Statistical Inference

Lecture 11b

ANU - RSFAS

Last Updated: Wed May 17 16:05:17 2017

The One Factor Gaussian Model

- In Tutorial 6, you examined the following data:
    - Three groups of six guinea pigs were each randomly injected, respectively, with 0.5 mg, 1.0 mg, and
      1.5 mg of a new tranquilizer. The following data present the number of minutes it took them to fall asleep:

0.5 mg	21	23	19	24	25	23
1.0 mg	19	21	20	18	22	20
1.5 mg	15	10	13	14	11	15

Fixed Effects

- For these data you the following model was suggested:

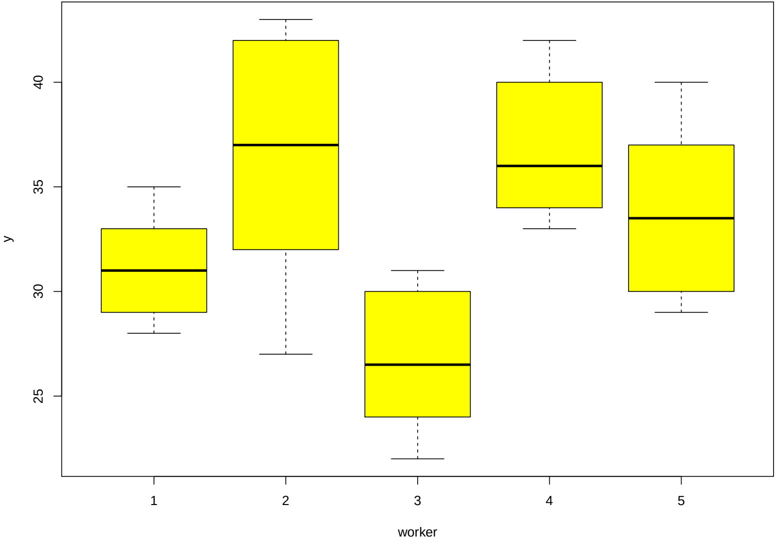
$$\begin{aligned} Y_{ij} &= \mu_i + \epsilon_{ij} \\ \epsilon_{ij} &\sim iid \quad \mathrm{normal}(0, \sigma^2), \quad j = 1, \ldots, m = 6; \; i = 1, \ldots, n = 3. \end{aligned}$$

- You we able to determine MLEs.

Random Effects - Assessable only for STAT4027/STAT8027

- The manager of an industrial plant wanted to determine whether workers with the same skill level have any discernible differences in the number of units of an automobile part that are manufactured during a fixed period of time. Five workers were randomly selected and the number of units produced by each worker for six equal length periods was recorded:

worker 1	30	32	28	33	35	29
worker 2	27	40	43	34	32	42
worker 3	24	22	31	30	27	26
worker 4	42	34	37	35	33	40
worker 5	33	29	40	30	34	37



y
25
30
35
40
1
2
3
4
5
worker

Random Effects

- In our example, the workers are randomly selected from a population of workers, and thus a random effects model seems appropriate for the data. Let's consider the following model for the data:

$$
\begin{aligned}
y_{ij}|a_i &\overset{\text{indep.}}{\sim} \text{normal}(\mu + a_i, \sigma^2) \\
a_i &\overset{\text{iid}}{\sim} \text{normal}(0, \sigma_a^2) \\
& i = 1, \ldots, m = 5; j = 1, \ldots, n_i = 6 \\
& N = m \times n = 30
\end{aligned}
$$

- Random effects:
    - If it is reasonable to assume that the levels of a factor come from a probability distribution (randomly sampled from a population), then use a random effect for that factor - traditional view.
    - Random effects are used to create for dependencies for data or factors that may not be readily seen as being sampled from a larger population.
    - Random effects can be seen as smoothed versions of fixed effects.
    - Random effects do not lead to a loss of degrees-of-freedom as with fixed effects.

Is this Bayesian?

- While it might appear that we are being Bayesian to put a distribution on a_i, here a distribution is placed on a tangible population (which Bayesians also do).
- Here we just have a *hierarchical model* that can be estimated via many approaches.
- A Bayesian approach would further place distributions on the unknown parameters $\mu, \sigma^2, \sigma_a^2$ and then inference would proceed through the use of Bayes' Theorem and Markov chain Monte Carlo
(e.g. Metropolis-Hastings is and MCMC algorithm).

