Markus Reiß

Vorlesung
Einführung in die Statistik
Sommersemester 2005

1. Übungsblatt

1. Es sei Ω eine abzählbare Menge. Beweise:

   (a) Ist $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $\mathfrak{P}(\Omega)$, so ist $p(\omega) := \mathbb{P}(\{\omega\})$, $\omega \in \Omega$, eine Zähldichte.

   (b) Ist p eine Zähldichte auf Ω, so definiert $\mathbb{P}(A) := \sum_{\omega \in A} p(\omega)$, $A \subseteq \Omega$, ein Wahrscheinlichkeitsmaß.

   (c) Die Funktion $\mathbb{Q} : \mathfrak{P}(\Omega) \to [0,1]$ sei eine Inhaltsfunktion, d.h. $\mathbb{Q}(\Omega) = 1$ und $\mathbb{Q}(A \cup B) = \mathbb{Q}(A) + \mathbb{Q}(B)$ gelte für alle disjunkten Mengen A, $B \subseteq \Omega$. Ist $\mathbb{Q}$ σ-stetig, so ist $\mathbb{Q}$ ein Wahrscheinlichkeitsmaß.

2. Löse die folgenden Textaufgaben jeweils mit vollständiger Angabe und Begründung der wahrscheinlichkeitstheoretischen Modellierung:

   (a) Wie viele Rosinen müssen in 500g Teig vorhanden sein, damit ein 50g-Brötchen mit mindestens 99% Wahrscheinlichkeit eine Rosine enthält?

   (b) Ein gewisser Chevalier de Méré wunderte sich, dass er beim Werfen mit drei Würfeln die Augensumme 11 häufiger beobachtet hatte als die Augensumme 12, obwohl doch 11 durch die Kombinationen $6-4-1$, $6-3-2$, $5-5-1$, $5-4-2$, $5-3-3$, $4-4-3$ und die Augensumme 12 durch ebensoviele (welche?) Kombinationen erzeugt würde. Kann diese Beobachtung als „vom Zufall bedingt“ angesehen werden oder ist die Argumentation falsch?

3. Beweise folgendes Konvergenzresultat für die Einzelwahrscheinlichkeiten von hypergeometrischer und Binomial-Verteilung:
   Für Parameter n, $w \in \mathbb{N}_0$ sowie N, $W(N) \in \mathbb{N}$ mit $\lim_{N\to\infty} W(N)/N = p \in [0,1]$ gilt $\lim_{N\to\infty} \mathrm{Hyp}_{N,W(N),n}(w) = \mathrm{Bin}_{n,p}(w)$.
   Erkläre dieses Resultat anhand eines Urnenmodells anschaulich.

4. Ein Teich enthält eine unbekannte Zahl N von Fischen. Um N zu schätzen, werden W Fische gefangen, markiert und wieder ausgesetzt. Es wird eine Weile gewartet und dann werden n Fische gefangen, und die Zahl w der markierten Fische darunter ermittelt.

   (a) Begründe, weshalb die Anzahl der markierten Fische im zweiten Fang näherungsweise als hypergeometrisch verteilt angesehen werden kann.

   (b) Was kann man über N mit Sicherheit auf Grund der Kenntnis von W, n und w aussagen?

   (c) Nun sei $W = 50$, $n = 50$, $w = 3$. Für die Ausgabe von Angellizenzen soll N möglichst nicht überschätzt werden. Bestimme dazu $\hat{N}_{0.05}$, den maximal möglichen Wert von N, für den die Wahrscheinlichkeit, höchstens drei markierte Fische im zweiten Fang zu fischen, weniger als 5% beträgt. Vergleiche $\hat{N}_{0.05}$ mit der sicheren unteren Schranke für N aus (b).

   Es ist gestattet, die Approximation $\mathrm{Hyp}_{N,W,n}(w) \approx \mathrm{Bin}_{n,W/N}(w)$ aus Aufgabe 3 zu benutzen.

Abgabe in der Vorlesung am Freitag, 22. April.

Markus Reiß

Vorlesung
Einführung in die Statistik
Sommersemester 2005

2. Übungsblatt

1. Zu einer Tanzstunde kommen n Paare. Um für Abwechslung zu sorgen, wird jeder Dame rein zufällig einer der Herren zugelost. Wie groß ist die Wahrscheinlichkeit, dass mindestens ein ursprüngliches Paar miteinander tanzen wird? Bestimmen Sie den Grenzwert dieser Wahrscheinlichkeit für $n \to \infty$.
   Anleitung:
   Sei A_k das Ereignis „Dame k wird ursprünglicher Partner zugelost".
   Beweise die *Einschluss-Ausschluss-Formel*:
   $$\mathbb{P}\Big(\bigcup_{k=1}^{n} A_k\Big) = \sum_{l=1}^{n}\Big((-1)^{l-1} \sum_{\{k_1,\dots,k_l\}\subseteq\{1,\dots,n\}} \mathbb{P}\left(A_{k_1}\cap\cdots\cap A_{k_l}\right)\Big).$$
   Bestimme die rechte Seite z.B. mittels der Ergebnisse für Urnenmodelle.

