Untitled

MSRR 6.44

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

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

$$E(X_i) = \mu$$

$$var(X_i) = \sigma^2$$

Generelle regler:

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

Hvor middelværdier

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

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

Finder middelværiden af tælleren: Husk at X bar er en konstant.

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

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

$$= E[\sum X_i^2 - 2\bar{X}\sum X_I + n\bar{X}^2] = E[\sum X_i^2 - 2\bar{X} \cdot n\bar{X} + n\bar{X}^2]$$

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

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

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

Sålees 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.

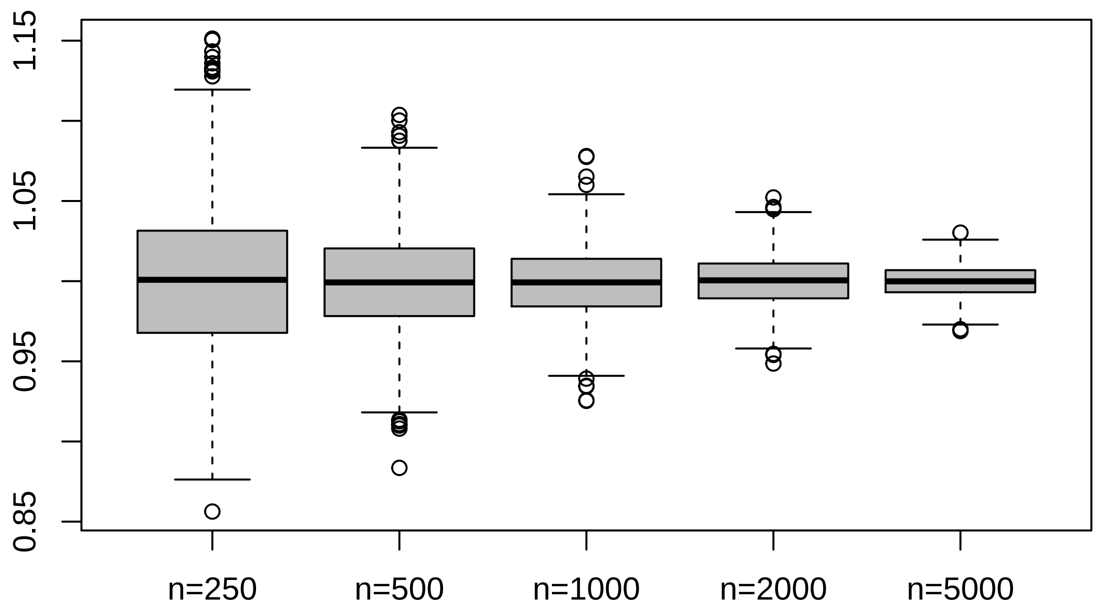
Hvis n bliver uendelig stor, så betyder det -1 ikke noget. Derfor vil estimatoren konvergere mod den populationsværdien.

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

simulering

```
set.seed(1234)
# sætter parametre til simulation og til plot
nrep <- 1000
theta <- 5
epsi <-  .05

# 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] <-  sd(rnorm(n = 250,   mean = 0, sd = 1))
  res.500[i] <-  sd(rnorm(n = 500,   mean = 0, sd = 1))
  res.1000[i] <- sd(rnorm(n = 1000, mean = 0, sd = 1))
```

```r
  res.2000[i] <- sd(rnorm(n = 2000, mean = 0, sd = 1))
  res.5000[i] <- sd(rnorm(n = 5000, mean = 0, sd = 1))
}

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




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

p.250 <- mean((res.250 > theta - epsi) & (res.250 < theta + epsi))
p.500 <- mean((res.500 > theta - epsi) & (res.500 < theta + epsi))
p.1000 <- mean((res.1000 > theta - epsi) & (res.1000 < theta + epsi))
p.2000 <- mean((res.2000 > theta - epsi) & (res.2000 < theta + epsi))
p.5000 <- mean((res.5000 > theta - epsi) & (res.5000 < theta + epsi))

# 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
```