A NEW SAEM ALGORITHM FOR ORDERED-CATEGORICAL AND COUNT DATA MODELS: IMPLEMENTATION AND EVALUATION

Marc Lavielle[1] and Radojka Savic[2]

(1) INRIA Saclay and University Paris 11

(2) Currently Postdoc at INSERM U738 and University Paris 7

OUTLINE

- Discrete data models
- Performance of existing algorithms
- The SAEM algorithm in MONOLIX
- Simulation examples
- Real data example
- Conclusions

CATEGORICAL AND COUNT DATA

- Often PD outcome of clinical trials

- Examples of categorical outcomes:
  - ✓ Pain scale (severe, moderate, mild, none)
  - ✓ Common toxicity criteria (Grade 0-4)
  - ✓ Adverse reactions (Yes/No)

- Examples of count outcomes:

Measured outcome within a time frame

  - ✓ Daily counts of epileptic seizuers
  - ✓ Weekly counts of hot flashes
  - ✓ Hourly counts of emetic episodes

CATEGORICAL DATA MODELS

- Proportional odds model (K categories)
  - Modeling of cumulative probabilities
    (e.g. P(Y>1), P(Y>2), P(Y>K-1))
  - From cumulative probabilities, all probabilities can be assessed
  - Parameterization so that P(Y>1) > P(Y>2)

$$k=1,\ldots,K\text{-}1.$$

$$\text{logit}\left[P(y_{ij} \geq k)\right] = \alpha_k + \beta x_{ij} + \eta_i$$

α_k - baseline probabilities for each category

β - an effect that is the same for all categories, e.g. E_{max} and EC_{50}

x_{ij} - the predictor vector, e.g. concentrations, dose, time

COUNT DATA MODELS

1. Poisson model: $\psi_i = \lambda_i$

$$\mathbb{P}_{\psi_i}(y_{ij} = k) = \frac{e^{-\lambda_i}\lambda_i^k}{k!}$$

y_{ij} - observation
k - count
ψ_i - inividual parameter = ψ*exp(η_i)

2. Poisson model with Markov elements: $\psi_i = (\lambda_{1i}, \lambda_{2i})$
3. Mixture Poisson model: $\psi_i = (\lambda_{1i}, \lambda_{2i}, p_i)$
4. Zero inflated Poisson model: $\psi_i = (\lambda_i, p_i)$
5. Negative Binomial model: $\psi_i = (\lambda_i, d_i)$
6. Generalized Poisson model: $\psi_i = (\lambda_i, \delta_i)$

ESTIMATION ALGORITHMS

Likelihood approximations:

- LAPLACE[1] (NONMEM, SAS)
  - ✓ Linearization using LAPLACE (second order Taylor expansion)
  - ✓ Unstable, problems with initial values

[1]Wolfinger (1993). Laplace's approximation for nonlinear mixed models. *Biometrika*, 80:791-5

- Gaussian quadrature[2,3,4] (SAS)
  - ✓ Integration of the likelihood by Adaptive Gaussian Quadrature
  - ✓ Limited to models with small number of random effects

[2]Pinheiro & Bates (1995). Approximations to the Log-Likelihood function in the nonlinear mixed-effects model. J Comput Graph Stat, 1:12-35

[3]Guedj, Thiebaut & Commenges (2007). Maximum likelihood estimation in dynamical models of HIV. Biometrics, 63: 1198-1206

[4]Molenberghs & Verbeke (2005). Models for discrete longitudinal data, Springer

Performance of the algorithms: Categorical & count data

Estimating Bias in Population Parameters for Some Models for Repeated Measures Ordinal Data using NONMEM and NLMIXED

Siv Jönsson,[1,*] Maria C. Kjellsson,[1] and Mats O. Karlsson[1]

J Pharmacokinet Pharmacodyn. 31:299-320 (2004)

Maximum Likelihood Approximations: Performance in Population Models for Count Data

Elodie L. Plan (1), Alan Maloney (1,2), Iñaki F. Trocóniz (3), Mats O. Karlsson (1)

In review - J Pharmacokinet Pharmacodyn.