Random Effects

- A nice property of the random effects model is that data within a worker are now correlated!

$$\begin{aligned} Cov\left(Y_{i,j}, Y_{i,l}\right) &= Cov\left[E\left(Y_{i,j}|a_i\right), E\left(Y_{i,l}|a_i\right)\right] + E\left[Cov\left(Y_{i,j}, Y_{i,l}|a_i\right)\right] \\ &= Cov\left[\mu + a_i, \mu + a_i\right] + 0 \\ &= \sigma_a^2 \end{aligned}$$

$$\begin{aligned} V\left(Y_{i,j}\right) &= V\left[E\left(Y_{i,j}|a_i\right)\right] + E\left[V\left(Y_{i,j}|a_i\right)\right] \\ &= V\left[\mu + a_i\right] + E\left[\sigma^2\right] \\ &= \sigma_a^2 + \sigma^2 \end{aligned}$$

Random Effects

$$
\begin{aligned}
Cor\left(Y_{i,j}, Y_{i,l}\right) &= \frac{Cov\left(Y_{i,j}, Y_{i,l}\right)}{\sqrt{V(Y_{i,j})}\sqrt{V(Y_{i,l})}} \\
&= \frac{\sigma_a^2}{\sqrt{\sigma_a^2+\sigma^2}\sqrt{\sigma_a^2+\sigma^2}} \\
&= \frac{\sigma_a^2}{\sigma_a^2+\sigma^2}
\end{aligned}
$$

Random Effects

- Also note:

$$
\begin{aligned}
E\left(Y_{i,j}\right) &= E\left[E\left(Y_{i,j} \middle| a_i\right)\right] \\
&= E\left[\mu + a_i\right] \\
&= \mu
\end{aligned}
$$

Random Effects

- These results suggest that we can write the model as follows:

$$y_i \overset{\text{indep.}}{\sim} \text{multivariate normal}_{n_i}(\mu 1_{n_i}, V_i),$$

where 1_{n_i} is a vector of ones of length n_i. Also,

$$V_i = \begin{bmatrix} \sigma^2 + \sigma_a^2 & \sigma_a^2 & \cdots & \sigma_a^2 \\ \sigma_a^2 & \sigma^2 + \sigma_a^2 & \cdots & \sigma_a^2 \\ \vdots & & & \vdots \\ \sigma_a^2 & \cdots & \cdots & \sigma^2 + \sigma_a^2 \end{bmatrix}$$

Random Effects

- What did we really do? We integrated out the random variable a_i!

$$L(\mu, \sigma_a^2, \sigma^2; \boldsymbol{y}) = \prod_{i=1}^{m} \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^{n_i} f_Y(y_{i,j}; \mu, a_i, \sigma^2) \right\} f_a(a_i; 0, \sigma_a^2) da_i.$$

- Why did we integrate out the random effects? Because the frequentist philosophy states that parameters are **fixed** but unknown quantities. The a_i's are random variables and thus are not fixed!

Random Effects

- If we do the integration and get the log likelihood we have:

$$\begin{aligned} l(\mu, \sigma^2, \sigma_a^2; y) \quad &= \quad -\frac{N}{2} log(2\pi) - \frac{1}{2} \sum_i^m log(\sigma^2 + n_i \sigma_a^2) - \frac{1}{2}(N-m) log(\sigma^2) \\ &\quad -\frac{1}{2\sigma^2} \sum_i \sum_j (y_{ij} - \mu)^2 + \frac{\sigma_a^2}{2\sigma^2} \sum_i^m \frac{(y_{i.} - n_i \mu)^2}{\sigma_2 + n_i \sigma_a^2} \end{aligned}$$

Random Effects - Balanced Data Case

- Now if we have balanced data so that $n_i = n \ \ \forall i$ then:

$$\begin{aligned} l(\mu, \sigma^2, \sigma_a^2; y) &= -\frac{N}{2} log(2\pi) - \frac{1}{2} m log(\sigma^2 + n\sigma_a^2) - \frac{1}{2}(N - m) log(\sigma^2) \\ &\quad - \frac{1}{2\sigma^2} \sum_i \sum_j (y_{ij} - \mu)^2 + \frac{\sigma_a^2 n^2}{2\sigma^2(\sigma^2 + n\sigma_a^2)} \sum_i^m (\bar{y}_{i.} - \mu)^2 \end{aligned}$$

