Mixed models in R using the lme4 package
Part 3: Inference based on profiled deviance

Douglas Bates

8th International Amsterdam Conference
on Multilevel Analysis
<Bates@R-project.org>

2011-03-16

Outline

1 Profiling the deviance

Profiling

Outline

Outline

Outline

1 Profiling the deviance

2 Plotting the profiled deviance

3 Density plots

4 Profile pairs

Outline

Outline

Likelihood ratio tests and deviance

- In section 2 we described the use of likelihood ratio tests (LRTs) to compare a reduced model (say, one that omits a random-effects term) to the full model.
- The test statistic in a LRT is the change in the deviance, which is negative twice the log-likelihood.
- We always use maximum likelihood fits (i.e. REML=FALSE) to evaluate the deviance.
- In general we calculate p-values for a LRT from a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters in the models.
- The important thing to note is that a likelihood ratio test is based on fitting the model under each set of conditions.

Profiling the deviance versus one parameter

- There is a close relationship between confidence intervals and hypothesis tests on a single parameter. When, e.g. $H_0 : \beta_1 = \beta_{1,0}$ versus $H_a : \beta_1 \neq \beta_{1,0}$ is **not** rejected at level α then $\beta_{1,0}$ is in a $1 - \alpha$ confidence interval on the parameter β_1.

- For linear fixed-effects models it is possible to determine the change in the deviance from fitting the full model only. For mixed-effects models we need to fit the full model and all the reduced models to perform the LRTs.

- In practice we fit some of them and use interpolation. The `profile` function evaluates such a "profile" of the change in the deviance versus each of the parameters in the model.

Transforming the LRT statistic

- The LRT statistic for a test of a fixed value of a single parameter would have a χ_1^2 distribution, which is the square of a standard normal.
- If a symmetric confidence interval were appropriate for the parameter, the LRT statistic would be quadratic with respect to the parameter.
- We plot the square root of the LRT statistic because it is easier to assess whether the plot looks like a straight line than it is to assess if it looks like a quadratic.
- To accentuate the straight line behavior we use the signed square root transformation which returns the negative square root to the left of the estimate and the positive square root to the right.
- This quantity can be compared to a standard normal. We write it as ζ
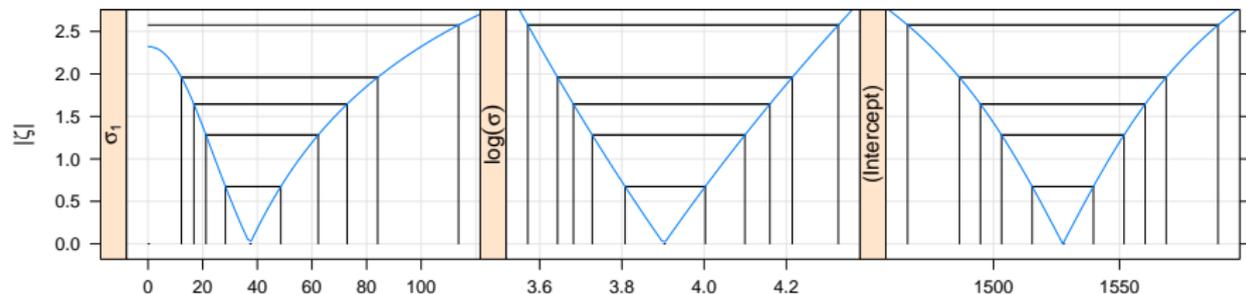
Evaluating and plotting the profile

```
> pr1 <- profile(fm1M <- lmer(Yield ~ 1+(1|Batch), Dyestuff, RI
> xyplot(pr1, aspect=1.3)
```
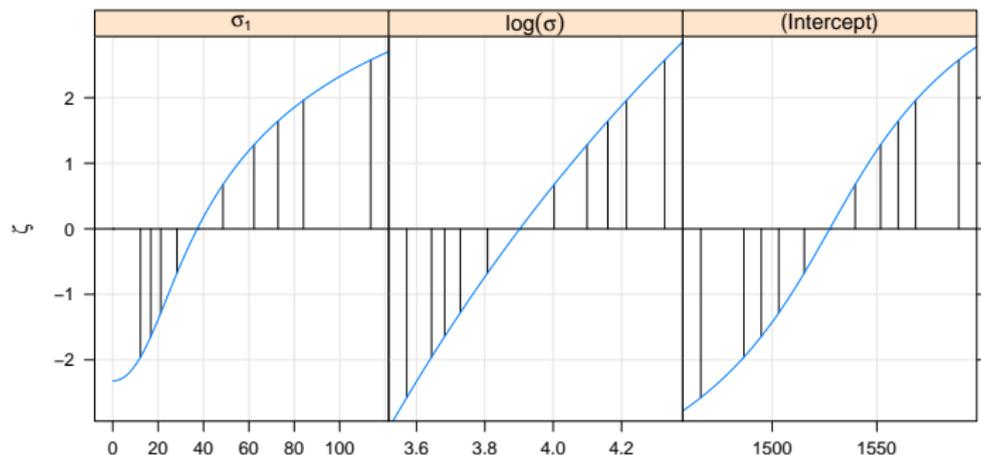

