



**HAL**  
open science

# Bayesian estimation in a high dimensional parameter framework

Denis Bosq, Maria Dolores Ruiz-Medina

► **To cite this version:**

Denis Bosq, Maria Dolores Ruiz-Medina. Bayesian estimation in a high dimensional parameter framework. *Electronic Journal of Statistics*, 2014, 8, pp.1604-1640. 10.1214/14-EJS935 . hal-01316654

**HAL Id: hal-01316654**

**<https://hal.sorbonne-universite.fr/hal-01316654>**

Submitted on 17 May 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Bayesian estimation in a high dimensional parameter framework\*

Denis Bosq

*LSTA, Université Pierre et Marie Curie-Paris 6*

*4 place Jussieu, 75005, Paris, France*

*e-mail: [denis.bosq@upmc.fr](mailto:denis.bosq@upmc.fr)*

and

María D. Ruiz-Medina

*University of Granada*

*Campus Fuente Nueva s/n E-18071 Granada, Spain*

*e-mail: [mruiz@ugr.es](mailto:mruiz@ugr.es)*

**Abstract:** Sufficient conditions are derived for the asymptotic efficiency and equivalence of componentwise Bayesian and classical estimators of the infinite-dimensional parameters characterizing  $l^2$  valued Poisson process, and Hilbert valued Gaussian random variable models. Conjugate families are considered for the Poisson and Gaussian univariate likelihoods, in the Bayesian estimation of the components of such infinite-dimensional parameters. In the estimation of the functional mean of a Hilbert valued Gaussian random variable, sufficient and necessary conditions, that ensure a better performance of the Bayes estimator with respect to the classical one, are also obtained for the finite-sample size case. A simulation study is carried out to provide additional information on the relative efficiency of Bayes and classical estimators in a high-dimensional framework.

**AMS 2000 subject classifications:** Primary 62M10, 62M20; secondary 62C10, 65F15.

**Keywords and phrases:** Asymptotic relative efficiency, Bayesian estimation, Hilbert valued Gaussian random variable, Hilbert valued Poisson process.

Received May 2013.

## 1. Introduction

In the last two decades, there have been numerous contributions for the statistical inference in function spaces, motivated by the analysis of high-dimensional data (see, for example, [3, 6, 7, 11, 14, 15, 16], among others). We particularly refer to the general overview on statistical estimation in high-dimensional spaces provided in Bosq and Blanke [7], and the references therein. In the cited references, special attention has been paid to the Hilbert valued random variable

---

\*This work has been supported in part by projects MTM2012-32674 of the DGI (co-funded with FEDER funds), MEC, Spain.

context. Specifically, Bosq [6] addresses the problem of infinite-dimensional parameter estimation and prediction of autoregressive Hilbert valued processes, deriving their asymptotic properties. In this framework, alternative projection methodologies have been considered in [1], for wavelet bases, in [4], for spline bases, in [17], for the autocorrelation operator's left and right eigenvector systems, among others. Recent developments in the context of directed acyclic graphs (also known as Bayesian networks) allow the efficient estimation of the adjacency matrix closely related with the influence matrix (see [18]). In particular, penalized likelihood estimation is considered in [19]. These results can be applied to the estimation of the autocorrelation operator eigenvalues (defining the influence matrix) of an autoregressive Hilbertian process. We finally refer to the Bayesian framework adopted by [10] in the statistical analysis of high-dimensional longitudinal, spatial and history data, in terms of suitable priors for the coefficients obtained after projection in general bases, e.g., spline bases. Special attention is paid to the semiparametric regression, and the generalized linear mixed effect model frameworks (see also [5], in relation to Bayesian prediction for stochastic processes).

The present paper provides new developments in relation to the comparison of componentwise Bayesian and classical parameter estimators, in terms of their asymptotic efficiency, for the  $l^2$  valued Poisson process and Hilbert valued Gaussian random variable framework. Specifically, the main goal of the paper is to provide sufficient conditions, that ensure the same asymptotic efficiency of componentwise Bayesian and classical estimators of the infinite-dimensional parameters characterizing such models. For the  $l^2$  valued Poisson process, conjugate Gamma priors on the intensity components are considered for their Bayesian estimation. Theorem 1 provides the asymptotic equivalence of the Bayesian and classical estimators of the infinite-dimensional intensity vector, as well as of the corresponding plug-in predictors of  $l^2$  valued Poisson process.

In the Hilbert valued Gaussian random variable context, suitable ranges of the values of the projections of the mean function, with respect to the auto-covariance eigenvector system, are obtained, for a more efficient performance of the functional mean Bayes estimator with respect to the classical one, when Bayesian estimation is achieved under the Gaussian conjugate family framework, in the finite sample size case. The asymptotic efficiency of Bayesian and classical estimators of the mean in the one-dimensional and infinite-dimensional cases is respectively obtained in Theorems 2 and 3, where the asymptotic equivalence of both estimators is also proved. Gamma priors are considered in the Bayesian estimation of the eigenvalues of the auto-covariance operator of a Hilbert valued Gaussian random variable. Sufficient conditions for the asymptotic equivalence of the Bayesian and classical estimators of the covariance operator are established in Theorem 4, with the explicit derivation of the common limit of their functional mean-square errors as well.

A simulation study is undertaken illustrating the results on asymptotic equivalence of Bayes, and classical estimators of the infinite-dimensional intensity vector, in the  $l^2$  valued Poisson process case. Specifically, the empirical approximations of the functional mean-square errors of Bayes and classical estimators

of the infinite-dimensional vector of intensities are computed. Different rates of convergence of the intensity components are analyzed. The relative empirical efficiency is calculated as well, showing that its value is one for the increasing sequence of sample sizes tested, in all the cases studied. The same statistics are computed for comparing classical and Bayesian estimators of the mean function of a Gaussian Hilbert valued random variable, when different rates of convergence of the eigenvalues of the covariance operator are considered, obtaining similar results in relation to the rate of convergence to zero of the empirical mean-square errors, when the sample size increases, as well as in relation to the empirical relative efficiency, which again is equal to one.

Special attention is devoted to the numerical results obtained in relation to the Bayesian and classical estimation of the covariance operator of a Hilbert valued zero-mean Gaussian random variable, since in that case the empirical results seem to depend on the rate of convergence of the auto-covariance operator eigenvalues. Specifically, for the finite sample sizes and for the truncation orders tested, the equivalence between Bayesian and classical estimators holds when the eigenvalues of the auto-covariance operator display an integer-order polynomial rate of convergence. However, for fractional-order polynomial and negative exponential rates, the relative efficiency of Bayes and classical estimators depends on the truncation order and the finite sample size tested. Namely, classical estimator outperforms the Bayesian estimator in the high-dimensional case corresponding to large truncation orders.

The remainder of the paper is organized as follows: Section 2 provides the fundamental concepts needed in the development of the paper, in relation to Hilbert valued parameter estimation and asymptotic efficiency. Section 3 shows the asymptotic equivalence of Bayesian and classical parameter estimators of the infinite-dimensional intensity vector of the  $l^2$  valued Poisson process. The problem of estimation of the functional mean from the observation of a sample of independent and identically distributed (i.i.d.) Hilbert valued Gaussian random variables is addressed in Section 4.2, in a Bayesian and classical framework. Sufficient conditions for the asymptotic equivalence of the Bayes and classical estimators of the functional mean are derived as well. The Bayesian and classical estimation of the auto-covariance operator is studied in Section 5, where the asymptotic equivalence of both estimators is also established. Finally, a simulation study is undertaken in Section 6 providing additional information of the relative efficiency of Bayes and classical parameter estimators, when truncation is performed, and an increasing sequence of finite functional sample sizes is tested, for different rate of convergence to zero on the components of the infinite-dimensional parameters approximated.

## 2. Estimation and prediction in Hilbert spaces

In the following all the functions are supposed to be measurable and defined on a basic probability space  $(\Omega, \mathcal{A}, P)$ . Also, given some real separable Hilbert space (RSH), say  $G$ , its scalar product will be denoted by  $\langle \cdot, \cdot \rangle_G$  and its norm by  $\| \cdot \|_G$ .

Now, let  $X$  be an observed random variable taking values in some real separable Hilbert space  $H$ . The distribution  $P_\theta$  of  $X$  depends on an unknown parameter  $\theta \in \Theta \subseteq \Theta_0$ , with  $\Theta_0$  being also a RSH. In this paper we address the problem of approximating  $g(X, \theta)$  from  $s(X)$ . If  $g(X, \theta) = g(\theta)$ , we will refer to it as an estimation problem. It becomes a prediction problem if  $g(X, \theta) = E_\theta(Y | X)$  where  $Y$  is also an  $H$  valued random variable such that  $E\|Y\|_H^2 < \infty$ , and where  $E_\theta(\cdot | X)$  denotes conditional expectation with respect to  $X$ . It can be seen that the prediction problem consisting of the estimation of the functional response  $Y$  coincides with the problem of approximating  $E_\theta(Y | X)$  (see [7]). Here, we will refer to the prediction problem only in the case of the  $l^2$  valued Poisson process, since the asymptotic properties of the predictor directly follow from the ones derived for the corresponding infinite-dimensional intensity estimator.

In this paper, the quality of the approximation is measuring in the sense of the following preference relation

$$s_1(X) \prec s_2(X) \iff E_\theta \|s_1(X) - g(X, \theta)\|^2 \leq E_\theta \|s_2(X) - g(X, \theta)\|^2, \quad \theta \in \Theta, \tag{1}$$

where the norm  $\|\cdot\|$  is taken in the RSH  $\Theta_0$ , when the estimation problem is addressed, i.e., when  $g(X, \theta) = g(\theta)$ . The norm  $\|\cdot\|$  is taken in the space  $H$ , when the prediction problem is considered, that is, when  $g(X, \theta) = E_\theta(Y | X)$ . However, a somewhat different point of view is to ask *for which values of  $\theta$ ,  $s_1(X)$  is better than  $s_2(X)$ ?* This question will be addressed in Section 4.2. Specifically, sufficient and necessary conditions that ensure that the Bayesian estimator of the infinite-dimensional mean outperforms the classical estimator, in the sense of equation (1), are provided.

Through this paper we will say that two functional estimators,  $s_{1n}(X_1, \dots, X_n)$  and  $s_{2n}(X_1, \dots, X_n)$  are asymptotically equivalent if

$$\begin{aligned} & \lim_{n \rightarrow \infty} nE_\theta \|s_{1n}(X_1, \dots, X_n) - g_n(X_1, \dots, X_n, \theta)\|^2 \\ & = \lim_{n \rightarrow \infty} nE_\theta \|s_{2n}(X_1, \dots, X_n) - g_n(X_1, \dots, X_n, \theta)\|^2 < \infty. \end{aligned} \tag{2}$$

In this paper, we will refer to the case where  $s_{1n}$  and  $s_{2n}$  are componentwise classical and Bayesian estimators, respectively. In the next section, using criterion (2), we will compare classical and Bayesian estimators of the intensity of  $l^2$  valued Poisson process. As commented, asymptotic comparison of the corresponding plug-in predictors then follows.

### 3. Estimation and prediction of a Poisson process in a Hilbert space

Let  $\{N_{t,j}, t \in \mathbb{R}_+, j \geq 1\}$  be a sequence of independent homogeneous Poisson processes with respective intensities  $\lambda_j > 0, j \geq 1$ , such that  $\sum_j \lambda_j < \infty$ . Since

$$E \left( \sum_j N_{t,j}^2 \right) = \sum_j E(N_{t,j}^2) = \sum_j [\lambda_j t + (\lambda_j t)^2] < \infty,$$

it follows that  $\sum_j N_{t,j}^2 < \infty$  almost surely. Then,  $M_t = \{N_{t,j}, j \geq 0\}$  defines a random variable with values in  $l^2$  and satisfying  $E\|M_t\|^2 < \infty$ . Thus,  $\{M_t, t \geq 0\}$  is a continuous time  $l^2$  valued process.

### 3.1. Classical prediction

Assume that  $M_T$  is observed and we want to predict  $M_{T+h}$  ( $h > 0$ ). For this purpose, we first define a MLE for  $\{\lambda_j, j \geq 1\}$ . Let  $\mathcal{N} \subset l^2$  be the family of sequences  $\{x_j, j \geq 1\} = (x_j)$  such that  $x_j$  is an integer for each  $j$  and  $x_j = 0$ , for sufficiently large  $j$ . We may write

$$\mathcal{N} = \bigcup_k \mathcal{N}_k,$$

where  $\mathcal{N}_k = \{(x_j) : (x_1, \dots, x_k) \in \mathbb{N}^k; x_j = 0, j > k\}$ . Clearly,  $\mathcal{N}_k$  is countable for every  $k$ . It then follows that  $\mathcal{N}$  is countable since it is the countable union of countable sets. Thus, one may define the counting measure  $\mu$  on  $\mathcal{N}$  and extend it by setting  $\mu(l^2 - \mathcal{N}) = 0$ . The obtained measure is then  $\sigma$ -finite.

Now  $M_t$  is considered as  $\mathcal{N}$  valued random variable. Actually, since  $N_{t,j}^2$  is an integer,  $\sum_j N_{t,j}^2 < \infty$  (a.s.) implies that  $N_{t,j} = 0$ , a.s. for  $j$  large enough. More precisely: There exists  $\Omega_0$  such that  $P(\Omega_0) = 1$ , and for all  $\omega \in \Omega_0$ , there exists a  $j_0(\omega, \lambda, T)$  such that  $N_{T,j}(\omega) = 0$ , for  $j > j_0$ . Therefore, one may define the likelihood of  $M_T$  with respect to  $\mu$  by setting

$$L(M_T(\omega), \lambda) = \prod_{j=1}^{j_0(\omega, \lambda, T)} \exp(-\lambda_j T) \frac{(\lambda_j T)^{N_{T,j}(\omega)}}{N_{T,j}(\omega)!},$$

hence, the MLE of  $(\lambda)$  is given by:

$$(\hat{\lambda})_T = \left( \frac{N_{T,j}}{T}, j \geq 1 \right) = (\hat{\lambda}_{j,T}, j \geq 1). \quad (3)$$

Clearly,  $\sum_{j=1}^{\infty} \hat{\lambda}_{j,T} < \infty$  (a.s.) and  $(\hat{\lambda})_T$  is unbiased ( $(\lambda)$  being considered as a parameter with values in  $l^2$ ). It follows that

$$f(M_T) = \frac{T+h}{T} M_T \quad (4)$$

is an unbiased predictor of  $M_{T+h}$  and of

$$E_{\lambda}(M_{T+h}|M_T) = h(\lambda) + M_T, \quad (\lambda) \in l^2,$$

since

$$E_{\lambda}[M_{T+h}] = E_{\lambda}[E_{\lambda}(M_{T+h}|M_T)] = (\lambda)(T+h) = E_{\lambda}[f(M_T)].$$

Concerning efficiency, one may write

$$\begin{aligned}
 E_\lambda \|f(M_T) - h(\lambda) - M_T\|^2 &= \sum_j E \left( \frac{T+h}{T} N_{T,j} - h\lambda_j - N_{T,j} \right)^2 \\
 &= \sum_j E \left[ \left( \frac{T+h}{T} - 1 \right) N_{T,j} - h\lambda_j \right]^2 \\
 &= \sum_j \left( \frac{h}{T} \right)^2 E[N_{T,j} - \lambda_j T]^2 = \frac{h^2}{T} \sum_j \lambda_j. \quad (5)
 \end{aligned}$$

Thus,  $f(M_T)$  is an unbiased efficient predictor provided the following definition:  $f(M_T)$  is said to be efficient if and only if each component of  $f(M_T)$  is an efficient predictor of the associated component of  $M_{T+h}$ .

