Properties of Variations of Power-Expected-Posterior Priors

ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

Ioannis Ntzoufras

(ntzoufras@aueb.gr)

Joint work with: **Dimitris Fouskakis** (Dep. of Mathematics, NTUA)

Konstantinos Perrakis (DZNE, German Center for Neurodegenerative Diseases)

14-17 July 2017: Greek Stochastics ι' @ Milos Island

Synopsis

1. Introduction: Bayesian Model Selection and Power-Expected-Posterior (PEP) Priors
2. PEP-priors and Variations
3. Properties of PEP priors
   - Connections with G-priors and parsimony
   - Predictive Matching
   - Model Selection Consistency
4. Illustrations
5. General Framework and Concluding Remarks

1 Introduction: Model Selection and Expected-Posterior Priors

Within the Bayesian framework the comparison between models M_0 and M_1 is evaluated via the **Posterior Odds** (PO)

$$PO_{01} \equiv \frac{\pi(M_0|\boldsymbol{y})}{\pi(M_1|\boldsymbol{y})} = \frac{m_0(\boldsymbol{y})}{m_1(\boldsymbol{y})} \times \frac{\pi(M_0)}{\pi(M_1)} = BF_{01} \times O_{01} \tag{1}$$

which is a function of the **Bayes Factor** (BF_{01}) and the **Prior Odds** (O_{01}).

In the above $m_\ell(\boldsymbol{y})$ is the marginal likelihood under model M_ℓ and $\pi(M_\ell)$ is the prior probability of model M_ℓ.

The marginal likelihood of model M_ℓ is given by

$$m_\ell(\boldsymbol{y}) = \int f_\ell(\boldsymbol{y}|\boldsymbol{\theta}_\ell)\pi_\ell(\boldsymbol{\theta}_\ell)d\boldsymbol{\theta}_\ell, \tag{2}$$

where $f_\ell(\boldsymbol{y}|\boldsymbol{\theta}_\ell)$ is the likelihood under model M_ℓ with parameters $\boldsymbol{\theta}_\ell$ and $\pi_\ell(\boldsymbol{\theta}_\ell)$ is the prior distribution of model parameters given model M_ℓ.

The Lindley-Bartlett-Jeffreys Paradox

For a single model inference $\Rightarrow$ a highly diffuse prior on the model parameters is often used (to represent ignorance).

$\Rightarrow$ Posterior density takes the shape of the likelihood and is insensitive to the exact value of the prior density function.

For multiple models inference $\Rightarrow$ BFs (and POs) are quite sensitive to the choice of the prior variance of model parameters.

$\Rightarrow$ For nested models, we support the simplest model with the evidence increasing as the variance of the parameters increase ending up to support of more parsimonious model no matter what data we have.

$\Rightarrow$ Under this approach, the procedure is quite informative since the data do not contribute to the inference.

$\Rightarrow$ Improper priors cannot be used since the BFs depend on the undefined normalizing constants of the priors.

Model Formulation

PEP methodology is general but here we work within the **generalized linear models (GLM)** setup.

- Response distribution: member of the exponential family (normal regression, binomial logistic regression, Poisson log-linear models)
- Linear predictor of the form

$$\eta_{\boldsymbol{\gamma}(i)} = \mathbf{X}_{\boldsymbol{\gamma}(i)}\boldsymbol{\beta}_{\gamma}$$

  - p: total number of covariates under consideration
  - $\boldsymbol{\gamma}$ covariate inclusion indicators (indicating active covariates)
  - $p_{\boldsymbol{\gamma}} = \sum_{j=1}^{p} \gamma_j$: number of active covariates
  - $\boldsymbol{\beta}_{\gamma}$ vector of covariate coefficients for the active covariates
  - ϕ dispersion parameter
  - $\mathbf{X}_{\boldsymbol{\gamma}}$ design/data matrix of dimension of $n \times d_m$ ($d_m = p_{\boldsymbol{\gamma}} + 1$ if the constant is included)
  - Index i indicates i observation $i = 1, \ldots, n$.

From EP to PEP

Expected-Posterior priors (EP; Pérez and Berger, 2002, *Bka)*

$\Rightarrow$ Power-Expected-Posterior Priors (PEP; Fouskakis, Ntzoufras and Draper, 2015, *BA)*.

$$\underbrace{\pi_\ell^{EPP}(\boldsymbol{\theta}_\ell)}_{\Downarrow} = \int \underbrace{\pi_\ell^{N}(\boldsymbol{\theta}_\ell|\boldsymbol{y}^*)}_{\Downarrow} \underbrace{m_0^{N}(\boldsymbol{y}^*)}_{\Downarrow} d\boldsymbol{y}^*$$

$$\pi_\ell^{PEP}(\boldsymbol{\theta}_\ell;\delta) = \int \underbrace{\pi_\ell^{N}(\boldsymbol{\theta}_\ell|\boldsymbol{y}^*,\delta)}_{\Downarrow} \underbrace{m_0^{N}(\boldsymbol{y}^*|\delta)}_{\Downarrow} d\boldsymbol{y}^*$$

we substitute the likelihood terms with powered-versions of the likelihoods

(i.e. they are raised to the power of $1/\delta$).

Features of PEP

PEP priors method amalgamates ideas from Intrinsic Priors, EPPs, Unit Information Priors and Power Priors, to unify ideas of Non-Data Objective Priors.

