Dr. Christof Luchsinger

Übungsblatt 2 zur Vorlesung

"Statistische Methoden"

R/S-PLUS, CLT und Grundlagen der Statistik

Herausgabe des Übungsblattes: Woche 09, Abgabe der Lösungen: Woche 10 (bis Freitag, 1615 Uhr), Besprechung: Woche 12

Must

Aufgabe 7 [individuelle Einführung Statistik-Paket]

Kapitel 1 und 2 im Dalgaard durcharbeiten

Aufgabe 8 [einfache Berechnungen I]

Berechnen Sie in R/S-PLUS folgende Wahrscheinlichkeiten (ohne in Tabellen aus Büchern nachzuschlagen):

a) $P[\mathcal{N}(0,1) > 3]$ (das ist kurz für X sei standardnormalverteilt; $P[X > 3]$)

b) $P[\mathcal{N}(35,36) > 42]$

c) Wahrscheinlichkeit für 10 Erfolge bei 10 unabhängigen Versuchen mit Erfolgswahrscheinlichkeit 0.8.

d) $P[\chi_2^2 > 6.5]$

Aufgabe 9 [einfache Berechnungen II]

Bei einer Normalverteilung gilt approximativ die Regel, dass 5 % der Realisationen ausserhalb von 2 Standardabweichungen um den Mittelwert liegen.

a) Wie ist die genaue Schranke; d.h. finde a: $P[|\mathcal{N}(0,\sigma^2)| > a\sigma] = 0.05$.

b) finde b: $P[|\mathcal{N}(0,\sigma^2)| > b\sigma] = 0.01$.

c) finde d: $P[|\mathcal{N}(0,\sigma^2)| > d\sigma] = 0.001$.

d) finde u: $P[|\mathcal{N}(0,\sigma^2)| > u\sigma] = 0.1$.

Standard

Aufgabe 10 [CLT] [3+3 Punkte]

Unser WTS-Theorem 5.4 [Zentraler Grenzwertsatz] setzt voraus, dass die Zufallsgrössen "iid" sind (unabhängig, identisch verteilt). Diese strenge Forderung kann in Verallgemeinerungen dieses Satzes gelockert werden. Wir wollen zeigen, dass nachfolgende Verallgemeinerung zu weit geht - also falsch ist. Konstruieren Sie dazu ein - einfach aufgebautes - Gegenbeispiel, dass zeigt, dass der folgende Satz so nicht richtig sein kann:

Sei $X_k, k \geq 1$, eine Folge von unabhängigen Zufallsgrössen mit gleichem Erwartungswert μ und Varianzen $0 < V[X_k] =: \sigma_k^2 < \infty$ für alle $k \geq 1$. Dann gilt für $a \in \mathbb{R}$ beliebig:

$$P\left[\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{\sum_{k=1}^n \sigma_k^2}} \leq a\right] \longrightarrow P[\mathcal{N}(0,1) \leq a]$$

wenn $n \to \infty$.

a) 3 Punkte für Konstruktion des Gegenbeispiels
b) 3 Punkte für mathematisch exakten Beweis (zum Beispiel an Hand des Gegenbeispiel's aus a)), dass obiger Satz falsch ist.

Aufgabe 11 [Aktionsraum, Entscheidungsfunktion, Verlustfunktion & Risiko] [1+1+3 Punkte]

Sei $(X)_{i=1}^n$ eine Folge von iid Be(p)-verteilten Zufallsgrössen, $p \in [0,1]$. Sie möchten p schätzen. Dazu haben Sie eine Realisation $(x_1, \ldots, x_n)$.

a) Geben Sie den Aktionsraum $\mathcal{A}$ für dieses Problem an.

b) Geben Sie eine sinnvolle Entscheidungsfunktion an.

c) Berechnen Sie bei quadratischer Verlustfunktion das Risiko obiger Entscheidungsfunktion.

