WS 2008/2009 *Name:*
Matrikel-Nr.:

ERSTE KLAUSUR
STOCHASTIK I – WAHRSCHEINLICHKEITSTHEORIE

3. FEBRUAR 2009
HILFSMITTEL: A4 BLATT MIT NOTIZEN, TR

Die Prüfung besteht aus zwei Teilen, die insgesamt 300 Punkte ergeben:
Teil 1: 10 Multiple Choice (MC) Aufgaben mit jeweils 5 Punkten
Teil 2: 5 von 8 Aufgaben mit jeweils 50 Punkten

Bitte ankreuzen, welche Aufgaben bewertet werden sollen (5 von 8)!

Aufgabe:	1	2	3	4	5	6	7	8	MC	Gesamt
Bewerten:										
Punkte:										

Schreiben Sie bitte auf jedes Blatt oben rechts Ihren Namen!

Viel Erfolg!

Teil 1. **Multiple Choice**

Zu jeder Frage ist genau eine richtige Antwortmöglichkeit vorgegeben. Tragen Sie Ihre Lösungen in die Kästchen auf der **übernächsten** Seite ein. Die Rückseiten der Blätter können Sie für Berechnungen sowie zu Anmerkungen und Erläuterungen Ihrer Lösungen verwenden.

(1) Der `R`-Befehl `mean(sample(1:6, 100, replace = TRUE))`
  (a) berechnet den Mittelwert der Zahlen $1, 2, \ldots, 6$ und multipliziert das Ganze mit 100.
  (b) berechnet den Mittelwert einer Zufallsstichprobe aus $\{1, 2, \ldots, 6\}$ der Länge 100 ohne Zurücklegen.
  (c) berechnet den Mittelwert von 100 simulierten Würfelwürfen.
  (d) gibt eine Fehlermeldung zurück, da `replace = TRUE` als Argument nicht erlaubt ist.

(2) Welche Aussage ist richtig?
  (a) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) + P(A \cap B)$
  (b) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - P(A \cap B)$
  (c) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$
  (d) $\mathbb{P}(A \cup B) = \mathbb{P}(A)\mathbb{P}(B)$

(3) Der Zentrale Grenzwertsatz besagt, dass
  (a) die Summenvariable in Verteilung gegen die Standard-Normalverteilung konvergiert.
  (b) die Summenvariable in Verteilung gegen die Gleichverteilung konvergiert.
  (c) die standardisierte Summenvariable in Verteilung gegen den Erwartungswert konvergiert.
  (d) die standardisierte Summenvariable in Verteilung gegen die Standard-Normalverteilung konvergiert.

(4) Für die Korrelation zwischen den Zufallsvariablen X und Y gilt nicht?
  (a) Sie mißt die lineare Abhängigkeit von X und Y.
  (b) Für X und Y linear unabhängig ist die Korrelation 0.
  (c) $\rho_{xy} = \frac{Cov(Y,X)}{\sigma_x \sigma_y}$
  (d) $-1 < \rho_{xy} < 1$

(5) Welches Ergebnis liefert `ppois(3,2)`?
  (a) 0.857 (b) 0.423 (c) 0.180 (d) 0.722

(6) Man wirft mehrfach zwei Würfel. Wie groß ist die Wahrscheinlichkeit, dass das Ereignis `Augensumme gleich 7` vor dem Ereignis `Augensumme gleich 5` erscheint?
  (a) 0.6 (b) $\frac{17}{18}$ (c) $1 - \frac{24}{36^2}$ (d) $\lim_{n\to\infty}(\frac{6}{36} + (1 - (\frac{4}{36})))^n$

(7) $\varphi_X(t) := E(e^{itX})$ ist die charakteristische Funktion der reellen Zufallsvariablen X mit Dichtefunktion f. Welche der folgenden Aussagen ist nicht richtig?
  (a) $|\varphi_X(t)| \geq 1$
  (b) $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$ für alle $a, b \in \mathbb{R}$
  (c) $E(X) = \varphi'_X(0)/i$
  (d) $E(X^2) = -\varphi''_X(0)$

