# SAS/STAT® 15.1 User's Guide

# The FMM Procedure

This document is an individual chapter from *SAS/STAT® 15.1 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2018. *SAS/STAT® 15.1 User's Guide*. Cary, NC: SAS Institute Inc.

**SAS/STAT® 15.1 User's Guide**

Copyright © 2018, SAS Institute Inc., Cary, NC, USA

All Rights Reserved. Produced in the United States of America.

**For a hard-copy book**: No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

**For a web download or e-book**: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

**U.S. Government License Rights; Restricted Rights:** The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication, or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a), and DFAR 227.7202-4, and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

November 2018

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

# Chapter 43
# The FMM Procedure

## Contents


Overview: FMM Procedure . . . **2874**
- Basic Features . . . 2875
- Assumptions . . . 2876
- Notation for the Finite Mixture Model . . . 2876
  - Homogeneous Mixtures . . . 2877
  - Special Mixtures . . . 2877
- PROC FMM Contrasted with Other SAS Procedures . . . 2877

Getting Started: FMM Procedure . . . **2878**
- Mixture Modeling for Binomial Overdispersion: “Student,” Pearson, Beer, and Yeast . 2878
- Modeling Zero-Inflation: Is It Better to Fish Poorly or Not to Have Fished at All? . . . 2884
- Looking for Multiple Modes: Are Galaxies Clustered? . . . 2892
  - Comparison with Roeder’s Method . . . 2899

Syntax: FMM Procedure . . . **2902**
- PROC FMM Statement . . . 2902
- BAYES Statement . . . 2915
- BY Statement . . . 2925
- CLASS Statement . . . 2926
- FREQ Statement . . . 2926
- ID Statement . . . 2926
- MODEL Statement . . . 2927
  - Response Variable Options . . . 2928
  - Model Options . . . 2930
- OUTPUT Statement . . . 2935
- PERFORMANCE Statement . . . 2938
- PROBMODEL Statement . . . 2939
- RANDSTART Statement . . . 2940
- RESTRICT Statement . . . 2942
- WEIGHT Statement . . . 2944

Details: FMM Procedure . . . **2945**
- A Gentle Introduction to Finite Mixture Models . . . 2945
  - The Form of the Finite Mixture Model . . . 2945
  - Mixture Models Contrasted with Mixing and Mixed Models: Untangling the Terminology Web . . . 2945
  - Overdispersion . . . 2947
- Log-Likelihood Functions for Response Distributions . . . 2948
- Bayesian Analysis . . . 2956

Conjugate Sampling . . . . . 2956
Metropolis-Hastings Algorithm . . . . . 2956
Latent Variables via Data Augmentation . . . . . 2957
Prior Distributions . . . . . 2958
Parameterization of Model Effects . . . . . 2959
Random Starting Values . . . . . 2959
Default Output . . . . . 2960
Model Information . . . . . 2960
Class Level Information . . . . . 2960
Number of Observations . . . . . 2960
Response Profile . . . . . 2961
Default Output for Maximum Likelihood . . . . . 2961
Default Output for Bayesian Estimation . . . . . 2963
ODS Table Names . . . . . 2964
ODS Graphics . . . . . 2967
Examples: FMM Procedure . . . . . **2967**
Example 43.1: Modeling Mixing Probabilities: All Mice Are Equal, but Some Mice Are More Equal Than Others . . . . . 2967
Example 43.2: The Usefulness of Custom Starting Values: When Do Cows Eat? . . . 2975
Example 43.3: Enforcing Homogeneity Constraints: Count and Dispersion—It Is All Over! . . . . . 2984
Example 43.4: Modeling Multinomial Overdispersion: Town and Country . . . . . 2989
References . . . . . **2994**

# Overview: FMM Procedure

The FMM procedure fits statistical models to data for which the distribution of the response is a finite mixture of distributions—that is, each response is drawn with unknown probability from one of several distributions. You can use PROC FMM to model the component distributions in addition to the mixing probabilities; see the section "A Gentle Introduction to Finite Mixture Models" on page 2945 for more precise definitions and discussion of similar but distinct modeling methodologies.

Classical statistical models are a special case of the finite mixture models in which the distribution of the data has only a single component.

Finite mixture models are useful for the following applications:

- estimating multimodal or heavy-tailed densities
- fitting zero-inflated or hurdle models to count data with excess zeros
- modeling overdispersed data
- fitting regression models with complex error distributions
- classifying observations according to predicted component probabilities

- accounting for unobservable, omitted variables
- estimating switching regressions

The FMM procedure is designed to fit finite mixtures of regression models or finite mixtures of generalized linear models in which the covariates and regression structure can be the same across components or might be different. You can fit finite mixture models by maximum likelihood or Bayesian methods.

For more information about the differences between the FMM procedure and other statistical modeling procedures in SAS/STAT software, see the section "PROC FMM Contrasted with Other SAS Procedures" on page 2877.

## Basic Features

The FMM procedure estimates the parameters in univariate finite mixture models and produces various statistics to evaluate parameters and model fit. The following list summarizes some basic features of the FMM procedure:

- maximum likelihood estimation for all models
- Markov chain Monte Carlo estimation for many models, including zero-inflated Poisson models
- many built-in link and distribution functions for modeling, including the beta, shifted *t*, Weibull, beta-binomial, and generalized Poisson distributions, in addition to many standard members of the exponential family of distributions
- specialized built-in mixture models such as binomial cluster models and multinomial cluster models (Morel and Nagaraj 1993; Morel and Neerchal 1997; Neerchal and Morel 1998)
- acceptance of multiple MODEL statements to build mixture models in which the model effects, distributions, or link functions vary across mixture components
- model-building syntax using CLASS and effect-based MODEL statements familiar from many other SAS/STAT procedures (for example, the GLM, GLIMMIX, and MIXED procedures)
- evaluation of sequences of mixture models when you specify ranges for the number of components
- simple syntax to impose linear equality and inequality constraints among parameters
- ability to model regression and classification effects in the mixing probabilities through the PROB-MODEL statement
- ability to incorporate full or partially known component membership into the analysis through the PARTIAL= option in the PROC FMM statement
- OUTPUT statement that produces a SAS data set with important statistics for interpreting mixture models, such as component log likelihoods and prior and posterior probabilities
- ability to add zero-inflation to any univariate model
- output data set with posterior parameter values for the Markov chain

- high degree of multithreading for high-performance optimization and Monte Carlo sampling
- ability to generate and evaluate multiple sets of random starting values

The FMM procedure uses ODS Graphics to create graphical output. For general information about ODS Graphics, see Chapter 21, "Statistical Graphics Using ODS." For specific information about the statistical graphics available with the FMM procedure, see the PLOTS options in the PROC FMM statement.

## Assumptions

The FMM procedure makes the following assumptions in fitting statistical models:

- The number of components $k$ in the finite mixture is known a priori and is not a parameter to be estimated.
- The parameters of the components are distinct a priori.
- The observations are uncorrelated.

## Notation for the Finite Mixture Model

The general expression for the finite mixture model fitted with the FMM procedure is as follows:

$$f(y) = \sum_{j=1}^{k} \pi_j(\mathbf{z}, \boldsymbol{\alpha}_j) p_j(y; \mathbf{x}_j' \boldsymbol{\beta}_j, \phi_j)$$

The number of components in the mixture is denoted as $k$. The mixture probabilities $\pi_j$ can depend on regressor variables $\mathbf{z}$ and parameters $\boldsymbol{\alpha}_j$. By default, the FMM procedure models these probabilities using a logit transform if $k = 2$ and as a generalized logit model if $k > 2$. The component distributions $p_j$ can also depend on regressor variables in $\mathbf{x}_j$, regression parameters $\boldsymbol{\beta}_j$, and possibly scale parameters $\phi_j$. Notice that the component distributions $p_j$ are indexed by $j$ because the distributions might belong to different families. For example, in a two-component model, you might model one component as a normal (Gaussian) variable and the second component as a variable with a $t$ distribution with low degrees of freedom to manage overdispersion.

The mixture probabilities $\pi_j$ satisfy $\pi_j \geq 0$, for all $j$, and

$$\sum_{j=1}^{k} \pi_j(\mathbf{z}, \boldsymbol{\alpha}_j) = 1$$

### Homogeneous Mixtures

If the component distributions are of the same distributional form, the mixture is called homogeneous. In most applications of homogeneous mixtures, the mixing probabilities do not depend on regression parameters. The general model then simplifies to

$$f(y) = \sum_{j=1}^{k} \pi_j \, p(y; \mathbf{x}'\boldsymbol{\beta}_j, \phi_j)$$

Because the component distributions depend on regression parameters $\boldsymbol{\beta}_j$, this model is known as a homogeneous regression mixture. A homogeneous regression mixture assumes that the regression effects are the same across the components, although the FMM procedure does not impose such a restriction. If the component distributions do not contain regression effects, the model

$$f(y) = \sum_{j=1}^{k} \pi_j \, p(y; \mu_j, \phi_j)$$

is *the* homogeneous mixture model. A classical case is the estimation of a continuous density as a $k$-component mixture of normal distributions.

### Special Mixtures

The FMM procedure enables you to fit several special mixture models. The Morel-Neerchal binomial cluster model (Morel and Nagaraj 1993; Morel and Neerchal 1997; Neerchal and Morel 1998) is a mixture of binomial distributions in which the success probabilities depend on the mixing probabilities. The multinomial cluster model is a generalization of the binomial cluster model. It is a mixture of multinomial distributions in which the outcome probability vector depends on the mixing probabilities.

Zero-inflated count models are obtained as two-component mixtures where one component is a classical count model—such as the Poisson or negative binomial model—and the other component is a distribution that is concentrated at zero. If the nondegenerate part of this special mixture is a zero-truncated model, the resulting two-component mixture is known as a hurdle model (Cameron and Trivedi 1998).

## PROC FMM Contrasted with Other SAS Procedures

Because the FMM procedure fits finite mixtures of generalized linear models, it can also fit standard forms of these models in which the distribution of the data does not follow a mixture. This enables you to use the FMM procedure to estimate parameters in models that can be fit with the CATMOD, LOGISTIC, GENMOD, or GLIMMIX procedures. However, the FMM procedure does not fit models that have random effects.

The FMM procedure has limited postprocessing capabilities compared to some other statistical procedures that are based on linear models. Concepts that are well understood and commonplace in linear models, such as (linear) estimable functions, estimability, and least squares means, do not apply to mixture models in the same way. For example, even the computation of a predicted value is not without ambiguity. You can estimate the means in the component distributions in addition to the overall mean of the mixture.

The FMM procedure provides a limited number of built-in distributions and link functions. User-defined distributions or link functions are not supported. Mixture models with component distributions that are not supported by the FMM procedure can be fit with the NLMIXED procedure.

For Bayesian estimation, the FMM procedure implements a small number of highly specialized sampling algorithms. These algorithms are very efficient and specifically designed for generalized linear models and their mixtures. This limits, for example, the allowable specifications for prior distributions of the model parameters. Models that do not fit the targeted algorithms of the FMM procedure can be fit with the MCMC procedure.

# Getting Started: FMM Procedure

## Mixture Modeling for Binomial Overdispersion: "Student," Pearson, Beer, and Yeast

The following example demonstrates how you can model a complicated, two-component binomial mixture distribution, either with maximum likelihood or with Bayesian methods, with a few simple PROC FMM statements.

William Sealy Gosset, a chemist at the Arthur Guinness Son and Company brewery in Dublin, joined the statistical laboratory of Karl Pearson in 1906–1907 to study statistics. At first Gosset—who published all but one paper under the pseudonym "Student" because his employer forbade publications by employees after a co-worker had disclosed trade secrets—worked on the Poisson limit to the binomial distribution, using haemacytometer yeast cell counts. Gosset's interest in studying small-sample (and limit) problems was motivated by the small sample sizes he typically saw in his work at the brewery.

Subsequently, Gosset's yeast count data have been examined and revisited by many authors. In 1915, Karl Pearson undertook his own examination and realized that the variability in "Student's" data exceeded that consistent with a Poisson distribution. Pearson (1915) bemoans the fact that if this were so, "it is certainly most unfortunate that such material should have been selected to illustrate Poisson's limit to the binomial."

Using a count of Gosset's yeast cell counts on the 400 squares of a haemacytometer (Table 43.1), Pearson argues that a mixture process would explain the heterogeneity (beyond the Poisson).

**Table 43.1** "Student's" Yeast Cell Counts

| **Number of Cells** | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| **Frequency** | 213 | 128 | 37 | 18 | 3 | 1 |

Pearson fits various models to these data, chief among them a mixture of two binomial series

$$\nu_1(p_1 + q_1)^\theta + \nu_2(p_2 + q_2)^\theta$$

where $\theta$ is real-valued and thus the binomial series expands to

$$(p+q)^\theta = \sum_{k=0}^{\infty} \frac{\Gamma(\theta+1)}{\Gamma(k+1)\Gamma(\theta-k+1)}\, p^k\, q^{\theta-k}$$

Pearson's fitted model has $\theta = 4.89997$, $\nu_1 = 356.986$, $\nu_2 = 43.014$ (corresponding to a mixing proportion of $356.986/(43.014 + 356.986) = 0.892$), and estimated success probabilities in the binomial components of 0.1017 and 0.4514, respectively. The success probabilities indicate that although the data have about a 90% chance of coming from a distribution with small success probability of about 0.1, there is a 10% chance of coming from a distribution with a much larger success probability of about 0.45.

If $\theta$ is an integer, the binomial series is the cumulative mass function of a binomial random variable. The value of $\theta$ suggests that a suitable model for these data could also be constructed as a two-component mixture of binomial random variables as follows:

$$f(y) = \pi \,\mathrm{binomial}(5, \mu_1) + (1 - \pi)\,\mathrm{binomial}(5, \mu_2)$$

The binomial sample size $n$=5 is suggested by Pearson's estimate of $\theta = 4.89997$ and the fact that the largest cell count in Table 43.1 is 5.

The following DATA step creates a SAS data set from the data in Table 43.1.

```
data yeast;
   input count f;
   n = 5;
   datalines;
   0     213
   1     128
   2      37
   3      18
   4       3
   5       1
;
```

The two-component binomial model is fit with the FMM procedure with the following statements:

```
proc fmm data=yeast;
   model count/n =  / k=2;
   freq f;
run;
```

Because the events/trials syntax is used in the MODEL statement, PROC FMM defaults to the binomial distribution. The K=2 option specifies that the number of components is fixed and known to be two. The FREQ statement indicates that the data are grouped; for example, the first observation represents 213 squares on the haemacytometer where no yeast cells were found.

The "Model Information" and "Number of Observations" tables in Figure 43.1 convey that the fitted model is a two-component homogeneous binomial mixture with a logit link function. The mixture is *homogeneous* because there are no model effects in the MODEL statement and because both component distributions belong to the same distributional family. By default, PROC FMM estimates the model parameters by maximum likelihood.

Although only six observations are read from the data set, the data represent 400 observations (squares on the haemacytometer). Because a constant binomial sample size of 5 is assumed, the data represent 273 successes (finding a yeast cell) out of 2,000 Bernoulli trials.

**Figure 43.1** Model Information for Yeast Cell Model

**The FMM Procedure**

| Model Information | |
|---|---|
| **Data Set** | WORK.YEAST |
| **Response Variable (Events)** | count |
| **Response Variable (Trials)** | n |
| **Frequency Variable** | f |
| **Type of Model** | Homogeneous Mixture |
| **Distribution** | Binomial |
| **Components** | 2 |
| **Link Function** | Logit |
| **Estimation Method** | Maximum Likelihood |

| | |
|---|---|
| **Number of Observations Read** | 6 |
| **Number of Observations Used** | 6 |
| **Sum of Frequencies Read** | 400 |
| **Sum of Frequencies Used** | 400 |
| **Number of Events** | 273 |
| **Number of Trials** | 2000 |

The estimated intercepts (on the logit scale) for the two binomial means are –2.2316 and –0.2974, respectively. These values correspond to binomial success probabilities of 0.09695 and 0.4262, respectively (Figure 43.2). The two components mix with probabilities 0.8799 and 1–0.8799 = 0.1201. These values are generally close to the values found by Pearson (1915) using infinite binomial series instead of binomial mass functions.

**Figure 43.2** Maximum Likelihood Estimates

**Parameter Estimates for Binomial Model**

| Component | Parameter | Estimate | Standard Error | z Value | Pr > \|z\| | Inverse Linked Estimate |
|---|---|---|---|---|---|---|
| 1 | Intercept | -2.2316 | 0.1522 | -14.66 | <.0001 | 0.09695 |
| 2 | Intercept | -0.2974 | 0.3655 | -0.81 | 0.4158 | 0.4262 |

**Parameter Estimates for Mixing Probabilities**

| | | Linked Scale | | | |
|---|---|---|---|---|---|
| Component | Mixing Probability | Logit(Prob) | Standard Error | z Value | Pr > \|z\| |
| 1 | 0.8799 | 1.9913 | 0.5725 | 3.48 | 0.0005 |
| 2 | 0.1201 | -1.9913 | | | |

To obtain fitted values and other observationwise statistics under the stipulated two-component model, you can add the OUTPUT statement to the previous PROC FMM run. The following statements request componentwise predicted values and the posterior probabilities:

```
proc fmm data=yeast;
   model count/n =  / k=2;
   freq f;
   output out=fmmout pred(components) posterior;
```

```
run;
data fmmout;
   set fmmout;
   PredCount_1 = post_1 * f;
   PredCount_2 = post_2 * f;
run;
proc print data=fmmout;
run;
```

The DATA step following the PROC FMM step computes the predicted cell counts in each component (Figure 43.3). The predicted means in the components, 0.48476 and 2.13099, are close to the values determined by Pearson (0.4983 and 2.2118), as are the predicted cell counts.

**Figure 43.3** Predicted Cell Counts

| Obs | count | f | n | Pred_1 | Pred_2 | Post_1 | Post_2 | PredCount_1 | PredCount_2 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 213 | 5 | 0.096951 | 0.42620 | 0.98606 | 0.01394 | 210.030 | 2.9698 |
| 2 | 1 | 128 | 5 | 0.096951 | 0.42620 | 0.91089 | 0.08911 | 116.594 | 11.4058 |
| 3 | 2 | 37 | 5 | 0.096951 | 0.42620 | 0.59638 | 0.40362 | 22.066 | 14.9341 |
| 4 | 3 | 18 | 5 | 0.096951 | 0.42620 | 0.17598 | 0.82402 | 3.168 | 14.8323 |
| 5 | 4 | 3 | 5 | 0.096951 | 0.42620 | 0.02994 | 0.97006 | 0.090 | 2.9102 |
| 6 | 5 | 1 | 5 | 0.096951 | 0.42620 | 0.00444 | 0.99556 | 0.004 | 0.9956 |

Gosset, who was interested in small-sample statistical problems, investigated the use of prior knowledge in mathematical-statistical analysis—for example, deriving the sampling distribution of the correlation coefficient after having assumed a uniform prior distribution for the coefficient in the population (Aldrich 1997). Pearson also was not opposed to using prior information, especially uniform priors that reflect "equal distribution of ignorance." Fisher, on the other hand, would not have any of it: the best estimator in his opinion is obtained by a criterion that is absolutely independent of prior assumptions about probabilities of particular values. He objected to the insinuation that his derivations in the work on the correlation were deduced from Bayes theorem (Fisher 1921).

The preceding analysis of the yeast cell count data uses maximum likelihood methods that are free of prior assumptions. The following analysis takes instead a Bayesian approach, assuming a beta prior distribution for the binomial success probabilities and a uniform prior distribution for the mixing probabilities. The changes from the previous FMM run are the addition of the ODS GRAPHICS, PERFORMANCE, and BAYES statements and the SEED=12345 option.

```
ods graphics on;
proc fmm data=yeast seed=12345;
   model count/n = / k=2;
   freq f;
   performance cpucount=2;
   bayes;
run;
ods graphics off;
```

With ODS Graphics enabled, PROC FMM produces diagnostic trace plots for the posterior samples. Bayesian analyses are sensitive to the random number seed and thread count; the SEED= and CPUCOUNT= options ensure consistent results for the purposes of this example. The SEED=12345 option in the PROC FMM statement determines the random number seed for the random number generator used in the analysis. The

CPUCOUNT=2 option in the PERFORMANCE statement sets the number of available processors to two. The BAYES statement requests a Bayesian analysis.

The "Bayes Information" table in Figure 43.4 provides basic information about the Markov chain Monte Carlo sampler. Because the model is a homogeneous mixture, the FMM procedure applies an efficient conjugate sampling algorithm with a posterior sample size of 10,000 samples after a burn-in size of 2,000 samples. The "Prior Distributions" table displays the prior distribution for each parameter along with its mean and variance and the initial value in the chain. Notice that in this situation all three prior distributions reduce to a uniform distribution on (0, 1).

**Figure 43.4** Basic Information about MCMC Sampler

**The FMM Procedure**

| Bayes Information | |
|---|---|
| **Sampling Algorithm** | Conjugate |
| **Data Augmentation** | Latent Variable |
| **Initial Values of Chain** | Data Based |
| **Burn-In Size** | 2000 |
| **MC Sample Size** | 10000 |
| **MC Thinning** | 1 |
| **Parameters in Sampling** | 3 |
| **Mean Function Parameters** | 2 |
| **Scale Parameters** | 0 |
| **Mixing Prob Parameters** | 1 |
| **Number of Threads** | 2 |

**Prior Distributions**

| Component | Parameter | Distribution | Mean | Variance | Initial Value |
|---|---|---|---|---|---|
| 1 | **Success Probability** | Beta(1, 1) | 0.5000 | 0.08333 | 0.1365 |
| 2 | **Success Probability** | Beta(1, 1) | 0.5000 | 0.08333 | 0.1365 |
| 1 | **Probability** | Dirichlet(1, 1) | 0.5000 | 0.08333 | 0.6180 |

The FMM procedure produces a log note for this model, indicating that the sampled quantities are not the linear predictors on the logit scale, but are the actual population parameters (on the data scale):

```
NOTE: Bayesian results for this model (no regressor variables,
      non-identity link) are displayed on the data scale, not the
      linked scale. You can obtain results on the linked (=linear)
      scale by requesting a Metropolis-Hastings sampling algorithm.
```

The trace panel for the success probability in the first binomial component is shown in Figure 43.5. Note that the first component in this Bayesian analysis corresponds to the second component in the MLE analysis. The plots in this panel can be used to diagnose the convergence of the Markov chain. If the chain has not converged, inferences cannot be made based on quantities derived from the chain. You generally look for the following:

- a smooth unimodal distribution of the posterior estimates in the density plot displayed on the lower right

- good mixing of the posterior samples in the trace plot at the top of the panel (good mixing is indicated when the trace traverses the support of the distribution and appears to have reached a stationary distribution)

**Figure 43.5** Trace Panel for Success Probability in First Component

The autocorrelation plot in Figure 43.5 shows fairly high and sustained autocorrelation among the posterior estimates. While this is generally not a problem, you can affect the degree of autocorrelation among the posterior estimates by running a longer chain and thinning the posterior estimates; see the NMC= and THIN= options in the BAYES statement.

Both the trace plot and the density plot in Figure 43.5 are indications of successful convergence.

Figure 43.6 reports selected results that summarize the 10,000 posterior samples. The arithmetic means of the success probabilities in the two components are 0.3884 and 0.0905, respectively. The posterior mean of the mixing probability is 0.1771. These values are similar to the maximum likelihood parameter estimates in Figure 43.2 (after swapping components).

**Figure 43.6** Summaries for Posterior Estimates

| Posterior Summaries | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | Percentiles | | |
| Component | Parameter | N | Mean | Standard Deviation | 25 | 50 | 75 |
| 1 | Success Probability | 10000 | 0.3884 | 0.0861 | 0.3254 | 0.3835 | 0.4457 |
| 2 | Success Probability | 10000 | 0.0905 | 0.0162 | 0.0811 | 0.0923 | 0.1017 |
| 1 | Probability | 10000 | 0.1771 | 0.0978 | 0.1073 | 0.1534 | 0.2227 |

| Posterior Intervals | | | | | | |
|---|---|---|---|---|---|---|
| | | | Equal-Tail | | | |
| Component | Parameter | Alpha | Interval | | HPD Interval | |
| 1 | Success Probability | 0.050 | 0.2355 | 0.5663 | 0.2224 | 0.5494 |
| 2 | Success Probability | 0.050 | 0.0538 | 0.1171 | 0.0572 | 0.1187 |
| 1 | Probability | 0.050 | 0.0564 | 0.4311 | 0.0424 | 0.3780 |

Note that the standard errors in Figure 43.2 are not comparable to those in Figure 43.6, because the standard errors for the MLEs are expressed on the logit scale and the Bayesian estimates are expressed on the data scale. You can add the METROPOLIS option in the BAYES statement to sample the quantities on the logit scale.

The "Posterior Intervals" table in Figure 43.6 displays 95% credible intervals (equal-tail intervals and intervals of highest posterior density). It can be concluded that the component with the higher success probability contributes less than 40% to the process.

## Modeling Zero-Inflation: Is It Better to Fish Poorly or Not to Have Fished at All?

The following example shows how you can use PROC FMM to model data that contain more zero values than expected.

Many count data show an excess of zeros relative to the frequency of zeros expected under a reference model. An excess of zeros leads to overdispersion because the process is more variable than a standard count data model. Different mechanisms can lead to excess zeros. For example, suppose that the data are generated from two processes with different distribution functions—one process generates the zero counts, and the other process generates nonzero counts. In the vernacular of Cameron and Trivedi (1998), such a model is called a *hurdle* model. With a certain probability—the probability of a nonzero count—a hurdle is crossed, and events are being generated. Hurdle models are useful, for example, to model the number of doctor visits per year. Once the decision to see a doctor has been made—the hurdle has been overcome—a certain number of visits follow.

Hurdle models are closely related to zero-inflated models. Both can be expressed as two-component mixtures in which one component has a degenerate distribution at zero and the other component is a count model. In a hurdle model, the count model follows a zero-truncated distribution. In a zero-inflated model, the count

model has a nonzero probability of generating zeros. Formally, a zero-inflated model can be written as

$$\Pr(Y = y) = \pi p_1 + (1 - \pi) p_2(y, \mu)$$
$$p_1 = \begin{cases} 1 & y = 0 \\ 0 & \text{otherwise} \end{cases}$$

where $p_2(y, \mu)$ is a standard count model with mean $\mu$ and support $y \in \{0, 1, 2, \ldots\}$.

The following data illustrates the use of a zero-inflated model. In a survey of park attendees, randomly selected individuals were asked about the number of fish they caught in the last six months. Along with that count, the gender and age of each sampled individual was recorded. The following DATA step displays the data for the analysis:

```
data catch;
   input gender $ age count @@;
   datalines;
  F   54  18   M   37   0   F   48  12   M   27   0
  M   55   0   M   32   0   F   49  12   F   45  11
  M   39   0   F   34   1   F   50   0   M   52   4
  M   33   0   M   32   0   F   23   1   F   17   0
  F   44   5   M   44   0   F   26   0   F   30   0
  F   38   0   F   38   0   F   52  18   M   23   1
  F   23   0   M   32   0   F   33   3   M   26   0
  F   46   8   M   45   5   M   51  10   F   48   5
  F   31   2   F   25   1   M   22   0   M   41   0
  M   19   0   M   23   0   M   31   1   M   17   0
  F   21   0   F   44   7   M   28   0   M   47   3
  M   23   0   F   29   3   F   24   0   M   34   1
  F   19   0   F   35   2   M   39   0   M   43   6
;
```

At first glance, the prevalence of zeros in the DATA set is apparent. Many park attendees did not catch any fish. These zero counts are made up of two populations: attendees who do not fish and attendees who fish poorly. A zero-inflation mechanism thus appears reasonable for this application because a zero count can be produced by two separate distributions.

The following statements fit a standard Poisson regression model to these data. A common intercept is assumed for men and women, and the regression slope varies with gender.

```
proc fmm data=catch;
   class gender;
   model count = gender*age / dist=Poisson;
run;
```

Figure 43.7 displays information about the model and data set. The "Model Information" table conveys that the model is a single-component Poisson model (a Poisson GLM) and that parameters are estimated by maximum likelihood. There are two levels in the CLASS variable gender, with females preceding males.

**Figure 43.7** Model Information and Class Levels in Poisson Regression

**The FMM Procedure**

| Model Information | |
|---|---|
| **Data Set** | WORK.CATCH |
| **Response Variable** | count |
| **Type of Model** | Generalized Linear (GLM) |
| **Distribution** | Poisson |
| **Components** | 1 |
| **Link Function** | Log |
| **Estimation Method** | Maximum Likelihood |

| Class Level Information | | |
|---|---|---|
| **Class** | **Levels** | **Values** |
| **gender** | 2 | F M |

| | |
|---|---|
| **Number of Observations Read** | 52 |
| **Number of Observations Used** | 52 |

The "Fit Statistics" and "Parameter Estimates" tables from the maximum likelihood estimation of the Poisson GLM are shown in Figure 43.8. If the model is not overdispersed, the Pearson statistic should roughly equal the number of observations in the data set minus the number of parameters. With $n$=52, there is evidence of overdispersion in these data.

**Figure 43.8** Fit Results in Poisson Regression

| Fit Statistics | |
|---|---|
| **-2 Log Likelihood** | 182.7 |
| **AIC (Smaller is Better)** | 188.7 |
| **AICC (Smaller is Better)** | 189.2 |
| **BIC (Smaller is Better)** | 194.6 |
| **Pearson Statistic** | 85.9573 |

| Parameter Estimates for Poisson Model | | | | | |
|---|---|---|---|---|---|
| **Effect** | **gender** | **Estimate** | **Standard Error** | **z Value** | **Pr > \|z\|** |
| **Intercept** | | -3.9811 | 0.5439 | -7.32 | <.0001 |
| **age*gender** | F | 0.1278 | 0.01149 | 11.12 | <.0001 |
| **age*gender** | M | 0.1044 | 0.01224 | 8.53 | <.0001 |

Suppose that the cause of overdispersion is zero-inflation of the count data. The following statements fit a zero-inflated Poisson model.

```
proc fmm data=catch;
   class gender;
   model count = gender*age / dist=Poisson ;
   model       +            / dist=Constant;
run;
```

There are two MODEL statements, one for each component of the mixture. Because the distributions are different for the components, you cannot specify the mixture model with a single MODEL statement. The first MODEL statement identifies the response variable for the model (count) and defines a Poisson model with intercept and gender-specific slopes. The second MODEL statement uses the continuation operator ("+") and adds a model with a degenerate distribution by using DIST=CONSTANT. Because the mass of the constant is placed by default at zero, the second MODEL statement adds a zero-inflation component to the model. It is sufficient to specify the response variable in one of the MODEL statements; you use the "=" sign in that statement to separate the response variable from the model effects.

Figure 43.9 displays the "Model Information" and "Optimization Information" tables for this run of the FMM procedure. The model is now identified as a zero-inflated Poisson (ZIP) model with two components, and the parameters continue to be estimated by maximum likelihood. The "Optimization Information" table shows that there are four parameters in the optimization (compared to three parameters in the Poisson GLM model). The four parameters correspond to three parameters in the mean function (intercept and two gender-specific slopes) and the mixing probability.

**Figure 43.9** Model and Optimization Information in the ZIP Model

**The FMM Procedure**

| Model Information | |
|---|---|
| **Data Set** | WORK.CATCH |
| **Response Variable** | count |
| **Type of Model** | Zero-inflated Poisson |
| **Components** | 2 |
| **Estimation Method** | Maximum Likelihood |

| Optimization Information | |
|---|---|
| **Optimization Technique** | Dual Quasi-Newton |
| **Parameters in Optimization** | 4 |
| **Mean Function Parameters** | 3 |
| **Scale Parameters** | 0 |
| **Mixing Prob Parameters** | 1 |
| **Number of Threads** | 4 |

Results from fitting the ZIP model by maximum likelihood are shown in Figure 43.10. The –2 log likelihood and the information criteria suggest a much-improved fit over the single-component Poisson model (compare Figure 43.10 to Figure 43.8). The Pearson statistic is reduced by factor 2 compared to the Poisson model and suggests a better fit than the standard Poisson model.

**Figure 43.10** Maximum Likelihood Results for the ZIP model

| Fit Statistics | |
|---|---|
| **-2 Log Likelihood** | 145.6 |
| **AIC (Smaller is Better)** | 153.6 |
| **AICC (Smaller is Better)** | 154.5 |
| **BIC (Smaller is Better)** | 161.4 |
| **Pearson Statistic** | 43.4467 |
| **Effective Parameters** | 4 |
| **Effective Components** | 2 |

**Figure 43.10** *continued*

| Parameter Estimates for Poisson Model | | | | | | |
|---|---|---|---|---|---|---|
| Component | Effect | gender | Estimate | Standard Error | z Value | Pr > \|z\| |
| 1 | Intercept | | -3.5215 | 0.6448 | -5.46 | <.0001 |
| 1 | age*gender | F | 0.1216 | 0.01344 | 9.04 | <.0001 |
| 1 | age*gender | M | 0.1056 | 0.01394 | 7.58 | <.0001 |

| Parameter Estimates for Mixing Probabilities | | | | | |
|---|---|---|---|---|---|
| | | Linked Scale | | | |
| Component | Mixing Probability | Logit(Prob) | Standard Error | z Value | Pr > \|z\| |
| 1 | 0.6972 | 0.8342 | 0.4768 | 1.75 | 0.0802 |
| 2 | 0.3028 | -0.8342 | | | |

The number of effective parameters and components shown in Figure 43.8 equals the values from Figure 43.9. This is not always the case because components can collapse (for example, when the mixing probability approaches zero or when two components have identical parameter estimates). In this example, both components and all four parameters are identifiable. The Poisson regression and the zero process mix, with a probability of approximately 0.6972 attributed to the Poisson component.

The FMM procedure enables you to fit some mixture models by Bayesian techniques. The following statements add the BAYES statement to the previous PROC FMM statements:

```
proc fmm data=catch seed=12345;
   class gender;
   model count = gender*age / dist=Poisson;
   model       +            / dist=constant;
   performance cpucount=2;
   bayes;
run;
```

The "Model Information" table indicates that the model parameters are estimated by Markov chain Monte Carlo techniques, and it displays the random number seed (Figure 43.11). This is useful if you did not specify a seed to identify the seed value that reproduces the current analysis. The "Bayes Information" table provides basic information about the Monte Carlo sampling scheme. The sampling method uses a data augmentation scheme to impute component membership and then the Gamerman (1997) algorithm to sample the component-specific parameters. The 2,000 burn-in samples are followed by 10,000 Monte Carlo samples without thinning.

**Figure 43.11** Model, Bayes, and Prior Information in the ZIP Model

**The FMM Procedure**

| Model Information | |
|---|---|
| **Data Set** | WORK.CATCH |
| **Response Variable** | count |
| **Type of Model** | Zero-inflated Poisson |
| **Components** | 2 |
| **Estimation Method** | Markov Chain Monte Carlo |
| **Random Number Seed** | 12345 |

| Bayes Information | |
|---|---|
| **Sampling Algorithm** | Gamerman |
| **Data Augmentation** | Latent Variable |
| **Initial Values of Chain** | ML Estimates |
| **Burn-In Size** | 2000 |
| **MC Sample Size** | 10000 |
| **MC Thinning** | 1 |
| **Parameters in Sampling** | 4 |
| **Mean Function Parameters** | 3 |
| **Scale Parameters** | 0 |
| **Mixing Prob Parameters** | 1 |
| **Number of Threads** | 2 |

**Prior Distributions**

| Component | Effect | gender | Distribution | Mean | Variance | Initial Value |
|---|---|---|---|---|---|---|
| 1 | **Intercept** | | Normal(0, 1000) | 0 | 1000.00 | -3.5215 |
| 1 | **age*gender** | F | Normal(0, 1000) | 0 | 1000.00 | 0.1216 |
| 1 | **age*gender** | M | Normal(0, 1000) | 0 | 1000.00 | 0.1056 |
| 1 | **Probability** | | Dirichlet(1, 1) | 0.5000 | 0.08333 | 0.6972 |

The "Prior Distributions" table identifies the prior distributions, their parameters for the sampled quantities, and their initial values. The prior distribution of parameters associated with model effects is a normal distribution with mean 0 and variance 1,000. The prior distribution for the mixing probability is a Dirichlet(1,1), which is identical to a uniform distribution (Figure 43.11). Because the second mixture component is a degeneracy at zero with no associated parameters, it does not appear in the "Prior Distributions" table in Figure 43.11.