2. Beim Skatspiel erhält jeder der drei Spieler zehn Karten, zwei Karten werden im Skat abgelegt. Mit welcher Wahrscheinlichkeit erhält jeder Spieler genau ein As?
   Hinweis: Unter den 32 Karten gibt es insgesamt vier Asse.

3. Es seien $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum sowie $B, C \subseteq \Omega$ Ereignisse mit $\mathbb{P}(B)$, $\mathbb{P}(B\cap C)$, $\mathbb{P}(B \cap C^c) > 0$. Beweise oder widerlege:
   (a) $\mathbb{Q}(A) := \mathbb{P}(A \,|\, B)$ definiert ein Wahrscheinlichkeitsmaß auf $\mathcal{P}(\Omega)$.
   (b) $\mathbb{Q}(A) := \mathbb{P}(B \,|\, A)$ definiert ein Wahrscheinlichkeitsmaß auf $\mathcal{P}(\Omega)$.
   (c) Für alle $A \subseteq \Omega$ gilt: $\mathbb{P}(A \,|\, B) = 0 \iff A \subseteq B^c$.
   (d) Für alle $A \subseteq \Omega$ gilt: $\mathbb{P}(A \,|\, B) = \mathbb{P}(A \,|\, B\cap C)\,\mathbb{P}(C) + \mathbb{P}(A \,|\, B \cap C^c)\,\mathbb{P}(C^c)$.

4. Die amerikanische Journalistin Marilyn vos Savant (mit angeblich dem höchsten IQ der Welt) bekam 1990 für ihre Denksportkolumne im „Parade Magazine" von einem Leser folgende Aufgabe:
   Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say number one, and the host, who knows what's behind the doors, opens another door, say number 3, which has a goat. He says to you "Do you want to pick door number 2?" Is it to your advantage to switch your choice of doors?
   Ihre Antwort lautete: "*Yes, you should switch. The first door has a 1/3 chance of winning, but the second door has a 2/3 chance.*" Hat sie recht?
   (Bitte mathematisch exakt begründen!)

Abgabe in der Vorlesung am Freitag, 29. April.

Markus Reiß

Vorlesung
Einführung in die Statistik
Sommersemester 2005

3. Übungsblatt

1. Zeige:

   (a) Die Ereignisse $\varnothing$ und Ω sind von jedem Ereignis A stochastisch unabhängig.

   (b) Sind A und B disjunkte Ereignisse mit $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$, so sind A und B nicht stochastisch unabhängig.

   (c) Sind die Ereignisse A, B und C stochastisch unabhängig, so sind sowohl $A \cap B$ und C als auch $A \cup B$ und C jeweils stochastisch unabhängig.

   (d) Es gibt Ereignisse A, B und C, die paarweise, aber nicht insgesamt stochastisch unabhängig sind.
   Tipp: betrachte den Wurf zweier Münzen und die drei Ereignisse, dass die erste Münze Kopf zeigt, dass die zweite Münze Kopf zeigt und dass beide Münzen dieselbe Seite zeigen.

2. Ein System besteht aus n gleichartigen, voneinander unabhängigen Komponenten $K_1, \dots, K_n$. Es funktioniert nur, solange alle Komponenten funktioniert. Für jede einzelne Komponente sei die Zeit bis zum Ausfall geometrisch verteilt mit demselben Parameter $p \in (0, 1)$.

   (a) Bestimme für jedes feste $i = 1, \dots, n$ und $m \in \mathbb{N}_0$ die Wahrscheinlichkeit des Ereignisses $A_{i,m}$, dass K_i zur Zeit m noch funktionsfähig ist, d.h. die Zeit bis zum Ausfall von K_i mindestens $m + 1$ beträgt.

   (b) Konstruiere einen diskreten Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ und gib Teilmengen $B_{i,m}$, $i = 1, \dots, n$, von Ω an, die jeweils dem Ereignis $A_{i,m}$ entsprechen, so dass $B_{1,m}, \dots, B_{n,m}$ stochastisch unabhängig sind.

   (c) Zeige, dass die Zeit bis zum Ausfall des Gesamtsystems geometrisch verteilt ist mit Parameter $1 - (1 - p)^n$.
   Tipp: Das Ereignis, dass die Zeit bis zum Ausfall des Gesamtsystems m beträgt, lässt sich darstellen als $(\bigcap_i B_{i,m-1}) \setminus (\bigcap_i B_{i,m})$ (Begründung!).

3. Es sei $(\Omega, \mathbb{P})$ ein Bernoulli-Experiment der Länge n mit Erfolgswahrscheinlichkeit p. Bestimme für die Zufallsvariablen

$$X(\omega) = \sum_{k=1}^{n} \omega_k, \quad Y(\omega) = \min\{k \in \{1, \dots, n\} : \omega_k = 1 \text{ oder } k = n\}$$

   $X(\Omega)$ und $Y(\Omega)$ sowie ihre Verteilungen $\mathbb{P}^X$ und $\mathbb{P}^Y$.
   Zur Erinnerung: $\Omega = \{0, 1\}^n$, $\mathbb{P}(\{\omega\}) = p^{\sum_{i=1}^n \omega_i}(1 - p)^{n - \sum_{i=1}^n \omega_i}$

4. Es sei X eine geometrisch verteilte Zufallsvariable mit Parameter p. Zeige, dass für alle n, $k \in \mathbb{N}$ gilt:

$$\mathbb{P}(X = k) = \mathbb{P}(X = n - 1 + k \,|\, X \geqslant n).$$

Erkläre, warum diese Eigenschaft *Gedächtnislosigkeit der geometrischen Verteilung* genannt wird, anhand der Interpretation von X als Zeit des ersten Erfolgs in einem Bernoulli-Experiment.

