Lösungsskizze zu Blatt 10

Aufgabe 1

Poisson-Verteilung:

$$P_\lambda(k) = \frac{\lambda^k}{k!}\, e^{-\lambda}$$

Nullhypothese $H_0: \ \lambda = 5$

Um den Ablehnungsbereich zu bestimmen, berechnen wir die kritischen Werte. Dies sind genau die Grenzwerte, für die die Poisson-Verteilung innerhalb der Grenzen 0,025 und 0,975 liegt.

$$\sum_{j=0}^{k} P_{\lambda=5}(j) = \sum_{j=0}^{k} \frac{5^j}{j!}\, e^{-5} \overset{!}{\geq} 0,975 \Leftrightarrow$$

$$\sum_{j=0}^{k} \frac{5^j}{j!} \overset{!}{\gtrsim} 144,7 \Leftrightarrow$$

$$1 + 5 + \frac{5^2}{2!} + \frac{5^3}{3!} + \frac{5^4}{4!} + \frac{5^5}{5!} + \frac{5^6}{6!} + \frac{5^7}{7!} + \frac{5^8}{8!} + \frac{5^9}{9!} + \frac{5^{10}}{10!} + ... \simeq 146,4 + ... \overset{!}{\gtrsim} 144,7 \Rightarrow k \geq 10$$

$$\sum_{j=0}^{k} P_{\lambda=5}(j) = \sum_{j=0}^{k} \frac{5^j}{j!}\, e^{-5} \overset{!}{\leq} 0,025 \Leftrightarrow$$

$$\sum_{j=0}^{k} \frac{5^j}{j!} \overset{!}{\lesssim} 3,71 \Leftrightarrow$$

$$1 + ... \overset{!}{\lesssim} 3,71 \Rightarrow k \leq 0$$

Der Ablehnungsbereich der Nullhypothese ist also

$$\{0\} \cup [10, \infty).$$

Aufgabe 2:

Es seien $X_i = \begin{cases} 1, & \text{falls } i\text{-te Ziehung die 13 ist} \\ 0, & \text{sonst} \end{cases}$ die Zufallsvariablen, die das i-te einfache Ziehen der Lotto-Kugel beschreiben. Die Zufallsvariable $S_{4900} = \sum_{i=1}^{4900} X_i$ ist dann binomialverteilt und beschreibt wie oft die 13 bei 4900 einfachen Ziehungen gezogen wird.

<u>Beachte:</u> Falls $P(X_i = 1) = p \quad \forall i \in \{1, \dots, 4900\}$, so gilt

$$\mu_p = \mathbb{E}_p(S_{4900}) = 4900 \cdot p$$

$$\sigma_p = \sqrt{\mathrm{Var}_p(S_{4900})} = \sqrt{4900\, p \cdot q} = \sqrt{4900\, p(1-p)}$$

Wir nehmen an, dass die Binomialverteilung durch die Normalverteilung angenähert werden kann und erhalten damit:

$$\mathbb{P}_p(S_{4900} \leq k) = \Phi^{4900p}_{4900 \cdot p(1-p)}(k) =$$

$$= \int_{-\infty}^{k} \frac{1}{\sqrt{2\pi \cdot 4900 \cdot p(1-p)}} \exp\left[\frac{-(x - 4900 \cdot p)^2}{2 \cdot 4900\, p(1-p)}\right]$$

Unsere Nullhypothese lautet $p \leq \frac{1}{49}$. Dies gibt uns eine Familie von Normalverteilungen. Wir wollen nun ein $k \in \{1, 2, \dots, 4900\}$ finden, sodass

$$\mathbb{P}_p(S_{4900} > k) < 0{,}05 \quad \forall p \leq \frac{1}{49}.$$

Aufgabe 2:

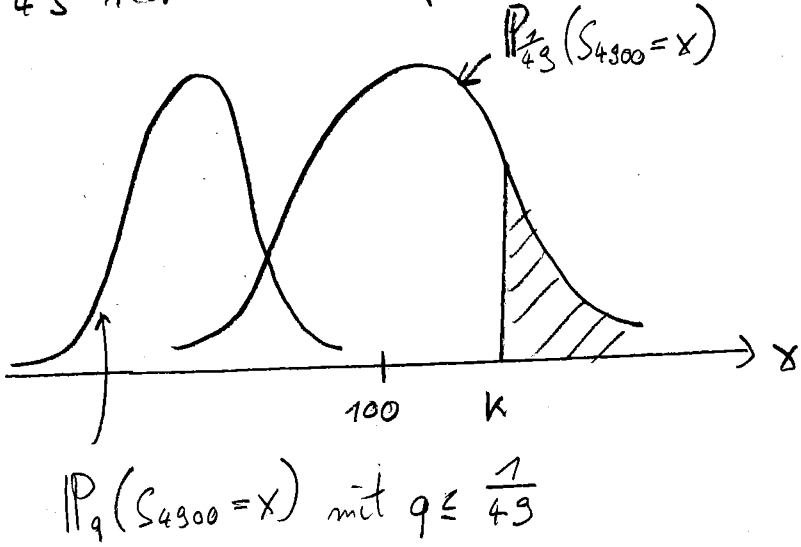
Da $\mu_p = 4900p$ streng monoton wachsend auf $[0, \frac{1}{49}]$ ist, ist die Normalverteilung von $p = \frac{1}{49}$ am weitesten nach rechts verschoben.

Ebenso ist $\frac{d}{dp}\sigma_p^2 = \frac{d}{dp}(4900(p-p^2)) = 4900(1-2p) > 0$

$\forall p \in [0, \frac{1}{49}]$.

Somit hat die Normalverteilung mit $p = \frac{1}{49}$ die größte Varianz und das Problem $\mathbb{P}_p(S_{4900} > k) < 0{,}05 \;\forall p \leq \frac{1}{49}$ reduziert sich auf

$$\mathbb{P}_{\frac{1}{49}}(S_{4900} > k) < 0{,}05.$$



Da $\mu_{\frac{1}{49}} = \frac{4900}{49} = 100$ und $\sigma_{\frac{1}{49}} = \sqrt{4900 \cdot \frac{1}{49} \cdot \frac{48}{49}} = \sqrt{100 \cdot \frac{48}{49}}$

gilt für $p = \frac{1}{49}$ $S_{4900} \sim \mathcal{N}(100, 100 \cdot \frac{48}{49})$.

Aufgrund der Symmetry der Normalverteilung gilt ($\mu = \mu_{\frac{1}{49}}$, $\sigma = \sigma_{\frac{1}{49}}$)

$$\mathbb{P}_{\frac{1}{49}}(S_{4900} > \mu + z \cdot \sigma) = \frac{1}{2}\mathbb{P}_{\frac{1}{49}}(S_{4900} \notin [\mu - z \cdot \sigma, \mu + z \cdot \sigma]) \overset{!}{=} 0{,}05$$

Aus einer Tabelle mit den Werten der Normalverteilung bekommen wir, dass $z = 1{,}645$. Wir erhalten somit

$$k = 100 + 1{,}645 \cdot \sqrt{100 \frac{48}{49}} = 116{,}28.$$

<u>Aufgabe 2:</u>

Die 13 muss also mindestens 117 mal erscheinen, damit die Behauptung $p = \frac{1}{49}$ mit einer Irrtumswahrscheinlichkeit von weniger als 5 % abgelehnt werden kann.

Aufgabe 3:

Das Ziel dieser Aufgabe ist es herauszufinden, wie viele Erbsen N der Großhändler mindestens testen muss um einen α- und β-Fehler von jeweils unter 5% zu bekommen.

Unsere Hypothesen lauten:

Nullhypothese H_0: Erbse gehört zu Saatgut 1. Wahl,
d.h. $p(\text{Erbse keimt}) = \frac{9}{10}$.

Alternativhypothese H_1: Erbse gehört zu Saatgut 2. Wahl, d.h.
$p(\text{Erbse keimt}) = \frac{7}{10}$.

Es seien $X_i = \begin{cases} 1 & i\text{-te Erbse keimt} \\ 0 & i\text{-te Erbse keimt nicht} \end{cases}$ $i \in \{1, \dots N\}$ die Zufallsvariable, die das keimen der i-ten Erbse beschreiben. Dann ist die Zufallsvariable $S_n = \sum_{i=1}^{N} X_i$ binomialverteilt und beschreibt wie viele Erbsen bei einer Stichprobe mit N-Stück keimen.