Figure 43.12 displays descriptive statistics about the 10,000 posterior samples. Recall from Figure 43.10 that the maximum likelihood estimates were –3.5215, 0.1216, 0.1056, and 0.6972, respectively. With this choice of prior, the means of the posterior samples are generally close to the MLEs in this example. The "Posterior Intervals" table displays 95% intervals of equal-tail probability and 95% intervals of highest posterior density (HPD) intervals.

**Figure 43.12** Posterior Summaries and Intervals in the ZIP Model

| Posterior Summaries | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Percentiles | | |
| Component | Effect | gender | N | Mean | Standard Deviation | 25 | 50 | 75 |
| 1 | Intercept | | 10000 | -3.5524 | 0.6509 | -3.9922 | -3.5359 | -3.0875 |
| 1 | age*gender | F | 10000 | 0.1220 | 0.0136 | 0.1124 | 0.1218 | 0.1314 |
| 1 | age*gender | M | 10000 | 0.1058 | 0.0140 | 0.0961 | 0.1055 | 0.1153 |
| 1 | Probability | | 10000 | 0.6938 | 0.0945 | 0.6293 | 0.6978 | 0.7605 |

| Posterior Intervals | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | Equal-Tail Interval | | HPD Interval | |
| Component | Effect | gender | Alpha | | | | |
| 1 | Intercept | | 0.050 | -4.8693 | -2.3222 | -4.8927 | -2.3464 |
| 1 | age*gender | F | 0.050 | 0.0960 | 0.1494 | 0.0961 | 0.1494 |
| 1 | age*gender | M | 0.050 | 0.0792 | 0.1339 | 0.0796 | 0.1341 |
| 1 | Probability | | 0.050 | 0.5041 | 0.8688 | 0.5025 | 0.8666 |

You can generate trace plots for the posterior parameter estimates by enabling ODS Graphics:

```
ods graphics on;
ods select TADPanel;
proc fmm data=catch seed=12345;
   class gender;
   model count = gender*age / dist=Poisson;
   model       +            / dist=constant;
   performance cpucount=2;
   bayes;
run;
ods graphics off;
```

A separate trace panel is produced for each sampled parameter, and the panels for the gender-specific slopes are shown in Figure 43.13. There is good mixing in the chains: the modest autocorrelation that diminishes after about 10 successive samples. By default, the FMM procedure transfers the credible intervals for each parameter from the "Posterior Intervals" table to the trace plot and the density plot in the trace panel.

**Figure 43.13** Trace Panels for Gender-Specific Slopes

**Figure 43.13** *continued*

## Looking for Multiple Modes: Are Galaxies Clustered?

Mixture modeling is essentially a generalized form of one-dimensional cluster analysis. The following example shows how you can use PROC FMM to explore the number and nature of Gaussian clusters in univariate data.

Roeder (1990) presents data from the Corona Borealis sky survey with the velocities of 82 galaxies in a narrow slice of the sky. Cosmological theory suggests that the observed velocity of each galaxy is proportional to its distance from the observer. Thus, the presence of multiple modes in the density of these velocities could indicate a clustering of the galaxies at different distances.

The following DATA step recreates the data set in Roeder (1990). The computed variable v represents the measured velocity in thousands of kilometers per second.

```
title "FMM Analysis of Galaxies Data";
data galaxies;
   input velocity @@;
   v = velocity / 1000;
```

```
   datalines;
9172  9350  9483  9558  9775  10227 10406 16084 16170 18419
18552 18600 18927 19052 19070 19330 19343 19349 19440 19473
19529 19541 19547 19663 19846 19856 19863 19914 19918 19973
19989 20166 20175 20179 20196 20215 20221 20415 20629 20795
20821 20846 20875 20986 21137 21492 21701 21814 21921 21960
22185 22209 22242 22249 22314 22374 22495 22746 22747 22888
22914 23206 23241 23263 23484 23538 23542 23666 23706 23711
24129 24285 24289 24366 24717 24990 25633 26960 26995 32065
32789 34279
;
```

Analysis of potentially multimodal data is a natural application of finite mixture models. In this case, the modeling is complicated by the question of the variance for each of the components. Using identical variances for each component could obscure underlying structure, but the additional flexibility granted by component-specific variances might introduce spurious features.

You can use PROC FMM to prepare analyses for equal and unequal variances and use one of the available fit statistics to compare the resulting models. You can use the model selection facility to explore models that have varying numbers of mixture components—say, from three to seven as investigated in Roeder (1990). The following statements select the best unequal-variance model by using Akaike's information criterion (AIC), which has a built-in penalty for model complexity:

```
title2 "Three to Seven Components, Unequal Variances";
ods graphics on;
proc fmm data=galaxies criterion=AIC;
   model v = / kmin=3 kmax=7;
run;
```

The KMIN= and KMAX= options indicate the smallest and largest number of components to consider. Figure 43.14, Figure 43.15, and Figure 43.16 show a selection of the output from this unequal variances model.

**Figure 43.14** Model Selection for Galaxy Data Assuming Unequal Variances

**FMM Analysis of Galaxies Data**
**Three to Seven Components, Unequal Variances**

**The FMM Procedure**

| Model Information | |
|---|---|
| **Data Set** | WORK.GALAXIES |
| **Response Variable** | v |
| **Type of Model** | Homogeneous Mixture |
| **Distribution** | Normal |
| **Min Components** | 3 |
| **Max Components** | 7 |
| **Link Function** | Identity |
| **Estimation Method** | Maximum Likelihood |

**Figure 43.14** *continued*

| Component Evaluation for Mixture Models | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Number of Components | | Number of Parameters | | | | | | | |
| Model ID | Total | Eff. | Total | Eff. | -2 Log L | AIC | AICC | BIC | Pearson | Max Gradient |
| 1 | 3 | 3 | 8 | 8 | 406.96 | 422.96 | 424.94 | 442.22 | 82.00 | 0.000024 |
| 2 | 4 | 4 | 11 | 11 | 406.96 | 428.96 | 432.74 | 455.44 | 82.00 | 0.00012 |
| 3 | 5 | 5 | 14 | 14 | 406.96 | 434.96 | 441.23 | 468.66 | 82.00 | 0.000040 |
| 4 | 6 | 6 | 17 | 17 | 406.96 | 440.96 | 450.53 | 481.88 | 82.00 | 0.000026 |
| 5 | 7 | 7 | 20 | 20 | 406.96 | 446.96 | 460.73 | 495.10 | 82.00 | 0.00018 |

**The model with 3 components (ID=1) was selected as 'best' based on the AIC statistic.**

| Fit Statistics | |
|---|---|
| -2 Log Likelihood | 407.0 |
| AIC (Smaller is Better) | 423.0 |
| AICC (Smaller is Better) | 424.9 |
| BIC (Smaller is Better) | 442.2 |
| Pearson Statistic | 82.0002 |
| Effective Parameters | 8 |
| Effective Components | 3 |

| Parameter Estimates for Normal Model | | | | | |
|---|---|---|---|---|---|
| Component | Parameter | Estimate | Standard Error | z Value | Pr > \|z\| |
| 1 | Intercept | 9.7101 | 0.1597 | 60.80 | <.0001 |
| 2 | Intercept | 33.0444 | 0.5322 | 62.09 | <.0001 |
| 3 | Intercept | 21.4039 | 0.2597 | 82.41 | <.0001 |
| 1 | Variance | 0.1785 | 0.09542 | | |
| 2 | Variance | 0.8496 | 0.6937 | | |
| 3 | Variance | 4.8567 | 0.8098 | | |

| Parameter Estimates for Mixing Probabilities | | | | | |
|---|---|---|---|---|---|
| | | Linked Scale | | | |
| Component | Mixing Probability | GLogit(Prob) | Standard Error | z Value | Pr > \|z\| |
| 1 | 0.0854 | -2.3308 | 0.3959 | -5.89 | <.0001 |
| 2 | 0.0366 | -3.1781 | 0.5893 | -5.39 | <.0001 |
| 3 | 0.8781 | 0 | | | |

**Figure 43.15** Density Plot for Best (Three-Component) Model Assuming Unequal Variances

**Figure 43.16** Criterion Panel Plot for Model Selection Assuming Unequal Variances

This example uses the AIC for model selection. Figure 43.16 shows the AIC and other model fit criteria for each of the fitted models.

To require that the separate components have identical variances, add the EQUATE=SCALE option in the MODEL statement:

```
title2 "Three to Seven Components, Equal Variances";
proc fmm data=galaxies criterion=AIC gconv=0;
   model v = / kmin=3 kmax=7  equate=scale;
run;
```

The GCONV= convergence criterion is turned off in this PROC FMM run to avoid the early stoppage of the iterations when the relative gradient changes little between iterations. Turning the criterion off usually ensures that convergence is achieved with a small absolute gradient of the objective function.

The output for equal variances is shown in Figure 43.17 and Figure 43.18.

**Figure 43.17** Model Selection for Galaxy Data Assuming Equal Variances

**FMM Analysis of Galaxies Data**
**Three to Seven Components, Equal Variances**

**The FMM Procedure**

| Model Information | |
|---|---|
| **Data Set** | WORK.GALAXIES |
| **Response Variable** | v |
| **Type of Model** | Homogeneous Mixture |
| **Distribution** | Normal |
| **Min Components** | 3 |
| **Max Components** | 7 |
| **Link Function** | Identity |
| **Estimation Method** | Maximum Likelihood |

**Component Evaluation for Mixture Models**

| Model ID | Number of Components Total | Eff. | Number of Parameters Total | Eff. | -2 Log L | AIC | AICC | BIC | Pearson | Max Gradient |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 3 | 6 | 6 | 478.74 | 490.74 | 491.86 | 505.18 | 82.00 | 1.197E-6 |
| 2 | 4 | 4 | 8 | 8 | 416.49 | 432.49 | 434.47 | 451.75 | 82.00 | 2.98E-6 |
| 3 | 5 | 5 | 10 | 10 | 416.49 | 436.49 | 439.59 | 460.56 | 82.00 | 4.312E-6 |
| 4 | 6 | 6 | 12 | 12 | 416.49 | 440.49 | 445.02 | 469.37 | 82.00 | 6.294E-6 |
| 5 | 7 | 7 | 14 | 14 | 416.49 | 444.49 | 450.76 | 478.19 | 82.00 | 4.884E-6 |

**The model with 4 components (ID=2) was selected as 'best' based on the AIC statistic.**

| Fit Statistics | |
|---|---|
| **-2 Log Likelihood** | 416.5 |
| **AIC (Smaller is Better)** | 432.5 |
| **AICC (Smaller is Better)** | 434.5 |
| **BIC (Smaller is Better)** | 451.7 |
| **Pearson Statistic** | 82.0000 |
| **Effective Parameters** | 8 |
| **Effective Components** | 4 |

**Parameter Estimates for Normal Model**

| Component | Parameter | Estimate | Standard Error | z Value | Pr > \|z\| |
|---|---|---|---|---|---|
| 1 | Intercept | 23.5058 | 0.3460 | 67.93 | <.0001 |
| 2 | Intercept | 33.0440 | 0.7610 | 43.42 | <.0001 |
| 3 | Intercept | 20.0086 | 0.3029 | 66.06 | <.0001 |
| 4 | Intercept | 9.7103 | 0.4981 | 19.50 | <.0001 |
| 1 | Variance | 1.7354 | 0.3905 | | |
| 2 | Variance | 1.7354 | 0.3905 | | |
| 3 | Variance | 1.7354 | 0.3905 | | |
| 4 | Variance | 1.7354 | 0.3905 | | |

**Figure 43.17** *continued*

| Parameter Estimates for Mixing Probabilities | | | | | |
|---|---|---|---|---|---|
| | | Linked Scale | | | |
| Component | Mixing Probability | GLogit(Prob) | Standard Error | z Value | Pr > \|z\| |
| **1** | 0.3503 | 1.4118 | 0.4497 | 3.14 | 0.0017 |
| **2** | 0.0366 | -0.8473 | 0.6901 | -1.23 | 0.2195 |
| **3** | 0.5277 | 1.8216 | 0.4205 | 4.33 | <.0001 |
| **4** | 0.0854 | 0 | | | |

**Figure 43.18** Density Plot for Best (Four-Component) Model Assuming Equal Variances

Not surprisingly, the two variance specifications produce different optimal models. The unequal variance specification favors a three-component model while the equal variance specification favors a four-component model. Comparison of the AIC fit statistics, 423.0 and 432.5, indicates that the three-component, unequal variance model provides the best overall fit.

## Comparison with Roeder's Method

It is important to note that Roeder's original analysis proceeds in a different manner than the finite mixture modeling presented here. The technique presented by Roeder first develops a "best" range of scale parameters based on a specific criterion. Roeder then uses fixed scale parameters taken from this range to develop optimal equal-scale Gaussian mixture models.

You can reproduce Roeder's point estimate for the density by specifying a five-component Gaussian mixture. In addition, use the EQUATE=SCALE option in the MODEL statement and a RESTRICT statement fixing the first component's scale parameter at 0.9025 (Roeder's $h = 0.95$, scale$= h^2$). The combination of these options produces a mixture of five Gaussian components, each with variance 0.9025. The following statements conduct this analysis:

```
title2 "Five Components, Equal Variances = 0.9025";
proc fmm data=galaxies;
   model v = / K=5 equate=scale;
   restrict int 0 (scale 1) = 0.9025;
run;
```

The output is shown in Figure 43.19 and Figure 43.20.

**Figure 43.19** Reproduction of Roeder's Five-Component Analysis of Galaxy Data

**FMM Analysis of Galaxies Data**
**Five Components, Equal Variances = 0.9025**

**The FMM Procedure**

| Model Information | |
|---|---|
| **Data Set** | WORK.GALAXIES |
| **Response Variable** | v |
| **Type of Model** | Homogeneous Mixture |
| **Distribution** | Normal |
| **Components** | 5 |
| **Link Function** | Identity |
| **Estimation Method** | Maximum Likelihood |

| Fit Statistics | |
|---|---|
| **-2 Log Likelihood** | 412.2 |
| **AIC (Smaller is Better)** | 430.2 |
| **AICC (Smaller is Better)** | 432.7 |
| **BIC (Smaller is Better)** | 451.9 |
| **Pearson Statistic** | 82.5549 |
| **Effective Parameters** | 9 |
| **Effective Components** | 5 |

| Linear Constraints at Solution | | |
|---|---|---|
| **k = 1** | | **Constraint Active** |
| Variance = | 0.90 | Yes |

**Figure 43.19** *continued*

| Parameter Estimates for Normal Model | | | | | |
|---|---|---|---|---|---|
| Component | Parameter | Estimate | Standard Error | z Value | Pr > \|z\| |
| 1 | Intercept | 26.3266 | 0.7778 | 33.85 | <.0001 |
| 2 | Intercept | 33.0443 | 0.5485 | 60.25 | <.0001 |
| 3 | Intercept | 9.7101 | 0.3591 | 27.04 | <.0001 |
| 4 | Intercept | 23.0295 | 0.2294 | 100.38 | <.0001 |
| 5 | Intercept | 19.7187 | 0.1784 | 110.55 | <.0001 |
| 1 | Variance | 0.9025 | 0 | | |
| 2 | Variance | 0.9025 | 0 | | |
| 3 | Variance | 0.9025 | 0 | | |
| 4 | Variance | 0.9025 | 0 | | |
| 5 | Variance | 0.9025 | 0 | | |

| Parameter Estimates for Mixing Probabilities | | | | | |
|---|---|---|---|---|---|
| | | Linked Scale | | | |
| Component | Mixing Probability | GLogit(Prob) | Standard Error | z Value | Pr > \|z\| |
| 1 | 0.0397 | -2.4739 | 0.7084 | -3.49 | 0.0005 |
| 2 | 0.0366 | -2.5544 | 0.6016 | -4.25 | <.0001 |
| 3 | 0.0854 | -1.7071 | 0.4141 | -4.12 | <.0001 |
| 4 | 0.3678 | -0.2466 | 0.2699 | -0.91 | 0.3609 |
| 5 | 0.4706 | 0 | | | |

**Figure 43.20** Density Plot for Roeder's Analysis

# Syntax: FMM Procedure

The following statements are available in the FMM procedure:

```
PROC FMM <options> ;
    BAYES bayes-options ;
    BY variables ;
    CLASS variables </ TRUNCATE> ;
    FREQ variable ;
    ID variables ;
    MODEL response<(response-options)> = <effects> </ model-options> ;
    MODEL events/trials = <effects> </ model-options> ;
    MODEL              + <effects> </ model-options> ;
    OUTPUT <OUT=SAS-data-set>
           <keyword<(keyword-options)> <=name>> ...
           <keyword<(keyword-options)> <=name>> </ options> ;
    PERFORMANCE performance-options ;
    PROBMODEL <effects> </ probmodel-options> ;
    RANDSTART <randstart-options> ;
    RESTRICT <'label'> constraint-specification <, ..., constraint-specification>
             <operator <value>> </ option> ;
    WEIGHT variable ;
```

The PROC FMM statement and at least one MODEL statement is required. The CLASS, RESTRICT, and MODEL statements can appear multiple times. If a CLASS statement is specified, it must precede the MODEL statements. The RESTRICT statements must appear after the MODEL statements.

## PROC FMM Statement

```
PROC FMM <options> ;
```

The PROC FMM statement invokes the FMM procedure. Table 43.2 summarizes the *options* available in the PROC FMM statement. These and other *options* in the PROC FMM statement are then described fully in alphabetical order.

**Table 43.2** PROC FMM Statement Options

| Option | Description |
|---|---|
| **Basic Options** | |
| DATA= | Specifies the input data set |
| EXCLUSION= | Specifies how the procedure responds to support violations in the data |
| NAMELEN= | Specifies the length of effect names |
| ORDER= | Determines the sort order of CLASS variables |
| SEED= | Specifies the random number seed for analyses that require random number draws |

**Table 43.2** *continued*

| Option | Description |
|---|---|
| **Displayed Output** | |
| COMPONENTINFO | Displays information about the mixture components |
| CORR | Displays the asymptotic correlation matrix of the maximum likelihood parameter estimates or the empirical correlation matrix of the Bayesian posterior estimates |
| COV | Displays the asymptotic covariance matrix of the maximum likelihood parameter estimates or the empirical covariance matrix of the Bayesian posterior estimates |
| COVI | Displays the inverse of the covariance matrix of the parameter estimates |
| FITDETAILS | Displays fit information for all examined models |
| ITDETAILS | Adds estimates and gradients to the "Iteration History" table |
| NOCLPRINT | Suppresses the "Class Level Information" table completely or partially |
| NOITPRINT | Suppresses the "Iteration History Information" table |
| NOPRINT | Suppresses tabular and graphical output |
| PARMSTYLE= | Specifies how parameters are displayed in ODS tables |
| PLOTS | Produces ODS statistical graphics |
| **Computational Options** | |
| CRITERION= | Specifies the criterion used in model selection |
| NOCENTER | Prevents centering and scaling of the regressor variables |
| PARTIAL= | Specifies a variable that defines a partial classification |
| **Options Related to Optimization** | |
| ABSCONV= | Tunes an absolute function convergence criterion |
| ABSFCONV= | Tunes an absolute function difference convergence criterion |
| ABSGCONV= | Tunes the absolute gradient convergence criterion |
| FCONV= | Specifies a relative function convergence criterion that is based on a relative change of the function value |
| FCONV2= | Specifies a relative function convergence criterion that is based on a predicted reduction of the objective function |
| GCONV= | Tunes the relative gradient convergence criterion |
| MAXITER= | Specifies the maximum number of iterations in any optimization |
| MAXFUNC= | Specifies the maximum number of function evaluations in any optimization |
| MAXTIME= | Specifies the upper limit of CPU time in seconds for any optimization |
| MINITER= | Specifies the minimum number of iterations in any optimization |
| TECHNIQUE= | Selects the optimization technique |

**Table 43.2** *continued*

| Option | Description |
|---|---|
| **Singularity Tolerances** | |
| INVALIDLOGL= | Tunes the value assigned to an invalid component log likelihood |
| SINGCHOL= | Tunes singularity for Cholesky decompositions |
| SINGRES= | Tunes singularity for the residual variance |
| SINGULAR= | Tunes general singularity criterion |
| ZEROPROB= | Tunes component weight threshold for number of effective components |

You can specify the following *options* in the PROC FMM statement.

**ABSCONV=***r*

**ABSTOL=***r*

specifies an absolute function convergence criterion. For minimization, the termination criterion is $f(\boldsymbol{\psi}^{(k)}) \leq r$, where $\boldsymbol{\psi}$ is the vector of parameters in the optimization and $f(\cdot)$ is the objective function. The default value of *r* is the negative square root of the largest double-precision value, which serves only as a protection against overflows.

**ABSFCONV=***r*

**ABSFTOL=***r*

specifies an absolute function difference convergence criterion. For all techniques except NMSIMP, the termination criterion is a small change of the function value in successive iterations:

$$|f(\boldsymbol{\psi}^{(k-1)}) - f(\boldsymbol{\psi}^{(k)})| \leq r$$

Here, $\boldsymbol{\psi}$ denotes the vector of parameters that participate in the optimization, and $f(\cdot)$ is the objective function. The same formula is used for the NMSIMP technique, but $\boldsymbol{\psi}(k)$ is defined as the vertex with the lowest function value, and $\boldsymbol{\psi}^{(k-1)}$ is defined as the vertex with the highest function value in the simplex. The default value is *r*=0.

**ABSGCONV=***r*

**ABSGTOL=***r*

specifies an absolute gradient convergence criterion. The termination criterion is a small maximum absolute gradient element:

$$\max_j |g_j(\boldsymbol{\psi}^{(k)})| \leq r$$

Here, $\boldsymbol{\psi}$ denotes the vector of parameters that participate in the optimization, and $g_j(\cdot)$ is the gradient of the objective function with respect to the *j*th parameter. This criterion is not used by the NMSIMP technique. The default value is *r*=1E–5.

**COMPONENTINFO**
**COMPINFO**
**CINFO**
produces a table with additional details about the fitted model components.

**COV**
produces the covariance matrix of the parameter estimates. For maximum likelihood estimation, this matrix is based on the inverse (projected) Hessian matrix. For Bayesian estimation, it is the empirical covariance matrix of the posterior estimates. The covariance matrix is shown for all parameters, even if they did not participate in the optimization or sampling.

**COVI**
produces the inverse of the covariance matrix of the parameter estimates. For maximum likelihood estimation, the covariance matrix is based on the inverse (projected) Hessian matrix. For Bayesian estimation, it is the empirical covariance matrix of the posterior estimates. This matrix is then inverted by sweeping, and rows and columns that correspond to linear dependencies or singularities are zeroed.

**CORR**
produces the correlation matrix of the parameter estimates. For maximum likelihood estimation this matrix is based on the inverse (projected) Hessian matrix. For Bayesian estimation, it is based on the empirical covariance matrix of the posterior estimates.

**CRITERION=***keyword*
**CRIT=***keyword*
specifies the criterion by which the FMM procedure ranks models when multiple models are evaluated during maximum likelihood estimation. You can choose from the following criteria to rank models by specifying the appropriate *keyword*:

| | |
|---|---|
| **AIC** | uses Akaike's information criterion. |
| **AICC** | uses the bias-corrected AIC criterion. |
| **BIC** | uses the Bayesian information criterion. |
| **GRADIENT** | uses the largest element of the gradient (in absolute value). |
| **LOGL \| LL** | uses the mixture log likelihood. |
| **PEARSON** | uses the Pearson statistic. |

The default for maximum likelihood estimation is CRITERION=BIC.

In Bayesian model selection, you can choose from the following criteria to rank models:

| | |
|---|---|
| **DIC** | uses the deviance information criterion. |
| **LOGL \| LL** | uses the mixture log likelihood |

The default for Bayesian estimation is CRITERION=DIC.

**DATA=***SAS-data-set*
names the SAS data set to be used by PROC FMM. The default is the most recently created data set.

**EXCLUSION=NONE | ANY | ALL**
**EXCLUDE=NONE | ANY | ALL**

specifies how the FMM procedure handles support violations of observations. For example, in a mixture of two Poisson variables, negative response values are not possible. However, in a mixture of a Poisson and a normal variable, negative values are possible, and their likelihood contribution to the Poisson component is zero. An observation that violates the support of one component distribution of the model might be a valid response with respect to one or more other component distributions. This requires some nuanced handling of support violations in mixture models.

The default exclusion technique, EXCLUSION=ALL, removes an observation from the analysis only if it violates the support of all component distributions. The other extreme, EXCLUSION=NONE, permits an observation into the analysis regardless of support violations. EXCLUSION=ANY removes observations from the analysis if the response violates the support of any component distributions. In the single-component case, EXCLUSION=ALL and EXCLUSION=ANY are identical.

**FCONV=***r*
**FTOL=***r*

specifies a relative function convergence criterion that is based on the relative change of the function value. For all techniques except NMSIMP, PROC FMM terminates when there is a small relative change of the function value in successive iterations:

$$\frac{|f(\boldsymbol{\psi}^{(k)}) - f(\boldsymbol{\psi}^{(k-1)})|}{|f(\boldsymbol{\psi}^{(k-1)})|} \leq r$$

Here, $\boldsymbol{\psi}$ denotes the vector of parameters that participate in the optimization, and $f(\cdot)$ is the objective function. The same formula is used for the NMSIMP technique, but $\boldsymbol{\psi}^{(k)}$ is defined as the vertex with the lowest function value, and $\boldsymbol{\psi}^{(k-1)}$ is defined as the vertex with the highest function value in the simplex.

The default is $r = 10^{-\mathrm{FDIGITS}}$, where FDIGITS is by default $-\log_{10}\{\epsilon\}$, and $\epsilon$ is the machine precision.

**FCONV2=***r*
**FTOL2=***r*

specifies a relative function convergence criterion that is based on the predicted reduction of the objective function. For all techniques except NMSIMP, the termination criterion is a small predicted reduction

$$df^{(k)} \approx f(\boldsymbol{\theta}^{(k)}) - f(\boldsymbol{\theta}^{(k)} + \mathbf{s}^{(k)})$$

of the objective function. The predicted reduction

$$\begin{aligned} df^{(k)} &= -\mathbf{g}^{(k)\prime}\mathbf{s}^{(k)} - \frac{1}{2}\mathbf{s}^{(k)\prime}\mathbf{H}^{(k)}\mathbf{s}^{(k)} \\ &= -\frac{1}{2}\mathbf{s}^{(k)\prime}\mathbf{g}^{(k)} \\ &\leq r \end{aligned}$$

is computed by approximating the objective function $f$ by the first two terms of the Taylor series and substituting the Newton step:

$$\mathbf{s}^{(k)} = -[\mathbf{H}^{(k)}]^{-1}\mathbf{g}^{(k)}$$

For the NMSIMP technique, the termination criterion is a small standard deviation of the function values of the $n+1$ simplex vertices $\boldsymbol{\theta}_l^{(k)}$, $l = 0, \ldots, n$,

$$\sqrt{\frac{1}{n+1}\sum_l \left[f(\boldsymbol{\theta}_l^{(k)}) - \overline{f}(\boldsymbol{\theta}^{(k)})\right]^2} \le r$$

where $\overline{f}(\boldsymbol{\theta}^{(k)}) = \frac{1}{n+1}\sum_l f(\boldsymbol{\theta}_l^{(k)})$. If there are $n_{act}$ boundary constraints active at $\boldsymbol{\theta}^{(k)}$, the mean and standard deviation are computed only for the $n+1-n_{act}$ unconstrained vertices.

The default value is *r* = 1E–6 for the NMSIMP technique and *r* = 0 otherwise.

**FITDETAILS**

requests that the "Optimization Information," "Iteration History," "Convergence Status," and "Fit Statistics" tables be produced for all optimizations when models with different number of components are evaluated. For example, the following statements fit a binomial regression model with up to three components and produces fit and optimization information for all three:

```
proc fmm fitdetails;
   model y/n  = x / kmax=3;
run;
```

Without the FITDETAILS option, only the "Fit Statistics" table for the selected model is displayed.

In Bayesian estimation, the FITDETAILS option displays the following tables for each model that the procedure fits: "Bayes Information," "Iteration History," "Prior Information," "Fit Statistics," "Posterior Summaries," "Posterior Intervals," and any requested diagnostics tables. The "Iteration History" table appears only if the BAYES statement includes the INITIAL=MLE option.

Without the FITDETAILS option, these tables are listed only for the selected model.

**GCONV=***r*

**GTOL=***r*

specifies a relative gradient convergence criterion. For all techniques except CONGRA and NMSIMP, the termination criterion is a small normalized predicted function reduction:

$$\frac{\mathrm{g}(\boldsymbol{\psi}^{(k)})'[\mathbf{H}^{(k)}]^{-1}\mathrm{g}(\boldsymbol{\psi}^{(k)})}{|f(\boldsymbol{\psi}^{(k)})|} \le r$$

Here, $\boldsymbol{\psi}$ denotes the vector of parameters that participate in the optimization, $f(\cdot)$ is the objective function, and $\mathrm{g}(\cdot)$ is the gradient. For the CONGRA technique (where a reliable Hessian estimate $\mathbf{H}$ is not available), the following criterion is used:

$$\frac{\| \mathrm{g}(\boldsymbol{\psi}^{(k)}) \|_2^2 \quad \| \mathrm{s}(\boldsymbol{\psi}^{(k)}) \|_2}{\| \mathrm{g}(\boldsymbol{\psi}^{(k)}) - \mathrm{g}(\boldsymbol{\psi}^{(k-1)}) \|_2 \, |f(\boldsymbol{\psi}^{(k)})|} \le r$$

This criterion is not used by the NMSIMP technique. The default value is *r* = 1E–8.

**HESSIAN**

displays the Hessian matrix of the model. This option is not available for Bayesian estimation.

**INVALIDLOGL=***r*

specifies the value assumed by the FMM procedure if a log likelihood cannot be computed (for example, because the value of the response variable falls outside of the response distribution's support). The default value is –1E20.

**ITDETAILS**

adds parameter estimates and gradients to the "Iteration History" table. If the FMM procedure centers or scales the model variables (or both), the parameter estimates and gradients reported during the iteration refer to that scale. You can suppress centering and scaling with the NOCENTER option.

**MAXFUNC=***n*

**MAXFU=***n*

specifies the maximum number of function calls in the optimization process. The default values are as follows, depending on the optimization technique:

- TRUREG, NRRIDG, and NEWRAP: 125
- QUANEW and DBLDOG: 500
- CONGRA: 1000
- NMSIMP: 3000

The optimization can terminate only after completing a full iteration. Therefore, the number of function calls that are actually performed can exceed the number that is specified by the MAXFUNC= option. You can choose the optimization technique with the TECHNIQUE= option.

**MAXITER=***n*

**MAXIT=***n*

specifies the maximum number of iterations in the optimization process. The default values are as follows, depending on the optimization technique:

- TRUREG, NRRIDG, and NEWRAP: 50
- QUANEW and DBLDOG: 200
- CONGRA: 400
- NMSIMP: 1000

These default values also apply when *n* is specified as a missing value. You can choose the optimization technique with the TECHNIQUE= option.

**MAXTIME=***r*

specifies an upper limit of *r* seconds of CPU time for the optimization process. The time is checked only at the end of each iteration. Therefore, the actual run time might be longer than the specified time. By default, CPU time is not limited.

**MINITER=***n*

**MINIT=***n*

specifies the minimum number of iterations. The default value is 0. If you request more iterations than are actually needed for convergence to a stationary point, the optimization algorithms can behave strangely. For example, the effect of rounding errors can prevent the algorithm from continuing for the required number of iterations.

**NAMELEN=***number*

specifies the length to which long effect names are shortened. The default and minimum value is 20.

**NOCENTER**

requests that regressor variables not be centered or scaled. By default the FMM procedure centers and scales columns of the **X** matrix if the models contain intercepts. If NOINT options in MODEL statements are in effect, the columns of **X** are scaled but not centered. Centering and scaling can help with the stability of estimation and sampling algorithms. The FMM procedure does not produce a table of the centered and scaled coefficients and provides no user control over the type of centering and scaling that is applied. The NOCENTER option turns any centering and scaling off and processes the raw values of the continuous variables.

**NOCLPRINT< =***number* **>**

suppresses the display of the "Class Level Information" table if you do not specify *number*. If you specify *number*, the values of the classification variables are displayed for only those variables whose number of levels is less than *number*. Specifying a *number* helps to reduce the size of the "Class Level Information" table if some classification variables have a large number of levels.

**NOITPRINT**

suppresses the display of the "Iteration History Information" table.

**NOPRINT**

suppresses the normal display of tabular and graphical results. The NOPRINT option is useful when you want to create only one or more output data sets with the procedure. This option temporarily disables the Output Delivery System (ODS); see Chapter 20, "Using the Output Delivery System," for more information.

**ORDER=***order-type*

specifies the sort order for the levels of CLASS variables. This ordering determines which parameters in the model correspond to each level in the data.

You can specify the following values for *order-type*:

**DATA**

sorts the levels by order of appearance in the input data set.

**FORMATTED**

sorts the levels by external formatted value, except for numeric variables with no explicit format, which are sorted by their unformatted (internal) value.

**FREQ**

sorts the levels by descending frequency count; levels with the most observations come first in the order.

**INTERNAL**
sorts the levels by unformatted value.

**FREQDATA**
sorts the levels by order of descending frequency count, and within counts by order of appearance in the input data set when counts are tied.

**FREQFORMATTED**
sorts the levels by order of descending frequency count, and within counts by formatted value (as above) when counts are tied.

**FREQINTERNAL**
sorts the levels by order of descending frequency count, and within counts by unformatted value when counts are tied.

When the default ORDER=FORMATTED is in effect for numeric variables for which you have supplied no explicit format, the levels are ordered by their internal values. To order numeric class levels with no explicit format by their BEST12. formatted values, you can specify this format explicitly for the CLASS variables.

When FORMATTED and INTERNAL values are involved, the sort order is machine-dependent.

When the response variable appears in a CLASS statement, the ORDER= option in the PROC FMM statement applies to its sort order. For example, in the following statements the sort order of the wheeze variable is determined by the order of appearance in the input data set because the response variable appears in the CLASS statement:

```
proc fmm order=data;
   class city wheeze;
   model wheeze = city age / dist=binary s;
run;
```

However, in the following statements the sort order of the wheeze variable is determined by the formatted value (the default *response-option* in the MODEL statement):

```
proc fmm order=data;
   class city;
   model wheeze = city age / dist=binary s;
run;
```

The ORDER= option in the PROC FMM statement has no effect on the sort order of the wheeze variable because it does not appear in the CLASS statement.

When you specify a *response-option* in the MODEL statement, it overrides the ORDER= option in the PROC FMM statement.

For more information about sort order, see the chapter on the SORT procedure in the *Base SAS Procedures Guide* and the discussion of BY-group processing in *SAS Language Reference: Concepts*.

**PARMSTYLE=EFFECT | LABEL**

specifies the display style for parameters and effects. The FMM procedure can display parameters in two styles:

**EFFECT** displays parameters and effects in the EFFECT style (which is used by the MIXED and GLIMMIX procedure, for example), which identifies a parameter with an "Effect" column and adds separate columns for the CLASS variables in the model.

**LABEL** displays parameters and effects in the LABEL style, which creates one column, named Parameter, that combines the relevant information about a parameter into a single column. If your model contains multiple CLASS variables, this style might use space more economically.

The EFFECT style is the default for models that contain effects; otherwise the LABEL style is used (for example, in homogeneous mixtures). You can change the display style by using the PARMSTYLE= option. Regardless of the display style, ODS output data sets that contain information about parameter estimates contain columns for both styles.

**PARTIAL=***variable*

**MEMBERSHIP=***variable*

specifies a variable in the input data set that identifies component membership. You can specify missing values for observations whose component membership is undetermined; this is known as a partial classification (McLachlan and Peel 2000, p. 75). For observations with known membership, the likelihood contribution is no longer a mixture. If observation $i$ is known to be a member of component $m$, then its log likelihood contribution is

$$\log\left\{\pi_m(\mathbf{z},\boldsymbol{\alpha}_m)\ \ p_m(y;\mathbf{x}_m'\boldsymbol{\beta}_m,\phi_m)\right\}$$

Otherwise, if membership is undetermined, it is

$$\log\left\{\sum_{j=1}^{k}\ \pi_j(\mathbf{z},\boldsymbol{\alpha}_j)p_j(y;\mathbf{x}_j'\boldsymbol{\beta}_j,\phi_j)\right\}$$

The *variable* that you specify in this option can be numeric or character. In case of a character variable, the *variable* must also appear in the CLASS statement. If it also appears in the CLASS statement, the membership assignment is made on the basis of the levelized values of the variable, as shown in the "Class Level Information" table. If the *variable* is numeric, any nonmissing values of the *variable* must take values in the range $1, 2, \ldots, k$, where $k$ is the number of mixture components. Invalid *variable* values are ignored.

