Fachbereich Mathematik
Prof. Dr. W. Stannat
B. Debrabant

TECHNISCHE UNIVERSITÄT DARMSTADT

17./18. Januar 2008

Statistik 1 für WInf, WI(MB), WI(ET), WI(BI), Übung 6

Gruppenübung

G 18 Maximum-Likelihood-Methode

Aus Erfahrung sei bekannt, dass die störungsfreie Betriebsdauer (in 1000 Stunden) eines bestimmten Systems durch eine stetig verteilte Zufallsvariable X mit der Dichte

$$f(x) = \begin{cases} 2\theta x e^{-\theta \cdot x^2} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

mit einem unbekannten Parameter $\theta > 0$ beschrieben werden kann. Bestimmen Sie aus den folgenden 20 Betriebsdauern (in 1000 Stunden) den Maximum-Likelihood-Schätzwert für $\tau(\theta) = \theta$.

1.530	1.173	1.832	1.075	1.539
0.998	2.083	0.693	2.529	1.693
1.325	1.487	1.298	1.743	1.432
1.369	0.987	2.222	1.818	1.505

G 19 Konfidenzintervalle

In einer Stadt liegen für 161 Jahre Niederschlagsmessungen im Monat April vor. Die Meßreihe $x_1, x_2, \ldots, x_{161}$ (x_i = Niederschlagshöhe in mm im i-ten Jahr) hat das arithmetische Mittel $\bar{x}_{(161)} = 53.68$ und die empirische Standardabweichung $s_{(161)} = 6.13$. Es wird angenommen, dass die Werte $x_1, x_2, \ldots, x_{161}$ eine Realisierung von 161 unabhängigen, identisch $N(\mu, \sigma^2)$ - verteilten Zufallsvariablen sind. Bestimmen Sie mit Konfidenzschätzverfahren zum Niveau $1 - \alpha = 0.95$ je ein konkretes Schätzintervall

a) für μ,

b) für σ^2,

c) für μ unter der Voraussetzung $\sigma^2 = 6.13^2$.

G 20 Bisher ist der Betreiber des öffentlichen Verkehrsnetzes in einer Großstadt davon ausgegangen, dass 35% der Fahrgäste Zeitkarteninhaber sind. Bei einer Fahrgastbefragung geben 112 der insgesamt 350 Befragten an, dass sie eine Zeitkarte benutzen. Testen Sie zum Niveau $\alpha = 0.05$, ob sich der Anteil der Zeitkarteninhaber verändert hat. Formulieren Sie die Fragestellung zunächst als statistisches Testproblem.

G 21 Es wird angenommen, dass die vorliegende Messreihe $x_1, \ldots, x_{25}$ (s. Tabelle) eine Realisierung von 25 unabhängigen $N(\mu, \sigma^2)$-verteilten Zufallsvariablen ist.

i	1	2	3	4	5	6	7	8	9	10	11	12	13
x_i	704	695	703	693	698	694	701	706	705	697	701	705	693
i	14	15	16	17	18	19	20	21	22	23	24	25	
x_i	702	686	696	703	693	712	706	693	702	699	693	695	

Für die Messreihe gilt $\sum_{i=1}^{25} x_i = 17475$ und $\sum_{i=1}^{25} x_i^2 = 12\,215\,867$.

a) Überprüfen Sie die Hypothese $H_0 : \mu \geq 700$ gegen $H_1 : \mu < 700$ mit einem geeigneten Test zum Niveau $\alpha = 0.05$.

b) Es sei nun vorausgesetzt, dass für obige Messwerte $\sigma^2 = 25$ bekannt ist. Überprüfen Sie nun mit einem anderen Test dieselbe Nullhypothese wie in Aufgabenteil a) zum Niveau $\alpha = 0.05$.

Hausübung

H 23 Maximum-Likelihood-Methode

Ein Tierpark besitzt sieben Exemplare einer seltenen Tierart. In einem biologischen Forschungsinstitut wurde eine bisher unbekannte, nicht ansteckende Krankheit an Tieren dieser Rasse entdeckt. Um herauszufinden, wie viele seiner Tiere von der Krankheit befallen sind, lässt der Leiter des Tierparks an drei aufeinander folgenden Tagen je ein Tier fangen und auf die Krankheit hin untersuchen, um es anschließend wieder ins Gehege zu entlassen. Die Zufallsvariable X_i nehme den Wert 1 an, falls das am i-ten Tag gefangene Tier für krank befunden wird und sonst den Wert 0 ($i = 1, \dots, 3$). Die unbekannte Anzahl der kranken Tiere sei θ.

a) Bestimmen Sie die Verteilung der Zufallsvariablen X_1, indem Sie die Wahrscheinlichkeiten $P_\theta(X_1 = 1)$ und $P_\theta(X_1 = 0)$ angeben.

b) Die Krankheit wurde lediglich bei den an den ersten beiden Tagen untersuchten Tieren diagnostiziert. Geben Sie auf der Basis dieser Beobachtung an, welche Werte für θ in Frage kommen und welcher dieser Werte am plausibelsten für die Anzahl an kranken Tieren ist.

