Using poisson.r to model clinical trial recruitment

Kristian Brock

25 January 2015

This notebook demonstrates the functionality of `poisson.r`, a collection of R functions for simulating homogeneous and non-homogenenous Poisson processes. We will first illustrate the main functionality using an imaginary clinical trial that aims to recruit 50 patients over 5 years. Then we will look at a real-life example in the *National Lung Matrix Trial*.

```
source("poisson.r")
```

```
## Creating a generic function for 'plot' from package 'graphics' in the global environment
```

```
target = 50
t0 <- 0  # Start time
t1 <- 6  # End time
```

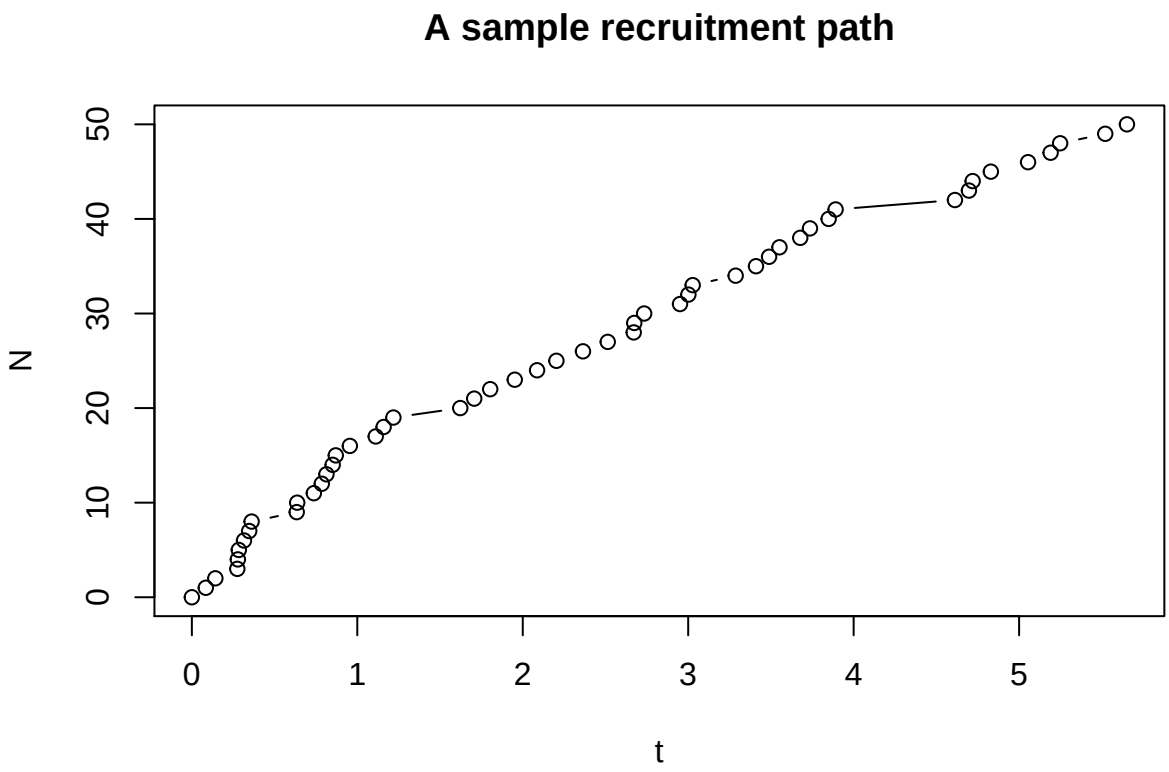
Let's set the random seed so that the results are reproducible on different computers. In real life, you should not execute this step if you want to get genuine random samples!

```
set.seed(123)
```

