Hierarchical Generalized Linear Models Have A Great Potential in Genetics

Lars Rönnegård[1,2] and Youngjo Lee [3]

[1]Dalarna University, Sweden

[2]Swedish University of Agricultural Sciences

[3]Seoul National University, Korea

WCGALP, 2010

Definition of the h-likelihood

The hierarchical log-likelihood (**h-likelihood**): $h(\beta, \theta, u) = \log f(y|u) + \log f(u)$.

Example:
The log of the normal density function for n iid observations with mean 0 and variance σ^2 is: $log(f(\mathbf{y})) = -\frac{n}{2} log(\sigma^2) - \frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{y}$

For a linear mixed model

$$h(\beta, \theta, \mathbf{u}) = \{-\frac{n}{2} log(\sigma_e^2) - \frac{1}{2\sigma_e^2}\mathbf{e}^T\mathbf{e}\} + \{-\frac{k}{2} log(\sigma_u^2) - \frac{1}{2\sigma_u^2}\mathbf{u}^T\mathbf{u}\}$$

where $\mathbf{e} = \mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{u}$

Definition of the h-likelihood

The hierarchical log-likelihood (**h-likelihood**): $h(\beta, \theta, u) = \log f(y|u) + \log f(u)$.

Example:
The log of the normal density function for n iid observations with mean 0 and variance σ^2 is: $log(f(\mathbf{y})) = -\frac{n}{2} log(\sigma^2) - \frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{y}$

For a linear mixed model

$$h(\beta, \theta, \mathbf{u}) = \{-\frac{n}{2} log(\sigma_e^2) - \frac{1}{2\sigma_e^2}\mathbf{e}^T\mathbf{e}\} + \{-\frac{k}{2} log(\sigma_u^2) - \frac{1}{2\sigma_u^2}\mathbf{u}^T\mathbf{u}\}$$

where $\mathbf{e} = \mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{u}$

Definition of the h-likelihood

The hierarchical log-likelihood (**h-likelihood**): $h(\beta, \theta, u) = \log f(y|u) + \log f(u)$.

Example:
The log of the normal density function for n iid observations with mean 0 and variance σ^2 is: $log(f(\mathbf{y})) = -\frac{n}{2} log(\sigma^2) - \frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{y}$

For a linear mixed model

$$h(\beta, \theta, \mathbf{u}) = \{-\frac{n}{2} log(\sigma_e^2) - \frac{1}{2\sigma_e^2}\mathbf{e}^T\mathbf{e}\} + \{-\frac{k}{2} log(\sigma_u^2) - \frac{1}{2\sigma_u^2}\mathbf{u}^T\mathbf{u}\}$$

where $\mathbf{e} = \mathbf{y} - \mathbf{X}\beta - \mathbf{Zu}$

h-likelihood estimation for HGLM

Estimating fixed and random effects: $\frac{\partial h}{\partial \beta} = 0$ and $\frac{\partial h}{\partial u} = 0$

Estimating variance components:

- MLE are biased. Solution: Adjusted profile likelihood

$$h_p = \left(h + \frac{1}{2} log|2\pi H^{-1}| \right)_{\beta=\hat{\beta}, u=\hat{u}}$$

h-likelihood estimation for HGLM

Estimating fixed and random effects: $\frac{\partial h}{\partial \beta} = 0$ and $\frac{\partial h}{\partial u} = 0$

Estimating variance components:

- MLE are biased. Solution: Adjusted profile likelihood

$$h_p = \left(h + \frac{1}{2} log|2\pi H^{-1}| \right)_{\beta=\hat{\beta}, u=\hat{u}}$$

Why is this useful?

- Estimation and inference of generalized linear models with random effects
- Both **y** and **u** can come from a wide range of distributions.
- Approximate h-likelihood estimates can obtained by iterating between a relatively simple set of GLM
  - Implemented in the R package **hglm** (Rönnegård, Alam & Shen)
- Estimates can also be computed using iterative Newton-Raphson directly on the h-likelihood
  - Implemented in the R package **HGLMMM** (Marek Molas)

Why is this useful?

- Estimation and inference of generalized linear models with random effects
- Both **y** and **u** can come from a wide range of distributions.
- Approximate h-likelihood estimates can obtained by iterating between a relatively simple set of GLM
  - Implemented in the R package **hglm** (Rönnegård, Alam & Shen)
- Estimates can also be computed using iterative Newton-Raphson directly on the h-likelihood
  - Implemented in the R package **HGLMMM** (Marek Molas)

Why is this useful?

