

ST 762, HOMEWORK 5 SOLUTIONS, FALL 2009

1. You should have written one program that computes all of the estimators. It is not necessarily appropriate to have a separate program for each estimator; in order that the comparison be sound, all estimators must be compared over the exact same set of data sets. Thus, if one of the estimators fails to converge for a particular data set, that data set should not be included in the results for the other estimators for which it might have worked.

Here are the results I obtained using $S = 1000$. For each distribution, I did not have any problems with convergent solutions for at least one of the estimators. As long as the percentage of time this happens is very small, it is not likely to affect the relevance of the results. In a simulation, one must automate the fitting and starting values, so this does often happen; in real life, one would have the luxury of refitting with “better” starting values, etc., and would very possibly achieve a convergent solution.

NORMAL DATA
SIMULATION RESULTS FROM 1000 MONTE CARLO DATA SETS

beta1

Bias

-5.7e-05 -5.8e-05 -6.3e-05 -8.4e-05

MSE

6e-06 6e-06 6e-06 6e-06

MSE ratios

0.931535 0.923771 0.874846

MSE relative to PL

0.991665 0.939145

beta2

Bias

2.4e-05 3.1e-05 1.8e-05 -9.0e-06

MSE

3e-06 3e-06 3e-06 4e-06

MSE ratios

0.854741 0.845497 0.748060

MSE relative to PL

0.989184 0.875189

theta

Means

0.828678 0.823082 0.806702

SDs

0.107718 0.110689 0.138236

Bias

0.028678 0.023082 0.006702

MSE

```
0.012426 0.012785 0.019154
MSE ratios
0.971899 0.648719
```

Here are the results I obtained using $S = 1000$ for standardized contaminated normal data:

```
CONTAMINATED NORMAL
SIMULATION RESULTS FROM 1000 MONTE CARLO DATA SETS
```

```
beta1
```

```
Bias
```

```
-7.5e-05 -8.0e-05 -7.9e-05 -9.3e-05
```

```
MSE
```

```
6e-06 6e-06 6e-06 6e-06
```

```
MSE ratios
```

```
0.929188 0.939991 0.878962
```

```
MSE relative to PL
```

```
1.011626 0.945946
```

```
beta2
```

```
Bias
```

```
-1.3e-05 1.0e-06 -1.0e-06 -2.1e-05
```

```
MSE
```

```
3e-06 3e-06 3e-06 4e-06
```

```
MSE ratios
```

```
0.870489 0.876712 0.756127
```

```
MSE relative to PL
```

```
1.007149 0.868623
```

```
theta
```

```
Means
```

```
0.830092 0.819481 0.793819
```

```
SDs
```

```
0.130470 0.124935 0.146926
```

```
Bias
```

```
0.030092 0.019481 -0.006181
```

```
MSE
```

```
0.017928 0.015988 0.021625
```

```
MSE ratios
```

```
1.121320 0.829025
```

If one plots histograms of the estimators (not shown here; you, of course, did this), one finds that those for the elements of β look pretty normal. Those for θ look pretty normal, too, with perhaps some tendency for a long right tail indicating some skewness.

(a) From the above results, for both distributions, the estimators of β_1 and β_2 are obviously virtually unbiased. The biases are very small in every case. From the histograms, the sampling distribution does appear at least approximately normal.

(b) These ratios are given above. According to the usual first-order theory, all of these MSE ratios should be equal to 1.

For the normal distribution, there does indeed appear to be a loss of efficiency when θ is estimated rather than known for both β_1 and β_2 , with it being worse for β_2 , which is the more “nonlinear” parameter in the model. Interestingly, for the contaminated normal scenario, the efficiency loss relative to knowing θ is about the same. Also interesting is the fact that the least efficiency loss for both parameters is obtained with PL, which is normality-based, which we might expect given the second order theory, although the dominance over the identity transformation is virtually negligible. However, it does seem overall that the “folklore” result is not relevant in this finite sample situation; there is some efficiency loss due to estimating θ .

(c) According to the usual first-order theory, all of these MSE ratios should be equal to 1, as it shouldn’t matter which θ estimator was used. From the above, these MSE ratios are virtually 1 for the comparison of the identity transformation to PL in both normal and contaminated normal cases, but are somewhat less than 1 for the log transformation in both cases, especially for β_2 . The amount of efficiency loss does seem to depend on the θ estimator used; although the effect appears modest in both cases.

(d) Clearly, for this scenario and this sample size ($n = 18$), the first-order theory does not seem to overall be a great approximation to the true sampling properties of the GLS estimators for β . This is in the sense that, although the theory says that whether or not one estimates or knows θ doesn’t affect the precision of estimation of β and thus now one estimates θ also does not matter in this regard, obviously the precision is indeed dependent on estimation of θ , both in terms of having to estimate it and how it is estimated. Happily, however, all the estimators are indeed virtually unbiased, so that the consistency results seem to be realistic.