PEP priors solve the following problems:

- Dependence of training sample size.
- Lack of robustness with respect to the sample irregularities.
- Excessive weight of the prior when the number of parameters is close to the number of data.

At the same time the PEP prior is a fully objective method and shares the advantages of Intrinsic Priors and EPPs.

- We choose $\delta = n^*$, $n^* = n$ and therefore $\mathrm{X}_\ell^* = \mathrm{X}_\ell$; by this way we dispense with the selection of the training samples.

2 Power-Expected-Posterior (PEP) Priors

Following Fouskakis and Ntzoufras (2016*b*, *JCGS)*, the conditional PEP (PCEP) prior in the GLM setup, under the null-reference model M_0, is defined as follows

$$\pi_{\boldsymbol{\gamma}}^{\mathrm{PEP}}(\boldsymbol{\beta}_\gamma, \phi|\boldsymbol{\delta}) = \pi_{\boldsymbol{\gamma}}^{\mathrm{PEP}}(\boldsymbol{\beta}_\gamma|\phi, \boldsymbol{\delta})\pi_{\boldsymbol{\gamma}}^{\mathrm{N}}(\phi), \tag{3}$$

where $\boldsymbol{\delta} = (\delta_1, \delta_2)$ and

$$\pi_{\boldsymbol{\gamma}}^{\mathrm{PEP}}(\boldsymbol{\beta}_\gamma|\phi, \boldsymbol{\delta}) = \int \pi_{\boldsymbol{\gamma}}^{\mathrm{N}}(\boldsymbol{\beta}_\gamma|\boldsymbol{y}^*, \phi, \delta_1) m_0^{\mathrm{N}}(\boldsymbol{y}^*|\phi, \delta_0)\mathrm{d}\boldsymbol{y}^*, \tag{4}$$

$$\pi_{\boldsymbol{\gamma}}^{\mathrm{N}}(\boldsymbol{\beta}_\gamma|\boldsymbol{y}^*, \phi, \delta) = \frac{f_{\boldsymbol{\gamma}}(\boldsymbol{y}^*|\boldsymbol{\beta}_\gamma, \phi, \delta)\pi_{\boldsymbol{\gamma}}^{\mathrm{N}}(\boldsymbol{\beta}_\gamma|\phi)}{m_{\boldsymbol{\gamma}}^{\mathrm{N}}(\boldsymbol{y}^*|\phi, \delta)}, \tag{5}$$

$$m_{\boldsymbol{\gamma}}^{\mathrm{N}}(\boldsymbol{y}^*|\phi, \delta) = \int f_{\boldsymbol{\gamma}}(\boldsymbol{y}^*|\boldsymbol{\beta}_\gamma, \phi, \delta)\pi_{\boldsymbol{\gamma}}^{\mathrm{N}}(\boldsymbol{\beta}_\gamma|\phi)\mathrm{d}\boldsymbol{\beta}_\gamma, \tag{6}$$

$$f_{\boldsymbol{\gamma}}(\boldsymbol{y}^*|\boldsymbol{\beta}_\gamma, \phi, \delta) = \frac{f_{\boldsymbol{\gamma}}(\boldsymbol{y}^*|\boldsymbol{\beta}_\gamma, \phi)^{1/\delta}}{k_{\boldsymbol{\gamma}}(\boldsymbol{\beta}_\gamma, \phi, \delta)}, \tag{7}$$

Original PEP prior of Fouskakis et al. (2015, *BA)* for normal regression models:

$$k_{\boldsymbol{\gamma}}(\boldsymbol{\beta}_\gamma, \phi, \delta) = \int f_{\boldsymbol{\gamma}}(\boldsymbol{y}^*|\boldsymbol{\beta}_\gamma, \phi)^{1/\delta} d\boldsymbol{y}^* \text{ and } \delta_1 = \delta_0 = \delta$$

for all models $\boldsymbol{\gamma} \in \mathcal{M}$.

- This choice results corresponds to the density normalized likelihood used in the original PEP prior setup for normal regression models (Fouskakis, Ntzoufras and Perrakis, 2017, *BA)*.
- For normal models, density normalized power likelihood $\Rightarrow N(0, \delta\sigma^2)$.
- This nice property (obtaining sampling distribution of the same type) is not generally applicable.
- Hence, the selection of $k_{\boldsymbol{\gamma}}(\boldsymbol{\beta}_\gamma, \phi, \delta)$ and $k_0(\beta_0, \phi, \delta)$ was revised on our work by Fouskakis, Ntzoufras and Perrakis (2017, *BA, min.rev.)* for GLMs.

2.1 Variations of PEP for GLM

- To work on GLM we have introduced (Fouskakis, Ntzoufras and Perrakis, 2017, *BA, min.rev.)* new variations of PEP depending on the selection of $k_{\boldsymbol{\gamma}}$ and k_0 and δ_1, δ_0.
- Two alternative definitions
  - DR-PEP: $k_{\boldsymbol{\gamma}} = k_0 = 1$ (unnormalized likelihoods) and $\delta_1 = \delta_0 = \delta$.
  - CR-PEP: δ_1 and $k_{\boldsymbol{\gamma}} = 1$ (unnormalized likelihoods) and $\delta_0 = 1 \Rightarrow k_0 = 1$ (original likelihood).
- For the density-normalized version in regression: Full PEP (method on $\boldsymbol{\beta}_\gamma$ and $\phi = \sigma^2$) and PCEP (method on $\boldsymbol{\beta}_\gamma$ conditionally on $\phi = \sigma^2$)
- For CR/DR-PEP: We have worked the conditional version.
- Placing a hyper-prior is an option explored in Fouskakis, Ntzoufras and Perrakis (2017, *BA, min.rev.)*.

