

Ue Statistik u. Wahrscheinlichkeitsth. f. Inf. 107.251 W 2005 6 http://www.statistik.tuwien.ac.at/RISueb/	Di 12:00 – 18:00
	10.Blatt
Werner GURKER Tel.: 58801-10724 Spr.: Di u. Do 11-12 e-mail: W.Gurker@tuwien.ac.at	17. Jän. 2006

1. X_1, \dots, X_n sei eine Stichprobe aus einer Verteilung mit Mittelwert μ und Varianz σ^2 ; \bar{X}_n sei der Stichprobenmittelwert und S_n^2 die Stichprobenvarianz.

(a) Zeigen Sie („Verschiebungssatz“):

$$\sum_{i=1}^n (X_i - c)^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 + n(\bar{X} - c)^2, \quad c \in \mathbb{R}$$

(b) Zeigen Sie mit Hilfe von (a), daß S_n^2 ein unverzerrter Schätzer für σ^2 ist.

Hinweis: Die Konstante c kann beliebig gewählt werden; klarerweise wird man sie aber so wählen, daß die Erwartungswerte der Ausdrücke von (a) einfach zu bestimmen sind.

2. Die folgende Tabelle ist die Zusammenfassung einer Stichprobe des Umfangs 50 aus einer Poissonverteilung P_μ :

x	0	1	2	3	4	5	6
Häufigkeit	3	8	19	9	7	2	2

- (a) Bestimmen Sie den plausiblen Schätzer von μ .
(b) Ist der plausible Schätzer unverzerrt und konsistent?
(c) Bestimmen Sie den plausiblen Schätzwert von μ .
*(d) Bestimmen Sie den plausiblen Schätzwert von $W(X \geq 2)$.
3. Die folgenden Daten sind Zeiten (Betriebsstunden) bis zum Ausfall eines bestimmten Typs von Transistoren (Einheit = 10^3 Stunden):

0.12, 0.61, 2.09, 2.94, 3.20, 3.32, 3.59, 5.08, 5.71, 5.90,
7.53, 8.04, 8.83, 9.18, 11.54, 12.35

(a) Ermitteln und zeichnen Sie die empirische Verteilungsfunktion.

Hinweis für R-User: Nehmen Sie die Funktion `ecdf` (Package: `stats`).

(b) Ermitteln Sie unter der Annahme, daß es sich um eine Stichprobe aus einer Verteilung mit der Dichte:

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

handelt, den plausiblen Schätzer von λ . Überprüfen Sie, ob der gefundene Schätzer tatsächlich die (logarithmierte) Plausibilitätsfunktion maximiert.

(c) Bestimmen Sie den plausiblen Schätzwert von λ .

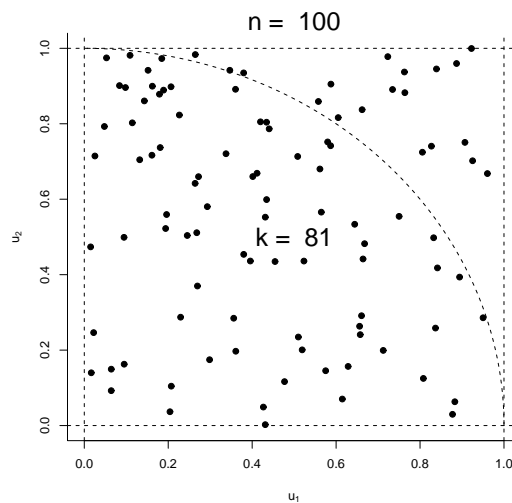
*(d) Zeichnen Sie über die empirische Verteilungsfunktion die auf Basis von (c) geschätzte Verteilungsfunktion. Bestimmen Sie graphisch und/oder rechnerisch Stelle und Größe des maximalen Abstands zwischen den beiden Funktionen.

