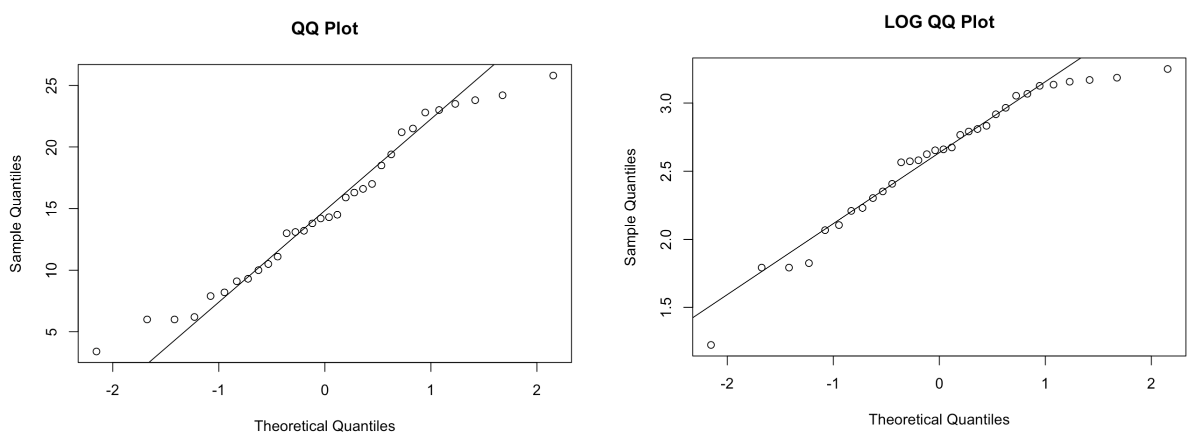
Del 1: Lineære normale modeller

Spørgsmål 1.

Vi arbejder her med de to faktorer køn og dosisgruppe, hvor køn er en blokfaktor, da det ikke nødvendigvis er muligt at lade han og hun rotter høre sammen under et, og hvor dosisgruppen er en interessefaktor. Altså er det oplagt at se på en tosidet variansanalysemodel. Under forsøget undersøges der desuden om kønnet vil have indflydesle på vækstraten og derfor er det oplagt at have en model med vekselvirkning.
For at undersøge om hvorvidt `growth_rate` skal transformeres med logaritmen vil vi se på QQ-plotene

```
qqnorm(growthdata$growth_rate, main = "QQ Plot")
qqline(growthdata$growth_rate)
qqnorm(log(growthdata$growth_rate),  main = "LOG QQ Plot")
qqline(log(growthdata$growth_rate))
```



Vi startede med at undersøge det første QQ-plot og ser at det, ud over enderne, følger en ret linje tilnærmelsesværdigt. Vi har da valgt at lave et QQ-plot med log-transformationen og ser at dette QQ-plot følger en ret linje i lige så stor grad som den forrige og altså er det ikke nødvendigt at log-transformere.

Spørgsmål 2.

Lad $\mathbb{G} = \{1, \texttt{koen}, \texttt{dosisgruppe}, \texttt{koen} \times \texttt{dosisgruppe}\}$. For at $\mathbb{G}$ opfylder betingelserne for EH sætning 14.21 skal $\mathbb{G}$ være et ortogonaltdesign og systemet skal være $\wedge$-stabilt.

For at $\mathbb{G}$ er et ortogonalt design, skal faktorerene i systemet være parvist geometrisk ortogonale.
Vi ser først på produktet $\texttt{koen} \times \texttt{dosis}$ og ser at denne produktfaktor er surjektiv, da vi kan finde alle kombinationer, hvorefter vi får fra eksempel 13.8 i EH at grafen er sammenhængene og altså må designet pr. definition være sammenhængene.
Desuden ser vi at alle $\texttt{koen} \times \texttt{dosisgruppe}$-grupper vil have præcis 4 elementer og pr. eksemplerne 13.12 og 13.13 i EH vil vi da have at designet er balanceret og at produktfaktoren dermed også opfylder balanceligningen hvilket giver at de to faktorer `koen` og `dosisgruppe` er geometrisk ortogonale.

Vi ønsker nu at vise at systemet er $\wedge$-stabilt og vi husker at vi i øvelse 14.1 har vist at hvis $T \times F$ er surjektiv, så vil $T \wedge F = 1$ og eftersom vi ovenfor har at $\texttt{koen} \times \texttt{dosisgruppe}$ er surjektiv, må $\texttt{koen} \wedge \texttt{dosisgruppe} = 1$
Af dette kan vi da se at

$$1 = \texttt{koen} \wedge \texttt{dosisgruppe} \leq \texttt{koen} \leq \texttt{koen} \times \texttt{dosisgruppe}$$

$$1 = \texttt{koen} \wedge \texttt{dosisgruppe} \leq \texttt{dosisgruppe} \leq \texttt{koen} \times \texttt{dosisgruppe}$$

Hvorefter vi får fra lemma 14.11 at faktorerne må være parvist ortogonale og at $\mathbb{G}$ dermed også må være et ortogonalt design. Men da vi jo ved at alle faktorerne er ordnede i forhold til hinanden har vi at minimumsdannelser af disse netop må være lig med den groveste faktor og at alle disse minimumsdannelser må ligge i $\mathbb{G}$. Altså må $\mathbb{G}$ være $\wedge$-stabil og betingelserne for sætning 14.21 er da opfyldt som ønsket.

