A goodness-of-fit test for the random effects distribution in mixed models

Roula Tsonaka

Jointly with: Geert Molenberghs, Dimitris Rizopoulos and Geert Verbeke

Interuniversity Institute for Biostatistics and statistical Bioinformatics
Catholic University Leuven & Hasselt University, Belgium

spyridoula.tsonaka@med.kuleuven.be

30th Annual Conference of the International Society for Clinical Biostatistics
Prague, August 26th, 2009
Mixed models invaluable tool for the analysis of correlated data structures

Key component: random effects

Standard statistical software assume normally distributed random effects
The impact of misspecifying random effects distribution extensively studied

- **Robustness** of mixed models regarding fixed effects estimates
  (Rizopoulos *et al.*, Bka:2008)

- **Asymptotic bias** in fixed effects estimates and variance components
  (Heagerty and Kurland, Bka:2001)

- Estimation of the random effects

- Impact on type I error and power for fixed effects
  (Litière *et al.*, 2007, Bcs:2007)
Robust approaches

- Relax the common normality assumptions for the mixing distribution
  - Spline-based approaches (Ghidey et al., Bcs:2004; Komarek and Lesaffre, CSDA:2008)

- Non-parametric alternatives (Tsonaka et al., Bcs:2009)

- But computationally intensive & limited applicability with standard statistical software
Diagnostic tools to test common parametric assumptions are necessary

We develop a goodness-of-fit test based on the directional derivative

Verbeke and Molenberghs (2009) studied the directional derivative as a graphical tool to assess assumptions for random effects
A directional derivative based diagnostic test

- Consider a mixture model setting (e.g., mixed model)

$$\ell(G \mid Y) = \sum_{i=1}^{n} \log \int f(y_i \mid b_i) dG(b_i)$$

- $G$ is the normal cdf

- Let $G_0$ be the true distribution then

$$H_0: G_0 = G$$
$$H_a: G_0 \neq G,$$

- We can test this hypothesis via the directional derivative
Directional derivative

- The directional derivative of the log-likelihood for $G$ and $Q$

$$\mathcal{D}(G, Q) = \lim_{\alpha \to 0} \frac{\ell\{ (1 - \alpha)G + \alpha Q \} - \ell(G)}{\alpha}$$

- Under $H_0 : G_0 = G$
  - $\Rightarrow$ no other $Q$ fits better to the data
  - $\Rightarrow \mathcal{D}(G, Q) \leq 0$ for all $Q \in \Omega_B$ with $b \in B$

- Under $H_a : G_0 \neq G$
  - $\Rightarrow$ there is at least one $Q \in \Omega_B$ that fits better to the data
  - $\Rightarrow \mathcal{D}(G, Q) > 0$
For a random sample of size $n$ we use

$$T = \frac{1}{n} \mathcal{D}(\hat{G}, Q) = \frac{1}{n} \sum_{i=1}^{n} \frac{f(y_i \mid Q)}{f(y_i \mid \hat{G})} - 1$$

We cannot compute $\mathcal{D}(\hat{G}, Q)$ for all $Q \in \Omega_B$

Instead we consider degenerate distributions $Q = \delta(b), \ b \in B$

Search if $\mathcal{D}(\hat{G}, \delta(b)) > 0$ for any $\delta(b) \Rightarrow \hat{G}$ needs misfits in the direction of $\delta(b)$
We do not take all $\delta(b)$ over $B$ but over $B_C \subseteq B$ (Lindsay, 1995)

We take $K$ degenerate distributions $\delta(b_k)$ with $k = 1, \ldots, K$ over $B_C$

We test the hypothesis for every $\delta(b_k)$

$H_0$: the fit cannot be further improved in the direction of $\delta(b_k)$

$H_a$: the fit can be improved in the direction of at least one $\delta(b_k)$

Compute $D(\hat{G}, \delta(b_k))$ for each $\delta(b_k)$ and use as statistic

$$T = \frac{1}{n} D(\hat{G}, \delta(b_k)) = \frac{1}{n} \sum_{i=1}^{n} \frac{f(y_i \mid \delta(b_k))}{f(y_i \mid \hat{G})} - 1$$
Under $H_0 \Rightarrow T = 0$ for all $\delta(b_k)$

Under $H_a \Rightarrow T > 0$ for at least one $\delta(b_k)$
Illustration

Directional derivative plots

Null hypothesis

Alternative hypothesis
The hypothesis will be tested for every $\delta(b_k)$

Issues

1. Correlated hypotheses $\Rightarrow$ type I error inflation if tested separately
2. Simultaneous testing while accounting for the inter-dependencies
3. The distribution of $T$ cannot be derived easily
- $T$ is in fact a sample mean of subject-specific contributions

$$T = \frac{1}{n} \sum_{i=1}^{n} \frac{f(y_i | \delta(b_k))}{f(y_i | \hat{G})} - 1 = \frac{1}{n} \sum_{i=1}^{n} y_{ik}^*$$

- A multivariate data setting arises for the $n \times K$ gradient data $Y^*$

- GEE approach to capture within-subject correlations
  - $\rightarrow$ correlations between the $K$ columns in $Y^*$

- Let $E(y_{ik}^*) = \mu_k$, $\mu = (\mu_1, \ldots, \mu_K)$
The hypothesis is formulated as

\[ H_0: \mu = 0 \]
\[ H_a: \mu > 0, \]

Based on the asymptotic normality of \( \hat{\mu} \) we use multivariate Wald test

\[ W = \hat{\mu}^T \hat{\text{Var}}(\hat{\mu})^{-1} \hat{\mu}. \]

with \( \hat{\text{Var}}(\hat{\mu})^{-1} \) the sandwich estimator

\[ W \overset{H_0}{\sim} \chi^2_K \]

\( H_0 \) is rejected when \( W > \chi^2_{2\alpha}, K \) and \( \sum_{k=1}^{K} \hat{\mu}_k > 0 \) (Follmann, JASA:1996)
Randomized study on 291 patients with toenail dermatophyte onychomycosis

2 treatments for a 3-month period

Unaffected nail length (mm) measured at 7 planned visits

Linear mixed effects model with

\[ E(y_{ij}) = \gamma_0 + \gamma_1 Time_{ij} + \gamma_2 Time_{ij}^2 + \gamma_3 \text{Treat}_i \text{Time}_{ij} + \gamma_4 \text{Treat}_i \text{Time}_{ij}^2, \]

with \( b_i \sim \text{Gaussian} \)

Test the normality in the random-intercepts case: \( T = 4.5, df = 11, p\text{-value} > 0.999 \)
LMM with random-intercepts: Dirichlet Process Mixture of Normals prior

A goodness-of-fit test for the random effects distribution
Randomized study on 895 patients with Rheumatoid Arthritis

5 treatments for a 3-month period

Visual Analogue Score at 5 planned visits

Linear mixed effects model with

\[ E(y_{ij}) = \gamma_0 + \gamma_1 \text{Time}_{ij} + \gamma_2 \text{Time}_{ij}^2 + \gamma_3 \text{Treat}_i \text{Time}_{ij} + \gamma_4 \text{Treat}_i \text{Time}_{ij}^2, \]

with \( b_i \sim \text{Gaussian} \)

Test the normality in the random-intercepts case: \( T = 121.9, df = 11, p\)-value < 0.001
LMM with random-intercepts: Dirichlet Process Mixture of Normals prior

![Graph showing density distribution with Dirichlet Process Mixture of Normals prior]
Thank you for your attention!