2019 SISG Module 8: Bayesian Statistics for Genetics
Lecture 7: Generalized Linear Modeling

Jon Wakefield

Departments of Statistics and Biostatistics
University of Washington

Outline

Introduction and Motivating Examples

Generalized Linear Models
- Bayes Linear Model
- Bayes Logistic Regression

Generalized Linear Mixed Models
- Temporal Smoothing
- ASE Example

Approximate Bayes Inference
- The Approximation

Appendix: Hierarchical Modeling of Allele-Specific Expression Data
- Motivation
- Modeling

Introduction

Introduction

In this lecture we will discuss Bayesian modeling in the context of Generalized Linear Models (GLMs).

This discussion will include the addition of random effects, i.e. we'll consider the class of Generalized Linear Mixed Models (GLMMs).

Estimation via the quick INLA technique will be demonstrated, along with its R implementation.

An approximation technique that is useful (in particular) in the context of Genome Wide Association Studies (GWAS) (in which the number of rows of data to analyze is large) will also be introduced.

The accompanying R code allows the analyses presented here to be replicated.

A complex mixture model for ASE is included in the Appendix, to illustrate some of the flexibility of Bayes modeling.

Motivating Example: Logistic Regression

We consider case-control data for the disease Leber Hereditary Optic Neuropathy (LHON) disease with genotype data for marker rs6767450:

	CC	CT	TT	Total
	$x = 0$	$x = 1$	$x = 2$	
Cases	6	8	75	89
Controls	10	66	163	239
Total	16	74	238	328

Let $x = 0, 1, 2$ represent the number of T alleles, and $p(x)$ the probability of being a case, given x copies of the T allele.

Motivating Example: Logistic Regression

For such case-control data one may fit the multiplicative odds model:

$$\frac{p(x)}{1-p(x)} = \exp(\alpha) \times \exp(\theta x),$$

with a binomial likelihood.

Interpretation:

- $\exp(\alpha)$ is of little interest given the case-control sampling.
- $\exp(\theta)$ is the odds ratio describing the multiplicative change in risk for one T allele versus zero T alleles.
- $\exp(2\theta)$ is the odds ratio describing the multiplicative change in risk for two T alleles versus zero T alleles.
- Odds ratios approximate the relative risk for a rare disease.

A Bayesian analysis adds a prior on α and θ.

Motivating Example: FTO Data

Recall

- $Y =$ weight
- $x_{\text{g}} =$ fto heterozygote $\in \{0, 1\}$
- $x_{\text{a}} =$ age in weeks $\in \{1, 2, 3, 4, 5\}$

We will examine the fit of the model

$$\mathrm{E}[Y|x_{\text{g}}, x_{\text{a}}] = \beta_0 + \beta_{\text{g}} x_g + \beta_{\text{a}} x_a + \beta_{\text{int}} x_{\text{g}} x_{\text{a}},$$

with independent normal errors, and compare with a Bayesian analysis.

Motivating Example: RNA Seq with Replicates

- We report an experiment carried out in a previous collaboration, see Connelly *et al.* (2014) for further details.
- Start with two haploid yeast strains (individuals).
- From these we obtain RNA-Seq data, where we isolate RNA from the two individuals, fragment and sequence it using next-generation sequencing, and map the sequencing reads back to the genome to generate RNA levels in the form of counts of the number of sequencing reads mapping at each gene.
- Also mate the two haploid yeast strains together to form a diploid hybrid. We again isolate RNA, fragment, and sequence it.
- Then take advantage of polymorphisms between the two strains in order to map reads to either of the two haploid individuals, giving us counts for the number of reads mapping to either one of the parental genomes in the diploid hybrid for each gene.

Motivating Example: RNA Seq with Replicates

- We are interested in two questions from this data. First, we want to look for evidence of trans effects at each gene; in biological terms, this means that polymorphisms located far from the gene are responsible for differences in RNA levels.
- To detect this, look for genes where the difference between RNA levels in the haploids differs from the difference between RNA levels for the two parental strains in the diploid.
- The question we concentrate on looking for cis effects, these are polymorphisms near the gene itself that are responsible for differences in RNA levels.
- We can detect cis effects as a difference in the count of reads mapping to each of the parental strains in the diploid at a gene, is the probability of arising from the two parents, 0.5?

Motivating Example: RNA Seq Data, Statistical Model

There are two replicates and so for each of N genes we obtain two sets of counts.

For the diploid hybrid let :

- Y_{ij} be the number of A alleles for gene i and replicate j, and
- N_{ij} is the total number of counts, so that $N_{ij} - Y_{ij}$ is the number of T alleles $j = 1, 2$.

Motivating Example: RNA Seq Data, Statistical Model

We fit a random effects logistic regression model starting with first stage:

$$Y_{ij} | N_{ij}, p_{ij} \sim \text{Binomial}(N_{ij}, p_{ij})$$

so that p_{ij} is the probability of seeing an A read for gene i and replicate j.

At the second stage:

$$\frac{p_{ij}}{1 - p_{ij}} = \exp(\theta_i + \epsilon_{ij}),$$

where

$$\epsilon_{ij} \sim \text{N}(0, \sigma^2),$$

represent random effects that allow for excess-binomial variation.

In the model θ_i is a parameter of interest – if a (say) 95% posterior interval estimate contains 0 then we have evidence of cis effects.

GLMs

Generalized Linear Models

- Generalized Linear Models (GLMs) provide a very useful extension to the linear model class.
- GLMs have three elements:
  1. The responses follow an exponential family.
  2. The mean model is linear in the covariates on some scale.
  3. A link function relates the mean of the data to the covariates.
- In a GLM the response y_i are independently distributed and follow an exponential family[1], $i = 1, \ldots, n$.
- Examples: Normal, Poisson, binomial.

[1] so that the distribution is of the form $p(y_i|\theta_i, \alpha) = \exp(\{y_i\theta_i - b(\theta_i)\}/\alpha + c(y_i, \alpha))$, where θ_i and α are scalars

Generalized Linear Models

- The link function $g(\cdot)$ provides the connection between the mean $\mu = \mathrm{E}[Y]$ and the linear predictor $\boldsymbol{x\beta}$, via

  $$g(\mu) = \boldsymbol{x\beta},$$

  where $\boldsymbol{x}$ is a vector of explanatory variables and $\boldsymbol{\beta}$ is a vector of regression parameters.
- For normal data, the usual link is the identity

  $$g(\mu) = \mu = \boldsymbol{x\beta}.$$

- For binary data, a common link is the logistic

  $$g(\mu) = \log\left(\frac{\mu}{1-\mu}\right) = \boldsymbol{x\beta}.$$

- For Poisson data, a common link is the log

  $$g(\mu) = \log(\mu) = \boldsymbol{x\beta}.$$

Bayesian Modeling with GLMs

- For a generic GLM, with regression parameters $\boldsymbol{\beta}$ and a scale parameter α, the posterior is

$$p(\boldsymbol{\beta}, \alpha | \boldsymbol{y}) \propto p(\boldsymbol{y} | \boldsymbol{\beta}, \alpha) \times p(\boldsymbol{\beta}, \alpha).$$

- An immediate question is: How to specify a prior distribution $p(\boldsymbol{\beta}, \alpha)$?
- How to perform the computations required to summarize the posterior distribution (including the calculation of Bayes factors)?

Bayesian Computation

Various approaches to computation are available:

- Conjugate analysis — the prior combines with likelihood in such a way as to provide analytic tractability (at least for some parameters).
- Analytical Approximations — asymptotic arguments used (e.g. Laplace).
- Numerical integration.
- Direct (Monte Carlo) sampling from the posterior, as we have already seen.
- Markov chain Monte Carlo — very complex models can be implemented, for example with `WinBUGS`, `JAGS` or `Stan`.
- Integrated nested Laplace approximation (INLA). Cleverly combines analytical approximations and numerical integration: we illustrate the use of this method in some detail.

Integrated Nested Laplace Approximation (INLA)

- The homepage of the INLA software is here:
  http://www.r-inla.org/home
- There are also lots of example links at this website.
- The fitting of many common models is described here:
  http://www.r-inla.org/models/likelihoods
- INLA can fit GLMs, GLMMs and many other useful model classes.

