Adaptive prior weighting in generalized regression

Leonhard Held

18. September 2015

University of Zurich[UZH]

Adaptive Prior Weighting

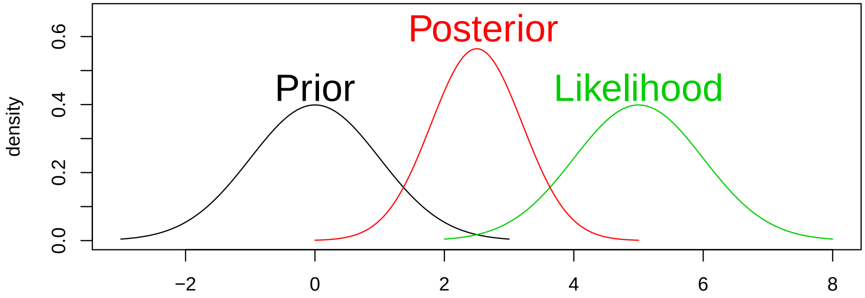
I. Power prior in clinical trials (joint with Isaac Gravestock)

II. Generalized Regression (joint with Rafael Sauter)

Prior-Data Conflict

> "Bayesian: One who, vaguely expecting a horse and catching a glimpse of a donkey, strongly concludes that he has seen a mule."

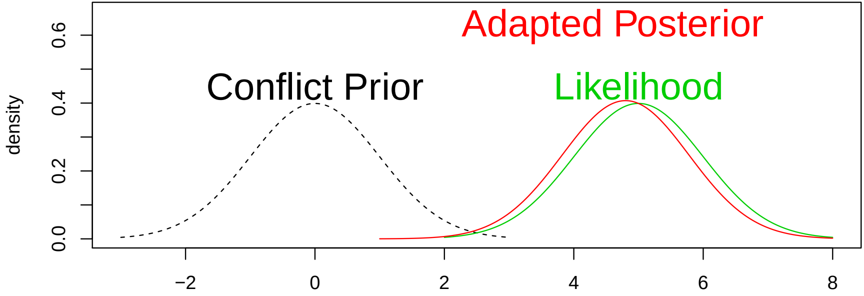
Senn (2007)



Adaptive Prior Weighting

"Bayesian: One who, vaguely expecting a horse and catching a glimpse of a donkey, concludes that he has seen . . . most likely a donkey!"

not Senn (2007)



Where Does the Prior Come From?

1. Historical data

   Spiegelhalter *and others* (2004)

2. Expert opinion

   O'Hagan *and others* (2006)

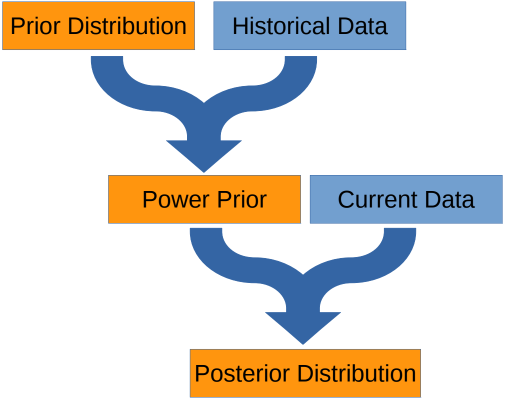
3. Structural considerations

   Greenland (2006, 2007)

   e.g. $\Pr(1/10 \leq \mathrm{OR} \leq 10) = 0.95$ in logistic regression

4. Proper default priors in regression
   - Ridge prior
   - g-prior
   - Lasso

The Power Prior



The Power Prior

1. Start with a prior on the parameter θ, maybe uninformative or uniform

$$p_0(\theta)$$

2. Update the prior based on the likelihood $L(\theta; x_0)$ of historical data x_0 raised to a power δ between 0 and 1

$$p(\theta \mid \delta, x_0) \propto L(\theta; x_0)^{\delta} p_0(\theta)$$

Ibrahim and Chen (2000)

→ If the historical study has n_0 patients, then the prior sample size is δn_0

3. Current study has $n_\star$ patients and data $x_\star$, say.
4. Current study is combined with power prior to posterior

→ Total sample size is $n_\star + \delta n_0$

Unknown Power Parameter

- Treat δ as unknown and include prior $p_0(\delta)$
- Requires normalisation:

$$
\begin{aligned}
p(\theta, \delta \mid x_0) &= p(\theta \mid \delta, x_o)\, p_0(\delta) \\
&= \frac{L(\theta; x_0)^\delta p_0(\theta)}{\int L(\theta; x_0)^\delta p_0(\theta) d\theta}\, p_0(\delta)
\end{aligned}
$$

Duan *and others* (2006)
Neuenschwander *and others* (2009)

- Joint posterior:

$$
p(\theta, \delta \mid x_\star, x_0) \propto L(\theta; x_\star)\, p(\theta, \delta \mid x_0)
$$

Choosing δ

Possible approaches for choosing δ from the literature:

- Pick some fixed values and do a sensitivity analysis afterwards
- Don't. Integrate it out of joint posterior and use a fully Bayesian approach

