Joint modelling workshop: Software training session Solutions

Ruwanthi Kolamunnage-Dona

6th September 2017, University of Calgary, Canada

Question 1

Question instructions

```
# Note: Load the joint modelling packages first
library(joineR)
library(joineRML)
data(pbc2)
```

Part (a)

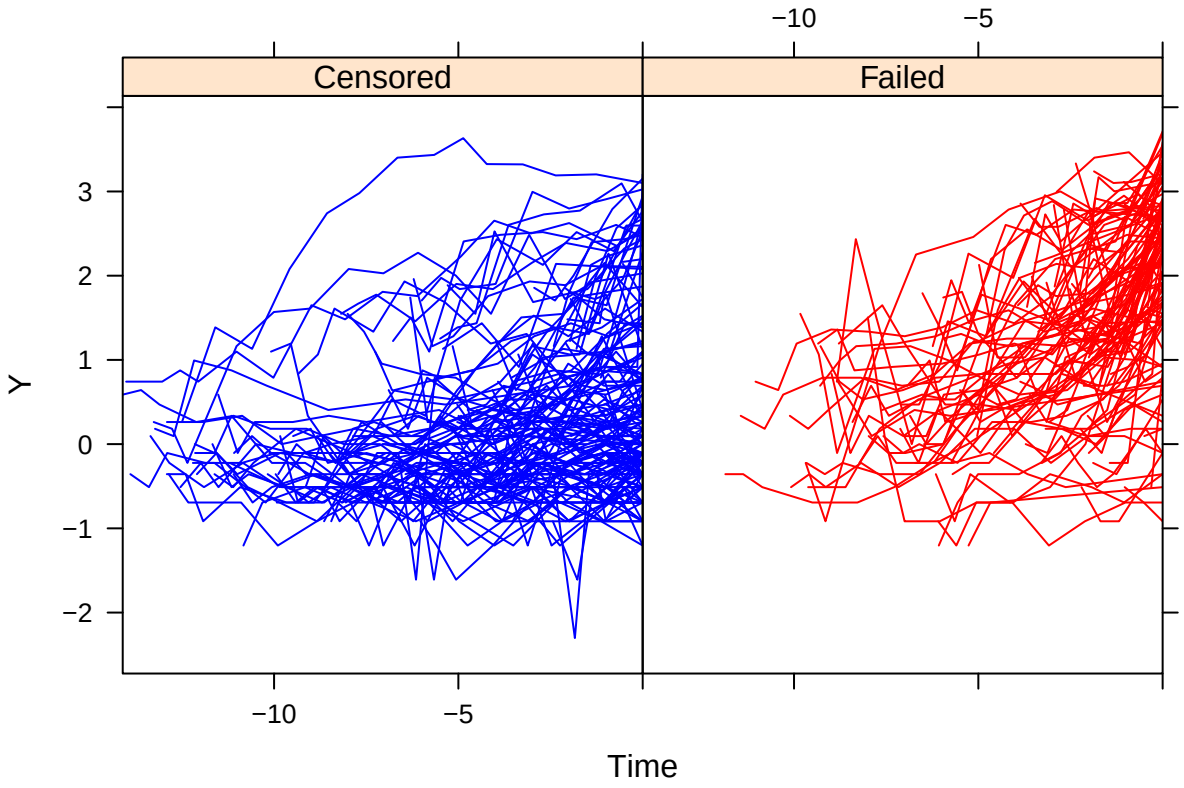
```
pbc2$log.b <- log(pbc2$serBilir)
pbc.surv <- UniqueVariables(data = pbc2,
                            var.col = c("years", "status2"),
                            id.col = "id")
pbc.long <- pbc2[, c("id", "year", "log.b")]
pbc.cov <- UniqueVariables(data = pbc2,
                           var.col = "drug",
                           id.col = "id")

pbc.jd <- jointdata(longitudinal = pbc.long,
                    baseline = pbc.cov,
                    survival = pbc.surv,
                    id.col = "id",
                    time.col = "year")
```

Note: In this dataset "year" is the longitudinal time and "years" is the survival time.

Part (b)

```
jointplot(pbc.jd,
          Y.col = "log.b",
          Cens.col = "status2")
```



Part (c)

```
fit1 <- joint(data = pbc.jd,
    long.formula = log.b ~ year + drug,
    surv.formula = Surv(years, status2) ~ drug,
    model = "intslope")

summary(fit1)
```

```
##
## Call:
## joint(data = pbc.jd, long.formula = log.b ~ year + drug, surv.formula = Surv(years,
##     status2) ~ drug, model = "intslope")
##
## Random effects joint model
##  Data: pbc.jd
##  Log-likelihood: -2270.467
##
## Longitudinal sub-model fixed effects: log.b ~ year + drug
## (Intercept)    0.5514713
## year           0.1806740
## drugD-penicil -0.1298256
##
```

```
## Survival sub-model fixed effects: Surv(years, status2) ~ drug
## drugD-penicil -0.08469162
##
## Latent association:
## gamma_0 1.231331
##
## Variance components:
##        U_0        U_1   Residual
## 1.00011123 0.03206622 0.12063642
##
## Convergence at iteration: 17
##
## Number of observations: 1945
## Number of groups: 312
```

Part (d)

```
fit2 <- update(fit1, model = "int")
summary(fit2)
```

```
##
## Call:
## joint(data = pbc.jd, long.formula = log.b ~ year + drug, surv.formula = Surv(years,
##     status2) ~ drug, model = "int")
##
## Random effects joint model
##  Data: pbc.jd
##  Log-likelihood: -2658.463
##
## Longitudinal sub-model fixed effects: log.b ~ year + drug
## (Intercept)    0.62660069
## year           0.09808779
## drugD-penicil -0.11108700
##
## Survival sub-model fixed effects: Surv(years, status2) ~ drug
## drugD-penicil -0.102991
##
## Latent association:
## gamma_0 1.282793
##
## Variance components:
##       U_0  Residual
## 1.2231540 0.2412634
##
## Convergence at iteration: 17
##
## Number of observations: 1945
## Number of groups: 312
```

```
fit3 <- update(fit1, model = "quad")
summary(fit3)
```

```
##
## Call:
## joint(data = pbc.jd, long.formula = log.b ~ year + drug, surv.formula = Surv(years,
##     status2) ~ drug, model = "quad")
##
## Random effects joint model
##  Data: pbc.jd
##  Log-likelihood: -2170.061
##
## Longitudinal sub-model fixed effects: log.b ~ year + drug
## (Intercept)    0.5691374
## year           0.1709158
## drugD-penicil -0.1269617
##
## Survival sub-model fixed effects: Surv(years, status2) ~ drug
## drugD-penicil -0.1212103
##
## Latent association:
## gamma_0 1.296074
##
## Variance components:
##          U_0          U_1          U_2     Residual
## 0.9938883681 0.0969264151 0.0006602518 0.0918742041
##
## Convergence at iteration: 19
##
## Number of observations: 1945
## Number of groups: 312
```

