1 SAMGEP E-step Algorithm

Our data, dat, is a data frame consisting of T_i timepoints for each of $i = \{1, \dots N\}$ patients. Features include:

1. ID: Patient IDs
2. T: Timepoints
3. Tlog: $\log T + 1$
4. H: Measure of a patient's healthcare utilization
5. Hlog: $H + 1$
6. $\{X_1, X_2, \dots X_p\}$: p features

Other variables include:

1. likeModel: List consisting of
   (a) β: $(p \times 8)$-dimensional matrix of estimated GLS coefficients
   (b) α: p-dimensional vector of estimated variance exponents
   (c) σ: p-dimensional vector of estimated model standard deviations
   (d) std.errors: $(8 \times 8 \times p)$-dimensional array of GLS coefficient standard errors
   (e) corrMat: $(p \times p)$-dimensional feature correlation matrix

The E-step algorithm goes as follows:

Algorithm 1: SAMGEP E-step

for $i = 1$ *to N patients* **do**

  Compute $\mu \mid \mathbf{Y} = \mathbf{0}, H_i, \mathbf{T}_i$

  Compute $\mu \mid \mathbf{Y} = \mathbf{1}, H_i, \mathbf{T}_i$

  Compute $\sigma^2_{like} = \sigma^2 \times H^{2\alpha}$

  Compute σ^2_{prior}

  Compute $\sigma = \sqrt{\sigma^2_{prior} + \sigma^2_{like}}$

  Compute $\Sigma = [corrMat]\sigma^T\sigma$ **for** $t = 1$ *to* T_i *timepoints* **do**

    if *length*$(T_i) = 1$ **then**

      Compute log prior of $Y_{i,t}$

      Compute log likelihood of $\mathbf{X}_{i,t} \mid Y_{i,t}$

      Compute posterior of $Y_{i,t} = 1 \mid \mathbf{X}_{i,t}$

    else if $t = 1$ **then**

      Compute log prior of $\{Y_{i,t}, Y_{i,t+1}\}$

      Compute log likelihood of $\mathbf{X}_{i,t}, \mathbf{X}_{i,t+1} \mid \{Y_{i,t}, Y_{i,t+1}\}$

      Compute posterior of $Y_{i,t} = 1 \mid \mathbf{X}_{i,t}, \mathbf{X}_{i,t+1}$

    else if $t = T_i$ **then**

      Compute log prior of $\{Y_{i,t-1}, Y_{i,t}\}$

      Compute log likelihood of $\mathbf{X}_{i,t-1}, \mathbf{X}_{i,t} \mid \{Y_{i,t-1}, Y_{i,t}\}$

      Compute posterior of $Y_{i,t} = 1 \mid \mathbf{X}_{i,t-1}, \mathbf{X}_{i,t}$

    else

      Compute log prior of $\{Y_{i,t-1}, Y_{i,t}, Y_{i,t+1}\}$

      Compute log likelihood of $\mathbf{X}_{i,t-1}, \mathbf{X}_{i,t}, \mathbf{X}_{i,t+1} \mid \{Y_{i,t-1}, Y_{i,t}, Y_{i,t+1}\}$

      Compute posterior of $Y_{i,t} = 1 \mid \mathbf{X}_{i,t-1}, \mathbf{X}_{i,t}, \mathbf{X}_{i,t+1}$

    Append estimated posterior and transition probabilities to output vector