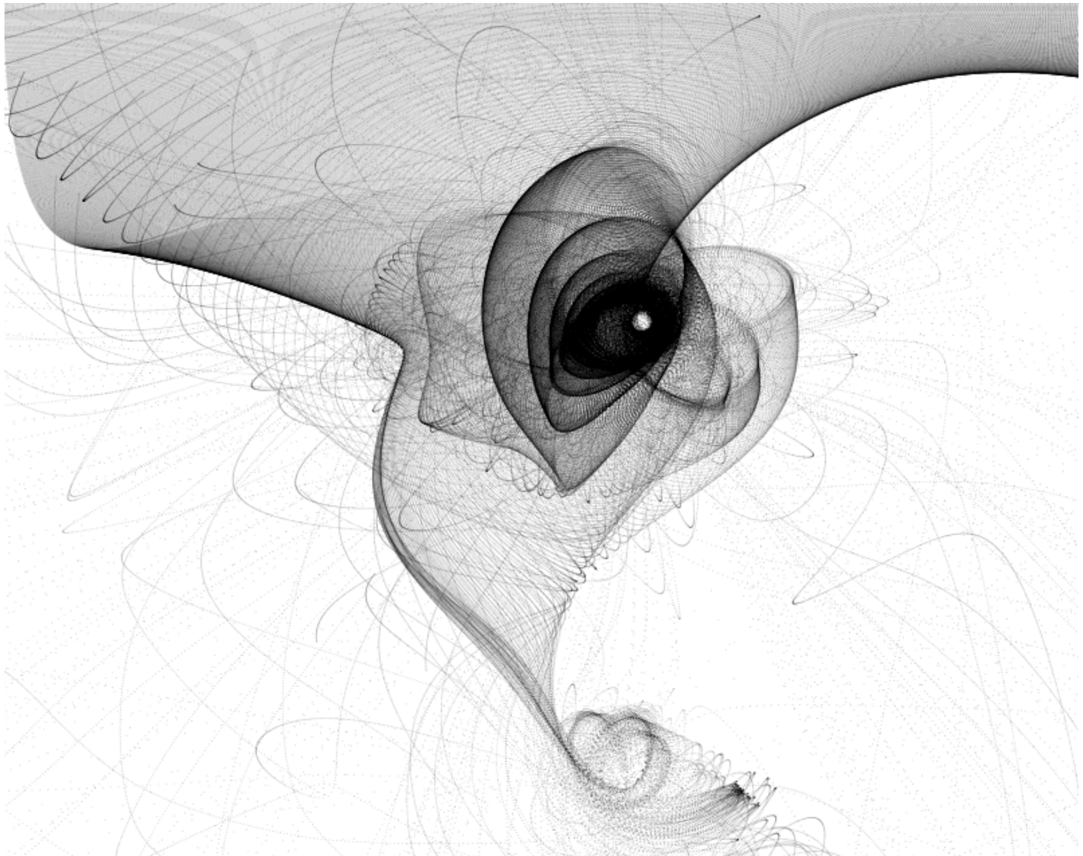
Calcul de la vraisemblance marginale: rappels et méthode des posteriors de puissance de température



Jean-Louis Foulley
INRA/Dept de Génétique Animale/GABI/PSGen
78350 Jouy-en-Josas, France (jean-louis.foulley@jouy.inra.fr)

Sommaire

- Objectifs
- Rappels des méthodes de calcul de base
  - Monte Carlo direct
  - Newton-Raphtery, Gelfand & Dey, Chib
  - Bridge sampling
- Méthode PP de Friel & Pettitt
- Lien avec l'approche du BF fractionnaire d'O Hagan
- Algorithme
- Exemples
- Conclusion

Objectifs

Vraisemblance marginale

$$m(\mathbf{y}) = \int_{\Theta} f(\mathbf{y} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \, d\boldsymbol{\theta}$$

-Constante de normalisation de $\pi^*(\boldsymbol{\theta} \mid \mathbf{y})$

$$\pi(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{\pi^*(\boldsymbol{\theta})}{m(\mathbf{y})} \text{ où } \pi^*(\boldsymbol{\theta} \mid \mathbf{y}) = f(\mathbf{y} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta})$$

-Composante du facteur de Bayes

$$BF_{12} = \frac{\pi(\mathrm{M}_1 \mid \mathbf{y}) / \pi(\mathrm{M}_2 \mid \mathbf{y})}{\pi(\mathrm{M}_1) / \pi(\mathrm{M}_2)} = \frac{m_1(\mathbf{y})}{m_2(\mathbf{y})}$$

$$\Delta D_{m,12} = -2 \ln BF_{12} = D_{m,1} - D_{m,2}$$

$D_{m,j} = -2 \ln m_j(\mathbf{y})$: Deviance marginale

Calibrage: Jeffreys & Turing (Deciban: $10\log_{10}$BF)

Méthodes de calcul/rappels Guihenneuc (2009)

1) Monte Carlo direct $\hat{m}_{MC}(\mathbf{y}) = \frac{1}{G}\sum_{g=1}^{G} f\left(\mathbf{y} \mid \boldsymbol{\theta}^{(g)}\right)$

$\boldsymbol{\theta}^{(1)},\ldots,\boldsymbol{\theta}^{(g)}$: tirages dans $\pi(\boldsymbol{\theta})$

Converge en théorie mais très très peu efficace

2)Moyenne harmonique (Newton & Raftery, 1994)

$$\hat{m}_{NR}(\mathbf{y}) = \left[\frac{1}{G}\sum_{g=1}^{G} \frac{1}{f\left(\mathbf{y} \mid \boldsymbol{\theta}^{(g)}\right)}\right]^{-1}$$

