

Power-Expected-Posterior Methodology with Baseline Shrinkage Priors



G. Tzoumerkas and D. Fouskakis

Abstract The Power-Expected-Posterior (PEP) prior gives us a convenient and objective method to deal with variable selection problems, under the Bayesian perspective, in regression models. The PEP prior inherits all of the advantages of Expected-Posterior-Prior (EPP) and furthermore it drops the need of selection over the imaginary data and decreases their effect over the final prior. Under the PEP prior methodology an initial (usually default) baseline prior is updated using imaginary data. This work focuses on normal regression models when the number of observations n is smaller than the number of explanatory variables p . We introduce the PEP prior methodology using different baseline shrinkage priors and we perform some comparisons in simulated data sets.

Keywords Bayesian variable selection · imaginary training sample · objective priors · shrinkage priors · sparse datasets

1 Introduction

We consider the variable selection problem for normal regression models, where the number of observations n is smaller than the number of explanatory variables p . Suppose the model space consists of all combinations of available covariates. Then for every model M_ℓ , in model space \mathcal{M} , the likelihood is given by

$$f_\ell(\mathbf{y}|X_\ell, \boldsymbol{\beta}_\ell, \sigma^2) = f_{N_n}(\mathbf{y}; X_\ell \boldsymbol{\beta}_\ell, \sigma^2 \mathbf{I}_n),$$

where $f_{N_d}(\mathbf{y}; \boldsymbol{\mu}, \Sigma)$ is denoting the d -dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ . Furthermore, $\mathbf{y} = (y_1, \dots, y_n)^T$ denotes the response data, X_ℓ is the $n \times p_\ell$ design matrix; where p_ℓ is the number of explanatory variables

G. Tzoumerkas (✉) · D. Fouskakis

Department of Mathematics, National Technical University of Athens, Athens, Greece
e-mail: tzoumg@mail.ntua.gr

D. Fouskakis

e-mail: fouskakis@math.ntua.gr

under model M_ℓ , β_ℓ is a vector of length p_ℓ of the effects of each covariate on the response variable, I_n is the $n \times n$ identity matrix and σ^2 is the error variance. We assume that \mathbf{y} and the columns of the design matrix of the full model (including all available explanatory variables) have been centered on zero, so there is no intercept in our model.

Under the Bayesian model choice perspective, we have to set priors both for the model space and the parameter space of each model. Regarding the prior on the model space, for sparsity reasons, we consider the uniform prior on model size, as a special case of the beta-binomial prior; see [18]. With respect to the prior distribution on the coefficients in each model, because we are not confident about any given set of regressors as explanatory variables, little prior information on their regression coefficients can be expected. This argument alone justifies the need for an objective model choice approach in which vague prior information is assumed. Furthermore, we need to use a prior capable to deal with the $n < p$ scenario. Finally, regarding the (common across models) error variance, the reference prior will be used, i.e. $\pi(\sigma^2) \propto \sigma^{-2}$.

1.1 Shrinkage Priors

A common way to deal with normal regression problems, when $n < p$, is by using shrinkage methods. Under the Bayesian perspective this can be done using a shrinkage prior on the model coefficients. By the term shrinkage, it is declared that the covariates that correspond to explanatory variables that do not affect the response variable will shrink towards zero. Shrinkage priors share eminent theoretical properties, compelling computational complexity and great empirical performance (e.g. [5, 17]).

A shrinkage prior can often be conceived as a scale-mixture prior, which is placed on the regression coefficients of every possible model. Something that characterizes such shrinkage priors, is their hyperparameters: the global shrinkage hyperparameter, that determines the overall sparsity in the whole parameter vector and the local shrinkage hyperparameter, where a distinct shrinkage parameter is considered specifically for every single effect and controls the shrinkage of this individual effect. Depending on the shrinkage prior, the global parameter or the local parameters may be absent from the formation.

