

Untitled

MSRR 6.44

Let X_1, X_2, \dots, X_n be i.i.d. from a normal distribution with mean μ and variance σ^2 . Show that $\hat{\sigma}_n^2 = (1/n) \sum_{i=1}^n (X_i - \bar{X})^2$ is a consistent estimator of σ^2 . *Hint:* Theorem B.10.5.

Vi har ud fra den er en normal fordeling at middelværdien og variansen er:

$$E(X_i) = \mu$$

$$\text{var}(X_i) = \sigma^2$$

Generelle regler:

$$E[\sum X_i] = \sum E[X_i]E[xX] = xE[X]\text{var}(X) = E(X^2) - [E(X)]^2$$

Middelværdien

Hvor middelværdier

$$E[X^2] = \sigma^2 + \mu^2 \text{var}(\bar{X}) = E[\bar{X}^2] - [E(\bar{X})]^2 E[\bar{X}^2] = \sigma^2/n + \mu^2$$

$$E[S^2] = E\left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}\right]$$

Finder middelværdien af tælleren: Husk at X bare er en konstant.

$$\begin{aligned} & E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] \\ &= E\left[\sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2)\right] = E\left[\sum X_i^2 - \sum 2X_i\bar{X} + \sum \bar{X}^2\right] \\ &= E\left[\sum X_i^2 - 2\bar{X} \sum X_i + n\bar{X}^2\right] = E\left[\sum X_i^2 - 2\bar{X} \cdot n\bar{X} + n\bar{X}^2\right] \\ &= E\left[\sum X_i^2 - n\bar{X}^2\right] = \sum E[X_i^2] - E(n\bar{X}^2) \end{aligned}$$

Nu kan vi bruge de forhold som jeg opskrev tidligere

$$E[X_i^2] = \sigma^2 + \mu^2 E[\bar{X}^2] = \frac{\sigma^2}{n} + \mu^2$$

Hermed kan vi substituere de værdier ind:

$$= \sum E[X_i^2] - E(n\bar{X}^2) = \sum (\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2) = n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2$$

Vi har således

$$= (n-1)\sigma^2$$

tilbage

$$\frac{1}{n}(n-1)\sigma^2$$

$$\frac{(n-1)}{n}\sigma^2$$

Således kan vi se at vi har en biased estimator.

Men lad os se om den er konsistent. Her skal vi se på

$$n \longrightarrow \infty$$

hvad sker der med vores bias når n går mod uendelig.

Hvis n bliver uendelig stor, så betyder det -1 ikke noget. Derfor vil estimatoren konvergere mod den populationsværdien.

Så på trods af vi ikke havde en unbiased estimator så var den konsistent.

Variansen

Vi skal også se på variansen.

$$var[\hat{\sigma}_n^2] = var[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2]$$

$$\frac{1}{n^2} var[\sum_{i=1}^n (X_i - \bar{X})^2]$$

$$\frac{1}{n^2} \sum_{i=1}^n var[(X_i - \bar{X})^2]$$

$$\frac{1}{n^2} (\sum_{i=1}^n E[(X_i - \bar{X})^4] - \sum_{i=1}^n E[(X_i - \bar{X})^2]^2)$$

$$\frac{1}{n^2} \left(\sum_{i=1}^n 3\sigma^4 - \sum_{i=1}^n (\sigma^2)^2 \right)$$

$$\frac{1}{n^2} \left(\sum_{i=1}^n 2\sigma^4 \right)$$

$$\frac{1}{n^2} n 2\sigma^4$$

$$\frac{2\sigma^4}{n}$$

Så den er heller ikke unbiased.

