Lecture 5
Estimation of Variance Components

Guilherme J. M. Rosa
University of Wisconsin-Madison

Mixed Models in Quantitative Genetics
SISG, Seattle
18 – 20 September 2018

Estimation of Variance Components
ANOVA Estimation

Consider the data set below, related to observations of half-sib families of k unrelated sires. The following model can be used to represent these data:

$$y_{ij} = \mu + s_i + e_{ij}$$

Sire			
1	2	…	k
y_{11}	y_{21}	…	y_{k1}
y_{12}	y_{22}	…	y_{k2}
⋮	⋮		⋮
y_{1n_1}	y_{2n_2}	…	y_{kn_k}

where y_{ij} represents the phenotypic trait observation of progeny j (j = 1, 2,…, n_i) in family i, μ is a mean, s_i is an effect common to all animals having sire i, and e_{ij} is a residual term

Estimation of Variance Components
ANOVA Estimation

The sire effect s_i is equivalent to the transmitting ability (which is equal to one-half additive genetic value) of sire i, as one-half of its genes are (randomly) transmitted to each of its n_i progeny.

The residual terms e_{ij} refer to additional genetics effects (such as the effect of dams) and environmental components.

It is assumed that $s_i \overset{ind}{\sim} (0, \sigma_s^2)$ and $e_{ij} \overset{ind}{\sim} (0, \sigma_e^2)$

From the model settings discussed before we have that

$$E[y_{ij}] = \mu \quad \text{and} \quad Var[y_{ij}] = \sigma_s^2 + \sigma_e^2$$

The overall sample mean is given by $\bar{y}_{\bullet\bullet} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} = \frac{1}{N}\sum_{i=1}^{k} y_{i\bullet}$

where $N = \sum_{i=1}^{k} n_i$, and $\bar{y}_{i\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$ are sire-specific means.

The ANOVA approach consists of an orthogonal decomposition of the total sum of squares (TSS) into between classes (or, in our case, sires) and within classes (or residual) components. The corrected (in terms of the general mean) TSS is given by: $TSS = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{\bullet\bullet})^2$

By adding and subtracting $\bar{y}_{i\bullet}$ within the parentheses, the TSS can be expressed as:

$$TSS = \sum_{i=1}^{k}\sum_{j=1}^{n_i}[(y_{ij} - \bar{y}_{i\bullet}) + (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})]^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\bullet})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 + 2\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\bullet})(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})$$

It is seen that the last part of this expression is equal to zero, so that TSS can be written as two components:

$$SSS = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 \quad \text{and} \quad RSS = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\bullet})^2$$

which are the sire and the residual sum of squares, respectively. The SSS term measures the variation of each progeny family around the overall mean, while the RSS term measures the extra variation related to each observation around its sire average

It can be shown that the expectation of these sums of squares terms are:

$$E[SSS] = \left(N - \frac{1}{N}\sum_{i=1}^{n_i} n_i^2\right)\sigma_s^2 + (k-1)\sigma_e^2 \quad \text{and} \quad E[RSS] = (N-k)\sigma_e^2$$

so that the ANOVA estimators of the sire and residual variance components are given by:

$$\hat{\sigma}_s^2 = \left(N - \frac{1}{N}\sum_{i=1}^{n_i} n_i^2\right)^{-1}[SSS - (k-1)\hat{\sigma}_e^2] \quad \text{and} \quad \hat{\sigma}_e^2 = \frac{1}{(N-k)}RSS$$

In the specific case of balanced data, i.e. the same progeny size for all sires, $n_i = n = N/k$ and the ANOVA estimators become:

$$\hat{\sigma}_s^2 = \frac{1}{n}\left[\frac{1}{(k-1)}SSS - \hat{\sigma}_e^2\right] \quad \text{and} \quad \hat{\sigma}_e^2 = \frac{1}{k(n-1)}RSS$$

Appendix: Calculating E(MS)