By assuming a shrinkage prior, on the vector of regression coefficients β_ℓ , in most of the cases a prior with heavy mass around zero is being produced and by so, non-true effects shrink towards zero. Furthermore, heavy tails are important, as they avert true effects to get shrunked. In Table 1, we mention some, often used, shrinkage priors, where by τ we refer to local shrinkage hyperparameters and by λ to global shrinkage hyperparameters. In all of the cases that a global shrinkage hyperparameter exists in the formation of a shrinkage prior (except Ridge g-prior), we consider a half-Cauchy prior on λ , which is a common choice in Bayesian hierarchical models (e.g. [17]). Furthermore, except Ridge g-prior, independent conditional priors for the

Table 1 A list of shrinkage priors

#	Name	Conditional prior of β_ℓ	Shrinkage hypparameters
1	LASSO [15]	$\beta_j \tau_j^2, \sigma^2 \sim N(0, \sigma^2\tau_j^2)$	$\tau_j^2 \lambda \sim \text{Exp}(\frac{\lambda^2}{2})$ $\lambda \sim \text{HC}(0, 1)^a$
2	Horseshoe [2]	$\beta_j \lambda, \tau_j, \sigma^2 \sim N(0, \sigma^2\lambda^2\tau_j^2)$	$\tau_j \sim \text{HC}(0, 1)$ $\lambda \sim \text{HC}(0, 1)$
3	Ridge [11]	$\beta_j \lambda, \sigma^2 \sim N(0, \sigma^2\frac{1}{\lambda})$	$\lambda \sim \text{HC}(0, 1)$
4	Local Student's t [20]	$\beta_j \tau_j^2, \sigma^2 \sim N(0, \sigma^2\tau_j^2)$	$\tau_j^2 \lambda \sim \text{IG}(\frac{k}{2}, \frac{k}{2\lambda})^b$ $\lambda \sim \text{HC}(0, 1)$ k fixed
5	Elastic Net [13]	$\beta_j \lambda_2, \tau_j, \sigma^2 \sim N(0, \sigma^2\frac{1}{\lambda_2+\tau_j^2})$	$\tau_j^2 \lambda_1 \sim \text{Exp}(\frac{\lambda_1^2}{2})$ $\lambda_1, \lambda_2 \sim \text{HC}(0, 1)$
6	Beta Prime [1]	$\beta_j \tau_j^2, \sigma^2 \sim N(0, \tau_j^2\sigma^2)$	$\tau_j^2 \sim \text{Inv} - \text{Beta}(a, b)$ a, b fixed
7	Ridge g-prior [10]	$\beta_\ell \lambda, \sigma^2 \sim N_{p_\ell}(\mathbf{0}, \sigma^2 V_\ell), V_\ell = g(X_\ell^T X_\ell + \lambda I_{p_\ell})^{-1}$	$g = \max\{n, p_\ell^2\}, \lambda$ fixed

^a $\text{HC}(x_0, \gamma)$, (half-Cauchy) is the truncated Cauchy distribution with location parameter x_0 , scale parameter γ and support (x_0, ∞)

^b $\text{IG}(\alpha, \beta)$, denotes the Inverse Gamma distribution, with shape parameter α and scale parameter β

coefficients of model M_ℓ are used and therefore, for those cases, we only present the marginal prior for $j = 1, \dots, p_\ell$.

1.2 Power-Expected-Posterior Priors

A principal approach to define objective priors is the use of random imaginary training data [4]. Power-Expected-Posterior (PEP) prior [6, 7], uses this methodology. In particular the PEP prior is defined as

$$\pi_\ell^{\text{PEP}}(\beta_\ell|\sigma^2, \delta, X_\ell^*) = \int \pi_\ell^N(\beta_\ell|\mathbf{y}^*, \sigma^2, \delta, X_\ell^*)m_0^N(\mathbf{y}^*|\sigma^2, \delta, X_0^*)d\mathbf{y}^*, \quad (1)$$

$$\pi_\ell^{\text{PEP}}(\sigma^2) = \pi^N(\sigma^2) \propto \frac{1}{\sigma^2},$$

with

$$\pi_\ell^N(\beta_\ell|\mathbf{y}^*, \sigma^2, \delta, X_\ell^*) \propto f_\ell(\mathbf{y}^*|\beta_\ell, \sigma^2, \delta, X_\ell^*)\pi_\ell^N(\beta_\ell|\sigma^2, X_\ell^*) \quad (2)$$

and

$$f_\ell(\mathbf{y}^*|\boldsymbol{\beta}_\ell, \sigma^2, \delta, X_\ell^*) = \frac{f_\ell(\mathbf{y}^*|\boldsymbol{\beta}_\ell, \sigma^2, X_\ell^*)^{1/\delta}}{\int f_\ell(\mathbf{y}^*|\boldsymbol{\beta}_\ell, \sigma^2, X_\ell^*)^{1/\delta} d\mathbf{y}^*}. \quad (3)$$

