facts that under B-3, $|I\pm G(\epsilon)|\to 1$ as $\epsilon\to 0$ and $\mu_\theta^{\hat{r}}(H(\theta_n,\delta_n)|\hat{r})\leq 1$ for every n. The equality holds if and only if $\lim_{n\to\infty}\mu_\theta^{\hat{r}}(H(\theta_n,\delta_n)|\hat{r}) = 1$ in $P_{n,*}^{f_*}$-probability, which is assured under A-1, A-2 and A-3 by theorem 3.2.

□

Theorem 3.3. *Assume that A-1, A-2, A-3, B-1, B-2, B-3 and B-4 hold. Then, for every $\theta_1,\theta_2\in\Theta$,*

$$
\int_{J_{\theta_1,\theta_2}} d\mu_\theta^{\hat{r}}(\theta|\hat{r}) \to \int_{\theta_1}^{\theta_2}\phi(u)du
$$

in $P_{n,*}^{f_*}$-probability, where $\phi(\cdot)$ denotes the standard Normal pdf and $J_{\theta_1,\theta_2} := \{\theta;\ \Delta_n^{-1/2}(\theta-\theta_n)\in(\theta_1,\theta_2)\}$.

Proof. Denote $D_{\hat{r}}$ the denominator of $\mu_\theta^{\hat{r}}$. For any $\theta_1,\theta_2\in\Theta$ we write $\theta_2\geq\theta_1$ (or $\theta_2\leq\theta_1$) if every component of $\theta_2-\theta_1$ is nonnegative. Let $Z\sim\mathcal{N}(0,1)$; as stated in the proof of

Theorem 2.1 in Chen (1985) it is sufficient to show that for every $\theta_1 \leq 0$ and $\theta_2 \geq 0$, the probability $\mu_\theta^{\hat{r}}([\theta_1, \theta_2]|\hat{r})$ ($\equiv \mu_\theta^{\hat{r}}(\theta_1 \leq \theta \leq \theta_2|\hat{r})$) converges to $\Phi((\theta_1, \theta_2))$ in $P_{n,*}^{f_*}$-probability, where $\Phi(\cdot)$ denotes the *cdf* of a $\mathcal{N}(0,1)$ distribution.

For sufficiently large n, $J_{\theta_1,\theta_2} \subset H(\theta_n, \delta_n)$ by B-3. Similarly as in the proof of lemma 3.1 the probability

$$\mu_\theta^{\hat{r}}(J_{\theta_1,\theta_2}|\hat{r}) \equiv \int_{J_{\theta_1,\theta_2}} d\mu_\theta^{\hat{r}}(\theta|\hat{r})$$

is upper bounded by

$$|I - G(\epsilon)|^{-1/2}|\Delta_n|^{1/2} D_{\hat{r}}^{-1} p_{n\theta}\mu_\theta(\theta_n) \int_{H^+} e^{-z'z/2}dz$$

with $H^+ := \{z; [I + G(\epsilon)]^{1/2}\theta_1 \leq z \leq \theta_2[I + G(\epsilon)]^{1/2}\}$ and lower bounded by

$$|I + G(\epsilon)|^{-1/2}|\Delta_n|^{1/2} D_{\hat{r}}^{-1} p_{n\theta}\mu_\theta(\theta_n) \int_{H^-} e^{-z'z/2}dz$$

with $H^- := \{z; [I - G(\epsilon)]^{1/2}\theta_1 \leq z \leq \theta_2[I - G(\epsilon)]^{1/2}\}$. By letting $\epsilon \to 0$ we have that $|I \pm G(\epsilon)| \to 1$ and

$$\lim_{n\to\infty} \mu_\theta^{\hat{r}}(J_{\theta_1,\theta_2}|\hat{r}) = \lim_{n\to\infty} |\Delta_n|^{1/2} D_{\hat{r}}^{-1} p_{n\theta}\mu_\theta(\theta_n) \int_{\theta_1}^{\theta_2} e^{-z'z/2}dz.$$

Finally, by the results of lemma 3.1 and theorem 3.2, $\lim_{n\to\infty} |\Delta_n|^{1/2} D_{\hat{r}}^{-1} p_{n\theta}\mu_\theta(\theta_n) = |2\pi|^{-k/2}$ in $P_{n,*}^{f_*}$-probability so that

$$\lim_{n\to\infty} \mu_\theta^{\hat{r}}(J_{\theta_1,\theta_2}|\hat{r}) = \Phi((\theta_1, \theta_2))$$

in $P_{n,*}^{f_*}$-probability.

