Statistical Theory and Related Fields

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Power-expected-posterior prior Bayes factor consistency for nested linear models with increasing dimensions

D. Fouskakis, J. K. Innocent & L. Pericchi

To cite this article: D. Fouskakis, J. K. Innocent & L. Pericchi (2020) Power-expected-posterior prior Bayes factor consistency for nested linear models with increasing dimensions, Statistical Theory and Related Fields, 4:2, 162-171, DOI: 10.1080/24754269.2020.1719355

To link to this article: https://doi.org/10.1080/24754269.2020.1719355

Published online: 30 Jan 2020.

STATISTICAL THEORY AND RELATED FIELDS
2020, VOL. 4, NO. 2, 162–171
https://doi.org/10.1080/24754269.2020.1719355

Taylor & Francis
Taylor & Francis Group

Power-expected-posterior prior Bayes factor consistency for nested linear models with increasing dimensions

D. Fouskakis [a], J. K. Innocent [b] and L. Pericchi [b]

[a]Department of Mathematics, National Technical University of Athens, Athens, Greece; [b]Department of Mathematics, University of Puerto Rico, San Juan, USA

ABSTRACT
The power-expected-posterior prior is used in this paper for comparing nested linear models. The asymptotic behaviour of the method is investigated for different values of the power parameter of the prior. Focus is given on the consistency of the Bayes factor of comparing the full model M_p versus a generic submodel M_ℓ. In each case, we allow the true generating model to be either M_p or M_ℓ and we keep the dimension of M_ℓ fixed, while the dimension of M_p can be either fixed or (grow as) $O(n)$, with n denoting the sample size.

ARTICLE HISTORY
Received 13 May 2019
Revised 24 October 2019
Accepted 14 January 2020

KEYWORDS
Bayesian model selection; Bayes factor; consistency; expected-posterior prior; Gaussian linear models; increasing dimension; power-expected-posterior prior

1. Introduction

Pérez and Berger (2002) developed priors for objective Bayesian model comparison, through the utilisation of the device of 'imaginary training samples'. The *expected-posterior prior (EPP)* for the parameter under a model is an expectation of the posterior distribution given imaginary observations $\boldsymbol{y}^*$ of size n^*. The expectation is taken with respect to a suitable probability measure of a reference model M_0, while the posterior distribution is computed via Bayes's theorem starting from a default, typically improper, prior. One of the advantages of using EPPs is that impropriety of baseline priors causes no indeterminacy in the computation of Bayes factors. On the other hand, the EPPs depend on the training sample size and particularly in variable selection problems, imaginary design matrices should also be introduced, under each competing model, and therefore the resulting prior will further depend on this choice (for a detailed discussion on this issue, see Fouskakis, Ntzoufras, & Draper, 2015). The selection of a *minimal training sample*, of size n^*, has been proposed (see, for example, Berger & Pericchi, 2004), to make the information content of the prior as small as possible, and this is an appealing idea. But even under this set-up, the resulting prior can be influential when the sample size n is not much larger than the total number of parameters under the full model (see Fouskakis et al., 2015).

The *power-expected-posterior (PEP)* prior, introduced by Fouskakis et al. (2015), is an objective prior which amalgamates ideas from the power prior (Ibrahim & Chen, 2000), the expected-posterior prior (Pérez & Berger, 2002) and the unit-information-prior approach of Kass and Wasserman (1995) to simultaneously (a) produce a minimally informative prior and (b) diminish the effect of training samples under the EPP methodology. The main idea is to substitute the likelihood by a *density-normalised version of a power-likelihood* in EPP. Fouskakis et al. (2015) and Fouskakis and Ntzoufras (2016b) studied in detailed the PEP priors under the variable selection problem in Gaussian regression models. In the first paper, they introduced the PEP prior by considering as parameter of interest both the coefficients of the model and the error variance while in the second paper they studied the conditional version of PEP, named PCEP, where they considered only the coefficients as the parameter of interest and the error variance as a common nuisance parameter. Here we focus in the former case. Under this approach, for every model M_ℓ in $\mathcal{M}$ (the set of all models under consideration) the sampling distribution $f_\ell(\cdot \mid \boldsymbol{\beta}_\ell, \sigma_\ell^2)$ is specified by

$$(\boldsymbol{Y} \mid \mathbf{X}_\ell, \boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell) \sim N_n(\mathbf{X}_\ell \boldsymbol{\beta}_\ell, \sigma_\ell^2 \mathbf{I}_n), \qquad (1)$$

where $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ is a vector containing the responses for all subjects, $\mathbf{X}_\ell$ is an $n \times d_\ell$ design matrix containing the values of the explanatory variables in its columns, $\mathbf{I}_n$ is the $n \times n$ identity matrix, $\boldsymbol{\beta}_\ell$ is a vector of length d_ℓ summarising the effects of the covariates in model M_ℓ on the response $\boldsymbol{Y}$ and σ_ℓ^2 is the error variance for model M_ℓ. Finally, by p we denote the

CONTACT D. Fouskakis fouskakis@math.ntua.gr

total number of the explanatory variables under consideration and by M_p the full model, including all p covariates.

Furthermore, we denote by $\pi_\ell^{\mathrm{N}}(\boldsymbol{\beta}_\ell, \sigma_\ell^2)$ the baseline prior to the parameters of model M_ℓ. Here we use the independence Jeffreys prior (or reference prior) as the baseline prior distribution. Hence, for any $M_\ell \in \mathcal{M}$, we have

$$\pi_\ell^N(\boldsymbol{\beta}_\ell, \sigma_\ell^2) = \frac{c_\ell}{\sigma_\ell^2}, \tag{2}$$

where c_ℓ is an unknown normalising constant.

We assume that in $\mathcal{M}$ there exists a model M_0, with parameters $\boldsymbol{\beta}_0$ and σ_0^2, sampling distribution $f_0(\cdot \mid \boldsymbol{\beta}_0, \sigma_0^2)$ and baseline prior $\pi_0^{\mathrm{N}}(\boldsymbol{\beta}_0, \sigma_0^2) \propto \sigma_0^{-2}$, which is nested into each of the remaining models and we consider it as a reference model. This is the typical case in the variable selection problem, studied in this paper. Given then a set of imaginary data $\mathbf{y}^* = (y_1^*, \ldots, y_{n^*}^*)^{\mathrm{T}}$ and a positive power parameter δ, that is used to regulate, essentially, the contribution of the imaginary data on the 'final' prior, we introduce the density-normalised power-likelihood, under model M_ℓ, given by

$$f_\ell(\mathbf{y}^* \mid \boldsymbol{\beta}_\ell, \sigma_\ell^2, \delta, \mathbf{X}_\ell^*) = \frac{f_\ell(\mathbf{y}^* \mid \boldsymbol{\beta}_\ell, \sigma_\ell^2, \mathbf{X}_\ell^*)^{1/\delta}}{\int f_\ell(\mathbf{y}^* \mid \boldsymbol{\beta}_\ell, \sigma_\ell^2, \mathbf{X}_\ell^*)^{1/\delta}\, \mathrm{d}\mathbf{y}^*}. \tag{3}$$

The above density-normalised power-likelihood is still a normal distribution with variance inflated by a factor of δ; in the above, $\mathbf{X}_\ell^*$ denotes the imaginary design matrix under model M_ℓ. In a similar manner, under the reference model, the density-normalised power-likelihood takes the form of (3) but using now the likelihood $f_0(\mathbf{y}^* \mid \boldsymbol{\beta}_0, \sigma_0^2, \mathbf{X}_0^*)$ of M_0.