- Now if we re-write: $y_{ij} - \mu$ as $y_{ij} - \bar{y}_{i.} + \bar{y}_{i.} - \mu$ and simplify, we can show:

$$
\begin{aligned}
l(\mu, \sigma^2, \sigma_a^2; y) &= -\frac{N}{2} log(2\pi) - \frac{1}{2} m log(\lambda) - \frac{1}{2}(N-m) log(\sigma^2) \quad (1) \\
&\quad -\frac{1}{2\sigma^2} SSE - \frac{1}{2\lambda} SSA - \frac{N(\bar{y}_{..} - \mu)^2}{2\lambda}
\end{aligned}
$$

$$
SSE = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2; \quad SSA = \sum_i n(\bar{y}_{i.} - \bar{y}_{..})^2; \quad \lambda = \sigma^2 + n\sigma_a^2
$$

Random Effects - Balanced Data Case - Estimation

- Now that we have a model for the data, we need to determine an approach for parameter estimation. Again we will use maximum-likelihood.

$$
\begin{aligned}
\ell'_\mu &= \frac{N(\bar{y}_{..} - \mu)}{\lambda} \\
\ell'_{\sigma^2} &= -\frac{N - m}{2\sigma^2} + \frac{SSE}{2(\sigma^2)^2} \\
\ell'_\lambda &= -\frac{m}{2\lambda} + \frac{SSA}{2\lambda^2} + \frac{N(\bar{y}_{..} - \mu)}{2\lambda^2}
\end{aligned}
$$

Random Effects - Balanced Data Case - Estimation

- So now we set the equations equal to zero and solve for the three parameters. Here I use will use $(\cdot)$ as I am not sure I have the MLEs:

$$
\begin{aligned}
\dot{\mu} &= \frac{1}{nm}\sum_i\sum_j y_{ij} = \frac{1}{N}\sum_i\sum_j y_{ij} = \bar{y}_{..} \\
\dot{\sigma}^2 &= \frac{SSE}{N-m} = MSE \\
\dot{\lambda} &= \frac{SSA + N(\bar{y}_{..} - \dot{\mu})^2}{m} = \frac{SSA}{m}
\end{aligned}
$$

Random Effects - Balanced Data Case - Estimation

- By the invariance property of MLE's:

$$\sigma_a^2 = f(\lambda, \sigma^2) = \frac{\lambda - \sigma^2}{n}$$

So by the property we can plug in the estimates:

$$\dot{\sigma}_a^2 = \frac{\dot{\lambda} - \dot{\sigma}^2}{n} = \frac{SSA/m - SSE/(N-m)}{n} = \frac{(1 - 1/m)MSA - MSE}{n}$$

Random Effects - Balanced Data Case - Estimation

- We do have to be careful that $\hat{\sigma}_a^2 \geq 0$, so the MLEs are:

$$\hat{\mu} = \frac{1}{nm}\sum_i \sum_j y_{ij} = \frac{1}{N}\sum_i \sum_j y_{ij} = \bar{y}_{..}$$

$$\hat{\sigma}^2 = \begin{cases} MSE & \text{if } \dot{\sigma}_a^2 > 0 \\ SST/N & \text{if } \dot{\sigma}_a^2 \leq 0 \end{cases}$$

$$\hat{\sigma}_a^2 = \begin{cases} \dot{\sigma}^2{}_a = \frac{(1-1/m)MSA - MSE}{n} & \text{if } \dot{\sigma}_a^2 > 0 \\ 0 & \text{if } \dot{\sigma}_a^2 \leq 0 \end{cases}$$