$\square$

4 Implementation

In this section we show, through the illustration of several examples, how our method can be implemented in practice. We start with toy examples that can be treated also with non-parametric priors different from the Gaussian prior. The interest in using Gaussian priors will be made evident in the more complicated examples where there are overidentifying restrictions which we show can be easily dealt with by using Gaussian priors.

4.1 Just identification and prior on θ through μ_f

Let the parameter θ of interest be the population mean with respect to f, that is, $\theta = \int xf(x)dx$ and $h(\theta, x) = (\theta - x)$. This example considers the just identified case where the prior on θ is deduced from the prior distribution of f, denoted by μ_f. The prior μ_f is a Gaussian measure which is unrestricted except for the fact that it must generate trajectories that integrate to 1 almost surely. To guarantee that, the prior mean function f_0 must be a *pdf* and the prior covariance operator Ω_0 must be such that $\Omega_0^{\frac{1}{2}}1 = 0$. Summarizing, the Bayesian experiment is

$$\begin{cases} f & \sim \quad \mu_f \sim \mathcal{N}(f_0, \Omega_0), \qquad \Omega_0^{\frac{1}{2}}1 = 0 \\ \hat{r}|f & \sim \quad P^f \sim \mathcal{N}(Kf, \Sigma_n). \end{cases} \tag{4.1}$$

This implies a prior and posterior distribution for θ as the following lemma states.

Lemma 4.1. *The Bayesian experiment (4.1) implies that the prior distribution for $\theta = \int xf(x)dx$ is Gaussian with mean $< f_0, x >$ and variance $< \Omega_0 x, x >$ and its posterior distribution is*

$$\theta|\hat{r} \sim \mathcal{N}(< f_0, x > + < \Omega_0 K^* C_n^{-1}(\hat{r} - Kf_0), x >, < [\Omega_0 - \Omega_0 K^* C_n^{-1} K \Omega_0]x, x >)$$

where $C_n^{-1} = (n^{-1}\Sigma + K\Omega_0 K^*)^{-1}$

This approach is appealing because it avoids the specification of two prior distributions while keeping the specification of the sampling distribution completely nonparametric. The prior is specified for the parameter with the highest dimension, that is f, and it implies a prior on the parameter θ.

We illustrate now how to construct in practice the covariance operator Ω_0 in (4.1). Let us suppose that $m = 1$, $S = [-1, 1]$ and Π be the Lebesgue measure. Then, the Legendre polynomials $\{P_n\}_{n\geq 0}$ are suitable to construct the eigenfunctions of Ω_0. The first few Legendre polynomials are $\{1, x, (3x^2 - 1)/2, (5x^3 - 3x)/2, \ldots\}$ and an important property of these polynomials is that they are orthogonal with respect to the L^2 inner product on $[-1, 1]$: $\int_{-1}^{1} P_m(x)P_n(x)dx = 2/(2n + 1)\delta_{mn}$, where δ_{mn} is equal to 1 if $m = n$ and to 0 otherwise. Moreover, the Legendre polynomial obey the recurrence relation $(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x)$ which is useful for computing Ω_0 in practice. The normalized Legendre polynomials form a basis for $L^2[-1, 1]$ so that we can construct the operator Ω_0 as

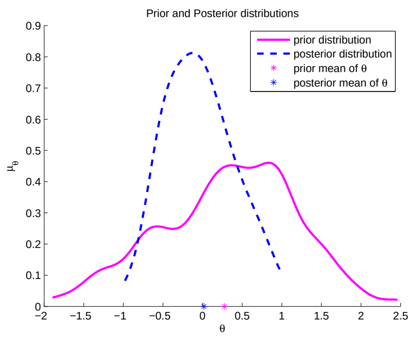
$$\Omega_0 \cdot = \sigma_0 \sum_{n=0}^{\infty} \lambda_n \frac{2n+1}{2} < P_n, \cdot > P_n$$

where $\lambda_0 = 0$ and the $\{\lambda_n,\ n \geq 1\}$ can be chosen in an arbitrary way provided that $\sum_{j\geq 1} \lambda_j < \infty$. The constant σ_0 can be set to an arbitrary value and has the purpose of tuning the size of the prior covariance. This construction of Ω_0 and the fact that f_0 is a *pdf* guarantee that the prior distribution generates functions that integrate to 1 almost surely.

