### PAGANZ 2024 Pumas Workshop

#### Mohamed Tarek

Senior Product Engineer at PumasAl Inc.

Research Affiliate at University of Sydney Business School







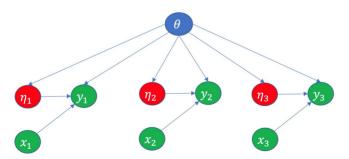
### Contents I

- 1 NLME fitting algorithms
- 2 Weighted residuals
- 3 Standard error estimation
- 4 Continuous visual predictive check
- 5 Time to event models

# NLME fitting algorithms



#### Assume there are 3 subjects



```
@model begin
                  @param begin
                     θ € VectorDomain(4, lower = zeros(4))
                    Ω ∈ PSDDomain(2)
                     Σ ∈ RealDomain(lower = 0.0)
                     a ∈ RealDomain(lower = 0.0, upper = 1.0)
                  end
                  @random begin
      \eta_i \mid \theta
                    n ~ MvNormal(Ω)
                  @covariates sex wt etn
                   @pre begin
                     \theta 1 := \theta [1]
                     CL = \theta[2] * ((wt / 70)^0.75) * (\theta[4]^sex) *
                     V_C = 0.031 * exp(n.021)
                   end
y_i | \theta, \eta_i, x_i
                  @dynamics begin
                     Depot' = -Ka * Depot
                     Central' = Ka * Depot - (CL / Vc) * Central
                     Res' = Depot - Central
                   end
                  @derived begin
                     conc = @. Central / Vc
                     dv ~ @. Normal(conc. conc * Σ)
                     T max = maximum(t)
                   end
                  @observed begin
                     obs cmax = maximum(dv)
                  end
                end
```

 $m{\theta}$ : all the population parameters (typical values, BSV, BOV and residual variability)

- $m{\theta}$ : all the population parameters (typical values, BSV, BOV and residual variability)
- lacksquare  $\eta_i$ : random effects of subject i

- $\theta$ : all the population parameters (typical values, BSV, BOV and residual variability)
- lacksquare  $\eta_i$ : random effects of subject i
- N: number of subjects

- $\theta$ : all the population parameters (typical values, BSV, BOV and residual variability)
- lacksquare  $\eta_i$ : random effects of subject i
- N: number of subjects
- $\bullet$   $\eta = \{\eta_i, \forall i \in 1 ... N\}$ : random effects of all the subjects

- $m{\theta}$ : all the population parameters (typical values, BSV, BOV and residual variability)
- lacksquare  $\eta_i$ : random effects of subject i
- N: number of subjects
- $\quad \blacksquare \ \eta = \{\eta_i, \, \forall i \in 1 \dots N\} \colon$  random effects of all the subjects
- $y_i$ : observations of subject i

- $\theta$ : all the population parameters (typical values, BSV, BOV and residual variability)
- lacksquare  $\eta_i$ : random effects of subject i
- N: number of subjects
- $\eta = \{\eta_i, \, \forall i \in 1 \dots N\}$ : random effects of all the subjects
- $lackbox{} y_i$ : observations of subject i
- $\mathbf{x}_i$ : covariates of subject i



■  $p(y=y_i \mid \theta=\theta, \eta=\eta_i, x=x_i) = p(y_i \mid \theta, \eta_i, x_i)$ : likelihood of  $(\theta, \eta_i)$  given subject i's data  $(x_i, y_i)$ . Also known as the conditional probability of  $y_i$  given  $\theta, \eta_i, x_i$ . Or just the conditional likelihood of  $\eta_i$  given  $\theta$  and  $(x_i, y_i)$ .

- $p(y=y_i \mid \theta=\theta, \eta=\eta_i, x=x_i) = p(y_i \mid \theta, \eta_i, x_i)$ : likelihood of  $(\theta, \eta_i)$  given subject i's data  $(x_i, y_i)$ . Also known as the conditional probability of  $y_i$  given  $\theta, \eta_i, x_i$ . Or just the conditional likelihood of  $\eta_i$  given  $\theta$  and  $(x_i, y_i)$ .
- $p(\eta = \eta_i \mid \theta = \theta) = p(\eta_i \mid \theta)$ : prior probability of the random effects  $\eta_i$  given the population parameters  $\theta$ .

