

SS 2013

Zentralübung Diskrete Wahrscheinlichkeitstheorie (zur Vorlesung Prof. Esparza)

Dr. Werner Meixner

Fakultät für Informatik
TU München

<http://www14.in.tum.de/lehre/2013SS/dwt/uebung/>

21. Juni 2013

ZÜ VII

Übersicht:

1. Thema: TA 8.2
2. Tipps zu HA Blatt 9

1. Thema: TA 8.2

Aufgabenstellung:

Seien X_1, X_2, \dots, X_n unabhängige stetige ZVN mit identischer Verteilung bzw. Verteilungsfunktion $F(t)$.

Für $k \in [n]$ gebe die ZV $X_{(k)}$ den k -kleinsten Wert in der Sequenz (X_1, X_2, \dots, X_n) an.

- (a) Stellen Sie die Verteilungsfunktion von $X_{(1)}$ und $X_{(n)}$ mit Hilfe von F dar.
- (b) Bestimmen Sie nun die Verteilungsfunktion von $X_{(k)}$ für beliebiges $k \in [n]$.
Betrachten Sie hierfür für alle $t \in \mathbb{R}$ die ZV $S_t = \sum_{j=1}^n I_{[X_j \leq t]}$.

Lösung (a):

$$\begin{aligned}\Pr[X_{(n)} \leq t] &= \Pr[(X_1 \leq t) \wedge (X_2 \leq t) \wedge \dots \wedge (X_n \leq t)] \\ &= \Pr[X_1 \leq t] \cdot \Pr[X_2 \leq t] \cdot \dots \cdot \Pr[X_n \leq t] \\ &= F(t)^n.\end{aligned}$$

$$\begin{aligned}\Pr[X_{(1)} > t] &= \Pr[(X_1 > t) \wedge (X_2 > t) \wedge \dots \wedge (X_n > t)] \\ &= \Pr[X_1 > t] \cdot \Pr[X_2 > t] \cdot \dots \cdot \Pr[X_n > t] \\ &= (1 - F(t))^n.\end{aligned}$$

$$\Pr[X_{(1)} \leq t] = 1 - (1 - F(t))^n.$$

Lösung (b):

Die n ZV $I_{[X_j \leq t]}$ sind unabhängige Bernoulli-verteilte ZV mit gleicher W'keit $p = \Pr[X_j \leq t] = F(t)$.

Daraus folgt

$$S_t \sim \text{Bin}(n, F(t)) .$$

S_t zählt also wie viele X_j einen Wert $\leq t$ besitzen.

Daraus folgt.

$$\begin{aligned} \Pr[X_{(k)} \leq t] &= \Pr[S_t \geq t] \\ &= \sum_{j=k}^n \binom{n}{j} F(t)^j (1 - F(t))^{n-j} . \end{aligned}$$

2. Tipps zu HA Blatt 9

Die folgenden Hinweise und Tipps zu Hausaufgaben sind für die Bearbeitung nicht notwendig, möglicherweise aber hilfreich. Man sollte zunächst versuchen, die Hausaufgaben ohne Hilfestellung zu lösen.

2.1 Tipps zu HA 9.1

Aufgabenstellung:

Wir betrachten nochmals die Problemstellung aus HA 7.1:

Anhand von n Ja-Nein-Fragen soll der Kenntnisstand eines Studenten ermittelt werden. Wir nehmen idealisierend an, dass ein Student bei jeder Frage – unabhängig von allen anderen Fragen – mit W'keit p die korrekte Antwort gibt, womit die Anzahl der korrekten Antworten K gerade $\text{Bin}(n, p)$ -verteilt ist.

Die Note soll so vergeben werden, dass gilt:

Ein Student mit $p \in (\frac{10-i}{10}, \frac{11-i}{10}]$ soll – zumindest mit hoher „Wahrscheinlichkeit“ – die Note i mit $i \in \{1, 2, 3, 4\}$ erhalten; gilt $p \leq 0.6$, so soll der Student eine 5 bekommen.

Da p nicht direkt messbar ist, soll p anhand von (n -maliger Ausführung von) K abgeschätzt werden.