### 3.2. Bayesian prediction

In order to define a Bayesian parameter estimator of  $(\lambda)$ , and a Bayesian predictor of  $M_{T+h}$ , we set the a-priori distribution  $\Gamma(a_j, b_j)$  on each  $\lambda_j$ , with  $a_j, b_j > 0$ , for each  $j \geq 1$ . The posterior distribution is then given by the Gamma distribution  $\Gamma(a_j + N_{T,j}, b_j + T)$ , for each  $j \geq 1$  (see, for example, [13]). Therefore, the Bayesian estimator is

$$\tilde{\lambda}_{j,T} = \frac{a_j + N_{T,j}}{b_j + T}, \quad j \geq 1, \quad (6)$$

and the corresponding predictor

$$f_0(M_T) = h(\tilde{\lambda})_T + M_T, \quad (\tilde{\lambda})_T = (\tilde{\lambda}_{j,T}). \quad (7)$$

This predictor is well defined if  $\sum_j \tilde{\lambda}_{j,T} < \infty$  almost surely, a natural condition (even if  $\sum_j \tilde{\lambda}_{j,T}^2 < \infty$  a.s. should be enough). A sufficient condition for this is

$$\sum_j a_j < \infty,$$

since, clearly,

$$\sum_j \tilde{\lambda}_{j,T} \leq \frac{1}{T} \left( \sum_j a_j + \sum_j N_{T,j} \right) < \infty \quad \text{a.s.}$$

#### 3.2.1. Asymptotic equivalence of classical and Bayesian parameter estimators and predictors

This section provides the asymptotic equivalence of the classical and Bayes parameter estimators and predictors in the sense of equation (2). In particular, the

asymptotic efficiency of the Bayes estimator, and the resulting plug-in predictor is also proved. In the derivation of the next theorem, the following assumption is made.

**Assumption A0.** Consider  $\sum_{j \geq 1} a_j < \infty$ , and  $\sup_{j \geq 1} b_j < \infty$ .

**Theorem 1.** Let  $(\hat{\lambda})_T$  be the classical parameter estimator introduced in (3), and  $(\tilde{\lambda})_T$  be the Bayes estimator of  $(\lambda)$  constructed from (6). As before, we denote by  $f_0(M_T)$  and  $f(M_T)$  the respective Bayesian and classical predictors. Under **Assumption A0**, the following identities hold:

$$\lim_{T \rightarrow \infty} TE \|(\hat{\lambda})_T - (\lambda)\|_{l^2}^2 = \lim_{T \rightarrow \infty} T \sum_{j=1}^{\infty} E[\hat{\lambda}_{j,T} - \lambda_j]^2 < \infty \quad (8)$$

$$\lim_{T \rightarrow \infty} TE \|(\tilde{\lambda})_T - (\lambda)\|_{l^2}^2 = \lim_{T \rightarrow \infty} T \sum_{j=1}^{\infty} E[\tilde{\lambda}_{j,T} - \lambda_j]^2 < \infty \quad (9)$$

$$\lim_{T \rightarrow \infty} TE \|f_0(M_T) - h(\lambda) - M_T\|^2 = \lim_{T \rightarrow \infty} TE \|f(M_T) - h(\lambda) - M_T\|^2 < \infty. \quad (10)$$

**Proof.** From equations (3) and (5), equation (8) follows in a direct way with

$$\lim_{T \rightarrow \infty} TE \|(\hat{\lambda})_T - (\lambda)\|_{l^2}^2 = \sum_{j=1}^{\infty} \lambda_j.$$

Hence, the classical estimator  $(\hat{\lambda})_T$  of  $(\lambda)$  is asymptotically efficient.

Equation (9) is now proved. Specifically, for each  $j \geq 1$ , from equation (6),

$$\begin{aligned} E[\tilde{\lambda}_{j,T} - \lambda_j]^2 &= \frac{1}{(b_j + T)^2} [a_j^2 + \lambda_j T + (\lambda_j b_j)^2 - 2\lambda_j a_j b_j] \\ &= \frac{\lambda_j T}{(b_j + T)^2} + \frac{(a_j - \lambda_j b_j)^2}{(b_j + T)^2} = A_j(T) + B_j(T). \end{aligned} \quad (11)$$

We then study the limit as  $T \rightarrow \infty$  of  $\sum_{j=1}^{\infty} T A_j(T)$ . Since

$$\frac{\lambda_j T^2}{(b_j + T)^2} < \lambda_j,$$

(i.e.,  $T^2 < (b_j + T)^2$ ,  $T > 1$ ,  $b_j > 0$ ), and  $\sum_j \lambda_j < \infty$ , we can applied Dominated Convergence Theorem with dominant summable sequence  $\lambda_j$ ,  $j \geq 1$ , to conclude that, for every  $j \geq 1$ ,

$$\lim_{T \rightarrow \infty} T \sum_{j=1}^{\infty} A_j(T) = \sum_{j=1}^{\infty} \lim_{T \rightarrow \infty} T A_j(T) = \sum_{j=1}^{\infty} \lambda_j < \infty. \quad (12)$$

Secondly, we prove that  $\lim_{T \rightarrow \infty} \sum_{j=1}^{\infty} T B_j(T) = 0$ . Indeed, we have that

$$\sum_{j=1}^{\infty} \frac{(a_j - \lambda_j b_j)^2}{(b_j + T)^2} \leq \frac{1}{T^2} \sum_{j=1}^{\infty} (a_j - \lambda_j b_j)^2$$

$$\leq \frac{1}{T^2} \left( \sum_{j=1}^{\infty} a_j^2 + \left[ \sup_{j \geq 1} b_j \right]^2 \sum_{j=1}^{\infty} \lambda_j^2 - 2 \min_j b_j \sum_{j=1}^{\infty} \lambda_j a_j \right). \tag{13}$$

As there exist positive constants  $c_1$  and  $c_2$  such that  $\sum_{j=1}^{\infty} a_j^2 \leq c_1 \sum_{j=1}^{\infty} a_j < \infty$ ,  $\sum_{j=1}^{\infty} \lambda_j^2 \leq c_2 \sum_{j=1}^{\infty} \lambda_j < \infty$ , and  $\sum_{j=1}^{\infty} a_j \lambda_j \leq \sqrt{\sum_{j=1}^{\infty} a_j^2} \sqrt{\sum_{j=1}^{\infty} \lambda_j^2} < \infty$ , the upper bound in equation (13) is given by  $\frac{C}{T^2}$ , with  $C$  being a finite constant, which leads to  $\lim_{T \rightarrow \infty} \sum_{j=1}^{\infty} T B_j(T) = 0$ .

From equation (11),

$$\lim_{T \rightarrow \infty} TE[\tilde{\lambda}_{j,T} - \lambda_j]^2 = \lambda_j, \quad \forall j \geq 1,$$

and from (12) and (13),

$$\lim_{T \rightarrow \infty} TE_{\lambda} \|(\tilde{\lambda})_T - (\lambda)\|^2 = \sum_{j=1}^{\infty} \lim_{T \rightarrow \infty} TE[\tilde{\lambda}_{j,T} - \lambda_j]^2 = \sum_{j=1}^{\infty} \lambda_j. \tag{14}$$

Therefore, the Bayesian estimator  $(\tilde{\lambda})_T$  of  $(\lambda)$  is asymptotically efficient.

Finally, equation (10) is obtained from the following identities, keeping in mind equation (5)

$$\begin{aligned} & \lim_{T \rightarrow \infty} TE_{\lambda} \|f_0(M_T) - h(\lambda) - M_T\|^2 \\ &= \lim_{T \rightarrow \infty} h^2 \sum_j TE_{\lambda} (\tilde{\lambda}_{j,T} - \lambda_j)^2 = h^2 \sum_j \lambda_j. \end{aligned} \tag{15}$$

#### 4. Bayesian and classical estimation of the mean in the $H$ valued Gaussian random variable framework

Sufficient conditions, for a better performance of the Bayes estimator of the mean of a Hilbert valued Gaussian random variable against the classical one, are derived. Conjugate Gaussian priors are considered in the Bayesian estimation of the mean projections with respect to the eigenvectors of the autocovariance operator. The asymptotic equivalence, in the sense of (2), of the Bayes and classical estimators of the functional mean is also obtained.

##### 4.1. Bayesian and classical mean estimators in the real valued case

Let us first begin with the one-dimensional Gaussian case, where we denote by  $\bar{X}_n$  and  $\tilde{\theta}_n$  the classical and Bayesian estimators of the mean, respectively. Let also  $X_1, \dots, X_n$  be a sample of independent and identically distributed real valued Gaussian random variables with mean  $\theta_1$  and variance  $\lambda_1 > 0$ . That is,  $X_i \sim \mathcal{N}(\theta_1, \lambda_1)$ ,  $i = 1, \dots, n$ . In the Bayesian estimation, we consider as prior for  $\theta_1$ , a zero-mean Gaussian distribution, i.e.,  $\theta_1 \sim \mathcal{N}(0, \gamma_1)$ ,  $\gamma_1 > 0$ . Then, the

Bayesian estimator of  $\theta_1$ , the mean of the a-posteriori distribution, is given by

$$\tilde{\theta}_{1n} = \frac{\bar{X}_n}{1 + \frac{1}{n} \frac{\lambda_1}{\gamma_1}} := \alpha_n \bar{X}_n$$

(see [13], p. 234). The values of  $\theta_1$  for which the Bayesian estimator outperforms the classical one are given in the following lemma.

**Lemma 1.** *For a finite sample size  $n$ ,*

$$E_{\theta_1}(\tilde{\theta}_{1n} - \theta_1)^2 \leq V_{\theta_1}(\bar{X}_n) = \frac{\lambda_1}{n} \Leftrightarrow |\theta_1| \leq \left( \frac{\lambda_1}{n} + 2\gamma_1 \right)^{1/2}. \quad (16)$$

**Proof.** By direct computation

$$E_{\theta_1}(\tilde{\theta}_{1n} - \theta_1)^2 = \alpha_n^2 \left[ \frac{\lambda_1}{n} + \theta_1^2 \right] + \theta_1^2 - 2\alpha_n \theta_1^2. \quad (17)$$

Replacing, in equation (17),

$$\alpha_n = \frac{\gamma_1}{\gamma_1 + \frac{\lambda_1}{n}},$$

the condition  $E_{\theta_1}(\tilde{\theta}_{1n} - \theta_1)^2 \leq V_{\theta_1}(\bar{X}_n) = \frac{\lambda_1}{n}$  is equivalent to

$$\gamma_1^2 \frac{\lambda_1}{n} + \gamma_1^2 \theta_1^2 + \theta_1^2 \left[ \gamma_1 + \frac{\lambda_1}{n} \right]^2 - 2\gamma_1 \left[ \gamma_1 + \frac{\lambda_1}{n} \right] \theta_1^2 \leq \left[ \gamma_1 + \frac{\lambda_1}{n} \right]^2 \left[ \frac{\lambda_1}{n} \right]. \quad (18)$$

That is, denoting  $x = \lambda_1/n$ , equation (18) holds if and only if

$$\theta_1^2 \leq \frac{(\gamma_1 + x)^2 x - \gamma_1^2 x}{\gamma_1^2 + [\gamma_1 + x]^2 - 2\gamma_1[\gamma_1 + x]} \Leftrightarrow |\theta_1| \leq \sqrt{x + 2\gamma_1},$$

which yields to the desired condition in terms of the upper bound  $\sqrt{\frac{\lambda_1}{n} + 2\gamma_1}$ .

The asymptotic equivalence in the sense of (2) of  $\bar{X}_n$  and  $\tilde{\theta}_{1n}$  is now established in Theorem 2.

**Theorem 2.** *The following identities hold:*

$$\lim_{n \rightarrow \infty} nE_{\theta_1}|\bar{X}_n - \theta_1|^2 = \lambda_1 = \lim_{n \rightarrow \infty} nE_{\theta_1}|\tilde{\theta}_{1n} - \theta_1|^2. \quad (19)$$

**Proof.** First, we recall

$$E_{\theta_1}|\bar{X}_n|^2 = \frac{\lambda_1}{n} + \theta_1^2,$$

and

$$E_{\theta_1}|\bar{X}_n - \theta_1|^2 := \frac{\lambda_1}{n}.$$

Hence,

$$nE_{\theta_1}|\bar{X}_n - \theta_1|^2 = \lambda_1.$$

Now

$$nE_{\theta_1}|\tilde{\theta}_{1n} - \theta_1|^2 = \lambda_1\alpha_n^2 + n(1 - \alpha_n)^2\theta_1^2 := A_n + nB_n,$$

where  $A_n \rightarrow \lambda_1$ , as  $n \rightarrow \infty$ , and

$$\begin{aligned} B_n &= (1 - \alpha_n)^2\theta_1^2 = \frac{\left(\frac{\lambda_1}{n\gamma_1}\right)^2}{\left(\frac{n\gamma_1 + \lambda_1}{n\gamma_1}\right)^2}\theta_1^2 = \frac{(\lambda_1\theta_1)^2}{(n\gamma_1 + \lambda_1)^2} = \frac{(\lambda_1\theta_1)^2}{(n\gamma_1)^2 + \lambda_1^2 + 2n\gamma_1\lambda_1} \\ &\leq \frac{1}{n^2} \frac{(\lambda_1\theta_1)^2}{\gamma_1^2} = \mathcal{O}\left(\frac{1}{n^2}\right). \end{aligned} \tag{20}$$

Thus, equation (19) holds.