Part (e)

```
fit4 <- update(fit1, sepassoc = TRUE)
summary(fit4)
```

```
##
## Call:
## joint(data = pbc.jd, long.formula = log.b ~ year + drug, surv.formula = Surv(years,
##     status2) ~ drug, model = "intslope", sepassoc = TRUE)
##
## Random effects joint model
##  Data: pbc.jd
##  Log-likelihood: -2270.136
##
## Longitudinal sub-model fixed effects: log.b ~ year + drug
## (Intercept)    0.5488091
## year           0.1817519
## drugD-penicil -0.1307232
##
## Survival sub-model fixed effects: Surv(years, status2) ~ drug
## drugD-penicil -0.08747518
##
## Latent association:
```

```
## gamma_0 1.178347
## gamma_1 1.355136
##
## Variance components:
##        U_0        U_1   Residual
## 0.99942772 0.03228746 0.12066524
##
## Convergence at iteration: 19
##
## Number of observations: 1945
## Number of groups: 312
```

Part (f)

```
fit1.se <- jointSE(fit1, n.boot = 100)
fit1.se
```

```
##      Component     Parameter Estimate     SE 95%Lower 95%Upper
## 1 Longitudinal   (Intercept)   0.5515 0.0887   0.3794   0.7255
## 2                      year   0.1807 0.0150   0.1431   0.2033
## 3              drugD-penicil  -0.1298 0.1127  -0.3642   0.0726
## 4      Failure drugD-penicil  -0.0847 0.2457  -0.5833   0.3596
## 5  Association       gamma_0   1.2313 0.1103   1.0278   1.4339
## 6     Variance           U_0   1.0001 0.0824   0.8276   1.1670
## 7                        U_1   0.0321 0.0054   0.0220   0.0455
## 8                   Residual   0.1206 0.0105   0.0997   0.1386
```

Question 2

Part (a)

```
fit5 <- mjoint(
  formLongFixed = list("bil" = log(serBilir) ~ year + drug,
                       "alb" = albumin ~ year + drug),
  formLongRandom = list("bil" = ~ year | id,
                        "alb" = ~ year | id),
  formSurv = Surv(years, status2) ~ drug,
  data = pbc2,
  timeVar = "year",
  control = list(burnin = 300),
  verbose = FALSE)
```

```
## Running multivariate LMM EM algorithm to establish initial parameters...
```

```
## Finished multivariate LMM EM algorithm...
```

```
## EM algorithm has converged!
```

```
## Calculating post model fit statistics...
```

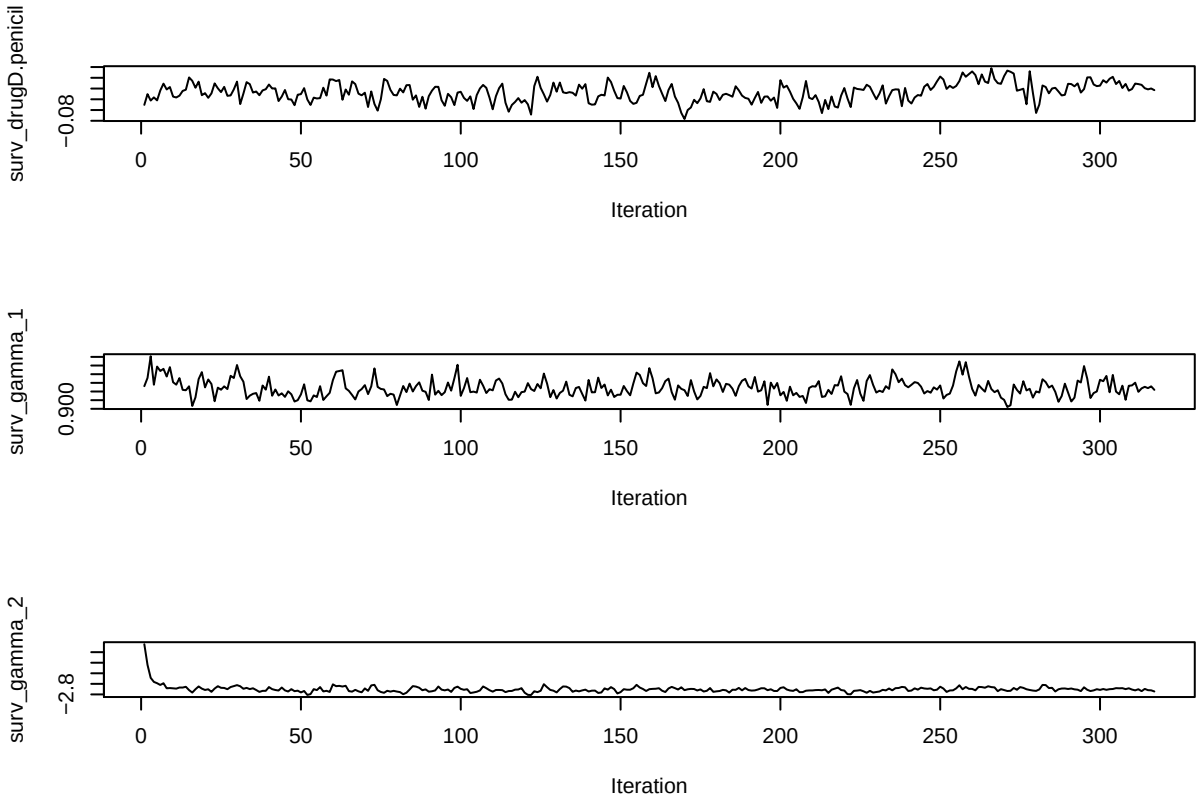
```
summary(fit5)
```

```
##
## Call:
## mjoint(formLongFixed = list(bil = log(serBilir) ~ year + drug,
##     alb = albumin ~ year + drug), formLongRandom = list(bil = ~year |
##     id, alb = ~year | id), formSurv = Surv(years, status2) ~
##     drug, data = pbc2, timeVar = "year", verbose = FALSE, control = list(burnin = 300))
##
## Data Descriptives:
##
## Event Process
##     Number of subjects: 312
##     Number of events: 140 (44.9%)
##
## Longitudinal Process
##     Number of longitudinal outcomes: K = 2
##     Number of observations:
##       Outcome 1 (bil): n = 1945
##       Outcome 2 (alb): n = 1945
##
## Joint Model Summary:
##
## Longitudinal Process: Multivariate linear mixed-effects model
##      log(serBilir) ~ year + drug, random = ~year | id
##      albumin ~ year + drug, random = ~year | id
## Event Process: Cox proportional hazards model
##      Surv(years, status2) ~ drug
## Model fit statistics:
##    log.Lik      AIC      BIC
##  -3087.871 6217.743 6296.346
##
## Variance Components:
##
## Random effects variance covariance matrix
##               (Intercept)_1    year_1 (Intercept)_2     year_2
## (Intercept)_1      0.993260  0.085397    -0.1922100 -0.0351190
## year_1             0.085397  0.034832    -0.0233150 -0.0110590
## (Intercept)_2     -0.192210 -0.023315     0.1251700  0.0056973
## year_2            -0.035119 -0.011059     0.0056973  0.0050031
##   Standard Deviations: 0.99662 0.18663 0.3538 0.070732
##
## Residual standard errors(s):
##  sigma2_1  sigma2_2
## 0.3467389 0.3195017
##
## Coefficient Estimates:
##
## Longitudinal sub-model:
##                   Value Std.Err  z-value p-value
## (Intercept)_1    0.5386  0.0752   7.1630 <0.0001
## year_1           0.1937  0.0104  18.6069 <0.0001
```