We propose an empirical Bayes (EB) method:

- Combines Bayesian and classical ideas
- Select the best prior based on the data
- Maximise the marginal likelihood to choose δ:

$$
\begin{aligned}
\hat{\delta}_{\text{EB}} &= \underset{\delta \in [0,1]}{\arg\max}\, L(\delta; x_0, x_\star) \\
&= \underset{\delta \in [0,1]}{\arg\max}\, \frac{\int L(\theta; x_\star) L(\theta; x_0)^\delta p_0(\theta) d\theta}{\int L(\theta; x_0)^\delta p_0(\theta) d\theta}.
\end{aligned}
$$

Binomial Model

Want to estimate true proportion θ

- Initial prior: $\theta \sim \text{Beta}(a, b)$
- Historical data: $X_0 \sim \text{Bin}(n_0, \theta)$
- Current data: $X_\star \sim \text{Bin}(n_\star, \theta)$
- Prior for power parameter δ (for fully Bayesian approach): $\delta \sim \text{Beta}(\alpha, \beta)$

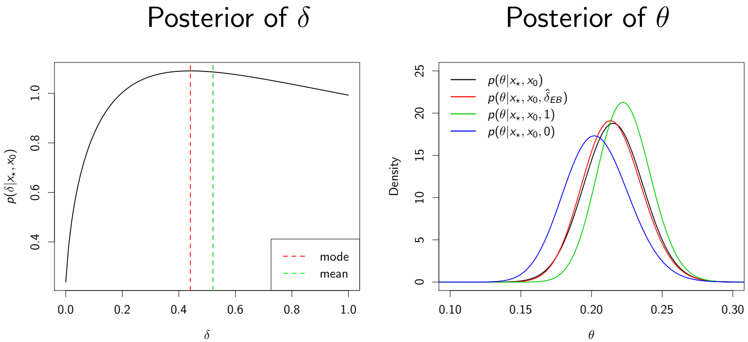
Antibiotics Trials

- Treating nosocomial pneumonia
- 2 studies comparing Linezolid and Vancomycin
- Binary outcome: all cause mortality

Study	Linezolid	Vancomycin
Rubinstein (2001)	36/203 (18%)	49/193 (25%)
Wunderink (2003)	64/321 (20%)	61/302 (20%)

Vancomycin Estimates

- Start with uniform priors on θ (and δ)
- Use Rubinstein results as historical data: $x_0/n_0 = 49/193$
- Empirical Bayes: $\hat{\delta} = 0.44 \rightarrow$ prior sample size $\hat{\delta} n_0 = 86$
- Posterior mean: $\hat{\delta} = 0.52 \rightarrow$ prior sample size $\hat{\delta} n_0 = 101$



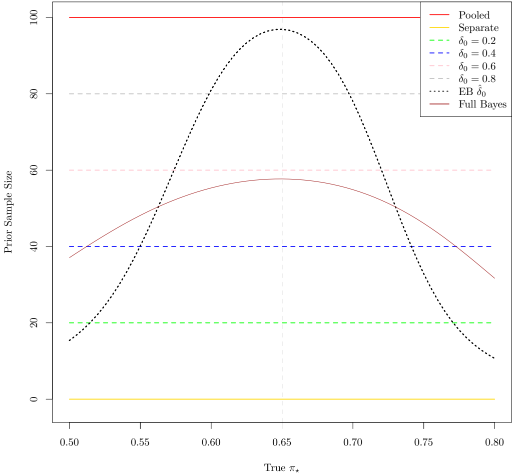
Operating Characteristics

- Viele et al. (2014) look at performance of various borrowing methods for the control arm of a randomized controlled clinical trial (RCT).
  - Expected Prior Sample Size
  - Mean Square Error
  - Power
  - Type I Error
- Binomial setting:

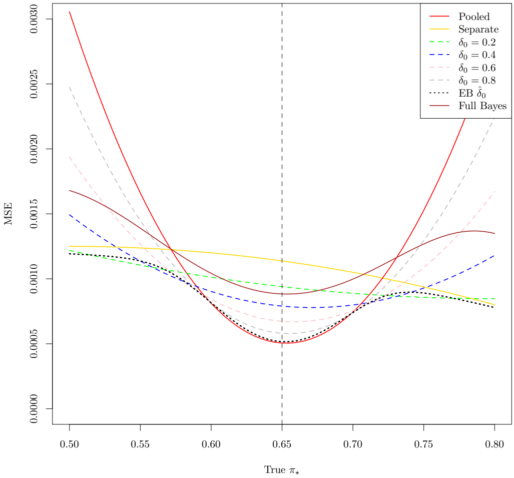
Historical control arm data:	$x_0 = 65,\ n_0 = 100$
Current control arm data:	$X_\star \sim \text{Bin}(n_\star = 200, \pi_\star)$
Current treatment arm data:	$X_T \sim \text{Bin}(n_T = 200, \pi_T)$

- "Bayesian Significance" if $\Pr(\pi_T > \pi_\star) > 0.975$
- Empirical Bayes compares favourably!