(8) Bei einem Messvorgang wird angenommen, dass er durch eine Zufallsvariable mit unbekanntem Erwartungswert μ und einer Varianz $\sigma^2 = 0.01$ angemessen beschrieben werden kann. Wieviele unabhängige Messungen müssen mindestens durchgeführt werden, so dass mit einer Wahrscheinlichkeit von größer 95% der Betrag der Differenz zwischen dem arithmetischen Mittel der Messwerte und μ kleiner als 0.02 ist? Man beantworte diese Frage durch Anwendung der Ungleichung von Tschebyschew.
  (a) 25 (b) 250 (c) 500 (d) 96

(9) Beim deutschen Lottosystem werden 6 aus 49 mit Zusatzzahl $(0, 1, \ldots, 9)$ gezogen. Jetzt soll das neue Euro-Lotto eingeführt werden. Hier werden 5 aus 50 und zusätzlich 2 aus 9 gezogen. Es ist daher

$$P_{alt}(\text{6 Richtige ohne Zusatzzahl})/P_{neu}(\text{5 Richtige mit 1 Zusatzzahl})$$

identisch
  (a) 1 (b) $\frac{\binom{43}{6}\binom{6}{0}\binom{9}{1}}{\binom{45}{5}\cdot\binom{5}{0}\binom{7}{1}\cdot 2}$ (c) $\frac{\binom{50}{5}}{\binom{49}{6}} \cdot \frac{18^2}{140}$ (d) $\frac{\binom{50}{5}\binom{9}{1}\binom{9}{2}}{\binom{49}{6}\cdot\binom{7}{1}\binom{10}{1}}$

(10) Michael Scofield versucht, seinen unschuldig zum Tode verurteilten Bruder Lincoln Burrows aus dem Fox-River-Gefängnis zu befreien. Kurz vor Ausbruch führt aber wegen eines Gefängnisumbaus nur einer von drei möglichen Fluchtwegen zur Freiheit, mit erwarteter Fluchtzeit von 5 Stunden, während die anderen zwei Wege zurück zum Anfangspunkt führen, mit jeweils erwarteter Fluchtzeit von 7 und 9 Stunden. Es sei angenommen, dass Michael und Lincoln genügend Zeit haben und die drei Fluchtwege jedesmal mit gleicher Wahrscheinlichkeit auswählen. Wie lange ist die erwartete Dauer bis beide aus dem Gefängnis herausfinden?
(a) 7 Stunden (b) 14 Stunden (c) 21 Stunden (d) 28 Stunden

Antworten:

1	2	3	4	5	6	7	8	9	10

Verteilungstabelle der Standardnormalverteilung

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.000	0.5000	0.750	0.7734	1.500	0.9332	2.250	0.9878
0.025	0.5100	0.775	0.7808	1.525	0.9364	2.275	0.9885
0.050	0.5199	0.800	0.7881	1.550	0.9394	2.300	0.9893
0.075	0.5299	0.825	0.7953	1.575	0.9424	2.325	0.9900
0.100	0.5398	0.850	0.8023	1.600	0.9452	2.350	0.9906
0.125	0.5497	0.875	0.8092	1.625	0.9479	2.375	0.9912
0.150	0.5596	0.900	0.8159	1.650	0.9505	2.400	0.9918
0.175	0.5695	0.925	0.8225	1.675	0.9530	2.425	0.9923
0.200	0.5793	0.950	0.8289	1.700	0.9554	2.450	0.9929
0.225	0.5890	0.975	0.8352	1.725	0.9577	2.475	0.9933
0.250	0.5987	1.000	0.8413	1.750	0.9599	2.500	0.9938
0.275	0.6083	1.025	0.8473	1.775	0.9621	2.525	0.9942
0.300	0.6179	1.050	0.8531	1.800	0.9641	2.550	0.9946
0.325	0.6274	1.075	0.8588	1.825	0.9660	2.575	0.9950
0.350	0.6368	1.100	0.8643	1.850	0.9678	2.600	0.9953
0.375	0.6462	1.125	0.8697	1.875	0.9696	2.625	0.9957
0.400	0.6554	1.150	0.8749	1.900	0.9713	2.650	0.9960
0.425	0.6646	1.175	0.8800	1.925	0.9729	2.675	0.9963
0.450	0.6736	1.200	0.8849	1.950	0.9744	2.700	0.9965
0.475	0.6826	1.225	0.8897	1.975	0.9759	2.725	0.9968
0.500	0.6915	1.250	0.8944	2.000	0.9772	2.750	0.9970
0.525	0.7002	1.275	0.8988	2.025	0.9786	2.775	0.9972
0.550	0.7088	1.300	0.9032	2.050	0.9798	2.800	0.9974
0.575	0.7174	1.325	0.9074	2.075	0.9810	2.825	0.9976
0.600	0.7257	1.350	0.9115	2.100	0.9821	2.850	0.9978
0.625	0.7340	1.375	0.9154	2.125	0.9832	2.875	0.9980
0.650	0.7422	1.400	0.9192	2.150	0.9842	2.900	0.9981
0.675	0.7502	1.425	0.9229	2.175	0.9852	2.925	0.9983
0.700	0.7580	1.450	0.9265	2.200	0.9861	2.950	0.9984
0.725	0.7658	1.475	0.9299	2.225	0.9870	2.975	0.9985

