

infectionAgeHIV Workflow

Erik Lundgren

July 2023

Overview

infectionAgeHIV is a small package that can produce probability distributions for infection times based on HIV biomarkers

Workflow

- format patient data into two data frames: one for data where there is one value per patient and one where there may be multiple
- use `find.infection.ages` to determine which biomarker values are usable in the multiple biomarker model and find infection age distributions

Load package and example data and set pRNG seed

```
library(infectionAgeHIV)
df.single <- df.single.example
df.multiple <- df.multiple.example
set.seed(12345)
```

Show format of example data

```
#truncate sequences to be able to be displayed on page
df.multiple.substr <- df.multiple.example
df.multiple.substr$value[c(3,7,9,11)] <- paste0(substr(df.multiple.substr$value[c(3,7,9,11)]), 1, 30), "
knitr::kable(df.single.example)
```

patient_id	last_neg_test_date	first_pos_test_date	ART_start_date
1	2003-07-06	2007-08-09	2010-02-02
2	NA	2012-03-16	2012-04-16
3	NA	2017-10-20	2017-11-20
4	NA	2015-04-17	2015-05-15

```
knitr::kable(df.multiple.substr)
```

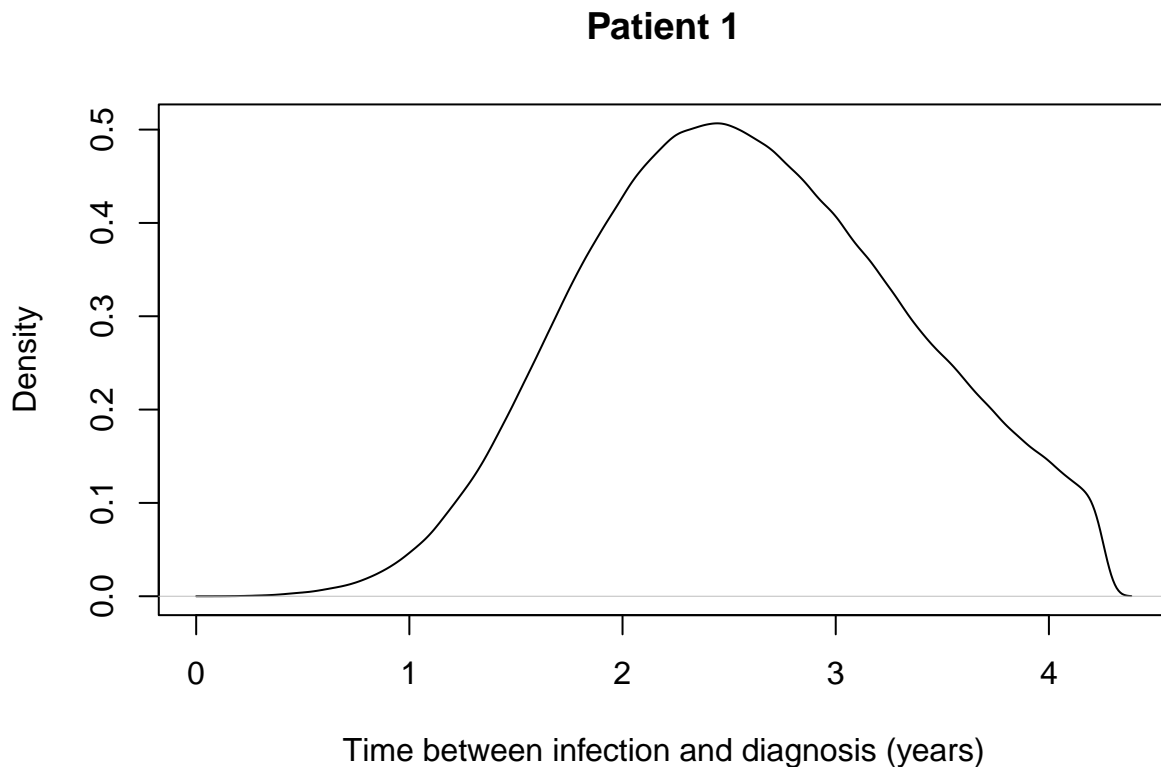
patient_id	name	date	value
1	CD4	2007-08-10	300
1	BED	2007-08-10	1.7
1	seq	2007-08-10	RAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA...
1	CD4	2009-12-20	150
1	CD4	2010-06-01	400
2	CD4	2012-03-18	600
2	seq	2012-09-18	RAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA...
3	CD4	2017-10-24	250
3	seq	2017-10-26	RAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA...
3	pol2	2017-10-27	0.01
3	seq	2019-10-20	RAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA...