In the above equations, we have set \mathbf{y}^* to be the imaginary observations of size n^* and X_ℓ^* the imaginary design matrix of model M_ℓ . By $\pi_\ell^N(\boldsymbol{\beta}_\ell|\mathbf{y}^*, \sigma^2, \delta, X_\ell^*)$ we denote the conditional on σ^2 posterior of $\boldsymbol{\beta}_\ell$, using a baseline prior $\pi_\ell^N(\boldsymbol{\beta}_\ell|\sigma^2, X_\ell^*)$ and data \mathbf{y}^* . In equation (3) the likelihood of imaginary observations is raised to the power of $1/\delta$ and density normalized. By doing this we decrease the effect of the imaginary data. For $\delta = 1$, Eq. (1) results to the Expected-Posterior-Prior (EPP) [16]. In order to have a unit information interpretation [12], we could set $\delta = n^*$ and in order to avoid any effect of the choice of imaginary design matrices, we set $n^* = n$ and we have that $X_\ell^* = X_\ell$. In Eq. (1), $m_0^N(\mathbf{y}^*|\sigma^2, \delta, X_0^*)$, is the prior predictive distribution (or the marginal likelihood), evaluated at \mathbf{y}^* , of the reference model M_0 , given σ^2 . As a reference model we consider, for reasons of parsimony, the model with no covariates (null model). Finally, for every model M_ℓ , the marginal likelihood under the baseline prior is given by

$$m_\ell^N(\mathbf{y}^*|\sigma^2, \delta, X_\ell^*) = \int f_\ell(\mathbf{y}^*|\boldsymbol{\beta}_\ell, \sigma^2, \delta, X_\ell^*)\pi_\ell^N(\boldsymbol{\beta}_\ell|\sigma^2, X_\ell^*)d\boldsymbol{\beta}_\ell. \quad (4)$$

2 PEP-Shrinkage Prior

In the above formulation, by choosing as a baseline prior $\pi_\ell^N(\boldsymbol{\beta}_\ell|\sigma^2, X_\ell^*)$ a shrinkage prior (see Table 1), a PEP-Shrinkage prior is created and thus we can apply the PEP prior methodology in shrinkage problems.

PEP priors can be considered as fully automatic, objective Bayesian methods for model comparison in regression models (see for example [4, 6]). They are developed through the utilization of the device of ‘‘imaginary’’ samples, coming from the simplest model under comparison. Therefore, PEP priors offer several advantages, among which they have an appealing interpretation based on imaginary training data coming from a prior predictive distribution and also provide an effective way to establish compatibility of priors among models (see [3]), through their dependence on a common marginal data distribution. Thus, the PEP methodology can be applied also with proper baseline prior distributions. Furthermore, by choosing the simplest model, as a reference model, to generate the imaginary samples, the PEP prior shares common ideas with the skeptical-prior approach described by Spiegelhalter et al. [19].

Under Eq. (3) the likelihood of the imaginary data \mathbf{y}^* , under model M_ℓ , is given by

$$f_\ell(\mathbf{y}^*|X_\ell^*, \boldsymbol{\beta}_\ell, \sigma^2, \delta) = f_{N_{n^*}}(\mathbf{y}^*; X_\ell^* \boldsymbol{\beta}_\ell, \delta \sigma^2 I_{n^*}).$$

From Table 1 it is obvious that all shrinkage priors that we will use as baseline priors under the PEP methodology, have the following general form

$$\pi_\ell^N(\boldsymbol{\beta}_\ell | \boldsymbol{\theta}_\ell, \sigma^2) = f_{N_{p_\ell}}(\boldsymbol{\beta}_\ell; \mathbf{0}, \sigma^2 \Omega_\ell),$$

where $\Omega_\ell \equiv \Omega_\ell(\boldsymbol{\theta}_\ell)$ is a $p_\ell \times p_\ell$ matrix, where its i -th main diagonal element is written as an equation of the global and the i -th local shrinkage hyperparameters. By $\boldsymbol{\theta}_\ell$ we denote the vector containing all the shrinkage hyperparameters of model M_ℓ , with a prior distribution denoted by $\pi(\boldsymbol{\theta}_\ell)$.