- Estimation and inference of generalized linear models with random effects
- Both **y** and **u** can come from a wide range of distributions.
- Approximate h-likelihood estimates can obtained by iterating between a relatively simple set of GLM
  - Implemented in the R package **hglm** (Rönnegård, Alam & Shen)
- Estimates can also be computed using iterative Newton-Raphson directly on the h-likelihood
  - Implemented in the R package **HGLMMM** (Marek Molas)

Double HGLM

- Easy to include predictors for the variance components in the h-likelihood
  $var(\mathbf{y}|\mathbf{u},\mathbf{u}_d) = \phi$ with $log(\phi) = \mathbf{X}_d b_d + \mathbf{Z}_d \mathbf{u}_d$

  $$h(\beta,\theta,u,u_d) = \log f(y|u,u_d) + \log f(u) + \log f(u_d)$$

- Estimation possible through a double set of GLM
  - Double Hierarchical GLM (**DHGLM**)

Double HGLM

- Easy to include predictors for the variance components in the h-likelihood
  $var(\mathbf{y}|\mathbf{u},\mathbf{u}_d) = \phi$ with $log(\phi) = \mathbf{X}_d b_d + \mathbf{Z}_d \mathbf{u}_d$

  $$h(\beta,\theta,u,u_d) = \log f(y|u,u_d) + \log f(u) + \log f(u_d)$$

- Estimation possible through a double set of GLM
  - Double Hierarchical GLM (**DHGLM**)

QTL analysis

Estimate effects of Single Nucleotide Polymorphic markers using a DHGLM

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$$\mathbf{e} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_\mathrm{n}\sigma^2_\mathrm{e})$$

$$\mathrm{u_i} \sim \mathrm{N}(\mathbf{0}, \sigma^2_{\mathrm{u,i}}),\ log(\sigma^2_u) = \mu + \mathbf{Z}_\mathrm{d}\mathbf{u}_d$$

$$\mathbf{u}_\mathrm{d} \sim \mathrm{N}(0, \mathbf{I}\sigma^2_\mathrm{d})$$

Analysis of the QTLMAS 2009 data

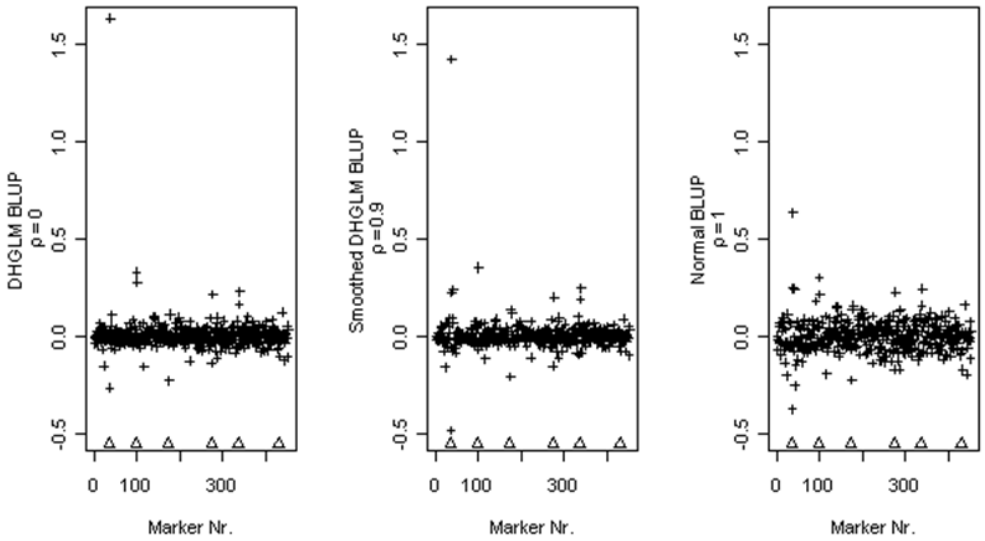


Figure 1: SNP-effects for the models: DHGLM, Smoothed DHGLM and a linear mixed model with constant variance. Simulated data from QTLMAS '09.

