Machine Learning and Statistics in Genetics and Genomics

V: Linear mixed models for Genetics

Christoph Lippert

Microsoft Research
eScience group

Los Angeles , USA

Current topics in computational biology
UCLA
Winter quarter 2014

GWAS Introduction

Population Structure

- Population structure

Population structure correction

- Genomic control
- Linear mixed models (LMM)
- FaST linear mixed models
- Dilution
- Proximal contamination

FaST-LMM-Set

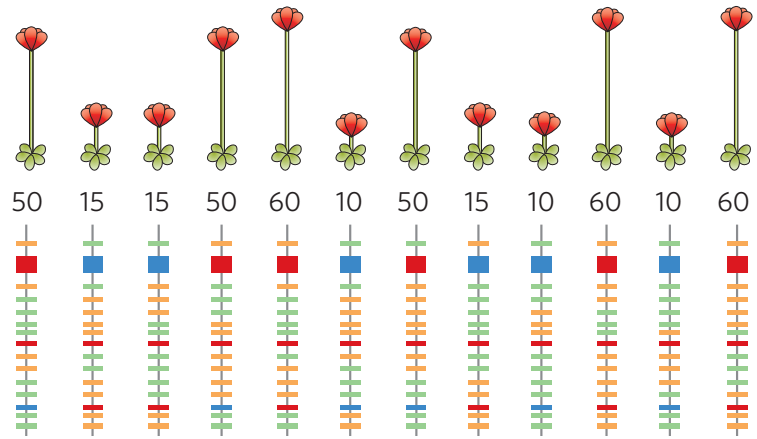
Outline

Outline

Genome-wide association studies (GWAS)

Given:

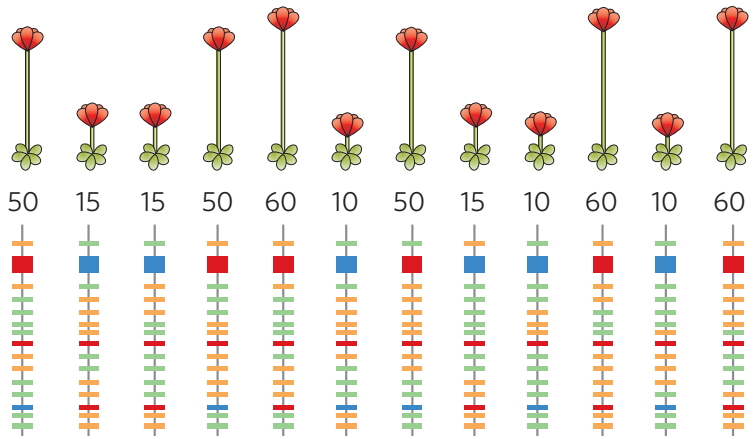
- Genotype for multiple individuals
  - Genome-wide single nucleotide polymorphism (SNP) markers
- Phenotypes for the same individuals
  - common diseases (qualitative)
  - height, BMI, ... (quantitative)



Genome-wide association studies (GWAS)

Given:

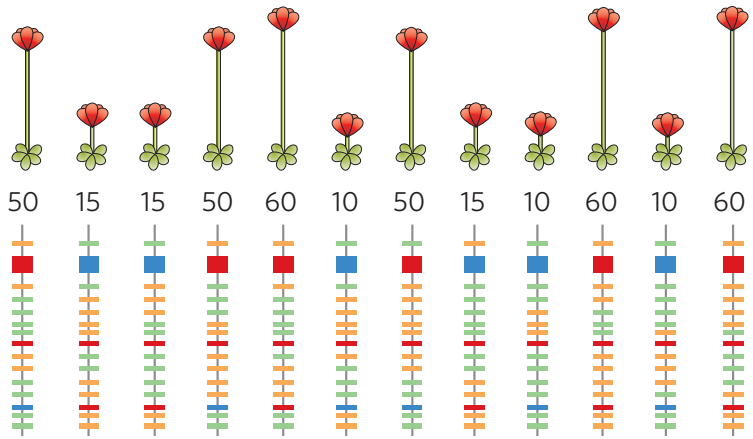
- Genotype for multiple individuals
  - Genome-wide single nucleotide polymorphism (SNP) markers
- Phenotypes for the same individuals
  - common diseases (qualitative)
  - height, BMI, ... (quantitative)



Genome-wide association studies (GWAS)

Given:

- Genotype for multiple individuals
  - Genome-wide single nucleotide polymorphism (SNP) markers
- Phenotypes for the same individuals
  - common diseases (qualitative)
  - height, BMI, ... (quantitative)



Genome-wide association studies (GWAS)

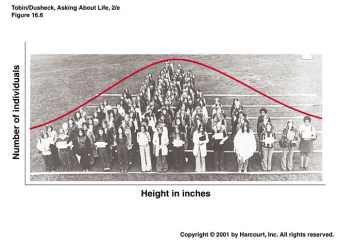
Given:

- Genotype for multiple individuals
  - Genome-wide single nucleotide polymorphism (SNP) markers
- Phenotypes for the same individuals
  - common diseases (qualitative)
  - height, BMI, ... (quantitative)

Genome-wide association studies (GWAS)

Given:

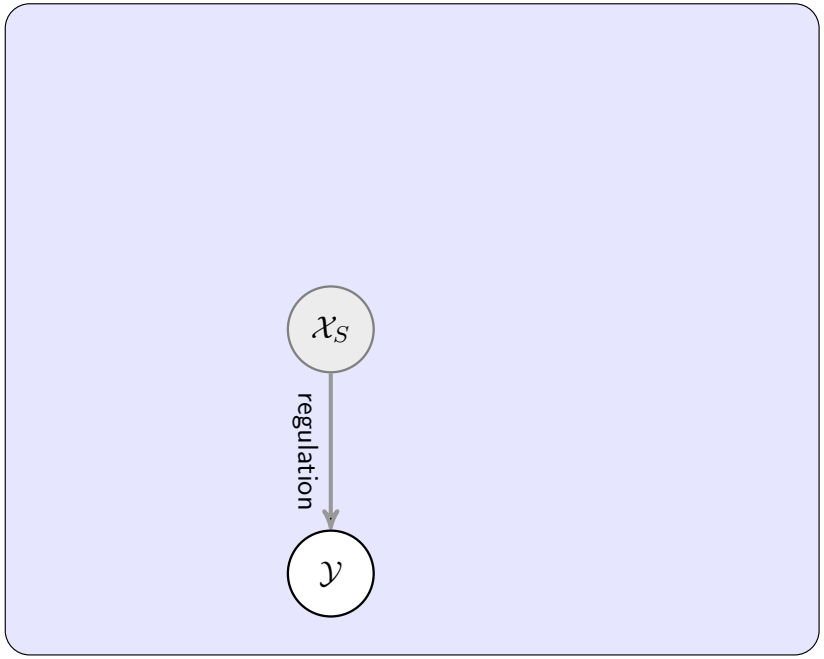
- Genotype for multiple individuals
  - Genome-wide single nucleotide polymorphism (SNP) markers
- Phenotypes for the same individuals
  - common diseases (qualitative)
  - height, BMI, . . . (quantitative)



Genome-wide association studies (GWAS)

Given:

- Genotype for multiple individuals
  - Genome-wide single nucleotide polymorphism (SNP) markers
- Phenotypes for the same individuals
  - common diseases (qualitative)
  - height, BMI, ... (quantitative)



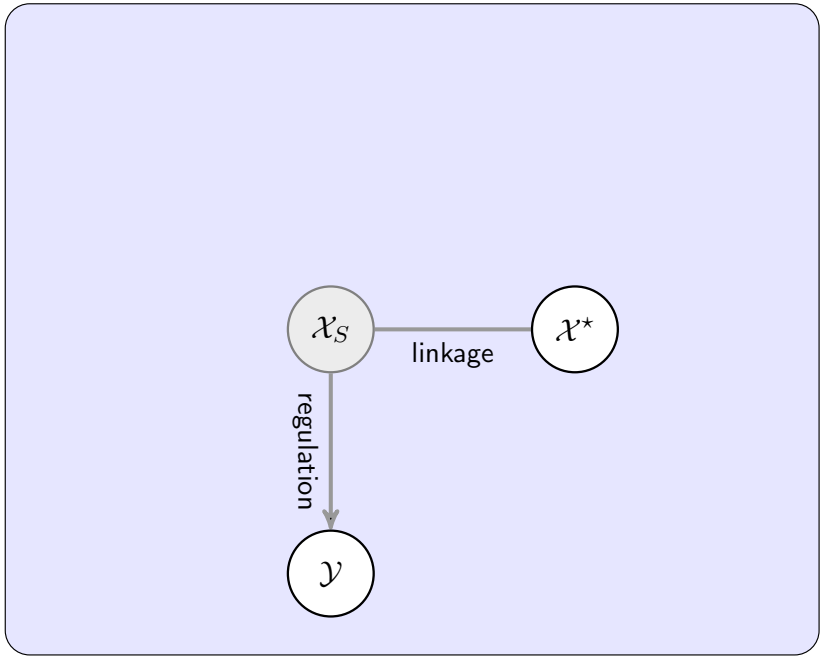
Goal:

- Identify causal loci that explain phenotypic differences.
- Use linked markers

Genome-wide association studies (GWAS)

Given:

- Genotype for multiple individuals
  - Genome-wide single nucleotide polymorphism (SNP) markers
- Phenotypes for the same individuals
  - common diseases (qualitative)
  - height, BMI, ... (quantitative)



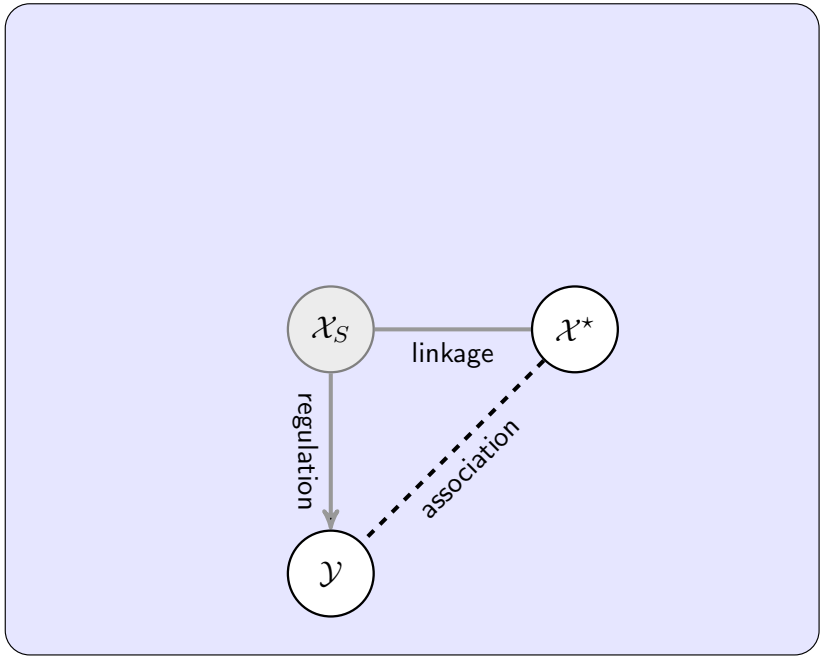
Goal:

- Identify causal loci that explain phenotypic differences.
- Use linked markers

Genome-wide association studies (GWAS)

Given:

- Genotype for multiple individuals
  - Genome-wide single nucleotide polymorphism (SNP) markers
- Phenotypes for the same individuals
  - common diseases (qualitative)
  - height, BMI, ... (quantitative)



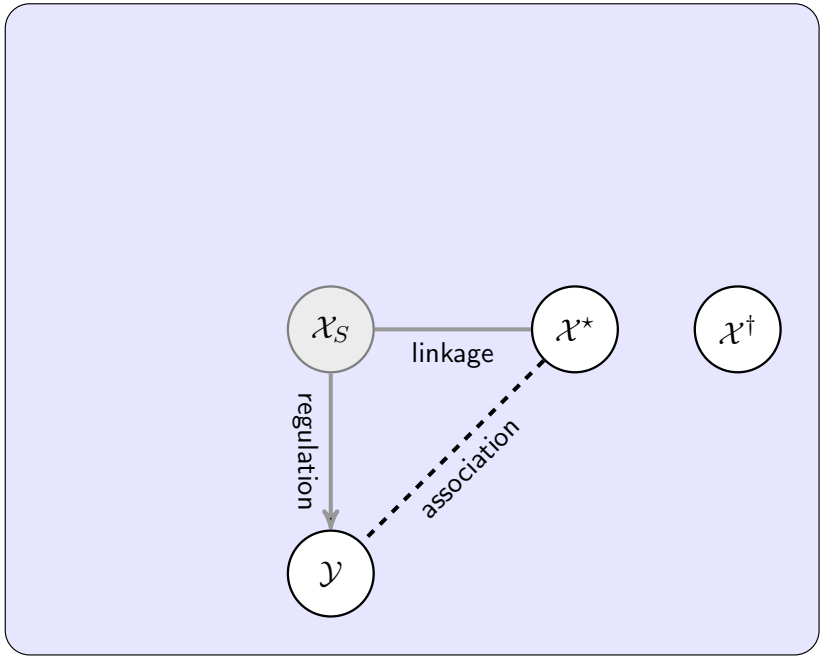
Goal:

- Identify causal loci that explain phenotypic differences.
- Use linked markers

Genome-wide association studies (GWAS)

Given:

- Genotype for multiple individuals
  - Genome-wide single nucleotide polymorphism (SNP) markers
- Phenotypes for the same individuals
  - common diseases (qualitative)
  - height, BMI, ... (quantitative)



Goal:

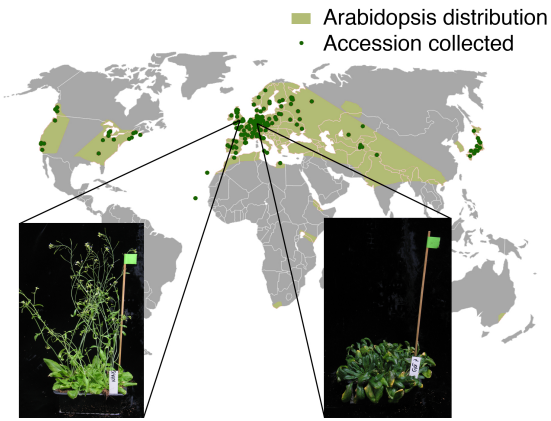
- Identify causal loci that explain phenotypic differences.
- Use linked markers

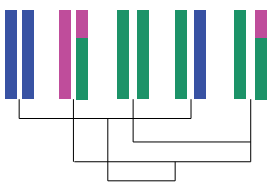
Challenges I

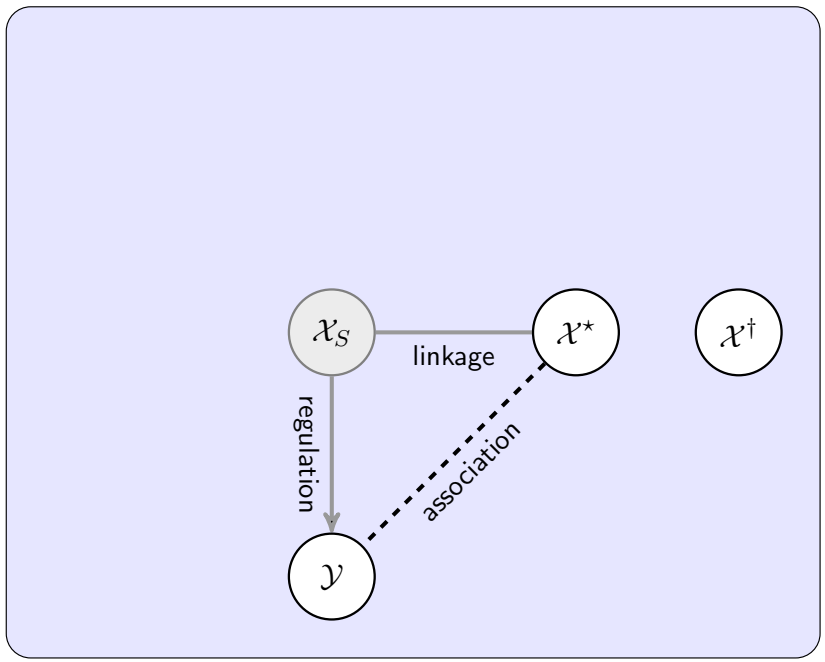
Population structure

Population-based sampling of humans, plants or animals.

- False positives due to population structure
- Take varying degrees of relatedness into account.





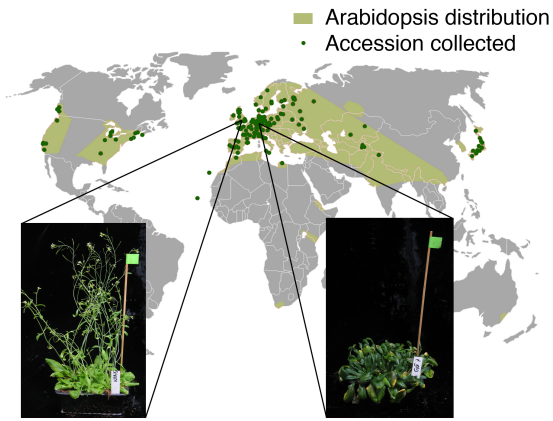


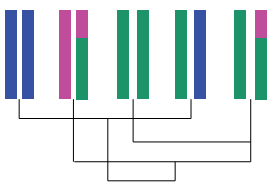
Challenges I

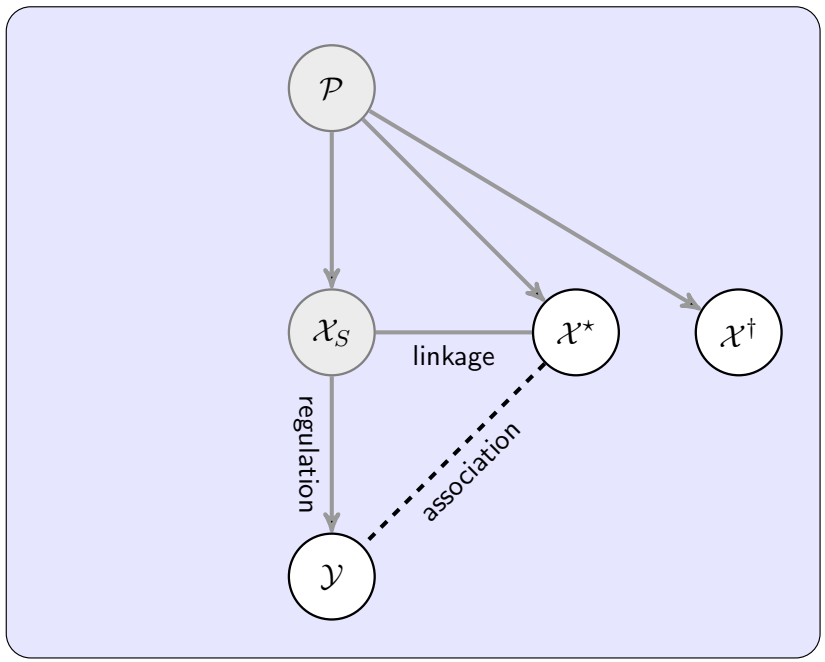
Population structure

Population-based sampling of humans, plants or animals.

- False positives due to population structure
- Take varying degrees of relatedness into account.





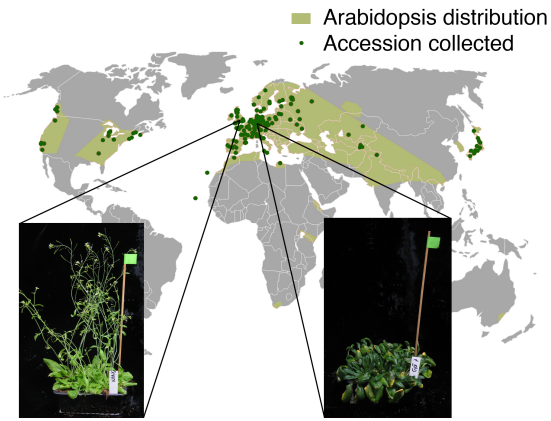


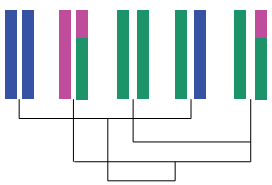
Challenges I

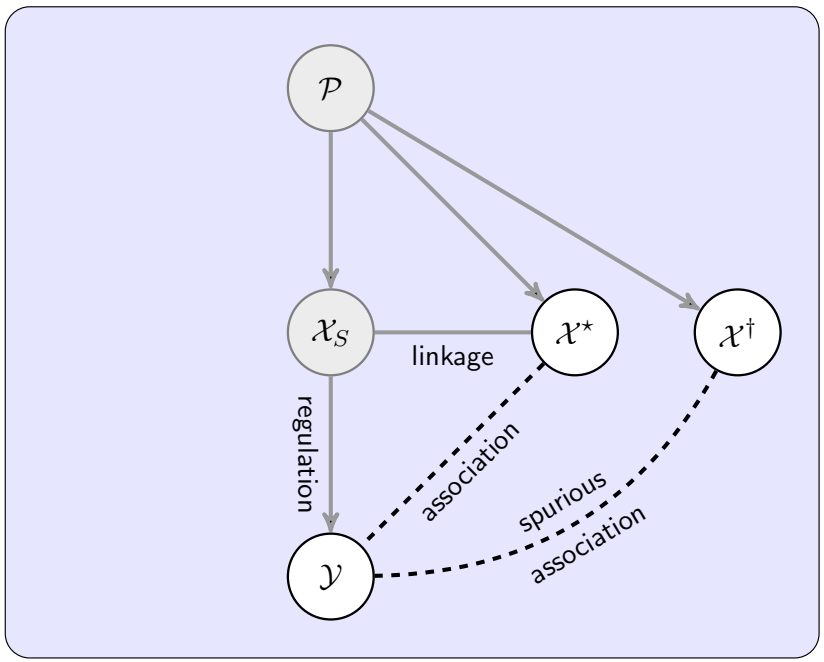
Population structure

Population-based sampling of humans, plants or animals.

- False positives due to population structure
- Take varying degrees of relatedness into account.





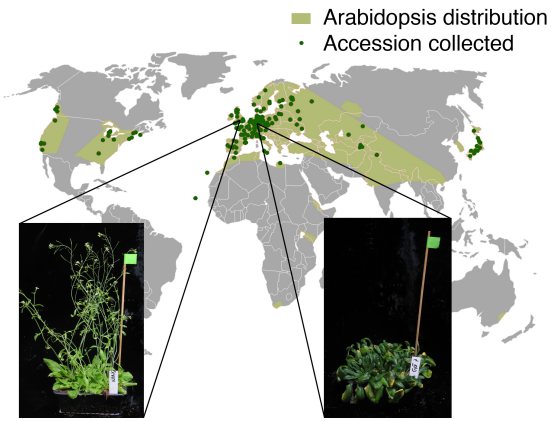


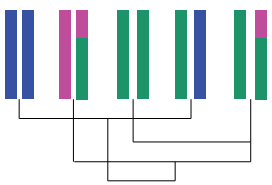
Challenges I

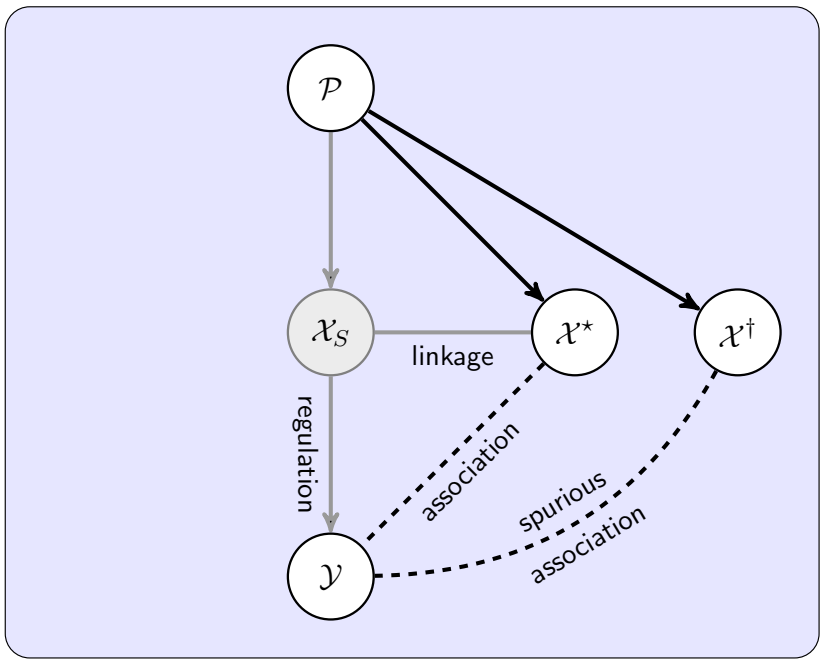
Population structure

Population-based sampling of humans, plants or animals.

- False positives due to population structure
- Take varying degrees of relatedness into account.





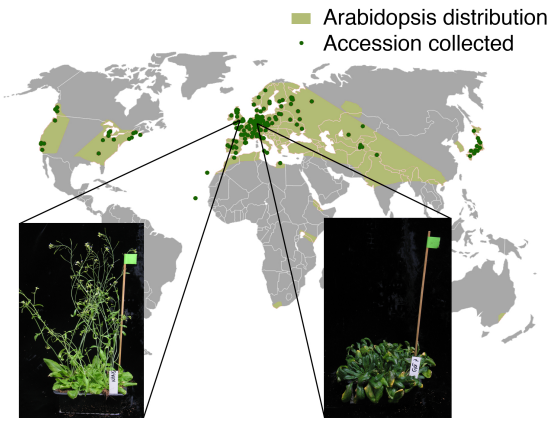


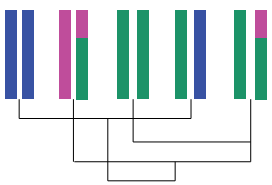
Challenges I

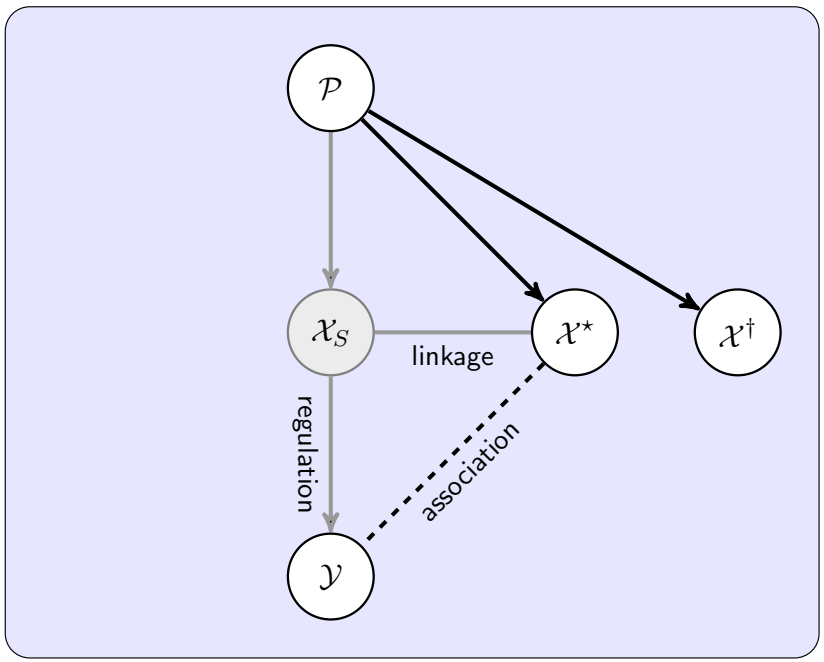
Population structure

Population-based sampling of humans, plants or animals.

- False positives due to population structure
- Take varying degrees of relatedness into account.