- $p(y=y_i \mid \theta=\theta, \eta=\eta_i, x=x_i) = p(y_i \mid \theta, \eta_i, x_i)$ : likelihood of  $(\theta, \eta_i)$  given subject i's data  $(x_i, y_i)$ . Also known as the conditional probability of  $y_i$  given  $\theta, \eta_i, x_i$ . Or just the conditional likelihood of  $\eta_i$  given  $\theta$  and  $(x_i, y_i)$ .
- $p(\eta = \eta_i \mid \theta = \theta) = p(\eta_i \mid \theta)$ : prior probability of the random effects  $\eta_i$  given the population parameters  $\theta$ .
- $p(y=y_i \mid \theta=\theta, x=x_i) = p(y_i \mid \theta, x_i) = \int p(y_i \mid \theta, \eta_i, x_i) \cdot p(\eta_i \mid \theta) \, d\eta_i$ : marginal likelihood of  $\theta$  given subject i's data  $(x_i, y_i)$ .



### Marginal likelihood maximization

$$\begin{split} \theta^* &= \arg \max_{\theta} \prod_{i=1}^{N} p(y_i \mid \theta, x_i) \\ &= \arg \max_{\theta} \prod_{i=1}^{N} \int p(y_i \mid \theta, \eta_i, x_i) \cdot p(\eta_i \mid \theta) \, d\eta_i \\ \text{EBE}_i &= \eta_i^* = \arg \max_{\eta_i} \Big( p(y_i \mid \theta = \theta^*, \eta_i, x_i) \cdot p(\eta_i \mid \theta = \theta^*) \Big) \end{split}$$

### General Laplace method

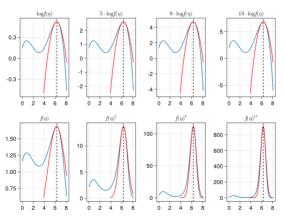
$$\int f(\eta)\,d\eta \approx f(\eta^*)\sqrt{(2\pi)^m/|-H|}$$

- f: a positive scalar-valued function of  $\eta$
- $\blacksquare$   $\eta$ : vector of m integration variables
- lacksquare  $\eta^*$ : global maximizer of  $\log f$ ,  $\frac{d \log f}{d \eta}(\eta^*) = 0$
- H: second derivative matrix of  $\log f$  wrt  $\eta$  at  $\eta^*$ , must be negative definite at  $\eta = \eta^*$
- | -H |: determinant of -H



### NLME Laplace method

Laplace uses a second order Taylor series approximation of  $\log f$  at  $\eta^*$ .



### NLME Laplace method

Consider the 2 local maximizers  $\eta_1$  (lower peak) and  $\eta_2$  (higher peak).

$$\begin{split} c &= \log f(\eta_2) - \log f(\eta_1) \\ n \cdot c &= n \cdot (\log f(\eta_2) - \log f(\eta_1)) \\ e^{n \cdot c} &= f(\eta_2)^n / f(\eta_1)^n \end{split}$$

#### Summary

Approximation error of  $n \cdot \log f$  away from the mode  $\eta^*$  is not significant as n increases.



### NLME Laplace method

- $\blacksquare$  There are N functions  $\{f_i: i=1\dots N\}$  to be integrated, one for each subject
- Laplace method

$$\int f_i(\eta_i)\,d\eta_i = f_i(\mathrm{EBE}_i)\sqrt{(2\pi)^m/|-H_i|}$$

■  $H_i$ : second derivative matrix of  $\log f_i$  wrt  $\eta_i$  at  $\mathsf{EBE}_i$ , must be negative definite at  $\eta_i = \mathsf{EBE}_i$ 



 $\mathsf{FOCE}(\mathsf{I})$  approximates the Hessian  $H_i$  for each subject i. Assume the following:

$$\begin{split} \log f_i(\eta_i) &= \log p(y_i \mid \theta, \eta_i, x_i) + \log p(\eta_i \mid \theta) \\ &= L_i(g_i(\eta_i)) + \log p(\eta_i \mid \theta) \end{split}$$

#### where:

- $g_i$  in the Gaussian case for example returns the vector of IPREDs  $\mu_i$  and the residual standard deviations  $\sigma_i$  (constant in the additive error model case), at all observed time points, and
- lacksquare  $L_i$  is the log probability of  $y_i$  given  $g_i$



 $g_i$  is usually the most expensive component of  $\log f_i$ , because it often involves solving a differential equation. So let's approximate it!