Teil 2. Bearbeiten Sie 5 der 8 Aufgaben!

1. Dreiecksverteilung

Gegeben sei folgende Dichte auf dem Intervall $[a, b]$:

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & a \le x \le c \\ \frac{2(b-x)}{(b-a)(b-c)} & c < x \le b \end{cases}$$

(a) Skizzieren Sie die obige Dichte.

(b) Zeigen Sie, dass der Erwartungswert $\frac{a+b+c}{3}$ beträgt und geben Sie den Modus an.

(c) Zeigen Sie, dass die zugehörige Verteilungsfunktion

$$F(x) = \begin{cases} \frac{(x-a)^2}{(b-a)(c-a)} & a \le x \le c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & c < x \le b \end{cases}$$

ist, und skizzieren Sie diese. Bestimmen Sie den Median der Dreiecksverteilung.

(d) Seien nun die die Parameter a, b, c wie folgt bekannt:

$$a = 10, b = 25, c = 20$$

Berechnen Sie die Wahrscheinlichkeit $\mathbb{P}(12 \le x < 21)$.

(e) Diskutieren Sie die folgenden Anwendungen einer Dreiecksverteilung. Gehen Sie hierbei insbesondere auf die Sinnhaftigkeit einer Verwendung der Dreiecksverteilung ein.

- Die Verteilung der Noten einer Klausur folgen einer Dreiecksverteilung.
- Die Kosten für eine Dienstleitung sollen als dreiecksverteilt angesehen werden, wobei nur das Minimum, Maximum und der Modus bekannt sind.
- Das Sterbealter in Deutschland soll als dreiecksverteilt modelliert werden.

2. Schneefall

Am 03.12.2008 war bei Spiegel Online zu lesen:

In Deutschland hat ein plötzlicher Wintereinbruch die Verkehrsteilnehmer offenbar völlig überrascht. Der Berufsverkehr auf Autobahnen und in vielen Städten kam am Morgen fast völlig zum Erliegen viele Autofahrer, beklagt die Polizei, hatten noch keine Winterbereifung aufgezogen.
Besonders in den Mittelgebirgen wie Eifel und Hunsrück, wo über Nacht 20 Zentimeter Neuschnee gefallen waren, sowie im Sauerland und Taunus stockte der Verkehr, weil Lastwagen an Steigungen steckenblieben.
In der Eifel war auf der B51 wegen querstehender Lastwagen zeitweise kein Durchkommen mehr, teilte die Polizei mit. Auf der Autobahn 3 entstand bei geschlossener Schneedecke ein Stau von mindestens 40 Kilometern. Im Hochsauerlandkreis kam es bereits vor 7 Uhr zu 15 Unfällen.

Bearbeiten Sie die nachfolgenden Aufgaben.

(a) Nehmen Sie an, die Menge an Neuschnee NS (in Zentimeter) ist normalverteilt mit Erwartungswert 15 und Varianz 9. Wie groß ist die Wahrscheinlichkeit, dass mehr als 20 Zentimeter Neuschnee fallen?

