Comparison of the precision of predicted dispensation length when using single or multiple index dates and saturated sampling

SG

2025-07-31

```
library(wtdr)
```

```
##
## Caricamento pacchetto: 'wtdr'

## I seguenti oggetti sono mascherati da 'package:stats':
##
##     dexp, dlnorm
```

```
library(bbmle)
```

```
## Caricamento del pacchetto richiesto: stats4
```

```
library(haven)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.2

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x dplyr::slice()  masks bbmle::slice()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(data.table)
```

```
##
## Caricamento pacchetto: 'data.table'
##
## I seguenti oggetti sono mascherati da 'package:lubridate':
##
```

```
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year
##
## I seguenti oggetti sono mascherati da 'package:dplyr':
##
##     between, first, last
##
## Il seguente oggetto è mascherato da 'package:purrr':
##
##     transpose
```

```
library(rlang)
```

```
##
## Caricamento pacchetto: 'rlang'
##
## Il seguente oggetto è mascherato da 'package:data.table':
##
##     :=
##
## I seguenti oggetti sono mascherati da 'package:purrr':
##
##     %@%, flatten, flatten_chr, flatten_dbl, flatten_int, flatten_lgl,
##     flatten_raw, invoke, splice
```

```
library(boot)
library(Deriv)
library(numDeriv)
```

```
set.seed(345)

df <- read.csv(file.path("..", "extdata", "df_sat_ex.csv"))

df <- as.data.table(df)

df <- df[, X:=NULL]
df <- df[, rxdate:=as.Date(rxdate)]


# id_sel <- sample(1:15000, 50) # filter df to 50 individuals
# df_sel <- df[df$pid %in% id_sel ,]
```

```
# 1) single index date - forward
forw_1 <- wtdttt(data = df,
                 rxdate ~ dlnorm(logitp, mu, lnsigma),
                 id = "pid",
                 start = as.Date('2014-01-01'),
                 end = as.Date('2014-12-31'),
                 reverse = F
)
```

```
## Warning in wtdttt(data = df, rxdate ~ dlnorm(logitp, mu, lnsigma), id = "pid",
```

```
## : Some dates are out of the window defined by start and end. Keeping only rows
## within the window.
```

```
summary(forw_1)
```

```
## Maximum likelihood estimation
##
## Call:
## mle2(minuslogl = form, start = init, fixed = list(delta = delta),
##     data = cpy, parameters = parameters_r)
##
## Coefficients:
##         Estimate Std. Error z value     Pr(z)
## logitp   1.22628    0.52837  2.3209  0.020293 *
## mu       4.08699    0.21002 19.4599 < 2.2e-16 ***
## lnsigma -1.12718    0.42059 -2.6800  0.007363 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 336.6988
##
##             Estimate Std. Error  z value Pr(z)  Lower.95  Upper.95
## prevalence 0.7731675  0.0926652 8.343666     0 0.5475307 0.9056694
```

```
predict(forw_1, distrx = "rxdate", quantile = 0.8)
```

```
##   Estimate       SE        z      p_value Lower.95 Upper.95
## 1 78.22788 12.03485 6.500115 8.025837e-11 54.64002 101.8157
```

```
# 2) single index date - reverse
rev_1 <- wtdttt(data = df,
                rxdate ~ dlnorm(logitp, mu, lnsigma),
                id = "pid",
                start = as.Date('2014-01-01'),
                end = as.Date('2014-12-31'),
                reverse = T
)
```