In a model in which label switching is a problem, you can sometimes avoid switching by assigning just a few observations to categories. For example, in a three-component model, switching might be prevented if you assign the observation that has the smallest response value to the first component and the observation that has the largest response value to the last component.

**PLOTS** < (*global-plot-options*) > < =*plot-request* < (*options*) > >

**PLOTS** < (*global-plot-options*) > < =(*plot-request* < (*options*) > < ... *plot-request* < (*options*) > >) >

controls the graphical output that is produced through ODS Graphics.

ODS Graphics must be enabled before plots can be requested. For example:

```
ods graphics on;
proc fmm data=yeast seed=12345;
   model count/n = / k=2;
   freq f;
   performance cpucount=2;
   bayes;
run;
ods graphics off;
```

For more information about enabling and disabling ODS Graphics, see the section "Enabling and Disabling ODS Graphics" on page 623 in Chapter 21, "Statistical Graphics Using ODS."

### Global Plot Options

The *global-plot-options* apply to all relevant plots generated by the FMM procedure. The *global-plot-options* supported by the FMM procedure are as follows:

**UNPACKPANEL**

**UNPACK**

displays each plot separately. (By default, some plots can appear together in a single panel.)

**ONLY**

produces only the specified plots. This option is useful if you do not want the procedure to generate all default plots, but only the ones specified.

### Specific Plot Options

The following listing describes the specific plots and their *options*.

**ALL**

requests that all plots appropriate for the analysis be produced.

**NONE**

requests that no ODS graphics be produced.

**DENSITY** < (*density-options*) >

requests a plot of the data histogram and mixture density function. This is a default plot in models that have no effects in the MODEL statements and is available only in these models. Furthermore, all distributions that are involved in the mixture must be continuous. You can specify the following *density-options* to modify the plot:

**CUMULATIVE**

**CDF**

displays the histogram and densities in cumulative form.

**NBINS=***n*

**BINS=***n*

specifies the number of bins in the histogram; *n* is greater than or equal to 0. By default, the FMM procedure computes a suitable bin width and number of bins, based on the range of the response and the number of usable observations. The option has no effect for binary data.

**NOCOMPONENTS**

**NOCOMP**

suppresses the component densities from the plot. If the component densities are displayed, they are scaled so that their sum equals the mixture density at any point on the graph. In single-component models, this option has no effect.

**NODENSITY**

**NODENS**

suppresses the computation of the mixture density and the component densities. If you specify the NOHISTOGRAM and the NODENSITY option, no plot is produced.

**NOLABEL**

suppresses the component identification with labels. By default, the FMM procedure labels component densities in the legend of the plot. If you do not specify a model label with the LABEL= option in the MODEL statement, an identifying label is constructed from the parameter estimates that are associated with the component. In this case the parameter values are not necessarily the mean and variance of the distribution; the values used to identify the densities on the plot are chosen to simplify linking between graphical and tabular results.

**NOHISTOGRAM**

**NOHIST**

suppresses the computation of the histogram of the raw values. If you specify the NOHISTOGRAM and the NODENSITY option, no plot is produced.

**NPOINTS=***n*

**N=***n*

specifies the number of values used to compute the density functions; *n* is greater than or equal to 0. The default is N=200.

**WIDTH=***value*

**BINWIDTH=***value*

specifies the bin width for the histogram. The *value* is specified in units of the response variable and must be positive. The option has no effect for binary data.

**TRACE < (***tadpanel-options***) >**

requests a trace panel with posterior diagnostics for a Bayesian analysis. If a BAYES statement is present, the trace panel plots are generated by default, one for each sampled parameter. You can specify the following *tadpanel-options* to modify the plot:

**BOX**
**BOXPLOT**
replaces the autocorrelation plot with a box plot of the posterior sample.

**SMOOTH=NONE | MEAN | SPLINE**
adds a reference estimate to the trace plot. By default, SMOOTH=NONE. SMOOTH=MEAN uses the arithmetic mean of the trace as the reference. SMOOTH=SPLINE adds a penalized B-spline.

**REFERENCE=***reference-style*
adds reference lines to the density plot, trace plot, and box plot. You can specify the following *reference-styles*:

| | |
|---|---|
| **NONE** | suppresses the reference lines. |
| **EQT** | requests equal-tail intervals. |
| **HPD** | requests intervals of highest posterior density. The level for the credible or HPD intervals is chosen based on the "Posterior Interval Statistics" table. |
| **PERCENTILES** | (or PERC) for percentiles. The largest and smallest percentiles in the "Posterior Summary Statistics" table are used. |

The default is REFERENCE=EQT.

**UNPACK**
unpacks the panel graphic and displays its elements as separate plots.

**CRITERIONPANEL < (***critpanel-options***) >**
requests a plot for comparing the model fit criteria for different numbers of components. This plot is available only if you also specify the KMAX option in at least one MODEL statement. The plot includes different criteria, depending on whether you are using maximum likelihood or Bayesian estimation. You can specify the following *critpanel-option* to modify the plot:

**UNPACK**
unpacks the panel plot and displays its elements as separate plots, one for each fit criterion.

**SEED=***n*
determines the random number seed for analyses that depend on a random number stream. If you do not specify a seed or if you specify a value less than or equal to zero, the seed is generated from reading the time of day from the computer clock. The largest possible value for the seed is $2^{31} - 1$. The seed value is reported in the "Model Information" table.

You can use the SYSRANDOM and SYSRANEND macro variables after a PROC FMM run to query the initial and final seed values. However, using the final seed value as the starting seed for a subsequent analysis does not continue the random number stream where the previous analysis left off. The SYSRANEND macro variable provides a mechanism to pass on seed values to ensure that the sequence of random numbers is the same every time you run an entire program.

Analyses that use the same (nonzero) seed are not completely reproducible if they are executed with a different number of threads because the random number streams in separate threads are independent. You can control the number of threads used by the FMM procedure with system options or through the PERFORMANCE statement in the FMM procedure.

**SINGCHOL=***number*
tunes the singularity criterion in Cholesky decompositions. The default is 1E4 times the machine epsilon; this product is approximately 1E–12 on most computers.

**SINGRES=***number*
sets the tolerance for which the residual variance or scale parameter is considered to be zero. The default is 1E4 times the machine epsilon; this product is approximately 1E–12 on most computers.

**SINGULAR=***number*
tunes the general singularity criterion applied by the FMM procedure in sweeps and inversions. The default is 1E4 times the machine epsilon; this product is approximately 1E–12 on most computers.

**TECHNIQUE=***keyword*
**TECH=***keyword*
specifies the optimization technique to obtain maximum likelihood estimates. You can choose from the following techniques by specifying the appropriate *keyword*:

| | |
|---|---|
| **CONGRA** | performs a conjugate-gradient optimization. |
| **DBLDOG** | performs a version of double-dogleg optimization. |
| **NEWRAP** | performs a Newton-Raphson optimization combining a line-search algorithm with ridging. |
| **NMSIMP** | performs a Nelder-Mead simplex optimization. |
| **NONE** | performs no optimization. |
| **NRRIDG** | performs a Newton-Raphson optimization with ridging. |
| **QUANEW** | performs a dual quasi-Newton optimization. |
| **TRUREG** | performs a trust-region optimization. |

The default is TECH=QUANEW.

For more details about these optimization methods, see the section "Choosing an Optimization Algorithm" on page 512 in Chapter 19, "Shared Concepts and Topics."

**ZEROPROB=***number*
tunes the threshold (a value between 0 and 1) below which the FMM procedure considers a component mixing probability to be zero. This affects the calculation of the number of effective components. The default is the square root of the machine epsilon; this is approximately 1E–8 on most computers.

## BAYES Statement

**BAYES** *bayes-options* **;**

The BAYES statement requests that the parameters of the model be estimated by Markov chain Monte Carlo sampling techniques. The FMM procedure can use maximum likelihood to estimate the parameters of all models that are supported by the procedure. Bayesian estimation is available for only a subset of these models.

In Bayesian analysis, it is essential to examine the convergence of the Markov chains before you proceed with posterior inference. With ODS Graphics turned on, the FMM procedure produces graphics at the end of the procedure output; you can visually examine the convergence of the chain in these graphics. You cannot base inferences on a Markov chain that has not converged.

The output produced for a Bayesian analysis is markedly different from that for a frequentist (maximum likelihood) analysis for the following reasons:

- Parameter estimates do not have the same interpretation in the two analyses. Parameters are fixed unknown constants in the frequentist context and random variables in a Bayesian analysis.
- The results of a Bayesian analysis are summarized through chain diagnostics and posterior summary statistics and intervals.
- The FMM procedure samples the mixing probabilities in Bayesian models directly, rather than mapping them onto a logistic (or other) scale.

The FMM procedure applies highly specialized sampling algorithms in Bayesian models. For single-component models without effects, a conjugate sampling algorithm is used where possible. For models in the exponential family that contain effects, the sampling algorithm is based on Gamerman (1997). For the normal and $t$ distributions, a conjugate sampler is the default sampling algorithm for models with and without effects. In multi-component models, the sampling algorithm is based on latent variable sampling through data augmentation (Frühwirth-Schnatter 2006) and the Gamerman or conjugate sampler. Because of this specialization, the options for controlling the prior distributions of the parameters are limited.

Table 43.3 summarizes the *bayes-options* available in the BAYES statement. The full assortment of options is then described in alphabetical order.

**Table 43.3** BAYES Statement Options

| Option | Description |
|---|---|
| **Options Related to Sampling** | |
| INITIAL= | Specifies how to construct initial values |
| NBI= | Specifies the number of burn-in samples |
| NMC= | Specifies the number of samples after burn-in |
| METROPOLIS | Forces a Metropolis-Hastings sampling algorithm even if conjugate sampling is possible |
| OUTPOST= | Generates a data set that contains the posterior estimates |
| THIN= | Controls the thinning of the Markov chain |
| **Specification of Prior Information** | |
| MIXPRIORPARMS | Specifies the prior parameters for the Dirichlet distribution of the mixing probabilities |
| BETAPRIORPARMS= | Specifies the parameters of the normal prior distribution for individual parameters in the $\boldsymbol{\beta}$ vector |
| MUPRIORPARMS= | Specifies the parameters of the prior distribution for the means in homogeneous mixtures without effects |

**Table 43.3** *continued*

| Option | Description |
|---|---|
| PHIPRIORPARMS= | Specifies the parameters of the inverse gamma prior distribution for the scale parameters in homogeneous mixtures |
| PRIOROPTIONS | Specifies additional options used in the determination of the prior distribution |
| **Posterior Summary Statistics and Convergence Diagnostics** | |
| DIAGNOSTICS= | Displays convergence diagnostics for the Markov chain |
| STATISTICS | Displays posterior summary information for the Markov chain |
| **Other Options** | |
| ESTIMATE= | Specifies which estimate is used for the computation of OUTPUT statistics and graphics |
| TIMEINC= | Specifies the time interval to report on sampling progress (in seconds) |

You can specify the following *bayes-options* in the BAYES statement.

**BETAPRIORPARMS=***pair-specification*

**BETAPRIORPARMS(***pair-specification* **...** *pair-specification***)**

specifies the parameters for the normal prior distribution of the parameters that are associated with model effects ($\beta$s). The *pair-specification* is of the form $(a, b)$, and the values $a$ and $b$ are the mean and variance of the normal distribution, respectively. This option overrides the PRIOROPTIONS option.

The form of the BETAPRIORPARMS with an equal sign and a single pair is used to specify one pair of prior parameters that applies to all components in the mixture. In the following example, the two intercepts and the two regression coefficients all have a $N(0, 100)$ prior distribution:

```
proc fmm;
   model y = x / k=2;
   bayes betapriorparms=(0,100);
run;
```

You can also provide a list of pairs to specify different sets of prior parameters for the various regression parameters and components. For example:

```
proc fmm;
   model y = x/ k=2;
   bayes betapriorparms( (0,10) (0,20) (.,.) (3,100) );
run;
```

The simple linear regression in the first component has a $N(0, 10)$ prior for the intercept and a $N(0, 20)$ prior for the slope. The prior for the intercept in the second component uses the FMM default, whereas the prior for the slope is $N(3, 100)$.

**DIAGNOSTICS=ALL | NONE | (*keyword-list*)**

**DIAG=ALL | NONE | (*keyword-list*)**

controls the computation of diagnostics for the posterior chain. You can request all posterior diagnostics by specifying DIAGNOSTICS=ALL or suppress the computation of posterior diagnostics by specifying DIAGNOSTICS=NONE. The following *keywords* enable you to select subsets of posterior diagnostics; the default is DIAGNOSTICS=(AUTOCORR).

**AUTOCORR < (LAGS= *numeric-list*) >**

computes for each sampled parameter the autocorrelations of lags specified in the LAGS= list. Elements in the list are truncated to integers, and repeated values are removed. If the LAGS= option is not specified, autocorrelations are computed by default for lags 1, 5, 10, and 50. See the section "Autocorrelations" on page 157 in Chapter 7, "Introduction to Bayesian Analysis Procedures," for details.

**ESS**

computes an estimate of the effective sample size (Kass et al. 1998), the correlation time, and the efficiency of the chain for each parameter. See the section "Effective Sample Size" on page 157 in Chapter 7, "Introduction to Bayesian Analysis Procedures," for details.

**GEWEKE < (*geweke-options*) >**

computes the Geweke spectral density diagnostics (Geweke 1992), which are essentially a two-sample $t$ test between the first $f_1$ portion and the last $f_2$ portion of the chain. The default is $f_1 = 0.1$ and $f_2 = 0.5$, but you can choose other fractions by using the following *geweke-options*:

**FRAC1=*value***

specifies the fraction $f_1$ for the first window.

**FRAC2=*value***

specifies the fraction $f_2$ for the second window.

See the section "Geweke Diagnostics" on page 151 in Chapter 7, "Introduction to Bayesian Analysis Procedures," for details.

**HEIDELBERGER < (*Heidel-options*) >**

**HEIDEL < (*Heidel-options*) >**

computes the Heidelberger and Welch diagnostic (which consists of a stationarity test and a half-width test) for each variable. The stationary diagnostic test tests the null hypothesis that the posterior samples are generated from a stationary process. If the stationarity test is passed, a half-width test is then carried out. See the section "Heidelberger and Welch Diagnostics" on page 153 in Chapter 7, "Introduction to Bayesian Analysis Procedures," for more details.

These diagnostics are not performed by default. You can specify the DIAGNOSTICS=HEIDELBERGER option to request these diagnostics, and you can also specify suboptions, such as DIAGNOSTICS=HEIDELBERGER(EPS=0.05), as follows:

**SALPHA=*value***

specifies the $\alpha$ level ($0 < \alpha < 1$) for the stationarity test. By default, SALPHA=0.05.

**HALPHA=***value*
specifies the $\alpha$ level ($0 < \alpha < 1$) for the half-width test. By default, HALPHA=0.05.

**EPS=***value*
specifies a small positive number $\epsilon$ such that if the half-width is less than $\epsilon$ times the sample mean of the retaining iterates, the half-width test is passed. By default, EPS=0.1.

**MCERROR**
**MCSE**
computes an estimate of the Monte Carlo standard error for each sampled parameter. See the section "Standard Error of the Mean Estimate" on page 158 in Chapter 7, "Introduction to Bayesian Analysis Procedures," for details.

**MAXLAG=***n*
specifies the largest lag used in computing the effective sample size and the Monte Carlo standard error. Specifying this option implies the ESS and MCERROR options. The default is MAXLAG=250.

**RAFTERY <(***Raftery-options***)>**
**RL <(***Raftery-options***)>**
computes the Raftery and Lewis diagnostics, which evaluate the accuracy of the estimated quantile ($\hat{\theta}_Q$ for a given $\mathrm{Q} \in (0, 1)$) of a chain. $\hat{\theta}_Q$ can achieve any degree of accuracy when the chain is allowed to run for a long time. The algorithm stops when the estimated probability $\hat{P}_Q = \Pr(\theta \leq \hat{\theta}_Q)$ reaches within $\pm$R of the value Q with probability S; that is, $\Pr(\mathrm{Q} - \mathrm{R} \leq \hat{\mathrm{P}}_Q \leq \mathrm{Q} + \mathrm{R}) = \mathrm{S}$. See the section "Raftery and Lewis Diagnostics" on page 154 in Chapter 7, "Introduction to Bayesian Analysis Procedures," for more details. The *Raftery-options* enable you to specify Q, R, S, and a precision level $\epsilon$ for a stationary test.

These diagnostics are not performed by default. You can specify the DIAGNOSTICS=RAFERTY option to request these diagnostics, and you can also specify suboptions, such as DIAGNOSTICS=RAFERTY(QUANTILE=0.05), as follows:

**QUANTILE=***value*
**Q=***value*
specifies the order (a value between 0 and 1) of the quantile of interest. By default, QUANTILE=0.025.

**ACCURACY=***value*
**R=***value*
specifies a small positive number as the margin of error for measuring the accuracy of estimation of the quantile. By default, ACCURACY=0.005.

**PROB=***value*
**S=***value*
specifies the probability of attaining the accuracy of the estimation of the quantile. By default, PROB=0.95.

**EPS=***value*
specifies the tolerance level (a small positive number between 0 and 1) for the stationary test. By default, EPS=0.001.

**MIXPRIORPARMS=K**

**MIXPRIORPARMS(***value-list***)**
specifies the parameters used in constructing the Dirichlet prior distribution for the mixing parameters. If you specify MIXPRIORPARMS=K, the parameters of the $k$-dimensional Dirichlet distribution are a vector that contains the number of components in the model ($k$), whatever that might be. You can specify an explicit list of parameters in *value-list*. If the MIXPRIORPARMS option is not specified, the default Dirichlet parameter vector is a vector of length $k$ of ones. This results in a uniform prior over the unit simplex; for $k$=2, this is the uniform distribution. See the section "Prior Distributions" on page 2958 for the distribution function of the Dirichlet as used by the FMM procedure.

**ESTIMATE=MEAN | MAP**
determines which overall estimate is used, based on the posterior sample, in the computation of OUTPUT statistics and certain ODS graphics. By default, the arithmetic average of the (thinned) posterior sample is used. If you specify ESTIMATE=MAP, the parameter vector is used that corresponds to the maximum log posterior density in the posterior sample. In any event, a message is written to the SAS log if postprocessing results depend on a summary estimate of the posterior sample.

**INITIAL=DATA | MLE | MODE | RANDOM**
determines how initial values for the Markov chain are obtained. The default when a conjugate sampler is used is INITIAL=DATA, in which case the FMM procedure uses the same algorithm to obtain data-dependent starting values as it uses for maximum likelihood estimation. If no conjugate sampler is available or if you use the METROPOLIS option to explicitly request that it not be used, then the default is INITIAL=MLE, in which case the maximum likelihood estimates are used as the initial values. If the maximum likelihood optimization fails, the FMM procedure switches to the default INITIAL=DATA.

The options INITIAL=MODE and INITIAL=RANDOM use the mode and random draws from the prior distribution, respectively, to obtain initial values. If the mode does not exist or if it falls on the boundary of the parameter space, the prior mean is used instead.

**METROPOLIS**
requests that the FMM procedure use the Metropolis-Hastings sampling algorithm based on Gamerman (1997), even in situations where a conjugate sampler is available.

**MUPRIORPARMS=***pair-specification*

**MUPRIORPARMS(***pair-specification* **. . .** *pair-specification***)**
specifies the parameters for the means in homogeneous mixtures without regression coefficients. The *pair-specification* is of the form $(a, b)$, where $a$ and $b$ are the two parameters of the prior distribution, optionally delimited with a comma. The actual distribution of the parameter is implied by the distribution selected in the MODEL statement. For example, it is a normal distribution for a mixture of normals, a gamma distribution for a mixture of Poisson variables, a beta distribution for a mixture of binary variables, and an inverse gamma distribution for a mixture of exponential variables. This option overrides the PRIOROPTIONS option.

The parameters correspond as follows:

Beta: The parameters correspond to the $\alpha$ and $\beta$ parameters of the beta prior distribution such that its mean is $\mu = \alpha/(\alpha + \beta)$ and its variance is $\mu(1 - \mu)/(\alpha + \beta + 1)$.

Normal: The parameters correspond to the mean and variance of the normal prior distribution.

Gamma: The parameters correspond to the $\alpha$ and $\beta$ parameters of the gamma prior distribution such that its mean is $\alpha/\beta$ and its variance is $\alpha/\beta^2$.

Inverse gamma: The parameters correspond to the $\alpha$ and $\beta$ parameters of the inverse gamma prior distribution such that its mean is $\mu = \beta/(\alpha - 1)$ and its variance is $\mu^2/(\alpha - 2)$.

The two techniques for specifying the prior parameters with the MUPRIORPARMS option are as follows:

- Specify an equal sign and a single pair of values:

```
proc fmm seed=12345;
   model y = / k=2;
   bayes mupriorparms=(0,50);
run;
```

- Specify a list of parameter pairs within parentheses:

```
proc fmm seed=12345;
   model y = / k=2;
   bayes mupriorparms( (.,.) (1.4,10.5));
run;
```

If you specify an invalid value (outside of the parameter space for the prior distribution), the FMM procedure chooses the default value and writes a message to the SAS log. If you want to use the default values for a particular parameter, you can also specify missing values in the *pair-specification*. For example, the preceding list specification assigns default values for the first component and uses the values 1.4 and 10.5 for the mean and variance of the normal prior distribution in the second component. The first example assigns a $N(0, 50)$ prior distribution to the means in both components.

**NBI=***n*

specifies the number of burn-in samples. During the burn-in phase, chains are not saved. The default is NBI=2000.

**NMC=***n*

**SAMPLE=***n*

specifies the number of Monte Carlo samples after the burn-in. Samples after the burn-in phase are saved unless they are thinned with the THIN= option. The default is NMC=10000.

**OUTPOST<(***outpost-options***)>=***data-set*

requests that the posterior sample be saved to a SAS data set. In addition to variables that contain log likelihood and log posterior values, the OUTPOST data set contains variables for the parameters. The variable names for the parameters are generic (Parm_1, Parm_2, . . ., Parm_p). The labels of the parameters are descriptive and correspond to the "Parameter Mapping" table that is produced when the OUTPOST= option is in effect.

You can specify the following *outpost-options* in parentheses:

**LOGPRIOR**
adds the value of the log prior distribution to the data set.

**NONSINGULAR | NONSING | COMPRESS**
eliminates parameters that correspond to singular columns in the design matrix (and were not sampled) from the posterior data set. This is the default.

**SINGULAR | SING**
adds columns of zeros to the data set in positions that correspond to singularities in the model or to parameters that were not sampled for other reasons. By default, these columns of zeros are not written to the posterior data set.

**PHIPRIORPARMS=***pair-specification*

**PHIPRIORPARMS(** *pair-specification* **...** *pair-specification***)**
specifies the parameters for the inverse gamma prior distribution of the scale parameters ($\phi$'s) in the model. The *pair-specification* is of the form $(a, b)$, and the values are chosen such that the prior distribution has mean $\mu = b/(a-1)$ and variance $\mu^2/(a-2)$.

The form of the PHIPRIORPARMS with an equal sign and a single pair is used to specify one pair of prior parameters that applies to all components in the mixture. For example:

```
proc fmm seed=12345;
   model y = / k=2;
   bayes phipriorparms=(2.001,1.001);
run;
```

The form with a list of pairs is used to specify different prior parameters for the scale parameters in different components. For example:

```
proc fmm seed=12345;
   model y = / k=2;
   bayes phipriorparms( (.,1.001)  (3.001,2.001) );
run;
```

If you specify an invalid value (outside of the parameter space for the prior distribution), the FMM procedure chooses the default value and writes a message to the SAS log. If you want to use the default values for a particular parameter, you can also specify missing values in the *pair-specification*. For example, the preceding list specification assigns default values for the first component $a$ prior parameter and uses the value 1.001 for the $b$ prior parameter. The second pair assigns 3.001 and 2.001 for the $a$ and $b$ prior parameters, respectively.

**PRIOROPTIONS < = >(***prior-options***)**

**PRIOROPTS < = >(***prior-options***)**
specifies options related to the construction of the prior distribution and the choice of their parameters. Some *prior-options* apply only in particular models. The BETAPRIORPARMS= and MUPRIORPARMS= options override this option.

You can specify the following *prior-options*:

**CONDITIONAL | COND**

chooses a conditional prior specification for the homogeneous normal and *t* distribution response components. The default prior specification in these models is an independence prior where the mean of the *h*th component has prior $\mu_h \sim N(a, b)$. The conditional prior is characterized by $\mu_h \sim N(a, \sigma_h^2/b)$.

**DEPENDENT | DEP**

chooses a data-dependent prior for the homogeneous models without effects. The prior parameters *a* and *b* are chosen as follows, based on the distribution in the MODEL statement:

Binary and binomial: $a = \bar{y}/(1-\bar{y})$, *b*=1, and the prior distribution for the success probability is $\mathrm{beta}(a, b)$.

Poisson: $a = 1$, $b = 1/\bar{y}$, and the prior distribution for $\mu$ is $\mathrm{gamma}(a, b)$. See Frühwirth-Schnatter (2006, p. 280) and Viallefont, Richardson, and Greene (2002).

Exponential: $a = 3$, $b = 2\bar{y}$, and the prior distribution for $\mu$ is inverse gamma with parameters *a* and *b*.

Normal and *t*: Under the default independence prior, the prior distribution for $\mu$ is $N(\bar{y}, fs^2)$ where *f* is the variance factor from the VAR= option and

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

Under the default conditional prior specification, the prior for $\mu_h$ is $N(a, \sigma_h^2/b)$ where $a = \bar{y}$ and $b = 2.6/(\max\{y\} - \min\{y\})$. The prior for the scale parameter is inverse gamma with parameters 1.28 and $0.36s^2$. For further details, see Raftery (1996) and Frühwirth-Schnatter (2006, p. 179).

**VAR=*f***

specifies the variance for normal prior distributions. The default is VAR=1000. This factor is used, for example, in determining the prior variance of regression coefficients or in determining the prior variance of means in homogeneous mixtures of *t* or normal distributions (unless a data-dependent prior is used).

**MLE< =*r* >**

specifies that the prior distribution for regression variables be based on a multivariate normal distribution centered at the MLEs and whose dispersion is a multiple *r* of the asymptotic MLE covariance matrix. The default is MLE=10. In other words, if you specify PRIOROPTIONS(MLE), the FMM procedure chooses the prior distribution for the regression variables as $N(\widehat{\beta}, 10\mathrm{Var}[\widehat{\beta}])$ where $\widehat{\beta}$ is the vector of maximum likelihood estimates. The prior for the scale parameter is inverse gamma with parameters 1.28 and $0.36s^2$ where

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

For further details, see Raftery (1996) and Frühwirth-Schnatter (2006, p. 179). If you specify PRIOROPTIONS(MLE) for the regression parameters, then the data-dependent prior is used for the scale parameter; see the PRIOROPTIONS(DEPENDENT) option above.

The MLE option is not available for mixture models in which the parameters are estimated directly on the data scale, such as homogeneous mixture models or mixtures of distributions without model effects for which a conjugate sampler is available. By using the METROPOLIS option, you can always force the FMM procedure to abandon a conjugate sampler in favor of a Metropolis-Hastings sampling algorithm to which the MLE option applies.

**STATISTICS** < (*global-options*) > = **ALL** | **NONE** | *keyword* | (*keyword-list*)

**SUMMARIES** < (*global-options*) > = **ALL** | **NONE** | *keyword* | (*keyword-list*)

controls the number of posterior statistics produced. Specifying STATISTICS=ALL is equivalent to specifying STATISTICS=(SUMMARY INTERVAL). To suppress the computation of posterior statistics, specify STATISTICS=NONE. The default is STATISTICS=(SUMMARY INTERVAL). See the section "Summary Statistics" on page 158 in Chapter 7, "Introduction to Bayesian Analysis Procedures," for more details.

The *global-options* include the following:

**ALPHA=***numeric-list*

controls the coverage levels of the equal-tail credible intervals and the credible intervals of highest posterior density (HPD) credible intervals. The ALPHA= values must be between 0 and 1. Each ALPHA= value produces a pair of $100(1-\alpha)\%$ equal-tail and HPD credible intervals for each sampled parameter. The default is ALPHA=0.05, which results in 95% credible intervals for the parameters.

**PERCENT=***numeric-list*

requests the percentile points of the posterior samples. The values in *numeric-list* must be greater than or equal to 0 and less than or equal to 100. The default is PERCENT=(25 50 75), which yields for each parameter the 25th, 50th, and 75th percentiles, respectively.

The list of *keywords* includes the following:

**SUMMARY**

produces the means, standard deviations, and percentile points for the posterior samples. The default is to produce the 25th, 50th, and 75th percentiles; you can modify this list with the global PERCENT= option.

**INTERVAL**

produces equal-tail and HPD credible intervals. The default is to produce the 95% equal-tail credible intervals and 95% HPD credible intervals, but you can use the ALPHA= *global-option* to request credible intervals for any probabilities.

**THIN=***n*

**THINNING=***n*

controls the thinning of the Markov chain after the burn-in. Only one in every $k$ samples is used when THIN=$k$, and if NBI=$n_0$ and NMC=$n$, the number of samples kept is

$$\left[\frac{n_0 + n}{k}\right] - \left[\frac{n_0}{k}\right]$$

where $[a]$ represents the integer part of the number $a$. The default is THIN=1—that is, all samples are kept after the burn-in phase.

**TIMEINC=***n*

specifies a time interval in seconds to report progress during the burn-in and sampling phase. The time interval is approximate, because the minimum time interval in which the FMM procedure can respond depends on the multithreading configuration.

## BY Statement

**BY** *variables* **;**

You can specify a BY statement in PROC FMM to obtain separate analyses of observations in groups that are defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables. If you specify more than one BY statement, only the last one specified is used.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data by using the SORT procedure with a similar BY statement.
- Specify the NOTSORTED or DESCENDING option in the BY statement in the FMM procedure. The NOTSORTED option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Create an index on the BY variables by using the DATASETS procedure (in Base SAS software).

Because sorting the data changes the order in which PROC FMM reads observations, the sort order for the levels of the CLASS variable might be affected if you have specified ORDER=DATA in the PROC FMM statement.

For more information about BY-group processing, see the discussion in *SAS Language Reference: Concepts*. For more information about the DATASETS procedure, see the discussion in the *Base SAS Procedures Guide*.

## CLASS Statement

**CLASS** *variables* < / **TRUNCATE** > **;**

The CLASS statement names the classification variables to be used in the model. Typical classification variables are Treatment, Sex, Race, Group, and Replication. If you use the CLASS statement, it must appear before the MODEL statement.

Classification variables can be either character or numeric. By default, class levels are determined from the entire set of formatted values of the CLASS variables.

**NOTE:** Prior to SAS 9, class levels were determined by using no more than the first 16 characters of the formatted values. To revert to this previous behavior, you can use the TRUNCATE option in the CLASS statement.

In any case, you can use formats to group values into levels. See the discussion of the FORMAT procedure in the *Base SAS Procedures Guide* and the discussions of the FORMAT statement and SAS formats in *SAS Formats and Informats: Reference*. You can adjust the order of CLASS variable levels with the ORDER= option in the PROC FMM statement.

You can specify the following *option* in the CLASS statement after a slash (/):

**TRUNCATE**
specifies that class levels should be determined by using only up to the first 16 characters of the formatted values of CLASS variables. When formatted values are longer than 16 characters, you can use this option to revert to the levels as determined in releases prior to SAS 9.

## FREQ Statement

**FREQ** *variable* **;**

The *variable* in the FREQ statement identifies a numeric variable in the data set that contains the frequency of occurrence for each observation. PROC FMM treats each observation as if it appears $f$ times, where $f$ is the value of the FREQ variable for the observation. If it is not an integer, the frequency value is truncated to an integer. If the frequency value is less than 1 or missing, the observation is not used in the analysis. When the FREQ statement is not specified, each observation is assigned a frequency of 1.

## ID Statement

**ID** *variables* **;**

The ID statement specifies a list of variables that are included in the OUT= data set of the OUTPUT statement. If no ID statement is specified, all variables from the input data set are copied into the output data set.

## MODEL Statement

**MODEL** *response* < **(***response-options***)** > **=** < *effects* > < / *model-options* > **;**

**MODEL** *events/trials* **=** < *effects* > < / *model-options* > **;**

**MODEL** **+** < *effects* > < / *model-options* > **;**

The MODEL statement defines elements of the mixture model, such as the model effects, the distribution, and the link function. At least one MODEL statement is required. You can specify more than one MODEL statement. Each MODEL statement identifies one or more components of a mixture. For example, if components differ in their distributions, link functions, or regressor variables, then you can use separate MODEL statements to define the components. If the finite mixture model is homogeneous—meaning that all components share the same regressors, distribution, and link function—then you can specify the mixture model by using a single MODEL statement and specifying the K= option.

The FMM procedure includes an intercept in each model by default. You can remove it by using the NOINT option.

You can specify the dependent variable by using either the *response* syntax or the *events/trials* syntax. The *events/trials* syntax is specific to models for binomial-type data. A binomial($n$, $\pi$) variable is the sum of $n$ independent Bernoulli trials with event probability $\pi$. Each Bernoulli trial results in either an event or a nonevent (with probability $1-\pi$). The value of the second variable, *trials*, gives the number $n$ of Bernoulli trials. The value of the first variable, *events*, is the number of events out of $n$. The values of both *events* and (*trials* – *events*) must be nonnegative, and the value of *trials* must be positive. Other distributions that allow the *events/trials* syntax are the beta-binomial distribution and the binomial cluster model.

If you use the *events/trials* syntax, the FMM procedure defaults to the binomial distribution. If you use the *response* syntax, the procedure defaults to the normal distribution unless the response variable is a character variable or is listed in the CLASS statement.

You use a similar syntax to fit multinomial models, Dirichlet-multinomial models, and multinomial cluster models, except that for these distributions you specify multiple dependent variables, one for each value of the multinomial response. For these models, you can specify either multiple *response* variables or multiple *event* variables. If you use multiple *response* or multiple *event* variables, the FMM procedure defaults to the multinomial distribution. If you use the multiple *response* syntax, PROC FMM treats the total of these responses as fixed.

The FMM procedure supports a continuation-style syntax in MODEL statements. Because a mixture has only one set of response variables, it is sufficient to specify the response variable in one MODEL statement. Other MODEL statements can use the continuation symbol "+" before the specification of effects. For example, the following statements fit a three-component binomial mixture model:

```
class A;
model y/n = x / k=2;
model     + A;
```

The first MODEL statement uses the "=" sign to separate response information from effect information and specifies the response variable by using the *events/trials* syntax. This determines that the distribution is binomial. This MODEL statement adds to the mixture model two components that have different intercepts and regression slopes. The second MODEL statement adds to the mixture model another component in which the mean is a function of the classification main effect for variable A. The response is also binomial; it is a continuation from the previous MODEL statement.

There are two sets of options in the MODEL statement. The *response-options* determine how the FMM procedure models probabilities for binary data. The *model-options* control other aspects of model formation and inference. Table 43.4 summarizes the *response-options* and *model-options* available in the MODEL statement. These are subsequently discussed in detail in alphabetical order by option category.

**Table 43.4** Summary of MODEL Statement Options

| Option | Description |
|---|---|
| **Response Variable Options** | |
| DESCENDING | Reverses the order of response categories |
| EVENT= | Specifies the event category in binary models |
| ORDER= | Specifies the sort order for the response variable |
| REFERENCE= | Specifies the reference category in binary models |
| **Model Building** | |
| DIST= | Specifies the response distribution |
| LINK= | Specifies the link function |
| K= | Specifies the number of mixture components |
| KMAX= | Specifies the maximum number of mixture components |
| KMIN= | Specifies the minimum number of mixture components |
| KRESTART | Requests that the starting values for each analysis be determined separately instead of sequentially |
| NOINT | Excludes fixed-effect intercept from model |
| OFFSET= | Specifies the offset variable for linear predictor |
| **Statistical Computations and Output** | |
| ALPHA=$\alpha$ | Determines the confidence level $(1-\alpha)$ |
| CL | Displays confidence limits for fixed-effects parameter estimates |
| EQUATE= | Imposes simple equality constraints on parameters in this model |
| LABEL= | Identifies the model |
| PARMS | Provides starting values for the parameters in this model |

## Response Variable Options

Response variable options determine how the FMM procedure models probabilities for binary data.