#### 4.2. Bayesian and classical estimators of the functional mean

Denote, as before, by  $H$  a real separable Hilbert space equipped with its scalar product  $\langle \cdot, \cdot \rangle_H$ , and its norm  $\|\cdot\|_H$ . Let us now consider  $\{X_n, n \geq 1\}$  to be a sequence of i.i.d  $H$  valued Gaussian random variables such that

$$E_{\theta}(X_n) = \theta, \quad \theta \in H,$$

and its auto-covariance operator  $C$  is given by

$$C = E_{\theta}((X_n - \theta) \otimes (X_n - \theta)), \tag{21}$$

where  $h \otimes g$  denotes the tensor product of functions  $h$  and  $g$  in  $H$ , defining a Hilbert-Schmidt operator on  $H$  as follows:

$$f \otimes g(h) = f \langle g, h \rangle_H, \quad \forall h \in H. \tag{22}$$

We suppose that  $C$  does not depend on  $\theta$ . From equations (21) and (22), it defines a Hilbert-Schmidt operator on  $H$  admitting a spectral decomposition in terms of a pure point spectrum  $\{\lambda_j, j \geq 1\}$ , with  $\lambda_j \geq 0, j \geq 1$ , and an associated system of eigenvectors  $\{v_j, j \geq 1\}$  in  $H$  satisfying

$$Cv_j = \lambda_j v_j, \quad j \geq 1.$$

Hence, from Hilbert-Schmidt Theorem,  $C$  can be represented as

$$C = \sum_{j=1}^{\infty} \lambda_j (v_j \otimes v_j), \tag{23}$$

in the sense that

$$C(h) = \sum_H \sum_{j=1}^{\infty} \lambda_j \langle v_j, h \rangle_H v_j, \quad \forall h \in H$$

(see [9], pp. 119 and 126). In the following we also assume that  $C$  is in the trace class, i.e.,  $\sum_{j=1}^{\infty} \lambda_j < \infty$ . Furthermore, the eigenvectors  $\{v_j\}$  of  $C$  constitute

an orthonormal complete system in  $H$ . Consequently,  $\theta$  admits the following representation in  $H$ :

$$\theta = \sum_{j=1}^{\infty} \langle \theta, v_j \rangle_H v_j = \sum_{j=1}^{\infty} \theta_j v_j.$$

Then,

$$E_{\theta} \|X_n\|_H^2 = \sum_{j=1}^{\infty} E_{\theta} [\langle X_n, v_j \rangle_H]^2 = \sum_{j=1}^{\infty} E_{\theta} [X_{nj}]^2 = \sum_{j=1}^{\infty} (\lambda_j + \theta_j^2) < \infty,$$

and

$$E_{\theta} \|X_n - \theta\|_H^2 = \sum_{j=1}^{\infty} E_{\theta} (X_{nj} - \theta_j)^2 = \sum_{j=1}^{\infty} \lambda_j < \infty,$$

since  $C$  is in the trace class. Note that independence of  $\{X_n, n \geq 1\}$  induces global independence of  $\{X_{nj}, n \geq 1, j \geq 1\}$ .

As an a-priori distribution of the unknown parameter  $\theta$  we consider  $\theta \sim \mathcal{N}(0, \Gamma)$ , where  $\Gamma$  denotes, as usual, the covariance operator of the  $H$  valued random variable  $\theta$ , i.e.,  $\Gamma = E[\theta \otimes \theta]$ . We also assume that  $\Gamma$  admits the following spectral representation in terms of the eigenvectors of the covariance operator  $C$  of  $X$ :

$$\Gamma = \sum_{j=1}^{\infty} \gamma_j (v_j \otimes v_j),$$

i.e.,

$$\Gamma(h) \stackrel{H}{=} \sum_{j=1}^{\infty} \gamma_j \langle v_j, h \rangle_H v_j, \quad \forall h \in H,$$

where  $\gamma_j > 0, j \geq 1$ , are the respective eigenvalues of  $\Gamma$  associated with the eigenvectors  $v_j, j \geq 1$ , satisfying  $\Gamma v_j = \gamma_j v_j, j \geq 1$ . To applying Kolmogorov extension theorem in the definition of the prior  $\theta \sim \mathcal{N}(0, \Gamma)$ , we assume that  $\sum_{j=1}^{\infty} \gamma_j < \infty$  (see, for example, [8], and [12]). Hence,  $\theta_j \sim \mathcal{N}(0, \gamma_j)$ . In order to compute  $E(\theta | X_1, \dots, X_n)$ , we suppose that  $\theta_j \perp (X_{i,j'}, i \geq 1, j' \neq j)$ , which leads to

$$\langle E(\theta | X_1, \dots, X_n), v_j \rangle_H = E(\theta_j | X_1, \dots, X_n) = E(\theta_j | X_{1,j}, \dots, X_{n,j}).$$

We can then perform an independent computation of the respective posterior distributions of the projections  $\theta_j, j \geq 1$ , with respect to the eigenvectors  $\{v_j, j \geq 1\}$ , of parameter  $\theta$ . Thus, for  $j \geq 1$ , since the prior  $\theta_j \sim \mathcal{N}(0, \gamma_j)$  is conjugate with the likelihood  $\bar{X}_{nj} | \theta_j \sim \mathcal{N}(\theta_j, \frac{\lambda_j}{n})$ , the Bayesian estimator  $\tilde{\theta}_{jn}$  of  $\theta_j$  is given by

$$\tilde{\theta}_{jn} = \frac{\bar{X}_{nj}}{1 + \frac{1}{n} \frac{\lambda_j}{\gamma_j}}, \quad j \geq 1, \quad (24)$$

provided that the posterior distribution

$$\theta_j | \bar{X}_{nj} \sim \mathcal{N} \left( \frac{\bar{X}_{nj}}{1 + \frac{1}{n} \frac{\lambda_j}{\gamma_j}}, \frac{\frac{\lambda_j \gamma_j}{n}}{\frac{\lambda_j}{n} + \gamma_j} \right)$$

(see [13], p. 234). Hence,

$$\tilde{\theta}_n = \sum_{j=1}^{\infty} \tilde{\theta}_{jn} v_j.$$

Clearly,

$$\|\tilde{\theta}_n\|_H^2 = \sum_{j=1}^{\infty} \tilde{\theta}_{jn}^2 < \infty \quad \text{a.s.},$$

since  $|\tilde{\theta}_{jn}| \leq |\bar{X}_{nj}|$ , and

$$\begin{aligned} E_{\theta} \left( \sum_{j=1}^{\infty} \bar{X}_{nj}^2 \right) &= \sum_{j=1}^{\infty} E_{\theta} \left( \bar{X}_{nj}^2 \right) = \sum_{j=1}^{\infty} \left( \frac{\lambda_j}{n} + \theta_j^2 \right) \\ &< \sum_{j=1}^{\infty} (\lambda_j + \theta_j^2) = E_{\theta} \|X_n\|_H^2 < \infty. \end{aligned} \tag{25}$$

Let us consider the problem of looking for conditions that provide the values of  $\theta$  for which the Bayes estimator outperforms the classical estimator, in the sense of variance, as given in Lemma 1 in the real valued case. The following proposition addresses this problem, considering additional conditions now formulated:

**Assumption A1.** There exists  $m, M > 0$ , such that

$$0 < m \leq \inf_j \frac{\lambda_j}{\gamma_j} \leq \sup_j \frac{\lambda_j}{\gamma_j} \leq M < \infty. \tag{26}$$

**Assumption A2.**  $\frac{\lambda_j}{\gamma_j} = M = m, j \geq 1$ .

**Proposition 1.** *The following assertions hold for a finite sample size  $n$ :*

(i) *A sufficient condition for  $\tilde{\theta}_n \prec \bar{X}_n$  is*

$$\text{Tr}(\mathcal{A}) = \text{Tr} \left( \Theta^2 - \frac{1}{n} C - 2\Gamma \right) \leq 0, \tag{27}$$

where

$$\begin{aligned} \Theta^2(f) &\stackrel{H}{=} \sum_{j=1}^{\infty} \theta_j^2 \langle v_j, f \rangle_H v_j, \quad \forall f \in H \\ C(f) &\stackrel{H}{=} \sum_{j=1}^{\infty} \lambda_j \langle v_j, f \rangle_H v_j, \quad \forall f \in H \end{aligned}$$

$$\Gamma(f) \stackrel{\text{H}}{=} \sum_{j=1}^{\infty} \gamma_j \langle v_j, f \rangle_H v_j, \quad \forall f \in H,$$

and hence,

$$\begin{aligned} \text{trace}(\Theta^2) &= \sum_{j=1}^{\infty} \theta_j^2 \\ \text{trace}(C) &= \sum_{j=1}^{\infty} \lambda_j \\ \text{trace}(\Gamma) &= \sum_{j=1}^{\infty} \gamma_j. \end{aligned}$$

(ii) Under **Assumption A1**,

$$\tilde{\theta}_n \prec \bar{X}_n \Leftrightarrow \|\theta\|_H^2 \leq \left( \frac{1 + \frac{M}{n}}{\frac{M}{n}} \right)^2 \frac{(1 + \frac{m}{n})^2 - 1}{(1 + \frac{m}{n})^2} \sum_{j=1}^{\infty} \frac{\lambda_j}{n}.$$

(iii) Under **Assumption A2**,

$$\tilde{\theta}_n \prec \bar{X}_n \Leftrightarrow \|\theta\|_H^2 \leq \frac{2}{m} \sum_{j=1}^{\infty} \lambda_j.$$

**Proof.**

(i) First, note that

$$E_{\theta} \|\tilde{\theta}_n - \theta\|_H^2 = \sum_{j=1}^{\infty} E_{\theta} (\tilde{\theta}_{nj} - \theta_j)^2 = \sum_{j=1}^{\infty} \left[ \alpha_{jn}^2 \frac{\lambda_j}{n} + (\alpha_{jn} - 1)^2 \theta_j^2 \right],$$

where

$$\alpha_{jn} = \frac{1}{1 + a_{jn}},$$

and  $a_{jn} = \frac{\lambda_j}{n\gamma_j}$ , for  $j \geq 1$ .

From equation (25),

$$E \|\bar{X}_n - \theta\|_H^2 = \sum_{j=1}^{\infty} \frac{\lambda_j}{n}.$$

For each  $j \geq 1$ , recall that similarly to the one-dimensional case (see Lemma 1):

$$\alpha_{jn}^2 \frac{\lambda_j}{n} + (\alpha_{jn} - 1)^2 \theta_j^2 \leq \frac{\lambda_j}{n} \Leftrightarrow \theta_j^2 \leq \frac{\lambda_j}{n} + 2\gamma_j.$$

In assertion (i) we have used the fact that  $\Theta^2$ ,  $C$  and  $\Gamma$  are trace operators. Specifically,  $\text{trace}(\Theta^2) = \sum_{j=1}^{\infty} \theta_j^2 < \infty$  by Parseval identity, keeping in mind

that  $\theta \in H$ , with  $v_j, j \geq 1$ , being an orthonormal basis of  $H$ . As we consider before,  $\text{trace}(C) = \sum_{j=1}^{\infty} \lambda_j < \infty$ , since  $\sum_{j=1}^{\infty} \lambda_j = E\|X - \theta\|_H^2 < \infty$ . Finally,  $\text{trace}(\Gamma) = \sum_{j=1}^{\infty} \gamma_j < \infty$  as we assumed in the definition of the Gaussian prior distribution for  $H$  valued random variable  $\theta$ . Therefore,  $\mathcal{A} = \Theta^2 - \frac{1}{n}C - 2\Gamma$  is also in the trace class. From the positiveness and trace property of operators  $\Theta^2, C$  and  $\Gamma$  inequality (27) implies

$$\theta_j^2 \leq \frac{\lambda_j}{n} + 2\gamma_j, \quad \forall j \geq 1, \tag{28}$$

which is equivalent (see Lemma 1) to the inequality

$$\alpha_{jn}^2 \frac{\lambda_j}{n} + (\alpha_{jn} - 1)^2 \theta_j^2 \leq \frac{\lambda_j}{n}, \quad j \geq 1,$$

leading to

$$E_{\theta} \|\tilde{\theta}_n - \theta\|_H^2 = \sum_{j=1}^{\infty} \alpha_{jn}^2 \frac{\lambda_j}{n} + (\alpha_{jn} - 1)^2 \theta_j^2 \leq \sum_{j=1}^{\infty} \frac{\lambda_j}{n} = V_{\theta}(\bar{X}_n) = E_{\theta} \|\bar{X}_n - \theta\|_H^2.$$

(ii) Under **Assumption A1**, the following inequalities hold:

$$\frac{1}{1 + \frac{M}{n}} \leq \alpha_{jn} \leq \frac{1}{1 + \frac{m}{n}},$$

and

$$\left(\frac{\frac{m}{n}}{1 + \frac{m}{n}}\right)^2 \leq (\alpha_{jn} - 1)^2 \leq \left(\frac{\frac{M}{n}}{1 + \frac{M}{n}}\right)^2.$$

Therefore,

$$\begin{aligned} E_{\theta} \|\tilde{\theta}_n - \theta\|_H^2 &= \sum_{j=1}^{\infty} \alpha_{jn}^2 \frac{\lambda_j}{n} + (\alpha_{jn} - 1)^2 \theta_j^2 \\ &\leq \sum_{j=1}^{\infty} \left[\frac{1}{1 + \frac{m}{n}}\right]^2 \frac{\lambda_j}{n} + \left(\frac{\frac{M}{n}}{1 + \frac{M}{n}}\right)^2 \theta_j^2. \end{aligned} \tag{29}$$

A sufficient condition can then be derived from (29) by imposing

$$\sum_{j=1}^{\infty} \left[\frac{1}{1 + \frac{m}{n}}\right]^2 \frac{\lambda_j}{n} + \left(\frac{\frac{M}{n}}{1 + \frac{M}{n}}\right)^2 \theta_j^2 \leq \sum_{j=1}^{\infty} \frac{\lambda_j}{n}, \tag{30}$$

which is equivalent to

$$\left(\frac{\frac{M}{n}}{1 + \frac{M}{n}}\right)^2 \sum_{j=1}^{\infty} \theta_j^2 \leq \left(1 - \left[\frac{1}{1 + \frac{m}{n}}\right]^2\right) \sum_{j=1}^{\infty} \frac{\lambda_j}{n}.$$

That is,

$$\begin{aligned} \left(\frac{\frac{M}{n}}{1+\frac{M}{n}}\right)^2 \|\theta\|_H^2 &\leq \frac{(1+\frac{m}{n})^2-1}{(1+\frac{m}{n})^2} \sum_{j=1}^{\infty} \frac{\lambda_j}{n}, \\ \Leftrightarrow \|\theta\|_H^2 &\leq \left(\frac{1+\frac{M}{n}}{\frac{M}{n}}\right)^2 \frac{(1+\frac{m}{n})^2-1}{(1+\frac{m}{n})^2} \sum_{j=1}^{\infty} \frac{\lambda_j}{n} \end{aligned} \quad (31)$$

leading to the sufficient condition for  $\tilde{\theta}_n \prec \bar{X}_n$  formulated in (ii). Note that this condition still depends on  $n$ , since the upper bound (31) decreases with  $n$ , i.e.,

$$\frac{1}{n} \left(\frac{1+\frac{M}{n}}{\frac{M}{n}}\right)^2 \frac{(1+\frac{m}{n})^2-1}{(1+\frac{m}{n})^2} = \left(\frac{1}{n} + \frac{n}{M^2} + \frac{2}{M}\right) \frac{m^2+2nm}{m^2+2nm+n^2}.$$

(iii) To find an upper bound independently of  $n$ , we consider **Assumption A2**. Under **A1–A2**, the upper bound (31) can be reformulated independently of  $n$ . Specifically, since

$$\frac{1}{n} \left(\frac{1+\frac{M}{n}}{\frac{M}{n}}\right)^2 \frac{(1+\frac{m}{n})^2-1}{(1+\frac{m}{n})^2} = \frac{1}{n} \left(\frac{M+n}{M}\right)^2 \frac{m(m+2n)}{(m+n)^2},$$

in the case where  $M = m$ , we have

$$\frac{1}{n} \left(\frac{M+n}{M}\right)^2 \frac{m(m+2n)}{(m+n)^2} = \frac{1}{n} \frac{m(m+2n)}{m^2} = \frac{1}{n} + \frac{2}{m} \geq \frac{2}{m},$$

yielding to the sufficient condition not depending on  $n$ ,

$$\|\theta\|_H^2 \leq \frac{2}{m} \sum_{j=1}^{\infty} \lambda_j,$$

ensuring that  $\tilde{\theta}_n \prec \bar{X}_n$ , for every  $n \geq 1$ .

