Bayesian Inference of REML

☺ **Variance Components:**

Example:

1. **Linear model:** $Y_i = X_i\alpha + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma^2 I)$

   Our variance component is σ^2.

2. **Linear mixed effects model:** $Y_i = X_i\alpha + Z_i b_i + \varepsilon_i,$

   $\varepsilon_i \sim N(0, R_i)$ **and** $b_i \sim N(0, D)$

   Our variance components: R_i **and D.**

☺ **REML (Restricted Maximum Likelihood Estimate):**

Estimators of Variance Components

Peichun Chen, Institute of Statistical Science, Academia Sinica

☺ **Simulation of Linear Model: ($V = I$)**

Model： $y_i = X_i\alpha + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma^2 I)$

$$\alpha = (5,2)^t，\sigma^2 = \mathbf{4}，\boldsymbol{p} = \mathbf{2}。X_i = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}^t$$

prior: $IG(4,6)$、$IG(3,8)$、$IG(4,20)$、$IG(3,16)$

	$IG(4,6)$	$IG(3,8)$	$IG(4,20)$	$IG(3,16)$
Mean	**2**	**4**	**6.67**	**8**
Mode	**1.2**	**2**	**4**	**4**
Variance	**2**	**16**	**22.22**	**64**

$$\text{posterior mode} = \tilde{\sigma}^2 = \frac{B + (y - X\hat{\alpha})^t (y - X\hat{\alpha})/2}{(N-p)/2 + A + 1}$$

Fixed Effects Model: $y \sim MVN\ (X\alpha, \sigma^2 V)$

a. Likelihood of ML:

$$L(\alpha, \sigma^2, \theta | y) \propto -\frac{1}{2}\left\{\sum_{i=1}^{m} n_i \times \log\sigma^2 + \log|V| + \sigma^{-2}(y - X\alpha)^t V^{-1}(y - X\alpha)\right\}$$

b. Likelihood of REML:

$$L^*(\eta) \propto -\frac{1}{2}\log|H| - \frac{1}{2}\log|X^t H^{-1} X| - \frac{1}{2}(y - X\hat{\alpha})^t H^{-1}(y - X\hat{\alpha})$$

$\boldsymbol{H} \equiv H(\eta)$，$\eta = (\sigma^2, \theta)$

Mixed Effects Model: $Y_i = X_i\alpha + Z_i b_i + \varepsilon_i,\ \varepsilon_i \sim N(0, R_i)$

and $b_i \sim N(0, D)$

→ **E-M algorithm.**

☺ **Relationship between REML and Bayesian Inference:**

For m=10 and 50, repeat 100 times.

☺ **Result:**

m＝10;

	REML	$IG(4,6)$	$IG(3,8)$	$IG(4,20)$	$IG(3,16)$
Mean	**4.0879**	**3.1616**	**3.4332**	**3.8616**	**3.8542**
Variance	**1.1915**	**0.5838**	**0.6469**	**0.5838**	**0.6469**
SE mean	**0.1092**	**0.0764**	**0.0804**	**0.0764**	**0.0804**
MSRE	**0.074205**	**0.08006**	**0.06104**	**0.03732**	**0.04135**

m=50;

	REML	$IG(4,6)$	$IG(3,8)$	$IG(4,20)$	$IG(3,16)$
Mean	**4.4063**	**3.8178**	**3.8915**	**3.9928**	**3.9927**
Variance	**0.2341**	**0.2003**	**0.2054**	**0.2003**	**0.2054**
SE mean	**0.0484**	**0.0448**	**0.0453**	**0.0448**	**0.0453**
MSRE	**0.01462**	**0.01447**	**0.01345**	**0.01240**	**0.01272**

$$\text{Where MSRE} = \frac{1}{100}\sum_{i=1}^{100}\left(\frac{\sigma_i^2 - \sigma^2}{\sigma^2}\right)^2$$

a. $p(\alpha,\eta)$ **flat** **(Harville, 1974)**

$$postr(\alpha,\eta|y)\propto f_y(y|\alpha,\eta)p(\alpha,\eta)\propto f_y(y|\alpha,\eta)$$

∴posterior mode (α,η)**=MLE**(α,η)

b. If we are only concerned about η**,** $p(\alpha,\eta)$ **flat,** $\alpha\perp\eta$

$$postr(\eta|y)=\int p(\alpha,\eta|y)d\alpha\propto\int f_y(y|\alpha,\eta)p(\alpha,\eta)d\alpha\propto f_z(B^T y|\eta)$$

∴posterior mode (η)**=REML**(η)

☺ Conjugate Prior:

a. Model: $y\sim$**MVN** $(X\alpha,\sigma^2V)$

1. If $V=I$

prior: $\sigma^2|A,B\sim IG(A,B)$, $\alpha|\theta_\alpha\sim f_\alpha(\alpha|\theta_\alpha)$

likelihood: $L(\alpha,\sigma^2,V|y)\propto|\sigma^2V|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(y-X\alpha)^t(\sigma^2V)^{-1}(y-X\alpha)\right\}$

