Discussion of "All About PEP" by Dimitris Fouskakis

Mark F. J. Steel

University of Warwick
Department of Statistics

O'Bayes 2019, Warwick

Principles

Thank you for a great presentation, based on 11(!) papers, and for the opportunity to discuss a “PEP talk”.

- the expected-posterior prior (Pérez and Berger, 2002)
- the unit-information prior approach (Kass and Wasserman, 1995)

and it is reassuring that the resulting prior satisfies many of the desiderata for objective priors of Bayarri et al. (2012)

Principles

Thank you for a great presentation, based on 11(!) papers, and for the opportunity to discuss a “PEP talk”.
It is really interesting to approach objective priors by combining a number of ideas:

- the expected-posterior prior (Pérez and Berger, 2002)
- the unit-information prior approach (Kass and Wasserman, 1995)

and it is reassuring that the resulting prior satisfies many of the desiderata for objective priors of Bayarri et al. (2012)

Principles

Thank you for a great presentation, based on 11(!) papers, and for the opportunity to discuss a “PEP talk”.
It is really interesting to approach objective priors by combining a number of ideas:

- the power-prior (Ibrahim and Chen, 2000)
- the expected-posterior prior (Pérez and Berger, 2002)
- the unit-information prior approach (Kass and Wasserman, 1995)

and it is reassuring that the resulting prior satisfies many of the desiderata for objective priors of Bayarri et al. (2012)

Principles

Thank you for a great presentation, based on 11(!) papers, and for the opportunity to discuss a “PEP talk”.
It is really interesting to approach objective priors by combining a number of ideas:

- the power-prior (Ibrahim and Chen, 2000)
- the expected-posterior prior (Pérez and Berger, 2002)
- the unit-information prior approach (Kass and Wasserman, 1995)

and it is reassuring that the resulting prior satisfies many of the desiderata for objective priors of Bayarri et al. (2012)

Principles

Thank you for a great presentation, based on 11(!) papers, and for the opportunity to discuss a “PEP talk”.
It is really interesting to approach objective priors by combining a number of ideas:

- the power-prior (Ibrahim and Chen, 2000)
- the expected-posterior prior (Pérez and Berger, 2002)
- the unit-information prior approach (Kass and Wasserman, 1995)

and it is reassuring that the resulting prior satisfies many of the desiderata for objective priors of Bayarri et al. (2012)

Principles

Thank you for a great presentation, based on 11(!) papers, and for the opportunity to discuss a “PEP talk”.
It is really interesting to approach objective priors by combining a number of ideas:

- the power-prior (Ibrahim and Chen, 2000)
- the expected-posterior prior (Pérez and Berger, 2002)
- the unit-information prior approach (Kass and Wasserman, 1995)

and it is reassuring that the resulting prior satisfies many of the desiderata for objective priors of Bayarri et al. (2012)

Sanity checks

Can we think of further “sanity checks” or properties that help us understand the priors proposed here?

Sanity checks

Can we think of further "sanity checks" or properties that help us understand the priors proposed here?

- What is the implied complexity prior induced by the Bayes factor? For example, in the case of PEP for the normal linear model, the expression for the BF can be used to derive this (using $RSS_p = RSS_l$ and $d_l = p - 1$) as a simple univariate integral, function of δ. For comparison, in the "standard" g-prior setup, this penalty for an extra regressor is $\sqrt{1+g}$.

Sanity checks

Can we think of further “sanity checks” or properties that help us understand the priors proposed here?

- What is the implied complexity prior induced by the Bayes factor? For example, in the case of PEP for the normal linear model, the expression for the BF can be used to derive this (using $RSS_p = RSS_l$ and $d_l = p - 1$) as a simple univariate integral, function of δ. For comparison, in the “standard” g-prior setup, this penalty for an extra regressor is $\sqrt{1+g}$.
- What is the meaning of the truncation on the implied prior for g? This can have practical consequences if it is truncated away from where most of the likelihood mass is. For example, $g > n$ (for PEP in the normal linear model) is quite dramatic.

Large p extensions?

- Most of the presentation requires the assumption that $n > p$. Applications in some areas often violate this.

Large p extensions?

- Most of the presentation requires the assumption that $n > p$. Applications in some areas often violate this.
- Using Ridge or Lasso baseline priors (or g-prior with Ridge component) this is circumvented for normal linear models.

Large p extensions?

- Most of the presentation requires the assumption that $n > p$. Applications in some areas often violate this.
- Using Ridge or Lasso baseline priors (or g-prior with Ridge component) this is circumvented for normal linear models.
- Can we, more generally, find an extension of PEP by either
  - Assigning zero prior probability to models for which $p_l > n$, which is the usual “informal” approach for g-priors

Large p extensions?