2.1 Conditional PEP-Shrinkage Prior

The conditional posterior distribution $\pi_\ell^N(\boldsymbol{\beta}_\ell | \mathbf{y}^*, \sigma^2, \delta, X_\ell^*, \boldsymbol{\theta}_\ell)$, using the baseline prior and the imaginary data is given by

$$\begin{aligned} \pi_\ell^N(\boldsymbol{\beta}_\ell | \mathbf{y}^*, \sigma^2, \delta, X_\ell^*, \boldsymbol{\theta}_\ell) &\propto f_\ell(\mathbf{y}^* | X_\ell^*, \boldsymbol{\beta}_\ell, \sigma^2, \delta) \pi_\ell^N(\boldsymbol{\beta}_\ell | \boldsymbol{\theta}_\ell, \sigma^2) \\ &= f_{N_{n^*}}(\mathbf{y}^*; X_\ell^* \boldsymbol{\beta}_\ell, \delta \sigma^2 I_{n^*}) f_{N_{p_\ell}}(\boldsymbol{\beta}_\ell; \mathbf{0}, \sigma^2 \Omega_\ell) \end{aligned}$$

and so we have have that

$$\pi_\ell^N(\boldsymbol{\beta}_\ell | \mathbf{y}^*, \sigma^2, \delta, X_\ell^*, \boldsymbol{\theta}_\ell) = f_{N_{p_\ell}}(\boldsymbol{\beta}_\ell; \delta^{-1} W_\ell X_\ell^{*T} \mathbf{y}^*, \sigma^2 W_\ell),$$

where $W_\ell = [\delta^{-1} X_\ell^{*T} X_\ell^* + \Omega_\ell^{-1}]^{-1}$. Moreover, from Eq. (4), for any model M_ℓ , the prior predictive distribution, under the baseline prior, conditional on σ^2 and $\boldsymbol{\theta}_\ell$ is

$$m_\ell^N(\mathbf{y}^* | \sigma^2, \delta, X_\ell^*, \boldsymbol{\theta}_\ell) = f_{N_{n^*}}(\mathbf{y}^*; \mathbf{0}, \sigma^2 \Lambda_\ell),$$

where $\Lambda_\ell = X_\ell^* \Omega_\ell X_\ell^{*T} + \delta I_{n^*}$. Thus, the conditional PEP-Shrinkage prior is

$$\begin{aligned} \pi_\ell^{PEP}(\boldsymbol{\beta}_\ell | \sigma^2, \delta, X_\ell^*, \boldsymbol{\theta}_\ell) &= \int \pi_\ell^N(\boldsymbol{\beta}_\ell | \mathbf{y}^*, \sigma^2, \delta, X_\ell^*, \boldsymbol{\theta}_\ell) m_0^N(\mathbf{y}^* | \sigma^2, \delta, X_0^*) d\mathbf{y}^* \\ &= \int f_{N_{p_\ell}}(\boldsymbol{\beta}_\ell; \delta^{-1} W_\ell X_\ell^{*T} \mathbf{y}^*, \sigma^2 W_\ell) f_{N_{n^*}}(\mathbf{y}^*; \mathbf{0}, \sigma^2 \Lambda_0) d\mathbf{y}^* \end{aligned}$$

and therefore we have that

$$\pi_\ell^{PEP}(\boldsymbol{\beta}_\ell | \sigma^2, \delta, X_\ell^*, \boldsymbol{\theta}_\ell) = f_{N_{p_\ell}}(\boldsymbol{\beta}_\ell; \mathbf{0}, \sigma^2 V_\ell),$$

where $V_\ell = [W_\ell^{-1} - \delta^{-2} X_\ell^{*T} Z_\ell X_\ell^*]^{-1}$ and $Z_\ell = [\delta^{-2} X_\ell^* W_\ell X_\ell^{*T} + \Lambda_0^{-1}]^{-1}$.

