Matematisk Statistik: Vejledende besvarelse af eksamen

Steffen Lauritzen og Niels Richard Hansen

18. juni, 2020

Spørgsmål 1.1

Vi har tætheden

$$f_{(\beta,\gamma)}(x,y) = \frac{1}{\beta}e^{-x/\beta}\frac{1}{\gamma}e^{-y/\gamma}$$

som vi omskriver på eksponentiel familie form ved at lade $\theta_1 = 1/\beta$ og $\theta_2 = 1/\gamma$ og derfor med $t(x,y) = -(x,y)^\top$

$$f_\theta(x,y) = \theta_1\theta_2 e^{\theta^\top t(x,y)} = e^{\theta^\top t(x,y) - \log\theta_1 - \log\theta_2}.$$

Idet $\lambda_1 x + \lambda_2 y = c$ for næsten alle (x,y) medfører $\lambda_1 = \lambda_2 = 0$, fremgår det, at den specificerede familie uden restriktioner på θ er en regulære og minimalt repræsenteret eksponentiel med dimension 2. Under hypotesen defineres en krum familie idet

$$\phi(\beta) = \begin{pmatrix} 1/\beta \\ 1/\beta^2 \end{pmatrix}$$

er en glat homeomorfi med Jacobi matrix

$$D\phi(\beta) = (-1/\beta, -2/\beta^2)$$

som har fuld rang 1.

Spørgsmål 1.2

Vi får likelihoodfunktionen

$$L_n(\beta) = \prod_{i=1}^n \frac{e^{-x_i/\beta}e^{-y_i/\beta}}{\beta^3}.$$

Lad $S_x = \sum_{i=1}^n X_i, \quad S_y = \sum_{i=1}^n Y_i$, så får vi log-likelihoodfunktionen

$$\ell_n(\beta) = 3n\log\beta + \frac{S_x}{\beta} + \frac{S_y}{\beta^2}$$

og videre scorefunktion ved differentiation

$$S_n(\beta) = \frac{3n}{\beta} - \frac{S_x}{\beta^2} - \frac{2S_y}{\beta^3}$$

og en gang til for at få informationsfunktionen

$$I_n(\beta) = -\frac{3n}{\beta^2} + \frac{2S_x}{\beta^3} + \frac{6S_y}{\beta^4}.$$

Spørgsmål 1.3

Scoreligningen $S_n(\beta) = 0$:

$$\frac{3n}{\beta} - \frac{S_x}{\beta^2} - \frac{2S_y}{\beta^3} = 0$$

omskrives til idet $\bar{x}_n = S_x/n$, $\bar{y}_n = S_y/n$

$$3\beta^2 - \bar{x}_n\beta - 2\bar{y}_n = 0$$

Med entydig løsning i området $\beta > 0$

$$\hat{\beta} = \frac{\bar{x}_n + \sqrt{\bar{x}_n^2 + 24\bar{y}_n}}{6}.$$

Idet

$$\ell_n(\beta) = 3n\log\beta + \frac{S_x}{\beta} + \frac{S_y}{\beta^2}$$

ser vi at $\ell_n(\beta) \to \infty$ for $\beta \to 0$ og $\beta \to \infty$, så det må være et minimum.

Spørgsmål 1.4

Vi har Fisherinformationen

$$\begin{aligned} i_n(\beta) &= \mathbf{E}_\beta\{I_n(\beta)\} \\ &= -\frac{3n}{\beta^2} + \frac{2\mathbf{E}_\beta\{S_x\}}{\beta^3} + \frac{6\mathbf{E}_\beta\{S_y\}}{\beta^4} \\ &= -\frac{3n}{\beta^2} + \frac{2n\beta}{\beta^3} + \frac{6n\beta^2}{\beta^4} = \frac{5n}{\beta^2} \end{aligned}$$