Analysis of the QTLMAS 2009 data

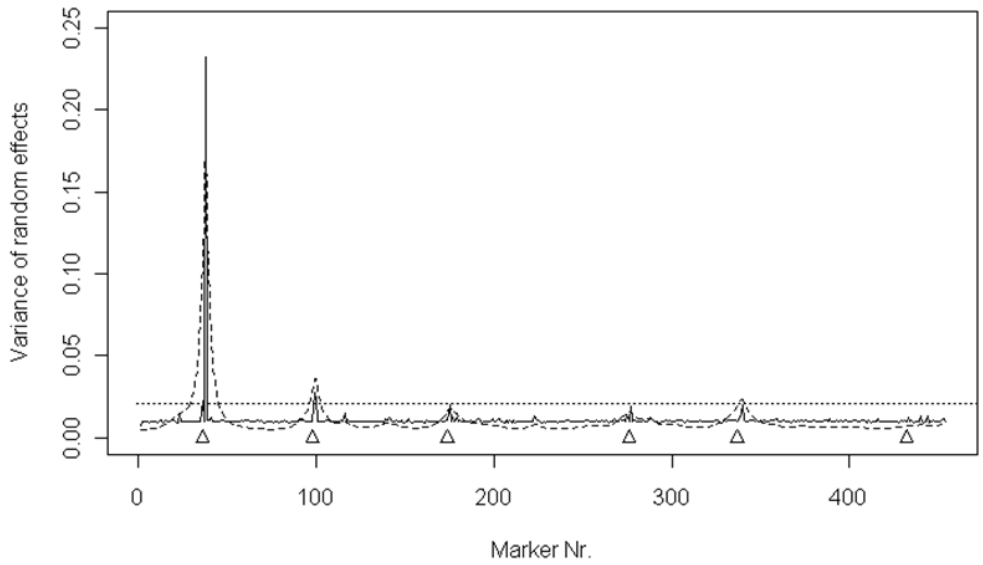
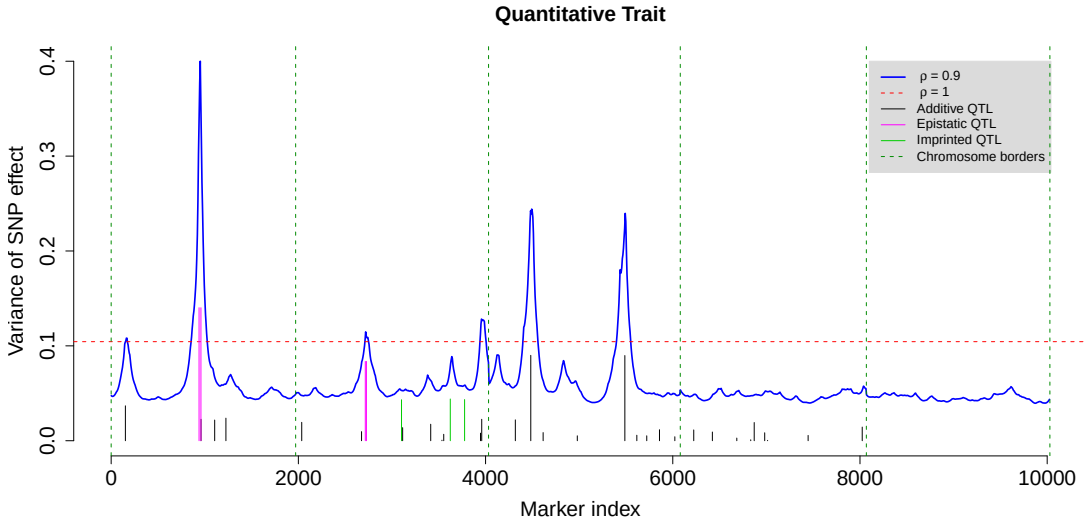
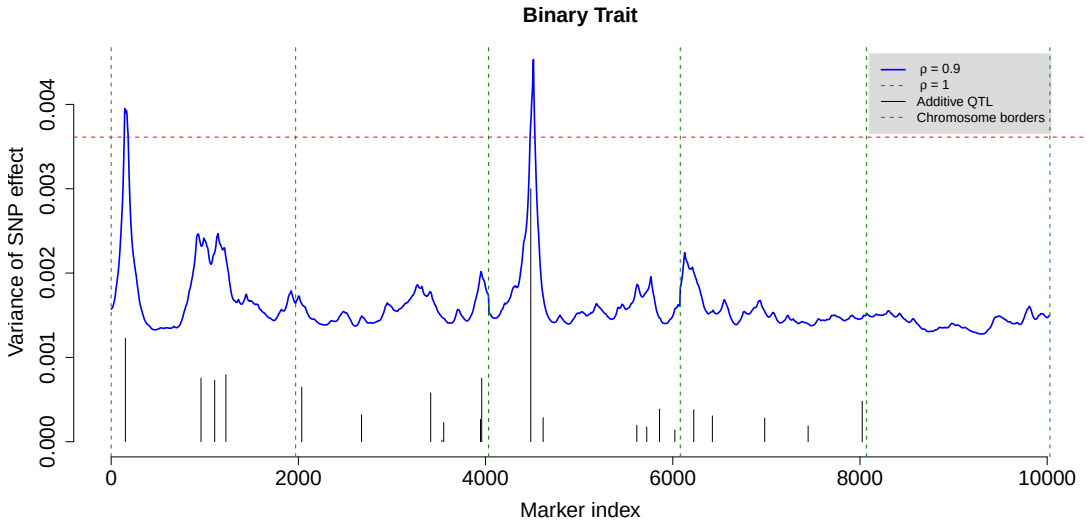


