Expectation-Maximization Methods in Population Analysis

Robert J. Bauer, Ph.D.
ICON plc.

Objective

- The objective of this tutorial is to briefly describe the statistical basis of Expectation-Maximization algorithms in Nonlinear Mixed Effects analysis
  - Monte Carlo Importance Sampling
    - Pseudo-Random (standard)
    - Quasi-Random
  - Stochastic approximation expectation-maximization
  - Linearized EM (Iterative two stage)
- How to use these methods
- For which models and data types are these methods useful

List of EM Methods and Their Origins (Quick Over-View)

- Monte Carlo (MC) Importance Sampling Expectation Maximization (EM)
  - First Implemented in PDx-MCPEM by Serge Guzy and in S-ADAPT by Bob Bauer. S-ADAPT is an extension of ADAPT II by D'Argenio and Schumitzky.
  - Quasi-Random variant by Robert Leary, first implemented in Phoenix NLME
- Markov Chain Monte Carlo (MCMC) Stochastic Approximation Expectation Maximization (SAEM)
  - Developed by Marc Lavielle, First implemented in Monolix
- Iterative Two Stage
  - Approximate EM method, described by Steimer,et al., extended by Mentre and Gomeni, first implemented in P-Pharm

Historical Perspective: First Order Method

- First Method was First Order (FO) (late 1970s)
  - – Statistical method that could simultaneously discern variability of measured levels of drug or drug response (residual variance) within a subject, and variability of PK/PD parameters between subjects (inter-subject variance).
  - – Method could determine how population PK/PD parameters related to patient characteristics
  - – Method could do this, even when there were few data points per subject.
  - – FO analysis could be accomplished using the computing power and memory that was available at the time

First Order Conditional Estimation Method

- First Order Conditional Estimation (FOCE) Method (Beal, 1992)
  - First order method was fast, but very approximate
  - Sometimes inaccurate assessments occurred if residual error and/or inter-subject variability were large
  - Conditional method, while also approximate, was more accurate for a larger variety of problems
  - In FOCE mixed effects modeling, an integral over all possible individual parameter values (etas, or random effects) is taken into consideration when determining the best fixed effects (thetas, omegas, and sigmas).
  - These integrations must be done for data of each individual separately.
  - Such integrations can be computationally expensive, and take longer than FO method

More Recent Statistical Methods: Expectation-Maximization (EM)

- Iterative Two Stage (deterministic expectation-maximization method, 1984)
  - Described by Steimer, Mallet, and Golmard, extended by Mentre and Gomeni, first implemented in P-Pharm
  - Approximate method that was able to analyze complex PK/PD problems with greater efficiency and incidence of success than FOCE
  - More accurate than FO, but not as accurate as FOCE

Monte Carlo EM Methods

Monte Carlo Expectation-Maximization Methods (early 2000's)

– An exact method that was able to analyze complex PK/PD problems with greater incidence of success than FOCE

– As in FOCE, the integral over all possible individual parameter values (etas, or random effects) is taken into consideration when determining the best fixed effects (thetas, omegas, and sigmas).

– These integrations are done by Monte Carlo integration techniques. This is called the expectation step.

– Although the Monte Carlo expectation step can be computationally expensive, and/or highly stochastic, the update of the fixed effect parameters can be efficiently carried out if the statistical model is structured in a Phi/Mu structure. This update of the fixed effects is called the maximization step

Monte Carlo EM Methods

- Monte Carlo Expectation-Maximization Methods, continued
  - More accurate than FOCE, especially for sparse data
  - Takes longer than FOCE for simple PK/PD problems (analytical), but more efficient than FOCE for complex PK/PD problems (ordinary differential equations)
  - Efficiency reduces considerably when model cannot be expressed in a particular fixed/random effect (Phi/Mu) format.
  - Can handle full Omega block models efficiently and with stability

Goal of Non-Linear Mixed Effects Methods

- Determine the set of THETAs, OMEGAs, and SIGMAs that best fit the population data, considering all possible values of individual parameters or ETAs.
- To do this, for each subject, an integration of the conditional density over all values of ETAs must be performed, for a given set of Thetas, Omegas, and Sigmas

Goal of Non-Linear Mixed Effects Methods

For observed data, a predictive function may be evaluated using the individual PK/PD parameters, derived from fixed parameters $\boldsymbol{\theta}$, and random variables $\boldsymbol{\eta}$:

$$\boldsymbol{\phi} = \boldsymbol{\mu}(\boldsymbol{\theta}) + \boldsymbol{\eta}$$

Where $\boldsymbol{\mu}(\boldsymbol{\theta})$ are “typical values”, and other fixed effects parameters, to produced predicted value $\mathbf{f}_i(\boldsymbol{\phi}, \boldsymbol{\theta})$.