$\boldsymbol{\theta}^{(1)},\ldots,\boldsymbol{\theta}^{(g)}$: tirages dans $\pi(\boldsymbol{\theta} \mid \mathbf{y})$

Très instable (variance infinie):à proscrire

Méthodes de calcul/rappels Guihenneuc (2009)

3) Moyenne harmonique généralisée (Gelfand & Dey, 1994)

$$\hat{m}_{GD}(\mathbf{y}) = \left[\frac{1}{G} \sum_{g=1}^{G} \frac{g\left(\boldsymbol{\theta}^{(g)}\right)}{f\left(\mathbf{y} \mid \boldsymbol{\theta}^{(g)}\right) \pi\left(\boldsymbol{\theta}^{(g)}\right)} \right]^{-1}$$

$\boldsymbol{\theta}^{(1)},\ldots,\boldsymbol{\theta}^{(g)}$: tirages dans $\pi(\boldsymbol{\theta} \mid \mathbf{y})$

g(.) à calibrer comme approx du posterior: pb en grde dimension

4)Méthode de Siddharta Chib (1995)

$$\ln m(\mathbf{y}) = \ln f(\mathbf{y} \mid \boldsymbol{\theta}) + \ln \pi(\boldsymbol{\theta}) - \ln \pi(\boldsymbol{\theta} \mid \mathbf{y}), \forall \boldsymbol{\theta}$$

$$\ln \hat{m}_{SC}(\mathbf{y}) = \ln f\left(\mathbf{y} \mid \boldsymbol{\theta}^{*}\right) + \ln \pi\left(\boldsymbol{\theta}^{*}\right) - \ln \hat{\pi}\left(\boldsymbol{\theta}^{*} \mid \mathbf{y}\right)$$

Estimation $\hat{\pi}(\boldsymbol{\theta} \mid \mathbf{y})$ et sélection $\boldsymbol{\theta}^{*} = ML, MAP, E(\boldsymbol{\theta} \mid \mathbf{y})$

Simple et souvent performant

Méthodes de calcul/Chib(suite)

4)Méthode de Chib (1995)

$$\ln \hat{m}_{SC}\left(\mathbf{y}\right) = \ln f\left(\mathbf{y} \mid \boldsymbol{\theta}^{*}\right) + \ln \pi\left(\boldsymbol{\theta}^{*}\right) - \ln \hat{\pi}\left(\boldsymbol{\theta}^{*} \mid \mathbf{y}\right)$$

1) Gibbs et RaoBlackwellization (Chib,1995)

2) Metropolis-Hastings (Chib & Jeliazkov,2001)

3) Estimateur à noyau (Ando, 2006)

Méthodes de calcul/Chib via Gibbs

Si $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$

$$\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \mid \mathbf{y}) = \underbrace{\pi(\boldsymbol{\theta}_1 \mid \mathbf{y}, \boldsymbol{\theta}_2)}_{connu} \underbrace{\pi(\boldsymbol{\theta}_2 \mid \mathbf{y})}_{estimé}$$

$$\pi(\boldsymbol{\theta}_2 \mid \mathbf{y}) = \int \underbrace{\pi(\boldsymbol{\theta}_2 \mid \mathbf{y}, \boldsymbol{\theta}_1)}_{connu} \underbrace{\pi(\boldsymbol{\theta}_1 \mid \mathbf{y})}_{tirages\ \text{MCMC}} d\boldsymbol{\theta}_1$$

"Calcul par Rao-Blackwellization"

$$\hat{\pi}(\boldsymbol{\theta}_2^* \mid \mathbf{y}) = \frac{1}{G} \sum_{g=1}^{G} \pi\left(\boldsymbol{\theta}_2^* \mid \mathbf{y}, \boldsymbol{\theta}_1^{(g)}\right)$$

$\boldsymbol{\theta}_1^{(g)}$: tirage dans $\pi(\boldsymbol{\theta}_1 \mid \mathbf{y})$

Méthodes de calcul / Bridge sampling

5)Bridge sampling (Meng & Wong, 1996)

$$m(\mathbf{y})=\frac{\int \alpha(\boldsymbol{\theta}) f(\mathbf{y}\mid\boldsymbol{\theta})\pi(\boldsymbol{\theta}) g(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int \alpha(\boldsymbol{\theta}) g(\boldsymbol{\theta})\pi(\boldsymbol{\theta}\mid\mathbf{y})d\boldsymbol{\theta}}=\frac{E^{g(\boldsymbol{\theta})}\left(\alpha(\boldsymbol{\theta}) f(\mathbf{y}\mid\boldsymbol{\theta})\pi(\boldsymbol{\theta})\right)}{E^{\pi(\boldsymbol{\theta}\mid\mathbf{y})}\left(\alpha(\boldsymbol{\theta}) g(\boldsymbol{\theta})\right)}$$

$\alpha(\boldsymbol{\theta})$"bridge function" tq 0<den<+∞, $g(\boldsymbol{\theta})$ = densité à calibrer

Pour $\alpha(\boldsymbol{\theta})=1/g(\boldsymbol{\theta})$

$$\hat{m}_{BS1}(\mathbf{y})=L^{-1}\sum_{l=1}^{L}\left[f\left(\mathbf{y}\mid\boldsymbol{\theta}^{(l)}\right)\pi\left(\boldsymbol{\theta}^{(l)}\right)/g\left(\boldsymbol{\theta}^{(l)}\right)\right]^{-1}(IS)$$

Pour $\alpha(\boldsymbol{\theta})=1/f(\mathbf{y}\mid\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ $\hat{m}_{BS2}(\mathbf{y})$ = Gelfand-Dey (1994)