2.2 Conditional Posterior Under the PEP-Shrinkage Prior

The posterior distribution, under the PEP prior, conditional on the shrinkage hyper-parameters θ_ℓ of model M_ℓ , is given by

$$\begin{aligned}\pi_\ell^{PEP}(\boldsymbol{\beta}_\ell, \sigma^2 | \mathbf{y}, \delta, X_\ell^*, X_\ell, \boldsymbol{\theta}_\ell) &\propto \pi_\ell^{PEP}(\boldsymbol{\beta}_\ell | \sigma^2, \delta, X_\ell^*, \boldsymbol{\theta}_\ell) \pi^N(\sigma^2) f_\ell(\mathbf{y} | X_\ell, \boldsymbol{\beta}_\ell, \sigma^2) \\ &= f_{N_{p_\ell}}(\boldsymbol{\beta}_\ell; \mathbf{0}, \sigma^2 V_\ell) \pi^N(\sigma^2) f_{N_n}(\mathbf{y}; X_\ell \boldsymbol{\beta}_\ell, \sigma^2 I_n).\end{aligned}$$

Using the reference prior for σ^2 (see Sect. 1), this joint posterior can be written as the product of

$$\pi_\ell^{PEP}(\boldsymbol{\beta}_\ell | \mathbf{y}, \sigma^2, \delta, X_\ell^*, X_\ell, \boldsymbol{\theta}_\ell) = f_{N_{p_\ell}}(\boldsymbol{\beta}_\ell; S_\ell X_\ell^T \mathbf{y}, \sigma^2 S_\ell)$$

and

$$\pi_\ell^{PEP}(\sigma^2 | \mathbf{y}, \delta, X_\ell^*, X_\ell, \boldsymbol{\theta}_\ell) = f_{IG}(\sigma^2; \alpha_\ell, b_\ell),$$

where $f_{IG}(x; \alpha, b)$ is denoting the Inverse Gamma distribution with shape parameter α and scale parameter b . Furthermore, we have set $S_\ell = (V_\ell^{-1} + X_\ell^T X_\ell)^{-1}$, $\alpha_\ell = \frac{n}{2}$ and $b_\ell = \frac{\mathbf{y}^T [I_n + X_\ell V_\ell X_\ell^T]^{-1} \mathbf{y}}{2}$.

2.3 Marginal Likelihood Under the PEP-Shrinkage Prior

The marginal likelihood, of model M_ℓ , under the PEP-Shrinkage prior, given the shrinkage parameter $\boldsymbol{\theta}_\ell$ is given by

$$\begin{aligned}m_\ell^{PEP}(\mathbf{y} | \delta, X_\ell^*, X_\ell, \boldsymbol{\theta}_\ell) &= \int \pi_\ell^{PEP}(\boldsymbol{\beta}_\ell | \sigma^2, \delta, X_\ell^*, \boldsymbol{\theta}_\ell) \pi^N(\sigma^2) f_\ell(\mathbf{y} | X_\ell, \boldsymbol{\beta}_\ell, \sigma^2) d\boldsymbol{\beta}_\ell d\sigma^2 \\ &\propto (\det(I_n + X_\ell V_\ell X_\ell^T))^{-\frac{1}{2}} (\mathbf{y}^T [I_n + X_\ell V_\ell X_\ell^T]^{-1} \mathbf{y})^{-\frac{n}{2}}.\end{aligned}$$

Therefore in cases where the shrinkage parameters of the baseline prior are fixed (e.g. Ridge g-prior), the above marginal likelihood can be calculated in closed form. The unknown normalizing constant, in the above expression, comes from the improper prior of the error variance, which is common in all compared models, and therefore we do not face any indeterminacy issues when calculating the Bayes factor.