Goal of Non-Linear Mixed Effects Methods

Often, the distribution of individual parameters $\boldsymbol{\phi}$ are modeled as a normal distribution:

$$h(\boldsymbol{\phi} \mid \boldsymbol{\mu}_i, \boldsymbol{\Omega})) \propto \frac{1}{\det(\boldsymbol{\Omega})^{1/2}} \exp\left(-\frac{1}{2} (\boldsymbol{\phi} - \boldsymbol{\mu}_i)' \boldsymbol{\Omega}^{-1} (\boldsymbol{\phi} - \boldsymbol{\mu}_i) \right) \qquad (0.1)$$

The individual parameter density among the population $h(\boldsymbol{\phi} \mid \boldsymbol{\mu}_i, \boldsymbol{\Omega})$ is the probability that the particular $\boldsymbol{\phi}$ would occur for an individual, given mean typical value parameters $\boldsymbol{\mu}_i$ and its inter-individual covariance $\boldsymbol{\Omega}$. The distribution of $\boldsymbol{\eta}$ is therefore centered about zero (**0**), and can be described as

$$h(\boldsymbol{\eta} \mid \mathbf{0}, \boldsymbol{\Omega})) \propto \frac{1}{\det(\boldsymbol{\Omega})^{1/2}} \exp\left(-\frac{1}{2} \boldsymbol{\eta}' \boldsymbol{\Omega}^{-1} \boldsymbol{\eta} \right) \qquad (0.2)$$

Goal of Non-Linear Mixed Effects Methods

For normally distributed data, a residual variance matrix **V** describes uncertainty of observed values: $\mathbf{V}_i(\mathbf{f}_i, \boldsymbol{\phi}, \boldsymbol{\theta})$. The density can be expressed as

$$l(\mathbf{y}_i \mid \boldsymbol{\phi}, \boldsymbol{\theta}) \propto \frac{\exp\left[-\frac{1}{2}(\mathbf{y}_i - \mathbf{f}_i)'\mathbf{V}_i^{-1}(\mathbf{y}_i - \mathbf{f}_i)\right]}{\sqrt{\det\left(\mathbf{V}_i\right)}} \qquad (0.1)$$

where $l(\mathbf{y}_i \mid \boldsymbol{\phi}, \boldsymbol{\theta})$ is the individual data density, the probability of data $\mathbf{y}_i$ occurring for individual *i*, given individual PK/PD parameters $\boldsymbol{\phi}$, and fixed effect parameters $\boldsymbol{\theta}$ that are not mu modeled.

Goal of Non-Linear Mixed Effects Methods

The joint density of data $\mathbf{y}_i$ and $\boldsymbol{\phi}$ for an individual is the combination of the data density and individual parameter density among the population, to form the joint density:

$$p(\mathbf{y}_i, \boldsymbol{\phi} \,|\, \boldsymbol{\theta}, \boldsymbol{\mu}_i(\boldsymbol{\theta}), \boldsymbol{\Omega}) = l(\mathbf{y}_i \,|\, \boldsymbol{\phi}, \boldsymbol{\theta}) h(\boldsymbol{\phi} \,|\, \boldsymbol{\mu}_i(\boldsymbol{\theta}), \boldsymbol{\Omega}) \qquad (0.1)$$

This is the joint likelihood density of parameters $\boldsymbol{\phi}$ and data $\mathbf{y}_i$ for a given individual.

Goal of Non-Linear Mixed Effects Methods

Because ϕ is unknown, the joint likelihood is integrated over all possible values of ϕ for each individual, so that the “best” population parameters θ and Ω are determined by taking into account the joint probability to an individual’s data over the entire parameter space of ϕ or equivalently, over all η, rather than at just one particular location, such as at the individual’s best fit.

Goal of Non-Linear Mixed Effects Methods

We are therefore interested in evaluating the marginal density of $\mathbf{y}_i$ for any given $\boldsymbol{\theta}$ and $\boldsymbol{\Omega}$:

$$p(\mathbf{y}_i \mid \boldsymbol{\theta}, \boldsymbol{\Omega}) = \int_{-\infty}^{\infty} p(\mathbf{y}_i, \boldsymbol{\phi} \mid \boldsymbol{\theta}, \boldsymbol{\mu}_i, \boldsymbol{\Omega}) d\boldsymbol{\phi} = \int_{-\infty}^{\infty} l(\mathbf{y}_i \mid \boldsymbol{\phi}, \boldsymbol{\theta}) h(\boldsymbol{\phi} \mid \boldsymbol{\mu}_i, \boldsymbol{\Omega}) d\boldsymbol{\phi}$$

(0.1)

for each subject *i*. The total marginal density for all *m* subjects is then

$$p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\Omega}) = \prod_{i=1}^{m} \int_{-\infty}^{\infty} p(\mathbf{y}_i, \boldsymbol{\phi} \mid \boldsymbol{\theta}, \boldsymbol{\mu}_i, \boldsymbol{\Omega}) d\boldsymbol{\phi} \qquad (0.2)$$

Goal of Non-Linear Mixed Effects Methods

It is convenient to use the negative logarithm of the density, and refer to this as the objective function, for each individual:

$$L_i = -\log\left(\int_{-\infty}^{+\infty} p(\mathbf{y}_i, \boldsymbol{\phi} \mid \boldsymbol{\theta}, \boldsymbol{\mu}_i, \boldsymbol{\Omega})\, d\boldsymbol{\phi}\right)$$

and for the total data set:

$$L = -\log(p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\Omega})) = \sum_{i=1}^{m} L_i$$

To fit a model with mean population parameters $\boldsymbol{\theta}$ and population variance $\boldsymbol{\Omega}$ to data **y**, the negative logarithm of the marginal density is minimized, which is equivalent to maximizing the marginal density.

