Generalized linear mixed models

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

April 16, 2012

Outline for today

- cherries data example
- logistic model with random effects
- generalized linear mixed models
- the likelihood function of a GLMM
- Laplace approximation and PQL

Cherry flower data

$$Y_{stBbf} = \begin{cases} 1 \text{ if bud is dead} \\ 0 \text{ if bud is alive} \end{cases}$$

$s = 1, \dots, 5$ STOCK (variety), $t = 1, \dots, 4$ TREE, $B = 1, 2, 3$ BRANCH, $b = 1, 2, \dots, 5$ BUD $f = 1, 2, \dots,$ FLOWER.

Are some stocks more sensitive to night frost than others ?

Logistic regression but observations on same tree/branch/bud may be correlated...

Logistic regression analysis

```
> cherfit=glm(cbind(STATUS,1-STATUS)~factor(STOCK)+factor(BRANCH
> summary(cherfit)
...
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3890  -1.2187   0.6165   0.8092   1.1367

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        0.09657    0.18383   0.525  0.59937
factor(STOCK)2     1.87325    0.29737   6.299 2.99e-10 ***
factor(STOCK)3     0.05564    0.21676   0.257  0.79741
factor(STOCK)4     0.67126    0.24026   2.794  0.00521 **
factor(STOCK)5     0.21711    0.22520   0.964  0.33503
factor(BRANCHNR)2  0.79626    0.18460   4.314 1.61e-05 ***
factor(BRANCHNR)3  0.82456    0.19346   4.262 2.03e-05 ***
...
Residual deviance:  997.92  on 946  degrees of freedom
```

Likelihood ratio tests using the drop1 function:

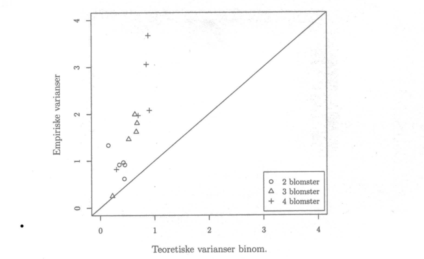
```
> drop1(cherfit,test="Chisq")
Single term deletions

Model:
cbind(STATUS, 1 - STATUS) ~ factor(STOCK) + factor(BRANCHNI
                 Df Deviance     AIC     LRT   Pr(Chi)
<none>                997.92 1011.92
factor(STOCK)     4  1061.17 1067.17   63.25 6.013e-13 ***
factor(BRANCHNR)  2  1023.49 1033.49   25.57 2.808e-06 ***
```


Overdispersion

Y_{stB} no. dead flowers on each bud.

Empirical variances of Y_{stB} versus binomial variances (within STOCK and splitting according to total numbers of flowers on buds).



Figur 2.4.: De empiriske varianser sammenlignet med de teoretiske for en binomialfordeling.

Evidence of overdispersion !

logistic mixed model for cherry flowers

p_{stBbf} probability of dead flower.

Linear mixed model for logit(p_{stBbf}):

$$\text{logit}(p_{stBbf}) = \mu + \alpha_s + \beta_B + U_{stBb} + U_{stB} + U_{st}$$

i.e. random effects for bud, branch and tree.

Analysis using lmer

```
> cherfitlmer=lmer(cbind(STATUS,1-STATUS)~factor(STOCK)+
 factor(BRANCHNR)+(1|BUDID) +(1|BRANCHID)+(1|TREEID),
 family=binomial(logit),method="Laplace")
> summary(cherfitlmer)
...
   AIC   BIC logLik deviance
 605.3 653.9 -292.7    585.3
Random effects:
 Groups   Name        Variance   Std.Dev.
 BUDID    (Intercept) 1.8118e+02 1.3460e+01
 BRANCHID (Intercept) 9.9348e+00 3.1520e+00
 TREEID   (Intercept) 5.0000e-10 2.2361e-05
number of obs: 953, groups: BUDID, 287; BRANCHID, 58; TREEID, 20
...
Estimated scale (compare to  1 )  0.3908327
```

Very large variance component for bud ! Almost zero variance for tree.

Very strong correlation within buds:

```
        0 1   #no of dead and alive for each bud
  3611  0 3   #TREEID 36 BRANCHNR 1 BUDNR 1 has 3 dead flo
  3612  0 4   #and 0 alive
  3613  3 0
  3614  0 3
  3615  0 2
  ...
```

Either all dead or all alive...

Suggests to cumulate data and create new variable dead/alive for each bud.

