Bias

Matt Kosko

2024-04-18

```
dat <- as.data.table(readRDS(file.path(base_dir, dat_path, paste0(base_nm, '_tmp'), 'coverage.rds')))
```

Coverage

Percentile

```
tmp <- dat[, .(lower = boot:::perc.ci(boot_reps)[4], upper = boot:::perc.ci(boot_reps)[5]),
           by = c('subset_num', 'replication', 'n')][
          ,.(lower = mean(lower), upper = mean(upper)), by = c('replication', 'n')
        ]
tmp[, .(coverage = mean(lower <= te & upper >= te)), by = c('n')]
```

```
##        n coverage
## 1: 20000    0.884
## 2: 50000    0.864
```

Normal

```
tmp <- dat[, .(lower = mean(boot_reps) - 1.96*sd(boot_reps), upper = mean(boot_reps) + 1.96*sd(boot_reps
           by = c('subset_num', 'replication', 'n')][
          ,.(lower = mean(lower), upper = mean(upper)), by = c('replication', 'n')
        ]
tmp[, .(coverage = mean(lower <= te & upper >= te)), by = c('n')]
```

```
##        n coverage
## 1: 20000    0.876
## 2: 50000    0.850
```

Bias

```
true_sd <- dat[, .(SD_full = mean(sd_full)), by = c('n')]
```

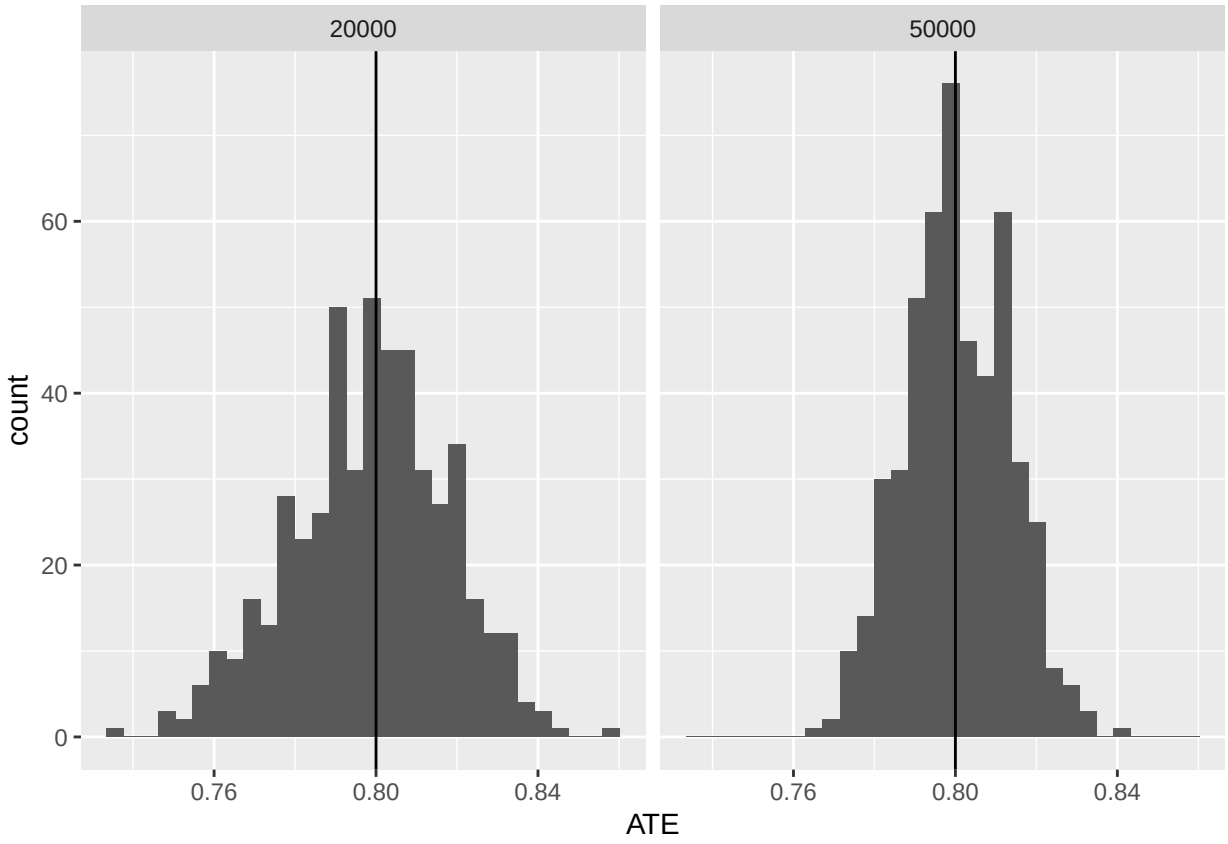
ATE

```
bias <- dat[, .(ATE = mean(boot_reps),
        SE = sd(boot_reps)), by = c('subset_num', 'replication', 'n')][
          , .(ATE = mean(ATE), SE = mean(SE)), by = c('replication', 'n')
        ]

ggplot(bias, aes(x = ATE)) +
```

```
  geom_histogram() +
  facet_wrap(~ n) +
  geom_vline(xintercept = te)
```