hvoraf vi slutter at MLE er asymptotisk normalfordelt

$$\hat{\beta}_n \sim N\left(\beta, \frac{\beta^2}{5n}\right).$$

Spørgsmål 1.5

(a): Vi at bruge et likelihood ratio test. Vi får

$$\begin{aligned} 2\ell(\hat{\beta}_{10}) - 2\ell(\hat{\theta}_{10}) &= 60\log(\hat{\beta}_{10}) + 20\frac{\bar{x}_{10}}{\hat{\beta}_{10}} + 20\frac{\bar{y}_{10}}{\hat{\beta}_{10}^2} \\ &\quad -20\log\bar{x}_{10} - 20\log\bar{y}_{10} - 20\frac{\bar{x}_{10}}{\bar{x}_{10}} - 20\frac{\bar{y}_{10}}{\bar{y}_{10}} \\ &= 60\log(\hat{\beta}_{10}) + 20\frac{\bar{x}_{10}}{\hat{\beta}_{10}} + 20\frac{\bar{y}_{10}}{\hat{\beta}_{10}^2} - 20\log\bar{x}_{10} - 20\log\bar{y}_{10} - 40 \\ &= 0.04. \end{aligned}$$

Denne skal vurderes i en χ^2-fordeling med $2-1=1$ frihedsgrader, hvilket giver en p-værdi på $p = 0.837$, så der er absolut ingen grund til at forkaste H_0.

(b): Denne gang vælger vi at bruge den ægte Wald størrelse for den simple hypotese. Idet $\hat{\beta}_{10} = 2.28$ fås

$$W_n = 5n(\hat{\beta}_n - 1)^2 = 50 \times 1.28^2 = 81.97$$

som skal vurderes i en χ^2 med 1 frihedsgrad, hvilket giver en p-værdi tæt på 0, så hypotesen kan ikke opretholdes.

Man kunne naturligvis også have brugt et LR test

$$\Lambda = 2(\ell(1) - \ell(\hat{\beta}_{10}) = 61.13$$

med samme resultat.

```
# data

xbar = 2.15
ybar = 5.349

# mle

hatbeta=(2.15+sqrt(xbar^2+24*ybar))/6

# LR sammensat hypotese

ell_0 =30*log(hatbeta)+10*xbar/hatbeta+10*ybar/(hatbeta^2)
ell_1 = 10*log(xbar) +10*log(ybar) +20
Lambda=2*(ell_0-ell_1)

# p vaerdi
1-pchisq(Lambda,1)
```

```
[1] 0.8375556
```

```
Lambda
```

```
[1] 0.04203367
```

```
# Wald
w=50*(hatbeta-1)^2
w
```

```
[1] 81.97329
```

```
# p vaerdi
1-pchisq(w,1)
```

```
[1] 0
```

```
# LR for simpel hypotese

ell_2=10*xbar+10*ybar
Lambda2= 2*(ell_2-ell_1)
Lambda2
```

```
[1] 61.13245
```

```
# p vaerdi
1-pchisq(Lambda2,1)
```

```
[1] 5.329071e-15
```

Spørgsmål 2.1

Da $\mathbf{E}_\theta(Z) = 0$ og X og Y er uafhængige, er

$$m(\theta) = \mathbf{E}_\theta(Z^2) = \mathbf{V}_\theta(X - Y) = \mathbf{V}_\theta(X) + \mathbf{V}_\theta(Y).$$

Da X og Y begge er poissonfordelte med middelværdi, og dermed varians, e^θ ser vi, at

$$m(\theta) = 2e^\theta.$$

Momentestimatoren er dermed givet som løsning til

$$2e^\theta = \frac{1}{n}\sum_{i=1}^{n} Z_i^2,$$

dvs.

$$\tilde{\theta}_n = \log\left(\frac{1}{2n}\sum_{i=1}^{n} Z_i^2\right),$$

som er veldefineret, hvis ikke alle Z_i-erne er 0.

Spørgsmål 2.2

Vi efterviser betingelserne for BMS, sætning 2.17. Da poissonfordelte variable har momenter af enhver orden, har $t(Z)$ specielt endelig varians, endvidere er momentfunktionen glat og injektiv, og

$$m'(\theta) = 2e^\theta \neq 0.$$

Heraf følger, at $\tilde{\theta}_n$ er konsistent og asymptotisk normalfordelt.

