Accurate Maximum Likelihood Estimation for Parametric Population Analysis

Bob Leary

UCSD/SDSC and LAPK, USC School of Medicine

Why use parametric maximum likelihood estimators?

- Consistency:

$$\hat{\theta}_{ML} \rightarrow \theta_{TRUE} \text{ as } N \rightarrow \infty$$

- Asymptotic Efficiency:

$$\frac{\operatorname{var} \hat{\theta}_{ML}}{\operatorname{var} \hat{\theta}_{OTHER}} \leq 1 \text{ as } N \rightarrow \infty$$

- Well-developed asymptotic theory allows hypothesis testing

But most current parametric methods only maximize approximate likelihoods (MAL)

- F.O.
- F.O.C.E
- Laplace

MAL estimation can cause serious degradation of statistical peformance

- Not Consistent:

$$\hat{\theta}_{MAL} \rightarrow \theta_{TRUE} + \text{bias as } N \rightarrow \infty$$

- Not Asymptotically Efficient:

$$\frac{\operatorname{var}\hat{\theta}_{MAL}}{\operatorname{var}\hat{\theta}_{ML}} > 1 \text{ as } N \rightarrow \infty$$

- No asymptotic theory – using ML asymptotic theory for MAL case can be VERY misleading!

Statistical efficiency definition

- Relative statistical efficiency:

$$\text{statistical efficiency}(\hat{\theta}_{MAL}) = \frac{\operatorname{var}\hat{\theta}_{ML}}{\operatorname{var}\hat{\theta}_{MAL}}$$

- Statistical efficiencies can be evaluated by Monte Carlo simulations

Simulation conditions

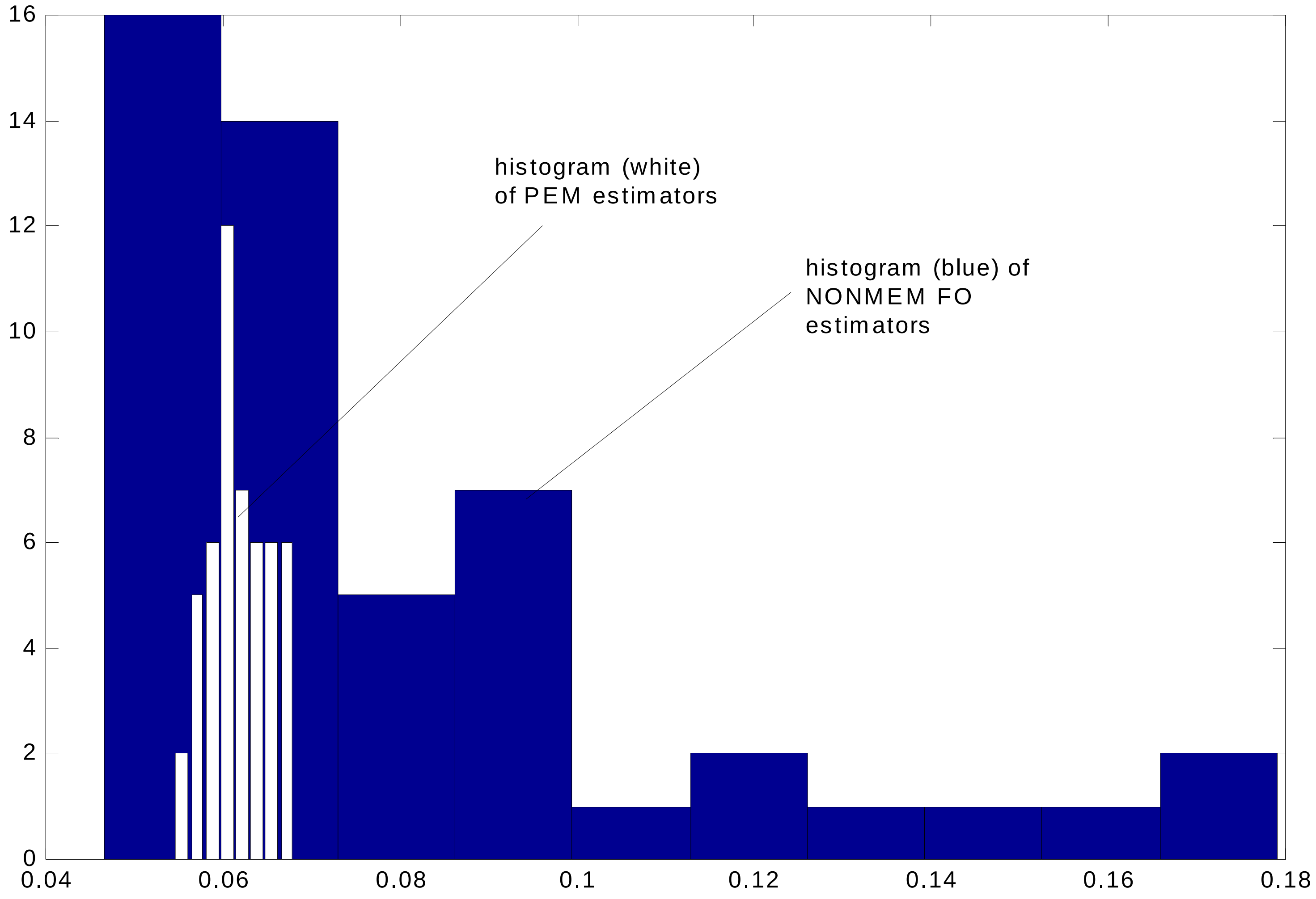
- Simple 1-compartment IV bolus model, two random effects V and K