$$SST = \sum_i \sum_j (y_{i,j} - \bar{y}_{.,.})^2$$

Random Effects - Balanced Data Case - Estimation

- Let's see for our data:

```
y1 <- c(30, 32, 28, 33, 35, 29)
y2 <- c(27, 40, 43, 34, 32, 42)
y3 <- c(24, 22, 31, 30, 27, 26)
y4 <- c(42, 34, 37, 35, 33, 40)
y5 <- c(33, 29, 40, 30, 34, 37)


y <- rbind(y1, y2, y3, y4, y5)


## Analytical results for balanced data
SSA <- 0
for(i in 1:m){
SSA <- SSA + n*(mean(y[i,]) - mean(y))^2
}

MSA <- SSA/(m-1)
```

```
SSE <- 0
for(i in 1:m){
for(j in 1:n){
SSE <- SSE + (y[i,j] - mean(y[i,]))^2
}}

MSE <- SSE/(m*(n-1))

##
mu.hat <- mean(y)

if( (1-(1/m))*MSA > MSE){
sigma.sq.hat <- MSE
sigma.sq.a.hat <- ( (1-(1/m))*MSA - MSE)/n
}
```

```r
if((1-(1/m))*MSA <= MSE){
SST <- 0
for(i in 1:m){
for(j in 1:n){
SST <- SST + (y[i,j] - mean(y))^2
}}

sigma.sq.hat <- SST/(m*n)
sigma.sq.a.hat <- 0
}
```

```r
mu.hat
```

```
## [1] 32.96667
```

```r
sigma.sq.hat
```

```
## [1] 17.80667
```

```r
sigma.sq.a.hat
```

```
## [1] 11.02556
```

```r
lambda.hat
```

```
## [1] 83.96
```

- For random effects models, you can use the `lme4` library in R:

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
y <- c(y1, y2, y3, y4, y5)
worker <- as.factor(rep(1:5, each=6))