Performance of the algorithms: Categorical & count data

Method	Data	Estimation properties	Drawbacks
LAPLACE	Categorical	Good performance if all categories are represented in similar proportions	Bias in fixed and random effects (>100%) if categories are skewed
	Count	Good performance if equidispersion	Bias in variance parameters if heterodispersion (<26%)
Adaptive Gaussian quadrature	Categorical	Good performance	Long run times
	Count	Good performance	Low stability

THE SAEM ALGORITHM in MONOLIX

- It is known to be accurate, precise and fast for continuous data models[1,2,3]
- No studies have been performed on discrete data models

Aim: To implement and investigate the performance of a new SAEM algorithm for discrete data

1. E. Kuhn, Lavielle, M. . Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics and Data Analysis. 49:1020 - 1038 (2005)

2. R.J. Bauer, S. Guzy, and C. Ng. A survey of population analysis methods and software for complex pharmacokinetic and pharmacodynamic models with examples. AAPS J. 9:E60-83 (2007)

3. P. Girard and F. Mentre. A comparison of estimation methods in nonlinear mixed effects models using a blind analysis, *PAGE*, Pamplona, Spain, 2005. PAGE 14 (2005) Abstr 834 [www.page-meeting.org/?abstract=834]

METHODS

- SAEM was implemented in working version of MONOLIX 3.1.
- Implementation both for population parameter estimation and estimation of the Fisher Information Matrix
- MC simulations according to the same scenarios and using the same models as in previous studies

SCENARIOS

- Categorical data:
  - ✓ Proportional odds model
  - ✓ 5 different scenarios - baseline
    - placebo model
    - drug model)
  - ✓ Each scenario with several subscenarios (different OMEGAs)
  - ✓ 1000 IDs with 4 obs/ID

- Count data:
  - ✓ 6 different models: Poisson,
    Negative binomial,
    Generalized Poisson,
    Zero inflated Poisson,
    Poisson with Markov elements,
    Mixture Poisson model
  - ✓ One scenario for each model
  - ✓ 551 IDs and 84 obs/ID

ESTIMATION ERRORS

100 simulations used for each scenario

- Parameters

For each parameter, the relative estimation error (REE) (%) is computed as follows for k=1,2,...100:

$$REE_k(\theta) = 100 \times \frac{\hat{\theta}_k - \theta^*}{\theta^*}$$

$\hat{\theta}_k$ - Parameter estimate using kth simulated file

θ^* - True parameter

- Standard errors

For each standard error, absolute estimation error (AEE) (%) is computed as:

$$AEE_{(k)} = \widehat{RSE}_{(k)} - RSE^*$$

$\widehat{RSE}_{(k)}$ - Estimated RSE

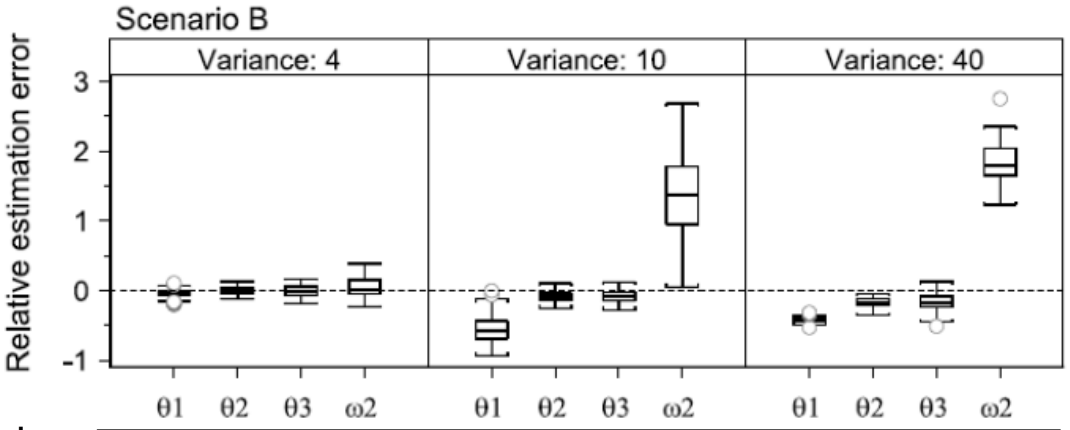
RSE^* - Empirical RSE – empirical standard deviation of the estimated parameter

