Ue Statistik u. Wahrscheinlichkeitsth. f. Inf. 107.251 W 2004|5 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	18. Jän. 2005

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

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Zeigen Sie:

(a) $\overline{X}_n$ ist ein unverzerrter Schätzer für μ. Varianz des Schätzers? Ist der Schätzer konsistent für μ ?

(b) S_n^2 ist ein unverzerrter Schätzer für σ^2.
Hinweis: Zeigen Sie zunächst (*empirischer Verschiebungssatz*):

$$S_n^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}_n^2\right]$$

2. Die folgende Tabelle ist die Zusammenfassung einer Stichprobe der Größe 55 aus einer Poissonverteilung P_μ:

x	0	1	2	3	4	5
Häufigkeit	7	14	12	13	6	3

(a) Bestimmen Sie den plausiblen Schätzer von μ. (Ist er unverzerrt?)

(b) Bestimmen Sie den plausiblen Schätzwert von μ.

(c) Bestimmen Sie den plausiblen Schätzwert von $W(X = 2)$.

3. Die folgende Tabelle ist die Zusammenfassung einer Stichprobe der Größe 50 aus einer Binomialverteilung $B_{n,p}$ mit $n = 5$:

x	0	1	2	3	4	5
Häufigkeit	6	10	14	13	6	1

(a) Bestimmen Sie den plausiblen Schätzer von p. (Ist er unverzerrt?)

(b) Bestimmen Sie den plausiblen Schätzwert von p.

(c) Bestimmen Sie den plausiblen Schätzwert von $W(X \geq 3)$.

4. Die folgenden Daten sind Zeiten (Betriebsstunden) zwischen aufeinander folgenden Ausfällen eines Systems:

67, 102, 47, 85, 34, 9, 4, 21, 229, 262, 43, 152, 1, 14, 79

(a) Ermitteln und zeichnen Sie (mit der Hand oder mittels R) die empirische Verteilungsfunktion.

Hinweis für R-User: Nehmen Sie beispielsweise die Funktion `ecdf` aus dem Package `stats` (wird ab R–Version 2.0.0 automatisch beim Start geladen).

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

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

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

*(c) Zeichnen Sie über die empirische Verteilungsfunktion von (a) die mit Hilfe von (b) geschätzte Verteilungsfunktion. Bestimmen Sie graphisch Stelle und Wert des größten Abstands zwischen den beiden Funktionen.

5. $X_1, \ldots, X_n$ sei eine Stichprobe aus einer Alternativverteilung A_p. Die folgende sG ist eine approximative Pivotgröße (*Begründung?):

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

(a) Entwickeln Sie auf Basis von Z_n ein approximatives $(1-\alpha)$–Konfidenzintervall für den Parameter p.

(b) Bei einer Befragung (dafür/dagegen) zu einem Projekt waren 168 von 300 Personen dafür. Gibt es eine Mehrheit für das Projekt?

6. Ermitteln Sie auf Basis der Daten von **Tabelle 2.1** (vgl. **2. Blatt**), wahlweise für `GE=M` oder `GE=W`, ein 95%–Konfidenzintervall für die mittlere Körpergröße. Gehen Sie dabei von normalverteilten Beobachtungen aus.

Lösungen zum 10. Blatt

1. (a) $\mathbb{E}(\overline{X}_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{n\mu}{n} = \mu$

$$\mathrm{Var}(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Die Konsistenz von $\overline{X}_n$ (für μ) ist eine Konsequenz des Gesetzes der Großen Zahlen (oder der Tschebyscheff'schen Ungleichung): $\overline{X}_n \xrightarrow{W} \mu$

(b) Empirischer Verschiebungssatz:

$$\begin{aligned} S_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2\right) \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n}X_i^2 - 2\overline{X}_n \underbrace{\sum_{i=1}^{n}X_i}_{=n\overline{X}_n} + n\,\overline{X}_n^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n}X_i^2 - n\,\overline{X}_n^2\right] \end{aligned}$$

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

Verschiebungssatz:

$$\mathbb{E}(X_i^2) = \mathrm{Var}(X_i) + \mathbb{E}^2(X_i) = \sigma^2 + \mu^2$$
$$\mathbb{E}(\overline{X}_n^2) = \mathrm{Var}(\overline{X}_n) + \mathbb{E}^2(\overline{X}_n) = \frac{\sigma^2}{n} + \mu^2$$

$$\Longrightarrow \quad \mathbb{E}(S_n^2) = \frac{1}{n-1}\left[n\,(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right] = \frac{1}{n-1}\,(n-1)\,\sigma^2 = \sigma^2$$

2. (a) Plausibilitätsfunktion:

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

Log–Plausibilitätsfunktion:

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

Bestimmung des Maximums: $\frac{d\,l(\mu;\mathrm{Daten})}{d\mu} = \frac{n\overline{x}_n}{\mu} - n = 0 \Longrightarrow \widehat{\mu} = \overline{x}_n$