Pour $\alpha(\boldsymbol{\theta})=1/\left[f(\mathbf{y}\mid\boldsymbol{\theta})\pi(\boldsymbol{\theta})g(\boldsymbol{\theta})\right]^{1/2}$ $\hat{m}_{BS3}(\mathbf{y})$ = Lopes-West (2004)

$$\hat{m}_{BS3}(\mathbf{y})=\frac{L^{-1}\sum_{l=1}^{L}\left[f\left(\mathbf{y}\mid\boldsymbol{\theta}^{(l)}\right)\pi\left(\boldsymbol{\theta}^{(l)}\right)/g\left(\boldsymbol{\theta}^{(l)}\right)\right]^{1/2}}{M^{-1}\sum_{m=1}^{M}\left[g\left(\boldsymbol{\theta}^{(m)}\right)/f\left(\mathbf{y}\mid\boldsymbol{\theta}^{(m)}\right)\pi\left(\boldsymbol{\theta}^{(m)}\right)\right]^{1/2}}$$

$\boldsymbol{\theta}^{(l)}$: tirages dans $g(\boldsymbol{\theta})$; $\boldsymbol{\theta}^{(m)}$: tirages dans $\pi(\boldsymbol{\theta}\mid\mathbf{y})$

Méthodes de calcul / Bridge sampling (suite)

5)Bridge sampling (Meng & Wong, 1996)

$$m(\mathbf{y})=\frac{\int \alpha(\boldsymbol{\theta}) f(\mathbf{y}\mid\boldsymbol{\theta})\pi(\boldsymbol{\theta}) g(\boldsymbol{\theta})\, d\boldsymbol{\theta}}{\int \alpha(\boldsymbol{\theta}) g(\boldsymbol{\theta})\pi(\boldsymbol{\theta}\mid\mathbf{y})\, d\boldsymbol{\theta}}=\frac{E^{g(\boldsymbol{\theta})}\left(\alpha(\boldsymbol{\theta}) f(\mathbf{y}\mid\boldsymbol{\theta})\pi(\boldsymbol{\theta})\right)}{E^{\pi(\boldsymbol{\theta}\mid\mathbf{y})}\left(\alpha(\boldsymbol{\theta}) g(\boldsymbol{\theta})\right)}$$

Pour $\alpha(\boldsymbol{\theta})=1/f(\mathbf{y}\mid\boldsymbol{\theta})\pi(\boldsymbol{\theta})g(\boldsymbol{\theta})$

$$\hat{m}_{BS4}(\mathbf{y})=\frac{L^{-1}\sum_{l=1}^{L}\left[1/g\left(\boldsymbol{\theta}^{(l)}\right)\right]}{M^{-1}\sum_{m=1}^{M}\left[1/f\left(\mathbf{y}\mid\boldsymbol{\theta}^{(m)}\right)\pi\left(\boldsymbol{\theta}^{(m)}\right)\right]^{1/2}}$$ (Lopes & West, 2004; Ando, 2010)

$\boldsymbol{\theta}^{(l)}$: tirages dans $g(\boldsymbol{\theta})$; $\boldsymbol{\theta}^{(m)}$: tirages dans $\pi(\boldsymbol{\theta}\mid\mathbf{y})$ Est. problématique (cf numérateur)

Pour $\alpha(\boldsymbol{\theta})\propto\left[s_M\pi(\boldsymbol{\theta}\mid\mathbf{y})+s_L g(\boldsymbol{\theta})\right]^{-1}$, **estimateur optimum** au sens E(RMSE)

(Meng & Wong, 1996; Lopes & West, 2004; Fruhwirth-Schnatter,2004)

$$\hat{m}_{BS5}^{(t+1)}(\mathbf{y})=\hat{m}_{BS5}^{(t)}\frac{L^{-1}\sum_{l=1}^{L}\dfrac{\hat{\pi}_t\left(\boldsymbol{\theta}^{(l)}\mid\mathbf{y}\right)}{s_M\hat{\pi}_t\left(\boldsymbol{\theta}^{(l)}\mid\mathbf{y}\right)+s_L g\left(\boldsymbol{\theta}^{(l)}\right)}}{M^{-1}\sum_{m=1}^{M}\dfrac{g\left(\boldsymbol{\theta}^{(m)}\right)}{s_M\hat{\pi}_t\left(\boldsymbol{\theta}^{(m)}\mid\mathbf{y}\right)+s_L g\left(\boldsymbol{\theta}^{(m)}\right)}}$$

où $\hat{\pi}_t(\boldsymbol{\theta}\mid\mathbf{y})=f(\mathbf{y}\mid\boldsymbol{\theta})\pi(\boldsymbol{\theta})/\hat{m}_{BS5}^{(t)}$ et $\hat{m}_{BS5}^{(0)}=\hat{m}_{BS1}$ ou $\hat{m}_{BS2}$

$s_M=1-s_L=M/(M+L)$

Méthode des posteriors de puissance/définition de base

Methode due à Friel and Petit (2008)

Posterior de puissance défini par

$$\pi(\boldsymbol{\theta} \mid \mathbf{y}, t) = \frac{f(\mathbf{y} \mid \boldsymbol{\theta})^t \, \pi(\boldsymbol{\theta})}{z_t(\mathbf{y})}$$

$$z_t(\mathbf{y}) = \int f(\mathbf{y} \mid \boldsymbol{\theta})^t \, \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

où $t \in \left]0,1\right[$ est une variable auxiliaire dite de température

A noter $z_0(\mathbf{y}) = 1$ and $z_1(\mathbf{y}) = m(\mathbf{y})$ densité marginale