Analysis with reduced data set

```
> cherredfit=glm(cbind(STATUS,1-STATUS)~factor(STOCK)
  +factor(BRANCHNR),family=binomial(logit))
> drop1(cherredfit,test="Chisq")
Single term deletions

Model:
cbind(STATUS, 1 - STATUS) ~ factor(STOCK) + factor(BRANCHNR)
                 Df Deviance    AIC    LRT   Pr(Chi)
<none>                313.07 327.07
factor(STOCK)     4   333.49 339.49  20.42 0.0004128 ***
factor(BRANCHNR)  2   321.76 331.76   8.69 0.0129405 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```



```
> cherredlmer=lmer(cbind(STATUS,1-STATUS)~factor(STOCK)+factor(B
+(1|BRANCHID),family=binomial(logit),method="Laplace",data=cherr
> summary(cherredlmer)
...
   AIC   BIC logLik deviance
 309.3 338.6 -146.6    293.3
Random effects:
 Groups   Name        Variance Std.Dev.
 BRANCHID (Intercept) 1.8156   1.3475
number of obs: 287, groups: BRANCHID, 58

Estimated scale (compare to  1 )  0.8506848
Fixed effects:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       -0.07162    0.58924 -0.1215  0.90326
factor(STOCK)2     2.51185    0.86824  2.8930  0.00382 **
factor(STOCK)3     0.20331    0.70205  0.2896  0.77213
factor(STOCK)4     0.93288    0.75724  1.2320  0.21797
factor(STOCK)5     0.44781    0.72471  0.6179  0.53664
factor(BRANCHNR)2  0.88478    0.57430  1.5406  0.12341
factor(BRANCHNR)3  1.27228    0.61601  2.0654  0.03889 *
```


Likelihood ratio test:

```
> cherredlmer2=lmer(cbind(STATUS,1-STATUS)~factor(BRANCHNR)+(1|B
> anova(cherredlmer2,cherredlmer)
Data: cherred
Models:
cherredlmer2: cbind(STATUS, 1 - STATUS) ~ factor(BRANCHNR) + (1
cherredlmer: cbind(STATUS, 1 - STATUS) ~ factor(STOCK) + factor(
cherredlmer2:     (1 | BRANCHID)
             Df     AIC     BIC  logLik  Chisq Chi Df Pr(>Chisq)
cherredlmer2  4  312.24  326.88 -152.12
cherredlmer   8  309.29  338.57 -146.64 10.952      4    0.02711
```

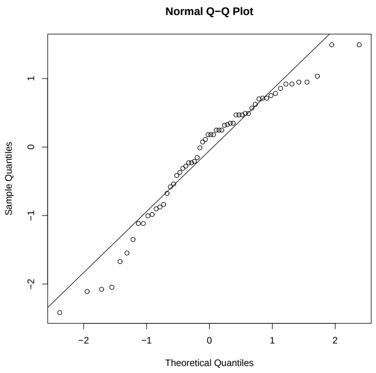
Likelihood for logistic model:

```
> logLik(cherredfit)
'log Lik.' -156.5337 (df=7)#likelihood much smaller than
                              #for logistic mixed model...
```

-2 log likelihood ratio for no branch random effect is $2(156,5 - 146.6) = 19.8$ which is highly significant when compared with $\chi^2(1)$.

Prediction of random effects for cherry flowers