4. X_1, \dots, X_n sei eine Stichprobe aus einer Alternativverteilung A_p :

*(a) Begründen Sie, warum die folgende sG eine approximative Pivotgröße ist.

$$Z_n = \frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1 - \bar{X}_n)/n}} \sim N(0, 1)$$

- (b) Entwickeln Sie auf Basis von Z_n ein approximatives $(1-\alpha)$ -Konfidenzintervall für den Parameter p .
- (c) Bei einem Zufallsexperiment, bei dem $n = 100$ uniform verteilte Punkte im Quadrat $[0, 1] \times [0, 1]$ erzeugt werden, liegen $k = 81$ innerhalb des Einheitskreises. Ermitteln Sie ein 95%-Konfidenzintervall für π .



5. X_1, \dots, X_n sei eine Stichprobe von $X \sim Ex_\tau$:

*(a) Zeigen Sie:

$$\frac{2n\bar{X}_n}{\tau} \sim \chi^2_{2n}$$

- (b) Entwickeln Sie auf Basis von (a) ein Konfidenzintervall für τ mit ÜDW $1-\alpha$.
- (c) Bestimmen Sie auf Basis der Daten von **Bsp 3** ein 90%-Konfidenzintervall für die mittlere Ausfallzeit dieses Transistortyps (Vs.: Exponentialverteilung).
6. Bestimmen Sie auf Basis des Datensatzes `normtemp.dat` (vgl. Ü-Homepage), wahlweise für `gender=M` oder `gender=W`:
- (a) ein 95%-Konfidenzintervall für die mittlere Körpertemperatur.
- (b) ein 90%-Konfidenzintervall für die Streuung der Körpertemperatur.

Gehen Sie dabei von normalverteilten Beobachtungen aus. Rechnen Sie zuerst die Angaben von °Fahrenheit in °Celsius um. (*Welche Auswirkungen hat diese Umrechnung auf die Konfidenzintervalle?)

Beispiel(teil)e mit (*) werden im Konversatorium behandelt.

Lösungen zum 10. Blatt

1. (a) Dies zeigt man durch Einfügen und Abziehen von \bar{X}_n :

$$\begin{aligned}\sum_{i=1}^n (X_i - c)^2 &= \sum_{i=1}^n (X_i - \bar{X}_n + \bar{X}_n - c)^2 \\&= \sum_{i=1}^n (X_i - \bar{X}_n)^2 + \sum_{i=1}^n (\bar{X}_n - c)^2 + 2(\bar{X}_n - c) \underbrace{\sum_{i=1}^n (X_i - \bar{X}_n)}_{=0} \\&= \sum_{i=1}^n (X_i - \bar{X}_n)^2 + n(\bar{X}_n - c)^2\end{aligned}$$

- (b) Speziell für $c = \mu = \mathbb{E}(X)$ ergibt sich:

$$\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X}_n - \mu)^2$$

Daraus folgt:

$$\begin{aligned}\mathbb{E} \left[\sum_{i=1}^n (X_i - \bar{X}_n)^2 \right] &= \sum_{i=1}^n \underbrace{\mathbb{E}(X_i - \mu)^2}_{\text{Var}(X_i) = \sigma^2} - n \underbrace{\mathbb{E}(\bar{X}_n - \mu)^2}_{\text{Var}(\bar{X}_n) = \sigma^2/n} \\&= n\sigma^2 - \sigma^2 \\&= (n-1)\sigma^2\end{aligned}$$

Damit folgt die Behauptung:

$$\mathbb{E}(S_n^2) = \frac{1}{n-1} \mathbb{E} \left[\sum_{i=1}^n (X_i - \bar{X}_n)^2 \right] = \sigma^2$$

2. (a) Zunächst bestimmt man die Plausibilitätsfunktion:

$$L(\mu; \text{Daten}) = \prod_{i=1}^n \frac{\mu^{x_i} e^{-\mu}}{(x_i)!} = \underbrace{\left[\prod_{i=1}^n \frac{1}{(x_i)!} \right]}_{=C} \mu^{\sum_{i=1}^n x_i} e^{-n\mu} = C \mu^{n\bar{x}_n} e^{-n\mu}$$

Durch Logarithmieren ergibt sich die Log-Plausibilitätsfunktion:

$$l(\mu; \text{Daten}) = \ln L(\mu; \text{Daten}) = \ln C + n\bar{x}_n \ln \mu - n\mu$$

Nun bestimmt man die Stelle des Maximums:

$$\frac{dl(\mu; \text{Daten})}{d\mu} = \frac{n\bar{x}_n}{\mu} - n = 0 \longrightarrow \hat{\mu} = \bar{x}_n$$

Der plausible Schätzer von μ ist also $\hat{\mu} = \bar{X}_n$.