Falls wir die Nullhypothese annehmen, so gilt:

$$\mu_{9/10} = \mathbb{E}_{9/10}(S_N) = N \cdot \frac{9}{10} \text{ und}$$

$$\sigma_{9/10} = \sqrt{Var_{9/10}(S_N)} = \sqrt{N \cdot \frac{9}{10} \cdot \frac{1}{10}} = \sqrt{N \frac{9}{100}}$$

Aufgabe 3:

Im Falle der Alternativhypothese H_1 bekommen wir:

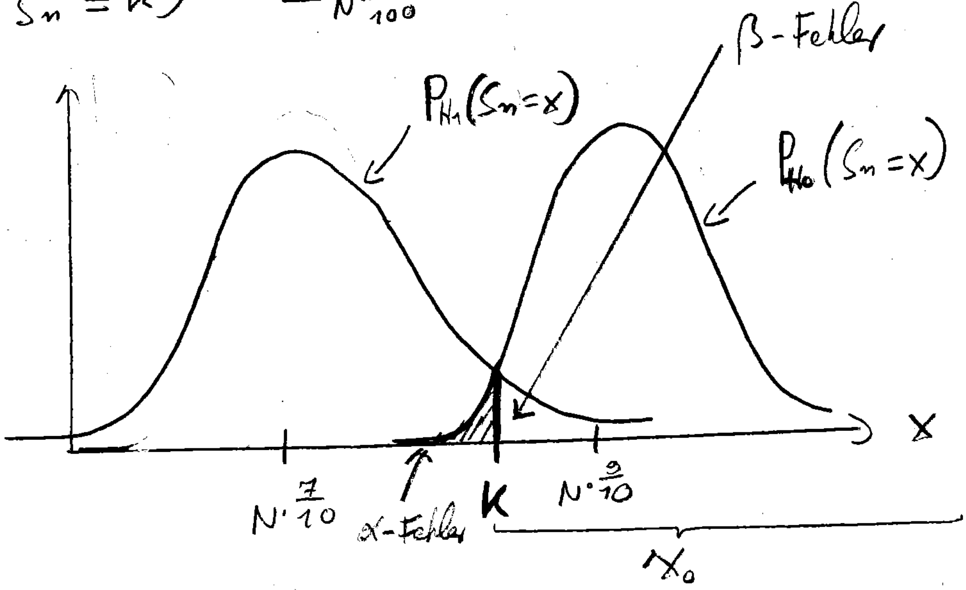
$$\mu_{7/10} = \mathbb{E}_{7/10}(S_N) = N \cdot \frac{7}{10} \quad \text{und}$$

$$\sigma_{7/10} = \sqrt{Var_{7/10}(S_N)} = \sqrt{N \cdot \frac{7}{10} \cdot \frac{3}{10}} = \sqrt{N \cdot \frac{21}{100}}$$

Im Folgenden nehmen wir an, dass wir die Binomialverteilung durch die Normalverteilung annähern können. Es gilt dann:

$$P_{H_0}(S_n \leq k) \approx \Phi_{N \cdot \frac{9}{100}}^{N \cdot \frac{9}{10}}(k)$$

$$P_{H_1}(S_n \leq k) \approx \Phi_{N \cdot \frac{21}{100}}^{N \cdot \frac{7}{10}}(k)$$



Es soll nun die minimale Stichprobenzahl N gefunden werden, sodass es einen Annahmebereich $X_0 = [k, N]$ gibt mit

α-Fehler: $P_{H_0}(H_0 \text{ wird abgelehnt}) = P_{H_0}(S_N < k) \overset{!}{<} 0{,}05$

β-Fehler: $P_{H_1}(H_1 \text{ wird abgelehnt}) = P_{H_1}(S_N \geq k) \overset{!}{<} 0{,}05$

Aufgabe 3:

Es gilt

$$P_{H_0}\left(S_N < \mu_{9/10} - z_0 \cdot \sigma_{9/10}\right) = \frac{1}{2} P_{H_0}\left(S_N \in \left[\mu_{9/10} - z_0 \cdot \sigma_{9/10}, \mu_{9/10} + z_0 \cdot \sigma_{9/10}\right]^c\right) \overset{\nabla}{=} 0{,}05$$

$$P_{H_1}\left(S_N > \mu_{7/10} + z_1 \cdot \sigma_{7/10}\right) = \frac{1}{2} P_{H_1}\left(S_N \in \left[\mu_{7/10} - z_1 \cdot \sigma_{7/10}, \mu_{7/10} + z_1 \cdot \sigma_{7/10}\right]^c\right) \overset{\nabla}{=} 0{,}05$$

Laut Tabelle gilt $z_0 = z_1 = 1{,}645$ und damit erhalten wir zwei Bendingungen

$$① \quad k \overset{\nabla}{\le} \mu_{9/10} - 1{,}645 \cdot \sigma_{9/10} = N \cdot \frac{9}{10} - 1{,}645 \cdot \sqrt{N \cdot \frac{9}{100}} \Leftrightarrow -k \geq -0{,}9N + 1{,}645 \cdot \sqrt{N \cdot \frac{9}{100}}$$

$$② \quad k \overset{\nabla}{\ge} \mu_{7/10} + 1{,}645 \cdot \sigma_{7/10} = N \cdot \frac{7}{10} + 1{,}645 \cdot \sqrt{N \cdot \frac{21}{100}} \Leftrightarrow k \geq 0{,}7N + 1{,}645 \sqrt{N \cdot \frac{21}{100}}.$$

Aus diesen folgt

$$0 \geq -0{,}2N + \sqrt{N} \cdot 1{,}645 \left(\sqrt{\frac{9}{100}} + \sqrt{\frac{21}{100}}\right) \Leftrightarrow$$

$$\sqrt{N} \geq 1{,}645 \cdot 5 \left(\sqrt{\frac{9}{100}} + \sqrt{\frac{21}{100}}\right) \Leftrightarrow$$

$$N \geq \left[1{,}645 \cdot 5 \left(\sqrt{\frac{9}{100}} + \sqrt{\frac{21}{100}}\right)\right]^2 \approx 38{,}89$$

Der Großhändler muss also mindestens 39 Erbsen testen.

Aufgabe 4:

Es wird angenommen, dass die Körpergröße des Menschen für ein Geschlecht normalverteilt ist. D.h., wenn sie die Körpergröße eines ihrer Freunde messen, ist anzunehmen, dass diese durch eine Zufallsvariable $X_i \sim \mathcal{N}(\mu, \sigma^2)$ mit $i \in \{3, 4, \dots, N\}$, $\mu \in \mathbb{R}$, $\sigma \geq 0$ beschrieben wird. Das Messen der Körpergrößen von n-Teilnehmern entspricht dem Ausführen von n-unabhängigen Experimenten $X = (X_1, \dots, X_n)$. Dies lässt darauf schließen, dass die Zufallsvariable X gemäß einer Produktdichte verteilt ist, d.h.

$$\varrho_{\mu,\sigma}(x_1, \dots, x_n) = \varrho_{\mu,\sigma}(x_1) \cdot \varrho_{\mu,\sigma}(x_2) \cdot \ldots \cdot \varrho_{\mu,\sigma}(x_n).$$

Für n spezifische Messergebnisse $(x_1, \dots, x_n) \in \mathbb{R}^n$ wollen wir nun $\tilde{\mu} \in \mathbb{R}, \tilde{\sigma} \geq 0$ finden, sodass $\varrho_{\tilde{\mu},\tilde{\sigma}}(x_1, \dots, x_n)$ für fixes $(x_1, \dots, x_n) \in \mathbb{R}^n$ maximal ist.

Da die X_i normalverteilt sind, betrachten wir $\varrho_X : \mathbb{R}^2 \to \mathbb{R}$, $(\mu, \sigma) \mapsto \varrho_X(\mu, \sigma)$ mit

$$\varrho_X(\mu, \sigma) = \varrho_{X_1}(\mu, \sigma) \cdot \varrho_{X_2}(\mu, \sigma) \cdot \ldots \varrho_{X_n}(\mu, \sigma) =$$

$$= \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{\sigma^n} \exp\left[- \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} \right]$$

und bestimmen dessen Maximum.