```
> branchre=ranef(cherredlmer)$BRANCHID
> qqnorm(branchre[[1]])
> qqline(branchre[[1]])
```



NB: even if model is correct, predictions only approximately normal. Predictions are modes of conditional distributions of random effects (more details later)

NB: predicted random effects can be used to rank subjects.

Residuals

Fitted values $\hat{p}_{stBb}$ obtained by plugging in parameter estimates and predictions for the unknown parameters and random effects.

Residuals $(y_{stBb} - \hat{p}_{stBb})/\sqrt{\hat{p}_{stBb}(1 - \hat{p}_{stBb})}$

Interpretation of variance components I

For a linear mixed model we can decompose the variance of an observation as the sum of the variance components.

For logistic and Poisson regression with random effects the decomposition is only possible at the level of the linear predictor.

If e.g.

$$\eta_{ij} = \text{logit}(p_{ij}) = \alpha + U_i + U_{ij}$$

and random effects independent, then

$$\mathbb{V}\text{ar}\eta_{ij} = \mathbb{V}\text{ar}U_i + \mathbb{V}\text{ar}U_{ij} = \tau^2 + \omega^2$$

Also note

$$\mathbb{E}p_{ij} = \mathbb{E}\frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})} \neq \frac{\exp(\mathbb{E}\eta_{ij})}{1 + \exp(\mathbb{E}\eta_{ij})}$$

Interpretation of variance components II

However, if $X \sim N(\mu, \sigma^2)$ then

$$\mathbb{E}\exp(tX) = \exp(t\mu + t^2\sigma^2/2).$$

(Laplace transform of X)

Thus if we consider odds

$$o_{ij} = p_{ij}/(1 - p_{ij}) = \exp(\eta_{ij})$$

then e.g.

$$\mathbb{E}o_{ij} = \exp(\alpha + (\tau^2 + \omega^2)/2)$$

Variance components and Poisson regression

For Poisson regression with

$$\eta_{ij} = \log \lambda_{ij} = \alpha + U_i$$

mean of observation Y_{ij} is

$$\mathbb{E}Y_{ij} = \mathbb{E}[Y_{ij}|\lambda_{ij}] = \mathbb{E}\lambda_{ij} = \exp(\alpha + \tau^2/2)$$

and variance is (note overdispersion!)

$$\begin{aligned}\mathbb{V}\mathrm{ar}Y_{ij} &= \mathbb{V}\mathrm{ar}\lambda_{ij} + \mathbb{E}\lambda_{ij} = \mathbb{E}\exp(2\eta_{ij}) - (\mathbb{E}\lambda_{ij})^2 + \mathbb{E}Y_{ij} \\ &= \mathbb{E}Y_{ij}[\exp(\alpha + 3\tau^2/2) - \mathbb{E}Y_{ij} + 1]\end{aligned}$$

(used general formulas $\mathbb{E}Y = \mathbb{E}\mathbb{E}[Y|X]$ and $\mathbb{V}\mathrm{ar}Y = \mathbb{V}\mathrm{ar}\mathbb{E}[Y|X] + \mathbb{E}\mathbb{V}\mathrm{ar}[Y|X]$)

Generalized linear mixed effects models

Consider stochastic variable $\mathbf{Y} = (Y_1, \ldots, Y_n)$ and random effects $\mathbf{U}$.

Two step formulation of GLMM:

- $\mathbf{U} \sim N(0, \Sigma)$.
- Given realization $\mathbf{u}$ of $\mathbf{U}$, Y_i independent and each follows density $f_i(\mathbf{y}|\mathbf{u})$ with mean $\mu_i = g^{-1}(\eta_i)$ and linear predictor $\eta = X\beta + Z\mathbf{u}$.

I.e. conditional on $\mathbf{U}$, Y_i follows a generalized linear model.

NB: GLMM specified in terms of marginal density of $\mathbf{U}$ and conditional density of $\mathbf{Y}$ given $\mathbf{U}$. But the likelihood is the marginal density of $f(\mathbf{y})$ which can be hard to evaluate !

Likelihood for generalized linear mixed model

For normal linear mixed models we could compute the marginal distribution of $\mathbf{Y}$ directly as a multivariate normal. That is, $f(\mathbf{y})$ is a density of a multivariate normal distribution.

For a generalized linear mixed model it is difficult to evaluate the integral:

$$f(\mathbf{y}) = \int_{\mathbb{R}^m} f(\mathbf{y}, \mathbf{u})\mathrm{d}\mathbf{u} = \int_{\mathbb{R}^m} f(\mathbf{y}|\mathbf{u})f(\mathbf{u})\mathrm{d}\mathbf{u}$$

since $f(\mathbf{y}|\mathbf{u})f(\mathbf{u})$ is a very complex function.

Non-normal example: logistic regression with random intercepts

$$\begin{aligned}&U_j \sim N(0, \tau^2),\ j = 1, \ldots, m \\ &Y_j|U_j = u_j \sim \mathrm{binomial}(n_j, p_j) \\ &\log(p_j/(1-p_j)) = \eta_j = \beta + U_j \\ &p_j = \exp(\eta_j)/(1 + \exp(\eta_j))\end{aligned}$$

Conditional density:

$$f(y|u;\beta) = \prod_j p_j^{y_j}(1-p_j)^{1-y_j} = \prod_j \frac{\exp(\beta + u_j)^{y_j}}{(1 + \exp(\beta + u_j))^{n_j}}$$

