TRAVELERS

## Finite Mixtures for Insurance Modeling


Matt Flynn

mjflynn@travelers.com 860-954-0894



---

### Outline - Finite Mixture Models (FMM)

- JMP 9 Distribution Platform – finite mixtures
- Interactive JMP Two-Component Normal mixture
- R – two packages - flexmix, gamlss
- SAS – Proc NLMIXED
- JMP's Nonlinear Platform
- STATA FMM module
- More Examples – Poisson counts, WC Losses

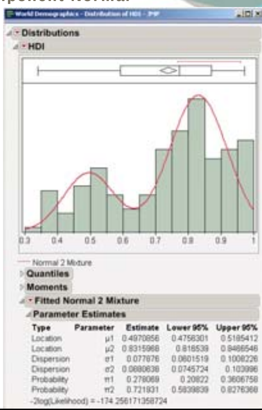


---

### Outline - Finite Mixture Models (FMM)

- FMM Background

## JMP 9 includes finite 2,3+ component Normal mixtures




JMP Sample Data

UN Health Development Index

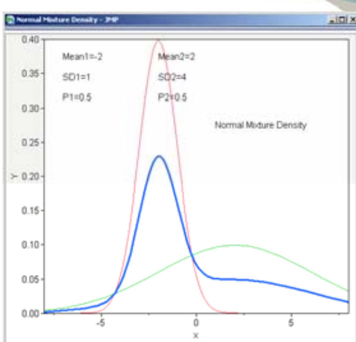
Health, Education, Living standards

http://hdr.undp.org/en/statistics/hdi/

TRAVELERS



## Interactive JMP Two-Component Normal mixture




C:\Documents and Settings\mjflynn\My Documents\JMP9\Normal2Mixture_dist.jsl

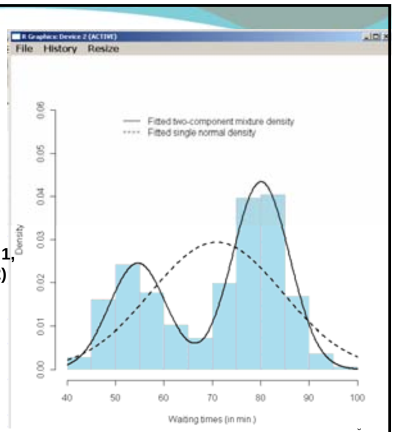
TRAVELERS



## Via R

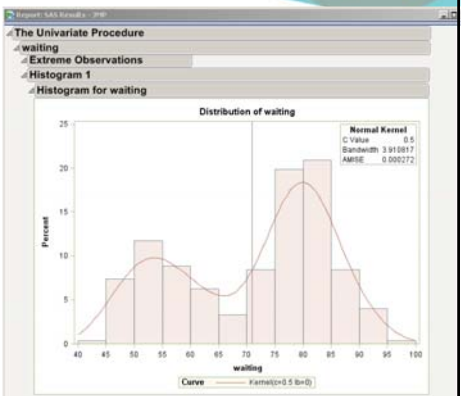
```
library(gamlss);
library(gamlss.mx);
m2 <- gamlssMX(
      waiting ~ 1,
     data=faithful,
       family=NO,
       k=2); m2

library("flexmix")
fl <- flexmix(waiting ~ 1,
   data = faithful, k = 2)
```




TRAVELERS

## Via SAS Proc UNIVARIATE




TRAVELERS



## Via SAS – obtain starting values

```
/*  two-component normal mixture */
proc sql;
select log(mean(waiting)-0.5*var(waiting)**0.5) as mu1start,
          log(mean(waiting)+0.5*var(waiting)**0.5) as mu2start
                into :mu1start, :mu2start
                      from faithful;
quit;
```

TRAVELERS



## Via SAS – obtain starting values

```
/*  two-component normal mixture */
proc sql;
select log(mean(waiting)-0.5*var(waiting)**0.5) as mu1start,
          log(mean(waiting)+0.5*var(waiting)**0.5) as mu2start
                into :mu1start, :mu2start
                      from faithful;
quit;
```