INLA for the Linear Model

- The model is

$$Y = \mathrm{E}[Y|x_{\mathrm{g}}, x_{\mathrm{a}}] = \beta_0 + \beta_{\mathrm{g}} x_g + \beta_{\mathrm{a}} x_a + \beta_{\mathrm{int}} x_{\mathrm{g}} x_{\mathrm{a}} + \epsilon$$

  where $\epsilon|\sigma^2 \sim_{iid} \mathrm{N}(0, \sigma^2)$.
- This model has five parameters: the four fixed effects are $\beta_0, \beta_{\mathrm{g}}, \beta_{\mathrm{a}}, \beta_{\mathrm{int}}$ and the error variance is σ^2 (note that in `inla` inference is reported for the precision σ^{-2}).
- In general, posterior distributions can be summarized graphically or via numerical summaries.
- In Figures 1 and 2 give posterior marginal distributions for the fixed effects and hyperparameter σ^{-2}, respectively, under an analysis with relatively flat priors.

Comparison of OLS and Bayess

```
# OLS
ols.fit <- lm(liny~linxg+linxa+linxint,data=ftodf)
# MLEs and SEs
cbind(coef(ols.fit),sqrt(diag(vcov(ols.fit))))
                   [,1]      [,2]
(Intercept) -0.06821632 1.4222970
linxg        2.94485495 2.0114316
linxa        2.84420729 0.4288387
linxint      1.72947648 0.6064695
# INLA
formula <- liny~linxg+linxa+linxint
lin.mod <- inla(formula,data=ftodf,family=”gaussian”)
# Posterior means and SDs
lin.mod$summary.fixed[c(1,2)]
                   mean        sd
(Intercept) -0.06162681 1.4255270
linxg        2.93325529 2.0135662
linxa        2.84237281 0.4298868
linxint      1.73261901 0.6073410
```

Virtually identical!

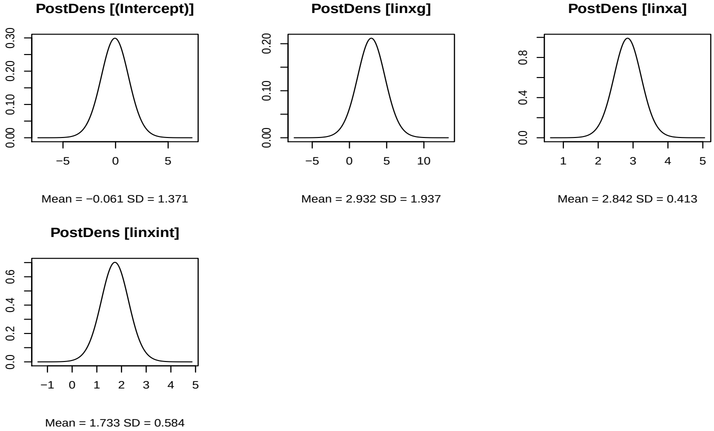


Figure 1: Marginal distributions of the intercept and regression coefficients.

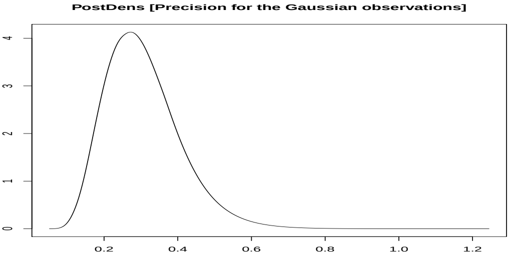


Figure 2: Marginal distribution of the error precision.

INLA for the Linear Model

- As with a non-Bayesian analysis, model checking is important and in Figure 3 we present a number of diagnostic plots.
- Plots:
  (a) Normality of residuals? Sample size is quite small.
  (b) Is the relationship with age linear?
  (c) Mean variance relationship?
  (d) Overall fit.
- For these data, the model assumptions look reasonable.

FTO Diagnostic Plots

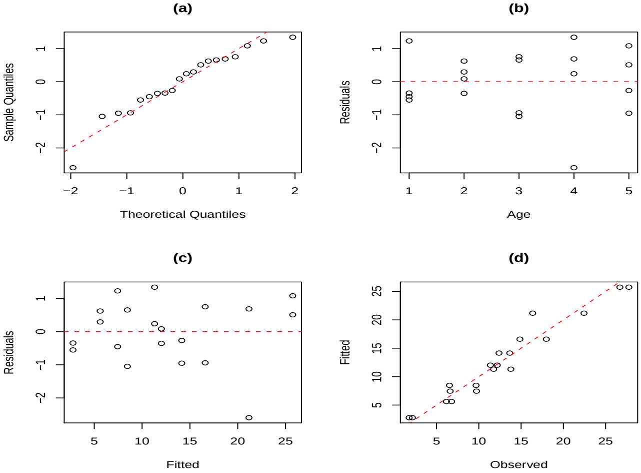


Figure 3: Plots to assess model adequacy: (a) Normal QQ plot, (b) residuals versus age, (c) residuals versus fitted, (d) fitted versus observed.

Bayes Logistic Regression

- The likelihood is

$$Y(x)|p(x) \sim \text{Binomial}(\,N(x), p(x)\,), \quad x = 0, 1, 2.$$

- Logistic link:

$$\log\left(\frac{p(x)}{1-p(x)}\right) = \alpha + \theta x$$

- The prior is

$$p(\alpha, \theta) = p(\alpha) \times p(\theta)$$

  with

  - $\alpha \sim \text{N}(\mu_\alpha, \sigma_\alpha)$ and
  - $\theta \sim \text{N}(\mu_\theta, \sigma_\theta)$. where $\mu_\alpha, \sigma_\alpha, \mu_\theta, \sigma_\theta$ are constant that are specified to reflect prior beliefs.

Comparison of MLE and Bayess

```
# MLE
logitmod <- glm(cbind(y,z)~x,family=”binomial”)
# MLEs and SEs
cbind(coef(logitmod),sqrt(diag(vcov(logitmod))))
                  [,1]      [,2]
(Intercept) -1.8076928 0.4553938
x            0.4787428 0.2504594
# INLA
cc.mod <- inla(y~x,family=”binomial”,data=cc.dat,Ntrials=y+z)
# Posterior mean and SD
cc.mod$summary.fixed[c(1,2)]
                  mean        sd
(Intercept) -1.8069628 0.4553857
x            0.4800092 0.2504597
```

Virtually identical!

Prior Choice for Positive Parameters

- It is convenient to specify lognormal priors for a positive parameter, for example $\exp(\beta)$ (the odds ratio) in a logistic regression analysis.
- One may specify two quantiles of the distribution, and directly solve for the two parameters of the lognormal.
- Denote by $\theta \sim \text{LogNormal}(\mu, \sigma)$ the lognormal distribution for a generic positive parameter θ with $\text{E}[\log \theta] = \mu$ and $\text{var}(\log \theta) = \sigma^2$, and let θ_1 and θ_2 be the q_1 and q_2 quantiles of this prior.
- In our example, $\theta = \exp(\beta)$.
- Then it is straightforward to show that

$$\mu = \log(\theta_1)\left(\frac{z_{q_2}}{z_{q_2} - z_{q_1}}\right) - \log(\theta_2)\left(\frac{z_{q_1}}{z_{q_2} - z_{q_1}}\right), \ \sigma = \frac{\log(\theta_1) - \log(\theta_2)}{z_{q_1} - z_{q_2}}. \tag{1}$$

Prior Choice for Positive Parameters

- As an example, suppose that for the odds ratio e^{β} we believe there is a 50% chance that the odds ratio is less than 1 and a 95% chance that it is less than 5; with

  $$q_1 = 0.5,\ \theta_1 = 1.0,\ q_2 = 0.95,\ \theta_2 = 5.0,$$

  we obtain lognormal parameters

  $$\begin{aligned} \mu &= 0 \\ \sigma &= (\log 5)/1.645 = 0.98. \end{aligned}$$

- The density is shown in Figure 4.

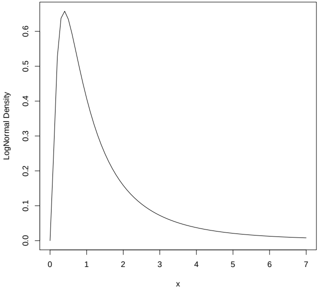


Figure 4: Lognormal density with 50% point 1 and 95% point 5.

Logistic Regression Example

- In the second analysis we specify

$$\begin{aligned}\alpha &\sim \mathrm{N}(0, 1/0.1)\\ \theta &\sim \mathrm{N}(0, W)\end{aligned}$$

  where W is such that the 97.5% point of the prior is $\log(1.5)$, i.e. we believe the odds ratio lies between $2/3$ and $3/2$ with probability 0.95.
- The marginal distributions are displayed.

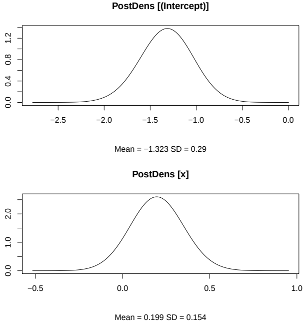


Figure 5: Posterior marginals for the intercept α and the log odds ratio θ.