- The parameters are σ_b, $\log(\sigma)$ (σ is the residual standard deviation) and μ. The vertical lines delimit 50%, 80%, 90%, 95% and 99% confidence intervals.

Alternative profile plot

```
> xyplot ( pr1 , aspect =0.7 , absVal = TRUE )
```


Numerical values of the confidence interval limits are obtained from the method for the confint generic

```
> confint ( pr1 )
```

```
                  2.5 %      97.5 %
.sig01        12.201753    84.06289
.lsig          3.643622     4.21446
(Intercept) 1486.451500  1568.54849
```

Changing the confidence level

As for other methods for the confint generic, we use level=α to obtain a confidence level other than the default of 0.95.

```
> confint ( pr1 , level =0.99)
```

```
                 0.5 %        99.5 %
.sig01               NA    113.692643
.lsig          3.571293      4.326347
(Intercept) 1465.874011   1589.126022
```

Note that the lower 99% confidence limit for σ_1 is undefined.

Interpreting the univariate plots

- A univariate profile ζ plot is read like a normal probability plot
  - ▶ a sigmoidal (elongated "S"-shaped) pattern like that for the (Intercept) parameter indicates overdispersion relative to the normal distribution.
  - ▶ a bending pattern, usually flattening to the right of the estimate, indicates skewness of the estimator and warns us that the confidence intervals will be asymmetric
  - ▶ a straight line indicates that the confidence intervals based on the quantiles of the standard normal distribution are suitable
- Note that the only parameter providing a more-or-less straight line is σ and this plot is on the scale of $\log(\sigma)$ not σ or, even worse, σ^2.
- We should expect confidence intervals on σ^2 to be asymmetric. In the simplest case of a variance estimate from an i.i.d. normal sample the confidence interval is derived from quantiles of a χ^2 distribution which is quite asymmetric (although many software packages provide standard errors of variance component estimates as if they were meaningful).

Profile ζ plots for $\log(\sigma)$, σ and σ^2
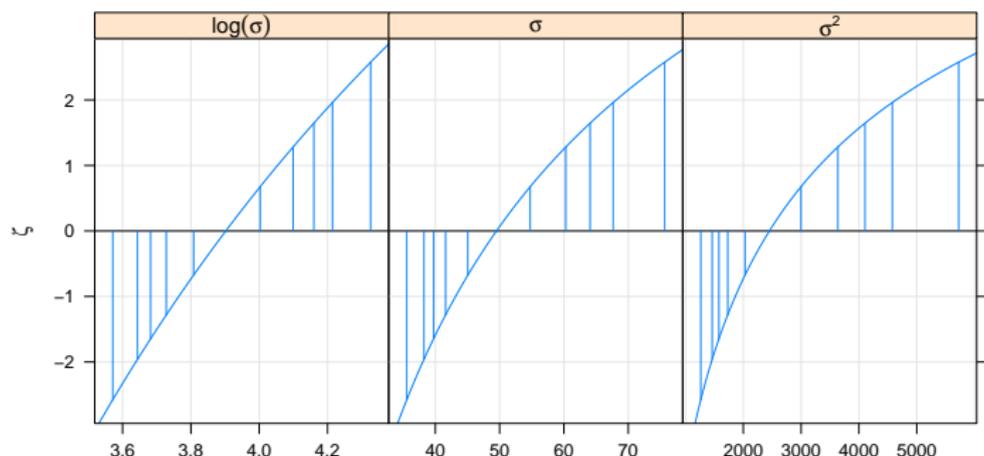
- We can see moderate asymmetry on the scale of σ and stronger asymmetry on the scale of σ^2.
- The issue of which of the ML or REML estimates of σ^2 are closer to being unbiased is a red herring. σ^2 is not a sensible scale on which to evaluate the expected value of an estimator.

