Musterlösung zur Übung 2

1. a) Unter H_0 ist $X - \theta Y = X - \theta_0 Y \sim \mathcal{N}(0, 1 + \theta_0^2)$ und folglich $\frac{X-\theta_0 Y}{\sqrt{1+\theta_0^2}} \sim \mathcal{N}(0,1)$. H_0 wird beibehalten, falls

$$\Phi^{-1}\big(\frac{\alpha}{2}\big) \leq \frac{X - \theta_0 Y}{\sqrt{1+\theta_0^2}} \leq \Phi^{-1}\big(1 - \frac{\alpha}{2}\big)$$

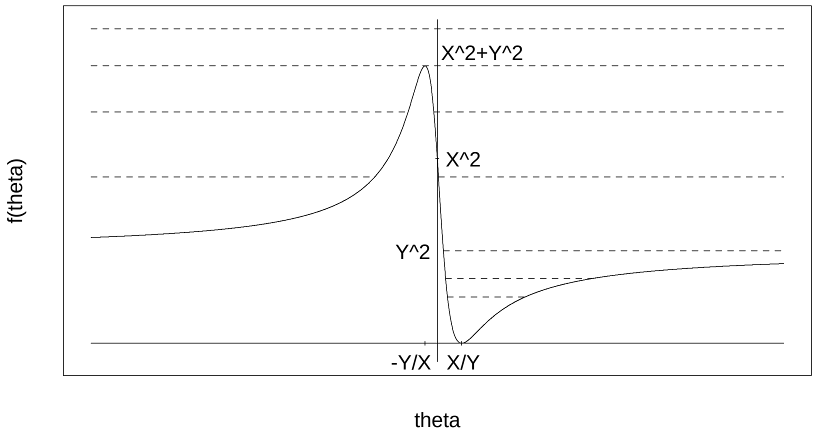
bzw. wenn

$$\Big|\frac{X - \theta_0 Y}{\sqrt{1+\theta_0^2}}\Big| \leq \Phi^{-1}\big(1 - \frac{\alpha}{2}\big) =: c_\alpha \,.$$

b) Der Konfidenzbereich (KB) enthält nach dem Dualitätslemma alle Nullhypothesen, welche vom obigen Test beibehalten werden:

$$KB = \Big\{\theta \,\Big|\, \Big|\frac{X - \theta Y}{\sqrt{1+\theta^2}}\Big| \leq c_\alpha\Big\} = \Big\{\theta \,\Big|\, \frac{(X-\theta Y)^2}{1+\theta^2} \leq c_\alpha^2\Big\} = \{\theta | f(\theta) \leq c_\alpha^2\}$$

Die Funktion $f(\theta)$ besitzt ein Minimum bei $\frac{X}{Y}$ (mit $f(\frac{X}{Y}) = 0$) und ein Maximum bei $-\frac{Y}{X}$ (mit $f(-\frac{Y}{X}) = X^2 + Y^2$). Ausserdem $\lim_{\theta\to\pm\infty} f(\theta) = Y^2$. Die Abbildung zeigt $f(\theta)$ für $X > Y > 0$:



Die Konfidenzbereiche für verschiedene kritische Werte c_α sind mit gestrichelten Linien angedeutet. Man unterscheidet vier Fälle:

	KB
$c_\alpha^2 < Y^2$	Intervall
$c_\alpha^2 = Y^2$	Halbgerade
$Y^2 < c_\alpha^2 < X^2 + Y^2$	zwei Halbgeraden
$X^2 + Y^2 \leq c_\alpha^2$	$\mathbb{R}$

