

Modellgestützte Analyse und Optimierung (SS 2008)

Übungsblatt 4

Aufgabe 4.1:

(3 Punkte)

Sei x_1, \dots, x_n eine Stichprobe mit n unabhängigen Beobachtungen einer Zufallsvariable X . Gesucht sind die Schätzwerte $\hat{\mu}$ und $\hat{\sigma}$ für die Parameter μ und σ der Normalverteilung $N(\mu, \sigma)$, so dass $N(\hat{\mu}, \hat{\sigma})$ „möglichst gut“ an die Stichprobe angepasst ist. Bestimmen sie $\hat{\mu}$ und $\hat{\sigma}$

- mit der *Momentenmethode*
- mit der *Maximum-Likelihood-Methode*.

(Hinweis zu b): Maximieren sie die log-Likelihoodfunktion indem sie die (partiellen) Ableitungen nach μ und σ betrachten.)

Aufgabe 4.2:

(6 Punkte)

Folgende Zwischenankunftszeiten von Bauteilen an einer Maschine wurden gemessen:

10.478	7.268	2.775	0.381	0.979	1.226	1.392	4.706
0.003	5.004	2.835	2.441	1.838	10.780	0.010	4.732
3.042	8.817	0.670	1.693	11.576	2.469	9.905	2.265
1.184	0.093	5.187	1.828	7.140	0.136	1.408	6.349
3.717	7.170	0.609	3.084	0.122	3.154	3.574	0.420

- Zeichnen sie ein *Histogramm* für diese Stichprobe. Unterteilen sie dazu das Intervall $[0,12]$ in 6 Teilbereiche.
- Die Form des Histogramms lässt vermuten, dass die beobachteten Zwischenankunftszeiten gut durch eine Exponentialverteilung beschrieben werden können. Bestimmen sie den Parameter λ der Exponentialverteilung, so dass diese bestmöglich an die Stichprobe angepasst wird und zeichnen sie die Dichtefunktion (richtig skaliert) über ihr Histogramm.
- Bestimmen sie zum Test der Anpassungsgüte die beiden Abstandsmaße D und D' wie auf Folie 4.21 „Bestimmung der Anpassgüte“ der Vorlesung beschrieben.
- Zeichnen sie den *Q-Q-Plot* und den *P-P-Plot*.
- Führen sie den *Chi-Quadrat Test* wie auf den Folien 4.24ff der Vorlesung beschrieben durch. Unterteilen sie dazu die angepasste theoretische Verteilungsfunktion in $k=4$ Intervalle, so dass Werte aus den Intervallen mit gleicher Wahrscheinlichkeit $p_i = 1/k$, $i=1, \dots, k$, angenommen werden. Berechnen sie die Teststatistik und vergleichen sie das Ergebnis mit den kritischen Werten der χ^2 -Verteilung (siehe Tabelle 1). Wie müssen die Freiheitsgrade gewählt werden? Akzeptieren sie die Exponentialverteilung als gute Approximation der Stichprobe oder lehnen sie sie ab?
- Führen sie den *Kolmogorov-Smirnov-Test* wie auf den Folien 4.27ff der Vorlesung beschrieben durch. Berechnen sie zuerst die Teststatistik D_n . Vergleichen sie dann die angepasste Teststatistik (siehe Tabelle 2) mit den kritischen Werten aus Tabelle 2. Akzeptieren sie die Exponentialverteilung als gute Approximation der Stichprobe oder lehnen sie sie ab?

(Hinweis zu d): Der Q-Q-Plot bzw. P-P-Plot ist die Darstellung der Werte $(y_i, F_t^{-1}(F_e(y_i)))$ bzw. $(F_e(y_i), F_t(y_i))$, wobei $F_e(y)$ die empirische und $F_t(y)$ die theoretische Verteilungsfunktionen ist.)

v	$\chi_{0.005}^2$	$\chi_{0.01}^2$	$\chi_{0.025}^2$	$\chi_{0.05}^2$	$\chi_{0.1}^2$
1	7.88	6.63	5.02	3.84	2.71
2	10.60	9.21	7.38	5.99	4.61
3	12.84	11.34	9.35	7.81	6.25
4	14.96	13.28	11.14	9.49	7.78

Tabelle 1. Kritische Werte χ_{α}^2 der χ^2 -Verteilung mit v Freiheitsgraden

Fall	Angepasste Teststatistik	1- α				
		0.850	0.900	0.950	0.975	0.990
Alle Verteilungsparameter bekannt	$\left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right) \cdot D_n$	1.138	1.224	1.358	1.480	1.628
Exponentialverteilung mit Parameter λ aus Stichprobe geschätzt	$\left(D_n - \frac{0.2}{n}\right) \cdot \left(\sqrt{n} + 0.26 + \frac{0.5}{\sqrt{n}}\right)$	0.926	0.990	1.094	1.190	1.308

Tabelle 2. Kritische Werte für angepasste Kolmogorov-Smirnov Teststatistiken

Aufgabe 4.3:

(3 Punkte)

Auf der Homepage der Übungen zu dieser Vorlesung finden sie das Archiv `SampleData.zip`, welches vier Dateien mit Messdaten enthält. Nutzen sie den *Arena-Input-Analyzer* um jeweils eine theoretische Verteilung zu bestimmen, welche die Daten bestmöglich approximiert. Den Input-Analyzer können sie aus dem Tools-Menu von Arena starten. Informationen zur Bedienung des Input-Analyzers finden sie in den `Arena-Slides2.pdf` auf der Homepage der Übungen zur Vorlesung oder auch im Arena Hilfe-Menu.

- Nennen sie für jede der vier Dateien den Verteilungstyp sowie die Parameter der Verteilung die Arena als bestmögliche Anpassung an die Stichprobe findet.
- Ändern sie die Anzahl der Intervalle im Histogramm und bestimmen sie erneut die bestmögliche Verteilung. Erklären sie warum (manchmal) die bestmögliche Verteilung, die der Input-Analyzer auswählt, abhängig ist von der Anzahl der Intervalle im Histogramm? Sind auch die Parameter der bestmöglichen Verteilung abhängig von der Anzahl Intervalle im Histogramm? Begründen sie ihre Antwort.