You can specify the following *response-options* by enclosing them in parentheses after the *response* variable. The default is ORDER=FORMATTED.

### DESCENDING

### DESC

reverses the order of the response categories. If both the DESCENDING and ORDER= options are specified, PROC FMM orders the response categories according to the ORDER= option and then reverses that order.

**EVENT=**'*category*' | *keyword*

specifies the event category for the binary response model. PROC FMM models the probability of the event category. You can specify the value (formatted, if a format is applied) of the event category in quotes, or you can specify one of the following *keywords*:

**FIRST**

designates the first ordered category as the event. This is the default.

**LAST**

designates the last ordered category as the event.

**ORDER=***order-type*

specifies the sort order for the levels of the response variable. You can specify the following values for *order-type*:

**DATA**

sorts the levels by order of appearance in the input data set.

**FORMATTED**

sorts the levels by external formatted value, except for numeric variables with no explicit format, which are sorted by their unformatted (internal) value.

**FREQ**

sorts the levels by descending frequency count; levels with the most observations come first in the order.

**INTERNAL**

sorts the levels by unformatted value.

**FREQDATA**

sorts the levels by order of descending frequency count, and within counts by order of appearance in the input data set when counts are tied.

**FREQFORMATTED**

sorts the levels by order of descending frequency count, and within counts by formatted value (as above) when counts are tied.

**FREQINTERNAL**

sorts the levels by order of descending frequency count, and within counts by unformatted value when counts are tied.

When ORDER=FORMATTED (the default) for numeric variables for which you have supplied no explicit format (that is, for which there is no corresponding FORMAT statement in the current PROC FMM run or in the DATA step that created the data set), the levels are ordered by their internal (numeric) value. If you specify the ORDER= option in the MODEL statement and the ORDER= option in the PROC FMM statement, the former takes precedence.

By default, ORDER=FORMATTED. For the FORMATTED and INTERNAL values, the sort order is machine-dependent.

For more information about sort order, see the chapter on the SORT procedure in the *Base SAS Procedures Guide* and the discussion of BY-group processing in *SAS Language Reference: Concepts*.

**REFERENCE=**'*category*' | *keyword*

**REF=**'*category*' | *keyword*

specifies the reference category for categorical models. For the binary response model, specifying one response category as the reference category is the same as specifying the other response category as the event category. You can specify the value (formatted if a format is applied) of the reference category in quotation marks, or you can specify one of the following *keywords*:

**FIRST**

designates the first-ordered category as the reference category.

**LAST**

designates the last-ordered category as the reference category. This is the default.

The reference category in multinomial, Dirichlet-multinomial, and multinomial cluster models is determined by the order in which you specify the response variables.

## Model Options

**ALPHA=***number*

requests that the procedure construct confidence intervals that have confidence level 1−*number* for each parameter estimate. The value of *number* must be between 0 and 1; the default is 0.05.

**CL**

requests that PROC FMM construct confidence limits for each parameter estimate. The confidence level is 0.95 by default; you can change it by using the ALPHA= option.

**DISTRIBUTION=***keyword*

**DIST=***keyword*

specifies the probability distribution for a mixture component.

If you specify the DIST= option and you do not specify a link function by using the LINK= option, a default link function is chosen (see Table 43.5). If you do not specify a distribution, the FMM procedure defaults to the normal distribution for continuous response variables and to the binary distribution for classification or character variables, unless you use the *events/trials* syntax in the MODEL statement. If you use the *events/trials* syntax, PROC FMM defaults to the binomial distribution.

Table 43.5 lists *keywords* that you can specify for the DISTRIBUTION= option and their corresponding default link functions. For generalized linear models with these distributions, you can find expressions for the log-likelihood functions in the section "Log-Likelihood Functions for Response Distributions" on page 2948.

**Table 43.5** Keyword Values of the DIST= Option

| *keyword* | Alias | Distribution | Default Link Function |
|---|---|---|---|
| **BETA** | | Beta | Logit |
| **BETABINOMIAL** | **BETABIN**,**BB** | Beta-binomial | Logit |
| **BINARY** | **BERNOULLI** | Binary | Logit |
| **BINOMIAL** | **BIN** | Binomial | Logit |
| **BINOMCLUSTER** | **BCLUS** | Binomial cluster | Logit |

**Table 43.5** *continued*

| DIST= | Alias | Distribution | Default Link Function |
|---|---|---|---|
| **CONSTANT** <(*c*)> | **DEGENERATE** <(*c*)> | Degenerate | N/A |
| **DIRICHLETMULTINOMIAL** | **DIRIMULT**,**DM** | Dirichlet-multinomial | Generalized logit |
| **EXPONENTIAL** | **EXPO** | Exponential | Log |
| **FOLDEDNORMAL** | **FNORMAL** | Folded normal | Identity |
| **GAMMA** | **GAM** | Gamma | Log |
| **GAUSSIAN** | **NORMAL** | Normal | Identity |
| **GENPOISSON** | **GPOISSON** | Generalized Poisson | Log |
| **GEOMETRIC** | **GEOM** | Geometric | Log |
| **INVGAUSS** | **IGAUSSIAN**, **IG** | Inverse Gaussian | Inverse squared (power(–2)) |
| **LOGNORMAL** | **LOGN** | Lognormal | Identity |
| **MULTINOMIAL** | **MULTI** | Multinomial | Generalized logit |
| **MULTINOMCLUSTER** | **MCLUS** | Multinomial cluster | Logit |
| **NEGBINOMIAL** | **NEGBIN**, **NB** | Negative binomial | Log |
| **POISSON** | **POI** | Poisson | Log |
| **T** <($\nu$)> | **STUDENT** <($\nu$)> | *t* | Identity |
| **TRUNCEXPO** <(*a*,*b*)> | **TEXPO** <(*a*,*b*)> | Truncated exponential | Log |
| **TRUNCLOGN** <(*a*,*b*)> | **TLOGN** <(*a*,*b*)> | Lognormal | Identity |
| **TRUNCNEGBIN** | **TNEGBIN**, **TNB** | Negative binomial | Log |
| **TRUNCNORMAL** <(*a*,*b*)> | **TNORMAL** <(*a*,*b*)> | Truncated normal | Identity |
| **TRUNCPOISSON** | **TPOISSON**, **TPOI** | Truncated Poisson | Log |
| **UNIFORM** <(*a*,*b*)> | **UNIF** <(*a*,*b*)> | Uniform | N/A |
| **WEIBULL** | | Weibull | Log |

Note that the PROC FMM default link for the gamma or exponential distribution is not the canonical link (the reciprocal link).

The binomial cluster and multinomial cluster models are multiple-component models that are described in Morel and Nagaraj (1993); Morel and Neerchal (1997); Neerchal and Morel (1998). See Example 43.1 for an application of the binomial cluster model in a teratological experiment. See Example 43.4 for an application of the multinomial cluster model to housing survey data.

If you use the *events/trials* syntax, then the default distribution is the binomial distribution and only the following choices are available: DIST=BINOMIAL, DIST=BETABINOMIAL, and DIST=BINOMCLUSTER. The *trials* variable is ignored for all other distributions. This enables you to fit models in which some components have a binomial or binomial-like distribution. For example, suppose that the variable n is a binomial denominator and the variable logn is its logarithm. Then the following statements model a two-component mixture of a binomial and Poisson count model:

```
model y/n  = ;
model      +  / dist=Poisson offset=logn;
```

You use the OFFSET= option in the second MODEL statement to specify that the Poisson counts refer to different base counts, because the trial variable *n* is ignored in the second model.

If DIST=BINOMIAL is specified without the *events/trials* syntax, then $n$=1 is used for the default number of trials.

Similarly, if you specify multiple dependent variables by using the *response ... response* and *event ... event /trials* notation, the default distribution is the multinomial and only DIST=MULTINOMIAL, DIST=DIRICHLETMULTINOMIAL, and DIST=MULTINOMCLUSTER are available.

DIST=TRUNCNEGBIN and DIST=TRUNCPOISSON are zero-truncated versions of DIST=NEGBINOMIAL and DIST=POISSON, respectively—that is, only the value of 0 is excluded from the support.

For DIST=TRUNCEXPO, DIST=TRUNCLOGN, and DIST=TRUNCNORMAL, you must specify the lower ($a$) and upper ($b$) truncation points of the distribution. For example:

DIST=TRUNCEXPO<($a$,$b$)>

DIST=TRUNCLOGN<($a$,$b$)>

DIST=TRUNCNORMAL<($a$,$b$)>

Each of these distributions is the conditional version of its corresponding nontruncated distribution that is confined to the support $[a, b]$ (inclusive). You can specify a missing value (.) for either $a$ or $b$ to truncate only on the other side; that is, $a$=. indicates a right-truncated distribution, and $b$=. indicates a left-truncated distribution.

For several distribution specifications you can provide additional optional parameters to further define the distribution. These optional parameters are listed in the following:

**CONSTANT**<($c$)> The number $c$ specifies the value where the mass is concentrated. The default is DIST=CONSTANT(0), so you can add zero-inflation to any model by adding a MODEL statement with DIST=CONSTANT.

**T**<($\nu$)> The number $\nu$ specifies the degrees of freedom for the (shifted) $t$ distribution. The default is DIST=T(3); this leads to a heavy-tailed distribution for which the variance is defined. See the section "Log-Likelihood Functions for Response Distributions" on page 2948 for the density function of the shifted $t_\nu$ distribution.

**UNIFORM**<($a$,$b$)> The values $a$ and $b$ define the support of the uniform distribution, $a < b$. By default, $a = 0$ and $b = 1$.

**EQUATE=MEAN | SCALE | NONE | EFFECTS(*effect-list*)**

specifies simple sets of parameter constraints across the components in a MODEL statement; the default is EQUATE=NONE. This option is available only for maximum likelihood estimation. If you specify EQUATE=MEAN, the parameters that determine the mean are reduced to a single set that is applicable to all components in the MODEL statement. If you specify EQUATE=SCALE, a single parameter represents the common scale for all components in the MODEL statement. The EFFECTS option enables you to force the parameters for the chosen model effects to be equal across components; however, the number of parameters is unaffected.

For example, the following statements fit a two-component multiple regression model in which the coefficients for variable logd vary by component and the intercepts and coefficients for variable dose are the same for the two components:

```
proc fmm;
   model num = dose logd / equate=effects(int dose) k=2;
run;
```

To fix all coefficients across the two components, you can write the MODEL statement as

```
model num = dose logd / equate=effects(int dose logd) k=2;
```

or

```
model num = dose logd / equate=mean k=2;
```

If you restrict all parameters in a *k*-component MODEL statement to be equal, the FMM procedure reduces the model to *k*=1.

**K=***n*

**NUMBER=***n*

specifies the number of components the MODEL statement contributes to the overall mixture. For binomial cluster models and multinomial cluster models, this option is not available because these are multiple-component models by definition.

**KMAX=***n*

specifies the maximum number of components the MODEL statement contributes to the overall mixture.

If the maximum number of components in the mixture, as determined by all KMAX= options, is larger than the minimum number of components, the FMM procedure fits all possible models and displays summary fit information for the sequence of evaluated models. The "best" model according to the CRITERION= option in the PROC FMM statement is then chosen, and the remaining output and analyses performed by PROC FMM pertain to this "best" model.

When you use MCMC methods to estimate the parameters of a mixture, you need to ensure that the chain for a given value of *k* has converged; otherwise, comparisons among models that have varying numbers of components might not be meaningful. You can use the FITDETAILS option to display summary and diagnostic information for the MCMC chains from each model.

If you specify the KMIN= option but not the KMAX= option, then the default value for the KMAX= option is the value of the KMIN= option (unless KMIN=0, in which case the KMAX= option is set to 1).

**KMIN=***n*

specifies the minimum number of components that the MODEL statement contributes to the overall mixture. When you use MCMC methods to estimate the parameters of a mixture, you need to ensure that the chain for a given value of *k* has converged; otherwise, comparisons among models that have varying numbers of components might not be meaningful.

**KRESTART**

requests that the starting values for each analysis (that is, for each unique number of components as determined by the KMIN= and KMAX= options) be determined separately, in the same way as if no other analyses were performed. If you do not specify the KRESTART option, then the starting values for each analysis are based on results from the previous analysis with one less component.

**LABEL='***label***'**

specifies an optional label for the model that is used to identify the model in printed output, on graphics, and in data sets created from ODS tables.

**LINK=***keyword*

specifies the link function in the model. The *keywords* and expressions for the associated link functions are shown in Table 43.6.

**Table 43.6** Link Functions in MODEL Statement of the FMM Procedure

| LINK= | Alias | Link Function | $g(\mu) = \eta =$ |
|---|---|---|---|
| **CLOGLOG** | **CLL** | Complementary log-log | $\log(-\log(1-\mu))$ |
| **IDENTITY** | **ID** | Identity | $\mu$ |
| **LOG** | | Log | $\log(\mu)$ |
| **LOGIT** | | Logit | $\log(\mu/(1-\mu))$ |
| **LOGLOG** | | Log-log | $-\log(-\log(\mu))$ |
| **PROBIT** | **NORMIT** | Probit | $\Phi^{-1}(\mu)$ |
| **POWER**($\lambda$) | **POW**($\lambda$) | Power with exponent $\lambda$= *number* | $\begin{cases} \mu^{\lambda} & \text{if } \lambda \neq 0 \\ \log(\mu) & \text{if } \lambda = 0 \end{cases}$ |
| **POWERMINUS2** | | Power with exponent –2 | $1/\mu^2$ |
| **RECIPROCAL** | **INVERSE** | Reciprocal | $1/\mu$ |

The default link functions for the various distributions are shown in Table 43.5.

**NOINT**

excludes the intercept from the model. An intercept is included by default, unless the distribution is DIST=CONSTANT or DIST=UNIFORM.

**OFFSET=***variable*

specifies the offset variable function for the linear predictor in the model. An offset variable can be thought of as a regressor variable whose regression coefficient is known to be 1. For example, you can use an offset in a Poisson model when counts have been obtained in time intervals of different lengths. With a log link function, you can model the counts as Poisson variables with the logarithm of the time interval as the offset variable.

**PARAMETERS(***parameter-specification***)**

**PARMS(***parameter-specification***)**

specifies starting values for the model parameters. If no PARMS option is specified, the FMM procedure determines starting values by using a data-dependent algorithm. To determine initial values

for the Markov chain in Bayesian estimation, see the INITIAL= option in the BAYES statement. The specification of the parameters takes the following form: parameters in the mean function precede the scale parameters, and parameters for different components are separated by commas.

The following statements specify starting parameters for a two-component normal model. The initial values for the intercepts are 1 and –3; the initial values for the variances are 0.5 and 4.

```
proc fmm;
   model y = / k=2 parms(1 0.5, -3 4);
run;
```

You can specify missing values for parameters whose starting values are to be determined by the default method. Only values for parameters that participate in the optimization are specified. The values for model effects are specified on the linear (linked) scale.

## OUTPUT Statement

**OUTPUT** < **OUT=***SAS-data-set* >
< *keyword*< (*keyword-options*) > < **=***name* > > . . .
< *keyword*< (*keyword-options*) > < **=***name* > > < / *options* > **;**

The OUTPUT statement creates a data set that contains observationwise statistics that are computed after fitting the model. By default, all variables in the original data set are included in the output data set. You can use the ID statement to limit the variables copied from the input data set to the output data set.

The output statistics are computed based on the parameter estimates of the converged model if the parameters are estimated by maximum likelihood. If a Bayesian analysis is performed, the output statistics are computed based on the arithmetic mean in the posterior sample. You can change to the maximum posterior estimate with the ESTIMATE=MAP option in the BAYES statement.

You can specify the following syntax elements in the OUTPUT statement before the slash (/).

**OUT=***SAS-data-set*
specifies the name of the output data set. If the OUT= option is omitted, the procedure uses the DATA*n* convention to name the output data set.

*keyword*< (*keyword-options*) > < **=***name* >
specifies a statistic to include in the output data set and optionally assigns the variable the name *name*. If you do not provide a name, the FMM procedure assigns a default name based on the type of statistic requested. If you provide a name for a statistic that leads to multiple output statistics, the name is modified to index the associated component number. For multivariate responses, names are also modified to index the associated response.

You can use the *keyword-options* to control which type of a particular statistic is computed. The following are valid values for *keyword* and *keyword-options*:

**PREDICTED< (COMPONENT | OVERALL) >**

**PRED< (COMPONENT | OVERALL) >**

**MEAN< (COMPONENT | OVERALL) >**

requests predicted values (predicted means) for the response variable. The predictions in the output data set are mapped onto the data scale in all cases except for a multinomial, binomial, or binary response with *events/trials* syntax and when PREDTYPE=COUNT has not been specified. In this case, for the binomial or binary response, the predictions are predicted success probabilities. For the multinomial response, the predictions are probabilities for each response category.

The default is to compute the predicted value for the mixture (OVERALL). You can request predictions for the means of the component distributions by adding the COMPONENT suboption in parentheses. The predicted values for some distributions are not identical to the parameter modeled as $\mu$. For example, in the lognormal distribution the predicted mean is $\exp\{\mu + 0.5\phi\}$ where $\mu$ and $\phi$ are the parameters of an underlying normal process; see the section "Log-Likelihood Functions for Response Distributions" on page 2948 for details.

**RESIDUAL< (COMPONENT | OVERALL) >**

**RESID< (COMPONENT | OVERALL) >**

requests residuals for the response or residuals in the component distributions. Only "raw" residuals on the data scale are computed (observed minus predicted).

**VARIANCE< (COMPONENT | OVERALL) >**

**VAR< (COMPONENT | OVERALL) >**

requests variances for the mixture or the component distributions. For multivariate responses, the output statistic represents the diagonal entry in the associated covariance matrix.

**LOGLIKE< (COMPONENT | OVERALL) >**

**LOGL< (COMPONENT | OVERALL) >**

requests values of the log-likelihood function for the mixture or the components. For observations used in the analysis, the overall computed value is the observations' contribution to the log likelihood; if a FREQ statement is present, the frequency is accounted for in the computed value. In other words, if all observations in the input data set have been used in the analysis, adding the value of the log-likelihood contributions in the OUTPUT data set produces the negative of the final objective function value in the "Iteration History" table. By default, the log-likelihood contribution to the mixture is computed. You can request the individual mixture component contributions with the COMPONENT suboption.

**MIXPROBS< (COMPONENT | MAX) >**

**MIXPROB< (COMPONENT | MAX) >**

**PRIOR< (COMPONENT | MAX) >**

**MIXWEIGHTS< (COMPONENT | MAX) >**

requests that the prior weights $\pi_j(\mathbf{z}, \boldsymbol{\alpha}_j)$ be added to the OUTPUT data set. By default, the probabilities are output for all components. You can limit the output to a single statistic, the largest mixing probability, with the MAX suboption.

**NOTE:** The keyword "prior" is used here because of long-standing practice to refer to the mixing probabilities as prior weights. This must not be confused with the prior distribution and its parameters in a Bayesian analysis.

**POSTERIOR< (COMPONENT | MAX) >**
**POST< (COMPONENT | MAX) >**
**PROB< (COMPONENT | MAX) >**

requests that the posterior weights

$$\frac{\pi_j(\mathbf{z}, \boldsymbol{\alpha}_j)\, p_j(y; \mathbf{x}'_j \boldsymbol{\beta}_j, \phi_j)}{\sum_{j=1}^{k} \pi_j(\mathbf{z}, \boldsymbol{\alpha}_j)\, p_j(y; \mathbf{x}'_j \boldsymbol{\beta}_j, \phi_j)}$$

be added to the OUTPUT data set. By default, the probabilities are output for all components. You can limit the output to a single statistic, the largest posterior probability, with the MAX suboption.

**NOTE:** The keyword "posterior" is used here because of long-standing practice to refer to these probabilities as posterior probabilities. This must not be confused with the posterior distribution in a Bayesian analysis.

**LINP**
**XBETA**

requests that the linear predictors for the models be added to the OUTPUT data set.

**CLASS | CATEGORY | GROUP**

adds the estimated component membership to the OUTPUT data set. An observation is associated with the component that has the highest posterior probability.

**MAXPOST | MAXPROB**

adds the highest posterior probability to the OUTPUT data set.

A *keyword* can appear multiple times. For example, the following OUTPUT statement requests predicted values for the mixture in addition to the predicted means in the individual components:

```
output out=fmmout pred=MixtureMean pred(component)=CompMean;
```

In a three-component model, this produces four variables in the fmmout data set: MixtureMean, CompMean_1, CompMean_2, and CompMean_3.

You can specify the following *options* in the OUTPUT statement after a slash (/).

**ALLSTATS**

requests that all statistics are computed. If you do not use a *keyword* to assign a name, the FMM procedure uses the default name.

**NOVAR**

requests that variables from the input data set not be added to the output data set. This option does not apply to variables listed in the BY statement or to variables listed in the ID statement.

**PREDTYPE=PROB | COUNT**

specifies the type of predicted values that are produced for a multinomial, binomial, or binary response with *events/trials* syntax. If PREDTYPE=PROB, the predicted values are success probabilities. If PREDTYPE=COUNT, the predicted values are success counts. The default is PREDTYPE=PROB. In multinomial models that do not use *events/trials* syntax, PREDTYPE=COUNT produces predicted counts only if all responses in an observation are nonmissing.

## PERFORMANCE Statement

**PERFORMANCE** < *performance-options* > **;**

The PERFORMANCE statement enables you to control the performance characteristics of the FMM procedure (for example, the number of CPUs, the number of threads for multithreading, and so on). By default, the FMM procedure performs many analyses in multiple threads, and the number of threads equals the number of CPUs. Certain system and configuration options also can control the number of CPUs available to a SAS session or whether multithreaded computations are permissible. For example, you can set the number of available processors to two with

```
options cpucount=2;
```

The FMM procedure then acts as though two processors were available, regardless of the number of physically available processors.

The FMM procedure applies multithreading to the following analytical tasks:

**Starting values**: all starting value computations that require a pass through the data.

**Optimization**: all evaluations of objective function, gradient, and Hessian; computation of covariance matrix.

**Bayesian analysis**: all sample passes through the data, formation of cross-product matrices, sampling of latent variables, and posterior diagnostics.

**Scoring and ODS Graphics**: computation of all output statistics and statistics for the construction of graphics that require passes through the data.

You can specify the following *performance-options*:

**CPUCOUNT=*n* | ACTUAL**
specifies the number of processors available to the FMM procedure; the number *n* must be between 1 and 1024. CPUCOUNT=ACTUAL sets the number of available processors equal to the number of physical processors.

**DETAILS**
requests a table with timing detail for the tasks performed by the FMM procedure.

**NOTHREADS**
disables multithreaded computations.

**THREADS=YES | NO**
enables or disables multithreaded processing. The number of threads used by the FMM procedure is displayed in the “Bayes Information” or “Optimization Information” table. It typically equals the number of available CPUs, which can be different from the number of physical CPUs, and can be modified with the global CPUCOUNT SAS option or with the CPUCOUNT= option in the PERFORMANCE statement.

## PROBMODEL Statement

**PROBMODEL** < *effects* > < / *probmodel-options* > **;**

The PROBMODEL statement defines the model effects for the mixing probabilities and their link function and starting values. Model effects (other than the implied intercept) are not supported with Bayesian estimation. By default, the FMM procedure models mixing probabilities on the logit scale for two-component models and as generalized logit models in situations with more than two components. The PROBMODEL statement is not required.

The generalized logit model with $k$ categories has a common vector of regressor or design variables, $\mathbf{z}$, $k-1$ parameter vectors that vary with category, and one linear predictor whose value is constant. The constant linear predictor is assigned by the FMM procedure to the last component in the model, and its value is zero ($\boldsymbol{\alpha}_k = \mathbf{0}$). The probability of observing category $1 \le j \le k$ is then

$$\pi_j(\mathbf{z}, \boldsymbol{\alpha}_j) = \frac{\exp\{\mathbf{z}'\boldsymbol{\alpha}_j\}}{\sum_{i=1}^{k} \exp\{\mathbf{z}'\boldsymbol{\alpha}_i\}}$$

For $k$=2, the generalized logit model reduces to a model with the logit link (a logistic model); hence the attribute *generalized* logit.

By default, an intercept is included in the model for the mixing probabilities. If you suppress the intercept by using the NOINT option, you must specify at least one effect in the statement.

You can specify the following *probmodel-options* in the PROBMODEL statement after the slash (/):

**ALPHA=***number*

requests that confidence intervals that have the confidence level $1 - \mathit{number}$ be constructed for the parameters in the probability model. The value of *number* must be between 0 and 1; the default is 0.05. If the probability model is simple—that is, it does not contain any effects—the confidence intervals are produced for the estimated parameters (on the logit scale) and for the mixing probabilities. This option has no effect when you perform Bayesian estimation. You can modify credible interval settings by specifying the STATISTICS(ALPHA=) option in the BAYES statement.

**CL**

requests that confidence limits be constructed for each of the parameter estimates. The confidence level is 0.95 by default; this can be changed with the ALPHA= option.

**LINK=***keyword*

specifies the link function in the model for the mixing probabilities. The default is a logit link for models with two components. For models with more than two components, only the generalized logit link is available. The *keywords* and expressions for the associated link functions for two-component models are shown in Table 43.7.

**Table 43.7** Link Functions in the PROBMODEL Statement

| LINK= | Link Function | $g(\mu) = \eta =$ |
|---|---|---|
| **CLOGLOG \| CLL** | Complementary log-log | $\log(-\log(1-\mu))$ |
| **LOGIT** | Logit | $\log(\mu/(1-\mu))$ |
| **LOGLOG** | Log-log | $-\log(-\log(\mu))$ |
| **PROBIT \| NORMIT** | Probit | $\Phi^{-1}(\mu)$ |

**NOINT**

excludes the intercept from the model for the mixing probabilities. An intercept is included by default. If you suppress the intercept by using this option, you must specify at least one other effect for the mixing probabilities, because an empty probability model is not meaningful.

**PARAMETERS(***parameter-specification***)**

**PARMS(***parameter-specification***)**

specifies starting values for the parameters. The specification of the parameters takes the following form: parameters in the mean function appear in a list, and parameters for different components are separated by commas. Starting values are given on the linked scale, not in terms of probabilities. Also, you need to specify starting values for each of the first $k$–1 components in a $k$-component model. The linear predictor for the last component is always assumed to be 0.

The following statements specify a three-component mixture of multiple regression models. The PROBMODEL statement does not list any effects; a standard "intercept-only" generalized logit model is used to model the mixing probabilities.

```
proc fmm;
   model y = x1 x2 / k=3;
   probmodel  / parms(2, 1);
run;
```

There are three linear predictors in the model for the mixing probabilities, $\alpha_1$, $\alpha_2$, and $\alpha_3$. If you have starting values of $\alpha_1 = 2$, $\alpha_2 = 1$, and $\alpha_3 = 0$, this leads to initial mixing probabilities of

$$\pi_1 = \frac{e^2}{e^2 + e^1 + e^0} = 0.24$$

$$\pi_2 = \frac{e^1}{e^2 + e^1 + e^0} = 0.66$$

$$\pi_3 = \frac{e^0}{e^2 + e^1 + e^0} = 0.1$$

You can specify missing values for parameters whose starting values are to be determined by the default method.

## RANDSTART Statement

**RANDSTART** < *randstart-options* > **;**

The RANDSTART statement enables you use randomly generated sets of starting values for mixture model analysis with the maximum likelihood method. This statement is not available for a Bayesian analysis that is performed by the FMM procedure, and it is not available for model selection.

If you specify the RANDSTART statement, PROC FMM determines the maximum likelihood estimates by using a two-stage process. In the first stage, the procedure generates *n1* sets of starting values. To generate the sets of starting values, PROC FMM uses its typical starting values. These values are determined either by the procedure's default method of finding starting values or from the values that you specify in the

PARAMETERS option in the MODEL and PROBMODEL statements. This set of values is used as the first starting set. The remaining starting sets are generated around this first set, with random variation controlled by a scale value *s*. You can control this variation by using the STD= option.

The procedure then uses each set of starting values to initialize a separate optimization. Each optimization is terminated according to an initial convergence criterion *r1*. The goal of the first stage is to identify the *n2* ($\leq$ *n1*) sets of starting values that have the largest maximum likelihood values after this optimization. In the second stage, the optimization for these *n2* sets is continued according to a stricter convergence criterion, which you specify by using the convergence criteria in the PROC FMM statement. For more information about this process, see the section "Random Starting Values" on page 2959.

To fine-tune the two-stage random starting values process, you can specify the following *randstart-options*:

**ALLITER**

includes the iteration history for the first stage of estimation in the "Iteration History" table. By default, only the iteration history for the second stage is displayed.

**MAXFIRST=***n*

specifies the maximum number of starting value sets to generate in the first stage.

This option is effective only if the number of starting sets that converge in the first stage is less than the number that you specify in the NSECOND= option. The procedure generates more starting sets than the number that you specify in the NFIRST= option if the number of starting value sets that converge in the first stage is less than the value of the NSECOND= option.

The procedure generates no more than *n* sets, where *n* must be greater than or equal to the values that you specify in the NFIRST= and NSECOND= options.

The default value of the MAXFIRST= option depends on the values of the NFIRST= and NSECOND= options. If you specify NFIRST=*n1*, the default is $2 \times n1$. If you specify NSECOND=*n2* but do not specify a value for the NFIRST= option, the default is $20 \times n2$. If you do not specify either the NFIRST= or NSECOND= option, then by default MAXFIRST=200.

**NFIRST=***n1*

specifies the number of starting value sets to generate for the first stage of analysis. The procedure generates at least *n1* starting sets to obtain *n2* starting sets that converge according to the criterion that you specify in the RANDGCONV= option.

When you specify this option, the procedure generates no fewer than *n1* starting sets, where *n1* must be greater than or equal to the value of the NSECOND= option and less than or equal to the value of the MAXFIRST= option.

The default value depends on the values of the NSECOND= and MAXFIRST= options. If you specify MAXFIRST=*n*, the default value is $n/2$. If you specify NSECOND=*n2* but do not specify the MAXFIRST= option, the default value is $10 \times n2$. If you do not specify either the NSECOND= or MAXFIRST= option, then by default NFIRST=100.

**NSECOND=***n2*

specifies the number of starting sets to carry forward into the second stage of optimization. These starting sets achieve convergence in the first stage by meeting the convergence criterion that you specify in the RANDGCONV= option. In the second stage, these starting sets continue to be optimized under a stricter convergence criterion that you specify in the PROC FMM statement. At the end of this second

stage, the starting set that has the best log likelihood is the set of maximum likelihood estimates that PROC FMM uses to produce all remaining tables and output.

The value of *n2* must be less than or equal to the values that you specify in the NFIRST= and MAXFIRST= options.

The default value depends on the values of the NFIRST= and MAXFIRST= options. If you specify NFIRST=*n1*, the default is *n1*/10. If you do not specify a value for the NFIRST= option but you do specify MAXFIRST=*n*, the default is *n*/20. If you do not specify either the NFIRST= or MAXFIRST= option, then by default NSECOND=10.

**RANDGCONV=***r1*

specifies the convergence criterion to use in the first stage of optimization. This option is defined similarly to the GCONV= option in the PROC FMM statement, but it is used only in the first stage of the two-stage random starting method. The value of *r1* must be greater than zero and greater than the value of the GCONV= option.

Each starting set in the first stage undergoes iterative optimization until it converges according to the criterion *r1*. If you do not specify the GCONV= option in the PROC FMM statement, then by default RANDGCONV=1E–2. If you specify GCONV=*r* and the value of *r* is between 0 and 1, the default RANDGCONV= value is the square root of *r*. If you specify GCONV=*r* and the value of *r* is greater than or equal to 1, then by default RANDGCONV=2*r*.

**STD=***s*

specifies the standard deviation for the random draw process that generates the starting sets. The value of *s* must be greater than or equal to 1E–1.

The random drawing process centers on the default starting values or the values that you specify in the PARAMETERS option of the MODEL or PROBMODEL statement.

In the random drawing, scale parameters are drawn from a lognormal distribution whose standard deviation is equal to *s*. Mean parameters are drawn from a normal distribution whose standard deviation is equal to *s*.

By default, STD=2.

## RESTRICT Statement

**RESTRICT** <'*label*'> *constraint-specification* <**,** . . . **,** *constraint-specification*>
<*operator* <*value*>> <*/ option*> **;**

The RESTRICT statement enables you to specify linear equality or inequality constraints among the parameters of a mixture model. These restrictions are incorporated into the maximum likelihood analysis. This statement is not available for a Bayesian analysis that is performed by the FMM procedure. It is not available for multinomial or Dirichlet-multinomial models that have more than two response variables or more than one event variable.

Following are reasons why you might want to place constraints and restrictions on the model parameters:

- to fix a parameter at a particular value

- to equate parameters in different components in a mixture
- to impose order conditions on the parameters in a model
- to specify contrasts among the parameters that the fitted model should honor

A restriction is composed of a left-hand side and a right-hand side, separated by an operator. If the operator and right-hand side are not specified, the restriction is assumed to be an equality constraint against zero. If the right-hand side is not specified, the value is assumed to be zero.

An individual *constraint-specification* is written in (nearly) the same form as estimable linear functions are specified in the ESTIMATE statement of the GLM, MIXED, or GLIMMIX procedure. The *constraint-specification* takes the form

*model-effect value-list* < . . . *model-effect value-list* > < **(SCALE=** *value***)** >

At least one *model-effect* must be specified followed by one or more values in the *value-list*. The values in the list correspond to the multipliers of the corresponding parameter that is associated with the position in the model effect. If you specify more values in the *value-list* than the *model-effect* occupies in the model design matrix, the extra coefficients are ignored.

To specify restrictions for effects in specific components in the model, separate the *constraint-specification* by commas. The following statements provide an example:

```
proc fmm;
   class A;
   model y/n = A x /  k = 2;
   restrict A 1 0 -1;
   restrict x 2, x -1 >= 0.5;
run;
```

The linear predictors for this two-component model can be written as

$$\eta_1 = \beta_{10} + \alpha_{11} A_1 + \cdots + \alpha_{1a} A_a + x\beta_{11}$$
$$\eta_2 = \beta_{20} + \alpha_{21} A_1 + \cdots + \alpha_{2a} A_a + x\beta_{21}$$

where $A_k$ is the binary variable associated with the $k$th level of A.

The first RESTRICT statement applies only to the first component and specifies that the parameter estimates that are associated with the first and third level of the A effect are identical. In terms of the linear predictor, the restriction can be written as

$$\alpha_{11} - \alpha_{13} = 0$$

Now suppose that A has only two levels. Then the FMM procedure ignores the value –1 in the first RESTRICT statement and imposes the restriction

$$\alpha_{11} = 0$$

on the fitted model.

The second RESTRICT statement involves parameters in two different components of the model. In terms of the linear predictors, the restriction can be written as

$$2\beta_{11} - \beta_{21} \geq \frac{1}{2}$$

When restrictions are specified explicitly through the RESTRICT statement or implied through the EQUATE=EFFECTS option in the MODEL statement, the FMM procedure lists all restrictions after the model fit in a table of linear constraints and indicates whether a particular constraint is active at the converged solution.

The following operators can be specified to separate the left- and right-hand sides of the restriction: **=**, **>**, **<**, **>=**, **<=**.

Some distributions involve scale parameters (the parameter $\phi$ in the expressions of the log likelihood) and you can also use the *constraint-specification* to involve a component's scale parameter in a constraint. To this end, assign a value to the keyword SCALE, separated from the model effects and value lists with parentheses. The following statements fit a two-component normal model and restrict the component variances to be equal:

```
proc fmm;
   model y = / k=2;
   restrict int 0 (scale  1),
            int 0 (scale -1);
run;
```

The intercept specification is necessary because each *constraint-specification* requires at least one model effect. The zero coefficient ensures that the intercepts are not involved in the restriction. Instead, the RESTRICT statement leads to $\phi_1 - \phi_2 = 0$.

You can specify the following *option* in the RESTRICT statement after a slash (/).

**DIVISOR=***value*
specifies a *value* by which all coefficients on the right-hand side and left-hand side of the restriction are divided.

## WEIGHT Statement

**WEIGHT** *variable* **;**

The WEIGHT statement is used to perform a weighted analysis. Consult the section "Log-Likelihood Functions for Response Distributions" on page 2948 for expressions on how weight variables are included in the log-likelihood functions. Because the probability structure of a mixture model is different from that of a classical statistical model, the presence of a weight variable in a mixture model *cannot* be interpreted as altering the variance of an observation.