Men idet den vil konvergere mod den sande populations værdi er den konsistent.

simulering

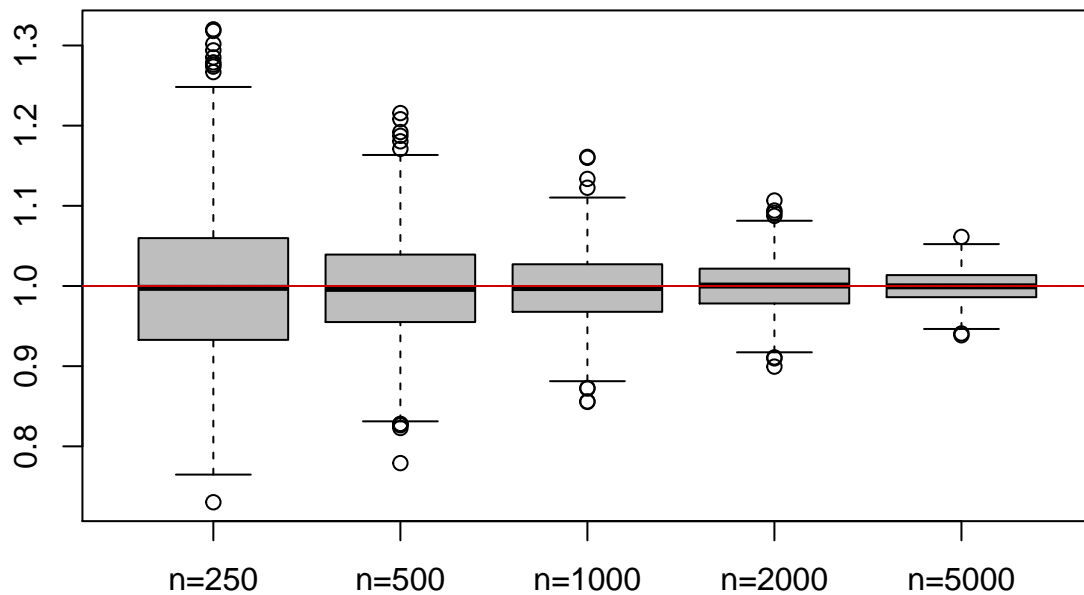
```
set.seed(1234)
# sætter parametre til simulation og til plot
nrep <- 1000

# reserverer plads til resultater,
# en vektor hver til stikprøvestørrelser n=250, 500, 1000, 2000, 5000
res.250 <- numeric(nrep)
res.500 <- numeric(nrep)
res.1000 <- numeric(nrep)
res.2000 <- numeric(nrep)
res.5000 <- numeric(nrep)

# kør simulationen

for (i in 1 : nrep){
  res.250[i] <- (250 - 1) / 250 * var(rnorm(250, mean = 0, sd = 1))
  res.500[i] <- (500 - 1) / 500 * var(rnorm(500, mean = 0, sd = 1))
  res.1000[i] <- (1000 - 1) / 1000 * var(rnorm(1000, mean = 0, sd = 1))
  res.2000[i] <- (2000 - 1) / 2000 * var(rnorm(2000, mean = 0, sd = 1))
  res.5000[i] <- (5000 - 1) / 5000 * var(rnorm(5000, mean = 0, sd = 1))
}

# pæn x-akse til boksplotten
boxnam <- c("n=250", "n=500", "n=1000", "n=2000", "n=5000")
theta = 1
epsi = 0
boxplot(res.250, res.500, res.1000, res.2000, res.5000, names = boxnam, col="gray")
abline(h = c(1, 1), col = "red", lty = "dashed")
abline(h = 1, col = "red3", lty = "solid")
```



```
# beregn, hvor mange værdier ligger er mellem theta +/- epsilon
```

```
p.250 <- var((res.250 > 1) & (res.250 < 1))
p.500  <- var((res.500 > 1) & (res.500 < 1))
p.1000 <- var((res.1000 > 1) & (res.1000 < 1))
p.2000 <- var((res.2000 > 1) & (res.2000 < 1))
p.5000 <- var((res.5000 > 1) & (res.5000 < 1))
```

```
# nu klistrer vi alle værdier sammen i en vektor, og navngiver dem som boxplottene.
# Så får vi navne med, når vektoren printes
```

```
alle.p <- c(p.250, p.500, p.1000, p.2000, p.5000)
names(alle.p) <- boxnam
alle.p
```

```
##  n=250  n=500 n=1000 n=2000 n=5000
##      0      0      0      0      0
```