Technische Universität München
Fakultät für Informatik
Lehrstuhl für Theoretische Informatik
Prof. Dr. Susanne Albers
Dennis Kraft, Richard Stotz

Sommersemester 2015
Lösungsblatt 11
13. Juli 2015

Diskrete Wahrscheinlichkeitstheorie

Tutoraufgabe 1

Der Hersteller ihrer Lieblingsschokokekse verspricht auf der Packung, dass mindestens jeder zweite Keks echte Schokolade enthält. Sie möchten diese Behauptung überprüfen und kaufen eine Packung mit vier Keksen. Dabei gehen Sie davon aus, dass jeder Keks unabhängig und mit identischer Wahrscheinlichkeit p echte Schokolade enthält. Befinden sich weniger als zwei Kekse mit echter Schokolade in der Packung, so lehnen Sie die Behauptung des Herstellers ab. Berechnen Sie die Fehlerwahrscheinlichkeit erster Art. Überprüfen Sie außerdem die Alternativhypothese, dass höchstens jeder vierte Keks echte Schokolade enthält und berechnen Sie die Fehlerwahrscheinlichkeit zweiter Art.

Lösungsvorschlag

Als Testgröße T betrachten wir die Anzahl der Kekse mit echter Schokolade. Nach der Angabe ist T binomialverteilt mit Parametern $n = 4$ und p. Ein Fehler erster Art tritt genau dann ein, wenn Sie die Behauptung des Herstellers ablehnen obwohl p größer als $\frac{1}{2}$ ist. Die maximale Fehlerwahrscheinlichkeit hierfür ist gegeben durch

$$\sup\left\{\Pr_p[T \leq 1] \middle| \frac{1}{2} \leq p \leq 1\right\} = \sup\left\{(1-p)^4 + 4p(1-p)^3 \middle| \frac{1}{2} \leq p \leq 1\right\}$$

Die Funktion $g(p) = (1-p)^4 + 4p(1-p)^3$ ist auf dem Intervall $[\frac{1}{2}, 1]$ streng monoton fallend, wie anhand der ersten Ableitung klar wird

$$g'(p) = -4(1-p)^3 + 4(1-p)^3 - 4p3(1-p)^2 = -12p(1-p)^2.$$

Nachdem $g(p)$ außerdem stetig ist, wird das Supremum der Funktion daher am linken Rand des Intervalls erreicht. Für die Fehlerwahrscheinlichkeit erster Art folgt somit ein Wert von

$$\left(1 - \frac{1}{2}\right)^3 \left(1 + 3\frac{1}{2}\right) = \frac{5}{16}.$$

Ein Fehler zweiter Art liegt vor, wenn die Alternativhypothese abgelehnt wird, obwohl diese stimmt. Da in unserem Fall die Alternativhypothese besagt, dass $p \leq \frac{1}{4}$ ist, folgt für die Fehlerwahrscheinlichkeit

$$\sup\left\{\Pr_p[T > 1] \middle| 0 \leq p \leq \frac{1}{4}\right\} = \sup\left\{6p^2(1-p)^2 + 4p^3(1-p) + p^4 \middle| 0 \leq p \leq \frac{1}{4}\right\}.$$

Wir betrachten also die Funktion $h(p) = 6p^2(1-p)^2 + 4p^3(1-p) + p^4$ und sehen an der zweiten Ableitung

$$h'(p) = 12p(1-p)^2 - 12p^2(1-p) + 12p^2(1-p) - 4p^3 + 4p^3 = 12p(1-p)^2,$$

dass $h(p)$ auf dem Intervall $[0, \frac{1}{4}]$ monoton steigt. Auch dieses Mal handelt es sich um eine stetige Funktion, und damit liegt das Supremum am rechten Rand des Intervalls und die Fehlerwahrscheinlichkeit zweiter Art beträgt

$$6 \cdot \frac{1}{4}^2 \left(1 - \frac{1}{4}\right)^2 + 4 \cdot \frac{1}{4}^3 \left(1 - \frac{1}{4}\right) + \left(\frac{1}{4}\right)^4 = \frac{67}{256}.$$