In order to apply the PEP methodology, the density-normalised power-likelihood (3) is used to evaluate, under the imaginary data and the baseline prior, the prior predictive distribution $m_0^{\mathrm{N}}(\mathbf{y}^* \mid \delta, \mathbf{X}_0^*)$ of model M_0 as well as the posterior distribution of the parameters of model M_ℓ

$$\pi_\ell^{\mathrm{N}}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 \mid \mathbf{y}^*, \delta, \mathbf{X}_\ell^*) = \frac{f_\ell(\mathbf{y}^* \mid \boldsymbol{\beta}_\ell, \sigma_\ell^2, \delta, \mathbf{X}_\ell^*)\pi_\ell^{\mathrm{N}}(\boldsymbol{\beta}_\ell, \sigma_\ell^2)}{m_\ell^{\mathrm{N}}(\mathbf{y}^* \mid \delta, \mathbf{X}_\ell^*)}, \tag{4}$$

where

$$m_j^{\mathrm{N}}(\mathbf{y}^* \mid \delta, \mathbf{X}_j^*) = \int\int f_j(\mathbf{y}^* \mid \boldsymbol{\beta}_j, \sigma_j^2, \delta, \mathbf{X}_j^*)\pi_j^{\mathrm{N}}(\boldsymbol{\beta}_j, \sigma_j^2)\, \mathrm{d}\boldsymbol{\beta}_j\, \mathrm{d}\sigma_j^2, \tag{5}$$

is the prior predictive distribution of model M_j for $j = \ell, 0$.

Finally, the imposed prior for the parameters of any model M_ℓ has the following form

$$\pi_\ell^{\mathrm{PEP}}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 \mid \delta, \mathbf{X}_\ell^*) = \int \pi_\ell^{\mathrm{N}}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 \mid \mathbf{y}^*, \delta, \mathbf{X}_\ell^*) m_0^{\mathrm{N}}(\mathbf{y}^* \mid \delta, \mathbf{X}_0^*)\, \mathrm{d}\mathbf{y}^*. \tag{6}$$

The default choice for δ is to set it equal to n^*, i.e. the sample size of the imaginary data, so that the overall information of the imaginary data in the posterior is equal to one data point. Furthermore, setting $n^* = n$ and, consequently, the design matrix of the imaginary data $\mathbf{X}_\ell^* \equiv \mathbf{X}_\ell$ simplifies significantly the overwhelming computations required when considering all possible 'minimal' training samples (Pérez & Berger, 2002) while it also avoids the complicated issue (in some cases) of defining the size of the minimal training samples (Berger & Pericchi, 2004). In addition, under the choice $n^* = n$, the PEP prior remains relatively noninformative even for models with dimension close to the sample size n, while the effect on the evaluation of each model is minimal since the resulting Bayes factors are robust over different values of n^*. Detailed information about the default specifications of the PEP prior is provided in Fouskakis et al. (2015). Finally, the null model (with no explanatory variables) is a standard choice for the reference model in regression problems; see, for example, Pérez and Berger (2002). In the above definition of PEP prior, the power parameter can also be model depended, and denoted by δ_ℓ.

Fouskakis and Ntzoufras (2016a) proved the consistency of the Bayes factor when using the PEP methodology, with the independence Jeffreys as a baseline prior, for Gaussian linear models, under very mild conditions on the design matrix, when the dimension of each model is fixed, the size of the training sample is equal to the sample size n and the power parameter is also set equal to n. In a similar manner as in Fouskakis and Ntzoufras (2016a), when comparing the full model M_p to a reduced model M_ℓ, the Bayes factor under the PEP prior is given by

$$BF_{p\ell}^{PEP} = 2\frac{\Gamma(n-p)}{\Gamma^2(\frac{n-p}{2})}\int_0^{\pi/2} \times \frac{(\sin\varphi)^{n-d_\ell-1}(\cos\varphi)^{n-p-1}(\delta_\ell + \sin^2\varphi)^{(n-p)/2}}{\left(\delta_\ell \frac{RSS_p}{RSS_\ell} + \sin^2\varphi\right)^{(n-d_\ell)/2}} \mathrm{d}\varphi, \tag{7}$$

with RSS_j denoting the residual sum of squares of model M_j ($j = \ell, p$). For large n, we can approximate the Bayes factor given in (7) as

$$BF_{p\ell}^{PEP} \approx \left(\frac{1}{\rho_{\ell p}}\right)^{(n-d_\ell)/2} \left(\frac{1}{\delta_\ell}\right)^{(p-d_\ell)/2} \left(\frac{1}{2}\right)^{(p-d_\ell)/2}, \tag{8}$$

if p is fixed constant; and as

$$BF_{p\ell}^{PEP} \approx \left(\frac{1}{\rho_{\ell p}}\right)^{(rp-d_\ell)/2} \left(\frac{1}{\delta_\ell}\right)^{(p-d_\ell)/2} 2^{(2(r-1)p-1)/2} \times \frac{(r-1)^{(r-1)p/2} r^{(rp-d_\ell-1)/2}}{(2r-1)^{((2r-1)p-d_\ell-1)/2}}, \quad (9)$$

if p increases as n grows to infinity and $(n-p)$ grows to infinity, with rate $r > 1$ so that $n = r \times p$ (for a detailed proof of (8) and (9) see Innocent, 2016).

In the rest of the paper, we denote by

$$\rho_{\ell p} = \frac{RSS_p}{RSS_\ell}$$

and by

$$\epsilon_{p\ell} = \frac{1}{\sigma_T^2} \boldsymbol{\beta}_T^t \frac{\mathbf{X}_T^t(\mathbf{I}_n - \mathbf{H}_\ell)\mathbf{X}_T}{n} \boldsymbol{\beta}_T,$$

where M_T denotes the 'true' model and $\mathbf{H}_\ell$ the hat matrix of model M_ℓ (see Casella, Girón, Martínez, & Moreno, 2009). Since the reduced model M_ℓ is nested in the full model M_p, we have that $\rho_{\ell p} \in (0, 1]$.