RESULTS: PROPORTIONAL ODDS MODEL

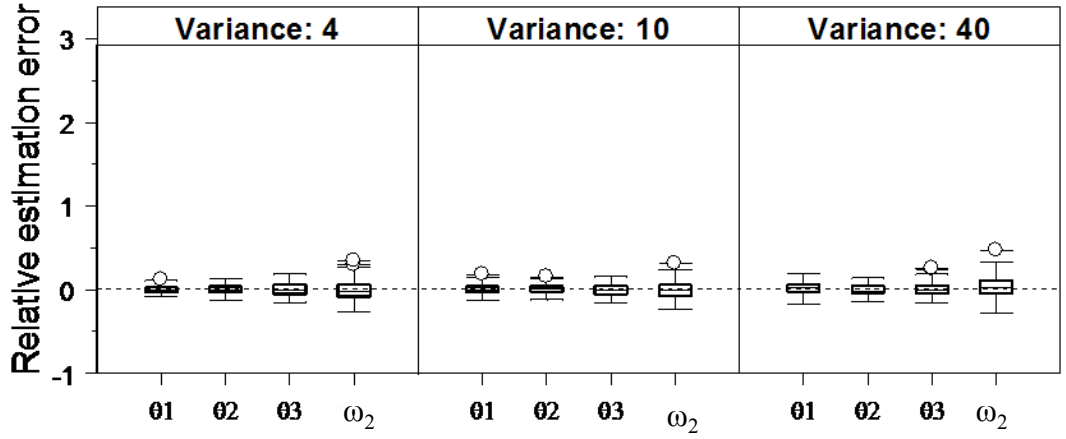
Scenario B:

- ✓Baseline model
- ✓Data with 4 categories
- ✓Proportions of observation equal to category 0/1/2/3 at baseline: 82.5/10/5/2.5

LAPLACE in NONMEM[1]



