Warped linear mixed models for the genetic analysis of transformed phenotypes

Habib Ganjgahi

January 22, 2015
Outline

1 Background

2 Warped LMM

3 Result
Background

- Why imaging genetic?
  - Heritability
  - Association Analysis
- Linear Mixed effect Models (LMM):

\[ Y = X\beta + g + \epsilon, \]

where

\[ \text{cov}(g) = \sigma_g^2(2\Phi), \text{cov}(\epsilon) = \Sigma = \sigma_g^2(2\Phi) + \sigma_e^2 I. \]

- Parameter Estimation: Likelihood function
- Hypothesis testing: LRT

Kinship matrix $\Phi$
LMM Limitations

- Likelihood optimisation:
  - Convergence Failure
  - Computationally Intensive
- Inference: Random effect, covariates
- Residuals follow Normal distribution
  - Increased false positive
  - Power decreasing
- Transform the data: marginal distribution is approximately Gaussian
- Transformations: log-transformation, Rank based
- No consensus, depends on data
- Limitations:
  - no criterion to select one transformation over another
  - Time consuming and multiple comparison
Intuition: fit transformation while performing genetic analyses.

- Estimate the Transformation based on the data.
- Incorporate transformation in likelihood function

\[ z = f(y, \Psi) \]

\( y \) observed non-normal phenotype, \( z \) corresponded normal distributed phenotype and \( f \) monotonic function with \( \Psi \) parameters

\[ z = x\beta + g + \epsilon \]

\[ z \sim N(x\beta, \sigma_g^2(2\Phi) + \sigma_e^2 I) \]
Parameter Estimation

Likelihood function for hidden $z$:

$$L = -\frac{1}{2} \left( \log |\Sigma| + (z - x\beta)'\Sigma^{-1}(z - x\beta) + N \log 2\pi \right)$$

Incorporating Jacobian:

$$L = -\frac{1}{2}(\log |\Sigma| + (f(y, \Psi) - x\beta)'\Sigma^{-1}(f(y, \Psi) - x\beta) + N \log 2\pi - \sum \log \frac{\partial f(y, \Psi)}{\partial y})$$
Any monotonic function:

$$f(y; \Psi) = d.y + \sum_{i=1}^{l} a_i \tanh(b_i(y + c_i))$$

where $\Psi = (d, a_1, b_1, c_1, \ldots, a_l, b_l, c_l)$, $f$ is sum up of step functions, $a_i$: step size, $b_i$: steepness, $c_i$: location and $d$ slope of linear part.

$$f_{\text{Box-Cox}}(y, \Psi) = \begin{cases} 
\frac{y^{\psi-1}}{\psi} & \text{if } \Psi \neq 0 \\
\ln(y) & \text{if } \Psi = 0
\end{cases}$$

Shifted logarithmic transformation or scaled arsinh.
Simulations

- Genetic effect: Hapmap
- $h^2$: 0.1, 0.2, 0.4, 0.7, 0.9
- Sample size: 200, 400, 600, 800, 1000.
- $z = ty + (1 - t)f(y)$, where $t$ determines intensity of transformation
Simulated Data Results

### Results

**Simulated heritability**

- **Legend:** LMM, Log-LMM, Box-Cox LMM, WarpedLMM
- **Graph:** Box plots showing the difference between estimated heritability ($\hat{h}^2$) and true heritability ($h^2$) across different simulated heritability levels (0.1 to 0.9).

**Number of samples**

- **Graph:** Box plots showing the difference between estimated heritability ($\hat{h}^2$) and true heritability ($h^2$) across different sample sizes (200 to 1,000).

**Linearity of transformation**

- **Graph:** Box plots showing the difference between estimated heritability ($\hat{h}^2$) and true heritability ($h^2$) across different linearity values (0.0 to 1.0).
Heritability Analysis of 52 phenotypes

![Graph showing the heritability analysis of 52 phenotypes](image)

- Consistency of our findings by comparing both models in an out-of-sample prediction task.
- We performed a tenfold cross-validation experiment, where each model was repeatedly trained and compared it to a standard LMM using out-of-sample validation.
- The heritability estimates of the WarpedLMM model were lower than those of the LMM (Fig. 2b), even for phenotypes where the corresponding associations were detected.
- WarpedLMM identified a total of six distinct QTL.
- WarpedLMM are invertible, we can assess the prediction accuracy of a genetic model on the natural scale of the raw phenotype prediction.
- As the warping functions fit by WarpedLMM are also more accurate on real data.
- WarpedLMM can help avoid under- or overfitting in the model.
- WarpedLMM are also more accurate on real data.
- WarpedLMM can help avoid under- or overfitting in the model.
Human Data Analysis

Application in Pleiotropy

(a) Without transforming the phenotypes

(b) Applying the transformation found by WarpedLMM