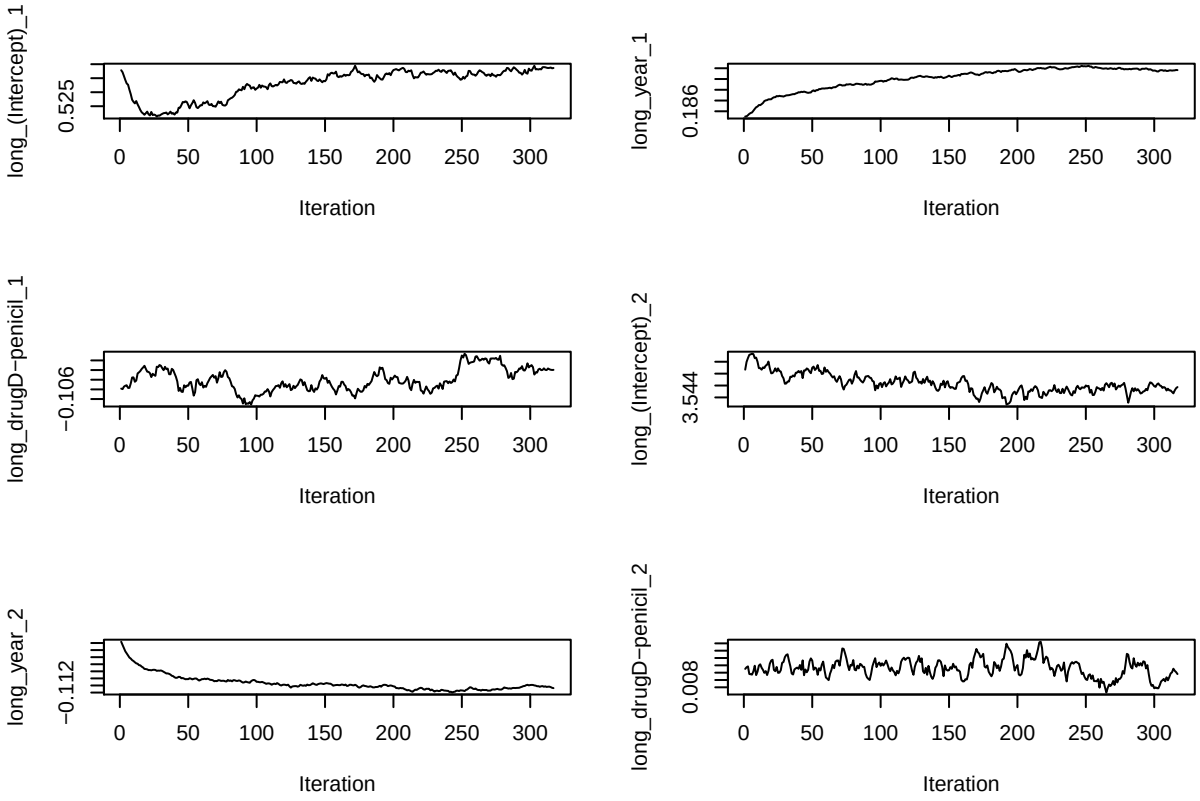
```
## drugD-penicil_1 -0.1000  0.1157  -0.8640  0.3876
## (Intercept)_2    3.5457  0.0320 110.9618 <0.0001
## year_2          -0.1114  0.0049 -22.8207 <0.0001
## drugD-penicil_2  0.0098  0.0442   0.2218  0.8245
##
## Time-to-event sub-model:
##                 Value Std.Err z-value p-value
## drugD.penicil -0.0515  0.2964 -0.1738  0.8621
## gamma_1        0.9109  0.1237  7.3669 <0.0001
## gamma_2       -2.7722  0.3369 -8.2278 <0.0001
##
## Algorithm Summary:
##     Total computational time: 5 mins
##     EM algorithm computational time: 4.6 mins
##     Convergence status: converged
##     Convergence criterion: sas
##     Final Monte Carlo sample size: 742
##     Standard errors calculated using method: approx
```

Part (b)

