

# Profiling the Deviance to Assess Variability of Parameter Estimates in Mixed Models

Douglas Bates

University of Wisconsin - Madison  
<Bates@Wisc.edu>

DAEW02 Workshop, Issac Newton Institute  
August 9, 2011

## 1 Overview

# Outline

- 1 Overview
- 2 Assessing variability of parameter estimates

# Outline

- 1 Overview
- 2 Assessing variability of parameter estimates
- 2 Likelihood ratio tests and deviance profiles

# The Dyestuff data set

- The Dyestuff, Penicillin and Pastes data sets all come from the classic book *Statistical Methods in Research and Production*, edited by O.L. Davies and first published in 1947.
- The Dyestuff data are a balanced one-way classification of the Yield of dyestuff from samples produced from six Batches of an intermediate product. See `?Dyestuff`.

```
> str(Dyestuff)
```

```
'data.frame': 30 obs. of 2 variables:
```

```
$ Batch: Factor w/ 6 levels "A","B","C","D",...: 1 1 1 1 1 2 2 ..
```

```
$ Yield: num 1545 1440 1440 1520 1580 ...
```

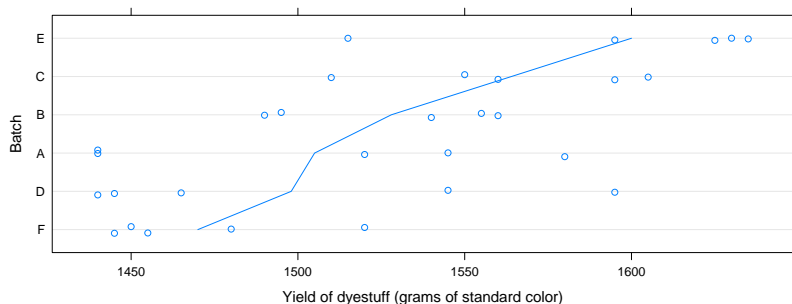
```
> summary(Dyestuff)
```

Batch	Yield
A:5	Min. :1440
B:5	1st Qu.:1469
C:5	Median :1530
D:5	Mean :1528
E:5	3rd Qu.:1575

# The effect of the batches

- To emphasize that `Batch` is categorical, we use letters instead of numbers to designate the levels.
- Because there is no inherent ordering of the levels of `Batch`, we will reorder the levels if, say, doing so can make a plot more informative.
- The particular batches observed are just a selection of the possible batches and are entirely used up during the course of the experiment.
- It is not particularly important to estimate and compare yields from these batches. Instead we wish to estimate the variability in yields due to batch-to-batch variability.
- The `Batch` factor will be used in *random-effects* terms in models that we fit.

# Dyestuff data plot



- The line joins the mean yields of the six batches, which have been reordered by increasing mean yield.
- The vertical positions are jittered slightly to reduce overplotting. The lowest yield for batch A was observed on two distinct preparations from that batch.

# A mixed-effects model for the dyestuff yield

```
> fm1 <- lmer(Yield ~ 1 + (1|Batch), Dyestuff)
> print(fm1)
```

Linear mixed model fit by REML [`'merMod'`]

Formula: `Yield ~ 1 + (1 | Batch)`

Data: `Dyestuff`

REML criterion at convergence: 319.6543

Random effects:

Groups	Name	Variance	Std.Dev.
Batch	(Intercept)	1764	42.00
	Residual	2451	49.51

Number of obs: 30, groups: Batch, 6

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1527.50	19.38	78.8

- Fitted model `fm1` has one fixed-effect parameter, the mean yield, and one random-effects term, generating a simple, scalar random effect for each level of `Batch`.