Comparison of MLE and Bayess

```
# MLE
logitmod <- glm(cbind(y,z)~x,family="binomial")
# MLEs and SEs
cbind(coef(logitmod),sqrt(diag(vcov(logitmod))))
                  [,1]      [,2]
(Intercept) -1.8076928 0.4553938
x            0.4787428 0.2504594
# INLA
W <- LogNormalPriorCh(1,1.5,0.5,0.975)$sigma^2
cc.mod2 <- inla(y~x,family="binomial",data=cc.dat,Ntrials=y+z,
   control.fixed=list(mean.intercept=c(0),prec.intercept=c(.1),
                      mean=c(0),prec=c(1/W)))
cc.mod2$summary.fixed[c(1,2)]
                 mean        sd
(Intercept) -1.322757 0.2895597
x            0.198683 0.1535503
```

Big changes!

GLMMs

Smoothing

When faced with estimation *n* different quantities of the **prevalence** under different conditions, there are three model choices:

- The true underlying prevalence risks are **ALL THE SAME**.
- The true underlying prevalence risks are **DISTINCT** but not linked.
- The true underlying prevalence risks are **SIMILAR IN SOME SENSE**.

The third option seems plausible when the conditions are related, but how do we model “similarity”?

Smoothing

There are a number of possibilities for **SMOOTHING** models:

- The prevalences are drawn from some **COMMON** probability distribution, but are not ordered in any way. We refer this as the independent and identically distributed, or **IID** model. We could think of this as saying we think the prevalences are likely to be of the same order of magnitude.
- The prevalences are **CORRELATED** over time.

These are both examples of **HIERARCHICAL** or **RANDOM EFFECTS MODELS** — a key element is estimating the **SMOOTHING PARAMETER**.

Smoothing over Time

Rationale and overview of models for temporal smoothing:

- We often expect that the true underlying prevalence in a study region will exhibit some degree of smoothness over time.
- A linear trend in time is unlikely to be suitable for more than a small number of years, and higher degree polynomials can produce erratic fits.
- Hence, local smoothing is preferred.
- Splines and random walk models have proved successful as local smoothers.
- And to emphasize again, in either approach, the choice of smoothing parameter is crucial.

Random Walk Models

We use random walk models which encourage the mean responses (e.g., prevalences) across time to not deviate too greatly from their neighbors.

The true underlying mean of the prevalence at time t is modeled as a function of its neighbors:

$$\mu_t \mid \mu_{\text{NE}(t)} \sim \text{N}(m_t, v_t),$$

where

- μ_t is the mean prevalence (or some function of it such as the logit) at time t.
- $\mu_{\text{NE}(t)}$ is the set of neighboring means – with the number of neighbors chosen depending on the model used – typically 2 or 4.
- m_t is the mean of some set of neighbors – for a **first order random walk** or **RW1** it is simply $\frac{1}{2}(\mu_{t-1} + \mu_{t+1})$.
- v_t is the variance, and depends on the number of neighbors – for the RW1 model it is $\sigma^2/2$, where σ^2 is a smoothing parameter – small values give large smoothing.

Random Walk Models

- The smoothing parameter σ^2 is estimated from the data, and determines the extent deviations from the mean are penalized.
- The penalty term for the RW1 model is:

$$p(\mu_t \mid \mu_{t-1}, \mu_{t+1}, \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}\left[\mu_t - \tfrac{1}{2}\left(\mu_{t-1} + \mu_{t+1}\right)\right]^2\right\}.$$

- Hence:
  - Values of μ_t that are close to $\frac{1}{2}(\mu_{t-1} + \mu_{t+1})$ are favored (higher density).
  - The relative favorability is governed by σ^2 – if this variance is small, then μ_t can't stray too far from its neighbors.
- Predictions from the RW1 are

$$\mu_{T+S}|\mu_1, \ldots, \mu_T, \sigma^2 \sim \mathrm{N}(\mu_T, \sigma^2 \times S).$$

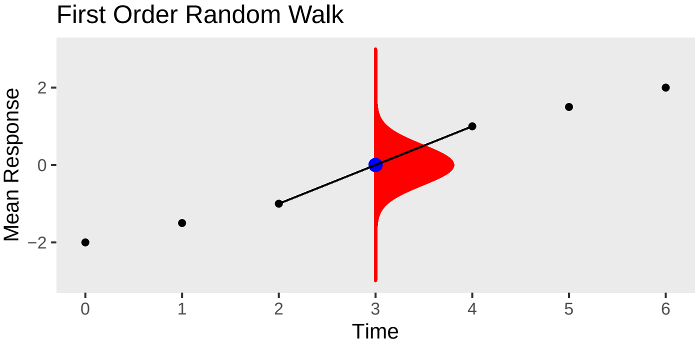


Figure 6: Illustration of the RW1 model for smoothing at time 3. The mean of the smoother is the average of the two adjacent points (and is highlighted as ●), and deviations from this mean are penalized via the normal distribution shown in **red**.

RW1 Model

- Form of the prior density is:

$$\begin{aligned}\pi(\boldsymbol{\mu}|\sigma^2) &\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T-1}(\mu_{t+1}-\mu_t)^2\right)\\ &= \exp\left(-\frac{1}{2\sigma^2}\sum_{t\sim t'}(\mu_t-\mu_{t'})^2\right) = \exp\left(-\frac{1}{2}\boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{Q}\boldsymbol{\mu}\right)\end{aligned}$$

  where $t \sim t'$ indicates t is a neighbor of t' and the precision is $\boldsymbol{Q} = \boldsymbol{R}/\sigma^2$ with

$$\boldsymbol{R} = \begin{bmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{bmatrix}$$

  and zeroes everywhere else.
- This sparsity leads to big gains in computational efficiency.

RW2 Model

- The second order RW (RW2) model produces smoother trajectories than the RW1, and has more reasonable short term predictions, which is desirable for modeling child prevalence.
- In terms of second differences:

$$(\mu_t - \mu_{t-1}) - (\mu_{t-1} - \mu_{t-2}) \sim \text{N}(\ 0, \sigma^2\ ),$$

  showing that deviations from linearity are discouraged.
- Forecasts S steps ahead have a normal distribution with mean:

$$\text{E}[\mu_{T+S} \mid \mu_1, \dots, \mu_T] = \mu_T + S(\mu_T - \mu_{T-1})$$

  which is a linear function of the values at the last two time points.
- The variance is

$$\text{var}(\mu_{T+S} \mid \mu_1, \dots, \mu_T) = \frac{\sigma^2}{6} \times S(S+1)(2S+1)$$

  which is cubic in the number of periods S, so blows up very quickly.

RW2 Model

- Form of the prior density is:

$$\begin{aligned}\pi(\boldsymbol{\mu}|\sigma^2) &\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T-2}(\mu_{t+2}-2\mu_{t+1}+\mu_t)^2\right)\\ &= \exp\left(-\frac{1}{2}\boldsymbol{\mu}^{\intercal}\boldsymbol{Q}\boldsymbol{\mu}\right)\end{aligned}$$

where the precision is $\boldsymbol{Q}=\boldsymbol{R}/\sigma^2$ with

$$\boldsymbol{R}=\begin{bmatrix} 1 & -2 & 1 & & & & \\ -2 & 5 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & 1 & -4 & 6 & -4 & 1 & \\ & & . & . & . & . & . \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 5 & -2 \\ & & & & 1 & -2 & 1 \end{bmatrix}$$

and zeroes everywhere else.

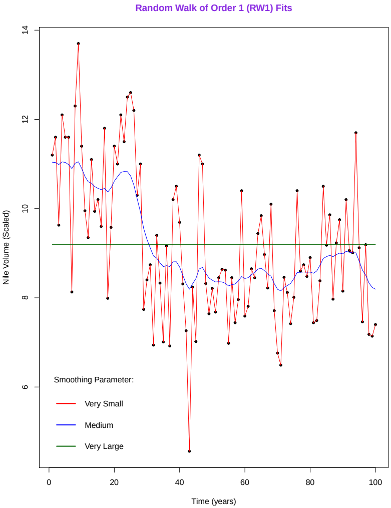


Figure 7: Nile data with RW1 fits under different priors for smoothing parameter σ^{-2}.

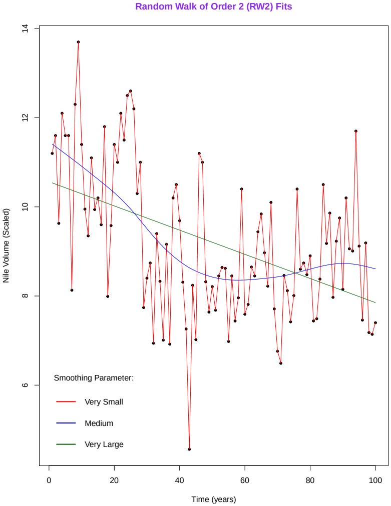


Figure 8: Nile data with RW2 fits under different priors for smoothing parameter σ^{-2}.