Homogeneous Poisson Processes

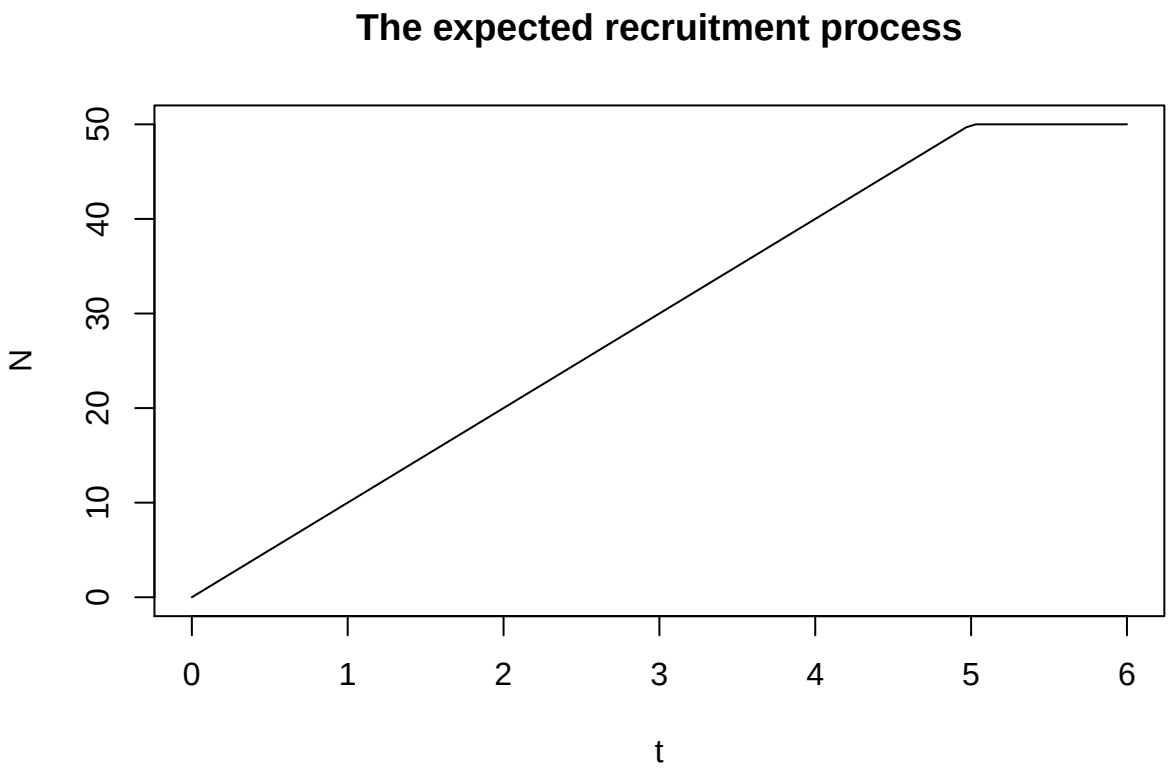
An homogeneous Poisson process is one where the average rate of events is the same at each point in time. Seeking to recruit 50 patients over 5 years in our notional trial, this suggests an annual average recruitment rate of 10 patients per year. Naturally, we expect real recruitment to exhibit random variation, but this is the average we expect. Let's simulate and plot a single recruitment path using the method `hpp.sim`, meaning *homogeneous Poisson process simulation*.

```
rate <- 10
x <- hpp.sim(rate, target)
plot(x, 0:target, main='A sample recruitment path', xlab='t', ylab='N', type='b')
```



We can also calculate the expected value process using `hpp.mean`. We seek to cease recruitment once we have recruited the target number of patients. We tell the function this through the `maximum` parameter.

```
n.steps <- 100
x.bar <- hpp.mean(rate, t0, t1, maximum=target)
plot(seq(t0, t1, length.out=n.steps), x.bar, main='The expected recruitment process',
     xlab='t', ylab='N', type='l')
```