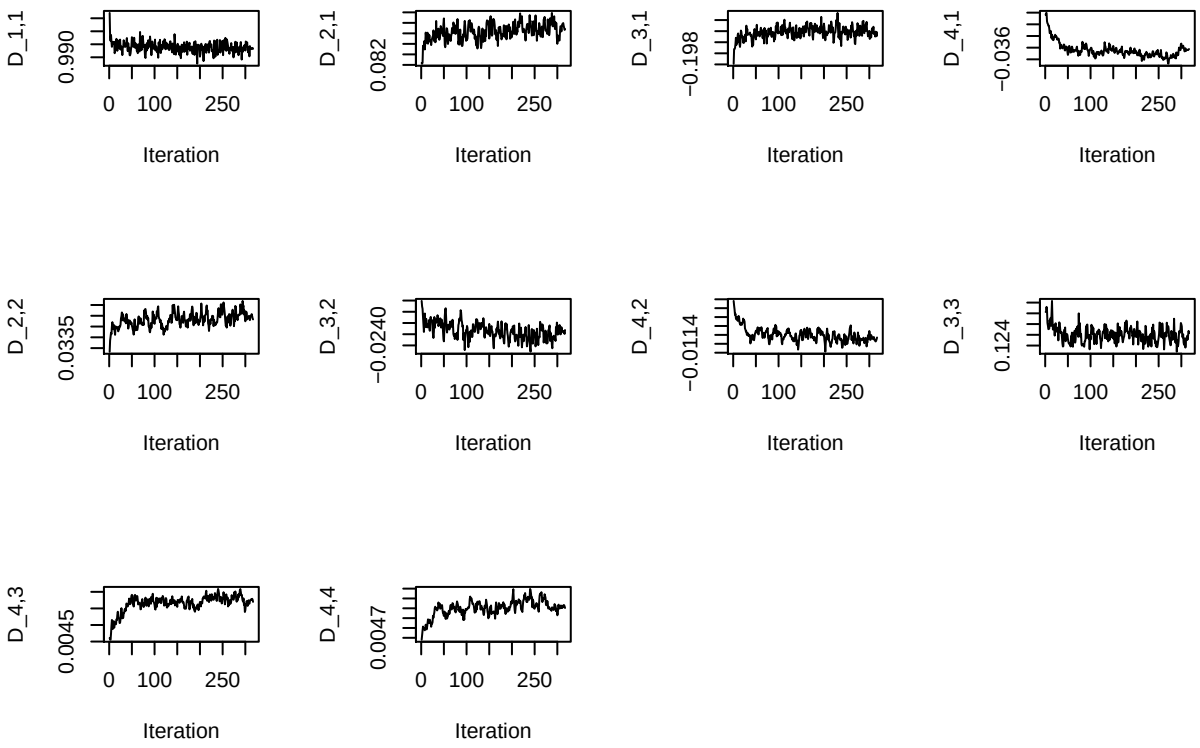
```
plot(fit5, param = "gamma")
```



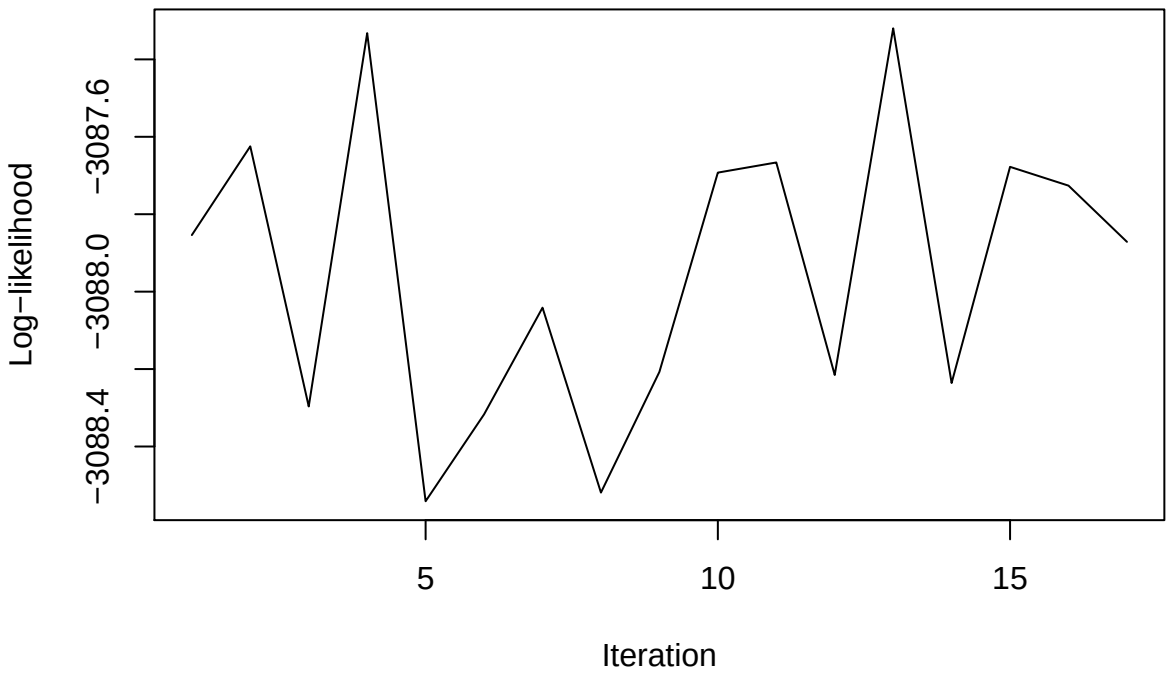
```
plot(fit5, param = "beta")
```




```
plot(fit5, param = "D")
```




```
plot(fit5, param = "loglik", discard = TRUE) # notice use of discard argument
```



Part (c)

```
fit6 <- update(fit5, formLongRandom = list("bil" = ~ 1 | id,
                                           "alb" = ~ 1 | id))
```

```
## Running multivariate LMM EM algorithm to establish initial parameters...

## Finished multivariate LMM EM algorithm...

## EM algorithm has converged!

## Calculating post model fit statistics...
```

```
AIC(fit5)
```

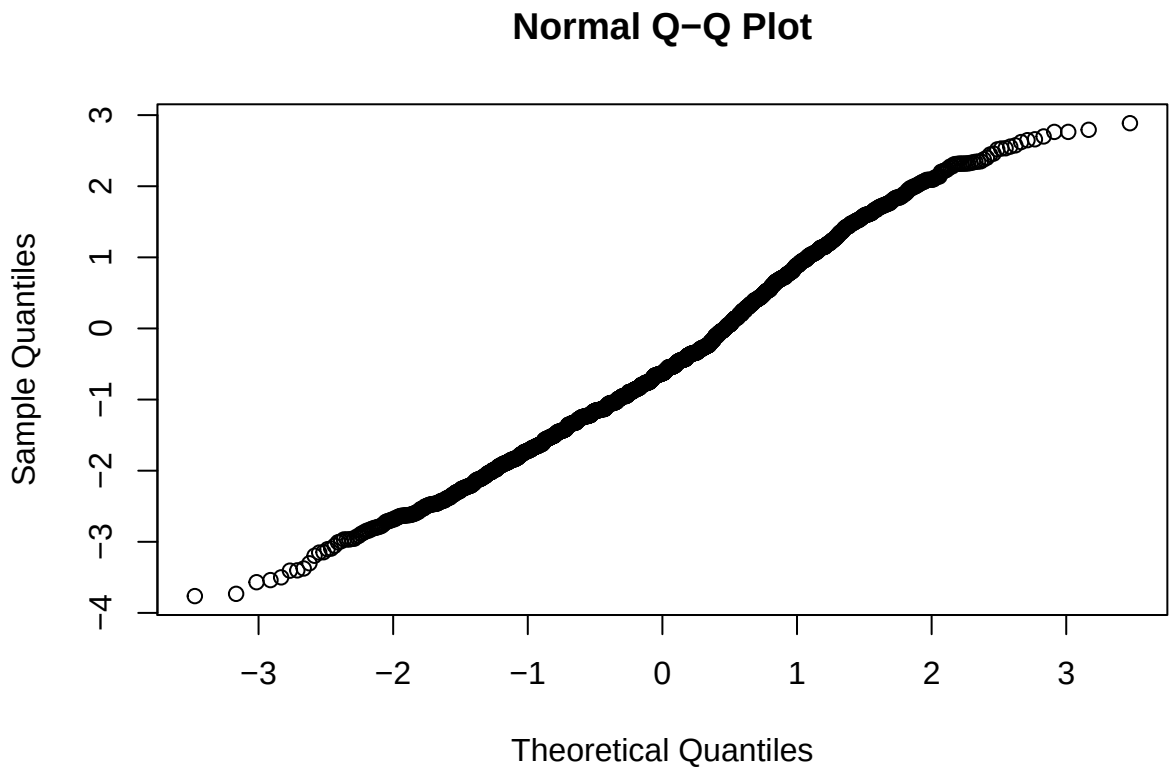
```
## [1] 6217.743
```

```
AIC(fit6)
```