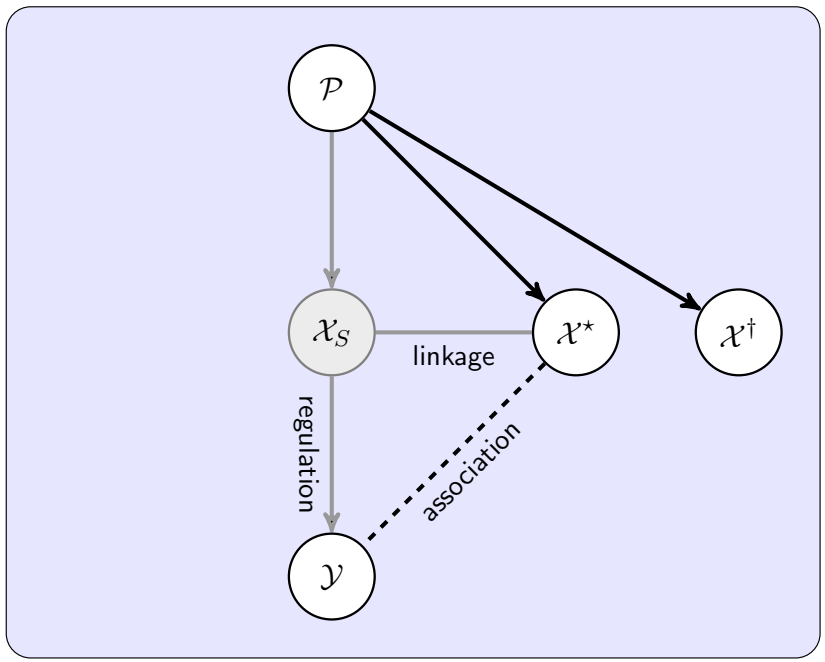




Challenges II

Study power

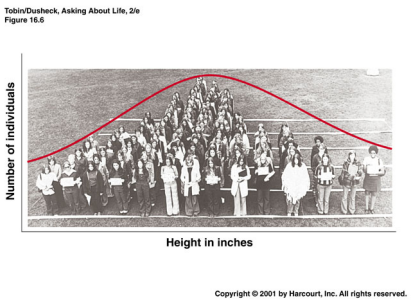
- Studies of tens of thousands of samples require efficient algorithms
- High polygenicity of quantitative traits and common diseases
  - better modeling of complex traits
  - Aggregating rare variants and effects of rare variants

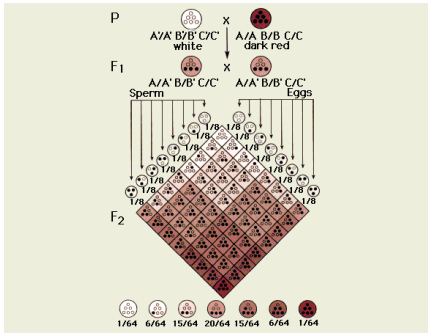


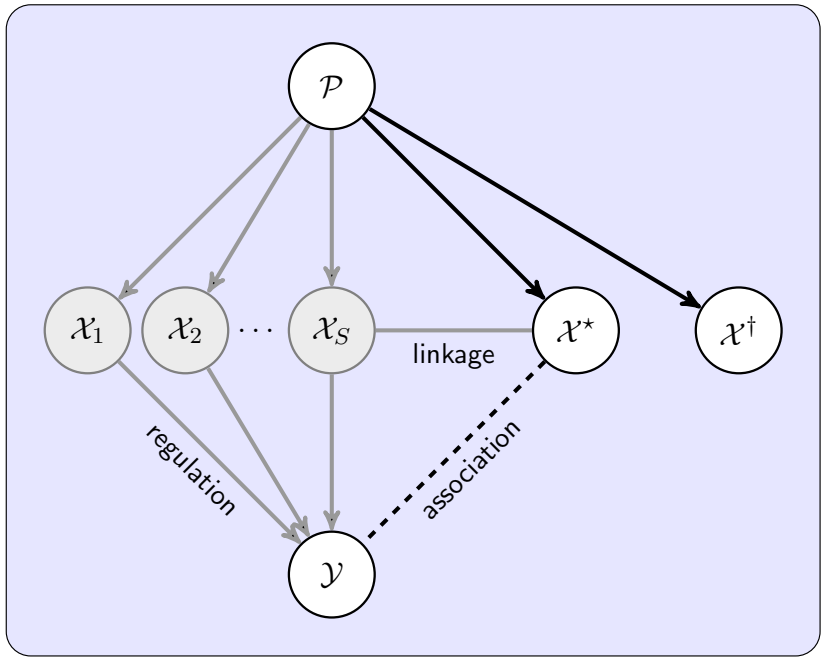
Challenges II

Study power

- Studies of tens of thousands of samples require efficient algorithms
- High polygenicity of quantitative traits and common diseases
  - better modeling of complex traits
  - Aggregating weak effects and effects of rare variants



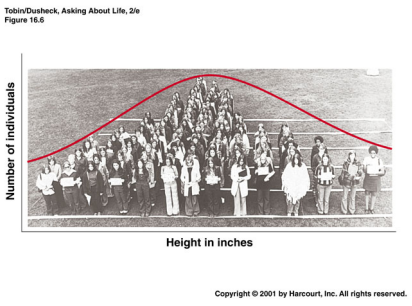


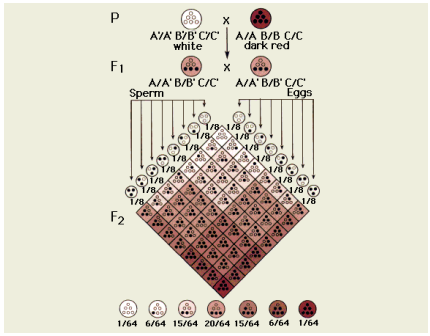


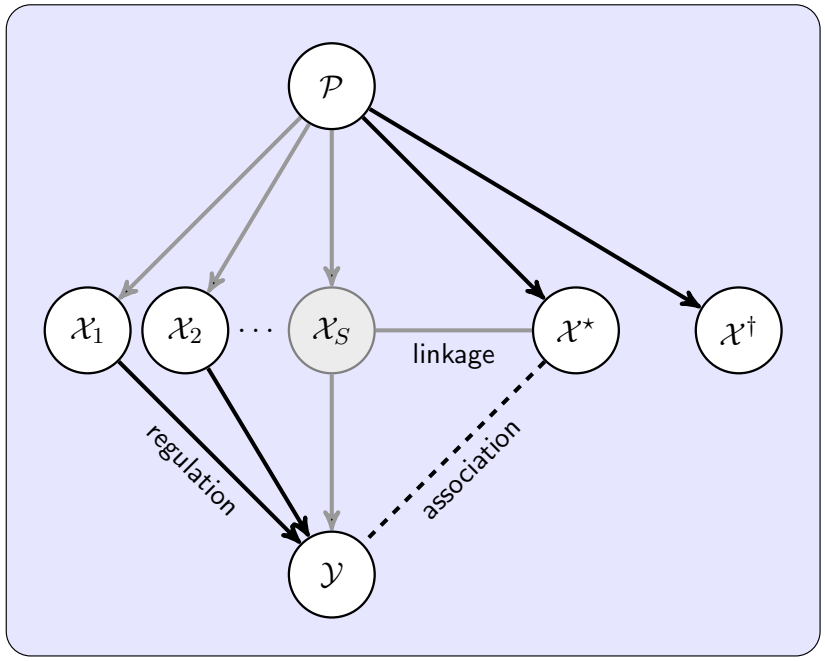
Challenges II

Study power

- Studies of tens of thousands of samples require efficient algorithms
- High polygenicity of quantitative traits and common diseases
  - better modeling of complex traits
  - Aggregating weak effects and effects of rare variants



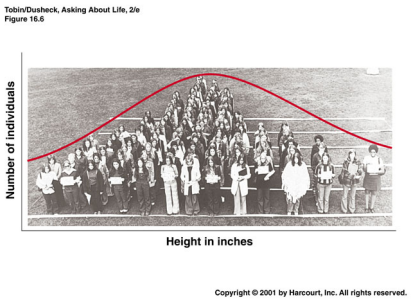


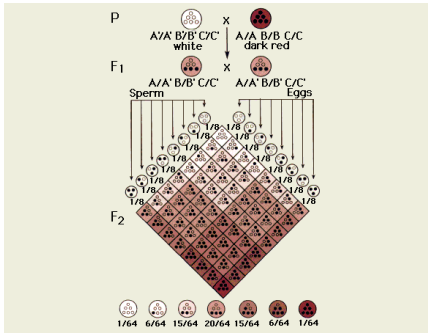


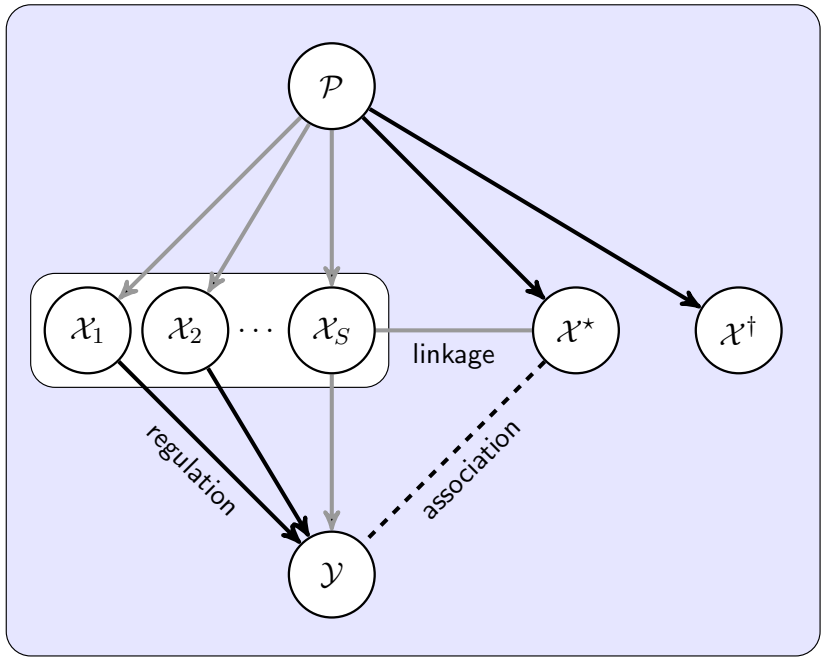
Challenges II

Study power

- Studies of tens of thousands of samples require efficient algorithms
- High polygenicity of quantitative traits and common diseases
  - better modeling of complex traits
  - Aggregating weak effects and effects of rare variants



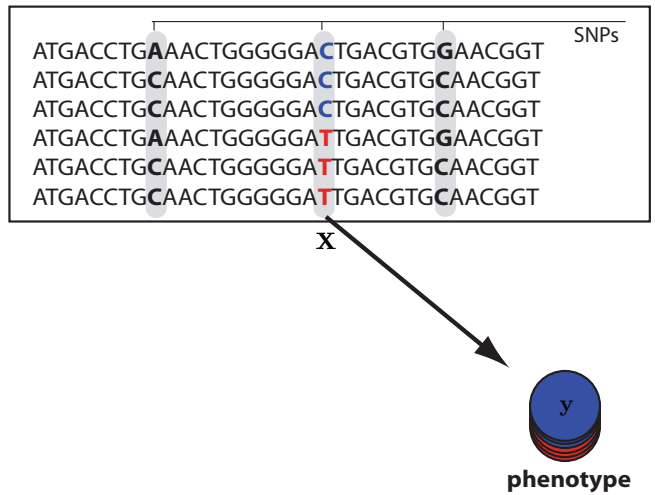




Outline

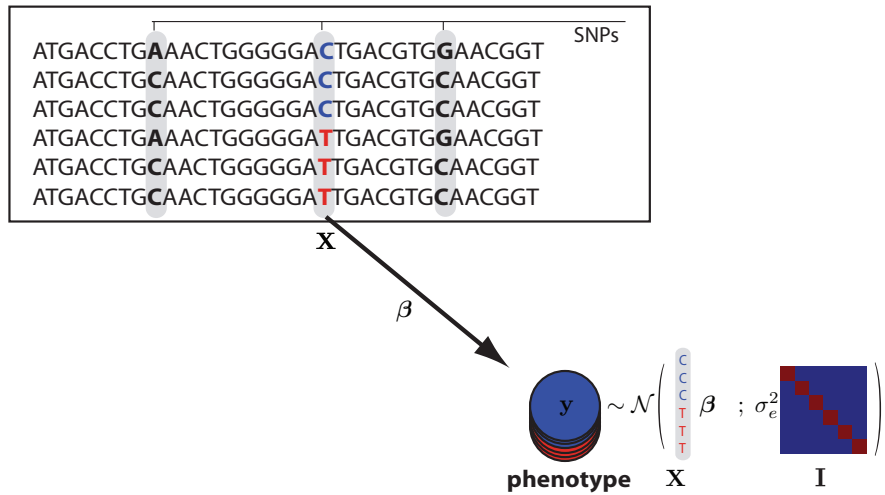
Genome wide association studies (GWAS)

- Identify associations between variable genetic loci and phenotypes.
  - Linear and logistic regression
  - Statistical dependence tests (F-test, likelihood ratio)



Genome wide association studies (GWAS)

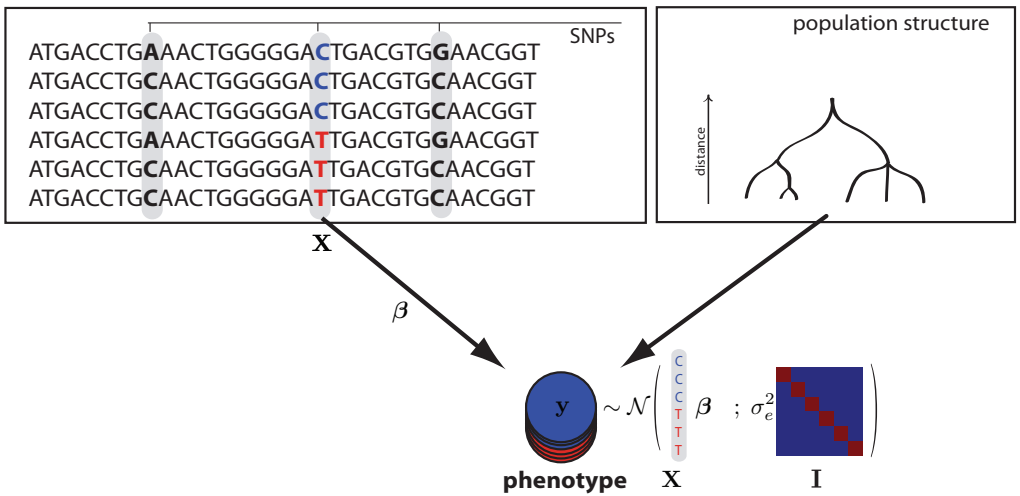
- Identify associations between variable genetic loci and phenotypes.
  - Linear and logistic regression
  - Statistical dependence tests (F-test, likelihood ratio)



$$\frac{\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}\right)}{\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma^2\boldsymbol{I}\right)}$$

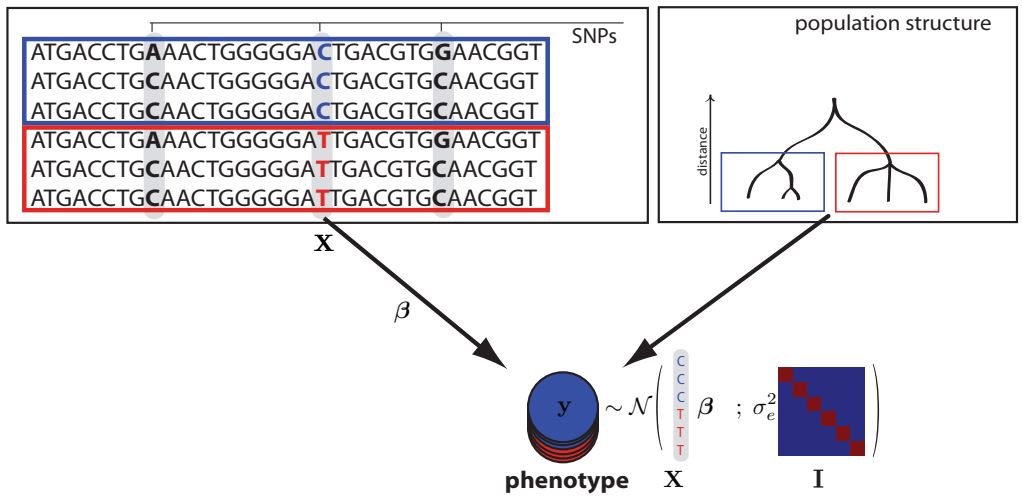
Population stratification

- Confounding structure leads to false positives.
  - Population structure
  - Family structure
  - Cryptic relatedness



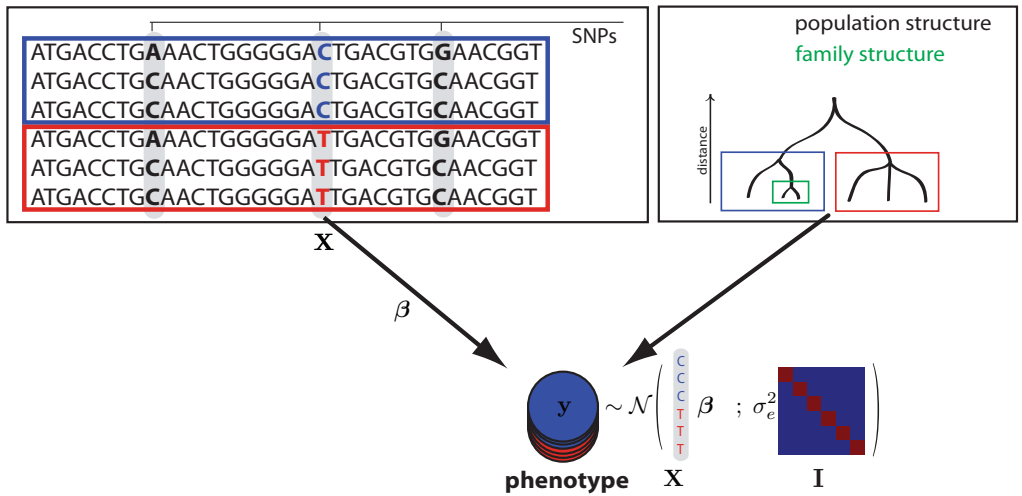
Population stratification

- Confounding structure leads to false positives.
  - Population structure
  - Family structure
  - Cryptic relatedness



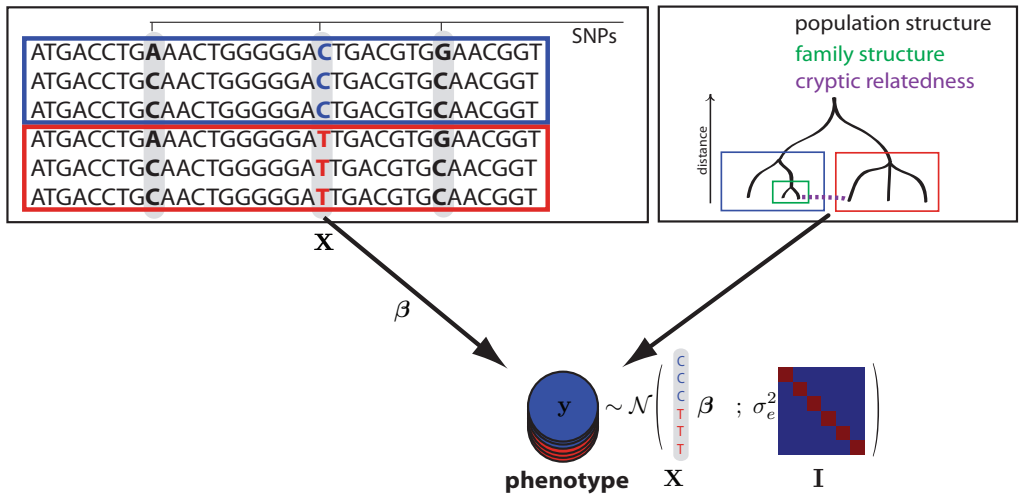
Population stratification

- Confounding structure leads to false positives.
  - Population structure
  - Family structure
  - Cryptic relatedness



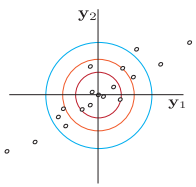
Population stratification

- Confounding structure leads to false positives.
  - Population structure
  - Family structure
  - Cryptic relatedness

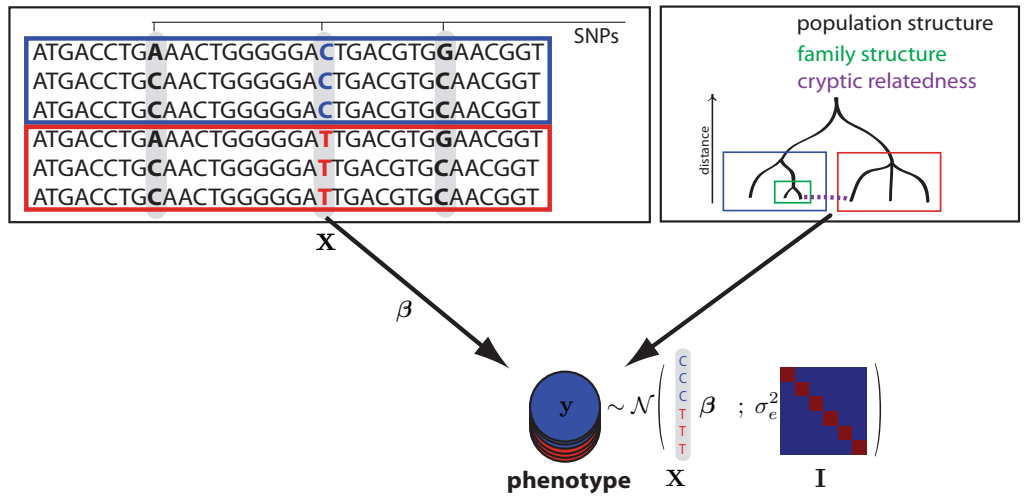


Population stratification

- Confounding structure leads to false positives.
  - Population structure
  - Family structure
  - Cryptic relatedness



$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_e^2\mathbf{I}\right)$



Population stratification

GWA on inflammatory bowel disease (WTCCC)

- 3.4k cases, 11.9k controls
- Methods
  - Linear regression
  - Likelihood ratio test

Population stratification

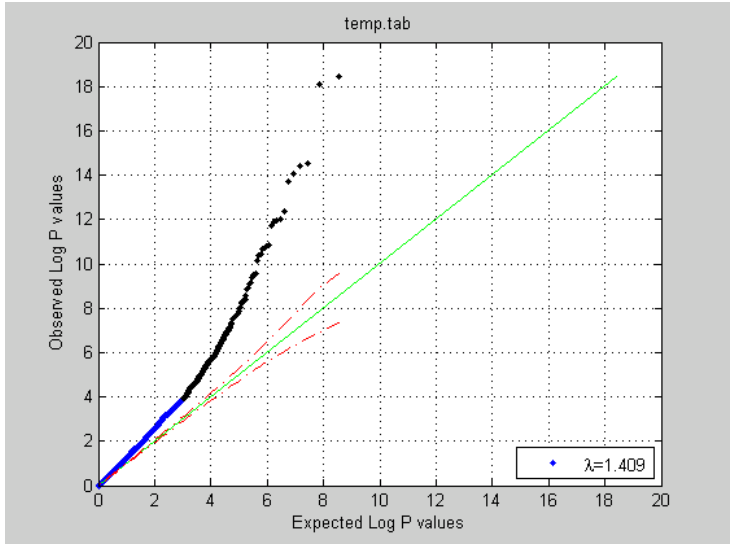
GWA on inflammatory bowel disease (WTCCC)

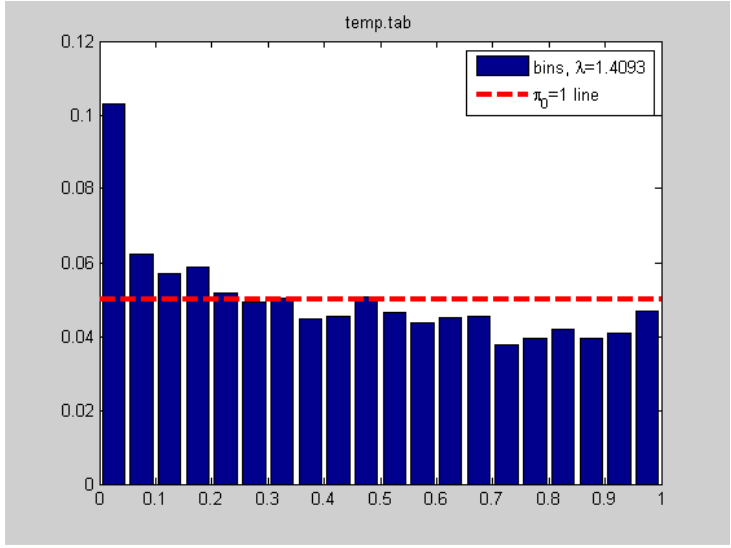
- 3.4k cases, 11.9k controls
- Methods
  - Linear regression
  - Likelihood ratio test

Population stratification

GWA on inflammatory bowel disease (WTCCC)

- 3.4k cases, 11.9k controls
- Methods
  - Linear regression
  - Likelihood ratio test





Outline

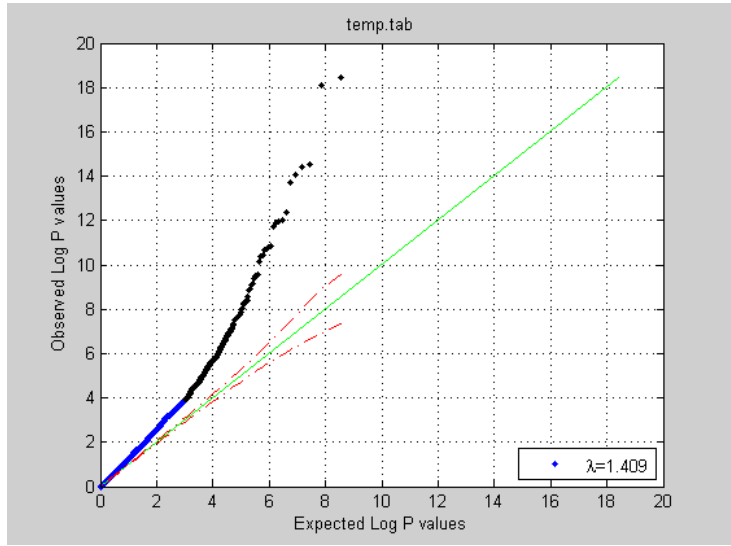
Genomic control [Devlin and Roeder, Biometrics 1999]

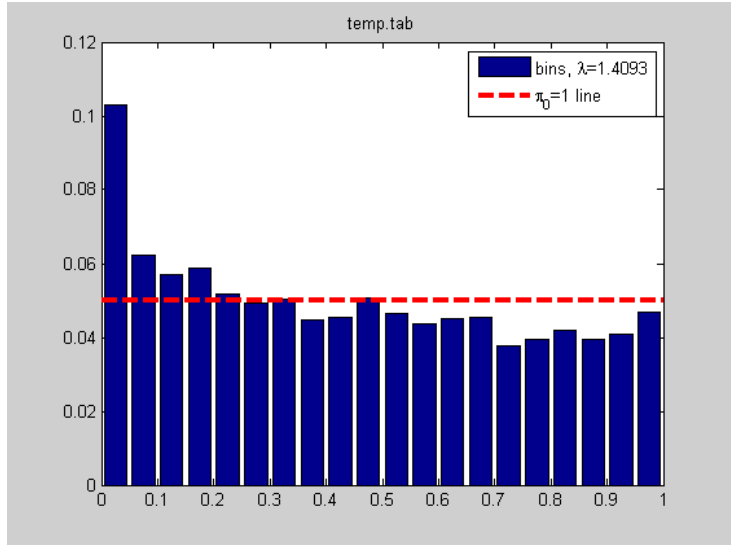
- Genomic control λ

$$\lambda = \frac{\text{median}(2LR)}{\text{median}(\chi^2)}.$$

  - $\lambda = 1$: Calibrated p-values
  - $\lambda > 1$: Inflation
  - $\lambda < 1$: Deflation

- Correct by dividing test statistic by λ.
- Applicable in combination with every method.
- Does not change (non-)uniformity of p-values.
- Very conservative.





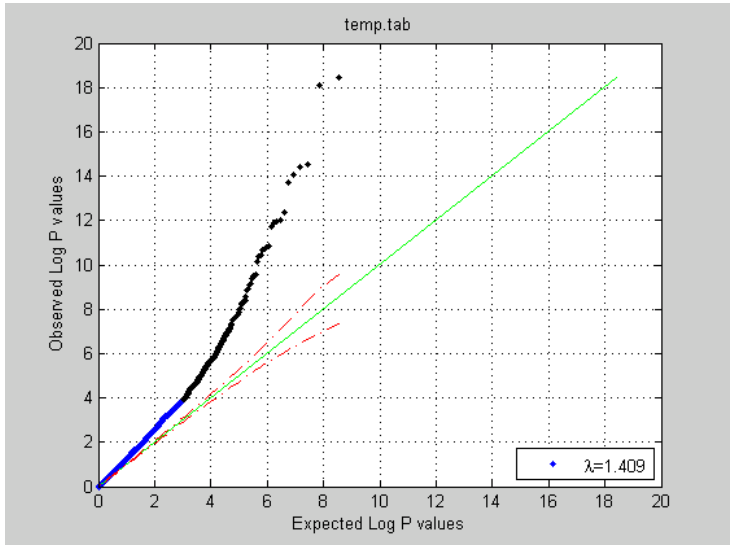
Genomic control [Devlin and Roeder, Biometrics 1999]

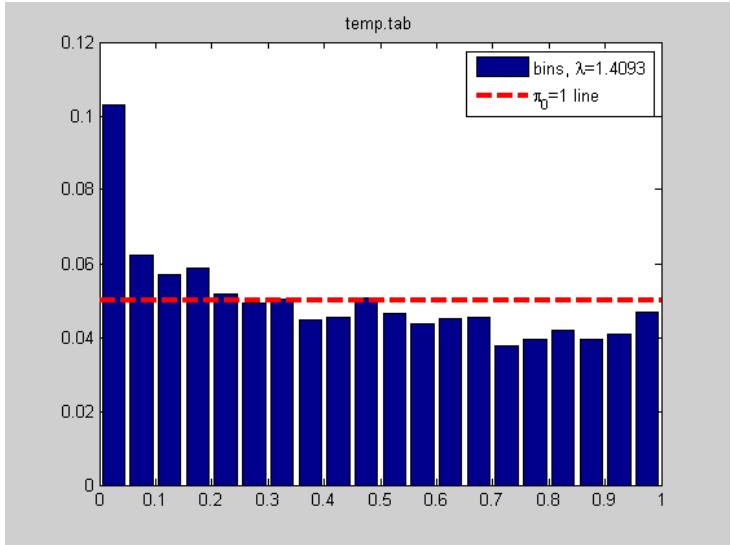
- Genomic control λ

$$\lambda = \frac{\text{median}(2LR)}{\text{median}(\chi^2)}.$$

  - $\lambda = 1$: Calibrated p-values
  - $\lambda > 1$: Inflation
  - $\lambda < 1$: Deflation

- Correct by dividing test statistic by λ.
- Applicable in combination with every method.
- Does not change (non-)uniformity of p-values.
- Very conservative.