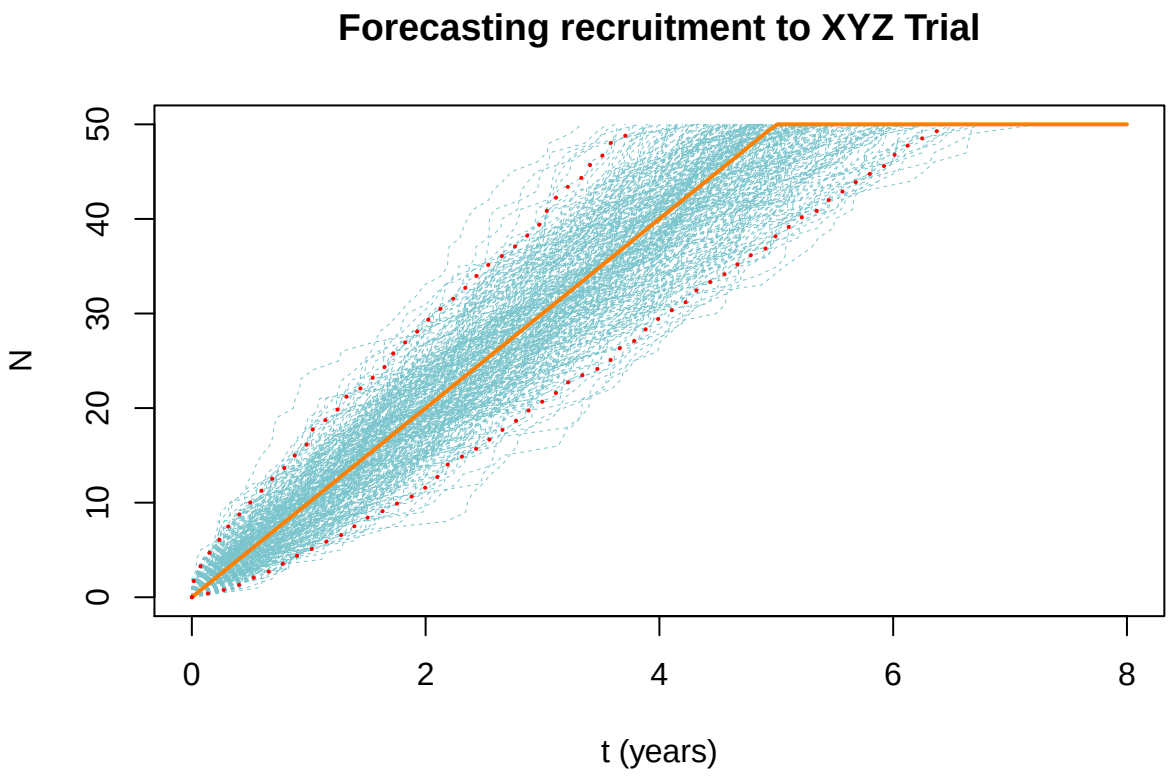


When analysing the variability of recruitment, it will be more informative to analyses *lots* of simulated paths. For this, we use `hpp.sim` again, this time providing the `num.sims` parameter.

```
x <- hpp.sim(rate, target, num.sims=50)
matplot(x, 0:target, main='Lots of simulated recruitment paths', xlab='t', ylab='N',
        type='l', lty=1, col=3)
```

Bringing this all together, the convenience function `hpp.plot` plots the expected recruitment path, lots of simulated paths, and quantile estimates of based on the simulations.

```
x <- hpp.plot(rate, target, num.sims=250, t1=t1+2,
              main="Forecasting recruitment to XYZ Trial")
```



The simulated processes, the mean and the quantiles are available in `x`:

```
names(x)
```

```
## [1] "x"     "x.bar" "x.q"
```

```
tail(x$x.q)
```

```
##           2.5%    97.5%
## [46,] 3.391039 5.832403
## [47,] 3.412933 5.982160
## [48,] 3.555212 6.021123
## [49,] 3.578955 6.157198
## [50,] 3.741158 6.330968
## [51,] 3.790482 6.481827
```