Create SAS Macro variables – note: separation

TRAVELERS

## Via SAS

```
Proc NLMIXED data=faithful;
 parms eta_mu1=&mu1start. eta_mu2=&mu2start. eta_sigma1=1.8
eta_sigma2=1.8 eta_p1=0.57 ;

  mu1 = exp(eta_mu1);
  mu2 = exp(eta_mu2);
  sigma1 = exp(eta_sigma1);
  sigma2 = exp(eta_sigma2);
  p1 = exp(eta_p1)/(1 + exp(eta_p1));
  p2 = 1 - p1;
  y = waiting;

  loglike = logpdf('NORMALMIX', y, 2, p1, p2, mu1, mu2, sigma1, sigma2) ;

  model y ~ general(loglike);
  estimate 'mu1' mu1;  estimate 'mu2' mu2;
  estimate 'sigma1' sigma1;  estimate 'sigma2' sigma2;
  estimate 'p1' p1;  estimate 'p2' p2;
run;
```



## Via SAS

Starting values (from above)

Log link functions

Normal 2 – Component Finite Mixture logLikelihood

```
Proc NLMIXED data=faithful;
 parms eta_mu1=&mu1start. eta_mu2=&mu2start. eta_sigma1=1.8
eta_sigma2=1.8 eta_p1=0.57 ;

  mu1 = exp(eta_mu1);
  mu2 = exp(eta_mu2);
  sigma1 = exp(eta_sigma1);
  sigma2 = exp(eta_sigma2);
  p1 = exp(eta_p1)/(1 + exp(eta_p1));
  p2 = 1 - p1;
  y = waiting;

  loglike = logpdf('NORMALMIX', y, 2, p1, p2, mu1, mu2, sigma1, sigma2) ;
*loglike = logpdf('NORMAL', y, mu1, sigma1)*p1 +
               (1 - p1)*logpdf('NORMAL', y, mu2, sigma2);
  model y ~ general(loglike);
  estimate 'mu1' mu1;  estimate 'mu2' mu2;
  estimate 'sigma1' sigma1;  estimate 'sigma2' sigma2;
  estimate 'p1' p1;  estimate 'p2' p2;
run;
```



## Via SAS NLMIXED

Report: SAS Results - JMP

**The Nlmixed Procedure**

**Iteration History**

| Iter | Calls | NegLogLike | Diff | MaxGrad | Slope |
|---|---|---|---|---|---|

**Convergence Status**

NOTE: GCONV convergence criterion satisfied.

**Fit Statistics**

| Description | Value |
|---|---|
| -2 Log Likelihood | 2068 |
| AIC (smaller is better) | 2078 |
| AICC (smaller is better) | 2078.23 |
| BIC (smaller is better) | 2096.03 |

**Parameter Estimates**

| Parameter | Estimate | Standard Error | DF | t Value | Pr > \|t\| | Alpha | Lower | Upper | Gradient |
|---|---|---|---|---|---|---|---|---|---|
| eta_mu1 | 4.00031 | 0.01249 | 272 | 320.309 | <.0001* | 0.05 | 3.97572 | 4.02489 | 0.00936 |
| eta_mu2 | 4.38316 | 0.00624 | 272 | 702.296 | <.0001* | 0.05 | 4.37088 | 4.39545 | -0.0024 |
| eta_sigma1 | 1.77008 | 0.08644 | 272 | 20.4776 | <.0001* | 0.05 | 1.59989 | 1.94024 | -0.0012 |
| eta_sigma2 | 1.76945 | 0.06821 | 272 | 25.9409 | <.0001* | 0.05 | 1.63517 | 1.90374 | -0.0032 |
| eta_p1 | -0.5715 | 0.09321 | 272 | -6.1314 | <.0001* | 0.05 | -0.755 | -0.388 | -0.0002 |

**Additional Estimates**

| Label | Estimate | Standard Error | DF | t Value | Pr > \|t\| | Alpha | Lower | Upper |
|---|---|---|---|---|---|---|---|---|
| mu1 | 54.615 | 0.68208 | 272 | 80.0712 | <.0001* | 0.05 | 53.2721 | 55.9578 |
| mu2 | 80.0911 | 0.49986 | 272 | 160.226 | <.0001* | 0.05 | 79.107 | 81.0752 |
| sigma1 | 5.87121 | 0.5075 | 272 | 11.5689 | <.0001* | 0.05 | 4.87208 | 6.87034 |
| sigma2 | 5.86765 | 0.40024 | 272 | 14.6604 | <.0001* | 0.05 | 5.07969 | 6.65561 |
| p1 | 0.36089 | 0.0215 | 272 | 16.7863 | <.0001* | 0.05 | 0.31856 | 0.40321 |
| p2 | 0.63911 | 0.0215 | 272 | 29.7276 | <.0001* | 0.05 | 0.59679 | 0.68144 |