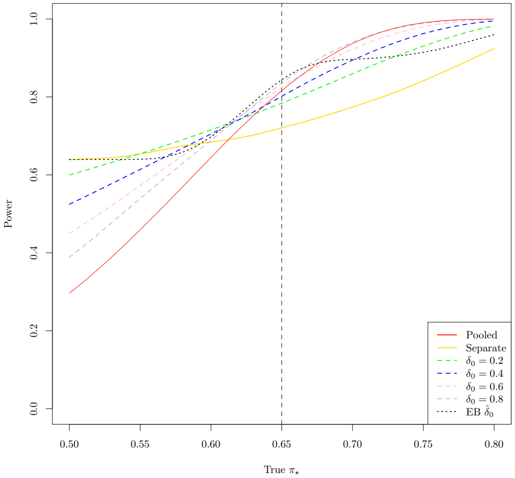
Expected Prior Sample Size



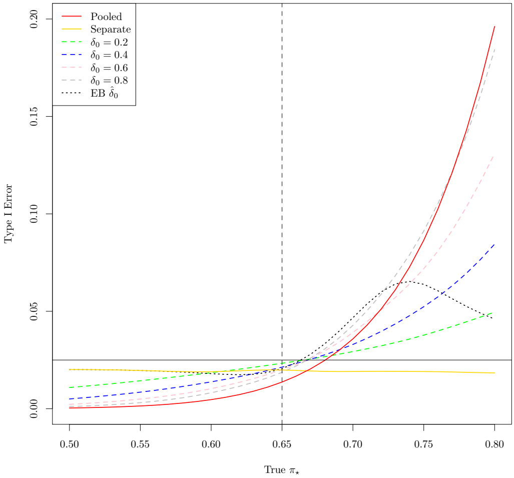
Mean Square Error



Power to Detect Difference $\pi_T - \pi_\star = 12\%$



Type I Error $\pi_T - \pi_\star$



Box's p-value

- Box (1980) defined a p-value to measure conflict between prior and data $X_\star = x_\star$:

$$\Pr\left(p(X_\star \mid x_0) \leq p(x_\star \mid x_0)\right)$$

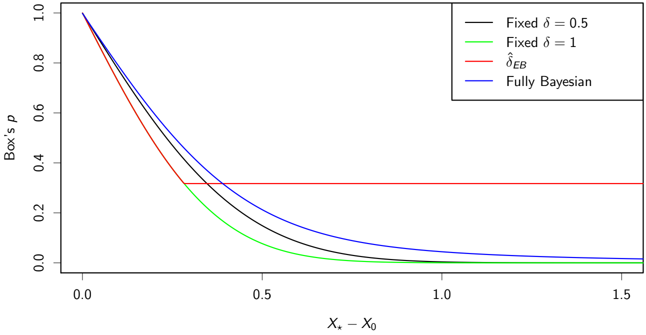
→ Probability that observed (or more extreme) data could come from prior predictive distribution $p(x_\star \mid x_0)$

- Low p-values suggest that prior is in conflict with data
- Calculate predictive distribution by

$$p(x_\star \mid x_0) = \int_\theta \int_\delta p(x_\star \mid \theta) \times p(\theta, \delta \mid x_0)\, d\delta d\theta$$

Box's p-value for Power Prior

- Power prior based on normal likelihood: $X_0 \sim \mathrm{N}(\theta, 0.2^2)$
- Current data also normal: $X_\star \sim \mathrm{N}(\theta, 0.2^2)$
- Uniform prior on θ (and δ)



$\rightarrow$ EB automatically adjusts compatibility between prior and data.

II. Adaptive Prior Weighting in Regression

- A multivariate normal prior distribution on the regression coefficients is a natural choice, especially if prior comes from historical data.
- However, careless specification of mean and covariance matrix may have strong effects on posterior inference.
- We discuss empirical and fully Bayesian approaches to avoid extreme prior-data disagreement and agreement, by adaptively weighting the prior distribution.
- The proposed methodology provides an alternative to the recently proposed Cauchy prior distributions for the regression coefficients of suitably standardized covariates.

Gelman *and others* (2008)

Preliminaries: Prior Weighting in the Linear Model

Consider the linear model with mean

$$\mathsf{E}(y_i) = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta}$$

with residual variance σ^2 and centred regression coefficients $\mathbf{X}^\top \mathbf{1} = \mathbf{0}$. An improper reference prior for α and σ^2, $f(\alpha, \sigma^2) \propto \sigma^{-2}$, is combined with a proper normal prior for $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} \,|\, \sigma^2 \sim \mathsf{N}_d(\boldsymbol{\nu}, g\,\sigma^2\,\boldsymbol{\Sigma})$$

→ Prior weight $\delta = 1/g > 0$

Note: Prior weight δ can be larger 1

→ Prior up- and downweighting possible

The g-prior

- Zellner's g-prior (Zellner, 1986) uses $\mathbf{\Sigma} = (\mathbf{X}^\top \mathbf{X})^{-1}$