For designet får vi følgende faktorstrukturdiagram

$$\begin{tikzcd}
\text{koen} \times \text{dosis} \arrow[r] \arrow[d] & \text{dosis} \arrow[d] \\
\text{koen} \arrow[r] & 1
\end{tikzcd}$$

Vi tester nu om den additive model giver en rimelig beskrivelse af data. Dette gøres ved at sammenligne den additive model med vekselvirkningsmodellen:

```
x_add <- lm(growth_rate ~ dosisgruppe + koen, data = growthdata)
x_mult <- lm(growth_rate ~ dosisgruppe*koen, data = growthdata)
anova(x_add, x_mult)
```

```
Analysis of Variance Table

Model 1: growth_rate ~ dosisgruppe + koen
Model 2: growth_rate ~ dosisgruppe * koen
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     27 121.95
2     24 118.55  3    3.4034 0.2297 0.8748
```

Vi får en p-værdi på 0.8748, og der er derved ikke evidens for at afvise hypotesen om, at den additive model beskriver data.

Spørgsmål 3.

Et estimat for den forventede vækstrate for en hun-rotte i dosisgruppe III finder vi ved

```
x_int <- lm(growth_rate ~ dosisgruppe + koen - 1, data = growthdata)
summary(x_int)
```

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
dosisgruppeI     7.7344     0.8401   9.207 8.11e-10
dosisgruppeII   12.6594     0.8401  15.069 1.15e-14
dosisgruppeIII  16.6969     0.8401  19.875  < 2e-16
dosisgruppeIV   23.6094     0.8401  28.104  < 2e-16
koenMale        -0.7688     0.7514  -1.023    0.315
```

Den forventede vækstrate for en hun-rotte i dosisgruppe III er altså 16.6969.

95%-konfidensintervallet for den forventede vækstrate for en hun-rotte i dosisgruppe III, finder vi ved

```
> confint(x_int)
                    2.5 %     97.5 %
dosisgruppeI     6.010673  9.4580770
dosisgruppeII   10.935673 14.3830770
dosisgruppeIII  14.973173 18.4205770
dosisgruppeIV   21.885673 25.3330770
koenMale        -2.310476  0.7729759
```

95%-konfidensintervallet er altså (14.9732, 18.4206)

Spørgsmål 4.

Vi tester nu hypotesen om at vækstraten er en lineær funktion af logaritmen til dosis. På baggrund af vores test i spørgsmål 2. vælger vi den additive model som udgangsmodel.

```
x_add <- lm(growth_rate ~ dosisgruppe + koen - 1, data = growthdata)
x_log <- lm(growth_rate ~ log(dosis), data = growthdata)
anova(x_add, x_log)

x_add$rank
x_log$rank
```

```
Analysis of Variance Table

Model 1: growth_rate ~ dosisgruppe + koen - 1
Model 2: growth_rate ~ log(dosis)
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     27 121.95
2     30 140.24 -3   -18.291 1.3499 0.2791
> x_add$rank
[1] 5
> x_log$rank
[1] 2
```

Vi ser den additive model har dimension 5, mens den log transformerede model har dimension 3. Der er altså et dimensionsfald på $5 - 2 = 3$. Da datasættet består af 32 observationer, får vi fordelingen af F-teststørrelsen har en F-fordeling med $(5 - 2, 32 - 5) = (3, 27)$ frihedsgrader. Vi ser at F-teststørrelsen er 1.3499 og vi får en p-værdi på 0.2791. Udfra F-teststørrelsen og p-værdien er der ikke evidens for at afvise hypotesen om at, vækstraten er en lineær funktion af logaritmen til dosis.

Spørgsmål 5.

Vi skal nu angive estimater for MLE under hypotesen om at vækstraten er en lineær funktion af logaritmen til dosis. Vi finder estimatorene ved

```
x_log <- lm(growth_rate ~ log(dosis), data = growthdata)
summary(x_log)
summary(x_log)$sigma^2
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.2118     1.0377   0.204     0.84
log(dosis)    7.4533     0.4932  15.112 1.44e-15 ***
```