mod <- lmer(y ~ (1|worker), REML=F)
```

```r
summary(mod)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: y ~ (1 | worker)
##
##      AIC      BIC   logLik deviance df.resid
##    185.3    189.5    -89.6    179.3       27
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0426 -0.6865 -0.1958  0.7731  1.7491
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  worker   (Intercept) 11.03    3.32
##  Residual             17.81    4.22
## Number of obs: 30, groups:  worker, 5
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   32.967      1.673   19.71
```

Random Effects - Balanced Data Case - Estimation

- Let's determine the asymptotic variances:

$$
\begin{aligned}
\ell''_{\mu,\mu} &= -N/\lambda \Rightarrow -E(-N/\lambda) = N/\lambda = N/(\sigma^2 + n\sigma_a^2) \\
\ell''_{\mu,\sigma^2} &= 0 \Rightarrow -E(0) = 0 \\
\ell''_{\mu,\lambda} &= -N\bar{y}_{..}/\lambda^2 + N\mu/\lambda^2 \\
&\Rightarrow -E(-N\bar{y}_{..}/\lambda^2 + N\mu/\lambda^2) = N\mu/\lambda^2 - N\mu/\lambda^2 = 0
\end{aligned}
$$

- We have an immediate result $V(\hat{\mu}) = (\sigma^2 + n\sigma_a^2)/N$ and is uncorrelated with the variance components in the model! This is a general result for linear models!!

$$
\begin{aligned}
\ell''_{\sigma^2,\sigma^2} &= \frac{N-m}{2\sigma^4} - \frac{2SSE}{2\sigma^6} \\
-E\left(\frac{N-m}{2\sigma^4} - \frac{2SSE}{2\sigma^6}\right) &= -\frac{N-m}{2\sigma^4} + 2/(2\sigma^6)E((N-m)MSE) \\
&= -\frac{N-m}{2\sigma^4} + \frac{2(N-m)\sigma^2}{(2\sigma^6)} = \frac{N-m}{2\sigma^4}
\end{aligned}
$$

$$
\ell''_{\sigma^2,\lambda} = 0
$$

$$\begin{aligned}
\ell''_{\lambda,\lambda} &= \frac{m}{2\lambda^2} - \frac{2SSA}{2\lambda^3} - \frac{2N(\bar{y}_{..} - \mu)^2}{2\lambda^3} \\
-E(\cdot) &= -\frac{m}{2\lambda^2} + \frac{2E(SSA)}{2\lambda^3} + \frac{2NE((\bar{y}_{..} - \mu)^2)}{2\lambda^3} \\
-E(\cdot) &= -\frac{m}{2\lambda^2} + \frac{2(m-1)\lambda}{2\lambda^3} + \frac{2N}{2\lambda^3} \times \frac{m(n\sigma^2 + n^2\sigma_a^2)}{m^2n^2} = \frac{m}{2\lambda^2}
\end{aligned}$$

Note: $V(\bar{y}_{..}) = \frac{m(n\sigma^2+n^2\sigma_a^2)}{m^2n^2}$.

This leads to the asymptotic variance:

$$V\begin{pmatrix} \hat{\sigma}^2 \\ \hat{\lambda} \end{pmatrix} = I(\theta)^{-1} = \begin{pmatrix} 2\sigma^4/(N-m) & 0 \\ 0 & 2\lambda^2/m \end{pmatrix}$$

Now:

$$\hat{\sigma}_a^2 = \frac{\hat{\lambda} - \hat{\sigma}^2}{n}$$

$$\begin{aligned} V(\hat{\sigma}_a^2) &= (1/n^2)V(\hat{\lambda}) + (1/n^2)V(\hat{\sigma}^2) - 2Cov((1/n)\hat{\lambda}, (1/n)\hat{\sigma}^2) \\ &= (2\lambda^2/(mn^2)) + (2\sigma^4/(n^2(N-m))) - 0 \end{aligned}$$

$$\begin{aligned} Cov(\hat{\sigma}_a^2, \hat{\sigma}^2) &= Cov(1/n(\hat{\lambda} - \hat{\sigma^2}), \hat{\sigma}^2) \\ &= Cov((1/n)\hat{\lambda}, \hat{\sigma}^2) - Cov((1/n)\hat{\sigma}^2, \hat{\sigma}^2) \\ &= 0 - 2\sigma^4/(n(N-m)) \end{aligned}$$

$$V\begin{pmatrix} \hat{\sigma}^2 \\ \hat{\sigma_a^2} \end{pmatrix} = I(\theta)^{-1} = \begin{pmatrix} 2\sigma^4/(N-m) & -2\sigma^4/(n(N-m)) \\ -2\sigma^4/(n(N-m)) & 2\lambda^2/(mn^2)) + (2\sigma^4/(n^2(N-m))) \end{pmatrix}$$

In order to get estimates for the variance-covariance (VCV) matrix, we must replace the true parameter values with their estimates:

$$V\begin{pmatrix} \hat{\sigma}^2 \\ \hat{\sigma_a^2} \end{pmatrix} = I(\theta)^{-1} = \begin{pmatrix} 2\hat{\sigma}^4/(N-m) & -2\hat{\sigma}^4/(n(N-m)) \\ -2\hat{\sigma}^4/(n(N-m)) & 2\hat{\lambda}^2/(mn^2) + (2\hat{\sigma}^4/(n^2(N-m))) \end{pmatrix}$$

```r
## Assyptotic variances
var.mu.hat <- (sigma.sq.hat + n*sigma.sq.a.hat)/(m*n)
var.sigma.sq.hat <- 2*sigma.sq.hat^2/(m*(n-1))
var.sigma.sq.a.hat <- 2*sigma.sq.hat^2*( (1/n^2)*
( (lambda.hat^2/sigma.sq.hat^2)/m + 1/(m*(n-1))))

##
var.mu.hat
```

```
## [1] 2.798667
```

```r
var.sigma.sq.hat
```

```
## [1] 25.36619
```

```r
var.sigma.sq.a.hat
```

```
## [1] 79.02997
```

Let's use `optim()`

```
y1 <- c(30, 32, 28, 33, 35, 29)
y2 <- c(27, 40, 43, 34, 32, 42)
y3 <- c(24, 22, 31, 30, 27, 26)
y4 <- c(42, 34, 37, 35, 33, 40)
y5 <- c(33, 29, 40, 30, 34, 37)


y <- rbind(y1, y2, y3, y4, y5)
```

```r
log.likelihood <- function(theta){

    mu <- theta[1]
    sigma.sq <- theta[2]
    sigma.sq.a <- theta[3]

    # n.i is the same for each i
    one <- matrix(1, nrow=n, ncol=1)
    Ident <- diag(n)
    J <- matrix(1, nrow=n, ncol=n)

