HawkesModel

Fitting Hawkes model to marathon record data

Produce estimates for model, exponential Hawkes process, using MLE: - baseline intensity, - reproduction mean
- exponential fertility function rate

Note: data is in days from the first world record which is set as time 0. The model is fit through day 40,300 (the last record is at day 40,231).

Three parameter estimates are first output.

NOTE: the results change if one doesn't set a seed. . . and by more than I would expect!!!! Maybe need to set starting values or something to ensure optimal?

```
record_table_mod<-read_rds("record_table_mod.rds")

days_between = as.numeric(diff(record_table_mod$Date_ymd))
daysfromstart <- cumsum(days_between)
daysfromstart <- c(0,daysfromstart)  ### get data in terms of days from first record (time 0)

set.seed(11)
optMarathon<-mle(daysfromstart,"Exponential",40300)  # end date picked number greater than longest time
optMarathon$par
```

```
## [1] 0.001157352 0.067176042 0.093735705
```

```
summary(optMarathon)
```

```
##        Length Class            Mode
## par     3     -none-           numeric
## model   1     Rcpp_Exponential S4
## events 50     -none-           numeric
## end     1     -none-           numeric
## opt    20     nloptr           list
```

```
optMarathon$events
```

```
##  [1]     0   161   203   288   290   403  1753  1772  2319  4412  6289  7651
## [13]  9736  9746  9749  9963 14148 16031 16395 16508 16773 17551 18293 19041
## [25] 19931 20049 20070 20413 20543 20777 21681 22225 22644 23932 25398 26209
## [37] 26798 27848 28029 29122 32930 33329 34232 34764 36227 36591 37683 38418
## [49] 38782 40231
```

```
optMarathon$end
```

```
## [1] 40300
```

```
optMarathon$model$param
```

```
##             [,1]
## [1,] 0.001157352
## [2,] 0.067176042
## [3,] 0.093735705
```

```
optMarathon$model$mean()  # expected value
```

```
## [1] 0.001240698
```

```
optMarathon$model$dmean()  # Jacobian matrix of expected value
```

```
##             [,1]
## [1,] 1.072013633
## [2,] 0.001330045
## [3,] 0.000000000
```

```
optMarathon$model$ddmean()  # Hessian matrix of expected value
```

```
##          [,1]        [,2] [,3]
## [1,] 0.000000 1.149213229    0
## [2,] 1.149213 0.002851652    0
## [3,] 0.000000 0.000000000    0
```

```
optMarathon$model$loglik(daysfromstart, optMarathon$end)  #  log-likelihood
```

```
## [1] -382.0092
```

```
optMarathon$model$dloglik(daysfromstart, optMarathon$end)  # Jacobian matrix of log-lik
```

```
##              [,1]
## [1,] 5.551819e-04
## [2,] 2.479034e-06
## [3,] 1.064309e-07
```

```
optMarathon$model$ddloglik(daysfromstart, optMarathon$end) # hessian matrix of log-lik
```

```
##               [,1]        [,2]      [,3]
## [1,] -33271354.494 -13848.3565 4739.6674
## [2,]    -13848.357   -505.7014  -81.7652
## [3,]      4739.667    -81.7652 -203.0887
```