Aufgabe 4:

- $$\underline{\underline{(\partial_\mu \varrho_x)(\mu,\sigma)}} = \overbrace{\frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{\sigma^n} \exp\left[-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right]}^{=\varrho_x(\mu,\sigma)} \cdot (-1) \sum_{i=1}^{n} \frac{2(x_i-\mu)}{2\sigma^2} \cdot (-1)$$

$$= \underline{\underline{\varrho_x(\mu,\sigma) \cdot \sum_{i=1}^{n} \frac{(x_i-\mu)}{\sigma^2}}}.$$

- $$\underline{\underline{(\partial_\sigma \varrho_x)(\mu,\sigma)}} = \frac{1}{(2\pi)^{n/2}} \cdot \frac{(-n)}{\sigma^{n+1}} \exp\left[-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right] +$$

$$+ \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{\sigma^n} \exp\left[-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right] \cdot (-1) \sum_{i=1}^{n} \frac{(-2)(x_i-\mu)^2}{2\sigma^3}$$

$$= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sigma^n} \exp\left[-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right] \cdot \left(\frac{-n}{\sigma} + \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{\sigma^3}\right)$$

$$= \underline{\underline{\varrho_x(\mu,\sigma) \cdot \left(\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{\sigma^3} - \frac{n}{\sigma}\right)}}.$$

$\Rightarrow$

$$(\partial_\mu \varrho_x)(\tilde{\mu},\tilde{\sigma}) = 0 \wedge (\partial_\sigma \varrho_x)(\tilde{\mu},\tilde{\sigma}) = 0 \quad \Leftrightarrow$$

$$① \sum_{i=1}^{n} \frac{(x_i-\tilde{\mu})}{\tilde{\sigma}^2} = 0 \wedge ② \left(\sum_{i=1}^{n} \frac{(x_i-\tilde{\mu})^2}{\tilde{\sigma}^3} - \frac{n}{\tilde{\sigma}}\right) = 0$$

$$① \Leftrightarrow \boxed{\tilde{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i}$$

$$② \Leftrightarrow \boxed{\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i-\tilde{\mu})^2}$$

Nun bleibt es noch zu zeigen, dass $\varrho_x(\tilde{\mu},\tilde{\sigma})$ ein lokales (und auch globales) Maximum besitzt. Dazu berechnen wir die Hessematrix $H\varrho(\tilde{\mu},\tilde{\sigma})$ an der Stelle $(\tilde{\mu},\tilde{\sigma})$ und zeigen, dass diese nur negative Eigenwerte besitzt.

Aufgabe 4:

$$\bullet\ \underline{(\partial_\mu^2 \varrho_x)(\mu,\sigma)} = \left(\partial_\mu(\partial_\mu \varrho_x)\right)(\mu,\nu) = (\partial_\mu \varrho_x)(\mu,\sigma)\cdot\sum_{i=1}^{n}\frac{(x_i-\mu)}{\sigma^2}$$

$$+ \varrho_x(\mu,\sigma)\cdot\sum_{i=1}^{n}\frac{1}{\sigma^2}\cdot(-1)$$

$$= (\partial_\mu \varrho_x)(\mu,\sigma)\cdot\sum_{i=1}^{n}\frac{(x_i-\mu)}{\sigma^2} + \varrho_x(\mu,\sigma)\cdot\frac{(-n)}{\sigma^2}$$