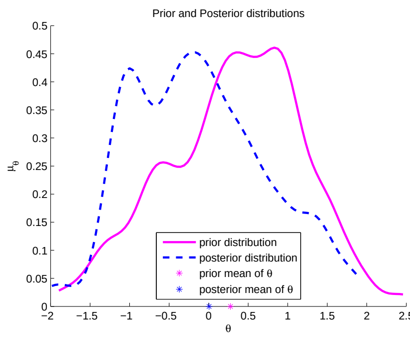
In our simulation exercise we generate n *i.i.d.* observations $(x_1, \ldots, x_n)$ from a $\mathcal{N}(0,1)$ distribution truncated to the interval $[-1,1]$ and construct the function $\hat{r}(t) = n^{-1}\sum_{i=1}^{n} 1\{x_i \leq t\}$ as the empirical *cdf*. Thus, the operators K and K^* take the form

$$\forall \phi \in \mathcal{E},\;\; K\phi = \int_{-1}^{1} 1\{x \leq t\} f(x) dx \quad \text{and } \forall \psi \in \mathcal{F}, \qquad K^*\psi = \int_{-1}^{1} 1\{x \leq t\}\psi(t) dt.$$

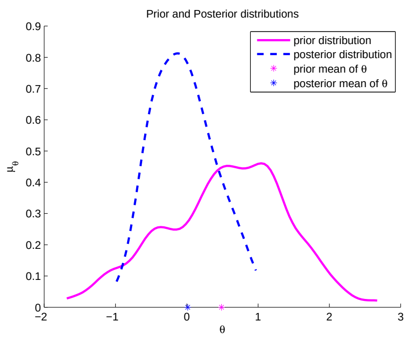
The eigenfunction of Ω_0 are set equal to the normalized Legendre polynomials $\{\sqrt{(2n+1)/2}P_n\}_{n\geq 0}$, the eigenvalues are set equal to $\sigma_0\lambda_n = 5 * n^{-a}$ for $n \geq 1$ and $a > 1$. The prior mean function f_0 is set equal to a $\mathcal{N}(\varrho, 1)$ distribution truncated to the interval $[-1,1]$. We show in Figure 1 the prior and posterior distribution of θ approximated by using a kernel smoothing and 100 drawings from the prior and posterior, respectively. The pictures are obtained for different values of ϱ and α. We also show the prior mean (magenta asterisk) and the posterior mean of θ (blue asterisk).



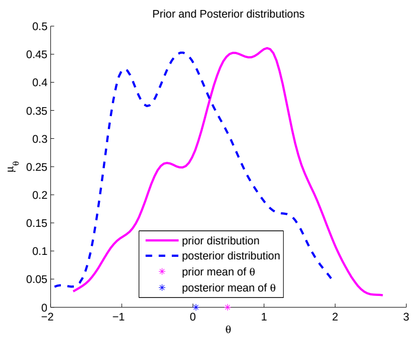
(a) $\mathbf{E}(\theta) = 0.276$, $\mathbf{E}(\theta|x^{(n)}) = 0.0173$, $f_0 = \mathcal{N}(0,1,-1,1)$, $\alpha = 0.01$ and $a = 1.1$.



(b) $\mathbf{E}(\theta) = 0.276$, $\mathbf{E}(\theta|x^{(n)}) = 0.0057$, $f_0 = \mathcal{N}(0,1,-1,1)$, $\alpha = 0.3$ and $a = 1.1$.



(c) $\mathbf{E}(\theta) = 0.4879$, $\mathbf{E}(\theta|x^{(n)}) = 0.0129$, $f_0 = \mathcal{N}(2,1,-1,1)$, $\alpha = 0.01$ and $a = 1.1$.



(d) $\mathbf{E}(\theta) = 0.4879$, $\mathbf{E}(\theta|x^{(n)}) = 0.0439$, $f_0 = \mathcal{N}(2,1,-1,1)$, $\alpha = 0.3$ and $a = 1.1$.