Figure 2: Estimated variance of random SNP-effects for QTLMAS '09 data.

Analysis of the QTLMAS 2010 data





Using DHGLM for estimating genetic heterogeneity in residual variance

In Rönnegård et al. (2010 GSE 42:8) we fit a DHGLM for the pig litter size data previously studied in Sorensen and Waagepetersen (2003)

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$$\mathbf{u} \sim \mathrm{N}(\mathbf{0}, \mathbf{A}\sigma^2_{\mathrm{u}})$$

$$\mathrm{e_i} \sim \mathrm{N}(\mathbf{0}, \sigma^2_{\mathrm{e,i}}),\ log(\sigma^2_e) = \mathbf{X}_d\beta_{\mathrm{d}} + \mathbf{Z}\mathbf{u}_d$$

$$\mathbf{u}_{\mathrm{d}} \sim \mathrm{N}(0, \mathbf{A}\sigma^2_{\mathrm{d}})$$

Data Description

Data from Danish Pig Production.

- Pig litter size from 4,149 sows
  - mean litter size 10.3
- The data includes 10,060 records from these 4,149 sows in 82 herds.
- Fixed effects: herd, season, type of insemination, parity

Data Description

Data from Danish Pig Production.

- Pig litter size from 4,149 sows
  - mean litter size 10.3
- The data includes 10,060 records from these 4,149 sows in 82 herds.
- Fixed effects: herd, season, type of insemination, parity

Fitted Model and Estimates

Model Random animal and sow effects included in the model for the mean and also in the model for the residual variance.

Table 1 - Comparison between DHGLM estimates and the estimates obtained by Sorensen & Waagepetersen (2003 Genetics)

	Mean model		Model for residual variance					
			Fixed effects			Variances		
	σ^2_a	σ^2_p	b_0	b_{ins}	b_{par}	$\sigma^2_{a_d}$	$\sigma^2_{p_d}$	ρ
DHGLM	1.36	0.44	1.72	-0.17	0.32	0.09	0.06	
S&W 2003	1.62	0.60	1.77	-0.17	0.35	0.09	0.06	-0.62

Fitted Model and Estimates

Model Random animal and sow effects included in the model for the mean and also in the model for the residual variance.

Table 1 - Comparison between DHGLM estimates and the estimates obtained by Sorensen & Waagepetersen (2003 Genetics)

	Mean model		Model for residual variance					
			Fixed effects			Variances		
	σ^2_a	σ^2_p	b_0	b_{ins}	b_{par}	$\sigma^2_{a_d}$	$\sigma^2_{p_d}$	ρ
DHGLM	1.36	0.44	1.72	-0.17	0.32	0.09	0.06	
S&W 2003	1.62	0.60	1.77	-0.17	0.35	0.09	0.06	-0.62

Summary

- The h-likelihood is an extension of Henderson's joint likelihood
- The h-likelihood can be used for estimation and inference for GLMs with random effects
- Estimation does not require MCMC nor numerical integration
- Random effects can be included in a model for the variance components (DHGLM)

Thank you!

Lars Rönnegård
lrn@du.se
www.larsronnegard.se

Special thanks to my students Majbritt Felleki and Xia Shen.