3 Properties of PEP priors

Connections with g-priors and hyper-g priors and Parsimony in Normal Regression

1. Full PEP (Fouskakis et al., 2015, *BA)*

   $\Rightarrow$ Mixture of g-prior placing a beta hyper-prior on $t = \delta/g$ (Fouskakis, Ntzoufras and Pericchi, 2017, *submitted)*.

   $\Rightarrow$ Information consistency is achieved (Fouskakis and Ntzoufras, 2017, *Metron)*.

2. Density Normalized Conditional PEP (PCEP) (Fouskakis and Ntzoufras, 2016*b*, *JCGS)* and DR-PEP coincide.

3. PCEP, DR-PEP and CR-PEP:
   - They are similar to g-priors (with more complicated covariance structure).
   - The prior variance volume is given by $\phi(\delta, n^*)|\boldsymbol{X}_{\boldsymbol{\gamma}}^{*T}\boldsymbol{X}_{\boldsymbol{\gamma}}^{*}|^{-1}$.
   - For the default choices $\delta = n^* = n$ and $\boldsymbol{X}_{\boldsymbol{\gamma}} = \boldsymbol{X}_{\boldsymbol{\gamma}}$, they are more parsimonious than the g-priors (for finite n) but asymptotically the same.
   - They suffer from the information paradox in normal regression.
4. Hyper-δ versions of PEP conditional priors are similar to hyper-g priors.

Details for point 3 can be found in Fouskakis, Ntzoufras and Perrakis (2016, *arXiv)*.

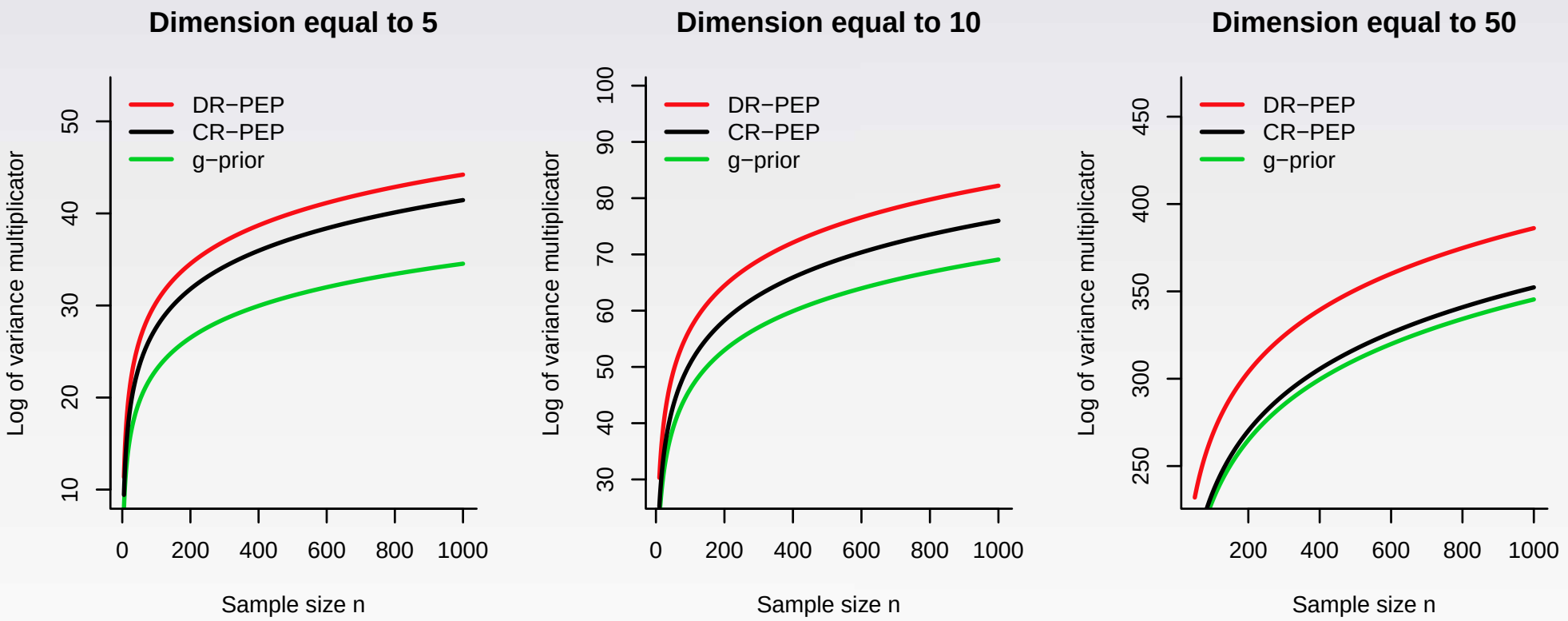


Figure 1: Log-variance multipliers of the DR-PEP, CR-PEP and g-priors versus sample size for $d_\ell = 5, 10, 50$.