Temporal Smoothing Model Summary

We have three models:

IID MODEL:

$$\mu_t \sim \text{N}(0, \sigma^2),$$

smooth towards zero.

RW1 MODEL:

$$\mu_t - \mu_{t-1} \sim \text{N}(0, \sigma^2),$$

smooth towards the previous value.

RW2 MODEL:

$$(\mu_t - \mu_{t-1}) - (\mu_{t-1} - \mu_{t-2}) \sim \text{N}(0, \sigma^2),$$

smooth towards the previous slope.

RW Fitting to Simulated Data

- We illustrate fitting with the RW2 model, using the simulated data seen earlier.
- The model is:

$$\begin{aligned}
Y_t | p_t &\sim \text{Binomial}(n_t, p_t) \\
\frac{p_t}{1-p_t} &= \exp(\alpha + \phi_t) \\
(\phi_1, \dots, \phi_T) &\sim \text{RW2}(\sigma^2) \\
\sigma^2 &\sim \text{Prior on Smoothing Parameter} \\
\alpha &\sim \text{Prior on Intercept}
\end{aligned}$$

RW Fitting to Simulated Data

- Fit using `R-INLA`.

```
n1 <- 10
p <- 0.2
time <- seq(1,60)
# Simulate data
y1 <- rbinom(length(time),n1,p)
inladf1 <- data.frame(y1=y1,time=time)
# Define model
formula1s = y1~f(time,model="rw2")
fit1s <- inla(formula1s,data=inladf1 ,
              family="binomial",Ntrials=n1,
              control.predictor=list(compute=TRUE))
```

- On Figures 9 and 10 the fitted values are shown in red – in both the constant prevalence and curved prevalence cases, the reconstruction is reasonable.

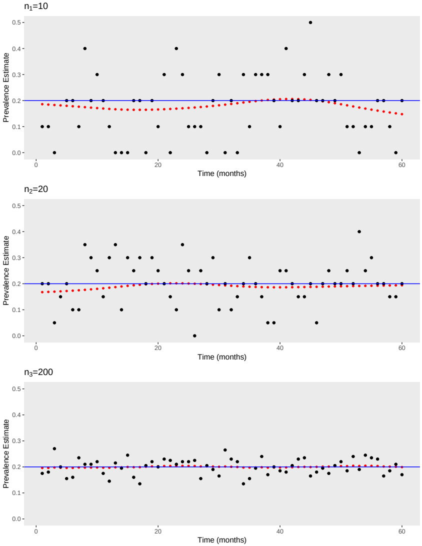


Figure 9: Prevalence estimates over time from simulated data, true prevalence $p = 0.2$ (blue solid lines). Smoothed random walk estimates in **red**.

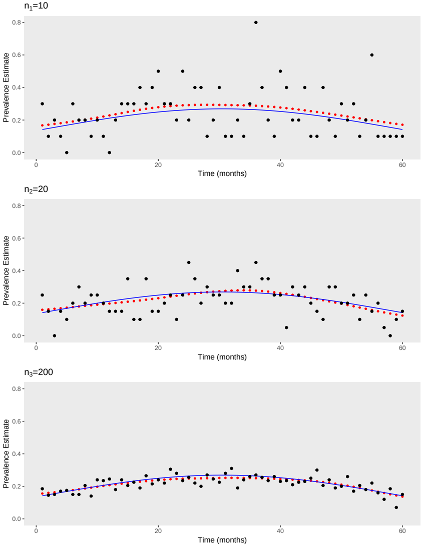


Figure 10: Prevalence estimates over time from simulated data, true prevalence corresponds to curved blue solid line. Smoothed random walk estimates in **red**.

The RNA-Seq Data: INLA Analysis

- Recall there are two replicates and so for each of N genes we obtain two sets of counts.
- For the diploid hybrid, let Y_{ij} be the number of A alleles for gene i and replicate j, and N_{ij} is the total number of counts, $j = 1, 2$.
- We fit a hierarchical logistic regression model starting with first stage:

  $$Y_{ij} | N_{ij}, p_{ij} \sim \text{Binomial}(N_{ij}, p_{ij})$$

  so that p_{ij} is the probability of seeing an A read for gene i and replicate j.
- At the second stage:

  $$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \theta_i + \epsilon_{ij}$$

  where $\epsilon_{ij} | \sigma^2 \sim \text{N}(0, \sigma^2)$ represent random effects that allow for excess-binomial variation; there are a pair for each gene.
- The θ_i parameters are taken as fixed effects with relatively flat priors.
- $\exp(\theta_i)$ is the odds of seeing an A read for gene i.
- Figures 11, 12 and 13 summarize inference.

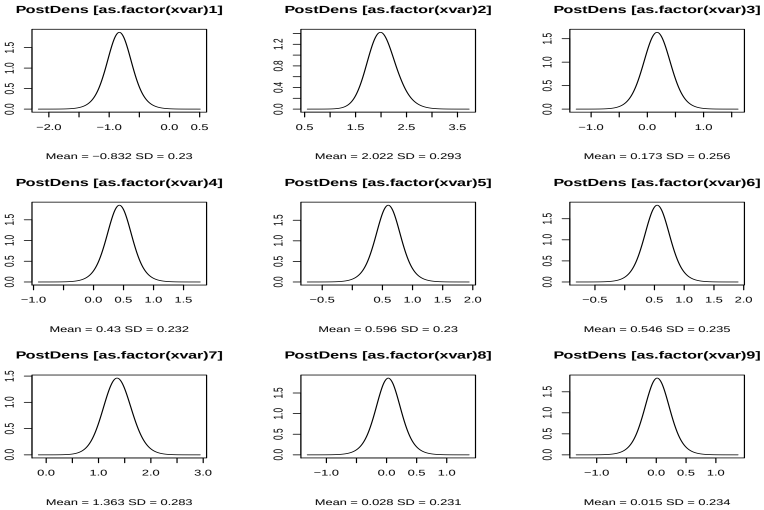


Figure 11: Posterior marginals for the first 9 gene effects θ_i (compare with zero for evidence of cis effects). We plot 9 rather than all 10 for display purposes.

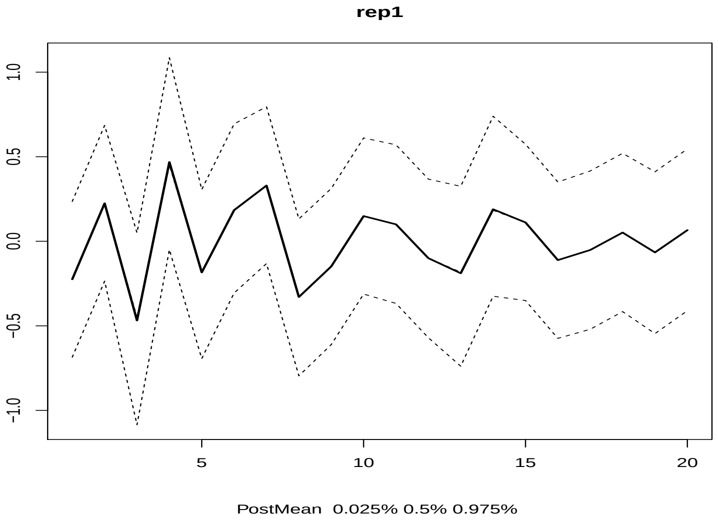


Figure 12: Posterior quantiles for 20 random effects ϵ_{ij}, which allow excess-binomial variation.

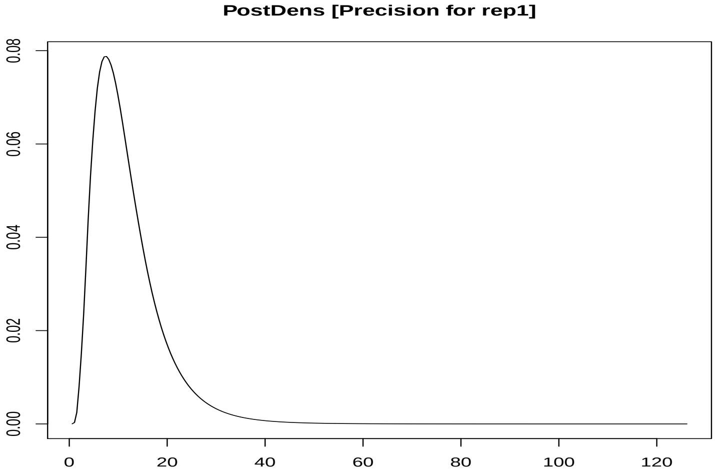


Figure 13: Posterior marginal for precision of random effects σ^{-2}.

