University of Zurich[UZH]

Division of Biostatistics, Institute of Social and Preventive Medicine

An infinitesimal sensitivity measure for Bayesian hierarchical models

Małgorzata Roos and Leonhard Held

Second Workshop on Bayesian Inference for Latent Gaussian Models with Applications

Trondheim, 30.05.-1.06.2012

Zurich, LGM2011

Roos and Held (2011):

– New sensitivity measure S based on the Hellinger distance with respect to changes in the prior distribution

– The whole amount of information contained in the prior densities and marginal posterior densities obtained from `inla` (Rue, Martino, and Chopin 2009) is used

– Calibration with respect to the Bernoulli distribution

Hellinger Distance

Le Cam (1986)

$$H(f_0, f_1) = \sqrt{\frac{1}{2}\int_{-\infty}^{\infty}\left(\sqrt{f_0(u)} - \sqrt{f_1(u)}\right)^2 du} = \sqrt{1 - BC(f_0, f_1)},$$

where

$$BC(f_0, f_1) = \int_{-\infty}^{\infty} \sqrt{f_0(u) f_1(u)} du$$

Hellinger Distance

Bernoulli:

$$H(\mathrm{Ber}(\pi_0), \mathrm{Ber}(\pi_1)) = \sqrt{1 - \sqrt{(1-\pi_0)(1-\pi_1)} - \sqrt{\pi_0 \pi_1}}$$

Normal:

$$H\Big(\mathrm{N}(\mu_0, \sigma_0^2), \mathrm{N}(\mu_1, \sigma_1^2)\Big) = \sqrt{1 - \sqrt{\frac{2\sigma_0\sigma_1}{\sigma_0^2+\sigma_1^2}} \exp\Big(-\frac{(\mu_0-\mu_1)^2}{4(\sigma_0^2+\sigma_1^2)}\Big)}$$

Gamma:

$$H\Big(\mathrm{G}(\alpha_0, \beta_0), \mathrm{G}(\alpha_1, \beta_1)\Big) = \sqrt{1 - \Gamma\Big(\frac{\alpha_0+\alpha_1}{2}\Big)\sqrt{\frac{\beta_0^{\alpha_0}\beta_1^{\alpha_1}}{\Gamma(\alpha_0)\Gamma(\alpha_1)(\frac{\beta_0+\beta_1}{2})^{\alpha_0+\alpha_1}}}}$$

Hellinger Distance

Bernoulli:

$$H(\mathrm{Ber}(\pi_0), \mathrm{Ber}(\pi_1)) = \sqrt{1 - \sqrt{(1-\pi_0)(1-\pi_1)} - \sqrt{\pi_0 \pi_1}}$$

Normal:

$$H\Big(\mathrm{N}(\mu_0, \sigma_0^2), \mathrm{N}(\mu_1, \sigma_1^2)\Big) = \sqrt{1 - \sqrt{\frac{2\sigma_0\sigma_1}{\sigma_0^2 + \sigma_1^2}} \exp\Big(-\frac{(\mu_0 - \mu_1)^2}{4(\sigma_0^2 + \sigma_1^2)}\Big)}$$

Gamma:

$$H\Big(\mathrm{G}(\alpha_0, \beta_0), \mathrm{G}(\alpha_1, \beta_1)\Big) = \sqrt{1 - \Gamma\Big(\frac{\alpha_0 + \alpha_1}{2}\Big)\sqrt{\frac{\beta_0^{\alpha_0}\beta_1^{\alpha_1}}{\Gamma(\alpha_0)\Gamma(\alpha_1)(\frac{\beta_0+\beta_1}{2})^{\alpha_0+\alpha_1}}}}$$

Calibration of the Hellinger distance with respect to the Normal distribution

$H(\mathrm{N}(0, 1), \mathrm{N}(\mu(h), 1)) = h$

$h = h(\mu)$	$\mu(h) = \mu$	$h = h(\mu)$	$\mu(h) = \mu$
0.000	0	0.0	0.000
0.343	1	0.1	0.284
0.627	2	0.2	0.571
0.822	3	0.3	0.869
0.930	4	0.4	1.181
0.978	5	0.5	1.517
0.994	6	0.6	1.890
0.999	7	0.7	2.321
1.000	8	0.8	2.859
1.000	9	0.9	3.645