Predictive Matching

Null and Dimension Predictive matching is valid for DR and CR-PEP priors under certain structural assumptions about the baseline priors (Fouskakis, Ntzoufras and Perrakis, 2017, *BA*, Propositions 5.1–5.4).

- Structure for null predictive matching: $\pi^{\mathrm{N}}_{\boldsymbol{\gamma}}(\boldsymbol{\beta}_{\gamma}|\phi) = \psi(\boldsymbol{\eta}_{\boldsymbol{\gamma}})\Psi_{\boldsymbol{\gamma}}(\boldsymbol{\beta}_{\setminus 0,\boldsymbol{\gamma}})$.
- Structure for dimension predictive matching: $\pi^{\mathrm{N}}_{\boldsymbol{\gamma}}(\boldsymbol{\beta}_{\gamma}|\phi) = \psi(\boldsymbol{\eta}_{\boldsymbol{\gamma}})$
- Both of these are restrictions are valid for the flat improper and the Jeffreys prior.
- Also valid for the g-prior but not for independent priors.

Model Selection Consistency

Mathematical proofs for consistency:

- Original PEP in Regression using the Jeffreys baseline prior (Fouskakis and Ntzoufras, 2016*a*, *Braz.JPS)*.
- PCEP $\Rightarrow$ in Regression (Fouskakis and Ntzoufras, 2016*b*, *JCGS)*.
- DR/CR-PEP $\Rightarrow$ in Regression (Fouskakis et al., 2016, *arXiv:1609.06926v2)*.

Empirically based evidence:

- DR/CR-PEP for GLMs (Fouskakis, Ntzoufras and Perrakis, 2017, *BA, min.rev.)*.
- Hyper-δ versions (Fouskakis, Ntzoufras and Perrakis, 2017, *BA, min.rev.)*.

4 Illustration 1 (Normal Regression)

- 100 simulated data-sets
- $p = 10$ $N(0, 1)$ covariates
- The response is generated from

$$Y_i \sim N(0.3X_{i3} + 0.5X_{i4} + X_{i5}, \sigma^2), \quad \text{for} \quad i = 1, \ldots, n. \tag{8}$$

1. $\sigma = 2.5$ and n = 30, 50, 100, 500, 1000
2. $n = 50$ and $\sigma = 2.5, 1.5, 0.5, 0.01 \Rightarrow$
   $R^2 \in \{[0.15, 0.66],\ [0.26, 0.71],\ [0.73, 0.94], [0.99, 1]\}.$

Figure 2: Posterior probability of the true (per 100 simulated datasets of different sample sizes).