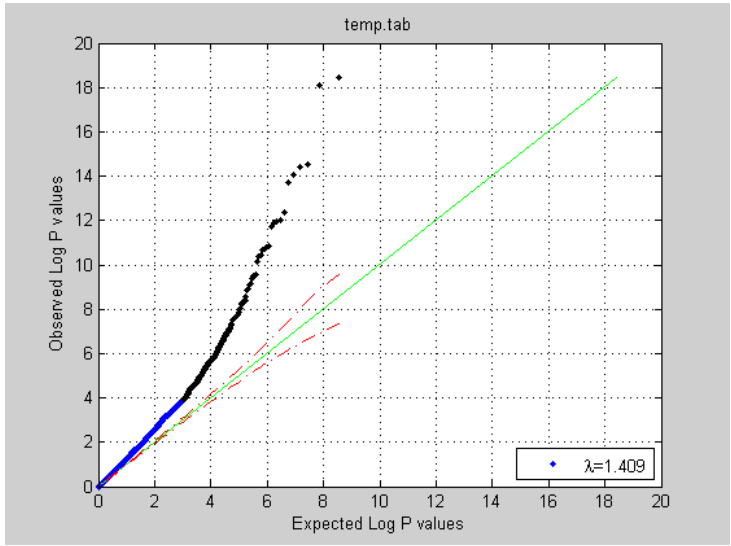


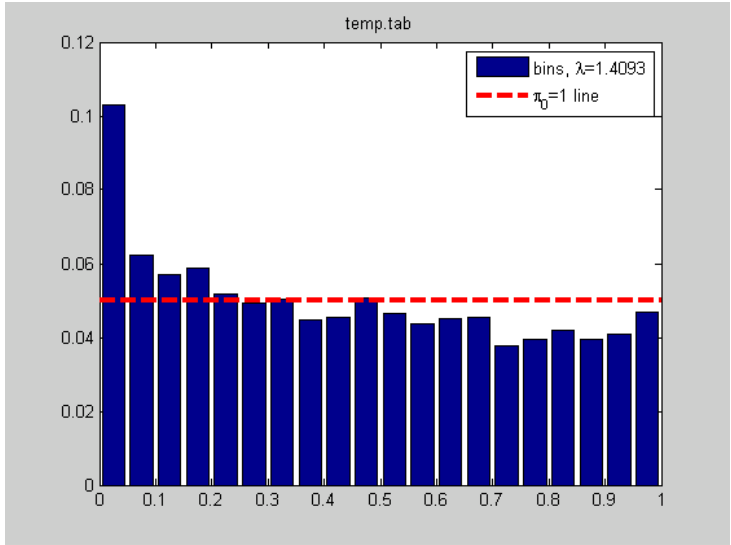
Genomic control [Devlin and Roeder, Biometrics 1999]

- Genomic control λ

$$\lambda = \frac{\text{median}(2LR)}{\text{median}(\chi^2)}.$$

  - $\lambda = 1$: Calibrated p-values
  - $\lambda > 1$: Inflation
  - $\lambda < 1$: Deflation
- Correct by dividing test statistic by λ.
- Applicable in combination with every method.
- Does not change (non-)uniformity of p-values.
- Very conservative.



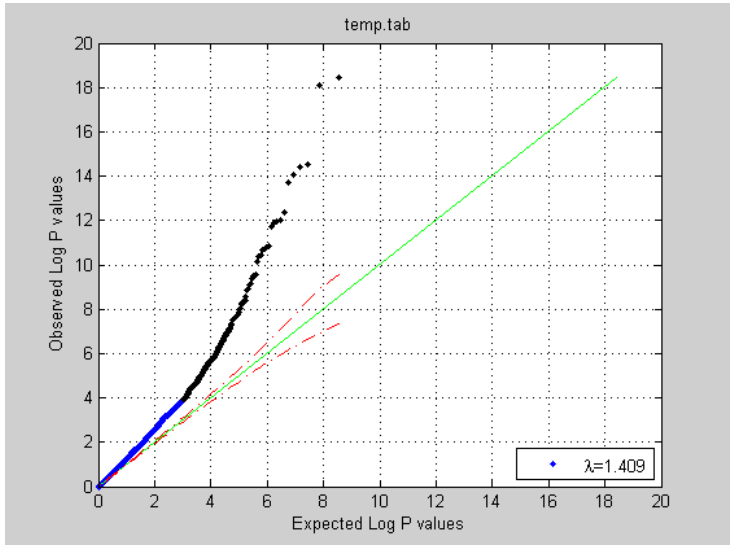


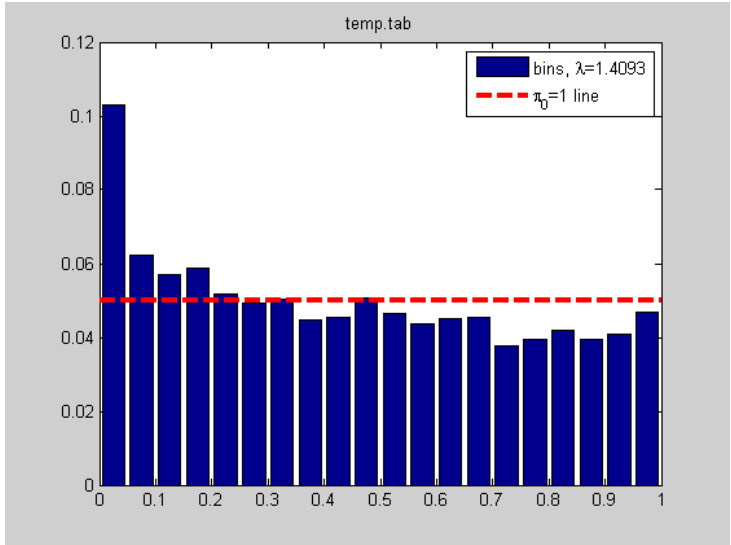
Genomic control [Devlin and Roeder, Biometrics 1999]

- Genomic control λ

$$\lambda = \frac{\text{median}(2LR)}{\text{median}(\chi^2)}.$$

  - $\lambda = 1$: Calibrated p-values
  - $\lambda > 1$: Inflation
  - $\lambda < 1$: Deflation
- Correct by dividing test statistic by λ.
- Applicable in combination with every method.
- Does not change (non-)uniformity of p-values.
- Very conservative.



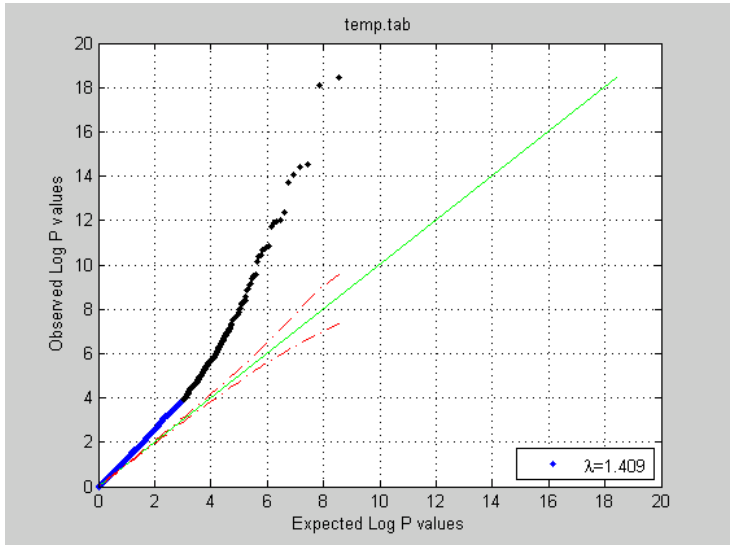


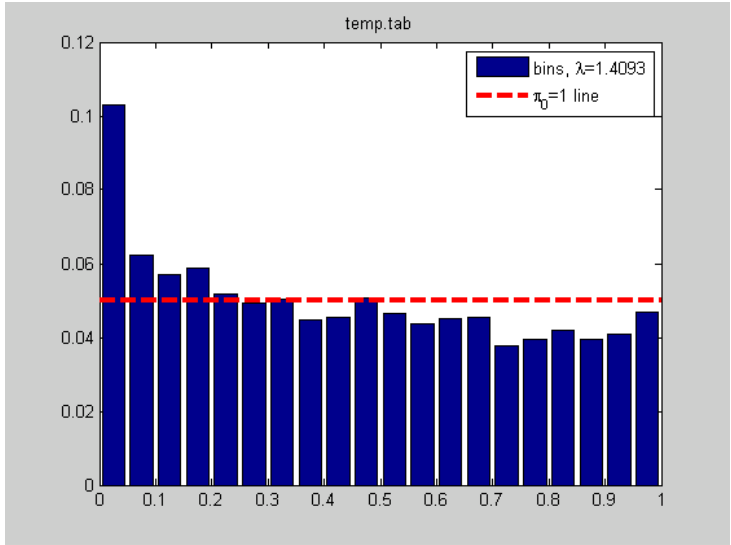
Genomic control [Devlin and Roeder, Biometrics 1999]

- Genomic control λ

$$\lambda = \frac{\text{median}(2LR)}{\text{median}(\chi^2)}.$$

  - $\lambda = 1$: Calibrated p-values
  - $\lambda > 1$: Inflation
  - $\lambda < 1$: Deflation
- Correct by dividing test statistic by λ.
- Applicable in combination with every method.
- Does not change (non-)uniformity of p-values.
- Very conservative.



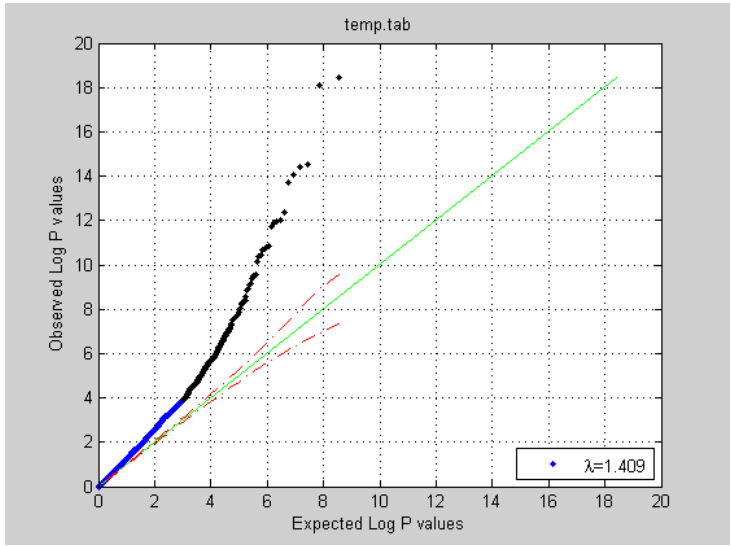


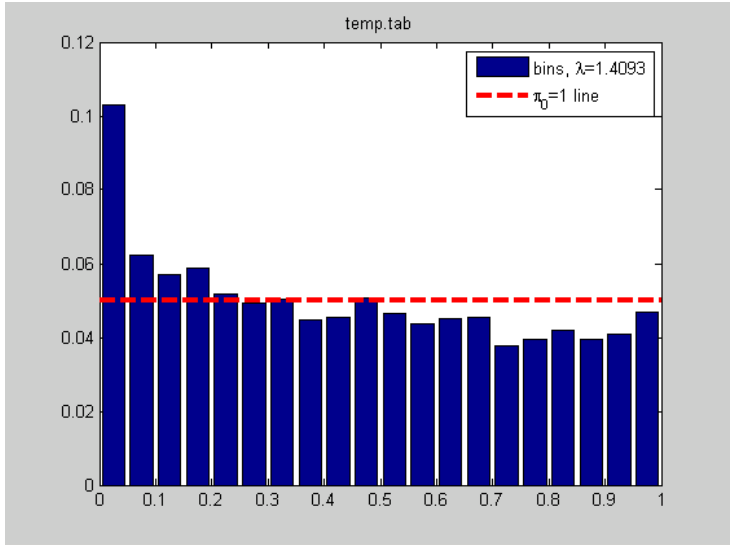
Genomic control [Devlin and Roeder, Biometrics 1999]

- Genomic control λ

$$\lambda = \frac{\text{median}(2LR)}{\text{median}(\chi^2)}.$$

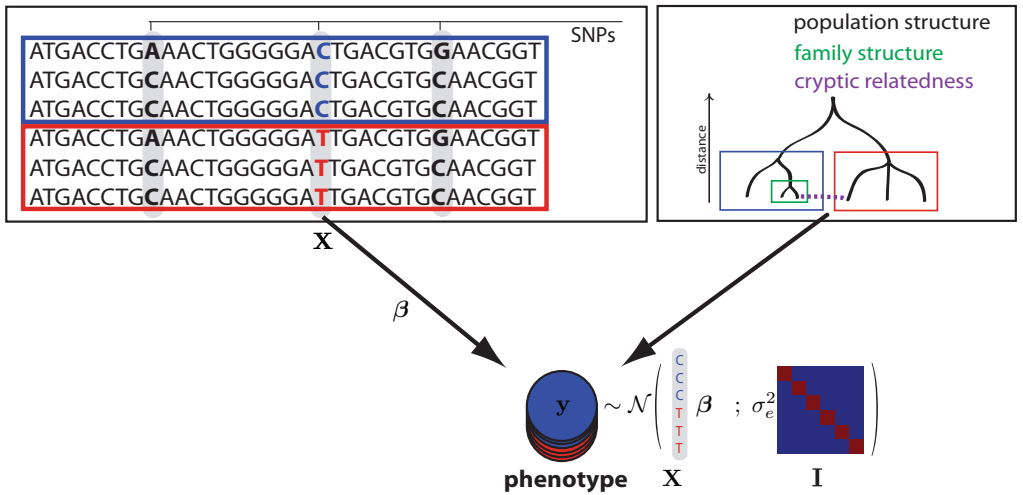
  - $\lambda = 1$: Calibrated p-values
  - $\lambda > 1$: Inflation
  - $\lambda < 1$: Deflation
- Correct by dividing test statistic by λ.
- Applicable in combination with every method.
- Does not change (non-)uniformity of p-values.
- Very conservative.





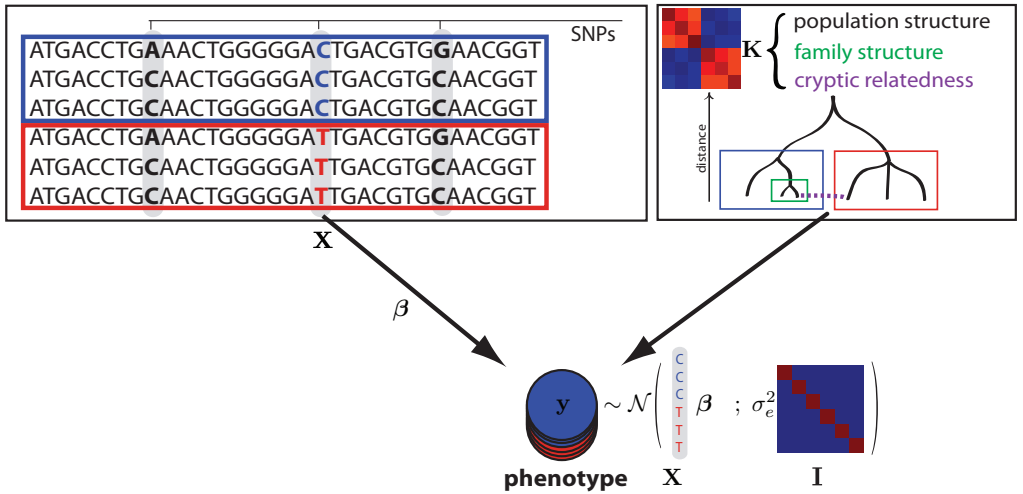
Linear mixed models (LMM)

- Kernel matrix $\boldsymbol{K}$
  - Estimated from SNP data
  - Identity by state
  - Identity by descent
  - Covariance
- Sample random effect u.
- Sample phenotype y.



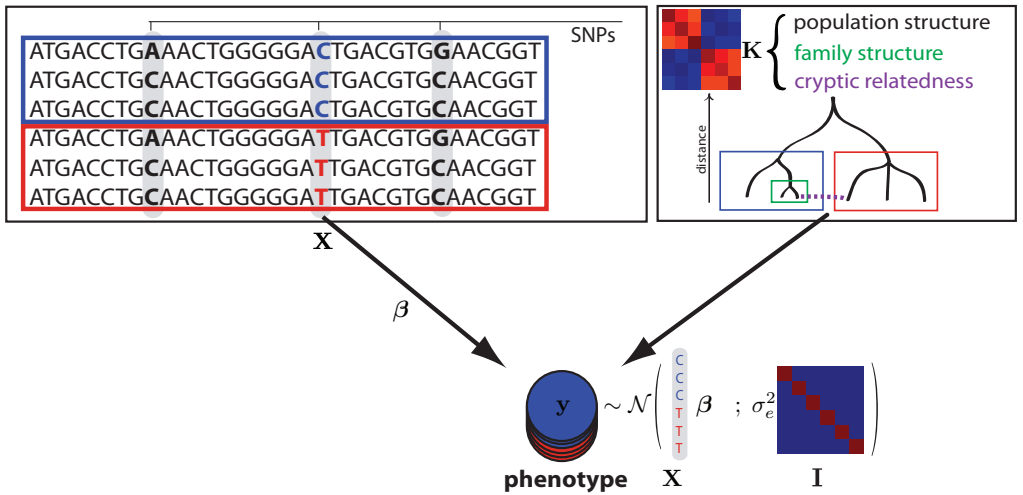
Linear mixed models (LMM)

- Kernel matrix $\boldsymbol{K}$
  - Estimated from SNP data
  - Identity by state
  - Identity by descent
  - Covariance
- Sample random effect u.
- Sample phenotype y.



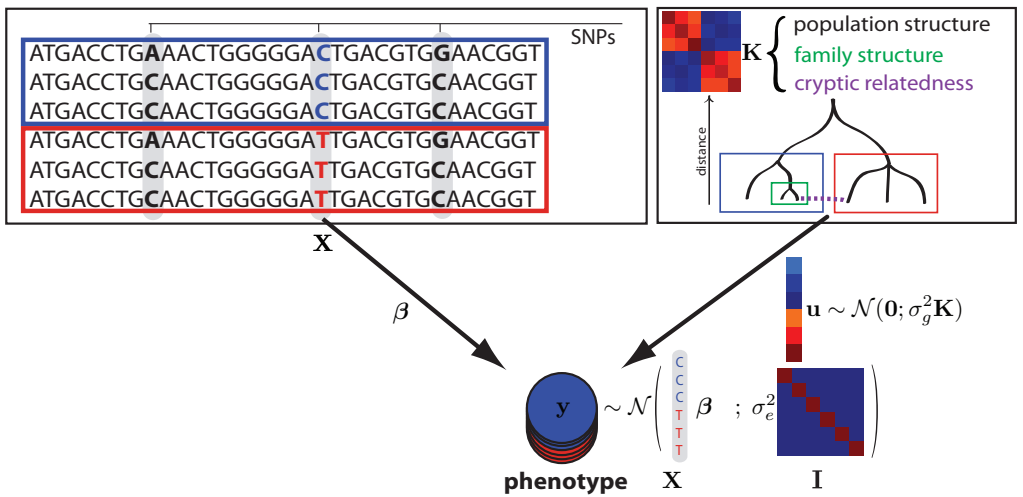
Linear mixed models (LMM)

- Kernel matrix $\boldsymbol{K}$
  - Estimated from SNP data
  - Identity by state
  - Identity by descent
  - Covariance
- Sample random effect u.
- Sample phenotype y.



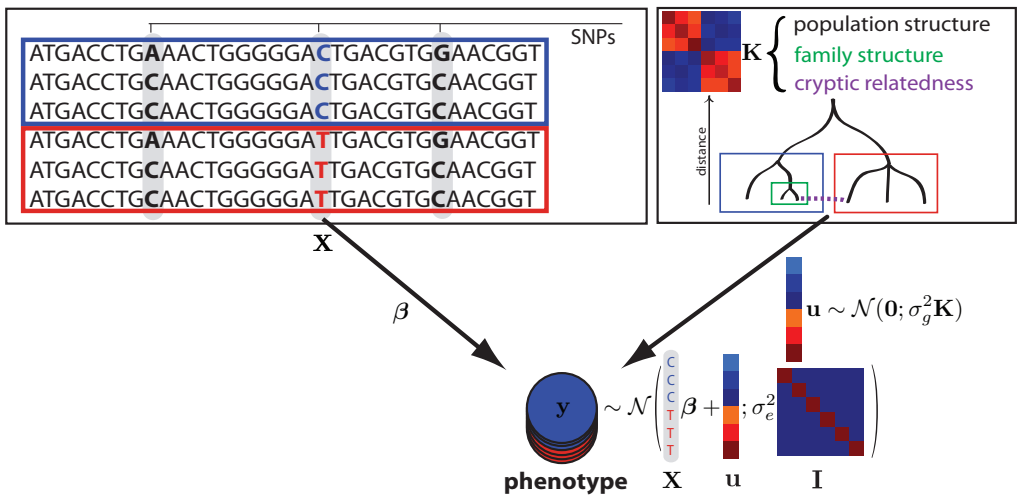
Linear mixed models (LMM)

- Kernel matrix $\boldsymbol{K}$
  - Estimated from SNP data
  - Identity by state
  - Identity by descent
  - Covariance
- Sample random effect $\boldsymbol{u}$.
- Sample phenotype $\boldsymbol{y}$.



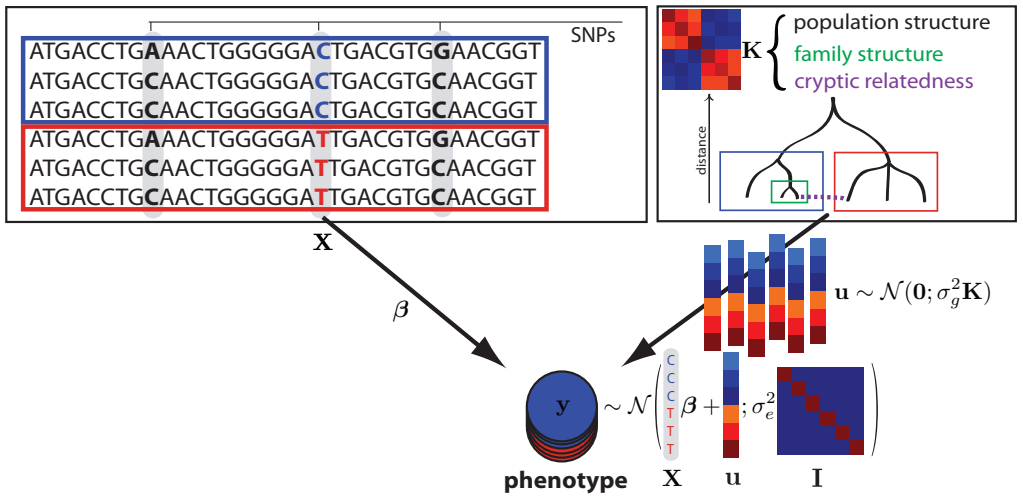
Linear mixed models (LMM)

- Kernel matrix $\boldsymbol{K}$
  - Estimated from SNP data
  - Identity by state
  - Identity by descent
  - Covariance
- Sample random effect $\boldsymbol{u}$.
- Sample phenotype $\boldsymbol{y}$.



Linear mixed models (LMM)

- Kernel matrix $\boldsymbol{K}$
  - Estimated from SNP data
  - Identity by state
  - Identity by descent
  - Covariance
- Sample random effect $\boldsymbol{u}$.
- Sample phenotype $\boldsymbol{y}$.

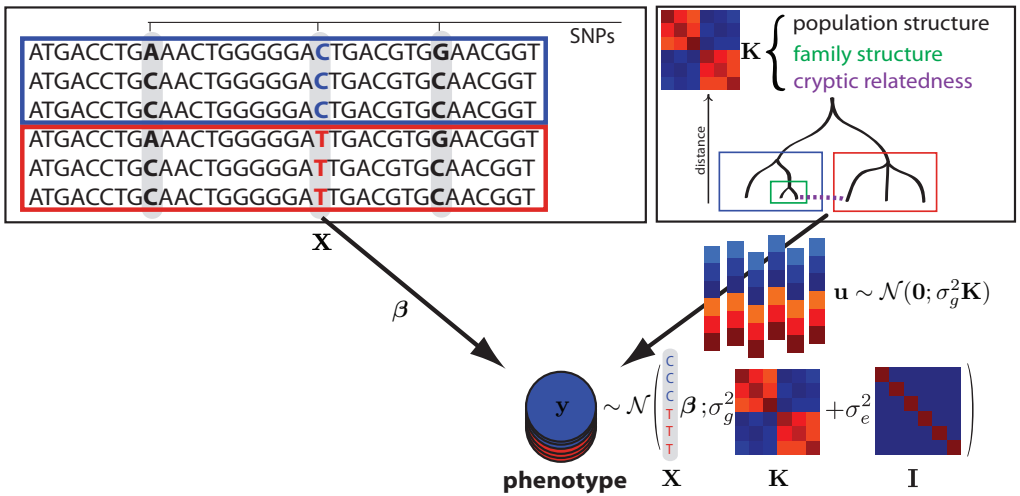


$$\int_{\boldsymbol{u}} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \sigma^2 \boldsymbol{I}\right) \mathcal{N}\left(\boldsymbol{u} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{K}\right)$$

Linear mixed models (LMM)

- Kernel matrix $\boldsymbol{K}$
  - Estimated from SNP data
  - Identity by state
  - Identity by descent
  - Covariance
- Sample random effect $\boldsymbol{u}$.
- Sample phenotype $\boldsymbol{y}$.

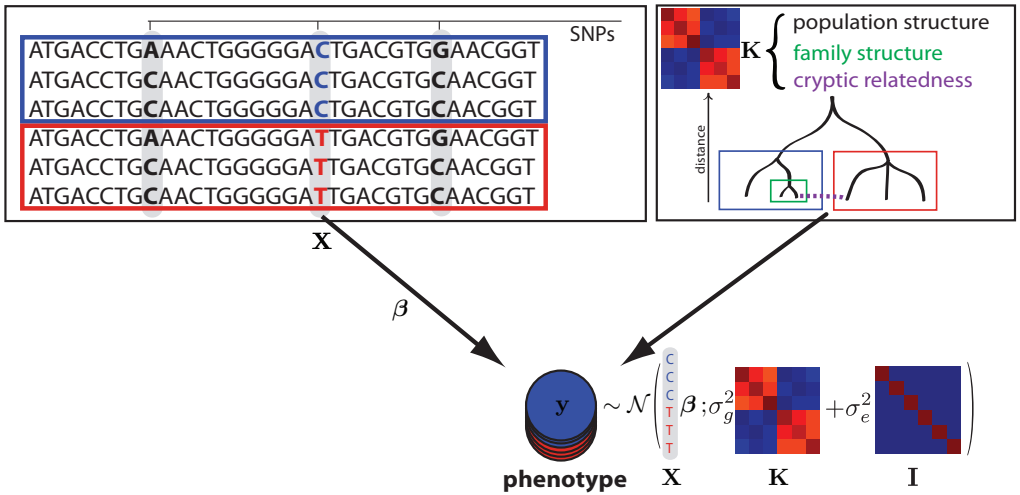
$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2\boldsymbol{K} + \sigma^2\boldsymbol{I}\right)$$



Linear mixed models (LMM)

- Corrects for all levels of population structure.
- ML estimation is computationally demanding
  - Non-convex in σ_g^2 and σ^2

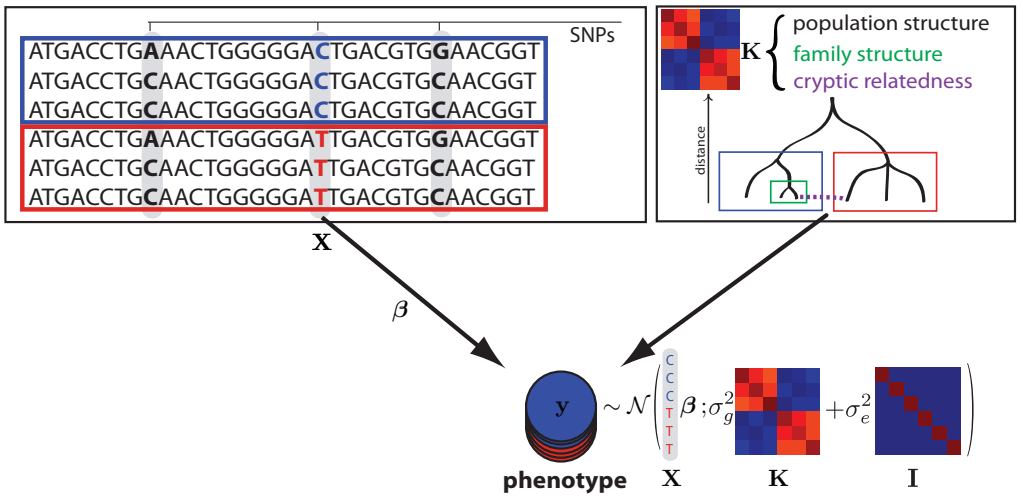
$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2\boldsymbol{K} + \sigma^2\boldsymbol{I}\right)$$

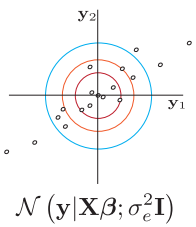


Linear mixed models (LMM)

- Corrects for all levels of population structure.
- ML estimation is computationally demanding
  - Non-convex in σ_g^2 and σ_e^2

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2\boldsymbol{K} + \sigma^2\boldsymbol{I}\right)$$

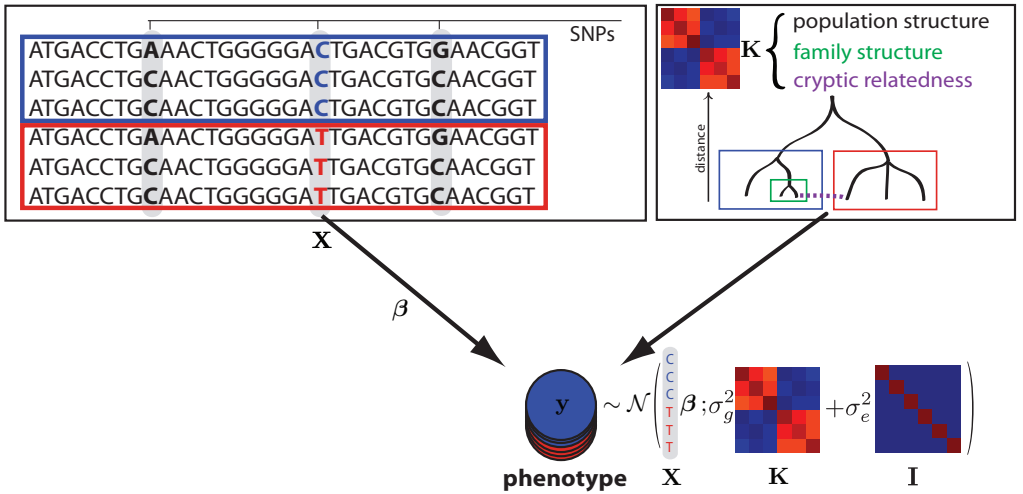


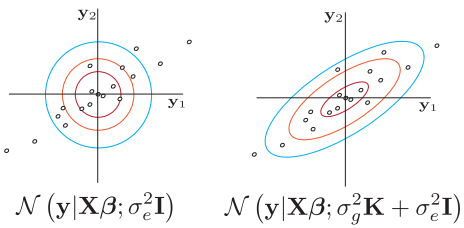


Linear mixed models (LMM)

- Corrects for all levels of population structure.
- ML estimation is computationally demanding
  - Non-convex in σ_g^2 and σ^2.