Observations with nonpositive or missing weights are not included in the PROC FMM analysis. If a WEIGHT statement is not included, all observations used in the analysis are assigned a weight of 1.

# Details: FMM Procedure

## A Gentle Introduction to Finite Mixture Models

### The Form of the Finite Mixture Model

Suppose that you observe realizations of a random variable $Y$, the distribution of which depends on an unobservable random variable $S$ that has a discrete distribution. $S$ can occupy one of $k$ states, the number of which might be unknown but is at least known to be finite. Because $S$ is not observable, it is frequently referred to as a latent variable.

Let $\pi_j$ denote the probability that $S$ takes on state $j$. Conditional on $S = j$, the distribution of the response $Y$ is assumed to be $f_j(y; \alpha_j, \boldsymbol{\beta}_j | S = j)$. In other words, each distinct state $j$ of the random variable $S$ leads to a particular distributional form $f_j$ and set of parameters $\{\alpha_j, \boldsymbol{\beta}_j\}$ for $Y$.

Let $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$ denote the collection of $\alpha_j$ and $\boldsymbol{\beta}_j$ parameters across all $j = 1$ to $k$. The marginal distribution of $Y$ is obtained by summing the joint distribution of $Y$ and $S$ over the states in the support of $S$:

$$
\begin{aligned}
f(y; \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \sum_{j=1}^{k} \Pr(S = j)\, f(y; \alpha_j, \boldsymbol{\beta}_j | S = j) \\
&= \sum_{j=1}^{k} \pi_j\, f(y; \alpha_j, \boldsymbol{\beta}_j | S = j)
\end{aligned}
$$

This is a mixture of distributions, and the $\pi_j$ are called the mixture (or prior) probabilities. Because the number of states $k$ of the latent variable $S$ is finite, the entire model is termed a finite mixture (of distributions) model.

The finite mixture model can be expressed in a more general form by representing $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in terms of regressor variables and parameters with optional additional scale parameters for $\boldsymbol{\beta}$. The section "Notation for the Finite Mixture Model" on page 2876 develops this in detail.

### Mixture Models Contrasted with Mixing and Mixed Models: Untangling the Terminology Web

Statistical terminology can have its limitations. The terms mixture, mixing, and mixed models are sometimes used interchangeably, causing confusion. Even worse, the terms arise in related situations. One application needs to be eliminated from the discussion in this documentation: mixture experiments, where design factors are the proportions with which components contribute to a blend, are not mixture models and do not fall under the purview of the FMM procedure. However, the data from a mixture experiment might be analyzed with a mixture model, a mixing model, or a mixed model, besides other types of statistical models.

Suppose that you observe realizations of random variable $Y$ and assume that $Y$ follows some distribution $f(y; \boldsymbol{\alpha}, \boldsymbol{\beta})$ that depends on parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Furthermore, suppose that the model is found to be deficient in the sense that the variability implied by the fitted model is less than the observed variability in the data, a condition known as overdispersion (see the section "Overdispersion" on page 2947). To tackle the problem the statistical model needs to be modified to allow for more variability. Clearly, one way of doing this is to introduce additional random variables into the process. Mixture, mixing, and mixed models are simply

different ways of adding such random variables. The section "The Form of the Finite Mixture Model" on page 2945 explains how mixture models add a discrete state variable $S$. The following two subsections explain how mixing and mixed models instead assume variation for a natural parameter or in the mean function.

### Mixing Models

Suppose that the model is modified to allow for some random quantity $U$, which might be one of the parameters of the model or a quantity related to the parameters. Now there are two distributions to cope with: the conditional distribution of the response given the random effect $U$,

$$f(y;\boldsymbol{\alpha},\boldsymbol{\beta}|u)$$

and the marginal distribution of the data. If $U$ is continuous, the marginal distribution is obtained by integration:

$$f(y;\boldsymbol{\alpha},\boldsymbol{\beta}) = \int f(y;\boldsymbol{\alpha},\boldsymbol{\beta}|u)\, f(u)\, du$$

Otherwise, it is obtained by summation over the support of $U$:

$$f(y;\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_u \Pr(U = u)\, f(y;\boldsymbol{\alpha},\boldsymbol{\beta}|u)$$

The important entity for statistical estimation is the marginal distribution $f(y;\boldsymbol{\alpha},\boldsymbol{\beta})$; the conditional distribution is often important for model description, genesis, and interpretation.

In a mixing model the marginal distribution is known and is typically of a well-known form. For example, if $Y|n$ has a binomial$(n,\mu)$ distribution and $n$ follows a Poisson distribution, then the marginal distribution of $Y$ is Poisson. The preceding operation is called *mixing* a binomial distribution with a Poisson distribution. Similarly, when mixing a Poisson$(\lambda)$ distribution with a gamma$(a,b)$ distribution for $\lambda$, a negative binomial distribution results as the marginal distribution. Other important mixing models involve mixing a binomial$(n,\mu)$ random variable with a beta$(a,b)$ distribution for the binomial success probability $\mu$. This results in a distribution known as the beta-binomial.

The finite mixtures have in common with the mixing models the introduction of random effects into the model to vary some or all of the parameters at random.

### Mixed Models

The difference between a mixing and a mixed model is that the conditional distribution is not that important in the mixing model. It matters to motivate the overdispersed reference model and to arrive at the marginal distribution. Inferences with respect to the conditional distribution, such as predicting the random variable $U$, are not performed in mixing models. In a mixed model the random variable $U$ typically follows a continuous distribution—almost always a normal distribution. The random effects usually do not model the natural parameters of the distribution; instead, they are involved in linear predictors that relate to the conditional mean. For example, a linear mixed model is a model in which the response and the random effects are normally distributed, and the random effects enter the conditional mean function linearly:

$$\begin{aligned}
\mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \boldsymbol{\epsilon} \\
\mathbf{U} &\sim N(\mathbf{0},\mathbf{G}) \\
\boldsymbol{\epsilon} &\sim N(\mathbf{0},\mathbf{R}) \\
\mathrm{Cov}[\mathbf{U},\boldsymbol{\epsilon}] &= \mathbf{0}
\end{aligned}$$

The conditional and marginal distributions are then

$$\mathbf{Y}|\mathbf{U} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \boldsymbol{\epsilon}, \mathbf{R})$$
$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R})$$

For this model, because of the linearity in the mean and the normality of the random effects, you could also refer to mixing the normal vector $\mathbf{Y}$ with the normal vector $\mathbf{U}$, because the marginal distribution is known. The linear mixed model can be fit with the MIXED procedure. When the conditional distribution is not normal and the random effects are normal, the marginal distribution does not have a closed form. In this class of mixed models, called generalized linear mixed models, model approximations and numerical integration methods are commonly used in model fitting; see for example, those models fit by the GLIMMIX and NLMIXED procedures. Chapter 6, "Introduction to Mixed Modeling Procedures," contains details about the various classes of mixed models and about the relevant SAS/STAT procedures.

The previous expression for the marginal variance in the linear mixed model, $\mathrm{var}[\mathbf{Y}] = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$, emphasizes again that the variability in the marginal distribution of a model that contains random effects exceeds the variability in a model without the random effects ($\mathbf{R}$).

The finite mixtures have in common with the mixed models that the marginal distribution is not necessarily a well-known model, but is expressed through a formal integration over the random-effects distribution. In contrast to the mixed models, in particular those involving nonnormal distributions or nonlinear elements, this integration is rather trivial; it reduces to a weighted and finite sum of densities or mass functions.

## Overdispersion

Overdispersion is the condition by which the data are more dispersed than is permissible under a reference model. Overdispersion arises only if the variability a model can capture is limited (for example, because of a functional relationship between mean and variance). For example, a model for normal data can never be overdispersed in this sense, although the reasons that lead to overdispersion also negatively affect a misspecified model for normal data. For example, omitted variables increase the residual variance estimate because variability that should have been modeled through changes in the mean is now "picked up" as error variability.

Overdispersion is important because an overdispersed model can lead to misleading inferences and conclusions. However, diagnosing and remedying overdispersion is complicated. In order to handle it appropriately, the source of overdispersion must be identified. For example, overdispersion can arise from any of the following conditions alone or in combination:

- omitted variables and model effects
- omitted random effects (a source of random variation is not being modeled or is modeled as a fixed effect)
- correlation among the observations
- incorrect distributional assumptions
- incorrectly specified mean-variance relationships
- outliers in the data

As discussed in the previous section, introducing randomness into a system increases its variability. Mixture, mixed, and mixing models have thus been popular in modeling data that appear overdispersed. Finite mixture models are particularly powerful in this regard, because even low-order mixtures of basic, symmetric distributions (such as two- or three-component mixtures of normal or $t$ distributions) enable you to model data with multiple modes, heavy tails, and skewness. In addition, the latent variable $S$ provides a natural way to accommodate omitted, unobservable variables into the model.

One approach to remedy overdispersion is to apply simple modifications of the variance function of the reference model. For example, with binomial-type data this approach replaces the variance of the binomial count variable $Y \sim \text{Binomial}(n, \mu)$, $\text{Var}[Y] = n \times \mu(1-\mu)$ with a scaled version, $\phi n \times \mu(1-\mu)$, where $\phi$ is called an overdispersion parameter, $\phi > 0$.

In addressing overdispersion problems, it is important to tackle the problem at its root. A missing scale factor on the variance function is hardly ever the root cause of overdispersion; it is only the easiest remedy.

## Log-Likelihood Functions for Response Distributions

The FMM procedure calculates the log likelihood that corresponds to a particular response distribution according to the following formulas. The response distribution is the distribution specified (or chosen by default) through the DIST= option in the MODEL statement. The parameterizations used for log-likelihood functions of these distributions were chosen to facilitate expressions in terms of mean parameters that are modeled through an (inverse) link functions and in terms of scale parameters. These are not necessarily the parameterizations in which parameters of prior distributions are specified in a Bayesian analysis of homogeneous mixtures. See the section "Prior Distributions" on page 2958 for details about the parameterizations of prior distributions.

The FMM procedure includes all constant terms in the computation of densities or mass functions. In the expressions that follow, $l$ denotes the log-likelihood function, $\phi$ denotes a general scale parameter, $\mu_i$ is the "mean", and $w_i$ is a weight from the use of a WEIGHT statement.

For some distributions (for example, the Weibull distribution) $\mu_i$ is not the mean of the distribution. The parameter $\mu_i$ is the quantity that is modeled as $g^{-1}(\mathbf{x}`\boldsymbol{\beta})$, where $g^{-1}(\cdot)$ is the inverse link function and the $\mathbf{x}$ vector is constructed based on the effects in the MODEL statement. Situations in which the parameter $\mu$ does not represent the mean of the distribution are explicitly mentioned in the list that follows.

The parameter $\phi$ is frequently labeled as a "Scale" parameter in output from the FMM procedure. It is not necessarily the scale parameter of the particular distribution.

**Beta** $(\mu, \phi)$

$$
\begin{aligned}
l(\mu_i, \phi; y_i, w_i) = \log &\left\{ \frac{\Gamma(\phi/w_i)}{\Gamma(\mu_i\phi/w_i)\Gamma((1-\mu_i)\phi/w_i)} \right\} \\
&+ (\mu_i\phi/w_i - 1)\log\{y_i\} \\
&+ ((1-\mu_i)\phi/w_i - 1)\log\{1-y_i\}
\end{aligned}
$$

This parameterization of the beta distribution is due to Ferrari and Cribari-Neto (2004) and has properties $\text{E}[Y] = \mu$, $\text{Var}[Y] = \mu(1-\mu)/(1+\phi)$, $\phi > 0$.

**Beta-binomial** $(n; \mu, \phi)$

$$
\begin{aligned}
\phi &= (1-\rho^2)/\rho^2 \\
l(\mu_i, \rho; y_i) &= \log\{\Gamma(n_i+1)\} - \log\{\Gamma(y_i+1)\} \\
&\quad - \log\{\Gamma(n_i - y_i + 1)\} \\
&\quad + \log\{\Gamma(\phi)\} - \log\{\Gamma(n_i+\phi)\} + \log\{\Gamma(y_i+\phi\mu_i)\} \\
&\quad + \log\{\Gamma(n_i - y_i + \phi(1-\mu_i))\} - \log\{\Gamma(\phi\mu_i)\} \\
&\quad - \log\{\Gamma(\phi(1-\mu_i))\} \\
l(\mu_i, \rho; y_i, w_i) &= w_i l(\mu_i, \rho; y_i)
\end{aligned}
$$

where $y_i$ and $n_i$ are the *events* and *trials* in the *events/trials* syntax and $0 < \mu < 1$. This parameterization of the beta-binomial model presents the distribution as a special case of the Dirichlet-Multinomial distribution—see, for example, Neerchal and Morel (1998). In this parameterization, $\mathrm{E}[Y] = n\mu$ and $\mathrm{Var}[Y] = n\mu(1-\mu)(1+(n-1)/(\phi+1))$, $0 \le \rho \le 1$. The FMM procedure models the parameter $\phi$ and labels it "Scale" on the procedure output. For other parameterizations of the beta-binomial model, see Griffiths (1973) or Williams (1975).

**Binomial** $(n; \mu)$

$$
\begin{aligned}
l(\mu_i; y_i) &= y_i \log\{\mu_i\} + (n_i - y_i)\log\{1-\mu_i\} \\
&\quad + \log\{\Gamma(n_i+1)\} - \log\{\Gamma(y_i+1)\} \\
&\quad - \log\{\Gamma(n_i - y_i + 1)\} \\
l(\mu_i; y_i, w_i) &= w_i\, l(\mu_i; y_i)
\end{aligned}
$$

where $y_i$ and $n_i$ are the *events* and *trials* in the *events/trials* syntax and $0 < \mu < 1$. In this parameterization $\mathrm{E}[Y] = n\mu$, $\mathrm{Var}[Y] = n\mu(1-\mu)$.

**Binomial Cluster** $(n; \mu, \pi)$

$$
\begin{aligned}
z &= \log\{\Gamma(n_i+1)\} - \log\{\Gamma(y_i+1)\} - \log\{\Gamma(n_i - y_i + 1)\} \\
\mu_i^* &= (1-\mu_i)\pi_i
\end{aligned}
$$

$$
\begin{aligned}
l(\mu_i, \pi_i; y_i) = z + \log\{&\pi_i(\mu_i^* + \mu_i)^{y_i}(1-\mu_i^*-\mu_i)^{n_i-y_i} \\
&+ (1-\pi_i)(\mu_i^*)^{y_i}(1-\mu_i^*)^{n_i-y_i}\}
\end{aligned}
$$

$$
l(\mu_i, \pi_i; y_i, w_i) = w_i l(\mu_i, \pi_i; y_i)
$$

In this parameterization, $\mathrm{E}[Y] = n\pi$ and $\mathrm{Var}[Y] = n\pi(1-\pi)\left\{1+\mu^2(n-1)\right\}$. The binomial cluster model is a two-component mixture of a binomial$(n, \mu^* + \mu)$ and a binomial$(n, \mu^*)$ random variable. This mixture is unusual in that it fixes the number of components and because the mixing probability $\pi$ appears in the moments of the mixture components. For further details, see Morel and Nagaraj (1993); Morel and Neerchal (1997); Neerchal and Morel (1998) and Example 43.1 in this chapter. The expressions for the mean and variance in the binomial cluster model are identical to those of the beta-binomial model shown previously, with $\pi_{bc} = \mu_{bb}$, $\mu_{bc} = \rho_{bb}$.

The FMM procedure models the parameter $\mu$ through the MODEL statement and the parameter $\pi$ through the PROBMODEL statement.

**Constant** ($c$)

$$l(y_i) = \begin{cases} 0 & |y_i - c| < \epsilon \\ -1\text{E}20 & |y_i - c| \geq \epsilon \end{cases}$$

The extreme value when $|y_i - c| \geq \epsilon$ is an approximation of $\log(0) = -\infty$, chosen so that $\exp\{l(y_i)\}$ yields a likelihood of 0. You can change this value by using the INVALIDLOGL= option in the PROC FMM statement. The constant distribution is useful for modeling overdispersion due to zero-inflation (or inflation of the process at support $c$).

The DIST=CONSTANT distribution is useful for modeling an inflated probability of observing a particular value (0, by default) in data from other discrete distributions, as demonstrated in the section "Modeling Zero-Inflation: Is It Better to Fish Poorly or Not to Have Fished at All?" on page 2884. Although it is syntactically valid to mix a constant distribution with a continuous distribution, such as DIST=LOGNORMAL, such a mixture is not mathematically appropriate, because the constant log likelihood is the logarithm of a probability, whereas a continuous log likelihood is the logarithm of a probability density function. If you want to mix a constant distribution with a continuous distribution, you could model the constant as a very narrow continuous distribution, such as DIST=UNIFORM($c-\delta$, $c+\delta$) for a small value $\delta$. However, fitting this type of mixture model can be sensitive to roundoff error and other numerical inaccuracies. Instead, the following approach is mathematically equivalent and more numerically stable:

1. Estimate the mixing probability $\text{P}(Y = c)$ as the proportion of observations in the data set such that $|y_i - c| < \epsilon$.
2. Estimate the parameters of the continuous distribution from the observations for which $|y_i - c| \geq \epsilon$.

**Dirichlet-multinomial** ($n, \boldsymbol{\mu}, \phi$)

$$\begin{aligned} l(\boldsymbol{\mu}_i, \phi; \mathbf{y}_i) &= \log\{\Gamma(n_i + 1)\} + \log\{\Gamma(\phi)\} + \sum_{j=1}^{M} \log\{\Gamma(y_{ij} + \phi\mu_{ij})\} \\ &\quad - \sum_{j=1}^{M} \log\{\Gamma(y_{ij} + 1)\} - \sum_{j=1}^{M} \log\{\Gamma(\phi\mu_{ij})\} - \log\{\Gamma(\sum_{j=1}^{M}[y_{ij} + \phi\mu_{ij}])\} \\ l(\boldsymbol{\mu}_i, \phi; \mathbf{y}_i, w_i) &= w_i l(\boldsymbol{\mu}_i, \phi; \mathbf{y}_i) \end{aligned}$$

where $M$ is the number of levels of the response variable, $n_i = \sum_{j=1}^{M} y_{ij}$, $0 < \mu_{ij} < 1$ for each $j$, and $\mu_{iM} = 1 - \sum_{j=1}^{M-1} \mu_{ij}$.

The mean and variance are $\text{E}[\mathbf{Y}] = n\boldsymbol{\mu}$ and $\text{Var}[\mathbf{Y}] = n\left(\text{Diag}(\boldsymbol{\mu}) - \boldsymbol{\mu}\boldsymbol{\mu}^T\right)\left(\frac{n+\phi}{1+\phi}\right)$.

**Exponential** ($\mu$)

$$l(\mu_i; y_i, w_i) = \begin{cases} -\log\{\mu_i\} - y_i/\mu_i & w_i = 1 \\ w_i \log\left\{\frac{w_i y_i}{\mu_i}\right\} - \frac{w_i y_i}{\mu_i} - \log\{y_i \Gamma(w_i)\} & w_i \neq 1 \end{cases}$$

In this parameterization, $\text{E}[Y] = \mu$ and $\text{Var}[Y] = \mu^2$.

**Folded Normal** $(\mu, \phi)$

$$l(\mu_i, \phi; y_i, w_i) = -\frac{1}{2}\log\{2\pi\} - \frac{1}{2}\log\{\phi/w_i\} + \log\left\{\exp\left\{\frac{-w_i(y_i-\mu_i)^2}{2\phi}\right\} + \exp\left\{\frac{-w_i(y_i+\mu_i)^2}{2\phi}\right\}\right\}$$

If $X$ has a normal distribution with mean $\mu$ and variance $\phi$, then $Y = |X|$ has a folded normal distribution and log-likelihood function $l(\mu, \phi; y, w)$ for $y \geq 0$. The folded normal distribution arises, for example, when normally distributed measurements are observed, but their signs are not observed. The mean and variance of the folded normal in terms of the underlying $\mathrm{N}(\mu, \phi)$ distribution are

$$\begin{aligned}\mathrm{E}[Y] &= \frac{1}{\sqrt{2\pi\phi}}\exp\left\{-\frac{\mu^2}{2/\phi}\right\} + \mu\left(1 - 2\Phi\left(-\mu/\sqrt{\phi}\right)\right)\\ \mathrm{Var}[Y] &= \phi + \mu^2 - \mathrm{E}[Y]^2\end{aligned}$$

The FMM procedure models the folded normal distribution through the mean $\mu$ and variance $\phi$ of the underlying normal distribution. When the FMM procedure computes output statistics for the response variable (for example when you use the OUTPUT statement), the mean and variance of the response $Y$ are reported. Similarly, the fit statistics apply to the distribution of $Y = |X|$, not the distribution of $X$. When you model a folded normal variable, the response input variable should be positive; the FMM procedure treats negative values of $Y$ as a support violation.

**Gamma** $(\mu, \phi)$

$$l(\mu_i, \phi; y_i, w_i) = w_i\phi\log\left\{\frac{w_i y_i \phi}{\mu_i}\right\} - \frac{w_i y_i \phi}{\mu_i} - \log\{y_i\} - \log\{\Gamma(w_i\phi)\}$$

In this parameterization, $\mathrm{E}[Y] = \mu$ and $\mathrm{Var}[Y] = \mu^2/\phi$, $\phi > 0$. This parameterization of the gamma distribution differs from that in the GLIMMIX procedure, which expresses the log-likelihood function in terms of $1/\phi$ in order to achieve a variance function suitable for mixed model analysis.

**Geometric** $(\mu)$

$$l(\mu_i; y_i, w_i) = y_i\log\left\{\frac{\mu_i}{w_i}\right\} - (y_i + w_i)\log\left\{1 + \frac{\mu_i}{w_i}\right\} + \log\left\{\frac{\Gamma(y_i + w_i)}{\Gamma(w_i)\Gamma(y_i + 1)}\right\}$$

In this parameterization, $\mathrm{E}[Y] = \mu$ and $\mathrm{Var}[Y] = \mu + \mu^2$. The geometric distribution is a special case of the negative binomial distribution with $\phi = 1$.

**Generalized Poisson** $(\mu, \phi)$

$$\begin{aligned}\xi_i &= 1 - \exp\{-\phi\}/w_i\\ \mu_i^* &= \mu_i - \xi(\mu_i - y_i)\\ l(\mu_i^*, \xi_i; y_i, w_i) &= \log\{\mu_i^* - \xi_i y_i\} + (y_i - 1)\log\{\mu_i^*\}\\ &\quad - \mu_i^* - \log\{\Gamma(y_i + 1)\}\end{aligned}$$

In this parameterization, $\mathrm{E}[Y] = \mu$, $\mathrm{Var}[Y] = \mu/(1-\xi)^2$, and $\phi \geq 0$. The FMM procedure models the mean $\mu$ through the effects in the MODEL statement and applies a log link by default. The generalized Poisson distribution provides an overdispersed alternative to the Poisson distribution; $\phi = \xi_i = 0$ produces the mass function of a regular Poisson random variable. For details about the generalized Poisson distribution and a comparison with the negative binomial distribution, see Joe and Zhu (2005).

**Inverse Gaussian** $(\mu, \phi)$

$$l(\mu_i, \phi; y_i, w_i) = -\frac{1}{2}\left[\frac{w_i(y_i-\mu_i)^2}{y_i\phi\mu_i^2} + \log\left\{\frac{\phi y_i^3}{w_i}\right\} + \log\{2\pi\}\right]$$

The variance is $\mathrm{Var}[Y] = \phi\mu^3$, $\phi > 0$.

**Lognormal** $(\mu, \phi)$

$$z_i = \log\{y_i\} - \mu_i$$

$$l(\mu_i, \phi; y_i, w_i) = -\frac{1}{2}\left(2\log\{y_i\} + \log\left\{\frac{\phi}{w_i}\right\} + \log\{2\pi\} + \frac{w_i z_i^2}{\phi}\right)$$

If $X = \log\{Y\}$ has a normal distribution with mean $\mu$ and variance $\phi$, then $Y$ has the log-likelihood function $l(\mu_i, \phi; y_i, w_i)$. The FMM procedure models the lognormal distribution and not the "shortcut" version you can obtain by taking the logarithm of a random variable and modeling that as normally distributed. The two approaches are not equivalent, and the approach taken by PROC FMM is the actual lognormal distribution. Although the lognormal model is a member of the exponential family of distributions, it is not in the "natural" exponential family because it cannot be written in canonical form.

In terms of the parameters $\mu$ and $\phi$ of the underlying normal process for $X$, the mean and variance of $Y$ are $\mathrm{E}[Y] = \exp\{\mu\}\sqrt{\omega}$ and $\mathrm{Var}[Y] = \exp\{2\mu\}\omega(\omega-1)$, respectively, where $\omega = \exp\{\phi\}$. When you request predicted values with the OUTPUT statement, the FMM procedure computes $\mathrm{E}[Y]$ and not $\mu$.

**Multinomial** $(n, \boldsymbol{\mu})$

$$l(\boldsymbol{\mu}_i; \mathbf{y}_i) = \sum_{j=1}^{M} y_{ij}\log(\mu_{ij}) + \log\{\Gamma(n_i+1)\} - \sum_{j=1}^{M}\log\{\Gamma(y_{ij}+1)\}$$

$$l(\boldsymbol{\mu}_i; \mathbf{y}_i, w_i) = w_i l(\boldsymbol{\mu}_i, \mathbf{y}_i)$$

where $M$ is the number of levels of the response variable, $n_i = \sum_{j=1}^{M} y_{ij}$, $0 < \mu_{ij} < 1$ for each $j$, and $\sum_{j=1}^{M}\mu_{ij} = 1$.

The mean and variance are $\mathrm{E}[\mathbf{Y}] = n\boldsymbol{\mu}$ and $\mathrm{Var}[\mathbf{Y}] = n(\mathrm{Diag}(\boldsymbol{\mu}) - \boldsymbol{\mu}\boldsymbol{\mu}^T)$, respectively, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{M-1})$.

**Multinomial Cluster** $(n, \mu, \boldsymbol{\pi})$

$$K(\mu_i, \boldsymbol{\pi}_i, \mathbf{y}_i) = \sum_{h=1}^{M}\left\{\pi_{ih}\prod_{g=1}^{M}[\pi_{ig}(1-\mu_i) + \mu_i 1(g=h)]^{y_{ig}}\right\}$$

$$l(\mu_i, \boldsymbol{\pi}_i, \mathbf{y}_i) = \log\{\Gamma(n_i+1)\} - \sum_{g=1}^{M}\log\{\Gamma(y_{ij}+1)\} + \log\{K(\mu_i, \boldsymbol{\pi}_i, \mathbf{y}_i)\}$$

$$l(\mu_i, \boldsymbol{\pi}_i, \mathbf{y}_i, w_i) = w_i l(\mu_i, \boldsymbol{\pi}_i, \mathbf{y}_i)$$

The multinomial cluster model for a response that has $M$ levels is an $M$-component mixture of multinomials. The parameter vector for each component multinomial is based on the mixing probability vector $\boldsymbol{\pi}$ and an additional factor $\mu$. As is true for the binomial cluster model, the number of components is fixed and the mixing probabilities in $\boldsymbol{\pi}$ appear in the moments for the mixture components. For more information, see Morel and Nagaraj (1993) and Example 43.4 in this chapter.

The mean and variance are $\mathrm{E}[\mathbf{Y}] = n\boldsymbol{\pi}$ and $\mathrm{Var}[\mathbf{Y}] = n(\mathrm{Diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T)\{1+\mu^2(n-1)\}$, respectively.

**Negative Binomial** $(\mu, \phi)$

$$l(\mu_i, \phi; y_i, w_i) = y_i \log\left\{\frac{\phi\mu_i}{w_i}\right\} - (y_i + w_i/\phi)\log\left\{1 + \frac{\phi\mu_i}{w_i}\right\} + \log\left\{\frac{\Gamma(y_i + w_i/\phi)}{\Gamma(w_i/\phi)\Gamma(y_i + 1)}\right\}$$

The variance is $\mathrm{Var}[Y] = \mu + \phi\mu^2$, $\phi > 0$.

For a given $\phi$, the negative binomial distribution is a member of the exponential family. The parameter $\phi$ is related to the scale of the data because it is part of the variance function. However, it cannot be factored from the variance, as is the case with the $\phi$ parameter in many other distributions.

**Normal** $(\mu, \phi)$

$$l(\mu_i, \phi; y_i, w_i) = -\frac{1}{2}\left[\frac{w_i(y_i - \mu_i)^2}{\phi} + \log\left\{\frac{\phi}{w_i}\right\} + \log\{2\pi\}\right]$$

The mean and variance are $\mathrm{E}[Y] = \mu$ and $\mathrm{Var}[Y] = \phi$, respectively, $\phi > 0$

**Poisson** $(\mu)$

$$l(\mu_i; y_i, w_i) = w_i(y_i \log\{\mu_i\} - \mu_i - \log\{\Gamma(y_i + 1)\})$$

The mean and variance are $\mathrm{E}[Y] = \mu$ and $\mathrm{Var}[Y] = \mu$.

**(Shifted) T** $(\nu; \mu, \phi)$

$$\begin{aligned} z_i &= -0.5\log\{\phi/w_i\} + \log\{\Gamma(0.5(\nu + 1)\} \\ &\quad - \log\{\Gamma(0.5\nu)\} - 0.5 \times \log\{\pi\nu\} \\ l(\mu_i, \phi; y_i, w_i) &= -\left(\frac{\nu + 1}{2}\right)\log\left\{1 + \frac{w_i}{\nu}\frac{(y_i - \mu_i)^2}{\phi}\right\} + z_i \end{aligned}$$

$\nu > 0, \phi > 0$. In this parameterization, $\mathrm{E}[Y] = \mu$ when $\nu > 1$ and $\mathrm{Var}[Y] = \phi\nu/(\nu - 2)$ when $\nu > 2$. Note that this form of the $t$ distribution is not a noncentral distribution, but that of a shifted central $t$ random variable.

**Truncated Exponential** $(\mu; a, b)$

$$l(\mu_i; a, b, y_i, w_i) = w_i \log\left\{\frac{w_i y_i}{\mu_i}\right\} - \frac{w_i y_i}{\mu_i} - \log\{y_i\Gamma(w_i)\} - \log\left[\frac{\gamma\left(w_i, \frac{w_i b}{\mu_i}\right)}{\Gamma(w_i)} - \frac{\gamma\left(w_i, \frac{w_i a}{\mu_i}\right)}{\Gamma(w_i)}\right]$$

where

$$\gamma(c_1, c_2) = \int_0^{c_2} t^{c_1-1} \exp(-t)\mathrm{d}t$$

is the lower incomplete gamma function. The mean and variance are

$$\mathrm{E}[Y] = \frac{(a+\mu_i)\exp(-a/\mu_i) - (b+\mu_i)\exp(-b/\mu_i)}{\exp(-a/\mu_i) - \exp(-b/\mu_i)}$$

$$\mathrm{Var}[Y] = \frac{(a^2 + 2a\mu_i + 2\mu_i^2)\exp(-a/\mu_i) - (b^2 + 2b\mu_i + 2\mu_i^2)\exp(-b/\mu_i)}{\exp(-a/\mu_i) - \exp(-b/\mu_i)} - (\mathrm{E}[Y])^2$$

**Truncated Lognormal** $(\mu, \phi; a, b)$

$$z_i = \log\{y_i\} - \mu_i$$

$$l(\mu_i, \phi; a, b, y_i, w_i) = -\frac{1}{2}\left(2\log\{y_i\} + \log\left\{\frac{\phi}{w_i}\right\} + \log\{2\pi\} + \frac{w_i z_i^2}{\phi}\right) - \log\left\{\Phi\left[\sqrt{w_i/\phi}(\log b - \mu_i)\right] - \Phi\left[\sqrt{w_i/\phi}(\log a - \mu_i)\right]\right\}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The mean and variance are

$$\mathrm{E}[Y] = \exp(\mu_i + 0.5\phi)\frac{\Phi\left(\sqrt{\phi} - \frac{\log a - \mu_i}{\sqrt{\phi}}\right) - \Phi\left(\sqrt{\phi} - \frac{\log b - \mu_i}{\sqrt{\phi}}\right)}{\Phi\left(\frac{\log b - \mu_i}{\sqrt{\phi}}\right) - \Phi\left(\frac{\log a - \mu_i}{\sqrt{\phi}}\right)}$$

$$\mathrm{Var}[Y] = \exp(2\mu_i + 2\phi)\frac{\Phi\left(2\sqrt{\phi} - \frac{\log a - \mu_i}{\sqrt{\phi}}\right) - \Phi\left(2\sqrt{\phi} - \frac{\log b - \mu_i}{\sqrt{\phi}}\right)}{\Phi\left(\frac{\log b - \mu_i}{\sqrt{\phi}}\right) - \Phi\left(\frac{\log a - \mu_i}{\sqrt{\phi}}\right)} - (\mathrm{E}[Y])^2$$

**Truncated Negative Binomial** $(\mu, \phi)$

$$l(\mu_i, \phi; y_i, w_i) = y_i \log\left\{\frac{\phi\mu_i}{w_i}\right\} - (y_i + w_i/\phi)\log\left\{1 + \frac{\phi\mu_i}{w_i}\right\} + \log\left\{\frac{\Gamma(y_i + w_i/\phi)}{\Gamma(w_i/\phi)\Gamma(y_i+1)}\right\} - \log\left\{1 - \left(\frac{\phi\mu_i}{w_i} + 1\right)^{-w_i/\phi}\right\}$$

The mean and variance are

$$\mathrm{E}[Y] = \mu_i\left\{1 - (\phi\mu_i + 1)^{-1/\phi}\right\}^{-1}$$

$$\mathrm{Var}[Y] = (1 + \phi\mu_i + \mu_i)\mathrm{E}[Y] - (\mathrm{E}[Y])^2$$

**Truncated Normal** $(\mu, \phi; a, b)$

$$l(\mu_i, \phi; a, b, y_i, w_i) = -\frac{1}{2}\left[\frac{w_i(y_i - \mu_i)^2}{\phi} + \log\left\{\frac{\phi}{w_i}\right\} + \log\{2\pi\}\right] - \log\left\{\Phi\left[\sqrt{w_i/\phi}(b - \mu_i)\right] - \Phi\left[\sqrt{w_i/\phi}(a - \mu_i)\right]\right\}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The mean and variance are

$$\mathrm{E}[Y] = \mu_i + \sqrt{\phi}\frac{\mathrm{phi}\left(\frac{a-\mu_i}{\sqrt{\phi}}\right) - \mathrm{phi}\left(\frac{b-\mu_i}{\sqrt{\phi}}\right)}{\Phi\left(\frac{b-\mu_i}{\sqrt{\phi}}\right) - \Phi\left(\frac{a-\mu_i}{\sqrt{\phi}}\right)}$$

$$\mathrm{Var}[Y] = \phi\left[1 + \frac{\frac{a-\mu_i}{\sqrt{\phi}}\mathrm{phi}\left(\frac{a-\mu_i}{\sqrt{\phi}}\right) - \frac{b-\mu_i}{\sqrt{\phi}}\mathrm{phi}\left(\frac{b-\mu_i}{\sqrt{\phi}}\right)}{\Phi\left(\frac{b-\mu_i}{\sqrt{\phi}}\right) - \Phi\left(\frac{a-\mu_i}{\sqrt{\phi}}\right)} - \left\{\frac{\mathrm{phi}\left(\frac{a-\mu_i}{\sqrt{\phi}}\right) - \mathrm{phi}\left(\frac{b-\mu_i}{\sqrt{\phi}}\right)}{\Phi\left(\frac{b-\mu_i}{\sqrt{\phi}}\right) - \Phi\left(\frac{a-\mu_i}{\sqrt{\phi}}\right)}\right\}^2\right]$$

where $\mathrm{phi}(\cdot)$ is the probability density function of the standard normal distribution.

**Truncated Poisson** $(\mu)$

$$l(\mu_i; y_i, w_i) = w_i(y_i \log\{\mu_i\} - \log\{\exp(\mu_i) - 1\} - \log\{\Gamma(y_i + 1)\})$$

The mean and variance are

$$\mathrm{E}[Y] = \frac{\mu_i}{1 - \exp(-\mu_i)}$$

$$\mathrm{Var}[Y] = \frac{\mu_i\,[1 - \exp(-\mu_i) - \mu_i \exp(-\mu_i)]}{[1 - \exp(-\mu_i)]^2}$$

**Uniform** $(a, b)$

$$l(\mu_i; y_i, w_i) = -\log\{b - a\}$$

The mean and variance are $\mathrm{E}[Y] = 0.5(a + b)$ and $\mathrm{Var}[Y] = (b - a)^2/12$.

