Musterlösung zur Übung 1

1. a) Die Veränderung des arithmethischen Mittels ist gegeben durch:

$$\begin{aligned}\hat{\mu}_{n+1} &= \frac{n}{n+1}\hat{\mu}_n + \frac{x}{n+1} \\ &= \hat{\mu}_n + \frac{1}{n+1}(x - \hat{\mu}_n).\end{aligned}$$

Für die Veränderung der empirischen Varianz berechnet man zunächst (z.B. mit Hilfe des Hinweis),

$$\sum_{i=1}^{n+1}(x_i - \hat{\mu}_{n+1})^2 = \sum_{i=1}^{n+1}(x_i - \hat{\mu}_n)^2 - (n+1)(\hat{\mu}_n - \hat{\mu}_{n+1})^2.$$

Die empirische Varianz lässt sich mit $x_{n+1} = x$ somit schreiben als

$$\begin{aligned}S_{n+1}^2 &= \frac{1}{n}\sum_{i=1}^{n+1}(x_i - \hat{\mu}_{n+1})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n+1}(x_i - \hat{\mu}_n)^2 - \frac{n+1}{n}(\hat{\mu}_n - \hat{\mu}_{n+1})^2 \\ &= \frac{n-1}{n}S_n^2 + \frac{1}{n}(x - \hat{\mu}_n)^2 - \frac{1}{n(n+1)}(x - \hat{\mu}_n)^2 \\ &= (1 - \frac{1}{n})S_n^2 + \frac{1}{n}(1 - \frac{1}{n+1})(x - \hat{\mu}_n)^2 \\ &= (1 - \frac{1}{n})S_n^2 + \frac{1}{n+1}(x - \hat{\mu}_n)^2\end{aligned}$$

b) Die t-Test Statistik ist gegeben durch

$$T_n(x_1, \ldots, x_n) = \frac{\sqrt{n}\,\hat{\mu}_n}{S_n}$$

Hinzufügen von x führt zu der Teststatistik (benütze bisherige Resultate)

$$\begin{aligned}T_{n+1}(x_1, \ldots, x_n, x) &= \frac{\sqrt{n+1}\,\hat{\mu}_{n+1}}{\sqrt{S_{n+1}^2}} \\ &= \frac{\sqrt{n+1}\hat{\mu}_n + \frac{x - \hat{\mu}_n}{\sqrt{n+1}}}{\sqrt{\frac{n-1}{n}S_n^2 + \frac{(x-\hat{\mu}_n)^2}{n+1}}}.\end{aligned}$$

Daher $T_{n+1} \to \pm 1$ für $x \to \infty$ und n fest.
Der Annahmebereich des Testes (zum 5% Niveau) mit $df = n$ Freiheitsgraden ist gegeben durch $[-t_{df;0.975}, t_{df;0.975}]$, wobei $t_{df;0.975}$ das 97.5 % Quantil einer t-Verteilung mit df Freiheitsgraden ist. Es gilt $t_{df;0.975} \geq t_{\infty,0.975} = \Phi^{-1}(0.975) \cong 1.96$. Das heisst, die Nullhypothese kann *nicht* verworfen werden, wenn eine extreme Beobachtung hinzugefügt wird.

c) Zur Erinnerung: Das α-gestutzte Mittel von $(X_1, \ldots, X_n)$ ist

$$\hat{\mu}_n = \frac{1}{n-2[n\alpha]} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{(i)}\,.$$

Es seien $X_{(1)} = x_1\,, X_{(2)} = x_2\ \ldots, X_{(n)} = x_n$ und x beobachtet worden. Von diesen $n+1$ (geordneten) Beobachtungen werden sowohl die kleinsten $k = [(n+1)\alpha]$ als auch die grössten k Beobachtungen weggestutzt. Falls $x \le x_k$, dann wird x selbst auch gestutzt, und es gilt

$$\hat{\mu}_{n+1}(x_1, \ldots, x_n, x) = \frac{1}{n+1-2k} \sum_{i=k}^{n-k} x_i \quad (\text{falls } x \le x_k)\,.$$