An Informative Summary for the RNA-Seq Data

- We extract the 95% intervals and posterior medians for the log odds of being an A allele.
- Comparison with 0 (in Figure 14) gives an indication of cis effects.
- Genes 1, 2, 5, 6, 7 show evidence of cis effects.

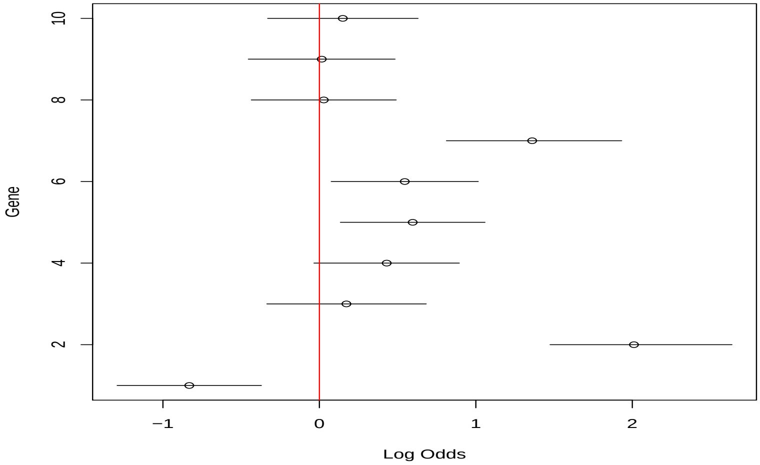


Figure 14: Posterior marginal intervals for posterior of interest θ_i. Genes with posterior intervals that do not include zero, show evidence of cis effects.

Approximate Bayes

Approximate Bayes Inference

- Particularly in the context of a large number of experiments, a quick and accurate model is desirable.
- We describe such a model in the context of a GWAS.
- This model is relevant when the sample size in each experiment is large.
- We first recap the normal-normal Bayes model.
- Subsequently, we describe the approximation and provide an example.

Recall: The Normal-Normal Model

The model:

- Prior: $\theta \sim \mathrm{N}(m, v)$ and
- Likelihood: $Y_1, \ldots, Y_n | \theta \sim \mathrm{N}(\theta, \sigma^2)$.

Posterior $p(\theta | y_1, \ldots, y_n)$ is normal with

$$\mathrm{var}(\theta | y_1, \ldots, y_n) = [1/v + n/\sigma^2]^{-1}$$

and

$$\begin{aligned} \mathrm{E}[\theta | y_1, \ldots, y_n] &= \frac{m/v + \bar{y} n/\sigma^2}{1/v + n/\sigma^2} \\ &= m \left(\frac{1/v}{1/v + n/\sigma^2} \right) + \bar{y} \left(\frac{n/\sigma^2}{1/v + n/\sigma^2} \right) \end{aligned}$$

A Normal-Normal Approximate Bayes Model

- Consider again the logistic regression model

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + x_i\theta$$

with interest focusing on θ.

- We require priors for α, θ, and some numerical/analytical technique for estimation/Bayes factor calculation.
- Wakefield (2007, 2009) considered replacing the likelihood by the asymptotic distribution of the MLE, to give posterior:

$$p(\theta|\widehat{\theta}) \propto p(\widehat{\theta}|\theta)p(\theta)$$

where

  - $\widehat{\theta}|\theta \sim \text{N}(\theta, V)$ – the asymptotic distribution of the MLE,
  - $\theta \sim \text{N}(0, W)$ – the prior on the log RR. Can choose W so that 95% of relative risks lie in some range, e.g. [2/3,1.5].

Posterior Distribution

- Under this model, the posterior distribution for the log odds ratio θ is
$$\theta|\widehat{\theta} \sim \mathrm{N}(r\widehat{\theta}, rV)$$
where
$$r = \frac{W}{V+W}.$$
- Hence, we have shrinkage to the prior mean of 0.
- The posterior median for the odds ratio is $\exp(r\widehat{\theta})$ and a 95% credible interval is
$$\exp(r\widehat{\theta} \pm 1.96\sqrt{rV}).$$
- Note that as $W \to \infty$ and/or $V \to 0$ (which occurs as we gather more data) the non-Bayesian point and interval estimates are recovered (since $r \to 1$).

A Normal-Normal Approximate Bayes Model

- We are interested in the hypotheses: $H_0 : \theta = 0,\ H_1 : \theta \neq 0$ and evaluation of the Bayes factor

$$\text{BF} = \frac{p(\widehat{\theta}|H_0)}{p(\widehat{\theta}|H_1)}.$$

- Using the approximate likelihood and normal prior we obtain:

$$\text{Approximate Bayes Factor } = \frac{1}{\sqrt{1-r}} \exp\left(-\frac{Z^2}{2} r\right),$$

with $Z = \frac{\widehat{\theta}}{\sqrt{V}}$, $r = \frac{W}{V+W}$.

A Normal-Normal Approximate Bayes Model

- The approximation can be combined with a Prior Odds $= \pi_0/(1-\pi_0)$ to give

$$\text{Posterior Odds on } H_0 = \frac{\text{BFDP}}{1-\text{BFDP}} = \text{ABF} \times \text{Prior Odds}$$

  where BFDP is the Bayesian False Discovery Probability.
- BFDP depends on the power, through r.
- For implementation, all that we need from the data is the Z-score and the standard error $\sqrt{V}$, or a confidence interval.
- Hence, published results that report confidence intervals can be converted into Bayes factors for interpretation.
- The approximation relies on sample sizes that are not too small, so the normal distribution of the estimator provides a good summary of the information in the data.

Combination of Data Across Studies

- Suppose we wish to combine data from two studies where we assume a common log odds ratio θ.
- The estimates from the two studies are $\widehat{\theta}_1, \widehat{\theta}_2$ with standard errors $\sqrt{V_1}$ and $\sqrt{V_2}$.
- The Bayes factor is

$$\frac{p(\widehat{\theta}_1, \widehat{\theta}_2 | H_0)}{p(\widehat{\theta}_1, \widehat{\theta}_2 | H_1)}.$$

- The approximate Bayes factor is

$$\text{ABF}(\widehat{\theta}_1, \widehat{\theta}_2) = \text{ABF}(\widehat{\theta}_1) \times \text{ABF}(\widehat{\theta}_2 | \widehat{\theta}_1) \qquad (2)$$

  where

$$\text{ABF}(\widehat{\theta}_2 | \widehat{\theta}_1) = \frac{p(\widehat{\theta}_2 | H_0)}{p(\widehat{\theta}_2 | \widehat{\theta}_1, H_1)}$$

  and

$$p(\widehat{\theta}_2 | \widehat{\theta}_1, H_1) = \text{E}_{\theta | \widehat{\theta}_1} \left[p(\widehat{\theta}_2 | \theta) \right]$$

  so that the density is averaged with respect to the posterior for θ.
- Important Point: The Bayes factors are not independent.

Combination of Data Across Studies

- This leads to an approximate Bayes factor (which summarizes the data from the two studies) of

$$\text{ABF}(\widehat{\theta}_1, \widehat{\theta}_2) = \sqrt{\frac{W}{RV_1 V_2}} \exp\left\{ -\frac{1}{2} \left(Z_1^2 RV_2 + 2Z_1 Z_2 R\sqrt{V_1 V_2} + Z_2^2 RV_1 \right) \right\}$$

  where

  - $R = W/(V_1 W + V_2 W + V_1 V_2)$
  - $Z_1 = \frac{\widehat{\theta}_1}{\sqrt{V_1}}$ and
  - $Z_2 = \frac{\widehat{\theta}_2}{\sqrt{V_2}}$ are the usual Z statistics.

- The ABF will be small (evidence for H_1) when the absolute values of Z_1 and Z_2 are large and they are of the same sign.

Stephens (2017) extends the ABF approach in an interesting way.