# What about models fit to smaller data sets?

- Computing power is useful even for models fit to small or medium-sized data sets. We can refit the model many, many times subject to certain restrictions to determine exactly, not through approximation, the quality of the fit.
- Most software for fitting mixed models produces estimates and their standard errors for the variance components. Why variances and not, say, standard deviations? Why standard errors for these quantities?
- The reason for variance estimates is because of Fisher's tables. The formulas for variance estimates are easier to write down.
- Why a standard error? Do we expect a confidence interval on a variance component to be symmetric?
- In the simplest situation, a single random sample from a Gaussian distribution, we use a  $\chi^2$  distribution to form an asymmetric confidence interval. It seems optimistic to expect that these quantities in much more complex models will be summarized by symmetric intervals.

## Profiling the deviance with respect to parameters

- Because of the variance component parameters in a linear mixed model the standard F and t statistics from linear models no longer have F and t distributions.
- However, if we wish to test the significance of a single term in a model, we can fit the model with and without it then form a *likelihood ratio test statistic* (which is actually the difference of the deviances, negative twice the log-likelihood).
- Likelihood ratio tests apply to any combination of model and submodel where the parameters of each have been estimated by maximum likelihood. They differ from t and F tests (and many other tests) in that they involve fitting both the “full” model (the alternative hypothesis) and the “restricted” model (the null hypothesis).
- They give a more realistic comparison of two models, at the expense of somewhat more computation. But extra computation may mean a few seconds on today’s machines and is a small expense to pay.

# Profiling the deviance versus one parameter

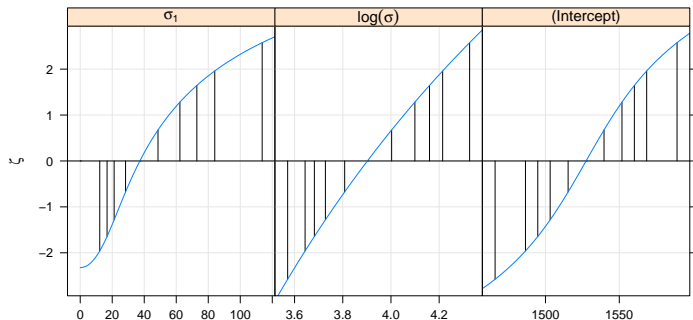
- There is a close relationship between confidence intervals and hypothesis tests on a single parameter. When, e.g.  $H_0 : \beta_1 = \beta_{1,0}$  versus  $H_a : \beta_1 \neq \beta_{1,0}$  is **not** rejected at level  $\alpha$  then  $\beta_{1,0}$  is in a  $1 - \alpha$  confidence interval on the parameter  $\beta_1$ .
- 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  $\chi_1^2$  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  $\zeta$

# Evaluating and plotting the profile

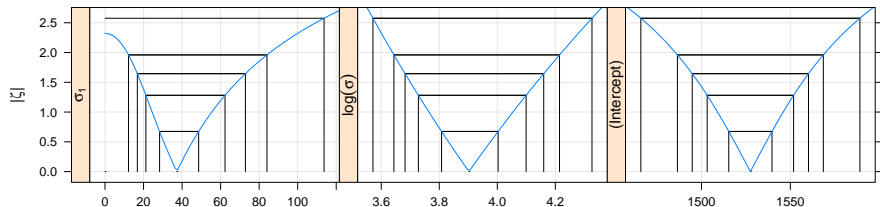
```
> pr1 <- profile(fm1M <- lmer(Yield ~ 1+(1|Batch), Dyestuff, REML = TRUE),
> xyplot(pr1, aspect=1.3)
```



- The parameters are  $\sigma_b$ ,  $\log(\sigma)$  ( $\sigma$  is the residual standard deviation) and  $\mu$ . 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= $\alpha$`  to obtain a confidence level other than the default of 0.95.