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2\boldsymbol{K} + \sigma^2\boldsymbol{I}\right)$$

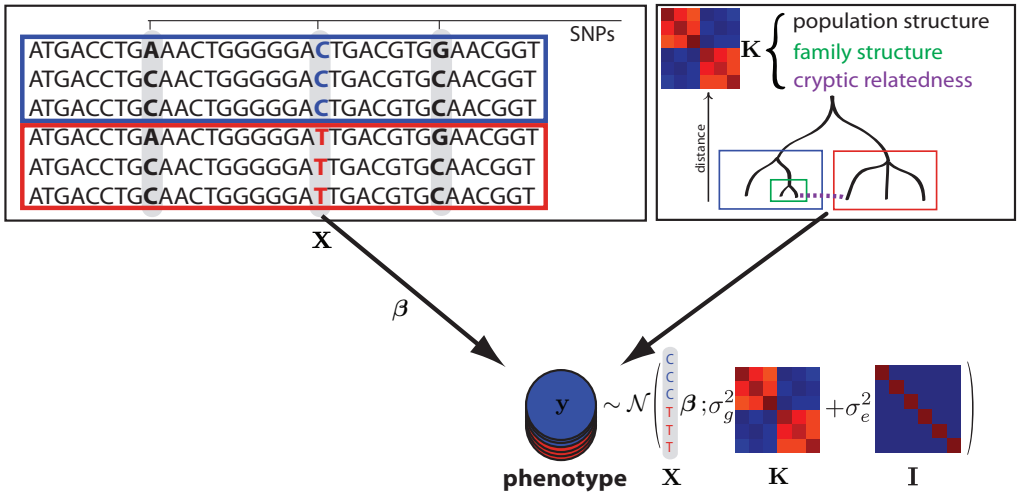


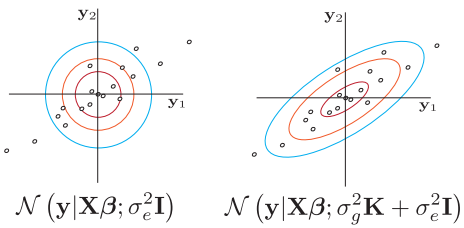


Linear mixed models (LMM)

- Corrects for all levels of population structure.
- ML estimation is computationally demanding
  - Non-convex in σ_g^2 and σ^2.

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_g^2 \boldsymbol{K} + \sigma^2 \boldsymbol{I}\right)$$

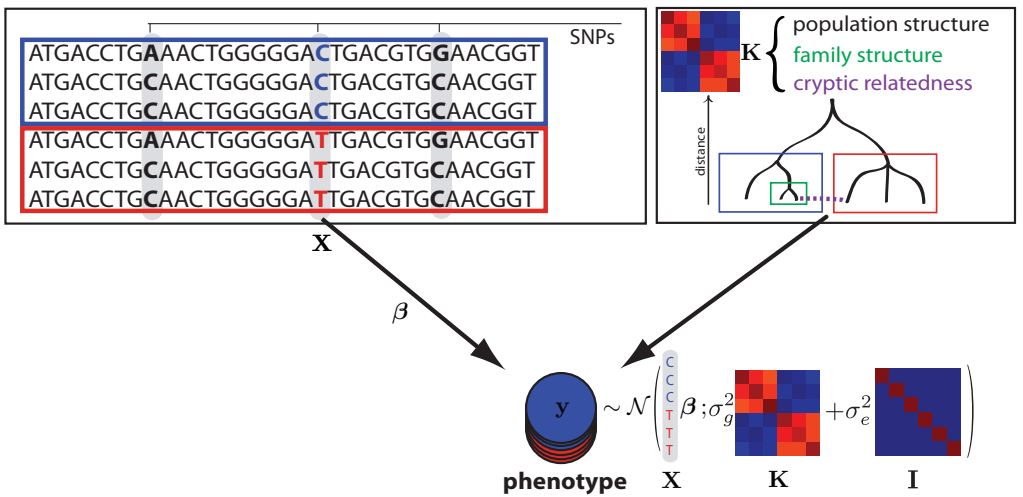


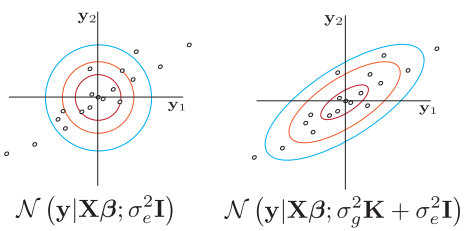


Linear mixed models (LMM)

- Corrects for all levels of population structure.
- ML estimation is computationally demanding
  - Non-convex in σ_{g}^2 and σ^2.

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\text{g}}^2\boldsymbol{K} + \sigma^2\boldsymbol{I}\right)$$





Linear mixed models (LMM)

EMMA [Kang et al., Genetics 2008]

- LMM log likelihood

$$LL(\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2, \sigma^2) = \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \boldsymbol{K} + \sigma^2 \boldsymbol{I}\right).$$

- Change of variables, introducing $\delta = \sigma^2 / \sigma_{\mathrm{g}}^2$:

$$LL(\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2, \delta) = \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)\right).$$

- ML-parameters $\boldsymbol{\beta}$ and σ_{g}^2 follow in closed form.
- Use optimizer to solve 1-dimensional optimization problem over δ.
- $O(N^3)$ per SNP.

Linear mixed models (LMM)

EMMA [Kang et al., Genetics 2008]

- LMM log likelihood

$$LL(\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2, \sigma^2) = \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \boldsymbol{K} + \sigma^2 \boldsymbol{I}\right).$$

- Change of variables, introducing $\delta = \sigma^2 / \sigma_{\mathrm{g}}^2$:

$$LL(\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2, \delta) = \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)\right).$$

- ML-parameters β and σ_{g}^2 follow in closed form.
- Use optimizer to solve 1-dimensional optimization problem over δ.
- $O(N^3)$ per SNP.

Linear mixed models (LMM)

EMMA [Kang et al., Genetics 2008]

- LMM log likelihood

$$LL(\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2, \sigma^2) = \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \boldsymbol{K} + \sigma^2 \boldsymbol{I}\right).$$

- Change of variables, introducing $\delta = \sigma^2 / \sigma_{\mathrm{g}}^2$:

$$LL(\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2, \delta) = \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)\right).$$

- ML-parameters $\boldsymbol{\beta}$ and σ_{g}^2 follow in closed form.
- Use optimizer to solve 1-dimensional optimization problem over δ.
- $O(N^3)$ per SNP.

Linear mixed models (LMM)

EMMA [Kang et al., Genetics 2008]

- LMM log likelihood

$$LL(\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2, \sigma^2) = \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \boldsymbol{K} + \sigma^2 \boldsymbol{I}\right).$$

- Change of variables, introducing $\delta = \sigma^2 / \sigma_{\mathrm{g}}^2$:

$$LL(\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2, \delta) = \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)\right).$$

- ML-parameters $\boldsymbol{\beta}$ and σ_{g}^2 follow in closed form.
- Use optimizer to solve 1-dimensional optimization problem over δ.
- $O(N^3)$ per SNP.

Linear mixed models (LMM)

EMMA [Kang et al., Genetics 2008]

- LMM log likelihood

$$LL(\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2, \sigma^2) = \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \boldsymbol{K} + \sigma^2 \boldsymbol{I}\right).$$

- Change of variables, introducing $\delta = \sigma^2 / \sigma_{\mathrm{g}}^2$:

$$LL(\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2, \delta) = \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)\right).$$

- ML-parameters $\boldsymbol{\beta}$ and σ_{g}^2 follow in closed form.
- Use optimizer to solve 1-dimensional optimization problem over δ.
- $O(N^3)$ per SNP.

Linear mixed models (LMM)

ML parameters

Gradient of the LMM log likelihood w.r.t. β

Linear mixed models (LMM)

ML parameters

Gradient of the LMM log likelihood w.r.t. $\boldsymbol{\beta}$

$$\begin{aligned}\nabla_{\boldsymbol{\beta}} \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)\right) &= \nabla_{\boldsymbol{\beta}} - \frac{1}{2\sigma_{\mathrm{g}}^2} \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^{\top} \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right) \\ &= \frac{1}{\sigma_{\mathrm{g}}^2} \left[-\boldsymbol{X}^{\top} \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{y} + \boldsymbol{X}^{\top} \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{X}\right]\end{aligned}$$

Linear mixed models (LMM)

ML parameters

Gradient of the LMM log likelihood w.r.t. $\boldsymbol{\beta}$

$$\begin{aligned}\nabla_{\boldsymbol{\beta}} \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)\right) &= \nabla_{\boldsymbol{\beta}} - \frac{1}{2\sigma_{\mathrm{g}}^2} \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right) \\ &= \frac{1}{\sigma_{\mathrm{g}}^2} \left[-\boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{y} + \boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{X}\right]\end{aligned}$$

set gradient to zero:

Linear mixed models (LMM)

ML parameters

Gradient of the LMM log likelihood w.r.t. $\boldsymbol{\beta}$

$$\begin{aligned}\nabla_{\boldsymbol{\beta}} \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)\right) &= \nabla_{\boldsymbol{\beta}} - \frac{1}{2\sigma_{\mathrm{g}}^2} \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right) \\ &= \frac{1}{\sigma_{\mathrm{g}}^2} \left[-\boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{y} + \boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{X}\right]\end{aligned}$$

set gradient to zero:

$$\begin{aligned}\boldsymbol{0} &= \frac{1}{\sigma_{\mathrm{g}}^2} \left[\boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{y} - \boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{X}\boldsymbol{\beta}\right] \\ \boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{X}\boldsymbol{\beta} &= \boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{y} \\ \boldsymbol{\beta}_{\mathrm{ML}} &= \left(\boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{y}\end{aligned}$$

Linear mixed models (LMM)

ML parameters

Gradient of the LMM log likelihood w.r.t. $\boldsymbol{\beta}$

$$\begin{aligned}\nabla_{\boldsymbol{\beta}} \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)\right) &= \nabla_{\boldsymbol{\beta}} - \frac{1}{2\sigma_{\mathrm{g}}^2} \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right) \\ &= \frac{1}{\sigma_{\mathrm{g}}^2} \left[-\boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{y} + \boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{X}\right]\end{aligned}$$

set gradient to zero:

$$\begin{aligned}\boldsymbol{0} &= \frac{1}{\sigma_{\mathrm{g}}^2} \left[\boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{y} - \boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{X}\boldsymbol{\beta}\right] \\ \boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{X}\boldsymbol{\beta} &= \boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{y} \\ \boldsymbol{\beta}_{\mathrm{ML}} &= \left(\boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1} \boldsymbol{y}\end{aligned}$$

Note that this solution is analogous to the ML solution of the linear regression $\left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}$.

Linear mixed models (LMM)

ML parameters

Derivative of the LMM log likelihood w.r.t. σ_{g}^2

- Note that For every SNP we need to calculate $(\boldsymbol{K} + \delta \boldsymbol{I})^{-1}$, which is an $O(N^3)$ operation.

Linear mixed models (LMM)

ML parameters

Derivative of the LMM log likelihood w.r.t. σ_{g}^2

$$\begin{aligned} & \mathrm{d}\sigma_{\mathrm{g}}^2 \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2\left(\boldsymbol{K}+\delta \boldsymbol{I}\right)\right) \\ = & -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2}-\frac{N}{\sigma_g^4}\left(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\right)^{\top}\left(\boldsymbol{K}+\delta \boldsymbol{I}\right)^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\right)\right] \end{aligned}$$

- Note that For every SNP we need to calculate $(\boldsymbol{K}+\delta \boldsymbol{I})^{-1}$, which is an $O(N^3)$ operation.

Linear mixed models (LMM)

ML parameters

Derivative of the LMM log likelihood w.r.t. σ_{g}^2

$$\mathrm{d}\sigma_{\mathrm{g}}^2 \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)\right)$$
$$= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)\right]$$

set derivative to zero:

- Note that For every SNP we need to calculate $\left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1}$, which is an $O(N^3)$ operation.

Linear mixed models (LMM)

ML parameters

Derivative of the LMM log likelihood w.r.t. σ_{g}^2

$$\mathrm{d}\sigma_{\mathrm{g}}^2 \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)\right)$$
$$= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)\right]$$

set derivative to zero:

$$\begin{aligned} 0 &= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)\right] \\ \sigma_{g\mathrm{ML}}^2 &= \frac{1}{N}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right) \end{aligned}$$

- Note that For every SNP we need to calculate $\left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1}$, which is an $O(N^3)$ operation.

Linear mixed models (LMM)

ML parameters

Derivative of the LMM log likelihood w.r.t. σ_{g}^2

$$\mathrm{d}\sigma_{\mathrm{g}}^2 \log \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)\right)$$
$$= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)\right]$$

set derivative to zero:

$$\begin{aligned} 0 &= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)\right] \\ \sigma_{g\mathrm{ML}}^2 &= \frac{1}{N}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\top \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\right) \end{aligned}$$

- Note that For every SNP we need to calculate $\left(\boldsymbol{K} + \delta \boldsymbol{I}\right)^{-1}$, which is an $O(N^3)$ operation.

FaST LMM

[Lippert et al., 2011]

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^{2}\left(\boldsymbol{K}+\delta \boldsymbol{I}\right)\right).$$

FaST LMM

[Lippert et al., 2011]

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \left(\boldsymbol{K} + \delta \boldsymbol{I}\right)\right).$$

$$= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 \left(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top + \delta \boldsymbol{I}\right)\right).$$

FaST LMM

[Lippert et al., 2011]

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2\left(\boldsymbol{K}+\delta\boldsymbol{I}\right)\right).$$

$$=\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2\left(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\top}+\delta\boldsymbol{I}\right)\right).$$

$$=\mathcal{N}\left(\boldsymbol{U}^{\top}\boldsymbol{y} \mid \boldsymbol{U}^{\top}\boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2\left(\boldsymbol{\Lambda}+\delta\boldsymbol{I}\right)\right).$$

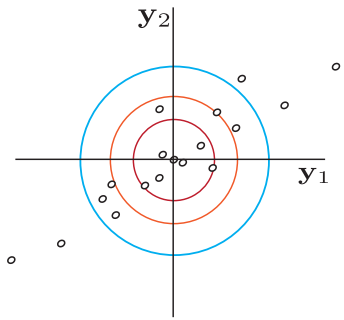
FaST LMM

[Lippert et al., 2011]

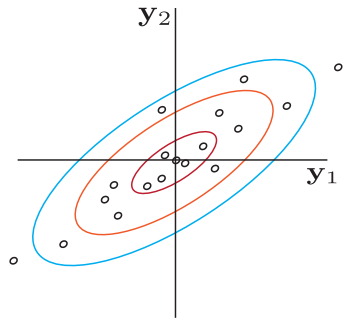
$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^{2}\left(\boldsymbol{K}+\delta \boldsymbol{I}\right)\right).$$

$$=\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^{2}\left(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\top}+\delta \boldsymbol{I}\right)\right).$$

$$=\mathcal{N}\left(\boldsymbol{U}^{\top}\boldsymbol{y} \mid \boldsymbol{U}^{\top}\boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^{2}\left(\boldsymbol{\Lambda}+\delta \boldsymbol{I}\right)\right).$$



$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_e^2\mathbf{I}\right)$



$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_g^2\mathbf{K}+\sigma_e^2\mathbf{I}\right)$

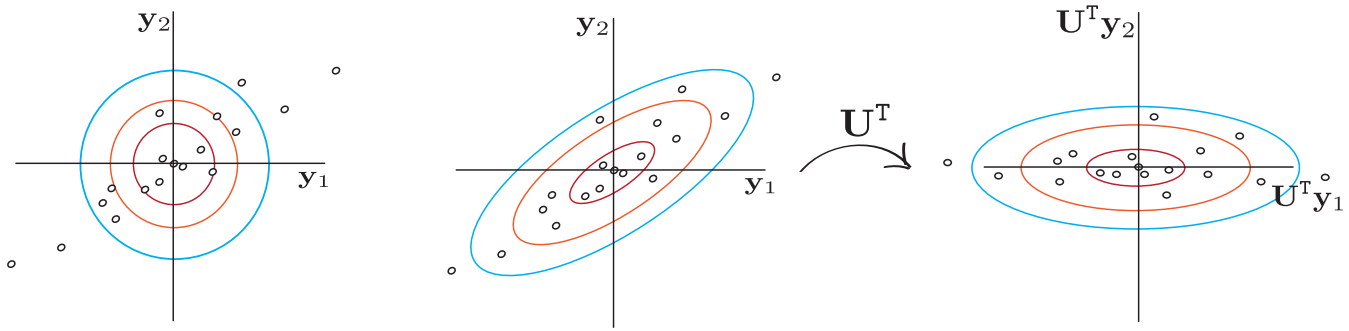
FaST LMM

[Lippert et al., 2011]

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^{2}\left(\boldsymbol{K}+\delta \boldsymbol{I}\right)\right).$$

$$=\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^{2}\left(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\top}+\delta \boldsymbol{I}\right)\right).$$

$$=\mathcal{N}\left(\boldsymbol{U}^{\top}\boldsymbol{y} \mid \boldsymbol{U}^{\top}\boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^{2}\left(\boldsymbol{\Lambda}+\delta \boldsymbol{I}\right)\right).$$



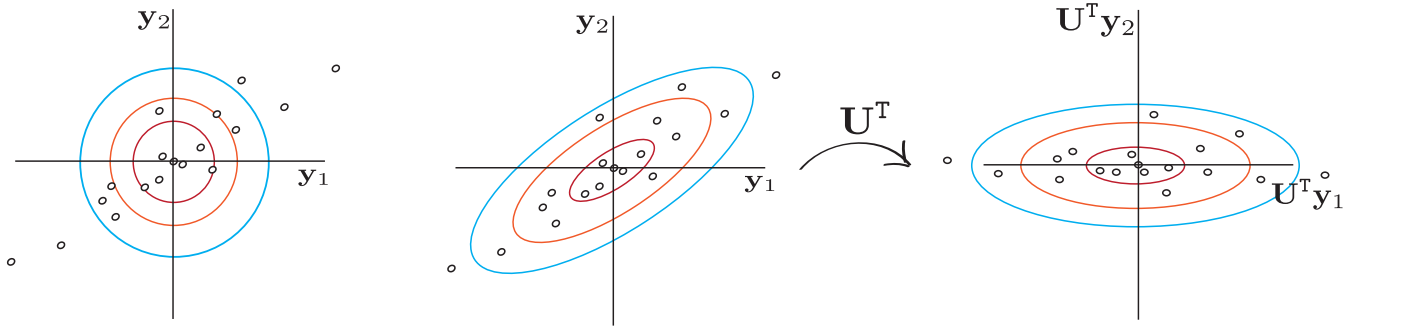
$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_e^2\mathbf{I}\right)$ $\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_g^2\mathbf{K}+\sigma_e^2\mathbf{I}\right)$ $\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_g^2(\mathbf{S}+\delta\mathbf{I})\right)$

FaST LMM

[Lippert et al., 2011]

$$\mathcal{N}\left(\boldsymbol{U}^\top \boldsymbol{y} \mid \boldsymbol{U}^\top \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 (\boldsymbol{\Lambda} + \delta \boldsymbol{I})\right).$$

- **Fa**ctored **S**pectrally **T**ransformed **LMM**
- $O(N^3)$ once for spectral decomposition.
- Exact LMM solution.
- Bottlenecks: $O(N^3)$ runtime, $O(N^2)$ memory for $\boldsymbol{K}$.



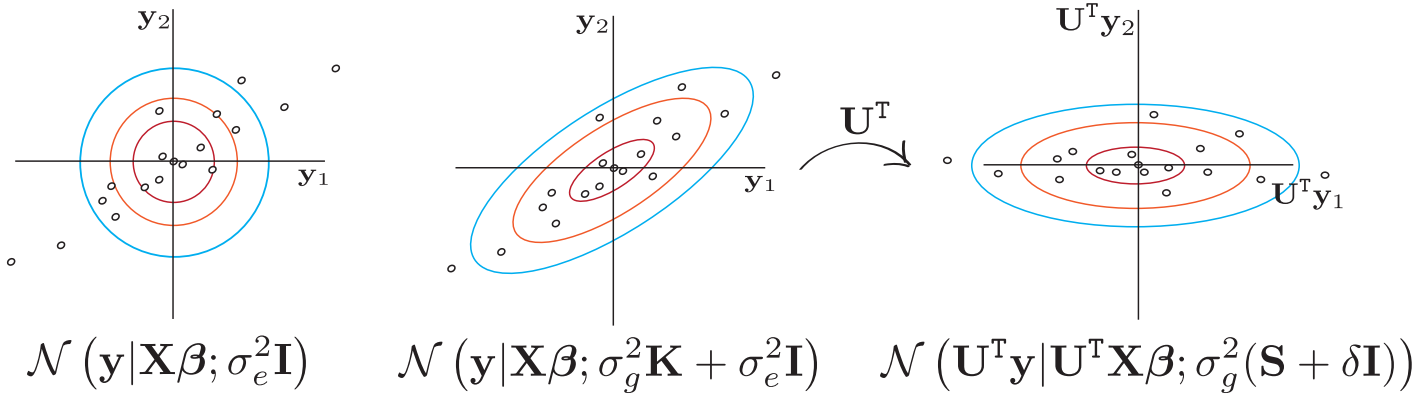
$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_e^2\mathbf{I}\right)$ $\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_g^2\mathbf{K} + \sigma_e^2\mathbf{I}\right)$ $\mathcal{N}\left(\mathbf{U}^\mathrm{T}\mathbf{y}|\mathbf{U}^\mathrm{T}\mathbf{X}\boldsymbol{\beta}; \sigma_g^2(\mathbf{S} + \delta\mathbf{I})\right)$

FaST LMM

[Lippert et al., 2011]

$$\mathcal{N}\left(\boldsymbol{U}^\top \boldsymbol{y} \mid \boldsymbol{U}^\top \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 (\boldsymbol{\Lambda} + \delta \boldsymbol{I})\right).$$

- **Fa**ctored **S**pectrally **T**ransformed **LMM**
- $O(N^3)$ once for spectral decomposition.
- Exact LMM solution.
- Bottlenecks: $O(N^3)$ runtime, $O(N^2)$ memory for $\boldsymbol{K}$.



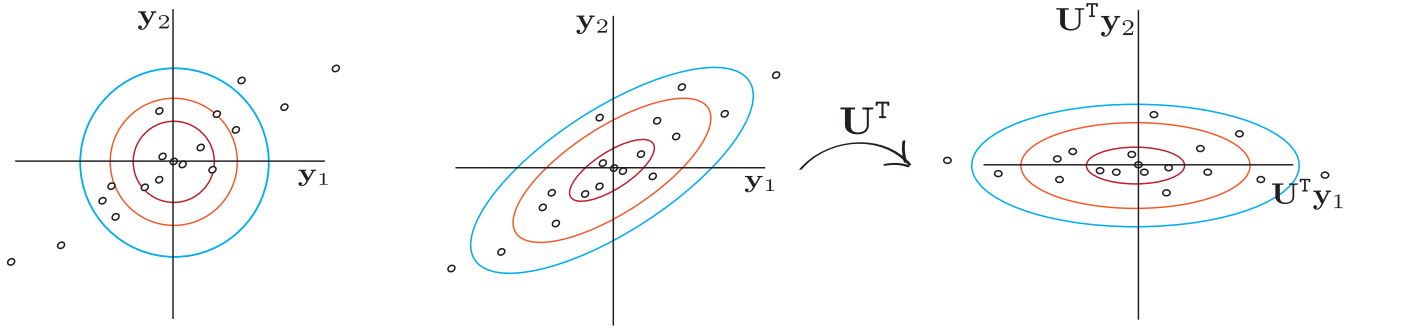
$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_e^2\mathbf{I}\right)$ $\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_g^2\mathbf{K} + \sigma_e^2\mathbf{I}\right)$ $\mathcal{N}\left(\mathbf{U}^\mathrm{T}\mathbf{y}|\mathbf{U}^\mathrm{T}\mathbf{X}\boldsymbol{\beta}; \sigma_g^2(\mathbf{S} + \delta\mathbf{I})\right)$

FaST LMM

[Lippert et al., 2011]

$$\mathcal{N}\left(\boldsymbol{U}^\top \boldsymbol{y} \mid \boldsymbol{U}^\top \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 (\boldsymbol{\Lambda} + \delta \boldsymbol{I})\right).$$

- **Fa**ctored **S**pectrally **T**ransformed **LMM**
- $O(N^3)$ once for spectral decomposition.
- Exact LMM solution.
- Bottlenecks: $O(N^3)$ runtime, $O(N^2)$ memory for $\boldsymbol{K}$.



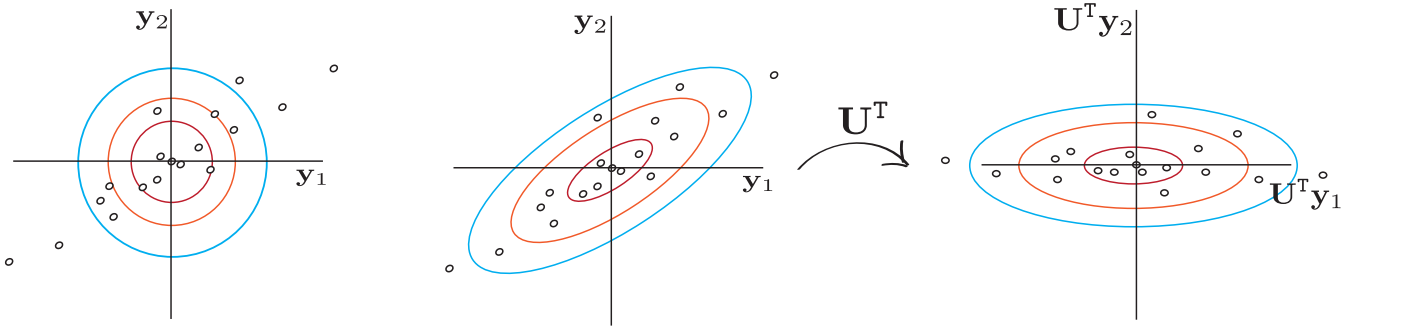
$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_e^2\mathbf{I}\right)$ $\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_g^2\mathbf{K} + \sigma_e^2\mathbf{I}\right)$ $\mathcal{N}\left(\mathbf{U}^\mathrm{T}\mathbf{y}|\mathbf{U}^\mathrm{T}\mathbf{X}\boldsymbol{\beta}; \sigma_g^2(\mathbf{S} + \delta\mathbf{I})\right)$

FaST LMM

[Lippert et al., 2011]

$$\mathcal{N}\left(\boldsymbol{U}^\top \boldsymbol{y} \mid \boldsymbol{U}^\top \boldsymbol{X}\boldsymbol{\beta}, \sigma_{\mathrm{g}}^2 (\boldsymbol{\Lambda} + \delta \boldsymbol{I})\right).$$

- **Fa**ctored **S**pectrally **T**ransformed **LMM**
- $O(N^3)$ once for spectral decomposition.
- Exact LMM solution.
- Bottlenecks: $O(N^3)$ runtime, $O(N^2)$ memory for $\boldsymbol{K}$.