Abgabe in der Vorlesung am Freitag, 6. Mai.

Markus Reiß

Vorlesung
Einführung in die Statistik
Sommersemester 2005

4. Übungsblatt

1. Es sei $(X_i)_{i\in I}$ mit einer Indexmenge I eine Familie unabhängiger Zufallsvariablen auf $(\Omega, \mathbb{P})$. Zeige:

   (a) Sind $g_i : X_i(\Omega) \to \mathbb{R}$, $i \in I$, beliebige Funktionen, so ist auch $(g_i(X_i))_{i\in I}$ eine Familie unabhängiger Zufallsvariablen.

   (b) Im Fall $I = \{1, 2, 3\}$ und für Funktionen $f : X_1(\Omega) \times X_2(\Omega) \to \mathbb{R}$, $g : X_3(\Omega) \to \mathbb{R}$ sind die Zufallsvariablen $Y = f(X_1, X_2)$ und $Z = g(X_3)$ unabhängig.

2. Es sei X ein r-dimensionaler Zufallsvektor auf $(\Omega, \mathbb{P})$ mit Koordinaten X_i, d.h. $X(\omega) = (X_1(\omega), \dots, X_r(\omega))$.

   (a) Beweise formal, dass die i-te Randverteilung von X durch folgende Zähldichte beschrieben wird:
   $$p^{X_i}(x_i) = \sum_{x_j \in X_j(\Omega);\, j\in\{1,\dots,r\}\setminus\{i\}} \mathbb{P}(X_1 = x_1, \dots, X_r = x_r), \quad x_i \in X_i(\Omega).$$

   (b) Nun sei $r = 3$ und X Multinomial-verteilt mit Parametern $n \geqslant 1$ und $p_1, p_2, p_3 \in [0, 1]$, so dass $p_1 + p_2 + p_3 = 1$. Weise nach, dass für $i = 1, 2, 3$ die i-te Randverteilung gerade die Binomialverteilung Bin_{n,p_i} ist. Sind X_1, X_2 und X_3 unabhängig?

3. Es sei X eine Poiss_λ-verteilte und Y eine Geo_p-verteilte Zufallsvariable mit Parameter $\lambda > 0$ bzw. $p \in (0, 1)$.

   (a) Bestimme $\mathbb{E}[X]$, $\mathbb{E}[X^2]$ sowie $\mathbb{E}[Y]$.

   (b) Zeige, dass X und Y im Raum L^p liegen für jedes $p \geqslant 1$.

4. Bei der Fußballweltmeisterschaft treten 32 Mannschaften mit einem Kader von jeweils 20 Spielern gegeneinander an. Es gibt von jedem Spieler ein Sammelbildchen. Am Kiosk wird ein sichtgeschützt verpacktes Bildchen für einen Cent verkauft. Wieviel wird ein Sammler im Mittel am Kiosk ausgeben, bis er von jedem Spieler (mindestens) ein Bildchen besitzt?
   Tipp: Bezeichne mit N_i die Anzahl der erworbenen Bildchen, bis man von i Spielern ein Bildchen besitzt, und bestimme die Verteilung von $D_i = N_i - N_{i-1}$.
   Freiwillige Zusatzaufgabe: Was ergibt sich, falls sich zwei Sammler zusammentun und Bildchen tauschen? (+3P)

Abgabe in der Vorlesung am Freitag, 13. Mai.

Markus Reiß

Vorlesung
Einführung in die Statistik
Sommersemester 2005

5. Übungsblatt

1. Es sei X eine Zufallsvariable in L^1 und $\psi : \mathbb{R} \to \mathbb{R}$ mit $\psi(x) = \mathbb{E}[|X - x|]$ die *mittlere absolute Abweichung*.

   (a) Weise nach, dass $\psi(x)$ für alle $x \in \mathbb{R}$ endlich ist.

   (b) Bestimme alle Werte $x \in \mathbb{R}$, an denen ψ sein Minimum annimmt.

   (c) Berechne diese Minimalwerte für den Fall, dass X Geo_p-verteilt ist mit $p \in (0, 1)$. Erkläre den Unterschied zum Erwartungswert.

2. Beim Würfelwurf mit zwei fairen Würfeln bezeichne X_1 die Augenzahl des ersten und X_2 die Augenzahl des zweiten Würfels. Setze $S := X_1 + X_2$ (Augensumme) und $M := \max(X_1, X_2)$ (höchste Augenzahl).

   (a) Bestimme die gemeinsame Verteilung von S und M. Zeichne die Bildmenge $M(\Omega) \times S(\Omega)$ in ein Koordinatensystem ein (M: Abszisse; S: Ordinate).