$\rightarrow$ Shrinkage of $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ towards $\boldsymbol{\nu}$:

$$\mathsf{E}(\boldsymbol{\beta} \,|\, \mathbf{y}) = \frac{\hat{\boldsymbol{\beta}}_{\mathrm{ML}} + 1/g \cdot \boldsymbol{\nu}}{1 + 1/g}$$

- For $\boldsymbol{\nu} = \mathbf{0}$ we have

$$\mathsf{E}(\boldsymbol{\beta} \,|\, \mathbf{y}) = \frac{g}{g+1}\, \hat{\boldsymbol{\beta}}_{\mathrm{ML}}$$

$\rightarrow$ Shrinkage factor $t = g/(g+1)$

Box's p-Value

- Take $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ as the "data" with distribution

$$\hat{\boldsymbol{\beta}}_{\mathrm{ML}} \,|\, \boldsymbol{\beta}, \sigma^2 \sim \mathrm{N}_d\left(\boldsymbol{\beta}, \sigma^2\,(\mathbf{X}^\top\mathbf{X})^{-1}\right)$$

- Combined with prior $\boldsymbol{\beta} \,|\, \sigma^2 \sim \mathrm{N}_d(\boldsymbol{\nu}, g\,\sigma^2\,\boldsymbol{\Sigma})$ we obtain the prior predictive distribution

$$\hat{\boldsymbol{\beta}}_{\mathrm{ML}} \,|\, \sigma^2 \sim \mathrm{N}_d\left(\boldsymbol{\nu}, \sigma^2\left\{(\mathbf{X}^\top\mathbf{X})^{-1} + g\,\boldsymbol{\Sigma}\right\}\right)$$

so

$$T(g) = \left(\hat{\boldsymbol{\beta}}_{\mathrm{ML}} - \boldsymbol{\nu}\right)^\top \left\{(\mathbf{X}^\top\mathbf{X})^{-1} + g\,\boldsymbol{\Sigma}\right\}^{-1} \left(\hat{\boldsymbol{\beta}}_{\mathrm{ML}} - \boldsymbol{\nu}\right)/\sigma^2$$

can be evaluated against a $\chi^2(d)$ distribution.

- Exact calculation based on F-distribution is also possible.

Box's p-Value: Some Properties

- Box's $p \to 1$ for the "uninformative" choice $g \to \infty$
- Copas (1983) has derived an empirical Bayes (EB) estimate of g under the g-prior:

$$\hat{g} = \max\{F_{\text{obs}} - 1, 0\},$$

where F_{obs} is the F statistic for $H_0 : \boldsymbol{\beta} = \boldsymbol{\nu}$.

→ For $F_{\text{obs}} > 1$, one can show that Box's p-value is $p \approx 0.5$, *i. e.* empirical Bayes automatically adjusts the compatibility between the prior and the data.

- This suggests to estimate g for arbitrary prior covariance matrix $\boldsymbol{\Sigma}$ by solving the equation $T(g) = \mathsf{E}\{\chi^2(d)\} = d$ for g, *i. e.*

$$\hat{g} = \begin{cases} T^{-1}(d) & \text{if } T(0) > d \\ 0 & \text{else} \end{cases}$$

Generalized Linear Model

- Consider now a generalized linear model (GLM) with outcomes y_i with mean $\mu_i = h(\eta_i)$ and linear predictor

$$\eta_i = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta}$$

- Prior $f(\alpha) \propto 1$ combined with $\boldsymbol{\beta} \sim \mathrm{N}_d(\boldsymbol{\nu}, g\,\boldsymbol{\Sigma})$
- MLE $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$, where $\hat{\boldsymbol{\beta}}_{\mathrm{ML}} \,|\, \boldsymbol{\beta} \sim \mathrm{N}_d(\boldsymbol{\beta}, \boldsymbol{\mathcal{T}})$, is used to evaluate

$$T(g) = (\hat{\boldsymbol{\beta}}_{\mathrm{ML}} - \boldsymbol{\nu})^\top (\boldsymbol{\mathcal{T}} + g\,\boldsymbol{\Sigma})^{-1} (\hat{\boldsymbol{\beta}}_{\mathrm{ML}} - \boldsymbol{\nu})$$

  against a $\chi^2(d)$ distribution to compute Box's p-value.
- EB estimate under the (generalized) g-prior based on the deviance z_{obs} (Copas, 1997):

$$\hat{g} = \max\{z_{\mathrm{obs}}/d - 1, 0\}.$$

- Can be extended to arbitrary $\mathrm{N}_d(\boldsymbol{\nu}, g\,\boldsymbol{\Sigma})$ prior by solving $T(g) = d$.