## Via JMP – nonlinear platform - setup

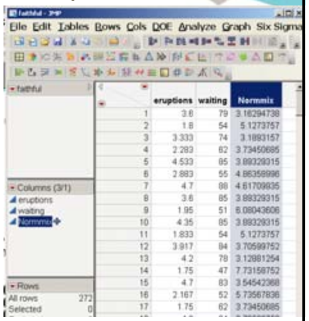
```
dt = Current Data Table();

// set up the  negative log likelihood with
// starting values
ll = dt << new column("Normmix");
Exform = expr(
ll << set formula(Parameter(
{ eta_mu1=4.160438,
  eta_mu2=4.352785,  eta_sigma1=1.8,
  eta_sigma2=1.8, eta_p1=-0.57 },

mu1 = exp(eta_mu1);
mu2 = exp(eta_mu2);
sigma1 = exp(eta_sigma1);
sigma2 = exp(eta_sigma2);

p1 = exp(eta_p1)/(1 + exp(eta_p1));
p2 = 1 - p1;

-log( Normal Mixture Density( :waiting,
        mu1 |/ mu2,  sigma1 |/ sigma2, p1 |/ p2 ) )
)
));
eval(eval Expr(exform));
```



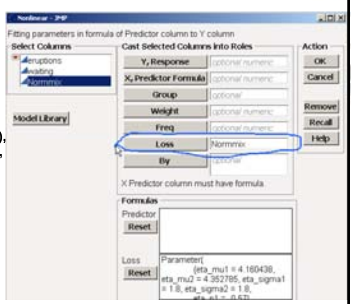

| | eruptions | waiting | Normmix |
|---|---|---|---|
| 1 | 3.6 | 79 | 3.16294738 |
| 2 | 1.8 | 54 | 5.1273757 |
| 3 | 3.333 | 74 | 3.1893157 |
| 4 | 2.283 | 62 | 3.73450685 |
| 5 | 4.533 | 85 | 3.89329315 |
| 6 | 2.883 | 55 | 4.86358996 |
| 7 | 4.7 | 88 | 4.61709835 |
| 8 | 3.6 | 85 | 3.89329315 |
| 9 | 1.95 | 51 | 6.08043606 |
| 10 | 4.35 | 85 | 3.89329315 |
| 11 | 1.833 | 54 | 5.1273757 |
| 12 | 3.917 | 84 | 3.70599752 |
| 13 | 4.2 | 78 | 3.12881254 |
| 14 | 1.75 | 47 | 7.73158752 |
| 15 | 4.7 | 83 | 3.54542368 |
| 16 | 2.167 | 52 | 5.73567836 |
| 17 | 1.75 | 62 | 3.73450685 |

TRAVELERS



## Via JMP
## Nonlinear Platform

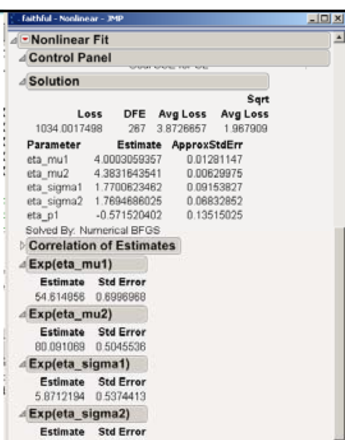
```
nl = Nonlinear(
  Loss( :Normmix ),
  Numeric Derivatives Only( 1 ),
  Loss is Neg LogLikelihood( 1 ),
  QuasiNewton BFGS,
  Finish,
  Custom Estimate( exp(eta_mu1)  ),
  Custom Estimate( exp(eta_mu2) ),
  Custom Estimate( exp(eta_sigma1)  ),
  Custom Estimate( exp(eta_sigma2) ),
  Custom Estimate(
    exp(eta_p1)/(1 + exp(eta_p1))  ),
  Custom Estimate(
     1 - exp(eta_p1)/(1 + exp(eta_p1))  ),
);
```




