Linear models for GWAS

II: Linear mixed models

Christoph Lippert

Microsoft Research, Los Angeles, USA

Microsoft Research

Current topics in computational biology
UCLA
October 15^{th}, 2012

Course structure

October 15^{th}

- Introduction
  - Terminology
  - Study design
  - Data preparation
  - Challenges and pitfalls
  - Course overview
- Linear regression
  - Parameter estimation
  - Statistical testing

October 17^{th}

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Phenotype prediction

Course structure

October 15^{th}

- Introduction
  - Terminology
  - Study design
  - Data preparation
  - Challenges and pitfalls
  - Course overview
- Linear regression
  - Parameter estimation
  - Statistical testing

October 17^{th}

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Phenotype prediction

Course structure

October 15^{th}

- Introduction
  - Terminology
  - Study design
  - Data preparation
  - Challenges and pitfalls
  - Course overview
- Linear regression
  - Parameter estimation
  - Statistical testing

October 17^{th}

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Permutation procedure

Course structure

October 15^{th}

- Introduction
  - Terminology
  - Study design
  - Data preparation
  - Challenges and pitfalls
  - Course overview
- Linear regression
  - Parameter estimation
  - Statistical testing

October 17^{th}

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Phenotype prediction

Course structure

October 15^{th}

- Introduction
  - Terminology
  - Study design
  - Data preparation
  - Challenges and pitfalls
  - Course overview
- Linear regression
  - Parameter estimation
  - Statistical testing

October 17^{th}

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Phenotype prediction

Course structure

October 15^{th}

- Introduction
  - Terminology
  - Study design
  - Data preparation
  - Challenges and pitfalls
  - Course overview
- Linear regression
  - Parameter estimation
  - Statistical testing

October 17^{th}

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Phenotype prediction

Course structure

October 15^{th}

- Introduction
  - Terminology
  - Study design
  - Data preparation
  - Challenges and pitfalls
  - Course overview
- Linear regression
  - Parameter estimation
  - Statistical testing

October 17^{th}

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Phenotype prediction

Course structure

October 15^{th}

- Introduction
  - Terminology
  - Study design
  - Data preparation
  - Challenges and pitfalls
  - Course overview
- Linear regression
  - Parameter estimation
  - Statistical testing

October 17^{th}

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Phenotype prediction

Course structure

October 15^{th}

- Introduction
  - Terminology
  - Study design
  - Data preparation
  - Challenges and pitfalls
  - Course overview
- Linear regression
  - Parameter estimation
  - Statistical testing

October 17^{th}

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Phenotype prediction

Probabilities

- Let X be a random variable, defined over a set $\mathcal{X}$ or measurable space.
- $P(X = x)$ denotes the probability that X takes value x, short $p(x)$.
  - Probabilities are positive, $P(X = x) \geq 0$
  - Probabilities sum to one

Probabilities

- Let X be a random variable, defined over a set $\mathcal{X}$ or measurable space.
- $P(X = x)$ denotes the probability that X takes value x, short $p(x)$.
  - Probabilities are positive, $P(X = x) \geq 0$
  - Probabilities sum to one

$$\int_{x \in \mathcal{X}} p(x)dx = 1 \qquad \sum_{x \in \mathcal{X}} p(x) = 1$$

Probabilities

- Let X be a random variable, defined over a set $\mathcal{X}$ or measurable space.
- $P(X = x)$ denotes the probability that X takes value x, short $p(x)$.
  - Probabilities are positive, $P(X = x) \geq 0$
  - Probabilities sum to one

$$\int_{x \in \mathcal{X}} p(x)dx = 1 \qquad \sum_{x \in \mathcal{X}} p(x) = 1$$

Probabilities

- Let X be a random variable, defined over a set $\mathcal{X}$ or measurable space.
- $P(X = x)$ denotes the probability that X takes value x, short $p(x)$.
  - Probabilities are positive, $P(X = x) \geq 0$
  - Probabilities sum to one

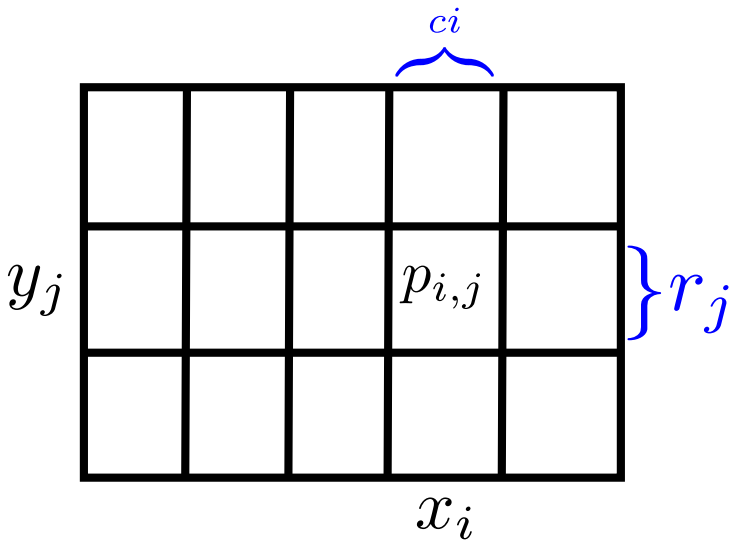
$$\int_{x \in \mathcal{X}} p(x)dx = 1 \qquad \sum_{x \in \mathcal{X}} p(x) = 1$$

Probabilities

- Let X be a random variable, defined over a set $\mathcal{X}$ or measurable space.
- $P(X = x)$ denotes the probability that X takes value x, short $p(x)$.
  - Probabilities are positive, $P(X = x) \geq 0$
  - Probabilities sum to one

$$\int_{x \in \mathcal{X}} p(x)dx = 1 \qquad \sum_{x \in \mathcal{X}} p(x) = 1$$

Probability Theory



Joint Probability

$$P(X = x_i, Y = y_j) = \frac{n_{i,j}}{N}$$

Marginal Probability

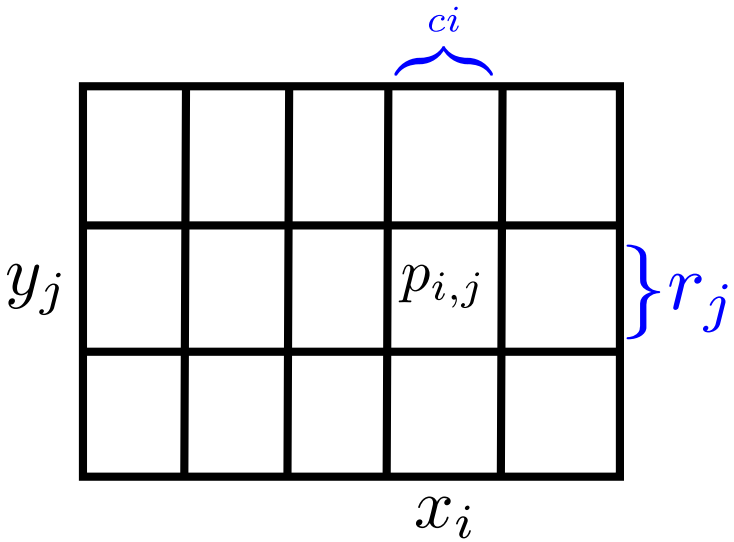
$$P(X = x_i) = \frac{c_i}{N}$$

Conditional Probability

$$P(Y = y_j \,|\, X = x_i) = \frac{n_{i,j}}{c_i}$$

(C.M. Bishop, Pattern Recognition and Machine Learning)

Probability Theory



Marginal Probability

$$P(X = x_i) = \frac{c_i}{N}$$

Conditional Probability

$$P(Y = y_j \,|\, X = x_i) = \frac{n_{i,j}}{c_i}$$

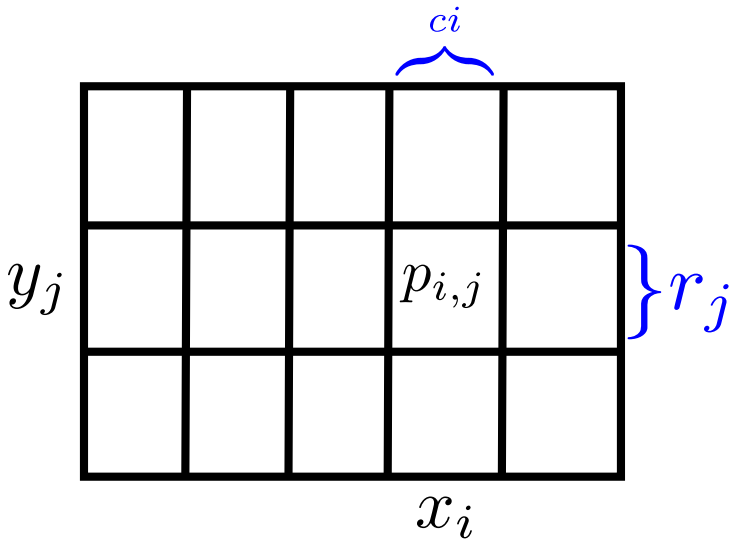
Product Rule

$$P(X = x_i, Y = y_j) = \frac{n_{i,j}}{N} = \frac{n_{i,j}}{c_i} \cdot \frac{c_i}{N}$$

$$= P(Y = y_j \,|\, X = x_i) P(X = x_i)$$

(C.M. Bishop, Pattern Recognition and Machine Learning)

Probability Theory



Sum Rule

$$P(X = x_i) = \frac{c_i}{N} = \frac{1}{N} \sum_{j=1}^{L} n_{i,j}$$

$$= \sum_{j} P(X = x_i, Y = y_j)$$

Product Rule

$$P(X = x_i, Y = y_j) = \frac{n_{i,j}}{N} = \frac{n_{i,j}}{c_i} \cdot \frac{c_i}{N}$$

$$= P(Y = y_j \,|\, X = x_i) P(X = x_i)$$

(C.M. Bishop, Pattern Recognition and Machine Learning)

The Rules of Probability

Sum & Product Rule

Sum Rule	$p(x) = \sum_y p(x, y)$
Product Rule	$p(x, y) = p(y \mid x)p(x)$

The Rules of Probability

Bayes Theorem

- Using the product rule we obtain

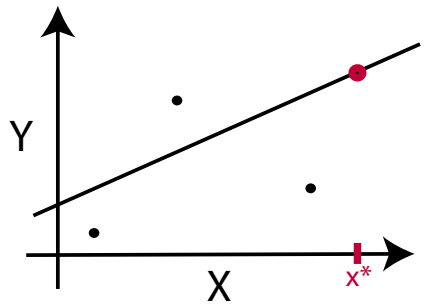
$$p(y \mid x) = \frac{p(x \mid y)p(y)}{p(x)}$$

$$p(x) = \sum_{y} p(x \mid y)p(y)$$

Bayesian probability calculus

- Bayes rule is the basis for inference and learning.
- Assume we have a model with parameters $\boldsymbol{\theta}$, e.g.

$$y = \theta_0 + \theta_1 \cdot x$$



- Goal: learn parameters $\boldsymbol{\theta}$ given Data $\mathcal{D}$.

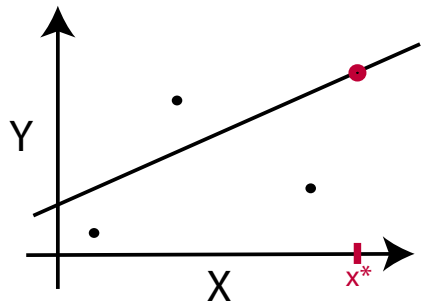
$$p(\boldsymbol{\theta} \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\mathcal{D})}$$

  - Posterior
  - Likelihood
  - Prior

Bayesian probability calculus

- Bayes rule is the basis for inference and learning.
- Assume we have a model with parameters $\boldsymbol{\theta}$, e.g.

$$y = \theta_0 + \theta_1 \cdot x$$



- Goal: learn parameters $\boldsymbol{\theta}$ given Data $\mathcal{D}$.

$$p(\boldsymbol{\theta} \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\mathcal{D})}$$

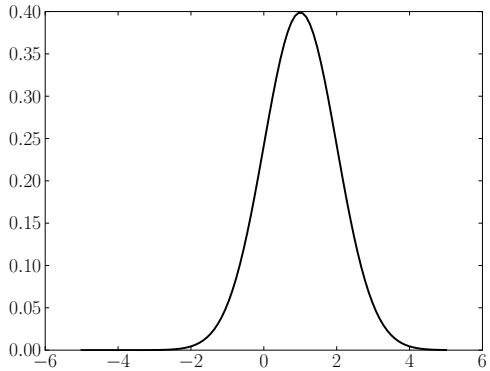
$$\text{posterior} \propto \text{likelihood} \cdot \text{prior}$$

- Posterior
- Likelihood
- Prior

Probability distributions

- Gaussian

$$p(x \mid \mu, \sigma^2) = \mathcal{N}(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$



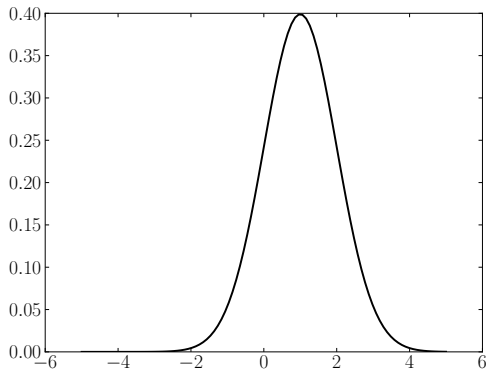
- Multivariate Gaussian

$$p(x \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$$
$$= \frac{1}{\sqrt{|2\pi\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

Probability distributions

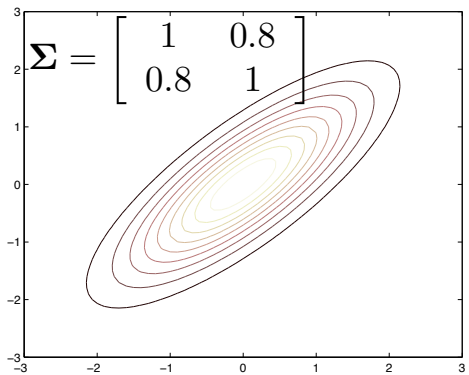
- Gaussian

$$p(x \mid \mu, \sigma^2) = \mathcal{N}(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$



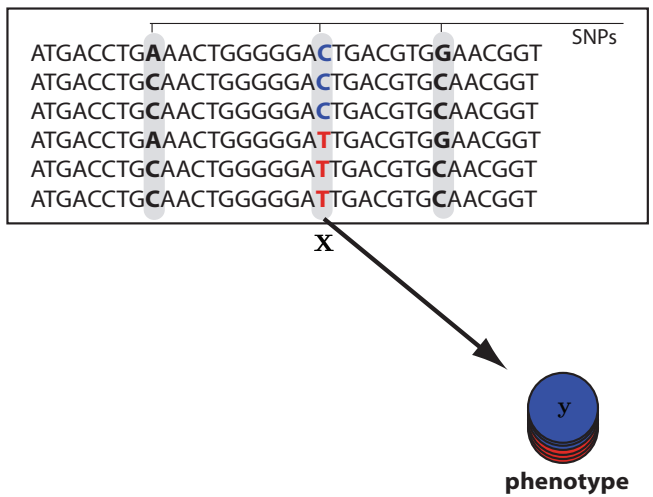
- Multivariate Gaussian

$$p(x \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$$
$$= \frac{1}{\sqrt{|2\pi\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$



Genome wide association studies (GWAS)

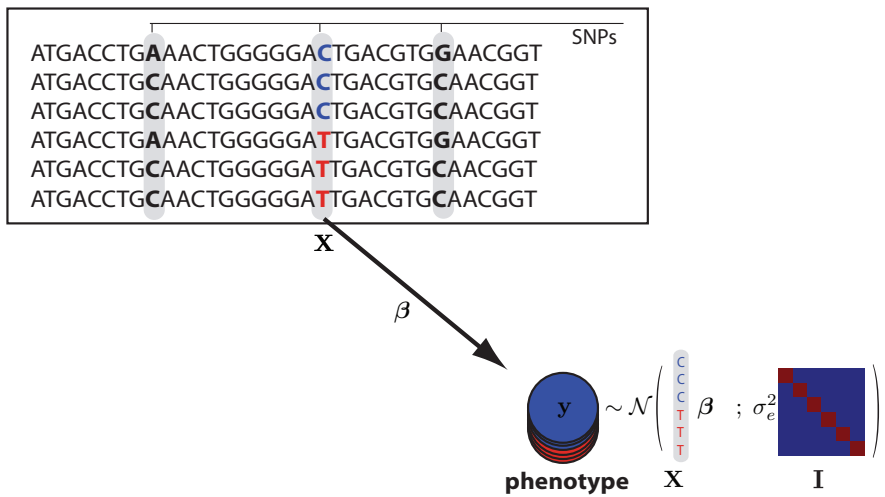
- Identify associations between variable genetic loci and phenotypes.
  - Linear and logistic regression
  - Statistical dependence tests (F-test, likelihood ratio)



Genome wide association studies (GWAS)

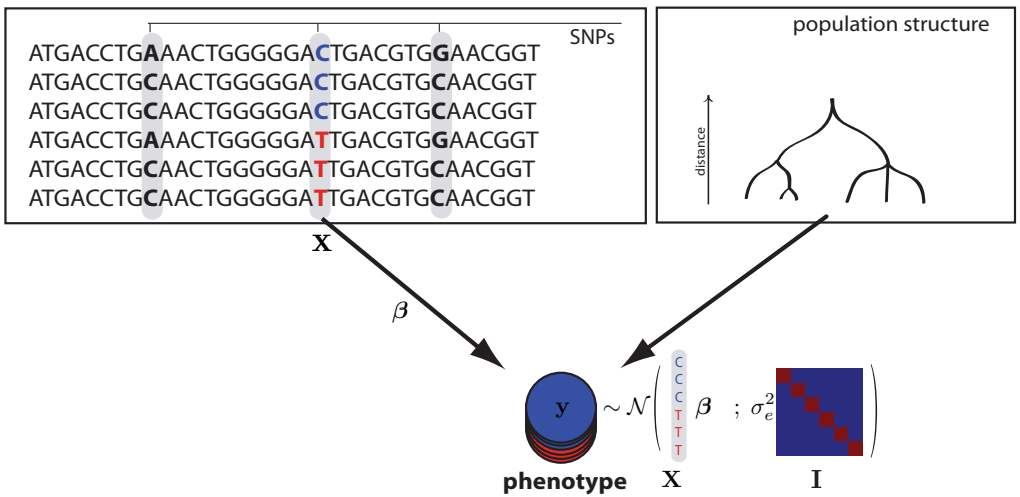
- Identify associations between variable genetic loci and phenotypes.
  - Linear and logistic regression
  - Statistical dependence tests (F-test, likelihood ratio)

$$\frac{\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{e}}^{2}\mathbf{I}\right)}{\mathcal{N}\left(\mathbf{y}|\mathbf{0};\sigma_{\mathrm{e}}^{2}\mathbf{I}\right)} \qquad (1)$$



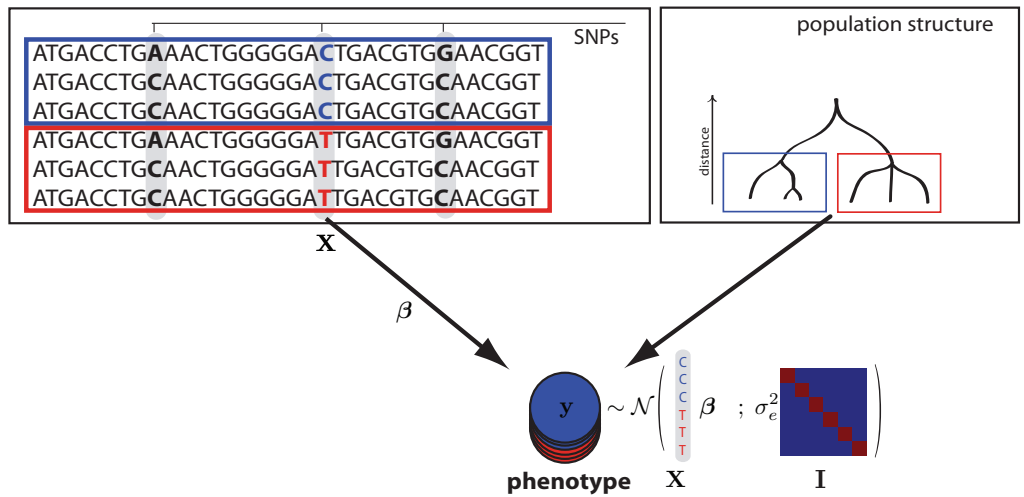
Population stratification

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



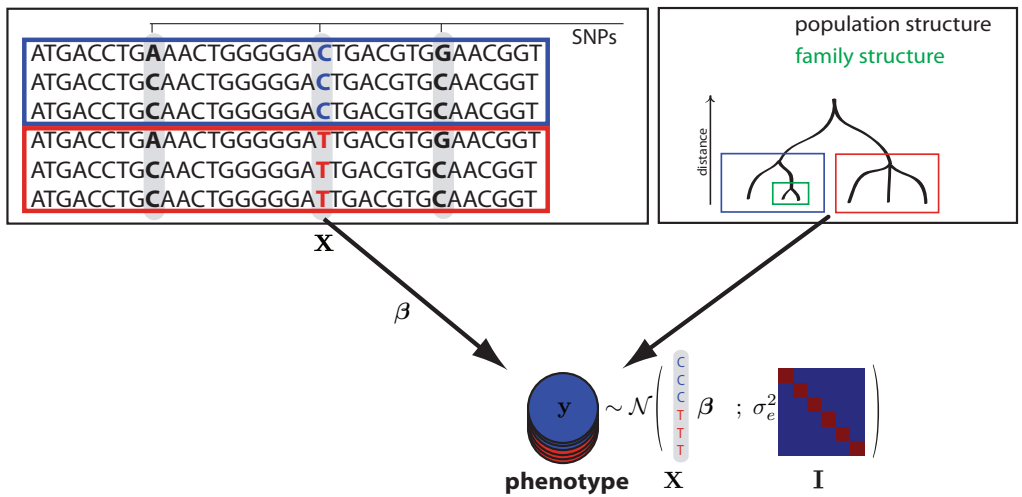
Population stratification

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



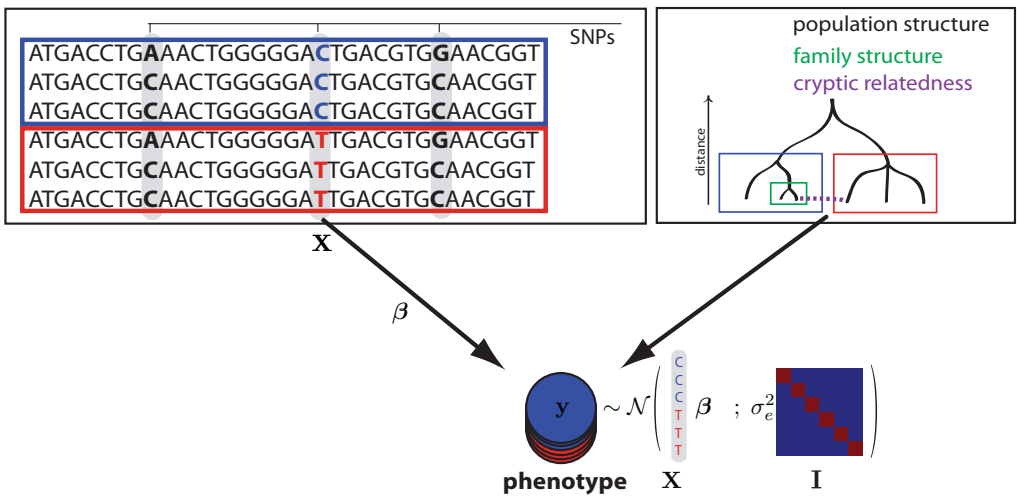
Population stratification

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



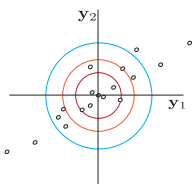
Population stratification

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

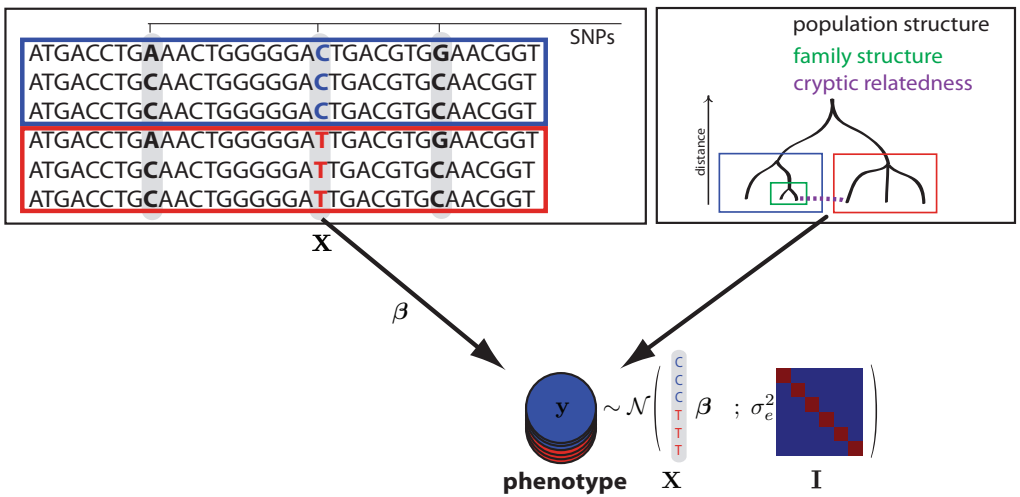


Population stratification

- 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

[Burton et al., 2007]

Population stratification

GWA on inflammatory bowel disease (WTCCC)

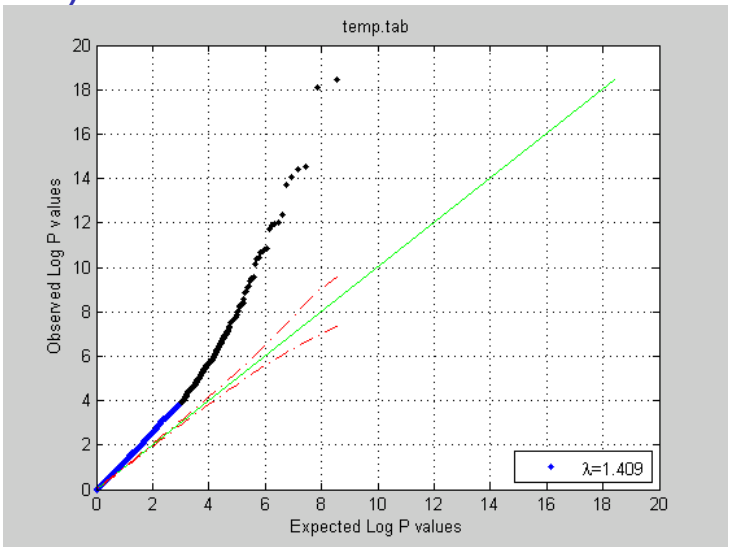
- 3.4k cases, 11.9k controls
- Methods
  - Linear regression
  - Likelihood ratio test