Application: Study on Obstetric Care and Neonatal Death

From Sullivan and Greenland (2013):

Table 1 Multiple logistic regressions of neonatal-death risk in a cohort of 2992 births with 17 deaths, intercept and 14 regressors in each model. Shown are the prior median odds ratio OR_{prior} and 95% limits; ML estimates with 95% Wald and profile-likelihood (profile) limits; approximate posterior medians from data augmentation including a prior on all 14 regressors with 95% Wald and profile limits, using prior data with ½ correction (A = 4.5) or with rescaled prior data (S = 10, A = 400); and simulated posterior medians and 95% limits (2.5th and 97.5th percentiles) from MCMC with normal priors

				Approximate posterior median (95% posterior limits)		
Regressor (X_j)	**Deaths with $X_j > 0$**	**Prior median OR_{prior} (95% prior limits)**	**ML estimate: A = 0 (95% Wald and profile limits)**	Data augmentation: A = 4.5[a]	Data augmentation with a rescaled ($S = 10$) prior[a]	MCMC[b]
Non-White	5	2 (0.5,8)	1.9 (0.55,6.5) (0.51,6.3)	1.8 (0.75,4.2) (0.72,4.1)	1.8 (0.73,4.3) (0.71,4.2)	1.8 (0.70,4.2)
Early age	3	2 (0.5,8)	1.6 (0.39,6.7) (0.32,6.1)	1.6 (0.65,4.1) (0.62,3.9)	1.6 (0.63,4.1) (0.61,4.0)	1.6 (0.59,4.0)
Nulliparity	8	2 (0.5,8)	1.5 (0.51,4.7) (0.50,4.9)	1.6 (0.69,3.5) (0.68,3.6)	1.5 (0.67,3.6) (0.67,3.6)	1.6 (0.67,3.6)
Gestational age	10	4 (1,16)	4.9 (2.4,9.8) (2.4,10.0)	4.5 (2.5,8.0) (2.5,8.1)	4.5 (2.5,8.1) (2.5,8.1)	4.6 (2.5,8.3)
Isoimmunization	1	2 (0.5,8)	3.0 (0.91,10) (0.62,8.5)	2.4 (0.95,6.0) (0.87,5.6)	2.4 (0.94,6.2) (0.85,5.7)	2.3 (0.81,5.6)
Past abortion	2	1 (0.25,4)	0.72 (0.18,2.9) (0.12,2.3)	0.84 (0.34,2.1) (0.31,1.9)	0.83 (0.33,2.1) (0.31,1.9)	0.79 (0.29,1.9)
Hydramnios	1	4 (1,16)	60 (5.7,635) (2.8,478)	5.8 (1.6,21) (1.6,22)	6.1 (1.6,23) (1.6,23)	6.0 (1.6,22)
Labour progress	2	2 (0.5,8)	0.50 (0.06,3.9) (0.04,2.8)	1.3 (0.45,3.5) (0.42,3.3)	1.2 (0.43,3.5) (0.41,3.3)	1.2 (0.40,3.3)
PCA	1	2 (0.5,8)	3.1 (0.33,29) (0.15,20)	2.2 (0.71,7.1) (0.67,7.0)	2.3 (0.68,7.5) (0.65,7.2)	2.2 (0.64,7.1)
No monitor	3	2 (0.5,8)	1.2 (0.32,4.9) (0.35,5.9)	1.8 (0.68,4.5) (0.70,4.7)	1.7 (0.66,4.6) (0.68,4.8)	1.8 (0.71,5.0)
Twin, triplet	3	4 (1,16)	8.2 (1.8,37) (1.5,33)	5.1 (1.9,14) (1.8,14)	5.2 (1.9,15) (1.8,14)	5.3 (1.8,14)
Public ward	6	2 (0.5,8)	0.86 (0.26,2.9) (0.25,2.8)	1.3 (0.56,3.0) (0.54,3.0)	1.3 (0.54,3.0) (0.53,3.0)	1.3 (0.53,3.0)
PROM	1	2 (0.5,8)	0.54 (0.06,4.8) (0.03,3.2)	1.3 (0.45,3.5) (0.41,3.3)	1.2 (0.43,3.5) (0.41,3.3)	1.2 (0.39,3.3)
Malpresented	3	4 (1,16)	3.9 (0.88,17) (0.73,15)	3.9 (1.5,10.0) (1.4,9.8)	3.9 (1.4,10) (1.4,9.9)	3.8 (1.4,10.0)

ML: Maximum likelihood; MCMC: Markov-chain Monte Carlo; PCA: placental or cord anomaly. PROM: premature rupture of membranes.
Variables are indicators except early age (0 = 20+, 1 = 15−19, 2 = under 15), gestational age (0 = no, 1 = 36−38 weeks, 2 = 33−36 weeks; under 33 weeks excluded), isoimmunization (0 = no, 1 = Rh, 2 = ABO), labour progress (0 = no, 0.33 = prolonged, 0.67 = protracted, 1 = arrested) and past abortion (0 = none, 1 = 1, 2 = 2+).
[a]Limits shown are Wald exp(estimate ∓ 1.96 × standard error) and profile likelihood from PROC LOGISTIC.
[b]Number of MCMC samples was 100 000

Assessment of Prior-Data Conflict

$$\begin{array}{lll} \text{Prior I:} & \mathrm{N}_{14}(\boldsymbol{\nu}, g\,\boldsymbol{\Sigma}) & \text{(Sullivan and Greenland (2013) prior)} \\ \text{Prior II:} & \mathrm{N}_{14}(\mathbf{0}, g\,\boldsymbol{\Sigma}) & \text{(ridge prior)} \end{array}$$

with

$$\begin{aligned} \boldsymbol{\nu} &= \log(2,2,2,4,2,1,4,2,2,2,4,2,2,4)^\top \\ \boldsymbol{\Sigma} &= \mathrm{diag}(1/2) \end{aligned}$$

- Prior I gives Box's $p = 0.91$ (!) for $g = 1$.
- Prior II gives Box's $p = 0.13$ for $g = 1$.