Model:

$$y_{ij} = \mu + s_i + e_{ij} \quad \text{with} \quad \begin{cases} \mu \text{ fixed} \rightarrow E[\mu] = \mu,\ E[\mu^2] = \mu^2,\ Var[\mu] = 0 \\ s_i \overset{iid}{\sim} N(0, \sigma_s^2) \rightarrow E[s_i] = 0,\ E[s_i^2] = Var[s_i] = \sigma_s^2 \\ e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2) \rightarrow E[e_{ij}] = 0,\ E[e_{ij}^2] = Var[e_{ij}] = \sigma_e^2 \\ Cov[s_i, s_{i'}] = Cov[s_i, e_{ij}] = Cov[e_{ij}, e_{i'j'}] = 0 \end{cases}$$

Sum of Squares:

$$SSS = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 = \frac{1}{n}\sum_{i=1}^{k} y_{i\bullet}^2 - \frac{1}{kn} y_{\bullet\bullet}^2$$

$$RSS = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\bullet})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \frac{1}{n}\sum_{i=1}^{k} y_{i\bullet}^2$$

Key Expectations: $E\left[\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2\right]$, $E\left[\frac{1}{kn} y_{\bullet\bullet}^2\right]$, and $E\left[\frac{1}{n}\sum_{i=1}^{k} y_{i\bullet}^2\right]$

$$E\left[\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2\right] = \sum_{i=1}^{k}\sum_{j=1}^{n} E\left[y_{ij}\right]^2 = \sum_{i=1}^{k}\sum_{j=1}^{n} E\left[\mu + s_i + e_{ij}\right]^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n} E\left[\mu^2 + s_i^2 + e_{ij}^2 + 2\mu s_i + 2\mu e_{ij} + 2 s_i e_{ij}\right]$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n}\left(\mu^2 + E[s_i^2] + E[e_{ij}^2] + 2\mu E[s_i] + 2\mu E[e_{ij}] + 2E[s_i]E[e_{ij}]\right)$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n}\left(\mu^2 + \sigma_s^2 + \sigma_e^2\right)$$

$$= kn\mu^2 + kn\sigma_s^2 + kn\sigma_e^2$$

$$E\left[\frac{1}{kn}y_{\bullet\bullet}^2\right] = \frac{1}{kn}E\left[\left(\sum_{i=1}^{k}\sum_{j=1}^{n}y_{ij}\right)^2\right] = \frac{1}{kn}E\left[\left(\sum_{i=1}^{k}\sum_{j=1}^{n}(\mu + s_i + e_{ij})\right)^2\right]$$

$$= \frac{1}{kn}E\left[\left(kn\mu + n\sum_{i=1}^{k}s_i + \sum_{i=1}^{k}\sum_{j=1}^{n}e_{ij}\right)^2\right]$$

$$= \frac{1}{kn}E\left[k^2n^2\mu^2 + n^2\left(\sum_{i=1}^{k}s_i\right)^2 + \left(\sum_{i=1}^{k}\sum_{j=1}^{n}e_{ij}\right)^2 + DPs\right]$$

$$= \frac{1}{kn}(k^2n^2\mu^2 + kn^2\sigma_s^2 + kn\sigma_e^2 + 0)$$

$$= kn\mu^2 + n\sigma_s^2 + \sigma_e^2$$

$$E\left[\frac{1}{n}\sum_{i=1}^{k}y_{i\bullet}^2\right] = \frac{1}{n}\sum_{i=1}^{k}E[y_{i\bullet}^2] = \frac{1}{n}\sum_{i=1}^{k}E\left[\left(\sum_{j=1}^{n}y_{ij}\right)^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{k}E\left[\left(n\mu + ns_i + \sum_{j=1}^{n}y_{ij}\right)^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{k}E\left[\left(n^2\mu^2 + n^2s_i^2 + \left(\sum_{j=1}^{n}y_{ij}\right)^2 + DPs\right)\right]$$

