Warped linear mixed models for the genetic analysis of transformed phenotypes
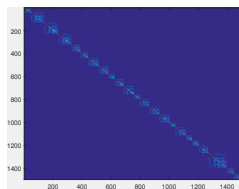
Habib Ganjgahi

January 22, 2015

Outline

Background

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

$$Y = X\beta + g + \epsilon,$$

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

Kinship matrix Φ

- Parameter Estimation: Likelihood function
- Hypothesis testing: LRT

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

Warped LMM

- 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 Ψ parameters

$$z = x\beta + g + \epsilon$$
$$z \sim N(x\beta, \sigma_g^2(2\Phi) + \sigma_\epsilon^2 I)$$

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:

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

Warping function

Any monotonic function:

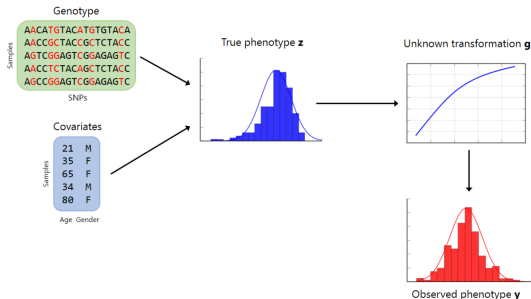$$f(y; \Psi) = d.y + \sum_{i=1}^{I} a_i \tanh(b_i(y + c_i))$$

where $\Psi = (d, a_1, b_1, c_1, \cdots, a_I, b_I, c_I)$, f is sum up of step functions, a_i: step size, b_i: steepness, c_i: location and d slope of linear part.

$$f_{\mathrm{Box-Cox}} \quad (y, \Psi) = \begin{cases} \frac{y^{\Psi} - 1}{\Psi} & \Psi \neq 0 \\ \ln(y) & \Psi = 0 \end{cases}$$
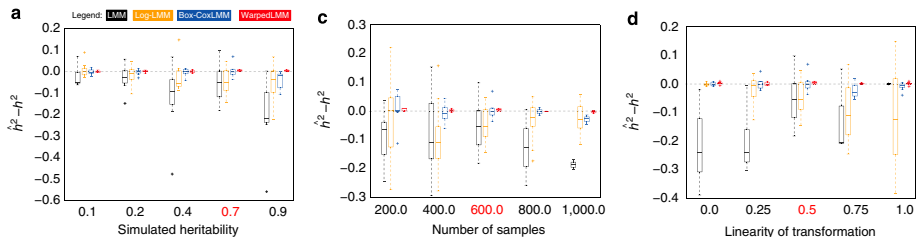
Shifted logarithmic transformation or scaled arsinh.

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
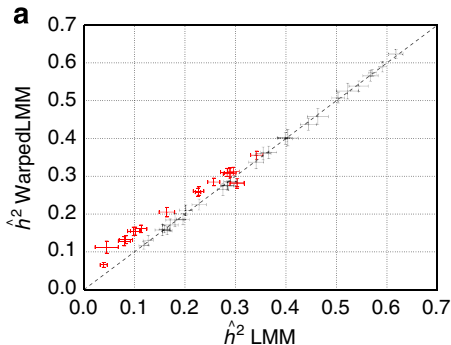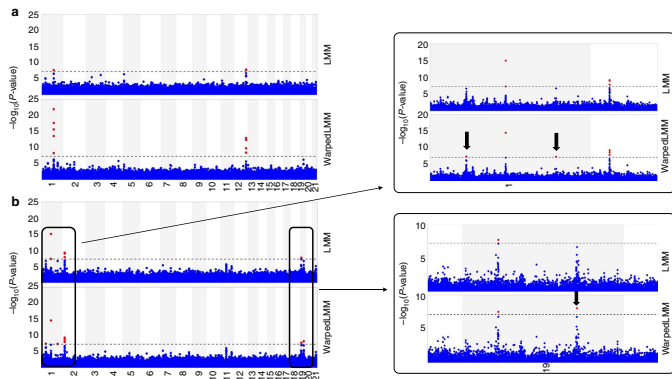
Results

Simulated Data Results

Real Data Analysis

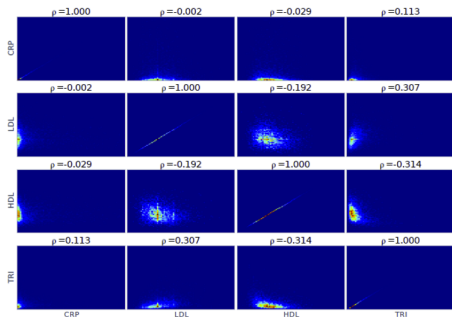Heritability Analysis of 52 phenotypes

Human Data Analysis

GWAS Analysis: Northern Finland birth cohort. Phenotypes: HDL, IDL, triglycerids, C-reactive protein. Transformations: logarithmic and linear scale.
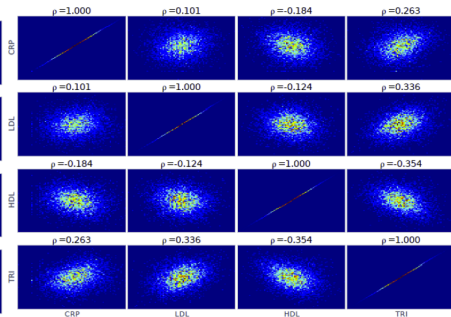
Human Data Analysis

Application in Pleiotropy

(a) Without transforming the phenotypes

(b) Applying the transformation found by WarpedLMM