Vi finder nu

$$\mathbf{V}_\theta(t(Z)) = \mathbf{V}_\theta(Z^4) = \mathbf{E}_\theta(Z^4) - (\mathbf{E}_\theta(Z^2))^2 = 2e^\theta + 12e^{2\theta} - 4e^{2\theta} = 2e^\theta + 8e^{2\theta}.$$

BMS, sætning 2.17, giver den asymptotiske varians

$$\sigma^2(\theta) = \mathbf{V}(t(Z))/m'(\theta)^2 = \frac{1}{4}e^{-2\theta}(2e^\theta + 8e^{2\theta}) = 2 + \frac{1}{2}e^{-\theta}.$$

Med andre ord er

$$\tilde{\theta}_n \overset{\text{as}}{\sim} N(\theta, (2 + e^{-\theta}/2)/n).$$

Spørgsmål 2.3

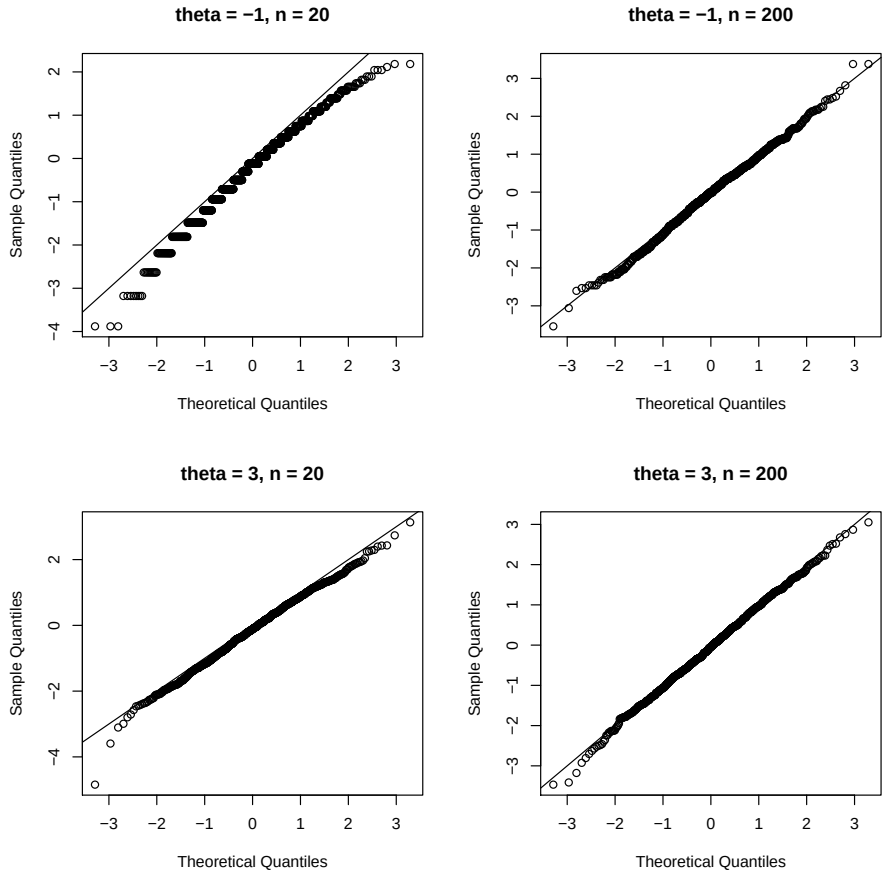
Det er mest interessant at undersøge den asymptotiske fordeling for negative værdier af θ og/eller små værdier af n, og eventuelt sammenholde med større værdier.