#### First order Taylor series approximation

$$\blacksquare$$
 FO 
$$g_i(\eta_i) \approx g_i(0) + \frac{dg_i}{d\eta_i}(0) \cdot \eta_i$$

FOCE(I)

$$g_i(\eta_i) \approx g_i(\mathrm{EBE}_i) + \frac{dg_i}{d\eta_i}(\mathrm{EBE}_i) \cdot (\eta_i - \mathrm{EBE}_i)$$



#### Summary

- FOCE(I) ensures that the approximation error in  $g_i$  (and  $\log f_i$  by extension) is low in the proximity of EBE $_i$ .
- FO does not ensure that so it only works well if:
  - EBE<sub>i</sub> is not far from 0, or
  - $g_i$  is close to linear in the interval  $[0, EBE_i]$ .
- FO requires a correction term in the Laplace method because the gradient of  $\log f_i$  wrt  $\eta_i$  at  $\eta_i = 0$  is not 0.



#### Chain rule for Hessians

$$(L_i \cdot g_i)''(\eta_i) = \frac{dg_i}{d\eta_i}^T \cdot \frac{\partial^2 L_i}{\partial g_i \cdot \partial g_i^T} \cdot \frac{dg_i}{d\eta_i} + \sum_{t=1}^d \left( \frac{\partial L_i}{\partial g_{i,t}} \cdot \underbrace{\frac{\partial^2 g_{i,t}}{\partial \eta_i \cdot \partial \eta_i^T}} \right)$$

where d is twice the number of observed time points (corresponding to  $\mu_i$  and  $\sigma_i$ ) and  $g_{i,t}$  is the  $t^{th}$  component of  $g_i$ .



#### Summary

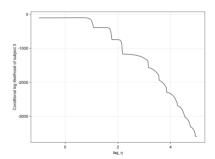
If  $g_i$  is linear in  $\eta_i$ :

$$lacksquare J_i = rac{\partial g_i}{\partial \eta_i}$$
 is constant

The Hessian simplifies to:

$$(L_i \cdot g_i)''(\eta_i) = J_i^T \cdot \frac{\partial^2 L_i}{\partial g_i \cdot \partial g_i^T} \cdot J_i$$

• One surprising advantage of FOCE(I) is that the Hessian approximation is often negative definite even when the exact Hessian is singular or not well defined at  $\eta_i = \eta_i^*$ .



- lacksquare  $J_i$  can be computed for each subject using finite difference at
  - $\bullet$   $\eta_i=0$  for FO, or
  - $\eta_i = \mathsf{EBE}_i \text{ for FOCE(I)}$
- For many data distributions,  $\frac{\partial^2 L_i}{\partial g_i \cdot \partial g_i^T}$  is both diagonal and has a closed form. Doesn't have to be Gaussian!

Recall

$$\begin{split} \log f_i(\eta_i) &= \log p(y_i \mid \theta, \eta_i, x_i) + \log p(\eta_i \mid \theta) \\ &= L_i(\underbrace{g_i(\eta_i)}_{\text{approx}}) + \log p(\eta_i \mid \theta) \end{split}$$

For many random effects distributions, the Hessian of  $\log p(\eta_i \mid \theta)$  wrt  $\eta_i$  has a closed form. Doesn't have to be Gaussian!