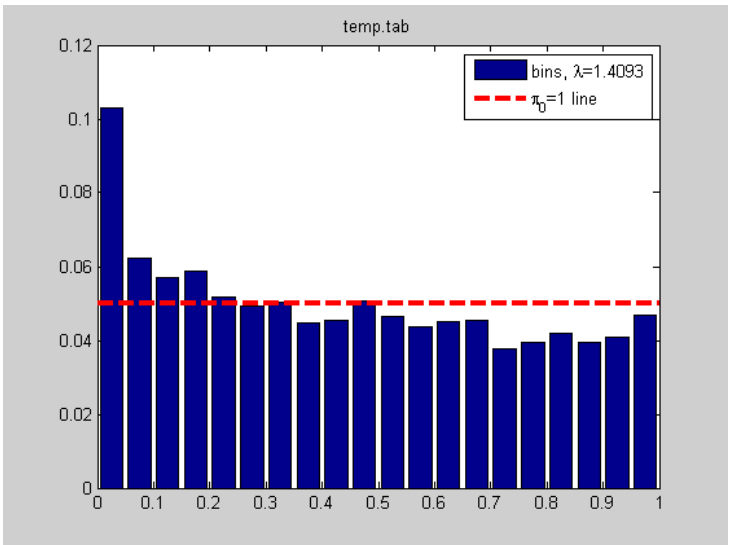
[Burton et al., 2007]

Population stratification

GWA on inflammatory bowel disease (WTCCC)

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





[Burton et al., 2007]

Outline

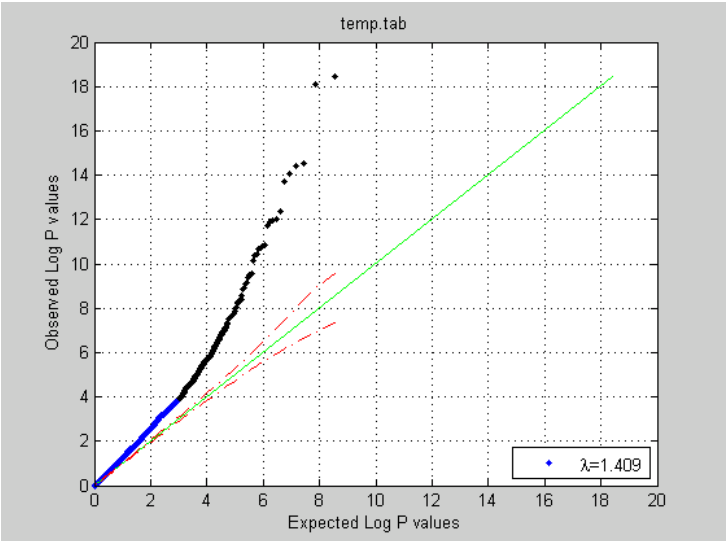
Genomic control [Devlin and Roeder, Biometrics 1999]

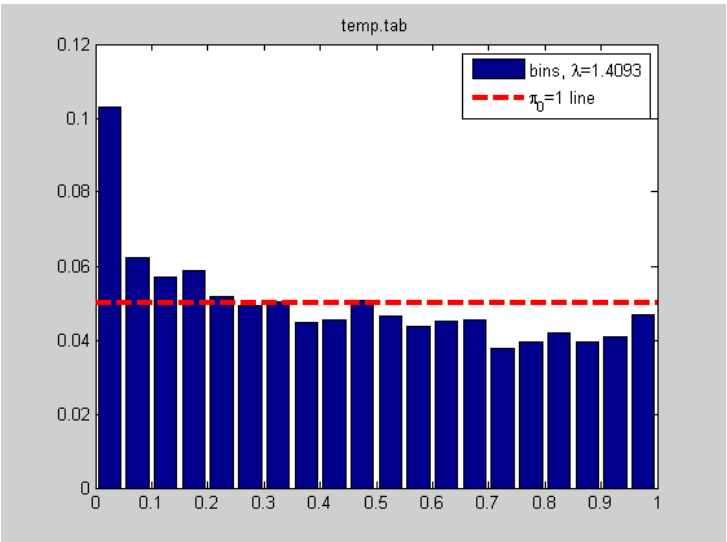
- Genomic control λ

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

  - $\lambda = 1$: 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.





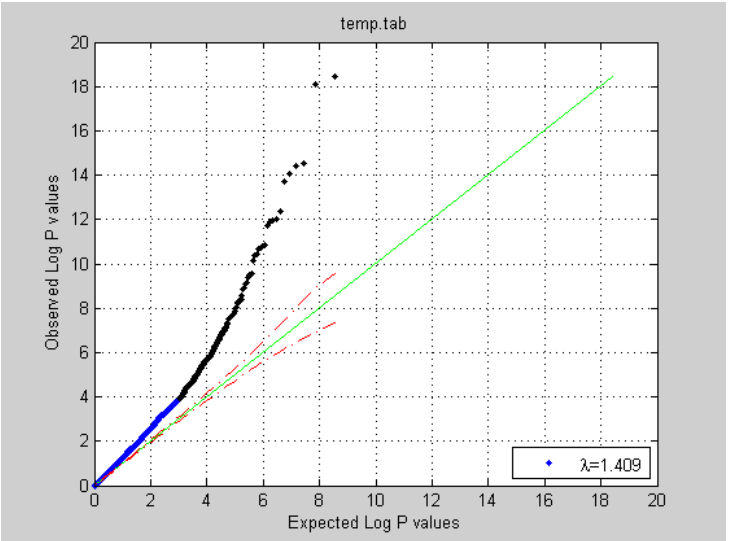
Genomic control [Devlin and Roeder, Biometrics 1999]

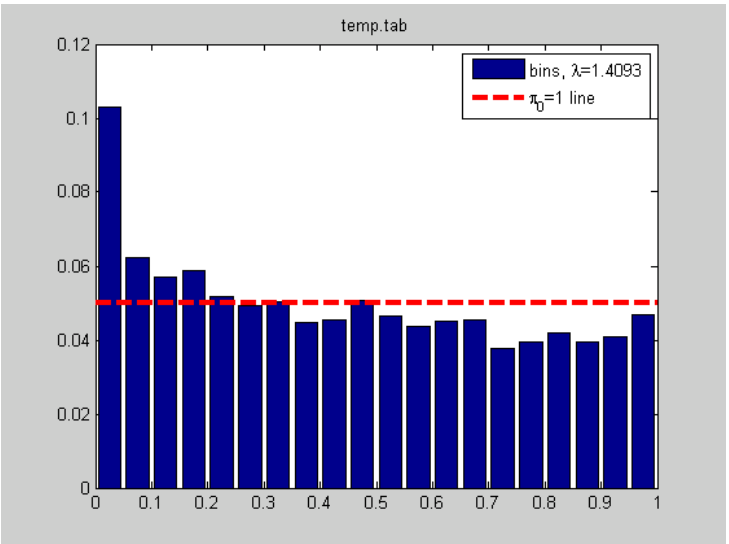
- Genomic control λ

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

  - $\lambda = 1$: 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.





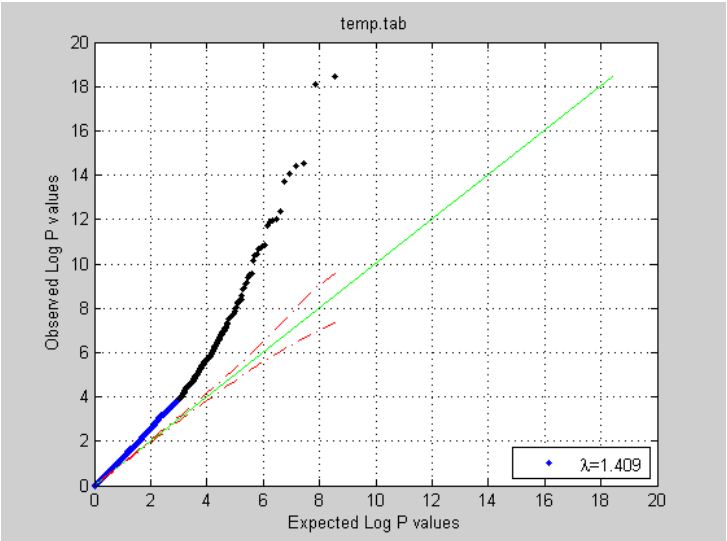
Genomic control [Devlin and Roeder, Biometrics 1999]

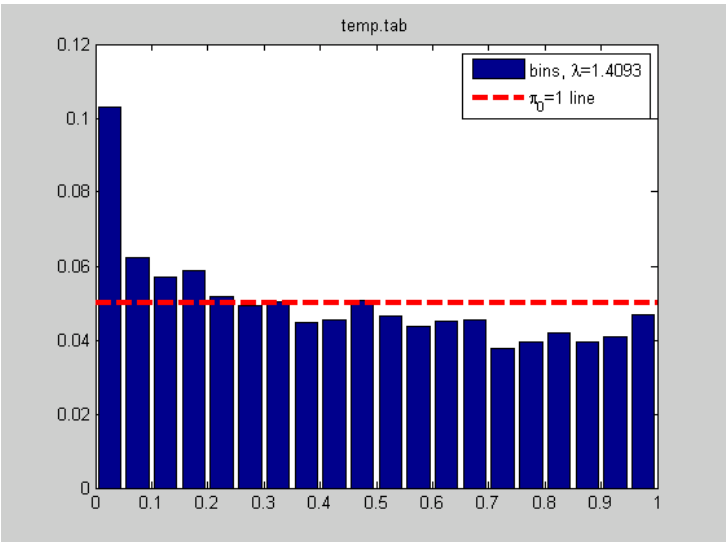
- Genomic control λ

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

  - $\lambda = 1$: 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.



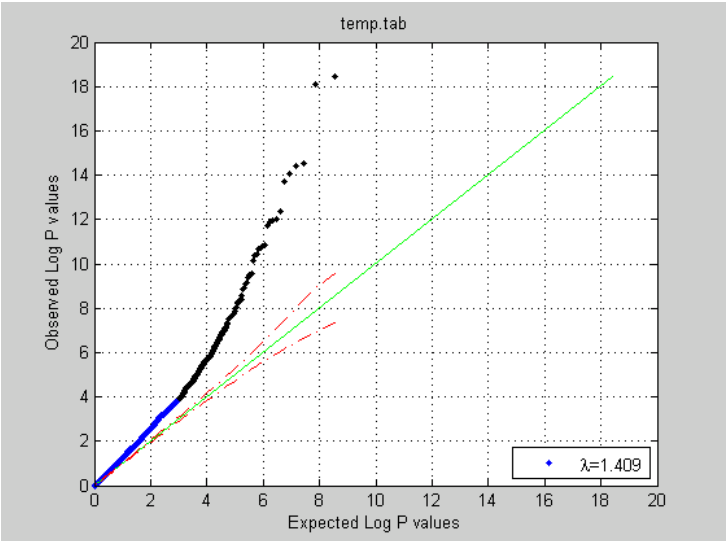


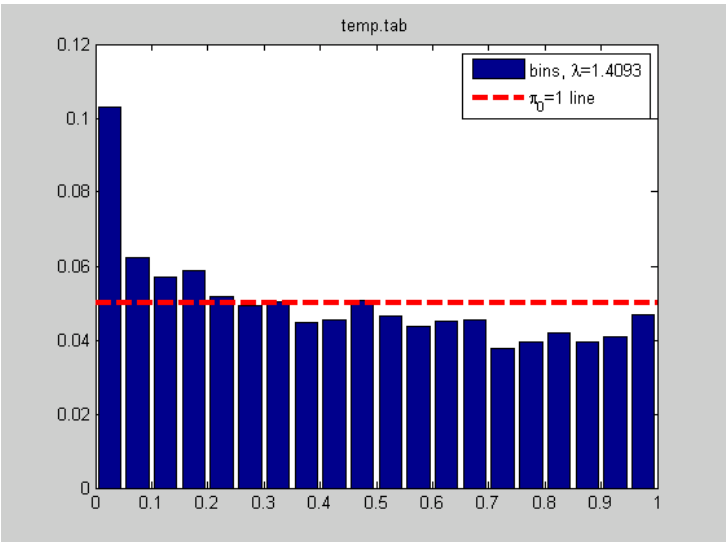
Genomic control [Devlin and Roeder, Biometrics 1999]

- 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.



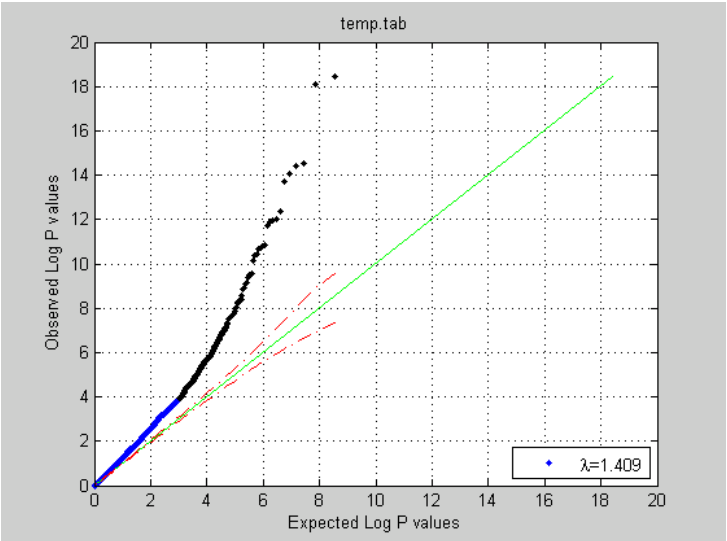


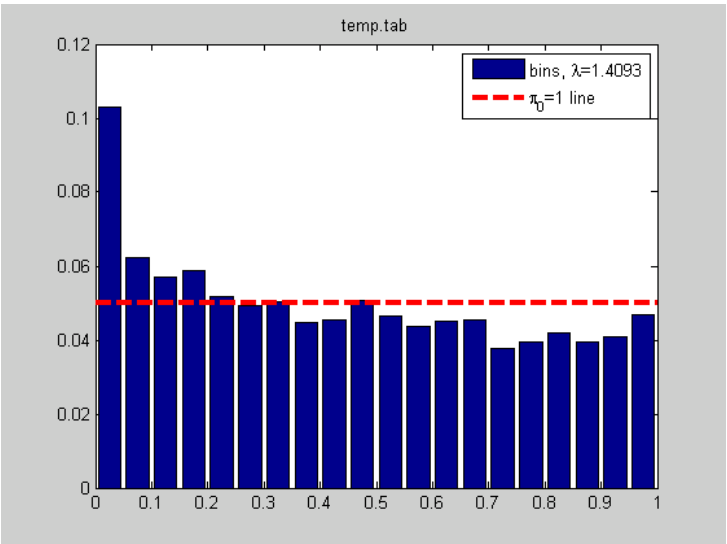
Genomic control [Devlin and Roeder, Biometrics 1999]

- 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.



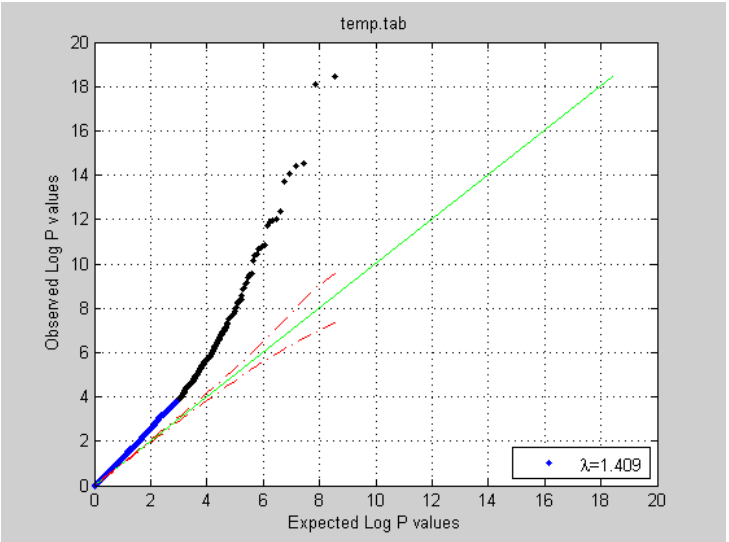


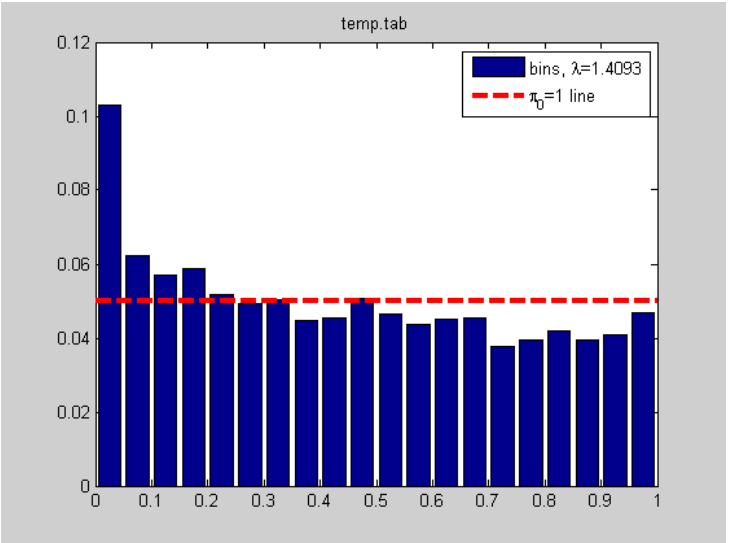
Genomic control [Devlin and Roeder, Biometrics 1999]

- Genomic control λ

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

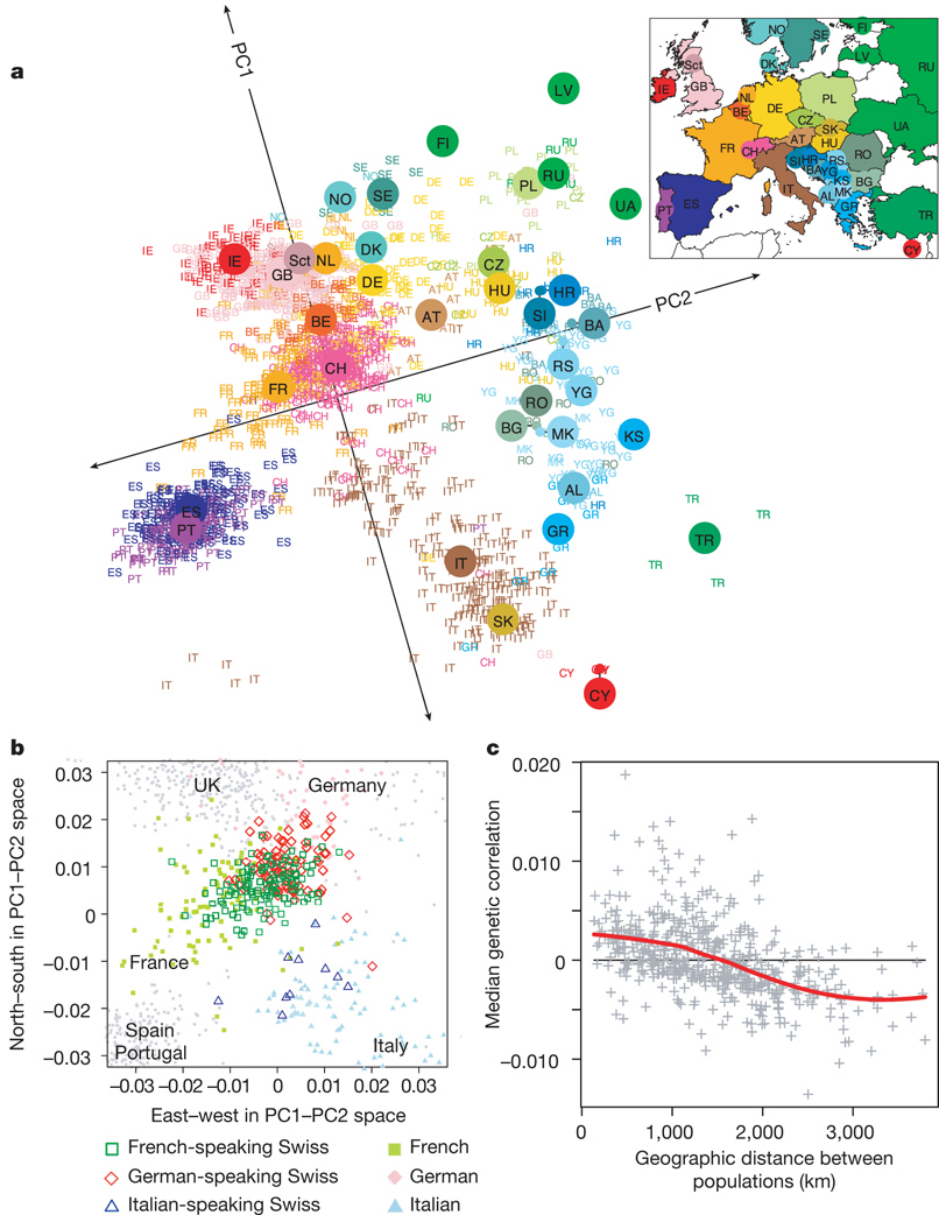
  - $\lambda = 1$: Calibrated P-values
  - $\lambda > 1$: Inflation
  - $\lambda < 1$: Deflation
- Correct by dividing test statistic by λ.
- Applicable in combination with every method.
- Does not change (non-)uniformity of P-values.
- Very conservative.





Eigenstrat

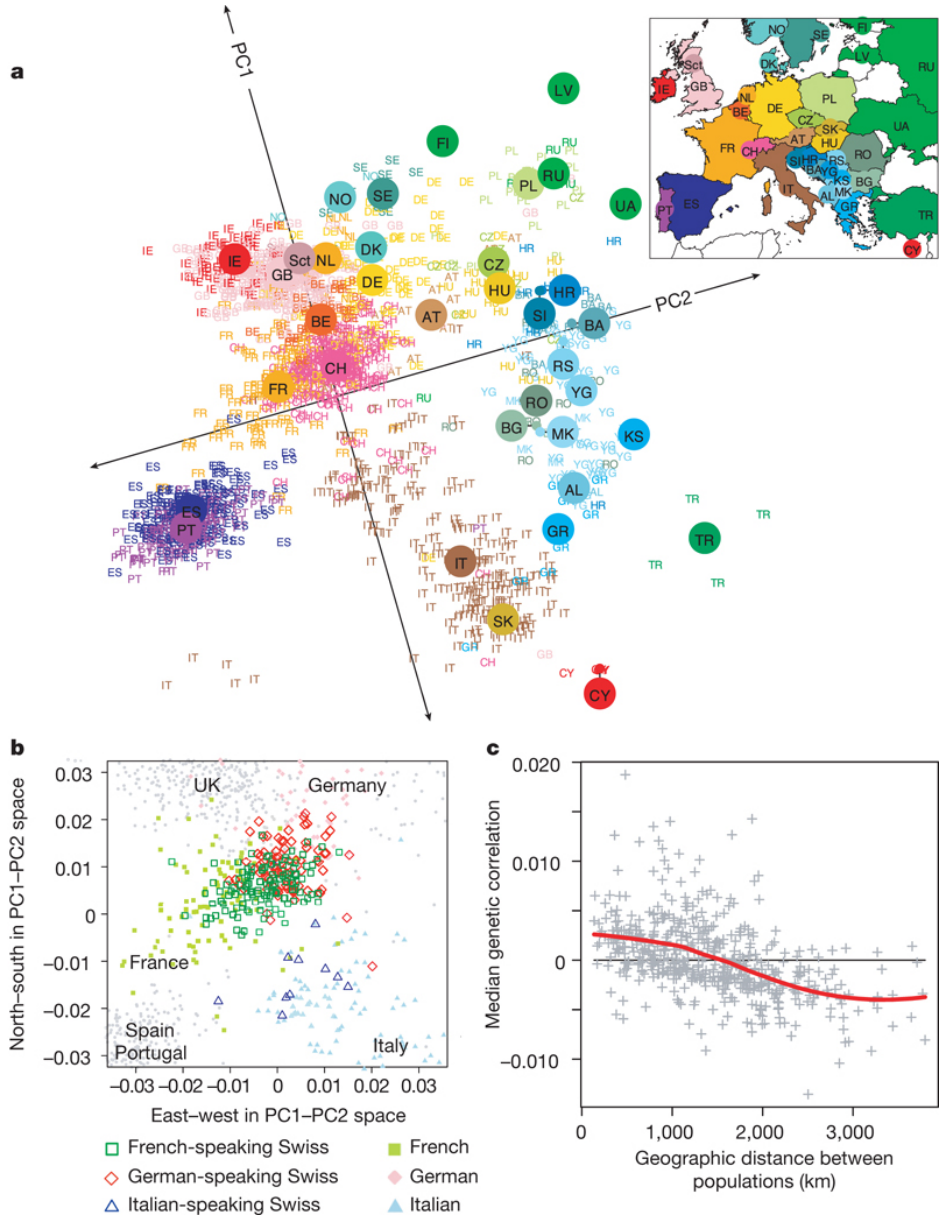
- Population structure causes genome-wide correlations between SNPs
- A large part of the total variation in the SNPs can be explained by population differences.
- Novembre et al. [2008] show that the eigenvectors of the SNP covariance matrix reflect population structure.
- Eigenstrat uses this property to correct for population structure in GWAS.