```
## Warning in wtdttt(data = df, rxdate ~ dlnorm(logitp, mu, lnsigma), id = "pid",
## : Some dates are out of the window defined by start and end. Keeping only rows
## within the window.
```

```
summary(rev_1)
```

```
## Maximum likelihood estimation
##
## Call:
## mle2(minuslogl = form, start = init, fixed = list(delta = delta),
##     data = cpy, parameters = parameters_r)
##
## Coefficients:
```

```
##         Estimate Std. Error z value     Pr(z)
## logitp   1.21660    0.60208  2.0207   0.04331 *
## mu       3.94455    0.53103  7.4281 1.102e-13 ***
## lnsigma -0.63158    0.67896 -0.9302   0.35226
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 343.3
##
##             Estimate Std. Error  z value Pr(z)  Lower.95  Upper.95
## prevalence 0.7714641  0.1061507 7.267631     0 0.5091354 0.9165713
```

```
predict(rev_1, distrx = "rxdate", quantile = 0.8)
```

```
##   Estimate      SE        z      p_value Lower.95 Upper.95
## 1 80.80868 22.7642 3.549814 0.0003855036 36.19166 125.4257
```

```
# 3) multiple random index date (m =5)
rev_ran_5 <- ranwtdttt(data = df,
                       rxdate ~ dlnorm(logitp, mu, lnsigma),
                       id = "pid",
                       start = as.Date('2014-01-01'),
                       end = as.Date('2014-12-31'),
                       reverse = T,
                       nsamp = 5,
                       robust = T
)

summary(rev_ran_5)
```

```
## Maximum likelihood estimation
##
## Call:
## mle2(minuslogl = form, start = init, fixed = list(delta = delta),
##     data = cpy, parameters = parameters_r)
##
## Coefficients:
##          Estimate Std. Error z value     Pr(z)
## logitp   1.040378   0.390403  2.6649  0.007702 **
## mu       4.296402   0.081841 52.4969 < 2.2e-16 ***
## lnsigma -1.500827   0.288854 -5.1958 2.038e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 1753.642
##
##             Estimate Std. Error  z value Pr(z)  Lower.95  Upper.95
## prevalence 0.7389229  0.0753149 9.811112     0 0.5683695 0.8588237
```

```
predict(rev_ran_5, distrx = "rxdate", quantile = 0.8)
```

```
##   Estimate      SE        z      p_value Lower.95 Upper.95
## 1 88.59163 6.00726 14.74743 3.196245e-49 76.81761 100.3656
```

```r
# 4) multiple random index date (m =50)
rev_ran_50 <- ranwtdttt(data = df,
                        rxdate ~ dlnorm(logitp, mu, lnsigma),
                        id = "pid",
                        start = as.Date('2014-01-01'),
                        end = as.Date('2014-12-31'),
                        reverse = T,
                        nsamp = 50,
                        robust = T
)

summary(rev_ran_50)
```

```
## Maximum likelihood estimation
##
## Call:
## mle2(minuslogl = form, start = init, fixed = list(delta = delta),
##     data = cpy, parameters = parameters_r)
##
## Coefficients:
##          Estimate Std. Error  z value     Pr(z)
## logitp   0.816647   0.374746   2.1792   0.02932 *
## mu       4.222950   0.030758 137.2961 < 2.2e-16 ***
## lnsigma -1.631253   0.304778  -5.3523 8.686e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 17614.93
##
##             Estimate Std. Error z value Pr(z)  Lower.95  Upper.95
## prevalence 0.6935241  0.0796516 8.70697     0 0.5205282 0.8250775
```

```r
predict(rev_ran_50, distrx = "rxdate", quantile = 0.8)
```

```
##   Estimate      SE       z      p_value Lower.95 Upper.95
## 1 80.45048 4.00058 20.1097 6.068548e-90 72.60949 88.29147
```

```r
# 5) saturated sampling
rev_sat_e <- satwtdttt(data = df,
                       rxdate ~ dlnorm(logitp, mu, lnsigma),
                       id = "pid",
                       start = as.Date('2014-01-01'),
                       end = as.Date('2014-12-31'),
                       robust = T,
                       reverse = F
)


summary(rev_sat_e)
```

```
## Maximum likelihood estimation
##
```

```
## Call:
## mle2(minuslogl = form, start = init, fixed = list(delta = delta),
##     data = cpy, parameters = parameters_r)
##
## Coefficients:
##         Estimate Std. Error  z value   Pr(z)
## logitp   0.93472    0.36930   2.5311 0.01137 *
## mu       4.18727    0.02151 194.6680 < 2e-16 ***
## lnsigma -1.45308    0.13551 -10.7230 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 126619.1
##
##             Estimate Std. Error  z value Pr(z)  Lower.95  Upper.95
## prevalence 0.7180311  0.0747688 9.603352     0 0.5525321 0.8400399
```

```
predict(rev_sat_e, distrx = "rxdate", quantile = 0.8)
```

```
##   Estimate       SE        z       p_value Lower.95 Upper.95
## 1 80.16418 2.704524 29.64077 4.459893e-193 74.86341 85.46494
```