→ no evidence for prior-data conflict under both priors.

- The EB estimates are $\hat{g} = 0$ ($p = 0.60$) and $\hat{g} = 2.10$ ($p = 0.45$)
- However, EB estimates $\hat{g} = 0$ are useless for parameter estimation.

Hyper-g Prior

- The EB approach avoids arbitrary choices of g which may be at odds with the data. However, the uncertainty about the estimate $\hat{g}$ is ignored and the posterior of $\boldsymbol{\beta}$ is degenerate for $\hat{g} = 0$.
- We propose to use the hyper-g prior with shrinkage factor

$$t = g/(1+g) \sim \mathrm{U}(0,1)$$

Liang *and others* (2008)

$\rightarrow$ Prior median of g is 1, distribution of g and $\delta = 1/g$ are the same.

$\rightarrow$ No preference regarding up- or downweighting

- Under the g-prior, posterior mode of t is equal to the corresponding EB estimate.

Held *and others* (2015)

$\rightarrow$ Hyper-g prior regularizes EB approach.

Other Choices

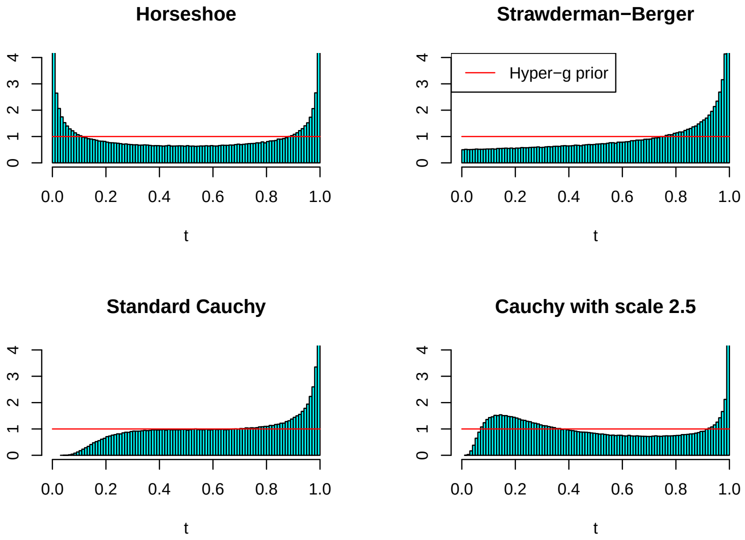
- Horseshoe prior: $t \sim \text{Be}(1/2, 1/2)$

  Carvalho *and others* (2010)
- Strawderman-Berger: $t \sim \text{Be}(1, 1/2)$

  Berger (1980)
- Cauchy prior distributions for the regression coefficients corresponds to a (possibly scaled) $\text{IG}(1/2, 1/2)$ prior for g

  Gelman *and others* (2008)

Other Choices



Implementation in INLA

- Rewrite linear predictor as

$$\eta_i = \alpha + \underbrace{\mathbf{x}_i^\top \boldsymbol{\nu}}_{\text{Offset}} + \mathbf{x}_i^\top \widetilde{\boldsymbol{\beta}} \text{ where } \widetilde{\boldsymbol{\beta}} \sim \mathrm{N}_d(\mathbf{0}, g\,\boldsymbol{\Sigma}).$$

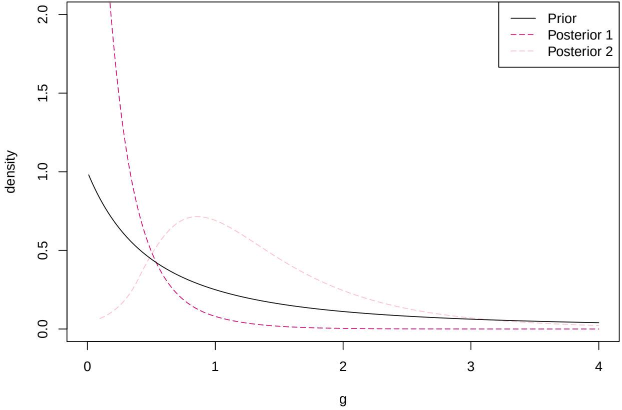
$\rightarrow$ Use `generic` Gaussian Markov random field (GMRF) with mean $\mathbf{0}$ and pre-specified precision matrix $\boldsymbol{\Sigma}^{-1}$ - up to the possibly unknown multiplicative weight factor $w = 1/g$.