Combination of Data Across Studies: The General Case

- Suppose we have K studies with estimates $\widehat{\theta}_k$ and asymptotic variances V_k, $k = 1, ..., K$.
- Assume a common underlying parameter θ.
- The Bayes factor is given by

$$
\begin{aligned}
\text{BF}_K &= \frac{p(\widehat{\theta}_1, \ldots, \widehat{\theta}_K | H_0)}{p(\widehat{\theta}_1, \ldots, \widehat{\theta}_K | H_1)} \\
&= \frac{\prod_{k=1}^{K} (2\pi V_k)^{-1/2} \exp\left(-\frac{\widehat{\theta}_k^2}{2V_k}\right)}{\int \prod_{k=1}^{K} (2\pi V_k)^{-1/2} \exp\left(-\frac{(\widehat{\theta}_k^2 - \theta)^2}{2V_k}\right) (2\pi W)^{-1/2} \exp\left(-\frac{\theta^2}{2V_k}\right) d\theta} \\
&= \sqrt{W\left(W^{-1} + \sum_{k=1}^{K} V_k^{-1}\right)} \exp\left[-\frac{1}{2}\left(\sum_{k=1}^{K} \frac{\widehat{\theta}_k}{V_k}\right)^2 \left(W^{-1} + \sum_{k=1}^{K} V_k^{-1}\right)^{-1}\right]
\end{aligned}
$$

Combination of Studies: The General Case

- The posterior is given by

$$\theta|\widehat{\theta}_1, \ldots, \widehat{\theta}_K \sim \mathsf{N}(\mu, \sigma^2)$$

where

$$\begin{aligned} \mu &= \left(\sum_{k=1}^{K} \frac{\widehat{\theta}_k}{V_k} \right) \left(W^{-1} + \sum_{k=1}^{K} V_k^{-1} \right)^{-1} \\ \sigma^2 &= \left(W^{-1} + \sum_{k=1}^{K} V_k^{-1} \right)^{-1} \end{aligned}$$

Example of Combination of Studies in a GWAS

- We illustrate how reported confidence intervals can be converted to Bayesian summaries.
- Frayling *et al.* (2007) report a GWAS for Type II diabetes.
- For SNP rs9939609:

				$\Pr(H_0\|\text{data})$ with prior:	
Stage	Estimate (CI)	p-value	$-\log_{10}$ BF	1/5,000	1/50,000
1st	1.27 (1.16–1.37)	6.4×10^{-10}	7.28	0.00026	0.0026
2nd	1.15 (1.09–1.23)	4.6×10^{-5}	2.72	0.905	0.990
Combined	–	–	13.8	8×10^{-11}	8×10^{-10}

- Combined evidence is stronger than each separately since the point estimates are in agreement.
- For summarizing inference the (5%, 50%, 95%) points for the RR are:

Prior	1.00 (0.67–1.50)
First Stage	1.26 (1.17–1.36)
Combined	1.21 (1.15–1.27)

Conclusions

- Computationally GLMs and GLMMs can now be fitted in a relatively straightforward way.
- `INLA` is very convenient and is being constantly improved.
- As with all analyses, it is crucial to check modeling assumptions (and there are usually more in a Bayesian analysis).
- Markov chain Monte Carlo provides an alternative for computation. Stan, `WinBUGS` and `JAGS` are possibilities.
- The mixture models required specialized code.

References

Blangiardo, M. and Cameletti, M. (2015). *Spatial and Spatio-Temporal Bayesian Models with R-INLA*. John Wiley and Sons.

Connelly, C., Wakefield, J., and Akey, J. (2014). Evolution and architecture of chromatin accessibility and function in yeast. *PLoS Genetics*. To appear.

Frayling, T., Timpson, N., Weedon, M., Zeggini, E., Freathy, R., and et al., C. L. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, **316**, 889–894.

Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2018). *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. Chapman and Hall/CRC.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, **71**, 319–392.

Skelly, D., Johansson, M., Madeoy, J., Wakefield, J., and Akey, J. (2011). A powerful and flexible statistical framework for testing hypothesis of allele-specific gene expression from RNA-Seq data. *Genome Research*, **21**, 1728–1737.

Stephens, M. (2017). False discovery rates: a new deal. *Biostatistics*, **18**, 275–294.

Wakefield, J. (2007). A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *American Journal of Human Genetics*, **81**, 208–227.

Wakefield, J. (2009). Bayes factors for genome-wide association studies: comparison with *p*-values. *Genetic Epidemiology*, **33**, 79–86.

Wang, X., Yue, Y., and Faraway, J. J. (2018). *Bayesian Regression Modeling with INLA*. Chapman and Hall/CRC.

Appendix: Hierarchical Modeling of Allele-Specific Expression Data

Specifics of ASE Experiment

Details of the data:

- Two “individuals” from genetically divergent yeast strains, BY and RM, are mated to produce a diploid hybrid.
- Three replicate experiments: same individuals, but separate samples of cells.
- Two technologies: Illumina and ABI SOLiD. Each of a few trillion cells are processed.
- Pre- and post-processing steps are followed by fragmentation to give millions of 200–400 base pair long molecules, with short reads obtained by sequencing.
- Strict criteria to call each read as a match are used, to reduce read-mapping bias.
- Data from 25,652 SNPs within 4,844 genes.

Allele Specific Expression via RNA-Seq

Additional data:

- Genomic DNA is sequenced in the diploid hybrid, which has one copy of each gene from BY and from RM.
- The only difference between the genomic DNA and the main experiment is that we expect the genomic DNA to always be present 50:50 (one copy each of BY and RM), whereas for the main experiment it is only 50:50 if there is no ASE.
- For both genomic DNA and RNA we obtain counts at SNPs, at each of BY and RM.

Genomic DNA and RNA

To clarify: we have the RNA measurements which are of primary interest and the genomic DNA which is like a control.

For genomic DNA, every cell will have one copy of each locus from “Mom” and one from “Dad” (assuming diploid).

If we could sequence the contents of the cell perfectly, then across a population of billions/trillions of cells we should see exactly 50% Mom alleles and 50% Dad alleles.

In reality there will be sampling noise due to differences in things like amplification efficiency during the sequencing library preparation process.

Genomic DNA and RNA

For the RNA, we are measuring transcribed molecules so there might be *x* copies of Moms locus and *y* copies of Dads.

The key is that the same sampling noise present in the DNA data is also present in the RNA data, because it undergoes the same sequencing library preparation process.

Actually there is one additional step for RNA which is converting to cDNA. This may add some noise as well but it is probably less than the combined effects of the other steps in the process.

So in this sense, using the DNA to model the sampling noise in molecule counts can serve as a useful baseline for calibrating our expectations of how much deviation in 50:50 we need to see at the RNA level before we believe the difference to be interesting.

Cartoon of the Experiment

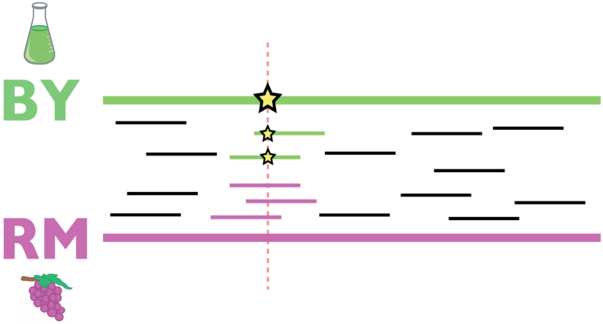


Figure 15: Mapping of RNA short reads to BY and RM.

Statistical Problem

- Aim of the Experiment: Estimate the proportion of genes that display ASE.
- Let p be the probability of a map to BY at a particular SNP.
- Additionally, we would like to classify genes into:
  - Genes that do not show ASE.
  - Genes that show:
    - Constant ASE across SNPs.
    - Variable ASE across SNPs, i.e. p varies within gene.

  Subsequently, we will examine genes displaying ASE to investigate the mechanism.
- A hierarchical model is feasible since we have within gene and between gene variability.
- Further, a mixture model is suggested, with a mixture of genes that do not display ASE (so there p's are 0.5) and that do display ASE.

Summaries for ASE Data

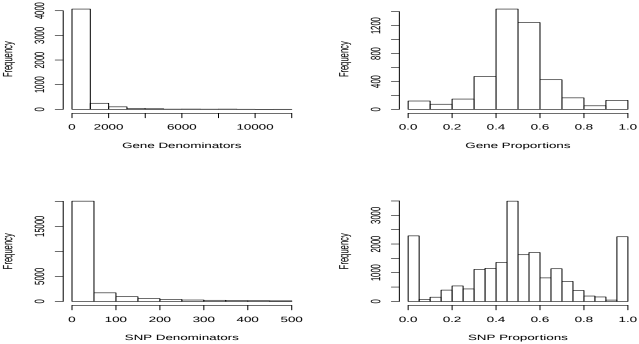


Figure 16: Summaries for RNA BY/RM yeast data; note that 739 SNP denominators are >500 and are not plotted.