2. a)

$$\overline{X}_n \sim \mathcal{N}(\mathbf{E}[\overline{X}_n], \operatorname{Var}(\overline{X}_n))$$

mit

$$\mathbf{E}[\overline{X}_n] = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}\, X_i [= X_i] \frac{1}{n}\, n\mu = \mu$$

und

$$\begin{aligned}
\operatorname{Var}(\overline{X}_n) &= \frac{1}{n^2}\operatorname{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}\operatorname{Var}\Big(\sum_{i=1}^{n}\varepsilon_i\Big) = \frac{1}{n^2}\operatorname{Cov}\Big(\sum_{i=1}^{n}\varepsilon_i, \sum_{i=1}^{n}\varepsilon_i\Big) \\
&= \frac{1}{n^2}\sum_{i,j=1}^{n}\operatorname{Cov}(\varepsilon_i,\varepsilon_j) = \frac{1}{n^2}\sum_{i,j=1}^{n}\sigma^2\rho^{|i-j|} \\
&\overset{h:=i-j}{=} \frac{\sigma^2}{n^2}\sum_{h=-n+1}^{n-1}\rho^{|h|}(n-|h|) = \frac{\sigma^2}{n}\sum_{h=-n+1}^{n-1}\rho^{|h|}\Big(1-\frac{|h|}{n}\Big)\,.
\end{aligned}$$

b)

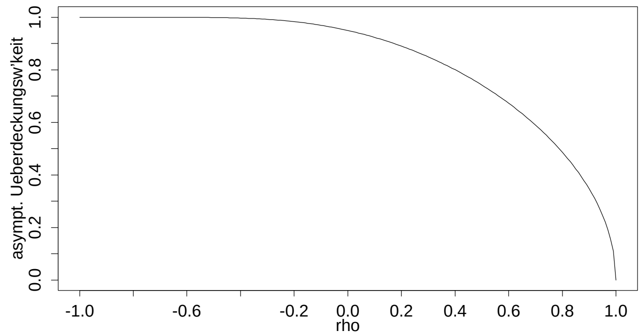
$$\begin{aligned}
P_{\mu,\rho}[I_n \ni \mu] &= P_{\mu,\rho}\Big[|\overline{X}_n-\mu| \le \frac{1.96}{\sqrt{n}}\Big] = 2P_{\mu,\rho}\Big[\overline{X}_n-\mu \le \frac{1.96}{\sqrt{n}}\Big]-1 \\
&= 2P_{\mu,\rho}\Big[\frac{\overline{X}_n-\mu}{\sqrt{\operatorname{Var}(\overline{X}_n)}} \le \frac{1.96}{\sqrt{n\operatorname{Var}(\overline{X}_n)}}\Big]-1 \\
&= 2\Phi\Big(\frac{1.96}{\sqrt{n\operatorname{Var}(\overline{X}_n)}}\Big)-1 \xrightarrow{\Phi\text{ stetig}} 2\Phi\Big(\frac{1.96}{\sigma_\infty(\rho)}\Big)-1
\end{aligned}$$

für $n\to\infty$, wobei

$$\begin{aligned}
\sigma_\infty^2(\rho) &= \lim_{n\to\infty} n\operatorname{Var}(\overline{X}_n) \overset{\sigma^2=1}{=} \lim_{n\to\infty}\sum_{h=-n+1}^{n-1}\rho^{|h|}\Big(1-\frac{|h|}{n}\Big) \\
&= \sum_{h=-\infty}^{\infty}\rho^{|h|} - \lim_{n\to\infty}\frac{1}{n}\sum_{h=-n+1}^{n-1}|h|\rho^{|h|} = 1+2\sum_{h=1}^{\infty}\rho^h - \lim_{n\to\infty}\frac{2}{n}\sum_{h=1}^{n-1}h\rho^h \\
&= 1+2\frac{\rho}{1-\rho} - \lim_{n\to\infty}\frac{2\rho}{n}\sum_{h=1}^{n-1}(\rho^h)' = \frac{1+\rho}{1-\rho} - \Big(\lim_{n\to\infty}\frac{2\rho}{n}\Big)\cdot\Big(\lim_{n\to\infty}\sum_{h=1}^{n-1}(\rho^h)'\Big) \\
&= \frac{1+\rho}{1-\rho} - 0\cdot\sum_{h=1}^{\infty}(\rho^h)' \overset{|\rho|<1}{=} \frac{1+\rho}{1-\rho} - 0\cdot\Big(\sum_{h=1}^{\infty}\rho^h\Big)' = \frac{1+\rho}{1-\rho} - 0\cdot\Big(\frac{\rho}{1-\rho}\Big)' \\
&= \frac{1+\rho}{1-\rho}\,.
\end{aligned}$$