When the shrinkage parameters are not fixed, the marginal likelihood is given by

$$m_\ell^{PEP}(\mathbf{y}) \equiv m_\ell^{PEP}(\mathbf{y} | \delta, X_\ell^*, X_\ell) = \int m_\ell^{PEP}(\mathbf{y} | \delta, X_\ell^*, X_\ell, \boldsymbol{\theta}_\ell) \pi(\boldsymbol{\theta}_\ell) d\boldsymbol{\theta}_\ell.$$

If the dimension of $\boldsymbol{\theta}_\ell$ is one (e.g. Ridge prior) the above integral can be easily numerically evaluated. Furthermore, in order to search the model space, MC^3 procedures [14] can be performed. If the dimension of $\boldsymbol{\theta}_\ell$ is greater than one (e.g. Horseshoe

prior), we perform an MC^3 procedure, conditionally on θ_ℓ , as in Algorithm 3 of the Appendix of [9], where each component of θ_ℓ is generated from its full conditional posterior distribution using a Metropolis-Hastings step.

3 Simulation Study

In this section we test the PEP-Shrinkage methodology (with $\delta = n = n^*$, $X_\ell^* = X_\ell$ and the reference model to be the null one) on simulated data. We use as a baseline prior, all the shrinkage priors listed in Table 1 and compare their results. Moreover we compare the results under the PEP-Ridge prior with the ones obtain by using the Ridge prior, without the PEP methodology.

We have simulated 100 different samples of length $n = 25$ with $p = 50$ predictors. The values of the explanatory variables have been generated from $N_{50}(\mathbf{0}, \Sigma)$, where the symmetrical matrix Σ has elements $\Sigma_{i,j} = (0.75)^{|i-j|}$, $i, j = 1, \dots, 50$. Finally, we center the columns of the design matrix on zero. For the predictor effects we have set $(\beta_1, \beta_2, \beta_{10})^T = (2, 0.8, 1.5)^T$ and for all of the rest, we set to be equal to 0. We have set $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N_{25}(\mathbf{0}, \sigma^2 I_{25})$, for $\sigma^2 = 1.5$. Finally, we center the values of the response variable on zero.

In Fig. 1 (left), we present the boxplots of the marginal posterior inclusion probabilities, for the true effects, of the 100 different samples, for the seven different PEP-Shrinkage priors. Regarding the two most influential variables, X_1 and X_{10} , under every baseline prior, we obtained high posterior inclusion probabilities with the majority of cases to be above 0.5. Furthermore, for these two effects, PEP-Ridge seems to outperform every other PEP-Shrinkage prior. On the contrary, PEP-(Ridge) g-prior seems to give the least satisfactory results. For the predictor X_2 , the median marginal posterior inclusion probabilities are above 0.5, for all baseline priors, except one. As before, PEP-Ridge gives the most satisfactory results, while PEP-(Ridge) g-prior produces posterior inclusion probabilities with a median value below 0.5. For the non-true effects, for brevity reasons, we present results in Fig. 1 (right) only for a subset of them. More specifically we present results only for variables X_3 , X_9 and X_{11} , which are the ones with the higher correlations with the true effects. For every selection of baseline prior, the median marginal posterior inclusion probabilities are below 0.5. It is distinct that, regardless the baseline prior we choose, only in a small percentage of occasions, the non-true effects would have been accepted as true effects of the model (posterior inclusion probabilities above 0.5). We notice that PEP-Ridge manages to give, in general, very small posterior inclusion probabilities with small variability also. For the rest of the non-true effects we get similar results.

In Fig. 2, we present the boxplots of the posterior inclusion probabilities of the true main effects (left) and the (previously made) selection of non-true effects (right) between the PEP-Ridge and the Ridge prior (without applying the PEP methodology). As for the true effects we notice similar results, as both priors manages to accept the true effects, in the vast majority of the cases. For predictor X_{10} we can observe slightly better results under the PEP-Ridge methodology. As for the non-true effects,