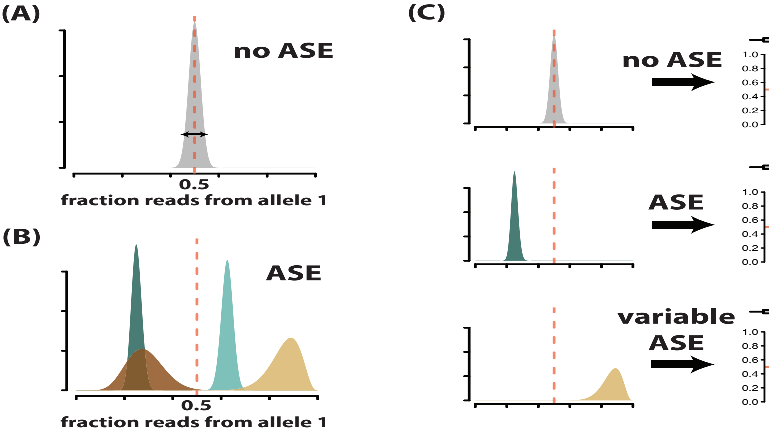


Figure 17: Schematic of the hierarchical model.

Approach to Modelling RNASeq Data

Overview, three models fitted:

1. Model 1: Two component mixture model to filter out aberrant SNPs using genomic DNA data.
2. Model 2: Using the filtered genomic DNA data, fit a hierarchical SNP within gene model, to determine the “null” distribution of counts.
   Specifically: “wobble” in p about 0.5, and SNP “wobble” in p within genes.
   Absence of ASE is not experimentally equivalent to $Y_i \sim \text{binomial}(N_i, p = 0.5)$ because of the steps involved in the experiment.
3. Model 3: For the RNA Seq data develop a two-component mixture model where each gene either displays no ASE, or ASE, with null component determined from the analysis of the genomic DNA data (Model 2).

Model 1: Filtering Model for Genomic DNA

Two-component mixture model for SNPs:

1. Majority of SNP counts arise from a beta-binomial distribution with p “close to” 0.5 (component 1).
2. Minority of SNP counts arise from a beta-binomial distribution with p “not close to” 0.5 due to sequencing bias at these SNPs (component 2).

- Data: y_j and N_j are counts at SNP j for $j = 1, \ldots, m$ SNPs.
- Note: Ignores gene information – don’t want to impose too much structure at this point.
- SNPs that are more likely to arise from component 2 are then removed from further analyses.

Filtering Model for Genomic DNA

- *Stage 1: SNP Count Likelihood:*

$$y_j|p_j \sim \text{binomial}(N_j, p_j), \quad j = 1, ..., N.$$

- *Stage 2: Between-SNP Prior:*

$$p_j|a, b, c, \pi_0 = \begin{cases} \text{beta}(a, a) & \text{with probability } \pi_0 \\ \text{beta}(b, c) & \text{with probability } 1 - \pi_0 \end{cases}$$

- *Stage 3: Hyperpriors:* Constrain $b < 1$, $c < 1$ to give U-shaped beta distribution.

$$\begin{aligned} a &\sim \text{lognormal}(4.3, 1.8)^{\star} \\ b &\sim \text{uniform}(0, 1) \\ c &\sim \text{uniform}(0, 1) \\ \pi_0 &\sim \text{uniform}(0, 1) \end{aligned}$$

*80% interval for $p : [0.43, 0.57]$. Separate a, b, c, π_0 for each technology.

Implementation for Genomic DNA

- Integrate p_j from model to give:

$$\begin{aligned} y_j | a, b, c, \pi_0 &\sim \pi_0 \times \text{beta-binomial}(N_j, a, a) \\ &+ (1 - \pi_0) \times \text{beta-binomial}(N_j, b, c). \end{aligned}$$

- This is a mixture of two distributions:
  1. The first distribution is for the majority of signals close to 0.5. The size of *a* denotes how close is close.
  2. The second distribution is for the minority of aberrant SNPs.

Implementation for Genomic DNA

- Likelihood:

$$\Pr(\boldsymbol{y}|a,b,c,\pi_0) = \prod_{j=1}^{N} \left(\begin{array}{c} N_j \\ Y_j \end{array} \right) \left\{ \pi_0 \frac{\Gamma(2a)}{\Gamma(a)^2} \frac{\Gamma(y_j+a)\Gamma(N_j-y_j+a)}{\Gamma(N_j+2a)} \right.$$

$$\left. + \quad (1-\pi_0)\frac{\Gamma(b+c)}{\Gamma(b)\Gamma(c)} \frac{\Gamma(y_j+b)\Gamma(N_j-y_j+c)}{\Gamma(N_j+b+c)} \right\}$$

- Posterior:

$$p(a,b,c,\pi_0|\boldsymbol{y}) \propto \Pr(\boldsymbol{y}|a,b,c,\pi_0) \times p(a)p(b)p(c)p(\pi_0).$$

- *Implementation:* Markov chain Monte Carlo.
  - Recall: Sequencing bias lead to aberrant SNPs, and these errors are likely to be repeated in the main experiment.
  - SNPs falling in the second mixture component were removed from further analyses.

Posterior Distributions

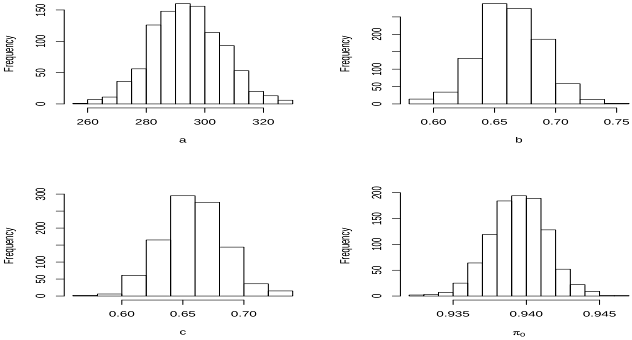


Figure 18: Posteriors for genomic filtering model for Illumina platform.

Posterior Filter

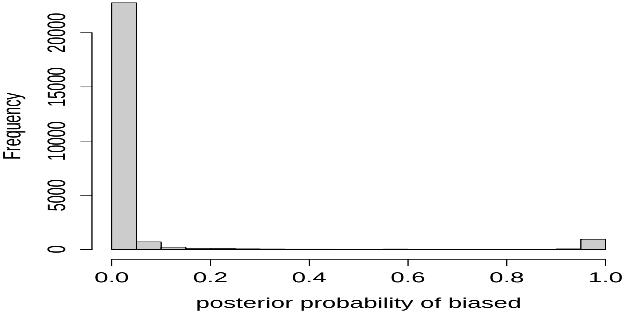


Figure 19: Posterior probabilities of biased genomic DNA SNPs: 1,295 removed from 25,262.

Effect of Filtering

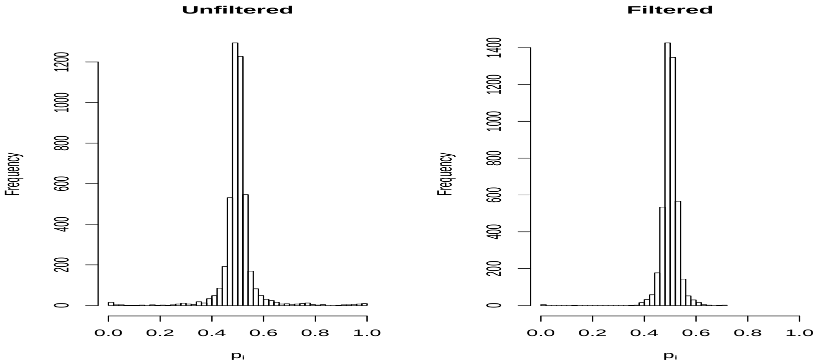


Figure 20: Original and filtered data, for Illumina platform.

Model 2: Calibration Model for Genomic Data

- With aberrant SNPs removed, the next step is to calibrate the null component.

- *Stage 1: Within-Gene Likelihood:*

  $$Y_{ij} | p_{ij} \sim \text{binomial}(N_{ij}, p_{ij}).$$

  where p_{ij} is the probability of an outcome from the first genetic background.

- *Stage 2: Within-Gene Prior:*

  $$p_{ij} | \alpha_i, \beta_i \sim \text{beta}(\alpha_i, \beta_i)$$

  so that α_i, β_i determine the distribution of variants within gene i.

Calibration Model for Genomic Data

- α_i and β_i are not straightforward to interpret.
- We reparameterize $(\alpha_i, \beta_i) \rightarrow (p_i, e_i)$ with mean and dispersion parameters (recall $\alpha_i + \beta_i$ is a prior sample size):

$$p_i = \frac{\alpha_i}{\alpha_i + \beta_i}$$
$$e_i = \frac{1}{1 + \alpha_i + \beta_i}$$

- Moments of ASE parameters:

$$\begin{aligned} E[p_{ij}|p_i, e_i] &= p_i \\ var(p_{ij}|p_i, e_i) &= p_i(1 - p_i)e_i \end{aligned}$$

- Moments of data:

$$\begin{aligned} E[Y_{ij}|p_i, e_i] &= N_{ij}p_i \\ var(Y_{ij}|p_i, e_i) &= N_{ij}p_i(1 - p_i)\left[1 + (N_{ij} - 1)e_i\right] \end{aligned}$$

  - As $e_i \rightarrow 0$ we approach the binomial model.
  - As $e_i \rightarrow 1$ we have more overdispersion (variability within gene).