Wenn x gross ist, wird es ebenfalls weggestutzt:

$$\hat{\mu}_{n+1}(x_1, \ldots, x_n, x) = \frac{1}{n+1-2k} \sum_{i=k+1}^{n-k+1} x_i \quad (\text{falls } x \ge x_{n-k+1})\,,$$

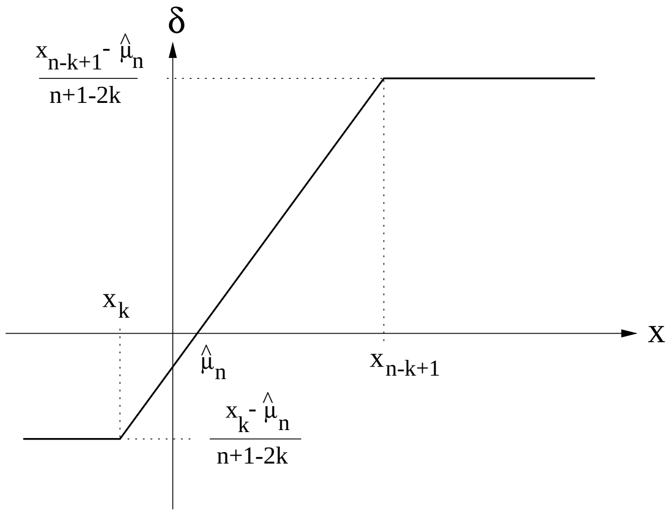
und in allen übrigen Fällen gilt

$$\hat{\mu}_{n+1}(x_1, \ldots, x_n, x) = \frac{1}{n+1-2k} \Big(\sum_{i=k+1}^{n-k} x_i + x \Big) \quad (\text{sonst})\,.$$

Als nächstes betrachten wir $\hat{\mu}_n$. Wir nehmen an, dass von der kleineren Stichprobe gleich viele Beobachtungen gestutzt werden wie von der grösseren, d. h. $[n\alpha] = [(n+1)\alpha] = k$,

$$\hat{\mu}_n = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} x_i.$$

(Möglich wäre ja auch $[n\alpha] = [(n+1)\alpha] - 1$. Die Ergebnisse sind jedoch dieselben.)
Die folgende Figur zeigt die gesuchte Veränderung $\delta = \hat{\mu}_{n+1}(x_1, \ldots, x_n, x) - \hat{\mu}_n$ als Funktion der zusätzlichen Beobachtung x (dies folgt aus Aufgabe a).



2. 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

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

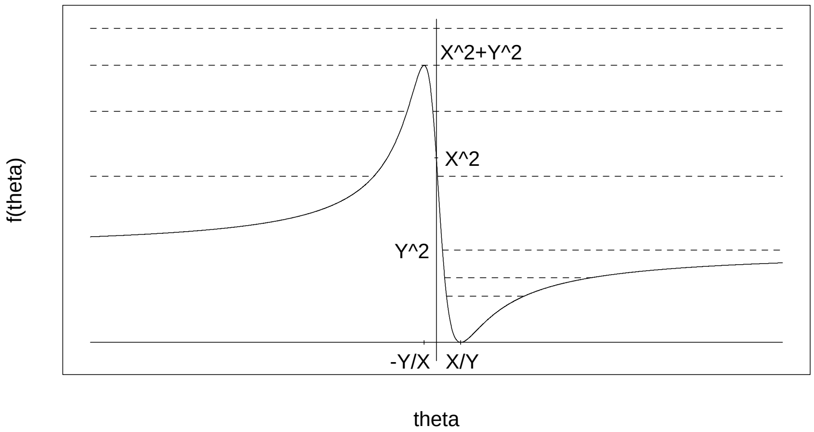
bzw. wenn

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

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

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

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
$c_\alpha^2 \geq X^2 + Y^2$	$\mathbb{R}$