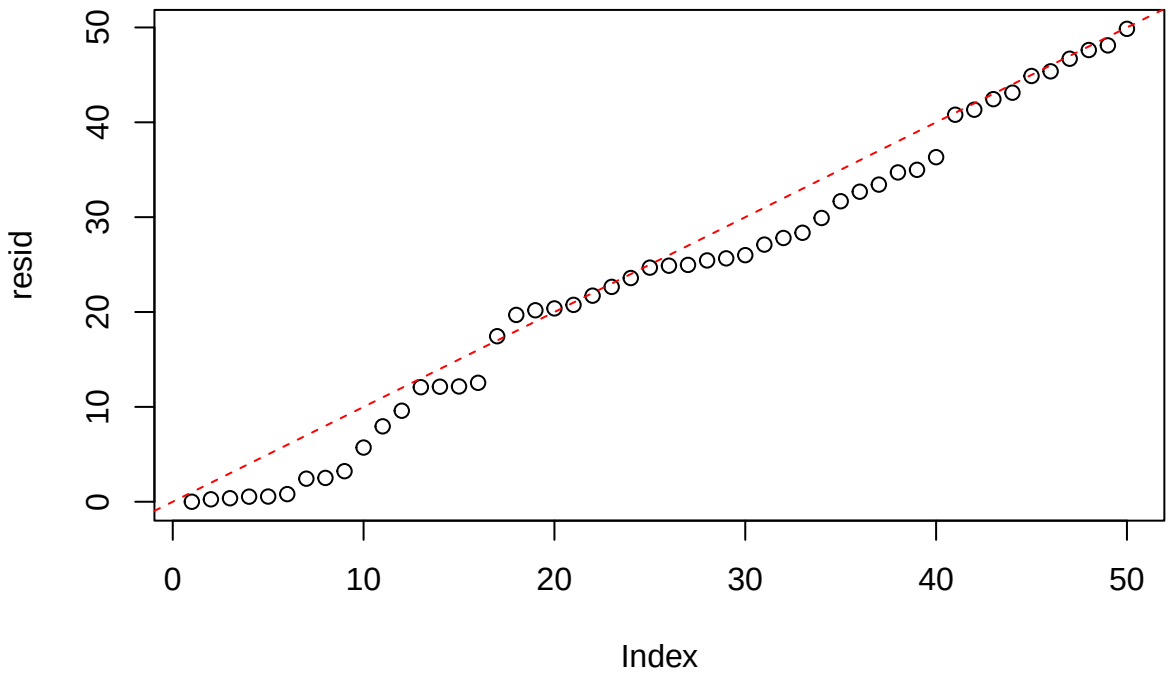
Residuals

From the help: "Outputs the residuals (values of the compensator at the times of arrival) of a Hawkes process. Useful function for diagnosis through the random time change theorem: the residuals should follow a unit rate Poisson process" Based on the example in the help I assume should follow the y=x line... we see divergence here suggesting an issue (in what direction?)

```
resid<-residuals(daysfromstart, fun = optMarathon$par[1], repr = optMarathon$par[2],
                 family = "exp", rate = optMarathon$par[3])
resid
```

```
##  [1]  0.0000000  0.2535098  0.3679841  0.5348219  0.5486242  0.8022901
##  [7]  2.4318950  2.5097434  3.2213085  5.7108231  7.9503496  9.5938396
## [13] 12.0740954 12.1265347 12.1529229 12.5383425 17.4490383 19.6955089
## [19] 20.1839612 20.3819164 20.7557925 21.7233887 22.6493202 23.5821959
## [25] 24.6794155 24.8831581 24.9652568 25.4387875 25.6564190 25.9944159
## [31] 27.1078385 27.8046142 28.3567209 29.9145668 31.6784215 32.6842103
## [37] 33.4330669 34.7154630 34.9921198 36.3242820 40.7986559 41.3276156
## [43] 42.4398808 43.1227684 44.8831510 45.3716033 46.7026081 47.6204382
## [49] 48.1088905 49.8530701
```