Her præsenteres resultaterne for fire kombinationer. De asymptotiske standardafvigelser (standard errors) beregnes endvidere for alle fire kombinationer.

```
set.seed(11)
B <- 1000
theta1 <- -1 ## Middelværdi exp(-1) = 0.37
n <- 20
se1 <- sqrt((2 + exp(-theta1) / 2) / n)
theta_tilde1 <- replicate(B, {
    x <- rpois(n, exp(theta1))
    y <- rpois(n, exp(theta1))
    log(mean((x - y)^2)/2)
})
theta2 <- 3 ## Middelværdi exp(3) = 20
se2 <- sqrt((2 + exp(-theta2) / 2) / n)
```

```r
theta_tilde2 <- replicate(B, {
    x <- rpois(n, exp(theta2))
    y <- rpois(n, exp(theta2))
    log(mean((x - y)^2)/2)
})
n <- 200
se3 <- sqrt((2 + exp(-theta1) / 2) / n)
theta_tilde3 <- replicate(B, {
    x <- rpois(n, exp(theta1))
    y <- rpois(n, exp(theta1))
    log(mean((x - y)^2)/2)
})
se4 <- sqrt((2 + exp(-theta2) / 2) / n)
theta_tilde4 <- replicate(B, {
    x <- rpois(n, exp(theta2))
    y <- rpois(n, exp(theta2))
    log(mean((x - y)^2)/2)
})
```

Vi sammenligner nu med den asymptotiske normalfordeling via qqplot.



Det er klart, at for $\theta = -1$ er estimatoren ikke normalfordelt for $n = 20$, mens normalfordelingen er en OK approksimation for $\theta = 3$, selv for $n = 20$. For $n = 200$ er normalfordelingen en god approksimation til fordelingen af estimatoren, selv for $\theta = -1$

Vi kan også sammenligne de asymptotiske standardafvigelser med de empiriske.

```
tibble(theta = c(theta1, theta2, theta1, theta2), n = c(20, 20, 200, 200),
             teo_se = c(se1, se2, se3, se4),
             emp_sd = c(sd(theta_tilde1), sd(theta_tilde2), sd(theta_tilde3), sd(theta_tilde4)))
```

```
# A tibble: 4 x 4
  theta     n teo_se emp_sd
  <dbl> <dbl>  <dbl>  <dbl>
1    -1    20  0.410  0.418
2     3    20  0.318  0.319
3    -1   200  0.130  0.133
4     3   200  0.101  0.101
```

Her ser vi en ret god overensstemmelse, selv for $\theta = -1$ og $n = 20$.

Spørgsmål 3.1

Vi indlæser data til opgaven.

```
gener <- read_csv("MatStat2020Juni_opg3.txt")
```

Opgavens første del løses ved krydstabulering af `C` og `S` Sålænge det er gjort korrekt, og argumenterne er rigtige, spiller det ingen rolle for bedømmelsen, hvordan det præcist er implementeret. Nedenfor følger en måde at løse opgaven på.

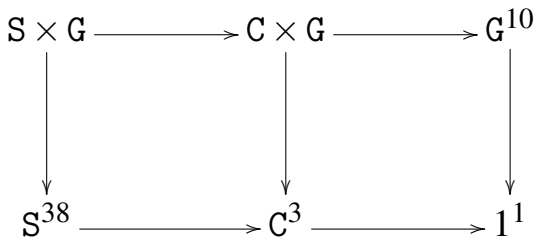
```
count(gener, C, S)
```

```
# A tibble: 38 x 3
   C     S                          n
   <chr> <chr>                  <int>
 1 bird  Aquila_chrysaetos          9
 2 bird  Cariama_cristata           6
 3 bird  Charadrius_vociferus       9
 4 bird  Eurypyga_helias            6
 5 bird  Gallus_gallus             10
 6 bird  Haliaeetus_leucocephalus  10
 7 bird  Pygoscelis_adeliae         9
 8 bird  Serinus_canaria           10
 9 fish  Ictalurus_punctatus        9
10 fish  Lepisosteus_oculatus       9
# ... with 28 more rows
```

I opgaven oplyses det, at der er 38 arter (hvilket i øvrigt også kan tjekkes ved tabulering), og da der ligeledes er 38 rækker i denne tabel følger det, at $\mathtt{C} \leq \mathtt{S}$. Enhver værdi af `S` må jo så forekomme netop en gang i tabellen, og bestemmer således værdien af `C`.