$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_e^2\mathbf{I}\right)$ $\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_g^2\mathbf{K} + \sigma_e^2\mathbf{I}\right)$ $\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta}; \sigma_g^2(\mathbf{S} + \delta\mathbf{I})\right)$

Linear mixed models

Linear regression view

- For linear similarities, a LMM is equivalent to a *linear regression*.
- All SNPs are used as regression covariates.
- Uncertainty about identity of true causal variants expressed by considering a distribution over the effect sizes.

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{x}\beta\,,\, \sigma_{\mathrm{g}}^2 \boldsymbol{K} + \sigma^2 \boldsymbol{I}\right).$$

Linear mixed models

Linear regression view

- For linear similarities, a LMM is equivalent to a *linear regression*.
- All SNPs are used as regression covariates.
- Uncertainty about identity of true causal variants expressed by considering a distribution over the effect sizes.

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{x}\beta\,,\, \sigma_{\mathrm{g}}^2 \tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^\top + \sigma^2 \boldsymbol{I}\right).$$

Linear mixed models

Linear regression view

- For linear similarities, a LMM is equivalent to a *linear regression*.
- All SNPs are used as regression covariates.
- Uncertainty about identity of true causal variants expressed by considering a distribution over the effect sizes.

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{x}\beta\,,\, \sigma_{\mathrm{g}}^2 \tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^\top + \sigma^2 \boldsymbol{I}\right).$$

$$\propto \int \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{x}\beta + \tilde{\boldsymbol{X}}\boldsymbol{\theta}\,,\, \sigma^2\delta\boldsymbol{I}\right) \cdot \mathcal{N}\left(\boldsymbol{\theta} \mid \mathbf{0}\,,\, \sigma_{\mathrm{g}}^2 \boldsymbol{I}\right) \mathrm{d}\boldsymbol{\theta}.$$

Linear mixed models

Linear regression view

- For linear similarities, a LMM is equivalent to a *linear regression*.
- All SNPs are used as regression covariates.
- Uncertainty about identity of true causal variants expressed by considering a distribution over the effect sizes.

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{x}\beta\,,\, \sigma_{\mathrm{g}}^2 \tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^\top + \sigma^2 \boldsymbol{I}\right).$$

$$\propto \int \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{x}\beta + \tilde{\boldsymbol{X}}\boldsymbol{\theta}\,,\, \sigma^2\delta\boldsymbol{I}\right) \cdot \mathcal{N}\left(\boldsymbol{\theta} \mid \boldsymbol{0}\,,\, \sigma_{\mathrm{g}}^2 \boldsymbol{I}\right) \mathrm{d}\boldsymbol{\theta}.$$

Linear mixed models

Linear regression view

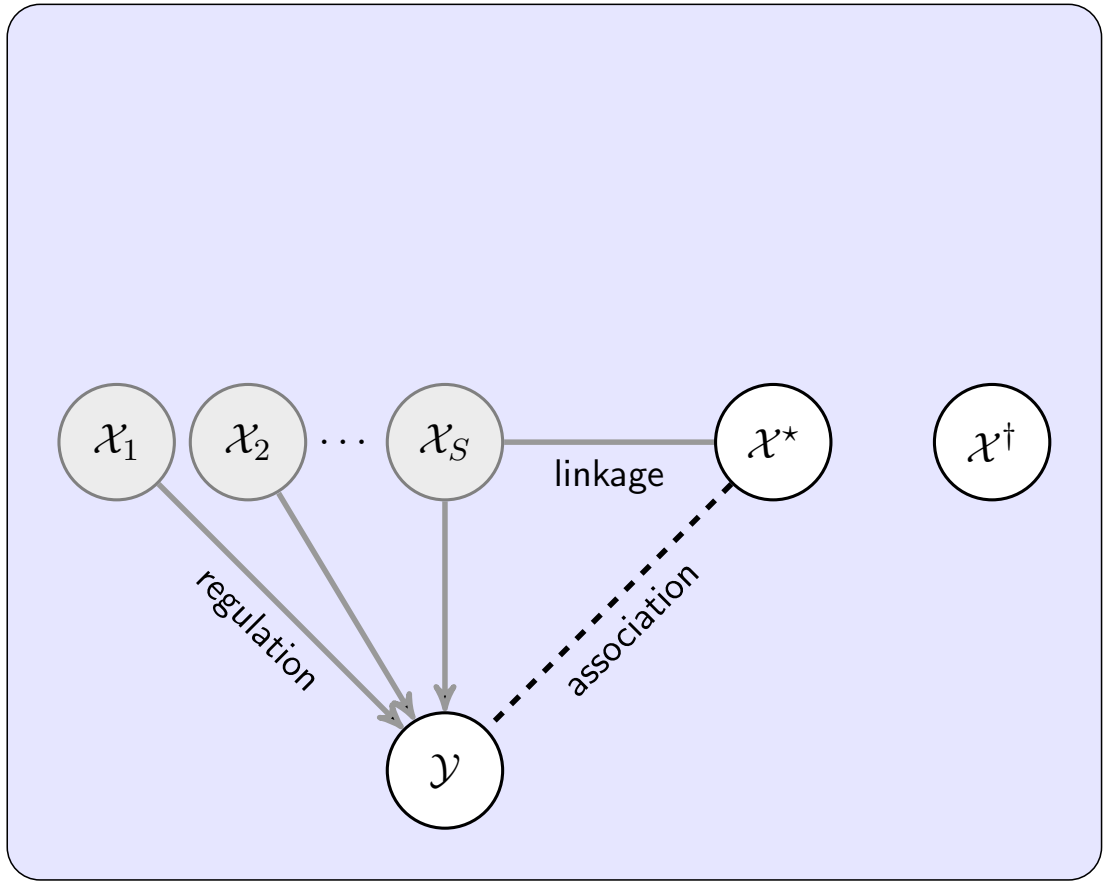
- For linear similarities, a LMM is equivalent to a *linear regression*.
- All SNPs are used as regression covariates.
- Uncertainty about identity of true causal variants expressed by considering a distribution over the effect sizes.

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{x}\beta\,,\, \sigma_{\mathrm{g}}^2 \tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^\top + \sigma^2 \boldsymbol{I}\right).$$

$$\propto \int \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{x}\beta + \tilde{\boldsymbol{X}}\boldsymbol{\theta}\,,\, \sigma^2\delta\boldsymbol{I}\right) \cdot \mathcal{N}\left(\boldsymbol{\theta} \mid \mathbf{0}\,,\, \sigma_{\mathrm{g}}^2 \boldsymbol{I}\right) \mathrm{d}\boldsymbol{\theta}.$$

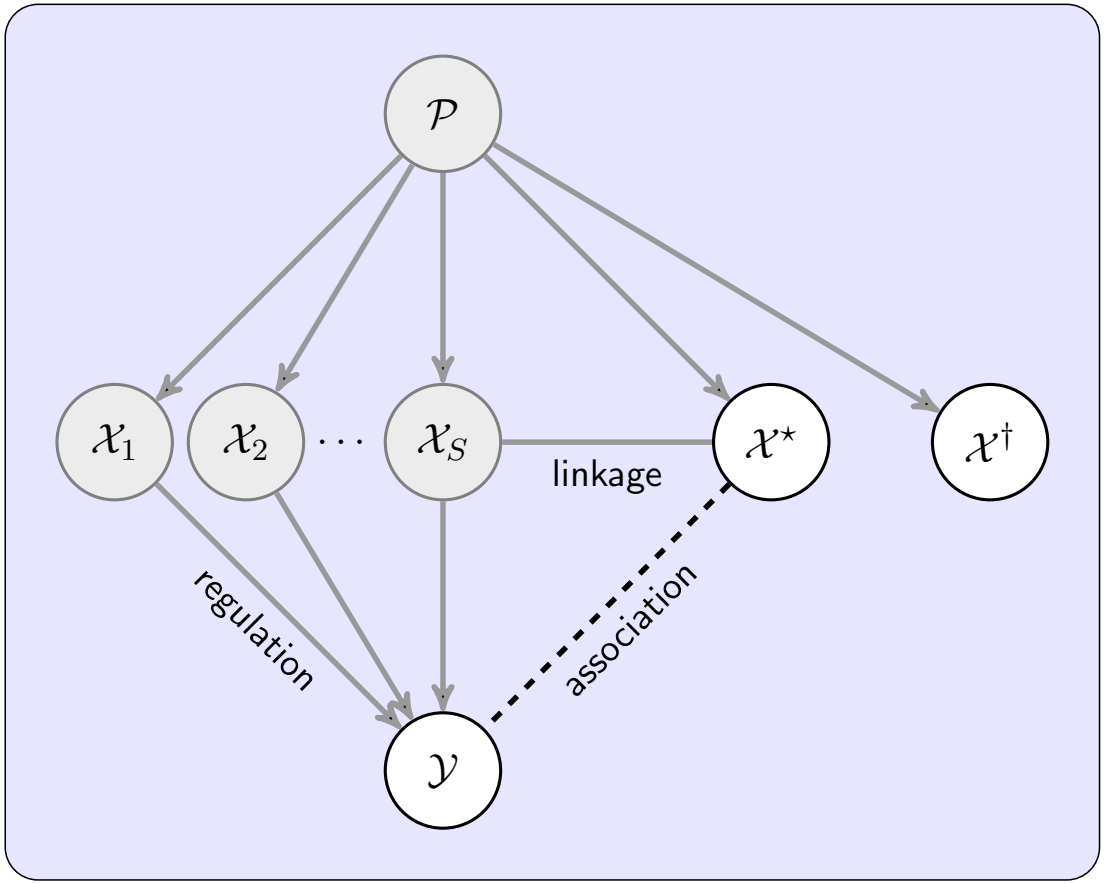
GWAS of complex traits

- Association by linkage to causal variant
- SNPs confounded by population structure creates correlation between physically unlinked markers
- Spurious associations if not taken into account
- Alternatively condition on causal/confounded markers



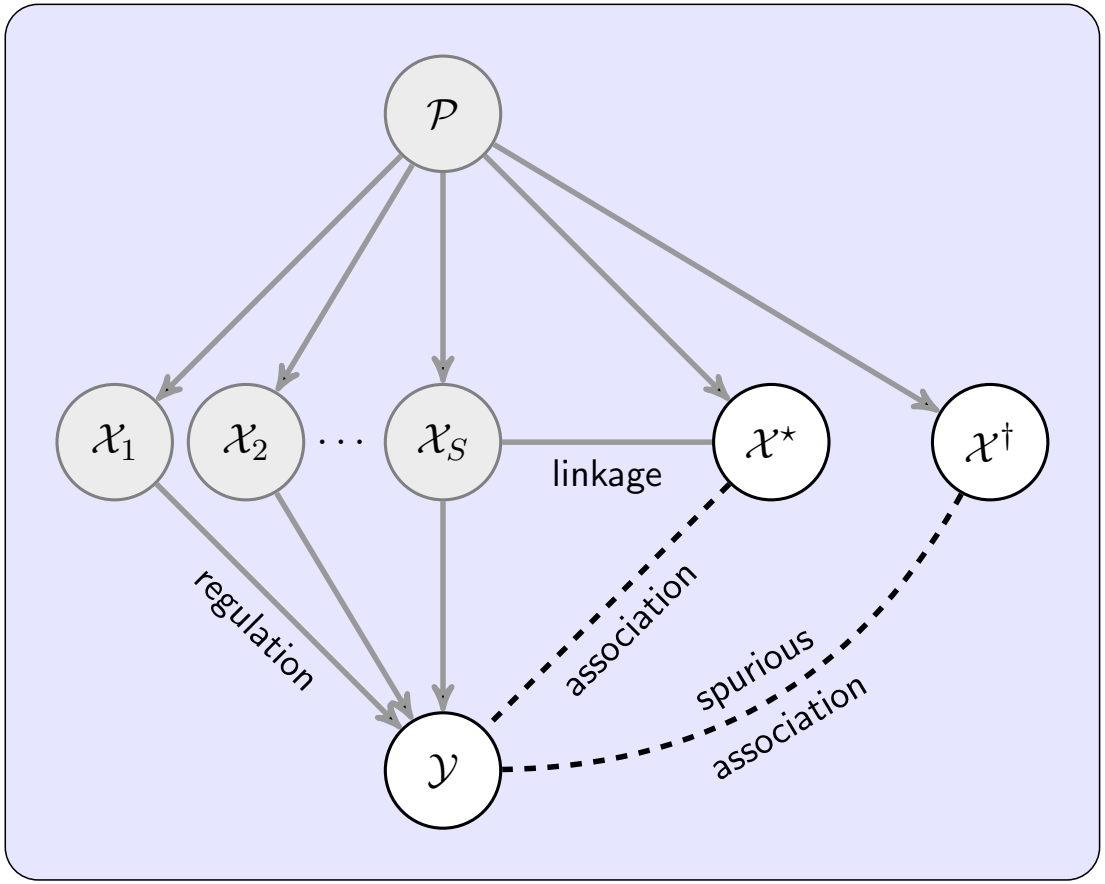
GWAS of complex traits

- Association by linkage to causal variant
- SNPs confounded by population structure creates correlation between physically unlinked markers
- Spurious associations if not taken into account
- Alternatively condition on causal/confounded markers



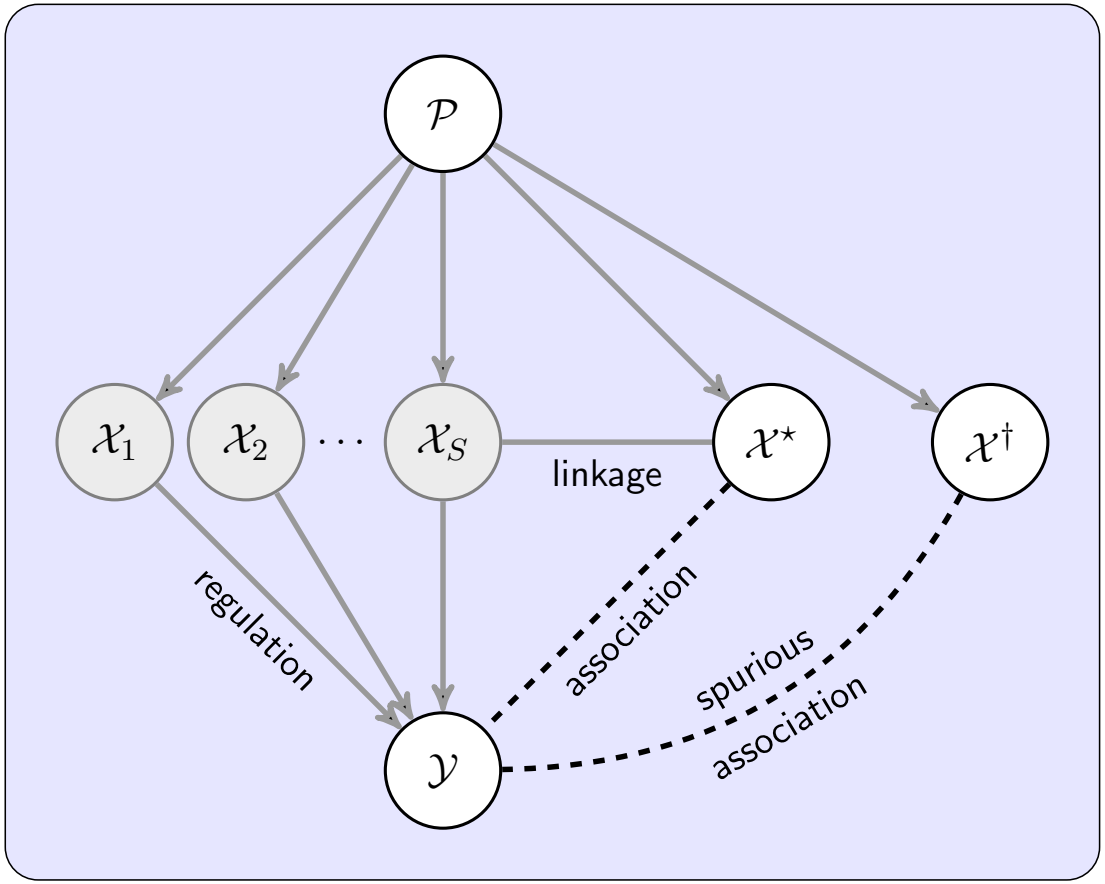
GWAS of complex traits

- Association by linkage to causal variant
- SNPs confounded by population structure creates correlation between physically unlinked markers
- Spurious associations if not taken into account
- Alternatively condition on causal/confounded markers



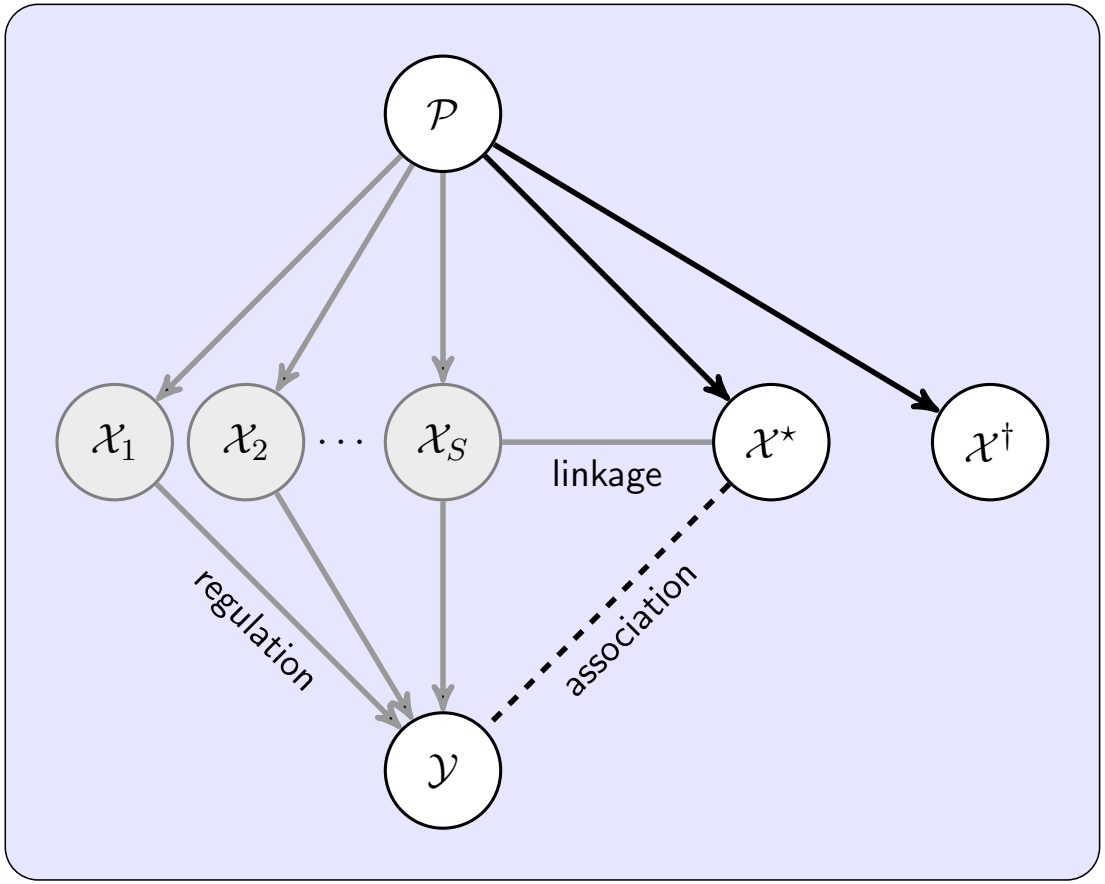
GWAS of complex traits

- Association by linkage to causal variant
- SNPs confounded by population structure creates correlation between physically unlinked markers
- Spurious associations if not taken into account
- Alternatively condition on causal/confounded markers



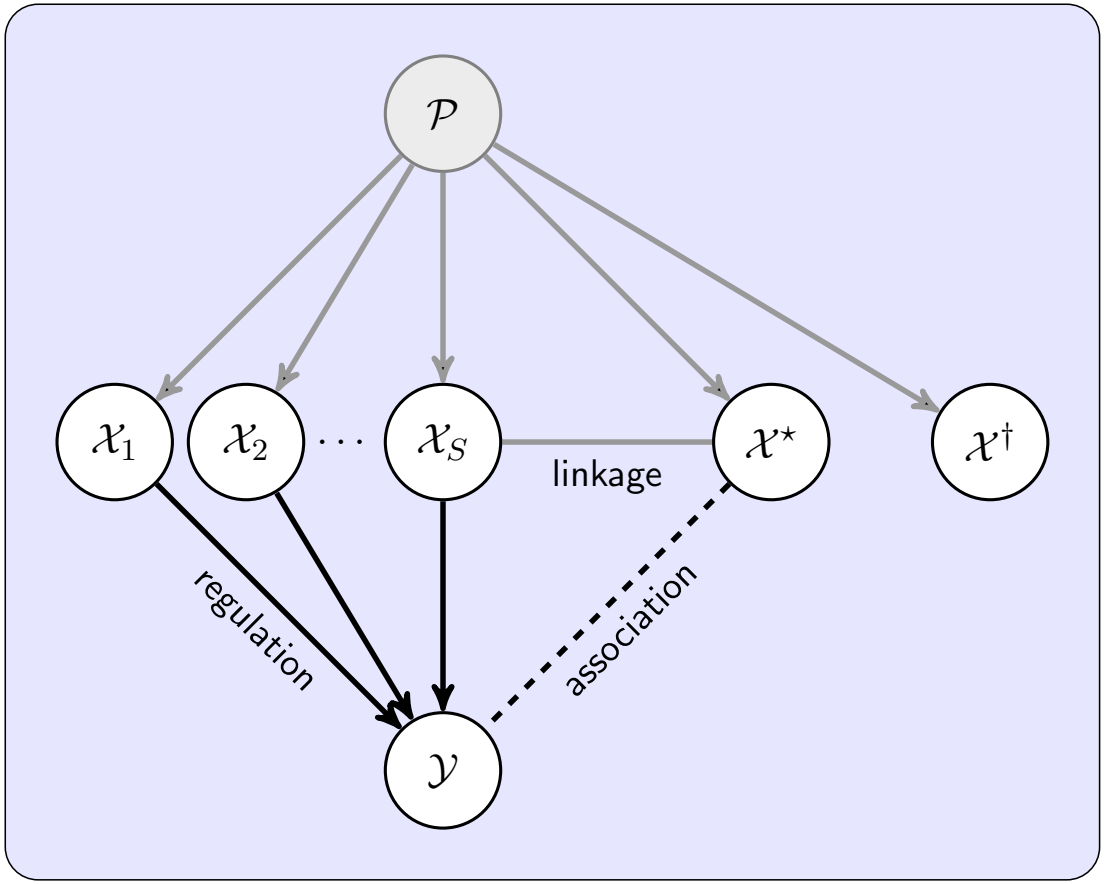
GWAS of complex traits

- Association by linkage to causal variant
- SNPs confounded by population structure creates correlation between physically unlinked markers
- Spurious associations if not taken into account
- Alternatively condition on causal/confounded markers



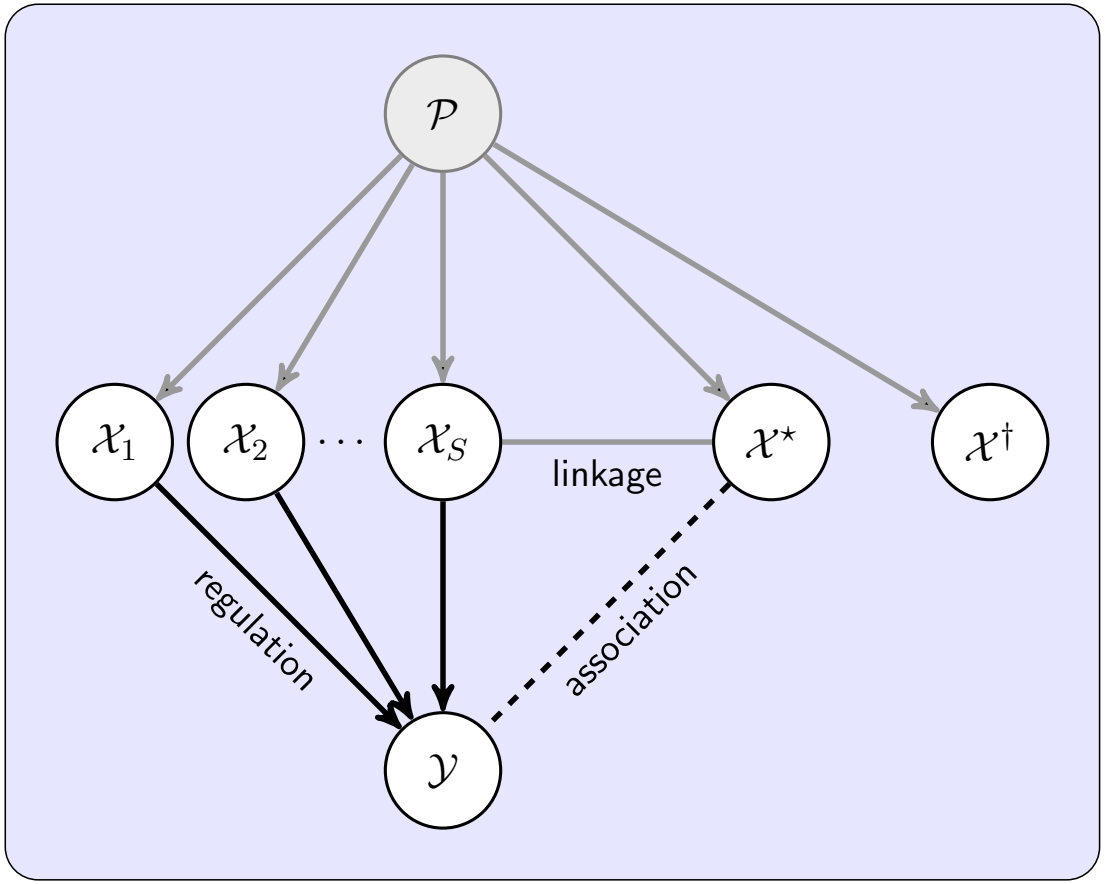
GWAS of complex traits

- Association by linkage to causal variant
- SNPs confounded by population structure creates correlation between physically unlinked markers
- Spurious associations if not taken into account
- Alternatively condition on causal/confounded markers



GWAS of complex traits

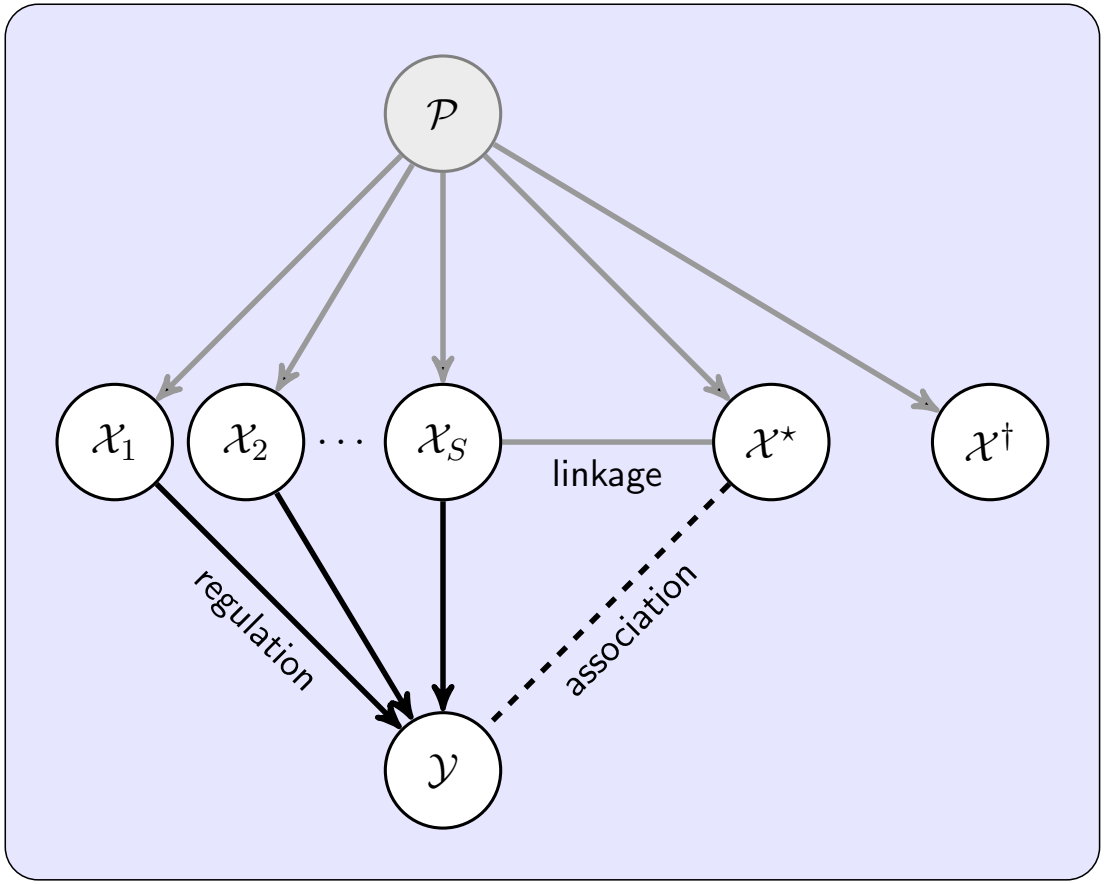
- Association by linkage to causal variant
- SNPs confounded by population structure creates correlation between physically unlinked markers
- Spurious associations if not taken into account
- Alternatively condition on causal/confounded markers



A LMM accounts for model misspecification when testing a univariate model when in reality the trait is multi-factorial.