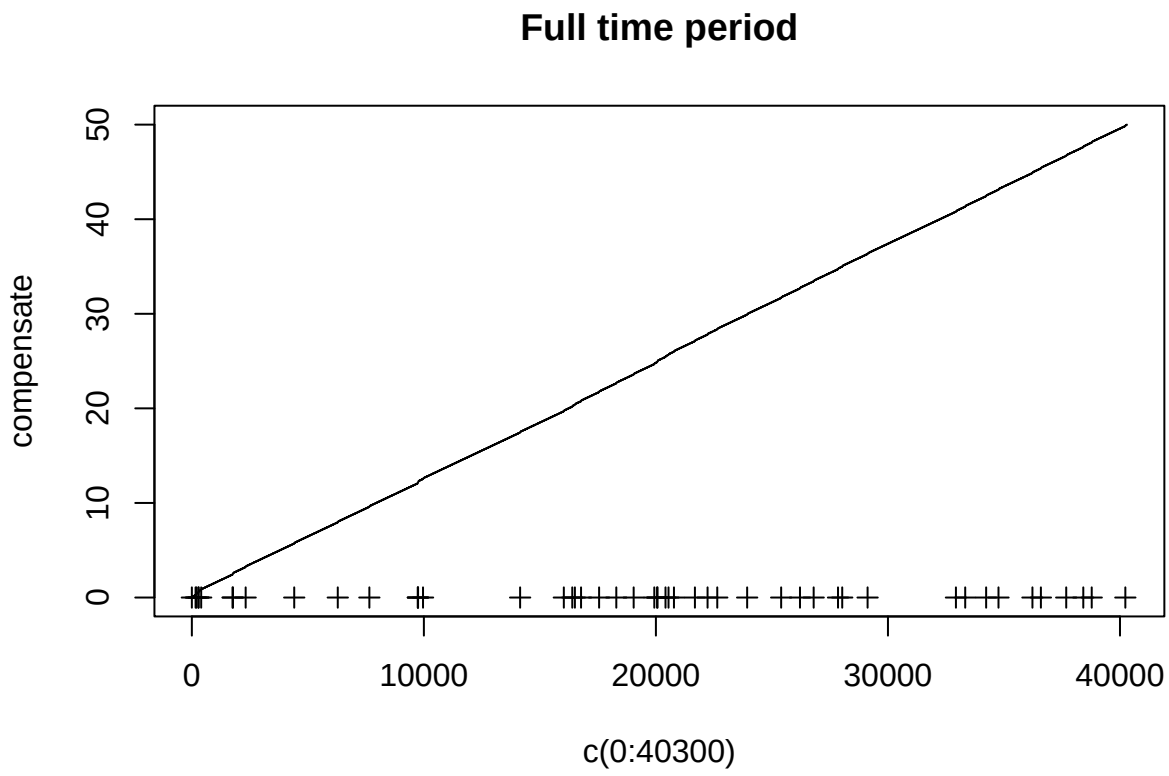
```
plot(resid)
abline(0, 1, col="red", lty="dashed")
```