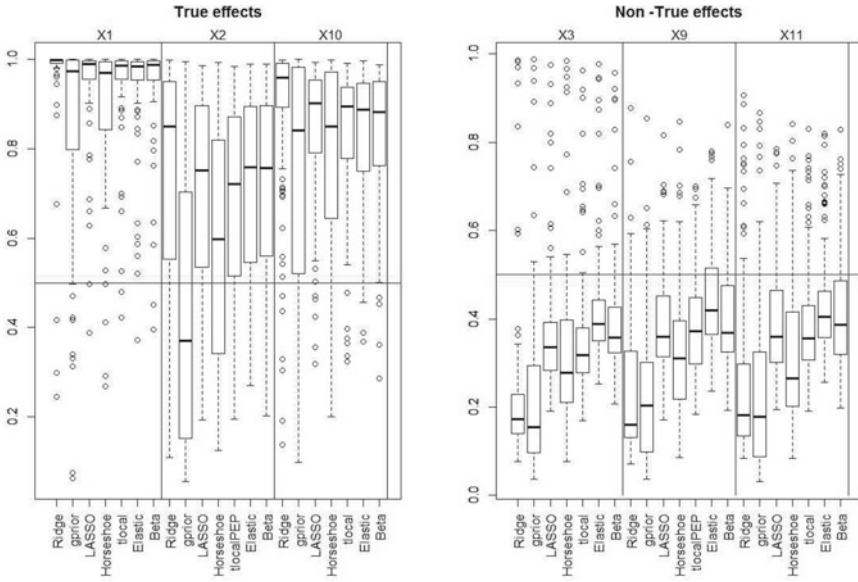


Fig. 1 Boxplots of posterior inclusion probabilities, across 100 simulated datasets, for the true effects—variables X_1, X_2, X_{10} (left) and for some of the non-true effects—variables X_3, X_9, X_{11} (right) using the PEP-Shrinkage methodology, for different baseline prior (X-axis).

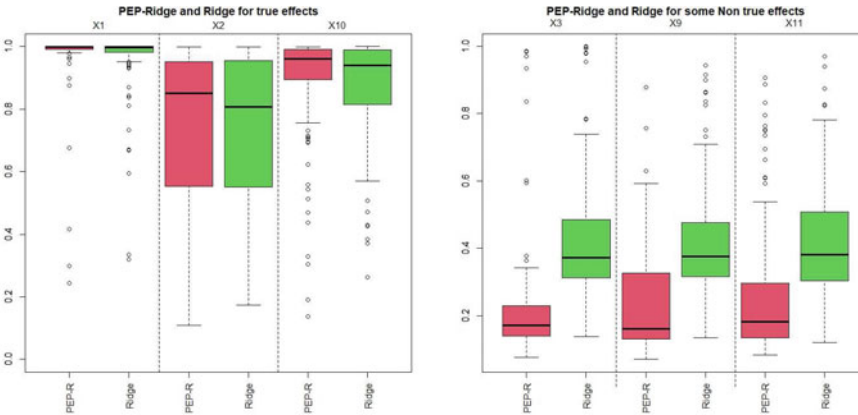


Fig. 2 Boxplots of posterior inclusion probabilities, across 100 simulated datasets, for the true effects—variables X_1, X_2, X_{10} (left) and for some of the non-true effects—variables X_3, X_9, X_{11} (right), using the PEP-Ridge prior (PEP-R) and the Ridge prior without the PEP methodology

the PEP-Ridge prior outperforms the Ridge prior, as it manages to restrain more cases to the desirable limits, that is, producing marginal posterior probabilities far below 0.5 with small variability. Thus we can conclude that the PEP methodology improves the initially chosen Ridge prior, as it produces more parsimonious results.

4 Discussion

In this paper we briefly present the model formulation and some preliminary results of an objective Bayesian prior distribution capable of dealing with variable selection problems in normal regression models when the number of observations is smaller than the number of explanatory variables. The proposed PEP-Shrinkage prior combines two approaches: the PEP prior methodology and the shrinkage priors. The resulting prior has a nice interpretation, based on imaginary data, and is compatible across models. Based on the simulation study, presented here, the PEP-Shrinkage priors, in the majority of cases, correctly identify the true model. Furthermore, under the Ridge prior, the PEP methodology improves the initial prior, by being more parsimonious, a property that is desirable on sparse regression problems.