The asymptotic equivalence of  $\tilde{\theta}_n$  and  $\bar{X}_n$  in the sense of (2) is now derived under **Assumptions A1** in the following theorem.

**Theorem 3.** *Under **Assumption A1**, the following identities hold:*

$$\lim_{n \rightarrow \infty} nE_{\theta} \|\bar{X}_n - \theta\|_H^2 = \sum_{j=1}^{\infty} \lambda_j = \lim_{n \rightarrow \infty} nE_{\theta} \|\tilde{\theta}_n - \theta\|_H^2. \quad (32)$$

**Proof.** First, we recall

$$E_{\theta} \|\bar{X}_n\|_H^2 = \sum_{j=1}^{\infty} E_{\theta} [\langle \bar{X}_n, v_j \rangle_H]^2 = \sum_{j=1}^{\infty} \left(\frac{\lambda_j}{n} + \theta_j^2\right) < \infty,$$

and

$$E_\theta \|\bar{X}_n - \theta\|_H^2 = \sum_{j=1}^{\infty} E_\theta (\bar{X}_{n_j} - \theta_j)^2 = \sum_{j=1}^{\infty} \frac{\lambda_j}{n}.$$

Hence,

$$nE_\theta \|\bar{X}_n - \theta\|_H^2 = \sum_{j=1}^{\infty} \lambda_j.$$

Now

$$nE_\theta \|\tilde{\theta}_n - \theta\|_H^2 = \sum_{j=1}^{\infty} \lambda_j \alpha_{jn}^2 + n \sum_{j=1}^{\infty} (1 - \alpha_{jn})^2 \theta_j^2 := A_n + nB_n.$$

Uniformly in  $n$ ,  $(\frac{n\gamma_j}{n\gamma_j + \lambda_j})^2 \lambda_j \leq \lambda_j = g(j)$ , with  $g$  independent of  $n$ , and  $\sum_{j=1}^{\infty} g(j) < \infty$ , due to the trace property of operator  $C$ . Thus, from Dominated Convergence Theorem, we can interchange the limit when  $n \rightarrow \infty$  with the sum in  $A_n$ , obtaining

$$\lim_{n \rightarrow \infty} A_n = \sum_{j=1}^{\infty} \lim_{n \rightarrow \infty} \left( \frac{n\gamma_j}{n\gamma_j + \lambda_j} \right)^2 \lambda_j = \sum_{j=1}^{\infty} \lambda_j.$$

Additionally,

$$\begin{aligned} B_n &= \sum_{j=1}^{\infty} (1 - \alpha_{jn})^2 \theta_j^2 = \sum_{j=1}^{\infty} \frac{\left(\frac{\lambda_j}{n\gamma_j}\right)^2}{\left(\frac{n\gamma_j + \lambda_j}{n\gamma_j}\right)^2} \theta_j^2 \\ &= \sum_{j=1}^{\infty} \frac{(\lambda_j \theta_j)^2}{(n\gamma_j + \lambda_j)^2} = \sum_{j=1}^{\infty} \frac{(\lambda_j \theta_j)^2}{(n\gamma_j)^2 + \lambda_j^2 + 2n\gamma_j \lambda_j} \\ &\leq \frac{1}{n^2} \sum_{j=1}^{\infty} \frac{(\lambda_j \theta_j)^2}{\gamma_j^2} = \mathcal{O}\left(\frac{1}{n^2}\right), \end{aligned} \tag{33}$$

provided that

$$\sum_{j=1}^{\infty} \frac{(\lambda_j \theta_j)^2}{\gamma_j^2} < \infty.$$

In particular, under **A1**,

$$\sum_{j=1}^{\infty} \frac{(\lambda_j \theta_j)^2}{\gamma_j^2} \leq M^2 \sum_{j=1}^{\infty} \theta_j^2 = (M \|\theta\|_H)^2 < \infty.$$

Thus,

$$\lim_{n \rightarrow \infty} nE_\theta \|\tilde{\theta}_n - \theta\|_H^2 = \sum_{j=1}^{\infty} \lambda_j = \lim_{n \rightarrow \infty} nE_\theta \|\bar{X}_n - \theta\|_H^2,$$

i.e.,  $\tilde{\theta}_n$  and  $\bar{X}_n$  are asymptotically equivalent.

**Remark 1.** Since  $C$  and  $\Gamma$  are assumed to be in the trace class, their respective sequences of eigenvalues  $\{\lambda_j\}_{j \geq 1}$  and  $\{\gamma_j\}_{j \geq 1}$  converge to zero when  $j \rightarrow \infty$ . Hence,  $1/\gamma_j \rightarrow \infty$  when  $j \rightarrow \infty$ . Thus, **A1** holds if and only if  $\{\lambda_j\}_{j \geq 1}$  and  $\{\gamma_j\}_{j \geq 1}$  go to zero with the same rate, i.e.,  $\lambda_j = \mathcal{O}(\gamma_j)$ ,  $j \rightarrow \infty$ . This condition is equivalent to

$$\lim_{\omega \rightarrow \infty} \frac{N_C(\omega)}{N_\Gamma(\omega)} = K,$$

for some positive constant  $K$ , where  $N_C(\lambda)$  and  $N_\Gamma(\gamma)$  denote the respective counting functions associated with the eigenvalues of operators  $C$  and  $\Gamma$ , given by

$$\begin{aligned} N_C(\lambda) &= \text{Card} \{n \geq 1 \text{ such that } \lambda_n \leq \lambda\} \\ N_\Gamma(\gamma) &= \text{Card} \{n \geq 1 \text{ such that } \gamma_n \leq \gamma\} \end{aligned}$$

with  $\text{Card}$  denoting the cardinality of a set. In particular,  $N_C(\lambda_k) = k$ , and  $N_\Gamma(\gamma_k) = k$ .

In practice, when  $C$  is known we can propose as prior distributions for the projections  $\theta_j$ ,  $j \geq 1$ , of  $\theta$  with respect to the eigenvectors  $v_j$ ,  $j \geq 1$ , a sequence of zero-mean Gaussian distributions whose variances  $\gamma_j$ ,  $j \geq 1$ , define a real positive summable sequence such that  $\gamma_j = L\lambda_j$ , for some positive constant  $L$ . In the case where  $\lambda_j$ ,  $j \geq 1$ , are unknown the problem becomes more difficult, requiring, for example, the consistent estimation of the covariance operator eigenvalues (see, for example, [6]). Furthermore, if  $v_j$ ,  $j \geq 1$ , are also unknown, suitable estimators should be computed (see, for example, [2]).

## 5. Estimation of the covariance operator

Suppose that for  $i = 1, \dots, n$ ,  $X_i \sim \mathcal{N}(0, C)$ , is a set of i.i.d. Hilbert valued zero-mean Gaussian random variables with, as before,

$$C = \sum_{j=1}^{\infty} \lambda_j (v_j \otimes v_j) \quad (34)$$

being in the trace class, i.e.,  $\sum_{j=1}^{\infty} \lambda_j < \infty$ . Assume that  $\{v_j, j \geq 1\}$  are known, but  $\{\lambda_j, j \geq 1\}$  are unknown. Recall that, for  $i = 1, \dots, n$ ,

$$\lambda_j = E [\langle X_i, v_j \rangle_H]^2, \quad j \geq 1.$$

For each  $j \geq 1$ , the joint density of  $X_{ij} = \langle X_i, v_j \rangle_H$ ,  $i = 1, \dots, n$ , is then given by

$$L(X_{1j}, \dots, X_{nj}, \tau_j) = C \tau_j^{n/2} \exp \left( -\tau_j \sum_{i=1}^n x_{ij}^2 \right), \quad \tau_j = \frac{1}{2\lambda_j}. \quad (35)$$

For each  $j \geq 1$ , as conjugate prior for parameter  $\tau_j = \frac{1}{2\lambda_j}$ , let us choose  $\tau_j \sim \Gamma(g_j, \frac{1}{\beta_j})$ , for  $g_j > 2$ . Note that, hence,  $E[\frac{1}{\tau_j}] = \frac{\beta_j}{g_j - 1}$ . The posterior

density of  $\tau_j$  given  $x_{ij}, i = 1, \dots, n$ , is

$$C \left( \sum_{i=1}^n x_{ij}^2 \right) \tau_j^{n/2+g_j-1} \exp \left( -\tau_j \left( \beta_j + \sum_{i=1}^n x_{ij}^2 \right) \right),$$

which is a Gamma distribution with parameters  $n/2 + g_j$  and  $1/(\beta_j + \sum_{i=1}^n x_{ij}^2)$ , for each  $j \geq 1$ . For the loss function defined in terms of the squared error, the Bayes estimator of  $2\lambda_j = 1/\tau_j$  is the posterior expectation of  $1/\tau_j$ , given by  $(\beta_j + \sum_{i=1}^n x_{ij}^2)/(n/2 + g_j - 1)$ , for each  $j \geq 1$ . The Bayes estimator  $\tilde{\lambda}_{jn}$  of  $\lambda_j = 1/2\tau_j$  is then

$$\tilde{\lambda}_{jn} = \frac{\beta_j + \sum_{i=1}^n x_{ij}^2}{n + 2g_j - 2}, \quad j \geq 1 \tag{36}$$

(see [13], p. 236). Hence, from (34), the componentwise Bayes estimator of  $C$  is constructed as

$$\tilde{C}_n = \sum_{j=1}^{\infty} \tilde{\lambda}_{jn} (v_j \otimes v_j)$$

(see also Theorem 1.2.1, p. 9, [8]).

Let us begin with the analysis of the classical estimator of the covariance operator

$$C_n = \sum_{j=1}^{\infty} \bar{\lambda}_{jn} (v_j \otimes v_j),$$

where

$$\bar{\lambda}_{jn} = \frac{1}{n} \sum_{i=1}^n x_{ij}^2 \tag{37}$$

is an unbiased estimator of  $\lambda_j$ . Clearly,

$$E \left( \sum_{j=1}^{\infty} \bar{\lambda}_{jn} \right) = \sum_{j=1}^{\infty} E(\bar{\lambda}_{jn}) = \sum_{j=1}^{\infty} \lambda_j < \infty.$$

Thus,

$$\sum_{j=1}^{\infty} \bar{\lambda}_{jn} < \infty, \quad \text{a.s.}$$

Now recall that, if  $N \sim \mathcal{N}(0, \sigma^2)$ , then  $EN^4 = 3\sigma^4$ , and

$$V(N^2) = 3\sigma^4 - \sigma^4 = 2\sigma^4.$$

Since  $C_n - C$ , as an element of the Hilbert space  $\mathcal{S}(H)$  of Hilbert-Schmidt operators on  $H$ , has coordinates  $\bar{\lambda}_{jn} - \lambda_j, j \geq 1$ , with respect to the orthonormal basis  $v_j \otimes v_j, j \geq 1$ , of  $\mathcal{S}(H)$ , by Parseval identity,

$$\|C_n - C\|_{\mathcal{S}(H)}^2 = \sum_{j=1}^{\infty} [\bar{\lambda}_{jn} - \lambda_j]^2, \quad j \geq 1.$$

Hence,

$$E\|C_n - C\|_{\mathcal{S}(H)}^2 = \sum_{j=1}^{\infty} E[\bar{\lambda}_{jn} - \lambda_j]^2 = \sum_{j=1}^{\infty} V(\bar{\lambda}_{jn}) = \frac{2}{n} \sum_{j=1}^{\infty} \lambda_j^2, \quad (38)$$

provided that there exists a positive constant  $c$  such that  $\sum_{j=1}^{\infty} \lambda_j^2 \leq c \sum_{j=1}^{\infty} \lambda_j < \infty$ , by assumption. As before,  $\|\cdot\|_{\mathcal{S}(H)}$  denotes the norm in  $\mathcal{S}(H)$ .

For each  $j \geq 1$ , let us now consider the Bayesian estimator  $\tilde{\lambda}_{jn}$  of  $\lambda_j$ . Then,

$$\tilde{\lambda}_{jn} - \lambda_j = \frac{1}{b_{j,n}}(\bar{\lambda}_{jn} - \lambda_j) + \frac{a_{j,n}}{b_{j,n}} + \frac{1 - b_{j,n}}{b_{j,n}}\lambda_j, \quad (39)$$

where

$$a_{j,n} = \frac{\beta_j}{n}, \quad b_{j,n} = 1 + \frac{2}{n}(g_j - 1).$$

From equations (38) and (39),

$$\begin{aligned} E[\tilde{\lambda}_{jn} - \lambda_j]^2 &= \frac{1}{b_{j,n}^2} E[\bar{\lambda}_{jn} - \lambda_j]^2 + \left[ \frac{a_{j,n}}{b_{j,n}} + \frac{1 - b_{j,n}}{b_{j,n}}\lambda_j \right]^2 \\ &\quad + 2 \frac{E[\bar{\lambda}_{jn} - \lambda_j] [a_{j,n} + (1 - b_{j,n})\lambda_j]}{b_{j,n}^2} \\ &= \frac{2}{nb_{j,n}^2} \lambda_j^2 + \frac{(a_{j,n} + (1 - b_{j,n})\lambda_j)^2}{b_{j,n}^2} \\ &= \frac{2}{nb_{j,n}^2} \lambda_j^2 + \frac{1}{n^2} \frac{(\beta_j - 2(g_j - 1)\lambda_j)^2}{b_{j,n}^2} \\ &= \frac{2n\lambda_j^2}{[n + 2(g_j - 1)]^2} + \frac{[\beta_j - 2(g_j - 1)\lambda_j]^2}{[n + 2(g_j - 1)]^2} \\ &= A_n(j) + B_n(j). \end{aligned} \quad (40)$$

Since  $g_j > 2$ , we have,

$$|nA_n(j)| = nA_n(j) = \frac{2(n\lambda_j)^2}{n^2 + 4(g_j - 1)^2 + 4n(g_j - 1)} \leq \frac{2n^2\lambda_j^2}{n^2} = 2\lambda_j^2,$$

and there exists a positive constant  $c$  such that  $2 \sum_{j=1}^{\infty} \lambda_j^2 \leq 2c \sum_{j=1}^{\infty} \lambda_j < \infty$ .