Méthode PP/résultat fondamental

$$\log m(\mathbf{y}) = \int_0^1 E_{\boldsymbol{\theta}|\mathbf{y},t}\left[\log f(\mathbf{y}|\boldsymbol{\theta})\right]dt$$

Posterior de puissance:

$$\pi(\boldsymbol{\theta}|\mathbf{y},t) = \frac{f(\mathbf{y}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta})}{z_t(\mathbf{y})}$$

Intégration thermodynamique (fin 70)

Méthode d'Ogata (1989)

Ripley (1988), Neal (1993)

"Path sampling" (Gelman & Meng, 1998)

Méthode PP/démonstration

Démonstration de

$$\log m(\mathbf{y}) = \int_0^1 E_{\boldsymbol{\theta}|\mathbf{y},t}\left[\log f(\mathbf{y}\mid\boldsymbol{\theta})\right]dt$$

Comme $z_0(\mathbf{y}) = 1$ and $z_1(\mathbf{y}) = m(\mathbf{y})$ densité marginale

$$\log m(\mathbf{y}) = \log \frac{z_1(\mathbf{y})}{z_0(\mathbf{y})} = \log \int_0^1 \left(\log z_t(\mathbf{y})\right)' dt$$

$$\left(\log z_t(\mathbf{y})\right)' = \left(z_t(\mathbf{y})\right)' / z_t(\mathbf{y})$$

$$\left(z_t(\mathbf{y})\right)' = \frac{d}{dt}\int \left[f(\mathbf{y}\mid\boldsymbol{\theta})\right]^t \pi(\boldsymbol{\theta})\,d\boldsymbol{\theta} = \int \frac{d}{dt}\int \left[f(\mathbf{y}\mid\boldsymbol{\theta})\right]^t \pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}$$

$$\frac{d}{dt}\left[f(\mathbf{y}\mid\boldsymbol{\theta})\right]^t = \left[f(\mathbf{y}\mid\boldsymbol{\theta})\right]^t \frac{d\log\left[f(\mathbf{y}\mid\boldsymbol{\theta})\right]^t}{dt} = \left[f(\mathbf{y}\mid\boldsymbol{\theta})\right]^t \log\left[f(\mathbf{y}\mid\boldsymbol{\theta})\right]$$

$$\left(\log z_t(\mathbf{y})\right)' = \int \log\left[f(\mathbf{y}\mid\boldsymbol{\theta})\right] \underbrace{\frac{\left[f(\mathbf{y}\mid\boldsymbol{\theta})\right]^t \pi(\boldsymbol{\theta})}{z_t(\mathbf{y})}}_{\pi(\boldsymbol{\theta}|\mathbf{y},t)} d\boldsymbol{\theta}$$

Méthode PP/Exemple

$$y_i \mid \theta \sim_{iid} N(\theta,1),\ i=1,..,N$$

$$\theta \sim N(\mu,\tau^2)$$

Alors $\theta \mid \mathbf{y},t \sim N(\mu_t,\tau_t^2)$

$$\mu_t = \frac{Nt\bar{y}+\mu\tau^{-2}}{Nt+\tau^{-2}};\ \tau_t^2 = \frac{1}{Nt+\tau^{-2}}$$

$$\underbrace{-2E_{\theta|\mathbf{y},t}\left[\log f(\mathbf{y}\mid\theta)\right]}_{\bar{D}_t(\theta)} =$$

$$N\left[\log 2\pi + \log s^2 + \frac{(\mu-\bar{y})^2}{(\mu\tau^2 t+1)^2} + \frac{1}{Nt+\tau^{-2}}\right]$$

$$\bar{y} = N^{-1}\sum_{i=1}^{N} y_i\ ; s^2 = N^{-1}\sum_{i=1}^{N}(y_i-\bar{y})^2$$