(b) Die Anzahl Autounfälle AU ist pro Autobahnabschnitt abhängig von der Menge an Neuschnee wie folgt:

$$\begin{aligned} AU_{p1} &\sim pois_{10}, \qquad NS < 20 \\ AU_{p2} &\sim pois_{20}, \qquad NS \geq 20 \end{aligned}$$

Man Berechne die Wahrscheinlichkeit für $NS \geq 20$ bei 15 vorliegenden Unfällen.

(c) Eine weitere Modellierung der Anzahl Autounfälle pro Autobahnabschnitt ist folgendermassen mittels der Exponentialverteilung definiert:

$$AU_e \sim exp_{1/NS}$$

Skizzieren Sie die Dichte der Exponentialverteilung für $NS = 2$. Berechnen Sie die Wahrscheinlichkeit für mehr als 15 Unfälle bei 20 Zentimeter Neuschnee.

(d) Welches Modell aus (b) und (c) erscheint realistischer? Begründen Sie Ihre Antwort.

(e) Mittels des Modells aus (b) wird die Wahrscheinlichkeit für mehr als 15 Unfälle bei 20 Zentimeter Neuschnee berechnet, diese beträgt 0.895. Berechnen Sie hiermit die Wahrscheinlichkeit, dass auf höchstens 2 von 9 Autobahnabschnitten mehr als 15 Autounfälle passieren. Welche Annahmen haben Sie hierbei getroffen und wie sinnvoll sind diese? Mit welchem `R`-Befehl könnte man diese Wahrscheinlichkeit berechnen?

3. Zentraler Grenzwertsatz

Der zentrale Grenzwertsatz soll bewiesen werden, der da lautet:

"Sei $X_1, X_2, \ldots$ eine Folge von reellwertigen, unabhängigen und identisch verteilten Zufallsvariablen mit $E[X_i] = \mu < \infty$ und $V[X_i] = \sigma^2 < \infty$, und sei $S_n = \sum_{i=1}^n X_i$. Dann gilt, dass

$$U_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

in Verteilung gegen $N(0,1)$ konvergiert."

(a) Zeigen Sie, dass $Y_i = (X_i - \mu)/\sigma$ die charakteristische Funktion (CF)

$$\zeta_Y(t) = 1 - \frac{1}{2}t^2 + o(t^2)$$

hat.

(b) Zeigen Sie, dass $U_n = (1/\sqrt{n}) \sum_{i=1}^n Y_i$ ist, und dass U_n die CF

$$\psi_n(t) = \left\{1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right\}^n$$

hat.

(c) Gegen welche Funktion konvergiert die CF $\psi_n(t)$ für $n \to \infty$? Letztere Grenzfunktion ist CF welcher Verteilung?

(d) Schließen Sie daher, dass U_n in Verteilung gegen $N(0,1)$ konvergiert. Welche Eigenschaften über charakteristische Funktionen sind hierbei zentral?

Zeigen Sie, dass aus dem zentralen Grenzwertsatz folgende Anwendung folgt:

(e) "Falls X binomialverteilt ist mit den Parametern N und p, so gilt für hinreichend großes N, dass X annähernd normalverteilt ist mit den Parametern $\mu = Np$ und $\sigma^2 = Np(1-p)$."

4. Sonographie

Im folgenden geht es um einen Datensatz zu Geburten im Klinikum Augsburg. Neben Geburtsposition, Schwangerschaftswoche und Kopfumfang des Kindes sind auch sonographische Messungen (Ultraschall) durchgeführt worden. Diese dienen unter anderem der besseren Vorhersage des Geburtsverlaufs.

(a) Im folgenden finden Sie eine (unvollständige) Tabelle der Art der Geburt V_G des Kindes und der Variable Frühgeburt $V_p, (1 = Ja)$. Nehmen Sie an, es sei bekannt, dass $P(V_p = 0|V_g = Gepl.K.schnitt) = 3.46796 \cdot P(V_p = 0) \cdot P(V_g = Gepl.K.schnitt)$ gilt. Schätzen Sie die nötigen Wahrscheinlichkeiten aus den Daten und ergänzen Sie die Tabelle. Sind diese Variablen in den Daten annähernd unabhängig?