(e) The estimators for θ all show acceptable bias (the worst is less than 4% of the true value). This is probably due to the slight skewness of the sampling distribution, which may make the normal approximation suspect.

(f) The relevant rows of Table 12.1 for comparison are the first row ($\alpha = 0$) for normality and the last row ($\alpha = 0.05$) for contaminated normal. We would expect that, if the theory is relevant, the MSE ratios would be similar to the ARE values in the table.

For normal data, the relevant AREs are 0.876 (ID) and 0.405 (LOG). From above, the MSE ratios are not as dramatic as the idealized values in the table, but they do follow the same pattern (0.972 for ID and 0.649 for LOG). The results thus qualitatively reflect what the theory would predict, although ID looks much better than predicted.

For contaminated normal data, however, the results are much less dramatic than in the table. Both ID and LOG should offer increased efficiency according to the theory. ID beats PL (1.121) but not as handily as suggest by the theory. Although the theory predicts that LOG should beat PL, albeit not so dramatically, the opposite is true (0.829).

(g) For this scenario with this small sample size, the relevance of the first-order theory is questionable in some aspects. The sampling distribution of the estimators seems possibly a little skewed. Moreover, the comparisons among θ estimators, at least for contaminated

normal data, do not seem to reflect what the theory would predict. The PL estimator seems to be more robust to nonnormality in this scenario than the theory would suggest.

2. (a) All baby mice within the same mother i got the same dose; thus, $f_{i1} = f_{i2} = f_{i3} = f_i = f(x_i, \beta)$, say. Thus, letting $g_i^2 = f_i$ the covariance matrix is

$$\sigma^2 g_i^2 \begin{pmatrix} 1 & \alpha & \alpha \\ & 1 & \alpha \\ & & 1 \end{pmatrix}.$$

(b) Here, there will be 6 terms in \mathbf{u}_i , three squared terms and three cross-product terms. For simplicity, we put the crossproduct terms first. Because all baby mice within a mother got the same dose, things simplify:

$$\mathbf{u}_i = \begin{pmatrix} (Y_{i1} - f_i)(Y_{i2} - f_i) \\ (Y_{i1} - f_i)(Y_{i3} - f_i) \\ (Y_{i2} - f_i)(Y_{i3} - f_i) \\ (Y_{i1} - f_i)^2 \\ (Y_{i2} - f_i)^2 \\ (Y_{i3} - f_i)^2 \end{pmatrix}, \quad \mathbf{v}_i = \begin{pmatrix} \sigma^2 \alpha g_i^2 \\ \sigma^2 \alpha g_i^2 \\ \sigma^2 \alpha g_i^2 \\ \sigma^2 g_i^2 \\ \sigma^2 g_i^2 \\ \sigma^2 g_i^2 \end{pmatrix}, \quad \mathbf{E}_i = \begin{pmatrix} \sigma^2 g_i^2 & 2\sigma \alpha g_i^2 \\ \sigma^2 g_i^2 & 2\sigma \alpha g_i^2 \\ \sigma^2 g_i^2 & 2\sigma \alpha g_i^2 \\ 0 & 2\sigma g_i^2 \\ 0 & 2\sigma g_i^2 \\ 0 & 2\sigma g_i^2 \end{pmatrix}.$$

(c) Here, we have to use the results on pages 380 and 382 (which specializes to the Gaussian working assumption) to deduce the elements of \mathbf{Z}_i , which is a (6×6) (symmetric) matrix. Luckily, because all baby mice within mother i got the same dose, $g_{i1} = g_{i2} = g_{i3} = g_i$, so that things simplify a lot: as follows. In general,

$$\sigma^4 \begin{pmatrix} (1 + \alpha^2)g_{i1}^2 g_{i2}^2 & \alpha(1 + \alpha)g_{i1}^2 g_{i2} g_{i3} & \alpha(1 + \alpha)g_{i2}^2 g_{i1} g_{i3} & 2\alpha g_{i1}^3 g_{i2} & 2\alpha g_{i2}^3 g_{i1} & 2\alpha g_{i1} g_{i2} g_{i3} \\ & (1 + \alpha^2)g_{i1}^2 g_{i3}^2 & \alpha(1 + \alpha)g_{i3}^2 g_{i1} g_{i2} & 2\alpha g_{i1}^3 g_{i3} & 2\alpha g_{i1} g_{i3} g_{i2}^2 & 2\alpha g_{i3}^3 g_{i1} \\ & & (1 + \alpha^2)g_{i2}^2 g_{i3}^2 & 2\alpha g_{i1} g_{i2} g_{i3} & 2\alpha g_{i2}^3 g_{i3} & 2\alpha g_{i3}^3 g_{i2} \\ & & & 2g_{i1}^4 & 2\alpha^2 g_{i1}^2 g_{i2}^2 & 2\alpha^2 g_{i1}^2 g_{i3}^2 \\ & & & & 2g_{i2}^4 & 2\alpha^2 g_{i2}^2 g_{i3}^2 \\ & & & & & 2g_{i3}^4 \end{pmatrix}.$$