$$= \frac{1}{n}\sum_{i=1}^{k}(n^2m^2 + n^2\sigma_s^2 + n\sigma_e^2 + 0)$$

$$= kn\mu^2 + kn\sigma_s^2 + k\sigma_e^2$$

Expected MS

$$E[SMS] = \frac{1}{k-1}E[SSS] = \frac{1}{k-1}E\left[\frac{1}{n}\sum_{i=1}^{k} y_{i\bullet}^2 - \frac{1}{kn}y_{\bullet\bullet}^2\right]$$

$$= \frac{1}{k-1}\left[(kn\mu^2 + kn\sigma_s^2 + kn\sigma_e^2) - (kn\mu^2 + n\sigma_s^2 + \sigma_e^2)\right]$$

$$= \frac{1}{k-1}\left[n(k-1)\sigma_s^2 + (k-1)\sigma_e^2\right] = n\sigma_s^2 + \sigma_e^2$$

$$E[RMS] = \frac{1}{k(n-1)}E[RSS] = \frac{1}{k(n-1)}E\left[\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \frac{1}{n}\sum_{i=1}^{k} y_{i\bullet}^2\right]$$

$$= \frac{1}{k(n-1)}E\left[(kn\mu^2 + kn\sigma_s^2 + kn\sigma_e^2) - (kn\mu^2 + kn\sigma_s^2 + k\sigma_e^2)\right]$$

$$= \frac{1}{k(n-1)}k(n-1)\sigma_e^2 = \sigma_e^2$$

Estimation of Variance Components

ANOVA approach works well for simple models (such as a one-way structure) or balanced data (such as data from designed experiments with no missing data), but they are not indicated for more complex models and data structures

Other proposed methods: expected mean squares approach of Henderson (1953), and the minimum norm quadratic unbiased estimation (Rao 1971a, 1971b), among others.

However, maximum likelihood based methods are currently the most popular, especially the restricted (or residual) maximum likelihood (REML) approach, which attempts to correct for the well-known bias in the classical maximum likelihood (ML) estimation of variance components. These two methods are briefly described next.

Estimation of Variance Components

Maximum Likelihood (ML) Estimator

Maximum likelihood estimates of the variance components can be obtained by maximizing the log-likelihood $L(\boldsymbol{\beta}, \mathbf{G}, \boldsymbol{\Sigma})$ with respect to each element of **G** and **Σ**, after replacing **β** by $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$

Alternatively, **G**, **Σ**, and **β** can be estimated simultaneously by maximizing their joint log-likelihood with respect to the variance components and the fixed effects.

As a simple example of maximum likelihood estimation of variance components, consider the balanced case (i.e., constant progeny sizes) half-sib families data set discussed previously, and the linear model:

$$y_{ij} = \mu + s_i + e_{ij}$$

with the same definitions as before, but with the additional assumption of normality of both the sire and the residual effects, i.e.:

$$s_i \overset{\text{ind}}{\sim} N(0, \sigma_s^2) \quad \text{and} \quad e_{ij} \overset{\text{ind}}{\sim} N(0, \sigma_e^2)$$

In matrix notation, this model can be expressed as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_k \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n \\ \mathbf{1}_n \\ \vdots \\ \mathbf{1}_n \end{bmatrix} \mu + \begin{bmatrix} \mathbf{1}_n & \mathbf{0}_n & \cdots & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{1}_n & \cdots & \mathbf{0}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_n & \mathbf{0}_n & \cdots & \mathbf{1}_n \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_k \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_k \end{bmatrix}$$

where $\mathbf{y}_i = [y_{i1}\ y_{i2} \cdots y_{ik}]^T$ represents the vector of observations of progeny i (i.e., relative to sire i); $\mathbf{1}_n$ and $\mathbf{0}_n$ represent n-dimensional column vectors of 1's and 0's, respectively; and $\mathbf{e}_i = [e_{i1}, e_{i2}, \ldots, e_{ik}]^T$ is the vector of residuals associated with progeny i