Tutoraufgabe 2

Die Pferde a, b und c aus Hausaufgabe 1 von Übungsblatt 1 treten erneut gegeneinander an. Von 1000 Rennen gewinnt a insgesamt 346 Rennen, b gewinnt 322 und c gewinnt 332. Untersuchen Sie die Vermutung, dass jedes Pferd mit der gleichen Wahrscheinlichkeit gewinnt, und überprüfen Sie insbesondere, ob sich die Hypothese mit einem Signifikanzniveau von 0,05 ablehnen lässt.

Hinweis: Sie dürfen $\chi^2_{k,\alpha}$ mit $k \cdot (1 - \frac{2}{9k} + z_\alpha \sqrt{\frac{2}{9k}})^3$ approximieren, wobei z_α das α-Quantil der Standardnormalverteilung ist.

Lösungsvorschlag

Sei X eine diskrete Zufallsvariable, die die Werte 1, 2 oder 3 annimmt, je nachdem ob Pferd a, b oder c gewinnt. Um die Nullhypothese

$$\Pr[X = 1] = \Pr[X = 2] = \Pr[X = 3] = \frac{1}{3}$$

zu Testen führen wir eine χ^2-Test durch. Der Wert unserer Testgröße ist folglich gegeben durch

$$\frac{(346 - \frac{1000}{3})^2}{\frac{1000}{3}} + \frac{(322 - \frac{1000}{3})^2}{\frac{1000}{3}} + \frac{(332 - \frac{1000}{3})^2}{\frac{1000}{3}} = \frac{109}{125}.$$

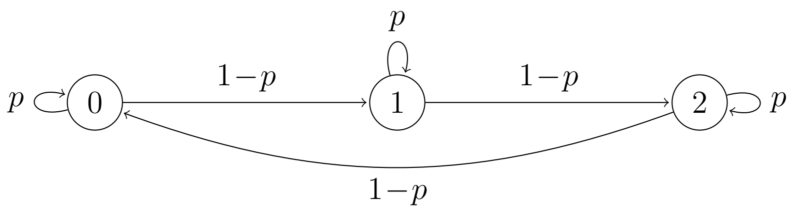
Ist dieser Wert größer als $\chi^2_{3-1,1-0,05}$, so müssen wir die Hypothese gemäß dem χ^2-Test ablehnen. Wir nutzen den Hinweis aus der Angabe und berechnen $\chi^2_{2,0,95}$ wie folgt

$$\chi^2_{2,0,95} \approx 2 \cdot \left(1 - \frac{2}{18} + z_{0,95}\sqrt{\frac{2}{18}}\right)^3 \approx 2 \cdot \left(1 - \frac{1}{9} + \frac{1,65}{3}\right)^3 \approx 5,96.$$

Die Hypothese lässt sich also nicht mit einem Signifikanzniveau von 0,05 ablehnen.

Tutoraufgabe 3

Wir betrachten eine Markov-Kette X_t mit $t \geq 0$ über der der Zustandsmenge $S = \{0, 1, 2\}$, die durch das folgende Übergangsdiagramm in Abhängigkeit eines Parameters $0 < p < 1$ gegeben ist.



1. Bestimmen Sie die Übergangsmatrix P.

2. Geben Sie die bedingte Wahrscheinlichkeit $\Pr[X_{t+2} = j \mid X_t = i]$ in Matrixschreibweise für alle i, j an.

3. Sei $T_{0,2}$ die Mindestanzahl an Schritten um von Zustand 0 den Zustand 2 zu erreichen. Bestimmen Sie die Wahrscheinlichkeit $\Pr[T_{0,2} = 3]$ sowie $\mathbb{E}[T_{0,2}]$.