**Weibull** $(\mu, \phi)$

$$l(\mu_i, \phi; y_i) = -\frac{\phi - 1}{\phi}\log\left\{\frac{y_i}{\mu_i}\right\} - \log\{\mu_i\phi\} - \exp\left\{\log\left\{\frac{y_i}{\mu_i}\right\}/\phi\right\}$$

In this particular parameterization of the two-parameter Weibull distribution, the mean and variance of the random variable $Y$ are $\mathrm{E}[Y] = \mu\Gamma(1 + \phi)$ and $\mathrm{Var}[Y] = \mu^2\left\{\Gamma(1 + 2\phi) - \Gamma^2(1 + \phi)\right\}$.

# Bayesian Analysis

## Conjugate Sampling

The FMM procedure uses Bayesian analysis via a conjugate Gibbs sampler if the model belongs to a small class of mixture models for which a conjugate sampler is available. See the section "Gibbs Sampler" on page 139 in Chapter 7, "Introduction to Bayesian Analysis Procedures," for a general discussion of Gibbs sampling. Table 43.8 summarizes the models for which conjugate and Metropolis-Hastings samplers are available.

**Table 43.8** Availability of Conjugate and Metropolis Samplers in the FMM Procedure

| Effects (exclusive of intercept) | Distributions | Available Samplers |
|---|---|---|
| No | Normal or T | Conjugate or Metropolis-Hastings |
| Yes | Normal or T | Conjugate or Metropolis-Hastings |
| No | Binomial, binary, Poisson, exponential | Conjugate or Metropolis-Hastings |
| Yes | Binomial, binary, Poisson, exponential | Metropolis-Hastings only |

The conjugate sampler enjoys greater efficiency than the Metropolis-Hastings sampler and has the advantage of sampling in terms of the natural parameters of the distribution.

You can always switch to the Metropolis-Hastings sampling algorithm in any model by adding the METROPOLIS option in the BAYES statement.

## Metropolis-Hastings Algorithm

If Metropolis-Hastings is the only sampler available for the specified model (see Table 43.8) or if the METROPOLIS option is specified in the BAYES statement, PROC FMM uses the Metropolis-Hastings approach of Gamerman (1997). See the section "Metropolis and Metropolis-Hastings Algorithms" on page 138 in Chapter 7, "Introduction to Bayesian Analysis Procedures," for a general discussion of the Metropolis-Hastings algorithm.

The Gamerman (1997) algorithm derives a specific density that is used to generate proposals for the component-specific parameters $\boldsymbol{\beta}_j$. The form of this proposal density is multivariate normal, with mean $\mathbf{m}_j$ and covariance matrix $\mathbf{C}_j$ derived as follows.

Suppose $\boldsymbol{\beta}_j$ is the vector of model coefficients in the $j$th component and suppose that $\boldsymbol{\beta}_j$ has prior distribution $N(\mathbf{a}, \mathbf{R})$. Consider a generalized linear model (GLM) with link function $g(\mu) = \eta = \mathbf{x}'\boldsymbol{\beta}$ and variance function $a(\mu)$. The pseudo-response and weight in the GLM for a weighted least squares step are

$$y^* = \eta + (y - \mu)/\frac{\partial \mu}{\partial \eta}$$

$$w = \frac{\partial \mu}{\partial \eta}/a(\mu)$$

If the model contains offsets or FREQ or WEIGHT statements, or if a *trials* variable is involved, suitable adjustments are made to these quantities.

In each component, $j = 1, \ldots, k$, form an adjusted cross-product matrix with a "pseudo" border

$$\begin{bmatrix} \mathbf{X}_j'\mathbf{W}_j\mathbf{X}_j + \mathbf{R}^{-1} & \mathbf{X}_j'\mathbf{W}_j\mathbf{y}_j^* + \mathbf{R}^{-1}\mathbf{a} \\ \mathbf{y}_j^{*\prime}\mathbf{W}_j\mathbf{X}_j + \mathbf{a}'\mathbf{R}^{-1} & c \end{bmatrix}$$

where $\mathbf{W}_j$ is a diagonal matrix formed from the pseudo-weights $w$, $\mathbf{y}^*$ is a vector of pseudo-responses, and $c$ is arbitrary. This is basically a system of normal equations with ridging, and the degree of ridging is governed by the precision and mean of the normal prior distribution of the coefficients. Sweeping on the leading partition leads to

$$\begin{aligned} \mathbf{C}_j &= \left(\mathbf{X}_j'\mathbf{W}_j\mathbf{X}_j + \mathbf{R}^{-1}\right)^- \\ \mathbf{m}_j &= \mathbf{C}_j\left(\mathbf{X}_j'\mathbf{W}_j\mathbf{y}_j^* + \mathbf{R}^{-1}\mathbf{a}\right) \end{aligned}$$

where the generalized inverse is a reflexive, $g_2$-inverse (see the section "Linear Model Theory" on page 55 in Chapter 3, "Introduction to Statistical Modeling with SAS/STAT Software," for details).

PROC FMM then generates a proposed parameter vector from the resulting multivariate normal distribution, and then accepts or rejects this proposal according to the appropriate Metropolis-Hastings thresholds.

## Latent Variables via Data Augmentation

In order to fit finite Bayesian mixture models, the FMM procedure treats the mixture model as a missing data problem and introduces an assignment variable **S** as in Dempster, Laird, and Rubin (1977). Because **S** is not observable, it is frequently referred to as a latent variable. The unobservable variable **S** assigns an observation to a component in the mixture model. The number of states, $k$, might be unknown, but it is known to be finite. Conditioning on the latent variable **S**, the component memberships of each observation is assumed to be known, and Bayesian estimation is straightforward for each component in the finite mixture model. That is, conditional on $S = j$, the distribution of the response is now assumed to be $f(y; \alpha_j, \beta_j | S = j)$. In other words, each distinct state of the random variable **S** leads to a distinct set of parameters. The parameters in each component individually are then updated using a conjugate Gibbs sampler (where available) or a Metropolis-Hastings sampling algorithm.

The FMM procedure assumes that the random variable **S** has a discrete multinomial distribution with probability $\pi_j$ of belonging to a component $j$; it can occupy one of $k$ states. The distribution for the latent variable **S** is

$$f(S_i = j | \pi_1, \ldots, \pi_k) = \text{multinomial}(1, \pi_1, \ldots, \pi_k)$$

where $f(\cdot|\cdot)$ denotes a conditional probability density. The parameters in the density $\pi_j$ denote the probability that $S$ takes on state $j$.

The FMM procedure assumes a conjugate Dirichlet prior distribution on the mixture proportions $\pi_j$ written as:

$$p(\boldsymbol{\pi}) \quad = \quad \text{Dirichlet}(a_1, \ldots, a_k)$$

where $p(\cdot)$ indicates a prior distribution.

Using Bayes' theorem, the likelihood function and prior distributions determine a conditionally conjugate posterior distribution of **S** and $\boldsymbol{\pi}$ from the multinomial distribution and Dirichlet distribution, respectively.

## Prior Distributions

The following list displays the parameterization of prior distributions for situations in which the FMM procedure uses a conjugate sampler in mixture models without model effects and certain basic distributions (binary, binomial, exponential, Poisson, normal, and *t*). You specify the parameters *a* and *b* in the formulas below in the MUPRIORPARMS= and PHIPRIORPARMS= options in the BAYES statement in these models.

**Beta**$(a, b)$

$$f(y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, y^{a-1}\,(1-y)^{b-1}$$

where $a > 0$, $b > 0$. In this parameterization, the mean and variance of the distribution are $\mu = a/(a+b)$ and $\mu(1-\mu)/(a+b+1)$, respectively. The beta distribution is the prior distribution for the success probability in binary and binomial distributions when conjugate sampling is used.

**Dirichlet**$(a_1, \ldots, a_k)$

$$f(\mathbf{y}) = \frac{\Gamma\left(\sum_{i=1}^{k} a_i\right)}{\prod_{i=1}^{k}\Gamma(a_i)}\, y_1^{a_1-1}\,\ldots\, y_k^{a_k-1}$$

where $\sum_{i=1}^{k} y_i = 1$ and the parameters $a_i > 0$. If any $a_i$ were zero, an improper density would result. The Dirichlet density is the prior distribution for the mixture probabilities. You can affect the choice of the $a_i$ through the MIXPRIORPARMS option in the BAYES statement. If *k*=2, the Dirichlet is the same as the beta$(a, b)$ distribution.

**Gamma**$(a, b)$

$$f(y) = \frac{b^a}{\Gamma(a)}\, y^{a-1}\,\exp\{-by\}$$

where $a > 0$, $b > 0$. In this parameterization, the mean and variance of the distribution are $\mu = a/b$ and $\mu/b$, respectively. The gamma distribution is the prior distribution for the mean parameter of the Poisson distribution when conjugate sampling is used.

**Inverse gamma**$(a, b)$

$$f(y) = \frac{b^a}{\Gamma(a)}\, y^{-a-1}\,\exp\{-b/y\}$$

where $a > 0$, $b > 0$. In this parameterization, the mean and variance of the distribution are $\mu = b/(a-1)$ if $a > 1$ and $\mu^2/(a-2)$ if $a > 2$, respectively. The inverse gamma distribution is the prior distribution for the mean parameter of the exponential distribution when conjugate sampling is used. It is also the prior distribution for the scale parameter $\phi$ in all models.

**Multinomial**$(1, \pi_1, \ldots, \pi_k)$

$$f(\mathbf{y}) = \frac{1}{y_1!\cdots y_k!}\pi_1^{y_1}\,\cdots\,\pi_k^{y_k}$$

where $\sum_{j=1}^{k} y_j = n$, $y_j \geq 0$, $\sum_{j=1}^{k}\pi_j = 1$, and $n$ is the number of observations included in the analysis. The multinomial density is the prior distribution for the mixture proportions. The mean and variance of $Y_j$ are $\mu_j = \pi_j$ and $\mu_j(1-\mu_j)$, respectively.

**Normal**$(a, b)$

$$f(y) = \frac{a}{\sqrt{2\pi b}} \exp\left\{-\frac{1}{2}\frac{(y-a)^2}{b}\right\}$$

where $b > 0$. The mean and variance of the distribution are $\mu = a$ and $b$, respectively. The normal distribution is the prior distribution for the mean parameter of the normal and $t$ distribution when conjugate sampling is used.

When a MODEL statement contains effects or if you specify the METROPOLIS option in the BAYES statement, the prior distribution for the regression parameters is multivariate normal, and you can specify the means and variances of the parameters in the BETAPRIORPARMS= option in the BAYES statement.

## Parameterization of Model Effects

PROC FMM constructs a finite mixture model according to the specifications in the CLASS, MODEL, and PROBMODEL statements. Each effect in the MODEL statement generates one or more columns in the matrix **X** for that model. The same **X** matrix applies to all components that are associated with the MODEL statement. Each effect in the PROBMODEL statement generates one or more columns in the matrix **Z** from which the linear predictors in the model for the mixture probability models is formed. The same **Z** matrix applies to all components.

The formation of effects from continuous and classification variables in the FMM procedure follows the same general rules and techniques as for other linear modeling procedures. See the section "GLM Parameterization of Classification Variables and Effects" on page 393 in Chapter 19, "Shared Concepts and Topics."

## Random Starting Values

The iterative optimization steps for maximum likelihood estimation of mixture models might not always find estimates that achieve a global maximum or even a best local maximum in likelihood. One way to alleviate this problem is to generate multiple sets of starting values for the optimization and find the best solution over those sets. PROC FMM supports this approach by automating a two-stage method that generates multiple sets of random starting values in maximum likelihood estimation. You can use this two-stage method by specifying the RANDSTART statement.

In the first stage, the procedure uses its default starting values as the first set of starting values. If you specify starting values by using the PARAMETERS option in either the MODEL statement or the PROBMODEL statement, PROC FMM uses those values as the first set of starting values instead. Then, the procedure randomly draws multiple sets of starting values around the values of the first random set; the variability is controlled by a standard deviation parameter, which you can adjust by using the STD= option. You can use the NFIRST= option to control the number of sets that the procedure generates in the first stage. The procedure repeatedly optimizes the model, using each set of random starting values in turn to start the optimization. It uses the convergence criterion that you specify in the RANDGCONV= option to terminate each fit in the first stage.

In the second stage, PROC FMM identifies the sets that have the best log likelihood at the end of the first stage and continues the fit for each of those sets. You can use the NSECOND= option to control the number

of sets to carry forward from the first stage. If not enough of the first-stage sets converge to supply the second stage, the procedure generates additional sets and optimizes by using those sets. The procedure will not generate more than the maximum number of sets that you specify in the MAXFIRST= option. At the end of the second stage, the procedure identifies the model that has the best fit, and it uses the corresponding converged estimates as the maximum likelihood estimates for the model.

The "Random Start Information" table displays the values that control the two-stage fitting process.

The "Random Start Convergence" table displays the random starting sets, the final objective function from optimization, and an indicator of convergence. This table contains identifiers for the starting sets and the stages of fitting. These identifiers are also displayed in the "Iteration History" table.

The "Parameter Mapping" table provides a mapping between the parameters in the "Random Start Convergence" table and their roles in the associated mixture component.

## Default Output

The following sections describe the output that PROC FMM produces by default. The output is organized into various tables, which are discussed in the order of appearance for maximum likelihood and Bayesian estimation, respectively.

### Model Information

The "Model Information" table displays basic information about the model, such as the response variable, frequency variable, link function, and the model category that the FMM procedure determined based on your input and options. The "Model Information" table is one of a few tables that are produced irrespective of estimation technique. Most other tables are specific to Bayes or maximum likelihood estimation.

If the analysis depends on generated random numbers, the "Model Information" table also displays the random number seed used to initialize the random number generators. If you repeat the analysis and pass this seed value in the SEED= option in the PROC FMM statement, an identical stream of random numbers results.

### Class Level Information

The "Class Level Information" table lists the levels of every variable specified in the CLASS statement. You should check this information to make sure that the data are correct. You can adjust the order of the CLASS variable levels with the ORDER= option in the PROC FMM statement. You can suppress the "Class Level Information" table completely or partially with the NOCLPRINT option in the PROC FMM statement.

### Number of Observations

The "Number of Observations" table displays the number of observations read from the input data set and the number of observations used in the analysis. If you specify a FREQ statement, the table also displays the sum of frequencies read and used. If the *events/trials* syntax is used for the response, the table also displays the number of events and trials used in the analysis.

Note that the number of observations "used" in the analysis is not unambiguous in a mixture model. An observation that is "unusable" for one component distribution (because the response value is outside of the support of the distribution) might still be usable in the mixture model when the response value is in the

support of another component distribution. You can affect the way in which PROC FMM handles exclusion of observations due to support violations with the EXCLUSION= option in the PROC FMM statement.

## Response Profile

For binary data, the "Response Profile" table displays the ordered value from which the FMM procedure determines the probability being modeled as an event for binary data. For each response category level, the frequency used in the analysis is reported.

## Default Output for Maximum Likelihood

### *Optimization Information*

The "Optimization Information" table displays basic information about the optimization setup to determine the maximum likelihood estimates, such as the optimization technique, the parameters that participate in the optimization, and the number of threads used for the calculations. This table is not produced during model selection—that is, if the KMAX= option is specified in the MODEL statement.

### *Iteration History*

The "Iteration History" table displays for each iteration of the optimization the number of function evaluations (including gradient and Hessian evaluations), the value of the objective function, the change in the objective function from the previous iteration, and the absolute value of the largest (projected) gradient element. The objective function used in the optimization in the FMM procedure is the negative of the mixture log likelihood; consequently, PROC FMM performs a minimization. This table is not produced if you specify the KMAX= option in the MODEL statement. If you wish to see the "Iteration History" table in this setting, you must also specify the FITDETAILS option in the PROC FMM statement.

### *Convergence Status*

The convergence status table is a small ODS table that follows the "Iteration History" table in the default output. In the listing, it appears as a message that identifies whether the optimization succeeded and which convergence criterion was met. If the optimization fails, the message indicates the reason for the failure. If you save the "Convergence Status" table to an output data set, a numeric Status variable is added that allows you to assess convergence programmatically. The values of the Status variable encode the following:

0 Convergence was achieved or an optimization was not performed (because of TECHNIQUE=NONE).

1 The objective function could not be improved.

2 Convergence was not achieved because of a user interrupt or because a limit was exceeded, such as the maximum number of iterations or the maximum number of function evaluations. To modify these limits, see the MAXITER=, MAXFUNC=, and MAXTIME= options in the PROC FMM statement.

3 Optimization failed to converge because function or derivative evaluations failed at the starting values or during the iterations or because a feasible point that satisfies the parameter constraints could not be found in the parameter space.

### Fit Statistics

The "Fit Statistics" table displays a variety of fit measures based on the mixture log likelihood in addition to the Pearson statistic. All statistics are presented in "smaller is better" form. If you are fitting a single-component normal, gamma, or inverse Gaussian model, the table also contains the unscaled Pearson statistic. If you are fitting a mixture model or the model has been fitted under restrictions, the table also contains the number of effective components and the number of effective parameters.

The calculation of the information criteria uses the following formulas, where $p$ denotes the number of effective parameters, $n$ denotes the number of observations used (or the sum of the frequencies used if a FREQ statement is present), and $l$ is the log likelihood of the mixture evaluated at the converged estimates:

$$\mathrm{AIC} = -2l + 2p$$

$$\mathrm{AICC} = \begin{cases} -2l + 2pn/(n-p-1) & n > p+2 \\ -2l + 2p(p+2) & \text{otherwise} \end{cases}$$

$$\mathrm{BIC} = -2l + p\log(n)$$

The Pearson statistic is computed simply as

$$\text{Pearson statistic} = \sum_{i=1}^{n} f_i \frac{(y_i - \widehat{\mu}_i)^2}{\widehat{\mathrm{Var}}[Y_i]}$$

where $n$ denotes the number of observations used in the analysis, $f_i$ is the frequency associated with the $i$th observation (or 1 if no frequency is specified), $\mu_i$ is the mean of the mixture, and the denominator is the variance of the $i$th observation in the mixture. Note that the mean and variance in this expression are not those of the component distributions, but the mean and variance of the mixture:

$$\mu_i = \mathrm{E}[Y_i] = \sum_{j=1}^{k} \pi_{ij}\mu_{ij}$$

$$\mathrm{Var}[Y_i] = -\mu_i^2 + \sum_{j=1}^{k} \pi_{ij}\left(\sigma_{ij}^2 + \mu_{ij}^2\right)$$

where $\mu_{ij}$ and $\sigma_{ij}^2$ are the mean and variance, respectively, for observation $i$ in the $j$th component distribution and $\pi_{ij}$ is the mixing probability for observation $i$ in component $j$.

The unscaled Pearson statistic is computed with the same expression as the Pearson statistic with $n$, $f_i$, and $\mu_i$ as previously defined, but the scale parameter $\phi$ is set to 1 in the $\widehat{\mathrm{Var}}[Y_i]$ expression.

The number of effective components and the number of effective parameters are determined by examining the converged solution for the parameters that are associated with model effects and the mixing probabilities. For example, if a component has an estimated mixing probability of zero, the values of its parameter estimates are immaterial. You might argue that all parameters should be counted towards the penalty in the information criteria. But a component with zero mixing probability in a $k$-component model effectively reduces the model to a $(k-1)$-component model. A situation of an overfit model, for which a parameter penalty needs to be taken when calculating the information criteria, is a different situation; here the mixing probability might be small, possibly close to zero.

### *Parameter Estimates*

The parameter estimates, their estimated (asymptotic) standard errors, and $p$-values for the hypothesis that the parameter is zero are presented in the "Parameter Estimates" table. A separate table is produced for each MODEL statement, and the components that are associated with a MODEL statement are identified with an overall component count variable that counts across MODEL statements. If you assign a label to a model with the LABEL= option in the MODEL statement, the label appears in the title of the "Parameter Estimates" table. Otherwise, the internal label generated by the FMM procedure is used.

If the MODEL statement does not contain effects and the link function is not the identity, the inversely linked estimate is also displayed in the table. For many distributions, the inverse linked estimate is the estimated mean on the data scale. For example, in a binomial or binary model, it represents the estimated probability of an event. For some distributions (for example, the Weibull distribution), the inverse linked estimate is not the component distribution mean.

If you request confidence intervals with the CL or ALPHA= option in the MODEL statement, confidence limits are produced for the estimate on the linear scale. If the inverse linked estimate is displayed, confidence intervals for that estimate are also produced by inversely linking the confidence bounds on the linear scale.

### *Mixing Probabilities*

If you fit a model with more than one component, the table of mixing probabilities is produced. If there are no effects in the PROBMODEL statement or if there is no PROBMODEL statement, the parameters are reported on the linear scale and as mixing probabilities. If model effects are present, only the linear parameters (on the scale of the logit, generalized logit, probit, and so on) are displayed.

## Default Output for Bayesian Estimation

### *Bayes Information*

This table provides basic information about the sampling algorithm. The FMM procedure uses either a conjugate sampler or a Metropolis-Hastings sampling algorithm based on Gamerman (1997). The table reveals, for example, how many model parameters are sampled, how many parameters associated with mixing probabilities are sampled, and how many threads are used to perform multithreaded analysis.

### *Prior Distributions*

The "Prior Distributions" table lists for each sampled parameter the prior distribution and its parameters. The mean and variance (if they exist) for those values of the parameters are also displayed, along with the initial value for the parameter in the Markov chain. The Component column in this table identifies the mixture component to which a particular parameter belongs. You can control how the FMM procedure determines initial values with the INITIAL= option in the BAYES statement.

### *Bayesian Fit Statistics*

The "Bayesian Fit Statistics" table shows three measures based on the posterior sample. The "Average -2 Log Likelihood" is derived from the average mixture log-likelihood for the data, where the average is taken over the posterior sample. The deviance information criterion (DIC) is a Bayesian measure of model fit and the effective number of parameters ($p_D$) is a penalization term used in the computation of the DIC. For more information about the DIC and $p_D$, see the section "Summary Statistics" on page 158 in Chapter 7, "Introduction to Bayesian Analysis Procedures."

### Posterior Summaries

The arithmetic mean, standard deviation, and percentiles of the posterior distribution of the parameter estimates are displayed in the "Posterior Summaries" table. By default, the FMM procedure computes the 25th, 50th (median), and 75th percentiles of the sampling distribution. You can modify the percentiles through suboptions of the STATISTICS option in the BAYES statement. If a parameter corresponds to a singularity in the design and was removed from sampling for that purpose, it is also displayed in the table of posterior summaries (and in other tables that relate to output from the BAYES statement). The posterior sample size for such a parameter is shown as N = 0.

### Posterior Intervals

The table of "Posterior Intervals" displays equal-tail intervals and intervals of highest posterior density for each parameter. By default, intervals are computed for an $\alpha$-level of 0.05, which corresponds to 95% intervals. You can modify this confidence level by providing one or more $\alpha$ values in the ALPHA= suboption of the STATISTICS option in the BAYES statement. The computation of these intervals is detailed in section "Summary Statistics" on page 158 in Chapter 7, "Introduction to Bayesian Analysis Procedures."

### Posterior Autocorrelations

Autocorrelations for the posterior estimates are computed by default for autocorrelation lags 1, 5, 10, and 50, provided that a sufficient number of posterior samples is available. See the section "Assessing Markov Chain Convergence" on page 144 in Chapter 7, "Introduction to Bayesian Analysis Procedures," for the computation of posterior autocorrelations and their utility in diagnosing convergence of Markov chains. You can modify the list of lags for which posterior autocorrelations are calculated with the AUTOCORR suboption of the DIAGNOSTICS= option in the BAYES statement.

### Effective Sample Sizes

The "Effective Sample Sizes" table displays the effective sample size (Kass et al. 1998) for each of the parameters in the model. The effective sample size (ESS) for an MCMC sample is another assessment of the mixing of the Markov chain. For more information about the ESS, see the section "Assessing Markov Chain Convergence" on page 144 in Chapter 7, "Introduction to Bayesian Analysis Procedures."

## ODS Table Names

Each table created by PROC FMM has a name associated with it, and you must use this name to reference the table when you use ODS statements. These names are listed in Table 43.9.

**Table 43.9** ODS Tables Produced by PROC FMM

| Table Name | Description | Required Statement / Option |
|---|---|---|
| AutoCorr | Autocorrelation among posterior estimates | BAYES |
| BayesInfo | Basic information about Bayesian estimation | BAYES |
| ClassLevels | Level information from the CLASS statement | CLASS |

**Table 43.9** *continued*

| Table Name | Description | Required Statement / Option |
|---|---|---|
| CompDescription | Component description in models with varying number of components | KMAX= in MODEL with ML estimation |
| CompEvaluation | Comparison of mixture models with varying number of components | KMAX= in MODEL with ML estimation |
| CompInfo | Component information | COMPONENTINFO option in PROC FMM statement |
| ConvergenceStatus | Status of optimization at conclusion of optimization | Default output |
| Constraints | Linear equality and inequality constraints | RESTRICT statement or EQUATE=EFFECTS option in MODEL statement |
| Corr | Asymptotic correlation matrix of parameter estimates (ML) or empirical correlation matrix of the Bayesian posterior estimates | CORR option in PROC FMM statement |
| Cov | Asymptotic covariance matrix of parameter estimates (ML) or empirical covariance matrix of the Bayesian posterior estimates | COV option in PROC FMM statement |
| CovI | Inverse of the covariance matrix of the parameter estimates | COVI option in PROC FMM statement |
| ESS | Effective sample sizes | BAYES |
| FitStatistics | Fit statistics | Default output |
| Geweke | Geweke diagnostics (Geweke 1992) for Markov chains | DIAG=GEWEKE option in BAYES statement |
| Heidelberger | Heidelberger and Welch diagnostics (Heidelberger and Welch 1981, 1983) for Markov chains | DIAG=HEIDELBERGER option in BAYES statement |
| Hessian | Hessian matrix from the maximum likelihood optimization, evaluated at the converged estimates | HESSIAN |
| IterHistory | Optimizer iteration history | Default output for maximum likelihood estimation; included in Bayesian estimation when BAYES statement includes the INITIAL=MLE option and PROC FMM statement includes the FITDETAILS option |

**Table 43.9** *continued*

| Table Name | Description | Required Statement / Option |
|---|---|---|
| MCSE | Monte Carlo standard errors | DIAG=MCERROR in BAYES statement |
| MixingProbs | Solutions for the parameter estimates associated with effects in PROBMODEL statements | Default output for ML estimation if number of components is greater than 1 |
| ModelInfo | Model information | Default output |
| NObs | Number of observations read and used, number of trials and events | Default output |
| OptInfo | Optimization information | Default output for ML estimation |
| ParameterEstimates | Solutions for the parameter estimates associated with effects in MODEL statements | Default output for ML estimation |
| ParameterMap | Mapping of parameter names to OUTPOST= data set or RandomStartConvergence table | OUTPOST= option in BAYES statement or RANDSTART statement |
| PostIntervals | Equal-tail and highest posterior density intervals for posterior estimates | BAYES |
| PostSummaries | Summary statistics for posterior estimates | BAYES |
| PriorInfo | Prior distributions and initial value of Markov chain | BAYES |
| Raftery | Raftery and Lewis diagnostics (Raftery and Lewis 1992, 1995) for Markov chains | DIAG=RAFTERY option in BAYES statement |
| RandomStartConvergence | Starting values and convergence indicator for each random starting set | RANDSTART statement |
| RandomStartInfo | Information about random starting values | RANDSTART statement |
| ResponseProfile | Response categories and category modeled | Default output in models with binary response |

## ODS Graphics

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, "Statistical Graphics Using ODS."

Before you create graphs, ODS Graphics must be enabled (for example, by specifying the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section "Enabling and Disabling ODS Graphics" on page 623 in Chapter 21, "Statistical Graphics Using ODS."

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section "A Primer on ODS Statistical Graphics" on page 622 in Chapter 21, "Statistical Graphics Using ODS."

You can reference every graph produced through ODS Graphics with a name. The names of the graphs that PROC FMM generates are listed in Table 43.10, along with the required statements and options.

**Table 43.10** Graphs Produced by PROC FMM

| ODS Graph Name | Plot Description | Option |
|---|---|---|
| TADPanel | Panel of diagnostic graphics to assess convergence of Markov chains | BAYES |
| DensityPlot | Histogram and density with component distributions | Default plot for homogeneous mixtures |
| CriterionPanel | Panel of plots showing progression of model fit criteria for mixtures with different numbers of components | KMIN= and KMAX= options in MODEL statement |

# Examples: FMM Procedure

## Example 43.1: Modeling Mixing Probabilities: All Mice Are Equal, but Some Mice Are More Equal Than Others

This example demonstrates how you can model the means and mixture proportions separately in a binomial cluster model. It also compares the binomial cluster model to the beta-binomial model.

In a typical teratological experiment, the offspring of animals that were exposed to a toxin during pregnancy are studied for malformation. If you count the number of malformed offspring in a litter of size $n$, then this count is typically not binomially distributed. The responses of the offspring from the same litter are not independent; hence their sum does not constitute a binomial random variable. Relative to a binomial model, data from teratological experiments exhibit overdispersion because ignoring positive correlation among the

responses tends to overstate the precision of the parameter estimates. Overdispersion mechanisms are briefly discussed in the section "Overdispersion" on page 2947.

In this application, the focus is on mixtures and models that involve a mixing mechanism. The mixing approach (Williams 1975; Haseman and Kupper 1979) supposes that the binomial success probability is a random variable that follows a $\text{beta}(\alpha, \beta)$ distribution:

$$
\begin{aligned}
Y|\mu &\sim \text{Binomial}(n, \mu) \\
\mu &\sim \text{Beta}(\alpha, \beta) \\
Y &\sim \text{Beta-binomial}(n, \mu, \phi) \\
\text{E}[Y] &= n\pi \\
\text{Var}[Y] &= n\pi(1-\pi)\left\{1 + \mu^2(n-1)\right\}
\end{aligned}
$$

If $\mu = 0$, then the beta-binomial distribution reduces to a standard binomial model with success probability $\pi$. The parameterization of the beta-binomial distribution used by the FMM procedure is based on Neerchal and Morel (1998), see the section "Log-Likelihood Functions for Response Distributions" on page 2948 for details.

Morel and Nagaraj (1993); Morel and Neerchal (1997); Neerchal and Morel (1998) propose a different model to capture dependency within binomial clusters. Their model is a two-component mixture that gives rise to the same mean and variance function as the beta-binomial model. The genesis is different, however. In the binomial cluster model of Morel and Neerchal, suppose there is a cluster of $n$ Bernoulli outcomes with success probability $\pi$. The number of responses in the cluster decomposes into $N \leq n$ outcomes that all respond with either "success" or "failure"; the important aspect is that they all respond identically. The remaining $n - N$ Bernoulli outcomes respond independently, so the sum of successes in this group is a binomial$(n - N, \pi)$ random variable. Denote the probability with which cluster members fall into the group of identical respondents as $\mu$. Then $1 - \mu$ is the probability that a response belongs to the group of independent Bernoulli outcomes.

It is easy to see how this process of dividing the individual Bernoulli outcomes creates clustering. The binomial cluster model can be written as the two-component mixture

$$\Pr(Y = y) = \pi \Pr(U = y) + (1 - \pi) \Pr(V = y)$$

where $U \sim \text{Binomial}(n, \mu^* + \mu)$, $V \sim \text{Binomial}(n, \mu^*)$, and $\mu^* = (1 - \mu)\pi$. This mixture model is somewhat unusual because the mixing probability $\pi$ appears as a parameter in the component distributions. The two probabilities involved, $\pi$ and $\mu$, have the following interpretation: $\pi$ is the unconditional probability of success for any observation, and $\mu$ is the probability with which the Bernoulli observations respond identically. The complement of this probability, $1 - \mu$, is the probability with which the Bernoulli outcomes respond independently. If $\mu = 0$, then the two-component mixture reduces to a standard binomial model that has success probability $\pi$. Because both $\pi$ and $\mu$ are involved in the success probabilities of the two binomial variables in the mixture, you can affect these binomial means by specifying effects in the PROBMODEL statement (for the $\pi$'s) or the MODEL statement (for the $\mu$'s). You would vary the success probabilities $\mu_1$ and $\mu_2$ by using the MODEL statement in a "straight" two-component binomial mixture,

$$\pi\text{Binomial}(n, \mu_1) + (1 - \pi)\text{Binomial}(n, \mu_2)$$

You can fit the beta-binomial model by specifying DIST=BETABIN and the binomial cluster model by specifying DIST=BINOMCLUSTER in the MODEL statement.

Morel and Neerchal (1997) report data from a completely randomized design that studies the teratogenicity of phenytoin in 81 pregnant mice. The treatment structure of the experiment is an augmented factorial. In addition to an untreated control, mice received 60 mg/kg of phenytoin (PHT), 100 mg/kg of trichloropropene oxide (TCPO), and their combination. The design was augmented with a control group that was treated with water. As in Morel and Neerchal (1997), the two control groups are combined here into a single group.

The following DATA step creates the data for this analysis as displayed in Table 1 of Morel and Neerchal (1997). The second DATA step creates continuous variables x1–x3 to match the parameterization of these authors.

```
data ossi;
   length tx $8;
   input tx$ n @@;
   do i=1 to n;
      input y m @@;
      output;
   end;
   drop i;
   datalines;
Control  18 8 8 9  9  7  9 0  5 3  3 5 8 9 10 5 8 5 8 1 6 0 5
            8 8 9 10  5  5 4  7 9 10 6 6 3  5
Control  17 8 9 7 10 10 10 1  6 6  6 1 9 8  9 6 7 5 5 7 9
            2 5 5  6  2  8 1  8 0  2 7 8 5  7
PHT      19 1 9 4  9  3  7 4  7 0  7 0 4 1  8 1 7 2 7 2 8 1 7
            0 2 3 10  3  7 2  7 0  8 0 8 1 10 1 1
TCPO     16 0 5 7 10  4  4 8 11 6 10 6 9 3  4 2 8 0 6 0 9
            3 6 2  9  7  9 1 10 8  8 6 9
PHT+TCPO 11 2 2 0  7  1  8 7  8 0 10 0 4 0  6 0 7 6 6 1 6 1 7
;

data ossi;
   set ossi;
   array xx{3} x1-x3;
   do i=1 to 3; xx{i}=0; end;
   pht  = 0;
   tcpo = 0;
   if (tx='TCPO') then do;
      xx{1} = 1;
      tcpo  = 100;
   end; else if (tx='PHT') then do;
      xx{2} = 1;
      pht   = 60;
   end; else if (tx='PHT+TCPO') then do;
      pht  = 60;
      tcpo = 100;
      xx{1} = 1; xx{2} = 1; xx{3}=1;
   end;
run;
```

The FMM procedure models the mean parameters $\mu$ through the MODEL statement and the mixing proportions $\pi$ through the PROBMODEL statement. In the binomial cluster model, you can place a regression structure on either set of probabilities, and the regression structure does not need to be the same. In the following statements, the unconditional probability of ossification is modeled as a two-way factorial, whereas

the intralitter effect—the propensity to group within a cluster—is assumed to be constant:

```
proc fmm data=ossi;
   class pht tcpo;
   model y/m = / dist=binomcluster;
   probmodel pht tcpo pht*tcpo;
run;
```

The CLASS statement declares the PHT and TCPO variables as classification variables. They affect the analysis through their levels, not through their numeric values. The MODEL statement declares the distribution of the data to follow a binomial cluster model. The FMM procedure then automatically assumes that the model is a two-component mixture. An intercept is included by default. The PROBMODEL statement declares the effect structure for the mixing probabilities. The unconditional probability of ossification of a fetus depends on the main effects and the interaction in the factorial.

The "Model Information" table displays important details about the model fit with the FMM procedure (Output 43.1.1). Although no K= option was specified in the MODEL statement, the FMM procedure recognizes the model as a two-component model. The "Class Level Information" table displays the levels and values of the PHT and TCPO variables. Eighty-one observations are read from the data and are used in the analysis. These observations comprise 287 events and 585 total outcomes.