The vector of observations $\mathbf{y} = [\mathbf{y}_1^T\ \mathbf{y}_2^T\ \ldots\ \mathbf{y}_k^T]^T$ has then a multivariate normal distr. with mean vector $\boldsymbol{\mu} = \mathbf{1}_N \mu$ and variance-covariance matrix given by $\mathbf{I}_s \otimes (\mathbf{1}_n \sigma_s^2 \mathbf{1}_n^T) + \mathbf{I}_N \sigma_e^2$, and its density function (from which the likelihood function obtained) can be written as:

$$p(\mathbf{y} \mid \mu, \sigma_s^2, \sigma_e^2) = \frac{1}{(2\pi)^{N/2} \, |\mathbf{I}_s \otimes \mathbf{J}_n \sigma_s^2 + \mathbf{I}_N \sigma_e^2|^{1/2}}$$

$$\times \exp\left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{1}_N \mu)^T (\mathbf{J}_n \sigma_s^2 + \mathbf{I}_n \sigma_e^2)^{-1} (\mathbf{y} - \mathbf{1}_N \mu) \right\}$$

$$= (2\pi)^{-\frac{N}{2}} (\sigma_e^2)^{-\frac{(N-k)}{2}} (\sigma_e^2 + n\sigma_s^2)^{-\frac{k}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{1}_N \mu)^T \left[\mathbf{I}_s \otimes \mathbf{J}_n \left(\frac{1}{n} \left(\frac{1}{\sigma_e^2 + n\sigma_s^2} - \frac{1}{\sigma_e^2} \right) \right) \right] (\mathbf{y} - \mathbf{1}_N \mu) \right\}$$

where $\mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n^T$ is an $(n \times n)$ matrix of 1's, and $\otimes$ is the Kronecker product

The log-likelihood function can be written then as:

$$l(\mu,\sigma_s^2,\sigma_e^2) \propto -\frac{(N-k)}{2}\log(\sigma_e^2) - \frac{k}{2}\log(\sigma_e^2 + n\sigma_s^2) - \frac{1}{2\sigma_e^2}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i\bullet})^2 - \frac{1}{2}\sum_{i=1}^{k}\frac{n(\bar{y}_{i\bullet} - \mu)^2}{\sigma_e^2 + n\sigma_s^2}$$

By taking the derivatives and setting them to 0, the following solutions are obtained:

$$\hat{\mu} = \bar{y}_{\bullet\bullet} \quad , \quad \hat{\sigma}_e^2 = \frac{1}{k(n-1)}RSS \quad \text{and} \quad \hat{\sigma}_s^2 = \frac{1}{n}\left[\frac{SSS}{k} - \hat{\sigma}_e^2\right]$$

from which ML estimates of the variance components are obtained, except if $\hat{\sigma}_s^2 < 0$, in which case the estimate is set to zero

ML estimates of variance components are biased downwards as they do not take into account the degrees of freedom used for estimating the fixed effects

Estimation of Variance Components

Residual Maximum Likelihood (REML) Estimator

Restricted (or residual) maximum likelihood approach (REML): corrects the bias associated with ML estimates by taking into account the degrees of freedom used for estimating the fixed effects

REML maximizes the likelihood function of a set of error contrasts $\mathbf{d} = \mathbf{L}^T\mathbf{y}$, where $\mathbf{L}$ is a [n x (n – p)] full-rank matrix with columns orthogonal to the columns of the incidence matrix $\mathbf{X}$

The vector $\mathbf{d}$ follows a multivariate normal distribution with null mean vector and variance-covariance matrix $\mathbf{L}^T\mathbf{V}\mathbf{L} = \mathbf{L}^T(\mathbf{Z}\mathbf{G}\mathbf{Z}^T + \boldsymbol{\Sigma})\mathbf{L}$. Note that the distribution of $\mathbf{d}$ does not depend on $\boldsymbol{\beta}$.

The residual likelihood function for the variance components is then:

$$L(\mathbf{G},\mathbf{\Sigma}\,|\,\mathbf{y}) = (2\pi)^{-(n-p)/2}\,|\mathbf{L}^{T}\mathbf{V}\mathbf{L}|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{d}^{T}(\mathbf{L}^{T}\mathbf{V}\mathbf{L})^{-1}\mathbf{d}\right\}$$

Another approach for obtaining the residual likelihood function for the variance components is by integrating the fixed effects out of the 'full' likelihood function, i.e.:

$$L(\mathbf{G},\mathbf{\Sigma}\,|\,\mathbf{y}) = \int L(\boldsymbol{\beta},\mathbf{G},\mathbf{\Sigma}\,|\,\mathbf{y})d\boldsymbol{\beta}$$

as illustrated in the following example.