Lösungsvorschlag

Die Übergangsmatrix P können wir sofort am Übergangsdiagramm ablesen. Der Eintrag $p_{i,j}$ entspricht dabei der Wahrscheinlichkeit $\Pr[X_{t+1} = j \mid X_t = i]$, also der Wahrscheinlichkeit vom Zustand i in einem Schritt in den Zustand j zu gehen. Somit folgt

$$P = \begin{pmatrix} p & 1-p & 0 \\ 0 & p & 1-p \\ 1-p & 0 & p \end{pmatrix}.$$

Um nun die geforderten Wahrscheinlichkeiten $\Pr[X_{t+2} = j \mid X_t = i]$ für Übergänge in zwei Schritten anzugeben, müssen wir die Matrix P lediglich quadrieren

$$P^2 = \begin{pmatrix} p^2 & 2p(1-p) & (1-p)^2 \\ (1-p)^2 & p^2 & 2p(1-p) \\ 2p(1-p) & (1-p)^2 & p^2 \end{pmatrix}.$$

Der Term $\Pr[T_{0,2} = 3]$ beschreibt die Wahrscheinlichkeit, ausgehend von Zustand 0 nach genau drei Schritten erstmals Zustand 2 zu erreichen. Im Diagramm können wir ablesen, dass es genau zwei Pfade gibt, über die dies möglich ist, nämlich $0, 0, 1, 2$ und $0, 1, 1, 2$. Die Wahrscheinlichkeit dieser Pfade ist das Produkt ihrer Übergangswahrscheinlichkeiten, also $p(1-p)^2$. Insgesamt folgt somit, dass

$$\Pr[T_{0,2} = 3] = 2p(1-p)^2.$$

Schließlich berechnen wir noch den Erwartungswert $\mathbb{E}[T_{0,2}]$, welchen wir im Folgenden abkürzen mit $h_{0,2}$. Nach Vorlesung gilt

$$h_{0,2} = 1 + p_{0,0} \cdot h_{0,2} + p_{0,1} \cdot h_{1,2} = 1 + p \cdot h_{0,2} + (1-p) \cdot h_{1,2},$$

wobei $h_{1,2}$ die erwartete Mindestanzahl an Schritten von Zustand 1 in Zustand 2 bezeichnet. Für diese Übergangszeit gilt wiederum

$$h_{1,2} = 1 + p_{1,0} \cdot h_{0,2} + p_{1,1} \cdot h_{1,2} = 1 + 0 \cdot h_{0,2} + p \cdot h_{1,2},$$

woraus folgt, dass $h_{1,2}$ gleich $\frac{1}{1-p}$ ist. Wir setzen dieses Ergebnis in unsere ursprüngliche Gleichung ein und erhalten

$$h_{0,2} = 1 + p \cdot h_{0,2} + \frac{1-p}{1-p} \Longleftrightarrow h_{0,2} = \frac{2}{1-p}.$$

Hausaufgabe 1 (5 Punkte)

Sie planen die Nachkommastellen von π als Pseudozufallsgenerator für die Zahlen von 0 bis 9 zu benutzen. Allerdings möchten Sie zunächst untersuchen, ob jede Ziffer mit der gleichen Wahrscheinlichkeit vorkommt. Führen Sie anhand der ersten 10000 Nachkommastellen einen geeigneten Hypothesentest mit Signifikanzniveau 0,05 durch.

Hinweis: Überlegen Sie sich eine geeignete Methode die ersten 10000 Nachkommastellen von π zu ermitteln. Sie dürfen auch elektronische Hilfsmittel zurate ziehen.

Lösungsvorschlag

Sei X eine diskrete Zufallvariable, welche ausschließlich Werte von 0 bis 9 annimmt. Wir führen einen χ^2-Test mit der Nullhypothese

$$\Pr[X = k] = \frac{1}{10}$$

für k von 0 bis 9 durch. Unsere Stichproben x_i entsprechen dabei den ersten 10000 Nachkommastellen von π. Die Häufigkeit der einzelnen Ziffern ist dabei in der folgenden Tabelle aufgeführt.