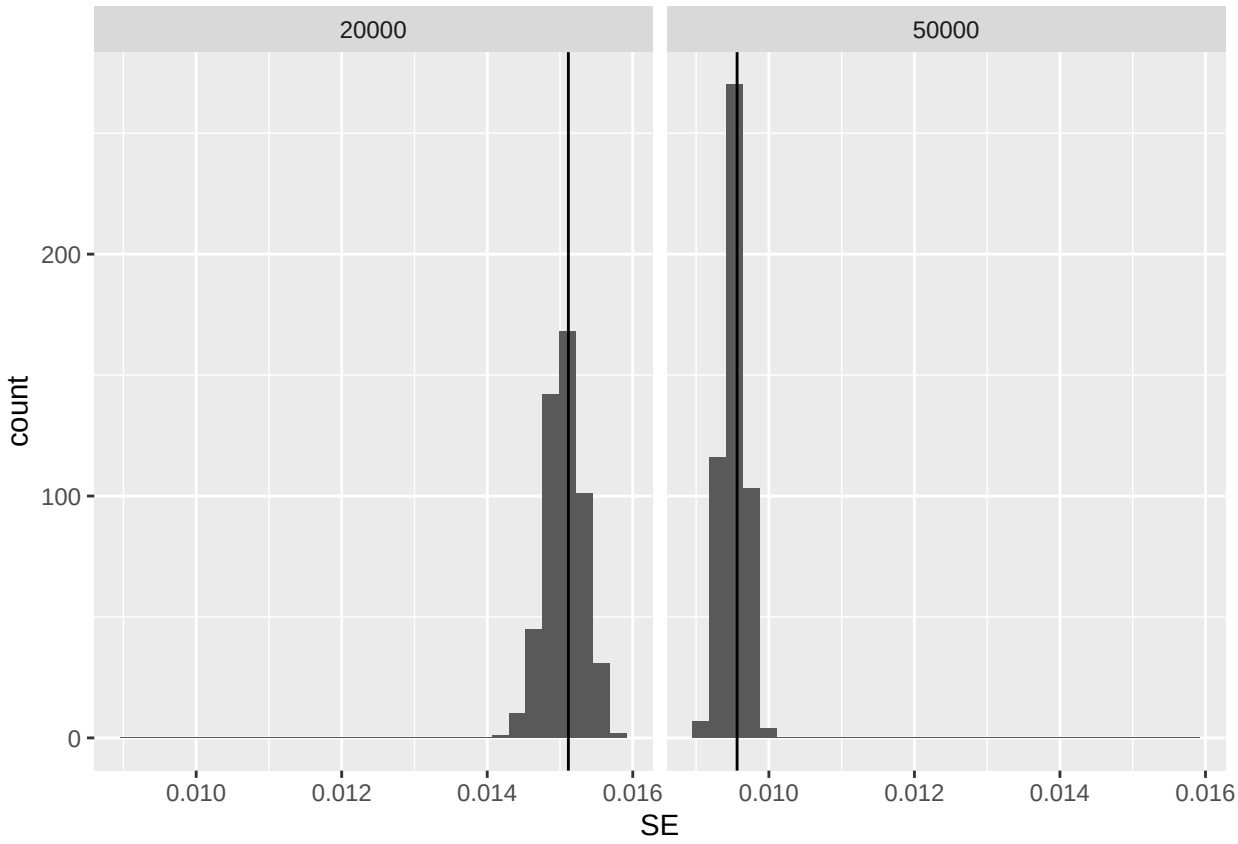
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Standard Error

```
ggplot(bias, aes(x = SE)) +
  geom_histogram() +
  facet_wrap(~ n) +
  geom_vline(data = true_sd, aes(xintercept = sqrt(SD_full)))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


20000
50000
200
100
0
count
0.010
0.012
0.014
0.016
SE