Übungsblatt 5 zur Vorlesung

"Statistische Methoden"

Testtheorie: $\theta \leq \theta_0$ vs $\theta > \theta_0$, MLQ, UMP, exponentielle Familie

Herausgabe des Übungsblattes: Woche 13, Abgabe der Lösungen: Woche 14 (bis Freitag, 1615 Uhr), Besprechung: Woche 15

Standard

Aufgabe 20 [exponentielle Familie, MLQ und minimal-suffiziente Statistik] [1 Punkt]

Zeigen Sie, dass die Gamma-Verteilung (1.4.2.3) und die Binomialverteilung (1.4.1.2) zur exponentiellen Familie gehören und berechnen Sie natürliche Parameter und minimal suffiziente Statistiken dazu. Tipp: n ist jeweils *nicht* selber Parameter und kann als bekannte Zahl aufgefasst werden.

Aufgabe 21 [Satz 4.1 bzw Satz 4.9] [4+1 Punkte]

Ein Hersteller von Glühbirnen behauptet, die von Ihnen produzierten Glühbirnen hätten eine durchschnittliche Lebensdauer von 1000 Stunden. Eine Konsumentenschutzorganisation bezweifelt dies. Bevor sie an die Öffentlichkeit geht, will sie aber mit dem Hersteller zusammen eine zufällige Stichprobe vom Umfang 2000 ausgiebig testen (brennen lassen bis kaputt). Man einigt sich darauf, davon auszugehen, dass die Glühbirnen unabhängig voneinander brennen und die Lebensdauer exponentialverteilt modelliert werden kann.

a) Entwickeln Sie mit Hilfe des Lemmas von Neyman-Pearson einen Test, indem Sie vorerst davon ausgehen, dass $\lambda_0 = 1/1000$ und $\lambda_1 = 1/950$ (Sie werden sehen, dass Sie λ_1 nie wirklich brauchen). Nehmen Sie $\alpha = 0.05$ und berechnen Sie das K, genauer das K'.

Tipps: Beispiel 1 aus 4.1.2, 1.4.2.3 und qgamma(0.05,2000,0.001).

b) In der darauffolgenden Untersuchung erhielt man eine durchschnittliche Brenndauer von 967.5 Stunden. Was raten Sie als statistischer Consultant der Konsumentenschutzorganisation?

Aufgabe 22 [NP-Lemma im diskreten Fall (Satz 4.2 bzw Satz 4.9)] [4 + 1 Punkte]

Herr Meier besucht einen Banker in der Bahnhofstrasse in Zürich. Herr Meier sagt, dass er in 60 % der Fälle voraussagen kann, ob der CHF / $-Kurs morgen höher oder tiefer liegt als heute (gleichen Kurs schliessen wir mal aus). Der Banker will Herrn Meier während 10 Handelstagen testen, bevor er ihm die Verantwortung für das Devisengeschäft überträgt. Für den Banker kann man gerade so gut eine Münze werfen, um zu prognostizieren, ob der Kurs morgen höher oder tiefer liegt. Der Banker versteht was von Statistik und wird auf dem 5 % - Niveau einen Test durchführen.

a) Wie wird dieser Test voraussichtlich aussehen? Sie werden die Befehle pbinom(7,10,0.5) und pbinom(8,10,0.5) brauchen. Tipp: Gleichung (4.2).

b) Herr Meier hat noch einen Bruder. Der sagt in genau 20 % der Fälle korrekt voraus, ob der Kurs sinkt oder steigt. Angenommen er kann das wirklich. Wie kann der Banker den Bruder geschickt einsetzen? Die Lösung dieses Problems ist nicht nur eine mathematische Spielerei, sondern ein praktisches statistisches Prinzip.

Honours

Aufgabe 23 [Nehmen Krankheitsfälle signifikant zu?] [1+1+2 Punkte]

Zur Modellierung von Krankheitsfällen (z.B. Creutzfeldt-Jakob CJD) pro Jahr in einem Land kann man zum Beispiel eine Poisson-Zufallsgrösse (vgl 1.4.1.5) einsetzen. In der Vorlesung Angewandte Stochastik werden wir sehen, dass dies nicht nur gut zu realen Daten passt, sondern auch aus theoretischen Gründen sinnvoll ist. Solche Übereinstimmung (praktisch passend und theoretisch fundiert) ist immer sehr wertvoll. Ansonsten hat man eine ad hoc Anpassung eines Modells an einen konkreten Datensatz - wenn wir einen neuen Datensatz erhalten, stimmt das Modell eventuell überhaupt nicht mehr, man spricht deshalb von "ad hoc"-erie.

Wir werden jetzt die Anzahl N_j von (gemeldeten) Krankheitsfällen in Jahr $j, 1 \leq j \leq n$, mit unabhängigen poissonverteilten Zufallsgrössen modellieren. Dabei sei der Parameter in Jahr j gleich θ^j (Potenz, "hoch j", nicht Index).

a) Was ist hier die minimal suffiziente Statistik der gemeinsamen Wahrscheinlichkeitsfunktion für θ über alle n Jahre?

b) Ist MLQ erfüllt ($\theta_0 = 1$ vs $\theta_1 > 1$ beliebig)?

c) Geben Sie einen UMP-Test der Hypothesen $\theta = 1$ vs $\theta > 1$ an (konkrete Zahlen nicht ausrechnen). Was raten Sie als statistischer Consultant der Gesundheitsbehörde, wenn Sie die Alternativ-Hypothese annehmen müssen (was ist dann konkret los)?

Technisches Detail: die gleiche Verteilung der Zufallsgrössen der verschiedenen Jahre wird nicht gefordert. Die bisherige Theorie kann trotzdem eingesetzt werden (freiwillige HA: gehen Sie dazu die bisherige Theorie durch).

Bemerkung zur Modellierung: Wir haben Unabhängigkeit der Anzahl Fälle pro Jahr gefordert. Das heisst unter anderem, dass wenn wir in Jahr i massiv mehr als die erwarteten θ^i Fälle haben, so heisst dies keineswegs, dass wir in Jahr $i+1$ ebenfalls massiv mehr als die erwarteten θ^{i+1} Fälle haben sollten. Damit eignet sich dieses Modell eindeutig nicht für ansteckende Krankheiten wie SARS; wenn wir dort in Woche i mehr als die ursprünglich erwarteten Fälle haben, so werden wir wohl auch in Woche $i+1$ mehr als die erwarteten Fälle haben, weil der "Überschuss" von Woche i auch fleissig "Nachkommen" produziert. CJD ist keine ansteckende Krankheit. Dass wir geometrisches Wachstum der erwarteten Fälle haben, kann hinterfragt werden - geometrisches Wachstum wäre bei ansteckenden Krankheiten ohne (oder mit ungenügenden) Gegenmassnahmen in der Anfangsphase eher angebracht.