Ziffer	0	1	2	3	4	5	6	7	8	9
Häufigkeit	968	1026	1021	974	1012	1046	1021	970	948	1014

Unsere Testgröße hat somit den Wert

$$\frac{(-32)^2 + 26^2 + 21^2 + (-26)^2 + 12^2 + 46^2 + 21^2 + (-30)^2 + (-52)^2 + 14^2}{10000 \cdot \frac{1}{10}} = \frac{4659}{500}.$$

Übersteigt dieser Wert $\chi^2_{9,0,95}$, so müssen wir die Hypothese ablehnen. Mit Hilfe des Hinweises aus Tutoraufgabe 2 von diesem Übungsblatt approximieren wir $\chi^2_{9,0,95}$ zu

$$\chi^2_{9,0,95} \approx 9 \cdot \left(1 - \frac{2}{81} + z_{0,95}\sqrt{\frac{2}{81}}\right)^3 \approx 9 \cdot \left(1 - \frac{2}{81} + \frac{1{,}65 \cdot \sqrt{2}}{9}\right)^3 \approx 16{,}94.$$

Damit liefert dieser Test also keinen Grund, die Nullhypothese abzulehnen.

Hausaufgabe 2 (5 Punkte)

In einem Langzeittest wollen die WG-Bewohner a und b ermitteln, was der schnellste Weg zur Universität ist. Hierbei nimmt a jeden Tag die U-Bahn, während b mit dem Fahrrad fährt. Beide fahren jeden Morgen gleichzeitig los. Nach 1000 Tagen war a insgesamt 550-mal früher an der Universität als b. Kann man auf einem Signifikanzniveau von 0,05 annehmen, dass Fahrradfahren mindestens so schnell ist, wie die öffentlichen Verkehrsmittel? Geben Sie hierzu einen Binomialtest an.

Lösungsvorschlag

Wir modellieren jeden einzelnen Tag als Bernoulli-Experiment, bei dem a mit Wahrscheinlichkeit p früher als b die Universität erreicht. Die Nullhypothese ist somit $p \leq \frac{1}{2}$. Wir führen also einen Binomialtest durch und erhalten die Testgröße

$$\frac{550 - \frac{1}{2} \cdot 1000}{\sqrt{1000 \cdot \frac{1}{2} \cdot (1 - \frac{1}{2})}} = \frac{50}{5\sqrt{10}} = \sqrt{10}$$

Um den Test anzunehmen, muss die Testgröße kleiner dem 0,95-Quantil der Standardnormalverteilung sein. Aus der Tabelle lesen wir ab, dass $z_{0,95}$ ungefähr 1,65 ist, wohingegen $\sqrt{10}$ größer als 3 ist. Folglich können wir die Nullhypothese mit einem Signifikanzniveau von 0,05 ablehnen.

Hausaufgabe 3 (5 Punkte)

Sie möchten den Erwartungswert μ einer normalverteilte Zufallsvariable X mit Varianz 1 ausgehend von n unabhängigen Stichproben auf $\mu = \mu_0$ gegen $\mu = \mu_1$ testen. Wählen Sie eine geeignete Testgröße T und bestimmen Sie eine untere Schranke für n in Abhängigkeit von μ_0 und μ_1, so dass sowohl der Fehler erster Art, als auch der Fehler zweiter Art kleiner als 0,01 sind.