```
> summary(x_log)$sigma^2
[1] 4.674721
```

Altså at den lineære funktion med hældning α og skærring β har MLE af middelværdien for α og β givet ved

$$\hat{\alpha} = 7.4533 \quad \text{og} \quad \hat{\beta} = 0.2118.$$

Da $\hat{\alpha}$ er positiv er dette et udtryk for, at vækstraten vokser, når dosis øges.

MLE af variansen er

$$\hat{\sigma}^2 = 4.6747.$$

Vi har nu fundet MLE $(\hat{\xi}, \hat{\sigma}^2)$, hvor $\hat{\xi} = (\hat{\beta},\ \hat{\alpha})^T$

Per EH korollar 10.21 er $\hat{\xi} \sim \mathcal{N}(\xi,\ \sigma^2(A^TA)^{-1})$ og $\hat{\sigma}^2$ er $\mathcal{X}^2$-fordelt med $32 - 2 = 30$ frihedgrader og skalaparameter $\frac{\sigma^2}{32}$.

Vi finder variansen for $\hat{\xi}$ til at være

```
> A <- model.matrix(x_log)
> solve(t(A) %*% A)
            (Intercept)  log(dosis)
(Intercept)   0.2303320 -0.10177954
log(dosis)   -0.1017795  0.05203422
```

$$V(\hat{\xi}) = \sigma^2 \begin{pmatrix} 0.2303 & -0.1018 \\ -0.1018 & 0.0520 \end{pmatrix}.$$

Vi har nu angivet estimatorer for MLE samt deres fordeling.

Spørgsmål 6.

Vi skal nu angive et 95%-prædikationsinterval for vækstraten for hunrotter i dosisgruppe III. Dette finder vi ved

```
x_add <- lm(growth_rate ~ dosisgruppe + koen, data = growthdata)
predict(x_add, newdata = data.frame(koen = "Female", dosisgruppe = "III"), interval = "prediction")
```