[Price et al., 2006, Patterson et al., 2006, Novembre et al., 2008]

Eigenstrat

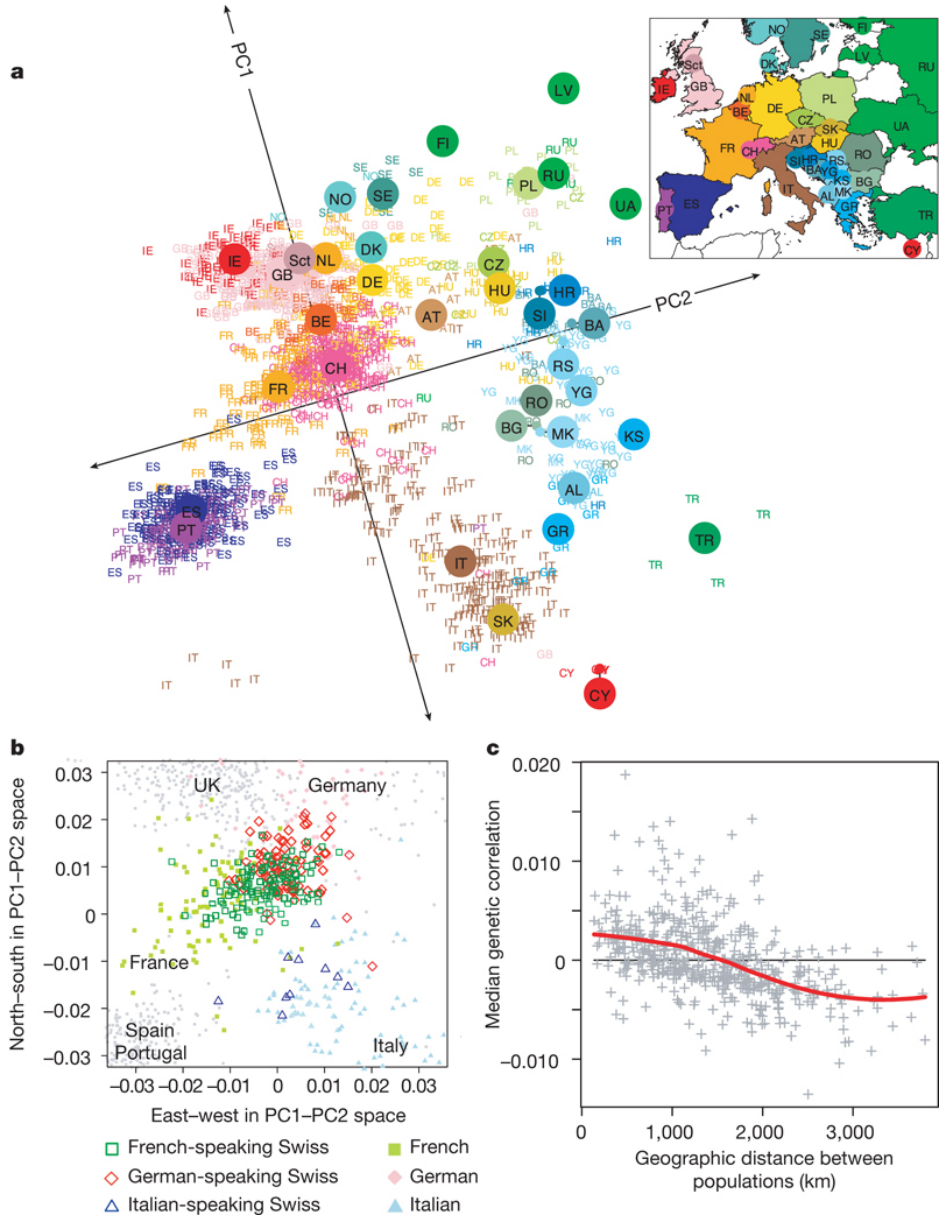
- Population structure causes genome-wide correlations between SNPs
- A large part of the total variation in the SNPs can be explained by population differences.
- Novembre et al. [2008] show that the eigenvectors of the SNP covariance matrix reflect population structure.
- Eigenstrat uses this property to correct for population structure in GWAS.



[Price et al., 2006, Patterson et al., 2006, Novembre et al., 2008]

Eigenstrat

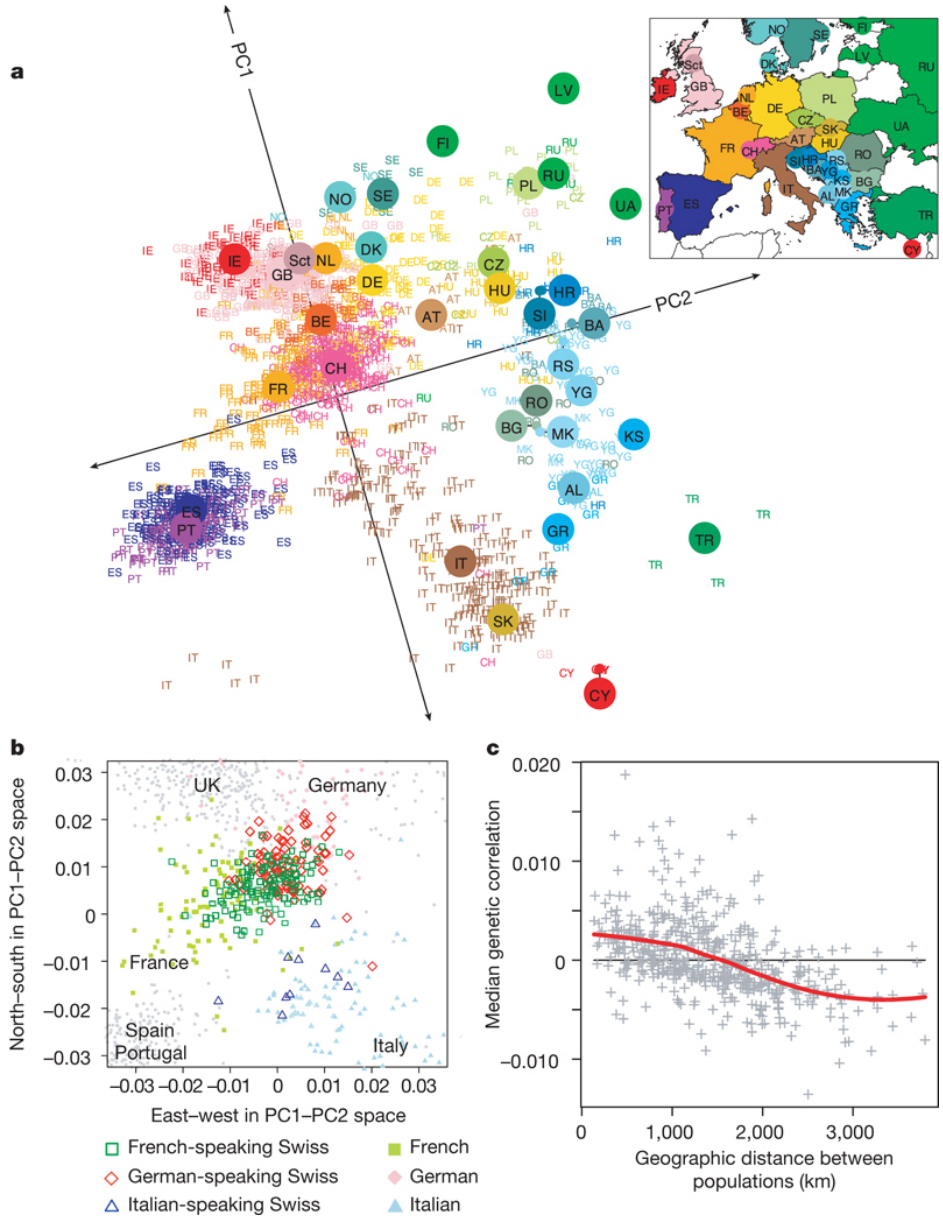
- Population structure causes genome-wide correlations between SNPs
- A large part of the total variation in the SNPs can be explained by population differences.
- Novembre et al. [2008] show that the eigenvectors of the SNP covariance matrix reflect population structure.
- Eigenstrat uses this property to correct for population structure in GWAS.



[Price et al., 2006, Patterson et al., 2006, Novembre et al., 2008]

Eigenstrat

- Population structure causes genome-wide correlations between SNPs
- A large part of the total variation in the SNPs can be explained by population differences.
- Novembre et al. [2008] show that the eigenvectors of the SNP covariance matrix reflect population structure.
- Eigenstrat uses this property to correct for population structure in GWAS.



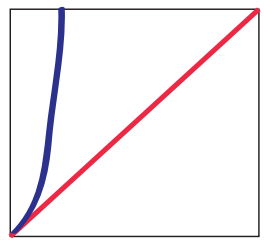
[Price et al., 2006, Patterson et al., 2006, Novembre et al., 2008]

Eigenstrat

Eigenstrat procedure:

- Compute covariance matrix based on SNPs
- Compute eigenvectors of covariance matrix
- Add largest eigenvector as covariate to regression.
- Repeat until P-values are uniform.

$$\mathcal{N}\left(\mathbf{y} \;\middle|\; \underbrace{\begin{matrix}C\\C\\C\\T\\T\\T\end{matrix}}_{\mathbf{X}}\beta\;;\;\sigma^2\underbrace{\begin{matrix}\blacksquare&&&&&\\&\blacksquare&&&&\\&&\blacksquare&&&\\&&&\blacksquare&&\\&&&&\blacksquare&\\&&&&&\blacksquare\end{matrix}}_{\mathbf{I}}\right)$$



[Price et al., 2006]

Eigenstrat

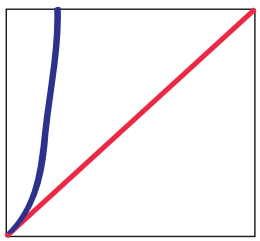
$$\frac{1}{S}\underbrace{\mathbf{X} - \mathbb{E}[\mathbf{X}]}_{S} \times \underbrace{\mathbf{X}^{\mathrm{T}} - \mathbb{E}[\mathbf{X}]^{\mathrm{T}}}_{N}$$

Genome-wide SNP covariance

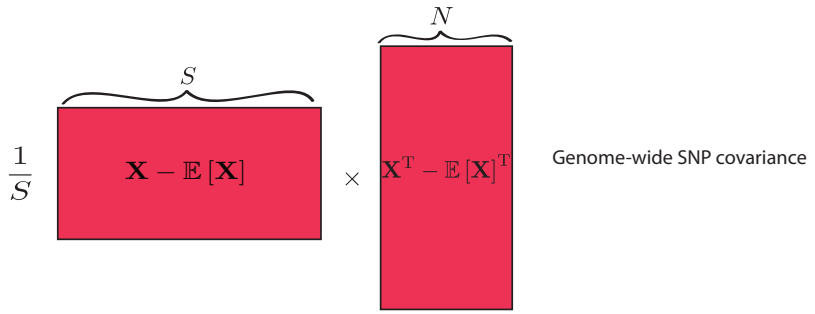
Eigenstrat procedure:

- Compute covariance matrix based on SNPs
- Compute eigenvectors of covariance matrix
- Add largest eigenvector as covariate to regression.
- Repeat until P-values are uniform.

$$\mathcal{N}\left(\mathbf{y} \,\middle|\, \underbrace{\begin{matrix}C\\C\\C\\T\\T\\T\end{matrix}}_{\mathbf{X}}\beta\,;\,\sigma^2\underbrace{\begin{matrix}\blacksquare&&\\&\blacksquare&\\&&\blacksquare\end{matrix}}_{\mathbf{I}}\right)$$

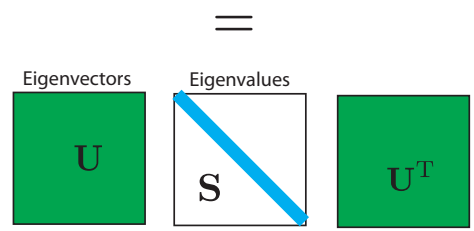


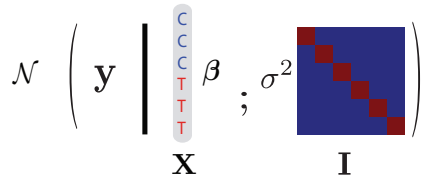
Eigenstrat

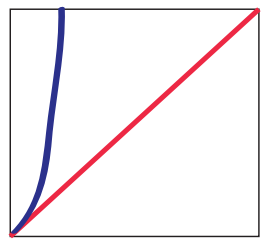


Eigenstrat procedure:

- Compute covariance matrix based on SNPs
- Compute eigenvectors of covariance matrix
- Add largest eigenvector as covariate to regression.
- Repeat until P-values are uniform.





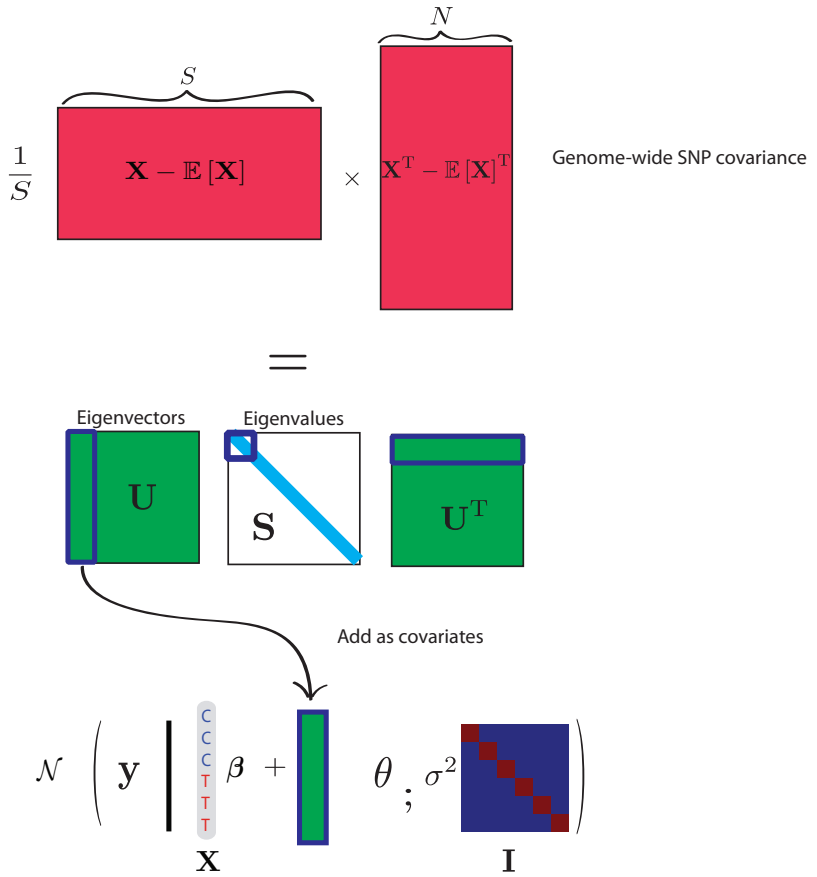


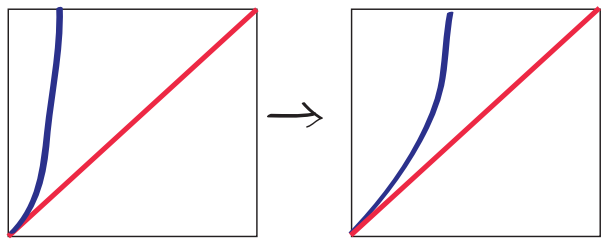
[Price et al., 2006]

Eigenstrat

Eigenstrat procedure:

- Compute covariance matrix based on SNPs
- Compute eigenvectors of covariance matrix
- Add largest eigenvector as covariate to regression.
- Repeat until P-values are uniform.



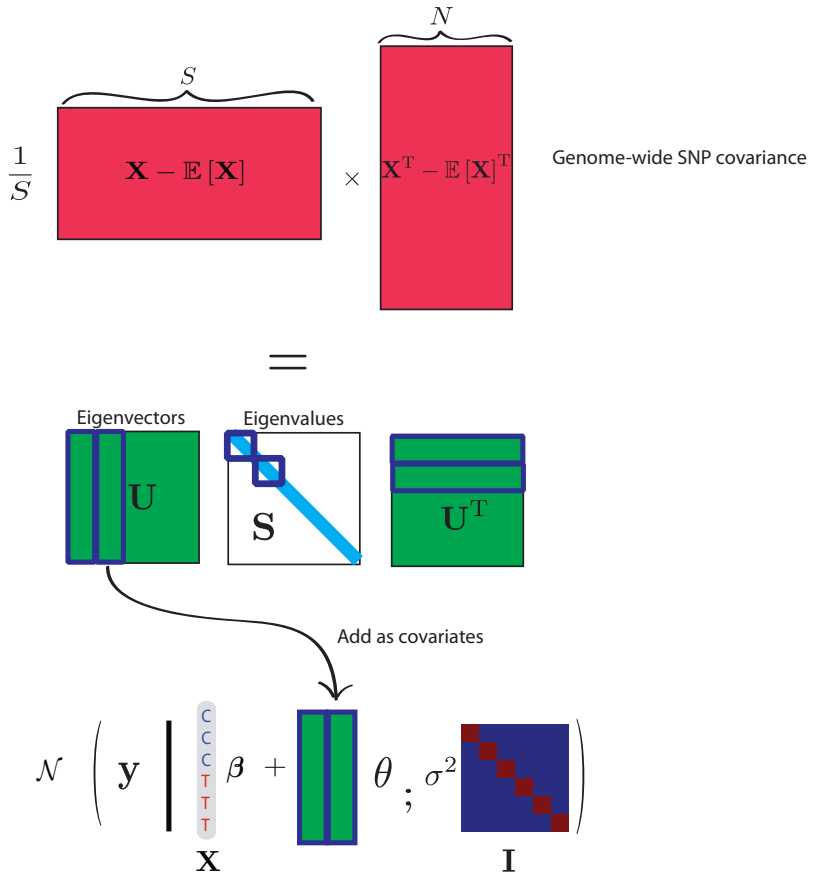


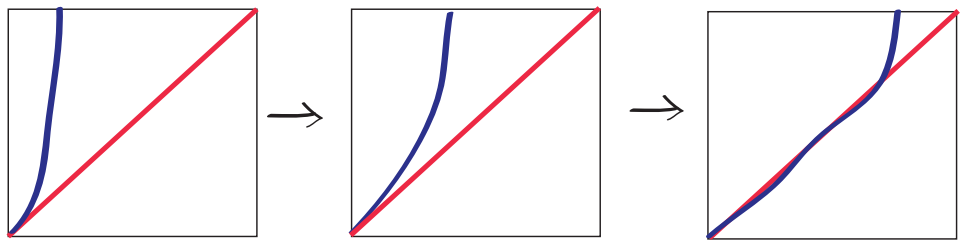
[Price et al., 2006]

Eigenstrat

Eigenstrat procedure:

- Compute covariance matrix based on SNPs
- Compute eigenvectors of covariance matrix
- Add largest eigenvector as covariate to regression.
- Repeat until P-values are uniform.

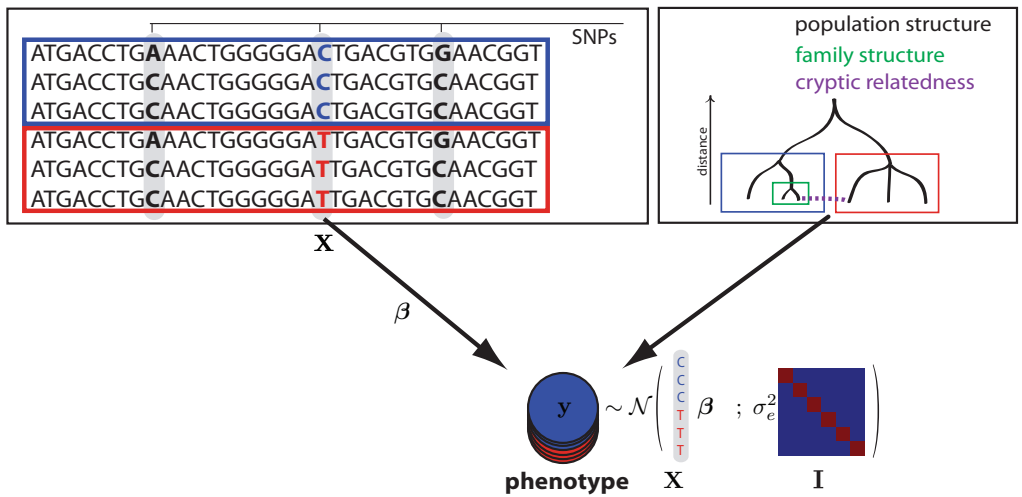




[Price et al., 2006]

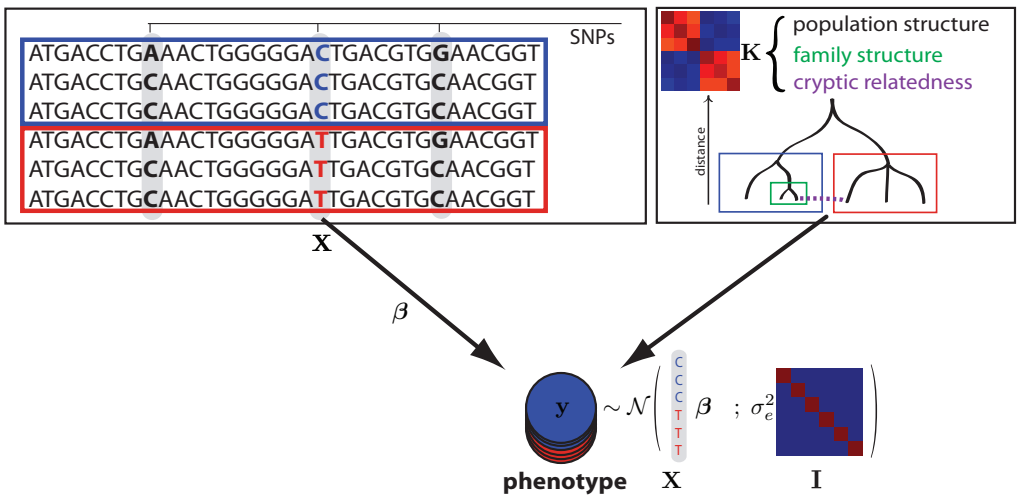
Linear mixed models (LMM)

- Covariance matrix $\mathbf{K}$
  - Estimated from SNP data
  - Kinship coefficients
    - Identity by state
    - Identity by descent
  - Realized relationship matrix (linear)
- Sample random effect $\mathbf{u}$.
- Sample phenotype $\mathbf{y}$.



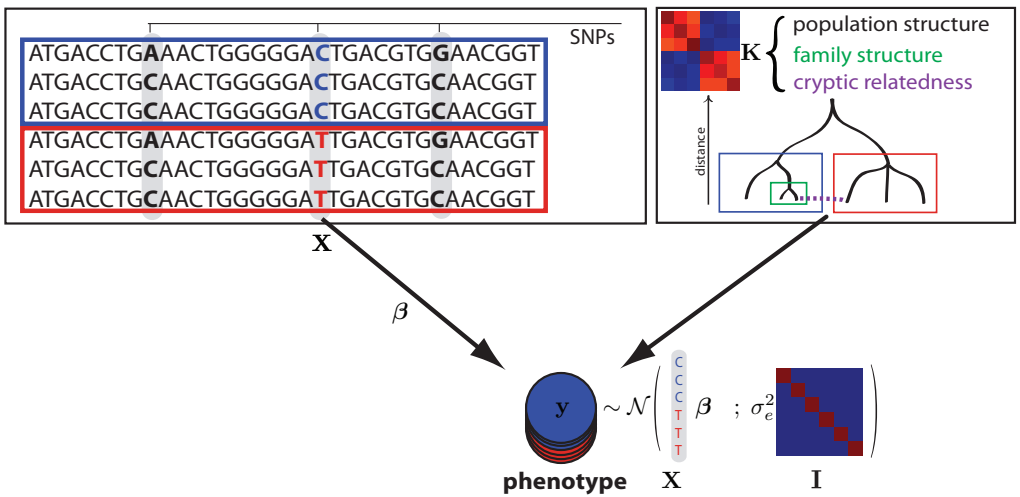
Linear mixed models (LMM)

- Covariance matrix $\mathbf{K}$
  - Estimated from SNP data
  - Kinship coefficients
    - Identity by state
    - Identity by descent
  - Realized relationship matrix (linear)
- Sample random effect $\mathbf{u}$.
- Sample phenotype $\mathbf{y}$.



Linear mixed models (LMM)

- Covariance matrix $\mathbf{K}$
  - Estimated from SNP data
  - Kinship coefficients
    - Identity by state
    - Identity by descent
  - Realized relationship matrix (linear)
- Sample random effect $\mathbf{u}$.
- Sample phenotype $\mathbf{y}$.

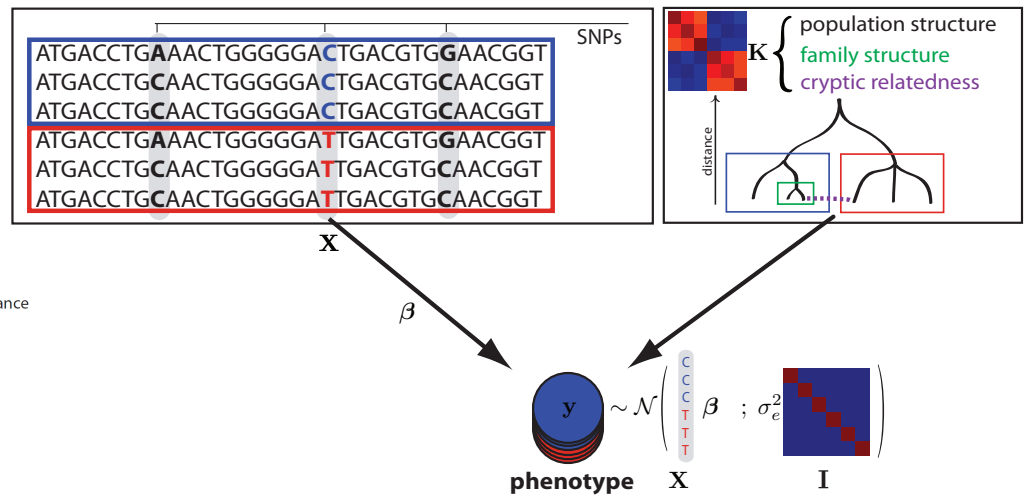


Linear mixed models (LMM)

- Covariance matrix $\mathbf{K}$
  - Estimated from SNP data
  - Kinship coefficients
    - Identity by state
    - Identity by descent
  - Realized relationship matrix (linear)

$$\frac{1}{S} \overbrace{\mathbf{X} - \mathbb{E}[\mathbf{X}]}^{S} \times \overbrace{\mathbf{X}^{\mathrm{T}} - \mathbb{E}[\mathbf{X}]^{\mathrm{T}}}^{N}$$

Genome-wide SNP covariance



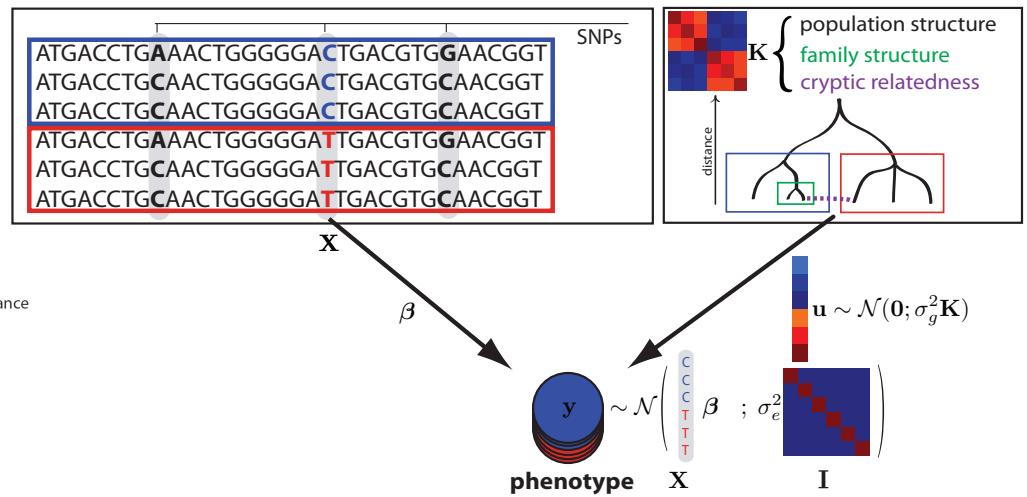
- Sample random effect $\mathbf{u}$.
- Sample phenotype $\mathbf{y}$.