(b) Der Schätzer ist unverzerrt:

$$\mathbb{E}(\hat{\mu}) = \mathbb{E}(\overline{X}_n) = \mathbb{E}(X) = \mu$$

Die Konsistenz folgt aus dem GGZ (= Gesetz der großen Zahlen):

$$\overline{X}_n \xrightarrow{W} \mathbb{E}(X) = \mu$$

(c) Der plausible Schätzwert von μ ist gegeben durch:

$$\hat{\mu} = \bar{x} = \frac{0 \cdot 3 + 1 \cdot 8 + \dots + 6 \cdot 2}{50} = \frac{123}{50} = 2.46$$

*(d) Zur Bestimmung des plausiblen Schätzers einer Funktion $g(\theta)$ des Parameters θ („geraffter“ Parameter), kann man sich die *Invarianz* des plausiblen Schätzers zunutze machen, d.h. für den plausiblen Schätzer von $g(\theta)$ gilt $\widehat{g(\theta)} = g(\hat{\theta})$. In der vorliegenden Situation gilt:

$$g(\mu) = W\{X \geq 2\} = 1 - W\{X \leq 1\} = 1 - e^{-\mu} - \mu e^{-\mu}$$

Der plausible Schätzer von $W\{X \geq 2\}$ ist also:

$$\widehat{g(\mu)} = 1 - e^{-\overline{X}_n} - \overline{X}_n e^{-\overline{X}_n}$$

Der plausible Schätzwert ist gegeben durch:

$$1 - e^{-\bar{x}} - \bar{x} e^{-\bar{x}} = 1 - e^{-2.46} - 2.46 e^{-2.46} = 0.7044$$

3. (a) Die empirische Verteilungsfunktion ist eine Treppenfunktion mit Sprüngen an den Stellen der (geordneten) Stichprobe; die Sprunghöhe beträgt jeweils $1/n$; bei mehreren identischen Beobachtungen (hier nicht der Fall) springt sie um das entsprechende Vielfache von $1/n$.

(b) Plausibilitätsfunktion:

$$L(\lambda; \text{Daten}) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i} = \lambda^n e^{-\lambda n \bar{x}_n}$$

Log-Plausibilitätsfunktion:

$$l(\lambda; \text{Daten}) = \ln L(\lambda; \text{Daten}) = n \ln \lambda - \lambda n \bar{x}_n$$

Bestimmung des Maximums:

$$\frac{dl(\lambda; \text{Daten})}{d\lambda} = \frac{n}{\lambda} - n \bar{x}_n = 0 \longrightarrow \hat{\lambda} = \frac{1}{\bar{x}_n}$$

Überprüfung auf (lok.) Maximum:

$$\left. \frac{d^2 l(\lambda; \text{Daten})}{d\lambda^2} \right|_{\lambda=\hat{\lambda}} = -n \bar{x}_n^2 < 0$$

Plausibler Schätzer:

$$\hat{\lambda} = \frac{1}{\overline{X}_n}$$

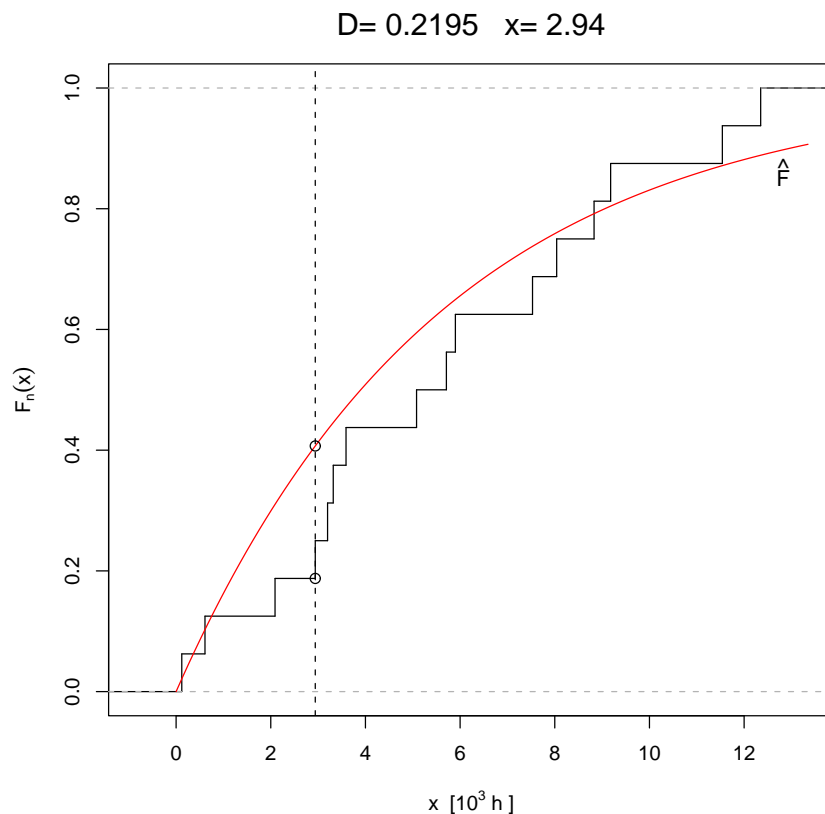
(c) Plausibler Schätzwert:

$$\hat{\lambda} = \frac{16}{0.12 + 0.61 + \dots + 12.35} = \frac{16}{90.03} = 0.1777$$

*(d) Die plausibel geschätzte Verteilungsfunktion ist gegeben durch (vgl. die Lösung von **Bsp 2(d)**):

$$\hat{F}(x) = 1 - e^{-\hat{\lambda}x}, \quad x > 0$$

Das Maximum (genauer: Supremum) D des Abstands zwischen $F_n(x)$ und $\hat{F}(x)$ kann nur an einer Sprungstelle von $F_n(x)$ auftreten; hier gilt $D \approx 0.22$ bei $x_{(4)} = 2.94$.



Bem.: Das R-Skript zur Erzeugung der obigen Abbildung (und der dazu nötigen Berechnungen) findet sich unter **b103.r** auf der Ü-Homepage.

4. *(a) Die Begründung dafür, daß Z_n eine approximative Pivotgröße ist, liegt im zentralen Grenzwertungssatz, und darin, daß $\overline{X}_n(1 - \overline{X}_n)/n$ ein konsistenter Schätzer von $p(1 - p)/n$ (= Varianz von \overline{X}_n) ist.

- (b) Einschließen von Z_n in zwei passende Quantile der Pivotverteilung ($= N(0, 1)$):

$$-z_{1-\alpha/2} \leq \frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1 - \bar{X}_n)/n}} \leq z_{1-\alpha/2}$$

Auflösen der Doppelungleichung nach p :

$$\bar{X}_n - z_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}} \leq p \leq \bar{X}_n + z_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}}$$

- (c) Sind U_1, U_2 zwei unabhängige $U_{(0,1)}$ -verteilte sGn, so ist die Wahrscheinlichkeit, daß (U_1, U_2) innerhalb des Einheitskreises liegt, gleich $\pi/4$. Die folgende Indikatorvariable:

$$X = \begin{cases} 1 & U_1^2 + U_2^2 < 1 \\ 0 & \text{sonst} \end{cases}$$

ist daher A_p -verteilt mit $p = \pi/4$. (Bem.: π wird hier als unbekannter Parameter interpretiert.) Ein Schätzwert für π ist also gegeben durch:

$$\hat{\pi} = 4\bar{x} = 4 \cdot \frac{81}{100} = 3.24$$

(Bem.: Dies ist auch der plausible Schätzwert von π .) Das 95%-Konfidenzintervall für p ist gegeben durch:

$$0.81 \mp \underbrace{z_{0.975}}_{1.96} \sqrt{\frac{0.81 \cdot 0.19}{100}} = [0.7331, 0.8869]$$

