variance_power.R

mollyow

2020-10-14

```r
## Power calculations to estimate variance
# Goal: what is the desired sample size to get an estimate of the variance where
# the width of the empirical 95% confidence interval is <= .15
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```r
set.seed(94305)
nsims <- 1e4
c <- .15

# Assumptions ----
# Sharing rates are 50/% for both true and false at baseline
# Assume maximum variance; true variance:
var(sample(c(-2, 1), 1e5, replace = TRUE))
```

```
## [1] 2.249998
```

```r
# hypothetical sample size
ss <- seq(500, 4000, 50)
# sample variance
svmat <- matrix(NA, nrow = nsims, ncol = length(ss))

for(j in 1:length(ss)){
  n <- floor(ss[j]/40)
  for(i in 1:nsims){
    newr <- sample(c(-2, 1), n, replace = TRUE)
    svmat[i,j] <- var(newr)
  }
}


df <- apply(svmat, 2, function(x) c(mean(x), quantile(x, c(0.025, 0.975))))

df <- data.frame(size = ss,
                 C = df[1,],
                 L = df[2,],
                 U = df[3,],
                 W = df[3,]-df[2,])

ggplot(df,
       aes(x = size, y = C)) +
  geom_point() +
  geom_point(aes(x = size, y = L)) +
```
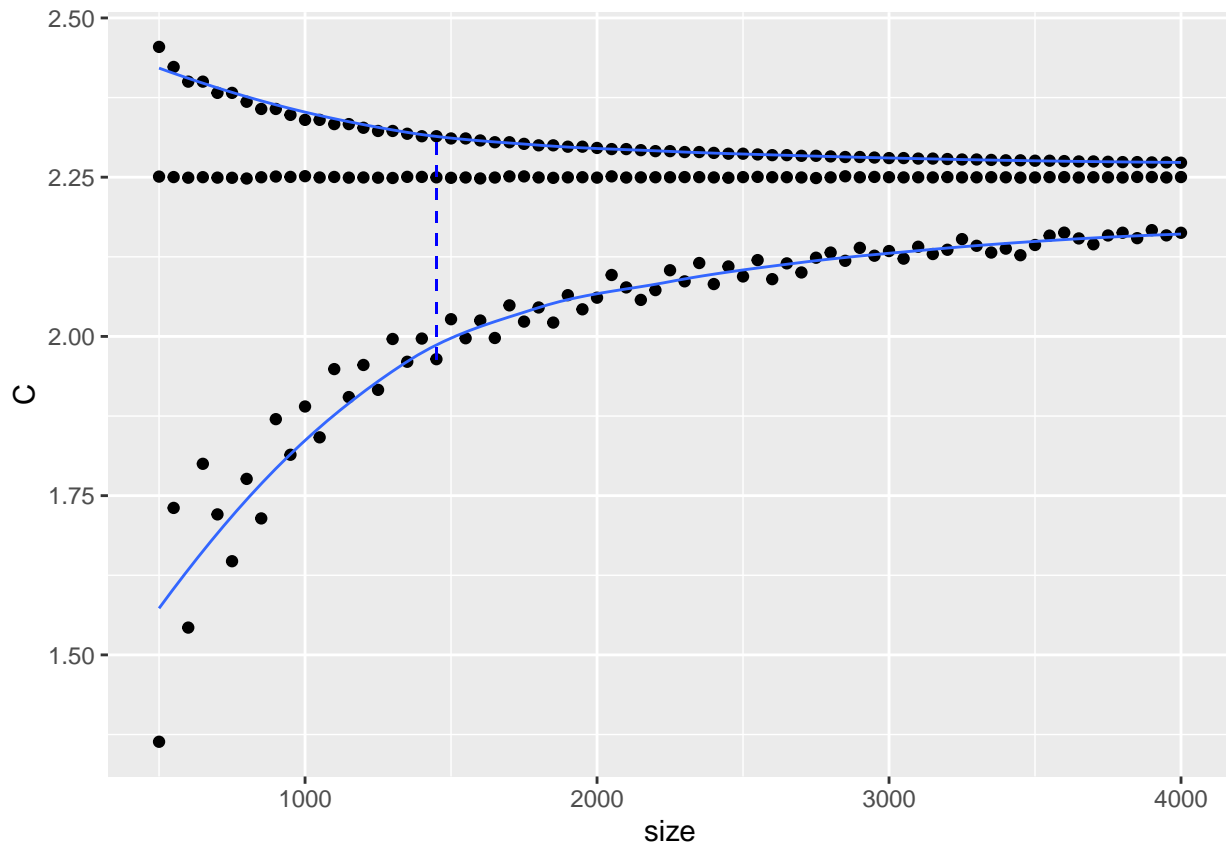
```
  geom_smooth(aes(x = size, y = L), se = FALSE, size = 0.5) +
  geom_point(aes(x = size, y = U)) +
  geom_smooth(aes(x = size, y = U), se = FALSE, size = 0.5) +
  geom_segment(aes(x = ss[max(which(W>2.25*c))],
                   y = L[max(which(W>2.25*c))],
                   xend = ss[max(which(W>2.25*c))],
                   yend = U[max(which(W>2.25*c))]),
               data = df, colour = 'blue', lty = 'dashed')
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
ss[max(which(df$W>2.25*c))]
```

```
## [1] 1450
```