Using biomarker data to predict infection times

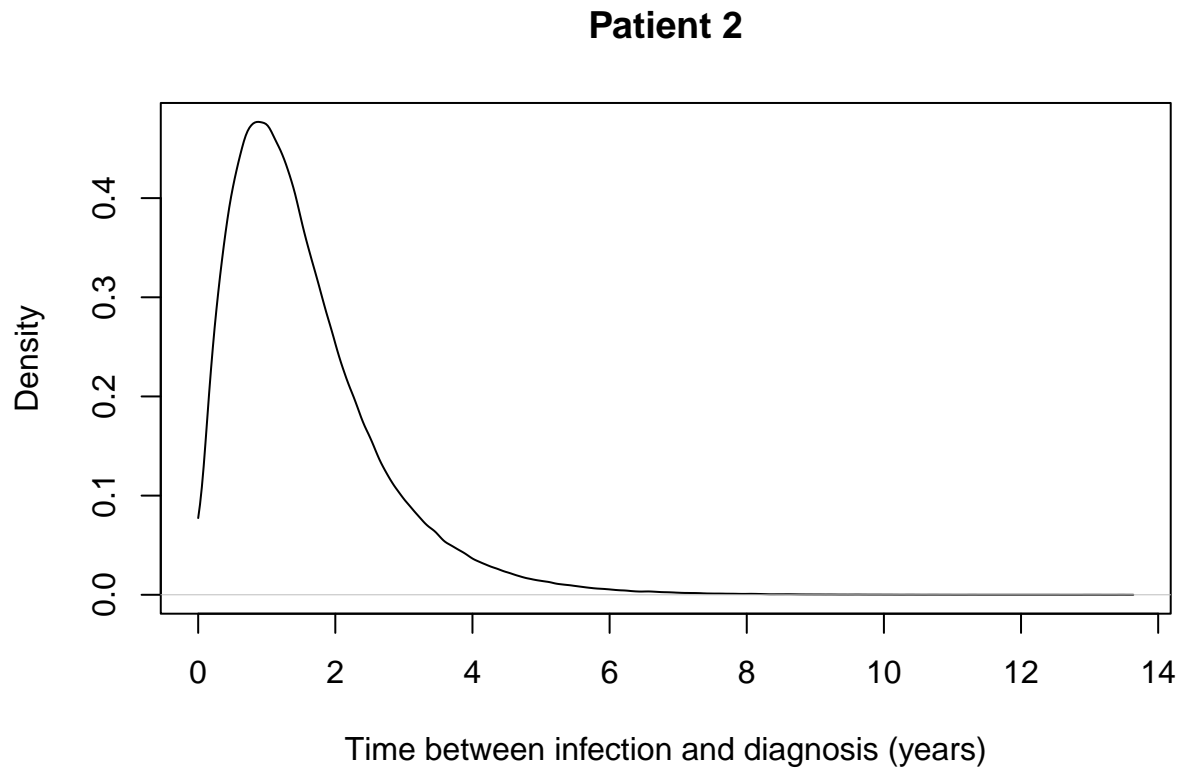
```
df <- find.infection.ages(df.single = df.single,
                        df.multiple = df.multiple,
                        n.adapt = 1e4, n.burn = 1e5, n.iter = 1e6)
```

We can plot the inferred distributions for the amount of time between infection and diagnosis.

```
plot(df$infection_age_dists_diag[[1]],
     xlab = "Time between infection and diagnosis (years)",
     main = paste0("Patient ", df.single$patient_id[[1]]))
```



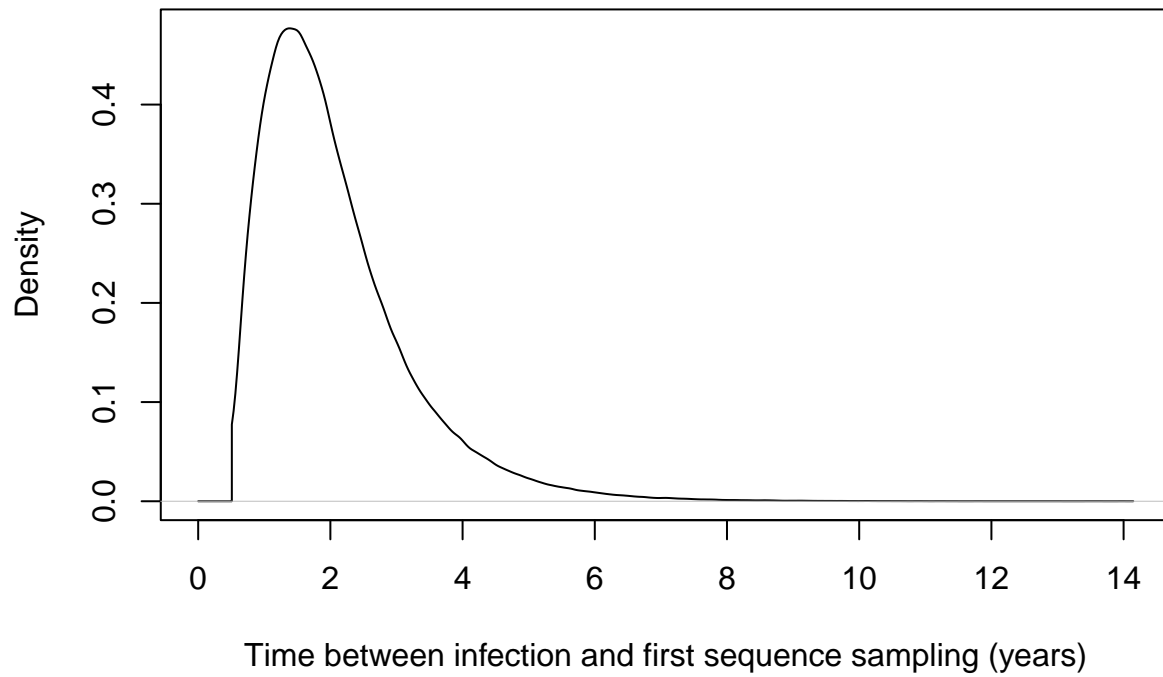
```
plot(df$infection_age_dists_diag[[2]],
     xlab = "Time between infection and diagnosis (years)",
     main = paste0("Patient ", df.single$patient_id[[2]]))
```