Also gilt

$$\lim_{n\to\infty} P_{\mu,\rho}[I_n\ni\mu] = 2\Phi\Big(1.96\sqrt{\frac{1-\rho}{1+\rho}}\Big)-1\,.$$



Aus der Figur ist schon erkennbar:

$$\begin{aligned}\lim_{\rho\to+1} 2\Phi\Big(1.96\sqrt{\frac{1-\rho}{1+\rho}}\Big)-1 &= 2\Phi(0)-1=0\\ \lim_{\rho\to-1} 2\Phi\Big(1.96\sqrt{\frac{1-\rho}{1+\rho}}\Big)-1 &= 2\Phi(\infty)-1=1\end{aligned}$$

Bei starker positiver Korrelation zwischen den X_i wird die asymptotische Überdeckungswahrscheinlichkeit verschwindend klein. Man darf die Abhängigkeit also nicht einfach vernachlässigen!

3. a) Betrachte die neue Stichprobe $z=(y,x_2,\ldots,x_n)$. Für $y\to\infty$ gilt:

$$\begin{aligned}\overline{z} &= \frac{1}{n}(y+\sum_{i=2}^n x_i)=\frac{1}{n}y+O(1)\\ s_z^2 &= \frac{1}{n-1}[(y-\overline{z})^2+\sum_{i=2}^n(x_i-\overline{z})^2]\\ &= \frac{1}{n-1}[(\frac{n-1}{n}y)^2+(n-1)(\frac{1}{n}y)^2+O(y)]\\ &= \frac{1}{n}y^2+O(y)\end{aligned}$$

Also gilt

$$\lim_{y\to\infty} T=\lim_{y\to\infty}\frac{\overline{z}\sqrt{n}}{s_z}=\lim_{y\to\infty}\frac{\frac{1}{n}y\sqrt{n}}{\frac{1}{\sqrt{n}}y}=1.$$

Der t-Test verwirft H_0, falls $|T|>t_{n-1}^{97.5\%}>1.96$. Im Falle eines groben Ausreissers verwirft also der t-Test die Nullhypothese nie!! Bei Daten mit Ausreissern ist die Macht des t-Tests somit sehr klein.

b) Analog wie in Teil a) gilt für die Stichprobe $z=(y,\ldots,y,x_{k+1},\ldots,x_n)$ und $y\to\infty$:

$$\begin{aligned}\overline{z} &= \frac{k}{n}y+O(1)\\ s_z^2 &= \frac{1}{n-1}[k(\frac{n-k}{n}y)^2+(n-k)(\frac{k}{n}y)^2+O(y)]\\ &= \frac{k(n-k)}{n(n-1)}y^2+O(y)\end{aligned}$$

Die Beobachtung y wird fälschlicherweise nicht als Ausreisser erkannt, falls

$$\lim_{y\to\infty} \frac{y-\overline{z}}{s_z} = \frac{(n-k)/n}{\sqrt{k(n-k)/(n(n-1))}} = \sqrt{\frac{(n-k)(n-1)}{nk}} \leq 3.$$

Auflösen nach k ergibt, dass y nicht verworfen wird, falls $k \geq \frac{n(n-1)}{10n-1} \approx 10\% \cdot n$. Weil 10% grobe Fehler in der Praxis durchaus auftreten können, ist die hier vorgestellte Verwerfungsregel nicht zu empfehlen. Besser wäre eine auf dem Median und dem MAD (statt auf dem Mittelwert und der Standardabweichung) basierende Verwerfungsregel.

Bemerkung: Wir haben hier nur gezeigt, dass der Bruchpunkt des vorgeschlagenen Schätzers kleiner oder gleich $\frac{n-1}{10n-1}$ ist. Es lässt sich jedoch auch zeigen, dass wir die Ausreisser so geschickt wie möglich gewählt haben, so dass der Bruchpunkt tatsächlich $\frac{n-1}{10n-1} \approx 10\%$ ist.