GLMMs and related

27 Nov 2024

Digression: Restricted maximum likelihood

- We usually do ML
- Remember the ML estimate for SD, $\sum(x-\bar{x})^2/n$ (biased), and the *Bessel correction* $\sum(x-\bar{x})^2/(n-1)$
- REML is analogous; integrates over the fixed effects, removes bias
- substitute REML estimate of $\hat{\theta}$ into ML expression to get "REML'' estimate of $\hat{\beta}$
- points to remember
  - usually gives a better estimate of RE variances (unbiased in simple cases)
  - harder to implement for GLMMs
  - **never compare REML-fitted models that differ in their fixed effects**

Estimation, continued

In general

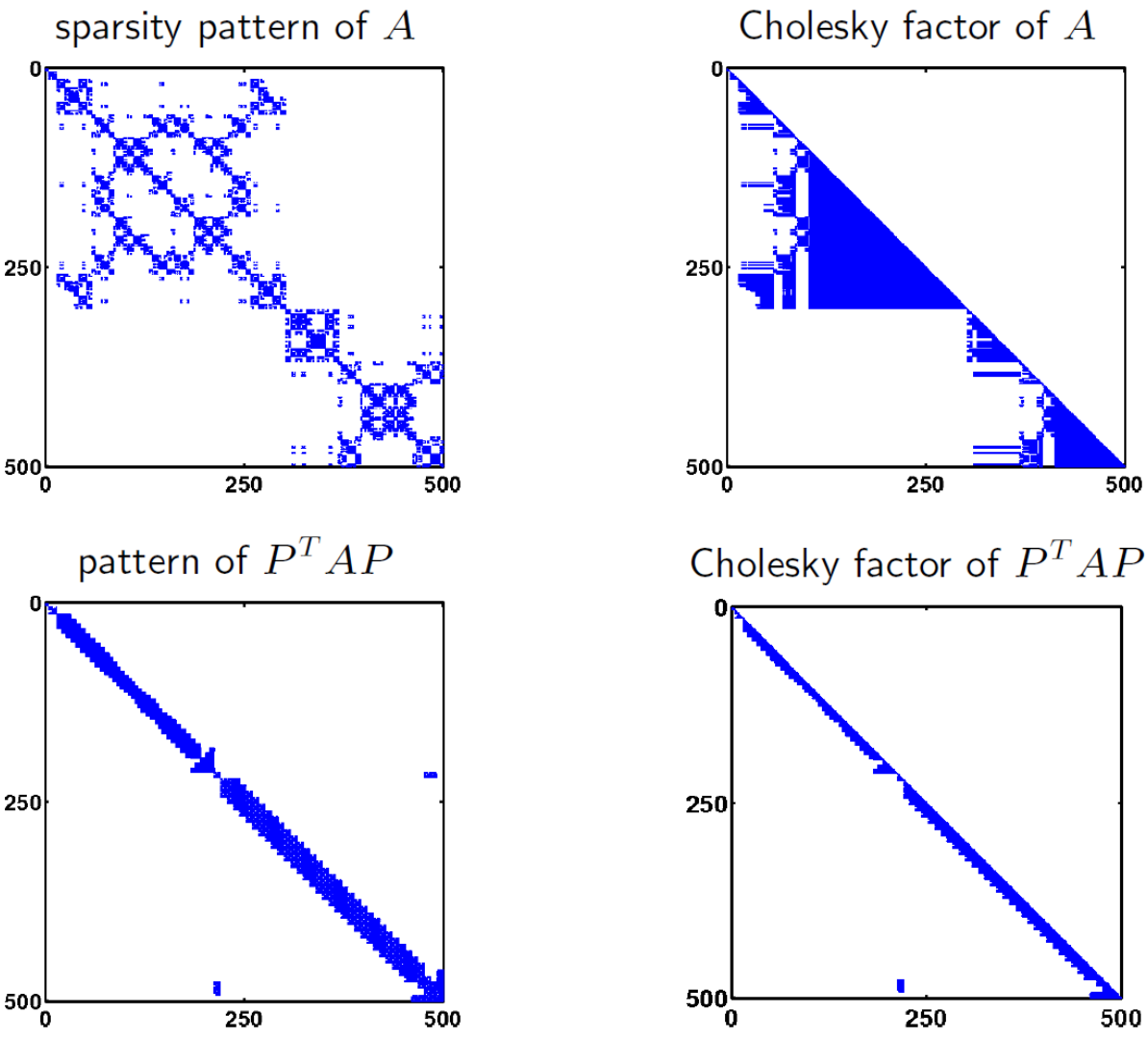
- We can evaluate the marginal likelihood $L(\theta,\beta) = P(y|\beta,\theta)$, with the latent variables integrated out, via standard linear algebra
- For the linear mixed model we can also profile out the fixed effects β so we have just $L(\theta)$
- we can do this *approximately* for GLMMs, for speed (see here