- Most of the presentation requires the assumption that $n > p$. Applications in some areas often violate this.
- Using Ridge or Lasso baseline priors (or g-prior with Ridge component) this is circumvented for normal linear models.
- Can we, more generally, find an extension of PEP by either
  - Assigning zero prior probability to models for which $p_l > n$, which is the usual “informal” approach for g-priors
  - Mimicking formal approaches to use g-priors in situations where $p > n$ such as Maruyama and George (2011) and Berger et al. (2016), based on different ways of generalizing the notion of inverse matrices?

Model Selection Consistency

Key desirable property

Model Selection Consistency

Key desirable property

- Why emphasis on consistency of the Bayes factor of the full model M_p versus a generic submodel M_l? Hyper-g (and other) priors are not consistent when null model is "truth", so why not include results for data generated under the null model?

Model Selection Consistency

Key desirable property

- Why emphasis on consistency of the Bayes factor of the full model M_p versus a generic submodel M_l? Hyper-g (and other) priors are not consistent when null model is “truth”, so why not include results for data generated under the null model?
- Consistency may depend on the prior on model space. What is the prior on model space recommended here?

Model Selection Consistency

Key desirable property

- Why emphasis on consistency of the Bayes factor of the full model M_p versus a generic submodel M_l? Hyper-g (and other) priors are not consistent when null model is "truth", so why not include results for data generated under the null model?
- Consistency may depend on the prior on model space. What is the prior on model space recommended here?
- Slides (p.30) state "we have showed **empirically** model selection consistency". For g-priors, simple conditions on g can be derived to ensure consistency. Can something similar not be done here (on δ)? Perhaps you do have such conditions for the normal linear model (but not for all GLMs)?

Model Selection Consistency

Key desirable property

- Why emphasis on consistency of the Bayes factor of the full model M_p versus a generic submodel M_l? Hyper-g (and other) priors are not consistent when null model is “truth”, so why not include results for data generated under the null model?
- Consistency may depend on the prior on model space. What is the prior on model space recommended here?
- Slides (p.30) state “we have showed **empirically** model selection consistency”. For g-priors, simple conditions on g can be derived to ensure consistency. Can something similar not be done here (on δ)? Perhaps you do have such conditions for the normal linear model (but not for all GLMs)?
- Does consistency extend in a sensible way to the $\mathcal{M}$-open setting?

GLMs

- Li and Clyde (2018) recommend prior structures for GLMs satisfying the Bayarri et al. (2012) desiderata. Very similar goal, but different approach (starting from the local geometry at the MLE).

GLMs

- Li and Clyde (2018) recommend prior structures for GLMs satisfying the Bayarri et al. (2012) desiderata. Very similar goal, but different approach (starting from the local geometry at the MLE).
- Would be nice to see a comparison of the recommended priors, both in terms of theoretical properties and in terms of empirical performance.

GLMs

- Li and Clyde (2018) recommend prior structures for GLMs satisfying the Bayarri et al. (2012) desiderata. Very similar goal, but different approach (starting from the local geometry at the MLE).
- Would be nice to see a comparison of the recommended priors, both in terms of theoretical properties and in terms of empirical performance.
- Li and Clyde (2018) use a truncated Compound Confluent Hypergeometric distribution on $1/(1+g)$. How does this relate to the shifted generalized beta prime distribution proposed here (on g) which implies a Beta truncated Pareto distribution for the shrinkage factor $g/(1+g)$? (covers the same special cases)

GLMs

- Li and Clyde (2018) recommend prior structures for GLMs satisfying the Bayarri et al. (2012) desiderata. Very similar goal, but different approach (starting from the local geometry at the MLE).
- Would be nice to see a comparison of the recommended priors, both in terms of theoretical properties and in terms of empirical performance.
- Li and Clyde (2018) use a truncated Compound Confluent Hypergeometric distribution on $1/(1+g)$. How does this relate to the shifted generalized beta prime distribution proposed here (on g) which implies a Beta truncated Pareto distribution for the shrinkage factor $g/(1+g)$? (covers the same special cases)

Some further comments

- Difference between PEP and PCEP seems to be conditioning on a common parameter, assigning a standard default prior to this. Why only condition on σ^2 and not on (β_0, σ^2)? Both are common and have convenient standard priors with nice properties.

Some further comments

- Difference between PEP and PCEP seems to be conditioning on a common parameter, assigning a standard default prior to this. Why only condition on σ^2 and not on (β_0, σ^2)? Both are common and have convenient standard priors with nice properties.
- Table 5 can be extended quite a bit. For example, the beta prime or gamma-gamma prior on *g* (possibly truncated) also covers the priors by
  - Maruyama and George (2011)
  - Bottolo and Richardson (2008) (right truncated)
  - horseshoe prior of Carvalho et al. (2010) (not consistent)
  - benchmark prior of Ley and Steel (2012)

  and the original Zellner-Siow (1980) prior corresponds to an inverted gamma prior on *g*.