Finally, the following results hold, as n increases, with respect to the distribution and the limiting behaviour of the statistic $\rho_{\ell p}$ (see Girón, Moreno, & Casella, 2010):

- If $\dim(M_\ell) = d_\ell = O(1)$ and $\dim(M_p) = p = O(1)$:
  - When sampling from model M_ℓ, the distribution of the statistic $\rho_{\ell p}$ is the central beta distribution $Be((n-p)/2, (p-d\ell)/2)$ and

    $$\lim_{n\to+\infty} \rho_{\ell p} = 1.$$

  - When sampling from model M_p, the distribution of the statistic $\rho_{\ell p}$ is the non-central beta distribution $Be((n-p)/2, (p-d_\ell)/2, 0, n\epsilon_{p\ell})$ and

    $$\lim_{n\to+\infty} \rho_{\ell p} = \frac{1}{1+\epsilon},$$

    with

    $$\lim_{n\to+\infty} \epsilon_{p\ell} = \epsilon > 0.$$

- If $\dim(M_\ell) = d_\ell = O(1)$ and $\dim(M_p) = p = O(n)$ with $r = \lim_{n,p\to+\infty} \frac{n}{p} > 1, p > d_\ell > 1$:
  - When sampling from model M_ℓ, the distribution of the statistic $\rho_{\ell p}$ is the central beta distribution $Be(p(r-1)/2, (p-d_\ell)/2)$ and

    $$\lim_{n\to+\infty} \rho_{\ell p} = \frac{r-1}{r}, r > 1.$$

  - When sampling from model M_p the distribution of the statistic $\rho_{\ell p}$ is the non-central beta distribution $Be(p(r-1)/2, (p-d_\ell)/2, 0, rp\epsilon_{p\ell})$ and

    $$\lim_{n\to+\infty} \rho_{\ell p} = \frac{r-1}{r(1+\epsilon)},$$

    where

    $$\lim_{n\to+\infty} \epsilon_{p\ell} = \epsilon > 0.$$

In this paper, we examine the consistency of the Bayes factor, for nested normal linear models, under the PEP methodology, using the pair of models M_ℓ and M_p. The number of parameters of the simpler model M_ℓ is always fixed, while for the full model is of order $O(n^\alpha)$, where $\alpha \in \{0, 1\}$. We investigate the effect of the power parameter δ_ℓ by examining four different scenarios. In each case, the 'true' model is set equal to either M_ℓ or M_p.

2. Bayes factor consistency under power-expected-posterior priors

In what follows we set the size of the training sample n^* equal to the sample size n as in Fouskakis et al. (2015).

2.1. When the power $\delta_\ell = n$

First, we consider the case where the power parameter is set equal to the sample size n, and studying the consistency when the dimension p of the full model M_p is either a fixed constant number or large and goes to infinity.

Then (7) becomes:

$$BF_{p\ell}^{PEP} = 2\frac{\Gamma(n-p)}{\Gamma^2(\frac{n-p}{2})} \int_0^{\pi/2} \times \frac{(\sin\varphi)^{n-d_\ell-1}(\cos\varphi)^{n-p-1}(n+\sin^2\varphi)^{(n-p)/2}}{(n\rho_{\ell p}+\sin^2\varphi)^{(n-d_\ell)/2}} d\varphi. \quad (10)$$

2.1.1. When $\dim(M_\ell) = O(1)$ and $\dim(M_p) = O(1)$

Theorem 1: *Let the sample size n increases and being strictly greater than the dimension of the full model M_p. Furthermore, suppose that the dimension of both models, under consideration, are fixed non-negative natural numbers, i.e. $\dim(M_\ell) = d_\ell = O(1)$ and $\dim(M_p) = p = O(1)$, where $p > d_\ell > 1$. Under the condition $\delta_\ell = n$, when sampling from model M_j, where j is either ℓ or p we have:*

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = \begin{cases} 0 & \text{if } j = \ell \\ +\infty & \text{if } j = p \end{cases}.$$

Proof: For $\delta_\ell = n$, (8) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{1}{2n}\right)^{(p-d_\ell)/2} \left(\frac{1}{\rho_{\ell p}}\right)^{(n-d_\ell)/2}. \quad (11)$$

(a) Suppose that the Reduced Model M_ℓ is true

Using the asymptotic results of $\rho_{\ell p}$ given in Section 1, (11) becomes:

$$BF_{p\ell}^{PEP} \approx \left(\frac{1}{2n}\right)^{(p-d\ell)/2}. \tag{12}$$

Since p and d_ℓ are constants and n goes to infinity we get

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = 0.$$

Thus, the Bayes factor of the full model M_p versus the reduced model M_ℓ is consistent under the reduced model M_ℓ.

(b) Suppose that the Full Model M_p is true

Using the asymptotic results of $\rho_{\ell p}$ given in Section 1, (11) becomes:

$$BF_{p\ell}^{PEP} \approx \left(\frac{1}{2n}\right)^{(p-d_\ell)/2}(1+\epsilon)^{n/2} \approx \left(\frac{1}{2}\right)^{(p-d_\ell)/2} \times \mathrm{e}^{-n((p-d_\ell)/2)(\log(n)/n)+(n/2)\log(1+\epsilon)}. \tag{13}$$

Thus

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = \mathrm{e}^{\lim_{n\to+\infty}(n/2)\log(1+\epsilon)} = +\infty,$$

since $\epsilon > 0$, and $(n/2)\log(1+\epsilon) \to +\infty$ as $n \to +\infty$. Therefore, the Bayes factor of the full model M_p versus the reduced model M_ℓ is consistent when sampling from the full model M_p. ■

2.1.2. When $\dim(M_\ell) = O(1)$ and $\dim(M_p) = O(n)$

Theorem 2: *Let $\delta_\ell = n$ and suppose that the reduced model M_ℓ has a fixed number of parameters, i.e. $\dim(M_\ell) = d_\ell = O(1)$, as the simple size n increases, and in the full model M_p the number of parameters increase with rate $\dim(M_p) = p = O(n)$ with*

$$r = \lim_{n,p\to+\infty} \frac{n}{p} > 1, \quad p > d_\ell > 1.$$

Then:

(1) *When sampling from model M_ℓ*

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = 0.$$

(2) *When sampling from model M_p*

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = \begin{cases} 0 & \text{if } r > 1 \text{ is a fixed constant} \\ \begin{cases} 0 & \text{if } \lim_{n\to+\infty}\epsilon_{p\ell} < \epsilon_p^2(r) \\ +\infty & \text{if } \lim_{n\to+\infty}\epsilon_{p\ell} \geq \epsilon_p^2(r) \end{cases} & \text{if } r > 1 \text{ is a large number} \end{cases}$$

for some function ϵ_p^2 given by $\epsilon_p^2(r) : (1,+\infty) \longrightarrow \mathbb{R}, r \longmapsto (2rp)^{1/r} - 1$.