Because $g_{i1} = g_{i2} = g_{i3} = g_i$, this simplifies to

$$\sigma^4 g_i^4 \begin{pmatrix} (1 + \alpha^2) & \alpha(1 + \alpha) & \alpha(1 + \alpha) & 2\alpha & 2\alpha & 2\alpha \\ \alpha(1 + \alpha) & (1 + \alpha^2) & \alpha(1 + \alpha) & 2\alpha & 2\alpha & 2\alpha \\ \alpha(1 + \alpha) & \alpha(1 + \alpha) & (1 + \alpha^2) & 2\alpha & 2\alpha & 2\alpha \\ 2\alpha & 2\alpha & 2\alpha & 2 & 2\alpha^2 & 2\alpha^2 \\ 2\alpha & 2\alpha & 2\alpha & 2\alpha^2 & 2 & 2\alpha^2 \\ 2\alpha & 2\alpha & 2\alpha & 2\alpha^2 & 2\alpha^2 & 2 \end{pmatrix}.$$

Of course, this would be much more complicated if x_i were instead a within-individual covariate.

(d) See the attached program. By taking `link=log` we get the loglinear mean model; by taking `dist=gamma` we get the desired marginal variance model. See the program and output.

(e) and (f) See the programs and output. Note that the estimates of β , $\hat{\beta} = (0.0.1017, -1.3398)^T$, are identical across all fits. Also, the estimates of α and σ are identical for the `genmod` and `nlinmix` with `method=m1` fits; here, $\hat{\alpha} = 0.8058$, and $\hat{\sigma} = \sqrt{0.008689 + 0.002094} = 0.1038$. are

identical. Similarly, the estimates of α and σ using `glimmix` and `nlinmix` with the REML default are also identical.

Why should all the β estimates be the same, and why should the estimates of α and σ have the relationships they do across these fits? Why the β estimates are the same is left as an exercise. From the `proc glimmix` documentation, the default is to use something called “`rspl`”, which under the conditions here estimates α and σ^2 using a REML method. In fact, in this simple setting, you should be able to do the algebra to see what the quadratic estimating equations give as the estimates under these conditions.

3. (a) From the output (`proc means`), the proportions are relatively constant over time for the placebo subjects, but rise to a new, approximately constant level from baseline for the subjects receiving active treatment.

(b) This model assumes that the probability of having good respiratory status at baseline is $e^{\beta_1}/(1 + e^{\beta_1})$ in both groups, which makes sense given this is a randomized study. In the placebo group, this probability changes to $e^{\beta_1+\beta_2}/(1 + e^{\beta_1+\beta_2})$ for all weeks post-baseline, so β_2 is the change in log odds after the study starts for these patients. In the drug group, the baseline probability changes to $e^{\beta_1+\beta_2+\beta_3}/(1 + e^{\beta_1+\beta_2+\beta_3})$ post-baseline; thus, $\beta_2 + \beta_3$ is the change in log odds after the study starts for these patients. Accordingly, β_3 represents the difference in log odds post-baseline between the two groups. Clearly, then, if $\beta_3 = 0$, the drug has no effect.

(c) See program and output. The estimate of the unstructured working correlation matrix suggests that the correlations among all pairs of time are in the same “ballpark.” Accordingly, a compound symmetric (exchangeable) working correlation structure seems like a reasonable choice.

(d) See programs and output. The `genmod` and `glimmix` analyses are identical, while that from `proc nlinmix` is slightly different. This is probably because `nlinmix` is estimating the correlation parameter differently.

(e) From (b), the question is whether or not $\beta_3 = 0$. A simple Wald test of the null hypothesis β_3 can be conducted based on any of the fits in (d); SAS conveniently prints out the test statistic and a p-value (based on a t critical value). In all cases, the test statistic is > 4 with a p-value 0 “ < 0.0001 ,” suggesting that there is strong evidence that $\beta_3 \neq 0$. The estimate of β_3 is positive, which of course suggests that the drug increases the probability of good respiratory status. This is of course consistent with the impression one gets from the sample proportions in (a).

4. See program and output.