$$\bar{D}_0(\theta) = N\left[Cte + (\mu-\bar{y})^2\right] + \boxed{N\tau^2}$$

Grande sensibilité à τ^2 $(\tau^2 \to \infty, \bar{D}_0(\theta) \to \infty)$

Méthode PP/Exemple/suite

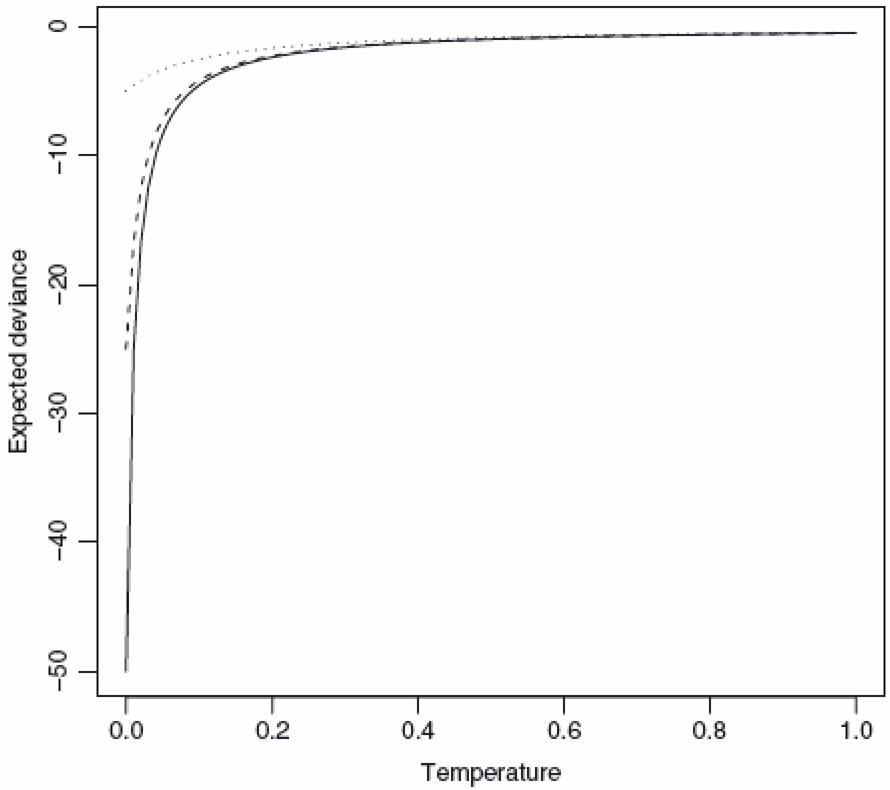


Fig. 1. Expected (half) deviance (6), under the distribution $(\theta|\mathbf{y}, t)$, plotted against t for prior variance equal to 1 (·······), 5 (-------) and 10 (———): as v increases so also does the rate at which the mean deviance changes with t

Méthode PP/Distance KL Prior-Posterior

$$KL\left(\pi(\boldsymbol{\theta}\mid\mathbf{y}),\pi(\boldsymbol{\theta})\right)=\int \ln\frac{\pi(\boldsymbol{\theta}\mid\mathbf{y})}{\pi(\boldsymbol{\theta})}\pi(\boldsymbol{\theta}\mid\mathbf{y})\,d\boldsymbol{\theta}$$

$$KL=\int \ln\frac{f(\mathbf{y}\mid\boldsymbol{\theta})\cancel{\pi}(\boldsymbol{\theta})}{m(\mathbf{y})\cancel{\pi}(\boldsymbol{\theta})}\pi(\boldsymbol{\theta}\mid\mathbf{y})\,d\boldsymbol{\theta}$$

$$KL=E_{\boldsymbol{\theta}\mid\mathbf{y}}\left[\ln f(\mathbf{y}\mid\boldsymbol{\theta})\right]-\ln m(\mathbf{y})$$

$$-2KL=\bar{D}-D_m \quad \text{(sous produit immédiat de PP)}$$

$$\boxed{D_m=\bar{D}+2KL} \quad Rappel: DIC=\bar{D}+p_D$$

Méthode PP/BF partiel

1) prior $\pi(\theta)$ impropre $\Rightarrow$ marginale $f(x)$ impropre

d'où des pb dans la définition du BF

2) grde sensibilité du BF au prior (ne disparait pas avec avec l'augmentation de la taille de l'échantillon)

Idée du BF "partiel" (Lempers,1971) $\mathbf{y} = (\mathbf{y}_P, \mathbf{y}_T)$

-échantillon d'apprentissage $\mathbf{y}_P$ ("pilote") pour étalonner le prior

-échantillon "test" $\mathbf{y}_T$ pour l'analyse des données

BF intrinsèque (Berger & Perrichi, 1996)

BF fractionnaire (O'Hagan, 1995)

Méthode PP/BF fractionnaire

Prior $\pi(\boldsymbol{\theta})$ impropre $\Rightarrow$ marginale $f(\mathbf{y})$ impropre

d'où des pb dans la définition du BF

$f(\mathbf{y}_P \mid \boldsymbol{\theta}) \approx f(\mathbf{y} \mid \boldsymbol{\theta})^b \quad b = m/N < 1$ (O'Hagan, 1995)

une fraction b de la vraisemblance sert à calibrer le prior

$$\pi(\boldsymbol{\theta}, b) \propto f(\mathbf{y} \mid \boldsymbol{\theta})^b \, \pi(\boldsymbol{\theta})$$

Méthode PP/BF fractionnaire/suite

$$\pi(\boldsymbol{\theta}, b) \propto f(\mathbf{y} \mid \boldsymbol{\theta})^b \, \pi(\boldsymbol{\theta})$$

$$m^F(\mathbf{y}, b) = \frac{\int f(\mathbf{y} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \, d\boldsymbol{\theta}}{\int f(\mathbf{y} \mid \boldsymbol{\theta})^b \, \pi(\boldsymbol{\theta}) \, d\boldsymbol{\theta}} = \frac{m(\mathbf{y}, 1)}{m(\mathbf{y}, b)}$$

Méthode PP fournit directement

$-\pi(\boldsymbol{\theta}, b) \, via \; \pi(\boldsymbol{\theta} \mid \mathbf{y}, t = b)$

$-\boxed{\log m^F(\mathbf{y}, b) = \int_{\boxed{b}}^{1} E_{\boldsymbol{\theta}|\mathbf{y},t} \left[\log f(\mathbf{y} \mid \boldsymbol{\theta}) \right] dt}$

Méthode PP/algorithme MCMC

Approche MCMC avec discrétisation de t sur $[0,1[$

$t_0 = 0 < t_1 < \ldots < t_i < \ldots < t_{n-1} < t_n = 1$

$t_i = (i/n)^c$ avec $i = 1,..,n;\ n = 20-100;\ c = 2-5$

1)Faire des tirages $\boldsymbol{\theta}^{(g_i)}$ MCMC dans $\pi\left(\boldsymbol{\theta} \mid \mathbf{y}, t_i\right)$

2)Estimer $\hat{E}_{\boldsymbol{\theta}|\mathbf{y},t=t_i}\left[\log p\left(\mathbf{y} \mid \boldsymbol{\theta}\right)\right] = \frac{1}{G}\sum_{g_i=1}^{G} \log p\left(\mathbf{y} \mid \boldsymbol{\theta}^{(g_i)}\right)$

Souvent ind. conditionnelle, $\log p\left(\mathbf{y} \mid \boldsymbol{\theta}\right) = \sum_{i=1}^{N} \log p\left(y_i \mid \boldsymbol{\theta}\right)$