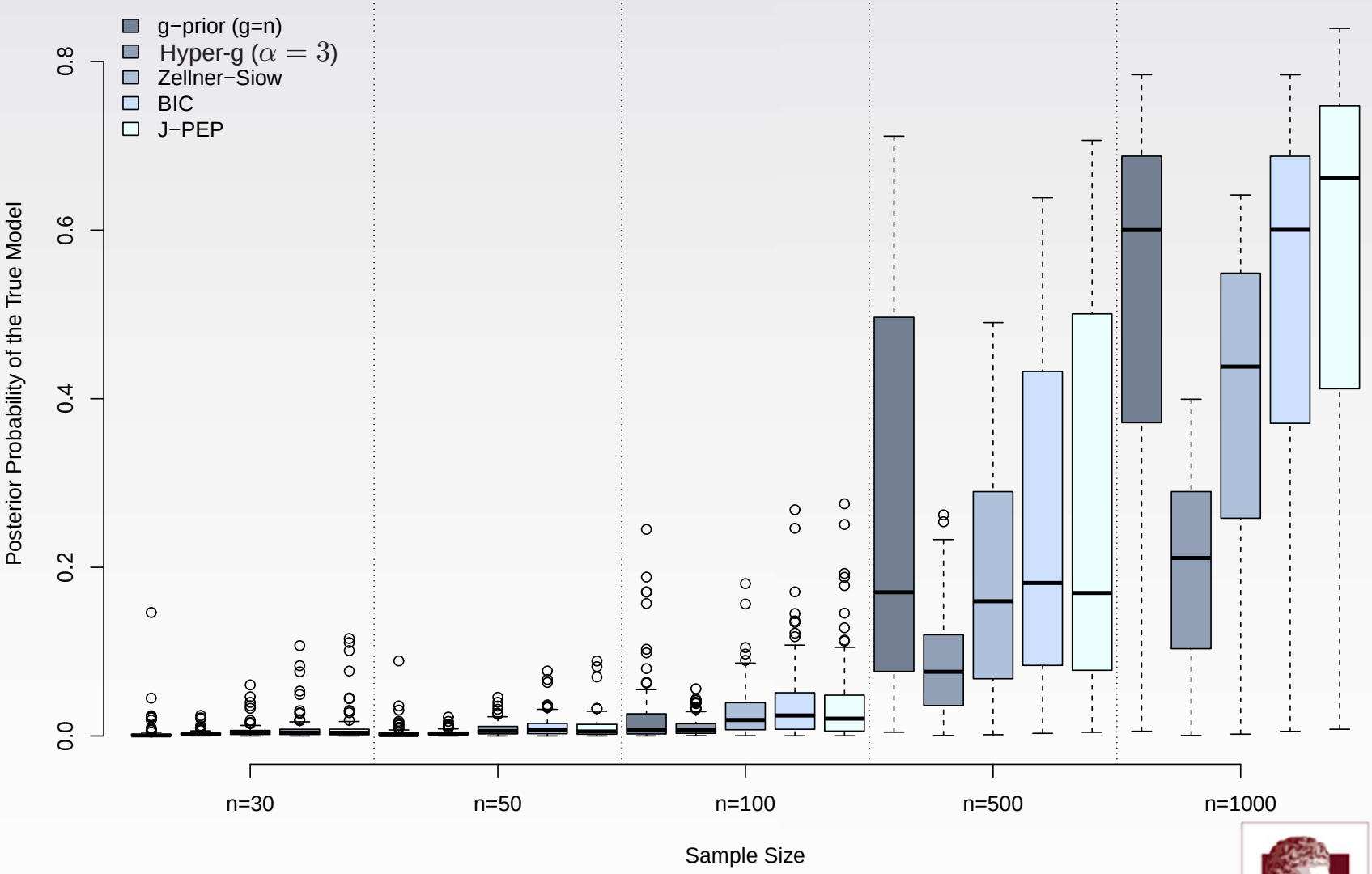
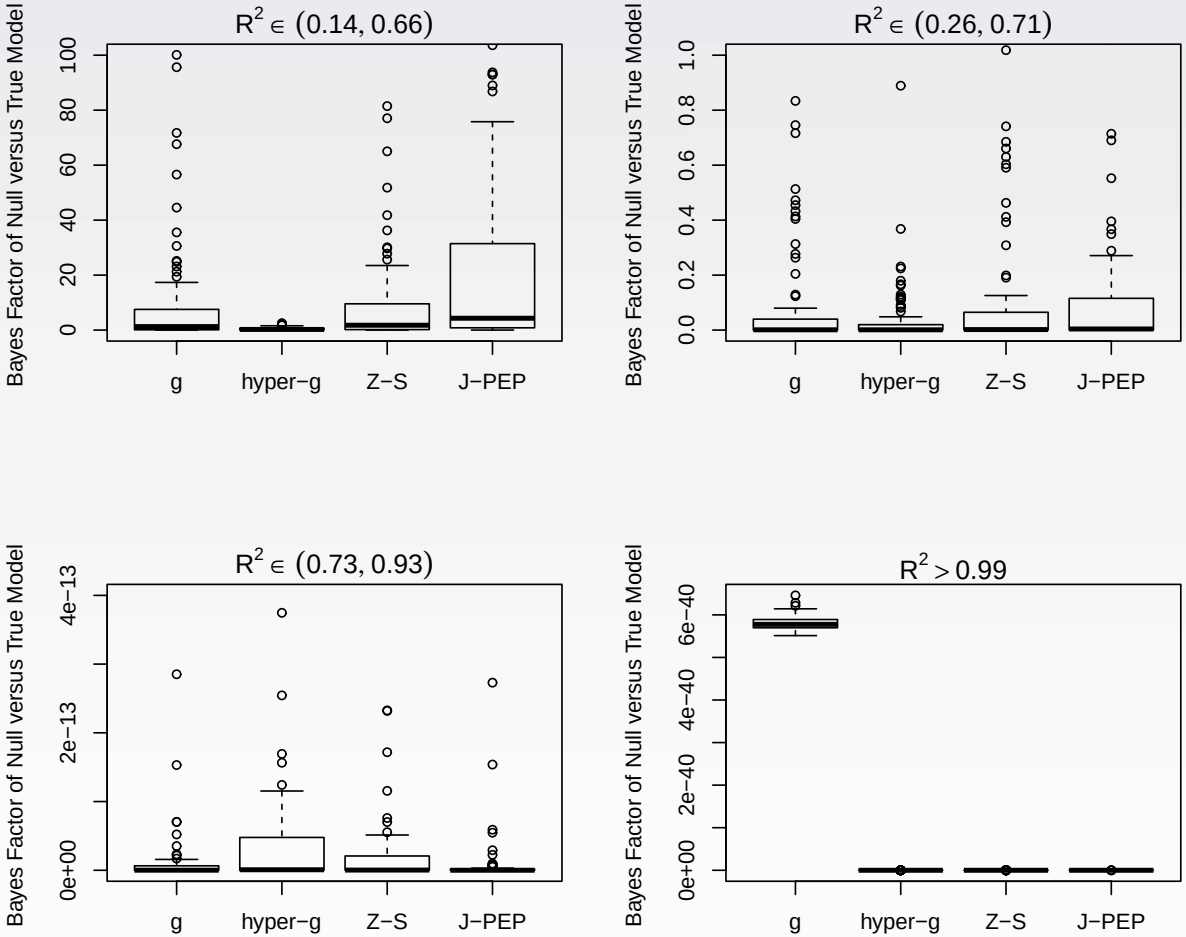


Figure 3: Bayes factor of the null model (only the intercept) versus the true model (per 100 simulated datasets with different coefficients of determination of the true model).



5 Illustration 2 (Poisson & Binomial Models)

- Also presented in Chen et al. (2008) and Li and Clyde (2015).
- $n = 100$, $p = 5$ and $p = 3$ predictors for logistic and Poisson scenarios respectively.
- Each simulation is repeated 100 times.
- Each predictor is drawn from a standard normal distribution with pairwise correlation given by

$$\text{corr}(X_i, X_j) = r^{|i-j|}, \;\; 1 \leq i < j \leq p.$$

  with (i) independent predictors ($r = 0$) and (ii) correlated predictors ($r = 0.75$).
- $n \in \{25, 100, 500, 1000, 10000\}$.