Lösungsvorschlag

Als Testgröße wählen wir das Stichprobenmittel $\bar{X}$. Außerdem nehmen wir ohne Beschränkung der Allgemeinheit an, dass μ_0 kleiner als μ_1 ist. Sei $\mu = \mu_0$ die Nullhypothese des Tests. Sollte $\bar{X}$ einen gewissen Wert t, den wir im Folgenden bestimmen werden, überschreiten, so lehnen wir die Nullhypothese ab. Anderenfalls akzeptieren wir sie. Die Wahrscheinlichkeit des Fehlers erster Art, also dass wir Nullhypothese fälschlicherweise ablehnen, beträgt

$$\Pr_{\mu_0}[\bar{X} > t] = 1 - \Phi(\sqrt{n}(t - \mu_0)),$$

wohingegen der Fehler zweiter Art, also dass wir Nullhypothese fälschlicherweise annehmen, gegeben ist durch

$$\Pr_{\mu_1}[\bar{X} \leq t] = \Phi(\sqrt{n}(t - \mu_1)).$$

Nachdem beide Fehler kleiner als $\frac{1}{100}$ sein sollen, erhalten wir das Ungleichungssystem

$$1 - \Phi(\sqrt{n}(t - \mu_0)) \leq 0{,}01 \iff \Phi(\sqrt{n}(t - \mu_0)) \geq 0{,}99 \iff t \geq \frac{z_{0,99}}{\sqrt{n}} + \mu_0$$

und

$$\Phi(\sqrt{n}(t - \mu_1)) \leq 0{,}01 \iff t \leq \frac{z_{0,01}}{\sqrt{n}} + \mu_1.$$

Wählt man t gleich $\frac{z_{0,99}}{\sqrt{n}} + \mu_0$, so ist die erste Ungleichung trivialerweise erfüllt. Wir müssen also lediglich ein n bestimmen so dass auch die zweite Ungleichung erfüllt ist. Dazu formen wir um zu

$$\begin{aligned}
& & t &\leq \frac{z_{0,01}}{\sqrt{n}} + \mu_1 \\
&\iff & \frac{z_{0,99}}{\sqrt{n}} + \mu_0 &\leq \frac{z_{0,01}}{\sqrt{n}} + \mu_1 \\
&\iff & z_{0,99} - z_{0,01} &\leq \sqrt{n}(\mu_1 - \mu_0) \\
&\iff & n &\geq \left(\frac{z_{0,99} - z_{0,01}}{\mu_1 - \mu_0}\right)^2,
\end{aligned}$$

wobei die letzte Umformung gültig ist, da μ_1 größer als μ_0 ist. Wir müssen somit insgesamt

$$\left\lceil \left(\frac{z_{0,99} - z_{0,01}}{\mu_1 - \mu_0}\right)^2 \right\rceil$$

Stichproben durchführen, um den Fehler erster und zweiter Art kleiner als 0,01 zu halten.

Hausaufgabe 4 (5 Punkte)

Das Mittel über die Durchschnittsnoten der Klausuren im Fach Diskrete Wahrscheinlichkeitstheorie in den letzten zwanzig Jahre beträgt 2,95 mit einer Stichprobenvarianz von 1,94. Im Gegensatz dazu beträgt das Mittel der Wiederholungsklausuren lediglich 4,19 mit einer Stichprobenvarianz von 2,26. Angenommen die Durchschnittsnoten der regulären Klausuren und der Wiederholungsklausuren sind unabhängig und jeweils identisch normalverteilt. Außerdem gehen wir davon aus, dass beide Klausurtypen die gleiche Varianz haben. Testen Sie mit einem Signifikanzniveau von 0,05, ob beide Klausuren in der Regel gleich gut ausfallen.

Lösungsvorschlag

Da es sich bei der Durchschnittsnot der regulären Klausur und der Wiederholungsklausur jeweils um eine normalverteilte Zufallsvariable X und Y handelt, wobei beide Variablen die gleiche Varianz haben, führen wir einen zwei Stichproben-t-Test durch. Dabei fällt die reguläre Klausur genau gleich gut aus wie die Wiederholungsklausur, wenn die Erwartungswerte von X und Y gleich sind. Laut Vorlesung ist der Wert unserer Testgröße demnach