Recall the balanced half-sib families data set, and its associated likelihood function:

$$L(\mu,\sigma_s^2,\sigma_e^2) = (2\pi)^{-\frac{N}{2}}(\sigma_e^2)^{-\frac{(N-k)}{2}}(\sigma_e^2+n\sigma_s^2)^{-\frac{k}{2}}$$
$$\times\exp\left\{-\frac{1}{2\sigma_e^2}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2-\frac{1}{2}\sum_{i=1}^{k}\frac{n(\bar{y}_{i\bullet}-\mu)^2}{\sigma_e^2+n\sigma_s^2}\right\}$$

Its residual likelihood is then:

$$L(\sigma_s^2,\sigma_e^2) = \int L(\mu,\sigma_s^2,\sigma_e^2)d\mu$$
$$= (2\pi)^{-\frac{N}{2}}(\sigma_e^2)^{-\frac{(N-k)}{2}}(\sigma_e^2+n\sigma_s^2)^{-\frac{k}{2}}$$
$$\times\exp\left\{-\frac{1}{2\sigma_e^2}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\bullet})^2\right\}\int\exp\left\{-\frac{1}{2}\sum_{i=1}^{k}\frac{n(\bar{y}_{i\bullet}-\mu)^2}{\sigma_e^2+n\sigma_s^2}\right\}d\mu$$

which is equal to:

$$L(\sigma_s^2, \sigma_e^2) = (2\pi)^{-\frac{N}{2}} (\sigma_e^2)^{-\frac{(N-k)}{2}} \lambda^{-\frac{k}{2}}$$

$$\times \exp\left\{-\frac{1}{2\sigma_e^2}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i\bullet})^2\right\}\exp\left\{-\frac{n}{2\lambda}\sum_{i=1}^{k}(\bar{y}_{i\bullet} - \mu)^2\right\}\sqrt{2\pi\frac{\lambda}{kn}}$$

where $\lambda = \sigma_e^2 + n\sigma_s^2$.

By taking the derivatives with respect to λ and σ_e^2, and by using the invariance property of maximum likelihood estimators, the following solutions are obtained:

$$\hat{\sigma}_e^2 = \frac{1}{k(n-1)}\text{RSS} \quad \text{and} \quad \hat{\sigma}_s^2 = \frac{1}{n}\left[\frac{1}{(k-1)}\text{SSS} - \hat{\sigma}_e^2\right]$$

which are the REML estimates of the variance components, except if $\hat{\sigma}_s^2 < 0$, i.e. if

$$\text{SSS} < \frac{(k-1)}{k(n-1)}\text{RSS}$$

Explicit forms of ML and REML estimators are often not available for more complex mixed effects models

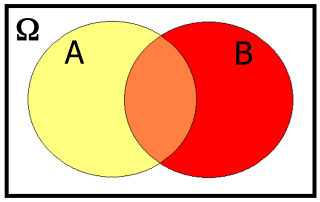
ML and REML estimates are then generally obtained by iterative approaches such as the expectation-maximization (EM) algorithm and Newton-Raphson-based procedures

Bayesian Data Analysis

Inferences using probability models for quantities we observe and for quantities about which we wish to learn

Explicit use of probability for quantifying uncertainty in inferences based on statistical data analysis

Conditional Probability (Bayes' Rule)