$$C(t) = (1/V)e^{-Kt}$$

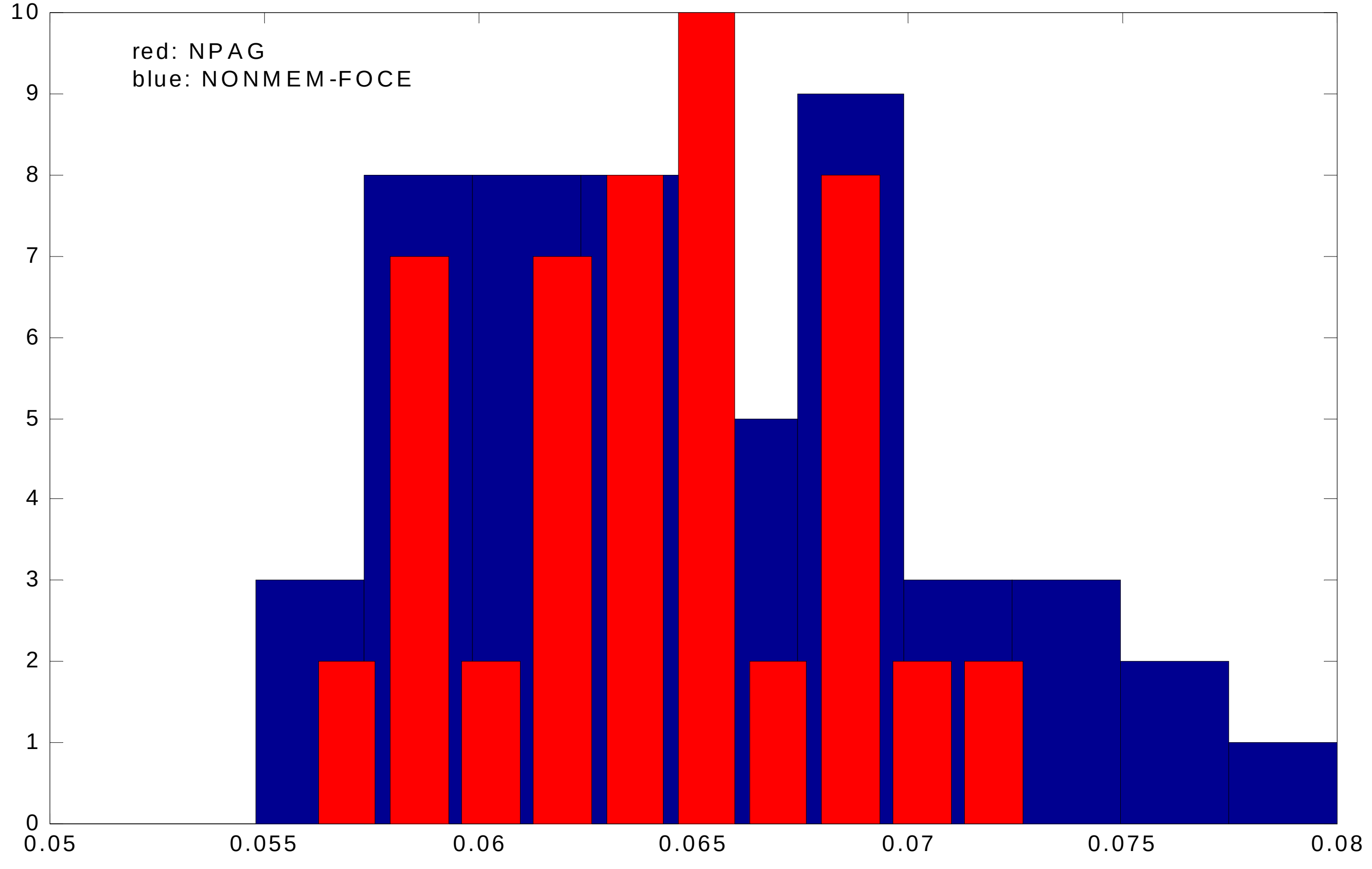
- $(V, K) \sim N(\mu, \Sigma)$, 25% inter-individual relative standard deviation in each of V, K.