Proof: By replacing $n \approx rp$, and $\delta_\ell = rp$, (9), becomes

$$BF_{p\ell}^{PEP} \approx \left[\frac{1}{2(r-1)p}\right]^{(p-d_\ell)/2} \times \left(\frac{2r}{2r-1}\right)^{((2r-1)p-d_\ell-1)/2}\left(\frac{r-1}{r\rho_{\ell p}}\right)^{(pr-d_\ell)/2}. \tag{14}$$

(a) Suppose that the Reduced Model M_ℓ is true

Using the asymptotic results of $\rho_{\ell p}$ given in Section 1, (14) becomes

$$BF_{p\ell}^{PEP} \approx \left[\frac{1}{2(r-1)p}\right]^{(p-d_\ell)/2}\left(\frac{2r}{2r-1}\right)^{((2r-1)p-d_\ell-1)/2}$$

and then

$$BF_{p\ell}^{PEP} \approx \left[\frac{1}{2(r-1)p}\left(\frac{2r}{2r-1}\right)^{2r-1}\right]^{p/2} \times \left(\frac{(2r-1)(r-1)p}{r}\right)^{d_\ell/2}\left(1-\frac{1}{2r}\right)^{1/2}.$$

So for large value of p, we have

$$BF_{p\ell}^{PEP} \approx \left[\frac{1}{2(r-1)p}\left(\frac{2r}{2r-1}\right)^{2r-1}\right]^{p/2}$$

and then

$$BF_{p\ell}^{PEP} \approx \begin{cases} \left(\frac{1}{p}\right)^{p/2} & \text{if } r > 1 \text{ is a fixed constant} \\ \left(\frac{1}{2rp}\right)^{p/2} & \text{if } r \text{ is a large number} \end{cases}. \tag{15}$$

In both cases, for large p, we get

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = 0,$$

since

$$\lim_{n\to+\infty}\left(\frac{1}{p}\right)^{p/2} = \lim_{n\to+\infty}\exp\left(-\frac{p}{2}\log p\right) = 0.$$

Thus the Bayes factor of the full model M_p against the reduced model M_ℓ is consistent under the reduced model M_ℓ.

(b) Suppose that the Full Model M_p is true

Using the asymptotic results of $\rho_{\ell p}$ given in Section 1, (14) becomes

$$BF_{p\ell}^{PEP} \approx \left[\frac{1}{2(r-1)p}\right]^{(p-d_\ell)/2} \times \left(\frac{2r}{2r-1}\right)^{((2r-1)p-d_\ell-1)/2}(1+\epsilon)^{(rp-d_\ell)/2}.$$

So for large p, we have

$$BF_{p\ell}^{PEP} \approx \begin{cases} \left(\dfrac{1}{p}\right)^{p/2} & \text{if } r > 1 \text{ is a fixed constant} \\ \left[\dfrac{(1+\epsilon)^r}{2rp}\right]^{p/2} (2rp)^{d_\ell/2} & \\ & \text{if } r \text{ is a large number} \end{cases}$$

and then

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} \approx \begin{cases} \left(\dfrac{1}{p}\right)^{p/2} & \text{if } r > 1 \text{ is a fixed constant} \\ \begin{cases} (2rp)^{d_\ell/2} & \text{if } \dfrac{(1+\epsilon)^r}{2rp} = 1 \\ \left[\dfrac{(1+\epsilon)^r}{2rp}\right]^{p/2} & \text{if } \dfrac{(1+\epsilon)^r}{2rp} \neq 1 \end{cases} & \\ \text{if } r > 1 \text{ is a large number} & \end{cases}.$$

Solving the equation $(1+\epsilon)^r/2rp = 1$ for ϵ, we get $\epsilon = (2rp)^{1/r} - 1$. Therefore using the function $\epsilon_p^2(r) : (1, +\infty) \longrightarrow \mathbb{R}, r \longmapsto (2rp)^{1/r} - 1$ we have

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = \begin{cases} 0 \quad \text{if } r > 1 \text{ is a fixed constant} \\ \begin{cases} 0 & \text{if } \lim_{n\to+\infty} \epsilon_{p\ell} < \epsilon_p^2(r) \\ +\infty & \text{if } \lim_{n\to+\infty}\epsilon_{p\ell} \geq \epsilon_p^2(r) \end{cases} \\ \text{if } r > 1 \text{ is a large number} \end{cases}.$$

Thus, the Bayes factor of the full model M_p versus the reduced model M_ℓ is consistent under the full model M_p if and only if $\lim_{n\to+\infty} \epsilon_{p\ell} \geq \epsilon_p^2(r)$ when r is large and goes to infinity. ∎

2.2. When the power $\delta_\ell = (n - p)$

Second, we consider the case where the power $\delta_\ell = (n - p)$ and studying the consistency when the dimension p of the full model M_p is either a fixed constant number or large and goes to infinity. Then (7) becomes:

$$BF_{p\ell}^{PEP} = 2\frac{\Gamma(n-p)}{\Gamma^2(\frac{n-p}{2})} \int_0^{\pi/2} \times \frac{(\sin\varphi)^{n-d_\ell-1}(\cos\varphi)^{n-p-1}((n-p) + \sin^2\varphi)^{(n-p)/2}}{\left[(n-p)\rho_{ip} + sin^2\varphi\right]^{(n-d_\ell)/2}} \mathrm{d}\varphi$$

2.2.1. When $\dim(M_\ell) = O(1)$ and $\dim(M_p) = O(1)$

Let the simple size n increases and being strictly greater than the dimension of the full model M_p. Furthermore, suppose that the dimension of both models, under consideration, are fixed non-negative natural numbers, i.e. $\dim(M_\ell) = d_\ell = O(1)$ and $\dim(M_p) = p = O(1)$, where $p > d_\ell > 1$.

For $\delta_\ell = (n-p)$, (8) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{1}{2}\right)^{(p-d_\ell)/2} \left(\frac{1}{n-p}\right)^{(p-d_\ell)/2} \left(\frac{1}{\rho_{\ell p}}\right)^{(n-d_\ell)/2},$$

and then since p and d_ℓ are fixed constants and for large values of n, we get

$$BF_{p\ell}^{PEP} \approx \left(\frac{1}{n}\right)^{p/2} \left(\frac{1}{\rho_{\ell p}}\right)^{n/2}. \tag{16}$$

Working as in the proof of Theorem 2.1, we conclude that the Bayes factor of the full model M_p versus the reduced model M_ℓ is consistent when sampling from either models.