Plausibler Schätzer: $\widehat{\mu} = \overline{X}_n$

$\mathbb{E}(\overline{X}_n) = \mathbb{E}(X) = \mu \quad \Longrightarrow \quad$ unverzerrt

(b) Plausibler Schätzwert von μ: $\widehat{\mu} = \dfrac{0 \cdot 7 + 1 \cdot 14 + \cdots + 5 \cdot 3}{55} = \dfrac{116}{55} = 2.11$

(c) Plausibler Schätzwert von $W(X = 2)$: $\dfrac{\widehat{\mu}^2 \, e^{-\widehat{\mu}}}{2!} = 0.270$

3. (a) Plausibilitätsfunktion: $L(p; \text{Daten}) = \displaystyle\prod_{i=1}^{n} \binom{5}{x_i} p^{x_i} (1-p)^{5-x_i} =$

$$= \underbrace{\left[\prod_{i=1}^{n} \binom{5}{x_i}\right]}_{=C} p^{\sum_{i=1}^{n} x_i} (1-p)^{5n - \sum_{i=1}^{n} x_i} = C \, p^{n \overline{x}_n} (1-p)^{n(5 - \overline{x}_n)}$$

Log–Plausibilitätsfunktion:

$$l(p; \text{Daten}) = \ln L(p; \text{Daten}) = \ln C + n \overline{x}_n \ln p + n(5 - \overline{x}_n) \ln(1-p)$$

Bestimmung des Maximums:

$$\frac{d\, l(p; \text{Daten})}{dp} = \frac{n \, \overline{x}_n}{p} - \frac{n(5 - \overline{x}_n)}{1-p} = 0 \implies \widehat{p} = \frac{\overline{x}_n}{5}$$

Plausibler Schätzer: $\widehat{p} = \dfrac{\overline{X}_n}{5}$

$$\mathbb{E}\left(\frac{\overline{X}_n}{5}\right) = \frac{\mathbb{E}(X)}{5} = \frac{5p}{5} = p \quad \implies \quad \text{unverzerrt}$$

(b) Plausibler Schätzwert von p: $\widehat{p} = \dfrac{0 \cdot 6 + 1 \cdot 10 + \cdots + 5 \cdot 1}{5 \cdot 50} = \dfrac{106}{250} = 0.424$

(c) Plausibler Schätzwert von $W(X \geq 3)$: $\displaystyle\sum_{i=3}^{5} \binom{5}{i} \widehat{p}^{\,i} (1 - \widehat{p})^{5-i} = 0.360$

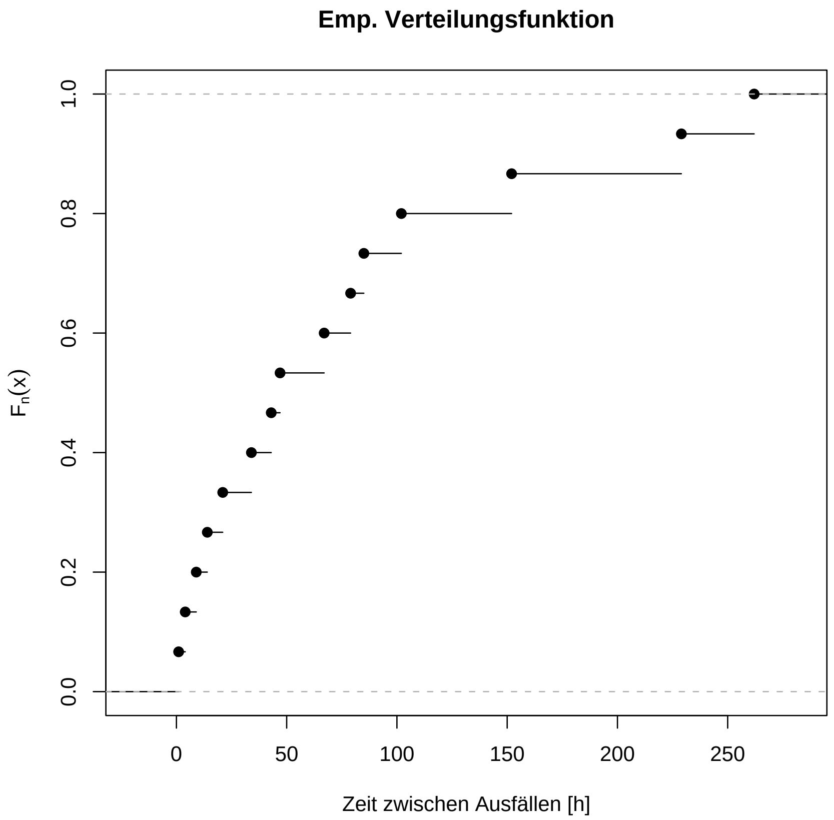
4. (a) Die empirische Verteilungsfunktion ist eine Treppenfunktion mit Sprüngen an den Stellen der (geordneten) Stichprobe; die Sprunghöhe beträgt jeweils $1/n$, bei mehreren identischen Beobachtungen (hier nicht der Fall) beträgt sie ein entsprechendes Vielfaches von $1/n$.