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

	0.5 %	99.5 %
<code>.sig01</code>	NA	113.692643
<code>.lsig</code>	3.571293	4.326347
<code>(Intercept)</code>	1465.874011	1589.126022

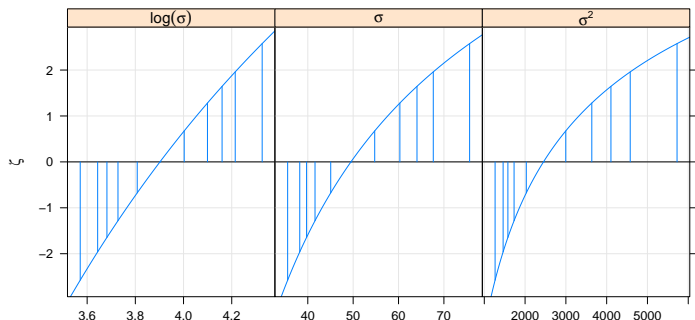
Note that the lower 99% confidence limit for  $\sigma_1$  is undefined.

# Interpreting the univariate plots

- A univariate profile  $\zeta$  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  $\sigma$  and this plot is on the scale of  $\log(\sigma)$  not  $\sigma$  or, even worse,  $\sigma^2$ .
- We should expect confidence intervals on  $\sigma^2$  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  $\chi^2$  distribution which is quite asymmetric (although many software packages provide standard errors of variance component estimates as if they were meaningful).

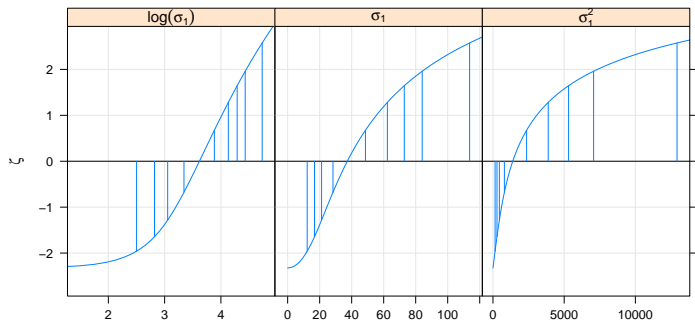


# Profile $\zeta$ plots for $\log(\sigma)$ , $\sigma$ and $\sigma^2$



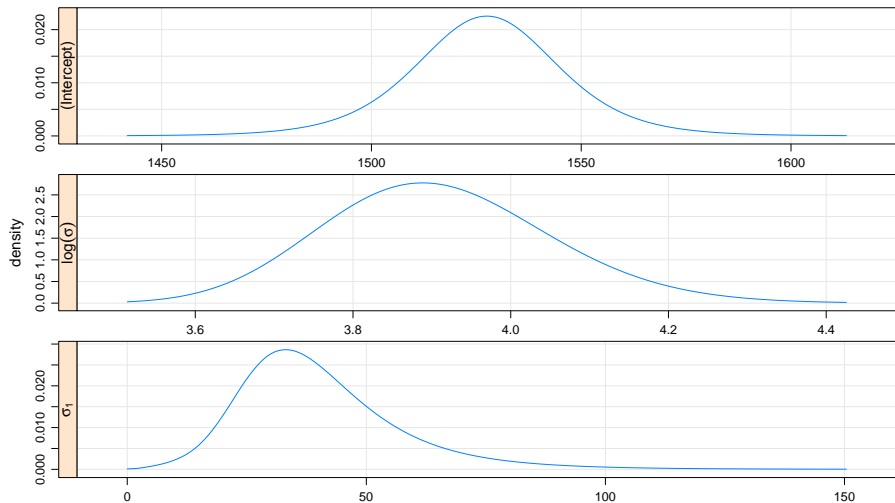
- We can see moderate asymmetry on the scale of  $\sigma$  and stronger asymmetry on the scale of  $\sigma^2$ .
- The issue of which of the ML or REML estimates of  $\sigma^2$  are closer to being unbiased is a red herring.  $\sigma^2$  is not a sensible scale on which to evaluate the expected value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