- Now use the copy feature in INLA to define d identical copies of $\widetilde{\boldsymbol{\beta}}$.

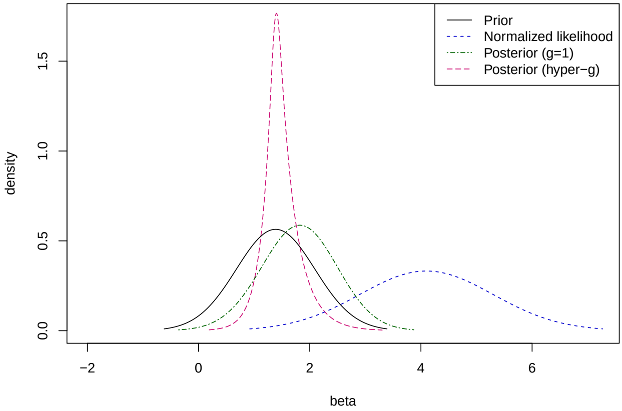
$\rightarrow$ The j-th component of the j-th copy of $\widetilde{\boldsymbol{\beta}}$ is then multiplied with the covariate values $\mathbf{x}_j = (x_{1j}, \ldots, x_{nj})^\top$ as a “weights vector”.

- Hyper-g (or any other) prior on g can be incorporated.

Hyper-g Prior and Posterior in Neonatal Death Study

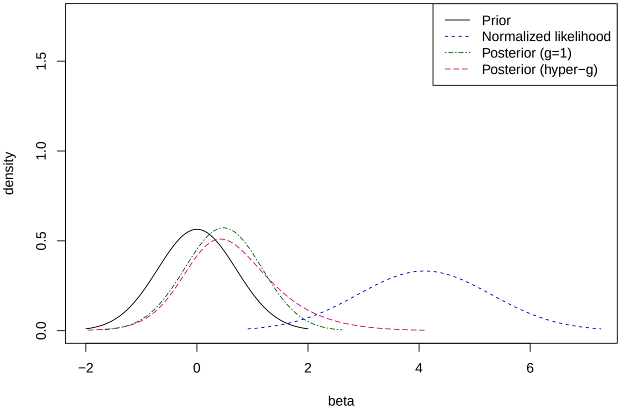


Posterior of $\beta_{\texttt{hydram}}$ (Prior I)



Method	OR Estimate	95% CI
ML	60	5.7 to 635
$g = 1$	6.1	1.6 to 22.8
Hyper-g	4.3	2.3 to 10.5

Posterior of $\beta_{\texttt{hydram}}$ (Prior II)

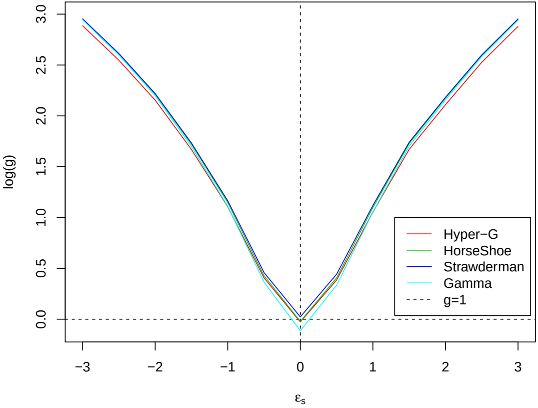


Method	OR Estimate	95% CI
ML	60	5.7 to 635
$g = 1$	1.6	0.4 to 6.3
Hyper-g	1.8	0.4 to 13.4

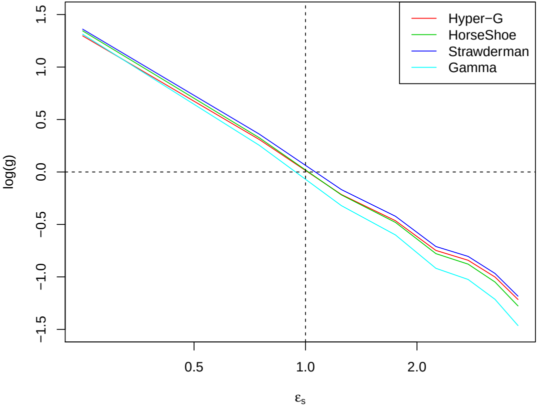
Simulation Study (preliminary results)

1. Based on the (centred) design matrix $\mathbf{X}$ of neonatal death study
2. Draw $k = 1, \ldots, 100$ times from misspecified prior
   a) $\boldsymbol{\beta}^{(k)} \sim \mathrm{N}(\boldsymbol{\nu} + \epsilon\,\mathbf{1}, \boldsymbol{\Sigma})$, $\epsilon \in \{-3, -2.5, \ldots, 0, 0.5, \ldots, 3\}$
   b) $\boldsymbol{\beta}^{(k)} \sim \mathrm{N}(\boldsymbol{\nu}, \epsilon\,\boldsymbol{\Sigma})$, $\epsilon \in \{1/4, 1/2, 1, 2, 4\}$
3. Simulate $\mathbf{y}^{(k)} \sim \mathrm{Bin}(1, \mathrm{expit}\{\mathbf{X}^\top \boldsymbol{\beta}^{(k)}\})$
4. Use `INLA` with $\boldsymbol{\beta} \sim \mathrm{N}(\boldsymbol{\nu}, g\,\boldsymbol{\Sigma})$ prior and different hyperpriors for g.