Scenario	Binomial Logistic Regression $(n = 100)$						Poisson $(n = 100)$			
	β_0	β_1	β_2	β_3	β_4	β_5	β_0	β_1	β_2	β_3
null	0.1	0	0	0	0	0	-0.3	0	0	0
sparse	0.1	0.7	0	0	0	0	-0.3	0.3	0	0
medium	0.1	1.6	0.8	-1.5	0	0	-0.3	0.3	0.2	0
full	0.1	1.75	1.5	-1.1	-1.4	0.5	-0.3	0.3	0.2	-0.15

Table 1: Simulation Binomial and Poisson regression scenarios.

Logistic Regression

Poisson Regression

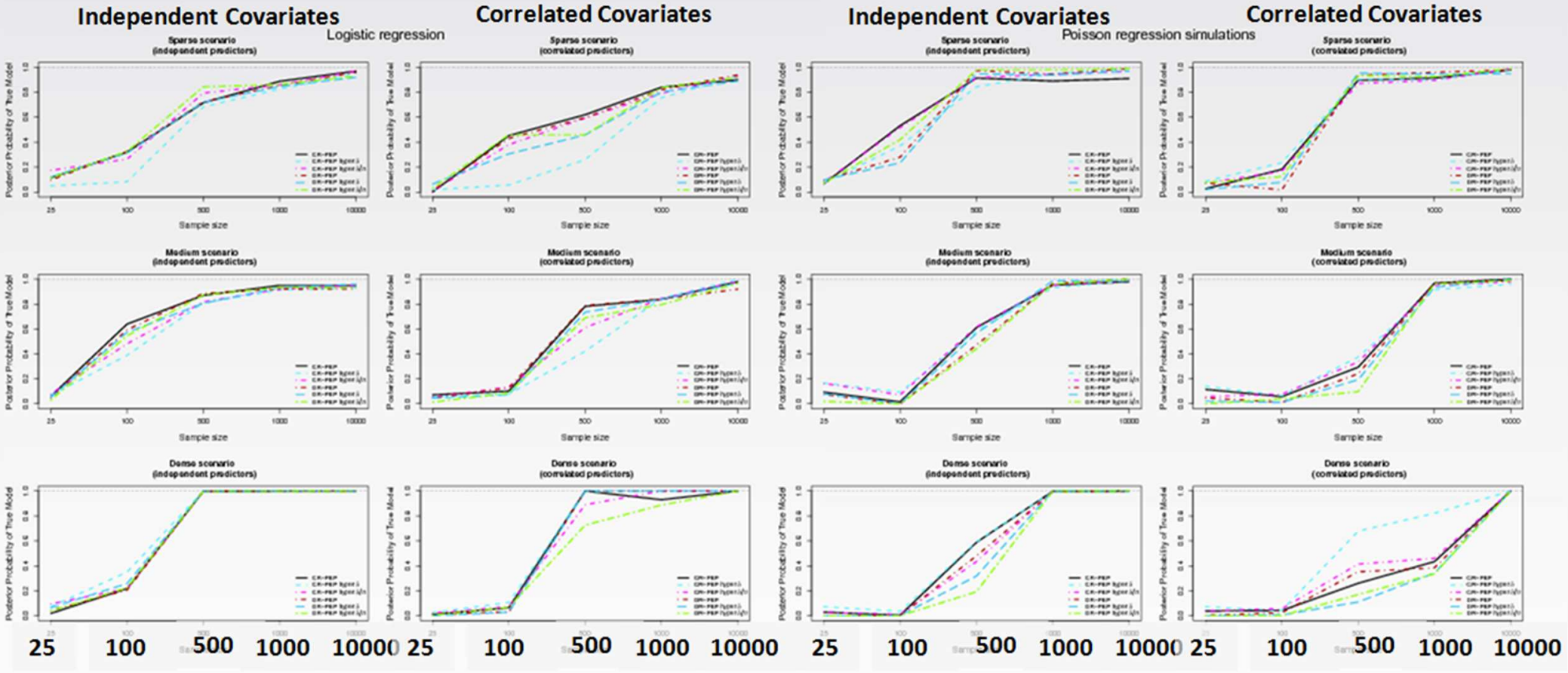
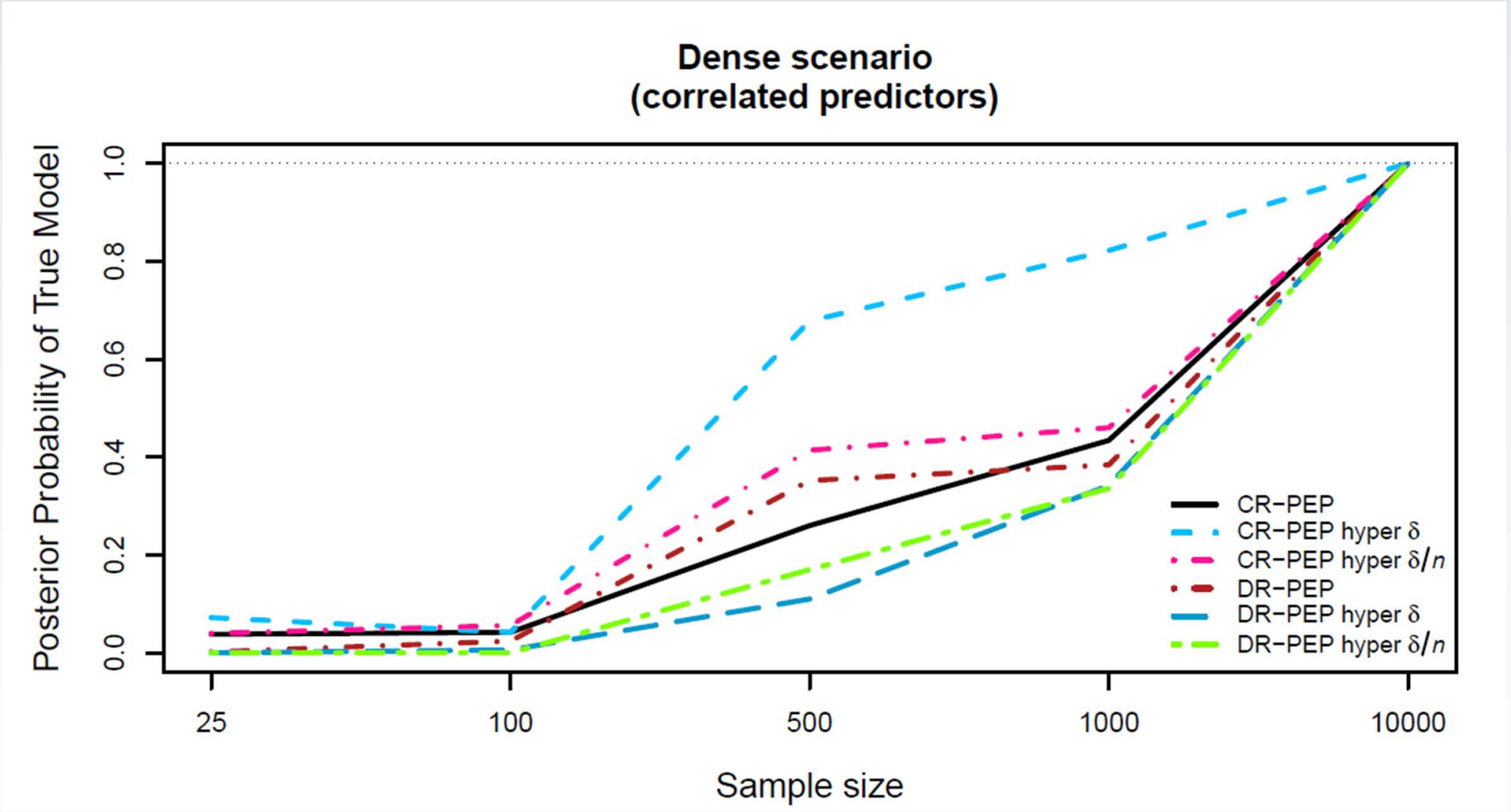