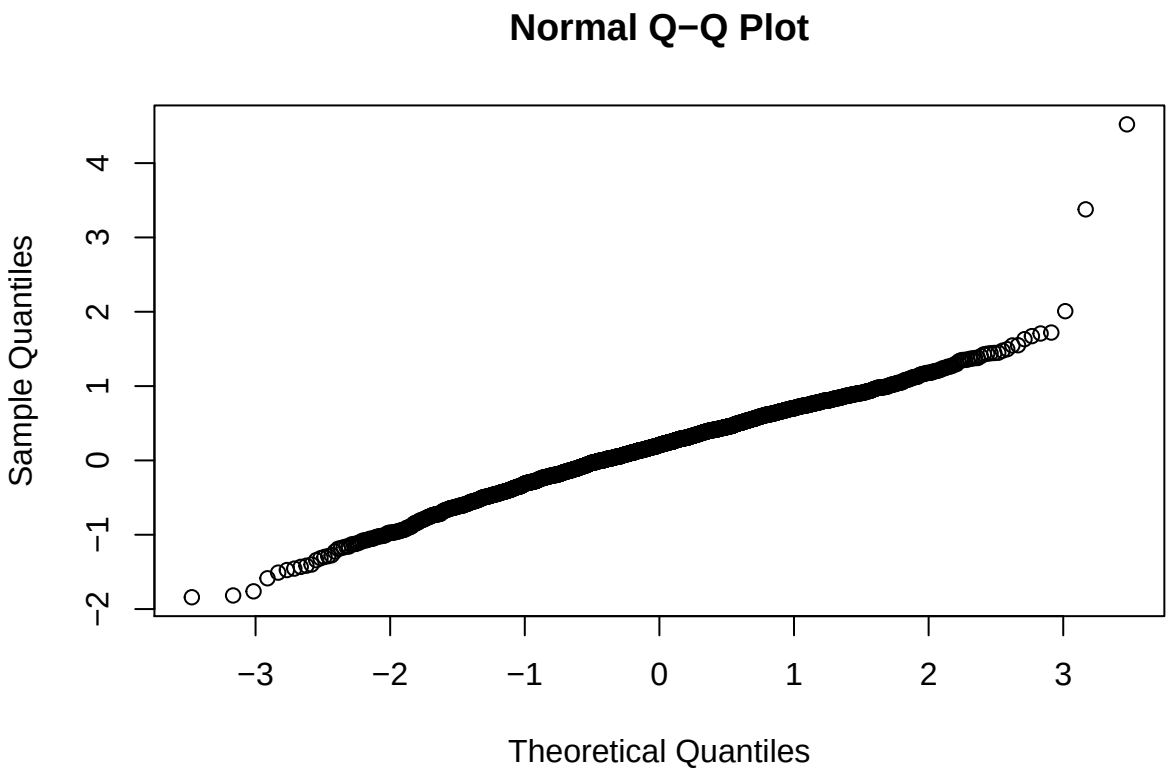
```
## [1] 7175.863
```

Part (d)

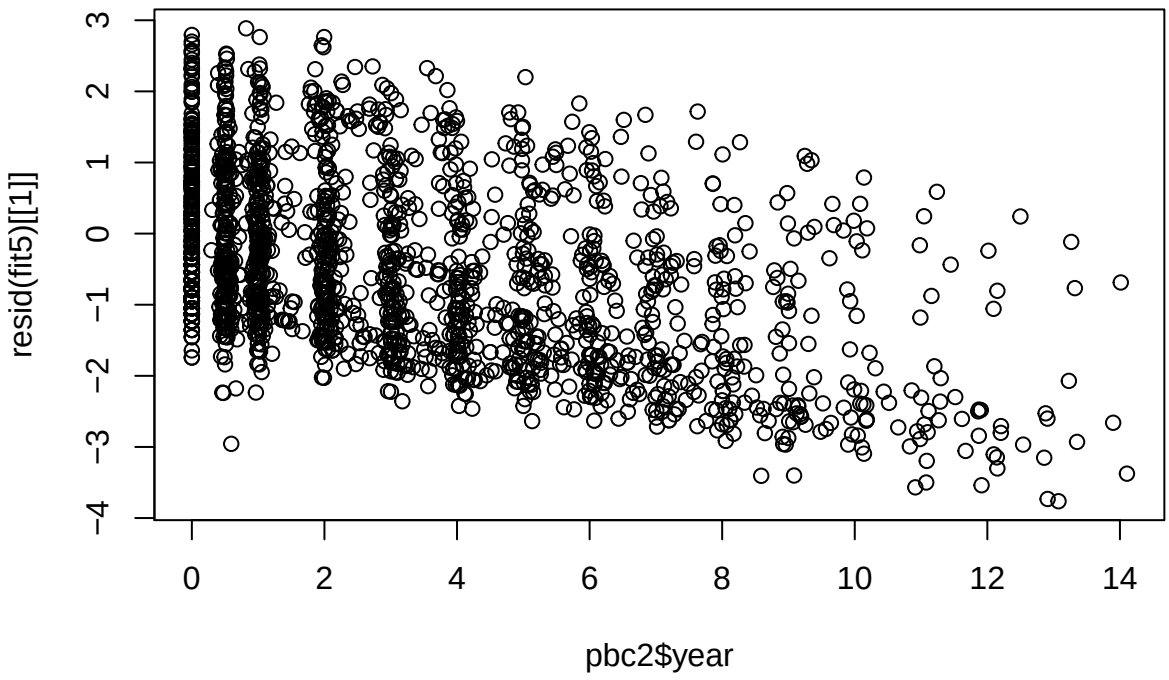
```
qqnorm(resid(fit5)[[1]])
```



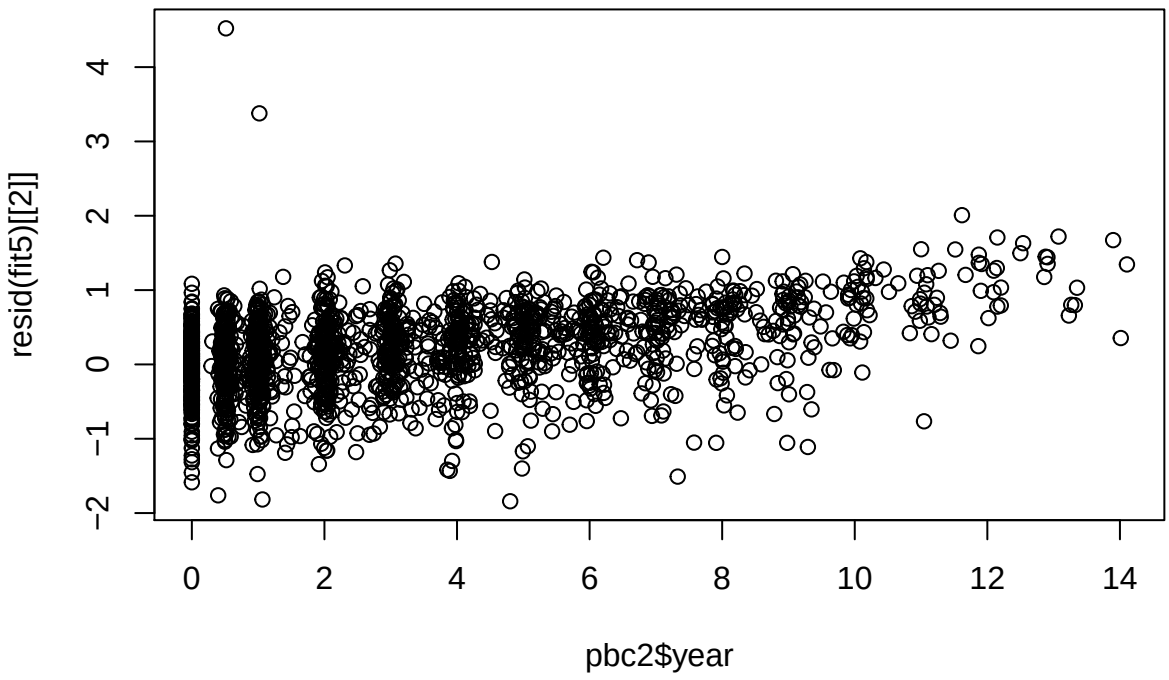

```
qqnorm(resid(fit5)[[2]])
```




```
plot(pbc2$year, resid(fit5)[[1]])
```




```
plot(pbc2$year, resid(fit5)[[2]])
```