Det følger også af (den fulde version af) tabellen ovenfor, at `C` optræder på tre niveauer (bird, fish, mammal), så $\dim(L_{\mathtt{C}}) = 3$.

Da $\mathtt{C} \leq \mathtt{S}$ er endvidere $\mathtt{C} \times \mathtt{G} \leq \mathtt{S} \times \mathtt{G}$, og $\mathtt{G} \leq \mathtt{C} \times \mathtt{G}$, så vi får, jf. også det tilsvarende design i eksempel 14.16 i EH, faktorstrukturdiagrammet



Diagrammet er ovenfor annoteret med dimensioner vi kender på nuværende tidspunkt.

Spørgsmål 3.2

Da $\mathtt{S} \wedge \mathtt{C} \times \mathtt{G} = \mathtt{C}$ og $\mathtt{C} \leq \mathtt{S}$ er der kun to ikke-trivielle minima, der skal undersøges, nemlig $\mathtt{C} \wedge \mathtt{G}$ og $\mathtt{S} \wedge \mathtt{G}$, jf. også det tilsvarende design i eksempel 14.20 i EH.

Det gøres ligeledes ved krydstabulering, hvor det her nok er lettest bare at bruge `table`.

```
table(gener$C, gener$G)
```

```
        1CQ7M 1CS4Z 1CSGF 1CV66 1CW69 1CXQA 1D1QW 1D229 1DD42 1DFZX
  bird      8     3     8     8     6     8     6     6     8     8
  fish      7     7     7     7     6     7     7     5     0     7
  mammal   23    22    21    23    23    22    18    17    23    21
```

Af tabellen ovenfor fremgår det, at der kun er en enkelt kombination af `C` og `G` (fish og 1DD42), som ikke forekommer. Designgrafen indeholder derfor alle kanter pånær denne ene, og er oplagt sammenhængende, hvorfor $\mathtt{C} \wedge \mathtt{G} = 1$.

Bemærk i øvrigt at tabellen viser, at produktfaktoren $\mathtt{C} \times \mathtt{G}$ forekommer på 29 niveauer, så $\dim(L_{\mathtt{C}\times\mathtt{G}}) = 29$.

```
table(gener$S, gener$G)
```

```
                        1CQ7M 1CS4Z 1CSGF 1CV66 1CW69 1CXQA 1D1QW 1D229 1DD42 1DFZX
  Acinonyx_jubatus          1     0     1     1     1     1     1     1     1     1
  Aotus_nancymaae           1     1     1     1     1     1     0     1     1     1
  Aquila_chrysaetos         1     0     1     1     1     1     1     1     1     1
  Callithrix_jacchus        1     1     1     1     1     1     1     0     1     1
  Cariama_cristata          1     0     1     1     0     1     0     0     1     1
  Cebus_capucinus           1     1     1     1     1     1     1     0     1     1
  Ceratotherium_simum       1     1     1     1     1     1     1     1     1     1
  Charadrius_vociferus      1     0     1     1     1     1     1     1     1     1
  Chinchilla_lanigera       1     1     1     1     1     1     1     1     1     1
  Chrysochloris_asiatica    1     1     1     1     1     1     0     1     1     1
  Elephantulus_edwardii     1     1     1     1     1     1     0     1     1     1
  Eptesicus_fuscus          1     1     1     1     1     1     1     1     1     1
  Eurypyga_helias           1     0     1     1     0     1     0     0     1     1
  Gallus_gallus             1     1     1     1     1     1     1     1     1     1
  Haliaeetus_leucocephalus  1     1     1     1     1     1     1     1     1     1
  Ictalurus_punctatus       1     1     1     1     1     1     1     1     0     1
  Lepisosteus_oculatus      1     1     1     1     1     1     1     1     0     1
  Leptonychotes_weddellii   1     1     1     1     1     1     1     0     1     0
  Lipotes_vexillifer        1     1     1     1     1     1     1     0     1     1
  Loxodonta_africana        1     1     1     1     1     1     1     1     1     1
```

```
  Manis_javanica              1     1     1     1     1     1     0     0     1     1
  Myotis_lucifugus            1     1     1     1     1     1     1     1     1     0
  Nannospalax_galili          1     1     1     1     1     1     1     1     1     1
  Neolamprologus_brichardi    1     1     1     1     1     1     1     0     0     1
  Octodon_degus               1     1     1     1     1     1     1     1     1     1
  Orycteropus_afer            1     1     1     1     1     1     0     1     1     1
  Pan_troglodytes             1     1     0     1     1     1     1     1     1     1
  Papio_anubis                1     1     0     1     1     1     1     1     1     1
  Poecilia_reticulata         1     1     1     1     0     1     1     1     0     1
  Pteropus_alecto             1     1     1     1     1     1     1     1     1     1
  Pygoscelis_adeliae          1     0     1     1     1     1     1     1     1     1
  Saimiri_boliviensis         1     1     1     1     1     1     1     1     1     1
  Sarcophilus_harrisii        1     1     1     1     1     0     1     0     1     1
  Scleropages_formosus        1     1     1     1     1     1     1     1     0     1
  Serinus_canaria             1     1     1     1     1     1     1     1     1     1
  Sinocyclocheilus_grahami    1     1     1     1     1     1     1     1     0     1
  Sorex_araneus               1     1     1     1     1     1     1     1     1     1
  Xiphophorus_maculatus       1     1     1     1     1     1     1     0     0     1
```