# Profile ζ plots for $\log(\sigma_1)$, σ_1 and σ_1^2
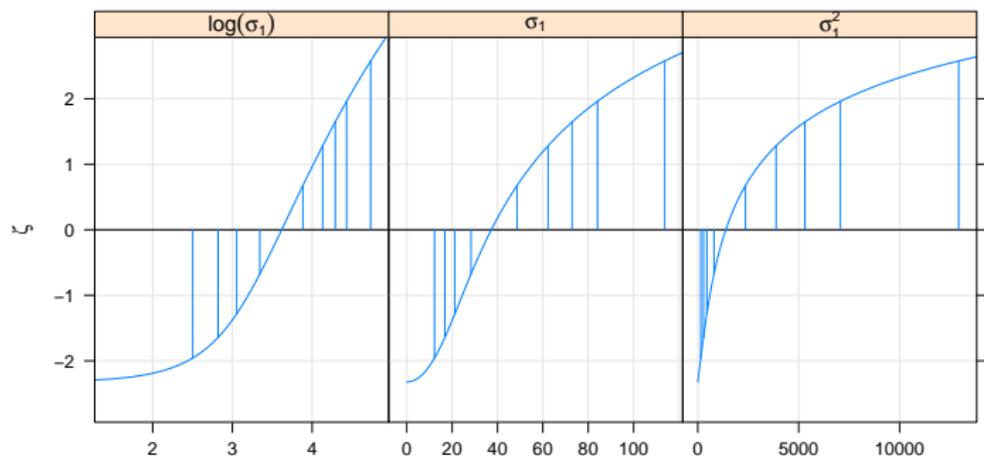
- For σ_1 the situation is more complicated because 0 is within the range of reasonable values. The profile flattens as $\sigma \to 0$ which means that intervals on $\log(\sigma)$ are unbounded.
- Obviously the estimator of σ_1^2 is terribly skewed yet most software ignores this and provides standard errors on variance component estimates.

Converting profile ζ to a density

- We speak of a profile ζ plot as showing skewness, especially for parameters such as σ_1 and σ.
- Often it is easier to envision symmetry or skewness in terms of a density plot.
- If ζ is compared to a standard Gaussian distribution then the corresponding cumulative distribution function is $\Phi(\zeta)$, from which we can derive a density function.

Profile-based densities for `fm1`
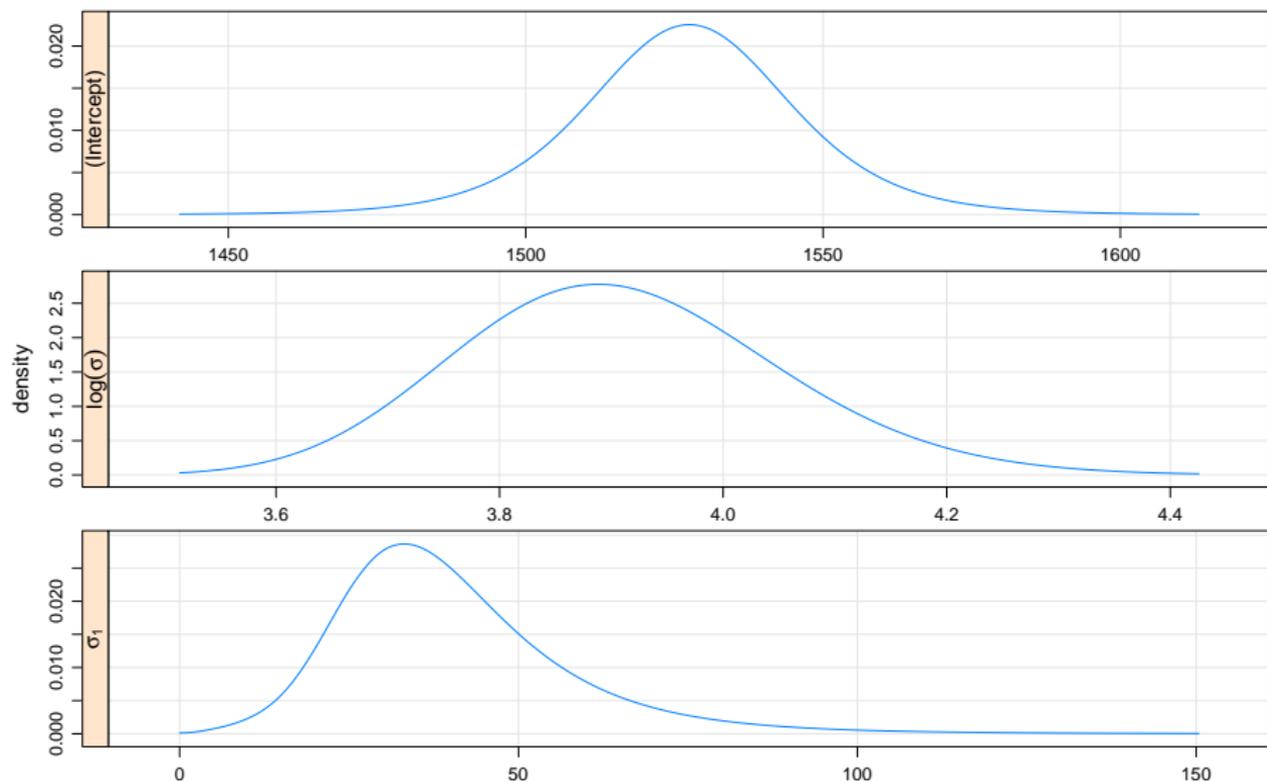
Profile pairs plots