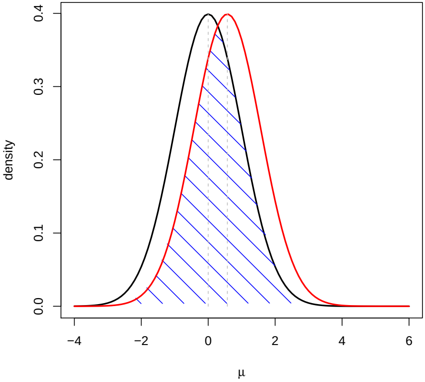
Calibration of the Hellinger distance with respect to the Normal distribution

$H(\mathrm{N}(0, 1), \mathrm{N}(\mu(h), 1)) = h$

$h = h(\mu)$	$\mu(h) = \mu$	$h = h(\mu)$	$\mu(h) = \mu$
0.000	0	0.0	0.000
0.343	1	0.1	0.284
0.627	2	0.2	0.571
0.822	3	0.3	0.869
0.930	4	0.4	1.181
0.978	5	0.5	1.517
0.994	6	0.6	1.890
0.999	7	0.7	2.321
1.000	8	0.8	2.859
1.000	9	0.9	3.645

Example Calibration

Densities: N(0,1) vs N(0.571,1) with $h = 0.02$



Hellinger distance for marginal posterior distributions

– INLA provides approximations of the marginal posterior distributions of the latent Gaussian field.

– `inla.hyperpar()`

– Function `inla.smarginal()` allows the evaluation and extrapolation of the posteriors at a fixed grid of values.

⇒ Numerical integration

$$H(post(\boldsymbol{\theta}_0), post(\boldsymbol{\theta})) \approx \sqrt{1 - \sum_k \sqrt{post(\boldsymbol{\theta}_0)(k) post(\boldsymbol{\theta})(k)} \Delta_k}$$

Sensitivity measure and its calibration

- Sensitivity measure:

  Default $\boldsymbol{\theta}_0$ and a shifted $\boldsymbol{\theta}$ values of the prior

  $pri(\boldsymbol{\theta})$: the prior (no-data) density

  $i(\boldsymbol{\theta}) = H(pri(\boldsymbol{\theta}_0), pri(\boldsymbol{\theta}))$

  $post(\boldsymbol{\theta})$: the corresponding posterior density

  $d(\boldsymbol{\theta}) = H(post(\boldsymbol{\theta}_0), post(\boldsymbol{\theta}))$

$$\boxed{S(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \frac{d(\boldsymbol{\theta})}{i(\boldsymbol{\theta})}}$$

- Calibration:

$$C(S, 0.5) = \mu(S \times h(0.5))$$

Infinitesimal Sensitivity

$$\boxed{S^*(\boldsymbol{\theta}_0) = \lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0} S(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0} \frac{d(\boldsymbol{\theta})}{i(\boldsymbol{\theta})}}$$

McCulloch (1989)

Let $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \boldsymbol{\epsilon}$ and apply the Taylor expansion to obtain

$$d(\boldsymbol{\theta}) \approx \boldsymbol{\epsilon}^T D^2 d(\boldsymbol{\theta}_0)\boldsymbol{\epsilon}/2, \qquad i(\boldsymbol{\theta}) \approx \boldsymbol{\epsilon}^T D^2 i(\boldsymbol{\theta}_0)\boldsymbol{\epsilon}/2$$

and

$$S(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0 + \boldsymbol{\epsilon}) \approx \boldsymbol{\epsilon}^T D^2 d(\boldsymbol{\theta}_0)\boldsymbol{\epsilon}/\boldsymbol{\epsilon}^T D^2 i(\boldsymbol{\theta}_0)\boldsymbol{\epsilon}$$

Infinitesimal Sensitivity

McCulloch (1989) (continued)

Let λ^* be the dominant eigenvalue and $\boldsymbol{\epsilon}^*$ the corresponding eigenvector of

$$(D^2 i(\boldsymbol{\theta}_0))^{-1} D^2 d(\boldsymbol{\theta}_0)$$

Take (Srivastava and Carter 1983)

$$\boxed{S^*(\boldsymbol{\theta}_0) \approx \lambda^*}$$

$\boldsymbol{\epsilon}^*$ indicates which linear combination of the elements of $\boldsymbol{\theta}_0$ is most influential.

$D^2 d(\boldsymbol{\theta}_0)$ and $D^2 i(\boldsymbol{\theta}_0)$ can be approximated numerically.