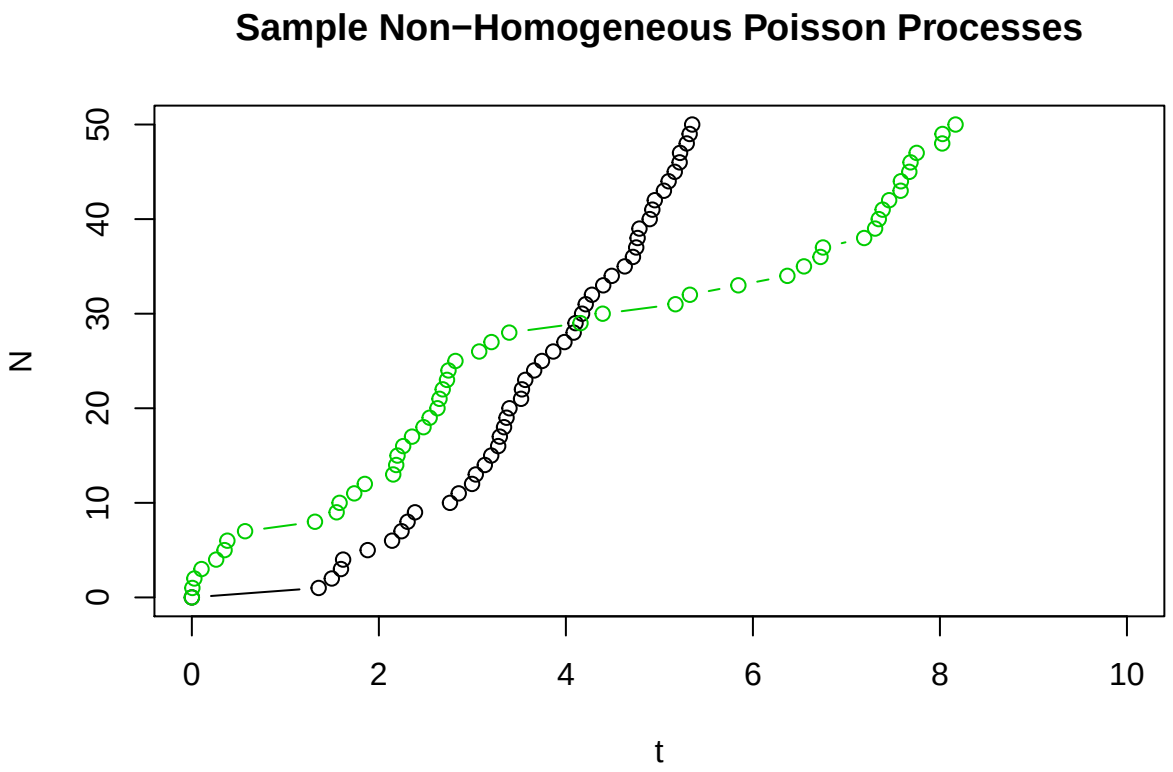
From this, we would infer that we are 95% likely to reach our target recruitment between 3.8 years and 6.5 years according to the homogeneous model, and that we are 95% likely to have recruited between 4 and 18 patients at the end of the first year. The inferences from the homogeneous model are only valid in the true recruitment scenario really is homogenenous. If it is not, the solution to allow time-varying recruitment rates using non-homogeneous processes.

Non-Homogeneous Poisson Processes

Non-homogenenous Poisson processes are so-called because they allow the rate of events to vary with time. For illustration, we will consider two ways in which recruitment varies. In the first scenario, we will consider cyclical recruitment using the `sin` function. This might be useful if recruitment is highly seasonal, e.g. to a trial of influenza treatments. In the second scenario, we will look at recruitment potential that starts at 10% when the trial opens and then increases linearly in time to 100% at the end of the fourth year. This scenario is useful when a trial recruits from many centres and the number of centres open increases with time as the site initiation process is completed.

