

## Prüfung - Stochastik (401-0603-00L)

Nachname: ..... Vorname: ..... Stud.-Nr: .....

---

### Regeln zum Prüfungsablauf:

- Bitte legen Sie Ihre Legi gut sichtbar auf den Tisch.
- Es dürfen sich nur erlaubte **Hilfsmittel** auf dem Tisch befinden:  
10 hand- oder maschinengeschriebene A4-Seiten, Taschenrechner ohne Kommunikationsmöglichkeiten. Neutrales (nicht fachspezifisches) Deutsch-Englisch Wörterbuch.
- Die Benutzung von Mobiltelefonen ist nicht gestattet. Diese müssen ausgeschaltet sein und dürfen sich nicht auf dem Tisch befinden.
- Schreiben Sie Ihren Namen und Ihre Studierendenummer auf dieses Deckblatt, das Antwortblatt und auf jede Seite, die Sie abgeben.
- Bitte verwenden Sie für jede Aufgabe, die Sie bearbeiten, ein neues Blatt Papier.
- Die für die Aufgaben benötigten Tabellen wurden mit der Prüfung ausgeteilt.
- Die Prüfung besteht aus 4 Aufgaben.
- Mit Bleistift, in rot oder in grün geschriebene Lösungen ergeben keine Punkte.
- Schreiben Sie bei Aufgaben 2, 3 und 4 alle **Zwischenschritte** und **-rechnungen** sowie **Begründungen** auf. Setzen Sie Zahlen erst am Schluss ein.
- Die meisten Unteraufgaben können **ohne** die Resultate der vorherigen Unteraufgaben gelöst werden. Verbringen Sie nicht zu viel Zeit mit einer einzelnen Aufgabe, die Ihnen Schwierigkeiten bereitet.
- Die Indikatorfunktion  $\mathbf{1}_{\{x \in A\}}$  hat den Wert 1, falls  $x \in A$  und sonst den Wert 0.
- Sie haben 2 Stunden Bearbeitungszeit für die Prüfung.

Viel Erfolg!

---

### Korrektur:

Aufgabe	1	2	3	4
Punkte				
Kontrolle				

Punktetotal:

**Siehe nächstes Blatt!**

**1. (10 Punkte)**

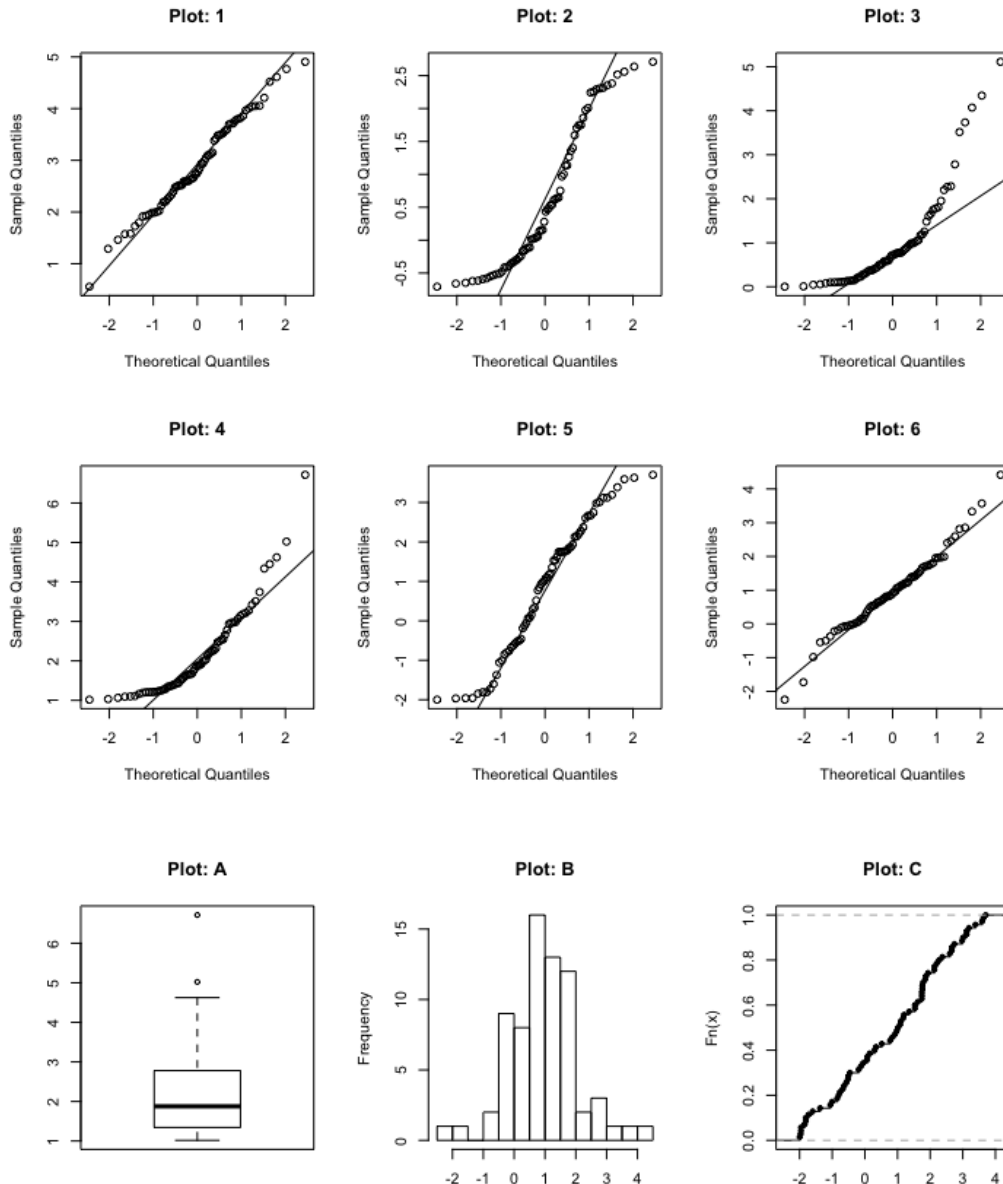
Bei den folgenden 10 Fragen ist jeweils genau eine Antwort richtig. Kreuzen Sie die richtigen Antworten auf dem **beiliegenden Antwortblatt** an.

*Beachten Sie: Jede richtige Antwort gibt einen Punkt. Es gibt keinen Abzug für falsche Antworten. Wir empfehlen daher alle Fragen zu beantworten.*