```
       fit     lwr      upr
1 16.69688 12.0079 21.38585
```

95%-prædikationsintervallet er altså $(12.0079, \quad 21.38585)$.

Del 2: Eksponentielle familier

Lad $(X_1, Y_1, \ldots, Y_n, X_n)$ være indbyrdes uafhængige og eksponentialfordelte stokastiske variable hvor X_j og Y_j har tætheder med $(\theta, \lambda) \in \mathbb{R}^2_+$.

$$f_j(x_j; \theta) = j \cdot \theta \cdot e^{-j\theta x_j}, \quad x_j > 0,$$

$$g_j(y_j; \lambda) = \lambda \cdot e^{-\lambda y_j}, \quad y_j > 0.$$

Spørgsmål 1.

Vi finder først den simultane tæthed for $(X, Y) := (X_1, Y_1, \ldots, X_n, Y_n)$

$$f(X,Y) = \prod_{j=1}^{n} f_j(x_j;\theta) \cdot g_j(y_j;\lambda) = \prod_{j=1}^{n} j \cdot \theta \cdot e^{-j\theta x_j} \cdot \lambda \cdot e^{-\lambda y_j} = n! \cdot \theta^n \lambda^n \prod_{j=1}^{n} e^{-j\theta x_j - \lambda y_j}$$

$$= n! \cdot \theta^n \lambda^n e^{\sum_{j=1}^n -j\theta x_j - \lambda y_j} = n! \cdot \theta^n \lambda^n e^{-\theta \sum_{j=1}^n j \cdot x_j - \lambda \sum_{j=1}^n j y_j}.$$

Vi genkender dette som tætheden for en eksponentiel familie med:
Lebesguemålet $m^{\otimes 2n}$ på $\mathbb{R}_+^2$ som grundmål, og normaliseringskonstant:

$$c(\theta, \lambda) = \lambda^{-n} \cdot \theta^{-n} \cdot n!^{-1},$$

og kanonisk stiksprøvefunktion

$$t(x_1, y_1, ..., x_n, y_n) = \left(-\sum_{j=1}^n j \cdot x_j, \quad -\sum_{j=1}^n y_j \right)^T.$$

Vi får da, at kumulantfunktionen er

$$\psi(\theta, \lambda) = \log c(\theta, \lambda) = \log \left(\lambda^{-n} \cdot \theta^{-n} \cdot n!^{-1} \right) = -n \log(\lambda) - n \log(\theta) - \log(n!).$$

Da parameterrummet $\mathbb{R}_+^2$ er en åben og konveks delmængde af $\mathbb{R}^2$ er familien regulær. Da for $\alpha \in \mathbb{R}_+^2$ er $\alpha^T t(x)$ kun konstant for $\alpha = 0$ og dermed er familien minimalt repræsenteret.

Spørgsmål 2.

Maksimaliseringsestimatoren $(\hat{\theta}, \hat{\lambda})$ findes ved scoreligningen $S_{X,Y}(\theta, \lambda) = 0$. Per korollar 2.21 er scorefunktionen givet ved

$$S_{X,Y}(\theta, \lambda) = t(x, y) - \tau(\theta, \lambda),$$

hvor $\tau(\theta, \lambda) = \nabla \psi(\theta, \lambda) = \left(-\frac{n}{\theta}, \quad -\frac{n}{\lambda} \right)^T$. Vi får da scoreligningen

$$\begin{pmatrix} \sum_{j=1}^n j \cdot x_j \\ \sum_{j=1}^n y_j \end{pmatrix} = \begin{pmatrix} \frac{n}{\theta} \\ \frac{n}{\lambda} \end{pmatrix}.$$

Vi får da at maksimaliseringsestimatoren er givet ved