Par ex si $\boldsymbol{\theta}$ est le plus proche parent stochastique de $\mathbf{y} = \left(y_i\right)$

3)Calcul de l'intégrale (ex méthode trapézoidale)

$\hat{\log}\, m\left(y\right) = ½\sum_{i=0}^{n} \left(t_{i+1} - t_i\right)\left(E_i + E_{i+1}\right)$

Algorithme d'échantillonnage dans θ *et t*

$$\log m(\mathbf{y}) = \int_0^1 \left[\log f(\mathbf{y} \mid \boldsymbol{\theta})\right] \pi(\boldsymbol{\theta} \mid \mathbf{y}, t)\, dt$$

$$\log m(\mathbf{y}) = \int_0^1 \frac{\log f(\mathbf{y} \mid \boldsymbol{\theta})}{p(t)} \underbrace{\pi(\boldsymbol{\theta} \mid \mathbf{y}, t)\, p(t)}_{\pi(\boldsymbol{\theta}, t \mid \mathbf{y})}\, dt$$

$$\log m(\mathbf{y}) = E_{\boldsymbol{\theta}, t \mid \mathbf{y}} \left[\frac{\log f(\mathbf{y} \mid \boldsymbol{\theta})}{p(t)} \right]$$

Supposant $p(t) \propto z_t(y) \Rightarrow \pi(t \mid \boldsymbol{\theta}, \mathbf{y}) \propto f(\mathbf{y} \mid \boldsymbol{\theta})^t$

Echantillonnant $(\boldsymbol{\theta}, t)$ dans ces conditions donne de mauvaises estimations (pas assez de tirages proches de 0)

Example/ Pothoff&Roy's data

Growth measurements in 11 girls and 16 boys: Pothoff and Roy,1964; Little and Rubin, 1987

	Age (years)					Age (years)			
Girl	8	10	12	14	Boy	8	10	12	14
1	210	200	215	230	1	260	250	290	310
2	210	215	240	255	2	215		230	265
3	205		245	260	3	230	225	240	275
4	235	245	250	265	4	255	275	265	270
5	215	230	225	235	5	200		225	260
6	200		210	225	6	245	255	270	285
7	215	225	230	250	7	220	220	245	265
8	230	230	235	240	8	240	215	245	255
9	200		220	215	9	230	205	310	260
10	165		190	195	10	275	280	310	315
11	245	250	280	280	11	230	230	235	250
					12	215		240	280
					13	170		260	295
					14	225	255	255	260
					15	230	245	260	300
					16	220		235	250

distance from the centre of the pituary to the pteryomaxillary fissure (unit 10^{-4}m)

Model comparison

i: subscript for individual $i = 1,..,I = 25$ (11girls+16boys)

j: subscript for measurement at age $t_j (8,10,12,14\, yrs)$

1)Purely Fixed Model

$$y_{ij} = \underbrace{(\alpha_0 + \alpha x_i)}_{\text{intercept}} + \underbrace{(\beta_0 + \beta x_i)}_{pente}(t_j - 8) + e_{ij}$$

2)Random intercept model

$$y_{ij} = (\alpha_0 + \alpha x_i + a_i) + (\beta_0 + \beta x_i)(t_j - 8) + e_{ij}$$

3)Random intercept & slope model assuming independent effects

$$y_{ij} = (\alpha_0 + \alpha x_i + a_i) + (\beta_0 + \beta x_i + b_i)(t_j - 8) + e_{ij}$$

or

$$y_{ij} = \phi_{i1} + \phi_{i2}(t_j - 8) + e_{ij},\ y_{ij} \sim_{\text{id}} \mathcal{N}(\eta_{ij}, \sigma_e^2)$$

with $\phi_i = \begin{pmatrix} \phi_{i1} \\ \phi_{i2} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \alpha_0 + \alpha x_i \\ \beta_0 + \beta x_i \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{pmatrix}\right]$

4)Random intercept & slope model assuming correlated effects

$$\phi_i = \begin{pmatrix} \phi_{i1} \\ \phi_{i2} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \alpha_0 + \alpha x_i \\ \beta_0 + \beta x_i \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}\right]$$

Model presentation:Hierarchical Bayes

1st level: $y_{ij} \sim_{\text{id}} \mathcal{N}\left(\eta_{ij}, \sigma_e^2\right)$ with $\eta_{ij} = \phi_{i1} + \phi_{i2}\left(t_j - 8\right)$

2nd level :

$$2a)\ \phi_i = \begin{pmatrix} \phi_{i1} \\ \phi_{i2} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \alpha_0 + \alpha x_i \\ \beta_0 + \beta x_i \end{pmatrix}, \underbrace{\begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}}_{\Sigma}\right]$$

$$2b)\ \sigma_e \sim \mathrm{U}\left(0, \Delta_e\right) or\ \sigma_e^2 \sim \mathrm{InvG}\left(1, \underline{\sigma}_e^2\right)$$

3rd level:

Fixed effects: $\alpha_0, \alpha, \beta_0, \beta \sim \mathrm{U}(\inf, \sup)$

Var (Covar) components:

$-If\ \sigma_{ab} = 0$, then i) $\sigma_a \sim \mathrm{U}\left(0, \Delta_a\right)$, same for $\sigma_b \sim \mathrm{U}\left(0, \Delta_b\right)$

or ii) $\sigma_a^2 \sim \mathrm{InvG}\left(1, \underline{\sigma}_a^2\right)$, same for $\sigma_b^2 \sim \mathrm{InvG}\left(1, \underline{\sigma}_b^2\right)$