SAEM in MONOLIX



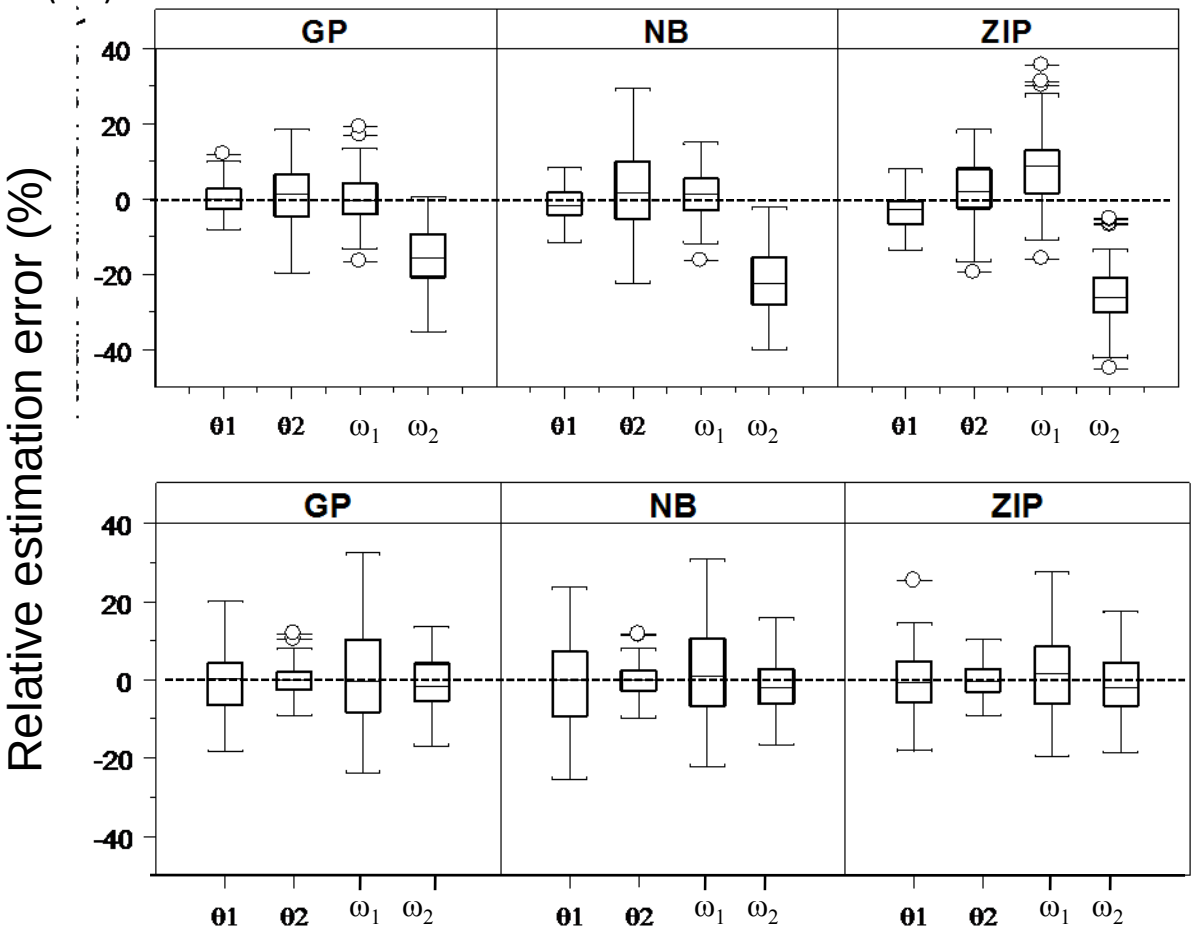
RESULTS: COUNT DATA MODEL

Models: Handlng overdispersion/underdispersion

- ✓Generalized Poisson (GP)
- ✓Negative binomial (NB)
- ✓Zero-inflated Poisson (ZIP)

LAPLACE in NONMEM[1]