$$\underline{\hat{\theta}} \equiv \begin{pmatrix} \hat{\theta} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} \frac{n}{\sum_{j=1}^n j \cdot x_j} \\ \frac{n}{\sum_{j=1}^n y_j} \end{pmatrix}.$$

Spørgsmål 3.

Lad $\phi : \mathbb{R}_{>0} \to \mathbb{R}_{>0}^2$ være defineret ved $\phi(\beta) = \left(\beta, \beta^{-1}\right)^T$. Bemærk at vi fra delspørgsmål 2.1 har at

$$\mathcal{Q} := \left\{ (X, Y)(P_\theta) \,|\, \theta \in \Theta \equiv \mathbb{R}_{>0}^2 \right\}$$

er en minimal og regulær todimensionel ekspotentiel familie med kanonisk parameterrum Θ, og at

$$\mathcal{Q}_{kr} := \left\{ (X, Y)(P_{\phi(\beta)}) \,|\, \beta \in \mathbb{R}_{>0} \right\} \subseteq \mathcal{Q}.$$

Hertil kan nævnes at

1. $\mathbb{R}_{>0} \subseteq \mathbb{R}$ åben
2. ϕ er kendt glat.
3. Jakobimatricen $J(\beta) = D\phi(\beta) = \frac{\partial}{\partial \beta} \phi(\beta) = \left(1, -\beta^{-2}\right)^T$ har fuld rang 1 $\forall \beta \in \mathbb{R}_{>0}$.
4. $\phi : \mathbb{R}_{>0} \to \mathbb{R}_{>0}^2$ er kendt kontinuert og injektiv. Da ϕ en $2-$involution, får vi dermed at ϕ er en homeomorfi på dets billedet.

Vi kan da erklære $\mathcal{Q}_{kr}$ en krum ekspotentiel familie af dimension 1 og orden 2.

Spørgsmål 4.

Fra delopgave 2.2 haves et udtryk for τ således at vi ved Proposition 2.30, BMS har at scorefunktionen i den krumme familie er at finde ved

$$\begin{aligned}
S_{x,y}(\beta) &= \left(t(x,y) - \tau(\phi(\beta))\right)^T D\phi(\beta) \\
&\doteq \left(\begin{pmatrix} -\sum_{j=1}^n jx_j \\ -\sum_{j=1}^n y_j \end{pmatrix} - \begin{pmatrix} -\frac{n}{\beta} \\ -n\beta \end{pmatrix} \right)^T \begin{pmatrix} 1 \\ -\beta^{-2} \end{pmatrix} \\
&\equiv \left(-\sum_{j=1}^n jx_j + \frac{n}{\beta}, -\sum_{j=1}^n y_j + n\beta \right) \begin{pmatrix} 1 \\ -\beta^{-2} \end{pmatrix} \\
&= -\sum_{j=1}^n jx_j + \frac{n}{\beta} + \frac{1}{\beta^2}\sum_{j=1}^n y_j - \frac{n}{\beta} \\
&= -\sum_{j=1}^n jx_j + \frac{1}{\beta^2}\sum_{j=1}^n y_j,
\end{aligned}$$

således at vi fra scoreligningen kan udregne;

$$\frac{1}{\beta^2}\sum_{j=1}^n y_j - \sum_{j=1}^n jx_j = 0$$
$$\Leftrightarrow$$
$$\sum_{j=1}^n y_j = \beta^2 \sum_{j=1}^n jx_j$$
$$\Leftrightarrow$$
$$\beta^2 = \frac{\sum_{j=1}^n y_j}{\sum_{j=1}^n jx_j}$$
$$\overset{\beta>0}{\Leftrightarrow}$$
$$\beta = \sqrt{\frac{\sum_{j=1}^n y_j}{\sum_{j=1}^n jx_j}},$$

hvormed vi får

$$\hat{\beta} = \sqrt{\frac{\sum_{j=1}^n y_j}{\sum_{j=1}^n jx_j}},$$