GWAS of complex traits

- Association by linkage to causal variant
- SNPs confounded by population structure creates correlation between physically unlinked markers
- Spurious associations if not taken into account
- Alternatively condition on causal/confounded markers



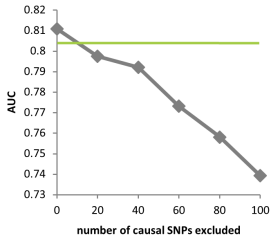
A LMM accounts for model misspecification when testing a univariate model when in reality the trait is multi-factorial.

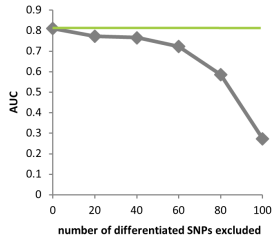
Can we do better than using **all** SNPs for correction?

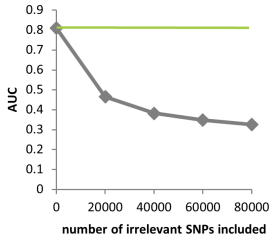
FaST-LMM **Select**

Dilution

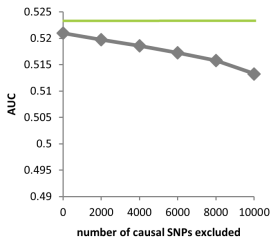
(a) Low polygenicity

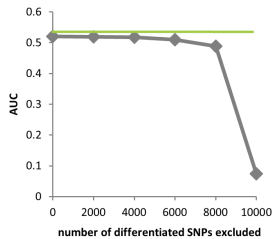


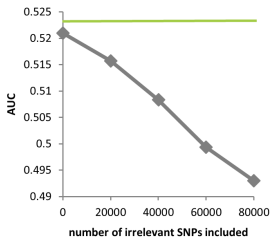




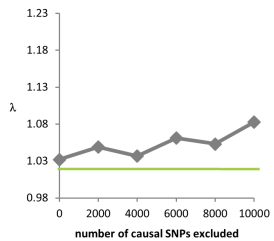
High polygenicity

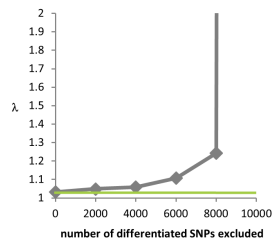


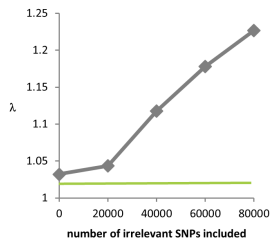




(b) High polygenicity





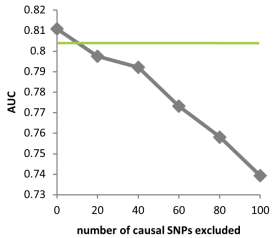


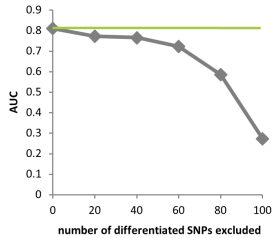
- phenotype differentiated between populations plus causal variants
- **Experiments:**
  - removal of causal SNPs
  - removal of differentiated SNPs
  - addition of irrelevant SNPs
- **Evaluation measures:**
  - AUC for case-control data
  - Genomic Control λ for quantitative data
- FaST-LMM-Select optimizing **phenotype prediction** generally performs well

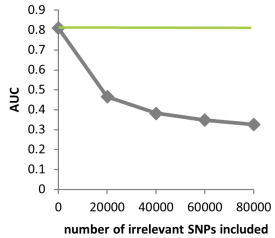
FaST-LMM **Select**

Dilution

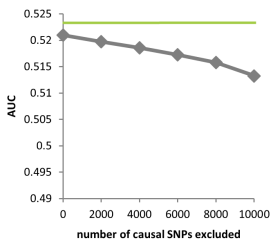
(a) Low polygenicity

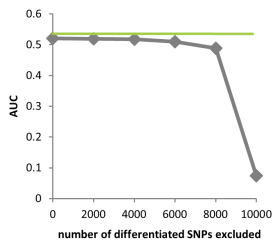


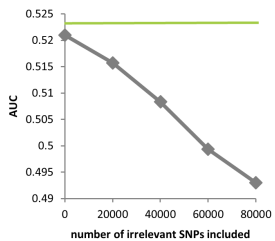




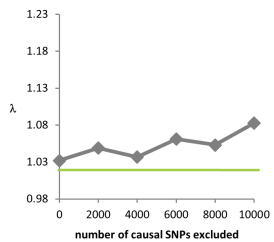
High polygenicity

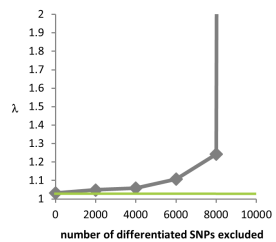


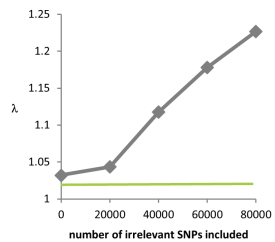




(b) High polygenicity





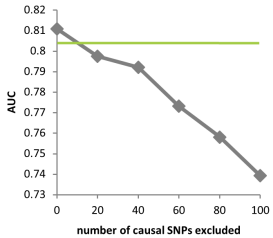


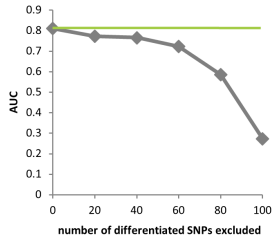
- phenotype differentiated between populations plus causal variants
- **Experiments:**
  - removal of causal SNPs
  - removal of differentiated SNPs
  - addition of irrelevant SNPs
- **Evaluation measures:**
  - AUC for case-control data
  - Generalized λ for quantitative data
- FaST-LMM-Select optimizing **phenotype prediction** generally performs well

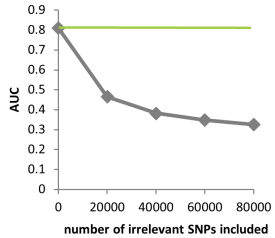
FaST-LMM **Select**

Dilution

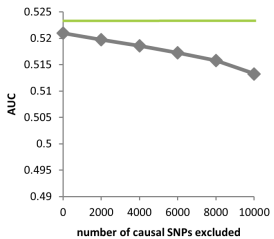
(a) Low polygenicity

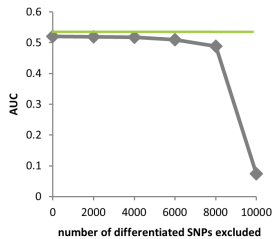


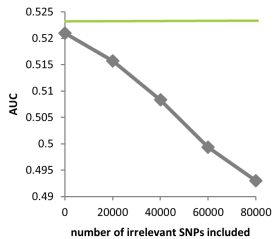




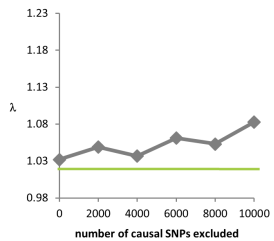
High polygenicity

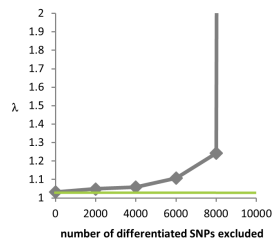


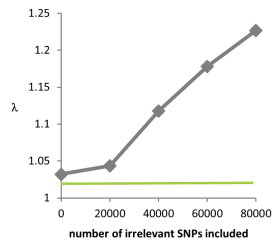




(b) High polygenicity





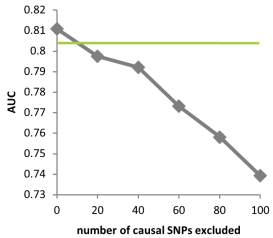


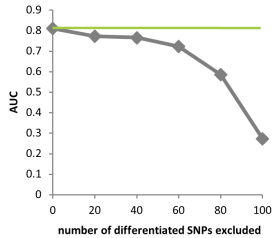
- phenotype differentiated between populations plus causal variants
- **Experiments:**
  - removal of causal SNPs
  - removal of differentiated SNPs
  - addition of irrelevant SNPs
- **Evaluation measures:**
  - AUC for discriminating SNPs
  - Genomic control λ for calibration
- FaST-LMM-Select optimizing **phenotype prediction** generally performs well

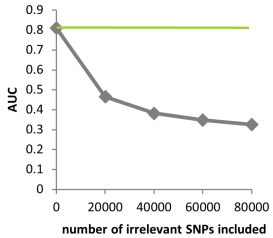
FaST-LMM **Select**

Dilution

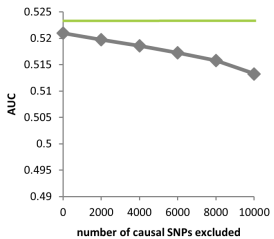
(a) Low polygenicity

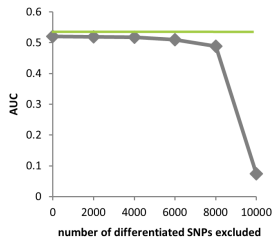


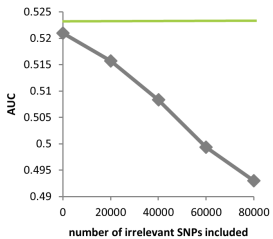




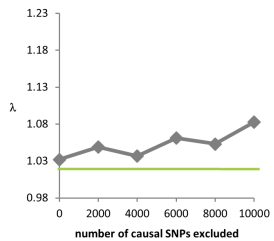
High polygenicity

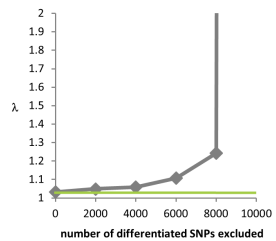


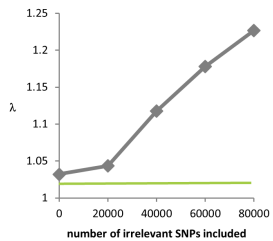




(b) High polygenicity





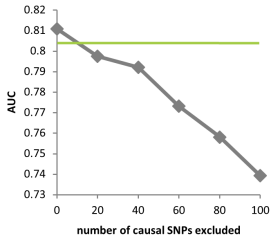


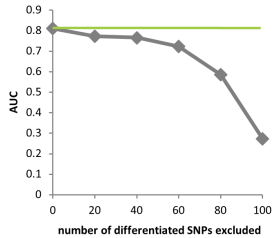
- phenotype differentiated between populations plus causal variants
- **Experiments:**
  - removal of causal SNPs
  - removal of differentiated SNPs
  - addition of irrelevant SNPs
- **Evaluation measures:**
  - AUC for binary phenotypes
  - Genomic control λ for quantitative phenotypes
- FaST-LMM-Select optimizing **phenotype prediction** generally performs well

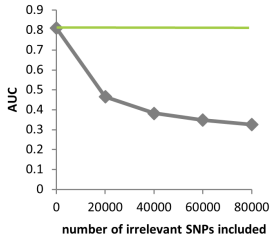
FaST-LMM **Select**

Dilution

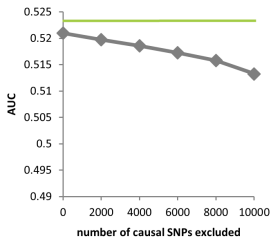
(a) Low polygenicity

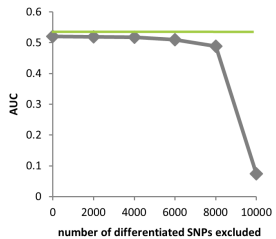


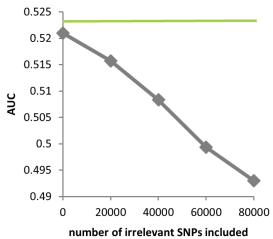




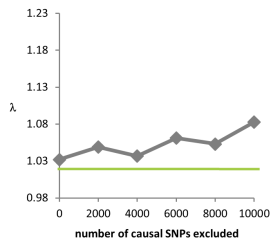
High polygenicity

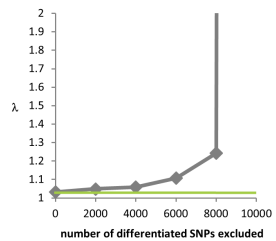


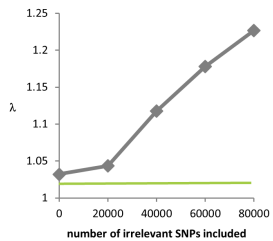




(b) High polygenicity





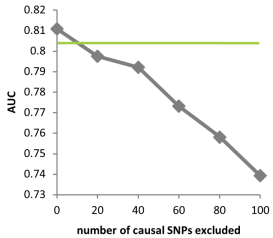


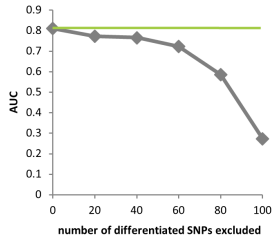
- phenotype differentiated between populations plus causal variants
- **Experiments:**
  - removal of causal SNPs
  - removal of differentiated SNPs
  - addition of irrelevant SNPs
- **Evaluation measures:**
  - AUC for detecting causal SNPs
  - Genomic Control λ for non-causal SNPs
- FaST-LMM-Select optimizing **phenotype prediction** generally performs well

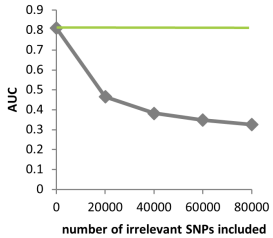
FaST-LMM **Select**

Dilution

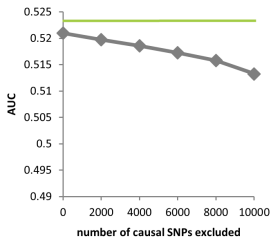
(a) Low polygenicity

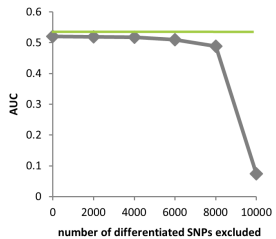


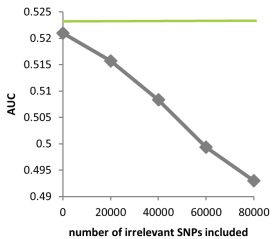




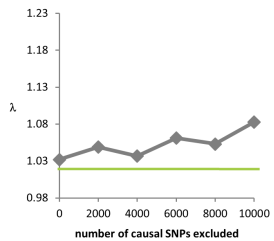
High polygenicity

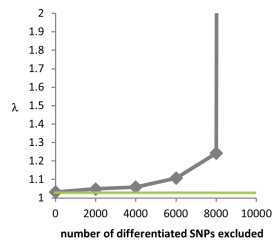


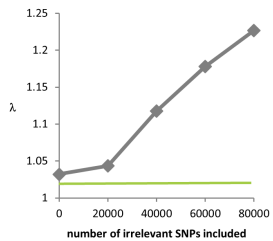




(b) High polygenicity





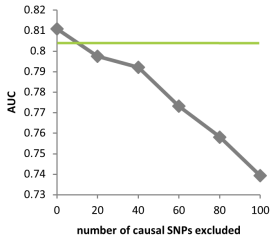


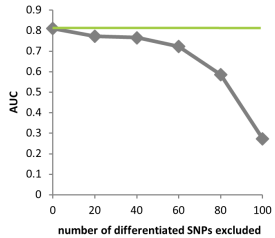
- phenotype differentiated between populations plus causal variants
- **Experiments:**
  - removal of causal SNPs
  - removal of differentiated SNPs
  - addition of irrelevant SNPs
- **Evaluation measures:**
  - **AUC** for detecting causal SNPs
  - **Genomic Control** λ on non-causal SNPs
- FaST-LMM-Select optimizing **phenotype prediction** generally performs well

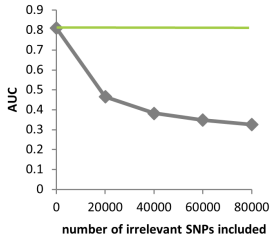
FaST-LMM **Select**

Dilution

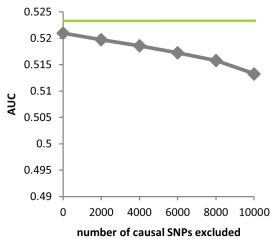
(a) Low polygenicity

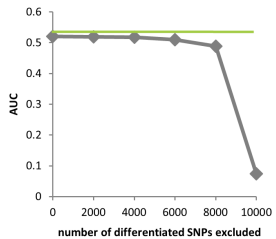


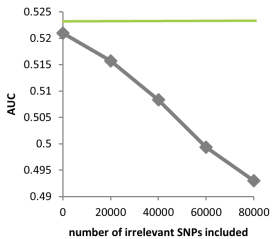




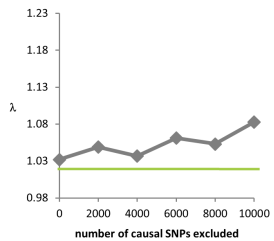
High polygenicity

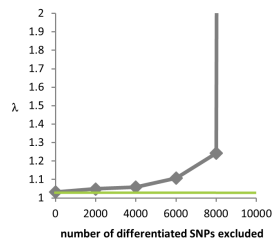


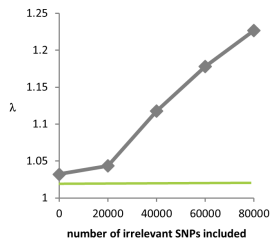




(b) High polygenicity





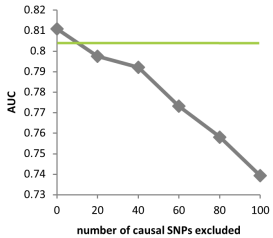


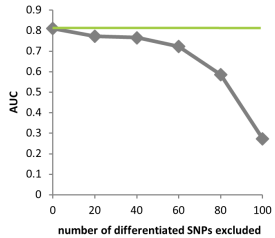
- phenotype differentiated between populations plus causal variants
- **Experiments:**
  - removal of causal SNPs
  - removal of differentiated SNPs
  - addition of irrelevant SNPs
- **Evaluation measures:**
  - **AUC** for detecting causal SNPs
  - **Genomic Control** λ on non-causal SNPs
- FaST-LMM-Select optimizing **phenotype prediction** generally performs well

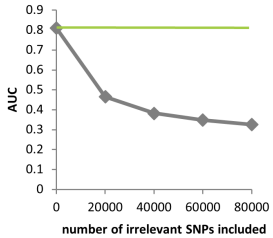
FaST-LMM **Select**

Dilution

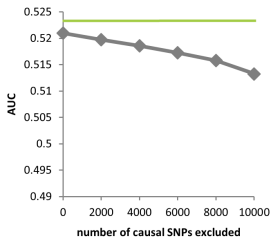
(a) Low polygenicity

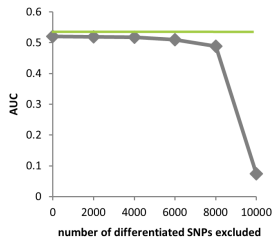


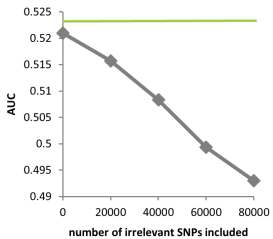




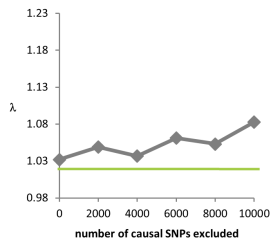
High polygenicity

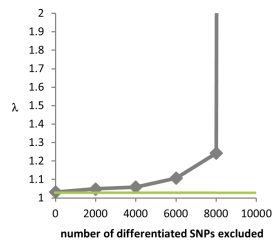


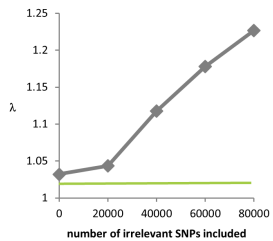




(b) High polygenicity





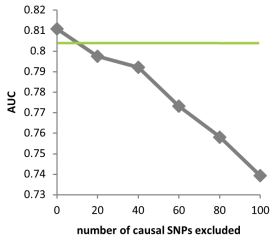


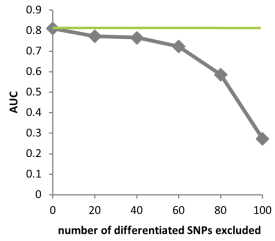
- phenotype differentiated between populations plus causal variants
- **Experiments:**
  - removal of causal SNPs
  - removal of differentiated SNPs
  - addition of irrelevant SNPs
- **Evaluation measures:**
  - **AUC** for detecting causal SNPs
  - **Genomic Control** λ on non-causal SNPs
- FaST-LMM-Select optimizing **phenotype prediction** generally performs well

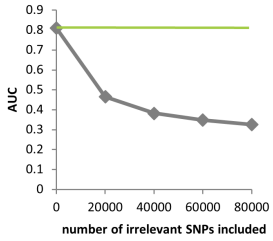
FaST-LMM **Select**

Dilution

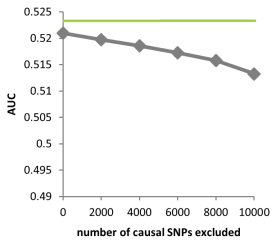
(a) Low polygenicity

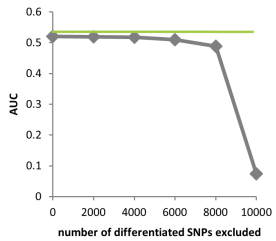


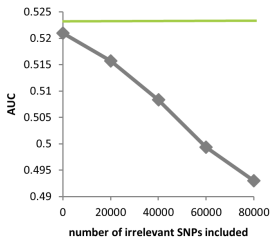




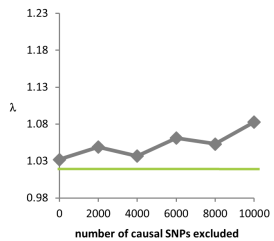
High polygenicity

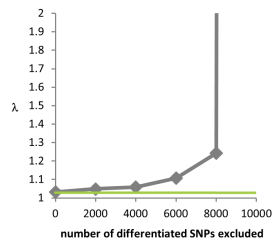


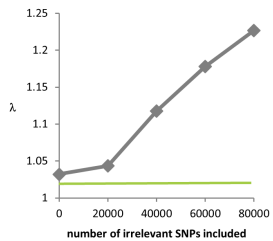




(b) High polygenicity



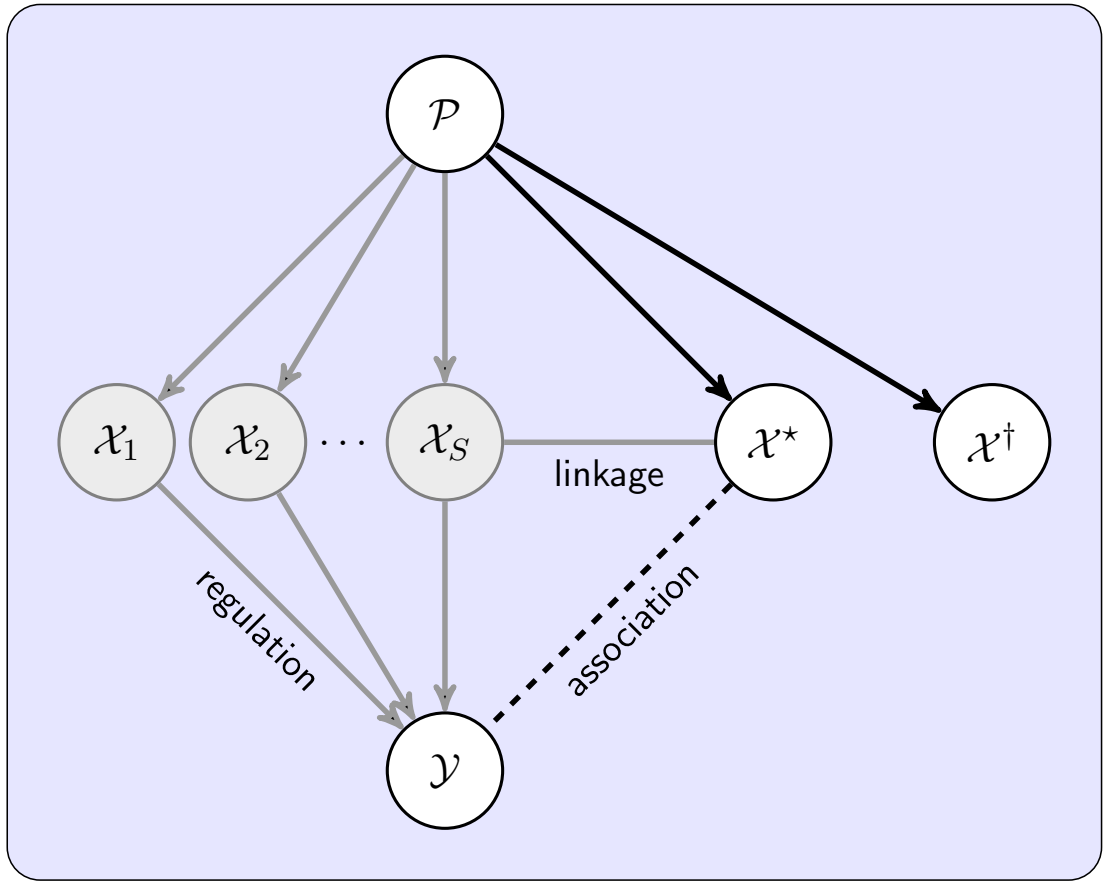




- phenotype differentiated between populations plus causal variants
- **Experiments:**
  - removal of causal SNPs
  - removal of differentiated SNPs
  - addition of irrelevant SNPs
- **Evaluation measures:**
  - **AUC** for detecting causal SNPs
  - **Genomic Control** λ on non-causal SNPs
- FaST-LMM-Select optimizing **phenotype prediction** generally performs well

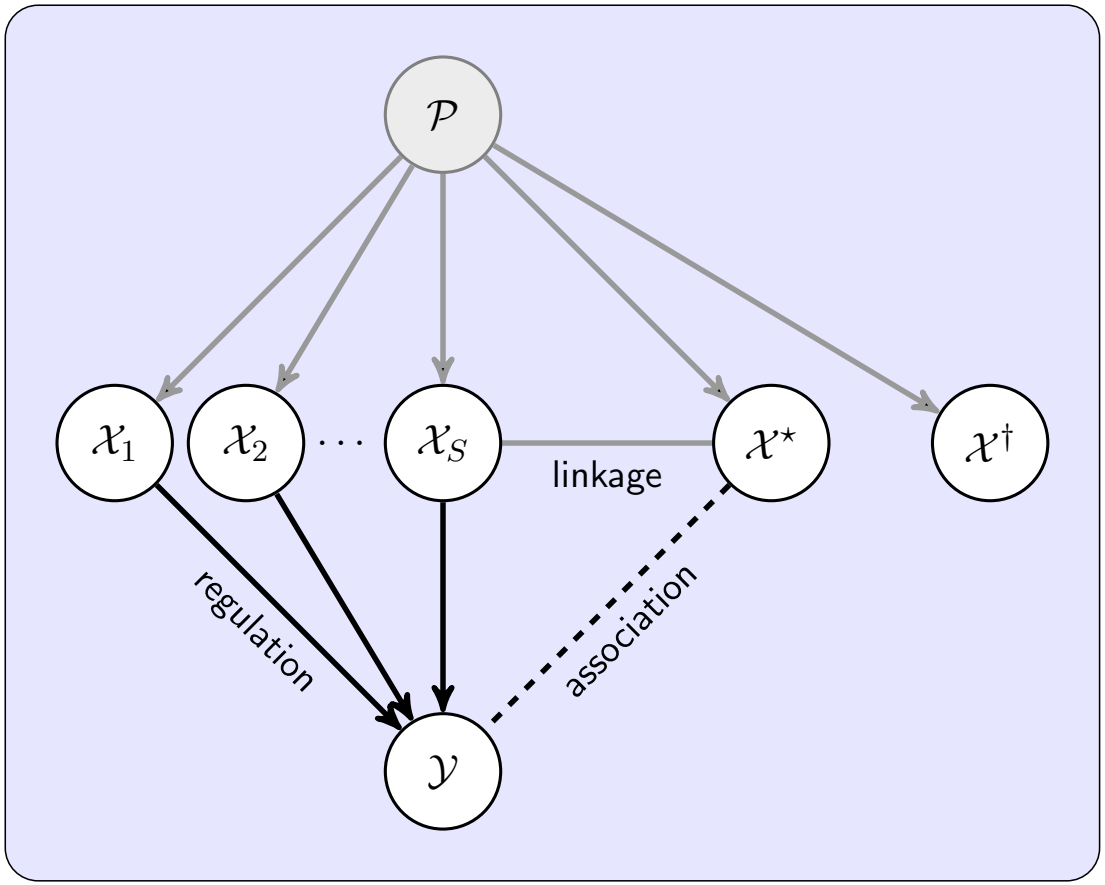
Proximal contamination

- Compute similarity matrix based on SNPs
- Equivalent to linear regression conditioned on these SNPs
- Conditioning on SNPs within linkage to test marker reduces association. $\rightarrow$ Loss in power!
- Remove SNPs within linkage from computation of similarity matrix (infeasible if done naively!)