Übungsblatt 2 zur Vorlesung "Statistische Methoden"

Olivier Warin

20. Februar 2011

Aufgabe 8 [einfache Berechungen I]

a) Es gilt:

$$P[\mathcal{N}(0,1) > 3] \underset{\text{R}}{\doteq} \texttt{pnorm(3,lower.tail=FALSE)} \underset{\text{R}}{\doteq} 0.001349898.$$

b) Analog berechnen wir

$$P[\mathcal{N}(35,36) > 42] \underset{\text{R}}{\doteq} \texttt{pnorm(42,35,sqrt(36),lower.tail=FALSE)} \underset{\text{R}}{\doteq} 0.1216725.$$

c) Die Anzahl Erfolge bei 10 unabhängigen Versuchen mit Erfolgswahrscheinlichkeit 0.8 ist Bin(10, 0.8) verteilt. Also können wir die Wahrscheinlichkeit für 10 Erfolge wie folgt berechnen:

$$P[\text{Bin}(10,0.8) \geqslant 10] = P[\text{Bin}(10,0.8) > 9] \underset{\text{R}}{\doteq} \texttt{pbinom(9,10,0.8,lower.tail=FALSE)} \underset{\text{R}}{\doteq} 0.1073742.$$

Mit der Rechnung

$$0.8^{10} \underset{\text{R}}{\doteq} \texttt{0.8**10} \underset{\text{R}}{\doteq} 0.1073742$$

erhält man natürlich dasselbe Resultat.

d) Auch die χ^2_2-Verteilung ist in R tabelliert. Zum Beispiel können wir berechnen:

$$P[\chi^2_2 > 6.5] \underset{\text{R}}{\doteq} \texttt{pchisq(6.5,2,lower.tail=FALSE)} \underset{\text{R}}{\doteq} 0.03877421.$$

Aufgabe 9 [einfache Berechnungen II]

Es sei σ eine positive reelle Zahl. In den folgenden Teilaufgaben müssen wir jeweils eine positive, reelle Zahl x bestimmen, so dass

$$P[|\mathcal{N}(0,\sigma^2)| > x\sigma] = \alpha,$$

für ein gewisses $\alpha \in \mathbb{R}$ gilt. Dies bereiten wir hier zuerst allgemein vor:

$$\begin{aligned}\alpha = P[|\mathcal{N}(0,\sigma^2)| > x\sigma] &= P[\mathcal{N}(0,\sigma^2) > x\sigma] + P[\mathcal{N}(0,\sigma^2) < -x\sigma] \underset{\text{metrie}}{\overset{\text{Sym-}}{=}} 2P[\mathcal{N}(0,\sigma^2) > x\sigma] \\ &= 2P\left[\frac{\mathcal{N}(0,\sigma)-0}{\sigma} > \frac{x\sigma - 0}{\sigma}\right] \underset{\text{formation}}{\overset{\text{Z-Trans-}}{=}} 2P[\mathcal{N}(0,1) > x].\end{aligned}$$

Wir schliessen also $P[\mathcal{N}(0,1) > x] = \alpha/2$.

a) Sei jetzt $a \in \mathbb{R}$ mit $0.05 = P[|\mathcal{N}(0,\sigma^2)| > a\sigma]$. Mit obiger Vorbereitung können wir direkt schliessen:

$$a \underset{\text{R}}{\doteq} \texttt{qnorm(0.05/2,lower.tail=FALSE)} \underset{\text{R}}{\doteq} 1.959964.$$

b) Nun suchen wir ein $b \in \mathbb{R}$ mit $0.01 = P[|\mathcal{N}(0,\sigma^2)| > b\sigma]$. Analog wie oben finden wir:

$$b \underset{\text{R}}{\doteq} \texttt{qnorm(0.01/2,lower.tail=FALSE)} \underset{\text{R}}{\doteq} 2.575829.$$

c) Hier geht es um ein $d \in \mathbb{R}$ mit $0.001 = P[|\mathcal{N}(0,\sigma^2)| > d\sigma]$. Wie zuvor folgt

$$d \underset{\text{R}}{\doteq} \texttt{qnorm(0.001/2,lower.tail=FALSE)} \underset{\text{R}}{\doteq} 3.290527.$$

d) Zum Schluss suchen wir noch ein $u \in \mathbb{R}$ mit $0.1 = P[|\mathcal{N}(0,\sigma^2)| > u\sigma]$. Wiederum schliessen wir wie in den vorigen Teilaufgaben :

$$u \underset{\mathrm{R}}{\doteq} \texttt{qnorm(0.1/2,lower.tail=FALSE)} \underset{\mathrm{R}}{\doteq} 1.644854.$$

Aufgabe 10 [CLT]
Sei X_k, $k \geqslant 1$ eine Folge von unabhängigen Zufallsgrössen mit gleichem Erwartungswert μ und Varianzen $0 < V[X_k] =: \sigma_k^2 < \infty$ für alle $k \geqslant 1$.