Linear mixed models (LMM)

- Covariance matrix $\mathbf{K}$
  - Estimated from SNP data
  - Kinship coefficients
    - Identity by state
    - Identity by descent
  - Realized relationship matrix (linear)

$$\frac{1}{S} \underbrace{\mathbf{X} - \mathbb{E}[\mathbf{X}]}_{S} \times \underbrace{\mathbf{X}^T - \mathbb{E}[\mathbf{X}]^T}_{N}$$

Genome-wide SNP covariance



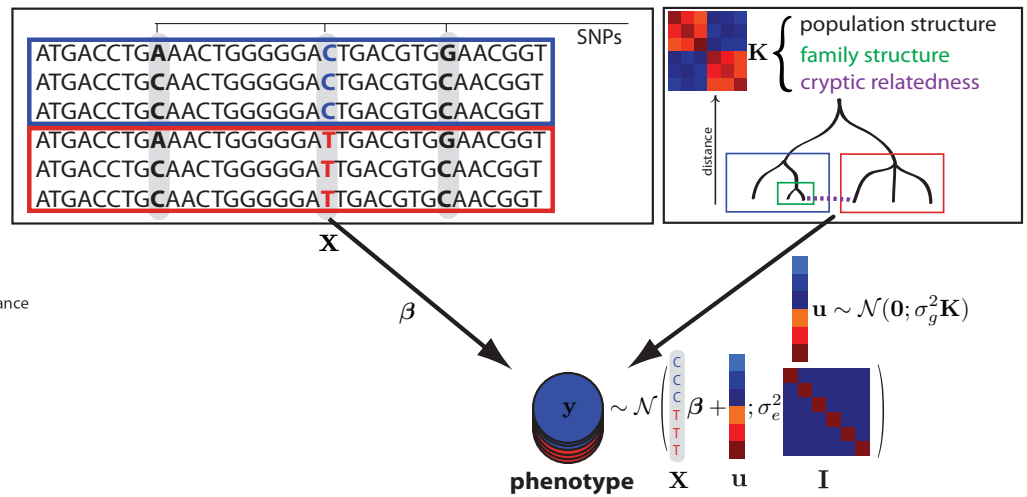
- Sample random effect $\mathbf{u}$.
- Sample phenotype $\mathbf{y}$.

Linear mixed models (LMM)

- Covariance matrix $\mathbf{K}$
  - Estimated from SNP data
  - Kinship coefficients
    - Identity by state
    - Identity by descent
  - Realized relationship matrix (linear)

$$\frac{1}{S}\underbrace{\mathbf{X} - \mathbb{E}[\mathbf{X}]}_{S} \times \underbrace{\mathbf{X}^{\mathrm{T}} - \mathbb{E}[\mathbf{X}]^{\mathrm{T}}}_{N}$$

Genome-wide SNP covariance



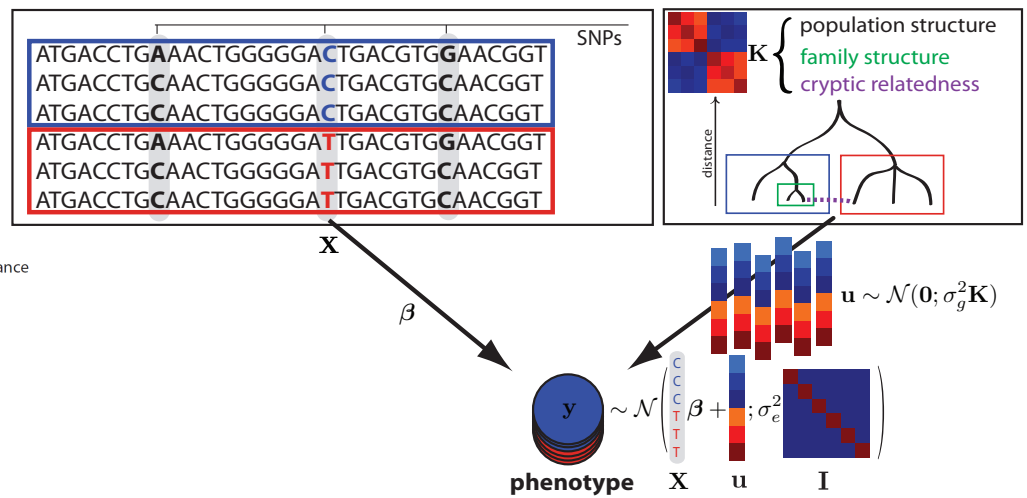
- Sample random effect $\mathbf{u}$.
- Sample phenotype $\mathbf{y}$.

Linear mixed models (LMM)

- Covariance matrix $\mathbf{K}$
  - Estimated from SNP data
  - Kinship coefficients
    - Identity by state
    - Identity by descent
  - Realized relationship matrix (linear)

$$\frac{1}{S}\underbrace{\mathbf{X} - \mathbb{E}[\mathbf{X}]}_{S} \times \underbrace{\mathbf{X}^{\mathrm{T}} - \mathbb{E}[\mathbf{X}]^{\mathrm{T}}}_{N}$$

Genome-wide SNP covariance

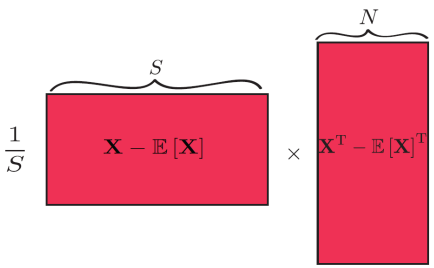


- Sample random effect $\mathbf{u}$.
- Sample phenotype $\mathbf{y}$.

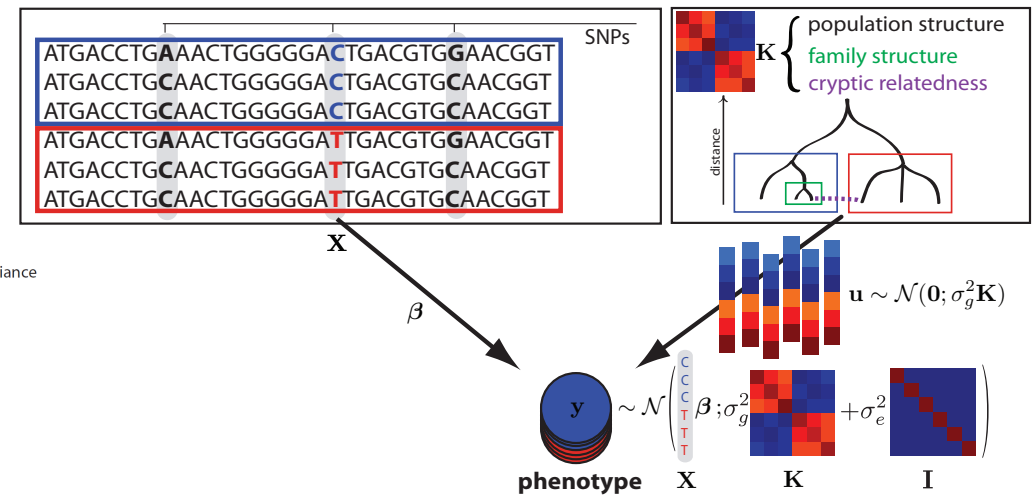
$$\int_{\mathbf{u}} \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta} + \mathbf{u}; \sigma_{\mathrm{e}}^2\mathbf{I}\right)\mathcal{N}\left(\mathbf{u}|\mathbf{0}; \sigma_{\mathrm{g}}^2\mathbf{K}\right)$$

Linear mixed models (LMM)

- Covariance matrix $\mathbf{K}$
  - Estimated from SNP data
  - Kinship coefficients
    - Identity by state
    - Identity by descent
  - Realized relationship matrix (linear)



Genome-wide SNP covariance



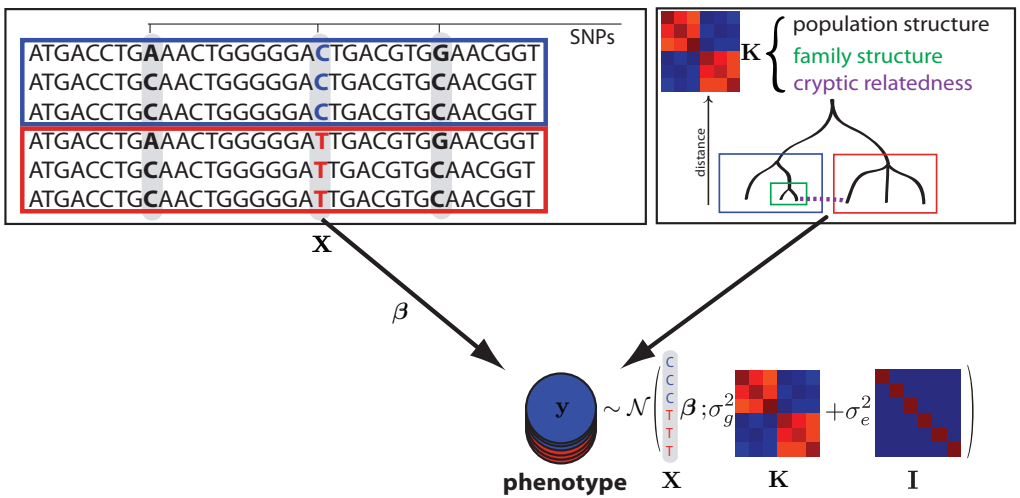
- Sample random effect $\mathbf{u}$.
- Sample phenotype $\mathbf{y}$.

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_{\mathrm{g}}^2\mathbf{K} + \sigma_{\mathrm{e}}^2\mathbf{I}\right)$$

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

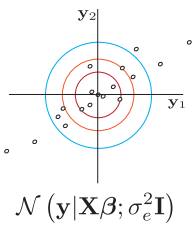
$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^2\mathbf{K}+\sigma_{\mathrm{e}}^2\mathbf{I}\right)$$

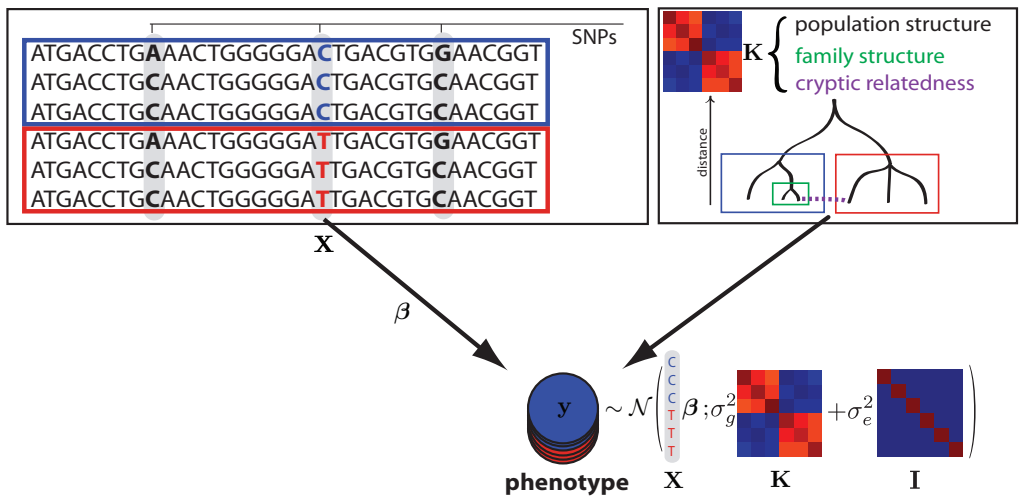


Linear mixed models (LMM)

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

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^2\mathbf{K}+\sigma_{\mathrm{e}}^2\mathbf{I}\right)$$

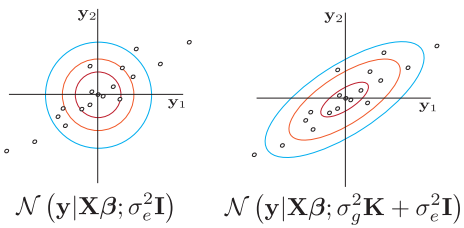




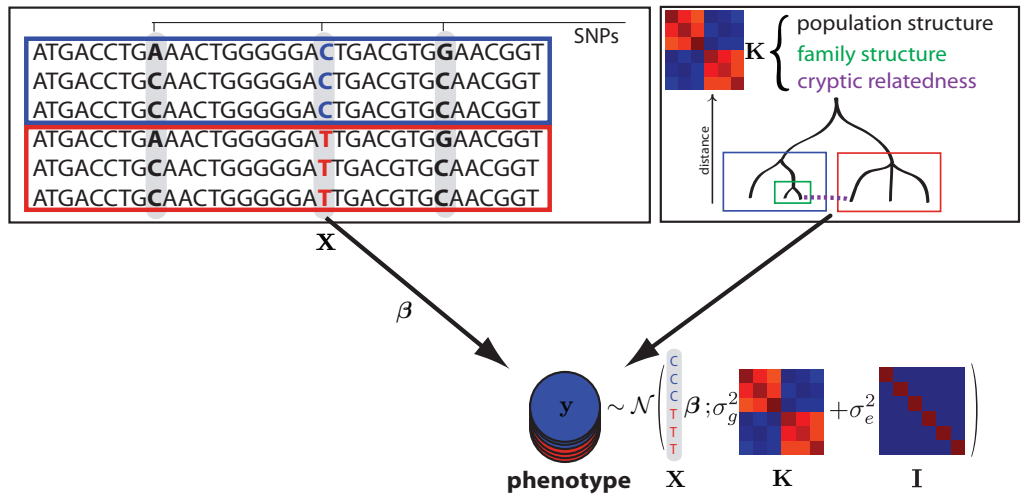
Linear mixed models (LMM)

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

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^2\mathbf{K}+\sigma_{\mathrm{e}}^2\mathbf{I}\right)$$



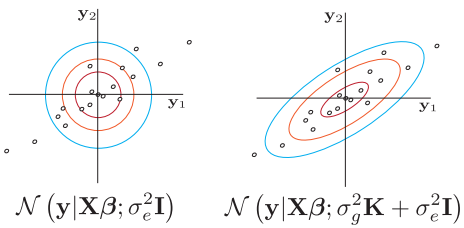
$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_e^2\mathbf{I}\right)$ $\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_g^2\mathbf{K}+\sigma_e^2\mathbf{I}\right)$

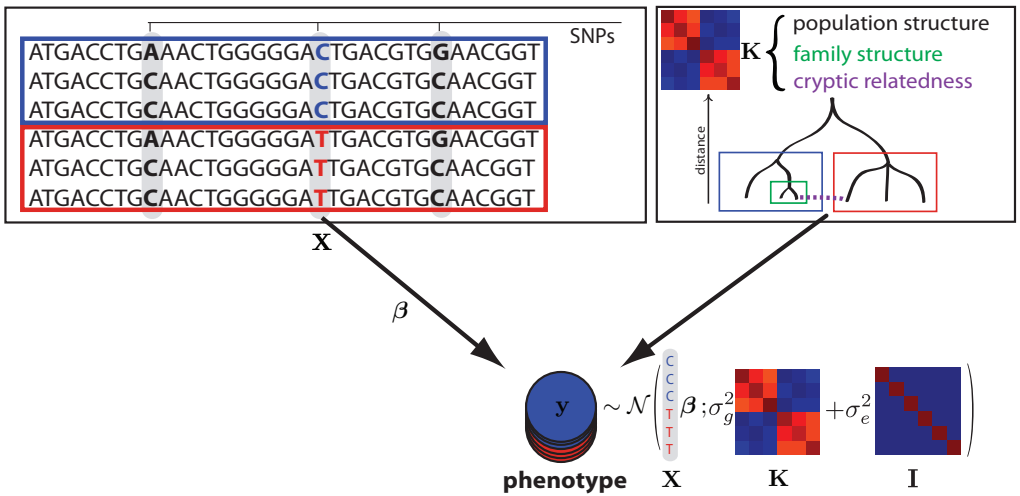


Linear mixed models (LMM)

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

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^2\mathbf{K}+\sigma_{\mathrm{e}}^2\mathbf{I}\right)$$

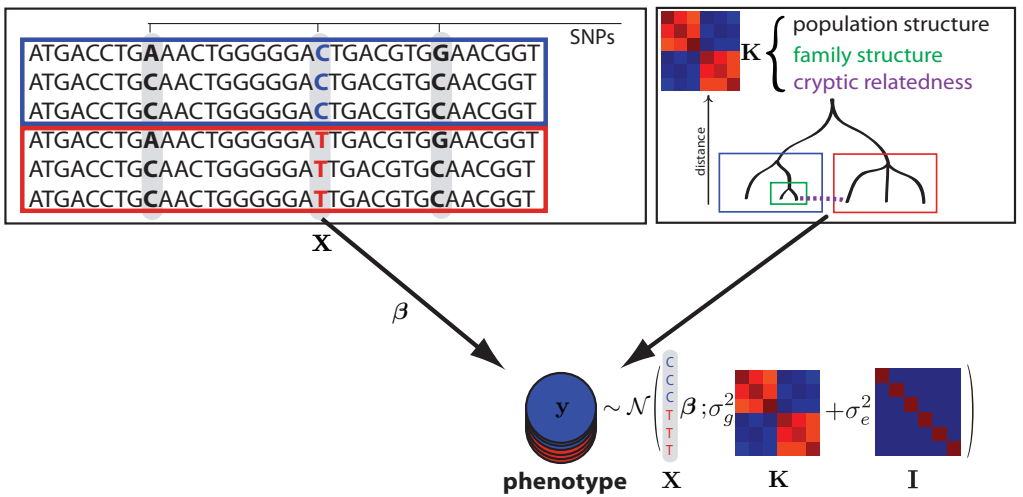


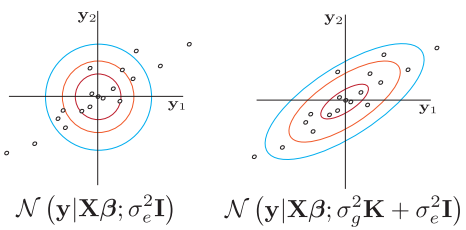


Linear mixed models (LMM)

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

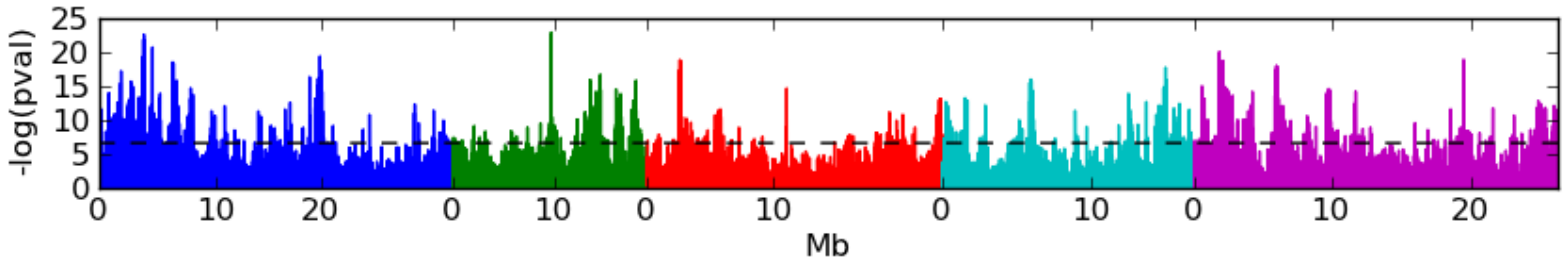
$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_g^2\mathbf{K}+\sigma_e^2\mathbf{I}\right)$$



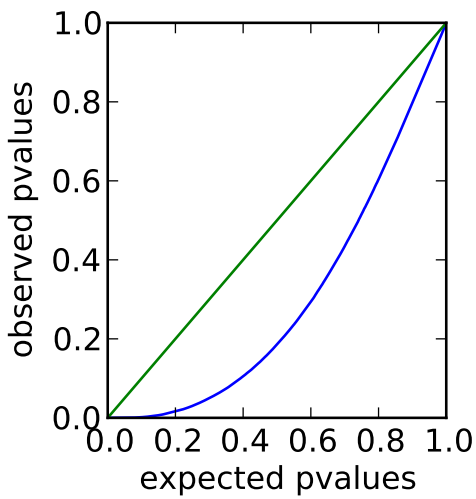


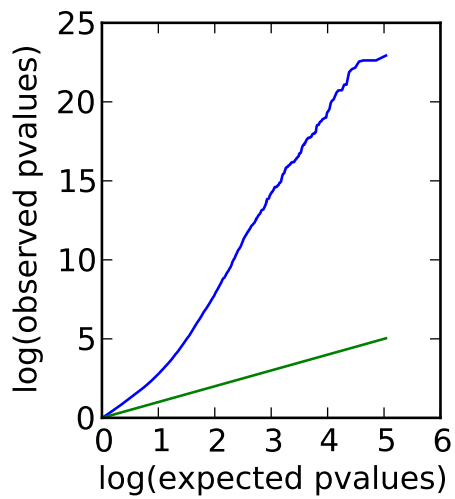
GWAS for Flowering Time in *Arabidopsis thaliana*

- Linear Model:



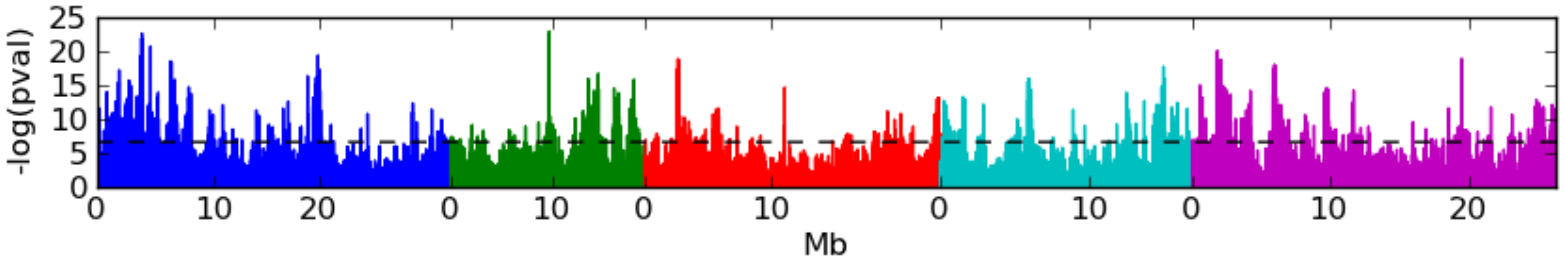
- QQ-plot:



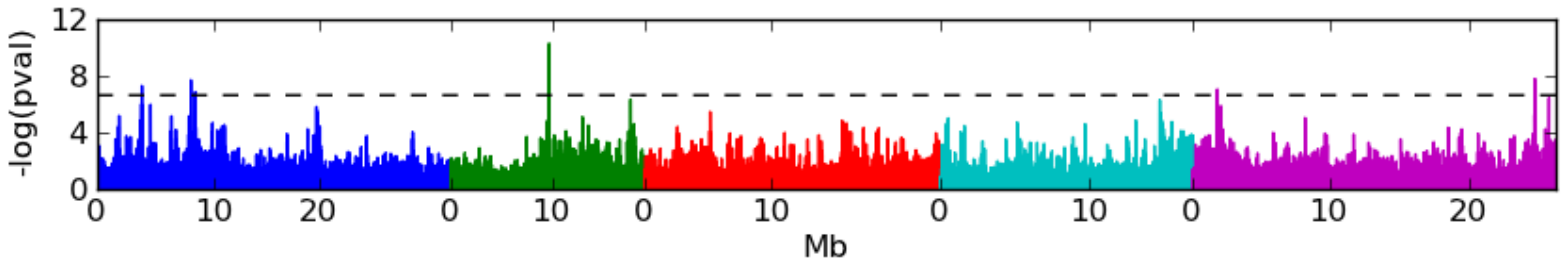


GWAS for Flowering Time in *Arabidopsis thaliana*

- Linear Model:

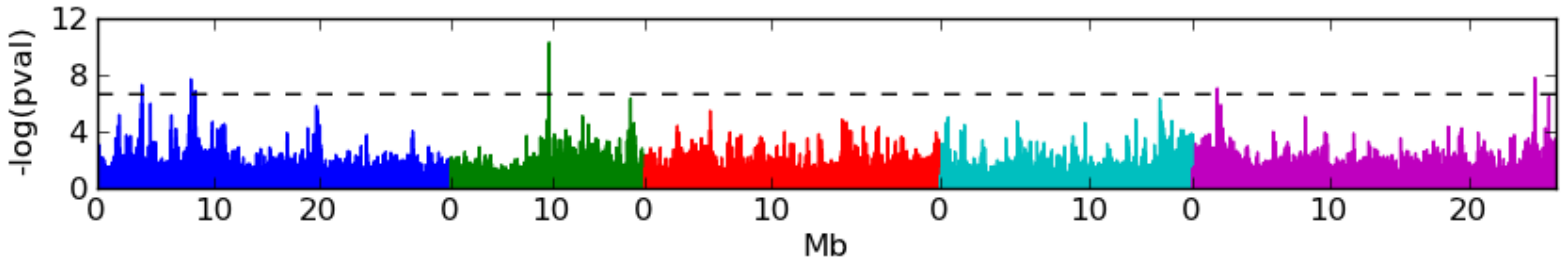


- Linear Mixed Model:

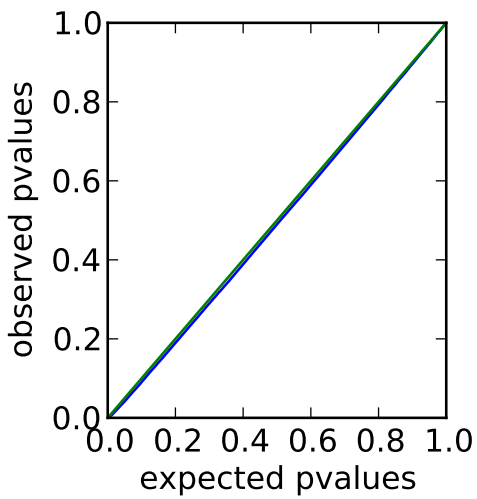


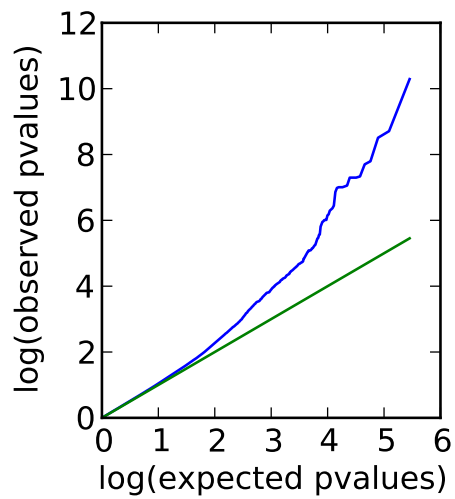
GWAS for Flowering Time in *Arabidopsis thaliana*

- Linear Mixed Model:



- QQ-plot:





Linear mixed models (LMM)

- LMM log likelihood

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

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

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

- ML-parameters $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma_{\mathrm{g}}^2}$ follow in closed form.
- Use optimizer to solve 1-dimensional optimization problem over δ.