Tabellen ovenfor viser, at der er visse art-gen kombinationer, der ikke forekommer, men f.eks. forekommer genet 1CQ7M sammen med alle arter, og da alle gener forekommer i kombination med mindst en art er der altid en vej i designgrafen mellem to knuder via 1CQ7M. Designgrafen er således sammenhængende, og $\mathtt{S} \wedge \mathtt{G} = 1$.

Designet er **ikke** ortogonalt, f.eks. fordi tabellen for $\mathtt{C} \times \mathtt{G}$ indeholder et 0, jf. lemma 13.11 i EH.

Vi kan nu endelig finde dimensionerne af sum-rummene ved at bruge formel (13.1) i EH sammen med lemma 14.6:

Da $L_{\mathtt{S}\wedge\mathtt{G}} = L_1$ er

$$\dim(L_\mathtt{S} + L_\mathtt{G}) = \dim(L_\mathtt{S}) + \dim(L_1) - \dim(L_{\mathtt{S}\wedge\mathtt{G}}) = 38 + 10 - 1 = 47$$

og da $L_{\mathtt{S}\wedge\mathtt{C}\times\mathtt{G}} = L_\mathtt{C}$ er

$$\dim(L_\mathtt{S} + L_{\mathtt{C}\times\mathtt{G}}) = \dim(L_\mathtt{S}) + \dim(L_{\mathtt{C}\times\mathtt{G}}) - \dim(L_\mathtt{C}) = 38 + 29 - 3 = 64$$

Bemærk at da designet ikke er ortogonalt, kan vi principielt ikke benytte teknikken baseret på sætning 14.21 til at finde dimensionerne af sum-rummene ovenfor, selvom det vil give de rigtige dimensioner i dette tilfælde.

Bonus: Minimum af S og C × G

Man kunne lave et tilsvarende argument som ovenfor via tabulering for at vise at $\mathtt{S} \wedge \mathtt{C} \times \mathtt{G} = \mathtt{C}$, men tabellen bliver uoverskuelig. I stedet kan vi finde minimum ved at finde sammenhængskomponenterne på følgende måde (som gennemgået ved forelæsningerne).

```
library(igraph)
tab <- count(gener, S, C, G)
g <- mutate(tab, CG  = paste(C, G, sep = "_")) %>%
    select(S, CG) %>%
    as.matrix() %>%
    graph_from_edgelist(directed = FALSE)
tab <- mutate(tab, min.S.CG = components(g)$membership[as.character(S)])
count(tab, C, min.S.CG)
```

```
# A tibble: 3 x 3
  C      min.S.CG     n
```

```
  <chr>     <dbl> <int>
1 bird          2    69
2 fish          3    60
3 mammal        1   213
```

Denne tabel viser, at minimum er identisk med `C`.