Behauptung: Im Allgemeinen gilt nun *nicht*, dass für jedes $a \in \mathbb{R}$ gilt:

$$P\left[\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{\sum_{k=1}^n \sigma_k^2}} \leqslant a\right] \overset{n\to\infty}{\longrightarrow} P[\mathcal{N}(0,1) \leqslant a].$$

Beweis: Wir beweisen die Behauptung natürlich mit einem Gegenbeispiel: Dazu sei X_k, $k \geqslant 1$ eine Folge von unabhängigen Zufallsgrössen mit folgender Verteilung:

$$P[X_k = {}^1/_{k^2}] \;=\; P[X_k = -{}^1/_{k^2}] \;=\; 0.5 \;(k \geqslant 1).$$

Es folgt sofort (für $k \geqslant 1$):

$$\begin{aligned} E[X_k] &= 0.5 \cdot {}^1/_{k^2} + 0.5 \cdot (-{}^1/_{k^2}) \;=\; 0 \;=:\; \mu \\ V[X_k] &= 0.5 \cdot ({}^1/_{k^2})^2 + 0.5 \cdot (-{}^1/_{k^2})^2 \;=\; {}^1/_{k^4} \;=:\; \sigma_k^2. \end{aligned}$$

Die geforderten Bedingungen sind also erfüllt.
Für eine natürliche Zahl n gilt aber

$$P\left[\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{\sum_{k=1}^n \sigma_k^2}} \leqslant 2\right] = P\left[\sum_{k=1}^n X_k \leqslant 2\sqrt{\sum_{k=1}^n \frac{1}{k^4}}\,\right] \;\geqslant\; P\left[\sum_{k=1}^n X_k \leqslant 2\right] \;=\; 1,$$

Denn es gilt klar für fast alle $\omega \in \Omega$:

$$\sum_{k=1}^n X_k(\omega) \;\leqslant\; \sum_{k=1}^n \frac{1}{k^2} \;\leqslant\; 2.$$

Wäre die Aussage nun korrekt, so müsste gelten

$$1 \neq P\left[\mathcal{N}(0,1) \leqslant 2\right] \;=\; \lim_{n\to\infty} P\left[\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{\sum_{k=1}^n \sigma_k^2}} \leqslant 2\right] \;=\; \lim_{n\to\infty} 1 \;=\; 1,$$

was natürlich falsch ist. Damit haben wir den gesuchten Widerspruch gefunden.

$\blacksquare$

Aufgabe 11 [Aktionsraum, Entscheidungsfunktion, Verlustfunktion & Risiko]
Sei $(X_i)_{i=1}^n$ eine Folge von iid Be(p)-verteilten Zufallsgrössen, wobei $p \in [0,1]$. Wir möchten nun p schätzen. Dazu haben wir eine Realisation $(x_1,\cdots,x_n) \in \{0,1\}^n$.

a) Der Aktionsraum $\mathcal{A}$ lautet hier wie folgt:

$$\mathcal{A} = [0,1].$$

b) Wir haben in der WTS-Vorlesung bereits gesehen, dass $\overline{x}$ ein erwartungstreuer und konsitenter Schätzer für $E[X_1] = p$ ist. Somit ist

$$\begin{array}{rccc} d: & \{0,1\}^n & \to & \mathcal{A} = [0,1] \\ & (x_1,\cdots,x_n) & \mapsto & d(x_1,\cdots,x_n) \;=\; \overline{x} \end{array}$$

eine sinnvolle Entscheidungsfunktion.

c) Für diese Entscheidungsfuktion berechnen wir nun noch das Risiko bei quadratischer Verlustfunktion L. In den Notationen der Vorlesung (wobei wir dem Zusammenhang entsprechend μ mit p ersetzen) gilt hier also

$$\begin{aligned}\mathcal{R}(p,d) &\overset{3.3.3}{=} E_p[L(p, d(X_1, \cdots, X_n)] \overset{3.3.3}{=} E_p[(p-\overline{X})^2] = V_p[\bar{X}] = \frac{1}{n^2} V_p\left[\sum_{i=1}^{n} X_i\right] \\ &\overset{\text{iid}}{=} \frac{1}{n^2} \cdot n V_p[X_1] = \frac{p(1-p)}{n}.\end{aligned}$$