Goal of Non-Linear Mixed Effects Methods

- When the individual's joint density for data and individual parameters is normalized, we have the posterior, or conditional density:

$$z(\phi) = \frac{l(\mathbf{y}_i \mid \phi, \theta) h(\phi \mid \mu, \Omega)}{\int l(\mathbf{y}_i \mid \phi, \theta) h(\phi \mid \mu, \Omega) d\phi}$$

- Often, the predicted function **f** that appears in the data density is non-linear with respect to individual parameters (phi)
- Therefore, while the data density is a normal distribution with respect to the data, and the individual parameter is normally distributed among the population, the conditional density is not normally distributed with respect to individual parameters

FOCE

- This integration is computationally difficult to do
- FOCE Evaluates the mode of the conditional density (most likely values of etas) and first order approximation of variances of etas
- This approximate integral of the Gaussian (Normal) function centered at the mode of the conditional density with the approximate variance can be easily calculated.
- This integral of the normal approximation serves as integrated objective function

FOCE

- The integral of the normal function over all etas or individual parameter values is equal to the value of the normal function at the mode or peak, (the “height”) multiplied by the determinant of the variance-covariance matrix of the normal function (the multi-dimensional “width”, so height x width=area):

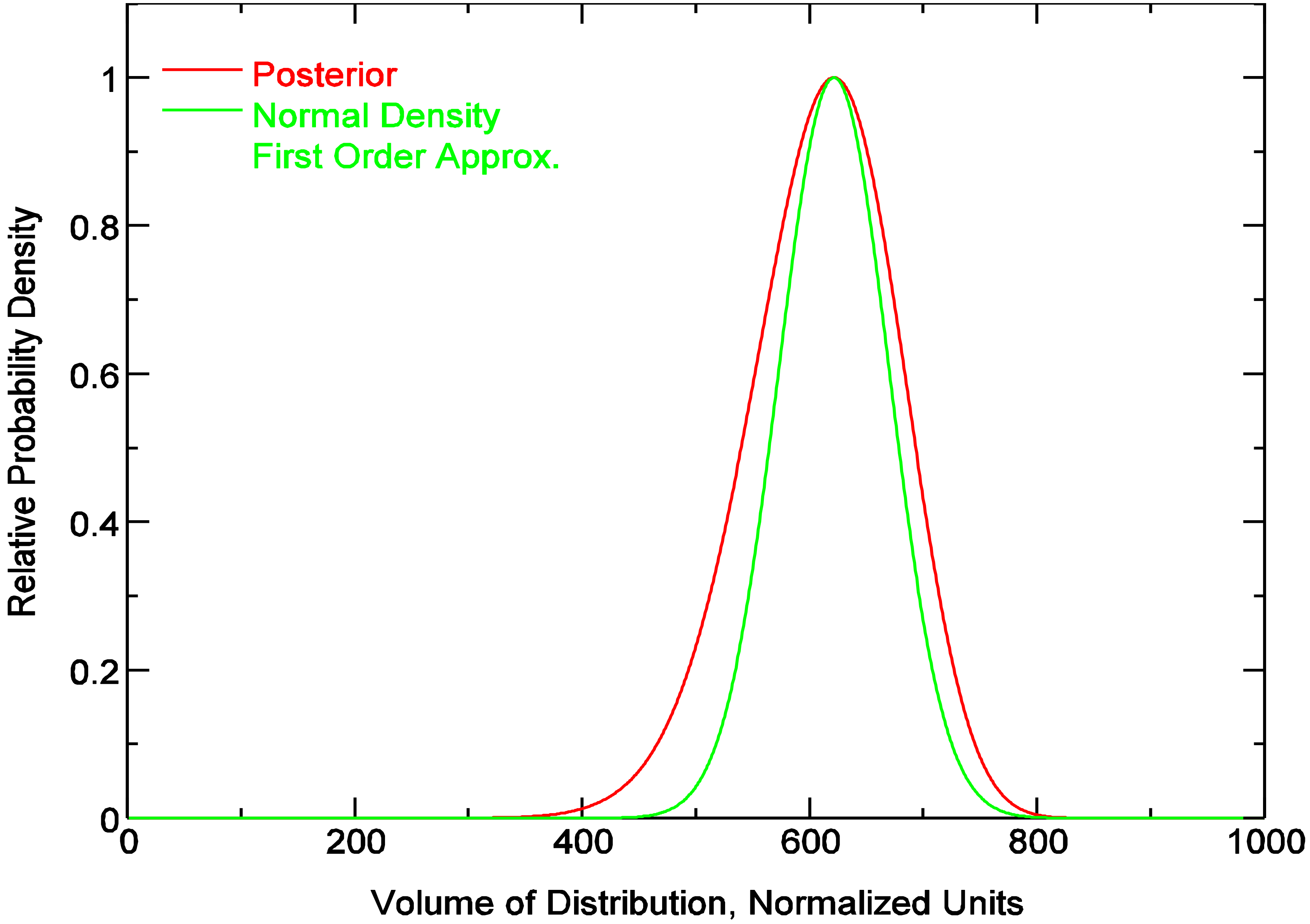
$\hat{\boldsymbol{\phi}}$ =individual parameters at mode (peak) of posterior density