Sampling non-homogeneous Poisson processes works by specifying the averate event rate when intensity is at 100%, and then a function that returns the event intensity between 0 and 1 at each point in time. We can think of these dampening functions as being probabilities, and will define them so:

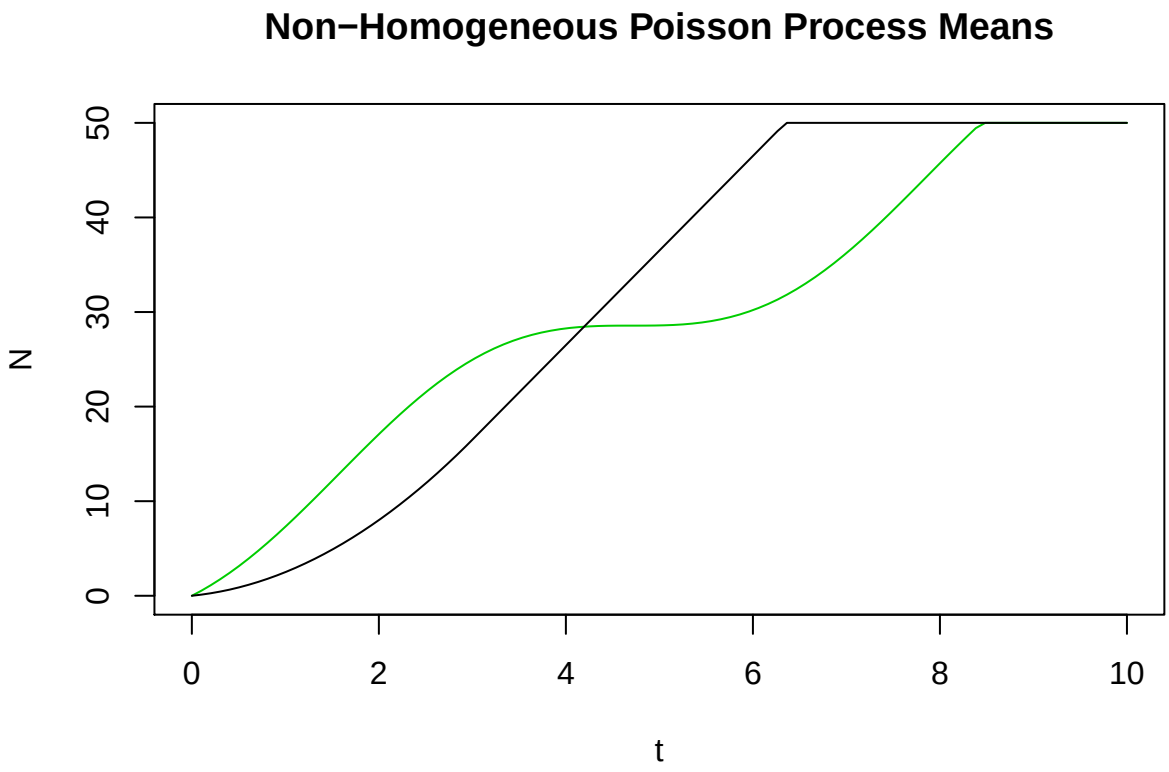
```
prob.func1 <- function(t) 0.5*(1+sin(t))
x1 <- nhpp.sim(rate, target, prob.func1)
prob.func2 <- function(t) pmin(1.0, 0.1 + 0.3*t)
x2 <- nhpp.sim(rate, target, prob.func2)
plot(x2, 0:target, main='Sample Non-Homogeneous Poisson Processes',
     xlab='t', ylab='N', type='b', xlim=c(0, 10))
points(x1, 0:target, type='b', col=3)
```



These simulated paths illustrate the cyclicality of recruitment in the first trial (green), and the acceleration of recruitment in the second trial (black).