Then, by applying Dominated Convergence Theorem,

$$\lim_{n \rightarrow \infty} n \sum_{j=1}^{\infty} A_n(j) = \sum_{j=1}^{\infty} \lim_{n \rightarrow \infty} nA_n(j) = 2 \sum_{j=1}^{\infty} \lambda_j^2 < \infty, \quad (41)$$

where the sum is interpreted as an integral with respect to a counting measure  $\mu(dx) = \sum_{j=1}^{\infty} \delta_{A_n(j)}(x)dx$ , with  $\delta_a(x)$  denoting the Dirac Delta measure at point  $a$ , and  $A_n(j)$ ,  $j \geq 1$ , are given in equation (40).

Additionally, again keeping in mind that  $g_j > 2$ ,

$$\begin{aligned} n \sum_{j=1}^{\infty} B_n(j) &= \sum_{j=1}^{\infty} \frac{n[\beta_j - 2(g_j - 1)\lambda_j]^2}{n^2 + 4(g_j - 1)^2 + 4n(g_j - 1)} \\ &\leq \frac{1}{n} \sum_{j=1}^{\infty} [\beta_j - 2(g_j - 1)\lambda_j]^2 = \mathcal{O}\left(\frac{1}{n}\right), \end{aligned} \tag{42}$$

provided that  $\sum_{j=1}^{\infty} [\beta_j - 2(g_j - 1)\lambda_j]^2 < \infty$ . In particular, the following assumption gives a sufficient condition for  $\sum_{j=1}^{\infty} [\beta_j - 2(g_j - 1)\lambda_j]^2 < \infty$  to hold:

**Assumption A3.** For  $j \geq 1$ ,

$$\frac{\beta_j}{2(g_j - 1) + 1} \leq \lambda_j.$$

**Remark 2.** Note that for  $g_j = 3$ ,  $j \geq 1$ , we have  $\beta_j \leq 5\lambda_j$ ,  $j \geq 1$ . Thus, **Assumption A3** leads to some upper bounds on the parameters of the a-priori Gamma distribution, in terms of the eigenvalues  $\lambda_j$ ,  $j \geq 1$ , of the covariance operator  $C$ . Some prior knowledge about  $\lambda_j$  is then required. To ensure that condition **A3** holds parameters  $g_j$ , and  $\beta_j$ ,  $j \geq 1$ , involved in the respective Gamma priors proposed for parameters  $\tau_j = \frac{1}{2\lambda_j}$ ,  $j \geq 1$ , could be chosen as follows:

$$\begin{aligned} g_j &= 2 + j\epsilon_1(j) \\ \beta_j &= (2(g_j - 1) + 1)\lambda_j - \epsilon_2(j). \end{aligned} \tag{43}$$

Here,  $\epsilon_i(j)$ ,  $j \geq 1$ ,  $i = 1, 2$ , are arbitrary positive constants that, in particular, could not depend on  $j$ .

In particular, from equation (43), since for  $\tau_j \sim \Gamma(g_j, \frac{1}{\beta_j})$ , and considering  $\epsilon_2(j) \geq \lambda_j$ ,

$$\begin{aligned} \mu_{\lambda_j} &= E[\lambda_j] = E\left[\frac{1}{2\tau_j}\right] = \frac{\beta_j}{2g_j - 2} \\ &= \frac{(2(g_j - 1) + 1)\lambda_j - \epsilon_2(j)}{2g_j - 2} \leq \lambda_j, \quad j \geq 1, \end{aligned} \tag{44}$$

we have

$$\sum_{j=1}^{\infty} \mu_{\lambda_j} = \sum_{j=1}^{\infty} E[\lambda_j] = \sum_{j=1}^{\infty} E\left[\frac{1}{2\tau_j}\right] \leq \sum_{j=1}^{\infty} \lambda_j < \infty.$$

Thus, the means  $\mu_{\lambda_j}$ ,  $j \geq 1$ , of parameters  $\lambda_j = 1/(2\tau_j)$ ,  $j \geq 1$ , with  $\tau_j \sim \Gamma(g_j, \frac{1}{\beta_j})$ ,  $j \geq 1$ , define a  $l^1$  sequence when  $\epsilon_2(j) \geq \lambda_j$  in equation (43).

For implementation of equation (43) in practice, since  $\lambda_j, j \geq 1$ , are unknown, a possible choice could be to replace in (43)  $\lambda_j, j \geq 1$ , by a sequence of positive real numbers, whose sum is smaller than the empirical variance computed from the data.

The following result is then formulated.

**Theorem 4.** *Under assumption A3, the estimators  $C_n$  and  $\tilde{C}_n$  satisfy*

$$\lim_{n \rightarrow \infty} nE\|\tilde{C}_n - C\|_{\mathcal{S}(H)}^2 = \lim_{n \rightarrow \infty} nE\|C_n - C\|_{\mathcal{S}(H)}^2 = 2 \sum_{j=1}^{\infty} \lambda_j^2.$$

*That is, they are asymptotically equivalent in the sense of equation (2).*

**Proof.** Under A3,

$$\sum_{j=1}^{\infty} [\beta_j - 2(g_j - 1)\lambda_j]^2 \leq \sum_{j=1}^{\infty} \lambda_j^2 < \infty.$$

Thus, equation (42) holds, which jointly with equations (38), (40) and (41) yields to the desired equality

$$\lim_{n \rightarrow \infty} nE\|\tilde{C}_n - C\|_{\mathcal{S}(H)}^2 = \lim_{n \rightarrow \infty} nE\|C_n - C\|_{\mathcal{S}(H)}^2 = 2 \sum_{j=1}^{\infty} \lambda_j^2.$$

## 6. Simulation study

The comparison of the efficiency of the finite-dimensional approximation of the studied Bayesian and classical estimators, in the  $l^2$  valued Poisson process and the Hilbert valued Gaussian random variable frameworks, is now performed through a simulation study, for an increasing sequence of functional sample sizes.

### 6.1. Equivalence of Bayesian and classical estimators and predictors of $l^2$ valued Poisson process

We study the following models of  $l^2$  valued Poisson process  $M_t = \{N_{t,j}, j \geq 0\}$ , with infinite-dimensional vectors of intensities  $\{\lambda_{ji}, j \geq 1\}$ ,  $i = 1, 2, 3$ , respectively given by:

$$\begin{aligned} \lambda_{k1} &= (1/k)^6, & k \geq 1 \\ \lambda_{k2} &= (1/10)^k, & k \geq 1 \\ \lambda_{k3} &= (1/k)^k, & k \geq 1. \end{aligned} \tag{45}$$

Note that in equation (45),  $\lambda_{ki}, k \geq 1$ , are selected in order to ensure that  $\sum_{k=1}^{\infty} \lambda_{ki} < \infty$ , testing different rates of convergence, for each model  $i$ , with  $i = 1, 2, 3$ .

TABLE 1  
 For  $i = 1, 2$   $MSEIC_i$  and  $MSEIB_i$  values in equation (47) are displayed for  $S \in [1000, 10000]$ , with  $DSS = 500$ , and for  $M = 75$

S.S	MSEIC <sub>1</sub>	MSEIB <sub>1</sub>	MSEIC <sub>2</sub>	MSEIB <sub>2</sub>
1000	0.1048e - 002	0.1052e - 002	0.1252e - 003	0.1249e - 003
1500	0.7038e - 003	0.7037e - 003	0.7449e - 004	0.7441e - 004
2000	0.4581e - 003	0.4562e - 003	0.5200e - 004	0.5192e - 004
2500	0.4173e - 003	0.4163e - 003	0.4447e - 004	0.4446e - 004
3000	0.3513e - 003	0.3537e - 003	0.3998e - 004	0.3993e - 004
3500	0.2881e - 003	0.2877e - 003	0.2882e - 004	0.2881e - 004
4000	0.2481e - 003	0.2480e - 003	0.2690e - 004	0.2689e - 004
4500	0.2412e - 003	0.2417e - 003	0.2473e - 004	0.2471e - 004
5000	0.1948e - 003	0.1950e - 003	0.2286e - 004	0.2286e - 004
5500	0.1895e - 003	0.1901e - 003	0.2055e - 004	0.2055e - 004
6000	0.1879e - 003	0.1887e - 003	0.1847e - 004	0.1846e - 004
6500	0.1583e - 003	0.1581e - 003	0.1502e - 003	0.1740e - 004
7000	0.1502e - 003	0.1504e - 003	0.1595e - 004	0.1595e - 004
7500	0.1308e - 003	0.1305e - 003	0.1415e - 004	0.1415e - 004
8000	0.1320e - 003	0.1316e - 003	0.1506e - 004	0.1506e - 004
8500	0.1267e - 003	0.1265e - 003	0.1336e - 004	0.1336e - 004
9000	0.1145e - 003	0.1149e - 003	0.1328e - 004	0.1327e - 004
9500	0.1158e - 003	0.1158e - 003	0.1213e - 004	0.1213e - 004
10000	0.8589e - 004	0.8597e - 004	0.1006e - 004	0.1006e - 004

The following parameters values are tested for the Gamma distributions defining the priors on the intensity components (45)

$$\begin{aligned}
 a_{k1} &= (1/2)^{2k}, & b_{k1} &= 3, & k &\geq 1 \\
 a_{k2} &= (1/5)^{6k}, & b_{k2} &= 1, & k &\geq 1 \\
 a_{k3} &= (1/4)^{3k}, & b_{k3} &= 2, & k &\geq 1.
 \end{aligned}
 \tag{46}$$

Note that  $\{a_{ki}\}_{k \geq 1}$  and  $\{b_{ki}\}_{k \geq 1}$  are selected in order to ensure that **Assumption A0** holds, since  $\sum_{k=1}^{\infty} a_{ki} < \infty$ , and  $\sup_{k \geq 1} b_{ki} < \infty$  for each  $i = 1, 2, 3$ . In addition, for each model  $i = 1, 2, 3$ , different rates of convergence to zero are tested in the selection of parameter sequence  $\{a_{ki}\}_{k \geq 1}$  in equation (46).

Tables 1-2 show the statistics MSEIC<sub>i</sub> and MSEIB<sub>i</sub> respectively denoting the integrated (truncated) empirical mean-square errors of the classical and Bayes estimators of the intensity vector of the Poisson process, obtained after truncation at term  $M$ , in Model  $i$ , for  $i = 1, 2, 3$ . Integration here is understood with respect to a counting measure given by  $\mu(dx) = \sum_{k=1}^M \delta_{1/R \sum_{l=1}^R [\tilde{\lambda}_{kS,i}^l - \lambda_{ki}]^2}(x)dx$ , in the case of MSEIC<sub>i</sub>, and given by  $\mu(dx) = \sum_{k=1}^M \delta_{1/R \sum_{l=1}^R [\tilde{\lambda}_{kS,i}^l - \lambda_{ki}]^2}(x)dx$ , in the case of MSEIB<sub>i</sub>, for  $i = 1, 2, 3$ , where, as before,  $\delta_a(x)$  denotes the Dirac delta measure at point  $a$ . The number of observed times is denoted as  $S$ . In these tables the values of  $S$  displayed are in the interval  $[1000, 10000]$ , i.e.,  $S \in [1000, 10000]$ , with discretization step size  $DSS = 500$ . The truncation order considered is  $M = 75$ . The showed integrated empirical mean-square errors are based on  $R = 500$  repetitions. Specifically, Tables 1-2 (see also Figure 1)

TABLE 2  
*MSEIC*<sub>3</sub> and *MSEIB*<sub>3</sub> values in equation (47) are displayed for  $S \in [1000, 10000]$ , with  $DSS = 500$ , and for  $M = 75$

S.S	MSEIC <sub>3</sub>	MSEIB <sub>3</sub>
1000	0.1270e - 002	0.1266e - 002
1500	0.8466e - 003	0.8477e - 003
2000	0.6272e - 003	0.6297e - 003
2500	0.4801e - 003	0.4810e - 003
3000	0.4668e - 003	0.4678e - 003
3500	0.3557e - 003	0.3547e - 003
4000	0.3462e - 003	0.3473e - 003
4500	0.2949e - 003	0.2955e - 003
5000	0.2745e - 003	0.2740e - 003
5500	0.2287e - 003	0.2287e - 003
6000	0.2173e - 003	0.2183e - 003
6500	0.2091e - 003	0.2088e - 003
7000	0.1816e - 003	0.1815e - 003
7500	0.1737e - 003	0.1738e - 003
8000	0.1656e - 003	0.1658e - 003
8500	0.1424e - 003	0.1422e - 003
9000	0.1424e - 003	0.1424e - 003
9500	0.1361e - 003	0.1357e - 003
10000	0.1439e - 003	0.1441e - 003

show the empirical values

$$\begin{aligned} \text{MSEIC}_i &= \sum_{k=1}^M \left[ \frac{1}{R} \sum_{l=1}^R [\hat{\lambda}_{kS,i}^l - \lambda_{ki}]^2 \right], \quad i = 1, 2, 3 \\ \text{MSEIB}_i &= \sum_{k=1}^M \left[ \frac{1}{R} \sum_{l=1}^R [\tilde{\lambda}_{kS,i}^l - \lambda_{ki}]^2 \right], \quad i = 1, 2, 3, \end{aligned} \quad (47)$$

where  $\hat{\lambda}_{kS,i}^l$  represents the Bayesian estimator of the  $k$ th component of the vector of intensities associated with the  $l^2$  valued Poisson process  $\{M_{t,i}, t \geq 0\}$ , based on the  $l$ th observed realization  $\{M_{t,i}^l, t = 1, \dots, S\}$  of the process generated in Model  $i$ , for  $i = 1, 2, 3$ , with  $M_{t,i}^l = \{N_{t,k,i}^l, k = 1, \dots, M\}$ , for  $t \in \{1, \dots, S\}$ , and for  $l = 1, \dots, R$ . Similarly,  $\tilde{\lambda}_{kS,i}^l$  denotes the classical estimator of the  $k$ th component of the vector of intensities associated with the  $l^2$  valued Poisson process  $\{M_{t,i}, t \geq 0\}$ , based on the  $l$ th observed realization  $\{M_{t,i}^l, t = 1, \dots, S\}$  of the process generated in Model  $i$ , for  $i = 1, 2, 3$ , with  $M_{t,i}^l = \{N_{t,k,i}^l, k = 1, \dots, M\}$ , for  $t \in \{1, \dots, S\}$ , and for  $l = 1, \dots, R$ .