- Pumas FOCE supports a number of data distributions:
  - Continuous: Normal, LogNormal, Gamma, Exponential, Beta
  - Discrete: NegativeBinomial, Bernoulli, Binomial, Poisson, Categorical

- Pumas supports a number of random effect distributions:
  - Unbounded: Cauchy, Gumbel, Laplace, Logistic, Normal, NormalCanon, NormalInverseGaussian, PGeneralizedGaussian, TDist
  - Positive: BetaPrime, Chi, Chisq, Erlang, Exponential, Frechet, Gamma, InverseGamma, InverseGaussian, Kolmogorov, LogNormal, NoncentralChisq, Rayleigh, Weibull
  - Between 0 and 1: Beta, LogitNormal
  - Other bounded: Uniform, Arcsine, Biweight, Cosine, Epanechnikov, LogUniform, Semicircle, SymTriangularDist, Triweight





#### Summary

- In Pumas, FOCE is always "with interaction".
- Use FOCE if supported, otherwise use Laplace.
- Avoid FO.

## Weighted residuals

Assuming Gaussian random effects and error model, for each subject i, the conditional distribution  $p(y_i \mid x_i, \theta)$  is given by:

$$\begin{split} \eta_i &\sim \mathcal{N}(0, \Omega) \\ (\mu_i, \sigma_i) &= g(\eta_i; x_i) \\ y_i &\sim \mathcal{N}(\mu_i, \sigma_i) \end{split}$$

where  $g(\eta_i; x_i) = g_i(\eta_i)$  (same functional form g for all subjects).

#### Alternative representation

$$\begin{split} \eta_i &\sim \mathcal{N}(0, \Omega) \\ (\mu_i, \sigma_i) &= g(\eta_i; x_i) \\ \epsilon_{i,t} &\sim \mathcal{N}(0, 1) \\ y_{i,t} &= \mu_{i,t} + \sigma_{i,t} \cdot \epsilon_{i,t} \end{split}$$

where the t is the index for the number of observations per subject.

The machine learning (ML) community call this class of models:

- (Conditional) generative models, or
- Latent variable models

Congratulations, you have been doing ML this whole time!



- $p(y_i \mid x_i, \theta)$  is the distribution we sample from when doing a visual predictive check (VPC) to compare the distribution of simulated  $y_i$  to the distribution of observed  $y_i$ .
- The weighted residual is

$$\mathsf{WRES}_{i,t} = \frac{y_{i,t} - E[y_{i,t} \mid x_i]}{\sqrt{\mathsf{Var}[y_{i,t} \mid x_i]}} \sim \mathcal{N}(0,1)$$

- **Problem**:  $p(y_i \mid x_i, \theta)$  (in general) has no closed form mean and variance.
- Solution: let's approximate it!



### Approximate distribution of response

#### First order Taylor series approximation

FO

$$\mu_i(\eta_i) \approx \mu_i(0) + \frac{d\mu_i}{d\eta_i}(0) \cdot \eta_i$$
  
$$\sigma_i(\eta_i) \approx \sigma_i(0) + \frac{d\sigma_i}{d\eta_i}(0) \cdot \eta_i$$

FOCEI

$$\begin{split} \mu_i(\eta_i) &\approx \mu_i(\mathsf{EBE}_i) + \frac{d\mu_i}{d\eta_i}(\mathsf{EBE}_i) \cdot (\eta_i - \mathsf{EBE}_i) \\ \sigma_i(\eta_i) &\approx \sigma_i(\mathsf{EBE}_i) + \underbrace{\frac{d\sigma_i}{d\eta_i}(\mathsf{EBE}_i) \cdot (\eta_i - \mathsf{EBE}_i)}_{\neq 0 \text{ in general}} \end{split}$$

#### Refresh

For dependent random variables

$$\begin{split} E[X \cdot Y] &= E[X] \cdot E[Y] + \mathsf{Cov}[X,Y] \\ \mathsf{Var}[X + Y] &= \mathsf{Var}[X] + \mathsf{Var}[Y] + 2\mathsf{Cov}[X,Y] \end{split}$$

For indenpendent random variables

$$\mathsf{Var}[X \cdot Y] = \mathsf{Var}[X] \cdot \mathsf{Var}[Y] + \mathsf{Var}[X] \cdot E[Y]^2 + E[X]^2 \cdot \mathsf{Var}[Y]$$