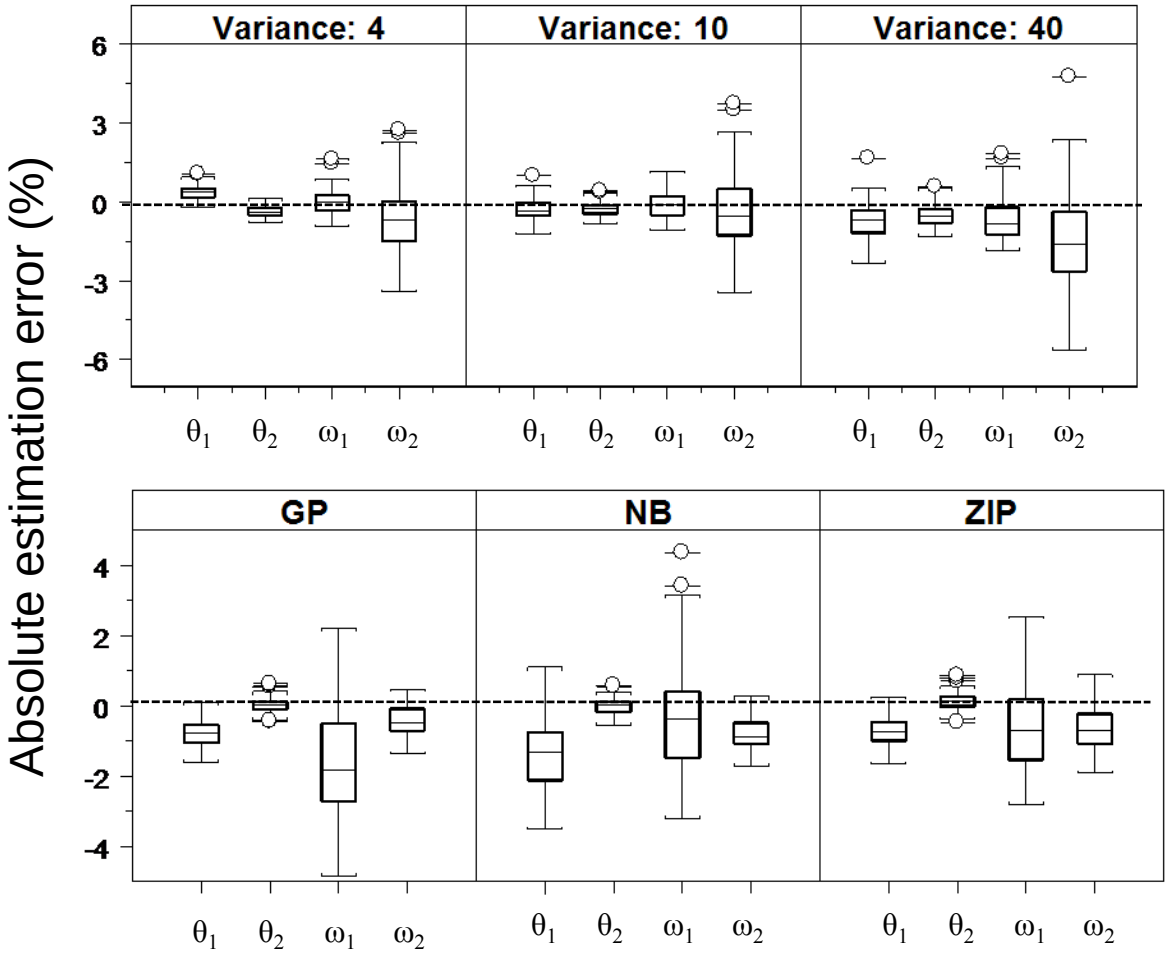
SAEM in MONOLIX



RESULTS:SAEM IN MONOLIX RELATIVE SE'S

Categorical data

Count data



RUN TIMES: NONMEM, SAS & SAEM

Model	LAPLACE in NONMEM*	Adaptive GQ* in SAS (1 point)	Adaptive GQ* in SAS (9 points)	SAEM in MONOLIX (C++)
ZIP	2min 32s	60s	12min 51s	20s + 19s (s.e.)
GP	1min 44s	59s	13min 37s	24s + 21s (s.e.)
NB	2min 40s	4min 27s	60min	27s + 25s (s.e.)

Models: Handlng overdispersion/underdispersion

- ✓Generalized Poisson (GP)
- ✓Negative binomial (NB)
- ✓Zero-inflated Poisson (ZIP)

*Run times for NONMEM and SAS kindly generated and provided by Plan E. and Maloney A.

MONOLIX IMPLEMENTATION (ORDERED CATEGORICAL AND COUNT DATA EXAMPLE)

Proportional odds model

$PROBLEM ; Ordered categorical model
$PSI th1 th2 th3 th4 th5
$REG OCC DOSE

$CATEGORICAL(0,3)
LOGIT1(Y>=1) = -th1 - th4*OCC - th5*DOSE
LOGIT1(Y>=2) = -th1 - th4*OCC - th5*DOSE - th2
LOGIT1(Y>=3) = -th1 - th4*OCC - th5*DOSE - th2 - th3