Figure 1: Prior and Posterior distribution and prior and posterior mean of θ. The true value of θ is 0.

4.2 Just identification and prior on θ

We consider the same framework as in the previous example where the parameter θ of interest is the population mean, that is, $\theta = \int x f(x)dx$ and $h(\theta, x) = (\theta - x)$ but now we are going to specify a joint proper prior distribution on (θ, f). We specify a marginal prior μ_θ on θ and a conditional prior on f given θ. While the first one can be arbitrarily chosen, the latter is specified as a Gaussian distribution constrained to generate functions that integrate to 1 and that have mean equal to θ almost surely. In particular, the prior mean function $f_{0\theta}$ must be a *pdf* and $\int x f_{0\theta}(x)dx = \theta$ must hold. The prior covariance operator Ω_0 must be such that $\Omega_0^{\frac{1}{2}} 1 = 0$ and $\Omega_0^{\frac{1}{2}} x = 0$. Together with the constraint on $f_{0\theta}$, the first constraint on Ω_0 guarantees that the trajectories of f generated by this prior integrate to 1 a.s. while the second one guarantees that $\int x f(x)dx = \theta$ a.s. Summarizing, the Bayesian experiment is

$$\begin{cases} \theta & \sim \quad \mu_\theta \\ f|\theta & \sim \quad \mu_f^\theta \sim \mathcal{N}(f_{0\theta}, \Omega_{0\theta}), \qquad \int x f_{0\theta}(x)dx = \theta \quad \text{and} \quad \Omega_0^{\frac{1}{2}}(1, x)' = 0 \\ \hat{r}|f & \sim \quad P^f \sim \mathcal{N}(Kf, \Sigma_n). \end{cases} \tag{4.2}$$

Compared to the approach in section 4.1, this approach allows to incorporate easily any prior information that an economist may have about θ. In fact, taking into account the information on θ through the prior distribution of f is complicated while to incorporate such an information directly in the prior distribution of θ results to be very simple.

Let us suppose that $m = 1$, $S = [-1, 1]$ and Π be the Lebesgue measure. Then, the covariance operator $\Omega_{0\theta}$ can be constructed in the same way as proposed in section 4.1 since the second Legendre polynomial $P_1(x) = x$ allows to implement the constraint on θ. The only difference concerns the number λ_1 which has to be equal to 0 in this case. Therefore, we construct the operator $\Omega_{0\theta}$ as:

$$\Omega_{0\theta} \cdot = \sigma_0 \sum_{n=2}^{\infty} \lambda_n \frac{2n+1}{2} < P_n, \cdot > P_n$$

where the λ_j, $j \geq 2$ can be chosen in an arbitrary way provided that $\sum_{j\geq 2} \lambda_j < \infty$. The constant σ_0 can be set to an arbitrary value and has the purpose of tuning the size of the prior covariance.

Many orthogonal polynomials are suitable for the construction of $\Omega_{0\theta}$ and they may be used to treat cases where S is different from $[-1, 1]$. Consider for instance the case $S = \mathbb{R}$, then, a suitable choice is the basis made of the Hermite polynomials. The Hermite polynomials $\{He_n\}_{n\geq 0}$ form an orthogonal basis of the Hilbert space $L^2(\mathbb{R}, \mathfrak{B}_S, \Pi)$ where $d\Pi(x) = e^{-x^2/2}dx$. It turns out that f will be the density of F with respect to Π and $f_{0\theta}$ the density of another probability measure with respect to Π instead of with respect to the

Lebesgue measure. The first Hermite polynomials are $\{1, x, (x^2-1), (x^3-3x), (x^4-6x^2+3), \ldots\}$ so that we can construct an $\Omega_{0\theta}$ that satisfies the constraints by setting $\lambda_0 = \lambda_1 = 0$ in the following way

$$\Omega_{0\theta}\cdot = \sum_{n=2}^{\infty} \lambda_n \frac{1}{\sqrt{2\pi} n!} < He_n, \cdot > He_n.$$

We performed two simulations exercise: one uses the Legendre polynomial and one makes use of Hermite polynomials. In both the simulations we use the empirical cumulative distribution function to construct $\hat{r}$: $\hat{r}(t) = n^{-1}\sum_{i=1}^{n} 1\{x_i \leq t\}$. In the first simulation, we generate n *i.i.d.* observations $(x_1, \ldots, x_n)$ from a $\mathcal{N}(0,1)$ distribution truncated to the interval $[-1,1]$ as in section 4.1. The prior distribution for θ is uniform over the interval $[-1,1]$. The prior mean function $f_{0\theta}$ is fixed equal to the *pdf* of a $\mathcal{N}(\theta, 1)$ distribution truncated to the interval $[-1,1]$. The covariance operator $\Omega_{0\theta}$ is constructed by using the Legendre polynomials and $\lambda_n = n^{-1.1}$.