TRAVELERS



## Via JMP
## Nonlinear Platform output




| Loss | DFE | Avg Loss | Sqrt Avg Loss |
|---|---|---|---|
| 1034.0017498 | 267 | 3.8726657 | 1.967909 |

| Parameter | Estimate | ApproxStdErr |
|---|---|---|
| eta_mu1 | 4.0003059357 | 0.01281147 |
| eta_mu2 | 4.3831643541 | 0.00629975 |
| eta_sigma1 | 1.7700623462 | 0.09153827 |
| eta_sigma2 | 1.7694686025 | 0.06832852 |
| eta_p1 | -0.571520402 | 0.13515025 |

| | Estimate | Std Error |
|---|---|---|
| Exp(eta_mu1) | 54.614856 | 0.6996968 |
| Exp(eta_mu2) | 80.091069 | 0.5045536 |
| Exp(eta_sigma1) | 5.8712194 | 0.5374413 |
| Exp(eta_sigma2) | 5.8677344 | 0.4009336 |

TRAVELERS

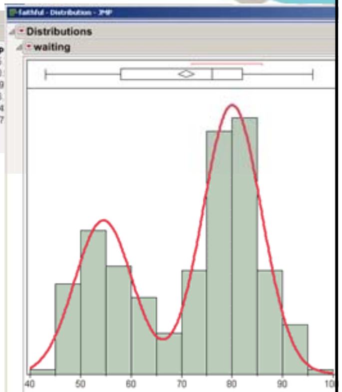
## Via JMP – Analyze, Distribution

**Fitted Normal 2 Mixture**

**Parameter Estimates**

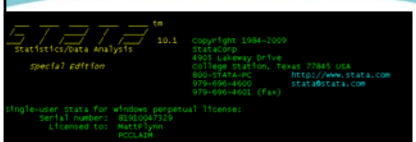
| Type | Parameter | Estimate | Lower 95% | Upp |
|---|---|---|---|---|
| Location | μ1 | 54.590831 | 53.460731 | 55 |
| Location | μ2 | 80.08455 | 79.223944 | 80. |
| Dispersion | σ1 | 5.7085839 | 4.7138081 | 6.9 |
| Dispersion | σ2 | 5.7916822 | 5.0062799 | 6. |
| Probability | π1 | 0.3603723 | 0.2925049 | 0.4 |
| Probability | π2 | 0.6396277 | 0.5402665 | 0.7 |

-2log(Likelihood) = 2068.18823356317




TRAVELERS

---




This file is part of IDEAS, which uses RePEc data

### FMM: Stata module to estimate finite mixture models

Author info | Abstract | Publisher info | Download info | Related research | Statistics

**Author Info**

Partha Deb (partha.deb@hunter.cuny.edu) (Hunter College)

Additional information is available for the following registered author(s):

- Partha Deb

**Abstract**

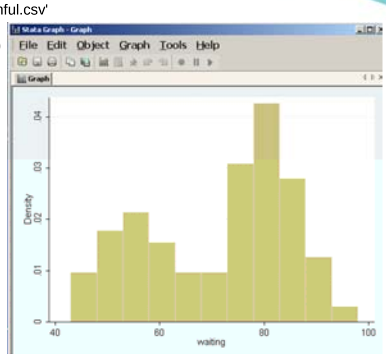
fmm fits a finite mixture regression model using maximum likelihood estimation. The model is a J-component finite mixture of densities, with the density within a class (j) allowed to vary in location and scale. Optionally, the mixing probabilities may be specified with covariates. fmm currently fits mixtures of the following distributions: Gamma, Normal (Gaussian), Negative binomial 1 & 2 (mean and constant dispersion), Poisson, Student-t (with fixed degrees of freedom)