Question 3

```
data(pbc2)
placebo <- subset(pbc2, drug == "placebo")
```

Part (a)

```
set.seed(123)
fit6 <- mjoint(
  formLongFixed = list("bil" = log(serBilir) ~ year,
                       "alb" = albumin ~ year),
  formLongRandom = list("bil" = ~ year | id,
                        "alb" = ~ year | id),
  formSurv = Surv(years, status2) ~ age,
  data = placebo,
  timeVar = "year",
  control = list(burnin = 300),
  verbose = FALSE)
```

```
## Running multivariate LMM EM algorithm to establish initial parameters...
```

```
## Finished multivariate LMM EM algorithm...

## EM algorithm has converged!

## Calculating post model fit statistics...
```

```
summary(fit6)
```

```
##
## Call:
## mjoint(formLongFixed = list(bil = log(serBilir) ~ year, alb = albumin ~
##     year), formLongRandom = list(bil = ~year | id, alb = ~year |
##     id), formSurv = Surv(years, status2) ~ age, data = placebo,
##     timeVar = "year", verbose = FALSE, control = list(burnin = 300))
##
## Data Descriptives:
##
## Event Process
##     Number of subjects: 154
##     Number of events: 69 (44.8%)
##
## Longitudinal Process
##     Number of longitudinal outcomes: K = 2
##     Number of observations:
##       Outcome 1 (bil): n = 967
##       Outcome 2 (alb): n = 967
##
## Joint Model Summary:
##
## Longitudinal Process: Multivariate linear mixed-effects model
##      log(serBilir) ~ year, random = ~year | id
##      albumin ~ year, random = ~year | id
## Event Process: Cox proportional hazards model
##      Surv(years, status2) ~ age
## Model fit statistics:
##    log.Lik      AIC      BIC
##  -1527.715 3093.429 3151.132
##
## Variance Components:
##
## Random effects variance covariance matrix
##               (Intercept)_1    year_1 (Intercept)_2     year_2
## (Intercept)_1      1.144200  0.088184    -0.1826400 -0.0508720
## year_1             0.088184  0.032242    -0.0175150 -0.0105570
## (Intercept)_2     -0.182640 -0.017515     0.1134200  0.0070106
## year_2            -0.050872 -0.010557     0.0070106  0.0056754
##   Standard Deviations: 1.0697 0.17956 0.33678 0.075335
##
## Residual standard errors(s):
##  sigma2_1  sigma2_2
## 0.3567699 0.3270099
##
## Coefficient Estimates:
```

```
##
## Longitudinal sub-model:
##                 Value Std.Err  z-value p-value
## (Intercept)_1  0.5548  0.0771   7.1978 <0.0001
## year_1         0.1909  0.0159  11.9783 <0.0001
## (Intercept)_2  3.5540  0.0330 107.6416 <0.0001
## year_2        -0.1195  0.0079 -15.1717 <0.0001
##
## Time-to-event sub-model:
##           Value Std.Err z-value p-value
## age      0.0484  0.0141  3.4332  0.0006
## gamma_1  0.9890  0.1469  6.7329 <0.0001
## gamma_2 -2.1935  0.3557 -6.1670 <0.0001
##
## Algorithm Summary:
##     Total computational time: 1.7 mins
##     EM algorithm computational time: 1.5 mins
##     Convergence status: converged
##     Convergence criterion: sas
##     Final Monte Carlo sample size: 3122
##     Standard errors calculated using method: approx
```

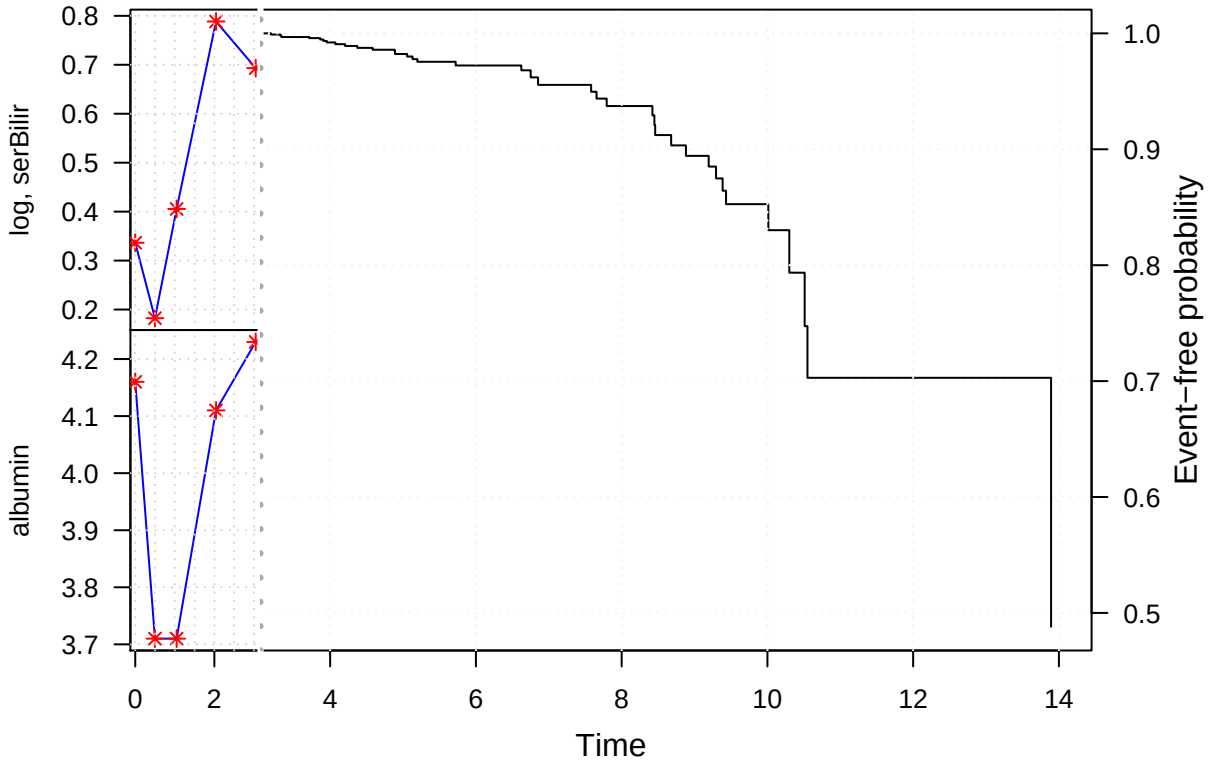
Part (b)