A Conjugate Example (Box, 1980)

$Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} \mathrm{N}(m, \kappa^{-1})$, where κ is fixed and $m \sim \mathrm{N}(\mu, \lambda^{-1})$ then

$$m \mid \mathbf{y} \sim \mathrm{N}\left(\frac{n\kappa\bar{y} + \lambda\mu}{n\kappa + \lambda}, (n\kappa + \lambda)^{-1}\right)$$

Assume $n = 4$, $\kappa = 1$ and $\bar{y} = 76$

Prior of interest: $\boldsymbol{\theta}_0 = (\mu_0, \lambda_0) = (70, 0.5)$

$S^*(\boldsymbol{\theta}_0) = \lambda^* = 1.79$ and $C(S^*(\boldsymbol{\theta}_0), 0.5) = 0.91$ are mainly caused by λ_0

as $\epsilon^* = (0.113, -0.994)$

Output from `inla`

```
Fixed effects:
const
mean    sd 0.025quant 0.5quant 0.975quant kld
75.33 0.47    74.40    75.33      76.26    0
```

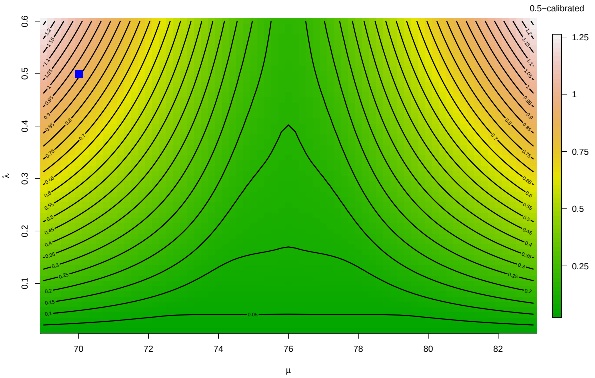
A possible output from `inla`

```
Fixed effects:

const

mean    sd 0.025quant 0.5quant 0.975quant kld sens c.sens

75.33 0.47    74.40    75.33      76.26    0  1.79  0.91

ev.mu ev.lambda

0.113 -0.994
```

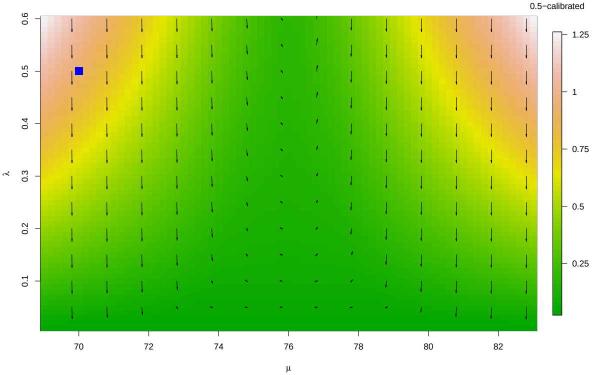
Prior used: $\boldsymbol{\theta}_0 = (\mu_0, \lambda_0) = (70, 0.5)$

Computational time: 3s

0.5-Calibrated Sensitivity Landscape from `inla`



0.5-Calibrated Sensitivity Landscape from `inla` with Eigenvectors



Prior-data conflict

– Box (1980):

Does the prior place most of its mass on parameter values that are not feasible in light of the data?

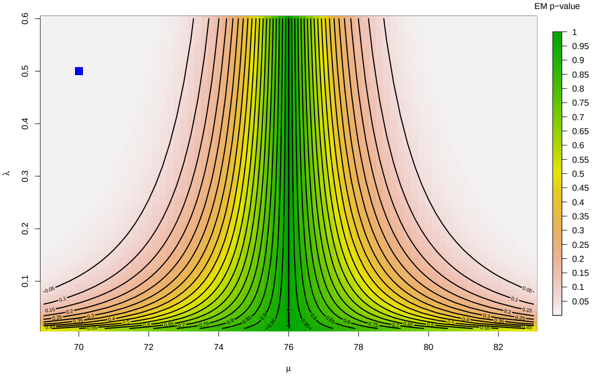
$$\text{p-value} = \mathrm{P}[p(\mathbf{x}) \leq p(\mathbf{x}_{obs})],$$

where $p(\mathbf{x})$ is the prior predictive distribution

– Evans and Moshonov (2006) restricted the Box's approach to minimal sufficient statistics.