- $Y_i=C(t)(1+e_i)$, $e_i \sim N(0,\sigma^2)$, intra-individual relative $\sigma=.10$, 2 observations/subject (sparse data)

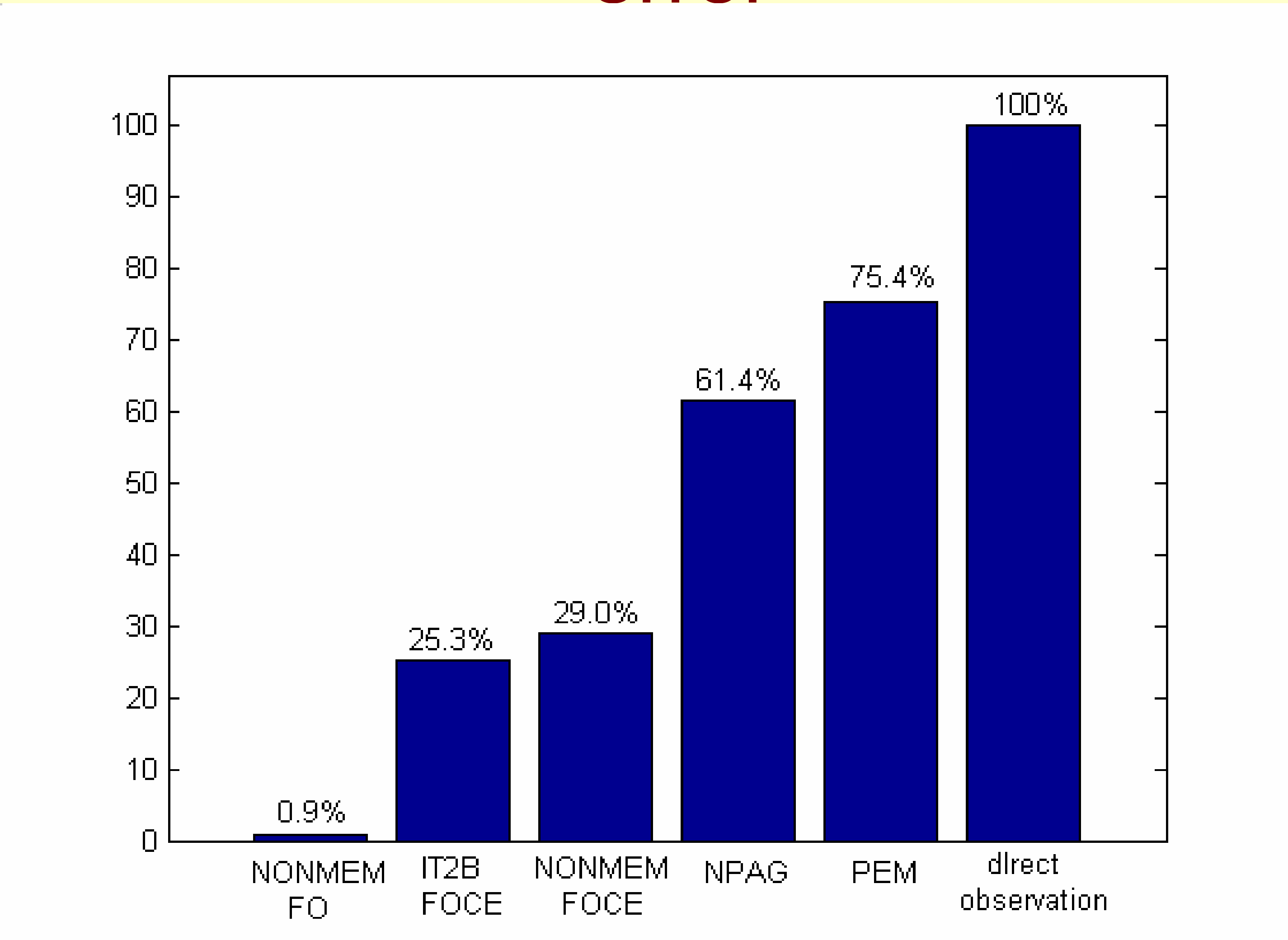
Approximate likelihoods can destroy statistical efficiency



FOCE does better, but still has <40% efficiency relative to ML



Comparative statistical efficiencies, 2 obs/subject, 10% observational error



Approximations add random error

$$\hat{\theta}_{MAL} = \hat{\theta}_{ML} + (\hat{\theta}_{MAL} - \hat{\theta}_{ML})$$

$$\mathrm{var}(\hat{\theta}_{FO} - \hat{\theta}_{ML}) = 80\,\mathrm{var}(\hat{\theta}_{ML})$$

$$\mathrm{var}(\hat{\theta}_{FOCE} - \hat{\theta}_{ML}) = 1.2\,\mathrm{var}(\hat{\theta}_{ML})$$

PEM Estimation

- Simplest case – just random effects (no fixed effects related to covariates)
- Population distribution of random effects vector η at the lowest level is $N(\mu, \Sigma)$, so parameters to be estimated are $\theta = (\mu, \Sigma)$
- We are given likelihood functions

$$l_i(y_i \mid \eta, \sigma^2)$$

PEM algorithm

- For current population distribution iterate $N(\mu^j, \Sigma^j)$, compute posterior distributions

$$f_i^{post} = \frac{l_i(y_i \mid \eta) f(\eta \mid \mu_j, \Sigma_j)}{Q_i}$$

- Compute $(\mu^{j+1}, \Sigma^{j+1})$ as the mean and covariance of the posterior mixture distribution