#### Approximate means

FO

$$E[\mu_i] \approx \mu_i(0)$$
  
 $E[\sigma_i] \approx \sigma_i(0)$ 

FOCEI

$$\begin{split} E[\mu_i] \approx \mu_i(\mathrm{EBE}_i) - \frac{d\mu_i}{d\eta_i}(\mathrm{EBE}_i) \cdot \mathrm{EBE}_i \\ E[\sigma_i] \approx \sigma_i(\mathrm{EBE}_i) - \frac{d\sigma_i}{d\eta_i}(\mathrm{EBE}_i) \cdot \mathrm{EBE}_i \end{split}$$



#### Approximate variances

FO

$$\begin{split} \text{Var}[\mu_i] &\approx \frac{d\mu_i}{d\eta_i}(0) \cdot \Omega \cdot \frac{d\mu_i}{d\eta_i}(0)^T \\ \text{Var}[\sigma_i] &\approx \frac{d\sigma_i}{d\eta_i}(0) \cdot \Omega \cdot \frac{d\sigma_i}{d\eta_i}(0)^T \end{split}$$

FOCEI

$$\begin{split} \text{Var}[\mu_i] &\approx \frac{d\mu_i}{d\eta_i} (\text{EBE}_i) \cdot \Omega \cdot \frac{d\mu_i}{d\eta_i} (\text{EBE}_i)^T \\ \text{Var}[\sigma_i] &\approx \frac{d\sigma_i}{d\eta_i} (\text{EBE}_i) \cdot \Omega \cdot \frac{d\sigma_i}{d\eta_i} (\text{EBE}_i)^T \end{split}$$



Recall

$$y_{i,t} = \mu_{i,t} + \sigma_{i,t} \cdot \epsilon_{i,t}$$

Mean

$$\begin{split} E[y_{i,t} \mid x_i] &= E[\mu_{i,t}] + E[\sigma_{i,t}] \cdot \overbrace{E[\epsilon_{i,t}]}^0 \\ &= E[\mu_{i,t}] \end{split}$$

Recall

$$y_{i,t} = \mu_{i,t} + \sigma_{i,t} \cdot \epsilon_{i,t}$$

#### Variance

$$\begin{split} \operatorname{Var}[y_{i,t} \mid x_i] &= \operatorname{Var}[\mu_{i,t}] + \operatorname{Var}[\sigma_{i,t} \cdot \epsilon_{i,t}] + 2 \cdot \underbrace{\operatorname{Cov}[\mu_{i,t}, \sigma_{i,t} \cdot \epsilon_{i,t}]}_{0} \\ &= \operatorname{Var}[\mu_{i,t}] + \operatorname{Var}[\sigma_{i,t} \cdot \epsilon_{i,t}] \\ &= \operatorname{Var}[\mu_{i,t}] + \operatorname{Var}[\sigma_{i,t}] + E[\sigma_{i,t}]^2 \end{split}$$

$$\begin{aligned} \operatorname{Cov}[\mu_{i,t},\sigma_{i,t}\cdot\epsilon_{i,t}] &= \overbrace{E[\mu_{i,t}\cdot\sigma_{i,t}\cdot\epsilon_{i,t}]}^0 - E[\mu_{i,t}] \cdot \overbrace{E[\sigma_{i,t}\cdot\epsilon_{i,t}]}^0 \\ E[\mu_{i,t}\cdot\sigma_{i,t}\cdot\epsilon_{i,t}] &= E[\mu_{i,t}\cdot\sigma_{i,t}] \cdot \overbrace{E[\epsilon_{i,t}]}^0 = 0 \\ E[\sigma_{i,t}\cdot\epsilon_{i,t}] &= E[\sigma_{i,t}] \cdot \overbrace{E[\epsilon_{i,t}]}^0 = 0 \end{aligned}$$

#### Summary

After the FO/FOCEI approximation, we were able to obtain closed form approximations of  $E[y_{i,t}\mid x_i]$  and  $\mathrm{Var}[y_{i,t}\mid x_i].$ 

$$\mathrm{WRES}_{i,t} = \frac{y_{i,t} - E[y_{i,t} \mid x_i]}{\sqrt{\mathrm{Var}[y_{i,t} \mid x_i]}}$$