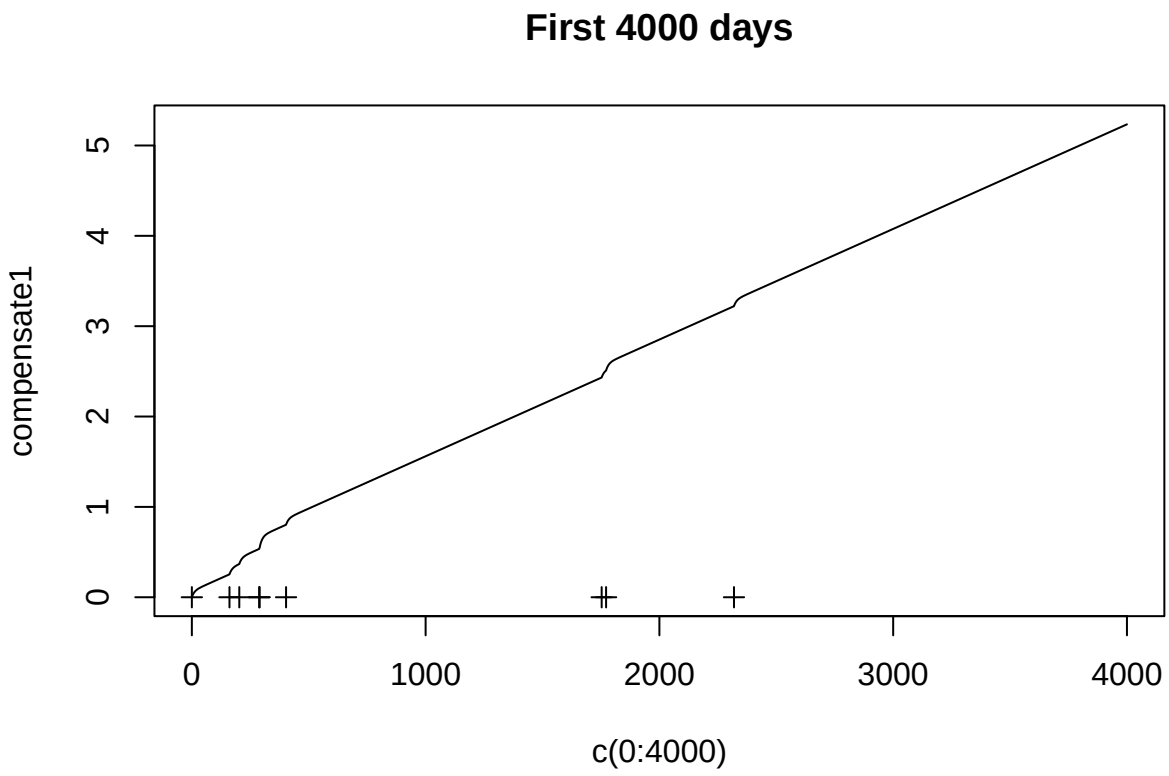
Compensator

From the help: the compensator (integrated intensity) of a Hawkes process. . . kind of cool, can see how the events (world records) up the intensity. Event times are the plus symbols plotted below the intensity function line.

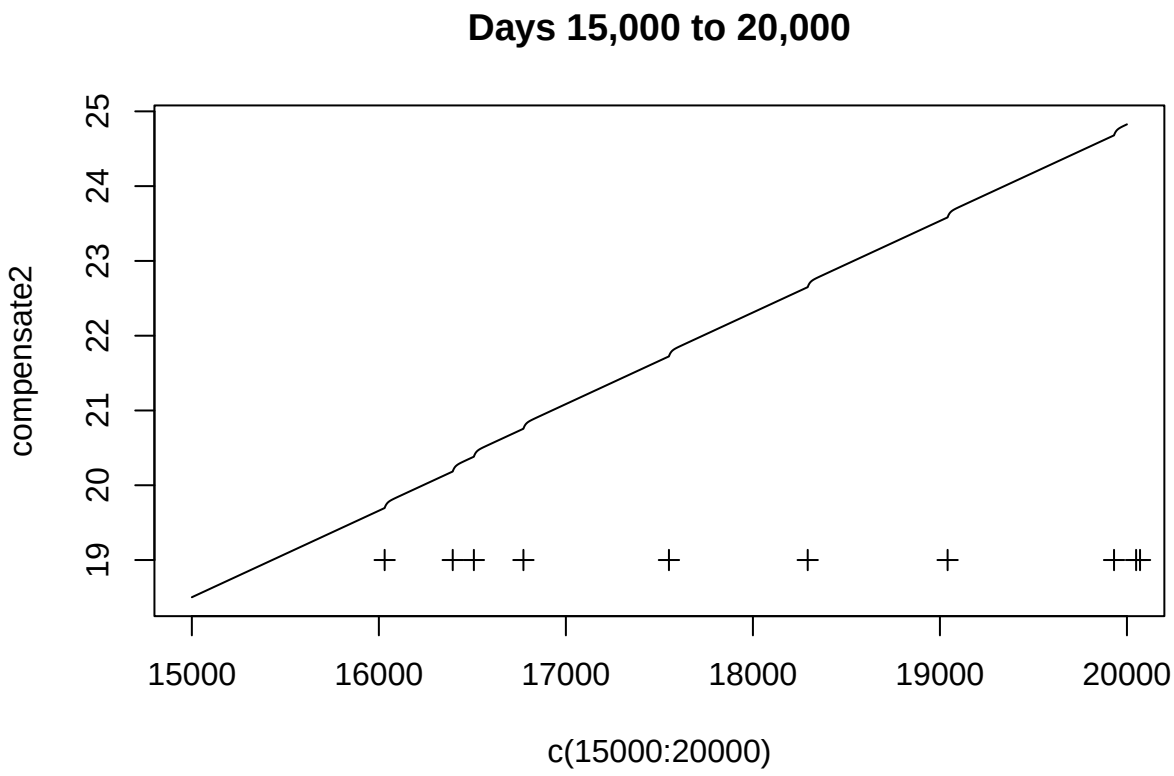
```
compensate<-compensator(daysfromstart, t=0:40300, fun = optMarathon$par[1], repr = optMarathon$par[2],
                  family = "exp", rate = optMarathon$par[3])
plot(c(0:40300),compensate,main="Full time period", type = "l")
points(daysfromstart,rep(0,length(daysfromstart)), pch=3)
```