- a) Seien  $A, B$  Ereignisse (also Teilmengen von einem Grundraum  $\Omega$ ). Falls  $A$  und  $B$  unabhängig sind, so sind  $A$  und  $B^c$
- abhängig.
  - unabhängig.
  - manchmal unabhängig, je nachdem wie  $A \cap B$  und  $A \cup B$  aussehen.
- b) Seien  $X_i \sim \text{Poisson}(\lambda_i)$ ,  $\lambda_1 > \lambda_2 > 0$  für  $1 \leq i \leq 2$  unabhängig. Dann gilt
- $X_1 - X_2 \sim \text{Poisson}(\lambda_1 - \lambda_2)$ .
  - $\text{Var}(X_1 - X_2 + 5) = \lambda_1 - \lambda_2$ .
  - $\mathbb{E}[X_1 X_2 + 4X_1 - 3X_2] = \lambda_1 \lambda_2 + 4\lambda_1 - 3\lambda_2$ .
- c) Wir haben die folgenden zufälligen Realisierungen einer Zufallsvariable:  $x_1 = 3.6$ ,  $x_2 = 4.5$ ,  $x_3 = 5.6$ ,  $x_4 = 4.2$ ,  $x_5 = 5.7$ ,  $x_6 = 3.9$ ,  $x_7 = 6.5$ ,  $x_8 = 5.0$ ,  $x_9 = 5.2$ ,  $x_{10} = 5.3$ . Welche der folgenden empirischen Quantile sind korrekt?
- $q_{0.1} = 4.05$ ,  $q_{0.5} = 4.8$ ,  $q_{0.75} = 5.0$ .
  - $q_{0.1} = 3.6$ ,  $q_{0.5} = 5.0$ ,  $q_{0.75} = 5.45$ .
  - $q_{0.1} = 3.75$ ,  $q_{0.5} = 5.1$ ,  $q_{0.75} = 5.6$ .
- d) Sei  $n \in \mathbb{N}$ ,  $n > 1$  und seien  $(x_i, y_i)$ ,  $1 \leq i \leq n$  paarweise beobachtete Werte, wobei alle  $x_i$  paarweise verschieden sind. Sei  $r_{x,y}$  die empirische Korrelation. Angenommen  $y_i = -\frac{1}{10}x_i$  für alle  $1 \leq i \leq n$ , dann gilt:
- $r_{x,y} = -\frac{1}{10}$ .
  - $r_{x,y} = -1$ .
  - Wenn  $r_{x,y} = 0$ , dann sind die Stichproben  $(x_i)$  und  $(y_i)$  unabhängig.
- e) Eine Maschine sollte jeweils 500ml einer Flüssigkeit in eine Flasche füllen. Eine fachlich relevante Abweichung liegt vor, falls der wahre Mittelwert der abgefüllte Menge um mindestens 1ml vom Sollwert abweicht. Mithilfe mehrerer Stichproben wird das arithmetische Mittel 503 und das Vertrauensintervall  $[500.7, 505.3]$  für den wahren Mittelwert in ml bestimmt. Das Ergebnis dieser Untersuchung ist ...
- ... signifikant, aber nicht relevant.
  - ... signifikant und eventuell relevant.
  - ... signifikant und relevant.
- f) Wir führen 15 statistische Tests mit exaktem Signifikanzniveau  $\alpha = 0.05$  durch, die alle unabhängig sind. Angenommen die Nullhypothese stimmt immer. Dann gilt für die Wahrscheinlichkeit  $p$ , mindestens einmal einen Fehler 1. Art zu machen:
- $p \leq 0.05$ .
  - $0.5 \leq p \leq 0.75$ .
  - $p \geq 0.75$ .

**Bitte wenden!**

- g) Seien  $X$  und  $Y$  stetige Zufallsvariablen mit gemeinsamer Dichte  $f_{X,Y}$  und Randdichten  $f_X$  und  $f_Y$ . Welche der folgenden Aussagen zu stetigen Verteilungen stimmt im Allgemeinen **nicht**?
- Die Funktion  $f(x) = \frac{2}{3}x\mathbf{1}_{\{x \in [-1,2]\}}$  definiert eine mögliche Dichtefunktion für  $X$ .
  - Falls  $X, Y$  unabhängig sind und  $f_X, f_Y$  bekannt sind, kann man  $f_{X,Y}$  bestimmen.
  - Sei  $f_{X,Y}$  stetig, dann folgt aus  $f_X(10) = 0$ , dass  $f_{X,Y}(10, y) = 0$  für alle  $y \in \mathbb{R}$ .
- h) Betrachte die folgenden Plots. Die Plots 1-6 sind Normalplots, die jeweils mit 70 Beobachtungen erstellt wurden. Welcher der Normalplots (1-6) passt am besten ...



- h1) ... zum Boxplot A?  
 i) Plot 1    ii) Plot 3    iii) Plot 4
- h2) ... zum Histogramm in Plot B?  
 i) Plot 1    ii) Plot 5    iii) Plot 6
- h3) ... zur empirischen kumulativen Verteilungsfunktion in Plot C?  
 i) Plot 2    ii) Plot 5    iii) Plot 6

Siehe nächstes Blatt!