[Kang et al., 2008]

Linear mixed models (LMM)

- LMM log likelihood

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

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

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

- ML-parameters $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma_{\mathrm{g}}^2}$ follow in closed form.
- Use optimizer to solve 1-dimensional optimization problem over δ.

[Kang et al., 2008]

Linear mixed models (LMM)

- LMM log likelihood

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

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

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

- ML-parameters $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma_{\mathrm{g}}^2}$ follow in closed form.
- Use optimizer to solve 1-dimensional optimization problem over δ.

[Kang et al., 2008]

Linear mixed models (LMM)

- LMM log likelihood

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

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

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

- ML-parameters $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma_{\mathrm{g}}^2}$ follow in closed form.
- Use optimizer to solve 1-dimensional optimization problem over δ.

[Kang et al., 2008]

Linear mixed models (LMM)

ML parameters

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

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

Linear mixed models (LMM)

ML parameters

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

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

Linear mixed models (LMM)

ML parameters

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

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

Linear mixed models (LMM)

ML parameters

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

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

set gradient to zero:

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

Linear mixed models (LMM)

ML parameters

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

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

set gradient to zero:

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

Linear mixed models (LMM)

ML parameters

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

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

set gradient to zero:

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

Linear mixed models (LMM)

ML parameters

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

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

set gradient to zero:

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

Linear mixed models (LMM)

ML parameters

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

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

set gradient to zero:

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

Linear mixed models (LMM)

ML parameters

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

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

set gradient to zero:

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

Note that this solution is analogous to the ML solution of the linear regression .

Linear mixed models (LMM)

ML parameters

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

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

set gradient to zero:

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

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

Linear mixed models (LMM)

ML parameters

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

$$\frac{\partial}{\partial \sigma_{\mathrm{g}}^2} \log \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_{\mathrm{g}}^2\left(\mathbf{K}+\delta\mathbf{I}\right)\right)$$

$$= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]$$

- **Bottleneck:** Inverting $N \times N$ matrix $\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}$
- If done naively, this is an $O(N^3)$ operation per SNP.

Linear mixed models (LMM)

ML parameters

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

$$\frac{\partial}{\partial \sigma_{\mathrm{g}}^2} \log \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_{\mathrm{g}}^2\left(\mathbf{K}+\delta\mathbf{I}\right)\right)$$

$$= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]$$

- **Bottleneck:** For every SNP we have to recalculate $(\mathbf{K}+\delta\mathbf{I})^{-1}$
- If done naively, this is an $O(N^3)$ operation per SNP.

Linear mixed models (LMM)

ML parameters

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

$$\frac{\partial}{\partial \sigma_{\mathrm{g}}^2} \log \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_{\mathrm{g}}^2\left(\mathbf{K}+\delta\mathbf{I}\right)\right)$$

$$= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]$$

set derivative to zero:

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

- **Bottleneck:** $(\mathbf{K}+\delta\mathbf{I})^{-1}$
- If done naively, this is an $O(N^3)$ operation per SNP.

Linear mixed models (LMM)

ML parameters

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

$$\frac{\partial}{\partial \sigma_{\mathrm{g}}^2} \log \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_{\mathrm{g}}^2\left(\mathbf{K}+\delta\mathbf{I}\right)\right)$$

$$= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]$$

set derivative to zero:

$$0 = -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]$$

$$\sigma_{g\mathrm{ML}}^2 = \frac{1}{N}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)$$

- **Bottleneck:** how to compute the trace without the matrix calculation $(\mathbf{K}+\delta\mathbf{I})^{-1}$
- If done naively, this is an $O(N^3)$ operation per SNP.

Linear mixed models (LMM)

ML parameters

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

$$\frac{\partial}{\partial \sigma_{\mathrm{g}}^2} \log \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_{\mathrm{g}}^2\left(\mathbf{K}+\delta\mathbf{I}\right)\right)$$

$$= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]$$

set derivative to zero:

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

- **Bottleneck:** $\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}$
- If done naively, this is an $O(N^3)$ operation per SNP.

Linear mixed models (LMM)

ML parameters

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

$$\frac{\partial}{\partial \sigma_{\mathrm{g}}^2} \log \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_{\mathrm{g}}^2\left(\mathbf{K}+\delta\mathbf{I}\right)\right)$$

$$= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]$$

set derivative to zero:

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

- **Bottleneck:** For every SNP that we test we need to calculate $(\mathbf{K}+\delta\mathbf{I})^{-1}$.
- If done naively, this is an $O(N^3)$ operation per SNP.

Linear mixed models (LMM)

ML parameters

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

$$\frac{\partial}{\partial \sigma_{\mathrm{g}}^2} \log \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_{\mathrm{g}}^2\left(\mathbf{K}+\delta\mathbf{I}\right)\right)$$

$$= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]$$

set derivative to zero:

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

- **Bottleneck:** For every SNP that we test we need to calculate $\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}$.
- If done naively, this is an $O(N^3)$ operation per SNP.

Linear mixed models (LMM)

ML parameters

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

$$\frac{\partial}{\partial \sigma_{\mathrm{g}}^2} \log \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_{\mathrm{g}}^2\left(\mathbf{K}+\delta\mathbf{I}\right)\right)$$

$$= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]$$

set derivative to zero:

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

- **Bottleneck:** For every SNP that we test we need to calculate $\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}$.
- If done naively, this is an $O(N^3)$ operation per SNP. (Very expensive!)

Linear mixed models (LMM)

ML parameters

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

$$\frac{\partial}{\partial \sigma_{\mathrm{g}}^2} \log \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_{\mathrm{g}}^2\left(\mathbf{K}+\delta\mathbf{I}\right)\right)$$

$$= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]$$

set derivative to zero:

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

- **Bottleneck:** For every SNP that we test we need to calculate $\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}$.
- If done naively, this is an $O(N^3)$ operation per SNP. (very expensive!)

Linear mixed models (LMM)

ML parameters

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

$$\frac{\partial}{\partial \sigma_{\mathrm{g}}^2} \log \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_{\mathrm{g}}^2\left(\mathbf{K}+\delta\mathbf{I}\right)\right)$$

$$= -\frac{1}{2}\left[\frac{N}{\sigma_{\mathrm{g}}^2} - \frac{N}{\sigma_g^4}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]$$

set derivative to zero:

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

- **Bottleneck:** For every SNP that we test we need to calculate $\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}$.
- If done naively, this is an $O(N^3)$ operation per SNP. (very expensive!)

FaST LMM

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

[Lippert et al., 2011]

FaST LMM

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

$$=\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}}+\delta\mathbf{I}\right)\right).$$

[Lippert et al., 2011]

FaST LMM

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

$$=\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}}+\delta\mathbf{I}\right)\right).$$

$$=\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{U}^{\mathrm{T}}\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}}\mathbf{U}+\delta\mathbf{U}^{\mathrm{T}}\mathbf{U}\right)\right).$$

[Lippert et al., 2011]

FaST LMM

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

$$=\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}}+\delta\mathbf{I}\right)\right).$$

$$=\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\underbrace{\mathbf{U}^{\mathrm{T}}\mathbf{U}}_{\mathbf{I}}\mathbf{S}\underbrace{\mathbf{U}^{\mathrm{T}}\mathbf{U}}_{\mathbf{I}}+\delta\underbrace{\mathbf{U}^{\mathrm{T}}\mathbf{U}}_{\mathbf{I}}\right)\right).$$

FaST LMM

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

$$=\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}}+\delta\mathbf{I}\right)\right).$$

$$=\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\underbrace{\mathbf{U}^{\mathrm{T}}\mathbf{U}}_{\mathbf{I}}\mathbf{S}\underbrace{\mathbf{U}^{\mathrm{T}}\mathbf{U}}_{\mathbf{I}}+\delta\underbrace{\mathbf{U}^{\mathrm{T}}\mathbf{U}}_{\mathbf{I}}\right)\right).$$

$$=\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{S}+\delta\mathbf{I}\right)\right).$$

[Lippert et al., 2011]

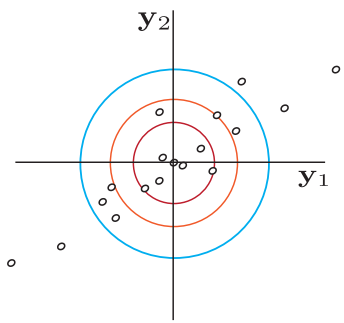
FaST LMM

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

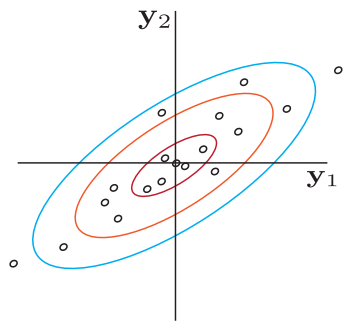
$$=\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^2\left(\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}}+\delta\mathbf{I}\right)\right).$$

$$=\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^2\left(\underbrace{\mathbf{U}^{\mathrm{T}}\mathbf{U}}_{\mathbf{I}}\mathbf{S}\underbrace{\mathbf{U}^{\mathrm{T}}\mathbf{U}}_{\mathbf{I}}+\delta\underbrace{\mathbf{U}^{\mathrm{T}}\mathbf{U}}_{\mathbf{I}}\right)\right).$$

$$=\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^2\left(\mathbf{S}+\delta\mathbf{I}\right)\right).$$



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



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

[Lippert et al., 2011]

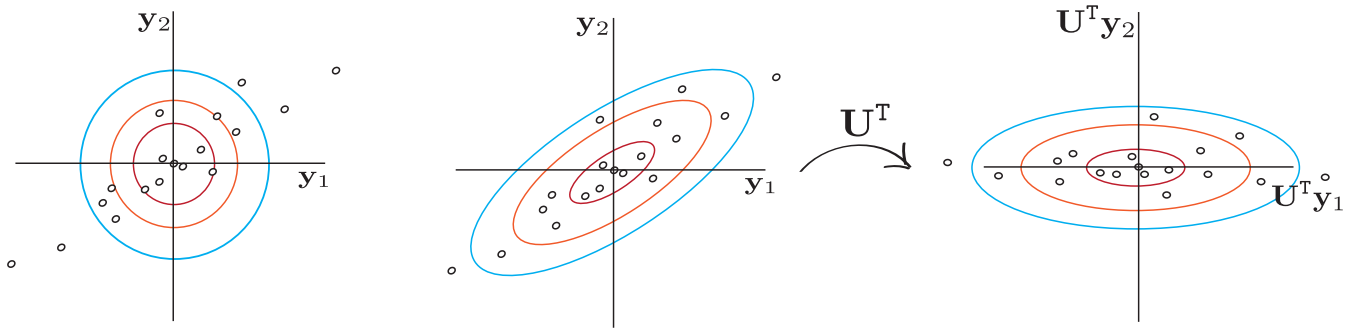
FaST LMM

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

$$=\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^2\left(\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}}+\delta\mathbf{I}\right)\right).$$

$$=\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^2\left(\underbrace{\mathbf{U}^{\mathrm{T}}\mathbf{U}}_{\mathbf{I}}\mathbf{S}\underbrace{\mathbf{U}^{\mathrm{T}}\mathbf{U}}_{\mathbf{I}}+\delta\underbrace{\mathbf{U}^{\mathrm{T}}\mathbf{U}}_{\mathbf{I}}\right)\right).$$

$$=\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^2\left(\mathbf{S}+\delta\mathbf{I}\right)\right).$$



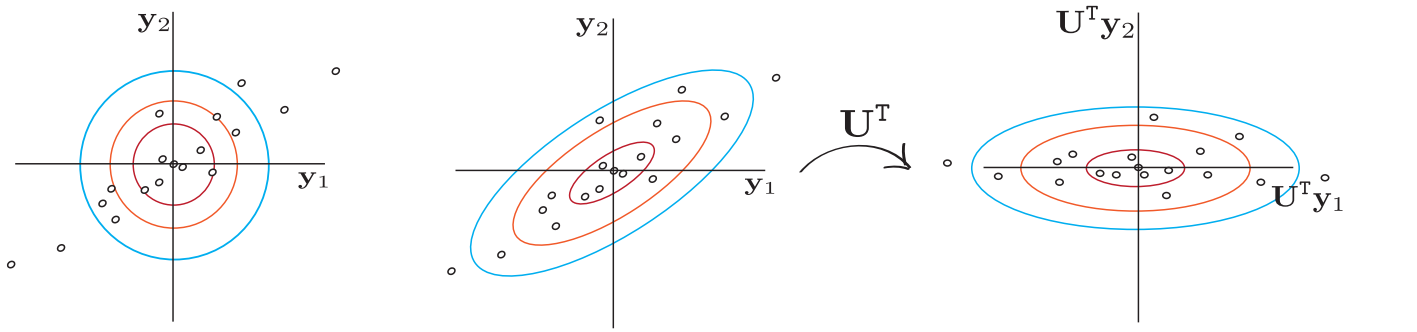
$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_e^2\mathbf{I}\right)$ $\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_g^2\mathbf{K}+\sigma_e^2\mathbf{I}\right)$ $\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_g^2(\mathbf{S}+\delta\mathbf{I})\right)$

[Lippert et al., 2011]

FaST LMM

$$\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{S}+\delta\mathbf{I}\right)\right). \tag{2}$$

- **Fa**ctored **S**pectrally **T**ransformed **LMM**
- $O(N^3)$ once for spectral decomposition.
- $O(N^2)$ runtime per SNP tested (multiplication with $\mathbf{U}^{\mathrm{T}}$).
- $O(N^2)$ memory for storing $\mathbf{K}$ and $\mathbf{U}$.

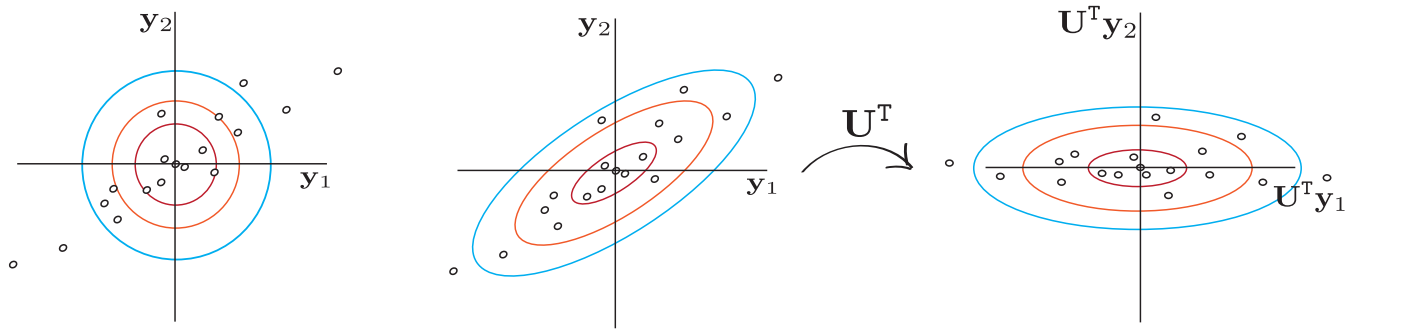


[Lippert et al., 2011]

FaST LMM

$$\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{S}+\delta\mathbf{I}\right)\right). \tag{2}$$

- **Fa**ctored **S**pectrally **T**ransformed **LMM**
- $O(N^3)$ once for spectral decomposition.
- $O(N^2)$ runtime per SNP tested (multiplication with $\mathbf{U}^{\mathrm{T}}$).
- $O(N^2)$ memory for storing $\mathbf{K}$ and $\mathbf{U}$.

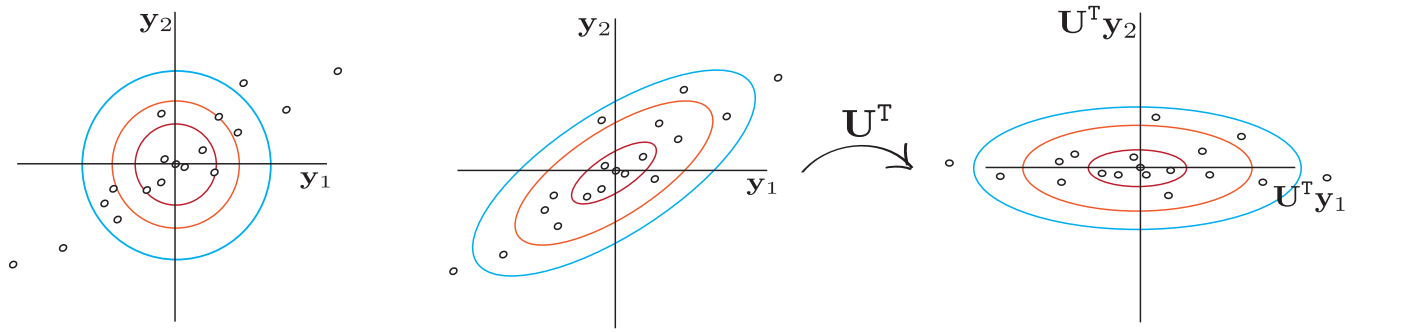


[Lippert et al., 2011]

FaST LMM

$$\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{S}+\delta\mathbf{I}\right)\right). \tag{2}$$

- **Fa**ctored **S**pectrally **T**ransformed **LMM**
- $O(N^3)$ once for spectral decomposition.
- $O(N^2)$ runtime per SNP tested (multiplication with $\mathbf{U}^{\mathrm{T}}$).
- $O(N^2)$ memory for storing $\mathbf{K}$ and $\mathbf{U}$.

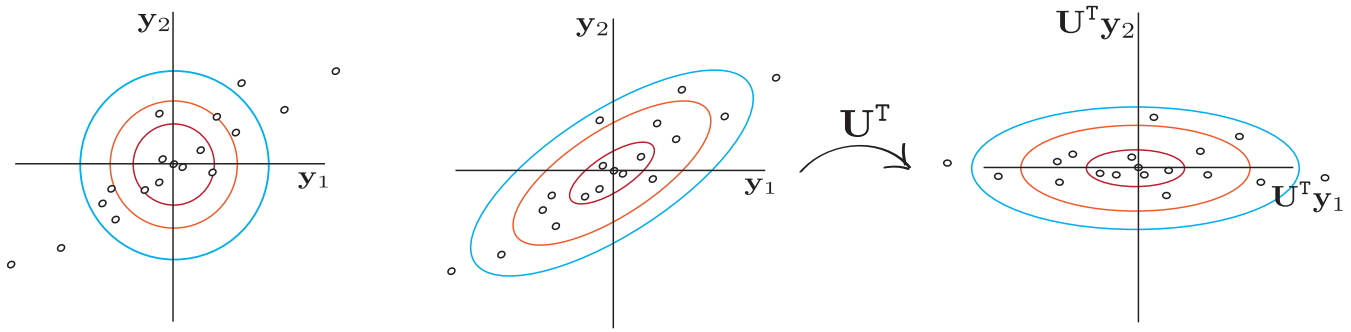


[Lippert et al., 2011]

FaST LMM

$$\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{S}+\delta\mathbf{I}\right)\right). \tag{2}$$

- **Fa**ctored **S**pectrally **T**ransformed **LMM**
- $O(N^3)$ once for spectral decomposition.
- $O(N^2)$ runtime per SNP tested (multiplication with $\mathbf{U}^{\mathrm{T}}$).
- $O(N^2)$ memory for storing $\mathbf{K}$ and $\mathbf{U}$.



[Lippert et al., 2011]

Summary

Population structure correction

- Genomic control
  - Simple method
  - Works with any statistical test
  - Can be combined with other correction methods
  - Very conservative!
- Eigenstrat (PCA)
  - Corrects well for differences on population level
  - Does not work well for closer relatedness
- Linear mixed models
  - Corrects well for most forms of relatedness.

Summary

Population structure correction

- Genomic control
  - Simple method
  - Works with any statistical test
  - Can be combined with other correction methods
  - Very conservative!
- Eigenstrat (PCA)
  - Corrects well for differences on population level
  - Does not work well for closer relatedness
- Linear mixed models
  - Corrects well for most forms of relatedness.

Summary

Population structure correction

- Genomic control
  - Simple method
  - Works with any statistical test
  - Can be combined with other correction methods
  - Very conservative!
- Eigenstrat (PCA)
  - Corrects well for differences on population level
  - Does not work well for closer relatedness
- Linear mixed models
  - Corrects well for most forms of relatedness.

Overview

- Single marker association model with random effect term

$$\mathbf{y} = \underbrace{\mathbf{x}_s\beta_s}_{\text{genetic effect}} + \underbrace{\mathbf{u}}_{\text{random effect covariates}} + \underbrace{\boldsymbol{\epsilon}}_{\text{noise}}$$

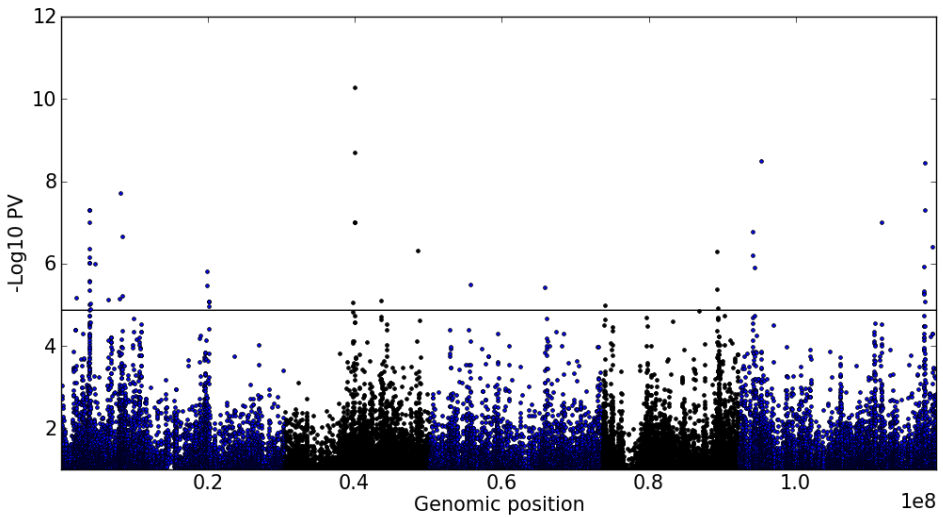
- Shortcomings
  - Weak effects are not captured by single-marker analysis.
  - Complex traits are controlled by more than a single SNP.

Overview

- Single marker association model with random effect term

$$\mathbf{y} = \underbrace{\mathbf{x}_s\beta_s}_{\text{genetic effect}} + \underbrace{\mathbf{u}}_{\text{random effect covariates}} + \underbrace{\boldsymbol{\epsilon}}_{\text{noise}}$$

- Shortcomings
  - Weak effects are not captured by single-marker analysis.
  - Complex traits are controlled by more than a single SNP.

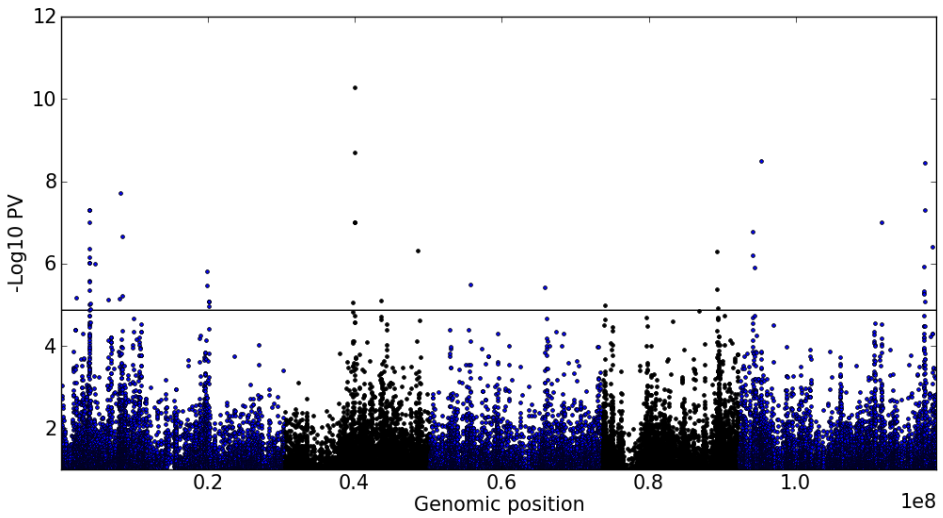


Overview

- Single marker association model with random effect term

$$\mathbf{y} = \underbrace{\mathbf{x}_s\beta_s}_{\text{genetic effect}} + \underbrace{\mathbf{u}}_{\text{random effect covariates}} + \underbrace{\boldsymbol{\epsilon}}_{\text{noise}}$$

- Shortcomings
  - Weak effects are not captured by single-marker analysis.
  - Complex traits are controlled by more than a single SNP.