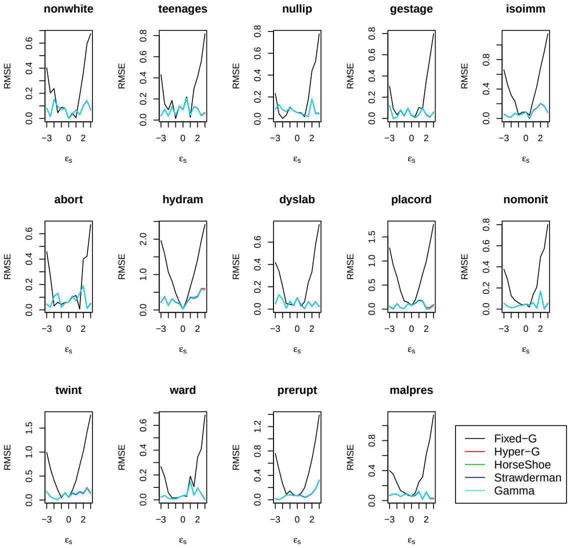
Simulation a): Mean estimate of g



Simulation b): Mean estimate of g



Simulation a): RMSE



Summary and Outlook

- Empirical Bayes is useful to downweight historical data using the power prior.
- In generalized regression, hyper-g prior regularizes EB and allows for both up- or downweighting of the prior distribution.
- Approach can be viewed as replacing a normal prior on the regression coefficients with a "robustified" scale mixture of normals prior.
- Implementation in INLA allows to extend the approach to more complex models, *e.g.* generalized linear mixed models or spatial models.
- Can also be generalized to several prior weight parameters for blocks of regression coefficients.

Literature I

Berger, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean, *Annals of Statistics* **8**(4): 716–761.

Carvalho, C. M., Polson, N. G. and Scott, J. G. (2010). The horseshoe estimator for sparse signals, *Biometrika* **97**(2): 465–480.

Copas, J. B. (1983). Regression, prediction and shrinkage, *Journal of the Royal Statistical Society. Series B (Methodological)* **45**(3): 311–354.

Copas, J. B. (1997). Using regression models for prediction: shrinkage and regression to the mean, *Statistical Methods in Medical Research* **6**(2): 167–183.

Duan, Y., Smith, E. P. and Ye, K. (2006). Using Power Priors to Improve the Binomial Test of Water Quality, *Journal of Agricultural, Biological, and Environmental Statistics* **11**(2): 151–168.

Gelman, A., Jakulin, A., Grazia, M. P. and Yu-Sung, S. (2008). A weakly informative default prior distribution for logistic and other regression models, *Annals of Applied Statisticts* **2**: 1360–1383.

Literature II

Greenland, S. (2006). Bayesian perspectives for epidemiological research: I. Foundations and basic methods, *International Journal of Epidemiology* **35**: 765–775.

Greenland, S. (2007). Bayesian perspectives for epidemiological research. II. Regression analysis, *International Journal of Epidemiology* **36**(1): 195–202.

Held, L., Sabanés Bové, D. and Gravestock, I. (2015). Approximate Bayesian model selection with the deviance statistic, *Statistical Science* **30**: 242–257.

Ibrahim, J. G. and Chen, M.-H. (2000). Power prior distributions for regression models, *Statistical Science* **15**(1): 46–60.

Liang, F., Paulo, R., Molina, G., Clyde, M. A. and Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection, *Journal of the American Statistical Association* **103**(481): 410–423.

Neuenschwander, B., Branson, M. and Spiegelhalter, D. J. (2009). A note on the power prior, *Statistics in Medicine* **28**(28): 3562–3566.

Literature III

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. and Rakow, T. (2006). *Uncertain Judgements; Eliciting Experts' Probabilities*, Wiley, Chichester.

Senn, S. (2007). *Statistical Issues in Drug Development*, 2nd edn, Wiley.

Spiegelhalter, D. J., Abrams, K. R. and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, Wiley, New York.

Sullivan, S. G. and Greenland, S. (2013). Bayesian regression in SAS software, *International Journal of Epidemiology* **42**(1): 308–317.

Viele, K., Berry, S., Neuenschwander, B., Amzal, B., Chen, F., Enas, N., Hobbs, B., Ibrahim, J. G., Kinnersley, N., Lindborg, S. et al. (2014). Use of historical control data for assessing treatment effects in clinical trials, *Pharmaceutical Statistics* **13**(1): 41–54.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions, *in* P. K. Goel and A. Zellner (eds), *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, Vol. 6 of *Studies in Bayesian Econometrics and Statistics*, North-Holland, Amsterdam, chapter 5, pp. 233–243.