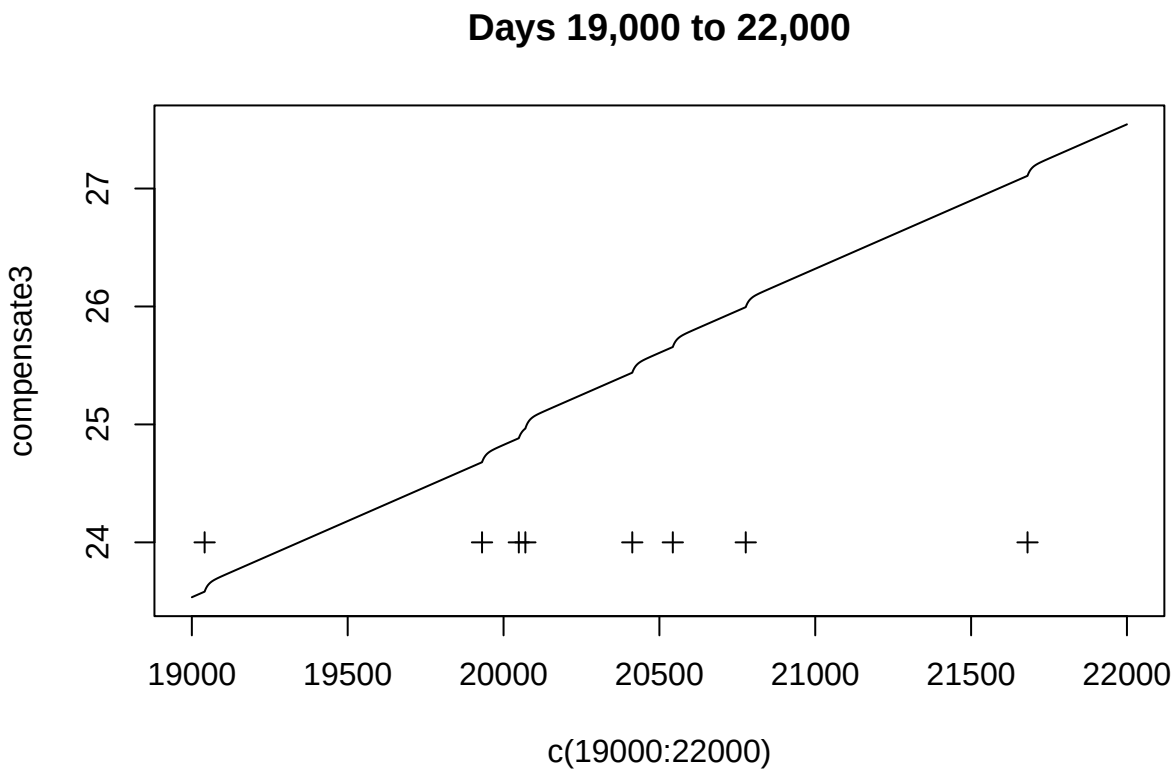

```
compensate1<-compensator(daysfromstart, t=0:4000, fun = optMarathon$par[1], repr = optMarathon$par[2],
                  family = "exp", rate = optMarathon$par[3])
plot(c(0:4000),compensate1,main="First 4000 days", type = "l")
points(daysfromstart,rep(0,length(daysfromstart)), pch=3)
```




```
compensate2<-compensator(daysfromstart, t=15000:20000, fun = optMarathon$par[1], repr = optMarathon$par
                  family = "exp", rate = optMarathon$par[3])
plot(c(15000:20000),compensate2,main="Days 15,000 to 20,000", type = "l")
points(daysfromstart,rep(19,length(daysfromstart)), pch=3)
```




```
compensate3<-compensator(daysfromstart, t=19000:22000, fun = optMarathon$par[1], repr = optMarathon$par
                  family = "exp", rate = optMarathon$par[3])
plot(c(19000:22000),compensate3,main="Days 19,000 to 22,000", type = "l")
points(daysfromstart,rep(24,length(daysfromstart)), pch=3)
```