Man beachte, dass die ideale Notenvergabe fest definiert und in den Alternativen (a) und (b) lediglich auf unterschiedliche Weise angenähert wird.

Die Wahrscheinlichkeitsverteilung für die Zufallsvariable K ist für jeden Studenten, d.h. jedes p , und jedes n fest definiert, allerdings unbekannt.

Approximation der Binomialverteilung $\text{Bin}(n, p)$ einer Zufallsvariablen X :

Es gelten

$$\mathbb{E}[X] = np = \mu \text{ und } \text{Var}[X] = np(1 - p) = \sigma^2.$$

Für „hinreichend große“ n wird X als normalverteilt $\mathcal{N}(\mu, \sigma^2)$ betrachtet.

Die Verteilungsfunktion F_X wird entsprechend durch $\Phi_{\mu, \sigma}$ ersetzt.

Durch **lineare Transformation** $Y = aX + b$

erhält man eine Variable Y mit Normalverteilung $\mathcal{N}(a\mu + b, a^2\sigma^2)$.

Transformation auf Standardnormalverteilung Φ :

$$Y = \frac{X - \mu}{\sigma}.$$

Dann gilt:

$$\Pr[X \leq t] = \Phi_{\mu, \sigma}(t),$$

$$\Pr[Y \leq t] = \Phi(t),$$

bzw.

$$\Pr[Y > t] = 1 - \Phi(t),$$

bzw. für symmetrische Intervalle:

$$\Pr[-t \leq Y \leq t] = 2\Phi(t) - 1.$$

Fortsetzung Aufgabenstellung:

- (a) Ein Ansatz für die Notenvergabe sieht vor, K/n stellvertretend für p zu verwenden:

Gilt am Ende der Prüfung $K/n \leq 0.6$, so erhält der Student eine 5;

ansonsten erhält er die Note i , für welche $K/n \in (\frac{10-i}{10}, \frac{11-i}{10}]$ gilt.

- (i) Es gelte $p = 0.75$ und $n = 100$.

Bestimmen Sie die Verteilung der Noten in diesem Fall.

- (ii) Es gelte $p = 0.75$.

Bestimmen Sie das kleinste n , so dass mit W'keit ≥ 0.99 die Note 3 vergeben wird.

Tipps dazu:

ad (i):

Man bestimmt approximativ für alle i und $b = \frac{10-i}{10}$ die W'keit $\Pr[K/n \leq b]$ unter Zurückführung mit linearer Transformation auf die Standardnormalverteilung.

Z.B. ist für $b = \frac{6}{10}$ die Funktion $\Phi(-6/\sqrt{3})$ auszuwerten.
(Mit Tabellen aus Schickinger/Steger ist das möglich.)

Die Dichtewerte der Notenverteilung ergeben sich durch Differenzbildung.

ad (ii):

Wichtig ist, dass die Wahrscheinlichkeitswerte für die Noten nun von n abhängen.

Nun ist die W'keit zu bestimmen, dass die Note 3 vergeben wird, also

$$\Pr\left[\frac{K}{n} \in \left(\frac{7}{10}, \frac{8}{10}\right]\right].$$

Man beachte, dass der angenommene Wert $p = 0.75$ genau in der Mitte des Intervalls $\left(\frac{7}{10}, \frac{8}{10}\right]$ liegt. Bei Transformation auf Standardnormalverteilung wird aus diesem Intervall ein zum Nullpunkt symmetrisches Intervall.

Die Umkehrfunktion von Φ reicht aus, um die Bestimmungsgleichung für n zu lösen.

Zur Probe: n sollte nicht wesentlich kleiner als 500 sein.

Fortsetzung Aufgabenstellung:

- (b) Ein alternativer Ansatz vergibt die beste (= kleinste) Note i , für welche die beobachtete Anzahl an korrekten Antworten noch sehr wahrscheinlich ist:
- (i) Bestimmen Sie in Abhängigkeit von p und n ein möglichst großes $u(n, p)$, so dass $\Pr[K \geq u(n, p)] \geq 0.99$ gilt.
 - (ii) Sei $n = 100$.
Bestimmen Sie nun für alle $i \in \{1, 2, 3, 4\}$ den expliziten Wert $u_i := u(100, \frac{10-i}{10})$.
 - (iii) Seien $n = 100$ und $u_5 = 0$.
Bestimmen Sie nun wieder die Verteilung der Note für einen Studenten mit $p = 0.75$.