Calibration Model for Genomic data

- *Stage 3: Within-Gene Likelihood:*

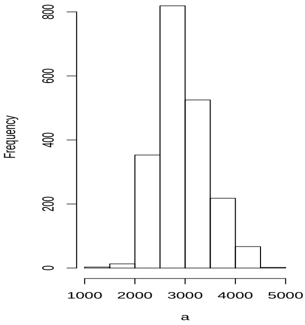
$$\begin{aligned} p_i | a &\sim \text{beta}(a, a) \\ e_i | d &\sim \text{beta}(1, d) \end{aligned}$$

  Note: prior on within-gene dispersion is monotonic decreasing from 0 (corresponding to no variability).

- *Stage 4: Hyperpriors:* Require priors on $a > 0, d > 0$.
- We take

$$\begin{aligned} a &\sim \text{lognormal}(4.3, 1.8) \\ d &\sim \text{exponential}(0.0001) \end{aligned}$$

- The latter prior determines the within-gene variability within-gene variability in genomic DNA – chosen by examination of resultant p_{ij}'s.
- Separate a, d for each technology.



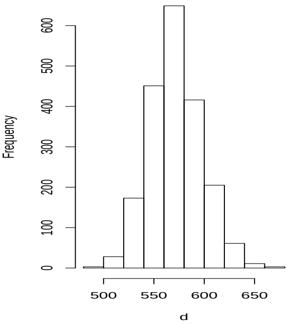


Figure 21: Posteriors for the RNA-Seq data, Illumina platform.

Model 3: Model for RNA-Seq Data

- Data are modeled as a two-component mixture: the first "null" component having a known distribution, from the genomic DNA analysis on the filtered data.
- *Stage 1: Within-Gene Likelihood:*

$$Y_{ij} | p_{ij} \sim \text{binomial}(N_{ij}, p_{ij}).$$

  where p_{ij} is the probability of an outcome from the first genetic background.
- *Stage 2: Within-Gene Prior:*

$$p_{ij} | \alpha_i, \beta_i \sim \text{beta}(\alpha_i, \beta_i)$$

  so that α_i, β_i determine the distribution of variants within gene i.
- *Stage 3: Between-Gene Prior:* We again reparameterize $(\alpha_i, \beta_i) \to (p_i, e_i)$:

$$p_i, e_i | f, g, h, \pi_0 \sim \begin{cases} \text{beta}(\widehat{a}, \widehat{a}) \times \text{beta}(1, \widehat{d}) & \text{with probability } \pi_0 \\ \text{beta}(f, g) \times \text{beta}(1, h) & \text{with probability } 1 - \pi_0 \end{cases}$$

  with $\widehat{a}, \widehat{d}$ from genomic DNA analysis.

Stage 4: Hyperpriors: Require priors on $\pi_0, f > 0, g > 0, h > 0$.

- Uniform prior on π_0.
- f and g describe beta distribution of p_i for genes displaying ASE – want this distribution to be centered on symmetry.
- Reparameterize as

  $$q = \frac{f}{f+g} \quad r = \frac{1}{1+f+g}$$

  so that $E[p_i] = q$, $var(p_i) = q(1-q)r$.
- Through experimentation:

  $$q \quad \sim \quad \text{beta}(100, 100) \quad r \sim \text{beta}(1, 20)$$

- For h, the distribution of within-gene variability in ASE:

  $$h \sim \text{exponential}(0.03).$$

- Separate f, g, h for each technology.

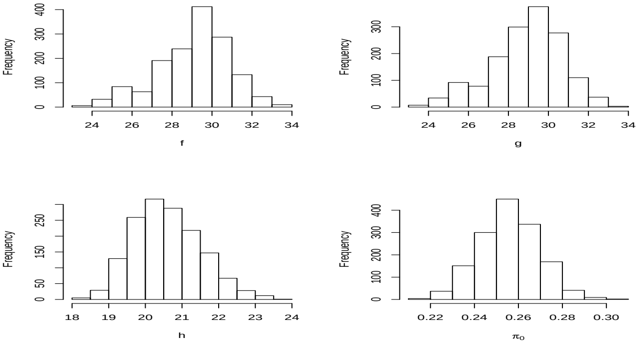


Figure 22: Posteriors for the RNA-Seq data, Illumina platform.

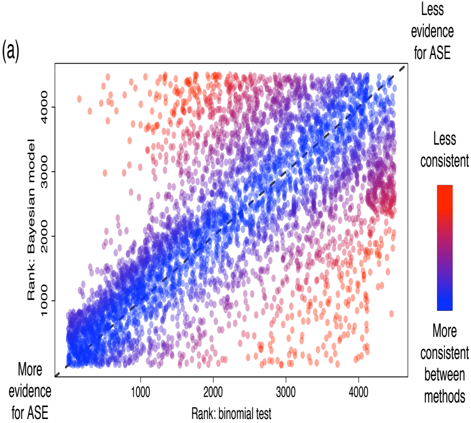


Figure 23: Comparison of rankings from binomial test and hierarchical model.

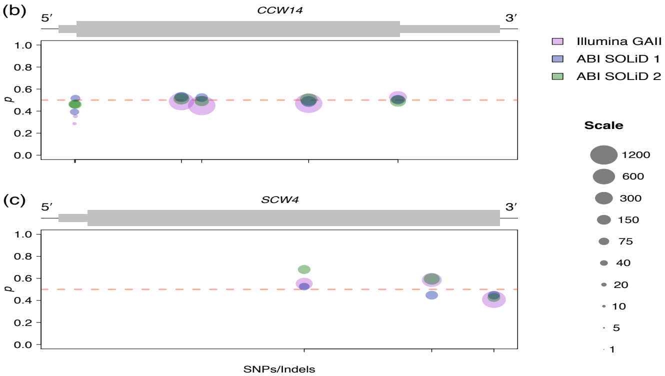


Figure 24: Examples of opposite conclusions: In (b) the *p*-value said ASE and Bayes not (large sample size, Bayes allows wobble). In (c) the *p*-value said no ASE, Bayes analysis yes.

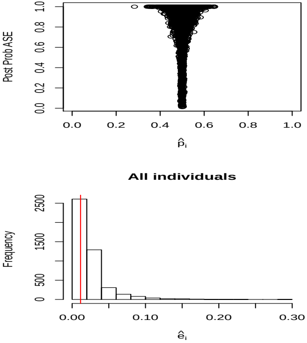


Figure 25: Between-gene variability p_i and within-gene variability e_i.

Varying ASE within genes

- One mechanism: Imagine a gene with an exon and an intron, and that we have SNPs in both.
- At each exonic SNP we see approximately the same number of BY and RM reads.
- Now suppose the intron is not spliced out for the BY allele, but it is spliced out efficiently for the RM allele. At each intronic SNP we will still see the same number of BY reads as in the exon (everything else being equal), but approximately 0 RM reads, leading to variable ASE across the gene
- In the figure: The “thin” part of the gene (YML024W) is an intron, while the “thick” part is an exon.
- For the RM allele (magenta) the intron is not spliced out, while it is mostly spliced out in the BY allele (green).

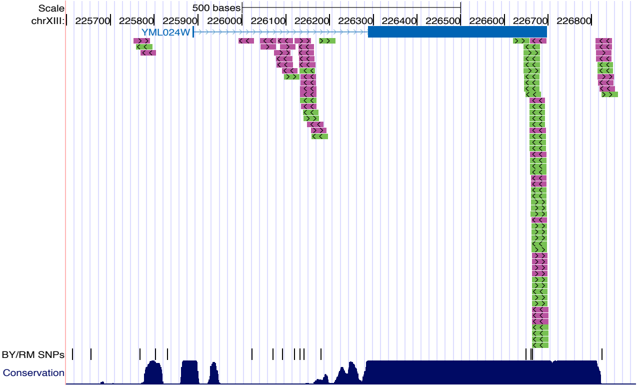


Figure 26: Example of a gene displaying variable ASE within a gene. Green = RM, magenta = BY.

Conclusions for Mixture Model

- For the ASE data we used the DNA experiment to calibrate the prior.
- More details of this experiment and the model can be found in Skelly *et al.* (2011).
- Implementation was via Markov chain Monte Carlo, but we had to write our own code.