$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B \mid A)}{P(B)}$$

Bayesian Inference

- y: observed data; y ~ p(y|θ)
- θ: parameters (all unobserved quantities)

$$p(\theta \mid y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y \mid \theta)}{p(y)}$$

$$p(\theta \mid y) \propto p(\theta)p(y \mid \theta)$$

posterior distribution — $p(\theta \mid y)$

prior distribution — $p(\theta)$

sampling distribution — $p(y \mid \theta)$

Prior Distributions

Informative and Noninformative

Proper and Improper

Conjugate and Nonconjugate

Jeffreys' Prior

Maximum Entropy

Reference Prior

Example 1: Binomial Distribution

Data: $y_1, y_2, \ldots, y_n \overset{iid}{\sim} \text{Bin}(n_i, \theta)$, θ = Prob(y = 1)

Sampling model: $p(\mathbf{y} \mid \theta) = \prod_{i=1}^{n} p(y_i \mid \theta) = \prod_{i=1}^{n} \binom{1}{y_i} \theta^{y_i} (1-\theta)^{1-y_i}$

$$\propto \theta^{\sum y_i} (1-\theta)^{n - \sum y_i}$$

Prior: $p(\theta) = \text{Beta}(a,b) \propto \theta^{a-1} (1-\theta)^{b-1}$

Posterior: $p(\theta \mid \mathbf{y}) \propto \theta^{a + \sum y_i - 1} (1-\theta)^{n+b-\sum y_i - 1}$

$$\boxed{\theta \mid \mathbf{y} \sim \text{Beta}\left(a + \sum y_i,\ n + b - \sum y_i\right)}$$

Example 1: Binomial Distribution

$$\boldsymbol{\theta} \mid \mathbf{y} \sim \text{Beta}\left(a + \sum y_i,\ n + b - \sum y_i\right)$$

Features of the posterior distribution:

Posterior mean: $E[\boldsymbol{\theta} \mid \mathbf{y}] = \dfrac{a + \sum y_i}{n + a + b}$

Posterior mode: $\text{Mode}[\boldsymbol{\theta} \mid \mathbf{y}] = \dfrac{a + \sum y_i - 1}{n + a + b - 2}$

Posterior variance: $\text{Var}[\boldsymbol{\theta} \mid \mathbf{y}] = \dfrac{\left(a + \sum y_i\right)\left(n + b - \sum y_i\right)}{(n + a + b)^2 (n + a + b + 1)}$

percentis, HPD, etc.

Example 1: Binomial Distribution

Setting, for example a = 1 and b = 1:

Prior: $p(\boldsymbol{\theta}) = \text{Uniform}(0,1)$

Posterior: $p(\boldsymbol{\theta} \mid \mathbf{y}) \propto \boldsymbol{\theta}^{\sum y_i - 1} (1 - \boldsymbol{\theta})^{n - \sum y_i}$

$$\boldsymbol{\theta} \mid \mathbf{y} \sim \text{Beta}\left(1 + \sum y_i,\ n + 1 - \sum y_i\right)$$

Note that in this case the posterior mode coincides with the maximum likelihood estimate of θ:

$$\text{Mode}[\boldsymbol{\theta} \mid \mathbf{y}] = \frac{1}{n} \sum y_i$$

Example 2: Normal Distribution

Data: $y_1, y_2, \ldots, y_n \overset{iid}{\sim} N(\mu, \sigma^2)$, with known σ^2

Sampling model: $p(y_i \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \mu)^2\right\}$

$$p(\mathbf{y} \mid \mu, \sigma^2) = \prod_{i=1}^{n} p(y_i \mid \mu, \sigma^2)$$

$$\propto \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right\}$$

$$\propto \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} n(\bar{y} - \mu)^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \bar{y})^2\right\}$$

Example 2: Normal Distribution

Prior (Conjugate): $\mu \sim N(\phi, \tau^2)$

$$p(\mu) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(\mu - \phi)^2\right\}$$

Joint posterior:

$$p(\mu \mid \mathbf{y}) \propto p(y \mid \mu, \sigma^2) \times p(\mu)$$

$$\propto \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} n(\bar{y} - \mu)^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \bar{y})^2\right\}$$

$$\times \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(\mu - \phi)^2\right\}$$