Prior-data conflict: Landscape

Box (1980): p = 3.2e-05 and Evans and Moshonov (2006): p = 6.3e-05



Non-conjugate example: Disease Mapping

Fong, Rue, and Wakefield (2010): Lip Cancer in Scotland (1975-1980)

$$Y_j \mid \mu_j \overset{ind}{\sim} \mathrm{Po}(\mu_j),\ j = 1, \ldots, J$$

$$\begin{aligned} \log \mu_j &= \log e_j + c + v_j & (1) \\ \log \mu_j &= \log e_j + c \qquad + u_j & (2) \\ \log \mu_j &= \log e_j + c + v_j + u_j & (3) \end{aligned}$$

const: $c \sim \mathrm{N}(\mu, \lambda^{-1})$

iid: $v_j \overset{iid}{\sim} \mathrm{N}(0, \kappa_v^{-1})$ with $\kappa_v \sim \mathrm{G}(\alpha, \beta)$

ICAR: $u_j \sim \mathrm{ICAR}(\kappa_u^{-1})$ with $\kappa_u \sim \mathrm{G}(\gamma, \delta)$ and `constr = TRUE`

Prior of interest:

$$\boldsymbol{\theta}_0 = (\mu_0, \lambda_0, \alpha_0, \beta_0, \gamma_0, \delta_0) = (\overbrace{0.1, 0.01}^{c}, \overbrace{1, 0.14}^{v}, \overbrace{1, 0.2/0.59}^{u})$$

Computation

– $i(\boldsymbol{\theta}_0)$, $d(\boldsymbol{\theta}_0)$:

Disjoint Support Decomposition (DSD)

$$BC(f_0 g_0 h_0, f_1 g_1 h_1) \stackrel{DSD}{=} BC(f_0, f_1) BC(g_0, g_1) BC(h_0, h_1)$$

– $d(\boldsymbol{\theta}_0)$: number of `inla` calls needed

$d = 4$ for "const + iid" and "const + ICAR" $\Rightarrow$ n.calls $= 33$

$d = 6$ for "const + iid + ICAR" $\Rightarrow$ n.calls $= 73$

In general n.calls $= 1 + 2d^2$

Results

Prior of interest:

$$\boldsymbol{\theta}_0 = (\mu_0, \lambda_0, \alpha_0, \beta_0, \gamma_0, \delta_0) = (\overbrace{0.1, 0.01}^{c}, \overbrace{1, 0.14}^{v}, \overbrace{1, 0.2/0.59}^{u})$$

iid	ICAR	$S^*(\theta_0) = \lambda^*$	ϵ^*						$C(S^*, 0.5)$
			$(\mu_0,$	$\lambda_0,$	$\alpha_0,$	$\beta_0,$	$\gamma_0,$	$\delta_0)$	
yes 31s	no	0.199	“0”	“0”	“1”	“0”			0.099
no 65s	yes	0.240	“0”	“0”			“-1”	“0”	0.119
yes 180s	yes	1.304	“0”	“0”	“0”	“-1”	“0”	“0”	0.656

Results (continued)

Sensitivities for each parameter can be obtained by a shortcut:

Look at

$$(D^2 i(\boldsymbol{\theta}_0))^{-1} D^2 d(\boldsymbol{\theta}_0),$$

where $D^2 i$ and $D^2 d$ are diagonal matrices i.e. no mixed second partial derivatives have to be computed!

iid	ICAR	$(\mu_0,$	$\lambda_0,$	$\alpha_0,$	$\beta_0,$	$\gamma_0,$	$\delta_0)$
yes	no	0.012	1e-04	0.198	0.065		
no	yes	0.005	5e-05			0.230	0.140
yes	yes	0.619	2e-04	0.403	1.284	0.274	0.237

Conclusions

- Calibration with respect to the normal distribution: $h \in (0, 1)$
- Relationship to prior-data conflict measures
- Perfect approximation of the analytical posterior distribution provided by `inla`
- An improvement of a few computational issues is necessary
- Approximate computation of infinitesimal sensitivity with `inla` is possible
- Non-conjugate models can be tackled

Thank you for your attention!

Box, G. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society, Series A. 143*(4), 383–430.

Evans, M. and H. Moshonov (2006). Checking for prior-data conflict. *Bayesian Analysis 4*(1), 893–914.

Fong, Y., H. Rue, and J. Wakefield (2010). Bayesian inference for generalized linear mixed models. *Biostatistics 11*(3), 397–412.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag.

McCulloch, R. (1989). Local model influence. *Journal of the American Statistical Association 84*(406), 473–478.

Roos, M. and L. Held (2011). Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Analysis 6*(2), 259–278.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B. 71*(2), 319–392.

Srivastava, M. and E. Carter (1983). *An Introduction to Applied Multivariate Statistics.* North–Holland.