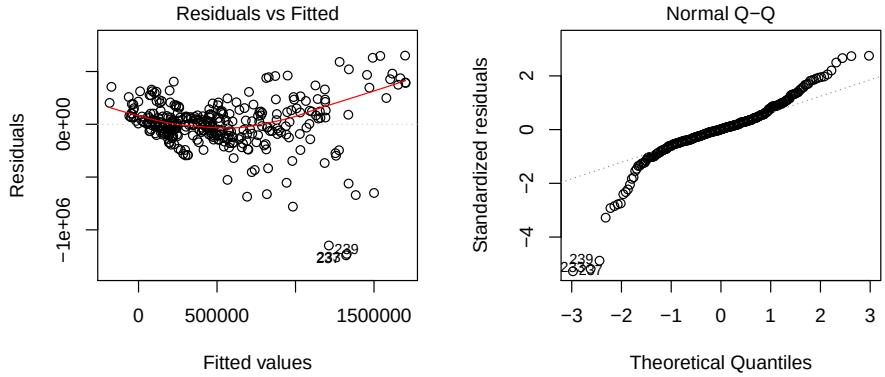
Spørgsmål 3.3

Vi fitter de to modeller i R.

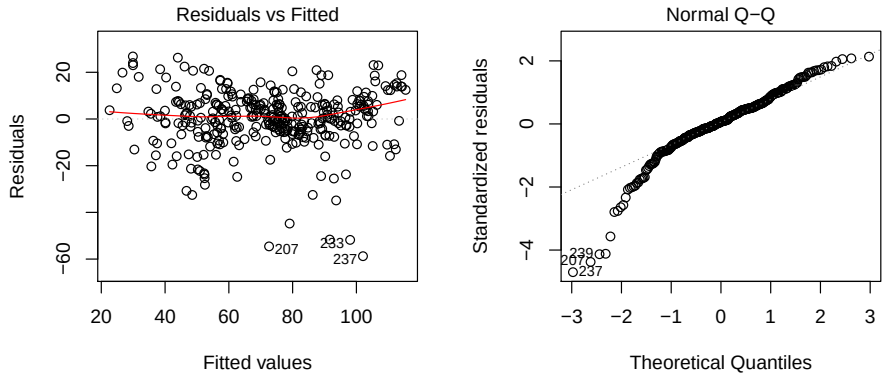
```
gen_lm <- lm(L ~ S + C * G, data = gener)
root_gen_lm <- lm(L^(1/3) ~ S + C * G, data = gener)
```

Dernæst ser vi på residualplot og qqplot for residualerne for de to modeller.

```
plot(gen_lm, 1:2)
```




```
plot(root_gen_lm, 1:2)
```



Det er klart, at modellen for `L` ikke fitter data særligt godt. Den krumme tragtform af residualplottet viser, at middelværdien ser misspecificeret ud, og at variansen ikke er konstant.

Modellen for $\sqrt[3]{\mathtt{L}}$ fitter bedre. Der er hverken nogen åbenlys misspecifikation af middelværdien, og variansen ser nogenlunde konstant ud på residualplottet. QQplottet viser dog, at residualerne ikke helt følger en normalfordeling, specielt ikke i den venstre hale. Og vi kan også identificere nogle ekstreme observationer.

Spørgsmål 3.4

Da $\mathtt{G} \leq \mathtt{C} \times \mathtt{G}$ er $L_{\mathtt{G}} \subseteq L_{\mathtt{C}\times\mathtt{G}}$, og deraf følger, at

$$L_{\mathtt{S}} + L_{\mathtt{G}} \subseteq L_{\mathtt{S}} + L_{\mathtt{C}\times\mathtt{G}}.$$

Den additive hypotese $\mathtt{S}+\mathtt{G}$ er således en hypotese i modellen specificeret ved $\mathtt{S}+\mathtt{C}\times\mathtt{G}$.