There are several directions of future extensions. The main aim is to create a unified approach; i.e. a new class of PEP-Shrinkage priors, that includes all the cases mentioned in this paper. To achieve this goal our aim is to write the PEP-Shrinkage prior as a scale mixture of normal distribution, with the mixing distribution denoting the different baseline prior distributions used. This representation will offer several advantages: faster evaluation of posterior distributions and Bayes factors, under all approaches considered, as well as, computational tractability. The performance of this new class of shrinkage prior distributions then have to be assessed in relation to: (a) computational efficiency, (b) frequentist assessment, especially in terms of the speed of concentration of the posterior parameter distribution, or functional thereof, to the true value, and in terms of coverage of credible sets, (c) ease of interpretation, (d) default set of tuning hyperparameters in scientific applications. Moreover, a very important aspect is to check and prove mathematical properties of the new class of prior distributions. Further research should be held, of what happens if we choose the size of the imaginary data, not to be equal to the number of the observations and how that affects the results. In the same manner, we should check what happens for different values of δ , or even set a prior distribution for it, as in [8]. Finally, more shrinkage methods could be considered, apart the ones presented in Table 1. Additional future extensions of our PEP-Shrinkage method include implementation in generalized linear models, where computation is more demanding.

Acknowledgements This work has received funding from the Research Program PEVE 2020 of the National Technical University of Athens.

References

1. Bai, R., Ghosh, M.: On the beta prime prior for scale parameters in high-dimensional bayesian regression models. *Stat. Sin.* **31**, 843–865 (2021)
2. Carvalho, C.M., Polson, N.G., Scott, J.G.: The horseshoe estimator for sparse signals. *Biometrika*. **97**, 465–480 (2010)
3. Consonni, G., Veronese, P.: Compatibility of prior specifications across linear models. *Stat. Sci.* **23**, 332–353 (2008)
4. Consonni, G., Fouskakis, D., Liseo, B., Ntzoufras, I.: Prior Distributions for objective Bayesian analysis. *Bayesian Anal.* **13**, 627–679 (2018)
5. Datta, J., Ghosh, J.K.: Asymptotic properties of Bayes risk for the horseshoe prior. *Bayesian Anal.* **8**, 111–132 (2013)
6. Fouskakis, D., Ntzoufras, I., Draper, D.: Power-expected-posterior priors for variable selection in Gaussian linear models. *Bayesian Anal.* **10**, 75–107 (2015)
7. Fouskakis, D., Ntzoufras, I.: Power-conditional-expected priors. Using g-priors with random imaginary data for variable selection. *J. Comput. Graph. Stat.* **25**, 647–664 (2016)
8. Fouskakis, D., Ntzoufras, I., Perrakis, K.: Power-expected-posterior priors in generalized linear models. *Bayesian Anal.* **13**, 721–748 (2018)
9. Fouskakis, D., Ntzoufras, I.: Power-expected-posterior priors as mixtures of g-Priors. *Bayesian Anal.* (accepted) (2021)
10. Gupta, M., Ibrahim, J.: An information matrix prior for Bayesian analysis in generalized linear models with high dimensional data. *Stat. Sin.* **19**, 1641–1663 (2009)
11. Hsiang, T.C.: A Bayesian view on ridge regression. *The Statist.* **24**, 267–268 (1975)
12. Kass, R.E., Wasserman, L.: A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *J. Am. Stat. Assoc.* **90**, 928–934 (1995)
13. Kyung, M., Gill, J., Ghosh, M., Casella, G.: Penalized regression, standard errors, and Bayesian lassos. *Bayesian Anal.* **5**, 369–411 (2010)
14. Madigan, D., York, J.: Bayesian graphical models for discrete data. *Int. Stat. Rev.* **63**, 215–232 (1995)
15. Park, T., Casella, G.: The Bayesian lasso. *J. Am. Stat. Assoc.* **103**, 681–687 (2008)
16. Pérez, J.M., Berger, J.O.: Expected—posterior prior distributions for model selection. *Biometrika* **89**, 491–511 (2002)
17. Polson, G., Scott, J.: On the half-Cauchy prior for a global scale parameter. *Bayesian Anal.* **7**, 887–902 (2011)
18. Scott, J.G., Berger, J.O.: Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Ann. Stat.* **38**, 2587–2619 (2010)
19. Spiegelhalter, D.J., Abrams, K.R., Myles, J.P.: *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, Chichester (2004)
20. Tipping, M.E.: Sparse Bayesian Learning and the Relevance Vector Machine. *J. Mach. Learn.* **1**, 211–244 (2001)