	$V_p = 0$	$V_p = 1$	TOTAL
Notfall-Kaiserschnitt	XXX	79	XXX
Geplanter Kaiserschnitt	434	378	812
Normale Geburt	2525	XXX	XXX
Saugglocke	316	18	334
TOTAL	XXX	XXX	4343

(b) Mittels Sonographie wird vor der Geburt der Kopfdurchmesser geschätzt. Der daraus berechnete Umfang betrage im Mittel $29.53cm$ mit einer Standardabweichung von $0.71cm$. Nach der Geburt wird der wahre Kopfumfang des Kindes gemessen und es ergibt sich durchschnittlich ein Wert von $34cm$ mit einer Standardabweichung von $2.35cm$.

(i) Wie gross ist die Wahrscheinlichkeit, dass der Kopfumfang unterschätzt wird? Treffen Sie geeignete Annahmen, um die Berechnung durchführen zu können.

(ii) Tatsächlich liegt der wahre Wert durchschnittlich um $4.47cm$ über der Vorhersage mit einer Standardabweichung von $1.39cm$. Wie ist der Unterschied zu Ihrem vorherigen Ergebnis zu erklären?

(c) Die Anzahl an Geburten pro Tag sei poissonverteilt mit Parameter $\lambda = 6$. Zudem handle es sich mit Wahrscheinlichkeit $p_z = 0.07$ bei einer Geburt um Zwillinge und mit $p_d = 0.01$ um Drillinge. Wieviele Kinder sind innerhalb einer Woche zu erwarten?

(d) Dr. Haus interessiert sich für die Schwangerschaftsdauer D bei Geburten, die nicht als Frühgeburten gelten (nach der 37. Schwangerschaftswoche, Abb. 1). Bekannt sind der Mittelwert $\mu = 39.7$ und die Standardabweichung $\sigma = 1.16$.
Er fragt am Lehrstuhl von Prof. U. nach Rat und wird an die Assistenten A_1, A_2 und A_3 verwiesen. Die drei sind sich einig, dass eine Normalverteilung benutzt werden sollte.
A_1 schlägt $D \sim N(\mu = 39.7, \sigma = 1.16)$ vor.
A_2 dagegen plädiert für $P(D = d) = P(Y = d - 37 | Y > 0)$ mit $Y \sim N(2.7, 1.16)$ für $d \geq 37$ und $P(D = d) = 0$ sonst.
A_3 schließt sich A_2 an, meint jedoch, man sollte lieber $Y \sim N(\mu - 37 - \tau, \sigma + \psi)$ mit geeigneten Werten $\tau, \psi > 0$ verwenden (ohne Berechnung!).

Beurteilen Sie alle genannten Vorschläge und geben Sie die von A_2 vorgeschlagene Dichte an. Welcher Empfehlung würden Sie sich am ehesten anschliessen?
Geben Sie außerdem eine eigene Empfehlung und begründen Sie diese!