### Standard error estimation

# Goal

Estimate the covariance matrix of the estimator  $\theta^*$ :

$$\theta^* = \arg \max_{\theta} \prod_{i=1}^N p(y_i \mid \theta, x_i)$$

#### Asymptotic covariance

$$\begin{split} \theta^* &\sim \mathcal{N}(\theta_0, V) = \mathcal{N}(\theta_0, A^{-1}BA^{-1}) \\ A &= \sum_{i=1}^N -\frac{\partial^2 \log p(y_i \mid \theta, x_i)}{\partial \theta \cdot \partial \theta^T}(\theta_0) \\ B &= \sum_{i=1}^N \frac{\partial \log p(y_i \mid \theta, x_i)}{\partial \theta}(\theta_0) \times \frac{\partial \log p(y_i \mid \theta, x_i)}{\partial \theta}(\theta_0)^T \end{split}$$

where  $\theta_0$  is the set of unknown true parameters.



#### Estimated covariance

$$\begin{split} & \theta^* \overset{a}{\sim} \mathcal{N}(\theta_0, \hat{V}) = \mathcal{N}(\theta_0, \hat{A}^{-1} \hat{B} \hat{A}^{-1}) \\ & \hat{A} = \sum_{i=1}^N -\frac{\partial^2 \log p(y_i \mid \theta, x_i)}{\partial \theta \cdot \partial \theta^T} (\theta^*) \\ & \hat{B} = \sum_{i=1}^N \frac{\partial \log p(y_i \mid \theta, x_i)}{\partial \theta} (\theta^*) \times \frac{\partial \log p(y_i \mid \theta, x_i)}{\partial \theta} (\theta^*)^T \end{split}$$

#### Standard error estiamtes

The square root of the diagonal elements of  $\hat{V}$  are the standard error estimates of  $\theta^*$ .

## Computing $\hat{V}$

- ullet  $\hat{A}$  and  $\hat{B}$  can be approximated with finite difference
- $\blacksquare \log p(y_i \mid \theta, x_i)$  itself needs to be approximated with Laplace/FOCE/FO
- $\hat{A}^{-1} \cdot \hat{B} \cdot \hat{A}^{-1}$  is computed using a generalized eigenvalue problem.

### Computing $\hat{V}$

Consider the following generalized eigenvalue problem:

$$\begin{split} \hat{B} \cdot U &= \hat{A} \cdot U \cdot \Lambda \\ I &= U^T \cdot \hat{A} \cdot U \end{split}$$

where U is the matrix of generalized eigenvectors and  $\Lambda$  is the diagonal matrix of generalized eigenvalues.

## Computing $\hat{V}$

The inverse of the matrix of eigenvectors U is obtained from the following constraint on U:

$$\begin{split} (U^T \cdot \hat{A}) \cdot U &= I \\ U^{-1} &= U^T \cdot \hat{A} \end{split}$$

## Computing $\hat{V}$

The following identity is true:

$$\begin{split} \hat{B} \cdot U &= \hat{A} \cdot U \cdot \Lambda \\ \hat{A}^{-1} \cdot \hat{B} \cdot U &= U \cdot \Lambda \\ \hat{A}^{-1} \cdot \hat{B} &= U \cdot \Lambda \cdot U^{-1} \\ \hat{A}^{-1} \cdot \hat{B} &= U \cdot \Lambda \cdot U^T \cdot \hat{A} \\ \hat{V} &= \hat{A}^{-1} \cdot \hat{B} \cdot \hat{A}^{-1} &= U \cdot \Lambda \cdot U^T \end{split}$$

#### Failed estimator

- If the computed  $\hat{A}$  is: a) singular, b) near singular, or c) has negative eigenvalues, the sandwich estimator will fail.
- This is a sign of poor identifiability of at least 1 parameter and/or significant numerical errors.
- Even if a single (IIV) parameter is not identifiable given the data,  $\hat{A}$  will be singular.



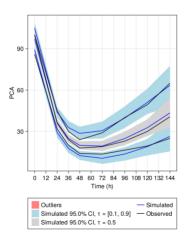
#### Failed estimator

Numerical errors in the finite difference or Laplace/FOCE/FO can also cause the computed approximate  $\hat{A}$  to be singular (or have small negative eigenvalues) even when the exact matrix  $\hat{A}$  may be only near singular and positive definite.