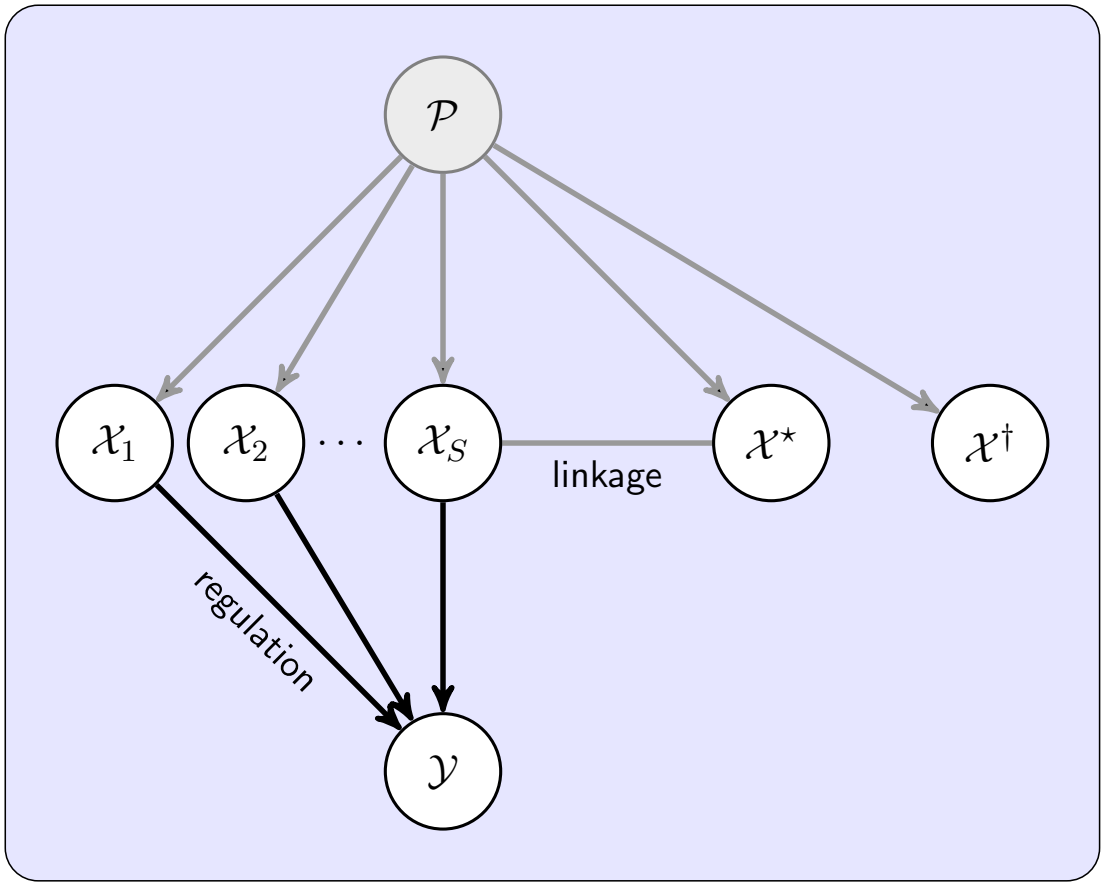
Proximal contamination

- Compute similarity matrix based on SNPs
- Equivalent to linear regression conditioned on these SNPs
- Conditioning on SNPs within linkage to test marker reduces association. $\rightarrow$ Loss in power!
- Remove SNPs within linkage from computation of similarity matrix (infeasible if done naively!)



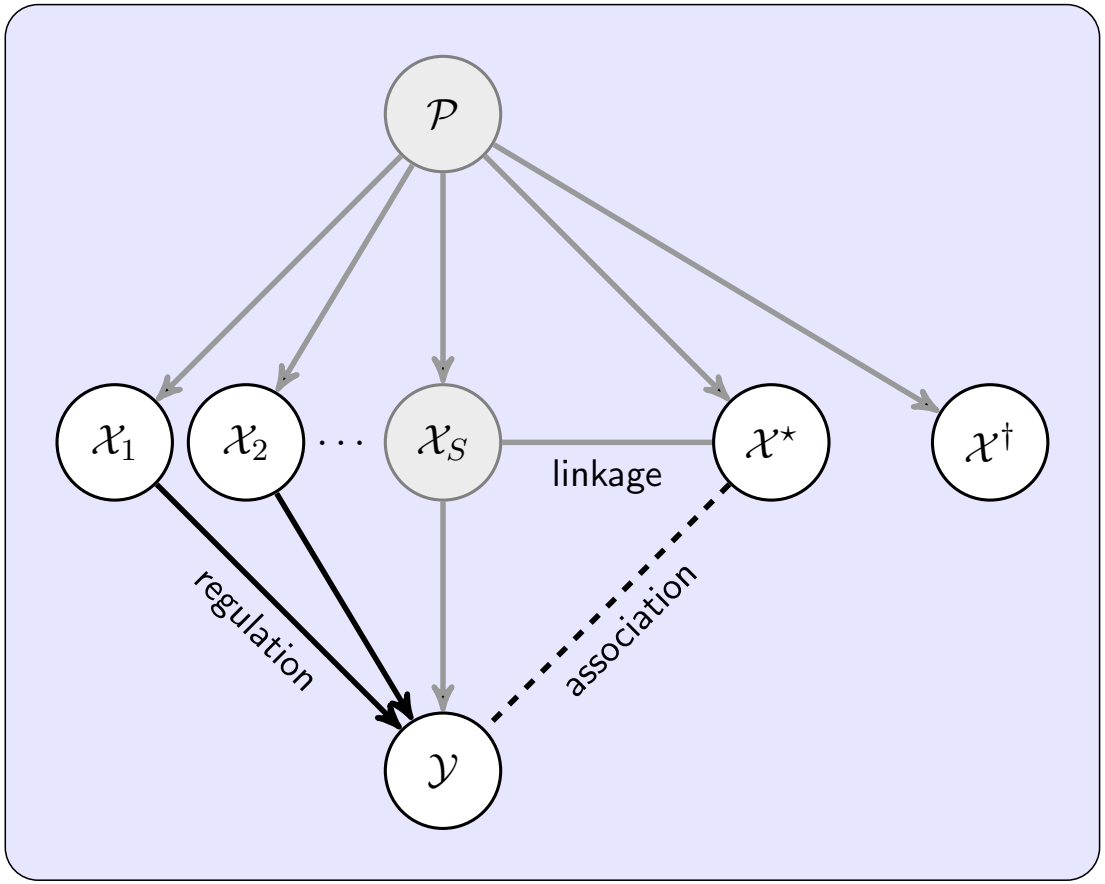
Proximal contamination

- Compute similarity matrix based on SNPs
- Equivalent to linear regression conditioned on these SNPs
- Conditioning on SNPs within linkage to test marker reduces association. $\rightarrow$ Loss in power!
- Remove SNPs within linkage from computation of similarity matrix (infeasible if done naively!)



Proximal contamination

- Compute similarity matrix based on SNPs
- Equivalent to linear regression conditioned on these SNPs
- Conditioning on SNPs within linkage to test marker reduces association. $\rightarrow$ Loss in power!
- Remove SNPs within linkage from computation of similarity matrix (infeasible if done naively!)

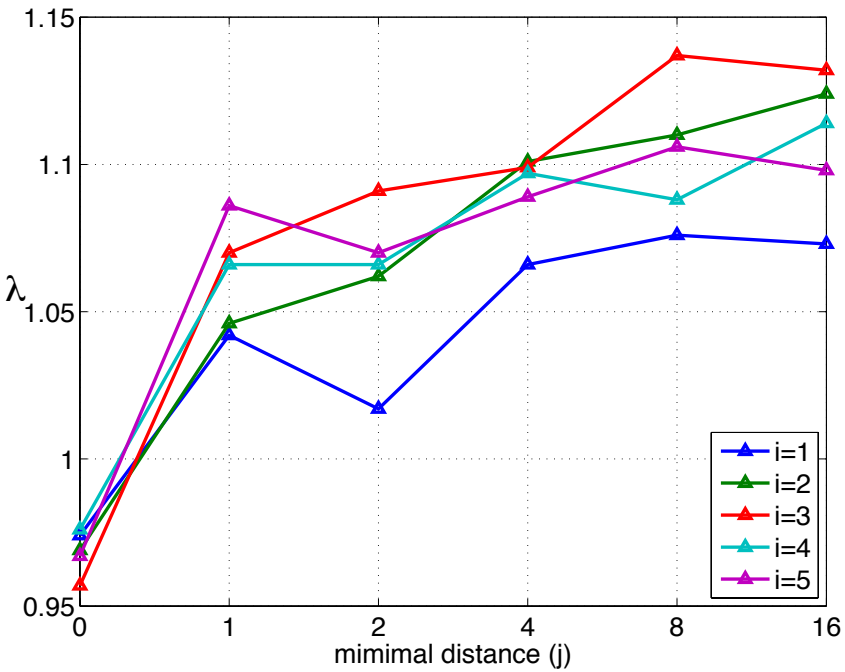


Proximal contamination

- Having a SNP in the similarity matrix that is linked to the SNP tested leads to loss in power.
- Correct by removing a sliding window around test-SNP from the similarity matrix.
- Correction is computed efficiently by subtracting a low-rank term.

Proximal contamination

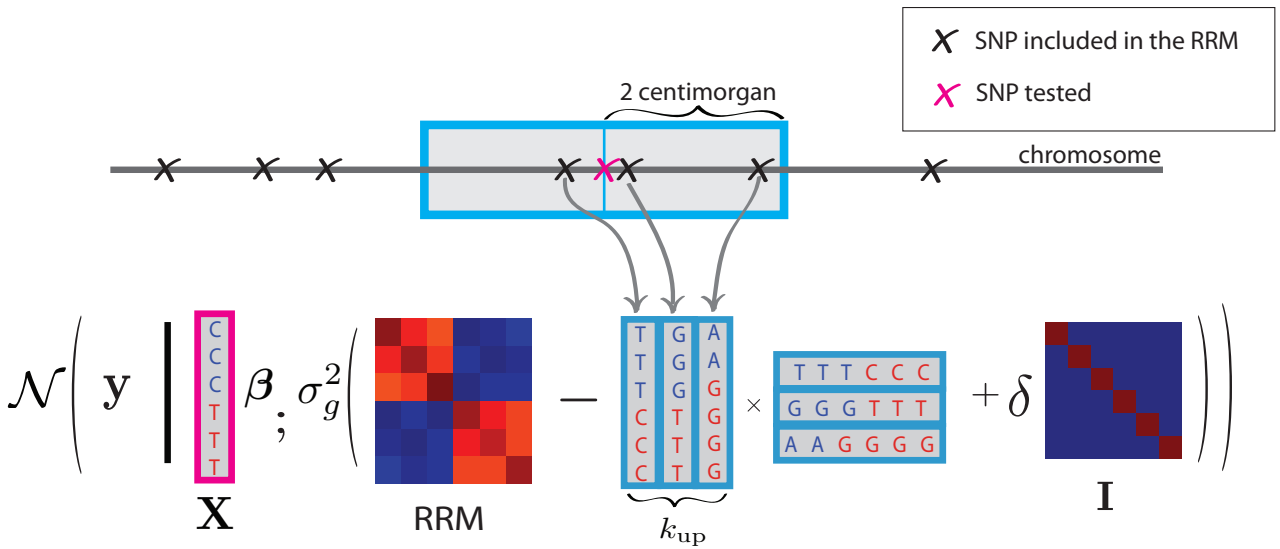
- Having a SNP in the similarity matrix that is linked to the SNP tested leads to loss in power.
- Correct by removing a sliding window around test-SNP from the similarity matrix.
- Correction is computed efficiently by subtracting a low-rank term.



- WTCCC data (Chrohn's)
- 6 genetic similarity matrices with equal number of markers at increasing distance to SNPs tested
- λ increases with distance

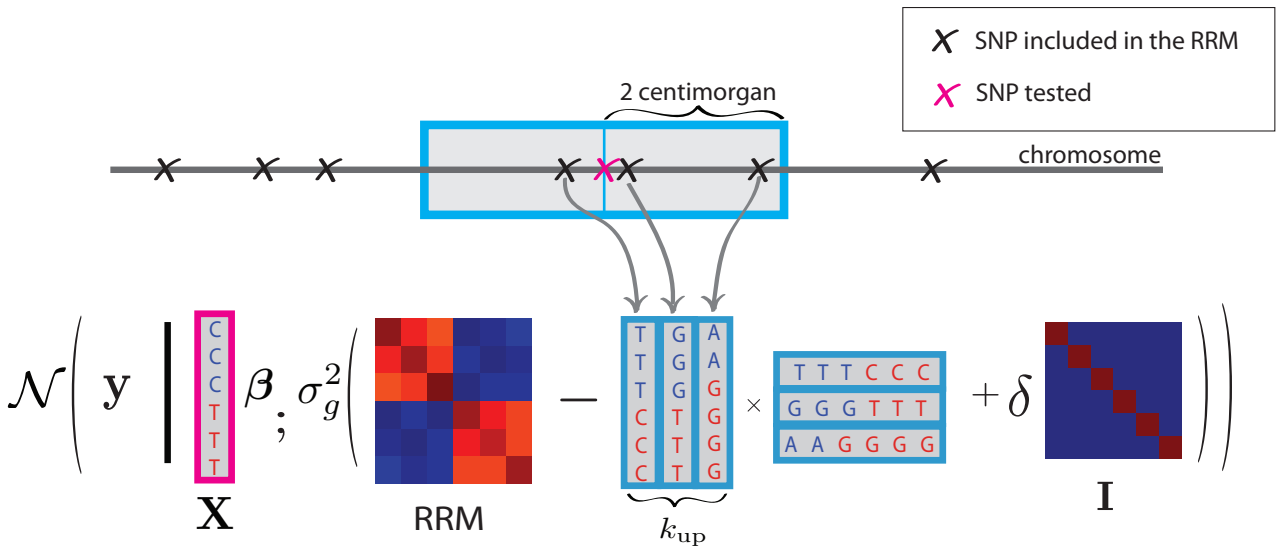
Proximal contamination

- Having a SNP in the similarity matrix that is linked to the SNP tested leads to loss in power.
- Correct by removing a sliding window around test-SNP from the similarity matrix.
- Correction is computed efficiently by subtracting a low-rank term.



Proximal contamination

- Having a SNP in the similarity matrix that is linked to the SNP tested leads to loss in power.
- Correct by removing a sliding window around test-SNP from the similarity matrix.
- Correction is computed efficiently by subtracting a low-rank term.



Inflammatory bowel disease [WTCCC, Nature 2007]

Algorithm parameters				**Algorithm performance**					
Name	SNP selection method	#SNPs in matrix	Avoid prox contam	λ_{GC}	False Positives	True Positives	Runtime (min) without speedup	Runtime (min) with speedup	Memory use (GB)
FaST-LMM-Select	Select	310	yes	1.08	0	100	1.3×10^3	45	<1
FaST-LMM all	All	All	yes	1.09	2	108	4.0×10^6	4567	86
FaST-LMM orig 310	Equi-spaced	310	yes	1.26	15	128	1.1×10^3	6	<1
FaST-LMM orig 4K	Equi-spaced	4000	yes	1.17	8	114	2.1×10^5	30	2
Traditional	All	All	no	0.97	2	64	42	NA	45

SNPs considered True Positive if:

- Reported in WTCCC paper [WTCCC, Nature 2007]
- Reported in meta analysis [Franke et al., Nat Gen 2010]
- In major histocompatibility complex (MHC) region

Marginal likelihood of variance component models

- Consider a linear model, accounting for a set of measured SNPs $\boldsymbol{X}$

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s, \sigma^2 \boldsymbol{I}\right)$$

- Choose identical Gaussian prior for all weights

$$p(\boldsymbol{\beta}) = \prod_{s=1}^{S} \mathcal{N}\left(\beta_s \mid 0, \sigma_{\mathrm{g}}^2\right)$$

- Marginal likelihood

$$p(\boldsymbol{y} \mid \boldsymbol{X},) = \int_{\boldsymbol{\beta}} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}\right) \mathcal{N}\left(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{I}\right)$$

- Number of hyperparameters independent of number of SNPs

Marginal likelihood of variance component models

- Consider a linear model, accounting for a set of measured SNPs $\boldsymbol{X}$

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s, \sigma^2 \boldsymbol{I}\right)$$

- Choose identical Gaussian prior for all weights

$$p(\boldsymbol{\beta}) = \prod_{s=1}^{S} \mathcal{N}\left(\beta_s \mid 0, \sigma_{\mathrm{g}}^2\right)$$

- Marginal likelihood

$$p(\boldsymbol{y} \mid \boldsymbol{X},) = \int_{\beta} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}\right) \mathcal{N}\left(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{I}\right)$$

- Number of hyperparameters independent of number of SNPs

Marginal likelihood of variance component models

- Consider a linear model, accounting for a set of measured SNPs $\boldsymbol{X}$

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s, \sigma^2 \boldsymbol{I}\right)$$

- Choose identical Gaussian prior for all weights

$$p(\boldsymbol{\beta}) = \prod_{s=1}^{S} \mathcal{N}\left(\beta_s \mid 0, \sigma_{\mathrm{g}}^2\right)$$

- Marginal likelihood

$$\begin{aligned} p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma^2, \sigma_{\mathrm{g}}^2) &= \int_{\boldsymbol{\beta}} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}\right) \mathcal{N}\left(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{I}\right) \\ &= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{X}\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I}\right) \end{aligned}$$

- Number of hyperparameters independent of number of SNPs

Marginal likelihood of variance component models

- Consider a linear model, accounting for a set of measured SNPs $\boldsymbol{X}$

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s, \sigma^2 \boldsymbol{I}\right)$$

- Choose identical Gaussian prior for all weights

$$p(\boldsymbol{\beta}) = \prod_{s=1}^{S} \mathcal{N}\left(\beta_s \mid 0, \sigma_{\mathrm{g}}^2\right)$$

- Marginal likelihood

$$\begin{aligned} p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma^2, \sigma_{\mathrm{g}}^2) &= \int_{\boldsymbol{\beta}} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}\right) \mathcal{N}\left(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{I}\right) \\ &= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{X}\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I}\right) \end{aligned}$$

- Number of hyperparameters independent of number of SNPs

Marginal likelihood of variance component models

Basis functions

- The analogous derivation can be repeated for a feature mapping ϕ

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{\phi}(\boldsymbol{x}_s)\beta_s, \sigma^2 \boldsymbol{I}\right) =$$

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}\right)$$

- Marginal likelihood

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma^2, \sigma_{\mathrm{g}}^2) = \int_{\boldsymbol{\beta}} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}\right) \mathcal{N}\left(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{I}\right)$$

$$= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \underbrace{\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})^{\top}}_{\boldsymbol{K}} + \sigma^2 \boldsymbol{I}\right)$$

- $\boldsymbol{K}$: (N x N) kernel or covariance induced by feature mapping ϕ.

Marginal likelihood of variance component models

Basis functions

- The analogous derivation can be repeated for a feature mapping ϕ

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{\phi}(\boldsymbol{x}_s)\beta_s, \sigma^2 \boldsymbol{I} \right) =$$

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\beta}, \sigma^2 \boldsymbol{I} \right)$$

- Marginal likelihood

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \sigma^2, \sigma_{\mathrm{g}}^2) = \int_{\boldsymbol{\beta}} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\beta}, \sigma^2 \boldsymbol{I} \right) \mathcal{N}\left(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{I} \right)$$

$$= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \underbrace{\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})^\top}_{\boldsymbol{K}} + \sigma^2 \boldsymbol{I} \right)$$

- $\boldsymbol{K}$: (N x N) kernel or covariance induced by feature mapping ϕ.

Marginal likelihood of variance component models

Basis functions

- The analogous derivation can be repeated for a feature mapping ϕ

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{\phi}(\boldsymbol{x}_s)\beta_s, \sigma^2 \boldsymbol{I} \right) =$$

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\beta}, \sigma^2 \boldsymbol{I} \right)$$

- Marginal likelihood

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \sigma^2, \sigma_{\mathrm{g}}^2) = \int_{\boldsymbol{\beta}} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\beta}, \sigma^2 \boldsymbol{I} \right) \mathcal{N}\left(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{I} \right)$$

$$= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \underbrace{\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})^\top}_{\boldsymbol{K}} + \sigma^2 \boldsymbol{I} \right)$$

- $\boldsymbol{K}$: (N x N) kernel or covariance induced by feature mapping ϕ.

Marginal likelihood of variance component models

Application to GWAS

The missing heritability paradox

- Complex traits are regulated by a large number of small effects
  - Human height: the best single SNP explains little variance.
  - But: height of the parents are highly predictive for the height of the child!

Marginal likelihood of variance component models

Application to GWAS

Linear additive models for complex traits

- Multiple linear regression model over causal SNPs

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \sum_{s \in \text{causal}} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Which SNPs are causal ?
  Approximation: consider all S available common SNPs [Yang et al. 2011]

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Causal SNPs either in the model or "tagged" by linkage disequilibrium to nearby common SNPs
- Uncertainty over causal SNPs: Prior on all SNP effects $p(\beta_s) = \mathcal{N}\left(\beta_s \mid 0, \sigma_{\text{g}}^2 / S\right)$
- Marginalize out weights

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma_{\text{g}}^2, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \boldsymbol{0} \,,\, \sigma_{\text{g}}^2 \sum_{s=1}^{S} \frac{1}{S} \boldsymbol{x}_s \boldsymbol{x}_s^\top + \sigma^2 \boldsymbol{I}\big)$$

- Perform maximum marginal likelihood estimation on σ_{g}^2 and σ^2.

Marginal likelihood of variance component models

Application to GWAS

Linear additive models for complex traits

- Multiple linear regression model over causal SNPs

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \sum_{s \in \text{causal}} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Which SNPs are causal ?
  Approximation: consider all S available common SNPs [Yang et al. 2011]

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Causal SNPs either in the model or "tagged" by linkage disequilibrium to nearby common SNPs
- Uncertainty over causal SNPs: Prior on all SNP effects $p(\beta_s) = \mathcal{N}\left(\beta_s \mid 0, \sigma_{\text{g}}^2 / S\right)$
- Marginalize out weights

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma_{\text{g}}^2, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \boldsymbol{0} \,,\, \sigma_{\text{g}}^2 \sum_{s=1}^{S} \frac{1}{S} \boldsymbol{x}_s \boldsymbol{x}_s^\top + \sigma^2 \boldsymbol{I}\big)$$

- Perform maximum marginal likelihood estimation on σ_{g}^2 and σ^2.

Marginal likelihood of variance component models

Application to GWAS

Linear additive models for complex traits

- Multiple linear regression model over causal SNPs

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \,|\, \sum_{s \in \text{causal}} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Which SNPs are causal ?
  Approximation: consider all S available common SNPs [Yang et al. 2011]

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \,|\, \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Causal SNPs either in the model or "tagged" by linkage disequilibrium to nearby common SNPs
- Uncertainty over causal SNPs: Prior on all SNP effects $p(\beta_s) = \mathcal{N}\left(\beta_s \mid 0, \sigma_{\text{g}}^2 / S\right)$
- Marginalize out weights

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \sigma_{\text{g}}^2, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \,|\, \boldsymbol{0} \,,\, \sigma_{\text{g}}^2 \sum_{s=1}^{S} \frac{1}{S} \boldsymbol{x}_s \boldsymbol{x}_s^\top + \sigma^2 \boldsymbol{I}\big)$$

- Perform maximum marginal likelihood estimation on σ_{g}^2 and σ^2.

Marginal likelihood of variance component models

Application to GWAS

Linear additive models for complex traits

- Multiple linear regression model over causal SNPs

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \,|\, \sum_{s \in \text{causal}} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Which SNPs are causal ?
  Approximation: consider all S available common SNPs [Yang et al. 2011]

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \,|\, \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Causal SNPs either in the model or "tagged" by linkage disequilibrium to nearby common SNPs
- Uncertainty over causal SNPs: Prior on all SNP effects $p(\beta_s) = \mathcal{N}\left(\beta_s \mid 0, \sigma_{\text{g}}^2 / S\right)$
- Marginalize out weights

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \sigma_{\text{g}}^2, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \,|\, \boldsymbol{0} \,,\, \sigma_{\text{g}}^2 \sum_{s=1}^{S} \frac{1}{S} \boldsymbol{x}_s \boldsymbol{x}_s^\top + \sigma^2 \boldsymbol{I}\big)$$

- Perform maximum marginal likelihood estimation on σ_{g}^2 and σ^2.

Marginal likelihood of variance component models

Application to GWAS

Linear additive models for complex traits

- Multiple linear regression model over causal SNPs

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \sum_{s \in \text{causal}} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Which SNPs are causal ?
  Approximation: consider all S available common SNPs [Yang et al. 2011]

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Causal SNPs either in the model or "tagged" by linkage disequilibrium to nearby common SNPs
- Uncertainty over causal SNPs: Prior on all SNP effects $p(\beta_s) = \mathcal{N}\left(\beta_s \mid 0, \sigma_{\text{g}}^2/S\right)$
- Marginalize out weights

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma_{\text{g}}^2, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \boldsymbol{0} \,,\, \sigma_{\text{g}}^2 \sum_{s=1}^{S} \frac{1}{S} \boldsymbol{x}_s \boldsymbol{x}_s^\top + \sigma^2 \boldsymbol{I}\big)$$

- Perform maximum marginal likelihood estimation on σ_{g}^2 and σ^2.