$OUTPUT
OUTPUT1 = LL1

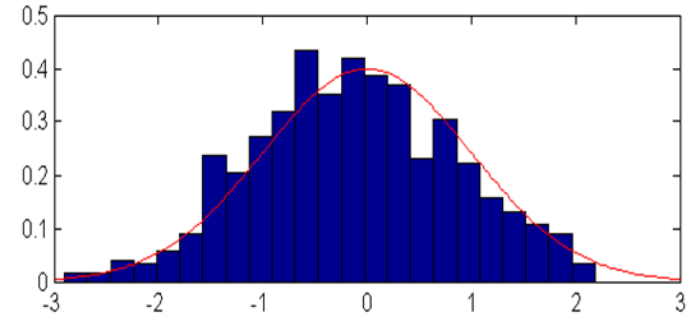
Poisson model

$PROBLEM ; Basic Poisson model
$PSI lambda
$COUNT
LL1(Y=k) = -lambda + k*log(lambda) - factln(k)

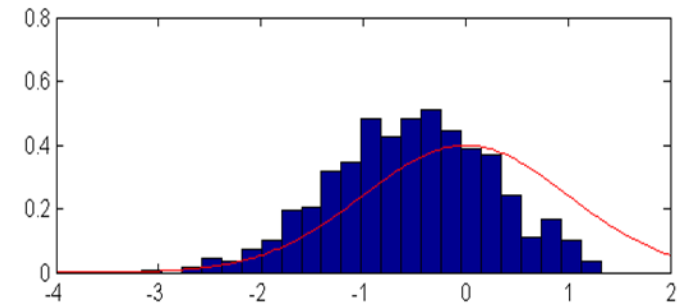
$OUTPUT
OUTPUT1 = LL1

Model Diagnostics: NPDEs

The correct model
Variance = TRUE



The misspecified model
Variance = 1.5 inflated



CATEGORICAL / COUNT DATA & SAEM

Overall performance:

- Parameter estimation:

Accurate: Bias < 8.13 % for all studied scenarios including ones with skewed distributions of response categories

Precise: RMSE < 30%

- Standard error estimation:

Accurate: AEE < 5.8%

Precise: RMSE< 5.6 %

- CPU time:

Fast: average CPU time 35s for parameter estimation and 17s for SE estimation)