$\hat{\mathbf{V}}$ =variance-covariance matrix of posterior density

$$p(\mathbf{y}_i \mid \boldsymbol{\theta},\boldsymbol{\Omega}) = \int_{-\infty}^{\infty} p(\mathbf{y}_i,\boldsymbol{\phi} \mid \boldsymbol{\theta},\boldsymbol{\mu}_i,\boldsymbol{\Omega})\,d\boldsymbol{\phi} \approx p(\mathbf{y}_i,\hat{\boldsymbol{\phi}} \mid \boldsymbol{\theta},\boldsymbol{\mu}_i,\boldsymbol{\Omega})\det\left(\hat{\mathbf{V}}\right)$$

Height x Width

Normal Density Serves as Approximation to Posterior Density in FOCE



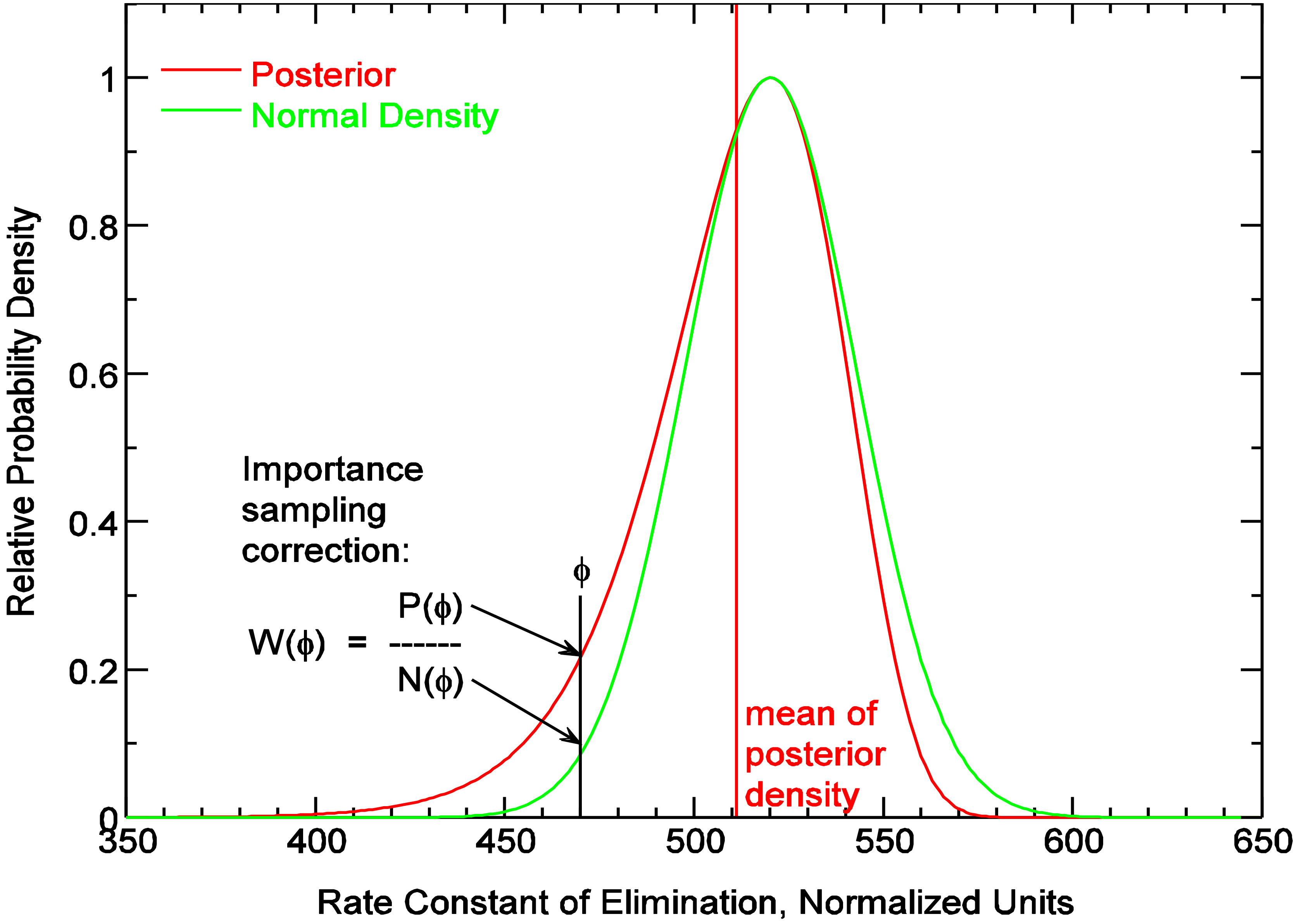
FOCE Approximation

- This linearized approximation of using the area of a Gaussian function as a substitute for the true area under the poster density is sufficiently accurate when residual variance within subjects is small, and/or the non-linearity (equivalently, deviation from Gaussian) of parameter distribution of the PK/PD model is not large.