Linear mixed models

- solve linear algebra, classically via *Henderson's equations* (Henderson Jr. 1982); rearranged by Bates et al. (2015): > In a complex model fit to a large data set, the dominant calculation in the evaluation of the profiled deviance (Equation 35) or REML criterion

(Equation 42) is this sparse Cholesky factorization (Equation 18). The factorization is performed in two phases; a symbolic phase and a numeric phase. The symbolic phase, in which the fill-reducing permutation P is determined along with the positions of the nonzeros in L_θ, does not depend on the value of θ. It only depends on the positions of the nonzeros in $Z\Lambda_\theta$. The numeric phase uses θ to determine the numeric values of the nonzeros in L_θ

Briefly, we can do this:



From Stack Overflow, ultimately from Lieven Vandenberghe's notes

GLMMs

Hierarchy of increasing accuracy/difficulty:

- Penalized quasi-likelihood
- Laplace approximation
- (adaptive) Gauss-Hermite quadrature
- Bayes! (Gibbs sampling, Hamiltonian Monte Carlo)

Accuracy of approximations depends on effective sample size *per cluster* [Stringer and Bilodeau (2022), Biswas (2015)

Key references:

Breslow (2004), Walker et al. (n.d.), Madsen and Thyregod (2011)

Penalized quasi-likelihood

Breslow (2004)

> As usual when software for complicated statistical inference procedures is broadly disseminated, there is potential for abuse and misinterpretation. In spite of the fact that PQL was initially advertised as a procedure for approximate inference in GLMMs, and its tendency to give seriously biased estimates of variance components and a fortiori regression parameters with binary outcome data was emphasized in multiple publications [5, 6, 24], some statisticians seemed to ignore these warnings and to think of PQL as synonymous with GLMM.

- Laplace approximation (see next section) applied to *quasi-likelihood*
- in practice: does the same iterative algorithm as IRLS, but with a weighted LMM rather than a weighted least-squares fit

Laplace approximation

- Second-order Taylor approximation of the conditional likelihood $L(\mathbf{b})$ around $bbest$
- Once we do this, doing the integral is easy
- procedure:
  - solve the "inner optimization" (find $\tilde{\mathbf{b}}$)
  - get the conditional log-likelihood and the determinant of the Hessian
- TMB and RTMB can do this automatically!

$$L(\theta, \mathbf{b}, \mathbf{y}) \approx L(\theta, \tilde{\mathbf{b}}, \mathbf{y}) - \frac{1}{2}(\mathbf{b} - \tilde{\mathbf{b}})^\top \mathbf{H} (\mathbf{b} - \tilde{\mathbf{b}})$$

We can write the integral as

$$\begin{aligned}
&\log \int \exp(L(\theta, \tilde{\mathbf{b}}) - \frac{1}{2}(\mathbf{b} - \tilde{\mathbf{b}})^\top \mathbf{H} (\mathbf{b} - \tilde{\mathbf{b}}))\, d\mathbf{b} \\
=&L(.) + \log \int \exp(-\frac{1}{2}(\mathbf{b} - \tilde{\mathbf{b}})^\top \mathbf{H} (\mathbf{b} - \tilde{\mathbf{b}}))\, d\mathbf{b} \\
=&L(.) + \log \left| \frac{(2\pi)^q}{\mathbf{H}(\tilde{\mathbf{b}})} \right|^{1/2} \int \exp(\frac{Q}{(2\pi)^{q/2} |\mathbf{H}^{-1}(\tilde{\mathbf{b}})|^{1/2}})\, d\mathbf{b} \\
=&L(.) - \frac{1}{2}\log(\mathbf{H}) + \frac{q}{2}\log(2\pi)
\end{aligned}$$