som MLE.

Spørgsmål 5.

Proposition 2.30, BMS giver at

$$i_n(\beta) = D\phi(\beta)^T \kappa(\phi(\beta)) D\phi(\beta) \equiv D\phi(\beta)^T D^2\psi(\phi(\beta)) D\phi(\beta),$$

således at idet

$$\kappa(\theta,\lambda) = D\tau(\theta,\lambda) \doteq D\left(-\frac{n}{\theta}, -\frac{n}{\lambda}\right) = \begin{pmatrix} \frac{n}{\theta^2} & 0 \\ 0 & \frac{n}{\lambda^2} \end{pmatrix},$$

hvormed

$$\kappa(\phi(\beta)) = \kappa(\beta,\beta^{-1}) \doteq \begin{pmatrix} \frac{n}{\beta^2} & 0 \\ 0 & \frac{n}{\left(\frac{1}{\beta}\right)^2} \end{pmatrix} \equiv \begin{pmatrix} \frac{n}{\beta^2} & 0 \\ 0 & n\beta^2 \end{pmatrix},$$

så

$$i_n(\beta) \doteq \left(1, -\beta^{-2}\right) \begin{pmatrix} \frac{n}{\beta^2} & 0 \\ 0 & n\beta^2 \end{pmatrix} \begin{pmatrix} 1 \\ -\beta^{-2} \end{pmatrix} = 2\frac{n}{\beta^2}.$$

Spørgsmål 6.

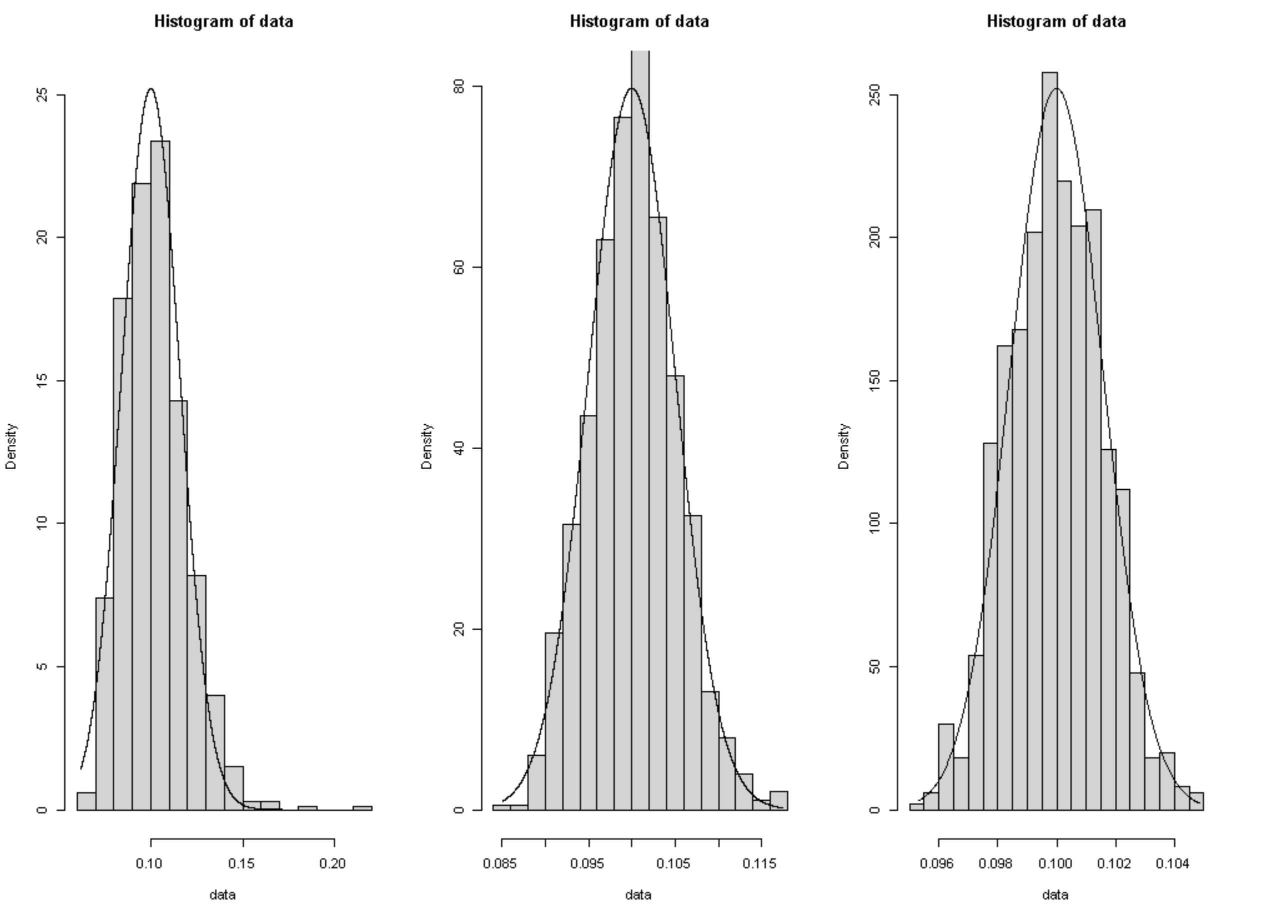
Bemærk at vi har regnet de relevante størrelser i tidligere opgaver. Vi laver vores simulation med seed 314, og $N = 10^3$ iterationer af $n = (20, 200, 2000)$ udtræk af $X_j \sim Exp(j\beta)$, $Y_j \sim Exp(\frac{1}{\beta})$, for $\beta = 0.1$ (arbitrært valg).
R koden kan findes i appendix. Specielt er det værd at bemærke at vi for $n = 20, 200, 2000$ får at

$$|E(data) - \beta| = 1.082893 \cdot 10^{-3},\ 2.150822 \cdot 10^{-4},\ 1.961224 \cdot 10^{-6},$$

og at

$$\left| V(data) - \frac{1}{i_n(\beta)} \right| = 6.061009 \cdot 10^{-6},\ 9.872272 \cdot 10^{-7},\ 3.517521 \cdot 10^{-8},$$

således at det umiddelbart virker ganske ganske muligt at vi ved Slutsky's sætning kunne få det ønskede. Vi laver også en visuel inspektion af overlap mellem en $\mathcal{N}(\beta, i_n(\beta))$ tæthed og vores data, for $n = 20, 200, 2000$:



som umiddelbart ser super fint ud. Vi konkludere at det ønskede virker ganske muligt, trods manglende identisk fordeling mellem X'erne og Y'erne.

Spørgsmål 7.

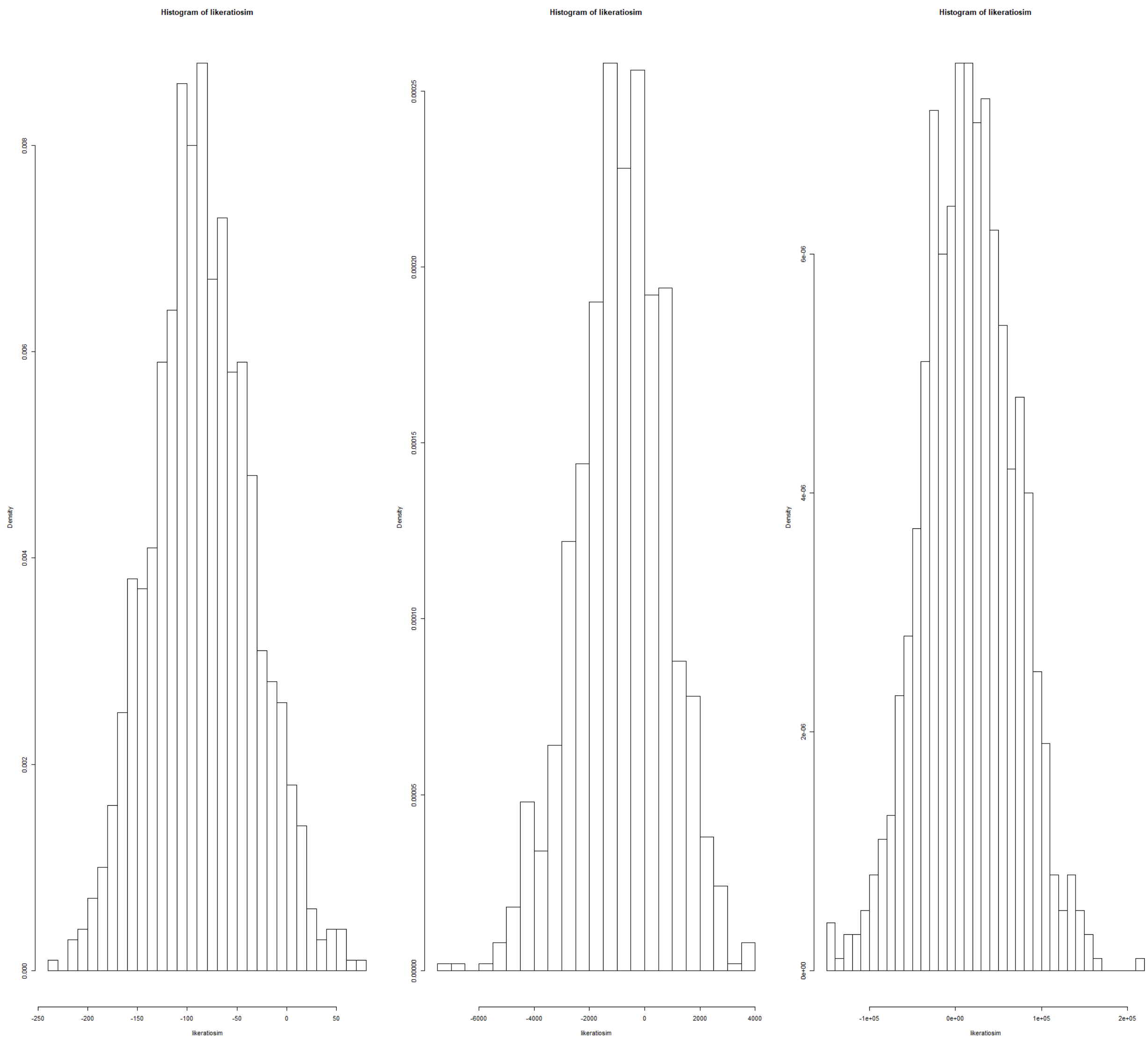
Vi bemærker at $H_0 : \theta\lambda = 1 \Leftrightarrow \lambda = \frac{1}{\theta}$ svarer til hypotesen om, at vi er på den overstående behandlede krumme ekspotentielle familie. Hertil kan bemærkes at vi både i den fulde model, og under hypotesen har fundet MLE'er, så vi kan dermed udregne at

$$\Lambda(x, y) = -2 \log \frac{\sup_{\beta \in B} L_{x,y}(\beta)}{\sup_{\underline{\theta} \in \mathbb{R}^2_{>0}} L_{x,y}(\underline{\theta})} = -2 \log \frac{L_{x,y}(\hat{\beta})}{L_{x,y(\hat{\underline{\theta}})}} = 2 \log L_{x,y}(\hat{\underline{\theta}}) - 2 \log L_{x,y}(\hat{\beta})$$