   (b) Berechne $\mathrm{Var}(S)$, $\mathrm{Var}(M)$ und $\mathrm{Cov}(S, M)$. Zeichne die beste lineare Vorhersage von S der Form $a^*M + b^*$ mit in das Koordinatensystem ein.

   (c) Bestimme die bedingten Erwartungswerte $\mathbb{E}[S \,|\, M = m]$, $m = 1, \dots, 6$, und zeichne sie ebenfalls in das Koordinatensystem ein.

   [Computereinsatz gestattet!]

*3. Bei der Eröffnungsveranstaltung einer Konferenz mit 500 Teilnehmern will der Organisator allen, die an diesem Tag Geburtstag haben, einen Blumenstrauß überreichen. Wie viele Sträuße muss er mindestens vorrätig haben, um mit Sicherheit (bzw. mit 99% Wahrscheinlichkeit bzw. mit 95% Wahrscheinlichkeit) alle „Geburtstagskinder“ beschenken zu können?

*4. Eine Krankheit kommt bei ca. $0,1\%$ der Bevölkerung vor. Ein Test zur Erkennung der Krankheit führt bei 97% der Kranken, aber auch bei 2% der Gesunden zu einer Reaktion. Wie groß ist die Wahrscheinlichkeit, dass eine Person, bei der der Test positiv reagiert, wirklich krank ist? Sollte man eher versuchen, die Erkennungsrate bei den Kranken um 1% zu steigern oder die Fehlerkennungsrate bei den Gesunden um 1% zu senken, um diese Wahrscheinlichkeit deutlich zu erhöhen?

Aufgaben 3 und 4 sind freiwillig und dienen nur zum „Auffüllen“ auf 50%.
Abgabe in der Vorlesung am Freitag, 20. Mai.

Markus Reiß

Vorlesung
Einführung in die Statistik
Sommersemester 2005

6. Übungsblatt

1. Es sei X eine reellwertige Zufallsvariable und $F_X(x) := \mathbb{P}(X \leqslant x)$, $x \in \mathbb{R}$, ihre *Verteilungsfunktion*. Zeige:

   (a) F_X ist monoton wachsend (d.h. $x_1 \leqslant x_2 \Rightarrow F_X(x_1) \leqslant F_X(x_2)$).

   (b) Es gilt $\mathbb{P}(X < x) = \lim_{y \uparrow x} F_X(y)$ (Tipp: σ-Stetigkeit).

   (c) F_X ist rechtsstetig (d.h. $x_n \downarrow x \Rightarrow F_X(x_n) \to F_X(x)$) (Tipp: σ-Stetigkeit).

   (d) Für jedes $\alpha \in (0,1)$ existiert ein Wert $q_\alpha \in \mathbb{R}$ mit $\mathbb{P}(X \leqslant q_\alpha) \geqslant \alpha$ und $\mathbb{P}(X \geqslant q_\alpha) \geqslant 1 - \alpha$ (ein α-*Quantil*). Wann ist q_α eindeutig bestimmt?

2. Ein Betrieb fertigt an zwei Maschinen elektrische Widerstände. Bei Maschine 1 gibt es erfahrungsgemäß 10% Ausschussstücke (1. Wahl), bei Maschine 2 30% Ausschussstücke (2. Wahl). Es wird eine große unbeschriftete Kiste mit Widerständen gefunden und es soll mittels einer Stichprobe von n Widerständen aus der Kiste getestet werden, ob es sich um eine Kiste aus der Produktion für 1. Wahl oder 2. Wahl handelt.

   (a) Formuliere dieses Testproblem mathematisch.

   (b) Der peinliche Irrtum, dass 2. Wahl irrtümlich als 1. Wahl deklariert wird, soll mit maximal 5% Wahrscheinlichkeit eintreten. Bestimme einen entsprechenden Neyman-Pearson-Test für die Fälle $n = 10$ und $n = 20$ und berechne die Wahrscheinlichkeit für den Fehler 2. Art.

   (c) Wie groß muss n mindestens gewählt werden, damit sowohl die Wahrscheinlichkeit für einen Fehler 1. Art als auch für einen Fehler 2. Art maximal 5% beträgt?

3. Zur Behandlung einer Krankheit gibt es ein bewährtes Medikament, dass in 80% der Fälle zu einer Heilung führt. Der Hersteller eines neuen Medikaments verspricht größere Heilungschancen, da nach Verabreichung an 50 erkrankten Personen 44 geheilt wurden. Ist diese Aussage gerechtfertigt durch einen gleichmäßig besten Test zum Irrtumsniveau 5% der Hypothese, dass das neue Medikament maximal gleiche Heilungschancen bietet, gegen die Alternative, dass es bessere bietet? Skizziere die Gütefunktion dieses Tests durch Interpolation einiger Werte.

4. Es seien $\alpha \in (0,1)$, $\Theta = [0,1]$ und $\mathcal{X} = \{0,1,\dots,n\}$ für ein $n \geqslant 2$. Betrachte für das statistische Modell $(\mathcal{X}, (\mathrm{Bin}_{n,p})_{p\in\Theta})$ das Problem des Tests der Hypothese $H_0 : p = 1/2$ gegen die Alternative $H_1 : p \in [0,1] \setminus \{1/2\}$. Zeige, dass es keinen gleichmäßig besten Test zum Niveau α für dieses Problem gibt.
Mögliche Beweisidee: Ein gleichmäßig bester Test müsste eine symmetrische Entscheidungsregel besitzen, aber für jedes $p \neq 1/2$ ist die Schärfe eines solchen symmetrischen Tests nicht optimal.