- The information from the profile can be used to produce pairwise projections of likelihood contours. These correspond to pairwise joint confidence regions.
- Such a plot (next slide) can be somewhat confusing at first glance.
- Concentrate initially on the panels above the diagonal where the axes are the parameters in the scale shown in the diagonal panels. The contours correspond to 50%, 80%, 90%, 95% and 99% pairwise confidence regions.
- The two lines in each panel are "profile traces", which are the conditional estimate of one parameter given a value of the other.
- The actual interpolation of the contours is performed on the ζ scale which is shown in the panels below the diagonal.

Profile pairs for model **fm1**

```
> splom(pr1)
```
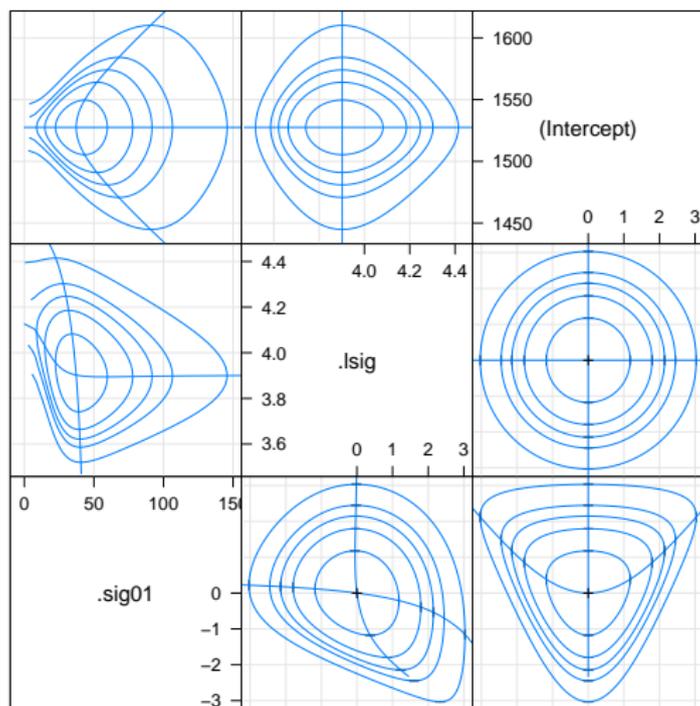

Scatter Plot Matrix

About those p-values

- Statisticians have been far too successful in propagating concepts of hypothesis testing and p-values, to the extent that quoting p-values is essentially a requirement for publication in some disciplines.

- When models were being fit by hand calculation it was important to use any trick we could come up with to simplify the calculation. Often the results were presented in terms of the simplified calculation without reference to the original idea of comparing models.

- We often still present model comparisons as properties of "terms" in the model without being explicit about the underlying comparison of models with the term and without the term.

- The approach I recommend for assessing the importance of particular terms in the fixed-effects part of the model is to fit with and without then use a likelihood ratio test (the `anova` function).

Hypothesis tests versus confidence intervals

- As mentioned earlier, hypothesis tests and confidence intervals are two sides of the same coin.
- For a categorical covariate, it often makes sense to ask "Is there a signficant effect for this factor?" which we could answer with a p-value. We may, in addition, want to know how large the effect is and how precisely we have estimated it, i.e. a confidence interval.
- For a continuous covariate we generally want to know the coefficient estimate and its precision (i.e. a confidence interval) in preference to a p-value for a hypothesis test.
- When we have many observations and only a moderate number of fixed and random effects, the distribution of the fixed-effects coefficients' estimators is well-approximated by a multivariate normal derived from the estimates, their standard errors and correlations.
- With comparatively few observations it is worthwhile using profiling to check on the sensistivity of the fit to the values of the coefficients.
- As we have seen, estimates of variance components can be poorly behaved and it is worthwhile using profiling to check their precision.