Importance Sampling EM: Expectation

- To improve accurate assessment of area under the posterior density, Consider Importance Sampling, which:
- Evaluates the conditional (posterior) mean and variance of individual parameters (etas) by Monte Carlo sampling (integration) (Expectation Step)
- Gaussian function positioned near the mean or mode of the posterior is used as a proposal (sampling) density, then weighted according to posterior density, which is not truly normally distributed

Normal Density Serves as Proposal to Posterior Density in Importance Sampling EM



Importance sampling correction:

$$W(\phi) = \frac{P(\phi)}{N(\phi)}$$

Importance Sampling EM: Maximization

- Simple single iteration maximization steps update parameters
  - stable, and statistically proven to improve the objective function
  - Example: Average of conditional parameters among all individuals serves as update to a relevant fixed (theta) parameter
  - Variance of conditional parameters among all individuals serves as update to a relevant omega parameter
- Population parameters converge towards the position of exact maximum likelihood
- Accurate marginal density based objective function
- Standard errors are rapidly obtained by inverting an information matrix constructed from Monte Carlo elements

Why is EM Efficient with Mu Modeling: Maximization Step

- In practice, only those structural parameters (thetas) associated with typical values to individual parameters that do not change for that subject (not time/record dependent) can be efficiently evaluated by phi/mu modeling. For subject *i*:

$$\boldsymbol{\phi}_i = \boldsymbol{\mu}_i(\boldsymbol{\theta}, \mathbf{x}_i) + \boldsymbol{\eta}$$

$$\boldsymbol{\phi} \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Omega})$$

Why is EM Efficient with Mu Modeling: Maximization Step

- In such cases, the objective function may be easily improved by simple maximization steps. For example, if

$$\boldsymbol{\mu}_i = \boldsymbol{\theta}$$

- Then the new theta that improves the objective function is evaluated as:

$$\overline{\boldsymbol{\phi}}_i = \int \boldsymbol{\phi} z_i d\boldsymbol{\phi} \rightarrow \frac{1}{R}\sum_{r=1}^{R} w_r \boldsymbol{\phi}_r \quad R = \mathrm{ISAMPLE}$$ (expectation)

$$\hat{\boldsymbol{\theta}}_{\mathrm{new}} = \hat{\boldsymbol{\mu}}_i = \frac{1}{m}\sum_{i=1}^{m} \overline{\boldsymbol{\phi}}_i$$ (maximization, for *m* subjects)

Why is EM Efficient with Mu Modeling: Maximization Step

- If a covariate model is used, for example:

$$\mu_i = \theta_1 + \theta_2 x_i$$

- Then the new theta that improves the objective function is evaluated as:

$$\text{update d } \hat{\theta}_1, \hat{\theta}_2 = \text{simple regression analysis on } (x_i, \overline{\phi}_i)$$

- If mu has linear relationship with thetas (as in above example), then the linear regression guarentees efficient improvement in objective function

- If not linear relation, such as

$$\mu_i = \log(\theta_1) + \theta_2 x_i$$

- then the non-linear regression is less efficient in improving in objective function

Efficiency of EM Methods

EM methods are most efficient when they can be expressed in this particular fixed/random effect (Phi/Mu) format, in which population parameters (theta) are used to model the mean (Mu) of a normal distribution of individual parameters (Phi).

Efficiency reduces when many population parameters are shared among subjects without inter-individual variability, that is, population parameter does not model a mean value to normally distributed individual parameters.

Efficiency of EM Methods

For parameters not mu-modeled, two methods could be used:

gradient method: Requires creation of Monte Carlo first-derivatives. To speed up evaluation, only a subset of Monte Carlo samples from each subject need be used for evaluation (recommended by Robert Leary).

Annealing method: Temporarily Mu-model the parameter (that is, add inter-subject variability), then constrain the variance to ever smaller values with each iteration, until variance is 0.

Why is EM Efficient with Mu Modeling: Maximization Step

- Similarly, the new Omega that improves the objective function is evaluated as:

$$\bar{\Omega}_i = \int (\phi - \mu_i)(\phi - \mu_i)' z_i d\phi$$

$$\rightarrow \frac{1}{R} \sum_{r=1}^{R} w_r (\phi - \mu_i)_r (\phi - \mu_i)'_r \quad R = \mathrm{ISAMPLE}$$ (expectation)

$$\hat{\Omega}_{\mathrm{new}} = \frac{1}{m} \sum_{i=1}^{m} \bar{\Omega}_i$$ (maximization)

- The Omega symbols above represent the entire matrix of individual Omega elements.

Full OMEGA Blocks in EM Analysis

- Full OMEGA blocks are more easily updated by EM methods than by FOCE.
- Full OMEGA blocks are preferred over diagonal Omegas in EM problems.
- Having off-diagonal elements does not necessarily make EM methods less stable.
- No need to fix off-diagonal elements to 0 even though SE is greater than estimate
- If original analysis was performed with FOCE, and if off-diagonal element was fixed to 0 because FOCE had round-off error problems in estimation, or $COV step could not complete, allow off-diagonal to be estimated when doing EM.
- Then, only if EM objective function is highly variable, or $COV step reports non-positive-definite issues, consider fixing off-diagonal elements back to 0.

Using Quasi-Random Importance Sampling

- One variant of importance sampling is to use a random sampler that generates vectors of etas in a Quasi-random (Sobol Sequence) manner, rather than in a pure random fashion.
- The quasi-random sample has a more even distribution across the multi-dimensional eta space than the usual pseudo-random number generator.
- The effect is to reduce the stochastic noise by 2-10 fold per N generated sample vectors.
- Fewer samples are required for a desired stochastic noise level.

Using Quasi-Random Importance Sampling

- Example: Two compartment model, with 5 data points per subject, 100 subjects.