Abgabe in der Vorlesung am Freitag, 27. Mai.

Markus Reiß

Vorlesung
Einführung in die Statistik
Sommersemester 2005

7. Übungsblatt

1. Ein Algorithmus zur Erzeugung von Pseudozufallsziffern soll getestet werden. Bei der Erzeugung von 10.000 Ziffern ergaben sich folgende Häufigkeiten:

Ziffer	0	1	2	3	4	5	6	7	8	9
Häufigkeit	1007	987	928	986	1010	1029	987	1006	1034	1026

   Führe zum Niveau $\alpha = 0,1$ einen χ^2-Anpassungstest auf Gleichverteilung durch.

2. Für $n \geqslant 1$ und $p_0 \in (0,1)$ soll im Binomialmodell $(\{0,\dots,n\}, (\mathrm{Bin}_{n,p})_{p\in[0,1]})$ die Hypothese $H_0: \ p = p_0$ gegen die Alternative $H_1: \ p \neq p_0$ getestet werden.

   (a) Zeige im Detail, dass die Identität
   $$R(x) := \frac{\sup_{p\neq p_0} \mathrm{Bin}_{n,p}(x)}{\mathrm{Bin}_{n,p_0}(x)} = \Big(\frac{x}{np_0}\Big)^x \Big(\frac{n-x}{n-np_0}\Big)^{n-x}, \quad x = 0,\dots,n,$$
   gilt (wie ist der Fall $x \in \{0, n\}$ zu behandeln?). Schließe, dass für $p_0 = 1/2$ der Likelihood-Quotiententest einen symmetrischen Ablehnungsbereich bezüglich $n/2$ besitzt.

   (b) Bei 300 Geburten einer Tierart wurden 168 männliche Nachkommen gezählt. Bestimme $2\log R(168)$ wie auch die Näherung mittels Pearsons χ^2-Statistik. Unter Verwendung der asymptotischen Verteilung dieser Größen entscheide jeweils, ob an der Hypothese, dass im Schnitt ebenso viele männliche wie weibliche Nachkommen geboren werden, mit großer Sicherheit festgehalten werden sollte.

3. Es sei $(\mathfrak{A}_i)_{i\in I}$ eine Familie von σ-Algebren auf einer Grundmenge Ω mit beliebiger Indexmenge I. Beweise:

   (a) $\bigcap_{i\in I} \mathfrak{A}_i$ ist stets wieder eine σ-Algebra auf Ω.

   (b) $\bigcup_{i\in I} \mathfrak{A}_i$ ist im Allgemeinen keine σ-Algebra auf Ω.

4. Es sei Ω eine überabzählbare Grundmenge und $\mathfrak{A}$ die kleinste σ-Algebra auf Ω, die die einelementigen Teilmengen $\{\omega\}$ für alle $\omega \in \Omega$ enthält. Zeige
   $$\mathfrak{A} = \{S \subseteq \Omega \,|\, S \text{ ist abzählbar oder } S^{\complement} \text{ ist abzählbar}\}.$$

Abgabe in der Vorlesung am Freitag, 3. Juni.

Markus Reiß

Vorlesung
Einführung in die Statistik
Sommersemester 2005

8. Übungsblatt

1. Bestimme Konstanten c_1, c_2 derart, dass

$$f_1(x) = \frac{c_1}{\alpha^2 + x^2}, \quad f_2(x) = \begin{cases} 0, & x \leqslant 1, \\ c_2 x^{-r}, & x > 1 \end{cases}$$

mit $\alpha > 0$, $r > 1$ Dichten auf $\mathbb{R}$ sind. Gib jeweils auch die entsprechenden Verteilungsfunktionen an.

2. Es sei φ_{μ,σ^2} die Dichte der Normalverteilung $\mathcal{N}(\mu, \sigma^2)$, also

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(x-\mu)^2}{2\sigma^2}\Big), \quad x \in \mathbb{R}\,.$$

   (a) Bestimme die Extrema und Wendepunkte von φ_{μ,σ^2} und skizziere den Funktionsgraphen.

   (b) Nun sei $\mu = 0$, $\sigma^2 = 1$. Zeige für alle $\eta > 0$ die Abschätzung

$$\tfrac{\eta}{\sqrt{2\pi}(1+\eta^2)} e^{-\eta^2/2} \leqslant \int_\eta^\infty \varphi_{0,1}(x)\,dx \leqslant \tfrac{1}{\sqrt{2\pi\eta^2}} e^{-\eta^2/2}.$$

   Anleitung: für die erste Ungleichung zeige $\int_\eta^\infty (1 + x^{-2}) e^{-x^2/2} dx = \eta^{-1} e^{-\eta^2/2}$ und schließe $\eta^{-1} e^{-\eta^2/2} \leqslant (1+\eta^{-2}) \int_\eta^\infty e^{-x^2/2} dx$, für die zweite Ungleichung benutze $e^{-x^2/2} \leqslant \frac{x}{\eta} e^{-x^2/2}$ falls $x \geqslant \eta$.