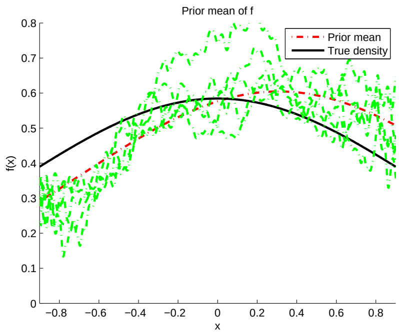
We represent in Figure 2a draws from the conditional prior distributions of f given θ (blue dashed-dotted line) together with the true f that has generated the data (black line) and the prior mean (dashed red line). Figure 2b shows draws from the conditional posterior distribution of f given θ (blue dashed-dotted line) together with the true f that generates the data (black line) and the posterior mean (dashed red line). Lastly, Figure 2c shows the posterior distribution of θ (marginalized with respect to f) approximated by using a kernel smoothing and 1000 drawings from the posterior together with the posterior mean of θ. All the pictures in Figure 2 are obtained for $\sigma_0 = 20$ and $\alpha = 0.1$. The posterior distribution of θ is obtained by integrating out f from the sampling distribution in the following way

$$\begin{cases} \theta & \sim & U[-1,1] \\ \hat{r}|\theta & \sim & \mathcal{N}(Kf_{0\theta}, \Sigma_n + K\Omega_{0\theta}K^*). \end{cases}$$

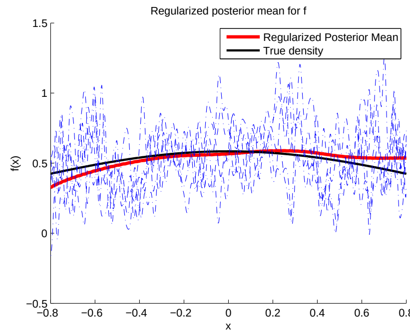
The posterior distribution of θ cannot be computed in a closed-form but we can easily simulate from it by using a *Metropolis-Hastings algorithm*, see for instance Robert (2002). To implement this algorithm we selected, as auxiliary distribution, a uniform distribution over $[-1-\theta, 1+\theta]$.

4.3 Overidentified case

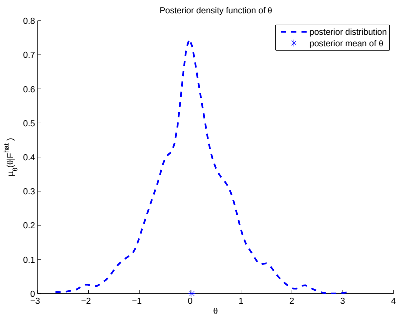
Let us consider the case in which the one-dimensional parameter of interest θ is characterized by the moment conditions $\mathbf{E}^F(h(\theta, x)) = 0$ with $h(\theta, x) = (\theta - x, \theta^2 - \frac{x^2}{2})'$. For instance, this arises when the true data generating process F is an exponential distribution with parameter θ. We specify a prior distribution for (θ, f). The prior μ_θ is chosen arbitrarily provided that the potential constraint on θ are satisfied. These are essentially constraint on the support of θ. The moment conditions affect the conditional prior distribution of f



(a) Draw from the conditional prior distribution μ_f^θ, Prior mean and true f, $\alpha = 0.1$ and $a = 1.1$.



(b) Draw from the conditional posterior distribution μ_f^θ, posterior mean and true f, $\alpha = 0.1$ and $a = 1.1$.



(c) Posterior distribution and mean of θ and $a = 1.1$.