As with the homogenenous scenario, we can calculate the expected recruitment processes

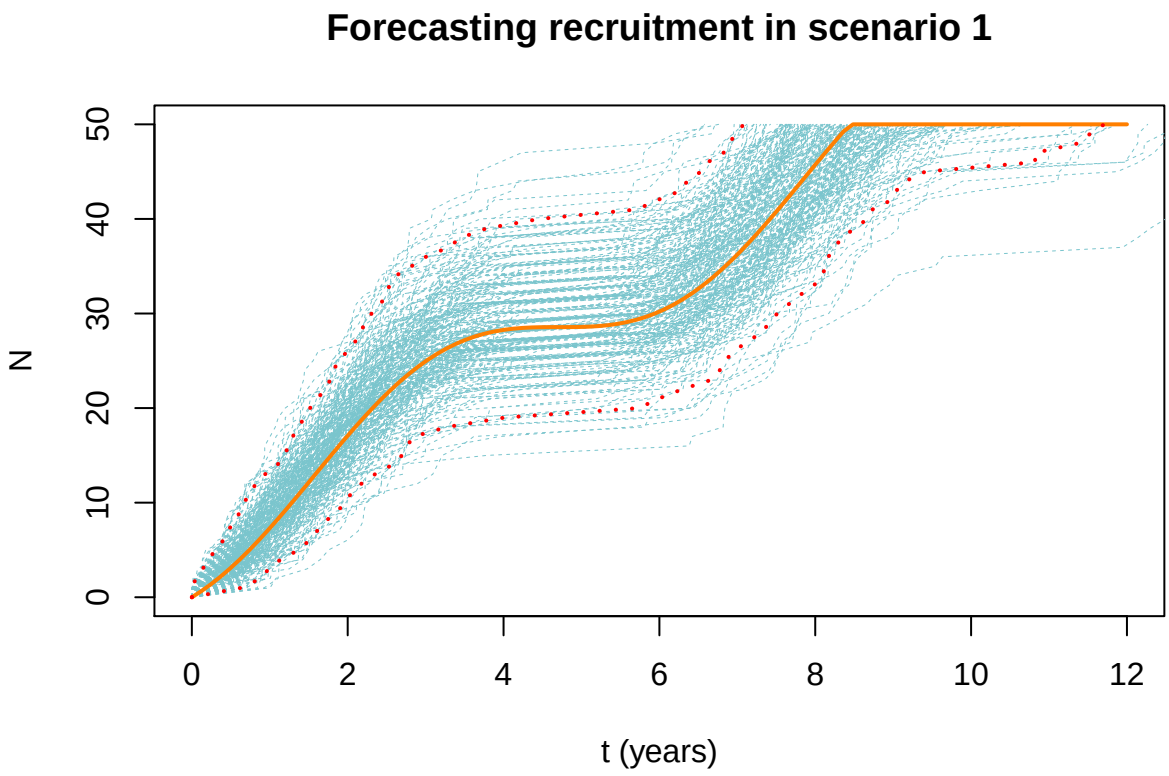
```
t1=10
x.points <- seq(t0, t1, length.out=n.steps)
x.bar1 = nhpp.mean(rate, prob.func1, t0, t1, n.steps, maximum=target)
plot(x.points, x.bar1, main='Non-Homogeneous Poisson Process Means',
     xlab='t', ylab='N', type='l', col=3)
x.bar2 = nhpp.mean(rate, prob.func2, t0, t1, n.steps, maximum=target)
lines(x.points, x.bar2, xlab='t', ylab='N', type='l')
```



This illustrates how the number of recruited patients in the second scenario (black) increases faster than linearly (quadratically, in fact).