Likelihood function ($u = (u_1, \ldots, u_m)$)

$$\int_{\mathbb{R}^m} f(y|u;\beta)f(u;\tau^2)\mathrm{d}u = \prod_j \int_{\mathbb{R}} \frac{\exp(\beta + u_j)^{y_j}}{(1 + \exp(\beta + u_j))^{n_j}} \frac{\exp(-u_j^2/(2\tau^2))}{\sqrt{2\pi\tau^2}}\mathrm{d}u_j$$

Integrals can not be evaluated in closed form.

One-dimensional case

Compute

$$L(\theta) = \int_{\mathbb{R}} f(y|u;\beta) f(u;\tau^2)\mathrm{d}u$$

Possibilities:

- Laplace approximation.
- Numerical integration/quadrature (e.g. Gaussian quadrature as in `PROC NLMIXED` (SAS) or `GLLAM` (STATA)) (one level of random effects, dimensions one or two).

Laplace approximation

Let $g(u) = \log(f(y|u)f(u))$ and choose $\hat{u}$ so $g'(\hat{u}) = 0$ ($\hat{u} = \arg\max g(u)$).

Taylor expansion around $\hat{u}$:

$$g(u) \approx \tilde{g}(u) =$$
$$g(\hat{u}) + (u-\hat{u})g'(\hat{u}) + \frac{1}{2}(u-\hat{u})^2 g''(\hat{u}) = g(\hat{u}) - \frac{1}{2}(u-\hat{u})^2(-g''(\hat{u}))$$

I.e. $\exp(\tilde{g}(u))$ proportional to normal density $N(\mu_{LP}, \sigma^2_{LP})$, $\mu_{LP} = \hat{u}$ $\sigma^2_{LP} = -1/g''(\hat{u})$.

$$L(\theta) = \int_{\mathbb{R}} \exp(g(u))\mathrm{d}u \approx \int_{\mathbb{R}} \exp(\tilde{g}(u))\mathrm{d}u$$
$$= \exp(g(\hat{u})) \int_{\mathbb{R}} \exp\big(-\frac{1}{2\sigma^2_{LP}}(u-\mu_{LP})^2\big)\mathrm{d}u = \exp(g(\hat{u}))\sqrt{2\pi\sigma^2_{LP}}$$

Laplace approximation also possible for higher dimensions (multivariate Taylor expansion).

NB:

$$f(u|y) = f(y|u)f(u)/f(y) \propto \exp(g(u)) \approx \mathit{const}\,\exp\big(-\frac{1}{2\sigma^2_{LP}}(u-\mu_{LP})^2\big)$$

where $\mu_{LP} = \hat{u}$ $\sigma^2_{LP} =, -1/g''(\hat{u})$.

Hence

$$U|Y = y \approx N(\mu_{LP}, \sigma^2_{LP})$$

Note: μ_{LP} is mode of conditional distribution - used for prediction of random effects in `lmer` (`ranef()`).

Penalized quasi-likelihood

One solution: do not use likelihood function but something simpler.

$\theta = (\beta, \tau^2)$

PQL estimates $\hat{\theta}$ and $\hat{u}$ maximize joint density

$$f(y,u;\theta) = f(y|u;\beta)f(u;\tau^2).$$

PQL estimates less accurate than ML.

Asymptotic results require increasing number of observations for each random effect.

Implemented in `lmer` and SAS macro `glimmix`.

Exercises

- Repeat the exercise regarding the wheezing data but now using introducing subject-specific random effects. How large is the standard deviation for these random effects ? Compute a likelihood ratio test for the significance of the subject specific random variation.
- The disruption data available on the web page were obtained by letting 24 subjects perform a memory task with sound playing in the background. The response is the number of errors committed. The five background sound conditions were: silence, white noise, continuous frequency-modulated (FM) tone at 8 Hz, FM tone at 0.25Hz interleaved with short periods of silence, and Korean speech. Consider Poisson regression models for the number of errors both with and without subject specific random effects. Is there a significant subject specific random variation ? - and is there a significant effect of the background noise ?

More exercises

- (continuation of previous exercise) Which subject is least susceptible to noise ? (hint: look at predicted random effects). Draw a qqplot of the random effects - do they appear to be normal ?
- try alternatively a linear mixed effects model for the disruption data (NB for large counts or square root transformed counts the normal distribution may be a good approximation). Assess the fitted linear mixed effects model using residuals and predicted random effects.

Optional exercises:

- (disruption data exercise) using the fitted Poisson mixed regression model, compute the marginal mean and variance for an randomly selected subject who is exposed to Korean speech.
- show that the conditional mean is the minimum mean square error predictor, ie. $\mathbb{E}(X - \hat{X})^2$ is minimal for $\hat{X} = \mathbb{E}(X|Y)$ where X denotes an unobserved random variable and Y the data.