Lin_ST625_HW2

Frances Lin

1/15/2022

h_t = #s of people dying per unit time in the interval / ((total # of patients surviving at t) x (interval width)) or

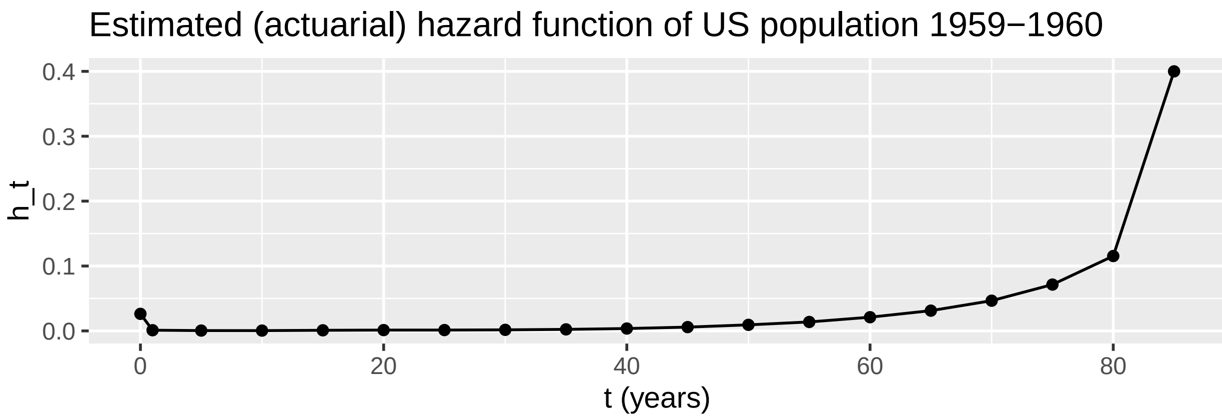
$h_t = \frac{f_t}{S_t}$.

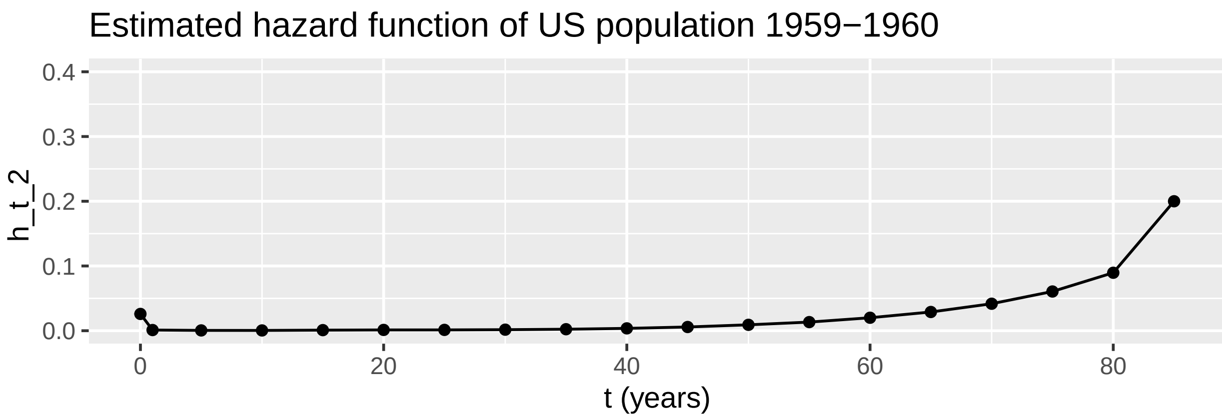
For example, $h(t=1) = \frac{2593}{100000*(1-0)} = 0.02593$, $h(t=2) = \frac{409}{97407*(5-1)} = 0.001049719$, etc.

$h(t=1) = \frac{0.0259300}{1} = 0.02593$, $h(t=2) = \frac{0.0010225}{0.97407} = 0.0010497193$, etc.