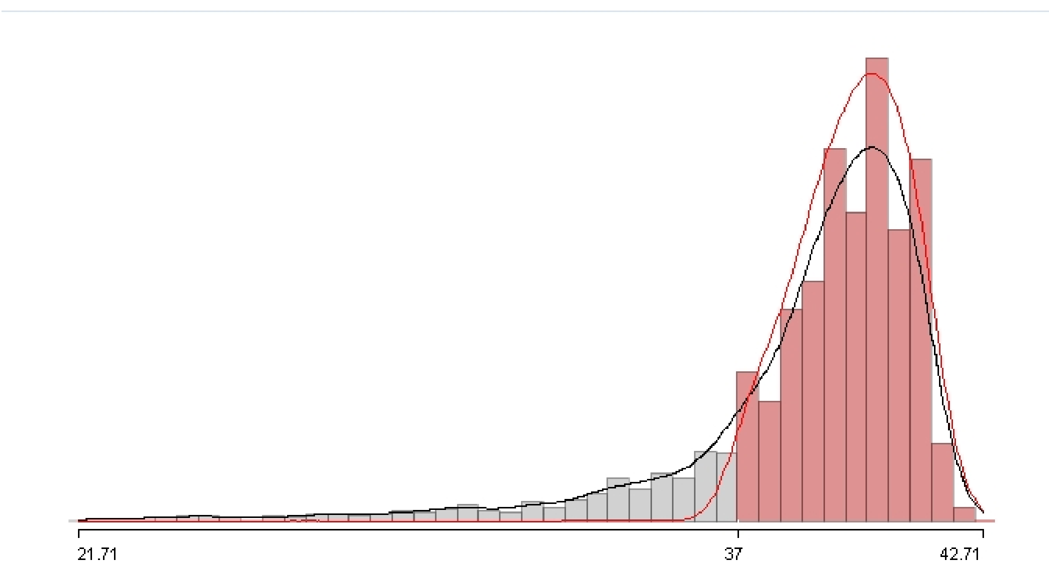


Abbildung 1. Histogramm der Schwangerschaftswoche. Rot gehighlighted sind die Nicht-Frühgeburten (37+)

5. Präsidentenfasching

Barack Obama und seine Frau beschließen ihre europäischen Freunde zum Staatsmänner-Fasching einzuladen. Die Gäste sind Angela Merkel (D), Silvio Berlusconi (I), Nicolas Sarkozy (F) und Gordon Brown (GB) jeweils mit Ehepartner.
Von jedem Gast wird eine Gesichtsmaske angefertigt. Vor Einlass in den Saal wählt jeder Gast eine der 10 Masken zufällig aus.
Am kreisrunden Tisch nehmen dann alle entsprechend der Masken neben Ihren Ehepartnern Platz.

(a) Wie groß ist die Wahrscheinlichkeit, dass die wahren Ehepaare alle nebeneinander sitzen?

(b) Angela Merkel hat aus ihrer Heimat 20 Krapfen mitgebracht. Es gibt zwei Füllungen (Pudding, Marmelade), die jeweils 10 mal vetreten sind. Nehmen Sie an, alle Gäste wählen jeweils zufällig zwei der Krapfen aus.
Wie groß ist die Wahrscheinlichkeit, dass jeder von ihnen beide Füllungen bekommt? (Die Angabe einer Formel genügt!)

(c) Die Zeit zum Verzehr eines Krapfens sei für alle Teilnehmer unabhängig normalverteilt. Der Erwartungswert liege für Frauen bei $\mu_F = 3$ Min., für Männer bei $\mu_M = 5$ Min., und die Standardabweichungen seien $\sigma_F = 1$ Min. sowie $\sigma_M = 2$ Min.

(i) Jeder Teilnehmer erhält zwei Krapfen. Wie groß ist die Wahrscheinlichkeit, dass nach 10 Minuten alle Krapfen verspeist sind?

(ii) Geben Sie eine allgemeine (in Dichte und Verteilungsfunktion ausgedrückte) Formel für die Wahrscheinlichkeit an, dass alle Frauen vor dem ersten Mann aufgegessen haben.

(iii) Wie wahrscheinlich ist es, dass eine Frau

- beide Krapfen jeweils in negativer Zeit verspeist?
- beide Krapfen zusammen in negativer Gesamtzeit verspeist?

Welche der vorhergehenden zwei Wahrscheinlichkeiten ist größer? Sind Frauen diesbezüglich besser modelliert als Männer? Wie kann man dies bereits an deren Verteilungen erkennen?

6. Gleichverteilung

$N = 10$ Studenten treffen sich zum gemeinsamen Lernen und vereinbaren als Treffpunkt 17.30Uhr. Man kann davon ausgehen, dass alle Studenten unabhängig voneinander, gleichverteilt zwischen 17.20Uhr und 17.40Uhr eintreffen.

(a) Welcher Verteilungsfunktion folgt der Ankunftszeitpunkt des als vorletzten eintreffenden Studenten?

(b) Wie groß muss N mindestens sein, damit die zu erwartende Ankunftszeit des letzten Studenten mindestens 17.39Uhr ist?