- Convergence: Stable: 100% convergence
- Implementation: Easy to implement

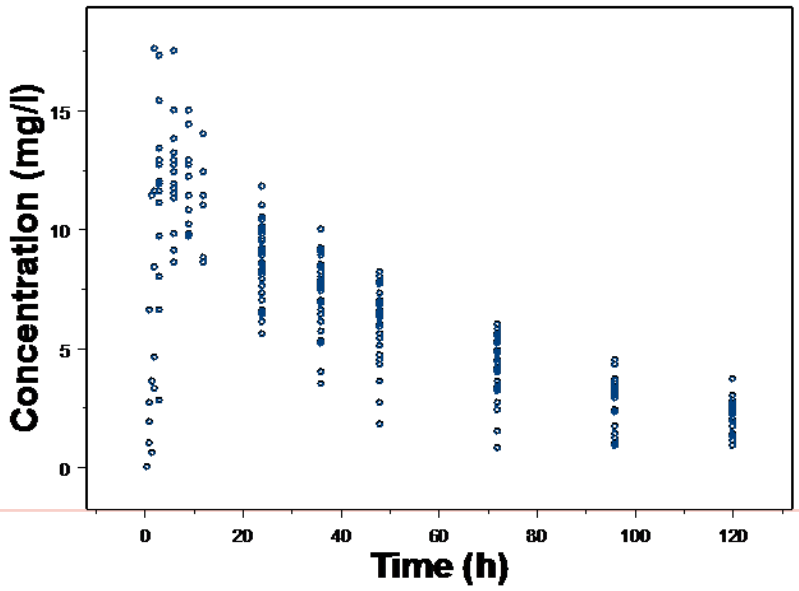
A REAL DATA EXAMPLE

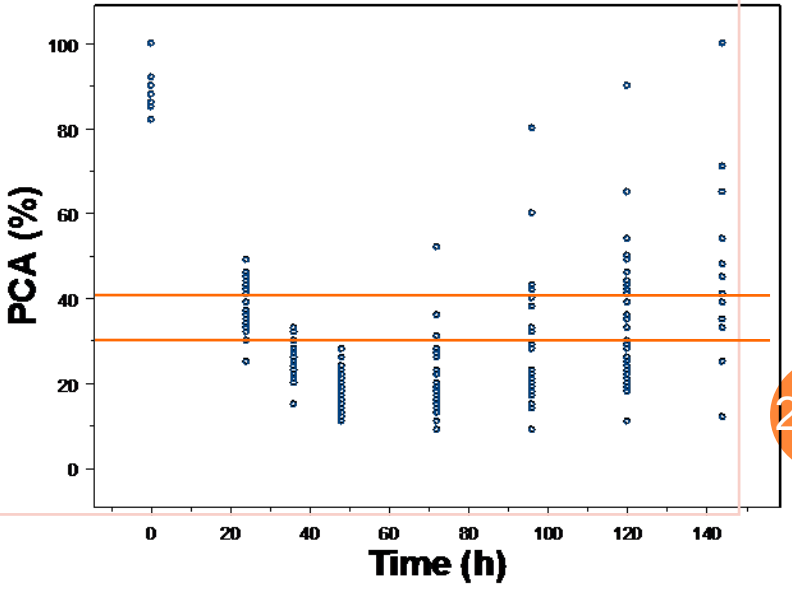
- Warfarin PKPD
- 33 IDs, 479 obs (PD categorized into 3 categories)

  Category 1: < 30% PCA

  Category 2: 30-40% PCA

  Category 3: > 40% PCA





MONOLIX IMPLEMENTATION

CONTINUOUS AND ORDERED CATEGORICAL DATA

$PROBLEM oral 1 (1 cpt with lag-time) and ordered categorical data

$MODEL

COMP = (Qc)

COMP = (Qe)

$PSI Tlag ka V Cl ke0 th1 th2 th3

$PK

KA1 = ka

ALAG1=Tlag

k=Cl/V

$ODE

DDT_Qc = -k*Qc

DDT_Qe = ke0*(Qc-Qe)

Cc=Qc/V

Ce=Qe/V

$CATEGORICAL(1,3)

LOGIT1(Y<=1)= -th1 + th2*Ce

LOGIT1(Y<=2)= -th1 + th2*Ce + th3

$OUTPUT

OUTPUT1 = Cc

OUTPUT2 = LL1

OUTPUT

results

```
**********************************************************************
*      warf_discrete2_project.mat
*      June 25, 2009 at 13:23:56
**********************************************************************
```

Estimation of the population parameters

		parameter	s.e. (s.a.)	r.s.e.(%)
Tlag	:	0.814	0.26	32
ka	:	1.55	0.46	30
V	:	7.94	0.32	4
Cl	:	0.132	0.007	5
ke0	:	0.0215	0.0018	8
th1	:	2.56	1.6	62
th2	:	0.686	0.38	56
th3	:	1.08	0.33	30
omega_Tlag	:	0.573	0.33	58
omega_ka	:	0.884	0.51	58
omega_V	:	0.215	0.013	6
omega_Cl	:	0.288	0.022	8
omega_ke0	:	0.0665	0.0032	5
omega_th1	:	0.183	0.036	20
omega_th2	:	0	-	-
omega_th3	:	0	-	-
a_1	:	0.256	0.27	104
b_1	:	0.058	0.048	82

correlation matrix of the estimates (stochastic approximation)

CONCLUSIONS

- The SAEM algorithm was extended for analysis of ordered categorical and count data
- It provides accurate & precise estimation of both, parameters and standard errors
- The estimation is significantly faster compared to other precise/accurate algorithms
- The estimation procedure is stable
- The algorithm is implemented in new MONOLIX 3.1 and supported by MLXTRAN
- The SAEM has been extended to the Hidden Markov Model (ACoP 2009)

ACKNOWLEDGMENTS

- Swedish Academy of Pharmaceutical Society
- MONOLIX group
- Uppsala University
- INSERM U738

The adventures of monolix continues…