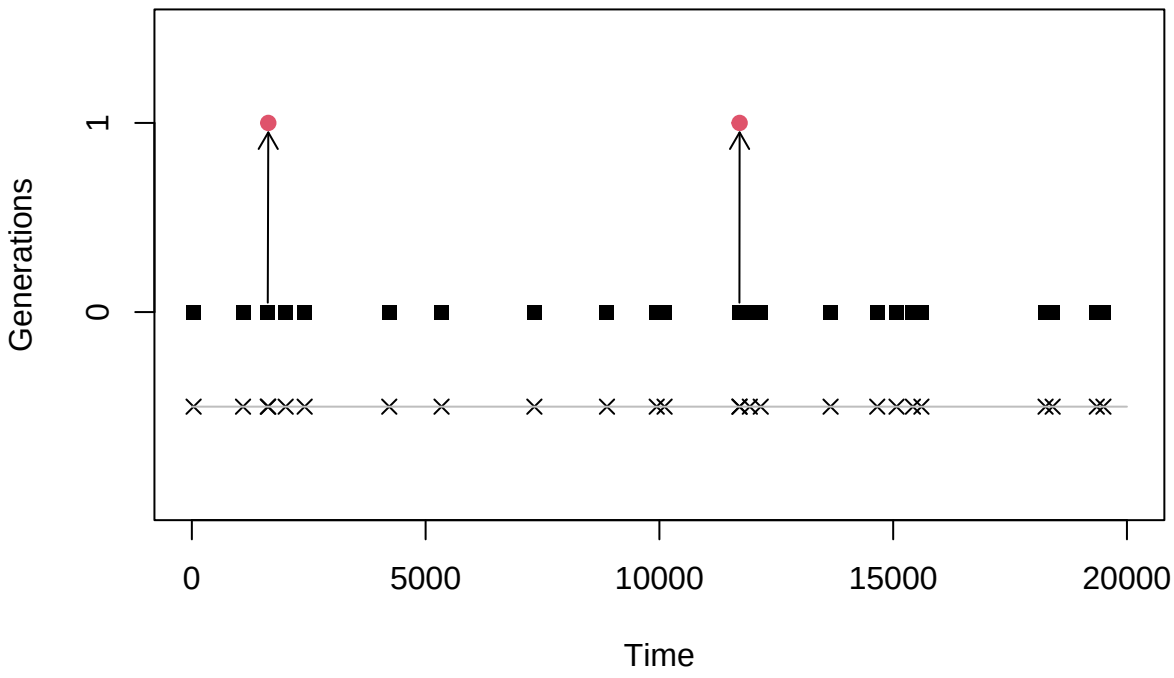


Simulation using Hawkes model with MLE parameter estimates from data

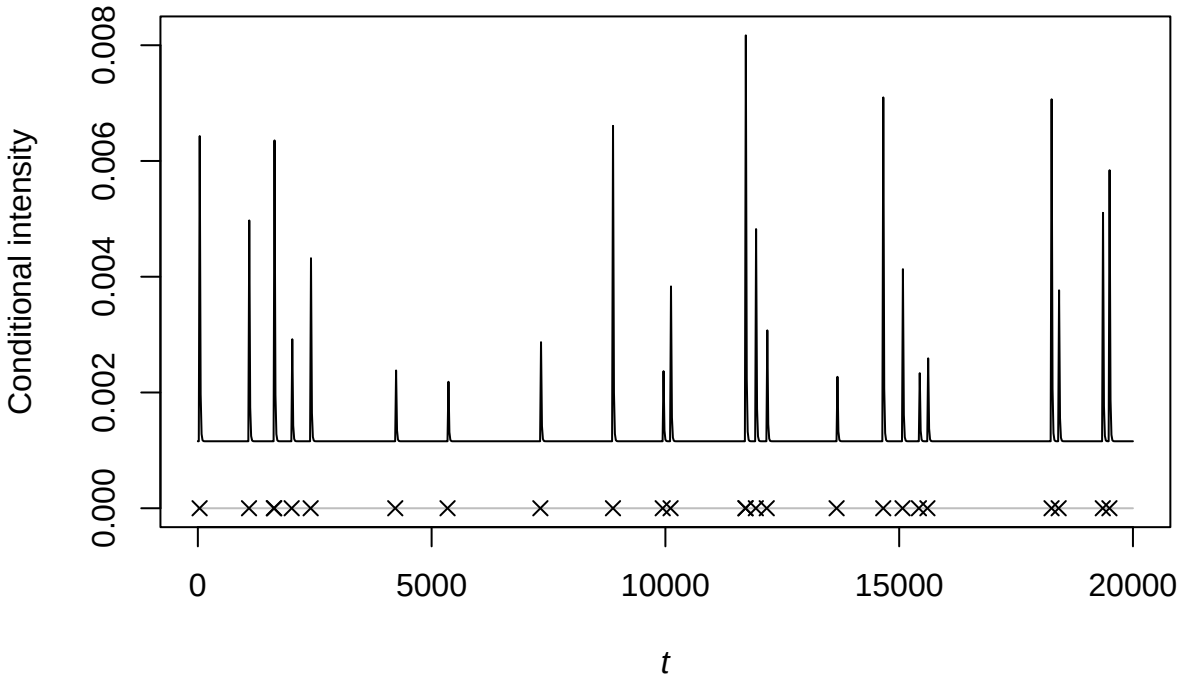
Simulation for the same number of days used in the estimation. Last plot is the residuals - totally unclear to me what these are. . . I guess of the simulated data vs the exact values from the model? Generally seem to follow the y=x line but some runs don't.