Gauss-Hermite quadrature

- only possible when we can separate the data into conditionally independent chunks

$$\begin{aligned}
&\int L(y|\mathbf{b}, \beta) L(\mathbf{b}|\theta)\, d\mathbf{b} \\
=&\prod_i \int L(y_i|\mathbf{b}_i, \beta) L(\mathbf{b}_i|\theta)\, d\mathbf{b}_i
\end{aligned}$$

where y_i, $\mathbf{b}_i$ are the data points in the i^{th} cluster and $\mathbf{b}_i$ are the corresponding latent variables

- only works for a *single* RE
- `lme4` can only handle *scalar* REs; `GLMMadaptive` does vector REs as well
- still need to know curvature: choose quadrature knots based on $\tilde{\mathbf{b}}$ and $1/H$

Bayes: Gibbs sampling

- *Gibbs sampling* relies on being able to efficiently sample from the *conditional* posterior distribution, e.g. Post($\mathbf{b}|y, \theta$) and Post($\theta|y, \mathbf{b}$)
- `MCMCglmm` package
  - fast(ish) sampling

  - flexible
  - priors must come from conjugate distributions (Gaussian, inverse-Gamma, inverse-Wishart)

Bayes: Hamiltonian Monte Carlo

- more efficient sampling
- relies on knowing gradients (autodiff!)
- not restricted to conjugate priors
- Stan, `tmbstan`

First- and second-order specifications (Hefley et al. 2017)

- "first-order" specification: all structure in $\mathbf{Z}$ matrix, i.e.

$$\begin{aligned}\mathbf{y} &\sim \text{MVN}(\mathbf{X}\beta + \mathbf{Z}\mathbf{b}, \sigma_r^2\mathbf{I}) \\ \mathbf{b} &\sim \text{MVN}(0, \sigma_g^2 I)\end{aligned}$$

- "second-order" specification: we can write this out as

$$\mathbf{y} \sim \text{MVN}(\mathbf{X}\beta, \sigma_g^2\mathbf{Z}\mathbf{Z}^\top + \sigma_r^2\mathbf{I})$$

In `lmer`, we distinguish between the *spherical random effects* $\mathbf{u}$ (iid $N(0,1)$) and the *non-spherical random effects $\mathbf{b} = \Lambda\mathbf{u}$ ($\Sigma = \Lambda\Lambda^\top$).

- example: phylogenetic mixed effects models

Covariance structures

- see covstruct vignette
- AR1, Ornstein-Uhlenbeck, Toeplitz, compound symmetric
- Gaussian processes: Matérn, Gaussian, exponential
- reduced-rank
- "proportional" (e.g. phylogenetic)