Joint posterior (cont'ed):

$$p(\mu \mid \mathbf{y}) \propto \exp\left\{-\frac{1}{2\sigma^2} n(\bar{y}-\mu)^2\right\} \exp\left\{-\frac{1}{2\tau^2}(\mu-\phi)^2\right\}$$

$$\propto \exp\left\{-\frac{n(\bar{y}-\mu)^2}{2\sigma^2} - \frac{(\mu-\phi)^2}{2\tau^2}\right\}$$

$$\propto \exp\left\{-\frac{\mu^2}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right) + \mu\left(\frac{n\bar{y}}{\sigma^2}+\frac{\phi}{\tau^2}\right) - \frac{1}{2\tau}\left(\frac{n\bar{y}^2}{\sigma^2}+\frac{\phi^2}{\tau^2}\right)\right\}$$

$$= \exp\left\{-\frac{1}{2\sigma_n^2}(\mu+\mu_n)^2\right\}$$

where $\frac{1}{\sigma_n^2} = \left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)$ and $\frac{\mu_n}{\sigma_n^2} = \left(\frac{n\bar{y}}{\sigma^2}+\frac{\phi}{\tau^2}\right)$

Hence: $\mu \mid \mathbf{y} \sim N\left(\frac{n\tau^2}{n\tau^2+\sigma^2}\bar{y} + \frac{\sigma^2}{n\tau^2+\sigma^2}\phi,\ \left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}\right)$

Multi Parameter Models

$$y \sim p(y \mid \theta_1, \theta_2, \ldots, \theta_p)$$

$$p(\theta_1, \theta_2, \ldots, \theta_p \mid y) \sim p(\theta_1, \theta_2, \ldots, \theta_p) p(y \mid \theta_1, \theta_2, \ldots, \theta_p)$$

Marginal Posterior Distributions

$$p(\theta_1 \mid y) \propto \int_{\theta \neq \theta_1} p(\theta_1, \theta_2, \ldots, \theta_p \mid y) d\theta_{\theta \neq \theta_1}$$

Marginal Posterior Distributions

Marginalization (i.e. integrals) in multi-dimensional models can be cumbersome and some times do not have analytical form

An alternative in this regard: Monte Carlo methods

Monte Carlo integration consists of sampling from the posterior distribution, and then using such sampled values to calculate features of interest on the (joint or marginal) posterior distribution

There are many algorithms that can be used to sample from a distribution; some are based on Markov chains, among which the Gibbs sampling is probably the most popular

Gibbs Sampling

$$\theta = (\theta_1, \theta_2, \ldots, \theta_r) \rightarrow p(\theta_i \mid \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_r)$$

$$\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_r^{(0)})$$

$$\theta_1^{(1)} \mid \theta_2^{(0)}, \theta_3^{(0)}, \ldots, \theta_r^{(0)}$$

$$\theta_2^{(1)} \mid \theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_r^{(0)}$$

$$\vdots$$

$$\theta_r^{(1)} \mid \theta_2^{(1)}, \theta_3^{(1)}, \ldots, \theta_{r-1}^{(1)}$$

Burn-in & Convergence

Tinning interval & Lag correlations

Sample size & Monte Carlo error

Monte Carlo Approximations

After convergence, each sampled vector is a sample from the joint posterior distribution, and so each sampled element (scalar) is a sample from the respective marginal posterior distribution

For each parameter (e.g., θ_1) we'll have then a series of values:

$$\theta_1^{(1)}, \theta_1^{(2)}, \theta_1^{(3)}, \ldots, \theta_1^{(N)}$$

from which features of its distribution (e.g., posterior mean) can be approximated, for example:

$$E[\theta_1 \mid \mathbf{y}] \cong \frac{1}{N} \sum_{j=1}^{N} \theta_1^{(j)}$$

Example

bayes linear regression

Information on phenotypes and genotypes for a specific marker

Marker Genotype	Phenotype (8 individuals per group)
MM	95.9, 108.0, 96.5, 92.9 101.0, 94.5, 93.7, 89.8
Mm	105.2, 107.9, 89.9, 113.4 109.7, 102.4, 97.1, 107.1
mm	117.1, 95.2, 106.4, 104.7 92.5, 123.9, 97.8, 100.5