$-If\ \sigma_{ab} \neq 0$, then i) $\sigma_a \sim \mathrm{U}\left(0, \Delta_a\right), \sigma_b \sim \mathrm{U}\left(0, \Delta_b\right), \rho \sim \mathrm{U}\left(-1, 1\right)$

or ii) $\Omega \sim W\left(\left(\nu \underline{\Sigma}\right)^{-1}, \nu\right)^*$ for $\Omega = \Sigma^{-1}$

with $\nu = \dim(\Omega) + 1$ and $\underline{\Sigma}$ known location parameter

*Take care as Winbugs uses another notation ie $W\left(\left(\nu \underline{\Sigma}\right), \nu\right)$

Resultats

Table : Criteria of comparison of models applied to Pothoff & Roy's data set
(missing data version of Little and Rubin, N=99)

Models	Residual Likelihood			Uniform priors		Inv-Gamma Wishart priors		Fractional priors	
	Do	AIC	BIC	Dm	DIC	Dm	DIC	Dm (0.05)	Dm (0.15)
Fixed	884.0.	886.0	888.6	919.5	907.6	919.4	907.4	866.8	770.6
Random Intercept	843.6	847.6	850.2	884.8	837.6	880.5	837.5	828.9	731.6
Intercept+Slope Indep	842.7	848.7	852.6	886.6	833.8	880.1	833.7	823.8	725.8
Intercept+Slope Correl	842.4	850.4	855.5	887.2	835.1	880.1	834.6	830.2	731.7

Priors on fixed effects are: $\alpha_0 \sim \mathrm{U}(0,500)$, $\alpha \sim \mathrm{U}(-50,20)$, $\beta_0 \sim \mathrm{U}(4,12)$, $\beta \sim \mathrm{U}(-10,4)$

Priors on variance covariance components are:

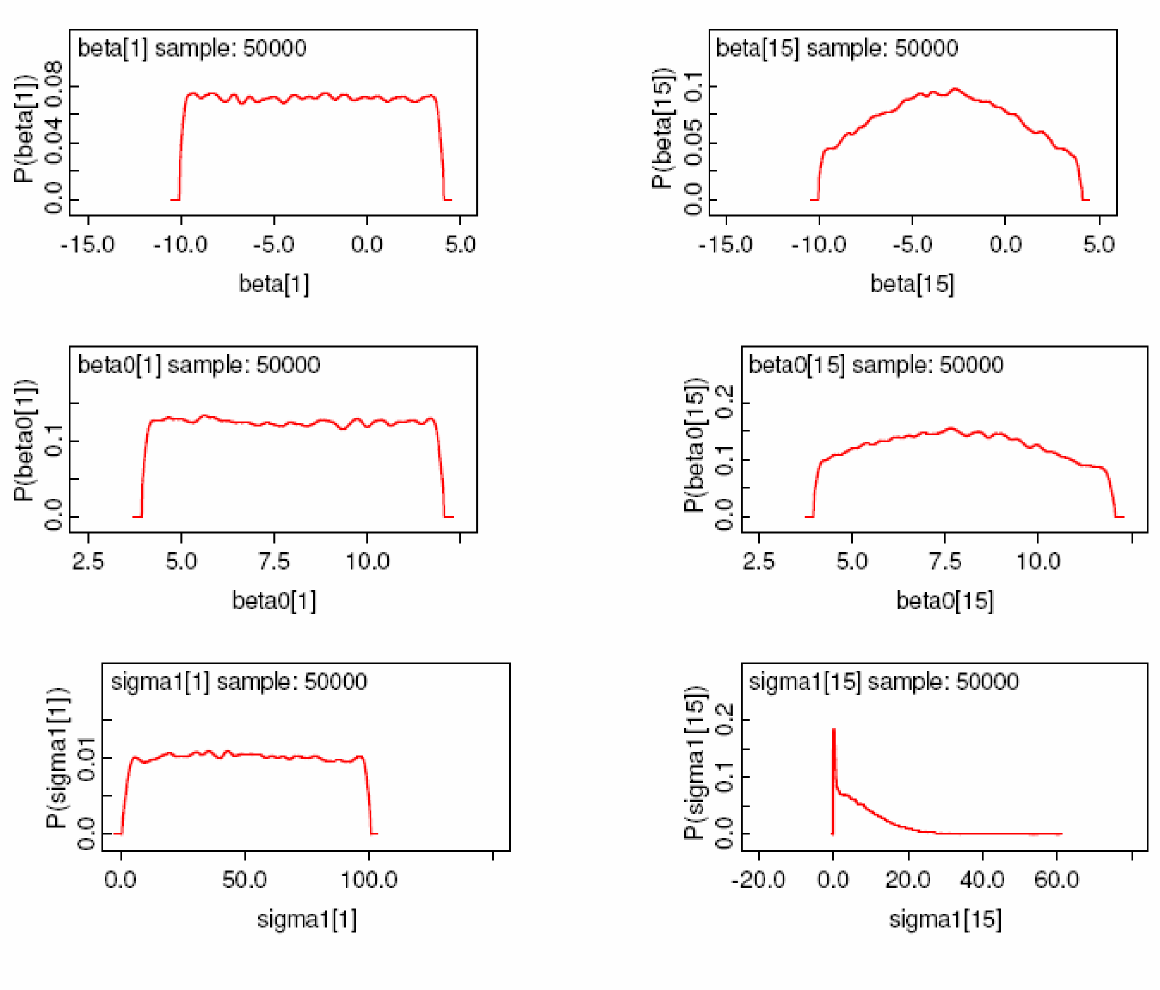
-uniform on standard deviations and correlation: $\sigma_e \sim \mathrm{U}(0,50)$, $\sigma_1 \sim \mathrm{U}(0,100)$, $\sigma_2 \sim \mathrm{U}(0,20)$, $\rho \sim \mathrm{U}(-1.0,1.0)$