(c) Angenommen, der letzte Student trifft um Punkt 17.37Uhr ein. Was ist nun die zu erwartende durchschnittliche Ankunftszeit der anderen neun Studenten?

Das Treffen findet sieben Mal im Semester statt.

(d) Wie wahrscheinlich ist es, dass die Studentin Julia immer vor Christiane eintrifft?

7. Pumpe

Eine Pumpe sei ununterbrochen in Betrieb, bis sie ausfalle. Die Zufallsvariable X, die die zufällige Dauer der Funktionsfähigkeit der Pumpe beschreibt, möge stetig verteilt sein mit einer Dichte der Form:

$$f(x) = \begin{cases} \lambda^2 x e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Weiter sei bekannt, dass Pumpen dieser Art im Mittel 100 Stunden laufen, bis sie ausfallen.

(a) Wie ist der Parameter λ zu wählen, damit der Erwartungswert von X gleich der mittleren Laufzeit ist?

(b) Man bestimme die folgenden Wahrscheinlichkeiten:

  (i) $\mathbb{P}(X \leq 100)$,

  (ii) $\mathbb{P}(X \leq 100(N+1) | X \geq 100N)$ für $N \to \infty$.

In einem Wohnblock sind 10 Pumpen unabhängig voneinander parallel im Einsatz. Alle 50 Stunden wird jeweils geprüft, ob die Pumpen noch laufen und die, die nicht mehr laufen, werden durch neue gleichartige ersetzt. Da der Block gerade frisch renoviert wurde sind alle Pumpen neu.

(c) Was ist die Wahrscheinlichkeit dafür, dass nach den ersten 50 Stunden keine Pumpe ausgetauscht werden muss.

(d) Gegeben Fall (c) ist eingetreten, was ist die Wahrscheinlichkeit dafür, dass nach 100 Stunden bereits 3 oder mehr Pumpen ausgefallen sind?