2.2.2. When $\dim(M_\ell) = O(1)$ and $\dim(M_p) = O(n)$

Theorem 3: *Let $\delta_\ell = (n-p)$ and suppose that the reduced model M_ℓ has a fixed number of parameters, i.e. $\dim(M_\ell) = d_\ell = O(1)$, as the simple size n increases, and in the full model M_p the number of parameters increase with rate $\dim(M_p) = p = O(n)$ with*

$$r = \lim_{n,p\to+\infty} \frac{n}{p} > 1, \quad p > d_\ell > 1.$$

Then:

(1) *When sampling from model M_ℓ*

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = 0.$$

(2) *When sampling from model M_p*

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = \begin{cases} 0 \\ \quad \text{if } r > 1 \text{ is a fixed constant} \\ \begin{cases} 0 & \text{if } \lim_{n\to+\infty}\epsilon_{p\ell} < \epsilon_p^2(r) \\ +\infty & \text{if } \lim_{n\to+\infty}\epsilon_{p\ell} \geq \epsilon_p^2(r) \end{cases} \\ \quad \text{if } r > 1 \text{ is a large number} \end{cases}$$

for some function ϵ_p^2 given by $\epsilon_p^2(r) : (1, +\infty) \longrightarrow \mathbb{R}, r \longmapsto (2rp)^{1/r} - 1$.

Proof: By replacing $n \approx rp$, and $\delta_\ell = rp - p$, (9), becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{2r}{2r-1}\right)^{((2r-1)p-d_\ell-1)/2} \times \left(\frac{r}{2p(r-1)^2}\right)^{(p-d_\ell)/2} \left(\frac{r-1}{r\rho_{\ell p}}\right)^{(pr-d_\ell)/2}. \tag{17}$$

(a) Suppose that the Reduced Model M_ℓ is true

Using the asymptotic results of $\rho_{\ell p}$ given in Section 1, (17) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{2r}{2r-1}\right)^{((2r-1)p-d_\ell-1)/2} \times \left(\frac{r}{2p(r-1)^2}\right)^{(p-d_\ell)/2}.$$

So for large value of p, we have

$$BF_{p\ell}^{PEP} \approx \begin{cases} \left(\frac{1}{p}\right)^{p/2} & \text{if } r>1 \text{ is a fixed constant} \\ \left(\frac{1}{2rp}\right)^{p/2} & \text{if } r \text{ is a large number} \end{cases}.$$

In both cases, for large p, we get

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = 0,$$

Thus the Bayes factor of the full model M_p against the reduced model M_ℓ is consistent under the reduced model M_ℓ.

(b) Suppose that the Full Model M_p is true

Using the asymptotic results of $\rho_{\ell p}$ given in Section 1, (17) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{r}{2p(r-1)^2}\right)^{(p-d_\ell)/2} \times \left(\frac{2r}{2r-1}\right)^{((2r-1)p-d_\ell-1)/2} (1+\epsilon)^{(rp-d_\ell)/2},$$

or

$$BF_{p\ell}^{PEP} \approx \left[\left(\frac{2r}{2r-1}\right)^{2r-1} \frac{r(1+\epsilon)^r}{2p(r-1)^2}\right]^{p/2} \times \left(\frac{(2r-1)(r-1)^2(1+\epsilon)^{-1}p}{r^2}\right)^{d_\ell/2} \times \left(1-\frac{1}{2r}\right)^{1/2}. \tag{18}$$

So for large p, we have

$$BF_{p\ell}^{PEP} \approx \begin{cases} \left(\frac{1}{p}\right)^{p/2} & \text{if } r>1 \text{ is a fixed constant} \\ \left(\frac{(1+\epsilon)^r}{2rp}\right)^{p/2} (2rp)^{d_\ell/2} & \\ \text{if } r \text{ is a large number} & \end{cases}$$

Thus working as in the proof of Theorem 2.2 we conclude that the Bayes factor of the full model M_p versus the reduced model M_ℓ is consistent under the full model M_p if and only if $\lim_{n\to+\infty} \epsilon_{p\ell} \ge \epsilon_p^2(r)$ when r is large and goes to infinity. ■

2.3. When the power $\delta_\ell = p$

Third, we consider the case where the power is equal to the dimension of the full model and studying the consistency when the dimension $p = dim(M_p)$ of the full model M_p is either a fixed constant number or large and goes to infinity.

Under this set-up, (7) becomes:

$$BF_{p\ell}^{PEP} = 2\frac{\Gamma(n-p)}{\Gamma^2(\frac{n-p}{2})} \times \int_0^{\pi/2} \frac{(\sin\varphi)^{n-d_\ell-1}(\cos\varphi)^{n-p-1} \, (p+\sin^2\varphi)^{(n-p)/2}}{(p\rho_{\ell p}+\sin^2\varphi)^{(n-d_\ell)/2}} \mathrm{d}\varphi.$$

2.3.1. When $\dim(M_\ell) = O(1)$ and $\dim(M_p) = O(1)$

Theorem 4: *Let $\delta_\ell = p$ and the sample size n increases and being strictly greater than the dimension of the full model M_p. Furthermore, suppose that the dimension of both models, under consideration, are fixed nonnegative natural numbers, i.e. $\dim(M_\ell) = d_\ell = O(1)$ and $\dim(M_p) = p = O(1)$, where $p > d_\ell > 1$. Then when sampling from model M_j, where j is either ℓ or p we have:*

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = \begin{cases} Constant > 0 & \text{if } j=\ell \\ +\infty & \text{if } j=p \end{cases}.$$

Proof: For $\delta_\ell = p$, (8) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{1}{2p}\right)^{(p-d_\ell)/2} \left(\frac{1}{\rho_{\ell p}}\right)^{(n-d_\ell)/2} \tag{19}$$

Then we consider the following two cases.

(a) Suppose that the Reduced Model M_ℓ is true

Using the asymptotic results of $\rho_{\ell p}$ given in Section 1, (19) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{1}{2p}\right)^{(p-d\ell)/2}.$$

Since p and d_ℓ are constants, with $p > d_\ell > 1$, we get

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = \lim_{n\to+\infty} \left(\frac{1}{2p}\right)^{(p-d\ell)/2} = Constant > 0.$$

Thus, the Bayes factor of the full model M_p versus the reduced model M_ℓ is inconsistent under the reduced model M_ℓ.