# Profile $\zeta$ plots for $\log(\sigma_1)$ , $\sigma_1$ and $\sigma_1^2$

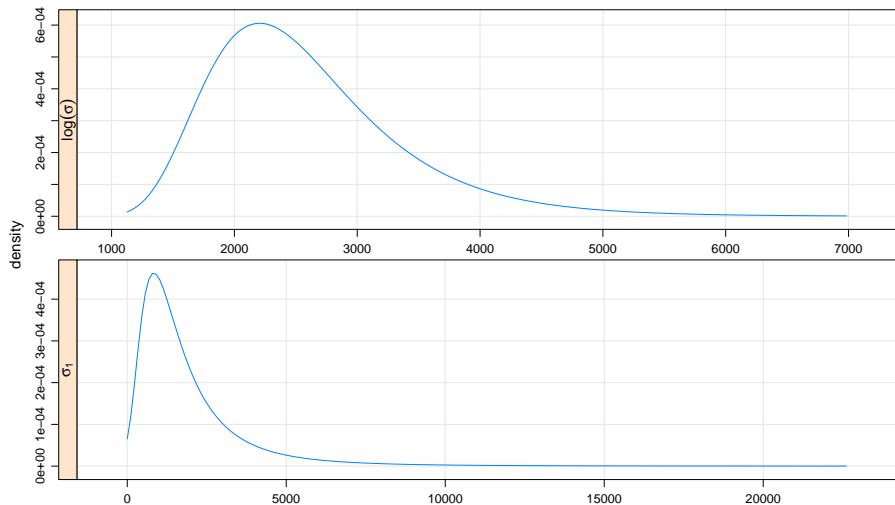


- For  $\sigma_1$  the situation is more complicated because 0 is within the range of reasonable values. The profile flattens as  $\sigma \rightarrow 0$  which means that intervals on  $\log(\sigma)$  are unbounded.
- Obviously the estimator of  $\sigma_1^2$  is terribly skewed yet most software ignores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

# Conversion to density plots



# Densities of variance components

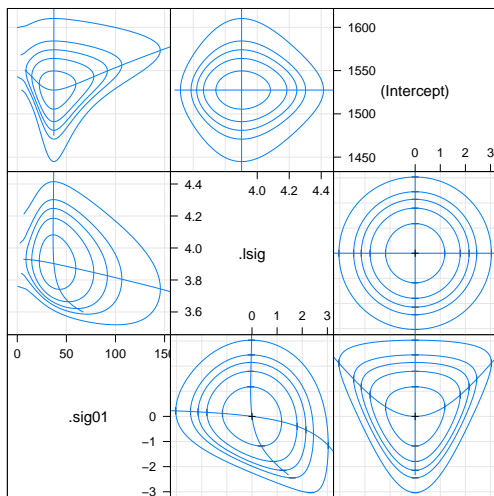


# Profile pairs plots

- 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  $\zeta$  scale which is shown in the panels below the diagonal.

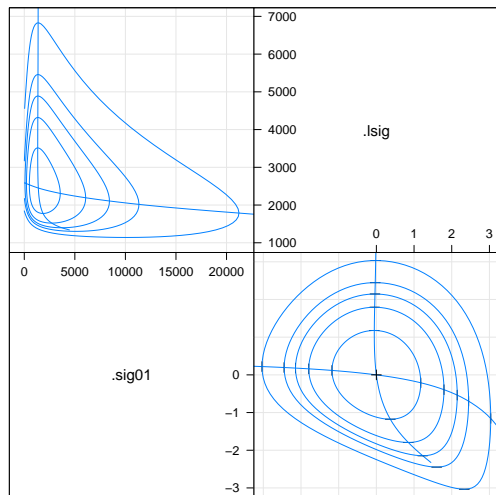
# Profile pairs for model **fm1**

```
> splom(pr1)
```



Scatter Plot Matrix

# Profile pairs for variance components in model **fm1**



Scatter Plot Matrix