   (c) Bestimme unter Verwendung von Teil (b) approximativ die Werte $\int_m^\infty \varphi_{0,1}(x)\,dx$ für $m = 0, 1, 2, 3, 4, 5$.

3. Es sei X eine Zufallsvariable mit Dichte f_X. Zeige, dass $Y = X^2$ folgende Dichte besitzt

$$f_Y(y) = \begin{cases} 0, & y < 0, \\ \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}, & y \geqslant 0. \end{cases}$$

Welche Dichte f_Y ergibt sich im Fall, dass X $\mathcal{N}(0,1)$-verteilt ist?

4. In einem Kreis vom Radius r werde „rein zufällig“ eine Sehne ausgewählt. Mit welcher Wahrscheinlichkeit ist die Länge dieser Sehne größer als r? Verwende folgende Zufallsbeschreibungen:

   (a) Die Sehne ist durch ihren Mittelpunkt eindeutig bestimmt. Die Lage des Mittelpunkts ist gleichmäßig in der Kreisscheibe verteilt.

   (b) Die Sehne ist durch ihre Endpunkte eindeutig bestimmt und aus Symmetriegründen wählen wir den einen Endpunkt fest. Der andere möge gleichmäßig auf dem Kreisrand verteilt sein.

   (c) Die Sehne ist durch ihren Abstand vom Kreismittelpunkt und die entsprechende Richtung eindeutig festgelegt. Aus Symmetriegründen kann die Richtung fest gewählt werden, der Abstand sei gleichmäßig auf $[0, r]$ verteilt.

Abgabe in der Vorlesung am Freitag, 10. Juni.

Markus Reiß

Vorlesung
Einführung in die Statistik
Sommersemester 2005

9. Übungsblatt

1. Es sei (X, Y) ein zweidimensionaler Zufallsvektor mit Dichte

$$f^{X,Y}(x,y) = \begin{cases} \lambda^2 \exp(-\lambda y), & \text{falls } 0 \leqslant x \leqslant y, \\ 0, & \text{sonst} \end{cases}$$

für einen Parameter $\lambda > 0$.

   (a) Überprüfe, ob $f^{X,Y}$ wirklich eine Wahrscheinlichkeitsdichte ist.

   (b) Bestimme die Randdichten f^X und f^Y von X bzw. Y.

   (c) Sind X und Y unabhängig?

2. Es seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit Verteilungsfunktionen $F_1, \ldots, F_n$.

   (a) Zeige, dass die Verteilungsfunktionen von $M = \max(X_1, \ldots, X_n)$ und $m = \min(X_1, \ldots, X_n)$ gegeben sind durch

$$F_M(x) = \prod_{i=1}^{n} F_i(x) \quad \text{bzw.} \quad F_m(x) = 1 - \prod_{i=1}^{n}(1 - F_i(x)).$$

   (b) Bestimme für den Fall, dass jedes X_i gleichmäßig auf $[0, 1]$ verteilt ist, die Dichte von M und m. Sind M und m unabhängig?

   (c) Die Zeit bis zum Zerfall eines radioaktiven Atoms wird durch die Exponentialverteilung $\mathrm{Exp}(\lambda)$ beschrieben. Wie ist die Zeit bis zum ersten Atomzerfall in einer Probe aus N solchen Atomen verteilt?

3. Es seien $X_1, \ldots, X_4$ unabhängige, gleichmäßig auf $[-1, 1]$ verteilte Zufallsvariablen. Bestimme die Dichte von

$$S_n^* := \left(\frac{3}{n}\right)^{1/2} \sum_{k=1}^{n} X_k$$

für jedes $n \in \{1, 2, 3, 4\}$. Zeichne die Dichtefunktionen zusammen mit der Dichte der Standardnormalverteilung in ein Koordinatensystem ein.

4. Ein Stab der Länge 1 wird zufällig in zwei Teile gebrochen, wobei der Bruchpunkt gleichmäßig verteilt ist. Danach wird ebenso zufällig das längere der beiden Teile in zwei Stücke gebrochen. Wie groß ist die Wahrscheinlichkeit, dass sich aus den entstandenen drei Stäben ein Dreieck legen lässt?

Abgabe in der Vorlesung am Freitag, 17. Juni.

Markus Reiß

Vorlesung
Einführung in die Statistik
Sommersemester 2005

10. Übungsblatt

1. Es sei X eine $\mathrm{Exp}(\lambda)$-verteilte Zufallsgröße. Bestimme $\mathbb{E}[X]$, $\mathrm{Var}(X)$ sowie die Korrelation zwischen X und X^2 (Tipp: partielle Integration).

2. Es sei X eine $\mathcal{N}(\mu, \sigma^2)$-verteilte Zufallsvariable. Benutze die Tschebyschew-Ungleichung, um $\mathbb{P}(|X - \mu| \geqslant k\sigma)$ für $k = 1, 2, 3$ abzuschätzen. Vergleiche mit den genaueren Werten, die sich ergeben aus

   $$\Phi(1) \approx 0,8413,\ \Phi(2) \approx 0,9772,\ \Phi(3) \approx 0,9987 \text{ mit } \Phi(x) = \int_{-\infty}^{x} \tfrac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

   Gibt es Zufallsvariablen, für die Gleichheit in der Tschebyschew-Ungleichung gilt?