(b) Suppose that the Full Model M_p is true

Using the asymptotic results of $\rho_{\ell p}$ given in Section 1, (19) becomes

$$BF_{p\ell}^{PEP} \approx e^{(n/2)\log(1+\epsilon)}.$$

and thus

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = +\infty.$$

Therefore, the Bayes factor of the full model M_p versus the reduced model M_ℓ is consistent when sampling from the full model M_p. ■

2.3.2. When $\dim(M_\ell) = O(1)$ and $\dim(M_p) = O(n)$

Theorem 5: *Let $\delta_\ell = p$ and suppose that the reduced model M_ℓ has a fixed number of parameters, i.e. $\dim(M_\ell) = d_\ell = O(1)$, as the simple size n increases, and in the full model M_p the number of parameters increase with rate $\dim(M_p) = p = O(n)$ with*

$$r = \lim_{n,p\to+\infty} \frac{n}{p} > 1, \quad p > d_\ell > 1.$$

Then:

(1) *When sampling from model M_ℓ*

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = 0.$$

(2) *When sampling from model M_p*

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = \begin{cases} 0 & \text{if } r > 1 \text{ is a fixed constant} \\ \begin{cases} 0 & \text{if } lim_{n\to+\infty}\epsilon_{p\ell} < \epsilon_p^1(r) \\ +\infty & \text{if } lim_{n\to+\infty}\epsilon_{p\ell} \geq \epsilon_p^1(r) \end{cases} \\ \quad \text{if } r > 1 \text{ is a large number} \end{cases}$$

for some function ϵ_p^1 given by $\epsilon_p^1(r) : (1, +\infty) \longrightarrow \mathbb{R}, r \longmapsto (2p)^{1/r} - 1$.

Proof: By replacing $n \approx rp$, and $\delta_\ell = p$, (9) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{r}{2(r-1)p}\right)^{(p-d_\ell)/2} \times \left(\frac{2r}{2r-1}\right)^{((2r-1)p-d_\ell-1)/2} \left(\frac{r-1}{r\rho_{\ell p}}\right)^{(pr-d_\ell)/2}. \tag{20}$$

(a) Suppose that the Reduced Model M_ℓ is true

Using the asymptotic results of $\rho_{\ell p}$ given in Section 1, (20) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{r}{2(r-1)p}\right)^{(p-d_\ell)/2} \times \left(\frac{2r}{2r-1}\right)^{((2r-1)p-d_\ell-1)/2}$$

and then

$$BF_{p\ell}^{PEP} \approx \left(\frac{r}{2(r-1)p}\left(\frac{2r}{2r-1}\right)^{2r-1}\right)^{p/2} \times \left(\frac{(2r-1)(r-1)p}{2r^2}\right)^{d_\ell/2}\left(1-\frac{1}{2r}\right)^{1/2}.$$

So for large value of p we have

$$BF_{p\ell}^{PEP} \approx \begin{cases} \left(\frac{1}{p}\right)^{p/2} & \text{if } r > 1 \text{ is a fixed constant} \\ \left(\frac{1}{2p}\right)^{p/2} & \text{if } r \text{ is a large number} \end{cases}. \tag{21}$$

In both cases, for large p, we get

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = 0.$$

Thus the Bayes factor of the full model M_p against the reduced model M_ℓ is consistent under the reduced model M_ℓ.

(b) Suppose that the Full Model M_p is true

Using the asymptotic results of $\rho_{\ell p}$ given in Section 1, (20) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{r}{2(r-1)p}\right)^{(p-d_\ell)/2} \times \left(\frac{2r}{2r-1}\right)^{((2r-1)p-d_\ell-1)/2} (1+\epsilon)^{(rp-d_\ell)/2}.$$

So for large p, we have

$$BF_{p\ell}^{PEP} \approx \begin{cases} \left(\frac{1}{p}\right)^{p/2} \\ \quad \text{if } r > 1 \text{ is a fixed constant} \\ \left(\frac{(1+\epsilon)^r}{2p}\right)^{p/2} (2p)^{d_\ell/2} \\ \quad \text{if } r \text{ is a large number} \end{cases}$$

and then

$$BF_{p\ell}^{PEP} \approx \begin{cases} \left(\frac{1}{p}\right)^{p/2} \\ \quad \text{if } r > 1 \text{ is a fixed constant} \\ \begin{cases} (2p)^{d_\ell/2} & \text{if } \frac{(1+\epsilon)^r}{2p} = 1 \\ \left(\frac{(1+\epsilon)^r}{2p}\right)^{p/2} & \text{if } \frac{(1+\epsilon)^r}{2p} \neq 1 \end{cases} \\ \quad \text{if } r > 1 \text{ is a large number} \end{cases}.$$

Solving the equation $(1+\epsilon)^r/2p = 1$ for ϵ, we get $\epsilon = (2p)^{1/r} - 1$. Therefore using the function $\epsilon_p^1(r) : (1, +\infty) \longrightarrow \mathbb{R}, r \longmapsto (2p)^{1/r} - 1$ we have

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = \begin{cases} 0 & \text{if } r > 1 \text{ is a fixed constant} \\ \begin{cases} 0 & \text{if } lim_{n\to+\infty}\epsilon_{p\ell} < \epsilon_p^1(r) \\ +\infty & \text{if } lim_{n\to+\infty}\epsilon_{p\ell} \geq \epsilon_p^1(r) \end{cases} \\ \quad \text{if } r > 1 \text{ is a large number} \end{cases}.$$

Thus, the Bayes factor of the full model M_p versus the reduced model M_ℓ is consistent under the full model M_p if and only if $\lim_{n\to+\infty} \epsilon_{p\ell} \geq \epsilon_p^1(r)$ when r is large and goes to infinity. ■

2.4. When the power $\delta_\ell = \delta$

Finally, we consider the case where the power parameter is set equal to a fixed non-negative constant δ, and studying the consistency when the dimension p of the full model M_p is either a fixed constant number or large and goes to infinity.

Then (7) becomes:

$$BF_{p\ell}^{PEP} = 2\frac{\Gamma(n-p)}{\Gamma^2(\frac{n-p}{2})}\int_0^{\pi/2} \times \frac{(\sin\varphi)^{n-d_\ell-1}(\cos\varphi)^{n-p-1}\,(\delta+\sin^2\varphi)^{(n-p)/2}}{(\delta\rho_{\ell p}+\sin^2\varphi)^{(n-d_\ell)/2}}\mathrm{d}\varphi. \tag{22}$$

2.4.1. When $\dim(M_\ell) = O(1)$ and $\dim(M_p) = O(1)$

Theorem 6: *Let the sample size n increases and being strictly greater than the dimension of the full model M_p. Furthermore, suppose that the dimension of both models, under consideration, are fixed non-negative natural numbers, i.e. $\dim(M_\ell) = d_\ell = O(1)$ and $\dim(M_p) = p = O(1)$, where $p > d_\ell > 1$. Under the condition $\delta_\ell = \delta > 0$, when sampling from model M_j, where j is either ℓ or p we have:*

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = \begin{cases} 0 & \text{if } j=\ell \text{ and } \delta \text{ is large} \\ Constant > 0 & \text{if } j=\ell \text{ and } \delta \text{ is not large} \\ +\infty & \text{if } j = p \end{cases}.$$

Proof: For $\delta_\ell = \delta$, (8) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{1}{2\delta}\right)^{(p-d_\ell)/2}\left(\frac{1}{\rho_{\ell p}}\right)^{(n-d_\ell)/2} \tag{23}$$

Then we consider the following two cases.