Multi locus models

- Generalization to multiple genetic factors

$$\mathbf{y} = \underbrace{\sum_{s=1}^{S} \mathbf{x}_s \beta_s}_{\text{genetic effect}} + \underbrace{\mathbf{u}}_{\text{random effect covariates}} + \underbrace{\boldsymbol{\epsilon}}_{\text{noise}}$$

- Challenge: $N << S$: explicit estimation of all β_s is not feasible.
- Solutions
  - Regularize β_s (Ridge regression, LASSO)
  - Variance component models

[Wu et al., 2011]

Multi locus models

- Generalization to multiple genetic factors

$$\mathbf{y} = \underbrace{\sum_{s=1}^{S} \mathbf{x}_s \beta_s}_{\text{genetic effect}} + \underbrace{\mathbf{u}}_{\text{random effect covariates}} + \underbrace{\boldsymbol{\epsilon}}_{\text{noise}}$$

- Challenge: $N << S$: explicit estimation of all β_s is not feasible.
- Solutions
  - Regularize β_s (Ridge regression, LASSO)
  - Variance component modeling

[Wu et al., 2011]

Multi locus models

- Generalization to multiple genetic factors

$$\mathbf{y} = \underbrace{\sum_{s=1}^{S} \mathbf{x}_s \beta_s}_{\text{genetic effect}} + \underbrace{\mathbf{u}}_{\text{random effect covariates}} + \underbrace{\boldsymbol{\epsilon}}_{\text{noise}}$$

- Challenge: $N << S$: explicit estimation of all β_s is not feasible.
- Solutions
  - Regularize β_s (Ridge regression, LASSO)
  - Variance component modeling

[Wu et al., 2011]

Multi locus models

- Generalization to multiple genetic factors

$$\mathbf{y} = \underbrace{\sum_{s=1}^{S} \mathbf{x}_s \beta_s}_{\text{genetic effect}} + \underbrace{\mathbf{u}}_{\text{random effect covariates}} + \underbrace{\boldsymbol{\epsilon}}_{\text{noise}}$$

- Challenge: $N << S$: explicit estimation of all β_s is not feasible.
- Solutions
  - Regularize β_s (Ridge regression, LASSO)
  - Variance component modeling

[Wu et al., 2011]

Outline

Outline

Multi locus models

Random effect models

- For now, let's drop the random effect term

$$\mathbf{y} = \sum_{s=1}^{S} \mathbf{x}_s \beta_s + \boldsymbol{\epsilon}.$$

- For mathematical convenience, we choose a shared Gaussian distribution on the weights and Gaussian noise

$$p(\beta_1, \ldots, \beta_S) = \prod_{s=1}^{S} \mathcal{N}\left(\beta_s \,|\, 0, \sigma_g^2\right) \quad p(\boldsymbol{\epsilon}) = \mathcal{N}\left(\boldsymbol{\epsilon} \,|\, \mathbf{0}, \sigma_e^2 \mathbf{I}\right)$$

- Marginalize out the weights $\beta_1, \ldots, \beta_S$

$$p(\mathbf{y} \,|\, \mathbf{X}, \sigma_g^2, \sigma_e^2) = \int_{\beta} \mathcal{N}\left(\mathbf{y} \,\middle|\, \sum_{s=1}^{S} \mathbf{x}_s \beta_s, \sigma_e^2 \mathbf{I}\right) \mathrm{d}\beta$$

$$= \mathcal{N}\left(\mathbf{y} \,|\, \mathbf{0}, \sum_{s=1}^{S} \mathbf{x}_s \mathbf{x}_s^T \sigma_g^2 + \sigma_e^2 \mathbf{I}\right)$$

Multi locus models

Random effect models

- For now, let's drop the random effect term

$$\mathbf{y} = \sum_{s=1}^{S} \mathbf{x}_s \beta_s + \boldsymbol{\epsilon}.$$

- For mathematical convenience, we choose a shared Gaussian distribution on the weights and Gaussian noise

$$p(\beta_1, \ldots, \beta_S) = \prod_{s=1}^{S} \mathcal{N}\left(\beta_s \,\middle|\, 0, \sigma_g^2\right) \quad p(\boldsymbol{\epsilon}) = \mathcal{N}\left(\boldsymbol{\epsilon} \,\middle|\, \mathbf{0}, \sigma_e^2 \mathbf{I}\right)$$

- Marginalize out the weights $\beta_1, \ldots, \beta_S$

$$p(\mathbf{y} \mid \mathbf{X}, \sigma_g^2, \sigma_e^2) = \int_{\beta} \underbrace{\mathcal{N}\left(\mathbf{y} \,\middle|\, \sum_{s=1}^{S} \mathbf{x}_s \beta_s, \sigma_e^2 \mathbf{I}\right)}_{\text{Data likelihood}} \mathrm{d}\beta$$

$$= \mathcal{N}\left(\mathbf{y} \,\middle|\, \mathbf{0}, \sum_{s=1}^{S} \mathbf{x}_s \mathbf{x}_s^T \sigma_g^2 + \sigma_e^2 \mathbf{I}\right)$$

Multi locus models

Random effect models

- For now, let's drop the random effect term

$$\mathbf{y} = \sum_{s=1}^{S} \mathbf{x}_s \beta_s + \boldsymbol{\epsilon}.$$

- For mathematical convenience, we choose a shared Gaussian distribution on the weights and Gaussian noise

$$p(\beta_1, \ldots, \beta_S) = \prod_{s=1}^{S} \mathcal{N}\left(\beta_s \,|\, 0, \sigma_g^2\right) \quad p(\boldsymbol{\epsilon}) = \mathcal{N}\left(\boldsymbol{\epsilon} \,|\, \mathbf{0}, \sigma_e^2 \mathbf{I}\right)$$

- Marginalize out the weights $\beta_1, \ldots, \beta_S$

$$p(\mathbf{y} \,|\, \mathbf{X}, \sigma_{\mathrm{g}}^2, \sigma_{\mathrm{e}}^2) = \int_{\beta} \underbrace{\mathcal{N}\left(\mathbf{y} \,\middle|\, \sum_{s=1}^{S} \mathbf{x}_s \beta_s, \sigma_{\mathrm{e}}^2 \mathbf{I}\right)}_{\text{Data likelihood}} \underbrace{\prod_{s=1}^{S} \mathcal{N}\left(\beta_s \,|\, 0, \sigma_{\mathrm{g}}^2\right)}_{\text{weight distribution}} \mathrm{d}\boldsymbol{\beta}$$

$$= \mathcal{N}\left(\mathbf{y} \,\middle|\, \mathbf{0}, \sigma_{\mathrm{g}}^2 \sum_{s=1}^{S} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)$$

Multi locus models

Random effect models

- For now, let's drop the random effect term

$$\mathbf{y} = \sum_{s=1}^{S} \mathbf{x}_s \beta_s + \boldsymbol{\epsilon}.$$

- For mathematical convenience, we choose a shared Gaussian distribution on the weights and Gaussian noise

$$p(\beta_1, \ldots, \beta_S) = \prod_{s=1}^{S} \mathcal{N}\left(\beta_s \,|\, 0, \sigma_g^2\right) \quad p(\boldsymbol{\epsilon}) = \mathcal{N}\left(\boldsymbol{\epsilon} \,|\, \mathbf{0}, \sigma_e^2 \mathbf{I}\right)$$

- Marginalize out the weights $\beta_1, \ldots, \beta_S$

$$p(\mathbf{y} \,|\, \mathbf{X}, \sigma_{\mathrm{g}}^2, \sigma_{\mathrm{e}}^2) = \int_{\boldsymbol{\beta}} \underbrace{\mathcal{N}\left(\mathbf{y} \,\middle|\, \sum_{s=1}^{S} \mathbf{x}_s \beta_s, \sigma_{\mathrm{e}}^2 \mathbf{I}\right)}_{\text{Data likelihood}} \underbrace{\prod_{s=1}^{S} \mathcal{N}\left(\beta_s \,|\, 0, \sigma_{\mathrm{g}}^2\right)}_{\text{weight distribution}} \mathrm{d}\boldsymbol{\beta}$$

$$= \mathcal{N}\left(\mathbf{y} \,\middle|\, \mathbf{0}, \sigma_{\mathrm{g}}^2 \sum_{s=1}^{S} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)$$

Multi locus models

Random effect models

- For now, let's drop the random effect term

$$\mathbf{y} = \sum_{s=1}^{S} \mathbf{x}_s \beta_s + \boldsymbol{\epsilon}.$$

- For mathematical convenience, we choose a shared Gaussian distribution on the weights and Gaussian noise

$$p(\beta_1, \ldots, \beta_S) = \prod_{s=1}^{S} \mathcal{N}\left(\beta_s \,|\, 0, \sigma_g^2\right) \quad p(\boldsymbol{\epsilon}) = \mathcal{N}\left(\boldsymbol{\epsilon} \,|\, \mathbf{0}, \sigma_e^2 \mathbf{I}\right)$$

- Marginalize out the weights $\beta_1, \ldots, \beta_S$

$$p(\mathbf{y} \,|\, \mathbf{X}, \sigma_{\mathrm{g}}^2, \sigma_{\mathrm{e}}^2) = \int_{\boldsymbol{\beta}} \underbrace{\mathcal{N}\left(\mathbf{y} \,\middle|\, \sum_{s=1}^{S} \mathbf{x}_s \beta_s, \sigma_{\mathrm{e}}^2 \mathbf{I}\right)}_{\text{Data likelihood}} \underbrace{\prod_{s=1}^{S} \mathcal{N}\left(\beta_s \,|\, 0, \sigma_{\mathrm{g}}^2\right)}_{\text{weight distribution}} \mathrm{d}\boldsymbol{\beta}$$

$$= \mathcal{N}\left(\mathbf{y} \,\middle|\, \mathbf{0}, \sigma_{\mathrm{g}}^2 \sum_{s=1}^{S} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)$$

Multi locus models

Random effect models

- For now, let's drop the random effect term

$$\mathbf{y} = \sum_{s=1}^{S} \mathbf{x}_s \beta_s + \boldsymbol{\epsilon}.$$

- For mathematical convenience, we choose a shared Gaussian distribution on the weights and Gaussian noise

$$p(\beta_1, \ldots, \beta_S) = \prod_{s=1}^{S} \mathcal{N}\left(\beta_s \,|\, 0, \sigma_g^2\right) \quad p(\boldsymbol{\epsilon}) = \mathcal{N}\left(\boldsymbol{\epsilon} \,|\, \mathbf{0}, \sigma_e^2 \mathbf{I}\right)$$

- Marginalize out the weights $\beta_1, \ldots, \beta_S$

$$p(\mathbf{y} \,|\, \mathbf{X}, \sigma_{\mathrm{g}}^2, \sigma_{\mathrm{e}}^2) = \int_{\boldsymbol{\beta}} \underbrace{\mathcal{N}\left(\mathbf{y} \,\middle|\, \sum_{s=1}^{S} \mathbf{x}_s \beta_s, \sigma_{\mathrm{e}}^2 \mathbf{I}\right)}_{\text{Data likelihood}} \underbrace{\prod_{s=1}^{S} \mathcal{N}\left(\beta_s \,|\, 0, \sigma_{\mathrm{g}}^2\right)}_{\text{weight distribution}} \mathrm{d}\boldsymbol{\beta}$$

$$= \mathcal{N}\left(\mathbf{y} \,\middle|\, \mathbf{0}, \sigma_{\mathrm{g}}^2 \sum_{s=1}^{S} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)$$

Multi locus models

Remarks

$$p(\mathbf{y} \mid \mathbf{X}, \sigma_{\mathrm{g}}^2, \sigma_{\mathrm{e}}^2) = \mathcal{N}\left(\mathbf{y} \mid \mathbf{0}, \sigma_{\mathrm{g}}^2 \underbrace{\sum_{s=1}^{S} \mathbf{s}_s \mathbf{x}_s^{\mathrm{T}}}_{\mathbf{K}_g} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right) \tag{3}$$

- $\mathbf{K}_g$ genotype covariance matrix.
  - Closely related to Kinship explaining population structure.
- Inference can be done my maximum likelihood.
- The ratio of σ_{g}^2 and σ_{e}^2 defines the narrow sense heritability of the trait

$$h^2 = \frac{\sigma_{\mathrm{g}}^2}{\sigma_{\mathrm{g}}^2 + \sigma_{\mathrm{e}}^2}.$$

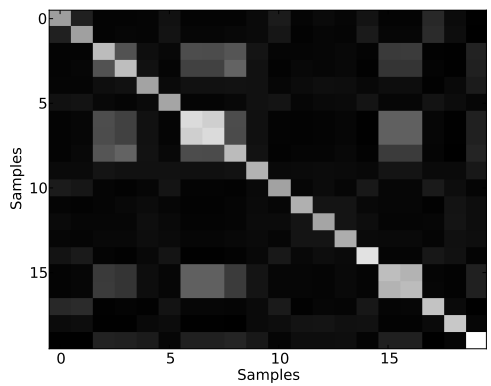
Multi locus models

Remarks

$$p(\mathbf{y} \mid \mathbf{X}, \sigma_{\mathrm{g}}^2, \sigma_{\mathrm{e}}^2) = \mathcal{N}\left(\mathbf{y} \mid \mathbf{0}, \sigma_{\mathrm{g}}^2 \underbrace{\sum_{s=1}^{S} \mathbf{s}_s \mathbf{x}_s^{\mathrm{T}}}_{\mathbf{K}_g} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right) \tag{3}$$

- $\mathbf{K}_g$ genotype covariance matrix.
  - Closely related to Kinship explaining population structure.
- Inference can be done my maximum likelihood.
- The ratio of σ_{g}^2 and σ_{e}^2 defines the narrow sense heritability of the trait

$$h^2 = \frac{\sigma_{\mathrm{g}}^2}{\sigma_{\mathrm{g}}^2 + \sigma_{\mathrm{e}}^2}.$$



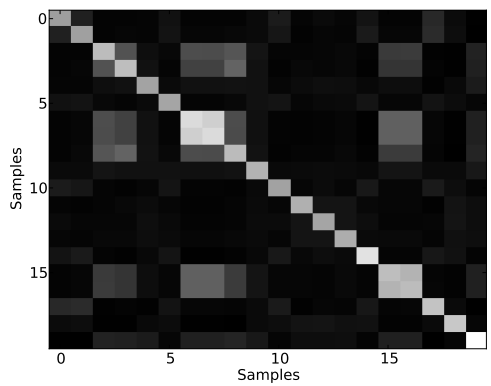
Multi locus models

Remarks

$$p(\mathbf{y} \mid \mathbf{X}, \sigma_{\mathrm{g}}^2, \sigma_{\mathrm{e}}^2) = \mathcal{N}\left(\mathbf{y} \mid \mathbf{0}, \sigma_{\mathrm{g}}^2 \underbrace{\sum_{s=1}^{S} \mathbf{s}_s \mathbf{x}_s^{\mathrm{T}}}_{\mathbf{K}_g} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right) \tag{3}$$

- $\mathbf{K}_g$ genotype covariance matrix.
  - Closely related to Kinship explaining population structure.
- Inference can be done my maximum likelihood.
- The ratio of σ_{g}^2 and σ_{e}^2 defines the narrow sense heritability of the trait

$$h^2 = \frac{\sigma_{\mathrm{g}}^2}{\sigma_{\mathrm{g}}^2 + \sigma_{\mathrm{e}}^2}.$$

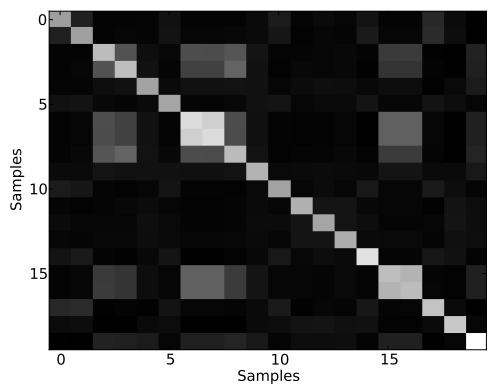


Multi locus models

Remarks

$$p(\mathbf{y} \mid \mathbf{X}, \sigma_{\mathrm{g}}^2, \sigma_{\mathrm{e}}^2) = \mathcal{N}\left(\mathbf{y} \mid \mathbf{0}, \sigma_{\mathrm{g}}^2 \underbrace{\sum_{s=1}^{S} \mathbf{s}_s \mathbf{x}_s^{\mathrm{T}}}_{\mathbf{K}_g} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right) \tag{3}$$

- $\mathbf{K}_g$ genotype covariance matrix.
  - Closely related to Kinship explaining population structure.
- Inference can be done my maximum likelihood.
- The ratio of σ_{g}^2 and σ_{e}^2 defines the narrow sense heritability of the trait

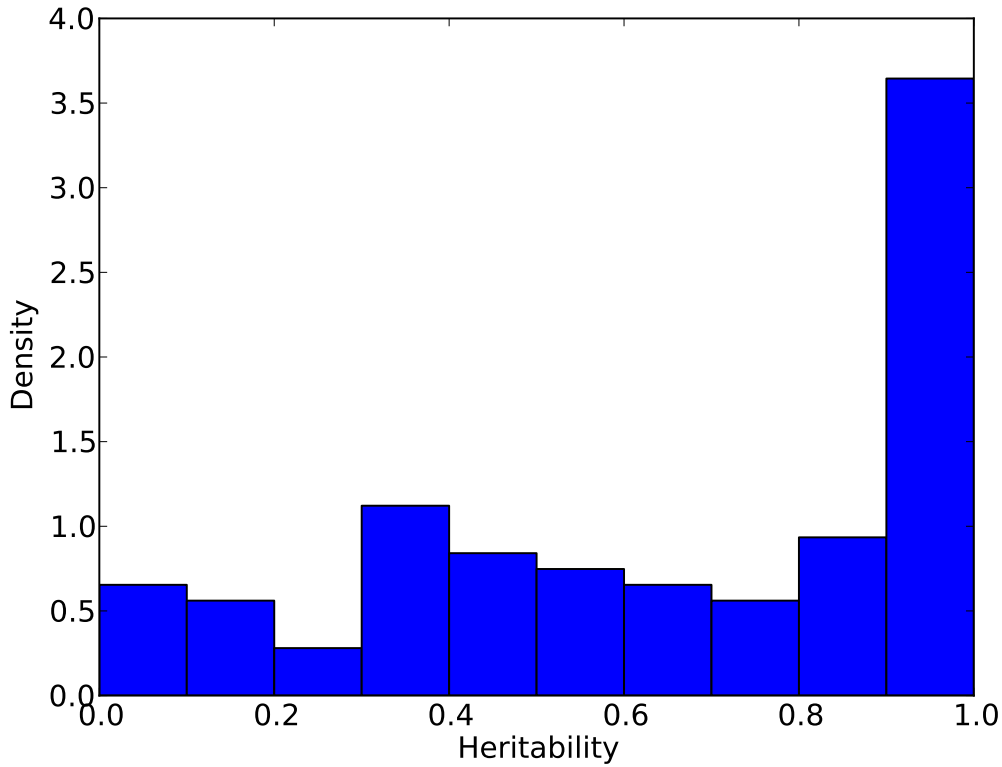


$$h^2 = \frac{\sigma_{\mathrm{g}}^2}{\sigma_{\mathrm{g}}^2 + \sigma_{\mathrm{e}}^2}.$$

Heritability

Heritability estimated on 107 *A. thaliana* phenotypes

Global genetic heritability

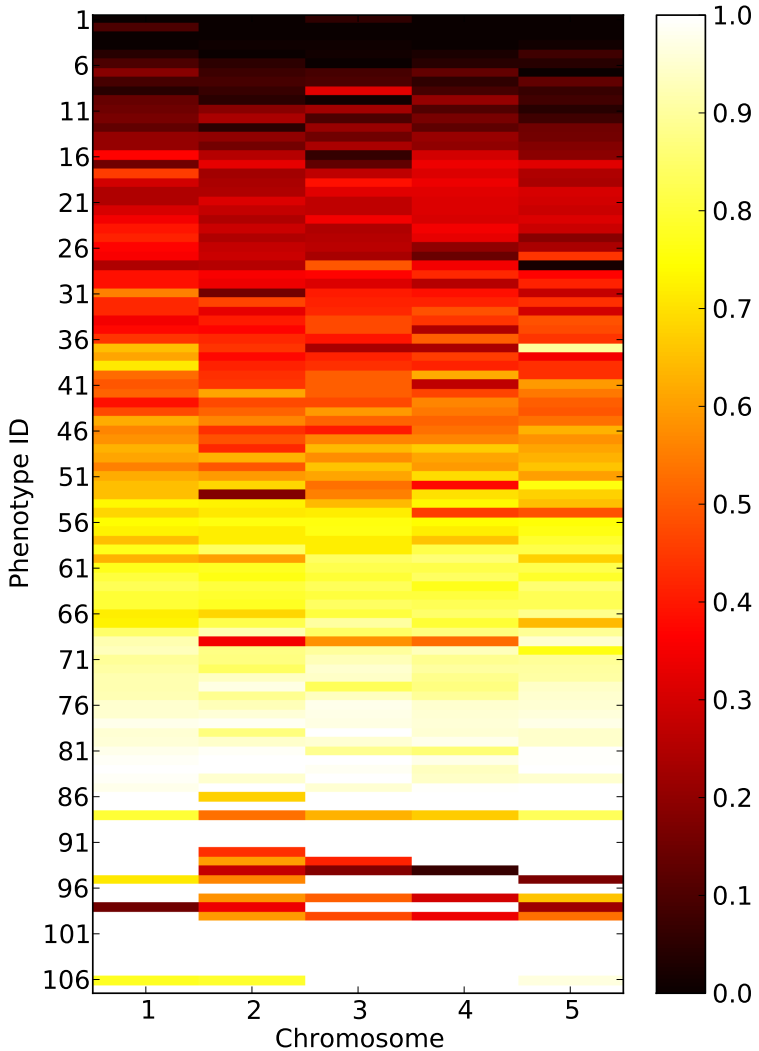


Heritability

Heritability estimated on 107 *A. thaliana* phenotypes

- Estimate can be restricted to a genomic region such as a single chromosome, etc.

$$\mathcal{N}\left(\mathbf{y} \,|\, \mathbf{0}, \sigma_{\mathrm{g}}^2 \sum_{s \in \mathrm{Chrom}} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)$$



Window-based composite variance analysis

Region-based testing

- Just fitting a particular region ignores the genome-wide context
- Variance dissection with region-based separation

$$p(\mathbf{y} \mid W) = \mathcal{N}(\mathbf{y} \mid \mathbf{0}, \sigma_w^2 \underbrace{\sum_{s \in W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}}}_{\mathbf{K}_w} + \sigma_g^2 \underbrace{\sum_{s \notin W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}}}_{\mathbf{K}_g} + \sigma_{\mathrm{e}}^2 \mathbf{I})$$

- Explained variance components can be read off subject to suitable normalization of the covariances $\mathbf{K}_w$ and $\mathbf{K}_g$.
- "Local" heritability

$$h^2(W) = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_g^2 + \sigma_e^2}$$

Window-based composite variance analysis

Region-based testing

- Just fitting a particular region ignores the genome-wide context
- Variance dissection with region-based separation

$$p(\mathbf{y} \mid W) = \mathcal{N}(\mathbf{y} \mid \mathbf{0}, \sigma_w^2 \underbrace{\sum_{s \in W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}}}_{\mathbf{K}_w} + \sigma_g^2 \underbrace{\sum_{s \notin W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}}}_{\mathbf{K}_g} + \sigma_{\mathrm{e}}^2 \mathbf{I})$$

- Explained variance components can be read off subject to suitable normalization of the covariances $\mathbf{K}_w$ and $\mathbf{K}_g$.
- "Local" heritability

$$h^2(W) = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_g^2 + \sigma_e^2}$$

Window-based composite variance analysis

Region-based testing

- Just fitting a particular region ignores the genome-wide context
- Variance dissection with region-based separation

$$p(\mathbf{y} \mid W) = \mathcal{N}(\mathbf{y} \mid \mathbf{0}, \sigma_w^2 \underbrace{\sum_{s \in W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}}}_{\mathbf{K}_w} + \sigma_g^2 \underbrace{\sum_{s \notin W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}}}_{\mathbf{K}_g} + \sigma_{\mathrm{e}}^2 \mathbf{I})$$

- Explained variance components can be read off subject to suitable normalization of the covariances $\mathbf{K}_w$ and $\mathbf{K}_g$.
- "Local" heritability

$$h^2(W) = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_g^2 + \sigma_e^2}$$

Window-based composite variance analysis

Region-based testing

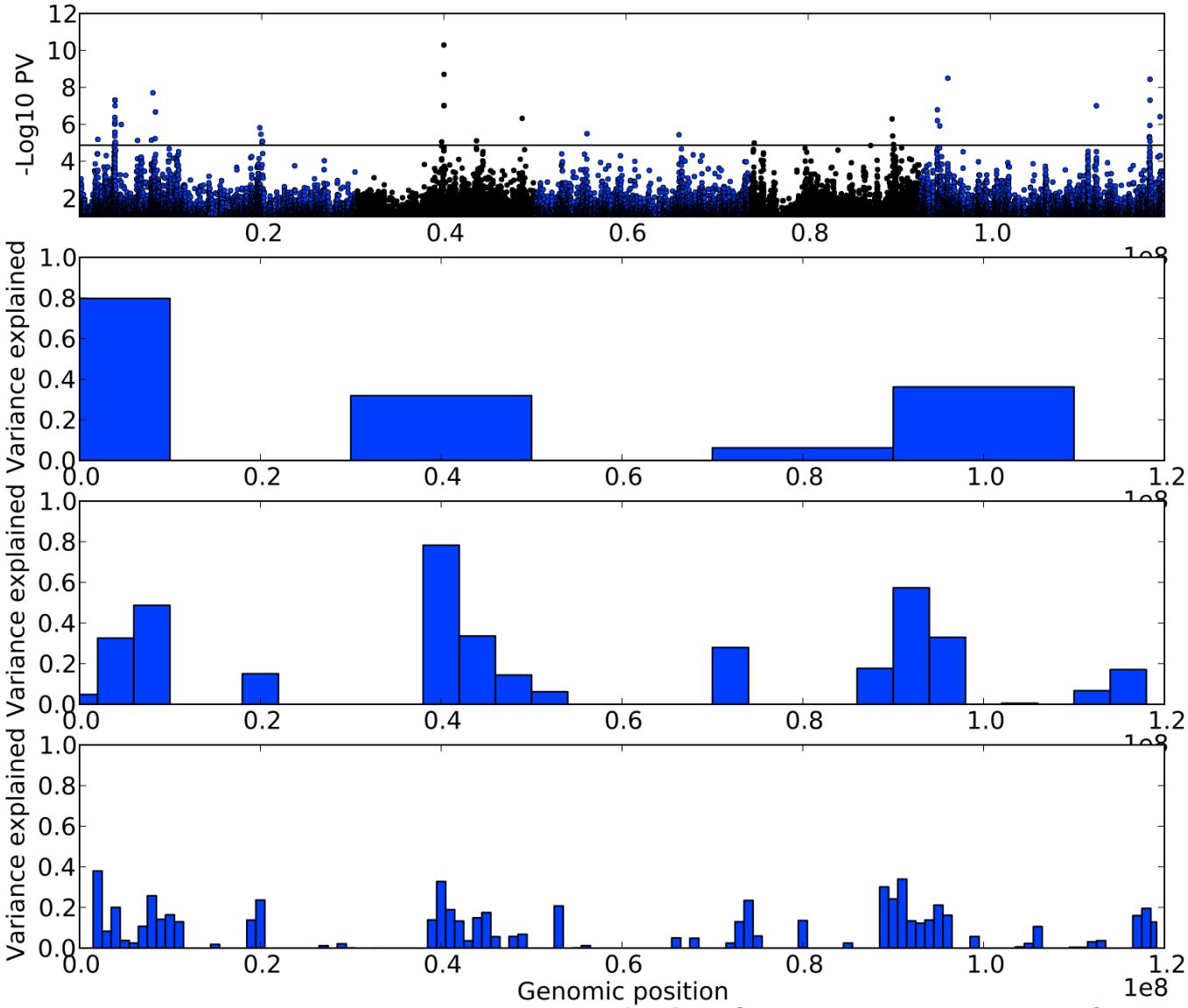
- Just fitting a particular region ignores the genome-wide context
- Variance dissection with region-based separation

$$p(\mathbf{y} \mid W) = \mathcal{N}(\mathbf{y} \mid \mathbf{0}, \sigma_w^2 \underbrace{\sum_{s \in W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}}}_{\mathbf{K}_w} + \sigma_g^2 \underbrace{\sum_{s \notin W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}}}_{\mathbf{K}_g} + \sigma_{\mathrm{e}}^2 \mathbf{I})$$

- Explained variance components can be read off subject to suitable normalization of the covariances $\mathbf{K}_w$ and $\mathbf{K}_g$.
- "Local" heritability

$$h^2(W) = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_g^2 + \sigma_e^2}$$

Window-based composite variance analysis



Window-based composite variance analysis

Significance testing

- Analogously to fixed effect testing, the significance of a specific window can be tested.
- Likelihood-ratio statistics to score the relevance of a particular genomic region W

$$\mathrm{LOD}(W) = \frac{\mathcal{N}\left(\mathbf{y} \middle| \mathbf{0}, \sigma_w^2 \sum_{s \in W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}} + \sigma_g^2 \sum_{s \notin W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)}{\mathcal{N}\left(\mathbf{y} \middle| \mathbf{0}, \sigma_g^2 \sum_{s \notin W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)}$$

- P-values can be obtained from permutation statistics or analytical approximation (variants of score tests or likelihood ratio tests).