H 24 Maximum-Likelihood-Schätzer, verschobene Exponentialverteilung

Die Zufallsvariablen $X_1, \dots, X_n$ seien unabhängig und identisch verteilt mit der Verteilungsfunktion

$$F_\theta(x) = \begin{cases} 1 - e^{-(x-\theta)} & \text{falls } x \geq \theta \\ 0 & \text{sonst.} \end{cases}$$

a) Bestimmen Sie den Maximum-Likelihood-Schätzer T_n für $\tau(\theta) = \theta$.

b) Geben Sie die Verteilungsfunktion von T_n an.

c) Bestimmen Sie den Bias und den mittleren quadratischen Fehler von T_n.

d) Ist die Folge der Schätzer $T_1, T_2, \dots$ asymptotisch erwartungstreu für $\tau(\theta) = \theta$?

Hinweis: Bestimmen Sie zur Bearbeitung der Teilaufgaben c) und d) zunächst die Verteilung der Zufallsvariablen $T_n - \theta$.

H 25 Konfidenzintervalle

Im Andenhochland wurde eine neue Sorte Tomaten gefunden. Um deren Ertrag zu bestimmen, wurde bei 70 Pflanzen der Ernteertrag in *kg* gemessen. Es ergab sich das arithmetische Mittel $\bar{x} = 3.73$ und die Stichprobenvarianz $s^2 = 1.27$.

a) Unter geeigneten Annahmen bestimme man ein Konfidenzintervall zum Niveau 0.99 für den Erwartungswert m und eines für die Varianz σ^2.

Ein Teil der Pflanzen ist von einer Krankheit befallen, was die Tomaten ungenießbar macht. Bei den 70 untersuchten Tomatenstöcken waren 14 von dieser Krankheit befallen.

b) Unter geeigneten Annahmen bestimme man ein (approximatives) Konfidenzintervall zum Niveau 0.95 für die Wahrscheinlichkeit p, dass eine Pflanze von dieser Krankheit befallen ist.

H 26

a) Ein Hersteller von Präzisionswaagen gibt an, dass die Messgenauigkeit einer neu entwickelten, besonders preisgünstigen Waage nicht schlechter sei als die des bisher gelieferten und bewährten Modells. Zur Überprüfung dieser Angabe führt ein potentieller Käufer mit einer dieser neuen Waagen 30 Messungen eines Gewichtes durch. Aus der entstandenen Messreihe $x_1, \dots, x_{30}$ wurde die empirische Varianz $s^2 = 2.07 \cdot 10^{-11}$ berechnet. Aufgrund langjähriger Praxis ist bekannt, dass sich die Wiegeergebnisse bei der bisher verwendeten Waage durch eine Zufallsvariable mit der Varianz $\sigma_0^2 = 1.5 \cdot 10^{-11}$ beschreiben lassen. Versuchen Sie, die Angabe des Herstellers mit einem geeigneten Test zum Niveau 0.05 zu widerlegen. Nehmen Sie dabei an, dass $x_1, \dots, x_{30}$ Realisierungen unabhängiger identisch normalverteilter Zufallsvariablen sind. Als Maß für die Messgenauigkeit ist die Varianz zu nehmen.

b) Wie würde die Überprüfung ausfallen, wenn sich die empirische Varianz $s^2 = 2.07 \cdot 10^{-11}$ auf der Grundlage von 100 Messwerten $x_1, \dots, x_{100}$ ergeben hätte?

Eventuell benötigte Quantile : $\chi^2_{29;0.95} = 42.557$ $\chi^2_{99;0.95} = 123.225$

H 27

Gegeben sei die folgende Messreihe $x_1, \dots, x_{16}$:

0.534	0.545	0.283	0.445	0.519	0.513	0.464	0.499
0.605	0.526	0.568	0.427	0.546	0.527	0.566	0.597

mit $\sum_{i=1}^{16} x_i = 8.164$ und $\sum_{i=1}^{16} x_i^2 = 4.257386$.

Dabei wird angenommen, dass $x_1, \dots, x_{16}$ Realisierungen von 16 unabhängigen, identisch $N(\mu, 0.01)$-verteilten Zufallsvariablen sind.

a) Überprüfen Sie die Hypothese $H_0 : \mu_0 = 0.5$ gegen die Alternative $H_1 : \mu \neq 0.5$ mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.05$.

b) Betrachten Sie nun die Hypothese $H_0 : \mu \geq 0.5$ gegen die Alternative $H_1 : \mu < 0.5$. Für welche Werte von μ wird die Wahrscheinlichkeit für einen Fehler 2. Art kleiner als 10 % auf gleichem Signifikanzniveau $\alpha = 0.05$?