Figure 2: Prior and posterior distribution of f and posterior distribution of θ. The true value of θ is 0.

conditionally on θ. This is a Gaussian distribution with mean function $f_{0\theta}$ whatever *pdf* with the same support as F that satisfies $\int x f_{0\theta}(x) d\Pi(x) = \theta$ and $\int x^2 f_{0\theta}(x) d\Pi(x) = 2\theta^2$. The covariance operator $\Omega_{0\theta}$ of f must be such that

$$\Omega_{0\theta}^{\frac{1}{2}} \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix} = 0. \tag{4.3}$$

In our simulation exercise we take $S = \mathbb{R}_+$ and $d\Pi(x) = e^{-x}dx$. We generate $N = 1000$ observations $x_1, \ldots, x_N$ independently from an exponential distribution with parameter $\theta_* = 2$. Therefore, the true f associated with this DGP is $f_*(x) = \theta_* e^{-(\theta_* - 1)x}$ which obviously satisfies the moment restrictions. The marginal prior distribution μ_θ for θ is a chi-squared distribution with 1 degree of freedom and, for every value of θ drawn from this μ_θ, the prior mean function $f_{0\theta}$ is fixed equal to $f_{0\theta} = \frac{1}{\theta} e^{-(1-\theta)x/\theta}$. We fix the eigenfunctions of $\Omega_{0\theta}$ proportional to the Laguerre polynomials $\{L_n\}_{n \geq 0}$. The first few Laguerre polynomials are $\{1, (1-x), \frac{1}{2}(x^2 - 4x + 2), \frac{1}{6}(-x^3 + 9x^2 - 18x + 6), \ldots\}$ and they are orthogonal in $L^2(\mathbb{R}_+, e^{-x})$. Remark that $x = L_0 - L_1$ and $\frac{x^2}{2} = L_2 - 2L_1 + L_0$. Therefore we construct the operator $\Omega_{0\theta}$ as:

$$\Omega_{0\theta}\cdot = \sigma_0 \sum_{n=0}^{\infty} \lambda_n < L_n, \cdot > L_n$$

with $\lambda_0 = \lambda_1 = \lambda_2 = 0$ to guarantee that (4.3) holds. The constant σ_0 and the λ_n, $n \geq 3$ can be arbitrarily set provided that $\sum_{n\geq 3} \lambda_n < \infty$. In our simulation exercise we take $\sigma_0 = 1$ and $\lambda_n = n^{-1.1}$ for $n \geq 3$.

We represent in Figure 3a draws from the conditional prior distributions of f given θ (blue dashed-dotted line) together with the true f_* that has generated the data (black line) and the prior mean (dashed red line). Figure 3b shows draws from the conditional posterior distribution of f given θ (blue dashed-dotted line) together with the true f_* having generated the data (black line) and the posterior mean (dashed red line). Lastly, Figure 3c shows the posterior distribution of θ (marginalized with respect to f) approximated by using a kernel smoothing and 1000 drawings from the posterior distribution together with the posterior mean of θ. All the pictures in Figure 3 are obtained for $\sigma_0 = 1$ and $\alpha = 0.1$. The posterior distribution of θ is obtained by integrating out f from the sampling distribution in the following way

$$\left\{ \begin{array}{lcl} \theta & \sim & \chi_1^2 \\ \hat{r}|\theta & \sim & \mathcal{N}(Kf_{0\theta}, \Sigma_n + K\Omega_{0\theta}K^*). \end{array} \right.$$

As the posterior distribution of θ cannot be computed in a closed-form we have used a *Metropolis-Hastings algorithm* to simulate from it. To implement this algorithm we selected, as auxiliary distribution, a $\chi^2_{\lceil\theta\rceil}$ distribution.

References

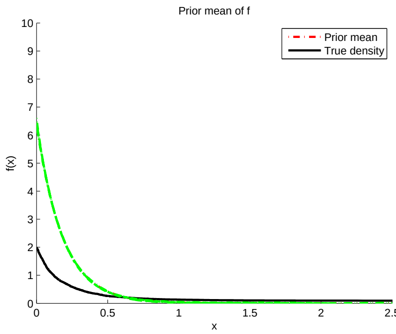
Bickel, P.J. and B.J.K., Kleijn (2012). The seminaprametric Bernstein-von Mises theorem. *Ann. Statist.* **40**, No.1, 206-237.