$$f^{post} = \frac{1}{N}\sum_{i=1}^{N} f_i^{post}$$

Likelihoods

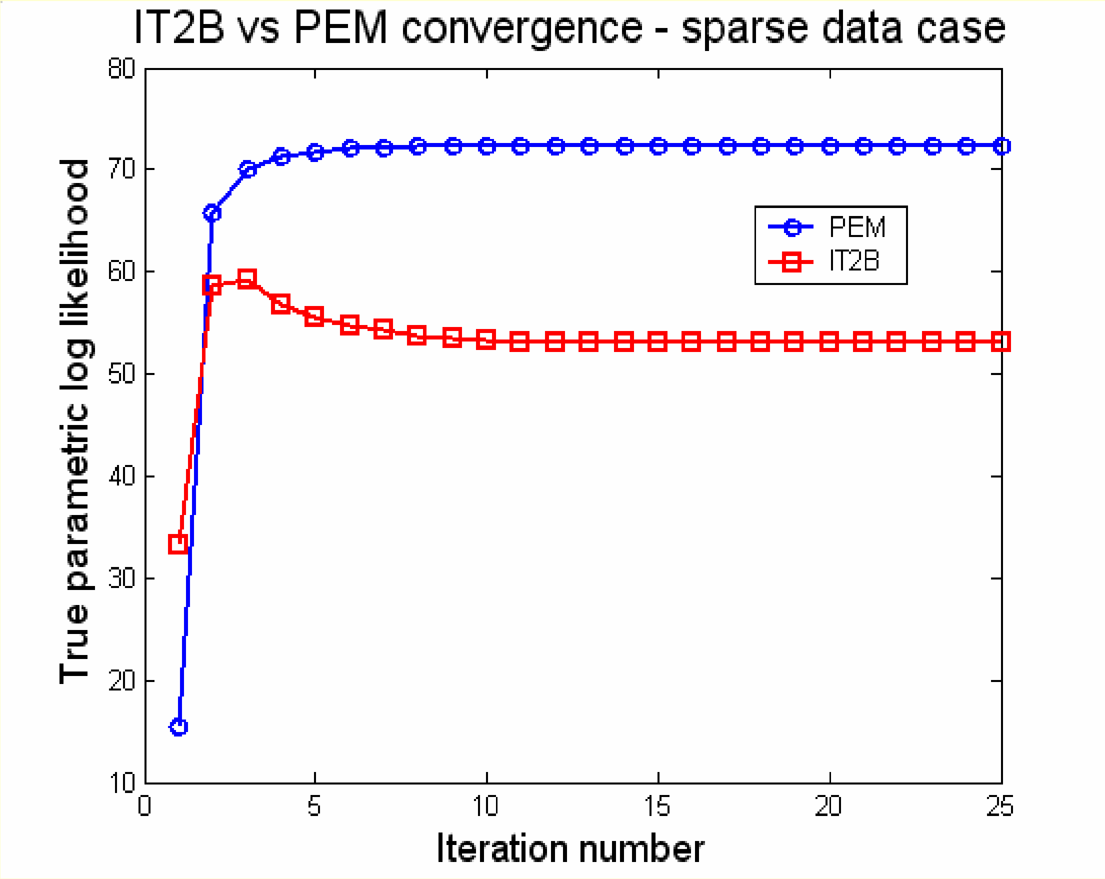
- Likelihood of j+1 iterate is

$$\log L(\mu^{j+1}, \Sigma^{j+1}) = \sum_{i=1}^{N} \log Q_i$$

- (Schumitzky, 1993) showed

$$\log L(\mu^{j+1}, \Sigma^{j+1}) \geq \log L(\mu^{j}, \Sigma^{j})$$

Likelihood convergence – FOCE vs PEM



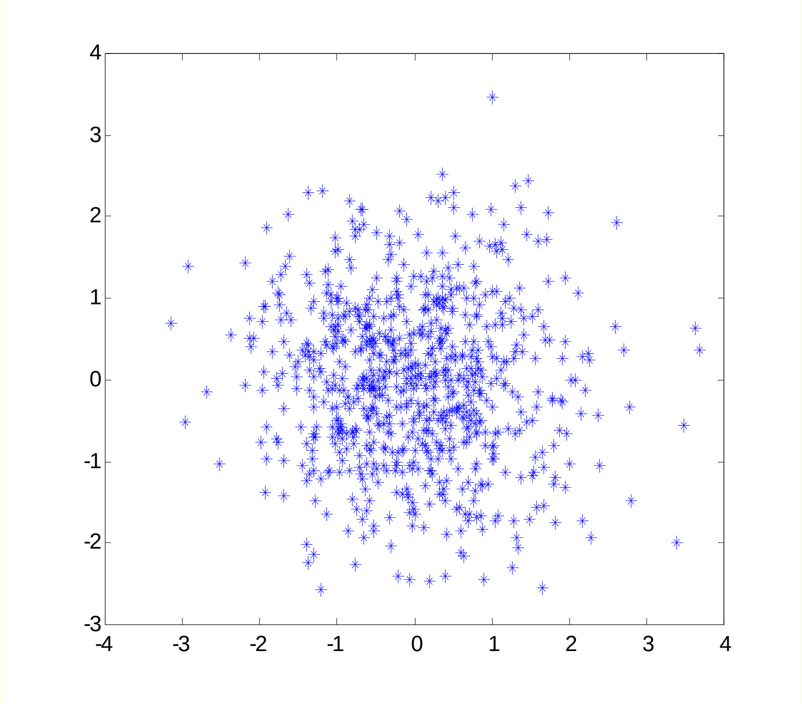
Numerical integration methods

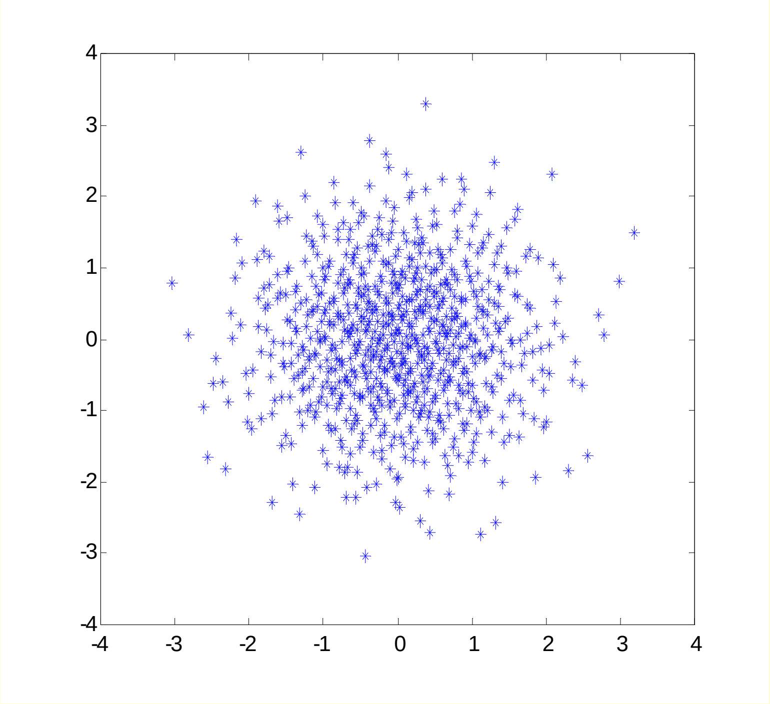
$$Q_i = \int l_i(y_i \mid \eta) f(\eta \mid \mu_j, \Sigma_j) d\eta$$

$$= \frac{1}{M} \sum_{k=1}^{M} l_i(y_i \mid \eta_k) \text{ for M samples } \eta_k$$

η_k can be Monte Carlo (pseudorandom) samples, Gauss-Hermite grid points, or "quasi random" samples (low discrepancy sequences)

1000 pseudo- vs quasi random samples from a bivariate N(0,**I**)





$$\Sigma_{pseudo} = \begin{matrix} 1.065 & -0.064 \\ -0.064 & 0.983 \end{matrix}$$