$$
\begin{aligned}
&\doteq 2\log\left(n!\hat{\theta}^n\hat{\lambda}^n e^{-\hat{\theta}\sum_{j=1}^n jx_j-\hat{\lambda}\sum_{j=1}^n jy_j}\right) - 2\log\left(n!\cdot\hat{\beta}^n\hat{\beta}^{-n}e^{-\hat{\beta}\sum_{j=1}^n jx_j-\hat{\beta}^{-1}\sum_{j=1}^n jy_j}\right)\\
&= 2\log n! + 2n\log\hat{\theta} + 2n\log\hat{\lambda} - 2\hat{\theta}\sum_{j=1}^n jx_j - 2\hat{\lambda}\sum_{j=1}^n jy_j - 2\log n! + 2\hat{\beta}\sum_{j=1}^n jx_j + 2\hat{\beta}^{-1}\sum_{j=1}^n jy_j\\
&= 2n\log\hat{\theta} + 2n\log\hat{\lambda} - 2\hat{\theta}\sum_{j=1}^n jx_j - 2\hat{\lambda}\sum_{j=1}^n jy_j + 2\hat{\beta}\sum_{j=1}^n jx_j + 2\hat{\beta}^{-1}\sum_{j=1}^n jy_j\\
&\doteq 2n\log\frac{n}{\sum_{j=1}^n jx_j} + 2n\log\frac{n}{\sum_{j=1}^n y_j} - 2\frac{n}{\sum_{j=1}^n jx_j}\sum_{j=1}^n jx_j - 2\frac{n}{\sum_{j=1}^n y_j}\sum_{j=1}^n jy_j + 2\hat{\beta}\sum_{j=1}^n jx_j + 2\hat{\beta}^{-1}\sum_{j=1}^n jy_j\\
&= 2n\log\frac{n}{\sum_{j=1}^n jx_j} + 2n\log\frac{n}{\sum_{j=1}^n y_j} - 2n - 2\frac{n}{\sum_{j=1}^n y_j}\sum_{j=1}^n jy_j + 2\hat{\beta}\sum_{j=1}^n jx_j + 2\hat{\beta}^{-1}\sum_{j=1}^n jy_j\\
&= 2n\log\frac{n}{\sum_{j=1}^n j\cdot x_j} + 2n\log\frac{n}{\sum_{j=1}^n y_j} - 2n - 2\frac{n}{\sum_{j=1}^n y_j}\sum_{j=1}^n jy_j + 2\sqrt{\frac{\sum_{j=1}^n y_j}{\sum_{j=1}^n jx_j}}\sum_{j=1}^n jx_j + 2\frac{1}{\sqrt{\frac{\sum_{j=1}^n y_j}{\sum_{j=1}^n jx_j}}}\sum_{j=1}^n jy_j\\
&= 2n\log\frac{n}{\sum_{j=1}^n j\cdot x_j} + 2n\log\frac{n}{\sum_{j=1}^n y_j} - 2n - 2\frac{n}{\sum_{j=1}^n y_j}\sum_{j=1}^n jy_j + 2\sqrt{\sum_{j=1}^n y_j}\sqrt{\sum_{j=1}^n jx_j} + 2\frac{\sqrt{\sum_{j=1}^n jx_j}}{\sqrt{\sum_{j=1}^n y_j}}\sum_{j=1}^n jy_j.
\end{aligned}
$$

Ved simulation baseret på R kode som det ses i appendix forekom for $n = 20, 200, 2000$ udtræk med $N = 10^3$ iterationer, seed 314 og $\beta = 0.1$ følgende tre plots.

Antager vi at vi overstående har regnet rigtigt på likelihood ratio static, og at simulationen er lavet korrekt, vil vi kunne afværve antagelser om at likelihood ratio static kan antages approksimativt χ^2-fordelt, idet simuleringerne, ud fra histogrammerne, lader til at have støtte som inkluderer negative dele af den reelle akse, hvilket går imod χ^2 fordelingerne, som kun har støtte på den positive halvakse.



Figur 1: Likelihood ratio test for hhv. $n = 20, 200, 2000$

Appendix: R kode