The classical and Bayes estimators of the intensity vector display the same behavior in relation to the rate of convergence to zero of their integrated empirical mean-square errors, when the sample size increases. A faster decreasing can be seen in Model 2, attaching the theoretical order of magnitude  $1/n = 10^{-4}$ . While the order of magnitude of the integrated empirical mean-square error in Models 1 and 3 is  $10^{-3}$ , presenting a slower rate of decrease for the samples sizes tested. These observed rates of decrease of the integrated empirical mean square

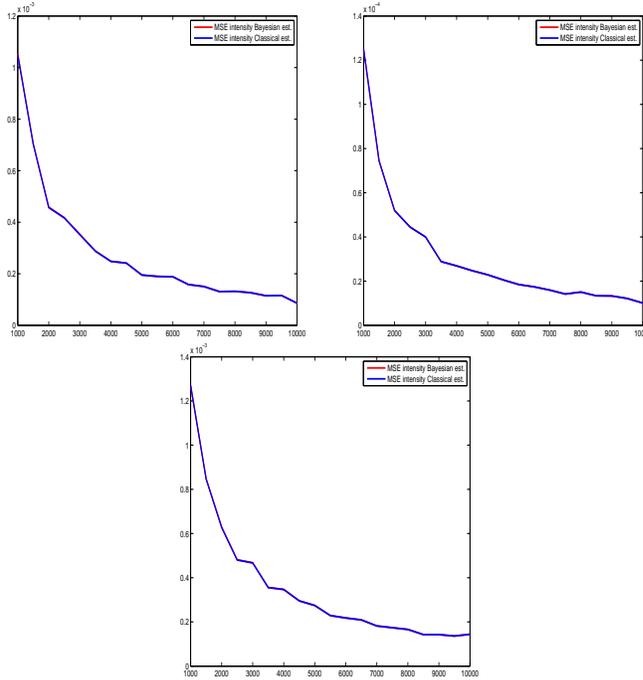


FIG 1. Results reflected in Tables 1–2 are represented here for Model 1 at the top-left, for Model 2 at the top-right, and for Model 3 at the bottom. Note that the red and blue lines coincide in all the cases.

errors in the three models analyzed reflect that the empirical approximations of  $A_k(T)$  and  $B_k(T)$ ,  $k = 1, \dots, M = 75$  (see equation (11)), providing upper bounds of the computed integrated empirical mean square errors, decrease when the parameters  $a_{ki}$ ,  $k \geq 1$ , and  $\lambda_{ki}$ ,  $k \geq 1$ , decrease, for  $i = 1, 2, 3$ . Therefore, the smallest values of such empirical quantities correspond to Model 2. Note that the differences between parameters involved in Models 1 and 3 are not significant in relation to the sample sizes tested. Summarizing, the rate of decrease of parameters  $a_{ki}$ ,  $k \geq 1$ , as well as of the components of the intensity vector  $\lambda_{ki}$ ,  $k \geq 1$ , for  $i = 1, 2, 3$  affects the rate of convergence to zero of the integrated empirical mean square errors of the Bayes and classical estimators of the intensity vector of  $l^2$  valued Poisson process.

To illustrate that the relative efficiency of Bayes and classical predictors is one, the absolute distances of their scaled, by the sample size, integrated empirical mean square errors to the theoretical limit derived in Theorem 1 (see also equation (15)) are displayed in Figure 2. Specifically, for model  $i = 1, 2, 3$ , by  $\text{AEBP}_i$  we denote the absolute distance of the integrated empirical mean square error, scaled by the sample size  $S$ , of the Bayes predictor to the theoretical value  $h^2 \sum_{k=1}^M \lambda_{ki}$ , obtained in Theorem 1. Similarly, for model  $i = 1, 2, 3$ ,  $\text{AECP}_i$  denotes the absolute distance of the integrated empirical mean square

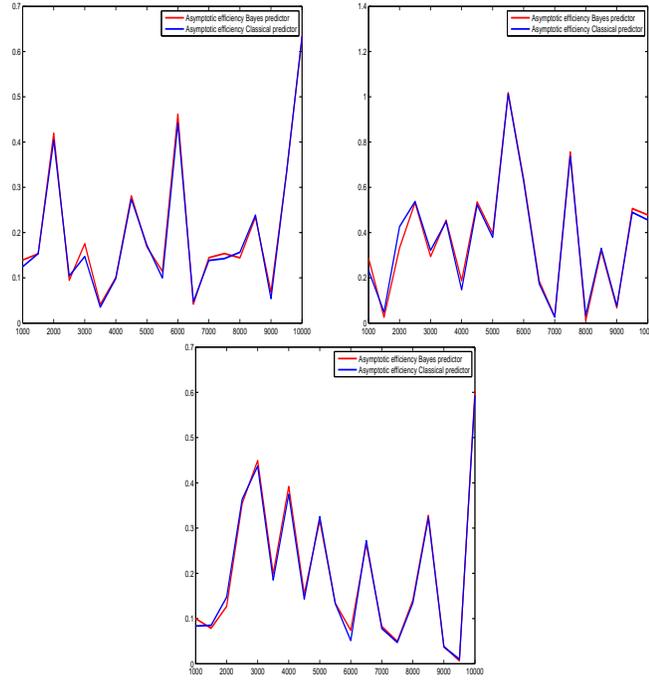


FIG 2.  $AEBP_i$ , and  $AECP_i$  values defined in equation (48), for  $i = 1$  at the top-left,  $i = 2$  at the top-right, and  $i = 3$  at the bottom. They are represented with respect to the sample sizes  $S \in [1000, 10000]$ , reflected in the horizontal axis, where discretization step size  $DSS = 500$  is considered.

error, scaled by the sample size  $S$ , of the classical predictor to the theoretical value  $h^2 \sum_{k=1}^M \lambda_{ki}$  given in Theorem 1. Equivalently, for  $i = 1, 2, 3$ , the following empirical quantities are displayed in Figure 2:

$$\begin{aligned}
 AEBP_i &= \left| S \sum_{k=1}^M \left[ \frac{1}{R} \sum_{l=1}^R [\tilde{\lambda}_{kS,i}^l h + N_{S,k,i}^l - (\lambda_{ki} h + N_{S,k,i}^l)]^2 \right] - h^2 \sum_{k=1}^M \lambda_{ki} \right| \\
 AECP_i &= \left| S \sum_{k=1}^M \left[ \frac{1}{R} \sum_{l=1}^R [h(\hat{\lambda}_{kS,i}^l - \lambda_{ki})]^2 \right] - h^2 \sum_{k=1}^M \lambda_{ki} \right|, \quad (48)
 \end{aligned}$$

where the value  $h = 2$  has been considered with the same notation as before. The sample sizes studied coincide with the ones tested in Tables 1–2,  $S \in [1000, 10000]$ , with discretization step size  $DSS = 500$ , and truncation order  $M = 75$ . The approximation of the integrated empirical mean-square error is computed from  $R = 500$  realizations of Poisson process over the interval  $(0, S]$ . The results displayed show almost the same values of the absolute distances to the fixed theoretical limit  $h^2 \sum_{k=1}^M \lambda_{ki}$ , given in Theorem 1, of the the integrated empirical mean-square errors of the Bayes and classical estimators, for the 20

samples sizes  $S$  tested between 1000 and 10000. It can also be observed that a larger number of realizations of Poisson process over the interval  $(0, S]$  should be generated for each  $S$  studied, in order to smoothing the absolute distances observed. However, memory restrictions, as well as running time have limited our computations to the case of 500 realizations of Poisson process, which allows us to illustrate our primary objective in relation to the value one of the relative efficiency of Bayes and classical estimators.

### 6.2. Equivalence of Bayesian and classical mean estimators for $H$ valued Gaussian random variables

To generate examples of trace, and, in particular, of Hilbert-Schmidt covariance operators, admitting the spectral decomposition given in equation (23), in terms of a system of eigenvalues and eigenvectors, we will consider the resolution of the identity  $\sum_{k=1}^{\infty} \phi_k \otimes \phi_k$ , defined in terms of the system of eigenfunctions of the Dirichlet negative Laplacian operator on  $L^2([0, T])$ , given by

$$\phi_k(t) = \sqrt{\frac{2}{T}} \sin\left(\frac{k\pi t}{T}\right), \quad k \geq 1, \quad t \in [0, T], \quad (49)$$

where  $L^2([0, T])$  denotes the space of square integrable functions on  $[0, T]$  (see, for example, [20]). As usual, by Dirichlet negative Laplacian operator on  $L^2([0, T])$ , denoted as  $(-\Delta)_{[0, T]}$ , we understand the operator defined by

$$(-\Delta)_{[0, T]}(f)(t) = -\frac{d^2}{dt^2}(f)(t), \quad \forall t \in (0, T),$$

with  $f(0) = f(T) = 0$ .

The following examples of covariance operators are then considered:

$$C_i = \sum_{k=1}^{\infty} \lambda_{ki} \phi_k \otimes \phi_k, \quad i = 1, 2, 3, \quad (50)$$

with eigenvectors  $\phi_k$ ,  $k \geq 1$ , given in equation (49), and with eigenvalues  $\lambda_{ki}$ ,  $k \geq 1$ ,  $i = 1, 2, 3$ , respectively defined by the following  $l^1$  sequences:

$$\begin{aligned} \lambda_{k1} &= \left[\frac{1}{10}\right]^k, \quad k \geq 1, \\ \lambda_{k2} &= \frac{1}{1+k^4}, \quad k \geq 1, \\ \lambda_{k3} &= \frac{1}{1+k^{6/5}}, \quad k \geq 1, \end{aligned} \quad (51)$$

where we have considered different rates of convergence to zero to investigate if they affect the relative efficiency of Bayes and classical estimators for finite truncation order and for the finite sample sizes tested below. As commented, we

have considered  $H = L^2([0, T])$ , with  $T = 200$ , and  $E[X_i] = \theta_i \in L^2([0, T])$  having Fourier coefficients  $\{\theta_{ki}, k \geq 1\}$ ,  $i = 1, 2, 3$ , with respect to the eigenfunction basis (49), given by

$$\begin{aligned}\theta_{k1} &= \left[\frac{1}{k}\right]^3, & k \geq 1, \\ \theta_{k2} &= \left[\frac{1}{k}\right]^2, & k \geq 1, \\ \theta_{k3} &= \left[\frac{1}{k}\right]^{3/4}, & k \geq 1,\end{aligned}\tag{52}$$

where again different rates of convergence to zero are tested. To ensure **Assumption A1** holds, for  $i = 1, 2, 3$ , let the Gaussian prior on  $\theta_i$  has covariance operator  $\Gamma_i$  with eigenvalues  $\gamma_{ki}$ ,  $k \geq 1$ , with respect to the eigenvectors (49), given by

$$\gamma_{ki} = L\lambda_{k,i}, \quad k \geq 1.\tag{53}$$

Here, any positive value of constant  $L$  ensures that **Assumption A1** holds with  $m = M = 1/L$ . In particular, we have considered  $L = 10$ . For  $i = 1, 2, 3$ ,  $\Gamma_i$  is then defined as

$$\Gamma_i = \sum_{k=1}^{\infty} \gamma_{ki} \phi_k \otimes \phi_k.$$

For  $i = 1, 2, 3$ , the Bayesian estimator of  $\theta_i$  is computed from equation (24), and compared with the classical estimate  $\bar{X}_{ni}$ , in terms of their integrated empirical mean square errors truncated at term  $M$ , respectively denoted as  $\text{MSEMB}_i$  and  $\text{MSEMC}_i$ , for each model  $i = 1, 2, 3$ . That is, they are compared in terms of the following empirical quantities

$$\begin{aligned}\text{MSEMB}_i &= \sum_{k=1}^M \left[ \frac{1}{R} \sum_{l=1}^R [\tilde{\theta}_{kn,i}^l - \theta_{ki}]^2 \right] \\ \text{MSEMC}_i &= \sum_{k=1}^M \left[ \frac{1}{R} \sum_{l=1}^R [\bar{X}_{kn,i}^l - \theta_{ki}]^2 \right],\end{aligned}\tag{54}$$

approximating the theoretical values  $E\|\tilde{\theta}_{ni} - \theta_i\|_{L^2([0,T])}^2$  and  $E\|\bar{X}_{ni} - \theta_i\|_{L^2([0,T])}^2$ , respectively. Here, for  $i = 1, 2, 3$ ,  $\tilde{\theta}_{kn,i}^l$  and  $\bar{X}_{kn,i}^l$  denote the respective Bayes and classical estimates of  $\theta_{ki}$ , based on the  $l$ th generation of a finite-dimensional approximation of sample  $X_{1i}^l, \dots, X_{ni}^l$  of size  $n$  of the Hilbert valued Gaussian random variable  $X_i$  of Model  $i$ , for  $l = 1, \dots, R$ .

Bayesian and classical estimators of the functional mean display the same efficiency in the three models studied for the sample sizes tested. That is, their relative efficiency is almost one as it can be seen in Tables 3–4. The values  $\text{MSEMB}_1$  and  $\text{MSEMC}_1$  are below the values of  $\text{MSEMB}_2$  and  $\text{MSEMC}_2$ , and  $\text{MSEMB}_3$  and  $\text{MSEMC}_3$  in all the sample sizes tested (see Tables 3–4 and

TABLE 3  
 For  $i = 1, 2$   $MSEMC_i$  and  $MSEMB_i$  values in equation (54) are displayed for  $n \in [1000, 10000]$ , with  $DSS = 500$ , and for  $M = 75$

S.S	MSEMC <sub>1</sub>	MSEMB <sub>1</sub>	MSEMC <sub>2</sub>	MSEMB <sub>2</sub>
1000	0.7809e - 004	0.7809e - 004	0.1142e - 003	0.1141e - 003
1500	0.5263e - 004	0.5266e - 004	0.8121e - 004	0.8128e - 004
2000	0.4011e - 004	0.4012e - 004	0.5222e - 004	0.5219e - 004
2500	0.3583e - 004	0.3581e - 004	0.4522e - 004	0.4521e - 004
3000	0.2876e - 004	0.2878e - 004	0.3778e - 004	0.3774e - 004
3500	0.2313e - 004	0.2314e - 004	0.3300e - 004	0.3300e - 004
4000	0.2053e - 004	2.0534e - 004	0.3008e - 004	0.3006e - 004
4500	0.1832e - 004	0.1831e - 004	0.2758e - 004	0.2758e - 004
5000	0.1485e - 004	0.1485e - 004	0.2258e - 004	0.2258e - 004
5500	0.1547e - 004	0.1546e - 004	0.2061e - 004	0.2060e - 004
6000	0.1456e - 004	0.1456e - 004	0.1771e - 004	0.1771e - 004
6500	0.1242e - 004	0.1241e - 004	0.1600e - 004	0.1599e - 004
7000	0.1076e - 004	0.1076e - 004	0.1473e - 004	0.1473e - 004
7500	0.1134e - 004	0.1134e - 004	0.1489e - 004	0.1489e - 004
8000	0.1053e - 004	0.1052e - 004	0.1229e - 004	0.1229e - 004
8500	0.9763e - 005	0.9758e - 005	0.1309e - 004	0.1309e - 004
9000	0.9747e - 005	0.9742e - 005	0.1174e - 004	0.1173e - 004
9500	0.8877e - 005	0.8874e - 005	0.9766e - 005	0.9769e - 005
10000	0.8234e - 005	0.8232e - 005	0.1104e - 004	0.1105e - 004