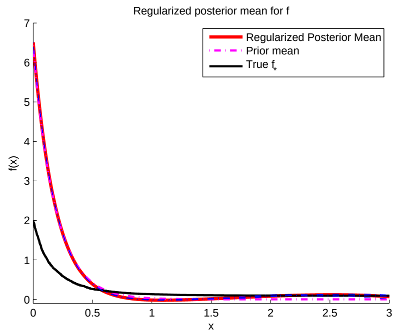
Carrasco, J.P. and M., Carrasco (2012). On the asymptotic efficiency of GMM, mimeo.

Castillo, I. (2012). A semiparametric Bernstein-von Mises theorem for Gaussian process priors. *Prob. Theory Rel. Fields* **152**, No.1, 53-99.

Cazals, C., Fève, F., Fève, P. and J-P., Florens (2004). Simple structural econometrics of price elasticities. *Economic Letter* **86**, 1-6.

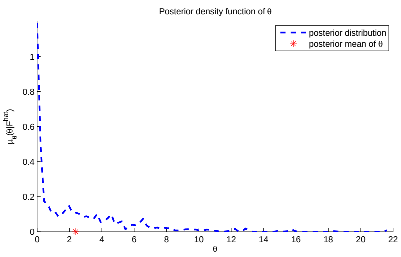
Chen, C.F. (1985). On asymptotic normality of limiting density functions with Bayesian implications. *J. R. Statist. Soc. B* **47**, No.3, pp.540-546.





(a) Draw from the conditional prior distribution μ_f^θ, Prior mean and true f, $\alpha = 0.1$ and $a = 1.1$.

(b) Draw from the conditional posterior distribution μ_f^θ, posterior mean and true f, $\alpha = 0.1$ and $a = 1.1$.



(c) Posterior distribution and mean of θ, $\alpha = 0.1$ and $a = 1.1$.

Figure 3: Prior and posterior distribution of f and posterior distribution of θ. The true value of θ is 2 and the posterior mean is $\mathbf{E}(\theta|X^{(n)}) = 2.3899$.

Cifarelli, D.M. and E., Regazzini (1990). Distribution functions of means of a Dirichlet process. *Ann. Statist.* **18**, No.1, 429-442.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, No.2, 209-230.

Ferguson, T.S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, No.4, 615-629.

Fève, F., Florens, J-P. and S., Richard (2006). Microeconomic demand modelling for price elasticities. In *Lineralization of the Postal and Delivery Sector*, (eds. M.A. Crew and P.R. Kleindorfer), Edouard Elgar, Northampton, USA, 357-368.

Florens, J.P. and J.M., Rolin (1994). Bayes, bootstrap, moments. Institut de Statistique DP9413, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Florens, J.P. and Simoni, A. Regularized posteriors in linear ill-posed inverse problems. *Scandinavian Journal of Statistics*, **39**, No.2, 214-235.

Florens, J.P. and A., Simoni (2012). Regularizing Priors for Linear Inverse Problems, preprint.

Hjort, N.L. and A., Ongaro (2005). Exact inference for random Dirichlet means. *Statistical Inference for Stochastic Processes*, **8**, 227 - 254.

Kim, J.-Y. (2002). Limited information likelihood and Bayesian analysis. *Journal of Econometrics*, **107**, 175-193.

Kitamura, Y. and T. Otsu (2011). Bayesian analysis of moment condition models using nonparametric priors. Mimeo.

Kress, R. (1999), 'Linear Integral Equation', Springer.

Kuo, H.H. (1975), 'Gaussian Measures in Banach Spaces', Springer.

Rivoirard, V. and J., Rousseau (2009). Bernstein-von Mises theorem for linear functionals of the density, preprint, arXiv:math.ST/0908.4167v1.

Robert, C.P. (2002). Méthodes de calcul en analyse Bayésienne. In *Méthodes Bayésiennes en statistique* (eds. J.J. Droesbeke, J. Fine & G. Saporta), 149-181. Editions TECHNIP.

Rolin, J.M. (1992). Some useful properties of the Dirichlet process. Institut de Statistique DP9202, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Rolin, J.M. (1993). On the distribution of jumps of the Dirichlet process. Institut de Statistique DP9302, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Schennach, S.M. (2005). Bayesian exponentially tilted empirical likelihood. *Biometrika*, **1**, 31-46.

Sethuraman, J. (1994). A constructive definition of the Dirichlet prior. *Statistica Sinica*, **2**, 639-650.