The distributions for the amount of time between infection and taking sequence samples and the infection time density in terms of real time are also returned.

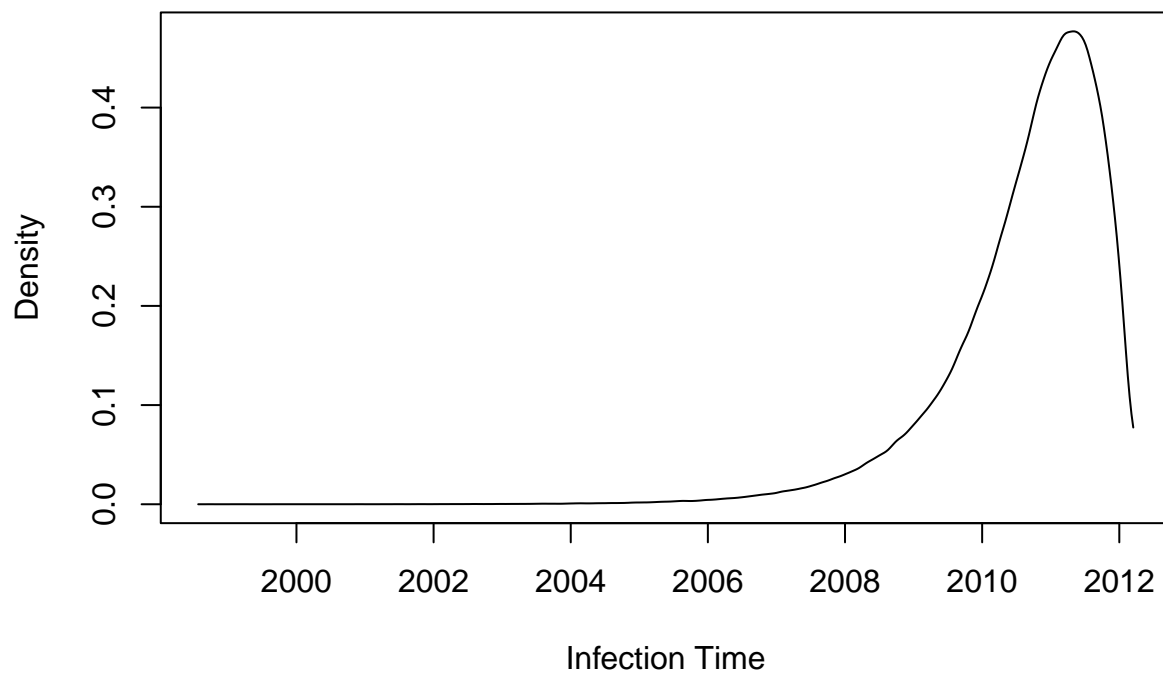
```
plot(df$infection_age_dists_seq[[2]],
     xlab = "Time between infection and first sequence sampling (years)",
     main = paste0("Patient ", df.single$patient_id[[2]]))
```

Patient 2



```
plot(df$pdf[[2]], type = "l",  
     xlab = "Infection Time",  
     ylab = "Density",  
     main = paste0("Patient ", df$single$patient_id[[2]]))
```

Patient 2



Drawing random infection ages

We can also draw from the inferred distributions by supplying a uniform $[0,1]$ random variable to the inverse CDF transform.

```
drawn.infection.ages <- df$icdf[[2]](runif(10))  
print(drawn.infection.ages)
```

```
## [1] 3.6882282 2.6216689 2.1774448 0.8939086 1.3958693 0.6991674 4.2342666  
## [8] 0.6798676 1.3378819 0.6588800
```

Using `prepare.HIV.data`

If more control is needed, `prepare.HIV.data` can be called without the wrapper `find.infection.ages`