    V.inv <- (1/sigma.sq)*(Ident - (sigma.sq.a/(sigma.sq + n*sigma.sq.a))*J)
    det.V <- (sigma.sq + n*sigma.sq.a)*(sigma.sq^(n-1))

    ll <- 0
    for(i in 1:m){
    ll <- ll + -(n/2)*log(2*pi) - (1/2)*log(det.V) +
        ((-1/2)*t(y[i,]-mu*one)%*%V.inv%*%(y[i,]-mu*one))
    }

   return(ll)
}
```

```r
## Let's test the function
y.bar <- mean(y)
s.sq <- var(c(y))
s.sq.a <- var(apply(y, 1, mean))

theta <- c(y.bar, s.sq, s.sq.a)
log.likelihood(theta)
```

```
##             [,1]
## [1,] -91.28965
```

```r
start <- c(y.bar, 10, 10)
out <- optim(start, log.likelihood, hessian=TRUE,
             control=list(fnscale=-1), method="BFGS")
out
```

```
## $par
## [1] 32.96667 17.80013 11.02457
##
## $value
## [1] -89.63867
##
## $counts
## function gradient
##       13       11
##
## $convergence
## [1] 0
##
## $message
## NULL
##
## $hessian
##               [,1]         [,2]          [,3]
## [1,] -3.573660e-01  0.000000000 -3.552714e-09
## [2,]  0.000000e+00 -0.039835335 -2.129145e-03
## [3,] -3.552714e-09 -0.002129145 -1.277484e-02
```

```r
ll.max <- out$value

assym.vcv <- solve(-out$hessian)
assym.vcv
```

```
##               [,1]          [,2]          [,3]
## [1,]  2.798251e+00  4.196761e-08 -7.851950e-07
## [2,]  4.196761e-08  2.532897e+01 -4.221505e+00
## [3,] -7.851950e-07 -4.221505e+00  7.898245e+01
```