Importance sampling was performed using 300 random samples per subject. The STD of the set of objective function values of the last 10 iterations was 1.13:

```
iteration           29 OBJ=   -1144.36
iteration           30 OBJ=   -1146.20
iteration           31 OBJ=   -1145.48
iteration           32 OBJ=   -1144.12
iteration           33 OBJ=   -1146.41
iteration           34 OBJ=   -1144.54
iteration           35 OBJ=   -1143.63
iteration           36 OBJ=   -1146.96
iteration           37 OBJ=   -1146.15
iteration           38 OBJ=   -1145.96
```

Using Quasi-Random Importance Sampling

Same problem performed with Sobol sequence sampling of 300 per subject, and the STD of set of objective function values of the last 10 iterations was 0.49:

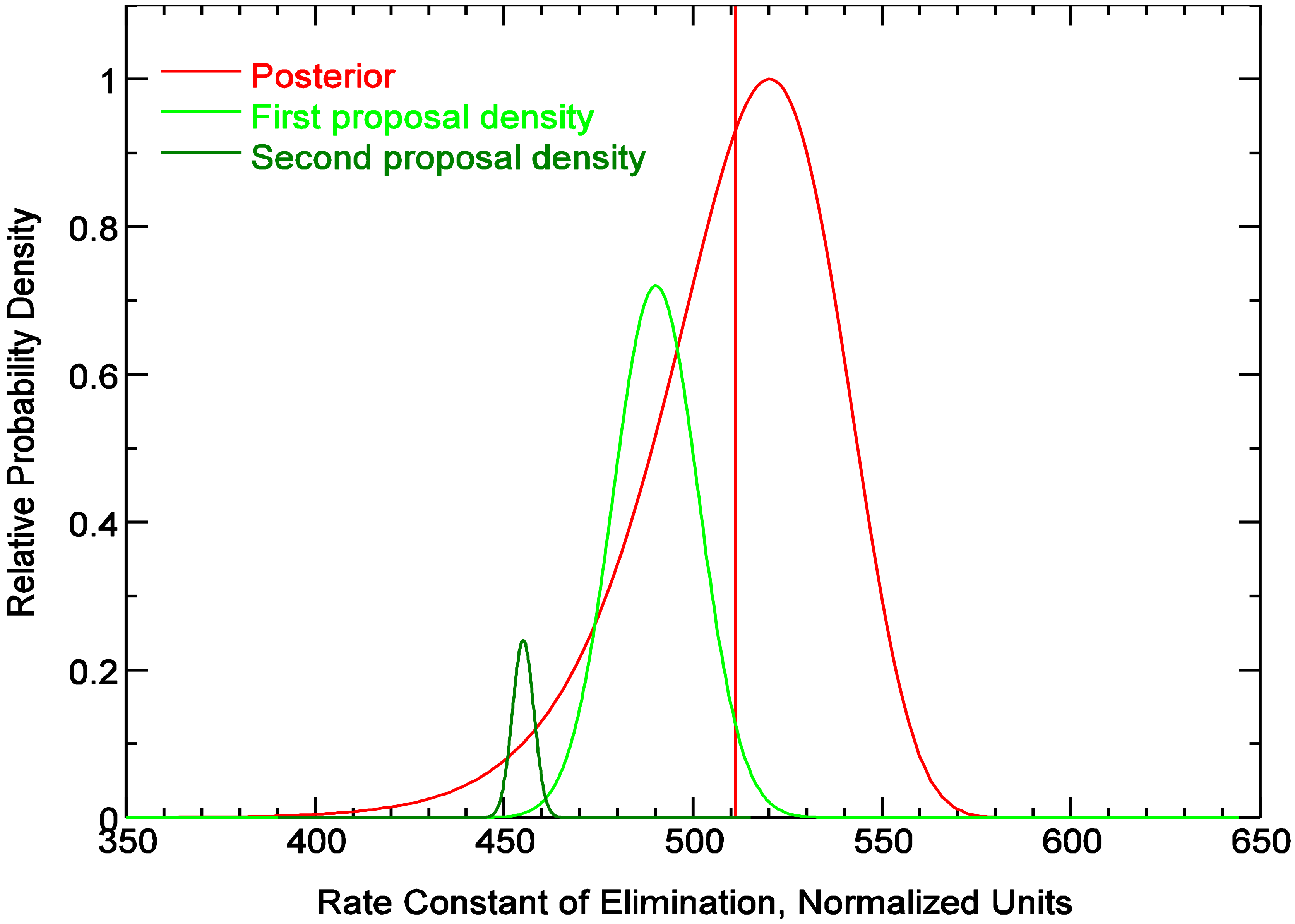
```
iteration             9 OBJ=   -1145.56
iteration            10 OBJ=   -1145.04
iteration            11 OBJ=   -1144.47
iteration            12 OBJ=   -1145.29
iteration            13 OBJ=   -1145.50
iteration            14 OBJ=   -1145.83
iteration            15 OBJ=   -1144.74
iteration            16 OBJ=   -1145.95
iteration            17 OBJ=   -1145.22
iteration            18 OBJ=   -1144.75
```

- This is an improvement of 2.3 fold over the standard random sampling. To get equivalent STD in standard random sampling, would need $300*2.3^2$=1500 samples per subject.