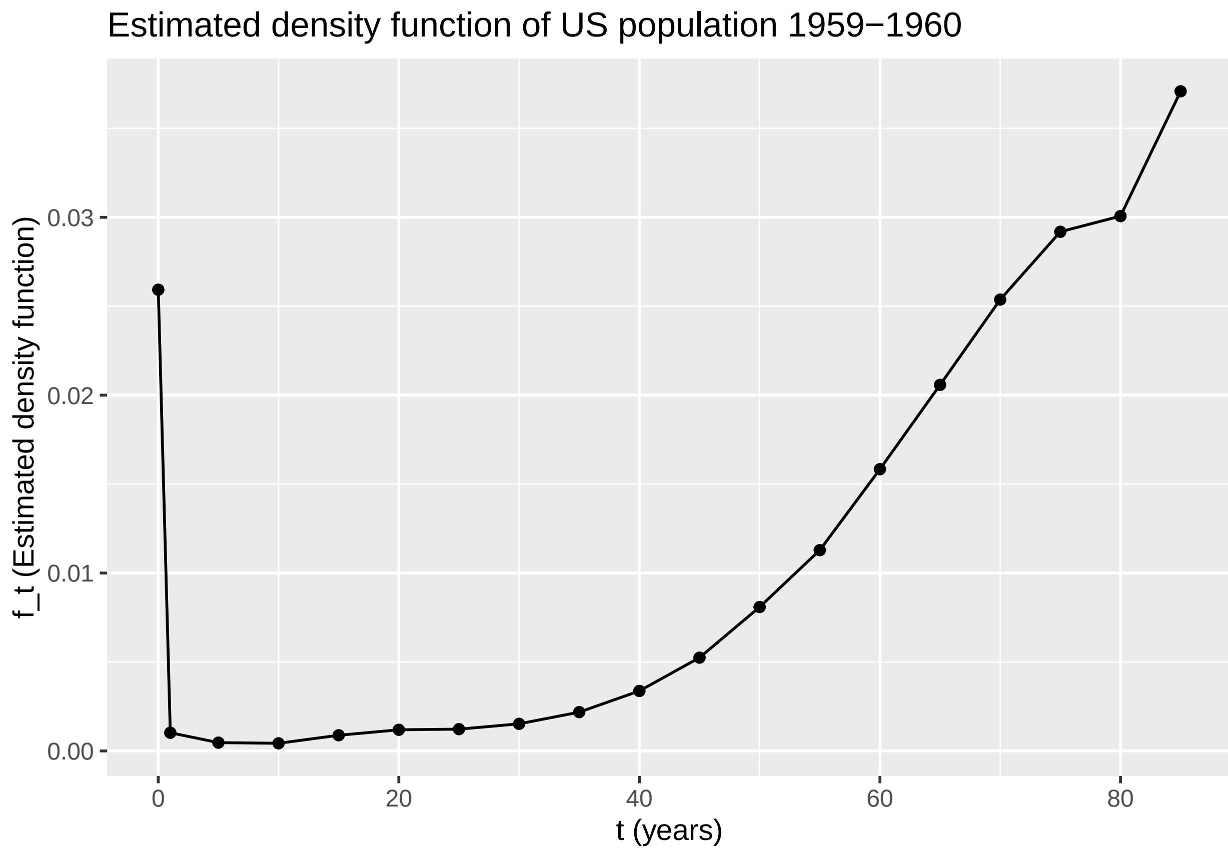
1a

Age	Age_2	Living	Dying	S_t	f_t	h_t	h_t_2
0	1	1e+05	2593	1	0.02593	0.02627	0.02593
1	5	97407	409	0.9741	0.001022	0.001052	0.00105
5	10	96998	233	0.97	0.000466	0.000481	0.0004804
10	15	96765	214	0.9677	0.000428	0.0004428	0.0004423
15	20	96551	440	0.9655	0.00088	0.0009135	0.0009114
20	25	96111	594	0.9611	0.001188	0.00124	0.001236


Estimated (actuarial) hazard function of US population 1959−1960
h_t
0.0
0.1
0.2
0.3
0.4
0
20
40
60
80
t (years)


Estimated hazard function of US population 1959−1960
h_t_2
0.0
0.1
0.2
0.3
0.4
0
20
40
60
80
t (years)

1b



1c

The hazard function follows a bathtub shape, although it is not quite symmetric. Initially, the risk is high, decreases a bit, and then gradually increases again.