**Output 43.1.1** Model Information in Binomial Cluster Model with Constant Clustering Probability

**The FMM Procedure**

| Model Information | |
|---|---|
| **Data Set** | WORK.OSSI |
| **Response Variable (Events)** | y |
| **Response Variable (Trials)** | m |
| **Type of Model** | Binomial Cluster |
| **Distribution** | Binomial Cluster |
| **Components** | 2 |
| **Link Function** | Logit |
| **Estimation Method** | Maximum Likelihood |

| Class Level Information | | |
|---|---|---|
| **Class** | **Levels** | **Values** |
| **pht** | 2 | 0 60 |
| **tcpo** | 2 | 0 100 |

| | |
|---|---|
| **Number of Observations Read** | 81 |
| **Number of Observations Used** | 81 |
| **Number of Events** | 287 |
| **Number of Trials** | 585 |

The "Optimization Information" table in Output 43.1.2 gives details about the maximum likelihood optimization. By default, the FMM procedure uses a quasi-Newton algorithm. The model contains five parameters, four of which are part of the model for the mixing probabilities. The fifth parameter is the intercept in the model for $\mu$.

**Output 43.1.2** Optimization in Binomial Cluster Model with Constant Clustering Probability

| Optimization Information | |
|---|---|
| **Optimization Technique** | Dual Quasi-Newton |
| **Parameters in Optimization** | 5 |
| **Mean Function Parameters** | 1 |
| **Scale Parameters** | 0 |
| **Mixing Prob Parameters** | 4 |
| **Number of Threads** | 4 |

| Iteration History | | | | |
|---|---|---|---|---|
| **Iteration** | **Evaluations** | **Objective Function** | **Change** | **Max Gradient** |
| **0** | 5 | 174.9272389 | | 43.78769 |
| **1** | 2 | 154.1318074 | 20.79543149 | 11.23460 |
| **2** | 3 | 153.2669361 | 0.86487133 | 6.888215 |
| **3** | 2 | 152.8497428 | 0.41719329 | 3.541977 |
| **4** | 3 | 152.6175603 | 0.23218248 | 2.783556 |
| **5** | 3 | 152.5479530 | 0.06960730 | 1.146807 |
| **6** | 3 | 152.5268493 | 0.02110374 | 0.034367 |
| **7** | 3 | 152.5267121 | 0.00013715 | 0.011511 |
| **8** | 3 | 152.5267080 | 0.00000415 | 0.000202 |
| **9** | 3 | 152.5267080 | 0.00000000 | 4.001E-6 |

Convergence criterion (GCONV=1E-8) satisfied.

| Fit Statistics | |
|---|---|
| **-2 Log Likelihood** | 305.1 |
| **AIC (Smaller is Better)** | 315.1 |
| **AICC (Smaller is Better)** | 315.9 |
| **BIC (Smaller is Better)** | 327.0 |
| **Pearson Statistic** | 89.2077 |
| **Effective Parameters** | 5 |
| **Effective Components** | 2 |

After nine iterations, the iterative optimization converges. The –2 log likelihood at the converged solution is 305.1, and the Pearson statistic is 89.2077. The FMM procedure computes the Pearson statistic as a general goodness-of-fit measure that expresses the closeness of the fitted model to the data.

The estimates of the parameters in the conditional probability $\mu$ and in the unconditional probability $\pi$ are given in Output 43.1.3. The intercept estimate in the model for $\mu$ is 0.3356. Because the default link in the binomial cluster model is the logit link, the estimate of the conditional probability is

$$\widehat{\mu} = \frac{1}{1 + \exp\{-0.3356\}} = 0.5831$$

This value is displayed in the "Inverse Linked Estimate" column. There is greater than a 50% chance that the individual fetuses in a litter provide the same response. The clustering tendency is substantial.

**Output 43.1.3** Parameter Estimates in Binomial Cluster Model with Constant Clustering Probability

| Parameter Estimates for Binomial Cluster Model | | | | | | |
|---|---|---|---|---|---|---|
| Component | Effect | Estimate | Standard Error | z Value | Pr > \|z\| | Inverse Linked Estimate |
| 1 | Intercept | 0.3356 | 0.1714 | 1.96 | 0.0503 | 0.5831 |

| Parameter Estimates for Mixing Probabilities | | | | | | | |
|---|---|---|---|---|---|---|---|
| Component | Effect | pht | tcpo | Estimate | Standard Error | z Value | Pr > \|z\| |
| 1 | Intercept | | | -1.2194 | 0.4690 | -2.60 | 0.0093 |
| 1 | pht | 0 | | 0.9129 | 0.5608 | 1.63 | 0.1036 |
| 1 | pht | 60 | | 0 | . | . | . |
| 1 | tcpo | | 0 | 0.3295 | 0.5534 | 0.60 | 0.5516 |
| 1 | tcpo | | 100 | 0 | . | . | . |
| 1 | pht*tcpo | 0 | 0 | 0.6162 | 0.6678 | 0.92 | 0.3561 |
| 1 | pht*tcpo | 0 | 100 | 0 | . | . | . |
| 1 | pht*tcpo | 60 | 0 | 0 | . | . | . |
| 1 | pht*tcpo | 60 | 100 | 0 | . | . | . |

The "Mixing Probabilities" table displays the estimates of the parameters in the model for $\pi$ on the logit scale (Output 43.1.3). Table 43.11 constructs the estimates of the unconditional probabilities of ossification.

**Table 43.11** Estimates of Ossification Probabilities

| PHT | TCPO | $\widehat{\eta}$ | $\widehat{\pi}$ |
|---|---|---|---|
| 0 | 0 | $-1.2194 + 0.9129 + 0.3295 + 0.6162 = 0.6392$ | 0.6546 |
| 60 | 0 | $-1.2194 + 0.3295 = -0.8899$ | 0.2911 |
| 0 | 100 | $-1.2194 + 0.9129 = -0.3065$ | 0.4240 |
| 60 | 100 | $-1.2194$ | 0.2280 |

Morel and Neerchal (1997) considered a model in which the intralitter effects also depend on the treatments. This model is fit with the FMM procedure with the following statements:

```
proc fmm data=ossi;
   class pht tcpo;
   model y/m = pht tcpo pht*tcpo / dist=binomcluster;
   probmodel   pht tcpo pht*tcpo;
run;
```

The –2 log likelihood of this model is much reduced compared to the previous model with constant conditional probability (compare 287.8 in Output 43.1.4 with 305.1 in Output 43.1.2). The likelihood-ratio statistic of 17.3 is significant, $\Pr(\chi^2_3 > 17.3 = 0.0006)$. Varying the conditional probabilities by treatment improved the model fit significantly.

**Output 43.1.4** Fit Statistics and Parameter Estimates in Binomial Cluster Model

**The FMM Procedure**

| Fit Statistics | |
|---|---|
| -2 Log Likelihood | 287.8 |
| AIC (Smaller is Better) | 303.8 |
| AICC (Smaller is Better) | 305.8 |
| BIC (Smaller is Better) | 323.0 |
| Pearson Statistic | 85.5998 |
| Effective Parameters | 8 |
| Effective Components | 2 |

**Parameter Estimates for Binomial Cluster Model**

| Component | Effect | pht | tcpo | Estimate | Standard Error | z Value | Pr > \|z\| |
|---|---|---|---|---|---|---|---|
| 1 | Intercept | | | 1.8213 | 0.5889 | 3.09 | 0.0020 |
| 1 | pht | 0 | | -1.4962 | 0.6630 | -2.26 | 0.0240 |
| 1 | pht | 60 | | 0 | . | . | . |
| 1 | tcpo | | 0 | -3.1828 | 1.1261 | -2.83 | 0.0047 |
| 1 | tcpo | | 100 | 0 | . | . | . |
| 1 | pht*tcpo | 0 | 0 | 3.3736 | 1.1953 | 2.82 | 0.0048 |
| 1 | pht*tcpo | 0 | 100 | 0 | . | . | . |
| 1 | pht*tcpo | 60 | 0 | 0 | . | . | . |
| 1 | pht*tcpo | 60 | 100 | 0 | . | . | . |

**Parameter Estimates for Mixing Probabilities**

| Component | Effect | pht | tcpo | Estimate | Standard Error | z Value | Pr > \|z\| |
|---|---|---|---|---|---|---|---|
| 1 | Intercept | | | -0.7394 | 0.5395 | -1.37 | 0.1705 |
| 1 | pht | 0 | | 0.4351 | 0.6203 | 0.70 | 0.4830 |
| 1 | pht | 60 | | 0 | . | . | . |
| 1 | tcpo | | 0 | -0.5342 | 0.5893 | -0.91 | 0.3646 |
| 1 | tcpo | | 100 | 0 | . | . | . |
| 1 | pht*tcpo | 0 | 0 | 1.4055 | 0.7080 | 1.99 | 0.0471 |
| 1 | pht*tcpo | 0 | 100 | 0 | . | . | . |
| 1 | pht*tcpo | 60 | 0 | 0 | . | . | . |
| 1 | pht*tcpo | 60 | 100 | 0 | . | . | . |

Table 43.12 computes the conditional probabilities in the four treatment groups. Recall that the previous model estimated a constant clustering probability of 0.5831.

**Table 43.12** Estimates of Clustering Probabilities

| PHT | TCPO | $\widehat{\eta}$ | $\widehat{\mu}$ |
|---|---|---|---|
| 0 | 0 | $1.8213 - 1.4962 - 3.1828 + 3.3736 = 0.5159$ | 0.6262 |
| 60 | 0 | $1.8213 - 3.1828 = -1.3615$ | 0.2040 |
| 0 | 100 | $1.8213 - 1.4962 = 0.3251$ | 0.5806 |
| 60 | 100 | 1.8213 | 0.8607 |

The presence of phenytoin alone reduces the probability of response clustering within the litter. The presence of trichloropropene oxide alone does not have a strong effect on the clustering. The simultaneous presence of both agents substantially increases the probability of clustering.

The following statements fit the binomial cluster model in the parameterization of Morel and Neerchal (1997).

```
proc fmm data=ossi;
   model y/m = x1-x3 / dist=binomcluster;
   probmodel   x1-x3;
run;
```

The model fit is the same as in the previous model (compare the "Fit Statistics" tables in Output 43.1.5 and Output 43.1.4). The parameter estimates change due to the reparameterization of the treatment effects and match the results in Table III of Morel and Neerchal (1997).

**Output 43.1.5** Fit Statistics and Estimates (Morel and Neerchal Parameterization)

**The FMM Procedure**

| Fit Statistics | |
|---|---|
| -2 Log Likelihood | 287.8 |
| AIC (Smaller is Better) | 303.8 |
| AICC (Smaller is Better) | 305.8 |
| BIC (Smaller is Better) | 323.0 |
| Pearson Statistic | 85.5999 |
| Effective Parameters | 8 |
| Effective Components | 2 |

Parameter Estimates for Binomial Cluster Model

| Component | Effect | Estimate | Standard Error | z Value | Pr > \|z\| |
|---|---|---|---|---|---|
| 1 | Intercept | 0.5159 | 0.2603 | 1.98 | 0.0475 |
| 1 | x1 | -0.1908 | 0.4006 | -0.48 | 0.6339 |
| 1 | x2 | -1.8774 | 0.9946 | -1.89 | 0.0591 |
| 1 | x3 | 3.3736 | 1.1953 | 2.82 | 0.0048 |

Parameter Estimates for Mixing Probabilities

| Component | Effect | Estimate | Standard Error | z Value | Pr > \|z\| |
|---|---|---|---|---|---|
| 1 | Intercept | 0.5669 | 0.2455 | 2.31 | 0.0209 |
| 1 | x1 | -0.8712 | 0.3924 | -2.22 | 0.0264 |
| 1 | x2 | -1.8405 | 0.3413 | -5.39 | <.0001 |
| 1 | x3 | 1.4055 | 0.7080 | 1.99 | 0.0471 |

The following sets of statements fit the binomial and beta-binomial models, respectively, as single-component mixtures in the parameterization akin to the first binomial cluster model. Note that the model effects that affect the underlying Bernoulli success probabilities are specified in the MODEL statement, in contrast to the binomial cluster model.

```
proc fmm data=ossi;
   model y/m = x1-x3 / dist=binomial;
run;
```

```
proc fmm data=ossi;
   model y/m = x1-x3 / dist=betabinomial;
run;
```

The Pearson statistic for the beta-binomial model (Output 43.1.6) indicates a much better fit compared to the single-component binomial model (Output 43.1.7). This is not surprising because these data are obviously overdispersed relative to a binomial model because the Bernoulli outcomes are not independent. The difference between the binomial cluster and the beta-binomial model lies in the mechanism by which the correlations are induced:

- a mixing mechanism in the beta-binomial model that leads to a common shared random effect among all offspring in a cluster
- a mixture specification in the binomial cluster model that divides the offspring in a litter into identical and independent responders

**Output 43.1.6** Fit Statistics in Binomial Model

**The FMM Procedure**

| Fit Statistics | |
|---|---|
| -2 Log Likelihood | 401.8 |
| AIC (Smaller is Better) | 409.8 |
| AICC (Smaller is Better) | 410.3 |
| BIC (Smaller is Better) | 419.4 |
| Pearson Statistic | 252.1 |

**Output 43.1.7** Fit Statistics in Beta-Binomial Model

**The FMM Procedure**

| Fit Statistics | |
|---|---|
| -2 Log Likelihood | 306.6 |
| AIC (Smaller is Better) | 316.6 |
| AICC (Smaller is Better) | 317.4 |
| BIC (Smaller is Better) | 328.5 |
| Pearson Statistic | 87.5379 |

## Example 43.2: The Usefulness of Custom Starting Values: When Do Cows Eat?

This example with a mixture of normal and Weibull distributions illustrates the benefits of specifying starting values for some of the components.

The data for this example were generously provided by Dr. Luciano A. Gonzalez of the Lethbridge Research Centre of Agriculture and Agri-Food Canada and his collaborator, Dr. Bert Tolkamp, from the Scottish Agricultural College.

The outcome variable of interest is the logarithm of a time interval between consecutive visits by cattle to feeders. The intervals fall into three categories:

- short breaks within meals—such as when an animal stops eating for a moment and resumes shortly thereafter
- somewhat longer breaks when eating is interrupted to go have a drink of water
- long breaks between meals

Modeling such time interval data is important to understand the feeding behavior and biology of the animals and to derive other biological parameters such as the probability of an animal to stop eating after it has consumed a certain amount of a given food. Because there are three distinct biological categories, data of this nature are frequently modeled as three-component mixtures. The point at which the second and third components cross over is used to separate feeding events into meals.

The original data set comprises 141,414 observations of log feeding intervals. For the purpose of presentation in this document, where space is limited, the data have been rounded to precision 0.05 and grouped by frequency. The following DATA step displays the modified data used in this example. A comparison with the raw data and the results obtained in a full analysis of the original data show that the grouping does not alter the presentation or conclusions in a way that matters for the purpose of this example.

```
data cattle;
   input LogInt Count @@;
   datalines;
 0.70   195  1.10   233  1.40   355  1.60   563
 1.80   822  1.95   926  2.10  1018  2.20  1712
 2.30  3190  2.40  2212  2.50  1692  2.55  1558
 2.65  1622  2.70  1637  2.75  1568  2.85  1599
 2.90  1575  2.95  1526  3.00  1537  3.05  1561
 3.10  1555  3.15  1427  3.20  2852  3.25  1396
 3.30  1343  3.35  2473  3.40  1310  3.45  2453
 3.50  1168  3.55  2300  3.60  2174  3.65  2050
 3.70  1926  3.75  1849  3.80  1687  3.85  2416
 3.90  1449  3.95  2095  4.00  1278  4.05  1864
 4.10  1672  4.15  2104  4.20  1443  4.25  1341
 4.30  1685  4.35  1445  4.40  1369  4.45  1284
 4.50  1523  4.55  1367  4.60  1027  4.65  1491
 4.70  1057  4.75  1155  4.80  1095  4.85  1019
 4.90  1158  4.95  1088  5.00  1075  5.05   912
 5.10  1073  5.15   803  5.20   924  5.25   916
 5.30   784  5.35   751  5.40   766  5.45   833
 5.50   748  5.55   725  5.60   674  5.65   690
 5.70   659  5.75   695  5.80   529  5.85   639
 5.90   580  5.95   557  6.00   524  6.05   473
 6.10   538  6.15   444  6.20   456  6.25   453
 6.30   374  6.35   406  6.40   409  6.45   371
 6.50   320  6.55   334  6.60   353  6.65   305
 6.70   302  6.75   301  6.80   263  6.85   218
 6.90   255  6.95   240  7.00   219  7.05   202
 7.10   192  7.15   180  7.20   162  7.25   126
 7.30   148  7.35   173  7.40   142  7.45   163
```

```
 7.50   152  7.55   149  7.60   139  7.65   161
 7.70   174  7.75   179  7.80   188  7.85   239
 7.90   225  7.95   213  8.00   235  8.05   256
 8.10   272  8.15   290  8.20   320  8.25   355
 8.30   307  8.35   311  8.40   317  8.45   335
 8.50   369  8.55   365  8.60   365  8.65   396
 8.70   419  8.75   467  8.80   468  8.85   515
 8.90   558  8.95   623  9.00   712  9.05   716
 9.10   829  9.15   803  9.20   834  9.25   856
 9.30   838  9.35   842  9.40   826  9.45   834
 9.50   798  9.55   801  9.60   780  9.65   849
 9.70   779  9.75   737  9.80   683  9.85   686
 9.90   626  9.95   582 10.00   522 10.05   450
10.10   443 10.15   375 10.20   342 10.25   285
10.30   254 10.35   231 10.40   195 10.45   186
10.50   143 10.55   100 10.60    73 10.65    49
10.70    28 10.75    36 10.80    16 10.85     9
10.90     5 10.95     6 11.00     4 11.05     1
11.15     1 11.25     4 11.30     2 11.35     5
11.40     4 11.45     3 11.50     1
;
```

If you scan the columns for the Count variable in the DATA step, the prevalence of values between 2 and 5 units of LogInt is apparent, as is a long right tail. To explore these data graphically, the following statements produce a histogram of the data and a kernel density estimate of the density of the LogInt variable.

```
ods graphics on;
proc kde data=cattle;
   univar LogInt / bwm=4;
   freq count;
run;
```

**Output 43.2.1** Histogram and Kernel Density for LogInt

Two modes are clearly visible in Output 43.2.1. Given the biological background, one would expect that three components contribute to the mixture. The histogram would suggest either a two-component mixture with modes near 4 and 9, or a three-component mixture with modes near 3, 5, and 9.

Following Dr. Gonzalez' suggestion, the process is modeled as a three-component mixture of two normal distributions and a Weibull distribution. The Weibull distribution is chosen because it can have long left and right tails and it is popular in modeling data that relate to time intervals.

```
proc fmm data=cattle gconv=0;
   model LogInt = / dist=normal k=2 parms(3 1, 5 1);
   model        + / dist=weibull;
   freq count;
run;
```

The GCONV= convergence criterion is turned off in this PROC FMM run to avoid the early stoppage of the iterations when the relative gradient changes little between iterations. Turning the criterion off usually ensures that convergence is achieved with a small absolute gradient of the objective function. The PARMS option in the first MODEL statement provides starting values for the means and variances for the parameters of the normal distributions. The means for the two components are started at $\mu = 3$ and $\mu = 5$, respectively. Specifying starting values is generally not necessary. However, the choice of starting values can play an

important role in modeling finite mixture models; the importance of the choice of starting values in this example is discussed further below.

The "Model Information" table shows that the model is a three-component mixture and that the FMM procedure considers the estimation of a density to be the purpose of modeling. The procedure draws this conclusion from the absence of effects in the MODEL statements. There are 187 observations in the data set, but these actually represent 141,414 measurements (Output 43.2.2).

**Output 43.2.2** Model Information and Number of Observations

**The FMM Procedure**

| Model Information | |
|---|---|
| **Data Set** | WORK.CATTLE |
| **Response Variable** | LogInt |
| **Frequency Variable** | Count |
| **Type of Model** | Density Estimation |
| **Components** | 3 |
| **Estimation Method** | Maximum Likelihood |

| | |
|---|---|
| **Number of Observations Read** | 187 |
| **Number of Observations Used** | 187 |
| **Sum of Frequencies Read** | 141414 |
| **Sum of Frequencies Used** | 141414 |

There are eight parameters in the optimization: the means and variances of the two normal distributions, the $\mu$ and $\phi$ parameter of the Weibull distribution, and the two mixing probabilities (Output 43.2.3). At the converged solution, the –2 log likelihood is 563,153 and all parameters and components are effective—that is, the model is not overspecified in the sense that components have collapsed during the model fitting. The Pearson statistic is close to the number of observations in the data set, indicating a good fit.

**Output 43.2.3** Optimization Information and Fit Statistics

| Optimization Information | |
|---|---|
| **Optimization Technique** | Dual Quasi-Newton |
| **Parameters in Optimization** | 8 |
| **Mean Function Parameters** | 3 |
| **Scale Parameters** | 3 |
| **Mixing Prob Parameters** | 2 |
| **Lower Boundaries** | 3 |
| **Upper Boundaries** | 0 |
| **Number of Threads** | 4 |

| Fit Statistics | |
|---|---|
| **-2 Log Likelihood** | 563153 |
| **AIC (Smaller is Better)** | 563169 |
| **AICC (Smaller is Better)** | 563169 |
| **BIC (Smaller is Better)** | 563248 |
| **Pearson Statistic** | 141458 |
| **Effective Parameters** | 8 |
| **Effective Components** | 3 |

Output 43.2.4 displays the parameter estimates for the three models and for the mixing probabilities. The order in which the "Parameter Estimates" tables appear in the output corresponds to the order in which the MODEL statements were specified.

**Output 43.2.4** Optimization Information and Fit Statistics

**Parameter Estimates for Normal Model**

| Component | Parameter | Estimate | Standard Error | z Value | Pr > \|z\| |
|---|---|---|---|---|---|
| 1 | Intercept | 3.3415 | 0.01260 | 265.16 | <.0001 |
| 2 | Intercept | 4.8940 | 0.05447 | 89.84 | <.0001 |
| 1 | Variance | 0.6718 | 0.01287 | | |
| 2 | Variance | 1.4497 | 0.05247 | | |

**Parameter Estimates for Weibull Model**

| Component | Parameter | Estimate | Standard Error | z Value | Pr > \|z\| | Inverse Linked Estimate |
|---|---|---|---|---|---|---|
| 3 | Intercept | 2.2531 | 0.000506 | 4452.11 | <.0001 | 9.5174 |
| 3 | Scale | 0.06848 | 0.000427 | | | |

**Parameter Estimates for Mixing Probabilities**

| Component | Mixing Probability | Linked Scale GLogit(Prob) | Standard Error | z Value | Pr > \|z\| |
|---|---|---|---|---|---|
| 1 | 0.4545 | 0.8106 | 0.03409 | 23.78 | <.0001 |
| 2 | 0.3435 | 0.5305 | 0.04640 | 11.43 | <.0001 |
| 3 | 0.2021 | 0 | | | |

The estimated means of the two normal components are 3.3415 and 4.8940, respectively. Note that the means are displayed here as Intercept. The inverse linked estimate is not produced because the default link for the normal distribution is the identity link; hence the Estimate column represents the means of the component distributions. The parameter estimates in the Weibull model are $\widehat{\beta}_0 = 2.2531$, $\widehat{\phi} = 0.06848$, and $\widehat{\mu} = \exp\{\widehat{\beta}_0\} = 9.5174$. In the Weibull distribution, the $\mu$ parameter does not estimate the mean of the distribution, the maximum likelihood estimate of the distribution's mean is $\widehat{\mu}\Gamma(\widehat{\phi} + 1) = 9.1828$.

The estimated mixing probabilities are $\widehat{\pi}_1 = 0.4545$, $\widehat{\pi}_2 = 0.3435$, and $\widehat{\pi_3} = 0.2021$. In other words, the estimated distribution of log feeding intervals is a 45:35:20 mixture of an $\mathrm{N}(3.3415, 0.6718)$, a $\mathrm{N}(4.8940, 1.4497)$, and a $\mathrm{Weibull}(9.5174, 0.06848)$ distribution.

You can obtain a graphical display of the observed and estimated distribution of these data by enabling ODS Graphics. The PLOTS option in the PROC FMM statement modifies the default density plot by adding the densities of the mixture components:

```
ods select DensityPlot;
proc fmm data=cattle gconv=0;
   model LogInt = / dist=normal k=2 parms(3 1, 5 1);
   model        + / dist=weibull;
   freq count;
run;
```

**Output 43.2.5** Observed and Estimated Densities in the Three-Component Model

The estimated mixture density matches the histogram of the observed data closely (Output 43.2.5). The component densities are displayed in such a way that, at each point in the support of the LogInt variable, their sum combines to the overall mixture density. The three components in the mixtures are well separated.

The excellent quality of the fit is even more evident when the distributions are displayed cumulatively by adding the CUMULATIVE option in the DENSITY option (Output 43.2.6):

```
ods select DensityPlot;
proc fmm data=cattle plot=density(cumulative) gconv=0;
   model LogInt = / dist=normal k=2 parms(3 1, 5 1);
   model        + / dist=weibull;
   freq count;
run;
```

The component cumulative distribution functions are again scaled so that their sum produces the overall mixture cumulative distribution function. Because of this scaling, the percentage reached at the maximum value of LogInt corresponds to the mixing probabilities in Output 43.2.4.

**Output 43.2.6** Observed and Estimated Cumulative Densities in the Three-Component Model

The importance of starting values for the parameter estimates was mentioned previously. Suppose that different starting values are selected for the three components (for example, the default starting values).

```
proc fmm data=cattle gconv=0;
   model LogInt = / dist=normal k=2;
   model        + / dist=weibull;
   freq count;
run;
ods graphics off;
```

The fit statistics and parameter estimates from this run are displayed in Output 43.2.7, and the density plot is shown in Output 43.2.8.

**Output 43.2.7** Fit Statistics and Parameter Estimates

**The FMM Procedure**

| Fit Statistics | |
|---|---|
| -2 Log Likelihood | 564431 |
| AIC (Smaller is Better) | 564447 |
| AICC (Smaller is Better) | 564447 |
| BIC (Smaller is Better) | 564526 |
| Pearson Statistic | 141228 |
| Effective Parameters | 8 |
| Effective Components | 3 |

**Parameter Estimates for Normal Model**

| Component | Parameter | Estimate | Standard Error | z Value | Pr > \|z\| |
|---|---|---|---|---|---|
| 1 | Intercept | 4.9106 | 0.02604 | 188.56 | <.0001 |
| 2 | Intercept | 9.2883 | 0.005031 | 1846.28 | <.0001 |
| 1 | Variance | 1.7410 | 0.02753 | | |
| 2 | Variance | 0.4158 | 0.005086 | | |

**Parameter Estimates for Weibull Model**

| Component | Parameter | Estimate | Standard Error | z Value | Pr > \|z\| | Inverse Linked Estimate |
|---|---|---|---|---|---|---|
| 3 | Intercept | 1.2908 | 0.002790 | 462.71 | <.0001 | 3.6358 |
| 3 | Scale | 0.2093 | 0.001311 | | | |

**Parameter Estimates for Mixing Probabilities**

| Component | Mixing Probability | Linked Scale GLogit(Prob) | Standard Error | z Value | Pr > \|z\| |
|---|---|---|---|---|---|
| 1 | 0.3745 | -0.1505 | 0.03678 | -4.09 | <.0001 |
| 2 | 0.1902 | -0.8280 | 0.01922 | -43.08 | <.0001 |
| 3 | 0.4353 | 0 | | | |

All components are active; no collapsing of components occurred. However, a closer look at the "Parameter Estimates" tables in Output 43.2.7 shows an important difference from the tables in Output 43.2.4. The means of the two normal distributions are now 4.9106 and 9.2883. Previously, the means were 3.3415 and 4.8940. The "position" of the Weibull distribution has moved from right to left, and the third component is now modeled by a symmetric normal distribution (Output 43.2.8). The mixture probabilities have also changed—in particular, for the first and third component.

**Output 43.2.8** Three-Component Model with Default Starting Values

Such switching is not uncommon in mixture modeling. As judged by the information criteria, the model in which the Weibull distribution is the component with the smallest mean does not fit the data as well as the first model in which the specification of the starting values guided the optimization towards placing the normal distributions first. The converged solution found in the last FMM run represents a local minimum of the log-likelihood surface. There are other local minima—for example, when components are removed from the model, which is tantamount to estimating the associated mixture probabilities as zero.

## Example 43.3: Enforcing Homogeneity Constraints: Count and Dispersion—It Is All Over!

The following example demonstrates how you can use either the EQUATE= option in the MODEL statement or the RESTRICT statement to impose homogeneity constraints on chosen model effects.

The data for this example were presented by Margolin, Kaplan, and Zeiger (1981) and analyzed by various authors applying a number of techniques. The following DATA step shows the number of revertant salmonella colonies (variable num) at six levels of quinoline dosing (variable dose). There are three replicate plates at each dose of quinoline.

```
data assay;
   label dose = 'Dose of quinoline (microg/plate)'
         num  = 'Observed number of colonies';
   input dose @;
   logd = log(dose+10);
   do i=1 to 3; input num@; output; end;
   datalines;
   0  15 21 29
  10  16 18 21
  33  16 26 33
 100  27 41 60
 333  33 38 41
1000  20 27 42
;
```

The basic notion is that the data are overdispersed relative to a Poisson distribution in which the logarithm of the mean count is modeled as a linear regression in dose (in $\mu g/\text{plate}$) and in the derived variable $\log\{\text{dose} + 10\}$ (Lawless 1987). The log of the expected count of revertants is thus

$$\beta_0 + \beta_1 \text{dose} + \beta_2 \log\{\text{dose} + 10\}$$

The following statements fit a standard Poisson regression model to these data:

```
proc fmm data=assay;
   model num = dose logd / dist=Poisson;
run;
```

The Pearson statistic for this model is rather large compared to the number of degrees of freedom ($18-3 = 15$). The ratio $46.2707/15 = 3.08$ indicates an overdispersion problem in the Poisson model (Output 43.3.1).

**Output 43.3.1** Result of Fitting Poisson Regression Models

**The FMM Procedure**

| | |
|---|---|
| **Number of Observations Read** | 18 |
| **Number of Observations Used** | 18 |

| Fit Statistics | |
|---|---|
| **-2 Log Likelihood** | 136.3 |
| **AIC (Smaller is Better)** | 142.3 |
| **AICC (Smaller is Better)** | 144.0 |
| **BIC (Smaller is Better)** | 144.9 |
| **Pearson Statistic** | 46.2707 |

**Parameter Estimates for Poisson Model**

| Effect | Estimate | Standard Error | z Value | Pr > \|z\| |
|---|---|---|---|---|
| **Intercept** | 2.1728 | 0.2184 | 9.95 | <.0001 |
| **dose** | -0.00101 | 0.000245 | -4.13 | <.0001 |
| **logd** | 0.3198 | 0.05700 | 5.61 | <.0001 |

Breslow (1984) accounts for overdispersion by including a random effect in the predictor for the log rate and applying a quasi-likelihood technique to estimate the parameters. Wang et al. (1996) examine these data using mixtures of Poisson regression models. They fit several two- and three-component Poisson regression mixtures. Examining the log likelihoods, AIC, and BIC criteria, they eventually settle on a two-component model in which the intercepts vary by category and the regression coefficients are the same. This mixture model can be written as

$$
\begin{aligned}
f(y) =& \pi \frac{1}{y!}\lambda_1^y \exp\{-\lambda_1\} + (1-\pi)\frac{1}{y!}\lambda_2^y \exp\{-\lambda_2\} \\
\lambda_1 =& \exp\{\beta_{01} + \beta_1 \text{dose} + \beta_2 \log[\text{dose} + 10]\} \\
\lambda_2 =& \exp\{\beta_{02} + \beta_1 \text{dose} + \beta_2 \log[\text{dose} + 10]\}
\end{aligned}
$$

This model is fit with the FMM procedure with the following statements:

```
proc fmm data=assay;
   model num = dose logd / dist=Poisson k=2
                           equate=effects(dose logd);
run;
```

The EQUATE= option in the MODEL statement places constraints on the optimization and makes the coefficients for dose and logd homogeneous across components in the model. Output 43.3.2 displays the "Fit Statistics" and parameter estimates in the mixture. The Pearson statistic is drastically reduced compared to the Poisson regression model in Output 43.3.1. With $18 - 5 = 13$ degrees of freedom, the ratio of the Pearson and the degrees of freedom is now $16.1573/13 = 1.2429$. Note that the effective number of parameters was used to compute the degrees of freedom, not the total number of parameters, because of the equality constraints.

**Output 43.3.2** Result for Two-Component Poisson Regression Mixture

**The FMM Procedure**

| Fit Statistics | |
|---|---|
| -2 Log Likelihood | 121.8 |
| AIC (Smaller is Better) | 131.8 |
| AICC (Smaller is Better) | 136.8 |
| BIC (Smaller is Better) | 136.3 |
| Pearson Statistic | 16.1573 |
| Effective Parameters | 5 |
| Effective Components | 2 |

**Parameter Estimates for Poisson Model**

| Component | Effect | Estimate | Standard Error | z Value | Pr > \|z\| |
|---|---|---|---|---|---|
| 1 | Intercept | 1.9097 | 0.2654 | 7.20 | <.0001 |
| 1 | dose | -0.00126 | 0.000273 | -4.62 | <.0001 |
| 1 | logd | 0.3639 | 0.06602 | 5.51 | <.0001 |
| 2 | Intercept | 2.4770 | 0.2731 | 9.07 | <.0001 |
| 2 | dose | -0.00126 | 0.000273 | -4.62 | <.0001 |
| 2 | logd | 0.3639 | 0.06602 | 5.51 | <.0001 |

**Output 43.3.2** *continued*

| Parameter Estimates for Mixing Probabilities | | | | | |
|---|---|---|---|---|---|
| | | Linked Scale | | | |
| Component | Mixing Probability | Logit(Prob) | Standard Error | z Value | Pr > \|z\| |
| 1 | 0.8173 | 1.4984 | 0.6875 | 2.18 | 0.0293 |
| 2 | 0.1827 | -1.4984 | | | |

You could also have used RESTRICT statements to impose the homogeneity constraints on the model fit, as shown in the following statements:

```
proc fmm data=assay;
   model num = dose logd / dist=Poisson k=2;
   restrict 'common dose' dose 1, dose -1;
   restrict 'common logd' logd 1, logd -1;
run;
```

The first RESTRICT statement equates the coefficients for the dose variable in the two components, and the second RESTRICT statement accomplishes the same for the coefficients of the logd variable. If the right-hand side of a restriction is not specified, PROC FMM defaults to equating the left-hand side of the restriction to zero. The "Linear Constraints" table in Output 43.3.3 shows that both linear equality constraints are active. The parameter estimates match the previous FMM run.