Marginal likelihood of variance component models

Application to GWAS

- Approximate variance model

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma_{\mathrm{g}}^2, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \frac{1}{S} \boldsymbol{X}\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I}\right)$$

- Genetic variance σ_{g}^2 across chromosomes
- (Narrow-sense) heritability

$$h^2 = \frac{\sigma_{\mathrm{g}}^2}{\sigma_{\mathrm{g}}^2 + \sigma^2} \approx \frac{\sum_{s=1}^{S} \beta_s^2}{\sum_{s=1}^{S} \beta_s^2 + \sigma^2}$$

- Narrow-sense refers to linear additive part of the heritability

Marginal likelihood of variance component models

Application to GWAS

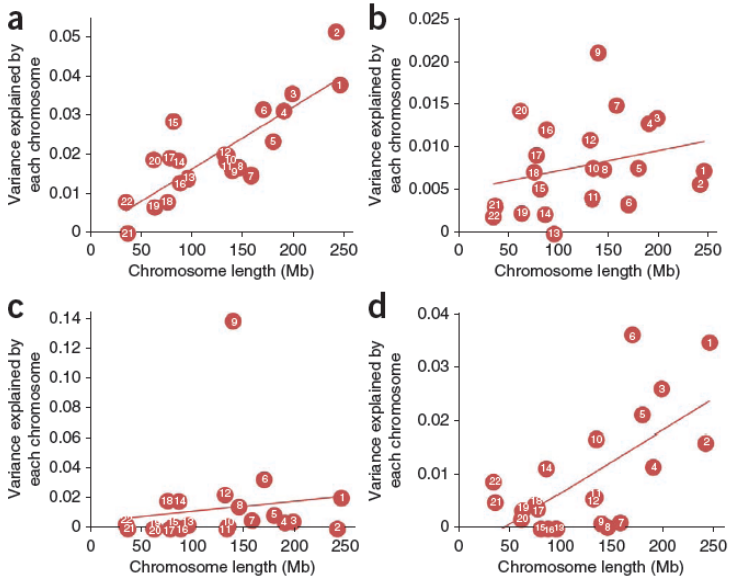
- Approximate variance model

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma_{\mathrm{g}}^2, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \frac{1}{S} \boldsymbol{X}\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I}\right)$$

- Genetic variance σ_{g}^2 across chromosomes
- (Narrow-sense) heritability

$$h^2 = \frac{\sigma_{\mathrm{g}}^2}{\sigma_{\mathrm{g}}^2 + \sigma^2} \approx \frac{\sum_{s=1}^{S} \beta_s^2}{\sum_{s=1}^{S} \beta_s^2 + \sigma^2}$$

- Narrow-sense refers to linear additive part of the heritability



[Yang et al. 2011]

Marginal likelihood of variance component models

Application to GWAS

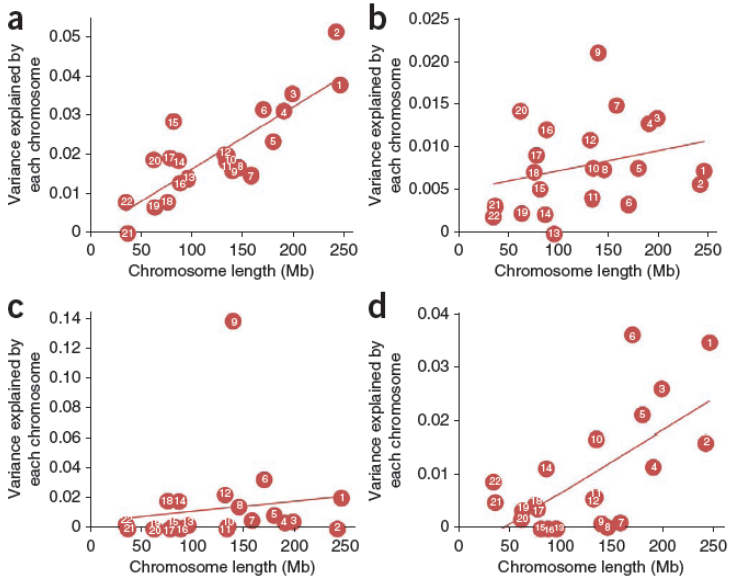
- Approximate variance model

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma_{\mathrm{g}}^2, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \frac{1}{S} \boldsymbol{X}\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I}\right)$$

- Genetic variance σ_{g}^2 across chromosomes
- (Narrow-sense) heritability

$$h^2 = \frac{\sigma_{\mathrm{g}}^2}{\sigma_{\mathrm{g}}^2 + \sigma^2} \approx \frac{\sum_{s=1}^{S} \beta_s^2}{\sum_{s=1}^{S} \beta_s^2 + \sigma^2}$$

- Narrow-sense refers to linear additive part of the heritability



[Yang et al. 2011]

Marginal likelihood of variance component models

Application to GWAS

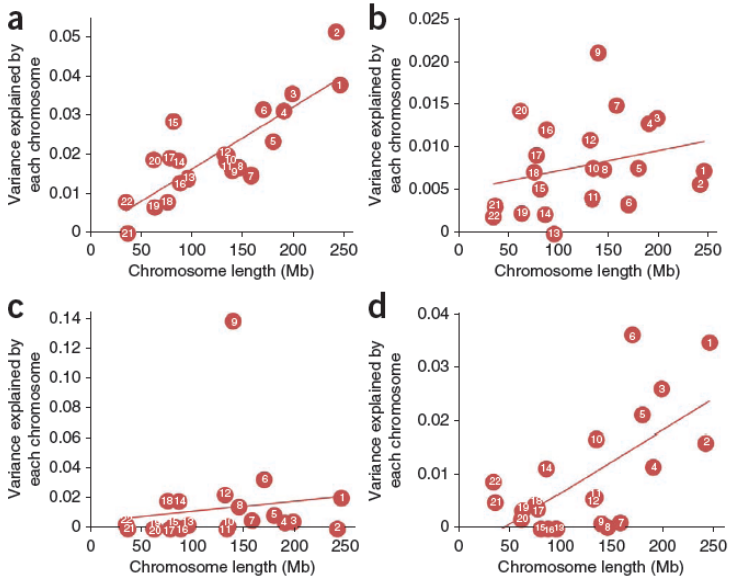
- Approximate variance model

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma_{\mathrm{g}}^2, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \frac{1}{S} \boldsymbol{X}\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I}\right)$$

- Genetic variance σ_{g}^2 across chromosomes
- (Narrow-sense) heritability

$$h^2 = \frac{\sigma_{\mathrm{g}}^2}{\sigma_{\mathrm{g}}^2 + \sigma^2} \approx \frac{\sum_{s=1}^{S} \beta_s^2}{\sum_{s=1}^{S} \beta_s^2 + \sigma^2}$$

- Narrow-sense refers to linear additive part of the heritability



[Yang et al. 2011]

Marginal likelihood of variance component models

- Consider a linear model, accounting for a set of measured SNPs $\boldsymbol{X}$

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s, \sigma^2 \boldsymbol{I}\right)$$

- Choose identical Gaussian prior for all weights

$$p(\boldsymbol{\beta}) = \prod_{s=1}^{S} \mathcal{N}\left(\beta_s \mid 0, \sigma_{\mathrm{g}}^2\right)$$

- Marginal likelihood

$$p(\boldsymbol{y} \mid \boldsymbol{X},) = \int_{\boldsymbol{\beta}} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}\right) \mathcal{N}\left(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{I}\right)$$

- Number of hyperparameters independent of number of SNPs

Marginal likelihood of variance component models

- Consider a linear model, accounting for a set of measured SNPs $\boldsymbol{X}$

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s, \sigma^2 \boldsymbol{I}\right)$$

- Choose identical Gaussian prior for all weights

$$p(\boldsymbol{\beta}) = \prod_{s=1}^{S} \mathcal{N}\left(\beta_s \mid 0, \sigma_{\mathrm{g}}^2\right)$$

- Marginal likelihood

$$p(\boldsymbol{y} \mid \boldsymbol{X},) = \int_{\beta} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}\right) \mathcal{N}\left(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{I}\right)$$

- Number of hyperparameters independent of number of SNPs

Marginal likelihood of variance component models

- Consider a linear model, accounting for a set of measured SNPs $\boldsymbol{X}$

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s, \sigma^2 \boldsymbol{I}\right)$$

- Choose identical Gaussian prior for all weights

$$p(\boldsymbol{\beta}) = \prod_{s=1}^{S} \mathcal{N}\left(\beta_s \mid 0, \sigma_{\mathrm{g}}^2\right)$$

- Marginal likelihood

$$\begin{aligned} p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma^2, \sigma_{\mathrm{g}}^2) &= \int_{\boldsymbol{\beta}} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}\right) \mathcal{N}\left(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{I}\right) \\ &= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{X}\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I}\right) \end{aligned}$$

- Number of hyperparameters independent of number of SNPs

Marginal likelihood of variance component models

- Consider a linear model, accounting for a set of measured SNPs $\boldsymbol{X}$

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s, \sigma^2 \boldsymbol{I}\right)$$

- Choose identical Gaussian prior for all weights

$$p(\boldsymbol{\beta}) = \prod_{s=1}^{S} \mathcal{N}\left(\beta_s \mid 0, \sigma_{\mathrm{g}}^2\right)$$

- Marginal likelihood

$$\begin{aligned} p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma^2, \sigma_{\mathrm{g}}^2) &= \int_{\boldsymbol{\beta}} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}\right) \mathcal{N}\left(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{I}\right) \\ &= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{X}\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I}\right) \end{aligned}$$

- Number of hyperparameters independent of number of SNPs

Marginal likelihood of variance component models

Basis functions

- The analogous derivation can be repeated for a feature mapping ϕ

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{\phi}(\boldsymbol{x}_s)\beta_s, \sigma^2 \boldsymbol{I}\right) =$$

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}\right)$$

- Marginal likelihood

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma^2, \sigma_{\mathrm{g}}^2) = \int_{\boldsymbol{\beta}} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}\right) \mathcal{N}\left(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{I}\right)$$

$$= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \underbrace{\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})^{\top}}_{\boldsymbol{K}} + \sigma^2 \boldsymbol{I}\right)$$

- $\boldsymbol{K}$: (N x N) kernel or covariance induced by feature mapping ϕ.

Marginal likelihood of variance component models

Basis functions

- The analogous derivation can be repeated for a feature mapping ϕ

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{\phi}(\boldsymbol{x}_s)\beta_s, \sigma^2 \boldsymbol{I} \right) =$$

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\beta}, \sigma^2 \boldsymbol{I} \right)$$

- Marginal likelihood

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \sigma^2, \sigma_{\mathrm{g}}^2) = \int_{\boldsymbol{\beta}} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\beta}, \sigma^2 \boldsymbol{I} \right) \mathcal{N}\left(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{I} \right)$$

$$= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \underbrace{\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})^\top}_{\boldsymbol{K}} + \sigma^2 \boldsymbol{I} \right)$$

- $\boldsymbol{K}$: (N x N) kernel or covariance induced by feature mapping ϕ.

Marginal likelihood of variance component models

Basis functions

- The analogous derivation can be repeated for a feature mapping ϕ

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{\phi}(\boldsymbol{x}_s)\beta_s, \sigma^2 \boldsymbol{I} \right) =$$

$$\mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\beta}, \sigma^2 \boldsymbol{I} \right)$$

- Marginal likelihood

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \sigma^2, \sigma_{\mathrm{g}}^2) = \int_{\boldsymbol{\beta}} \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\beta}, \sigma^2 \boldsymbol{I} \right) \mathcal{N}\left(\boldsymbol{\beta} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \boldsymbol{I} \right)$$

$$= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \underbrace{\boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{\Phi}(\boldsymbol{X})^\top}_{\boldsymbol{K}} + \sigma^2 \boldsymbol{I} \right)$$

- $\boldsymbol{K}$: (N x N) kernel or covariance induced by feature mapping ϕ.

Marginal likelihood of variance component models

Application to GWAS

The missing heritability paradox

- Complex traits are regulated by a large number of small effects
  - Human height: the best single SNP explains little variance.
  - But: height of the parents are highly predictive for the height of the child!

Marginal likelihood of variance component models

Application to GWAS

Linear additive models for complex traits

- Multiple linear regression model over causal SNPs

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \sum_{s \in \text{causal}} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Which SNPs are causal ?
  Approximation: consider all S available common SNPs [Yang et al. 2011]

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Causal SNPs either in the model or "tagged" by linkage disequilibrium to nearby common SNPs
- Uncertainty over causal SNPs: Prior on all SNP effects $p(\beta_s) = \mathcal{N}\left(\beta_s \mid 0, \sigma_{\text{g}}^2 / S\right)$
- Marginalize out weights

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma_{\text{g}}^2, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \boldsymbol{0} \,,\, \sigma_{\text{g}}^2 \sum_{s=1}^{S} \frac{1}{S} \boldsymbol{x}_s \boldsymbol{x}_s^\top + \sigma^2 \boldsymbol{I}\big)$$

- Perform maximum marginal likelihood estimation on σ_{g}^2 and σ^2.

Marginal likelihood of variance component models

Application to GWAS

Linear additive models for complex traits

- Multiple linear regression model over causal SNPs

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \sum_{s \in \text{causal}} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Which SNPs are causal ?
  Approximation: consider all S available common SNPs [Yang et al. 2011]

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Causal SNPs either in the model or "tagged" by linkage disequilibrium to nearby common SNPs
- Uncertainty over causal SNPs: Prior on all SNP effects $p(\beta_s) = \mathcal{N}\left(\beta_s \mid 0, \sigma_{\text{g}}^2 / S\right)$
- Marginalize out weights

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma_{\text{g}}^2, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \boldsymbol{0} \,,\, \sigma_{\text{g}}^2 \sum_{s=1}^{S} \frac{1}{S} \boldsymbol{x}_s \boldsymbol{x}_s^\top + \sigma^2 \boldsymbol{I}\big)$$

- Perform maximum marginal likelihood estimation on σ_{g}^2 and σ^2.

Marginal likelihood of variance component models

Application to GWAS

Linear additive models for complex traits

- Multiple linear regression model over causal SNPs

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \,|\, \sum_{s \in \text{causal}} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Which SNPs are causal ?
  Approximation: consider all S available common SNPs [Yang et al. 2011]

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \,|\, \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Causal SNPs either in the model or "tagged" by linkage disequilibrium to nearby common SNPs
- Uncertainty over causal SNPs: Prior on all SNP effects $p(\beta_s) = \mathcal{N}\left(\beta_s \mid 0, \sigma_{\text{g}}^2 / S\right)$
- Marginalize out weights

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \sigma_{\text{g}}^2, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \,|\, \boldsymbol{0} \,,\, \sigma_{\text{g}}^2 \sum_{s=1}^{S} \frac{1}{S} \boldsymbol{x}_s \boldsymbol{x}_s^\top + \sigma^2 \boldsymbol{I}\big)$$

- Perform maximum marginal likelihood estimation on σ_{g}^2 and σ^2.

Marginal likelihood of variance component models

Application to GWAS

Linear additive models for complex traits

- Multiple linear regression model over causal SNPs

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \,|\, \sum_{s \in \text{causal}} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Which SNPs are causal ?
  Approximation: consider all S available common SNPs [Yang et al. 2011]

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \,|\, \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Causal SNPs either in the model or "tagged" by linkage disequilibrium to nearby common SNPs
- Uncertainty over causal SNPs: Prior on all SNP effects $p(\beta_s) = \mathcal{N}\left(\beta_s \mid 0, \sigma_{\text{g}}^2 / S\right)$
- Marginalize out weights

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \sigma_{\text{g}}^2, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \,|\, \boldsymbol{0} \,,\, \sigma_{\text{g}}^2 \sum_{s=1}^{S} \frac{1}{S} \boldsymbol{x}_s \boldsymbol{x}_s^\top + \sigma^2 \boldsymbol{I}\big)$$

- Perform maximum marginal likelihood estimation on σ_{g}^2 and σ^2.

Marginal likelihood of variance component models

Application to GWAS

Linear additive models for complex traits

- Multiple linear regression model over causal SNPs

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \sum_{s \in \text{causal}} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Which SNPs are causal ?
  Approximation: consider all S available common SNPs [Yang et al. 2011]

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \sum_{s=1}^{S} \boldsymbol{x}_s \beta_s \,,\, \sigma^2 \boldsymbol{I}\big)$$

- Causal SNPs either in the model or "tagged" by linkage disequilibrium to nearby common SNPs
- Uncertainty over causal SNPs: Prior on all SNP effects $p(\beta_s) = \mathcal{N}\left(\beta_s \mid 0, \sigma_{\text{g}}^2/S\right)$
- Marginalize out weights

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma_{\text{g}}^2, \sigma^2) = \mathcal{N}\big(\boldsymbol{y} \mid \boldsymbol{0} \,,\, \sigma_{\text{g}}^2 \sum_{s=1}^{S} \frac{1}{S} \boldsymbol{x}_s \boldsymbol{x}_s^\top + \sigma^2 \boldsymbol{I}\big)$$

- Perform maximum marginal likelihood estimation on σ_{g}^2 and σ^2.

Marginal likelihood of variance component models

Application to GWAS

- Approximate variance model

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma_{\mathrm{g}}^2, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \frac{1}{S} \boldsymbol{X}\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I}\right)$$

- Genetic variance σ_{g}^2 across chromosomes
- (Narrow-sense) heritability

$$h^2 = \frac{\sigma_{\mathrm{g}}^2}{\sigma_{\mathrm{g}}^2 + \sigma^2} \approx \frac{\sum_{s=1}^{S} \beta_s^2}{\sum_{s=1}^{S} \beta_s^2 + \sigma^2}$$

- Narrow-sense refers to linear additive part of the heritability

Marginal likelihood of variance component models

Application to GWAS

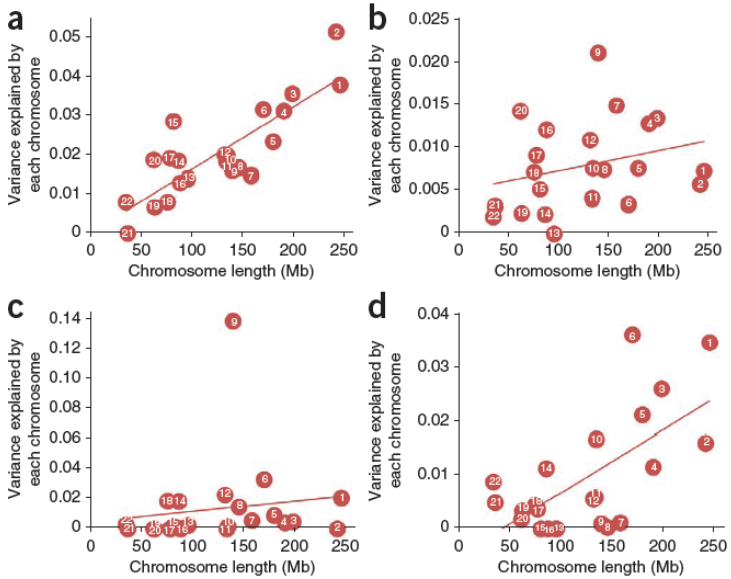
- Approximate variance model

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma_{\mathrm{g}}^2, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \frac{1}{S} \boldsymbol{X}\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I}\right)$$

- Genetic variance σ_{g}^2 across chromosomes
- (Narrow-sense) heritability

$$h^2 = \frac{\sigma_{\mathrm{g}}^2}{\sigma_{\mathrm{g}}^2 + \sigma^2} \approx \frac{\sum_{s=1}^{S} \beta_s^2}{\sum_{s=1}^{S} \beta_s^2 + \sigma^2}$$

- Narrow-sense refers to linear additive part of the heritability



[Yang et al. 2011]

Marginal likelihood of variance component models

Application to GWAS

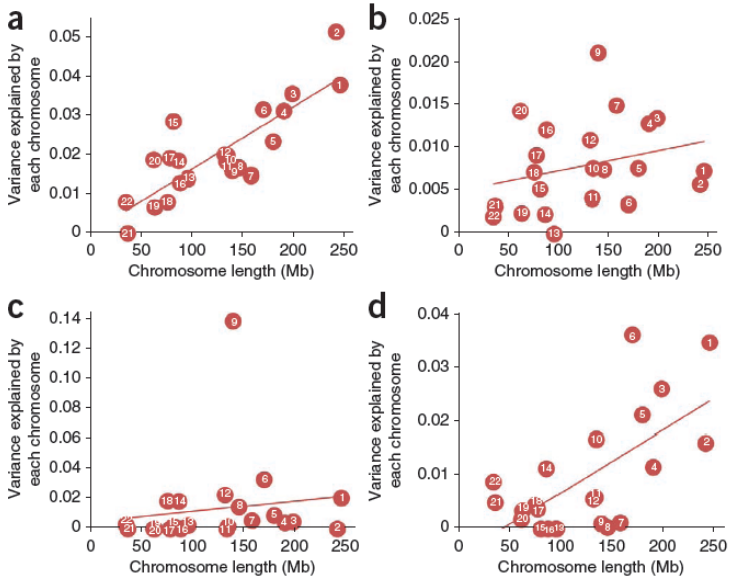
- Approximate variance model

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma_{\mathrm{g}}^2, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \frac{1}{S} \boldsymbol{X}\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I}\right)$$

- Genetic variance σ_{g}^2 across chromosomes
- (Narrow-sense) heritability

$$h^2 = \frac{\sigma_{\mathrm{g}}^2}{\sigma_{\mathrm{g}}^2 + \sigma^2} \approx \frac{\sum_{s=1}^{S} \beta_s^2}{\sum_{s=1}^{S} \beta_s^2 + \sigma^2}$$

- Narrow-sense refers to linear additive part of the heritability



[Yang et al. 2011]

Marginal likelihood of variance component models

Application to GWAS

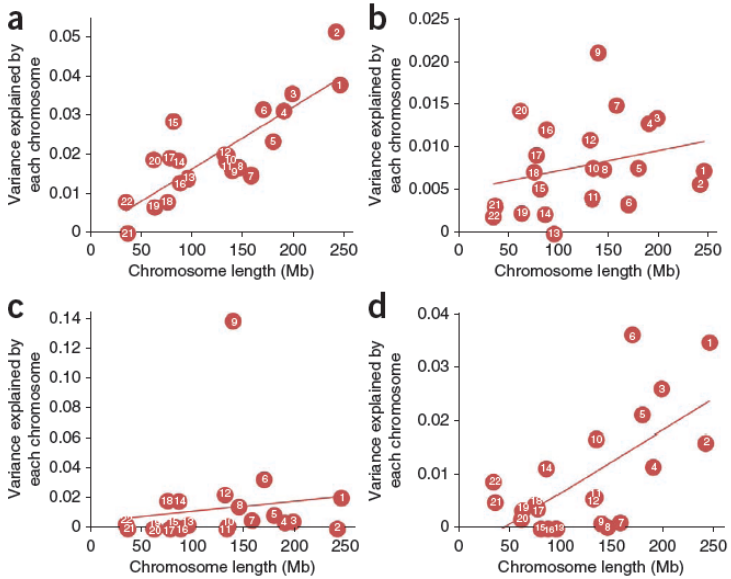
- Approximate variance model

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma_{\mathrm{g}}^2, \sigma^2) = \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \sigma_{\mathrm{g}}^2 \frac{1}{S} \boldsymbol{X}\boldsymbol{X}^\top + \sigma^2 \boldsymbol{I}\right)$$

- Genetic variance σ_{g}^2 across chromosomes
- (Narrow-sense) heritability

$$h^2 = \frac{\sigma_{\mathrm{g}}^2}{\sigma_{\mathrm{g}}^2 + \sigma^2} \approx \frac{\sum_{s=1}^{S} \beta_s^2}{\sum_{s=1}^{S} \beta_s^2 + \sigma^2}$$

- Narrow-sense refers to linear additive part of the heritability



[Yang et al. 2011]

ï≫¿

Outline

FaST-LMM-Set

- Aggregate effects within a gene or pathway
- Variance component test (e.g. [Wu et al 2011])
- Correct for population structure
- Perform a Likelihood ratio test

$$\mathcal{N}\left(\boldsymbol{y} \mid \underbrace{\sum_{j=1}^{s} \boldsymbol{X}_s \beta_s}_{\neq 0?}, \quad \underbrace{\sigma^2 \boldsymbol{I}}_{\text{noise}}\right)$$

FaST-LMM-Set

- Aggregate effects within a gene or pathway
- Variance component test (e.g. [Wu et al 2011])
- Correct for population structure
- Perform a Likelihood ratio test

$$\mathcal{N}\left(\boldsymbol{y} \mid \underbrace{\sum_{j=1}^{s} \boldsymbol{X}_s \beta_s}_{\neq 0?}, \quad \underbrace{\sigma^2 \boldsymbol{I}}_{\text{noise}}\right) \prod_{j=1}^{s} \mathcal{N}\left(\beta_s \mid 0, \sigma_c^2\right)$$

FaST-LMM-Set

- Aggregate effects within a gene or pathway
- Variance component test (e.g. [Wu et al 2011])
- Correct for population structure
- Perform a Likelihood ratio test

$$\int \mathcal{N}\left(\boldsymbol{y} \mid \underbrace{\sum_{j=1}^{s} \boldsymbol{X}_s \beta_s}_{\neq 0?}, \quad \underbrace{\sigma^2 \boldsymbol{I}}_{\text{noise}} \right) \prod_{j=1}^{s} \mathcal{N}\left(\beta_s \mid 0, \sigma_c^2 \right) \mathrm{d}\boldsymbol{\beta}$$

FaST-LMM-Set

- Aggregate effects within a gene or pathway
- Variance component test (e.g. [Wu et al 2011])
- Correct for population structure
- Perform a Likelihood ratio test

$$\int \mathcal{N}\left(\boldsymbol{y} \mid \underbrace{\sum_{j=1}^{s} \boldsymbol{X}_s \beta_s}_{\neq 0?}\,, \qquad \underbrace{\sigma^2 \boldsymbol{I}}_{\text{noise}} \right) \prod_{j=1}^{s} \mathcal{N}\left(\beta_s \mid 0\,, \sigma_c^2 \right) \mathrm{d}\boldsymbol{\beta}$$

$$= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}\,, \underbrace{\sigma_c^2}_{>0?} \underbrace{\left(\sum_{j=1}^{s} \boldsymbol{X}_s \boldsymbol{X}_s^\top \right)}_{\boldsymbol{K}_s} \qquad + \sigma^2 \boldsymbol{I} \right)$$

FaST-LMM-Set

- Aggregate effects within a gene or pathway
- Variance component test (e.g. [Wu et al 2011])
- Correct for population structure
- Perform a Likelihood ratio test

$$\int \mathcal{N}\left(\boldsymbol{y} \mid \underbrace{\sum_{j=1}^{s} \boldsymbol{X}_s \beta_s}_{\neq 0?}\,, \quad \underbrace{\sigma_{\mathrm{g}}^2 \boldsymbol{K}_c}_{\text{confounding structure}} + \underbrace{\sigma^2 \boldsymbol{I}}_{\text{noise}} \right) \prod_{j=1}^{s} \mathcal{N}\left(\beta_s \mid 0\,, \sigma_c^2 \right) \mathrm{d}\boldsymbol{\beta}$$

$$= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}\,, \underbrace{\sigma_c^2}_{>0?} \underbrace{\left(\sum_{j=1}^{s} \boldsymbol{X}_s \boldsymbol{X}_s^\top \right)}_{\boldsymbol{K}_s} + \sigma_{\mathrm{g}}^2 \boldsymbol{K}_c + \sigma^2 \boldsymbol{I} \right)$$

FaST-LMM-Set

- Aggregate effects within a gene or pathway
- Variance component test (e.g. [Wu et al 2011])
- Correct for population structure
- Perform a Likelihood ratio test

$$\int \mathcal{N}\left(\boldsymbol{y} \mid \underbrace{\sum_{j=1}^{s} \boldsymbol{X}_s \beta_s}_{\neq 0?}\,, \quad \underbrace{\sigma_{\mathrm{g}}^2 \boldsymbol{K}_c}_{\text{confounding structure}} + \underbrace{\sigma^2 \boldsymbol{I}}_{\text{noise}} \right) \prod_{j=1}^{s} \mathcal{N}\left(\beta_s \mid 0\,, \sigma_c^2 \right) \mathrm{d}\boldsymbol{\beta}$$

$$= \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}\,, \underbrace{\sigma_c^2}_{>0?} \underbrace{\left(\sum_{j=1}^{s} \boldsymbol{X}_s \boldsymbol{X}_s^\top \right)}_{\boldsymbol{K}_s} + \sigma_{\mathrm{g}}^2 \boldsymbol{K}_c + \sigma^2 \boldsymbol{I} \right)$$

Variance component testing

Likelihood-Ratio test vs. Score test

- For convenience most variance component tests use Score tests (e.g. [Wu et al. 2011])
- For the likelihood ratio test (LRT) no exact null distribution is known (Permutations are prohibitive!)
- Small number of permutations
- a parametric fit to get an accurate and efficient estimate of the null distribution of the LRT.
- Empirically, the LRT outperforms the score test in terms of power

Variance component testing

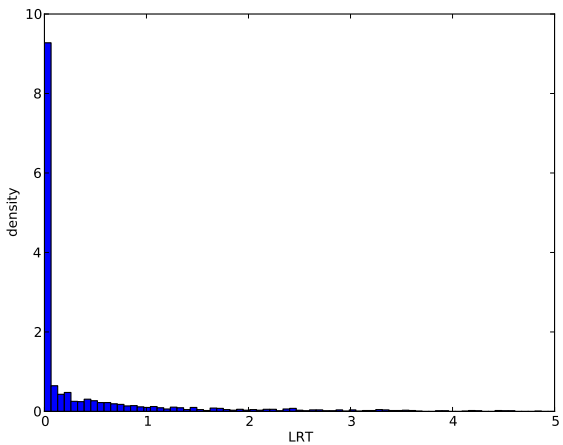
Likelihood-Ratio test vs. Score test

- For convenience most variance component tests use Score tests (e.g. [Wu et al. 2011])
- For the likelihood ratio test (LRT) no exact null distribution is known (Permutations are prohibitive!)
- Small number of permutations
- a parametric fit to get an accurate and efficient estimate of the null distribution of the LRT.
- Empirically, the LRT outperforms the score test in terms of power

Variance component testing

Likelihood-Ratio test vs. Score test

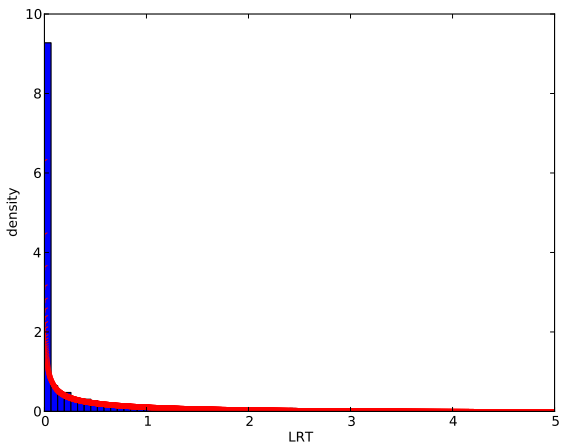
- For convenience most variance component tests use Score tests (e.g. [Wu et al. 2011])
- For the likelihood ratio test (LRT) no exact null distribution is known (Permutations are prohibitive!)
- Small number of permutations
- a parametric fit to get an accurate and efficient estimate of the null distribution of the LRT.
- Empirically, the LRT outperforms the score test in terms of power



Variance component testing

Likelihood-Ratio test vs. Score test

- For convenience most variance component tests use Score tests (e.g. [Wu et al. 2011])
- For the likelihood ratio test (LRT) no exact null distribution is known (Permutations are prohibitive!)
- Small number of permutations
- a parametric fit to get an accurate and efficient estimate of the null distribution of the LRT.
- Empirically, the LRT outperforms the score test in terms of power



Variance component testing

Likelihood-Ratio test vs. Score test

- For convenience most variance component tests use Score tests (e.g. [Wu et al. 2011])
- For the likelihood ratio test (LRT) no exact null distribution is known (Permutations are prohibitive!)
- Small number of permutations
- a parametric fit to get an accurate and efficient estimate of the null distribution of the LRT.
- Empirically, the LRT outperforms the score test in terms of power

Type 1 error:

Significance Level	$\alpha = 10^{-5}$	$\alpha = 10^{-4}$	$\alpha = 10^{-3}$
Fast-LMM-Set	1×10^{-5}	1.21×10^{-4}	1.01×10^{-3}
$0.5\chi_0^2 + 0.5\chi_1^2$	$5 \times 10^{-6}\star$	$4 \times 10^{-5}\star$	$4.55 \times 10^{-4}\star$

Variance component testing

Likelihood-Ratio test vs. Score test

- For convenience most variance component tests use Score tests (e.g. [Wu et al. 2011])
- For the likelihood ratio test (LRT) no exact null distribution is known (Permutations are prohibitive!)
- Small number of permutations
- a parametric fit to get an accurate and efficient estimate of the null distribution of the LRT.
- Empirically, the LRT outperforms the score test in terms of power