Integration error~1/$M^{1/2}$

$$\Sigma_{quasi} = \begin{matrix} 1.003 & 0.004 \\ 0.004 & 0.998 \end{matrix}$$

Integration error ~1/M

Recent blind trial Pop PK method comparison (Girard et al, 2004)

- Model

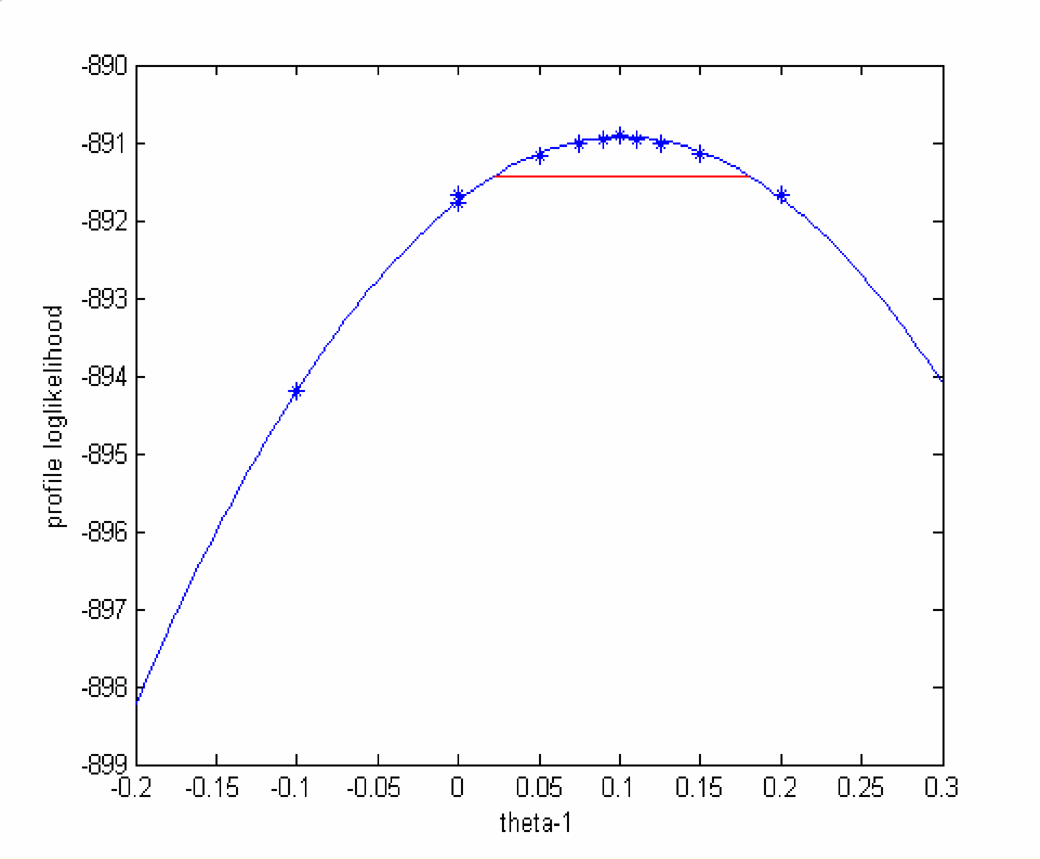
$$response = E_0 e^{\eta_1} + \frac{E_{\max} e^{\eta_2} dose}{ED50 * f(sex) e^{\eta_3} + dose}$$

$$f(male) = 1, f(female) = \theta$$

$$E_0 e^{\eta_1} = e^{\log E_0}$$

- Estimate all 8 parameters with standard errors and p-value for null hypothesis θ=1

Profile likelihood method



Conclusions

- Likelihood approximations such as FO and FOCE significantly degrade statistical results, particularly in the sparse data case
- Accurate likelihood methods such as PEM perform much better in the sparse data case than approximate methods
- Accurate likelihood methods such as PEM are feasible with current computational technology