The initial "spike" perhaps has to do with the high risk for newborns, though the uneven age interval for Age $0-1$ and Age $1-5$ may also contribute to it.

1d

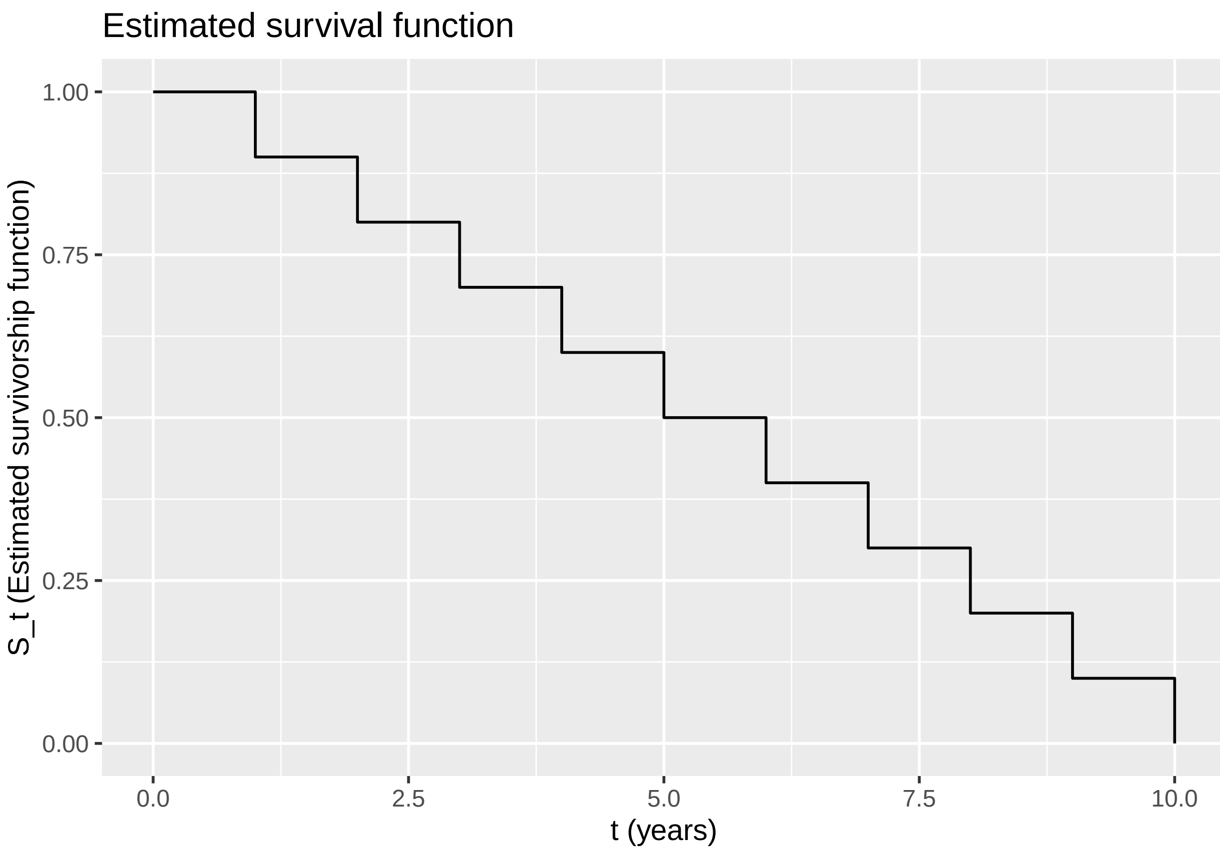
$h(t)$, the hazard function, is an conditional instantaneous failure rate at time t *given that* the individual survived to t, and it captures the expected number of events during one unit time period after t, given $T \geq t$.

$f(t)$, the density function, on the other hand, is an unconditional instantaneous failure rate at time t.