SAEM: Markov Chain Monte Carlo (MCMC)

- As in importance sampling, random samples generated from Normal proposal densities
- Instead of always centered at mode (or mean) of the posterior density, proposal density is centered at the previous (p) sample position
- New samples are accepted with probability $W(\varphi)/W(\varphi_p)$
- The variance of proposal density is adjusted to maintain a certain average acceptance rate.
- This method requires more elaborate sampling strategy, but is useful for highly non-normally distributed posterior densities

Normal Density Serves as Moving Proposal Density in MCMC



MCMC Stochastic Approximation EM

- In first mode (stochastic/burn-in) , SAEM evaluates an unbiased but highly stochastic approximation of individual parameters (semi integration, usually 2 samples per individual)
- Population parameters are updated from individual parameters by single iteration maximization steps that are very stable, and statistically proven to improve the objective function
- In second mode (reduced stochastic/accumulation), individual parameter samples from previous iterations are averaged together, converging towards the true conditional individual parameter means and variances

MCMC Stochastic Approximation, continued

- Leads to Population parameters converging towards the maximum of the exact likelihood
- Objective function best obtained by a single iteration of importance sampling at final population parameter values

Approximate (Linearized) Likelihood EM Method: Iterative Two Stage

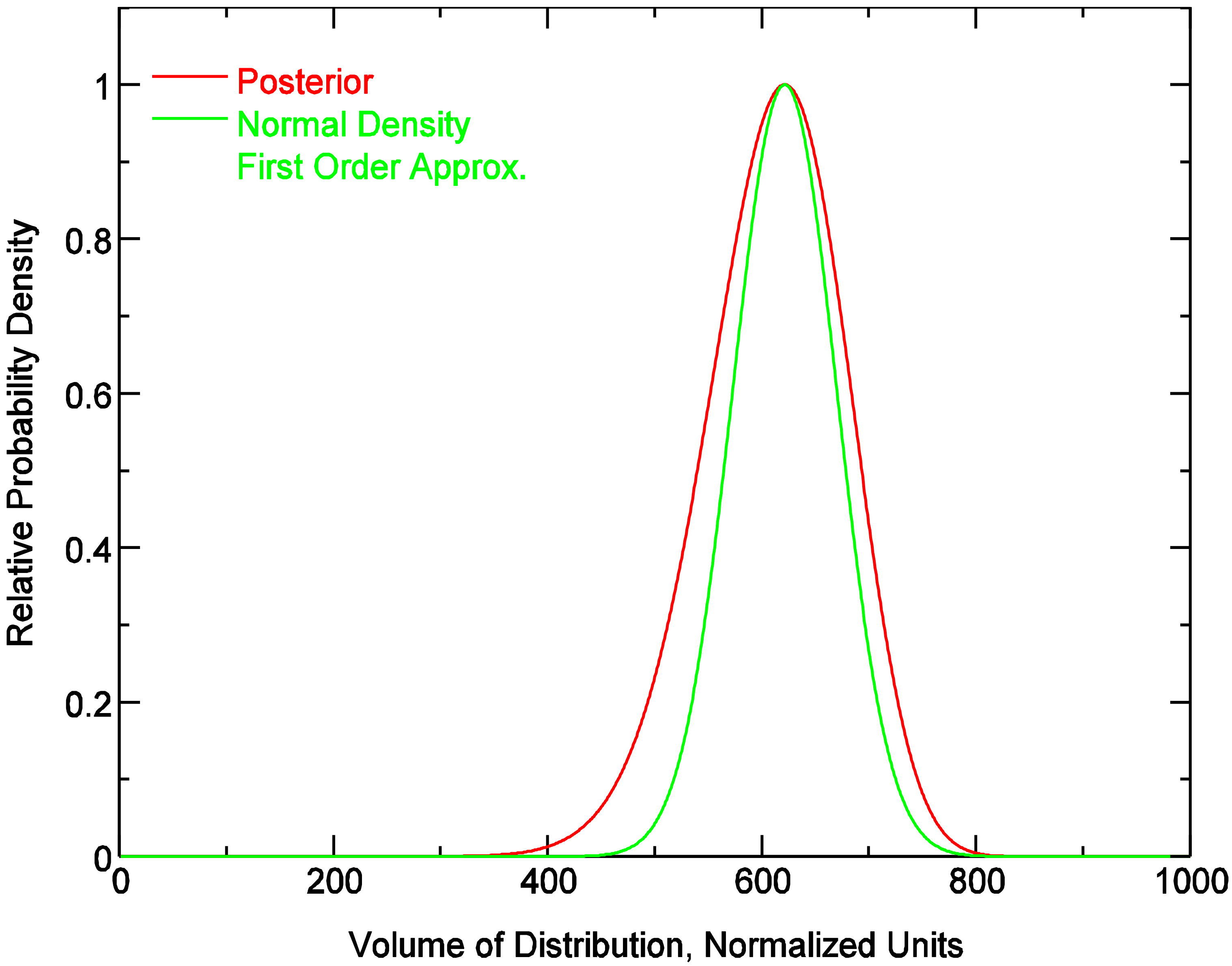
- Evaluates the conditional mode (not mean!) and first order approximation of the (expected) variance of parameters of individuals by maximizing the posterior density
  - This integration step is the same as in FOCE
- Parameters updated from conditional modes ($\hat{\boldsymbol{\phi}}_i$), and approximate individual variances ($\hat{\mathbf{V}}_i$)

$$\hat{\boldsymbol{\theta}}_{\text{new}} = \hat{\boldsymbol{\mu}}_i = \frac{1}{m}\sum_{i=1}^{m}\hat{\boldsymbol{\phi}}_i$$

$$\hat{\boldsymbol{\Omega}}_{\text{new}} = \frac{1}{m}\sum_{i=1}^{m}(\hat{\boldsymbol{\phi}}_i - \boldsymbol{\mu}_i)(\hat{\boldsymbol{\phi}}_i - \boldsymbol{\mu}_i)' + \frac{1}{m}\sum_{i=1}^{m}\hat{\mathbf{V}}_i$$