TABLE 4  
 $MSEMC_3$  and  $MSEMB_3$  values in equation (54) are displayed for  $n \in [1000, 10000]$ , with  $DSS = 500$ , and for  $M = 75$

S.S	MSEMC <sub>3</sub>	MSEMB <sub>3</sub>
1000	0.2681e - 002	0.2681e - 002
1500	0.1782e - 002	0.1782e - 002
2000	0.1348e - 002	0.1348e - 002
2500	0.1134e - 002	0.1134e - 002
3000	0.8825e - 003	0.8825e - 003
3500	0.7916e - 003	0.7915e - 003
4000	0.6959e - 003	0.6959e - 003
4500	0.5955e - 003	0.5955e - 003
5000	0.5454e - 003	0.5454e - 003
5500	0.4777e - 003	0.4777e - 003
6000	0.4519e - 003	0.4518e - 003
6500	0.4163e - 003	0.4162e - 003
7000	0.3824e - 003	0.3823e - 003
7500	0.3517e - 003	0.3517e - 003
8000	0.3316e - 003	0.3316e - 003
8500	0.3237e - 003	0.3237e - 003
9000	0.2955e - 003	0.2955e - 003
9500	0.2834e - 003	0.2834e - 003
10000	0.2629e - 003	0.2629e - 003

Figure 3). Moreover, the largest values of the integrated empirical mean square errors correspond to Model 3 (see Table 4). The results obtained reflect the fact that the functional mean-square errors of the Bayes and classical estimators, computed in the proof of Theorem 3, are monotone increasing functions with respect to the parameters  $\lambda_{ki}$ , and  $\theta_{ki}$ ,  $k \geq 1$ , for each  $i = 1, 2, 3$ , when these

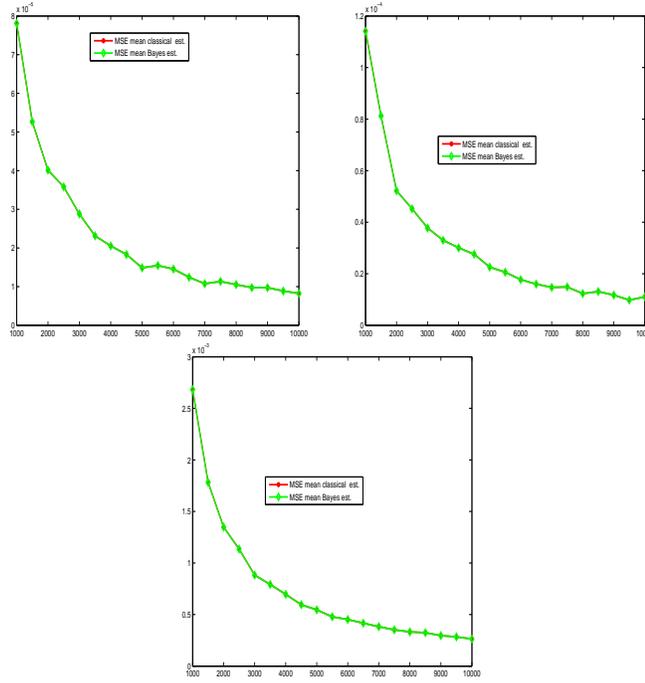


FIG 3.  $MSEMB_i$  and  $MSEMC_i$  values (vertical axis), for  $i = 1, 2$  at the top, and for  $i = 3$  at the bottom, considering the sample size  $n \in [1000, 10000]$ , with discretization step size  $DSS = 500$  (horizontal axis).

parameters are positive. Note that, in the models studied

$$0 < \lambda_{k1} \leq \lambda_{k2} \leq \lambda_{k3}, \quad 0 < \theta_{k1} \leq \theta_{k2} \leq \theta_{k3}, \quad k \geq 1. \tag{55}$$

In addition, we also consider, for  $i = 1, 2, 3$ , the absolute distances  $AEMB_i$  and  $AEMC_i$  of the respective integrated (truncated) empirical mean-square errors, scaled by the sample size  $n$ , of the Bayes and classical estimators to the limit quantity  $\sum_{k=1}^M \lambda_{ki}$  derived in Theorem 3. Equivalently, for  $i = 1, 2, 3$ , let us consider  $AEMB_i$  and  $AEMC_i$  respectively defined, for a given truncation order  $M$ , as:

$$\begin{aligned} AEMB_i &= \left| n \sum_{k=1}^M \left[ \frac{1}{R} \sum_{l=1}^R [\tilde{\theta}_{kn,i}^l - \theta_{ki}]^2 \right] - \sum_{k=1}^M \lambda_{ki} \right| \\ AEMC_i &= \left| n \sum_{k=1}^M \left[ \frac{1}{R} \sum_{l=1}^R [\bar{X}_{kn,i}^l - \theta_{ki}]^2 \right] - \sum_{k=1}^M \lambda_{ki} \right| \end{aligned} \tag{56}$$

where the values  $M = 75$ ,  $R = 500$ , and  $n \in [1000, 10000]$ , with discretization step size  $DSS = 500$ , have been tested. The order of magnitude of  $AEMB_i$  and  $AEMC_i$  is substantially smaller for  $i = 1$ , that is, for Model 1, which also displays

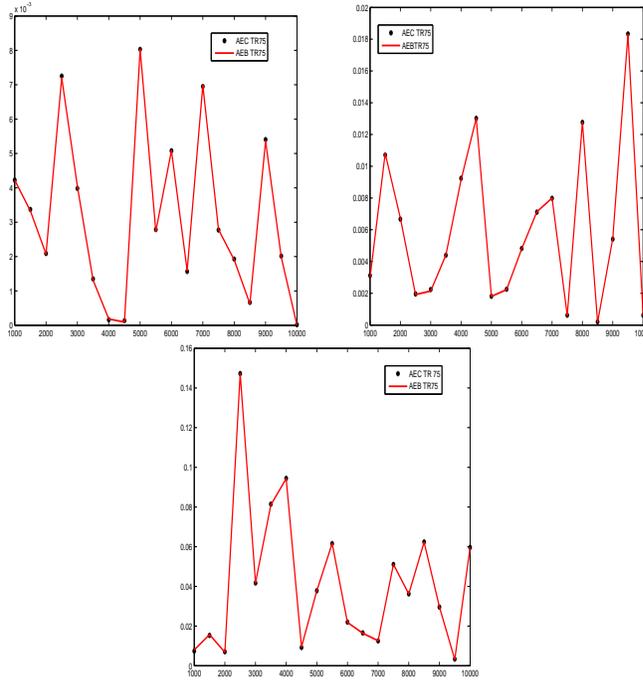


FIG 4.  $AEMB_i$  and  $AEMC_i$  values (vertical axis), for  $i = 1, 2$  at the top, and for  $i = 3$  at the bottom, considering the sample size  $n \in [1000, 10000]$ , with discretization step size  $DSS = 500$  (horizontal axis).

a faster decreasing to zero of the integrated (truncated) empirical mean square error. Regarding Models 2 and 3, again we have a faster decreasing to zero of the absolute distances of the integrated empirical mean-square error for Model 2 to the limit  $\sum_{k=1}^M \lambda_{k,2}$ , derived in Theorem 3, than for Model 3. As commented, this behavior is due to the relative magnitude of the covariance operator eigenvalues and mean function projections in the three models analyzed, as reflected in equation (55). A larger number of repetitions than  $R = 500$  should be considered in order to smoothing the results displayed in Figure 4. We have displayed the case of 500 repetitions by memory and running time restrictions, which, on the other hand, it is sufficient to show the equivalent behavior of Bayes and classical estimators in relation to their efficiency for the large finite sample sizes tested.

**6.3. Asymptotic equivalence of Bayesian and classic covariance operator estimators for  $H$  valued Gaussian random variables**

To analyze, for finite samples sizes and a given truncation order, the influence of the rate of convergence to zero of the eigenvalues of the covariance operator on the equivalent behavior of Bayes and classical estimators, we consider the

following models  $C_i$  of trace covariance operators for  $i = 1, \dots, 6$ :

$$C_i(h) = \sum_{k=1}^{\infty} \lambda_{ki} \langle \phi_k, h \rangle_H \phi_k, \quad \forall h \in H,$$

with  $H = L^2([0, T])$ ,  $\phi_k$ ,  $k \geq 1$ , are defined as in equation (49), and

$$\begin{aligned} \lambda_{j1} &= \frac{1}{(1+j^{6/5})}, & \lambda_{j2} &= \frac{1}{(1+j^2)}, & \lambda_{j3} &= \exp(-(1+j^2)), \\ \lambda_{j4} &= \left[\frac{1}{10}\right]^j, & \lambda_{j5} &= \frac{1}{(1+j^{3/2})}, & \lambda_{j6} &= \frac{1}{(1+j^{4/3})}, \quad j \geq 1. \end{aligned} \quad (57)$$

In order to ensure that the restrictions for the asymptotic equivalence of Bayes and classical estimators required in Section 5 are satisfied by the parameters of the Gamma priors for the eigenvalues of the covariance operator (see also **Assumption A3** and Remark 2, for  $i = 1, 2, 3$ , the following values of parameters  $g_{ki}$  and  $\beta_{ki}$ , characterizing the Gamma prior distributions on  $\tau_{ki} = \frac{1}{2\lambda_{ki}}$ ,  $k \geq 1$ , are selected (see also equation (43)):

$$\begin{aligned} g_{ki} &= 2 + k/10, & \beta_{ki} &= [2(g_{ki} - 1) + 1]\lambda_{ki} - 5/10 & i = 1, 2, 3, 4 \\ g_{ki} &= 3 + k/100 & \beta_{ki} &= [2(g_{ki} - 1) + 1]\lambda_{ki} - 5/10, & i = 5, 6. \end{aligned} \quad (58)$$

For  $i = 1, \dots, 6$ , the Bayesian  $\tilde{C}_i$  and classical  $\bar{C}_i$  estimators of the covariance operator are compared in terms of their integrated (truncated at term  $M$ ) empirical mean square errors, respectively denoted as  $\text{MSEB}_i$  and  $\text{MSEC}_i$ . That is, the following empirical quantities are computed for their comparison for the finite sample sizes  $n \in [1000, 10000]$  tested:

$$\begin{aligned} \text{MSEB}_i &= \sum_{k=1}^M \left[ \frac{1}{R} \sum_{l=1}^R [\tilde{\lambda}_{kn,i}^l - \lambda_{ki}]^2 \right], \\ \text{MSEC}_i &= \sum_{k=1}^M \left[ \frac{1}{R} \sum_{l=1}^R [\bar{\lambda}_{kn,i}^l - \lambda_{ki}]^2 \right] \end{aligned} \quad (59)$$

approximating the theoretical ones

$$\begin{aligned} E\|\tilde{C}_{ni} - C_i\|_{L^2([0,T])}^2, \\ E\|\bar{C}_{ni} - C_i\|_{L^2([0,T])}^2, \end{aligned} \quad (60)$$

where  $\tilde{\lambda}_{kn,i}^l$  and  $\bar{\lambda}_{kn,i}^l$  respectively denote the Bayes (36) and classical (37) estimates of  $\lambda_{ki}$ , based on the  $l$ th generation of a sample  $Y_{1i}^l, \dots, Y_{ni}^l$  of size  $n$  of the Gaussian Hilbert valued random variable  $Y_i$  of Model  $i$ , with  $l = 1, \dots, R$ . Specifically, we have considered  $H = L^2([0, T])$ , with  $T = 200$ , and  $R = 500$  to computing (59). In addition, for  $i = 1, \dots, 6$ , the absolute distances  $\text{AEB}_i$  and  $\text{AEC}_i$  to the limit  $2 \sum_{k=1}^M \lambda_{ki}^2$ , derived in Theorem 4, of the integrated

(truncated at term  $M$ ) empirical mean square errors of the Bayes and classical estimators are respectively computed as follows:

$$\begin{aligned} \text{AEB}_i &= \left| n \sum_{k=1}^M \left[ \frac{1}{R} \sum_{l=1}^R [\tilde{\lambda}_{kn,i}^l - \lambda_{ki}]^2 \right] - 2 \sum_{k=1}^M \lambda_{ki}^2 \right|, \quad i = 1, \dots, 6 \\ \text{AEC}_i &= \left| n \sum_{k=1}^M \left[ \frac{1}{R} \sum_{l=1}^R [\bar{\lambda}_{kn,i}^l - \lambda_{ki}]^2 \right] - 2 \sum_{k=1}^M \lambda_{ki}^2 \right|, \quad i = 1, \dots, 6, \end{aligned} \tag{61}$$

which they provide an empirical finite-dimensional approximation of

$$\begin{aligned} &\left| nE \|\tilde{C}_{ni} - C_i\|_{L^2([0,T])}^2 - 2 \sum_{k=1}^{\infty} \lambda_{ki}^2 \right|, \quad i = 1, \dots, 6 \\ &\left| nE \|\bar{C}_{ni} - C_i\|_{L^2([0,T])}^2 - 2 \sum_{k=1}^{\infty} \lambda_{ki}^2 \right|, \quad i = 1, \dots, 6. \end{aligned}$$

The behavior of the classical and Bayes estimators for the six covariance models introduced in (57) is investigated, when truncation is achieved at levels  $M = 5, 10, 25, 50, 75$ , for the sample sizes  $n \in [1000, 10000]$ , with discretization step size  $\text{DSS} = 500$ . The evaluated empirical approximations  $\text{MSEB}_i$  and  $\text{MSEC}_i$ ,  $i = 1, \dots, 6$ , of the functional mean-square errors of the Bayes and classical estimators display a hyperbolic rate of convergence to zero, with respect to the increasing sequence of sample sizes analyzed, in the six covariance models tested, for all the truncation orders studied (see Figure 5). It can be seen that the rate of convergence to zero of the eigenvalues of the covariance operator affects the rate of decrease of the integrated (truncated) empirical mean square errors for the finite sample sizes tested. Specifically, a slower rate of decrease to zero of the eigenvalues of the covariance operator leads to a slower rate of decrease to zero of  $\text{MSEB}_i$  and  $\text{MSEC}_i$ , as observed for  $i = 1$  and  $6$ , that is, for Model 1 and 6 (see top-left and bottom-right graphs in Figure 5). The fastest rate of decrease to zero of the eigenvalues of the covariance operator corresponds to Models 3 and 4, where the fastest decrease to zero is also observed for  $\text{MSEB}_i$  and  $\text{MSEC}_i$ , with  $i = 3, 4$  (see center-left and center-right graphs in Figure 5). The middle moderate case of decrease rate of the covariance operator eigenvalues corresponds to Models 2 and 5, where the same behavior is observed in the decrease rate of  $\text{MSEB}_i$  and  $\text{MSEC}_i$ , with  $i = 2, 5$  (see top-right and bottom-left graphs in Figure 5). In all the models studied, the Bayes and classical estimators of the covariance operator display an equivalent behavior, in the sense of the rate of convergence to zero of their integrated empirical mean square errors, for the truncation orders and finite sample sizes tested.

