from FF

Paper 332-2012

Tips and Strategies for Mixed Modeling with SAS/STAT® Procedures

Kathleen Kiernan, Jill Tao, and Phil Gibbs, SAS Institute Inc., Cary, NC, USA

ABSTRACT

Inherently, mixed modeling with SAS/STAT® procedures, such as GLIMMIX, MIXED, and NLMIXED is computationally intensive. Therefore, considerable memory and CPU time can be required. The default algorithms in these procedures might fail to converge for some data sets and models. This paper provides recommendations for circumventing memory problems and reducing execution times for your mixed modeling analyses. This paper also shows how the new HPMIXED procedure can be beneficial for certain situations, as with large sparse mixed models. Lastly, the discussion focuses on the best way to interpret and address common notes, warnings, and error messages that can occur with the estimation of mixed models in SAS software.

INTRODUCTION

Over the past 20 years, mixed-modeling methodology has expanded to many areas of statistical applications. Initially, the mixed-model capabilities in the SAS System depended on the MIXED procedure. Subsequently, the NLMIXED, HPMIXED, and GLIMMIX procedures were added. SAS/STAT software is a fully integrated component of the SAS System. PROC GLIMMIX and PROC MIXED are two of the most popular procedures in SAS/STAT software that fit mixed models. Most of the questions that SAS customers have about any of these mixed-modeling procedures can be categorized into the following areas:

- providing recommendations for improving performance
- providing methods for obtaining convergence
- providing explanations of various notes, warnings, or error messages in the SAS log

This paper addresses these specific areas. It also provides recommendations for standard practices that have shown significant benefit when using PROC GLIMMIX, PROC MIXED, and PROC NLMIXED.

There are three sections in this paper. The first section provides tips on how to make programs more efficient by reducing memory and execution time. The second section provides suggestions for troubleshooting convergence problems. The last section includes a brief discussion of some of the commonly reported notes, warnings, and errors that are reported in the SAS log for a mixed model analysis using PROC GLIMMIX, PROC MIXED, or PROC NLMIXED.

SECTION 1: IMPROVING PERFORMANCE

The GLIMMIX, MIXED, and NLMIXED procedures are computationally intensive, and execution times can be long. A model might be resource intensive (requiring a large amount of memory or time) for a variety of reasons:

- The input data set is large.
- There are many levels associated with the variables in the CLASS statement that might subsequently be used in the MODEL, RANDOM, or REPEATED statements.
- The model is complex.
- Certain options are specified in the PROC, MODEL, RANDOM, or REPEATED statements.

If you have a model that encounters an out of memory error or takes too long to run, the following suggestions might be helpful.

MAKE CHANGES TO THE RUNNING ENVIRONMENT

The following changes to the running environment are good practices to implement:

- To maximize available memory on your system, close all unnecessary applications when running your program.
- If your program generates a large number of results tables, use the ODS NORESULTS statement to prevent the tracking of output objects in the Results window. For example, this suggestion is useful when using BY processing with a large number of BY groups or when using a macro to run the procedure many times.

from Adobe

Paper 332-2012

Tips and Strategies for Mixed Modeling with SAS/STAT® Procedures

Kathleen Kiernan, Jill Tao, and Phil Gibbs, SAS Institute Inc., Cary, NC, USA

ABSTRACT

Inherently, mixed modeling with SAS/STAT® procedures, such as GLIMMIX, MIXED, and NLMIXED is computationally intensive. Therefore, considerable memory and CPU time can be required. The default algorithms in these procedures might fail to converge for some data sets and models. This paper provides recommendations for circumventing memory problems and reducing execution times for your mixed modeling analyses. This paper also shows how the new HPMIXED procedure can be beneficial for certain situations, as with large sparse mixed models. Lastly, the discussion focuses on the best way to interpret and address common notes, warnings, and error messages that can occur with the estimation of mixed models in SAS software.

INTRODUCTION

Over the past 20 years, mixed-modeling methodology has expanded to many areas of statistical applications. Initially, the mixed-model capabilities in the SAS System depended on the MIXED procedure. Subsequently, the NLMIXED, HPMIXED, and GLIMMIX procedures were added. SAS/STAT software is a fully integrated component of the SAS System. PROC GLIMMIX and PROC MIXED are two of the most popular procedures in SAS/STAT software that fit mixed models. Most of the questions that SAS customers have about any of these mixed-modeling procedures can be categorized into the following areas:

- providing recommendations for improving performance
- providing methods for obtaining convergence
- providing explanations of various notes, warnings, or error messages in the SAS log

This paper addresses these specific areas. It also provides recommendations for standard practices that have shown significant benefit when using PROC GLIMMIX, PROC MIXED, and PROC NLMIXED.

There are three sections in this paper. The first section provides tips on how to make programs more efficient by reducing memory and execution time. The second section provides suggestions for troubleshooting convergence problems. The last section includes a brief discussion of some of the commonly reported notes, warnings, and errors that are reported in the SAS log for a mixed model analysis using PROC GLIMMIX, PROC MIXED, or PROC NLMIXED.

SECTION 1: IMPROVING PERFORMANCE

The GLIMMIX, MIXED, and NLMIXED procedures are computationally intensive, and execution times can be long. A model might be resource intensive (requiring a large amount of memory or time) for a variety of reasons:

- The input data set is large.
- There are many levels associated with the variables in the CLASS statement that might subsequently be used in the MODEL, RANDOM, or REPEATED statements.
- The model is complex.
- Certain options are specified in the PROC, MODEL, RANDOM, or REPEATED statements.

If you have a model that encounters an out of memory error or takes too long to run, the following suggestions might be helpful.

MAKE CHANGES TO THE RUNNING ENVIRONMENT

The following changes to the running environment are good practices to implement:

- To maximize available memory on your system, close all unnecessary applications when running your program.
- If your program generates a large number of results tables, use the ODS NORESULTS statement to prevent the tracking of output objects in the Results window. For example, this suggestion is useful when using BY processing with a large number of BY groups or when using a macro to run the procedure many times.