Profiling a model for the `classroom` data

```
> pr8 <- profile(fm8 <- lmer(mathgain ~ mathkind + minority +
+                  ses + (1|classid) + (1|schoolid), classroom, l
```
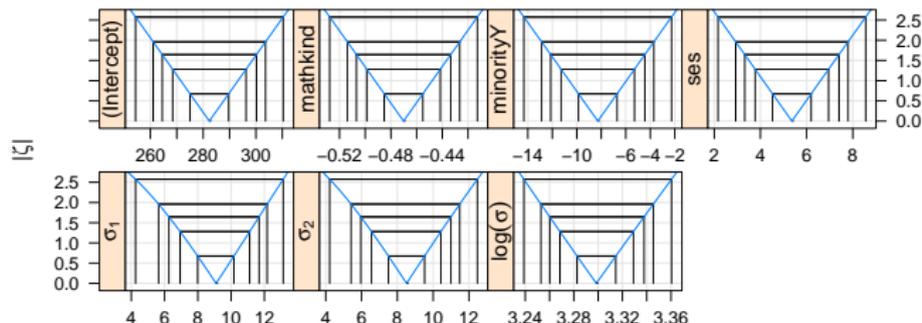

- The fixed-effects coefficient estimates (top row) have good normal approximations (i.e. a 95% confidence intervals will be closely approximated by estimate $\pm$ 1.96 $\times$ standard error).
- The estimators of σ_1, σ_2 and $\log(\sigma)$ are also well approximated by a normal. If anything, the estimators of σ_1 and σ_2 are skewed to the left rather than skewed to the right.

Profile pairs for many parameters
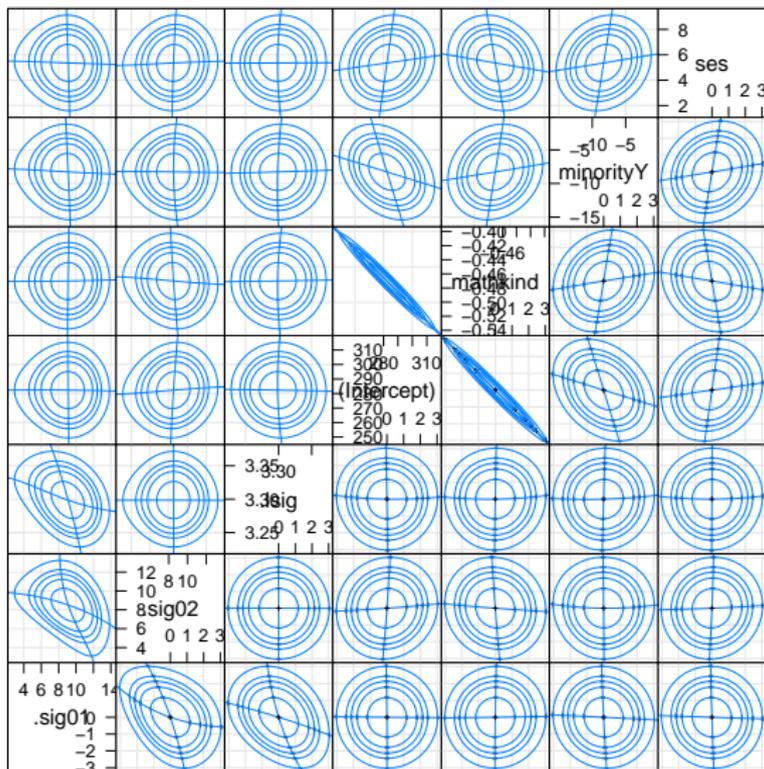
Scatter Plot Matrix

Summary

- Profile of the deviance with respect to the parameters in the model allow us to assess the variability in the parameters in terms of how well the model can be fit.

- We apply the signed square root transformation to the change in the deviance to produce ζ. When the Gaussian approximation to the distribution of the parameter estimate is appropriate, this function will be close to a straight line.

- Profile zeta plots and profile pairs plots provide visual assessment of the precision of parameter estimates.

- Typically the distribution of variance component estimates is highly skewed to the right and poorly approximated by a Gaussian, implying that standard errors of such estimates are of little value.