Baseret på resultaterne i spørgsmål 3.4 vælger vi at teste den additive hypotese i modellen for $\sqrt[3]{\mathtt{L}}$ ved hjælp af et F-test. Bemærk at de 17 frihedsgrader i testet netop er dimensionsfaldet på $64-47=17$, som kan beregnes på basis af spørgsmål 3.2.

```
add_root_gen_lm <- lm(L^(1/3) ~ S + G, data = gener)
anova(add_root_gen_lm, root_gen_lm)
```

```
Analysis of Variance Table

Model 1: L^(1/3) ~ S + G
Model 2: L^(1/3) ~ S + C * G
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1    295 72922
2    278 51747 17     21175 6.6916 2.227e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Her ser vi at p-værdien er meget lille (2.2×10^{-13}), så vi afviser den additive hypotese.

Havde vi udført testet i modellen for L, havde vi fået et tilsvarende resultat.

```
Analysis of Variance Table

Model 1: L ~ S + G
Model 2: L ~ S + C * G
  Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
1    295 2.5793e+13
2    278 1.8340e+13 17 7.4533e+12 6.6457 2.837e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Spørgsmål 4.1

Modellen er som i Eksempel 5.10 men middelværdi 0 men uden restriktioner på variansmatricen, og derfor er $\hat{\Sigma} = \frac{1}{n}\mathrm{S}$. Hypotesen H_0 er hypotesen om at Σ er diagonal, og det er den netop hvis Σ^{-1} er diagonal. Mængden af diagonalmatricer, $M_0 \subseteq \mathrm{Sym}_3$ udgør et underrum af dimension 3, så hypotesen er en lineær hypotese i den kanoniske parameter. Ortogonalprojektionen, q_0, på M_0 består i at sætte ikke-diagonalindgangene til 0. Det følger igen af Eksemple 5.10 at likelihoodligningen er

$$\begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} = q_0(\Sigma) = \frac{1}{n} q_0(S) = \frac{1}{n}\begin{pmatrix} \mathrm{S}_{11} & 0 & 0 \\ 0 & \mathrm{S}_{22} & 0 \\ 0 & 0 & \mathrm{S}_{33} \end{pmatrix},$$

hvoraf det følger, at $\hat{\sigma}_1^2 = \frac{1}{n}\mathrm{S}_{11}, \hat{\sigma}_2^2 = \frac{1}{n}\mathrm{S}_{22}, \hat{\sigma}_3^2 = \frac{1}{n}\mathrm{S}_{33}$.

Fra korollar 5.9, formel (14), følger det, at $\log Q = \frac{n}{2}(\log\det(\mathrm{S}) - \log\det(n\hat{\Sigma}_0)) = \frac{n}{2}(\log\det(\mathrm{S}) - \log(\mathrm{S}_{11}\mathrm{S}_{22}\mathrm{S}_{33}))$, eller

$$Q = \left(\frac{\det(\mathrm{S})}{\mathrm{S}_{11}\mathrm{S}_{22}\mathrm{S}_{33}}\right)^{n/2}.$$

Spørgsmål 4.2

Da Sym_3 har dimension 6 og hypotesen er en lineær hypotese af dimension 3 følger det af Wilks sætning (BMS, sætning 4.5) at $-2\log Q$ er asymptotisk χ^2-fordelt med $6-3=3$ frihedsgrader. Vi beregner p-værdien

```
z <- 30 * (sum(log(diag(S)))- determinant(S)$modulus)
c(test = z, pvalue = pchisq(z, 3, lower.tail = FALSE))
```

```
       test      pvalue
15.17512045  0.00167295
```

Da p-værdien er relativt lille afvises hypotesen H_0 om at variansmatricen er diagonal.