```
#Opgave 2.6
set.seed(314)

N <- 10^3
n <- c(20,200,2000)
beta <- 0.1

inbf <- function(n,beta) { #FisherInfoAfBeta
  2*n/(beta^2)
}

sim <- function (N,n,beta) { #Simulering
  MLE_n <- c() # Initialisering
  for (j in 1:N) {
```

```
    x <- c() #Initialisering
    y <- rexp(n = n, 1/beta)
    for (i in 1:n) {
      x[i] <- rexp(n=1, i*beta)
      x[i] <- i*x[i]
    }
    MLE_n[j] <- sqrt((sum(y))/(sum(x)))
  }
  MLE_n
}

meanvarsim <- function (N,n,beta) { #Middelv rdi og varians af data
    c(mean(sim(N,n,beta)), var(sim(N,n,beta)))
}

difsim <- function(N, n, beta) { #Differenser
  q <- meanvarsim(N, n, beta)
  difmean <- (abs(q[1]-beta))
  difvar <- abs(q[2]-1/(inbf(n,beta)))
  rbind(difmean, difvar)
}


meanvarsim(N=N, n=n[1], beta = beta)
meanvarsim(N=N, n=n[2], beta = beta)
meanvarsim(N=N, n=n[3], beta = beta)

difsim(N, n[1], beta) #(1.082893e-03, 6.061009e-06) <-> n = 20
difsim(N, n[2], beta) #(2.150822e-04, 9.872272e-07) <-> n = 200
difsim(N, n[3], beta) #(1.961224e-06, 3.517521e-08) <-> n = 2000

fMLE_nk <- function(n,beta,k,x) { #Taethed N(beta,fisher^(-1))
  dnorm(x, mean = beta, sd = sqrt(1/(inbf(n[k],beta))))
}

par(mfrow=c(1,3)) # Plotvindue til 3 plots

plottheplots <- function(N,n,beta,k,mea,vari) {
  data <- sim(N,n[k],beta)
  seqT <- seq(min(data),max(data), by = 1/(5*10^6*(max(data)-min(data))))
  f1T <- dnorm(seqT, mean=mea, sd=sqrt(vari))
  maxf1T <- max(f1T)+max(0.2,max(f1T)/100)
  hist(data, ylim = c(0,maxf1T), breaks = 15, prob = T)
  lines(seqT,f1T)
}

for (q in 1:3) {
  k=q
  plottheplots(N=N,n=n,beta=beta,q,mea=beta, vari=1/(inbf(n[q],beta)))
}

#Opgave 2.7
set.seed(314)

N <- 10^3
n <- c(20,200,2000)
beta <- 0.1

simJXmatYmatJYmat <- function (N,n,beta) {
```

```
  JXmat <- matrix(NA, nrow = n, ncol = N) #Setup
  Ymat <- matrix(NA, nrow = n, ncol = N)
  JYmat <- matrix(NA, nrow = n, ncol = N)
  for (j in 1:N) {
    jx <- c() #Setup
    jy <- c()
    y <- c()
    for (k in 1:n) { #Simulate jx, y and jy vectors
      #(notice that we use notation k to do this in R)
      y[k] <- rexp(n=1, 1/beta)
      x <- rexp(n=1, k*beta)
      jx[k] <- k*x
      jy[k] <- k*y[k]
    }
    JXmat[,j] <- jx #Having simulated n draws, we insert it into a matrix column
    Ymat[,j] <- y
    JYmat[,j] <- jy
  }
  #Sum the columns of each matrix (resulting in a vector of length N) that
  #we then row-bind to the other column sums, and output.
  rbind(apply((JXmat),2, sum), apply((Ymat),2, sum), apply((JYmat),2, sum))
}

simLikeRatio <- function(N,n,beta) {
  #Grabbing the output of simJXmatYmatJYmat
  data <- simJXmatYmatJYmat(N,n,beta)
  sumJXmat <- data[1,]
  sumYmat <- data[2,]
  sumJYmat <- data[3,]

  likeratio <- c()
  for (k in 1:N) {
    #Breaking down the six calculated terms of the likelihood ratio test static
    first <- 2*n*log((n)/(sumJXmat[k])) #First term ...
    second <- 2*n*log((n)/(sumYmat[k]))
    third <- -2*n
    fourth <- -2*(n/(sumYmat[k]))*sumJYmat[k]
    fifth <- 2*sqrt(sumYmat[k])*sqrt(sumJXmat[k])
    sixth <- 2*sqrt((sumJXmat)/(sumYmat))*sumJYmat
    likeratio[k] <- first+second+third+fourth+fifth+sixth #collecting the terms
    }
   likeratio #Outputting the resulting vector of N likelihood ratio statics
}
qq <- 1
likeratiosim <- simLikeRatio(N,n[qq],beta)
hist(likeratiosim, breaks=30, prob=T)
```