New patient

```
nd <- subset(pbc2, id == "11")
```

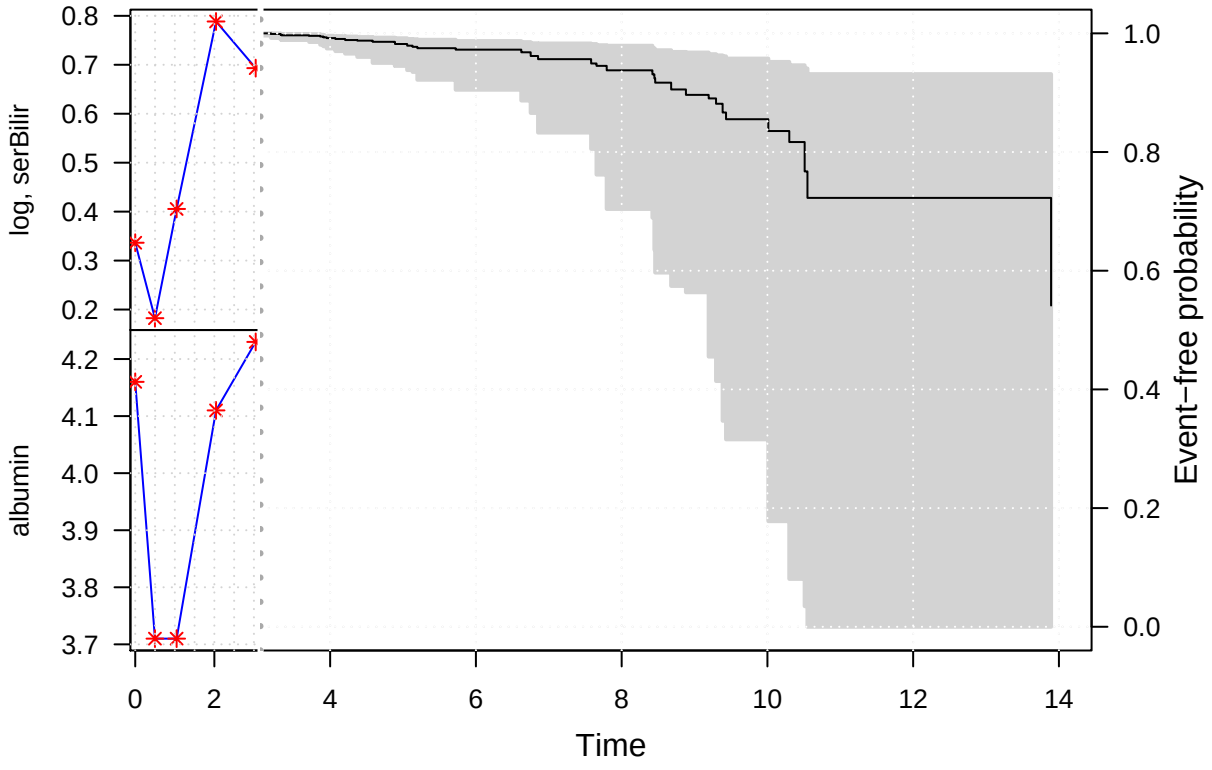
First-order prediction (default):

```
pred1 <- dynSurv(fit6, nd[1:5, ])
plot(pred1)
```



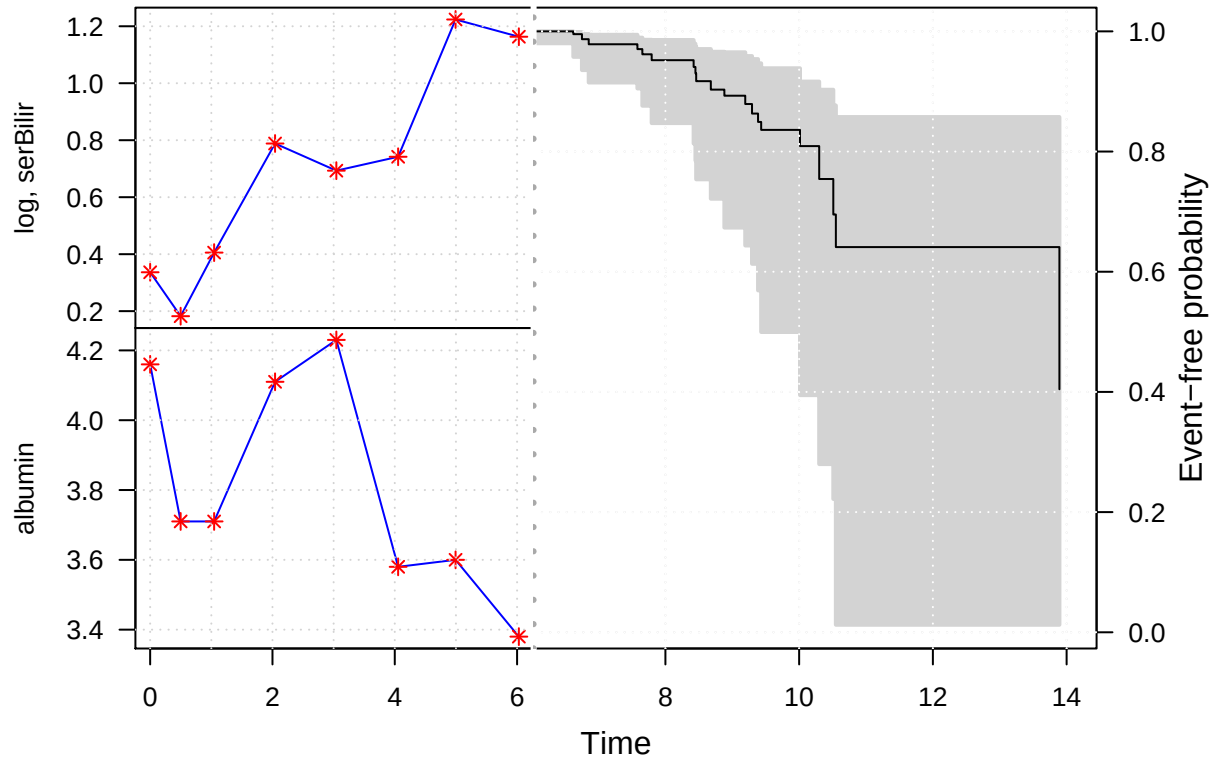
Simulated prediction:

```
pred2 <- dynSurv(fit6, nd[1:5, ], type = "simulated", scale = 2, progress = FALSE)
plot(pred2)
```