TRAVELERS

---

## Via STATA - FMM

```
insheet using 'C:/temp/faithful.csv'
summarize
histogram waiting, width(5)
```




TRAVELERS

18

## Via STATA - FMM

insheet using 'C:/temp/faithful.csv'
summarize
histogram waiting, width(5)
**fmm** waiting, components(2) mixtureof(normal)

```
2 component Normal regression                   Number of obs   =        272
                                                Wald chi2(0)    =          .
Log likelihood = -1034.0017                     Prob > chi2     =          .

     waiting |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
component1   |
       _cons |   54.61486    .699675    78.06   0.000     53.24352    55.98619
-------------+----------------------------------------------------------------
component2   |
       _cons |   80.09107   .5045947   158.72   0.000     79.10208    81.08006
-------------+----------------------------------------------------------------
 /imlogitpi1 |  -.5715204   .1351187    -4.23   0.000    -.8363481   -.3066927
   /lnsigma1 |   1.770062    .091518    19.34   0.000      1.59069    1.949434
   /lnsigma2 |   1.769469   .0683333    25.89   0.000     1.635538    1.903399
-------------+----------------------------------------------------------------
      sigma1 |   5.871219   .5373224                      4.907135    7.024713
      sigma2 |   5.867734   .4009615                      5.132218    6.708661
         pi1 |   .3608861   .0311648                      .3023045    .4239222
         pi2 |   .6391139   .0311648                      .5760778    .6976955
```

TRAVELERS

## Time Permitting – Additional Examples

**Proc FMM.sas – FMM(2) Poisson – Counts - regressors**

**Exp_mix.sas – FMM Exponential, Gamma dists**

**WC_Loss.sas – FMM Gamma with regressors**

TRAVELERS

20

## SAS recently announced experimental Proc FMM coming in SAS/STAT 9.3


TRAVELERS

21

## Further reading:


Deb, Partha and J. F. Burgess Jr., A quasi-experimental comparison of statistical models for health care expenditures, 2003, wp, http://urban.hunter.cuny.edu/RePEc/htr/papers/debburgess10.pdf

Grun, Bettina and Friedrich Leisch, Fitting Finite Mixtures of Generalized Linear Regressions in R, Computational Statistics and Data Analysis, 2006, http://statmath.wu.ac.at/projects/AASC/mixtures/Gruen+Leisch-2007b.pdf

Klugman, Stuart and Jacques Rioux, Toward a unified approach to fitting loss models, North American Actuarial Journal, Jan-06, 10, 1, 63-83, http://www.iowaactuariesclub.org/library/lossmodels.pdf

Lee, Andy H., Kui Wang, Kelvin K.W. Yau, Geoffrey J. McLachlan and S.K. Ng
Maternity length of stay modeling by gamma mixture regression with random effects
Biometrical Journal, Aug-2007, v49, n5, p750-764
http://www.maths.uq.edu.au/~gjm/lwymn_biomj07.doc

Leisch, Friederich and Bettina Gruen, "FlexMix Version 2: Finite mixtures with concomitant variables and varying and constant parameters", Journal of Statistical Software, 2007, 28(4), 1-35, http://cran.r-project.org/web/packages/flexmix/vignettes/mixture-regressions.pdf

Park, Byung-Jung and Dominique Lord, Application of Finite Mixture Models for Vehicle Crash Data Analysis, wp, Feb-2009, https://ceprofs.civil.tamu.edu/dlord/papers/park_lord_%20finite_mixture_model.pdf

Rempala, Grzegorz A. and Richard A. Derrig, Modeling Hidden Exposures in Claim Severity via the EM Algorithm, ASTIN Colloquia - Bergen , Norway Jun-2004, http://www.actuaries.org/ASTIN/Colloquia/Bergen/Rempala_Derrig.pdf


TRAVELERS



## Further reading:


Stokes, Maura E., Fang Chen, and Ying So, On Deck SAS/STAT 9.3, SAS Global Forum, 2011, 331, http://support.sas.com/resources/papers/proceedings11/331-2011.pdf

Teodorescu, Sandra, Different approaches to model the loss distribution of a real data set from motor third party liability insurance, Romanian Journal of Insurance, Apr-2010, 93-104, http://www.ima-imi.ro/en/publications/assets/pdf/Romanian%20Journal%20of%20Insurance%20Year%202010%20No.4.pdf#page=94


TRAVELERS



## Thank you – Questions?


Contact info:

Matt Flynn – Travelers

mjflynn@travelers.com

860.954.0894



TRAVELERS