[Wu et al., 2011, Listgarten et al., 2012]

Window-based composite variance analysis

Significance testing

- Analogously to fixed effect testing, the significance of a specific window can be tested.
- Likelihood-ratio statistics to score the relevance of a particular genomic region W

$$\mathrm{LOD}(W) = \frac{\mathcal{N}\left(\mathbf{y} \middle| \mathbf{0}, \sigma_w^2 \sum_{s \in W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}} + \sigma_g^2 \sum_{s \notin W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)}{\mathcal{N}\left(\mathbf{y} \middle| \mathbf{0}, \sigma_g^2 \sum_{s \notin W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)}$$

- P-values can be obtained from permutation statistics or analytical approximation (variants of score tests or likelihood ratio tests).

[Wu et al., 2011, Listgarten et al., 2012]

Window-based composite variance analysis

Significance testing

- Analogously to fixed effect testing, the significance of a specific window can be tested.
- Likelihood-ratio statistics to score the relevance of a particular genomic region W

$$\mathrm{LOD}(W) = \frac{\mathcal{N}\left(\mathbf{y} \middle| \mathbf{0}, \sigma_w^2 \sum_{s \in W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}} + \sigma_g^2 \sum_{s \notin W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)}{\mathcal{N}\left(\mathbf{y} \middle| \mathbf{0}, \sigma_g^2 \sum_{s \notin W} \mathbf{x}_s \mathbf{x}_s^{\mathrm{T}} + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)}$$

- P-values can be obtained from permutation statistics or analytical approximation (variants of score tests or likelihood ratio tests).

[Wu et al., 2011, Listgarten et al., 2012]

Making predictions with linear mixed models

- Linear model, accounting for a set of measured SNPs $\mathbf{X}$

$$p(\mathbf{y} \,|\, \mathbf{X}, \boldsymbol{\theta}, \sigma^2) = \mathcal{N}\left(\mathbf{y} \,\middle|\, \sum_{s=1}^{S} \mathbf{x}_s \theta_s, \sigma^2 \mathbf{I}\right)$$

- Prediction at unseen test input given max. likelihood weight:

$$p(y^\star \,|\, \mathbf{x}^\star, \hat{\boldsymbol{\theta}}) = \mathcal{N}\left(y^\star \,\middle|\, \mathbf{x}^\star \hat{\boldsymbol{\theta}}, \sigma^2\right)$$

- Marginal likelihood

$$p(\mathbf{y} \,|\, \mathbf{X}, \sigma^2, \sigma_g^2) = \int_{\boldsymbol{\theta}} \mathcal{N}\left(\mathbf{y} \,|\, \mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}\right) \mathcal{N}\left(\boldsymbol{\theta} \,|\, \mathbf{0}, \sigma_g^2 \mathbf{I}\right)$$

$$= \mathcal{N}\left(\mathbf{y} \,\middle|\, \mathbf{0}, \underbrace{\sigma_g^2 \mathbf{X}\mathbf{X}^{\mathrm{T}}}_{\mathbf{K}} + \sigma^2 \mathbf{I}\right)$$

- Making predictions with linear mixed models?

Making predictions with linear mixed models

- Linear model, accounting for a set of measured SNPs $\mathbf{X}$

$$p(\mathbf{y} \,|\, \mathbf{X}, \boldsymbol{\theta}, \sigma^2) = \mathcal{N}\left(\mathbf{y} \,\middle|\, \sum_{s=1}^{S} \mathbf{x}_s \theta_s, \sigma^2 \mathbf{I}\right)$$

- Prediction at unseen test input given max. likelihood weight:

$$p(y^\star \,|\, \mathbf{x}^\star, \hat{\boldsymbol{\theta}}) = \mathcal{N}\left(y^\star \,\middle|\, \mathbf{x}^\star \hat{\boldsymbol{\theta}}, \sigma^2\right)$$

- Marginal likelihood

$$p(\mathbf{y} \,|\, \mathbf{X}, \sigma^2, \sigma_g^2) = \int_{\boldsymbol{\theta}} \mathcal{N}\left(\mathbf{y} \,|\, \mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}\right) \mathcal{N}\left(\boldsymbol{\theta} \,|\, \mathbf{0}, \sigma_g^2 \mathbf{I}\right)$$

$$= \mathcal{N}\left(\mathbf{y} \,\middle|\, \mathbf{0}, \underbrace{\sigma_g^2 \mathbf{X}\mathbf{X}^{\mathrm{T}}}_{\mathbf{K}} + \sigma^2 \mathbf{I}\right)$$

- Making predictions with linear mixed models?

Making predictions with linear mixed models

- Linear model, accounting for a set of measured SNPs $\mathbf{X}$

$$p(\mathbf{y} \,|\, \mathbf{X}, \boldsymbol{\theta}, \sigma^2) = \mathcal{N}\left(\mathbf{y} \,\middle|\, \sum_{s=1}^{S} \mathbf{x}_s \theta_s, \sigma^2 \mathbf{I}\right)$$

- Prediction at unseen test input given max. likelihood weight:

$$p(y^\star \,|\, \mathbf{x}^\star, \hat{\boldsymbol{\theta}}) = \mathcal{N}\left(y^\star \,\middle|\, \mathbf{x}^\star \hat{\boldsymbol{\theta}}, \sigma^2\right)$$

- Marginal likelihood

$$p(\mathbf{y} \,|\, \mathbf{X}, \sigma^2, \sigma_g^2) = \int_{\boldsymbol{\theta}} \mathcal{N}\left(\mathbf{y} \,|\, \mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}\right) \mathcal{N}\left(\boldsymbol{\theta} \,|\, \mathbf{0}, \sigma_g^2 \mathbf{I}\right)$$

$$= \mathcal{N}\left(\mathbf{y} \,\middle|\, \mathbf{0}, \underbrace{\sigma_g^2 \mathbf{X}\mathbf{X}^{\mathrm{T}}}_{\mathbf{K}} + \sigma^2 \mathbf{I}\right)$$

- Making predictions with linear mixed models?

Making predictions with linear mixed models

- Linear model, accounting for a set of measured SNPs $\mathbf{X}$

$$p(\mathbf{y} \,|\, \mathbf{X}, \boldsymbol{\theta}, \sigma^2) = \mathcal{N}\left(\mathbf{y} \,\middle|\, \sum_{s=1}^{S} \mathbf{x}_s \theta_s, \sigma^2 \mathbf{I}\right)$$

- Prediction at unseen test input given max. likelihood weight:

$$p(y^\star \,|\, \mathbf{x}^\star, \hat{\boldsymbol{\theta}}) = \mathcal{N}\left(y^\star \,\middle|\, \mathbf{x}^\star \hat{\boldsymbol{\theta}}, \sigma^2\right)$$

- Marginal likelihood

$$p(\mathbf{y} \,|\, \mathbf{X}, \sigma^2, \sigma_g^2) = \int_{\boldsymbol{\theta}} \mathcal{N}\left(\mathbf{y} \,|\, \mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}\right) \mathcal{N}\left(\boldsymbol{\theta} \,|\, \mathbf{0}, \sigma_g^2 \mathbf{I}\right)$$

$$= \mathcal{N}\left(\mathbf{y} \,\middle|\, \mathbf{0}, \underbrace{\sigma_g^2 \mathbf{X}\mathbf{X}^{\mathrm{T}}}_{\mathbf{K}} + \sigma^2 \mathbf{I}\right)$$

- Making predictions with linear mixed models?

The Gaussian distribution

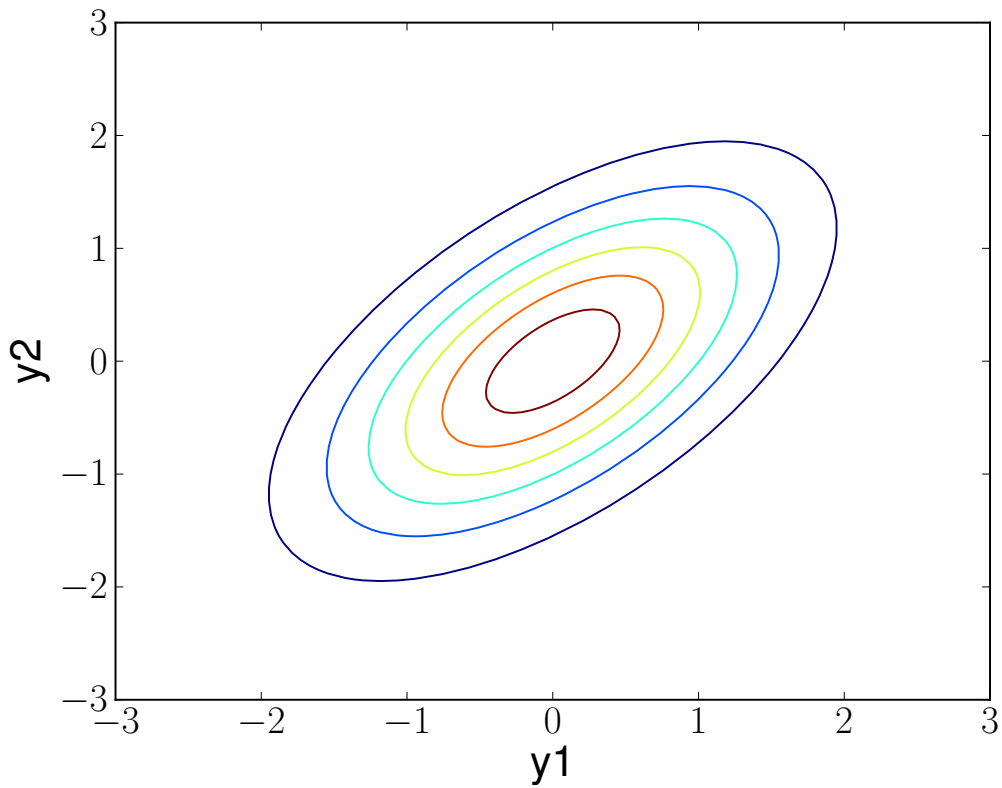
- Linear mixed models are merely based on the good old multivariate Gaussian

$$\mathcal{N}\left(\mathbf{x}\,\middle|\,\boldsymbol{\mu}, \mathbf{K}\right) = \frac{1}{\sqrt{|2\pi\,\mathbf{K}|}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\,\mathbf{K}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

- Covariance matrix or kernel matrix

A 2D Gaussian

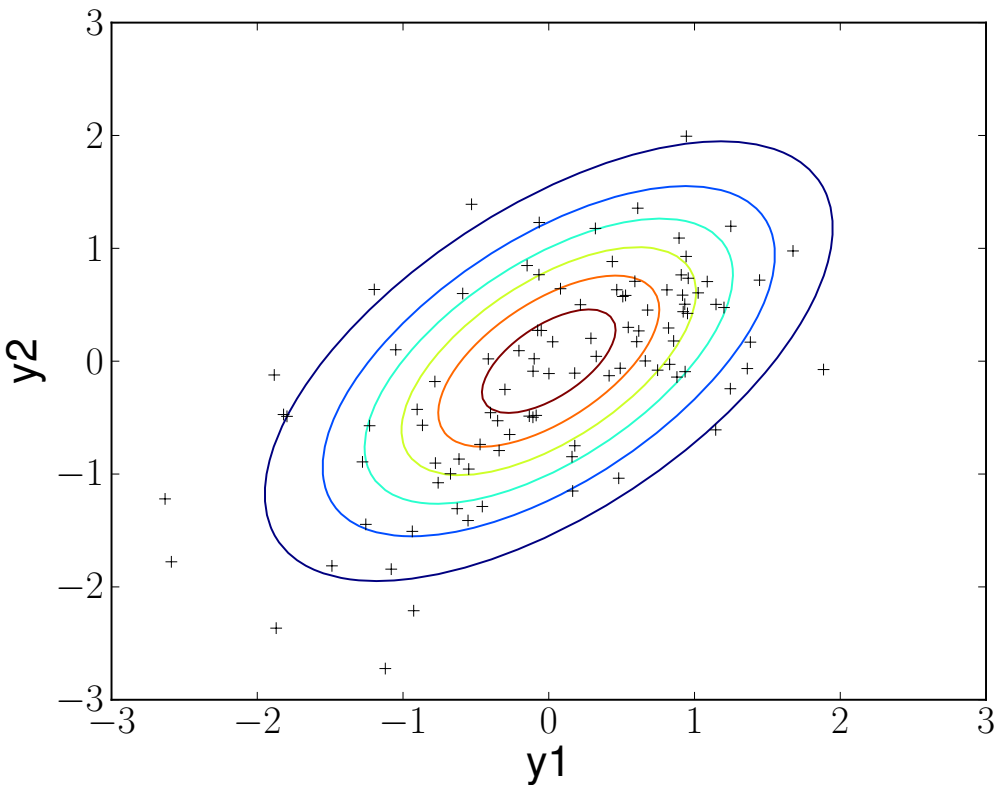
- Probability contour
- Samples



$$\mathbf{K} = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}$$

A 2D Gaussian

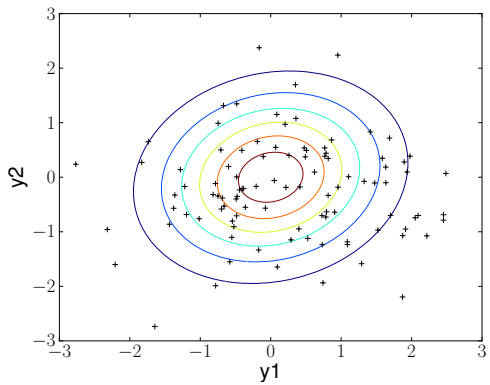
- Probability contour
- Samples



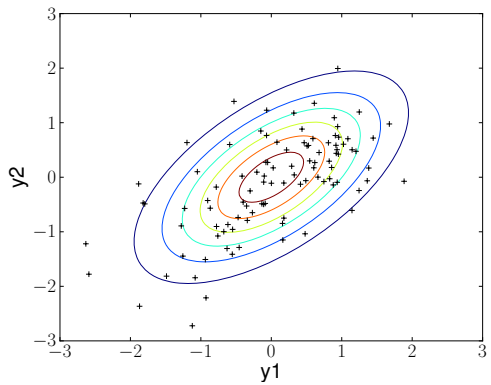
$$\mathbf{K} = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}$$

A 2D Gaussian

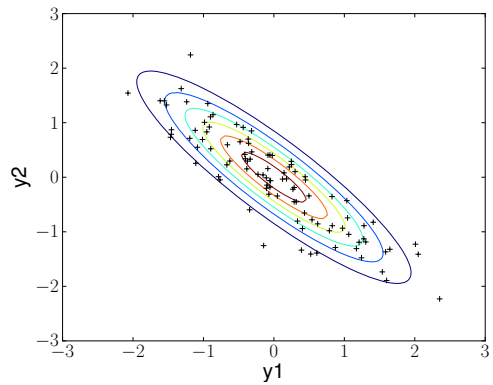
Varying the covariance matrix



$$\mathbf{K} = \begin{bmatrix} 1 & 0.14 \\ 0.14 & 1 \end{bmatrix}$$



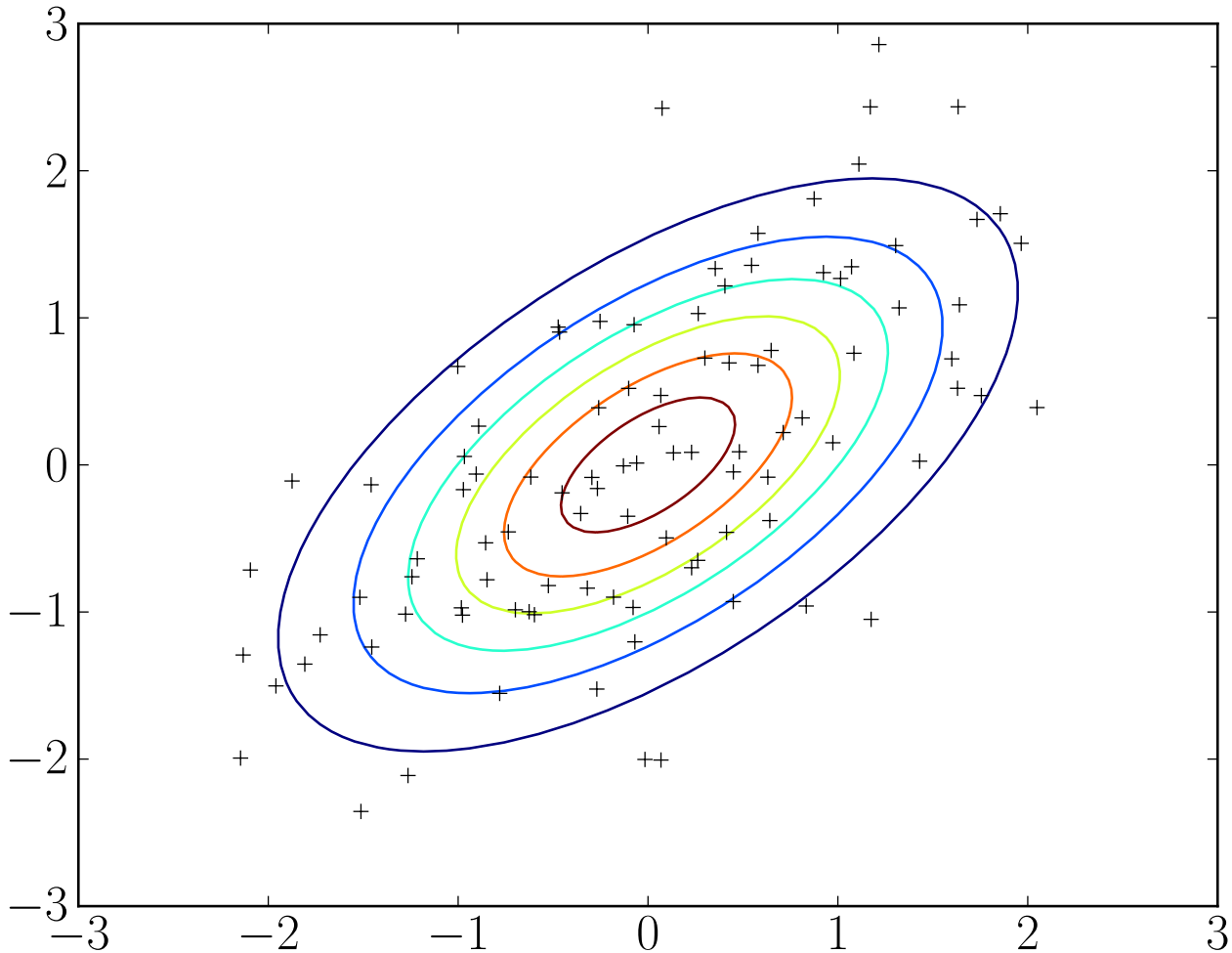
$$\mathbf{K} = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}$$



$$\mathbf{K} = \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}$$

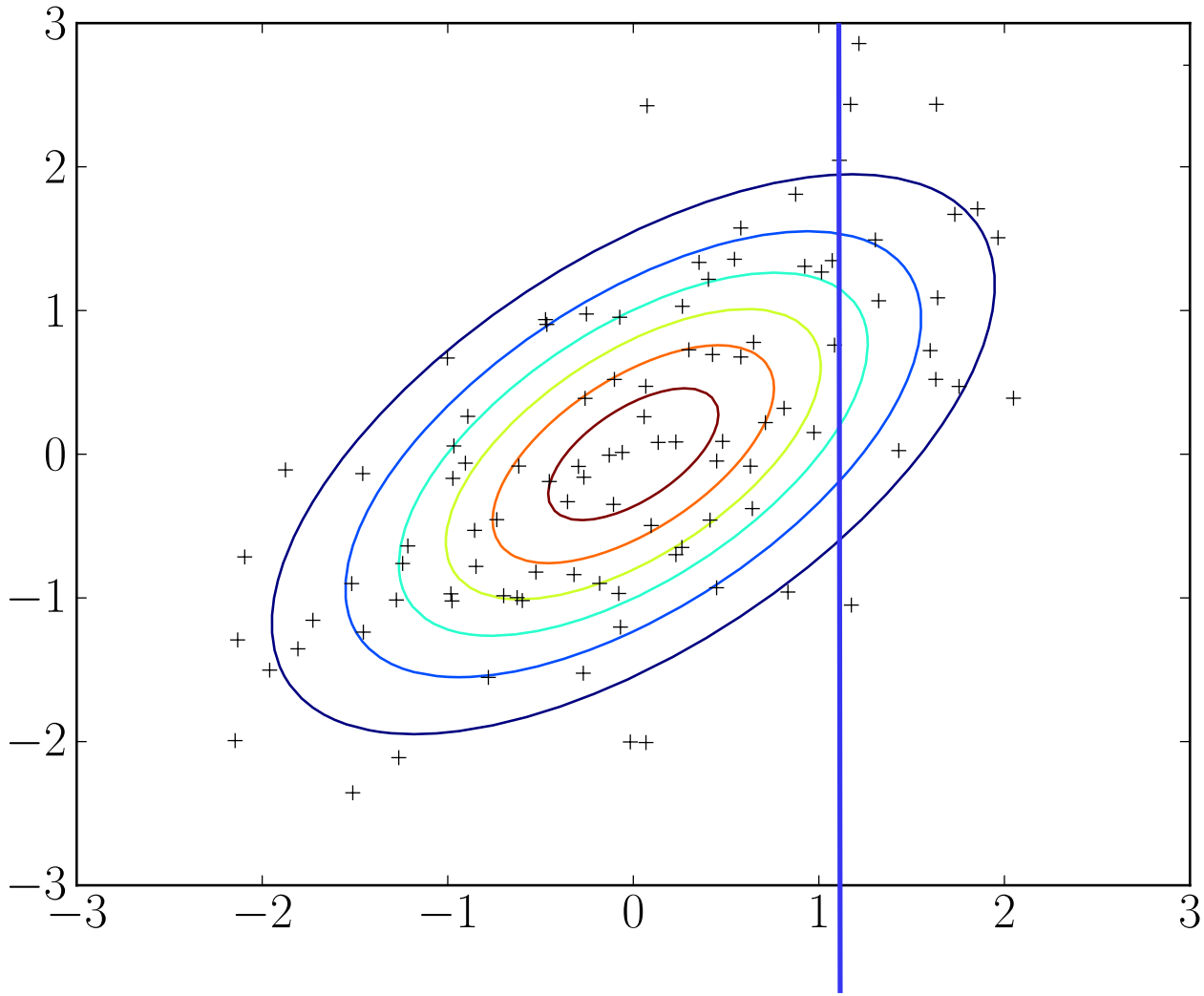
A 2D Gaussian

Inference



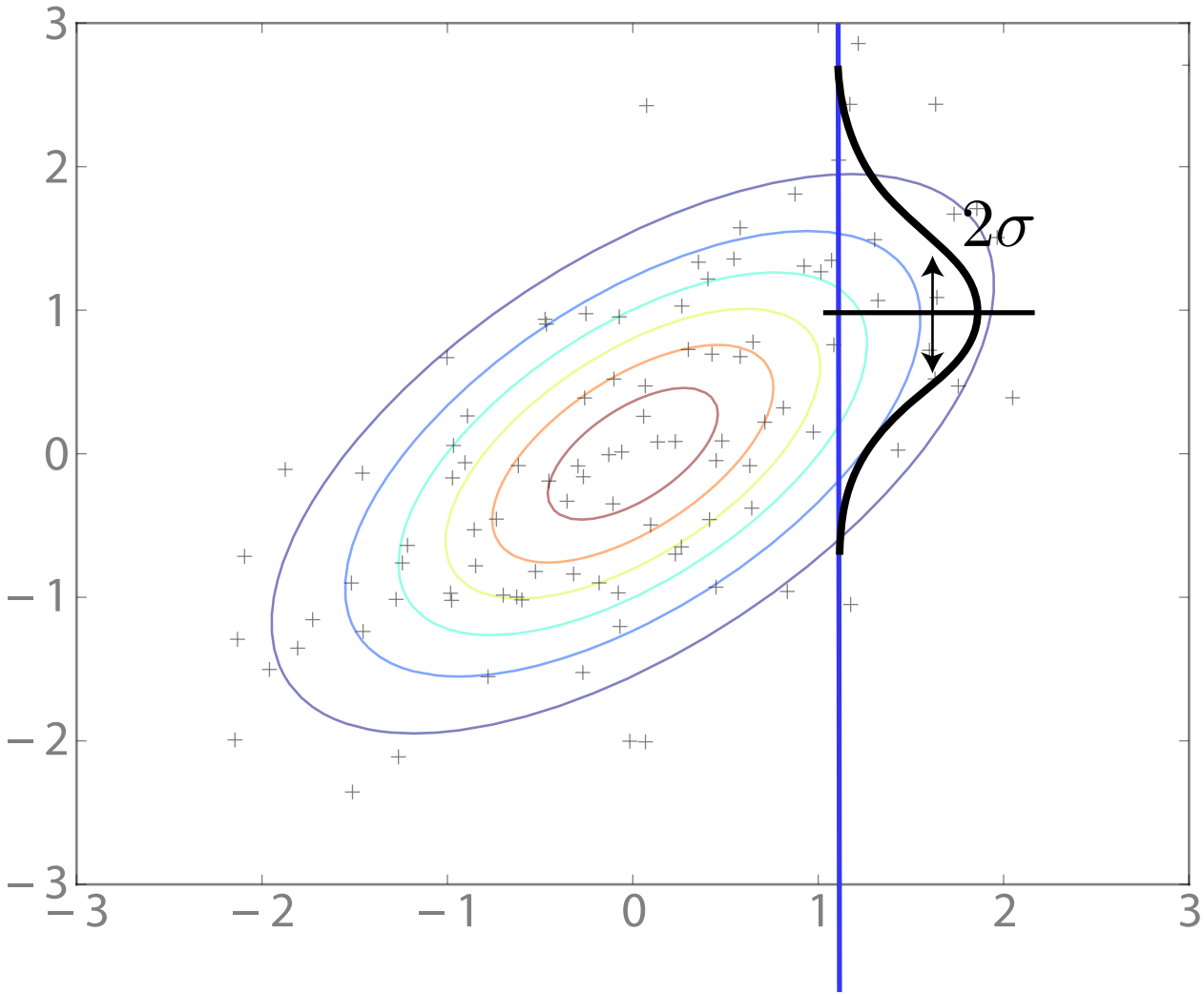
A 2D Gaussian

Inference



A 2D Gaussian

Inference



Phenotype prediction

Best linear unbiased prediction

- Given the phenotype values $\mathbf{y}$ of a set of individuals and the genetic relatedness, we can predict the genetic component of the phenotype of a new individual $\mathbf{y}^\star$.