Part (c)

```
pred2 <- dynSurv(fit6, nd[1:8, ], type = "simulated", scale = 2, progress = FALSE)
plot(pred2)
```



Question 4

Part (a)

```
set.seed(1)
sim <- simData(150, model = "int")
```

```
## 29.3% experienced event
```

```
fit7 <- mjoint(
    formLongFixed = list("Y1" = Y.1 ~ time + ctsxl + binxl,
                         "Y2" = Y.2 ~ time + ctsxl + binxl),
    formLongRandom = list("Y1" = ~ 1 | id,
                          "Y2" = ~ 1 | id),
    formSurv = Surv(survtime, cens) ~ ctsx + binx,
    data = sim$longdat,
    survData = sim$survdat,
    timeVar = "time",
    verbose = FALSE)
```

```
## Running multivariate LMM EM algorithm to establish initial parameters...
```

```
## Finished multivariate LMM EM algorithm...

## EM algorithm has converged!

## Calculating post model fit statistics...
```

Part (b)

```
summary(fit7)
```

```
##
## Call:
## mjoint(formLongFixed = list(Y1 = Y.1 ~ time + ctsxl + binxl,
##     Y2 = Y.2 ~ time + ctsxl + binxl), formLongRandom = list(Y1 = ~1 |
##     id, Y2 = ~1 | id), formSurv = Surv(survtime, cens) ~ ctsx +
##     binx, data = sim$longdat, survData = sim$survdat, timeVar = "time",
##     verbose = FALSE)
##
## Data Descriptives:
##
## Event Process
##     Number of subjects: 150
##     Number of events: 44 (29.3%)
##
## Longitudinal Process
##     Number of longitudinal outcomes: K = 2
##     Number of observations:
##       Outcome 1 (Y1): n = 547
##       Outcome 2 (Y2): n = 547
##
## Joint Model Summary:
##
## Longitudinal Process: Multivariate linear mixed-effects model
##      Y.1 ~ time + ctsxl + binxl, random = ~1 | id
##      Y.2 ~ time + ctsxl + binxl, random = ~1 | id
## Event Process: Cox proportional hazards model
##      Surv(survtime, cens) ~ ctsx + binx
## Model fit statistics:
##    log.Lik      AIC      BIC
##  -2069.522 4173.044 4224.225
##
## Variance Components:
##
## Random effects variance covariance matrix
##               (Intercept)_1 (Intercept)_2
## (Intercept)_1       1.20290       0.28193
## (Intercept)_2       0.28193       0.87753
##   Standard Deviations: 1.0968 0.93677
##
## Residual standard errors(s):
## sigma2_1 sigma2_2
```

```
## 1.077243 1.058093
##
## Coefficient Estimates:
##
## Longitudinal sub-model:
##                Value Std.Err z-value p-value
## (Intercept)_1 1.0178  0.1618  6.2890 <0.0001
## time_1        1.0121  0.0387 26.1783 <0.0001
## ctsxl_1       0.9029  0.1389  6.4984 <0.0001
## binxl_1       0.8416  0.2176  3.8672  0.0001
## (Intercept)_2 0.8667  0.1329  6.5194 <0.0001
## time_2        1.0641  0.0380 27.9743 <0.0001
## ctsxl_2       1.1470  0.1223  9.3751 <0.0001
## binxl_2       1.1229  0.1990  5.6426 <0.0001
##
## Time-to-event sub-model:
##           Value Std.Err z-value p-value
## ctsx     0.4101  0.1739  2.3578  0.0184
## binx     0.7585  0.3279  2.3130  0.0207
## gamma_1  0.2701  0.1961  1.3778  0.1683
## gamma_2 -0.7327  0.2265 -3.2354  0.0012
##
## Algorithm Summary:
##     Total computational time: 52.2 secs
##     EM algorithm computational time: 50.9 secs
##     Convergence status: converged
##     Convergence criterion: sas
##     Final Monte Carlo sample size: 3122
##     Standard errors calculated using method: approx
```

Part (c) [advanced + computationally slow]

```r
set.seed(12345)
coef.gamma <- matrix(NA, nrow = 10, ncol = 4)

for (i in 1:10) {

  # Print iteration
  print(paste("This is iteration", i))

  # Simulate data
  sim <- simData(150, model = "int")

  # Fit a model
  suppressMessages(
    fit.loop <- mjoint(
      formLongFixed = list("Y1" = Y.1 ~ time + ctsxl + binxl,
                           "Y2" = Y.2 ~ time + ctsxl + binxl),
      formLongRandom = list("Y1" = ~ 1 | id,
                            "Y2" = ~ 1 | id),
      formSurv = Surv(survtime, cens) ~ ctsx + binx,
      data = sim$longdat,
```

```
      survData = sim$survdat,
      timeVar = "time",
      verbose = FALSE)
  )

  coef.gamma[i, ] <- fixef(fit.loop, process = "Event")

}
```

```
## [1] "This is iteration 1"
## 35.3% experienced event
## [1] "This is iteration 2"
## 38% experienced event
## [1] "This is iteration 3"
## 33.3% experienced event
## [1] "This is iteration 4"
## 37.3% experienced event
## [1] "This is iteration 5"
## 29.3% experienced event
## [1] "This is iteration 6"
## 27.3% experienced event
## [1] "This is iteration 7"
## 36% experienced event
## [1] "This is iteration 8"
## 38.7% experienced event
## [1] "This is iteration 9"
## 34.7% experienced event
## [1] "This is iteration 10"
## 39.3% experienced event
```

```
coef.gamma
```

```
##             [,1]      [,2]      [,3]       [,4]
##  [1,] 1.1862781 1.0629874 0.7242445 -0.8719221
##  [2,] 1.2762230 1.8690539 0.7452022 -1.2600278
##  [3,] 1.2166291 1.7435443 0.6528479 -1.3746420
##  [4,] 0.6007634 0.4790311 0.5807761 -0.9501765
##  [5,] 1.0644930 1.0277224 0.5777170 -1.1733504
##  [6,] 1.0990309 0.8798908 0.1578589 -0.8562531
##  [7,] 1.4065074 1.1916455 0.4428832 -1.1603768
##  [8,] 0.9670713 0.7593213 0.4753206 -0.8230737
##  [9,] 1.1157401 1.2082889 0.4478894 -0.9434301
## [10,] 0.8775672 0.5627971 0.8869815 -1.1891641
```

```
apply(coef.gamma, 2, mean)
```

```
## [1]  1.0810304  1.0784283  0.5691721 -1.0602417
```