Durch Multiplikation mit 4 bekommt man das 95%-Konfidenzintervall für π :

$$[2.9324, 3.5476]$$

5. *(a) Nach dem Additionstheorem für (unabhängige) Gammaverteilungen gilt zunächst (Bem.: $Ex_\tau \equiv \text{Gam}(1, \tau)$):

$$S_n = \sum_{i=1}^n X_i \sim \text{Gam}(n, \tau)$$

Die Dichte dieser Gammaverteilung ist gegeben durch:

$$f_{S_n}(s) = \frac{s^{n-1} e^{-s/\tau}}{(n-1)! \tau^n}, \quad s > 0$$

Nach dem Transformationssatz für Dichten gilt für $Z := 2n\bar{X}_n/\tau = 2S_n/\tau$ ($s = \tau z/2$, $ds/dz = \tau/2$):

$$f_Z(z) = f_{S_n}\left(\frac{\tau z}{2}\right) \left| \frac{ds}{dz} \right| = \frac{\tau^{n-1} z^{n-1} e^{-z/2}}{(n-1)! 2^{n-1} \tau^n} \cdot \frac{\tau}{2} = \frac{z^{n-1} e^{-z/2}}{(n-1)! 2^n}, \quad z > 0$$

Dies ist die Dichte einer χ^2 -Verteilung mit $2n$ Freiheitsgraden. Da letztere – im Gegensatz zur Verteilung von S_n – nicht von τ abhängt, bedeutet dies insbesondere, daß $2n\bar{X}_n/\tau$ eine Pivotgröße ist.

(b) Auf Basis der Pivotgröße von (a) folgt:

$$W \left\{ \chi_{2n; \alpha/2}^2 \leq \frac{2n\bar{X}_n}{\tau} \leq \chi_{2n; 1-\alpha/2}^2 \right\} = 1 - \alpha$$

Durch Auflösen der Doppelungleichung nach τ („ τ in die Mitte“) erhält man ein $100(1 - \alpha)\%$ Konfidenzintervall für τ :

$$\left[\frac{2n\bar{X}}{\chi_{2n; 1-\alpha/2}^2}, \frac{2n\bar{X}}{\chi_{2n; \alpha/2}^2} \right]$$

(c) Auf Basis der Daten von **Bsp 3** ergibt sich als Konfidenzintervall für die mittlere Ausfallzeit ($\bar{x} = 5.627$, $\chi_{32; 0.05}^2 = 20.07$, $\chi_{32; 0.95}^2 = 46.19$):

$$\left[\frac{32 \cdot 5.627}{46.19}, \frac{32 \cdot 5.627}{20.07} \right] = [3.898, 8.971]$$

6. (a) Ein $(1 - \alpha)$ -Konfidenzintervall für den Mittelwert ist gegeben durch (Stichprobe aus Normalverteilung):

$$\bar{X}_n - t_{n-1; 1-\alpha/2} \frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + t_{n-1; 1-\alpha/2} \frac{S_n}{\sqrt{n}}$$

Umrechnen der Temperaturen in °C:

$$x[^\circ\text{F}] = \frac{5(x - 32)}{9} [^\circ\text{C}]$$

	n	\bar{x}_n	s_n	$t_{n-1; 0.975}$	95%-Konfidenzintervall (μ)	
					untere Grenze	obere Grenze
M	65	36.72	0.3882	1.998	36.63	36.82
W	65	36.89	0.4130	1.998	36.78	36.99

Bem.: Die Intervalle für °F und °C stehen in derselben Beziehung zueinander wie die Temperaturen.

(b) Ein $(1 - \alpha)$ -Konfidenzintervall für die Streuung ist gegeben durch (Stichprobe aus Normalverteilung):

$$\left[\sqrt{\frac{(n-1)S_n^2}{\chi_{n-1; 1-\alpha/2}^2}}, \sqrt{\frac{(n-1)S_n^2}{\chi_{n-1; \alpha/2}^2}} \right]$$

Mit $\chi_{64; 0.05}^2 = 46.595$ und $\chi_{64; 0.95}^2 = 83.675$ ergibt sich:

	90%-Konfidenzintervall (σ)	
	untere Grenze	obere Grenze
M	0.3395	0.4550
W	0.3612	0.4841

Bem.: Die Intervalle für °F und °C stehen in der Beziehung $I^{(C)} = 5/9 \cdot I^{(F)}$ zueinander.