- Parameters updated by single iteration maximization steps that are very stable (usually in 50-100 iterations)
- For rich data, almost as accurate as FOCE, but much faster

Normal Density Serves as Approximation to Posterior Density in ITS



When to Use Each EM Method

- MC Importance Sampling EM (IMP):
  - Sparse (few data points per subject) or rich data
  - Complex PK/PD problems with many parameters
- SAEM
  - Very sparse, sparse, or rich data
  - categorical data
- Iterative Two Stage (ITS)
  - Rich data
  - Rapid, exploratory method
- FOCE
  - Rich data
  - Many THETAS with no ETA’s associated with them
  - More accurate than ITS

When to Use Each EM Method

- MC Importance Sampling EM (IMP)
  - Complex PK/PD problems with many parameters are rapidly evaluated compared to FOCEI
  - Sparse (fewer data points per subject than etas to be estimated) or rich data
  - Can be less accurate than SAEM with highly categorical data or very sparse data
  - Can track progress of improvement in true objective function with each iteration
  - Results can vary stochastically, typically by about 25% of SE
  - Can handle full OMEGA blocks well
  - May become less efficient when some or many thetas may not be MU modeled

When to Use Each EM Method

- SAEM
  - categorical data
  - Very sparse, sparse, or rich data
  - Complex PK/PD problems with many parameters (may sometimes reach true objective function only within +/- 10 units of optimum, and can take longer than IMP)
  - Cannot assess true objective function during its progress, must finish analysis with Importance sampling assessment of objective function
  - Results can vary stochastically, typically by about 25% of SE
  - Can handle full OMEGA blocks well
  - May become less efficient when some or many thetas may not be MU modeled

When to Use Each EM Method

- Iterative Two Stage (ITS)
  - Rich data
  - Rapid, exploratory method
  - Can be used as pre-analysis to facilitate IMP or SAEM
  - Requires less fuss with adjusting options than SAEM or IMP
  - Results are highly reproducible to +- 4 digits
  - Can have large bias or instability for some problems
  - Can handle full OMEGA blocks well
  - Less efficient when some or many thetas may not be MU modeled

When to Use Each EM Method

- FOCE
  - – Rich and semi-rich data
  - – Good for many THETAS with no ETA’s associated with them
  - – More accurate than ITS
  - – Requires less fuss with adjusting options than SAEM or IMP
  - – Results are highly reproducible to +- 4 digits
  - – Does not handle full Omega blocks as easily as EM
  - – If convergence criterion reduced, then time of analysis can be reduced by 2-3 fold in some cases

How to Use Each EM Method (IMP)

- IMP

For standard PK in which data is normally distributed (continuous data, such as from quantitative assays), and there are more data points per subject than etas to be estimated, the following settings (which are close to default) are usually sufficient to begin with:

random samples per subject=300

number of iterations=50-100

Normal density sampler

Acceptance rate of 0.4-1.0

If stochastic noise of OFV is higher than 1-3 STD, increase number of samples per subject. Typically noise decreases by approximately $N^{1/2}$

How to Use Each EM Method (IMP)

Models in which the prediction function causes high degree of non-linearity on the etas (some PD models may do this), or as the data becomes more sparse (fewer data points per subject), or for large numbers of data below the quantifiable limit of the assay, or the observed data is non-normally distributed such as with categorical data, the posterior density becomes less Gaussian in shape and can have extended tails. To reach widened tails in the distribution, do one or more of the following:

increase number of samples per subject (1000-10000)

use t-distribution sampler with DF=1-4

Decrease acceptance rate (expanding sampler coverage) to 0.1-0.2

How to Use Each EM Method (SAEM)

- SAEM

  For standard PK in which data is normally distributed (continuous data, such as from quantitative assays), and there are more data points per subject than etas to be estimated, the following settings (which are close to default) are usually sufficient to begin with:

  random samples (chains) per subject=1-2 (may need more if there are fewer than 50 subjects)

  number of burn-in iterations=200-300

  number of accumulating iterations=200-300

How to Use Each EM Method (SAEM)

Models in which the prediction function causes high degree of non-linearity on the etas (some PD models may do this), or as the data becomes more sparse (fewer data points per subject), or for large numbers of data below the quantifiable limit of the assay, or the observed data is non-normally distributed such as with categorical data, the posterior density becomes less Gaussian in shape. To reach widened tails in the distribution, do at least:

increase number of samples per subject (3-10)

and maybe:

number of burn-in iterations=500-1000

number of accumulating iterations=500-1000

How to Use Each EM Method (SAEM)

If there are few subjects in the population analysis (<30 for example), then number of samples (chains) evaluated per subject should be increased to 5-10, so that the stochastic noise in assessing population parameters during the burn-in period is not too great.