(a) Suppose that the Reduced Model M_ℓ is true

Using the asymptotic results of $\rho_{\ell p}$ given in Section 1, (23) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{1}{2\delta}\right)^{(p-d\ell)/2}.$$

Since p and d_ℓ are constants, with $p > d_\ell > 1$, if δ is large, we get

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = 0,$$

while if δ is not large, we get

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = Constant > 0.$$

Thus, the Bayes factor of the full model M_p versus the reduced model M_ℓ is consistent under the reduced model M_ℓ, only for large values of δ.

(b) Suppose that the Full Model M_p is true

Using the asymptotic results of $\rho_{\ell p}$ given in Section 1, (23) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{1}{2\delta}\right)^{(p-d_\ell)/2}(1+\epsilon)^{(n-d_\ell)/2} \approx \mathrm{e}^{(n/2)(-((p-d_\ell)/n)\log(2\delta)+\log(1+\epsilon))}.$$

Thus

$$\lim_{n\to+\infty} BF_{pi}^{PEP} = +\infty.$$

Therefore, the Bayes factor of the full model M_p versus the reduced model M_ℓ is consistent when sampling from the full model M_p. ■

2.4.2. When $\dim(M_\ell) = O(1)$ and $\dim(M_p) = O(n)$

Theorem 7: *Let $\delta_\ell = \delta > 0$ and suppose that the reduced model M_ℓ has a fixed number of parameters, i.e. $\dim(M_\ell) = d_\ell = O(1)$, as the simple size n increases, and in the full model M_p the number of parameters increase with rate $\dim(M_p) = p = O(n)$ with*

$$r = \lim_{n,p\to+\infty}\frac{n}{p} > 1, \quad p > d_\ell > 1.$$

Then:

(1) *When sampling from model M_ℓ*

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = \begin{cases} 0 & \text{if}\delta > \beta_1(r) \\ +\infty & \text{if}\delta < \beta_1(r) \\ Constant > 1 & \text{if}\delta = \beta_1(r) \end{cases}$$

for a continuous and decreasing function $\beta_1 : (1,+\infty) \longrightarrow \mathbb{R}, r \longmapsto (2r/(2r-1))^{2r-1}(r/2(r-1))$.

(2) *When sampling from model M_p*

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = \begin{cases} 0 & \text{if } \delta > \beta_2(r) \\ +\infty & \text{if } \delta < \beta_2(r) \\ +\infty & \text{if } \delta = \beta_2(r) \text{ and large} \\ Constant > 0 & \text{if } \delta = \beta_2(r) \text{ and small} \end{cases}$$

for a continuous function $\beta_2 : (1,+\infty) \longrightarrow \mathbb{R}$, $r \longmapsto \beta_1(r)(1+r)^r$.

Proof: By replacing $n \approx rp$ and $\delta_\ell = \delta$, (9) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{r}{2(r-1)\delta}\right)^{(p-d_\ell)/2} \times \left(\frac{2r}{2r-1}\right)^{((2r-1)p-d_\ell-1)/2}\left(\frac{r-1}{r\rho_{\ell p}}\right)^{(pr-d_\ell)/2}. \tag{24}$$

(a) Suppose that the Reduced Model M_ℓ is true

Using the asymptotic results of $\rho_{\ell p}$ given in Section 1, (24) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{r}{2(r-1)\delta}\right)^{(p-d_\ell)/2} \times \left(\frac{2r}{2r-1}\right)^{((2r-1)p-d_\ell-1)/2}$$

and then

$$BF_{p\ell}^{PEP} \approx \left(\frac{r}{2(r-1)\delta} \left(\frac{2r}{2r-1} \right)^{2r-1} \right)^{p/2} \times \left(\frac{(2r-1)(r-1)\delta}{r^2} \right)^{d_\ell/2} \left(1 - \frac{1}{2r} \right)^{1/2}.$$

We consider the following cases:

- If $(r/2(r-1)\delta)(2r/(2r-1))^{2r-1} = 1 \Rightarrow \delta = \beta_1(r)$ then

$$BF_{p\ell}^{PEP} \approx \left(\frac{2r}{2r-1} \right)^{(r-1)d_\ell} \left(1 - \frac{1}{2r} \right)^{1/2}.$$

  Thus for any $r > 1$

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = Constant > 0.$$

- If $(r/2(r-1)\delta)(2r/(2r-1))^{2r-1} \neq 1$ for large values of p we get

$$BF_{p\ell}^{PEP} \approx \left(\frac{r}{2(r-1)\delta} \left(\frac{2r}{2r-1} \right)^{2r-1} \right)^{p/2}.$$

  Then if
  - $(r/2(r-1)\delta)(2r/(2r-1))^{2r-1} < 1 \Rightarrow \delta > \beta_1(r)$

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = 0.$$

  - $(r/2(r-1)\delta)(2r/(2r-1))^{2r-1} > 1 \Rightarrow \delta < \beta_1(r)$

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = +\infty.$$

Thus, the Bayes factor of the full model M_p versus the reduced model M_ℓ is consistent under the reduced model M_ℓ if and only if the power $\delta > \beta_1(r)$.