Figure 4: Posterior probabilities of the true model vs. sample size for the dense Poisson regression scenario.



6 General Framework and Concluding Remarks

$$\pi^{\mathrm{G}}(\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \boldsymbol{\omega}, \delta_0, \delta_1) = \pi^{\mathrm{G}}(\boldsymbol{\theta}_{\boldsymbol{\gamma}}|\boldsymbol{\omega}, \delta_0, \delta_1)\pi(\boldsymbol{\omega})\pi(\delta_0)\pi(\delta_1),$$

$$\pi^{\mathrm{G}}(\boldsymbol{\theta}_{\boldsymbol{\gamma}}|\boldsymbol{\omega}, \delta_0, \delta_1) = \frac{\pi^{\mathrm{N}}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}_{\boldsymbol{\gamma}}|\boldsymbol{\omega})}{k_{\boldsymbol{\gamma}}(\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \boldsymbol{\omega}, \delta_1)\mathcal{C}_0} \int \frac{m^{\mathrm{N}}_0(\boldsymbol{y}^*|\boldsymbol{\omega}, \delta_0)}{m^{\mathrm{N}}_{\boldsymbol{\gamma}}(\boldsymbol{y}^*|\boldsymbol{\omega}, \delta_1)} f_{\boldsymbol{\gamma}}(\boldsymbol{y}^*|\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \boldsymbol{\omega})^{1/\delta_1} d\boldsymbol{y}^*, \tag{9}$$

where

$$m^{\mathrm{N}}_{\boldsymbol{\gamma}}(\boldsymbol{y}^*|\boldsymbol{\omega}, \delta) = \int k_{\boldsymbol{\gamma}}(\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \boldsymbol{\omega}, \delta_1)^{-1} f_{\boldsymbol{\gamma}}(\boldsymbol{y}^*|\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \boldsymbol{\omega})^{1/\delta} \pi^{\mathrm{N}}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}_{\boldsymbol{\gamma}}|\boldsymbol{\omega}) d\boldsymbol{\theta}_{\boldsymbol{\gamma}}$$