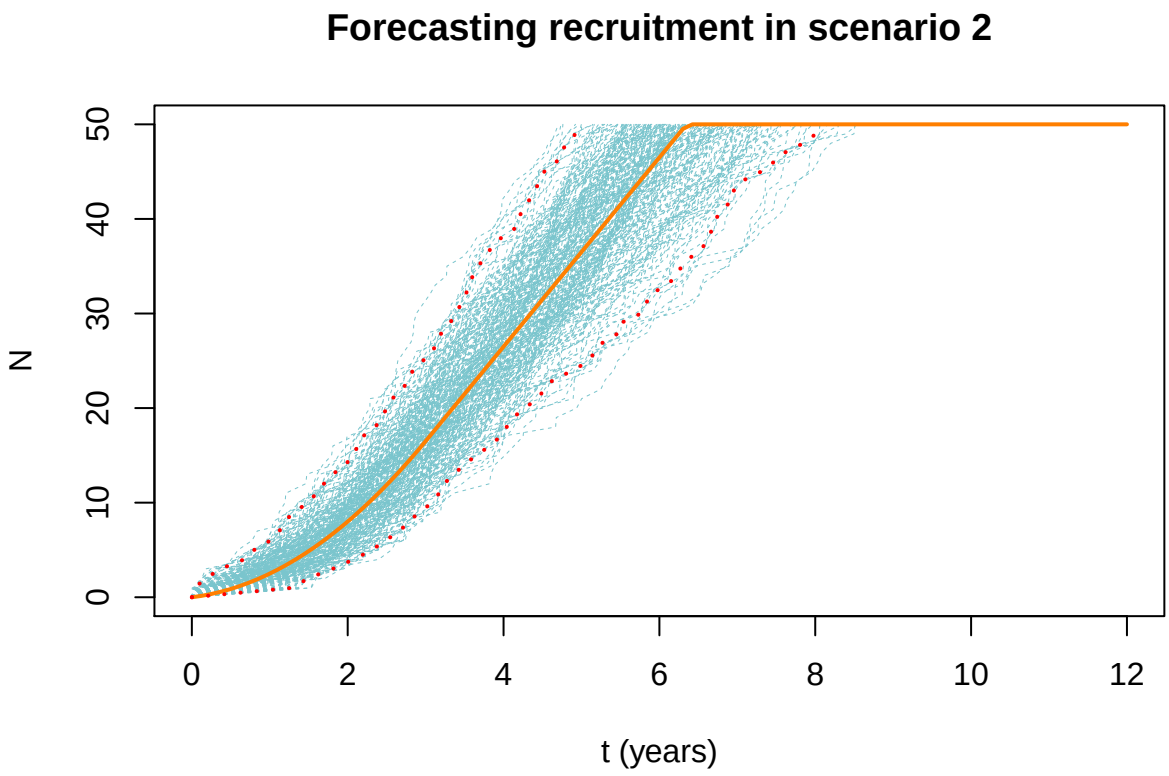
Once again, the convenience function `nhpp.plot` plots simulated paths and the mean and quantile processes. For the cyclical example, this yields:

```
x <- nhpp.plot(rate, target, prob.func1, t1=t1+2, num.sims=250,
               main="Forecasting recruitment in scenario 1")
```



and for the exponential example

```
x <- nhpp.plot(rate, target, prob.func2, t1=t1+2, num.sims=250,
               main="Forecasting recruitment in scenario 2")
```



Modelling Recruitment in National Lung Matrix Trial

Matrix expects to receive 960 patients per year that will be potentially suitable for the trial from the second stage of the NHS Stratified Medicine Programme (SMP2). Patients with *actionable targets*, the genomic changes that are the entry criteria for one of the trial cohorts in the National Lung Matrix Trial, will be registered.

For illustration, we will condider recruitment to cohort C1. This cohort seeks to recruit 30 patients and we expect 7.7% of the patients screened for registration on Matrix to be eligible for C1.

```
N <- 960
rate <- 0.077
lambda <- N * rate
```

Matrix is a multi-centre trial and the rate of recruitment will naturally depend on the number of centres open. Broadly, we expect:

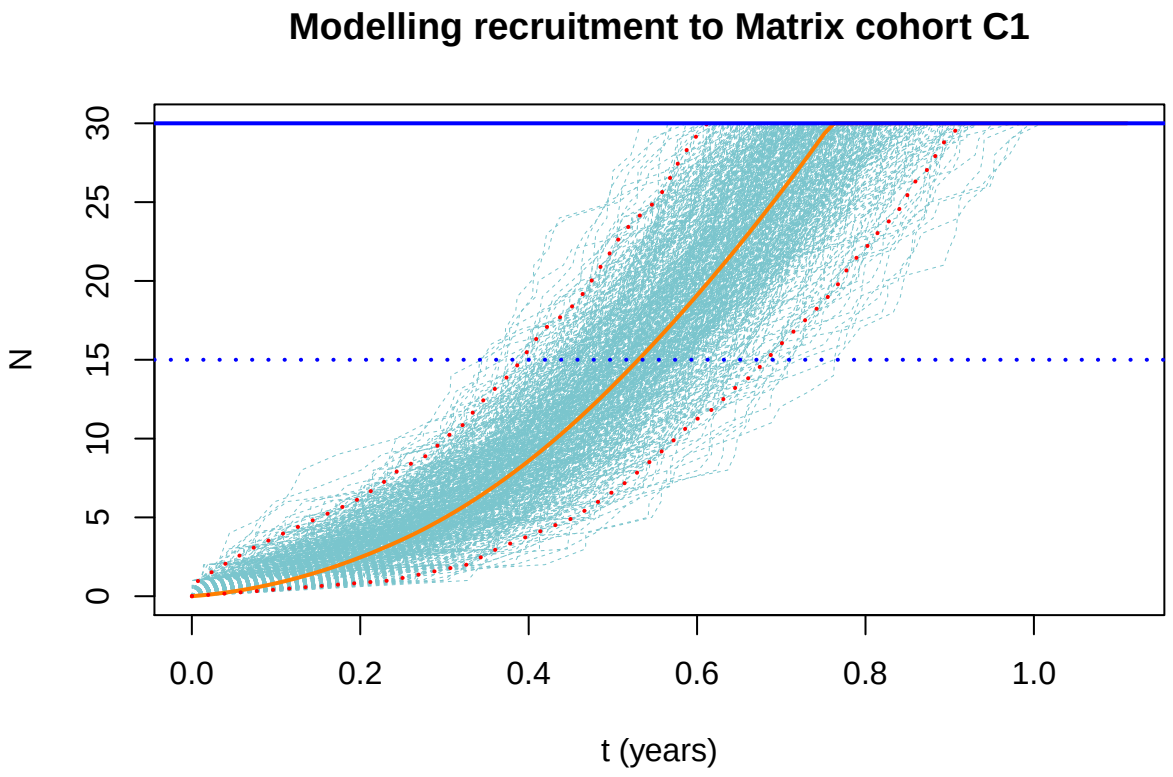
- to commence the trial with a single recruiting site open;
- to have opened 6 of 18 sites within 3 months;
- to have opened 13 of 18 sites within 6 months;
- to have opened all 18 sites within 9 months.

This is reflected in the recruitment intensity function:

```
matrix.recruitment.intensity <- function(t) {
  return(ifelse(t < 0.25,
                1/18 + 20*t/18,
                ifelse(t<0.5,
                       6/18 + 28*(t-0.25)/18,
                       ifelse(t<0.75,
                              13/18 + 20*(t-0.5)/18,
                              1)
                       )
                )
         )
}
```

Using this functon to model non-homogeneous recruitment scenario,

```
x = nhpp.plot(rate=lambda, num.events=30, prob.func=matrix.recruitment.intensity,
              num.sims=500, main='Modelling recruitment to Matrix cohort C1')
abline(h=30, col='blue', lwd=2.0, lty=1)
abline(h=15, col='blue', lwd=2.0, lty=3)
```



We plan to perform an interim analysis of this cohort as soon as we have recruited 15 patients. We also plan to perform the final analysis approximately 3 months after recruiting the 30th patient. Using our forecasts, after allowing for staggered site opening and the natural variation in patient arrivals, we expect to conduct the interim analysis after about 0.55 years and the final analysis 3 months from 0.75 years. It is not infeasible, however, that the cohort may take longer than one year to recruit.

References

For more information, please see:

- Forecasting Recruitment and Analysing Time-to-Event Outcomes in the National Lung Matrix Trial (working title) by Brock, Billingham, Crack, Popat & Middleton
- The National Lung Matrix Trial Protocol (v2.0) by Middleton, Billingham & Crack.

R Markdown

This natty format is called R Markdown and it is brilliant for illuminating code usage. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.