(b) Suppose that the Full Model M_p is true

Using the asymptotic results of $\rho_{\ell p}$ given in Section 1, (24) becomes

$$BF_{p\ell}^{PEP} \approx \left(\frac{r}{2(r-1)\delta} \right)^{(p-d_\ell)/2} \times \left(\frac{2r}{2r-1} \right)^{((2r-1)p-d_\ell-1)/2} (1+\epsilon)^{(pr-d_\ell)/2},$$

or

$$BF_{p\ell}^{PEP} \approx \left[\left(\frac{2r}{2r-1} \right)^{2r-1} \frac{r(1+\epsilon)^r}{2(r-1)\delta} \right]^{p/2} \times \left(\frac{(2r-1)(r-1)\delta}{r^2(1+\epsilon)} \right)^{d_\ell/2} \left(1 - \frac{1}{2r} \right)^{1/2}.$$

We consider the following cases

- If $(2r/(2r-1))^{2r-1}(r(1+\epsilon)^r/2(r-1)\delta) = 1 \Rightarrow \delta = \beta_2(r)$ then $BF_{p\ell}^{PEP} \approx ((2r-1)(r-1)\delta/r^2(1+\epsilon))^{d_\ell/2}(1-1/2r)^{1/2}$ and for large values of δ we have

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} \approx \lim_{n\to+\infty} \left(\frac{2\delta}{1+\epsilon} \right)^{d_\ell/2} = +\infty,$$

  while if δ is not large

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} \approx \lim_{n\to+\infty} \left(\frac{2\delta}{1+\epsilon} \right)^{d_\ell/2} = Constant > 0.$$

- If $(2r/(2r-1))^{2r-1}(r(1+\epsilon)^r/2(r-1)\delta) \neq 1$, for large value p we have $BF_{p\ell}^{PEP} \approx ((2r/(2r-1))^{2r-1} (r(1+\epsilon)^r/2(r-1)\delta))^{p/2}$. Then if
  (1) $(2r/(2r-1))^{2r-1}(r(1+\epsilon)^r/2(r-1)\delta) < 1 \Rightarrow \delta > \beta_2(r)$

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = 0.$$

  (2) $(2r/(2r-1))^{2r-1}(r(1+\epsilon)^r/2(r-1)\delta) > 1 \Rightarrow \delta < \beta_2(r)$

$$\lim_{n\to+\infty} BF_{p\ell}^{PEP} = +\infty.$$

Thus, the Bayes factor of the full model M_p versus the reduced model M_ℓ is inconsistent under the full model M_p if $\delta > \beta_2(r)$ or when $\delta = \beta_2(r)$ and δ is small. ■

3. Summary and conclusions

In this paper, we examined the asymptotic behaviour of the power-expected-posterior methodology when

Table 1. Consistency of $BF_{p\ell}^{PEP}$ when model M_ℓ has dimension $\dim(M_\ell) = i = O(1)$ and $\delta_\ell \in \{n, n-p\}$.

	M_ℓ is correct	M_p is correct
$p = O(1)$	Consistent	Consistent
$p = O(n)$	Consistent	Consistent if $\lim_{n\to+\infty} \epsilon_{p\ell} \geq \epsilon_p^2(r)$ for large r

Table 2. Consistency of $BF_{p\ell}^{PEP}$ when model M_ℓ has dimension $\dim(M_\ell) = i = O(1)$ and $\delta_\ell = p$.

	M_ℓ is correct	M_p is correct
$p = O(1)$	Inconsistent	Consistent
$p = O(n)$	Consistent	Consistent if $\lim_{n\to+\infty} \epsilon_{p\ell} \geq \epsilon_p^1(r)$ for large r

Table 3. Consistency of $BF_{p\ell}^{PEP}$ when model M_ℓ has dimension $\dim(M_\ell) = i = O(1)$ and $\delta_\ell = \delta > 0$.

	M_ℓ is correct	M_p is correct
$p = O(1)$	Consistent if δ is large	Consistent
$p = O(n)$	Consistent if $\delta > \beta_1(r)$	Consistent if $\delta < \beta_2(r)$ or if $\delta = \beta_2(r)$ and large

comparing nested normal linear models. Emphasis was given on the consistency of the Bayes factor of the full model M_p versus a generic submodel M_ℓ. The number of parameters of the simplest model M_ℓ was kept always fixed, while for the full model was set of order $O(n^\alpha)$, where $\alpha \in \{0, 1\}$. We investigated the effect of the prior power parameter δ_ℓ, by examining four different scenarios. In each case, the 'true' model was set equal to either M_ℓ or M_p. Tables 1–3 summarise our findings.

The consistency properties of the Power-Expected-Posterior (PEP) prior Bayes factors are eminently reasonable, assuming that we are sampling from either of the candidate models. It is always consistent for fixed dimensions of the candidate models and even in the difficult situation on which the alternative model can grow with the sample size, for the situations described in Tables 1– 3, the PEP Bayes factor is consistent, unless the alternative model is extremely close to the null model, in which case, we conjecture, the lack of consistency is not a critical issue, at least for prediction purposes.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

D. Fouskakis is an Associate Professor in the Department of Mathematics, at the National Technical University of Athens, in Greece. He is also the Director of the Stats Lab at the same University. His research mostly focuses on Bayesian model and variable selection, on objective priors and on stochastic optimization methods.

J. K. Innocent received a Ph.D in Mathematics at the University of Puerto Rico, Puerto Rico, USA in 2016. He is currently back to Haiti, where he teaches mathematical and Statistical courses at a university level. His main research areas are on Bayesian Statistics, Statistical Analysis, Biostatistics and Epidemiology.

L. Pericchi is a Full Professor in the Department of Mathematics of the University of Puerto Rico Rio Piedras, USA. He is also the Director of the Center of Biostatistics and Bioinformatics of the College of Natural Sciences. His research is in the Theory and Applications of Statistics, with emphasis in the Bayesian Approach.

ORCID

D. Fouskakis http://orcid.org/0000-0001-8514-6911

References

Berger, J., & Pericchi, L. (2004). Training samples in objective Bayesian model selection. *Annals of Statistics*, *32*, 841–869.

Casella, G., Girón, F. J., Martínez, M. L., & Moreno, E. (2009). Consistency of Bayesian procedures for variable selection. *Annals of Statistics*, *37*, 1207–1228.

Fouskakis, D., & Ntzoufras, I. (2016a). Limiting behaviour of the Jeffreys power-expected-posterior Bayes factor in Gaussian linear models. *Brazilian Journal of Probability and Statistics*, *30*, 299–320.

Fouskakis, D., & Ntzoufras, I. (2016b). Power-conditional-expected priors: Using g-priors with random imaginary data for variable selection power-conditional-expected priors. *Journal of Computational and Graphical Statistics*, *25*, 647–664.

Fouskakis, D., Ntzoufras, I., & Draper, D. (2015). Power-expected-posterior priors for variable selection in Gaussian linear models. *Bayesian Analysis*, *10*, 75–107.

Girón, F. J., Moreno, E., & Casella, G. (2010). Consistency of objective Bayes factors as the model dimension grows. *Annals of Statistics*, *38*, 1937–1952.

Ibrahim, J. G., & Chen, M. H.. (2000). Power prior distributions for regression models. *Statistical Science*, *15*, 46–60.

Innocent, J. K. (2016). *Bayes factors consistency for nested linear models with increasing dimensions* (Unpublished doctoral dissertation). University of Puerto Rico.

Kass, R. E., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, *90*, 928–934.

Pérez, J. M., & Berger, J. O. (2002). Expected-posterior prior distributions for model selection. *Biometrika*, *89*, 491–512.