In Figure 6, for  $M = 5, 10, 25, 50, 75$ , the corresponding values of  $\text{AEB}_i$  and  $\text{AEC}_i$ ,  $i = 1, 2, 3, 4$ , are displayed. In particular,  $\text{AEB}_i$ ,  $i = 1, 2, 3, 4$ , are represented with solid line, and  $\text{AEC}_i$ ,  $i = 1, 2, 3, 4$ , are given with dotted lines. We observe that Models 1 and 2 display a similar behavior. For  $i = 1, 2$  (see top-panels in Figure 6),  $\text{AEB}_i$  and  $\text{AEC}_i$  are very close for all the truncation orders

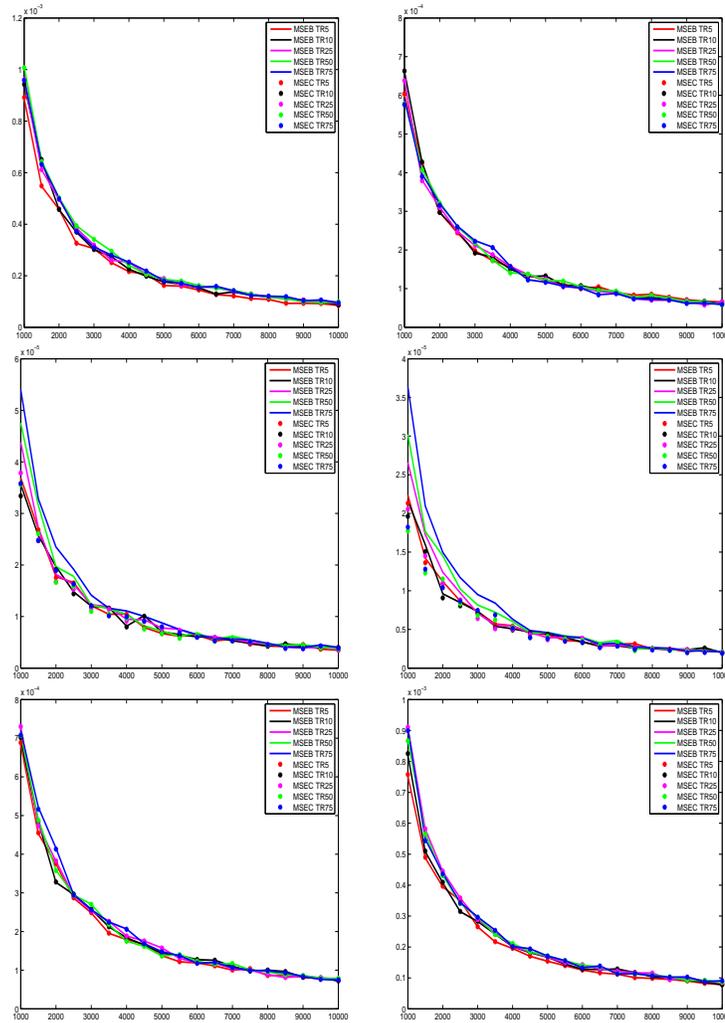


FIG 5. For  $M = 5, 10, 25, 50, 75$ , the corresponding values of  $MSEB_1$  and  $MSEC_1$  are represented at the top-left, of  $MSEB_2$  and  $MSEC_2$  at the top-right, of  $MSEB_3$  and  $MSEC_3$  at the center-left, of  $MSEB_4$  and  $MSEC_4$  at the center-right, of  $MSEB_5$  and  $MSEC_5$  at the bottom-left and of  $MSEB_6$  and  $MSEC_6$  at the bottom-right.

tested. Thus, Bayesian and classical estimators have relative efficiency close to one. For  $i = 3, 4$  (see bottom-panels in Figure 6),  $AEB_i$  and  $AEC_i$  are close, increasing the distance between their values for the truncation orders  $M = 50, 75$ . Note that, for these truncation orders, the classical estimator outperforms the Bayesian estimator, specially, for the smallest sample sizes studied. The relative efficiency of Bayes and classical estimators is farer from 1 in Model 5 and 6 than in the previous referred models (see Figure 7). The values of  $AEB_i$  and  $AEC_i$ ,

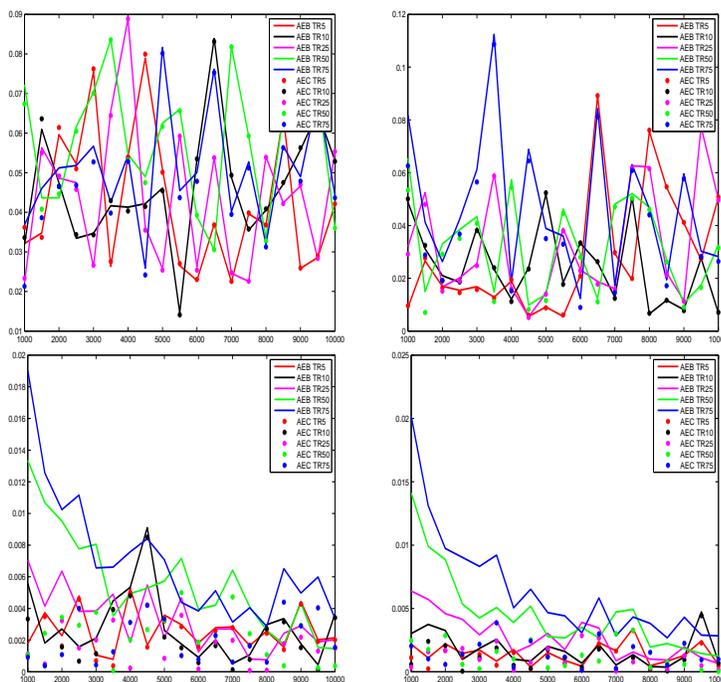


FIG 6. For  $M = 5, 10, 25, 50, 75$ , the corresponding values of  $AEB_1$  and  $AEC_1$  are represented at the top-left, of  $AEB_2$  and  $AEC_2$  at the top-right, of  $AEB_3$  and  $AEC_3$  at the bottom-left, and of  $AEB_4$  and  $AEC_4$  at the bottom-right.

$i = 5, 6$ , for the truncation orders  $M = 5, 10$ , are displayed at the left-hand side of Figure 7, and for  $M = 25, 50, 75$ , at the right-hand side of such a figure. For Models 5 and 6, the relative efficiency of Bayes and classical estimators is close to one for the truncation orders  $M = 5, 10$ , i.e.,  $AEB_i$  and  $AEC_i$  values for  $M = 5, 10$  are close when  $i = 5, 6$ . The classical estimator outperforms the Bayesian estimator for the truncation orders  $M = 25, 50, 75$ , when Model 5 and 6 are analyzed. Summarizing, the rate of convergence to zero of the eigenvalues of the covariance operators studied affects the relative efficiency of their Bayes and classical estimators for the highest truncation orders tested.

The theoretical results derived in Sections 3–5 provide sufficient conditions for the asymptotic equivalence of Bayes and classical estimators of the infinite-dimensional parameters, characterizing the  $l^2$  valued Poisson process and the Hilbert valued Gaussian models studied. In practice, for large finite sample sizes, the numerical results obtained in Section 6, in terms of the integrated (truncated) empirical mean square error, support the theoretical asymptotic results previously derived. In addition, for finite sample sizes, the rate of convergence to zero of the components of the infinite-dimensional parameters affects the rate of decrease of the integrated empirical mean square errors of their Bayes and

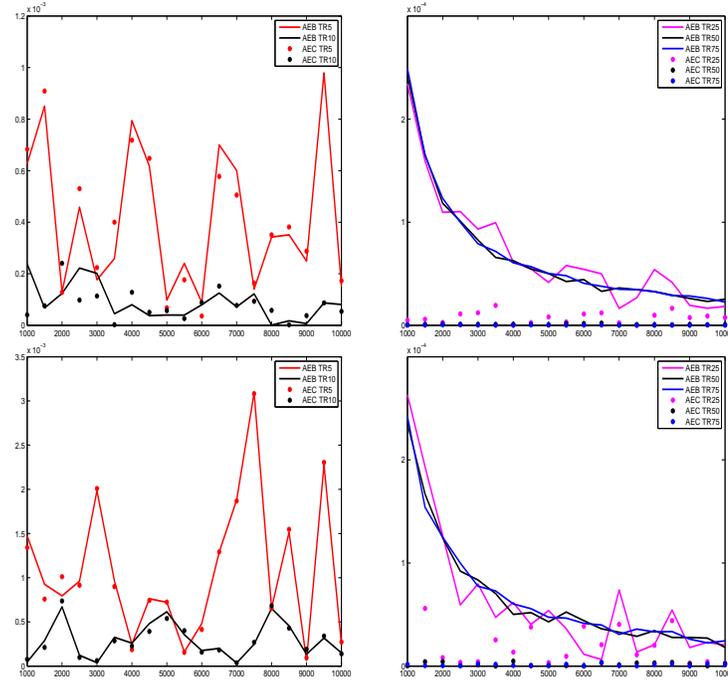


FIG 7. For  $M = 5, 10$ , the corresponding values of  $AEB_5$  and  $AEC_5$  are represented at the top-left, and of  $AEB_6$  and  $AEC_6$  at the bottom-left. For  $M = 25, 50, 75$ , the corresponding values of  $AEB_5$  and  $AEC_5$  are represented at the top-right, and of  $AEB_6$  and  $AEC_6$  at the bottom-right.

classical estimators. The absolute distance, to the theoretical limits derived in Theorems 1, 3 and 4, of the scaled, by the sample size, integrated empirical mean square errors of Bayes and classical estimators of the infinite-dimensional parameters considered constitutes a more unstable statistics, which requires a large number of repetitions to be run for its smoothing. In addition, it can be substantially affected by the truncation order in the case of the estimation of the covariance operator eigenvalues.

**7. Final comments**

The componentwise Bayesian and classical parameter estimators proposed for the infinite-dimensional vector of intensities, in the  $l^2$  valued Poisson process case, as well as for the functional mean, in the Hilbert valued Gaussian random variable context, are asymptotically equivalent under the conditions assumed. Under the same setting of conditions, for a given truncation order, they present an equivalent behavior in relation to their finite-sample-size efficiency. Note that the components of these estimators are linear functionals of the empirical projections of the data. However, the situation is different for the non-linear

componentwise Bayesian and classical covariance operator estimators. Although the asymptotic equivalence of both estimators is theoretically proved under suitable conditions, their relative efficiency for finite sample sizes, and for a given truncation order also depends on the rate of convergence to zero of the eigenvalues of the covariance operator.

In the case of correlated functional data, for example, in the case of observation of the functional values of an autoregressive Hilbertian process of order one (ARH(1) process), the problem becomes more complex. In the Bayesian autocorrelation operator estimation two priors can be considered related to the Gaussian and Beta families. In the last case, the conjugate Beta prior for the Gaussian likelihood leads to the definition of two possible Bayes estimators of the autocorrelation operator. Sufficient conditions can be derived to ensure a better asymptotic performance of the componentwise Bayesian estimator against the classical one in both cases. The rate of convergence to zero of the eigenvalues of the autocorrelation operator also affects the finite-sample size comparison of these estimators for a given truncation order in the two referred cases. That is, it affects the rate of decrease of the corresponding integrated (truncated) empirical mean square errors of the Bayes and classical estimators of the autocorrelation operator, but this subject will be addressed in a subsequent paper.

## Acknowledgments

This work has been supported in part by project MTM2012-32674 (co-funded with FEDER funds).

## References

- [1] ANTONIADIS, A. and SAPATINAS, T. (2003). Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *Journal of Multivariate Analysis* **87** 133–158. [MR2007265](#)
- [2] BALI, J. L., BOENTE, G., TYLER, D. E. AND WANG, J.-L. (2011). Robust functional principal components: A projection-pursuit approach. **39**, 2795–3443. [MR3012394](#)
- [3] BASSE, M., DIOP, A. and DABO-NIANG, S. (2008). Mean square properties of a class of kernel density estimates for spatial functional random variables. *Annales De L'I.S.U.P. Publications de l'Institut de Statistique de l'Université de Paris*. [MR2435043](#)
- [4] BESSE, P. C. and CARDOT, H. (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Canad. J. Statist.* **24** 467–487. [MR1438418](#)
- [5] BLANKE, D. and BOSQ, D. (2014). Bayesian prediction for stochastic processes. To appear in *Sankhya*.
- [6] BOSQ, D. (2000). *Linear Processes in Function Spaces*. Springer-Verlag, New York. [MR1783138](#)

- [7] BOSQ, D. and BLANKE, D. (2007). *Inference and Predictions in Large Dimensions*. John Wiley, Chichester, UK [MR2364006](#)
- [8] DA PRATO, G. and ZABCZYK, J. (2002). *Second Order Partial Differential Equations in Hilbert spaces*. Cambridge University Press, New York. [MR1985790](#)
- [9] DAUTRAY, R. and LIONS, J.-L. (1990). *Mathematical Analysis and Numerical Methods for Science and Technology*. Volume 3: Spectral Theory and Applications. Springer ISBN 978-3-540-66099-6. [MR1064315](#)
- [10] FAHRMEIR, L. and KNEIB, T. (2011). *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*. Oxford University Press, New York. [MR2850683](#)
- [11] FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis*. Springer, New York. [MR2229687](#)
- [12] KHOSHNEVISAN, D. (2007). *Probability*. American Mathematical Society, UE. [MR2296582](#)
- [13] LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*. Springer-Verlag, New York. [MR1639875](#)
- [14] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis* (2nd edition). Springer-Verlag, New York. [MR2168993](#)
- [15] RUIZ-MEDINA, M. D. (2011). Spatial autoregressive and moving average Hilbertian processes. *J. Multiv. Anal.* **102** 292–305. [MR2739116](#)
- [16] RUIZ-MEDINA, M. D. (2012). Spatial functional prediction from spatial autoregressive Hilbertian processes. *Environmetrics* **23** 119–128. [MR2873789](#)
- [17] RUIZ-MEDINA, M. D. and SALMERÓN, R. (2010). Functional maximum-likelihood estimation of ARH(p) models. *Stoch. Env. Res. Risk. Asses.* **24** 131–146.
- [18] SHOJAIE, A. and MICHAILIDIS, G. (2009). Analysis of gene sets based on the underlying regulatory network. *J. Comp. Biol.* **16**, 407–26. [MR2487566](#)
- [19] SHOJAIE, A. and MICHAILIDIS, G. (2010). Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs. *Biometrika* **97** 519–538. [MR2672481](#)
- [20] WALSH, J. B. (1984). Regularity properties of a stochastic partial differential equation. In *Seminar on Stochastic Processes* (Gainestville, FL, 1983; Progress Prob. Statist. 7), Birkhäuser, Boston, MA, pp. 257–290. [MR0902420](#)