-inverted gamma or Wishart on variances: $\sigma_e^{-2} \sim \mathrm{G}(1,200)$, $\sigma_1^{-2} \sim \mathrm{G}(1,300)$, $\sigma_2^{-2} \sim \mathrm{G}(1,2)$, $\Sigma^{-1} \sim \mathcal{W}\left((3\underline{\Sigma})^{-1},3\right)$

where $\underline{\Sigma} = Diag(300,2)$

Fractional "priors" (O'Hagan (1995) were based on a fraction (b=0.05, 0.15) of the likelihood to turn initial uniform into data-based priors.
Do : minimum frequentist deviance ; Dm : marginal (integrated) deviance, AIC, BIC computed as in SAS-proc mixed and DIC as in Winbugs.
Computations involved a trapezoidal integration with a grid of 30 points on the (0,1) defined as $t_i = (i/N)^3$ with $i = 1,..,30$ and $N = 30$

Resultats/Priors fractionnaires (b=0 vs 0.125)



Ex2: Modèles de différenciation génétique/Beta-Binomial

Modèle hiérarchique à plusieurs niveaux

i=locus; j=(sous)population

a_{ij}=Nbre de gènes d'un type allélique au locus i dans la pop j

p_{ij} = Fréquence de cet allèle au locus i dans la pop j

0) $\boxed{y_{ij} \mid \alpha_{ij} \sim_{id} \mathrm{B}\left(n_{ij}, \alpha_{ij}\right)}$

1) $\boxed{\alpha_{ij} \mid x_i, \lambda_{ij} \sim_{id} \mathrm{Beta}\left(\tau_j \pi_i, \tau_j \left(1-\pi_i\right)\right)}$ $\quad \tau_j = \dfrac{1-c_j}{c_j} \quad c_j = \textit{différenciation}$

2) $\pi_i \sim_{id} \mathrm{Beta}\left(a_\pi, b_\pi\right), c_j \sim_{id} \mathrm{Beta}\left(a_c, b_c\right)$

Modèle Migration-Dérive à l'équilibre (Balding)

Ex2: Modèle de Nicholson

Modèle Nicholson et al (2002) idem sauf

1) $\alpha_{ij} \mid x_i, \lambda_{ij} \sim_{id} N\left(\pi_i, c_j \pi_i \left(1-\pi_i\right)\right)$

Normale tronquée avec masses en 0 et 1

d'où $y_{ij} \mid \alpha_{ij} \sim_{id} \mathrm{B}\left(n_{ij}, \alpha_{ij}^*\right)$

où $\alpha_{ij}^* = \max(0, \min(1, \alpha_{ij}))$

2) $\pi_i \sim_{id} \mathrm{Beta}\left(a_\pi, b_\pi\right), c_j \sim_{id} \mathrm{Beta}\left(a_c, b_c\right)$

Modèle de dérive pure

Résultats

Table : Comparison of two models of distribution of gene frequencies on 495 SNPs in 14 cattle populations

Models	Dm	Dm(b=0.5)	D bar	KL	DIC
Nicholson et al	45370±5	16431	29960	7705	35130
Beta-Binomial	45094±7	16466	30160	7467	35360

Calculation via the PP method (Friel & Pettitt,2008) with 20 points defined as t=(i/20)**5
Dm: Marginal deviance=(Dbar+2KL) based on uniform (0,1) priors on c coefficients and pi gene frequencies.
Dm(b=0.5): Fractional Marginal Deviance with fraction of the likelihood b=0.5 used in the training sample.
Dbar: Posterior mean deviance
KL: Kullback-Leibler distance between posterior and prior =(MD-Dbar)/2
DIC: Deviance Information Criterion = Dbar+Pd (complexity)

Conclusion

- Idée basée sur l'intégration thermodynamique
- Connexion avec « bridge sampling » et « simulated tempering »
- Bien défini et général (modèles hiérarchiques comlexes)
- Les « theta » peuvent être définis comme les plus proches parents stochastiques (idem à DIC) pour bénéficier de l'indépendance conditionnelle
- Tirages uniquement dans les posteriors
- Facile à programmer (y compris Win/Openbugs)
- Attention à la discrétisation de t (au voisinage de 0)

Qlq références

- Friel N, Pettitt AN (2008) Marginal likelihood estimation via power posteriors, JRSS, B, 70, 589-607
- Frühwirth-Schnatter (2004) Estimating marginal likelihoods from mixtures & Markov switching models using bridge sampling techniques. Econometrics Journal, 7,143-167
- Gelman A, Meng X-L (1998) Simulating normalizing constants: from importance sampling to bridge sampling and path sampling, Statistical Science, 13, 163-185
- Marin JM, Robert CP (2009) Importance sampling methods for Bayesian discrimination between embedded models. arXiv:0910.2325v1
- Meng X-L, Wong WH (1996) Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. Statistica Sinica,6,831-860
- O Hagan A (1995) Fractional Bayes factors for model comparison. JRSS, B, 57, 99-138

Remerciements

- Nial Friel (U College, Dublin) pour l'intérêt porté à ce travail et ses précieux conseils et suggestions
- Tony O'Hagan pour ses explications sur le FBF
- Gilles Celeux, Mathieu Gautier & Yoan Soussan dans le cadre du stage Master Paris VI de ce dernier
- Christian Robert pour ses précieuses remarques et ses notes de blog
- Les groupes Applibugs & Babayes pour les discussions stimulantes sur DIC, BF,CPO et autres critères d'information (AIC,BIC)