The second plot is the intensity - not easy to see, I think maybe a scale issue (not sure why it appears 0 with a few spikes, and to me those spikes don't correspond to plot one events).

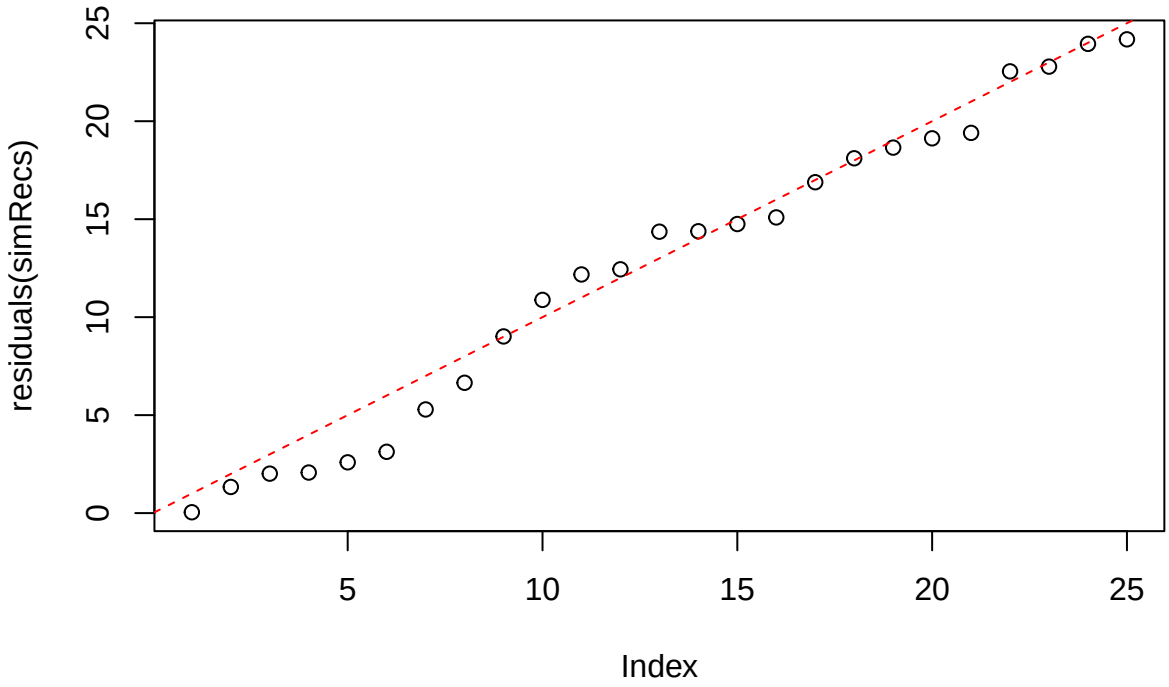
```
set.seed(96)
simRecs <- hawkes(20000, fun = optMarathon$par[1], repr = optMarathon$par[2],
                  family = "exp", rate = optMarathon$par[3])
plot(simRecs, intensity = FALSE)
```




```
plot(simRecs, intensity = TRUE)
```




```
plot(residuals(simRecs))
abline(0, 1, col="red", lty="dashed")
```



test