Table 2: Schematic presentation of all priors in $\mathcal{P}$

Prior (G)	$\boldsymbol{\theta}_{\boldsymbol{\gamma}}$	$\boldsymbol{\omega}$	δ_0	δ_1	Hyper-prior $\pi(\delta)$	$k_0(\boldsymbol{\theta}_0, \boldsymbol{\omega}, \delta_0)$	$k_{\boldsymbol{\gamma}}(\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \boldsymbol{\omega}, \delta_1)$	$\mathcal{C}_0$
EP	$\boldsymbol{\beta}_{\gamma}, \phi_{\boldsymbol{\gamma}}$	$\emptyset$	1	1		1	1	1
PEP	$\boldsymbol{\beta}_{\gamma}, \phi_{\boldsymbol{\gamma}}$	$\emptyset$	n^*	n^*		κ_0	κ_1	1
PCEP	$\boldsymbol{\beta}_{\gamma}$	ϕ	n^*	n^*		κ_0	κ_1	1
CR-PEP	$\boldsymbol{\beta}_{\gamma}$	ϕ	1	n^*		1	1	1
DR-PEP	$\boldsymbol{\beta}_{\gamma}$	ϕ	n^*	n^*		1	1	c_0
CR-PEP hyper-δ	$\boldsymbol{\beta}_{\gamma}$	ϕ	1	δ	$\frac{a-2}{2}(1+\delta)^{-a/2}$	1	1	1
DR-PEP hyper-δ	$\boldsymbol{\beta}_{\gamma}$	ϕ	δ	δ	$\frac{a-2}{2}(1+\delta)^{-a/2}$	1	1	c_0
CR-PEP hyper-δ/n	$\boldsymbol{\beta}_{\gamma}$	ϕ	1	δ	$\frac{a-2}{2n}(1+\frac{\delta}{n})^{-a/2}$	1	1	1
DR-PEP hyper-δ/n	$\boldsymbol{\beta}_{\gamma}$	ϕ	δ	δ	$\frac{a-2}{2n}(1+\frac{\delta}{n})^{-a/2}$	1	1	c_0

$\kappa_0 = \int f_0(\boldsymbol{y}^*|\boldsymbol{\theta}_0, \boldsymbol{\omega})^{1/\delta_0} d\boldsymbol{y}^*$; $\kappa_1 = \int f_{\boldsymbol{\gamma}}(\boldsymbol{y}^*|\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \boldsymbol{\omega})^{1/\delta_1} d\boldsymbol{y}^*$; $c_0 = \int \int f_0(\boldsymbol{y}^*|\boldsymbol{\theta}_0, \boldsymbol{\omega})^{1/\delta_0} \pi_0^N(\boldsymbol{\theta}_0|\boldsymbol{\omega}) d\boldsymbol{\theta}_0 d\boldsymbol{y}^*$.

Table 3: Issues and solutions of all priors in $\mathcal{P}$

Prior	**Issues**	**Solutions**
EP	– Selection of imaginary sample size n^* – Sub-sampling of $\boldsymbol{X}_{\boldsymbol{\gamma}}^*$ – Informative when using minimal training sample and p is close to n	– Issues are solved using PEP with $\delta = n^* = n$ and $\boldsymbol{X}_{\boldsymbol{\gamma}}^* = \boldsymbol{X}_{\boldsymbol{\gamma}}$
PEP	– Cumbersome normalized power likelihood in GLMs – Monte Carlo is needed for the marginal likelihood even in the normal model – Selection of δ	– Use of unnormalized power likelihoods that lead to the CR/DR-PEP priors – Use PCEP that (conjugate for the normal model) or mixture-t expression – Set $\delta = n^*$ for unit interpretation or consider random δ
PCEP	– Not information consistent	– Use PEP which is information consistent

Table 3 (cont'd): Issues and solutions of all priors in $\mathcal{P}$

Prior	**Issues**	**Solutions**
DR-PEP	– No clear definition of m_0^{N} under the unnormalized power likelihood – Selection of δ	– Use the density normalized m_0^{Z} under the unnormalized power likelihood – Set $\delta = n^*$ for unit interpretation or consider random δ
CR-PEP	– Selection of δ	– Set $\delta = n^*$ for unit interpretation or consider random δ
hyper-δ	– Demanding computation – Prior of δ is not centered to unit-information	– Use fixed-δ CR/DR-PEP versions – Use the hyper-δ/n prior
hyper-δ/n	– Demanding computation	– Use fixed-δ CR/DR-PEP versions


Be Patient....
Two more talks and couple of posters until the BEACH.....

References

Fouskakis, D. and Ntzoufras (2017), 'Information Consistency of the Jeffreys Power-Expected-Posterior Prior in Gaussian Linear Models', *Metron (forthcoming)* .

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

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

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

Fouskakis, D., Ntzoufras, I. and Pericchi, L. R. (2017), 'Priors via imaginary training samples of sufficient statistics for objective bayesia model comparison', *(submitted)* .

Fouskakis, D., Ntzoufras, I. and Perrakis (2017), 'Power-Expected-Posterior Priors for Generalized Linear Models', *Bayesian Analysis (under minor revision)* .

Fouskakis, D., Ntzoufras, I. and Perrakis, K. (2016), 'Variations of power-expected-posterior priors in normal regression models', *arXiv:1609.06926v2* .

Pérez, J. M. and Berger, J. O. (2002), 'Expected-posterior prior distributions for model selection', *Biometrika* **89**, 491–511.