2a

Since $f(t) = \frac{1}{10}$ for $t \in [0, 10]$, $F(t) = \int_0^t f(x)dx = \int_0^t \frac{1}{10}dx = \frac{1}{10}x\Big|_0^t = \frac{1}{10}t$. Therefore, $S(t) = 1 - F(t) = 1 - \frac{1}{10}t$. Or

$S(t) = \int_t^{10} f(s)ds = \int_t^{10} \frac{1}{10}ds$ would give the same answer.

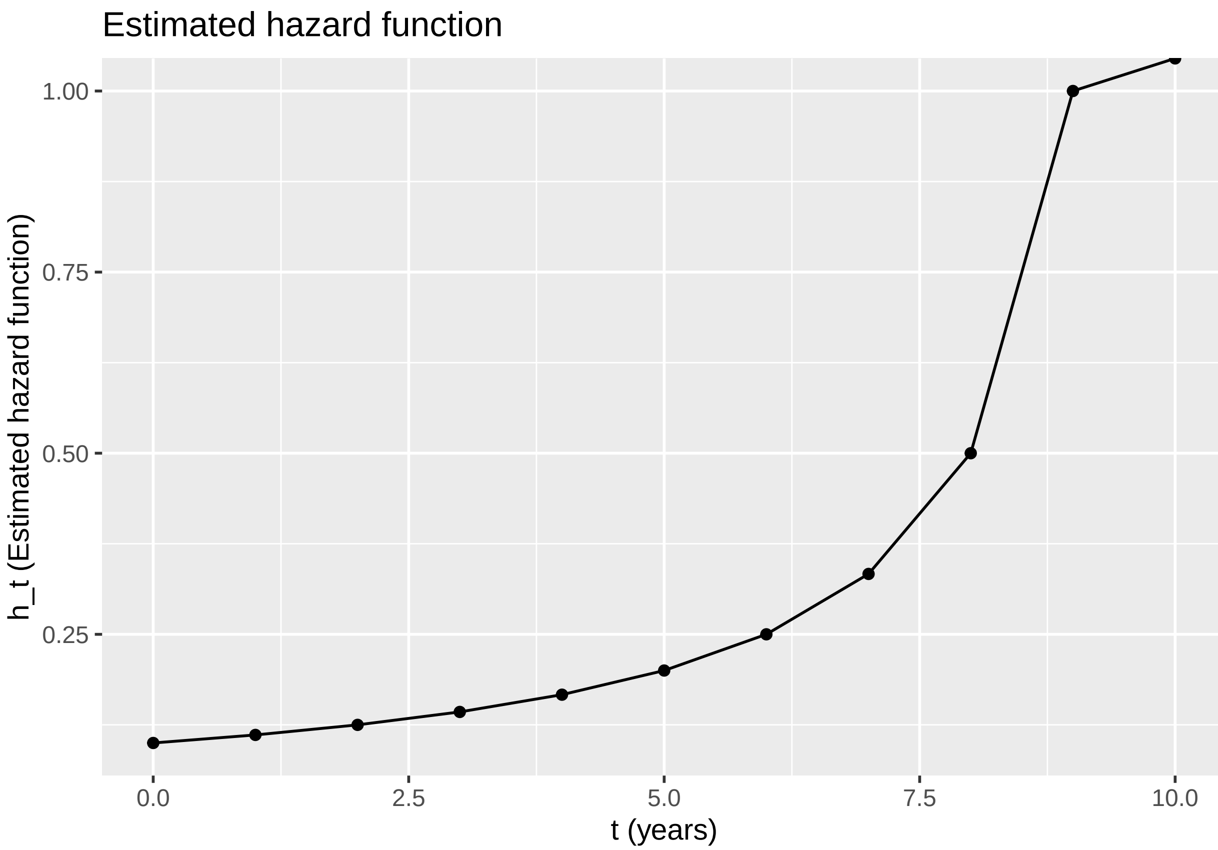


2b

0.4 since $S(6) = 1 - \frac{1}{10}(6) = 0.4$.

2c

$h(t) = \frac{f(t)}{S(t)} = \frac{\frac{1}{10}}{1-\frac{1}{10}t} = \frac{1}{10-t}$ for $t \in [0, 10]$.