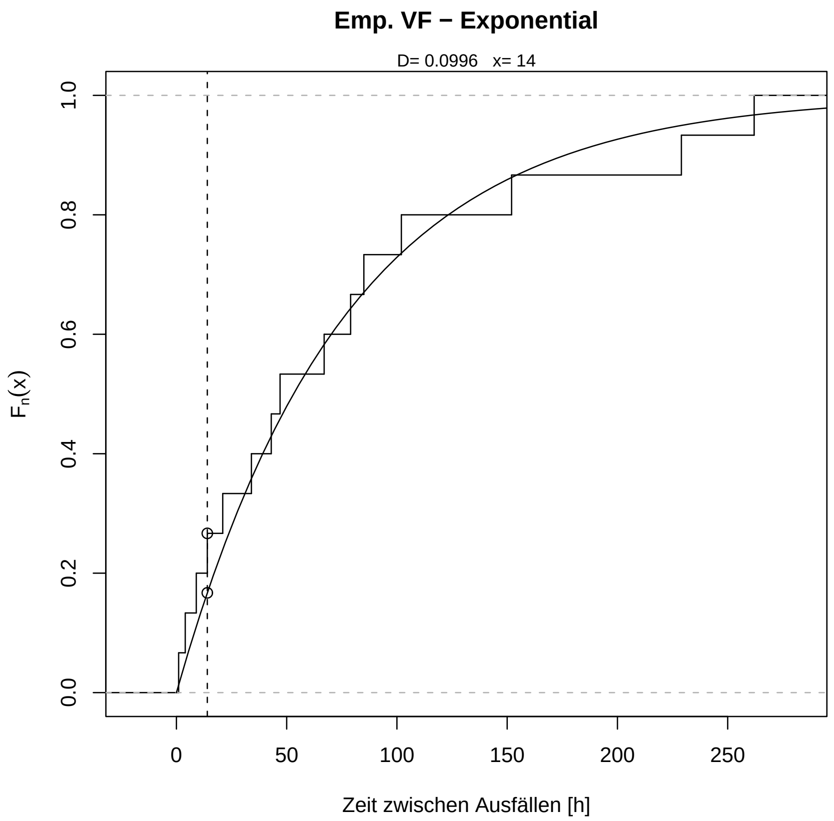
(b) Plausibilitätsfunktion: $L(\lambda; \text{Daten}) = \displaystyle\prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} = \lambda^n \, e^{-\lambda n \overline{x}_n}$

Log–Plausibilitätsfunktion: $l(\lambda; \text{Daten}) = \ln L(\lambda; \text{Daten}) = n \ln \lambda - \lambda n \overline{x}_n$

Bestimmung des Maximums: $\dfrac{d\, l(\lambda; \text{Daten})}{d\lambda} = \dfrac{n}{\lambda} - n \overline{x}_n = 0 \implies \widehat{\lambda} = \dfrac{1}{\overline{x}_n}$

Überprüfung auf (lok.) Maximum: $\left.\dfrac{d^2\, l(\lambda; \text{Daten})}{d\lambda^2}\right|_{\lambda = \widehat{\lambda}} = -n \overline{x}_n^2 < 0$





Plausibler Schätzer: $\widehat{\lambda} = \dfrac{1}{\overline{X}_n}$

Plausibler Schätzwert: $\widehat{\lambda} = \dfrac{15}{67 + 102 + \cdots + 79} = 0.0131$

*(c) Plausibel geschätzte Verteilungsfunktion: $\widehat{F}(x) = 1 - e^{-\widehat{\lambda}x}, \; x > 0$

Das Supremum D des Abstands zwischen $F_n(x)$ und $\widehat{F}(x)$ kann nur an einer Sprungstelle von $F_n(x)$ auftreten: $D \approx 0.1$ bei $x_{(4)} = 14$.

5. * Die Begründung dafür, daß Z_n eine approximative Pivotgröße ist, liegt im Zentralen Grenzverteilungssatz, 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.

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

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

Auflösen der Doppelungleichung nach p:

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

(b) $\overline{x}_n = \dfrac{168}{300} = 0.56$

	Konfidenzintervall	
ÜW	untere Grenze	obere Grenze
90%	0.511	0.608
95%	0.502	0.617
99%	0.484	0.633

Die Situation ist nicht ganz eindeutig: Bei der üblichen ÜW von 95% liegt 0.5 zwar (knapp) außerhalb des Intervalls („Mehrheit dafür“), bei 99% sind allerdings auch Werte von p unter 0.5 nicht auszuschließen.

6. $(1-\alpha)$–Konfidenzintervall für den Mittelwert (bei Stichproben aus Normalverteilungen):

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

					Konfidenzintervall	
GE	n	$\overline{x}_n$	s_n	$t_{n-1;0.975}$	untere Grenze	obere Grenze
M	39	182.51	6.54	2.02	180.39	184.63
W	30	168.27	5.23	2.05	166.32	170.22