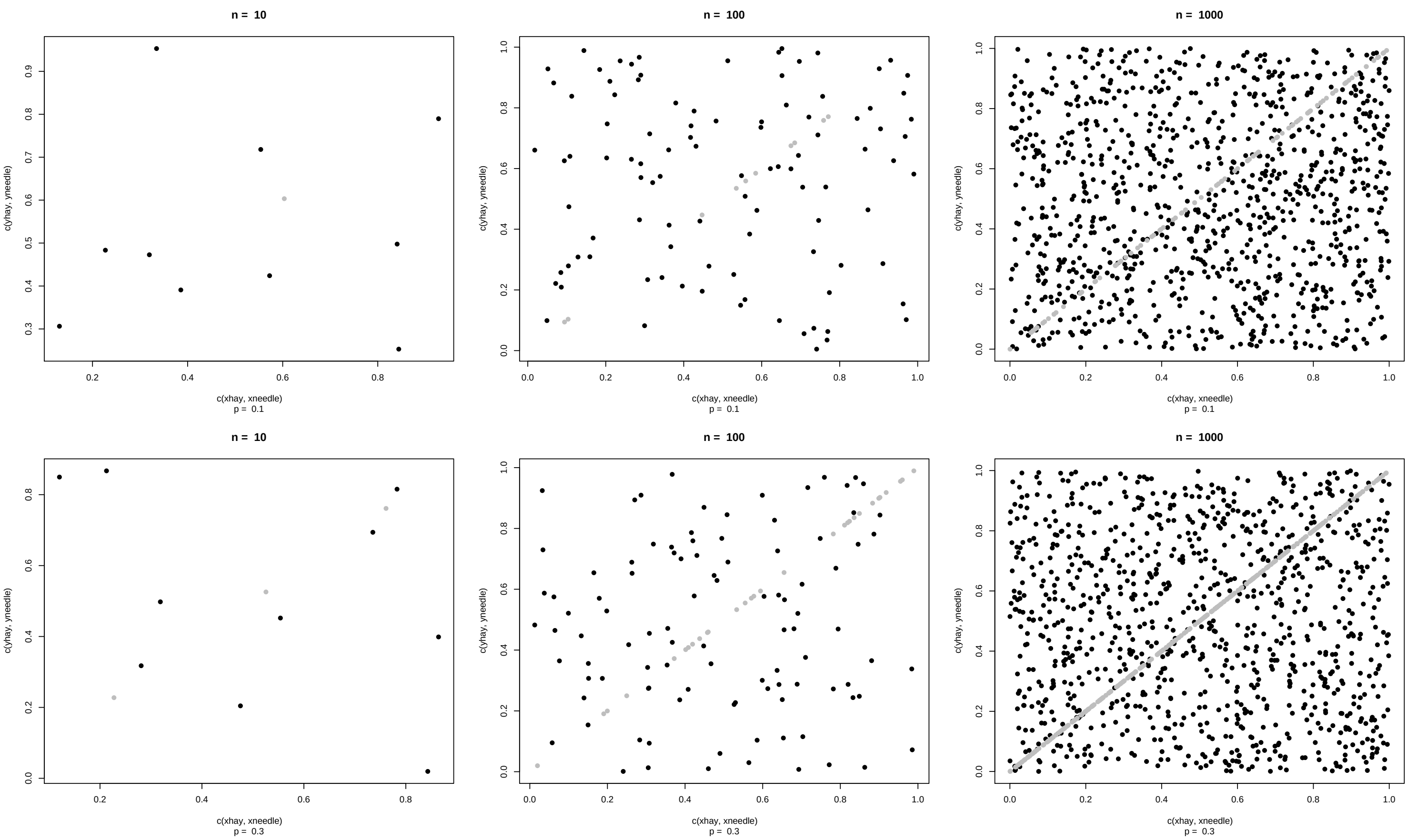
8. R Code und Output

In einer Simulationsstudie wird das Auffinden einer Stecknadel in einem Heuhaufen modelliert. Der Heuhaufen sei repräsentiert durch n zufällige Punkten im Einheitsquadrat; die Stecknadel hingegen durch eine kleine Anzahl, $p \cdot n$ (gerundet auf 0 Nachkommastellen), von zufälligen Punkten auf der Hauptdiagonalen des Einheitsquadrates.

Erläutern Sie den folgenden Code und Output in `R`. Folgender Fragekatalog kann hierbei hilfreich sein:

(a) Erklären Sie die Argumente und den Rückgabewert der definierten Funktion. Was macht diese Funktion?
(b) Interpretieren Sie die Plots, vorallem bei wachsendem n.
(c) Was machen in diesem Beispiel die `par` und `rep` Befehle? Erläutern Sie die Layoutbefehle im einzelnen.
(d) Gehen Sie auf die ausgegebenen Zahlen ein. Warum sind diese identisch bzw. wie lässt sich der Unterschied bei verschiedenen p-Werten interpretieren? Was geben diese Zahlen an?
(e) Durch welche Wahrscheinlichkeit sehen Sie in diesem Beispiel die Wahrscheinlichkeit für das Auffinden der Stecknadel modelliert?

```
> NeedleInTheHayStack <- function(n, p = 0.1){
+ xhay <- runif(n)
+ yhay <- runif(n)
+ xneedle <- runif(round(p * n))
+ yneedle <- xneedle
+
+ plot(x = c(xhay, xneedle), y = c(yhay, yneedle),
+ main = paste("n = ", n), sub = paste("p = ", p),
+ col = rep(c("black", "grey"), c(length(xhay), length(xneedle))),
+ pch = 19)
+
+ return(round(p*n)/(round(p*n) + n))
+ }
> par(mfrow = c(2, 3))
> NeedleInTheHayStack(10)
[1] 0.0909091
> NeedleInTheHayStack(100)
[1] 0.0909091
> NeedleInTheHayStack(1000)
[1] 0.0909091
> NeedleInTheHayStack(10, p = 0.3)
[1] 0.2307692
> NeedleInTheHayStack(100, p = 0.3)
[1] 0.2307692
> NeedleInTheHayStack(1000, p = 0.3)
[1] 0.2307692
```


n = 10
n = 100
n = 1000
c(yhay, yneedle)
c(xhay, xneedle)
p = 0.1
n = 10
n = 100
n = 1000
c(yhay, yneedle)
c(xhay, xneedle)
p = 0.3