2d

Recall that $h(t) = \frac{f(t)}{S(t)}$, since in this example, $f(t)$ (unconditional failure rate) is constant and $S(t)$ decreases over time, $h(t)$ (conditional failure rate) increases over time.

2e

$h(t = 6) = 0.25$.

For those who survive at the 6th year, the probability of them dying during one unit time period after the 6th year (aka the 7th year) is 0.25. (For those who are still alive at time 10, their instantaneous rate is 0.25).

t	S_t	f_t	h_t	h_t_2
6	0.4	0.1	0.25	0.25

3a

Exponential distribution (= Weibull with $\alpha = 0$) has constant hazard rates.

(e.g. risk of healthy persons between 18 and 40 years)

3b

Weibull distribution with $\alpha > 0$ has increasing hazard rates.

(e.g. cancer patients who do not receive treatment)

3c

Weibull distribution with $\alpha < 0$ has decreasing hazard rates.

(e.g. wounded solider who undergo surgery)

3d

Log-normal distribution has non-monotone rates.

(e.g. Ebola infection)

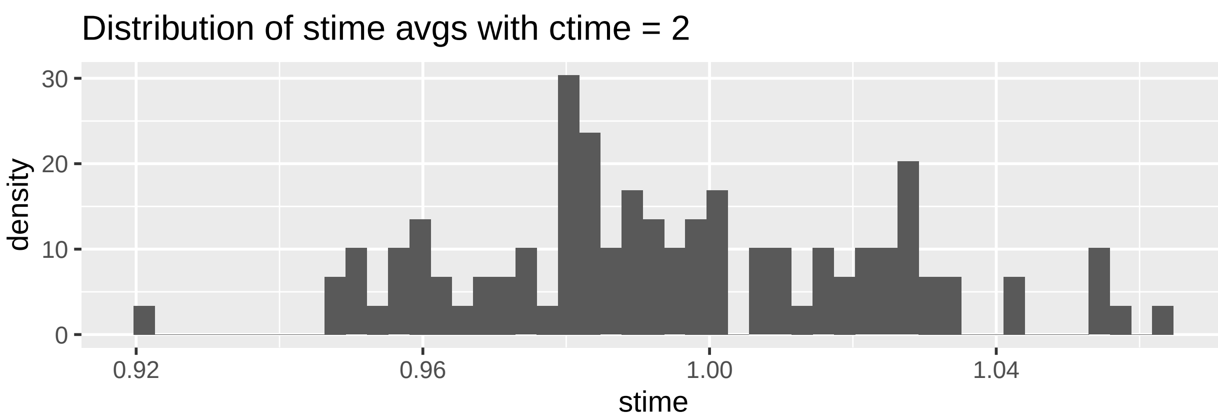
4a

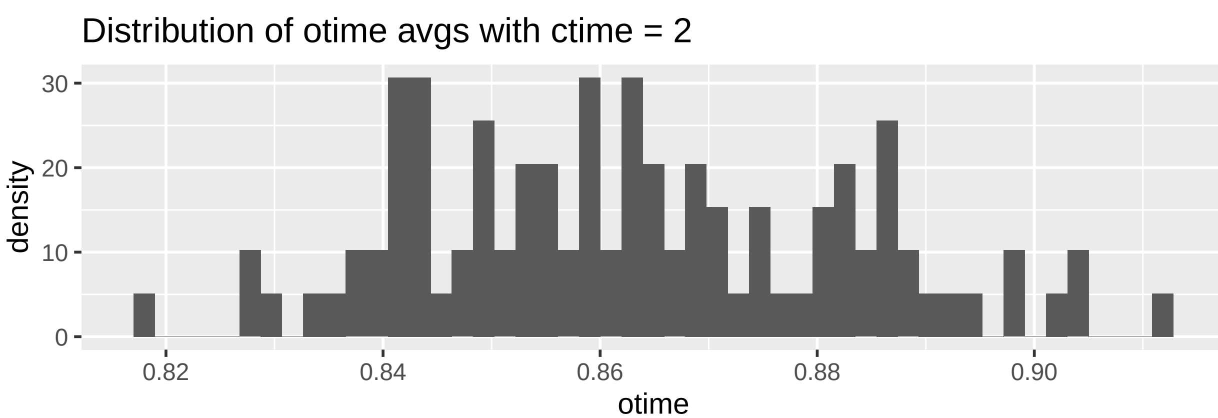
The mean and sd of these 100 stime averages:

mean	std.dev
0.9951	0.02881

The mean and sd of these 100 otime averages:

mean	std.dev
0.863	0.01975





Average censoring rate:

```
## [1] 0.1369563
```

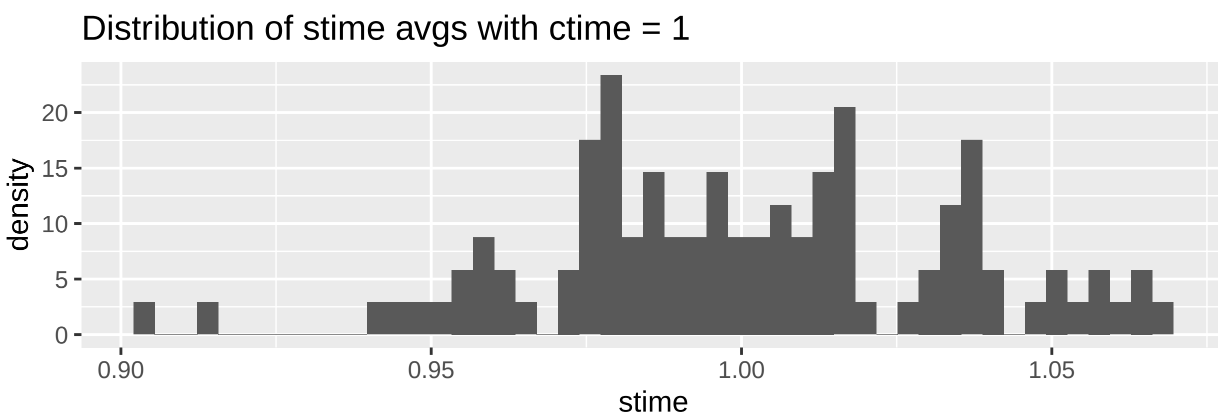
4b

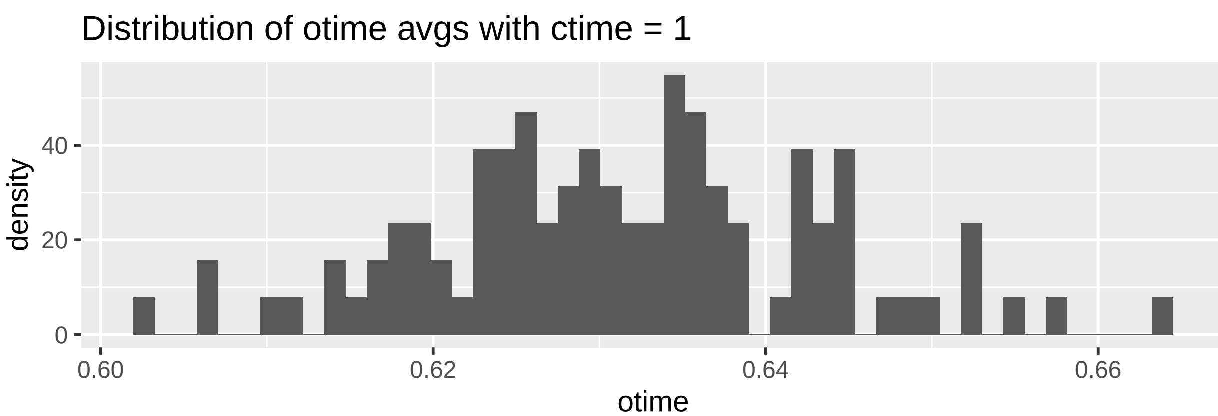
The mean and sd of these 100 stime averages:

mean	std.dev
1.001	0.03319

The mean and sd of these 100 otime averages:

mean	std.dev
0.6314	0.01159





Average censoring rate:

```
## [1] 0.3685652
```

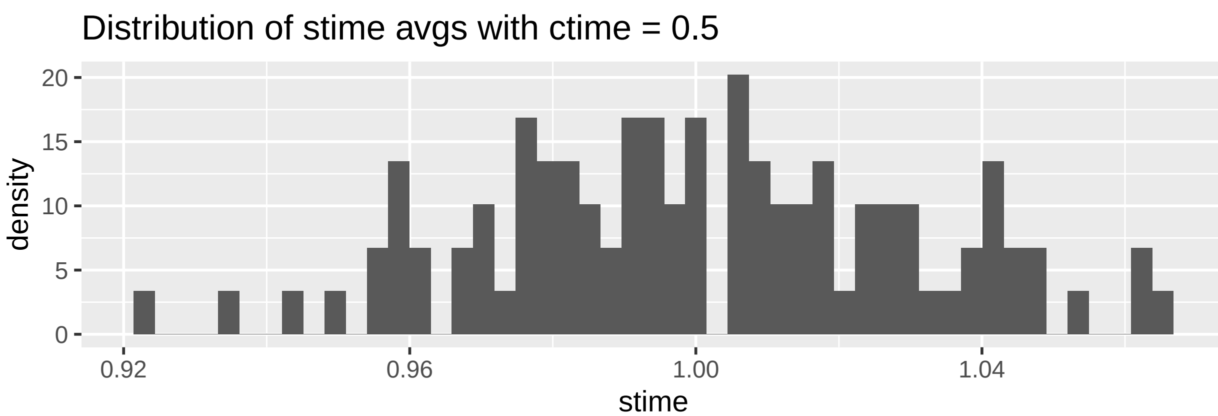
4c

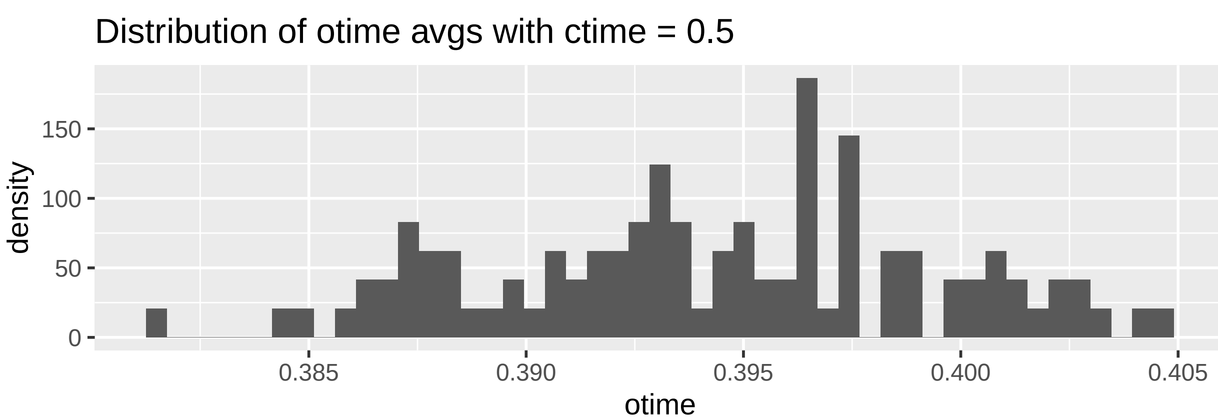
The mean and sd of these 100 stime averages:

mean	std.dev
1	0.02978

The mean and sd of these 100 otime averages:

mean	std.dev
0.3943	0.005078





Average censoring rate:

```
## [1] 0.6057391
```

4d

The resulting estimation and inference results are biased if we ignore censoring. (Sample averages of censored data (whether `ctime = 2`, `ctime = 1` or `ctime = 0.5`) are all smaller than the population mean with sample average of `ctime = 0.5` being the smallest among all.)

Higher censoring rate causes the distribution of otime averages to shift further to the left. (Censoring rate increases as `ctime` decreases with censoring rate of `ctime = 0.5` being the largest among all.)