3. Zur stochastischen Bestimmung von π wird folgendes Verfahren vorgeschlagen: Es seien $X_1, \dots, X_n, Y_1, \dots, Y_n$ unabhängige gleichmäßig auf $[0, 1]$ verteilte Zufallsvariablen (*Standard-Zufallszahlen*) sowie

   $$A_n := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{[0,1]}(X_i^2 + Y_i^2).$$

   (a) Weise nach, dass $\mathbb{E}[A_n] = \pi/4$ gilt (d.h. $4A_n$ gibt eine sinnvolle Schätzung von π an).

   (b) Berechne $\mathrm{Var}[A_n]$.

   (c) Bestimme für $k = 0, 1, 2, 3, 4$ eine möglichst kleine Zahl n derart, dass $\mathbb{P}(|A_n - \pi/4| \geqslant 10^{-k}) \leqslant 0,1$ gilt (z.B. mittels Tschebyschew-Ungleichung). Wie viele Glieder braucht man jeweils für die Genauigkeit 10^{-k} in der Leibnizschen Summenformel

   $$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + - \cdots ?$$

4. Es seien $(X_n)_{n \geqslant 1}$ Zufallsvariablen, die fast sicher gegen eine Zufallsvariable X konvergieren. Folgere schrittweise:

   (a) $\mathbb{P}(\exists \varepsilon > 0\ \forall n \geqslant 1 : \sup_{m \geqslant n} |X_m - X| > \varepsilon) = 0$.

   (b) $\forall \varepsilon > 0 :\ \mathbb{P}(\bigcap_{n=1}^{\infty} \{\sup_{m \geqslant n} |X_m - X| > \varepsilon\}) = 0$.

   (c) $\forall \varepsilon > 0 :\ \lim_{n \to \infty} \mathbb{P}(\sup_{m \geqslant n} |X_m - X| > \varepsilon) = 0$.

   (d) X_n konvergiert gegen X stochastisch.

Abgabe in der Vorlesung am Freitag, 24. Juni.

Markus Reiß

Vorlesung
Einführung in die Statistik
Sommersemester 2005

11. Übungsblatt

1. Ein Spieler startet mit dem Anfangskapital $K_0 = 1$. Bei jeder Runde $i = 1, \dots, n$ setzt er sein gesamtes Kapital ein, es wird eine faire Münze geworfen, und bei 'Kopf' erhält er den anderthalbfachen Einsatz zurück, bei 'Zahl' nur den halben.

   (a) Stelle das Kapital nach der n-ten Runde als $K_n = \prod_{i=1}^n R_i$ mit geeigneten unabhängigen Zufallsvariablen R_i dar.

   (b) Weise nach, dass das Spiel fair ist in dem Sinne, dass $\mathbb{E}[K_n] = 1$ gilt.

   (c) Zeige, dass trotzdem $\lim_{n\to\infty} K_n = 0$ (fast sicher) gilt.
   Tipp: Wende das starke Gesetz der großen Zahlen auf $\log(K_n)$ an.

2. Beweise für Zufallsvariablen $(X_n)_{n\geqslant 1}$, X:

   (a) Nehmen X_n, $n \geqslant 1$, und X nur Werte in $\mathbb{N}$ an, so gilt $X_n \xrightarrow{\mathcal{L}} X$ genau dann, wenn $\lim_{n\to\infty} \mathbb{P}(X_n = m) = \mathbb{P}(X = m)$ für alle $m \in \mathbb{N}$ gilt.

   (b) Besitzen X_n, $n \geqslant 1$, die Dichten f_n und X die Dichte f, so gilt die Implikation

   $$\lim_{n\to\infty} \int_{-\infty}^{\infty} |f_n(x) - f(x)|\, dx = 0 \Rightarrow X_n \xrightarrow{\mathcal{L}} X.$$

   (c) *Zusatzaufgabe:* Die Umkehrung dieser Implikation gilt nicht.
   Tipp: betrachte die Verteilungsfunktionen zu den Dichtefunktionen $f_n(x) = 2\sum_{k=1}^{2^n} \mathbf{1}_{[k2^{-n}-2^{-n-1}, k2^{-n}]}(x)$.

3. Um den Anteil p der Raucher in der Bevölkerung zu schätzen, werden n zufällig ausgewählte Bürger befragt und der Anteil S_n/n der Anzahl S_n von Rauchern an der Gesamtzahl als Schätzung verwendet.

   (a) Begründe, weshalb S_n näherungsweise als $\mathrm{Bin}_{n,p}$-verteilt angesehen werden kann.

   (b) Verwende die Normalapproximation, um folgende Näherung für einen Schätzfehler um mehr als einen Prozentpunkt herzuleiten:

   $$\mathbb{P}(|S_n/n - p| > 0,01) \approx 2\Big(1 - \Phi\Big(\frac{0,01n}{\sqrt{np(1-p)}}\Big)\Big)$$

   mit der Verteilungsfunktion Φ der Standardnormalverteilung.