- Use conditional probability distribution
- Note, that the result is again a Gaussian distribution!
- As a Gaussian distribution is always normalized, we can drop any constant terms, that do not contain $\mathbf{y}^\star$.
- Completing the square identifies the mean $\mu^\star$ and the variance $\Sigma^\star$ of $\mathbf{y}^\star$ given $\mathbf{y}$.

Phenotype prediction

Best linear unbiased prediction

- Given the phenotype values $\mathbf{y}$ of a set of individuals and the genetic relatedness, we can predict the genetic component of the phenotype of a new individual $\mathbf{y}^\star$.

$$P(\mathbf{y}^\star \,|\, \mathbf{y}) = ?$$

- Use conditional probability distribution
- Note, that the result is again a Gaussian distribution!
- As a Gaussian distribution is always normalized, we can drop any constant terms, that do not contain $\mathbf{y}^\star$.
- Completing the square identifies the mean $\mu^\star$ and the variance $\Sigma^\star$ of $\mathbf{y}^\star$ given $\mathbf{y}$.

Phenotype prediction

Best linear unbiased prediction

- Given the phenotype values $\mathbf{y}$ of a set of individuals and the genetic relatedness, we can predict the genetic component of the phenotype of a new individual $\mathbf{y}^\star$.

$$P(\mathbf{y}^\star \,|\, \mathbf{y}) = ?$$

- Use conditional probability distribution
- Note, that the result is again a Gaussian distribution!
- As a Gaussian distribution is always normalized, we can drop any constant terms, that do not contain $\mathbf{y}^\star$.
- Completing the square identifies the mean $\mu^\star$ and the variance $\Sigma^\star$ of $\mathbf{y}^\star$ given $\mathbf{y}$.

Phenotype prediction

Best linear unbiased prediction

- Given the phenotype values $\mathbf{y}$ of a set of individuals and the genetic relatedness, we can predict the genetic component of the phenotype of a new individual $\mathbf{y}^\star$.

$$P(\mathbf{y}^\star \,|\, \mathbf{y}) = \frac{P(\mathbf{y}, \mathbf{y}^\star)}{P(\mathbf{y})}$$

- Use conditional probability distribution
- Note, that the result is again a Gaussian distribution!
- As a Gaussian distribution is always normalized, we can drop any constant terms, that do not contain $\mathbf{y}^\star$.
- Completing the square identifies the mean $\mu^\star$ and the variance $\Sigma^\star$ of $\mathbf{y}^\star$ given $\mathbf{y}$.

Phenotype prediction

Best linear unbiased prediction

- Given the phenotype values $\mathbf{y}$ of a set of individuals and the genetic relatedness, we can predict the genetic component of the phenotype of a new individual $\mathbf{y}^\star$.

$$P(\mathbf{y}^\star \,|\, \mathbf{y}) = \frac{P(\mathbf{y}, \mathbf{y}^\star)}{P(\mathbf{y})} = \frac{\mathcal{N}\left(\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^\star \end{bmatrix} | \mathbf{0}, \sigma_g^2 \begin{bmatrix} \mathbf{K} & \mathbf{K}_{:,\star} \\ \mathbf{K}_{:,\star}^{\mathrm{T}} & \mathbf{K}^{\star,\star} \end{bmatrix} \sigma_e^2 \mathbf{I}\right)}{\mathcal{N}\left(\mathbf{0}, \sigma_g^2 \mathbf{K} + \sigma_e^2 \mathbf{I}\right)}$$

- Use conditional probability distribution
- Note, that the result is again a Gaussian distribution!
- As a Gaussian distribution is always normalized, we can drop any constant terms, that do not contain $\mathbf{y}^\star$.
- Completing the square identifies the mean $\mu^\star$ and the variance $\Sigma^\star$ of $\mathbf{y}^\star$ given $\mathbf{y}$.

Phenotype prediction

Best linear unbiased prediction

- Given the phenotype values $\mathbf{y}$ of a set of individuals and the genetic relatedness, we can predict the genetic component of the phenotype of a new individual $\mathbf{y}^\star$.

$$P(\mathbf{y}^\star \,|\, \mathbf{y}) = \frac{P(\mathbf{y}, \mathbf{y}^\star)}{P(\mathbf{y})} = \frac{\mathcal{N}\left(\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^\star \end{bmatrix} | \mathbf{0}, \sigma_g^2 \begin{bmatrix} \mathbf{K} & \mathbf{K}_{:,\star} \\ \mathbf{K}_{:,\star}^{\mathrm{T}} & \mathbf{K}^{\star,\star} \end{bmatrix} \sigma_e^2 \mathbf{I}\right)}{\mathcal{N}\left(\mathbf{0}, \sigma_g^2 \mathbf{K} + \sigma_e^2 \mathbf{I}\right)}$$

$$= \mathcal{N}\left(\mathbf{y}^\star \,|\, \boldsymbol{\mu}^\star, \boldsymbol{\Sigma}^\star\right)$$

- Use conditional probability distribution
- Note, that the result is again a Gaussian distribution!
- As a Gaussian distribution is always normalized, we can drop any constant terms, that do not contain $\mathbf{y}^\star$.
- Completing the square identifies the mean $\mu^\star$ and the variance $\Sigma^\star$ of $\mathbf{y}^\star$ given $\mathbf{y}$.

Phenotype prediction

Best linear unbiased prediction

- Given the phenotype values $\mathbf{y}$ of a set of individuals and the genetic relatedness, we can predict the genetic component of the phenotype of a new individual $\mathbf{y}^\star$.

$$P(\mathbf{y}^\star \,|\, \mathbf{y}) = \frac{P(\mathbf{y}, \mathbf{y}^\star)}{P(\mathbf{y})} = \frac{\mathcal{N}\left(\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^\star \end{bmatrix} | \mathbf{0}, \sigma_g^2 \begin{bmatrix} \mathbf{K} & \mathbf{K}_{:,\star} \\ \mathbf{K}_{:,\star}^{\mathrm{T}} & \mathbf{K}^{\star,\star} \end{bmatrix} \sigma_e^2 \mathbf{I}\right)}{\mathcal{N}\left(\mathbf{0}, \sigma_g^2 \mathbf{K} + \sigma_e^2 \mathbf{I}\right)}$$

$$= \mathcal{N}\left(\mathbf{y}^\star \,|\, \boldsymbol{\mu}^\star, \boldsymbol{\Sigma}^\star\right) \propto \mathcal{N}\left(\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^\star \end{bmatrix} | \mathbf{0}, \sigma_g^2 \begin{bmatrix} \mathbf{K} & \mathbf{K}_{:,\star} \\ \mathbf{K}_{:,\star}^{\mathrm{T}} & \mathbf{K}^{\star,\star} \end{bmatrix} \sigma_e^2 \mathbf{I}\right)$$

- Use conditional probability distribution
- Note, that the result is again a Gaussian distribution!
- As a Gaussian distribution is always normalized, we can drop any constant terms, that do not contain $\mathbf{y}^\star$.
- Completing the square identifies the mean $\mu^\star$ and the variance $\Sigma^\star$ of $\mathbf{y}^\star$ given $\mathbf{y}$.

Phenotype prediction

Best linear unbiased prediction

- Given the phenotype values $\mathbf{y}$ of a set of individuals and the genetic relatedness, we can predict the genetic component of the phenotype of a new individual $\mathbf{y}^\star$.

$$P(\mathbf{y}^\star \,|\, \mathbf{y}) = \frac{P(\mathbf{y}, \mathbf{y}^\star)}{P(\mathbf{y})} = \frac{\mathcal{N}\left(\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^\star \end{bmatrix} | \mathbf{0}, \sigma_g^2 \begin{bmatrix} \mathbf{K} & \mathbf{K}_{:,\star} \\ \mathbf{K}_{:,\star}^{\mathrm{T}} & \mathbf{K}^{\star,\star} \end{bmatrix} \sigma_e^2 \mathbf{I}\right)}{\mathcal{N}\left(\mathbf{0}, \sigma_g^2 \mathbf{K} + \sigma_e^2 \mathbf{I}\right)}$$

$$= \mathcal{N}\left(\mathbf{y}^\star \,|\, \boldsymbol{\mu}^\star, \boldsymbol{\Sigma}^\star\right) \propto \mathcal{N}\left(\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^\star \end{bmatrix} | \mathbf{0}, \sigma_g^2 \begin{bmatrix} \mathbf{K} & \mathbf{K}_{:,\star} \\ \mathbf{K}_{:,\star}^{\mathrm{T}} & \mathbf{K}^{\star,\star} \end{bmatrix} \sigma_e^2 \mathbf{I}\right)$$

- Use conditional probability distribution
- Note, that the result is again a Gaussian distribution!
- As a Gaussian distribution is always normalized, we can drop any constant terms, that do not contain $\mathbf{y}^\star$.
- Completing the square identifies the mean $\boldsymbol{\mu}^\star$ and the variance $\boldsymbol{\Sigma}^\star$ of $\mathbf{y}^\star$ given $\mathbf{y}$.

Inference

Gaussian conditioning in 2D

$$p(y_2 \mid y_1, \mathbf{K}) = \frac{p(y_1, y_2 \mid \mathbf{K})}{p(y_1 \mid \mathbf{K})} \propto \exp\left\{-\frac{1}{2}[y_1, y_2]\,\mathbf{K}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right\}$$

$$= \exp\{-\frac{1}{2}\left[y_1^2\mathbf{K}_{1,1}^{-1} + y_2^2\mathbf{K}_{2,2}^{-1} + 2y_1\mathbf{K}_{1,2}^{-1}y_2\right]\}$$

$$= \exp\{-\frac{1}{2}\left[y_2^2\mathbf{K}_{2,2}^{-1} + 2y_2\mathbf{K}_{1,2}^{-1}y_1 + C\right]\}$$

$$= Z\exp\{-\frac{1}{2}\mathbf{K}_{2,2}^{-1}\left[y_2^2 + 2y_2\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}\right]\}$$

$$= Z\exp\{-\frac{1}{2}\mathbf{K}_{2,2}^{-1}\left[y_2^2 + 2y_2\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}} + \frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}^2\right] + \frac{1}{2}\mathbf{K}_{2,2}^{-1}\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}^2\}$$

$$= Z'\exp\{-\frac{1}{2}\underbrace{\mathbf{K}_{2,2}^{-1}}_{\Sigma}\left[y_2 + \underbrace{\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}}_{-\mu}\right]^2\} \propto \mathcal{N}(y_2 \mid \mu, \Sigma)$$

Inference

Gaussian conditioning in 2D

$$p(y_2 \mid y_1, \mathbf{K}) = \frac{p(y_1, y_2 \mid \mathbf{K})}{p(y_1 \mid \mathbf{K})} \propto \exp\left\{-\frac{1}{2}[y_1, y_2]\,\mathbf{K}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right\}$$

$$= \exp\{-\frac{1}{2}\left[y_1^2\mathbf{K}_{1,1}^{-1} + y_2^2\mathbf{K}_{2,2}^{-1} + 2y_1\mathbf{K}_{1,2}^{-1}y_2\right]\}$$

$$= \exp\{-\frac{1}{2}\left[y_2^2\mathbf{K}_{2,2}^{-1} + 2y_2\mathbf{K}_{1,2}^{-1}y_1 + C\right]\}$$

$$= Z\exp\{-\frac{1}{2}\mathbf{K}_{2,2}^{-1}\left[y_2^2 + 2y_2\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}\right]\}$$

$$= Z\exp\{-\frac{1}{2}\mathbf{K}_{2,2}^{-1}\left[y_2^2 + 2y_2\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}} + \frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}^2\right] + \frac{1}{2}\mathbf{K}_{2,2}^{-1}\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}^2\}$$

$$= Z'\exp\{-\frac{1}{2}\underbrace{\mathbf{K}_{2,2}^{-1}}_{\Sigma}\left[y_2 + \underbrace{\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}}_{-\mu}\right]^2\} \propto \mathcal{N}(y_2 \mid \mu, \Sigma)$$

Inference

Gaussian conditioning in 2D

$$p(y_2 \mid y_1, \mathbf{K}) = \frac{p(y_1, y_2 \mid \mathbf{K})}{p(y_1 \mid \mathbf{K})} \propto \exp\left\{-\frac{1}{2}[y_1, y_2]\,\mathbf{K}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right\}$$

$$= \exp\{-\frac{1}{2}\left[y_1^2\mathbf{K}_{1,1}^{-1} + y_2^2\mathbf{K}_{2,2}^{-1} + 2y_1\mathbf{K}_{1,2}^{-1}y_2\right]\}$$

$$= \exp\{-\frac{1}{2}\left[y_2^2\mathbf{K}_{2,2}^{-1} + 2y_2\mathbf{K}_{1,2}^{-1}y_1 + C\right]\}$$

$$= Z\exp\{-\frac{1}{2}\mathbf{K}_{2,2}^{-1}\left[y_2^2 + 2y_2\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}\right]\}$$

$$= Z\exp\{-\frac{1}{2}\mathbf{K}_{2,2}^{-1}\left[y_2^2 + 2y_2\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}} + \frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}^2\right] + \frac{1}{2}\mathbf{K}_{2,2}^{-1}\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}^2\}$$

$$= Z'\exp\{-\frac{1}{2}\underbrace{\mathbf{K}_{2,2}^{-1}}_{\Sigma}\left[y_2 + \underbrace{\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}}_{-\mu}\right]^2\} \propto \mathcal{N}(y_2 \mid \mu, \Sigma)$$

Inference

Gaussian conditioning in 2D

$$
\begin{aligned}
p(y_2 \,|\, y_1, \mathbf{K}) = \frac{p(y_1, y_2 \,|\, \mathbf{K})}{p(y_1 \,|\, \mathbf{K})} &\propto \exp\left\{-\frac{1}{2}[y_1, y_2]\,\mathbf{K}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right\} \\
&= \exp\{-\frac{1}{2}\left[y_1^2\mathbf{K}_{1,1}^{-1} + y_2^2\mathbf{K}_{2,2}^{-1} + 2y_1\mathbf{K}_{1,2}^{-1}y_2\right]\} \\
&= \exp\{-\frac{1}{2}\left[y_2^2\mathbf{K}_{2,2}^{-1} + 2y_2\mathbf{K}_{1,2}^{-1}y_1 + C\right]\} \\
&= Z\exp\{-\frac{1}{2}\mathbf{K}_{2,2}^{-1}\left[y_2^2 + 2y_2\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}\right]\} \\
&= Z\exp\{-\frac{1}{2}\mathbf{K}_{2,2}^{-1}\left[y_2^2 + 2y_2\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}} + \frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}^2\right] + \frac{1}{2}\mathbf{K}_{2,2}^{-1}\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}^2\} \\
&= Z'\exp\{-\frac{1}{2}\underbrace{\mathbf{K}_{2,2}^{-1}}_{\Sigma}\left[y_2 + \underbrace{\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}}_{-\mu}\right]^2\} \propto \mathcal{N}\left(y_2 \,|\, \mu, \Sigma\right)
\end{aligned}
$$

Inference

Gaussian conditioning in 2D

$$
\begin{aligned}
p(y_2 \,|\, y_1, \mathbf{K}) = \frac{p(y_1, y_2 \,|\, \mathbf{K})}{p(y_1 \,|\, \mathbf{K})} &\propto \exp\left\{-\frac{1}{2}[y_1, y_2]\,\mathbf{K}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right\} \\
&= \exp\{-\frac{1}{2}\left[y_1^2\mathbf{K}_{1,1}^{-1} + y_2^2\mathbf{K}_{2,2}^{-1} + 2y_1\mathbf{K}_{1,2}^{-1}y_2\right]\} \\
&= \exp\{-\frac{1}{2}\left[y_2^2\mathbf{K}_{2,2}^{-1} + 2y_2\mathbf{K}_{1,2}^{-1}y_1 + C\right]\} \\
&= Z\exp\{-\frac{1}{2}\mathbf{K}_{2,2}^{-1}\left[y_2^2 + 2y_2\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}\right]\} \\
&= Z\exp\{-\frac{1}{2}\mathbf{K}_{2,2}^{-1}\left[y_2^2 + 2y_2\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}} + \frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}^2\right] + \frac{1}{2}\mathbf{K}_{2,2}^{-1}\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}^2\} \\
&= Z'\exp\{-\frac{1}{2}\underbrace{\mathbf{K}_{2,2}^{-1}}_{\Sigma}\left[y_2 + \underbrace{\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}}_{-\mu}\right]^2\} \propto \mathcal{N}\left(y_2 \,|\, \mu, \Sigma\right)
\end{aligned}
$$

Inference

Gaussian conditioning in 2D

$$
\begin{aligned}
p(y_2 \,|\, y_1, \mathbf{K}) = \frac{p(y_1, y_2 \,|\, \mathbf{K})}{p(y_1 \,|\, \mathbf{K})} &\propto \exp\left\{-\frac{1}{2}[y_1, y_2]\,\mathbf{K}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right\} \\
&= \exp\{-\frac{1}{2}\left[y_1^2\mathbf{K}_{1,1}^{-1} + y_2^2\mathbf{K}_{2,2}^{-1} + 2y_1\mathbf{K}_{1,2}^{-1}y_2\right]\} \\
&= \exp\{-\frac{1}{2}\left[y_2^2\mathbf{K}_{2,2}^{-1} + 2y_2\mathbf{K}_{1,2}^{-1}y_1 + C\right]\} \\
&= Z\exp\{-\frac{1}{2}\mathbf{K}_{2,2}^{-1}\left[y_2^2 + 2y_2\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}\right]\} \\
&= Z\exp\{-\frac{1}{2}\mathbf{K}_{2,2}^{-1}\left[y_2^2 + 2y_2\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}} + \frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}^2\right] + \frac{1}{2}\mathbf{K}_{2,2}^{-1}\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}^2\} \\
&= Z'\exp\{-\frac{1}{2}\underbrace{\mathbf{K}_{2,2}^{-1}}_{\Sigma}\left[y_2 + \underbrace{\frac{\mathbf{K}_{1,2}^{-1}y_1}{\mathbf{K}_{2,2}^{-1}}}_{-\mu}\right]^2\} \propto \mathcal{N}\left(y_2 \,|\, \mu, \Sigma\right)
\end{aligned}
$$

Phenotype prediction

Best linear unbiased prediction

- Given the phenotype values of a set of individuals and the genetic relatedness, we can predict the genetic component of the phenotype of a new individual.
- $P(\mathbf{y}^\star \mid \mathbf{y}) = \mathcal{N}\left(\mathbf{y}^\star \mid \boldsymbol{\mu}^\star, \sigma_{\mathrm{g}}^2 \mathbf{V}_g^\star + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)$
- Predictive mean: $\boldsymbol{\mu}^\star = \underbrace{\sigma_{\mathrm{g}}^2 \mathbf{K}_g^{\star,:} \left(\sigma_{\mathrm{g}}^2 \mathbf{K}_g + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)^{-1}}_{\text{BLUP}} \mathbf{y}$
- Predictive Variance: $\mathbf{V}_g^\star = \mathbf{K}_g^{\star,\star} - \sigma_{\mathrm{g}}^2 \mathbf{K}_g^{\star,:} \left(\sigma_{\mathrm{g}}^2 \mathbf{K}_g + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)^{-1} \mathbf{K}_g^{:,\star}$

Phenotype prediction

Best linear unbiased prediction

- Given the phenotype values of a set of individuals and the genetic relatedness, we can predict the genetic component of the phenotype of a new individual.
- $P(\mathbf{y}^\star \,|\, \mathbf{y}) = \mathcal{N}\left(\mathbf{y}^\star \,\middle|\, \boldsymbol{\mu}^\star, \sigma_{\mathrm{g}}^2 \mathbf{V}_g^\star + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)$
- Predictive mean: $\boldsymbol{\mu}^\star = \underbrace{\sigma_{\mathrm{g}}^2 \mathbf{K}_g^{\star,:} \left(\sigma_{\mathrm{g}}^2 \mathbf{K}_g + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)^{-1}}_{\text{BLUP}} \mathbf{y}$
- Predictive Variance: $\mathbf{V}_g^\star = \mathbf{K}_g^{\star,\star} - \sigma_{\mathrm{g}}^2 \mathbf{K}_g^{\star,:} \left(\sigma_{\mathrm{g}}^2 \mathbf{K}_g + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)^{-1} \mathbf{K}_g^{:,\star}$

Phenotype prediction

Best linear unbiased prediction

- Given the phenotype values of a set of individuals and the genetic relatedness, we can predict the genetic component of the phenotype of a new individual.
- $P(\mathbf{y}^\star \,|\, \mathbf{y}) = \mathcal{N}\left(\mathbf{y}^\star \,\middle|\, \boldsymbol{\mu}^\star, \sigma_{\mathrm{g}}^2 \mathbf{V}_g^\star + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)$
- Predictive mean: $\boldsymbol{\mu}^\star = \underbrace{\sigma_{\mathrm{g}}^2 \mathbf{K}_g^{\star,:} \left(\sigma_{\mathrm{g}}^2 \mathbf{K}_g + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)^{-1} \mathbf{y}}_{\text{BLUP}}$
- Predictive Variance: $\mathbf{V}_g^\star = \mathbf{K}_g^{\star,\star} - \sigma_{\mathrm{g}}^2 \mathbf{K}_g^{\star,:} \left(\sigma_{\mathrm{g}}^2 \mathbf{K}_g + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)^{-1} \mathbf{K}_g^{:,\star}$

Phenotype prediction

Best linear unbiased prediction

- Given the phenotype values of a set of individuals and the genetic relatedness, we can predict the genetic component of the phenotype of a new individual.
- $P(\mathbf{y}^\star \,|\, \mathbf{y}) = \mathcal{N}\left(\mathbf{y}^\star \,\middle|\, \boldsymbol{\mu}^\star, \sigma_{\mathrm{g}}^2 \mathbf{V}_g^\star + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)$
- Predictive mean: $\boldsymbol{\mu}^\star = \underbrace{\sigma_{\mathrm{g}}^2 \mathbf{K}_g^{\star,:} \left(\sigma_{\mathrm{g}}^2 \mathbf{K}_g + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)^{-1} \mathbf{y}}_{\text{BLUP}}$
- Predictive Variance: $\mathbf{V}_g^\star = \mathbf{K}_g^{\star,\star} - \sigma_{\mathrm{g}}^2 \mathbf{K}_g^{\star,:} \left(\sigma_{\mathrm{g}}^2 \mathbf{K}_g + \sigma_{\mathrm{e}}^2 \mathbf{I}\right)^{-1} \mathbf{K}_g^{:,\star}$

Summary

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Phenotype prediction

Summary

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Phenotype prediction

Summary

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Phenotype prediction

Summary

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Phenotype prediction

Summary

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Phenotype prediction

Summary

- Basic probability theory
- Linear mixed models
  - Population structure correction
  - Parameter estimation
  - Variance component modeling
  - Phenotype prediction

Acknowledgements

- **Joint course material**
  O. Stegle
- **FaST lmm**
  J. Listgarten, D. Heckerman

References I

P. Burton, D. Clayton, L. Cardon, N. Craddock, P. Deloukas, A. Duncanson, D. Kwiatkowski, M. McCarthy, W. Ouwehand, N. Samani, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145):661–678, 2007.

B. Devlin and K. Roeder. Genomic control for association studies. *Biometrics*, 55(4):997–1004, 1999.

H. M. Kang, N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, M. J. Daly, and E. Eskin. Efficient control of population structure in model organism association mapping. *Genetics*, 107, 2008.

C. Lippert, J. Listgarten, Y. Liu, C. Kadie, R. Davidson, and D. Heckerman. Fast linear mixed models for genome-wide association studies. *Nature Methods*, 8(10):838;835, 10 2011. doi: 10.1038/nmeth.1681.

J. Listgarten, C. Lippert, and D. Heckerman. An efficient group test for genetic markers that handles confounding. *arXiv preprint arXiv:1205.0793*, 2012.

J. Novembre, T. Johnson, K. Bryc, Z. Kutalik, A. R. Boyko, A. Auton, A. Indap, K. S. King, S. Bergmann, M. R. Nelson, M. Stephens, and C. D. Bustamante. Genes mirror geography within europe. *Nature*, 456(7218):98–101, Nov. 2008.

N. Patterson, A. L. Price, and D. Reich. Population structure and eigenanalysis. *PLoS Genetics*, 2(12):e190+, December 2006.

References II

A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, 38(8):904–909, August 2006.

M. Wu, S. Lee, T. Cai, Y. Li, M. Boehnke, and X. Lin. Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics*, 2011.