```r
diag(assym.vcv)
```

```
## [1]  2.798251 25.328975 78.982448
```

These do match with the analytical results quite well! So for 95% asymptotic confidence intervals `qnorm(0.975) = 1.96`:

$$\begin{aligned} \mu : 32.97 &\pm 1.96 \times \sqrt{2.798} \\ \sigma^2 : 17.81 &\pm 1.96 \times \sqrt{25.329} \\ \sigma_a^2 : 11.03 &\pm 1.96 \times \sqrt{78.98} \end{aligned}$$

Unbalaced Data

- With unbalanced data, we are back to this equation for the log likelihood:

$$
\begin{aligned}
l(\mu, \sigma^2, \sigma_a^2; y) &= -\frac{N}{2} log(2\pi) - \frac{1}{2} \sum_i^m log(\sigma^2 + n_i \sigma_a^2) - \frac{1}{2}(N - m) log(\sigma^2) \\
&\quad -\frac{1}{2\sigma^2} \sum_i \sum_j (y_{ij} - \mu)^2 + \frac{\sigma_a^2}{2\sigma^2} \sum_i^m \frac{(y_{i.} - n_i \mu)^2}{\sigma_2 + n_i \sigma_a^2}
\end{aligned}
$$

Unbalaced Data

- We can show, similarly to what we did before:

$$
\begin{aligned}
l(\mu, \sigma^2, \sigma_a^2; y) &= -\frac{N}{2} log(2\pi) - \frac{1}{2}\sum_i^m log(\sigma^2 + n_i\sigma_a^2) - \frac{1}{2}(N-m)log(\sigma^2) \\
&\quad -\frac{1}{2\sigma^2} SSE - \sum_i \frac{n_i(\bar{y}_{i.} - \mu)^2}{2(\sigma_2 + n_i\sigma_a^2)}
\end{aligned}
$$

- So now we will differentiate the log likelihood wrt $\mu, \sigma^2, \sigma_a^2$. Note $\lambda_i = \sigma^2 + n_i\sigma_a^2$.

Unbalaced Data

$$l_\mu = \sum_i \frac{n_i(\bar{y}_{i.} - \mu)}{2\lambda_i} \tag{2}$$

$$l_{\sigma^2} = -\frac{N - m}{2\sigma^2} - \frac{1}{2}\sum_i \frac{1}{\lambda_i} + \frac{SSE}{2(\sigma^2)^2} + \sum_i \frac{n_i(\bar{y}_{i.} - \mu)^2}{2\lambda_i^2} \tag{3}$$

$$l_{\sigma_a^2} = -\frac{1}{2}\sum_i \frac{1}{\lambda_i} + \sum_i \frac{n_i^2(\bar{y}_{i.} - \mu)^2}{2\lambda_i^2} \tag{4}$$

Unbalaced Data

- Now set the equations equal to zero and solve:

$$\dot{\mu} = \sum_i \frac{n_i \bar{y}_{i.}}{\dot{\sigma}_2 + n_i \dot{\sigma}_a^2} \Big/ \sum_i \frac{n_i}{\dot{\sigma}_2 + n_i \dot{\sigma}_a^2}$$

$$-\frac{N-m}{2\sigma^2} - \frac{1}{2}\sum_i \frac{1}{\lambda_i} + \frac{SSE}{2(\sigma^2)^2} + \sum_i \frac{n_i(\bar{y}_{i.} - \mu)^2}{2\lambda_i^2} = 0$$

$$\sum_i \frac{n_i^2(\bar{y}_{i.} - \mu)^2}{2\lambda_i^2} = \frac{1}{2}\sum_i \frac{1}{\lambda_i}$$

Unbalaced Data

- There is no closed form analytical solution for the estimators! Thus we must use a computational tool to get $\dot{\mu}, \dot{\sigma^2}, \dot{\sigma^2_a}$. We can use:
  - `optim()` as we have been doing;
  - the *BB* library in R. This library solves systems of non-linear equations. You can try this out.
- After finding $\dot{\mu}, \dot{\sigma^2}, \dot{\sigma^2_a}$ numerically:
  - when $\dot{\sigma}^2_a > 0 \Rightarrow \hat{\mu} = \dot{\mu}, \hat{\sigma^2} = \dot{\sigma^2}, \hat{\sigma^2_a} = \dot{\sigma^2_a}$;
  - when $\dot{\sigma}^2_a \leq 0 \Rightarrow \hat{\mu} = \dot{\mu}, \hat{\sigma^2} = SST/N, \hat{\sigma^2_a} = 0$.
- While we can't determine the MLEs analytically you still are able to determine the asymptotic VCV through the inverse of the negative expected value of the Hessian matrix! However, since we must use a computational tool, then we can retrieve the estimated asymptotic VCV!

Generalized Linear Models with Random Effects

- You may argue that the *worker data*, would more naturally be modeled using a Poisson distribution (as we are examining counts over a period of time).
- This leads to the model:

$$
\begin{aligned}
y_{ij}|a_i &\overset{\text{indep.}}{\sim} \text{Poisson}(\lambda_i) \\
log(\lambda_i) &= \mu + a_i \\
a_i &\overset{\text{iid}}{\sim} \text{normal}(0, \sigma_a^2) \\
& i = 1, \ldots, m = 5; j = 1, \ldots, n_i = 6 \\
& N = m \times n = 30
\end{aligned}
$$

- To conduct inference via MLE we must integrate out the random variable (a_i):

$$L(\mu, \sigma_a^2; \boldsymbol{y}) = \prod_{i=1}^{m} \int_{a_i} \prod_{j=1}^{n} p(y_{i,j}|a_i)p(a_i)d_{a_i}$$

- This integration must typically be done via numerical methods (See Computational Statistics Chapter 5).
- Another approach (actually used in the lme4 package) is Penalized Iteratively Reweighted Least Squares (PIRLS)
- Once MLEs are obtained, inference proceeds as usual.

```r
library(lme4)
y <- c(y1, y2, y3, y4, y5)
worker <- as.factor(rep(1:5, each=6))

mod <- glmer(y ~ (1|worker), family="poisson")
```

```r
summary(mod)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: y ~ (1 | worker)
##
##      AIC      BIC   logLik deviance df.resid
##    187.0    189.8    -91.5    183.0       28
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.35605 -0.50787 -0.06808  0.51230  1.34747
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  worker (Intercept) 0.008239 0.09077
## Number of obs: 30, groups:  worker, 5
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.49142    0.05166   67.58   <2e-16 ***
## ---
```