Da $(\partial_\mu \varrho_x)(\tilde\mu,\tilde\sigma) = 0$, gilt

$$\underline{(\partial_\mu^2 \varrho_x)(\tilde\mu,\tilde\sigma)} = \varrho_x(\tilde\mu,\tilde\sigma)\cdot\frac{(-n)}{\tilde\sigma^2}$$

$$\bullet\ \underline{(\partial_\sigma^2 \varrho_x)(\mu,\sigma)} = \left(\partial_\sigma(\partial_\sigma \varrho_x)\right)(\mu,\nu) = (\partial_\sigma \varrho_x)(\mu,\sigma)\cdot\left(\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^3} - \frac{n}{\sigma}\right) +$$

$$+ \varrho_x(\mu,\sigma)\cdot\left[\sum_{i=1}^{n}\frac{(-3)(x_i-\mu)^2}{\sigma^4} + \frac{n}{\sigma^2}\right]$$

$$= (\partial_\sigma \varrho_x)(\mu,\sigma)\left[\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^3} - \frac{n}{\sigma}\right] + \varrho_x(\mu,\sigma)\cdot\frac{1}{\sigma^2}\left[n - 3\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\right]$$

$$\Rightarrow$$

$$\underline{(\partial_\sigma^2 \varrho_x)(\tilde\mu,\tilde\sigma)} = 0 + \varrho_x(\tilde\mu,\tilde\sigma)\cdot\frac{1}{\tilde\sigma^2}\left[n - 3\sum_{i=1}^{n}\frac{(x_i-\tilde\mu)^2}{\tilde\sigma^2}\right] =$$

$$\overset{(2)}{=} \varrho_x(\tilde\mu,\tilde\sigma)\cdot\frac{1}{\tilde\sigma^2}\left[n - 3\cdot n\right]$$

$$\Rightarrow \underline{\underline{(\partial_\sigma^2 \varrho_x)(\tilde\mu,\tilde\sigma) = \varrho_x(\tilde\mu,\tilde\sigma)\cdot\frac{(-2\cdot n)}{\tilde\sigma^2}}}$$

Aufgabe 4:

$$\underline{\underline{(\partial_\mu \partial_\sigma \varrho_x)(\mu,\sigma)}} = (\partial_\sigma \partial_\mu \varrho_x)(\mu,\sigma) = (\partial_\sigma(\partial_\mu \varrho_x))(\mu,\sigma) =$$

$$= (\partial_\sigma \varrho_x)(\mu,\sigma)\cdot\left(\sum_{i=1}^{n} \frac{(x_i-\mu)}{\sigma^2}\right) + \varrho_x(\mu,\sigma)\cdot\left(\sum_{i=1}^{n} \frac{(-2)(x_i-\mu)}{\sigma^3}\right)$$

$$= (\partial_\sigma \varrho_x)(\mu,\sigma)\cdot\left(\sum_{i=1}^{n} \frac{(x_i-\mu)}{\sigma^2}\right) - \varrho_x(\mu,\sigma)\cdot\frac{2}{\sigma}\left(\sum_{i=1}^{n} \frac{(x_i-\mu)}{\sigma^2}\right).$$

Aus ① folgt, dass $\sum_{i=1}^{n} \frac{(x_i-\tilde{\mu})}{\tilde{\sigma}^2} = 0$ und wir erhalten somit

$$(\partial_\mu \partial_\sigma \varrho_x)(\tilde{\mu},\tilde{\sigma}) = (\partial_\sigma \partial_\mu \varrho_x)(\tilde{\mu},\tilde{\sigma}) = 0$$

und damit

$$H\varrho(\tilde{\mu},\tilde{\sigma}) = \varrho_x(\tilde{\mu},\tilde{\sigma})\begin{bmatrix} \frac{-n}{\tilde{\sigma}^2} & 0 \\ 0 & -\frac{2\cdot n}{\tilde{\sigma}^2} \end{bmatrix}.$$

Da $\varrho_x(\mu,\sigma) \geq 0$ für alle $\mu \in \mathbb{R}, \sigma \geq 0$ hat $H\varrho(\tilde{\mu},\tilde{\sigma})$ die zwei negativen Eigenwerte $-\varrho_x(\tilde{\mu},\tilde{\sigma})\cdot\frac{n}{\tilde{\sigma}^2}$ und $-\varrho_x(\tilde{\mu},\tilde{\sigma})\cdot\frac{2n}{\tilde{\sigma}^2}$.

Es handelt sich also bei $(\tilde{\mu},\tilde{\sigma})$ um ein lokales Maximum. Da $\varrho_x(\mu,\sigma)$ für große Werte von μ oder σ gegen 0 geht handelt es sich auch um ein globales Maximum.