Frank Bengtson **2015**
frank.bengtson@gmail.com

Tilføjelse til eks. *Uafhængighed mellem tøjkøb og køn?*

Eks. på anvendelse af programmel R på den ret omtalte metode omrøring.

Man havde

98	102	200
60	100	160
158	202	360

og skal bestemme løsninger $x = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$, der giver faste række- og søjlesummer. Det ses at x_{11} kan vælges som parameter i $0, 1, \ldots, 158$, men hvordan?

Rækker og søjler som sædvanlig

	1	2	
1	x_{11}	x_{12}	200
2	x_{21}	x_{22}	160
	158	202	360

Rækker og søjler repræsenteres som

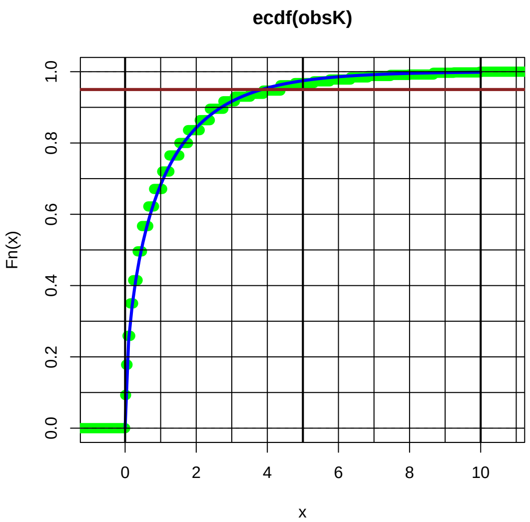
$$\begin{aligned} \text{rækker } r &: \overbrace{1,1,1,\ldots,1,1,1,1,1,1,\ldots,1,1,1}^{200}, \overbrace{2,2,2,\ldots,2,2,2}^{160}\,, \\ \text{søjler } s &: \underbrace{1,1,1,\ldots,1,1,1}_{158}, \underbrace{2,2,2,\ldots,2,2,2,2,2,2,\ldots,2,2,2}_{202}\,, \end{aligned}$$ og optælles som x_{rs}.

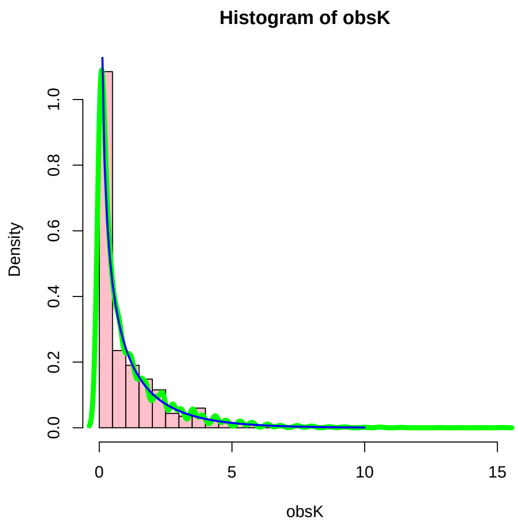
Begyndelsesværdierne for rækker r og søjler s ovenfor svarer til

158	42	200
0	160	160
158	202	360

.

sample(r) giver en blanding af tallene i *r*, *df=data.frame(r,s)* laver to lister/rækker og *table(df)* optæller $x_{11}, x_{12}, x_{21}, x_{22}$. K og eps1 som før eller af *chisq.test(table(df))$p.value* og X-squared af $sum(chisq.test(table(df))\$residuals^2)$. Grønne kraftige grafer observerede størrelser og blå tæthed og fordeling for χ^2_1. Spring i K værdier afspejler x_{11} diskret.



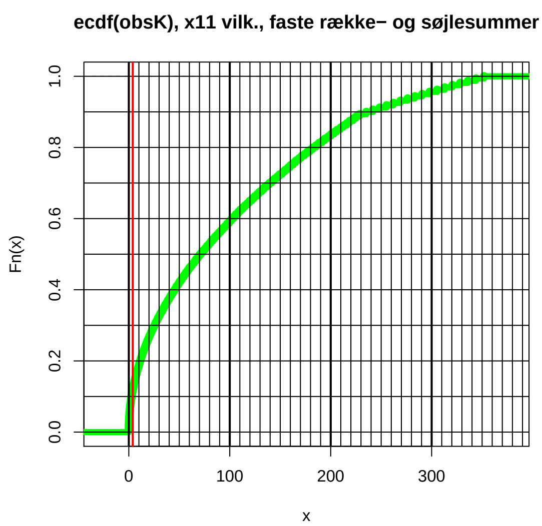


Omrøring giver en fordeling svarende til hypergeometrisk $\frac{\binom{158}{x_{11}}\binom{202}{200-x_{11}}}{\binom{360}{200}}$ og kan simuleres med x_{11}=*rhyper(1,158,202,200)*, $x_{21} = 158 - x_{11}, x_{12} = 200 - x_{11}, x_{22} = 2 + x_{11}$.

R program til omrøring af 2×2 tabel

```
r[1:200]<-1    # 1.række
r[201:360]<-2  # 2.række
r              # rækker
c[1:158]<-1    # 1.søjle # s er åbenbart reserveret
c[159:360]<-2  # 2.søjle
c              # søjler
data.frame(r,c)

r1<-sample(r)
c1<-sample(c)
df1<-data.frame(r1,c1)
df1
table(df1)
addmargins(table(df1))

r5<-sample(r) # test r
r5
table(r5)     # se at antallet passer
c5<-sample(c) # test c
c5
table(c5)

antaleksp<-1000
obsK<-rep(0,antaleksp)
#XX[1:2000]<-0

for (m in 1:antaleksp){
  r1<-sample(r1)
  c1<-sample(c1)
  df1<-data.frame(r1,c1)
  obsK[m]<-K(table(df1)) }

# test sample
t1<-seq(1:100)    # 1 til 100
t1
s2<-0
for (m in 1:10000){s2<-s2+sample(t1)}
s2/100

x11 hypergeometrisk, samme K værdier og grafer
for (m in 1:antaleksp){
  x11<-rhyper(1,158,202,200)
  x21<-158-x11
  x12<-200-x11
  x22<-2+x11
  row1<-c(x11,x12)
  row2<-c(x21,x22)  ...... }
```

Man kunne overveje, hvordan K ser ud, hvis x_{11} vælges vilkårligt blandt $0, 1, \ldots, 158$ og søjle- og rækkesummer holdes faste. Det ser ud til at K værdierne bliver mellem nul og 350 og ca. 12% er mindre end 3.84=qchisq(0.95,df=1).




```
# faste søjle- og rækkesummer, x11 vælges vilkårligt
for (m in 1:antaleksp){
  x11<-sample(0:158,1)
  x21<-158-x11
  x12<-200-x11
  x22<-2+x11  ...}
```