Continuous visual predictive check

## Continuous visual predictive check

## Example





## Continuous VPC procedure

- I Simulate a synthetic population a given number of samples (samples, default 499).
- 2 Stratify the observed and simulated populations by the stratification variable.
- 3 For each simulated population stratum, do smoothed quantile regression at nnodes nodes picked from the the data.
  - Default quantiles: 0.1, 0.5 and 0.9.
  - Default nnodes: 11
  - Default smoothing bandwidth: 2.0



## Continuous VPC procedure

- 4 Find the (hyper-)quantiles of the per-scenario population quantiles within each stratum,
  - Hyper-quantiles:
    - (1 level) / 2
    - 0.5 (simquantile\_medians hidden by default)
    - (1 + level) / 2
  - Default level: 0.95
- 5 For each observed population stratum, repeat step 3.
- 6 For each stratum, plot the population's quantiles and the hyper-quantiles of each simulated quantile.



└─Time to event models

#### Time to event models



#### **Definitions**

Instantaneous hazard

$$\lambda(t) > 0$$

Cumulative hazard

$$\Lambda(t) = \int_0^t \lambda(t') \, dt'$$

lacksquare Survival function: probability of survival up to time t

$$S(t) = \exp(-\Lambda(t))$$

#### **Definitions**

lacktriangle Failure function: probability of death/failure before time t

$$F(t) = 1 - S(t)$$

Probability density function of time of death t

$$f(t) = \frac{dF}{dt} = \lambda(t) \cdot \exp(-\Lambda(t))$$

lacksquare Expected time of death E[t]

$$E[t] = \int_0^\infty t \cdot f(t) dt = \int_0^\infty S(t) dt$$

## Log likelihood

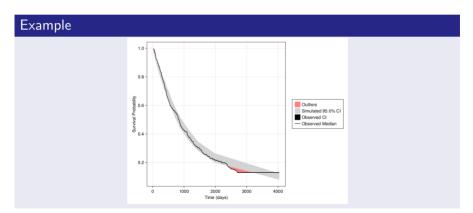
The log likelihood for censored survival data is given by the following 2 formulas:

lacksquare For censored subjects at time t (patient survived until time t)

$$\log \operatorname{likelihood} = \log S(t) = -\Lambda(t)$$

For subjects dead at time t

$$\log \text{likelihood} = \log f(t) = \log \lambda(t) - \Lambda(t)$$



- Simulate a synthetic population a given number of samples (samples, default 499). For each subject:
  - I Evaluate the cumulative hazard function  $\Lambda$  at nT (default 10) time points between minT and maxT.
  - 2 Use a cubic spline to interpolate between the  $\Lambda$  values.
  - 3 Use inverse CDF transform sampling to sample the time of death from the cumulative hazard function.

- 2 Stratify the observed and simulated populations by the stratification variable.
- 3 For each simulated population stratum:
  - **I** Estimate the Kaplan Meier (KM) curve.  $d_i$  is the number of deaths at  $t_i$  and  $n_i$  is the number of people at risk at time  $t_i$ .

$$\hat{S}(t) = \prod_{i:t_i < t} \left(1 - \frac{d_i}{n_i}\right)$$

- 2 Combine all simulated populations' KM curves into one data frame.
- 3 Do quantile regression with smoothing to get smooth curves for the quantiles at a number of nodes nnodes (default 11).



- 4 For each observed population stratum, estimate the KM curve.
- 5 Plot the observed KM curve against the smoothed quantiles for each stratum.

## Inverse CDF sampling

■ If  $R \sim \mathsf{Uniform}(0,1)$ , then  $-\log(1-R) \sim \mathsf{Exponential}(1)$ .

$$F(t) \le R$$
 
$$1 - S(t) \le R$$
 
$$\exp(-\Lambda(t)) \ge 1 - R$$
 
$$\Lambda(t) \le -\log(1 - R)$$

■ The sample t is obtained using a root finding algorithm to find the root for  $\Lambda(t) = -\log(1-R)$ .