**Output 43.3.3** Result for Two-Component Mixture with RESTRICT Statements

**The FMM Procedure**

| Linear Constraints at Solution | | | | | |
|---|---|---|---|---|---|
| Label | k = 1 | k = 2 | | | Constraint Active |
| **common dose** | dose | - dose | = | 0 | Yes |
| **common logd** | logd | - logd | = | 0 | Yes |

| Parameter Estimates for Poisson Model | | | | | |
|---|---|---|---|---|---|
| Component | Effect | Estimate | Standard Error | z Value | Pr > \|z\| |
| 1 | **Intercept** | 1.9097 | 0.2654 | 7.20 | <.0001 |
| 1 | **dose** | -0.00126 | 0.000273 | -4.62 | <.0001 |
| 1 | **logd** | 0.3639 | 0.06602 | 5.51 | <.0001 |
| 2 | **Intercept** | 2.4770 | 0.2731 | 9.07 | <.0001 |
| 2 | **dose** | -0.00126 | 0.000273 | -4.62 | <.0001 |
| 2 | **logd** | 0.3639 | 0.06602 | 5.51 | <.0001 |

| Parameter Estimates for Mixing Probabilities | | | | | |
|---|---|---|---|---|---|
| | | Linked Scale | | | |
| Component | Mixing Probability | Logit(Prob) | Standard Error | z Value | Pr > \|z\| |
| 1 | 0.8173 | 1.4984 | 0.6875 | 2.18 | 0.0293 |
| 2 | 0.1827 | -1.4984 | | | |

Wang et al. (1996) note that observation 12 with a revertant colony count of 60 is comparably high. The following statements remove the observation from the analysis and fit their selected model:

```
proc fmm data=assay(where=(num ne 60));
   model num = dose logd / dist=Poisson k=2
                           equate=effects(dose logd);
run;
```

**Output 43.3.4** Result for Two-Component Model without Outlier

**The FMM Procedure**

| Fit Statistics | |
|---|---|
| -2 Log Likelihood | 111.5 |
| AIC (Smaller is Better) | 121.5 |
| AICC (Smaller is Better) | 126.9 |
| BIC (Smaller is Better) | 125.6 |
| Pearson Statistic | 16.5987 |
| Effective Parameters | 5 |
| Effective Components | 2 |

**Parameter Estimates for Poisson Model**

| Component | Effect | Estimate | Standard Error | z Value | Pr > \|z\| |
|---|---|---|---|---|---|
| 1 | Intercept | 2.2272 | 0.3022 | 7.37 | <.0001 |
| 1 | dose | -0.00065 | 0.000445 | -1.46 | 0.1440 |
| 1 | logd | 0.2432 | 0.1045 | 2.33 | 0.0199 |
| 2 | Intercept | 2.5477 | 0.3331 | 7.65 | <.0001 |
| 2 | dose | -0.00065 | 0.000445 | -1.46 | 0.1440 |
| 2 | logd | 0.2432 | 0.1045 | 2.33 | 0.0199 |

**Parameter Estimates for Mixing Probabilities**

| | | Linked Scale | | | |
|---|---|---|---|---|---|
| Component | Mixing Probability | Logit(Prob) | Standard Error | z Value | Pr > \|z\| |
| 1 | 0.5777 | 0.3134 | 1.7261 | 0.18 | 0.8559 |
| 2 | 0.4223 | -0.3134 | | | |

The ratio of Pearson Statistic over degrees of freedom (12) is only slightly worse than in the previous model; the loss of 5% of the observations carries a price (Output 43.3.4). The parameter estimates for the two intercepts are now fairly close. If the intercepts were identical, then the two-component model would collapse to the Poisson regression model:

```
proc fmm data=assay(where=(num ne 60));
   model num = dose logd / dist=Poisson;
run;
```

**Output 43.3.5** Result of Fitting Poisson Regression Model without Outlier

**The FMM Procedure**

| | |
|---|---|
| Number of Observations Read | 17 |
| Number of Observations Used | 17 |

**Output 43.3.5** *continued*

| Fit Statistics | |
|---|---|
| -2 Log Likelihood | 114.1 |
| AIC (Smaller is Better) | 120.1 |
| AICC (Smaller is Better) | 121.9 |
| BIC (Smaller is Better) | 122.5 |
| Pearson Statistic | 27.8008 |

| Parameter Estimates for Poisson Model | | | | |
|---|---|---|---|---|
| Effect | Estimate | Standard Error | z Value | Pr > \|z\| |
| Intercept | 2.3164 | 0.2244 | 10.32 | <.0001 |
| dose | -0.00072 | 0.000258 | -2.78 | 0.0055 |
| logd | 0.2603 | 0.05996 | 4.34 | <.0001 |

Compared to the same model applied to the full data, the Pearson statistic is much reduced (compare 46.2707 in Output 43.3.1 to 27.8008 in Output 43.3.5). The outlier—or *over*count, if you will—induces at least some of the *over*dispersion.

## Example 43.4: Modeling Multinomial Overdispersion: Town and Country

This example illustrates how you can use the multinomial distribution to model a discrete response that has multiple levels, and how you can use the multinomial cluster model to address overdispersion in multinomial models. The data are survey results from random samples of neighborhoods in both rural and urban areas of Montevideo, Minnesota. There are 18 rural neighborhoods and 17 urban neighborhoods in the survey. In each sampled neighborhood, five households were selected to be interviewed about their level of satisfaction with their homes. The families rated their level of satisfaction as “Unsatisfied,” “Satisfied,” or “Very Satisfied.” These data have previously been analyzed in Brier (1980), Koehler and Wilson (1986), Wilson (1989), and Morel and Nagaraj (1993).

The data include a location type and the numbers of households that respond at each satisfaction level:

```
data housing;
   label us    = 'Unsatisfied'
         s     = 'Satisfied'
         vs    = 'Very Satisfied';
   input type $ us s vs @@;
   datalines;
rural 3 2 0  rural 3 2 0  rural 0 5 0  rural 3 2 0  rural 0 5 0
rural 4 1 0  rural 3 2 0  rural 2 3 0  rural 4 0 1  rural 0 4 1
rural 2 3 0  rural 4 1 0  rural 4 1 0  rural 1 2 2  rural 4 1 0
rural 1 3 1  rural 4 1 0  rural 5 0 0
urban 0 4 1  urban 0 5 0  urban 0 3 2  urban 3 2 0  urban 2 3 0
urban 1 3 1  urban 4 1 0  urban 4 0 1  urban 0 3 2  urban 1 2 2
urban 0 5 0  urban 3 2 0  urban 2 3 0  urban 2 2 1  urban 4 0 1
urban 0 4 1  urban 4 1 0
;
```

The following statements fit a single-component multinomial model to these data, including the location type in the mean model for the multinomial. The response variables are the counts for each observation in vector form.

```
proc fmm data=housing;
   class type;
   model us s vs = Type  / dist=multinomial;
   output out=Pred pred;
run;
```

The model includes the only available covariate, Type, as an explanatory variable for the mean of the multinomial distribution. You use the OUTPUT statement and the PRED keyword to direct PROC FMM to include predicted values for each observation in the Pred output data set.

The "Model Information" table in Output 43.4.1 lists the response variables and indicates that this is a single-component multinomial model. The "Fit Statistics" table shows the associated fit statistics for the model.

**Output 43.4.1** Model Information and Fit Statistics for the Multinomial Model

**The FMM Procedure**

| Model Information | |
|---|---|
| **Data Set** | WORK.HOUSING |
| **Response Variable** | us |
| **Response Variable** | s |
| **Response Variable** | vs |
| **Type of Model** | Homogeneous Regression Mixture |
| **Distribution** | Multinomial |
| **Components** | 1 |
| **Link Function** | Logit |
| **Estimation Method** | Maximum Likelihood |

| Fit Statistics | |
|---|---|
| **-2 Log Likelihood** | 194.1 |
| **AIC (Smaller is Better)** | 202.1 |
| **AICC (Smaller is Better)** | 203.4 |
| **BIC (Smaller is Better)** | 208.3 |
| **Pearson Statistic** | 107.3 |

The parameter estimates capture the relationship between the explanatory variable Type and the different response levels, "Unsatisfied," "Satisfied," and "Very Satisfied." To maintain identifiability, the FMM procedure uses two sets of parameters for the three response variables to parameterize this model. Output 43.4.2 shows the resulting parameter estimates.

**Output 43.4.2** Parameter Estimates for the Multinomial Model

| Parameter Estimates for Multinomial Model | | | | | | |
|---|---|---|---|---|---|---|
| Response | Effect | type | Estimate | Standard Error | z Value | Pr > \|z\| |
| 1 | Intercept | | 0.9163 | 0.3416 | 2.68 | 0.0073 |
| 1 | type | rural | 1.3244 | 0.5813 | 2.28 | 0.0227 |
| 1 | type | urban | 0 | . | . | . |
| 2 | Intercept | | 1.2763 | 0.3265 | 3.91 | <.0001 |
| 2 | type | rural | 0.7519 | 0.5770 | 1.30 | 0.1925 |
| 2 | type | urban | 0 | . | . | . |

The Response column indicates the level of the response that is associated with the parameter set. In this model, Response 1 corresponds to the "Unsatisfied" level and Response 2 corresponds to the "Satisfied" level. This corresponds to the order in which you specify the response variables in the MODEL statement. The "Very Satisfied" level does not appear because of identifiability constraints; the corresponding parameter estimates are set to 0, which means that you can treat the "Very Satisfied" level as the reference level. The estimates of the intercept and the rural effect are positive for both of the other levels, indicating that the estimated proportion at the "Very Satisfied" level is smaller than the proportion at the other two levels for both rural and urban locations.

The following statements compute the predicted proportions for the rural and urban locations from the Pred output data set by normalizing the predicted counts for each location:

```
data Pred; set Pred;
   Pred_1 = Pred_1 / (us + s + vs);
   Pred_2 = Pred_2 / (us + s + vs);
   Pred_3 = Pred_3 / (us + s + vs);
run;

proc sort data=Pred nodupkey;
  by type;
proc print data=pred noobs;
  var type pred:;
run;
```

Output 43.4.3 shows the predicted proportions at each response level for each location type. As in Output 43.4.2, the order reflects the order in which you specified the responses in the MODEL statement. Pred_1 corresponds to "Unsatisfied", Pred_2 corresponds to "Satisfied," and Pred_3 corresponds to "Very Satisfied."

**Output 43.4.3** Predicted Proportions for Multinomial

| type | Pred_1 | Pred_2 | Pred_3 |
|---|---|---|---|
| rural | 0.52222 | 0.42222 | 0.05556 |
| urban | 0.35294 | 0.50588 | 0.14118 |

The estimates of response proportions for the two location types indicate a difference in the distribution of satisfaction levels for the rural and urban populations. In particular, the urban population shows a smaller proportion of respondents in the "Unsatisfied" category (Pred_1).

The number of degrees of freedom is $N \times (R-1) - p$, where $N$ is the number of observations, $R$ is the number of levels in the multinomial response, and $p$ is the number of parameters in the model. The ratio of

the Pearson statistic to the degrees of freedom is then 107.3 / (35 × 2 – 4) = 1.625; this is larger than 1 and so indicates potential overdispersion.

One explanation for overdispersion might be correlation. It is likely that the families in these households meet and talk with one another, which might result in some influence of opinions about housing satisfaction. The observations are not independent in this case; if you model the proportion of each level of satisfaction based only on location type, you will miss this interhousehold influence.

The multinomial cluster model (Morel and Nagaraj 1993) is based on the idea of "clumping"; that is, some proportion $\mu$ of the observed population responds in the same way. In the context of the housing satisfaction data, this means that the clumped responders all express the same satisfaction level. The remaining households respond according to a multinomial distribution with parameter $\boldsymbol{\pi}$.

In this model, the clumped responders respond identically with one of the three levels of satisfaction, and that level is not observable. This discrete latent factor makes a mixture of three multinomials an appropriate method. The difference between this mixture and a general mixture of multinomials is the role of the clumping proportion $\mu$ and the use of the mixing probabilities in the mean model. In this model, the mixing probabilities $\boldsymbol{\pi}$ also define the multinomial distribution that governs the distribution of the non-clumped responses.

The following statements fit a multinomial cluster model to these data:

```
proc fmm data=housing;
   class type;
   model us s vs = Type / dist=multinomcluster;
   output out=Pred pred;
   probmodel Type;
run;
```

You include Type in the mean for the underlying multinomial distribution by using the PROBMODEL statement and also in the mean for the clumping parameter $\mu$ by using the MODEL statement. Output 43.4.4 shows model information and fit statistics for this multinomial cluster model. Because the model specifies three response variables, the resulting mixture model has three components.

**Output 43.4.4** Model Information and Fit Statistics for the Multinomial Cluster Model

**The FMM Procedure**

| Model Information | |
|---|---|
| **Data Set** | WORK.HOUSING |
| **Response Variable** | us |
| **Response Variable** | s |
| **Response Variable** | vs |
| **Type of Model** | Multinomial Cluster |
| **Distribution** | Multinomial Cluster |
| **Components** | 3 |
| **Link Function** | Logit |
| **Estimation Method** | Maximum Likelihood |

**Output 43.4.4** *continued*

| Fit Statistics | |
|---|---|
| **-2 Log Likelihood** | 182.9 |
| **AIC (Smaller is Better)** | 194.9 |
| **AICC (Smaller is Better)** | 197.9 |
| **BIC (Smaller is Better)** | 204.3 |
| **Pearson Statistic** | 61.9809 |
| **Effective Parameters** | 6 |
| **Effective Components** | 3 |

The fit statistics are generally better for the multinomial cluster model. However, Output 43.4.5 indicates that the parameters in the mean model for the clumping probability $\mu$ are not significantly different from 0. There does not appear to be strong evidence for a clumping effect as modeled by the multinomial cluster model.

**Output 43.4.5** Parameter Estimates for the Multinomial Cluster Model

| Parameter Estimates for Multinomial Cluster Model | | | | | | |
|---|---|---|---|---|---|---|
| **Component** | **Effect** | **type** | **Estimate** | **Standard Error** | **z Value** | **Pr > \|z\|** |
| **1** | **Intercept** | | -0.3696 | 0.4385 | -0.84 | 0.3992 |
| **1** | **type** | rural | 0.09401 | 0.6312 | 0.15 | 0.8816 |
| **1** | **type** | urban | 0 | . | . | . |

In the multinomial cluster model, the predicted proportions are the same as the mixing probabilities. Output 43.4.6 shows the parameter estimates for the mixing probabilities.

**Output 43.4.6** Mixing Probability Parameter Estimates for the Multinomial Cluster Model

| Parameter Estimates for Mixing Probabilities | | | | | | |
|---|---|---|---|---|---|---|
| **Component** | **Effect** | **type** | **Estimate** | **Standard Error** | **z Value** | **Pr > \|z\|** |
| **1** | **Intercept** | | 0.6383 | 0.4106 | 1.55 | 0.1201 |
| **1** | **type** | rural | 1.4138 | 0.6781 | 2.08 | 0.0371 |
| **1** | **type** | urban | 0 | . | . | . |
| **2** | **Intercept** | | 1.1077 | 0.3741 | 2.96 | 0.0031 |
| **2** | **type** | rural | 0.7900 | 0.6527 | 1.21 | 0.2262 |
| **2** | **type** | urban | 0 | . | . | . |

As in the multinomial example, the estimates for the intercept and rural effect are positive for both the "Unsatisfied" and "Satisfied" response levels, indicating that these levels have larger predicted proportions than the "Very Satisfied" level.

You can use the same approach as before to produce the predicted proportions.

Output 43.4.7 shows the predicted proportions at each level of the response for each location type.

**Output 43.4.7** Predicted Proportions for the Multinomial Cluster Model

| type | Pred_1 | Pred_2 | Pred_3 |
|---|---|---|---|
| rural | 0.50367 | 0.43163 | 0.06471 |
| urban | 0.31977 | 0.51133 | 0.16890 |

By comparing Output 43.4.7 with Output 43.4.3, you can see that the proportion estimates are not markedly different between the models. This is consistent with the lack of significance in the multinomial cluster model's clumping parameters.

# References


Aldrich, J. (1997). "R. A. Fisher and the Making of Maximum Likelihood, 1912–1922." *Statistical Science* 12:162–176.

Breslow, N. E. (1984). "Extra-Poisson Variation in Log-Linear Models." *Journal of the Royal Statistical Society, Series C* 33:38–44.

Brier, S. S. (1980). "Analysis of Contingency Tables under Cluster Sampling." *Biometrika* 67:591–596.

Cameron, A. C., and Trivedi, P. K. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. (2006). "Deviance Information Criteria for Missing Data Models." *Bayesian Analysis* 1:651–674.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society, Series B* 39:1–38.

Everitt, B. S., and Hand, D. J. (1981). *Finite Mixture Distributions*. London: Chapman & Hall.

Ferrari, S. L. P., and Cribari-Neto, F. (2004). "Beta Regression for Modelling Rates and Proportions." *Journal of Applied Statistics* 31:799–815.

Fisher, R. A. (1921). "On the 'Probable Error' of a Coefficient of Correlation Deduced from a Small Sample." *Metron* 1:3–32.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York: Springer.

Gamerman, D. (1997). "Sampling from the Posterior Distribution in Generalized Linear Models." *Statistics and Computing* 7:57–68.

Geweke, J. (1992). "Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments." In *Bayesian Statistics*, vol. 4, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 169–193. Oxford: Clarendon Press.

Griffiths, D. A. (1973). "Maximum Likelihood Estimation for the Beta-Binomial Distribution and an Application to the Household Distribution of the Total Number of Cases of a Disease." *Biometrics* 29:637–648.

Haseman, J. K., and Kupper, L. L. (1979). "Analysis of Dichotomous Response Data from Certain Toxicological Experiments." *Biometrics* 35:281–293.

Heidelberger, P., and Welch, P. D. (1981). "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations." *Communications of the ACM* 24:233–245.

Heidelberger, P., and Welch, P. D. (1983). "Simulation Run Length Control in the Presence of an Initial Transient." *Operations Research* 31:1109–1144.

Joe, H., and Zhu, R. (2005). "Generalized Poisson Distribution: The Property of Mixture of Poisson and Comparison with Negative Binomial Distribution." *Biometrical Journal* 47:219–229.

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M. (1998). "Markov Chain Monte Carlo in Practice: A Roundtable Discussion." *American Statistician* 52:93–100.

Koehler, K. J., and Wilson, J. R. (1986). "Chi-Square Tests for Comparing Vectors of Proportions for Several Cluster Samples." *Communications in Statistics—Theory and Methods* 15:2977–2990.

Lawless, J. F. (1987). "Negative Binomial and Mixed Poisson Regression." *Canadian Journal of Statistics* 15:209–225.

Margolin, B. H., Kaplan, N. L., and Zeiger, E. (1981). "Statistical Analysis of the Ames Salmonella Microsome Test." *Proceedings of the National Academy of Sciences* 76:3779–3783.

McLachlan, G. J., and Peel, D. (2000). *Finite Mixture Models*. New York: John Wiley & Sons.

Morel, J. G., and Nagaraj, N. K. (1993). "A Finite Mixture Distribution for Modelling Multinomial Extra Variation." *Biometrika* 80:363–371.

Morel, J. G., and Neerchal, N. K. (1997). "Clustered Binary Logistic Regression in Teratology Data Using a Finite Mixture Distribution." *Statistics in Medicine* 16:2843–2853.

Neerchal, N. K., and Morel, J. G. (1998). "Large Cluster Results for Two Parametric Multinomial Extra Variation Models." *Journal of the American Statistical Association* 93:1078–1087.

Pearson, K. (1915). "On Certain Types of Compound Frequency Distributions in Which the Components Can Be Individually Described by Binomial Series." *Biometrika* 11:139–144.

Raftery, A. E. (1996). "Hypothesis Testing and Model Selection." In *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, 163–188. London: Chapman & Hall.

Raftery, A. E., and Lewis, S. M. (1992). "One Long Run with Diagnostics: Implementation Strategies for Markov Chain Monte Carlo." *Statistical Science* 7:493–497.

Raftery, A. E., and Lewis, S. M. (1995). "The Number of Iterations, Convergence Diagnostics, and Generic Metropolis Algorithms." In *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, D. J. Spiegelhalter, and S. Richardson, 115–130. London: Chapman & Hall.

Richardson, S. (2002). "Discussion of Spiegelhalter et al." *Journal of the Royal Statistical Society, Series B* 64:631.

Roeder, K. (1990). "Density Estimation with Confidence Sets Exemplified by Superclusters and Voids in the Galaxies." *Journal of the American Statistical Association* 85:617–624.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van der Linde, A. (2002). "Bayesian Measures of Model Complexity and Fit." *Journal of the Royal Statistical Society, Series B* 64:583–616. With discussion.

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: John Wiley & Sons.

Viallefont, V., Richardson, S., and Greene, P. J. (2002). "Bayesian Analysis of Poisson Mixtures." *Journal of Nonparametric Statistics* 14:181–202.

Wang, P., Puterman, M. L., Cockburn, I., and Le, N. (1996). "Mixed Poisson Regression Models with Covariate Dependent Rates." *Biometrics* 52:381–400.

Williams, D. A. (1975). "The Analysis of Binary Responses from Toxicological Experiments Involving Reproduction and Teratogenicity." *Biometrics* 31:949–952.

Wilson, J. R. (1989). "Chi-Square Tests for Overdispersion with Multiparameter Estimates." *Journal of the Royal Statistical Society, Series C* 38:441–453.

# Subject Index


alpha level
  FMM procedure, 2930, 2939

Bayes information
  FMM procedure, 2963
Bayesian analysis
  FMM procedure, 2915
Bayesian fit statistics
  FMM procedure, 2963
Bernoulli distribution
  FMM procedure, 2930
beta distribution
  FMM procedure, 2930
beta-binomial distribution
  FMM procedure, 2930
binary distribution
  FMM procedure, 2930
binomial distribution
  FMM procedure, 2930

class level
  FMM procedure, 2909, 2960
confidence limits
  model parameters (FMM), 2930, 2939
constrained analysis
  FMM procedure, 2942
convergence criterion
  FMM procedure, 2904, 2907
convergence status
  FMM procedure, 2961

default output
  FMM procedure, 2960
Dirichlet-multinomial distribution
  FMM procedure, 2930

effect
  name length (FMM), 2909
effective sample sizes
  FMM procedure, 2964
examples, FMM
  binary data, sort order, 2910
  binomial data, 2970
  cattle feeding data, 2976
  housing satisfaction survey, 2989
  logistic model, binomial cluster, 2970
  multinomial model, 2990
  ossification data, 2969
  PROBMODEL specification, 2970
  salmonella assay, 2984
  three-component mixture, 2978
  Weibull distribution, 2978
exponential distribution
  FMM procedure, 2930

finite mixture models
  See FMM procedure, 2874
first stage convergence criterion
  FMM procedure, 2942
first stage iteration history
  FMM procedure, 2941
fit statistics
  FMM procedure, 2962
FMM procedure, 2874
  alpha level, 2930, 2939
  Bayes information, 2963
  Bayesian analysis, 2915
  Bayesian fit statistics, 2963
  Bernoulli distribution, 2930
  beta distribution, 2930
  beta-binomial distribution, 2930
  binary distribution, 2930
  binomial distribution, 2930
  centering and scaling, 2909
  class level, 2909, 2960
  compared to other procedures, 2877
  confidence limits, 2930, 2939
  constrained analysis, 2942
  convergence criterion, 2904, 2907
  convergence status, 2961
  default output, 2960
  Dirichlet-multinomial distribution, 2930
  effect name length, 2909
  effective sample sizes, 2964
  exponential distribution, 2930
  first stage convergence criterion, 2942
  first stage iteration history, 2941
  fit statistics, 2962
  folded normal distribution, 2930
  function-based convergence criteria, 2904, 2906
  gamma distribution, 2930
  Gaussian distribution, 2930
  generalized Poisson distribution, 2930
  geometric distribution, 2930
  gradient-based convergence criteria, 2904, 2907
  heavy-tailed density, 2874, 2932, 2948
  hurdle model, 2874, 2877, 2884

input data sets, 2905
introductory example, 2878
inverse Gaussian distribution, 2930
iteration details, 2908
iteration history, 2961
link function, 2934, 2939
lognormal distribution, 2930
maximum number of random starting sets, 2941
mixing probabilities, 2963
model information, 2960
multimodal density, 2874, 2892, 2893, 2948
multinomial cluster distribution, 2930
multinomial distribution, 2930
multithreading, 2938
negative binomial distribution, 2930
normal distribution, 2930
number of first stage starting sets, 2941
number of observations, 2960
number of second stage starting sets, 2941
ODS Graphics, 2912, 2967
ODS table names, 2964
offset variable, 2934
optimization information, 2961
ordering of CLASS variable levels, 2909
ordering of effects, 2909
overdispersion, 2874, 2876, 2878, 2884, 2886, 2945–2947, 2967, 2975, 2985, 2990
parameter estimates, 2963
parameterization, 2959
plots, 2876
Poisson distribution, 2930
posterior autocorrelations, 2964
posterior intervals, 2964
posterior summaries, 2964
prior distributions, 2963
random number seed, 2914
random start draw standard deviation, 2942
random starting values, 2940, 2959
residual variance tolerance, 2915
response level ordering, 2928
response profile, 2961
response variable options, 2928
restricted analysis, 2942
statistical graphics, 2967
*t* distribution, 2930
Weibull distribution, 2930
weighting, 2944
zero-inflated model, 2874, 2875, 2877, 2884, 2885, 2932, 2950
folded normal distribution
FMM procedure, 2930
frequency variable
FMM procedure, 2926
gamma distribution
FMM procedure, 2930
Gaussian distribution
FMM procedure, 2930
generalized Poisson distribution
FMM procedure, 2930
geometric distribution
FMM procedure, 2930

heavy-tailed density
finite mixture models (FMM), 2874, 2932, 2948
hurdle model
finite mixture models (FMM), 2874, 2877, 2884

inverse Gaussian distribution
FMM procedure, 2930
iteration details
FMM procedure, 2908
iteration history
FMM procedure, 2961

link function
FMM procedure, 2934, 2939
lognormal distribution
FMM procedure, 2930

maximum number of random starting sets
FMM procedure, 2941
mixing probabilities
FMM procedure, 2963
mixture model (FMM)
parameterization, 2959
model
information (FMM), 2960
multi-modal density
finite mixture models (FMM), 2874
multimodal density
finite mixture models (FMM), 2892, 2893, 2948
multinomial cluster distribution
FMM procedure, 2930
multinomial distribution
FMM procedure, 2930
multithreading
FMM procedure, 2938

negative binomial distribution
FMM procedure, 2930
normal distribution
FMM procedure, 2930
number of first stage starting sets
FMM procedure, 2941
number of observations
FMM procedure, 2960
number of second stage starting sets
FMM procedure, 2941

ODS Graphics
FMM procedure, 2912, 2967
offset variable
FMM procedure, 2934
optimization information
FMM procedure, 2961
options summary
BAYES statement, 2916
MODEL statement (FMM), 2928
PROC FMM statement, 2902
overdispersion
finite mixture models (FMM), 2874, 2876, 2878, 2884, 2886, 2945–2947, 2967, 2975, 2985, 2990

parameter estimates
FMM procedure, 2963
parameterization
FMM procedure, 2959
mixture model (FMM), 2959
plots
finite mixture models (FMM), 2876
Poisson distribution
FMM procedure, 2930
posterior autocorrelations
FMM procedure, 2964
posterior intervals
FMM procedure, 2964
posterior summaries
FMM procedure, 2964
prior distributions
FMM procedure, 2963
probability distributions
FMM procedure, 2930

random number seed
FMM procedure, 2914
random start draw standard deviation
FMM procedure, 2942
random starting values
FMM procedure, 2940, 2959
response level ordering
FMM procedure, 2928
response profile
FMM procedure, 2961
response variable options
FMM procedure, 2928
restricted analysis
FMM procedure, 2942
reverse response level ordering
FMM procedure, 2928

statistical graphics
FMM procedure, 2967

*t* distribution
FMM procedure, 2930

Weibull distribution
FMM procedure, 2930
weighting
FMM procedure, 2944

zero-inflated model
finite mixture models (FMM), 2874, 2875, 2877, 2884, 2885, 2932, 2950

# Syntax Index


ABSCONV option
    PROC FMM statement, 2904
ABSFCONV option
    PROC FMM statement, 2904
ABSGCONV option
    PROC FMM statement, 2904
ABSGTOL option
    PROC FMM statement, 2904
ABSTOL option
    PROC FMM statement, 2904
ALLITER option
    RANDSTART statement, 2941
ALLSTATS option
    OUTPUT statement (FMM), 2937
ALPHA= option
    MODEL statement (FMM), 2930
    PROBMODEL statement (FMM), 2939

BAYES statement
    FMM procedure, 2915
BETAPRIORPARMS option
    BAYES statement (FMM), 2917
BY statement
    FMM procedure, 2925

CINFO option
    PROC FMM statement, 2905
CL option
    MODEL statement (FMM), 2930
    PROBMODEL statement (FMM), 2939
CLASS statement
    FMM procedure, 2926
COMPINFO option
    PROC FMM statement, 2905
COMPONENTINFO option
    PROC FMM statement, 2905
CORR option
    PROC FMM statement, 2905
COV option
    PROC FMM statement, 2905
COVI option
    PROC FMM statement, 2905
CPUCOUNT= option
    PERFORMANCE statement (FMM), 2938
CRIT= option
    PROC FMM statement, 2905
CRITERION= option
    PROC FMM statement, 2905

DATA= option
    PROC FMM statement, 2905
DESCENDING option
    MODEL statement, 2928
DETAILS option
    PERFORMANCE statement (FMM), 2938
DIAGNOSTICS option
    BAYES statement (FMM), 2918
DIST= option
    MODEL statement (FMM), 2930
DISTRIBUTION= option
    MODEL statement (FMM), 2930
DIVISOR= option
    RESTRICT statement (FMM), 2944

EQUATE= option
    MODEL statement (FMM), 2932
ESTIMATE= option
    BAYES statement (FMM), 2920
EVENT= option
    MODEL statement, 2929
EXCLUDE= option
    PROC FMM statement, 2906
EXCLUSION= option
    PROC FMM statement, 2906

FCONV option
    PROC FMM statement, 2906
FCONV2 option
    PROC FMM statement, 2906
FITDETAILS option
    PROC FMM statement, 2907
FMM procedure, 2902
    BAYES statement, 2915
    FREQ statement, 2926
    ID statement, 2926
    MODEL statement, 2927
    OUTPUT statement, 2935
    PERFORMANCE statement, 2938
    PROBMODEL statement, 2939
    PROC FMM statement, 2902
    RANDSTART statement, 2940
    RESTRICT statement, 2942
    syntax, 2902
    WEIGHT statement, 2944
FMM procedure, BAYES statement, 2915
    BETAPRIORPARMS option, 2917
    DIAGNOSTICS option, 2918
    ESTIMATE= option, 2920

INITIAL= option, 2920
METROPOLIS option, 2920
MIXPRIORPARMS option, 2920
MUPRIORPARMS option, 2920
NBI= option, 2921
NMC= option, 2921
OUTPOST= option, 2921
PHIPRIORPARMS option, 2922
PRIOROPTIONS option, 2922
PRIOROPTS option, 2922
STATISTICS option, 2924
SUMMARIES option, 2924
THIN= option, 2925
THINNING= option, 2925
TIMEINC= option, 2925
FMM procedure, BY statement, 2925
FMM procedure, CLASS statement, 2926
TRUNCATE option, 2926
FMM procedure, FREQ statement, 2926
FMM procedure, ID statement, 2926
FMM procedure, MODEL statement, 2927
ALPHA= option, 2930
CL option, 2930
DESCENDING option, 2928
DIST= option, 2930
DISTRIBUTION= option, 2930
EQUATE= option, 2932
EVENT= option, 2929
K= option, 2933
KMAX= option, 2933
KMIN= option, 2933
KRESTART option, 2934
LABEL= option, 2934
LINK= option, 2934
NOINT option, 2934
NUMBER= option, 2933
OFFSET= option, 2934
ORDER= option, 2929
PARAMETERS option, 2934
PARMS option, 2934
REFERENCE= option, 2930
FMM procedure, OUTPUT statement, 2935
ALLSTATS option, 2937
keyword= option, 2935
NOVAR option, 2937
OUT= option, 2935
PREDTYPE option, 2937
FMM procedure, PERFORMANCE statement, 2938
CPUCOUNT option, 2938
DETAILS option, 2938
NOTHREADS option, 2938
THREADS option, 2938
FMM procedure, PROBMODEL statement, 2939
ALPHA= option, 2939
CL option, 2939
LINK= option, 2939
NOINT option, 2940
PARAMETERS option, 2940
PARMS option, 2940
FMM procedure, PROC FMM statement, 2902
ABSCONV option, 2904
ABSFCONV option, 2904
ABSFTOL option, 2904
ABSGCONV option, 2904
ABSGTOL option, 2904
ABSTOL option, 2904
CINFO option, 2905
COMPINFO option, 2905
COMPONENTINFO option, 2905
CORR option, 2905
COV option, 2905
COVI option, 2905
CRIT= option, 2905
CRITERION= option, 2905
DATA= option, 2905
EXCLUDE= option, 2906
EXCLUSION= option, 2906
FCONV option, 2906
FCONV2 option, 2906
FITDETAILS option, 2907
FTOL option, 2906
FTOL2 option, 2906
GCONV option, 2907
GTOL option, 2907
HESSIAN option, 2908
INVALIDLOGL= option, 2908
ITDETAILS option, 2908
MAXFUNC= option, 2908
MAXITER= option, 2908, 2909
MAXTIME= option, 2908
MEMBERSHIP= option, 2911
NAMELEN= option, 2909
NOCENTER option, 2909
NOCLPRINT option, 2909
NOITPRINT option, 2909
NOPRINT option, 2909
ORDER= option, 2909
PARMSTYLE= option, 2911
PARTIAL= option, 2911
PLOTS option, 2912
SEED= option, 2914
SINGCHOL= option, 2915
SINGULAR= option, 2915
TECHNIQUE= option, 2915
ZEROPROB= option, 2915
FMM procedure, RANDSTART statement, 2940–2942
FMM procedure, RESTRICT statement, 2942
DIVISOR= option, 2944

FMM procedure, WEIGHT statement, 2944
FREQ statement
FMM procedure, 2926
FTOL option
PROC FMM statement, 2906
FTOL2 option
PROC FMM statement, 2906

GCONV option
PROC FMM statement, 2907
GTOL option
PROC FMM statement, 2907

HESSIAN option
PROC FMM statement, 2908

ID statement
FMM procedure, 2926
INITIAL= option
BAYES statement (FMM), 2920
INVALIDLOGL= option
PROC FMM statement, 2908
ITDETAILS option
PROC FMM statement, 2908

K= option
MODEL statement (FMM), 2933
keyword= option
OUTPUT statement (FMM), 2935
KMAX= option
MODEL statement (FMM), 2933
KMIN= option
MODEL statement (FMM), 2933
KRESTART option
MODEL statement (FMM), 2934

LABEL= option
MODEL statement (FMM), 2934
LINK= option
MODEL statement (FMM), 2934
PROBMODEL statement (FMM), 2939

MAXFIRST= option
RANDSTART statement, 2941
MAXFUNC= option
PROC FMM statement, 2908
MAXITER= option
PROC FMM statement, 2908, 2909
MAXTIME= option
PROC FMM statement, 2908
MEMBERSHIP= option
PROC FMM statement, 2911
METROPOLIS option
BAYES statement (FMM), 2920
MIXPRIORPARMS option
BAYES statement (FMM), 2920
MODEL statement
FMM procedure, 2927
MUPRIORPARMS option
BAYES statement (FMM), 2920

NAMELEN= option
PROC FMM statement, 2909
NBI= option
BAYES statement (FMM), 2921
NFIRST= option
RANDSTART statement, 2941
NMC= option
BAYES statement (FMM), 2921
NOCENTER option
PROC FMM statement, 2909
NOCLPRINT option
PROC FMM statement, 2909
NOINT option
MODEL statement (FMM), 2934
PROBMODEL statement (FMM), 2940
NOITPRINT option
PROC FMM statement, 2909
NOPRINT option
PROC FMM statement, 2909
NOTHREADS option
PERFORMANCE statement (FMM), 2938
NOVAR option
OUTPUT statement (FMM), 2937
NSECOND= option
RANDSTART statement, 2941
NUMBER= option
MODEL statement (FMM), 2933

OFFSET= option
MODEL statement (FMM), 2934
ORDER= option
MODEL statement, 2929
PROC FMM statement, 2909
OUT= option
OUTPUT statement (FMM), 2935
OUTPOST= option
BAYES statement (FMM), 2921
OUTPUT statement
FMM procedure, 2935

PARAMETERS option
MODEL statement (FMM), 2934
PROBMODEL statement (FMM), 2940
PARMS option
MODEL statement (FMM), 2934
PROBMODEL statement (FMM), 2940
PARMSTYLE= option
PROC FMM statement, 2911
PARTIAL= option

PROC FMM statement, 2911
PERFORMANCE statement
FMM procedure, 2938
PHIPRIORPARMS option
BAYES statement (FMM), 2922
PLOTS option
PROC FMM statement, 2912
PREDTYPE option
OUTPUT statement (FMM), 2937
PRIOROPTIONS option
BAYES statement (FMM), 2922
PRIOROPTS option
BAYES statement (FMM), 2922
PROBMODEL statement
FMM procedure, 2939
PROC FMM procedure, PROC FMM statement
SINGRES= option, 2915
PROC FMM statement
FMM procedure, 2902

RANDGCONV= option
RANDSTART statement, 2942
RANDSTART statement
FMM procedure, 2940
REFERENCE= option
MODEL statement (FMM), 2930
RESTRICT statement
FMM procedure, 2942

SEED= option
PROC FMM statement, 2914
SINGCHOL= option
PROC FMM statement, 2915
SINGRES= option
PROC FMM statement, 2915
SINGULAR= option
PROC FMM statement, 2915
STATISTICS option
BAYES statement (FMM), 2924
STD= option
RANDSTART statement, 2942
SUMMARIES option
BAYES statement (FMM), 2924
syntax
FMM procedure, 2902

TECHNIQUE= option
PROC FMM statement, 2915
THIN= option
BAYES statement (FMM), 2925
THINNING= option
BAYES statement (FMM), 2925
THREADS= option
PERFORMANCE statement (FMM), 2938
TIMEINC= option
BAYES statement (FMM), 2925
TRUNCATE option
CLASS statement (FMM), 2926

WEIGHT statement
FMM procedure, 2944

ZEROPROB= option
PROC FMM statement, 2915