☺ Simulation of Linear Model with Correlation: ($V\neq I$)

$$\varepsilon_i\sim N(0,\sigma^2V)\qquad T=\sigma^2V=\begin{bmatrix}2&1&1\\1&2&1\\1&1&2\end{bmatrix}$$

By Bernstein-von Mises Theorem: under some conditions, posterior mean=REML

prior: $T^{-1}|k,\Sigma\sim W(S_1^{-1},n,5)$、$W(S_1^{-1},n,10)$、$W(S_2^{-1},n,5)$、$W(S_2^{-1},n,10)$

Where $S_1=\begin{bmatrix}2&1&1\\1&2&1\\1&1&2\end{bmatrix}$ **and** $S_2=\begin{bmatrix}2&0&0\\0&2&0\\0&0&2\end{bmatrix}$

	$W(S_1^{-1},n,5)$	$W(S_1^{-1},n,10)$	$W(S_2^{-1},n,5)$	$W(S_2^{-1},n,10)$
Mean	$\begin{bmatrix}2&1&1\\1&2&1\\1&1&2\end{bmatrix}$	$\begin{bmatrix}0.33&0.17&0.17\\0.17&0.33&0.17\\0.17&0.17&0.33\end{bmatrix}$	$\begin{bmatrix}2&0&0\\0&2&0\\0&0&2\end{bmatrix}$	$\begin{bmatrix}0.33&0&0\\0&0.33&0\\0&0&0.33\end{bmatrix}$

posterior mean= $\dfrac{\left[\Sigma^{-1}+(y-X\hat{\alpha})(y-X\hat{\alpha})'\right]}{m+k-2n-2}$

posterior: $p(\sigma^2|y,\alpha,\theta_\alpha,A,B)\propto L(\sigma^2|y,\alpha)p(\sigma^2)$

$$\propto(\sigma^2)^{-\left(\frac{N}{2}+A+1\right)}\exp\left\{-\frac{1}{\sigma^2}\left[\frac{1}{2}(y-X\alpha)^t(y-X\alpha)+B\right]\right\}$$

hence, $\sigma^2|y,\alpha,\theta_\alpha,A,B\sim IG\left(\frac{N}{2}+A,\frac{1}{2}(y-X\alpha)^t(y-X\alpha)+B\right)$

2. If $V\neq I$

prior: $T^{-1}|k,\Sigma\sim Wishart_N(\Sigma,N,k)$**, where** $T_{N\times N}$**=**σ^2V

posterior:

$$p(T^{-1}|k,\Sigma,y,\alpha,\theta_\alpha)\propto|T|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(y-X\alpha)^tT^{-1}(y-X\alpha)\right\}|T|^{-\frac{1}{2}(k-N-1)}\exp\left\{-\frac{1}{2}tr(\Sigma^{-1}T^{-1})\right\}$$

$$=|T|^{-\frac{1}{2}(k-N)}\exp\left\{-\frac{1}{2}tr\left(\Sigma^{-1}+(y-X\alpha)(y-X\alpha)^t\right)T^{-1}\right\}$$

so, $T^{-1}|k,\Sigma,y,\alpha,\theta_\alpha\sim Wishart\left(\left[\Sigma^{-1}+(y-X\alpha)(y-X\alpha)^t\right]^{-1},N,k+1\right)$

For m=50, repeat=100.

☺ Result:

$W(S_1^{-1},n,5)$

	T_{11}	T_{12}	T_{13}	T_{22}	T_{23}	T_{33}
Mean	**2.1442**	**1.0821**	**1.0983**	**2.1211**	**1.1177**	**2.2135**
Variance	**0.1911**	**0.1310**	**0.1413**	**0.2033**	**0.1369**	**0.1914**

$W(S_1^{-1},n,10)$

	T_{11}	T_{12}	T_{13}	T_{22}	T_{23}	T_{33}
Mean	**1.9393**	**0.9447**	**0.9635**	**1.9230**	**0.9467**	**1.9343**
Variance	**0.1571**	**0.1123**	**0.0859**	**0.1878**	**0.1100**	**0.1472**

$W(S_2^{-1},n,5)$

	T_{11}	T_{12}	T_{13}	T_{22}	T_{23}	T_{33}
Mean	**2.0869**	**1.0727**	**1.0540**	**2.1847**	**1.0685**	**2.1072**
Variance	**0.1717**	**0.1373**	**0.1206**	**0.1430**	**0.0685**	**0.2123**

$W(S_2^{-1},n,10)$

	T_{11}	T_{12}	T_{13}	T_{22}	T_{23}	T_{33}
Mean	**1.8462**	**0.9067**	**0.9218**	**1.8690**	**0.9276**	**1.9249**
Variance	**0.1051**	**0.0730**	**0.0779**	**0.1238**	**0.0859**	**0.1551**