Tipps dazu:

ad (i):

Überprüfen Sie, ob $\Pr[K \geq u] \approx \Phi\left(-\frac{u-np}{\sqrt{np(1-p)}}\right)$ gilt.

ad (ii):

Die allgemeine Formel wird nun benützt, um die Werte $u(100, \frac{10-i}{10})$ für alle i zu berechnen.

Zur Probe: $u_3 \approx 59$.

ad (iii):

Die Berechnung der Notenverteilung erfolgt wie in (a), allerdings mit den Rasterpunkten u_i .

2.2 Tipps zu HA 9.2

Aufgabenstellung:

Die Zeiten eines 100-Meter-Läufers seien als $\mathcal{N}(9.66s, 0.011s^2)$ -verteilt angenommen. Wir betrachten 10 unabhängige 100-Meter-Läufe desselben Läufers.

- (a) Bestimmen Sie das kleinste t , so dass die schnellste Zeit mit $W'keit \geq 0.99$ unter t Sekunden liegt.
- (b) Bestimmen Sie das kleinste t , so dass die 3 schnellsten Zeiten mit $W'keit \geq 0.99$ unter t Sekunden liegen.

Tipps:

Ohne Angabe der Maßeinheit in Sekunden ist also die gemeinsame Verteilung aller X_i gegeben durch die Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ mit dem Erwartungswert $\mu = 9.66$ und der Varianz $\sigma^2 = 0.011$.

Die Aufgabe schließt sich eng an Tutoraufgabe 8.2 von Blatt 8 an.

Insbesondere sollte man die dortigen Bezeichnungen $X_{(1)}, X_{(2)}, \dots, X_{(k)}, \dots$ übernehmen.

- (a) Gefragt ist also nach dem kleinsten t , so dass $\Pr[X_{(1)} \leq t] \geq 0.99$ gilt.

$\Pr[X_{(1)} \leq t] = 1 - (1 - \Phi_{\mu, \sigma}(t))^{10}$ ergibt sich unmittelbar aus TA 8.2.

$\sqrt[10]{x}$ berechnet man mit einem Taschenrechner.

- (b) Nun ist gefragt nach dem kleinsten t , so dass $\Pr[X_{(3)} \leq t] \geq 0.99$ gilt.

In TA 8.2 wurde die Verteilung $\Pr[X_{(k)} \leq t]$ berechnet, die nun für $k = 3$ spezialisiert wird. Allerdings ist es bequemer mit $\Pr[X_{(k)} > t]$ zu arbeiten, weil die betreffende Summenformel dann nur 3 Glieder besitzt.

Es ist hilfreich zu wissen, dass 0.3883 eine geeignete Nullstelle des Polynoms $x^{10} + 10x^9(1-x) + 45x^8(1-x)^2 - 0.01$ annähert.

2.3 Tipps zu HA 9.3

Aufgabenstellung:

Es seien X_1, X_2, X_3 unabhängige ZVn, jede davon gleichverteilt auf $[0, 1]$.

Bestimmen Sie die Dichte und die Verteilungsfunktion von $X_1 + X_2 + X_3$.

Tipps:

Bestimmen Sie zunächst die Dichte f_Y der Summe $Y = X_1 + X_2$ und anschließend die gesuchte Dichte f_Z von $Z = Y + X_3$.

In beiden Fällen wird die Faltungsformel angewandt.

Beachten Sie, dass die Integration letztendlich auf die Fallunterscheidungen $t < 0$ bzw. $0 \leq t < 1$ bzw. $1 \leq t < 2$ bzw. $2 \leq t \leq 3$ bzw. $3 < t$ des Arguments t der Dichtefunktion f_Z führt.

Zur *Ergebniskontrolle* für die gesuchte Verteilungsfunktion $F_Z(t)$:
Im Bereich $0 \leq t \leq 1$ gilt $F_Z(t) = t^3/6$.