Oppgaver, STK4040, uke 37

11. september 2005

1 Trekke fra $N_p(\mu, \Sigma)$

Ikke alle statistikkprogrammer har funksjoner for å trekke tilfeldige observasjoner direkte fra multinormalfordelingen. Da må man trekke fra (den univariate) normalfordelingen og transformere dataene slik at de får riktig fordeling: hvis X er en datamatrise fra $N_p(\mathbf{0}, \mathrm{I})$, så er $\mathrm{X}\Sigma^{1/2} + \mathbf{1}\mu^t$ en datamatrise fra $N_p(\mu, \Sigma)$.

a)

Trekk 1000 observasjoner fra $N_2(\mu, \Sigma)$, der $\mu^t = (1, 2)$ og $\Sigma = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, ved å trekke fra $N(0, 1)$ og bruke transformasjonen over. (Hint: bruk egenverdidekomposisjonen (`eigen()` i `R`) til å beregne $\Sigma^{1/2}$.

Beregn $\bar{\boldsymbol{x}}$ og S og sammenlign med μ og Σ som en sjekk.

b)

Trekk 1000 observasjoner fra samme fordeling ved å bruke en ferdiglagd funksjon for å trekke fra multinormale fordelinger. (I `R`: si `library(MASS)` for å laste pakken `MASS`, som inneholder funksjonen `mvrnorm()`.)

Beregn $\bar{\boldsymbol{x}}$ og S og sammenlign med resultatene i 1a).

Tilleggsspørsmål: Er den innebygde kovariansfunksjonen forventningsrett (det vil si beregner den S eller $\mathrm{S_u}$)?

2 Plotte multinormalfordelte data

Trekk $n = 100$ observasjoner (x_1, x_2) fra $N_2(\mu, \Sigma)$, med $\mu^t = (1, 2)$ og forskjellige Σ: I_2, $\begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$ og $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$.

I hvert tilfelle:

- Plott x_2 mot x_1
- Finn egenverdidekomposisjonen til Σ.

- Sammenlign egenvektorene med retningene på halvaksene i plottet
- Sammenlign egenverdiene med lengdene på halvaksene.
- Ekstraoppgave: estimer Σ fra dataene, og sammenlign egenverdidekomposisjonen med «sannheten».

Trekk flere ganger og se på variasjonene i plott og egenverdistruktur. Prøv gjerne med andre Σ og n.

3 Hotellings T^2

Trekk $n = 100$ observasjoner fra

$$N_3\left(\begin{pmatrix}1\\1\\1.5\end{pmatrix}, \begin{pmatrix}1 & 0.2 & 0.7\\0.2 & 1.5 & -0.4\\0.7 & -0.4 & 2\end{pmatrix}\right).$$

Test nullhypotesen H_0: $\mu = \mathbf{0}$ mot den alternative hypotesen H_1: $\mu \neq \mathbf{0}$. (Tips: `R` har ikke T^2-fordelingen, men den har F-fordelingen; se `?FDist`.) Trekk flere ganger. Hvorfor varierer p-verdien?

Noen kjekke `R`-kommandoer

- `?funk` eller `help(funk)` gir hjelp om en funksjon `funk`. `help.start()` åpner hele `R`-dokumentasjonen i en nettleser. `help.search(etellerannet")` søker etter `etellerannet` i `R`-dokumentasjonen.
- Tilordning i `R` gjøres med `<-`; f.eks. `mu <- c(1, 2)` lagrer vektoren $(1, 2)$ i variabelen `mu`.
- for å lage en $m \times n$ matrise fra an nm-vektor `v`: `matrix(v, nrow = n)`.
- bruk `solve(M)` for å invertere en matrise `M`
- matriseprodukt i `R` er `%*%`; f.eks. `X %*% S`.
- `rep(1, n)` gir deg $\mathbf{1}_n$, og `diag(n)` er I_n