$$\sqrt{\frac{20+20-2}{\frac{1}{20}+\frac{1}{20}}} \cdot \frac{2{,}95-4{,}19}{\sqrt{(20-1)\cdot 1{,}94+(20-1)\cdot 2{,}26}} \approx -2{,}706,$$

wobei wir unsere Hypothese ablehnen, wenn der absolute Wert der Testgröße größer als $t_{38,0{,}975}$. Aus der Vorlesung ist bekannt, dass $t_{n,\alpha}$ für Werte von n größer gleich 30 durch das α-Quantil der der Standardnormalverteilung ausreichend gut approximiert wird und wir erhalten

$$t_{20+20-2,1-\frac{0{,}05}{2}} = t_{38,0{,}975} \approx z_{0{,}975} \approx 1{,}96.$$

Folglich liegt unsere Testgröße im Ablehnungsbereich und wir können die Nullhypothese mit einem Signifikanzniveau von 0,05 verwerfen.

$\Phi(x)$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500	0,504	0,508	0,512	0,516	0,520	0,524	0,528	0,532	0,536
0,1	0,540	0,544	0,548	0,552	0,556	0,560	0,564	0,567	0,571	0,575
0,2	0,579	0,583	0,587	0,591	0,595	0,599	0,603	0,606	0,610	0,614
0,3	0,618	0,622	0,626	0,629	0,633	0,637	0,641	0,644	0,648	0,652
0,4	0,655	0,659	0,663	0,666	0,670	0,674	0,677	0,681	0,684	0,688
0,5	0,691	0,695	0,698	0,702	0,705	0,709	0,712	0,716	0,719	0,722
0,6	0,726	0,729	0,732	0,736	0,739	0,742	0,745	0,749	0,752	0,755
0,7	0,758	0,761	0,764	0,767	0,770	0,773	0,776	0,779	0,782	0,785
0,8	0,788	0,791	0,794	0,797	0,800	0,802	0,805	0,808	0,811	0,813
0,9	0,816	0,819	0,821	0,824	0,826	0,829	0,831	0,834	0,836	0,839
1,0	0,841	0,844	0,846	0,848	0,851	0,853	0,855	0,858	0,860	0,862
1,1	0,864	0,867	0,869	0,871	0,873	0,875	0,877	0,879	0,881	0,883
1,2	0,885	0,887	0,889	0,891	0,893	0,894	0,896	0,898	0,900	0,901
1,3	0,903	0,905	0,907	0,908	0,910	0,911	0,913	0,915	0,916	0,918
1,4	0,919	0,921	0,922	0,924	0,925	0,926	0,928	0,929	0,931	0,932
1,5	0,933	0,934	0,936	0,937	0,938	0,939	0,941	0,942	0,943	0,944
1,6	0,945	0,946	0,947	0,948	0,949	0,951	0,952	0,953	0,954	0,954
1,7	0,955	0,956	0,957	0,958	0,959	0,960	0,961	0,962	0,962	0,963
1,8	0,964	0,965	0,966	0,966	0,967	0,968	0,969	0,969	0,970	0,971
1,9	0,971	0,972	0,973	0,973	0,974	0,974	0,975	0,976	0,976	0,977
2,0	0,977	0,978	0,978	0,979	0,979	0,980	0,980	0,981	0,981	0,982
2,1	0,982	0,983	0,983	0,983	0,984	0,984	0,985	0,985	0,985	0,986
2,2	0,986	0,986	0,987	0,987	0,987	0,988	0,988	0,988	0,989	0,989
2,3	0,989	0,990	0,990	0,990	0,990	0,991	0,991	0,991	0,991	0,992
2,4	0,992	0,992	0,992	0,992	0,993	0,993	0,993	0,993	0,993	0,994
2,5	0,994	0,994	0,994	0,994	0,994	0,995	0,995	0,995	0,995	0,995
2,6	0,995	0,995	0,996	0,996	0,996	0,996	0,996	0,996	0,996	0,996
2,7	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997
2,8	0,997	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998
2,9	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,999	0,999	0,999