   (c) Verwende die Ungleichung $p(1-p) \leqslant 1/4$, um eine Stichprobengröße n zu bestimmen, für die gilt $\mathbb{P}(|S_n/n - p| > 0,01) \leqslant 0,05$.

4. Bei einem Experiment soll die Hypothese, dass eine Münze fair ist, durch das Zählen von 'Kopf' bei n Würfen getestet werden. Bestimme für $n = 10, 100, 300$ jeweils einen nicht-randomisierten Test vom ungefähren Niveau $\alpha = 0,05$ mit symmetrischem Ablehnungsbereich (für $n = 100, 300$ darf die Normalapproximation verwendet werden). Wie groß ist jeweils die Wahrscheinlichkeit für einen Fehler zweiter Art, falls die Münze unfair ist und mit Wahrscheinlichkeit $p = 0,6$ 'Kopf' zeigt?

Abgabe in der Vorlesung am Freitag, 1. Juli.

Markus Reiß

Vorlesung
Einführung in die Statistik
Sommersemester 2005

12. Übungsblatt

1. Es seien $X_1, \ldots, X_n$ unabhängige identisch verteilte Zufallsvariablen in L^2. Setze $\mu = \mathbb{E}[X_1]$, $\sigma^2 = \mathrm{Var}(X_1)$. Weise nach, dass folgende Schätzer erwartungstreu sind:

   (a) $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i$ als Schätzer von μ;

   (b) $\tilde{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$ als Schätzer von σ^2, falls μ bekannt ist;

   (c) $\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \hat{\mu}_n)^2$ als Schätzer von σ^2, falls μ unbekannt ist.

   Zeige ferner, dass $\hat{\mu}_n$ und $\tilde{\sigma}_n^2$ konsistent sind für $n \to \infty$.

2. Für $\vartheta > 0$ seien $X_1, \ldots X_n$ unabhängige Beobachtungen, die gleichmäßig auf dem Intervall $[0, \vartheta]$ verteilt sind. Es werden folgende drei Schätzer für ϑ vorgeschlagen:

$$\hat{\vartheta}_1 = \frac{2}{n}\sum_{i=1}^n X_i, \quad \hat{\vartheta}_2 = \max(X_1, \ldots, X_n), \quad \hat{\vartheta}_3 = \tfrac{n+1}{n}\hat{\vartheta}_2.$$

   (a) Zeige, dass $\hat{\vartheta}_2$ die Dichte $f_2(x) = n\vartheta^{-n}x^{n-1}\mathbf{1}_{[0,\vartheta]}(x)$ besitzt.

   (b) Zeige, dass $\hat{\vartheta}_1$ und $\hat{\vartheta}_3$ erwartungstreu sind, aber nicht $\hat{\vartheta}_2$.

   (c) Beweise, dass der mittlere quadratische Fehler $\mathbb{E}[(\hat{\vartheta}_k - \vartheta)^2]$, $k = 1, 2, 3$, für alle $\vartheta > 0$ und $n \geqslant 1$ im Fall $k = 3$ am kleinsten ist.

*3. Es seien X und Y unabhängige $\mathbb{N}_0$-wertige Zufallsvariablen mit Zähldichten p^X bzw. p^Y.

   (a) Zeige, dass $X + Y$ folgende Zähldichte besitzt:

$$p^{X+Y}(m) = \sum_{k=0}^m p^X(k)p^Y(m-k), \quad m \geqslant 0, \qquad (\textit{diskrete Faltung}).$$

   (b) Schließe, dass $X + Y$ $\mathrm{Poiss}_{\lambda+\mu}$-verteilt ist, wenn X Poiss_λ-verteilt und Y Poiss_μ-verteilt ist.

*4. Aufgabe 4 auf Blatt 5

*5. Unter 10.000 Geburten einer Tierart wurden 4982 männliche Jungtiere gezählt. Kann man auf Grund dieses Ergebnisses mit einer Sicherheit von 95% an der Hypothese festhalten, dass die Wahrscheinlichkeit einer Männchengeburt mindestens 50% beträgt? Formuliere zunächst das Testproblem mathematisch, begründe die Auswahl des verwendeten Tests und verwende zur Berechnung die Normalapproximation.

*6. Um die Strahlenbelastung bei Waldpilzen zu überprüfen, wird bei n unabhängigen Pilzproben die Anzahl X_i, $i = 1, \dots, n$, der Geigerzählerimpulse während einer festen Zeiteinheit gemessen.

   (a) Begründe, weshalb X_i als Poisson-verteilt angenommen werden kann.

   (b) Unter der Annahme, dass $X_1, \dots, X_n$ unabhängig und jeweils Poiss_λ-verteilt sind, konstruiere einen geeigneten erwartungstreuen Schätzer $\hat{\lambda}$ von λ.

   (c) Ist dieser Schätzer für $n \to \infty$ konsistent?

Abgabe in der Vorlesung am Freitag, 8. Juli. Die Aufgaben 3-6 sind freiwillig und dienen der Klausurvorbereitung.

Markus Reiß

Vorlesung
Einführung in die Statistik
Sommersemester 2005

13. Übungsblatt