RTMB code for mixed models

```
library(RTMB)       ## autodiff and Laplace approximation
library(Matrix)     ## handling sparse matrices
library(reformulas) ## processing RE formulas
data("sleepstudy", package = "lme4")
form <- Reaction ~ Days + (Days | Subject)
X <- model.matrix(Reaction ~ Days, data = sleepstudy)
fr <- model.frame(subbars(form), data = sleepstudy)  ## get the necessary variables
reTrms <- mkReTrms(findbars(form), fr = fr)
Z <- t(reTrms$Zt)
us <- unstructured(2)
nsubj <- length(levels(sleepstudy$Subject))
pars0 <- list(beta = c(250,0), cor = 0, logsd = c(1,1,1),
              b = rep(0, 2*nsubj))
tmbdata <- list(y = sleepstudy$Reaction, X = X, Z = Z)
ff <- function(pars) {
    getAll(pars,  tmbdata)   ## unpacking the data and parameters
    mu <- drop(X %*% beta + Z %*% b)  ## computing the linear predictor
    ## the conditional likelihood
    L1 <- -sum(dnorm(y, mean = mu, sd = exp(logsd[3]), log = TRUE))
    ## reshape the random effects
    bmat <- matrix(b, ncol=2, byrow=TRUE)
    ## likelihood of the latent variables
    L2 <- -sum(dmvnorm(bmat, mu = rep(0,2), Sigma = us$corr(cor),
                       scale = exp(logsd[1:2]), log = TRUE))
    return(L1 + L2)
}
ff(pars0)
```

```
[1] 67506.76
```

```
obj <- MakeADFun(ff, pars0)
obj$fn()
```

```
[1] 67506.76
```

```
obj$gr()
```

```
outer mgc:  134004.5
```

```
          [,1]      [,2] [,3] [,4] [,5]      [,6]      [,7]      [,8]     [,9]
[1,] -1181.669 -7421.152    0   18   18 -134004.5 -124.6896 -804.1095 47.05204
       [,10]    [,11]    [,12]     [,13]     [,14]     [,15]     [,16]     [,17]
[1,] 186.481 25.71199 47.4301 -72.02736 -357.7087 -80.43795 -420.7667 -77.54992
         [,18]     [,19]     [,20]     [,21]     [,22]       [,23]    [,24]
[1,] -455.7891 -89.53554 -504.9823 -61.30966 -412.7018 -0.09478883 31.74066
         [,25]    [,26]     [,27]     [,28]     [,29]     [,30]     [,31]
[1,] -170.1449 -978.08 -34.96315 -307.9961 -86.07687 -605.1114 -54.2664
         [,32]     [,33]     [,34]     [,35]     [,36]     [,37]     [,38]
[1,] -316.0299 -118.3122 -683.8776 -75.83464 -467.9593 -56.43721 -455.5673
         [,39]     [,40]     [,41]     [,42]
[1,] -60.87928 -376.5472 -91.87389 -539.5773
```

```
obj2 <- MakeADFun(ff, pars0,
                  ## treat b as a random effect
                  random = "b", silent = TRUE)
obj2$fn()
```

```
[1] 7690.738
attr(,"logarithm")
[1] TRUE
```

```
fit <- with(obj2, nlminb(par, fn, gr))
library(lme4)
fit2 <- lme4::lmer(Reaction ~ Days + (Days | Subject), data = sleepstudy)
fixef(fit2)
```

```
(Intercept)        Days
  251.40510    10.46729
```

```
fit$par
```

```
       beta         beta          cor        logsd        logsd        logsd
251.40509651  10.46728836   0.08159032   3.16886893   1.74341509   3.24227251
```

```
## check starting vals!
```

Key references

- Walker et al. (n.d.)

References

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using Lme4." *Journal of Statistical Software* 67 (October): 1–48. https://doi.org/10.18637/jss.v067.i01.

Biswas, Keya. 2015. "Performances of Different Estimation Methods for Generalized Linear Mixed Models." Master's thesis, McMaster University. https://macsphere.mcmaster.ca/bitstream/11375/17272/2/M.Sc_Thesis_final_Keya_Biswas.pdf.

Breslow, N. E. 2004. "Whither PQL?" In *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*, edited by Danyu Y. Lin and P. J. Heagerty, 1–22. Springer.

Hefley, Trevor J., Kristin M. Broms, Brian M. Brost, Frances E. Buderman, Shannon L. Kay, Henry R. Scharf, John R. Tipton, Perry J. Williams, and Mevin B. Hooten. 2017. "The Basis Function Approach for Modeling Autocorrelation in Ecological Data." *Ecology* 98 (3): 632–46. https://doi.org/10.1002/ecy.1674.

Henderson Jr., Charles R. 1982. "Analysis of Covariance in the Mixed Model: Higher-Level, Nonhomogeneous, and Random Regressions." *Biometrics* 38 (3): 623–40. http://www.jstor.org/stable/2530044.

Madsen, Henrik, and Poul Thyregod. 2011. *Introduction to General and Generalized Linear Models*. CRC Press.

Stringer, Alex, and Blair Bilodeau. 2022. "Fitting Generalized Linear Mixed Models Using Adaptive Quadrature." arXiv. https://doi.org/10.48550/arXiv.2202.07864.

Walker, Steven C., Rune Haubo Bojesen Christensen, Douglas Bates, and Martin Maechler. n.d. "Fitting Generalized Linear Mixed-Effects Models Using Lme4." https://github.com/lme4/lme4/blob/master/misc/glmer_JSS/glmer.pdf.