**Description:**

The following programs and data were utilized in the manuscript "Joint analysis of longitudinal and survival data measured on nested time-scales using shared parameter models: an application to fecundity data" by Alexander C. McLain, Rajeshwari Sundaram, and Germaine M. Buck Louis. The programs were written by A McLain. If you have any questions regarding the programs please contact A McLain (mclaina@mailbox.sc.edu). The analysis was originally completed in May of 2012 with R v2.15.0.The data in this file is:• Oxford\_train\_day\_level.csv and Oxford\_test\_day\_level.csv: Day level data used to estimate parameters and prediction error, respectively. Each row of this file corresponds to one day for one women. The variables included are: o Y = Intercourse indicator. o ind = Woman ID number. o Intercept = 1 for all rows. o FemaleAge = the woman's age minus 31. o Cigarettes/ Alcohol = indicator that the woman consumed at least one cigarette/alcoholic beverage on a given day. o Period = indicator that menstrual bleeding occurred on a given day. o Weekend = indicator that the given day was a Fri/Sat/Sun. o Parity = Indicator of one previous live birth. o Parity\*Age = interaction between parity and age. o Lag = Lag 1 intercourse value. o Cyc-1 = cycle number minus 1. o Day = day relative to ovulation (if known) o day\_miss = indicator that the day relative to ovulation was known. o RE\_ind = indicator of which rows will be used for the empirical bayes estimates (for prediction only).• Oxford\_train\_women\_level.csv and Oxford\_test\_women\_level.csv: Woman level data used to estimate parameters and prediction error, respectively. Each row of this file corresponds to one woman. The variables included are: o Age = standardized value for female age. o Parity = Indicator of one previous live birth. o TTP = minimum of the woman’s time-to-pregnancy and censoring. o CEN = indicator of censoring. o Pre\_ovJ = length of the follicular phase for cycle J. Set to ‘0’ if unknown and ‘-1’ if unobserved. o ind\_po = indicator that the woman had a missing follicular phase length.This file contains 3 programs:• R\_programs.R = contains all of the necessary R functions.• Oxford.Anal.R = used to estimate the parameters (will take ~ 3 weeks to run). • Oxford\_prediction.R = used to implement the prediction procedure.

**functions in R\_code** R\_programs.R

mjm.like.aq {}, mjm.EB {}

**USAGE**

mjm.like.aq(par,Dat,BK) in conjunction with an optimization program, such as optim or nls.

mjm.EB(par,Dat,nodes,knots,BK): used to get empirical Bayes estimates of the random effects.

**ARGUMENTS**

par: The values of the following parameters

beta: a vector of length q (= dim(X)[2]) of regression coefficients to be used in the intercourse model.

sigma: log standard deviation of the baseline random effect in the intercourse model.

sigK: log standard deviation of the peakedness random effect in the intercourse model.

sigY: log standard deviation of the baseline random effect in the pre-ovular model.

xi: vector of 2 parameters for the effect the intercourse random effects have on the TTP model.

lambda: vector of max(T) baseline intercepts (i.e., an intercept for each cycle)

betaT: a vector of length pT (i.e., dim(X\_T)[2]) to be used for regression coefficients in the TTP model

phi: coefficients for the cubic B-splines function.

rho: a vector coefficients of length pk for the peakedness function.

cov12: correlation between the baseline and peakedness random effects.

mu\_po,sigma\_po: coefficients for the location and scale of the lognormal preovular model.

Dat: This is a list of all data. Which includes:

$Y: a vector of intercourse indicators (length L= total number of days observed),

$ind: a vector of indicators of person number, should be consecutively numbered 1:n (length L),

$ind\_po: a vector of indicators of all cycles being observed (length n=number of subjects),

$X: intercourse covariate matrix (dimension L x q),

$X\_P: intercourse peakedness covariate matrix (dimension L x pk),

$day: vector of day relative to ovulation (length L, set to the day of cycle if unknown),

$day\_miss: vector of indicators that the day relative to ovulation known (length L),

$T: a vector of TTP or Censoring values (length n),

$CEN: a vector of censoring indicators (length n),

$X\_T: a matrix of covariates for the TTP model (dimension n x pT),

$nij: a matrix of pre-ovular lengths (dimension n x max(T), set to -1 if unobserved).

nodes: number of quadrature nodes.

knots: the values of the knots for the cubic b-spline function.

BK: boundary knots for the cubic b-spline function.

**OUTPUT**

The output of function mjm.like.aq(par,Dat,BK) is a sum(log(likelihood))

The output of function mjm.EB(par,Dat,nodes,knots,BK) is a n x 3 matrix of empirical Bayes Random Effects.

**DETAILS**

The function mjm.like.aq(par,Dat,BK) evaluates a joint model of intercourse and TTP. Pre-ovular length is also modeled, since the intercourse model is a function of pre-ovular length and pre-ovular length presents informative missingness. The likelihood is evaluated using adaptive Gaussian quadrature. First the people with full pre-ovular information are evaluated. Then those with missing pre-ovular information are evaluated

The function mjm.EB(par,Dat,nodes,knots,BK) produces empirical Bayes estimates of the Random Effects.

**Authors:**

Alexander C. McLain, Rajeshwari Sundaram, and Germaine M. Buck Louis

[mclaina@mailbox.sc.edu](mailto:mclaina@mailbox.sc.edu)

**Examples:**

See following R-programs:

* Oxford.Anal.R : used to estimate the parameters (will take ~ 3 weeks to run).
* Oxford\_prediction.R : used to implement the prediction procedure.