- 2. (9 Punkte)** Ein Antikörpertest wird verwendet, um Aussagen treffen zu können, ob eine Person eine Krankheit bereits hatte (und deswegen Antikörper dagegen gebildet hat) oder nicht. Der Test hat eine Falsch-Positiv-Rate von 2.5% (die Wahrscheinlichkeit beträgt 2.5%, dass eine Person, die die Krankheit noch nicht hatte, ein positives Testergebnis hat). Wir nehmen an, dass Bob die Krankheit noch nicht hatte und mehrmals hintereinander getestet wird. Sei  $X$  die Anzahl Tests, die an Bob durchgeführt werden, bis der erste Test positiv ist.
- (1 Punkt)** Treffen Sie passende Annahmen und geben Sie die Wahrscheinlichkeitsverteilung von  $X$  an (sowohl Name als auch alle benötigten Parameter).
  - (0.5 Punkte)** Berechnen Sie die erwartete Anzahl Tests, bis Bob das erste Mal positiv getestet wird.
  - (1.5 Punkte)** Ab welcher Anzahl durchgeführter Tests ist die Wahrscheinlichkeit kleiner als 5%, dass alle Tests negativ sind?
  - (0.5 Punkte)** Welchen Wert müsste die Falsch-Positiv-Rate mindestens haben, damit bereits nach 10 Tests die Wahrscheinlichkeit kleiner als 5% ist, dass alle Tests negativ sind?

Wir nehmen nun zusätzlich an, dass der Test eine Wahr-Positiv-Rate von 93% hat, also dass mit Wahrscheinlichkeit 93% der Test positiv ist, falls die getestete Person die Krankheit bereits hatte. Ausserdem nehmen wir an, dass der Anteil der Bevölkerung, der die Krankheit bereits hatte,  $q \in [0, 1]$  ist.

- (1 Punkt)** Bestimmen Sie die Wahrscheinlichkeit, dass eine zufällig aus der Bevölkerung ausgewählte Person ein positives Testergebnis hat, in Abhängigkeit von  $q$ . Berechnen Sie danach diese Wahrscheinlichkeit für  $q = 0.1$  und  $q = 0.5$ .
- (1.5 Punkte)** Angenommen eine Person hat ein positives Testergebnis. Bestimmen Sie die Wahrscheinlichkeit, dass die Person die Krankheit wirklich bereits hatte, in Abhängigkeit von  $q$ . Berechnen Sie danach diese Wahrscheinlichkeit für  $q = 0.1$  und  $q = 0.5$ .

Es wurde auch ein Schnelltest entwickelt, bei dem Personen, die die Krankheit bereits hatten, mit Wahrscheinlichkeit 90% ein positives Ergebnis haben (Wahr-Positiv) und Personen, die die Krankheit noch nicht hatten, mit Wahrscheinlichkeit 40% ein positives Ergebnis haben (Falsch-Positiv).

- (3 Punkte)** Insgesamt werden 330 Personen getestet, von denen 240 ein positives Testresultat haben. Wir wollen aus diesen Zahlen den Anteil  $q \in [0, 1]$  der Personen schätzen, die die Krankheit bereits hatten. Definieren Sie geeignete Zufallsvariablen und leiten Sie einen Momentenschätzer für  $q$  her. Nehmen Sie dazu an, dass diese 330 Testpersonen nur aus Personen bestehen, die entweder die Krankheit bereits hatten oder noch nicht (also keine akut erkrankten Personen) und dass diese alle unabhängig voneinander sind. Berechnen Sie auch den numerischen Wert des Momentenschätzers mit den gegebenen Zahlen.

**Bitte wenden!**

3. (10 Punkte) Jeden Tag verglühen Milliarden von Meteoroiden in der Erdatmosphäre. Wir verwenden die folgende Wahrscheinlichkeitsverteilung, um die Anzahl der von der Erde aus sichtbaren Meteoroiden (Sternschnuppen) in Millionen pro Tag zu modellieren

$$f(x) = f(x|(k, \lambda)) = \begin{cases} c \lambda^{-k} x^{k-1} \exp(-(x/\lambda)^k), & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Dabei sind  $k > 0$  und  $\lambda > 0$  Parameter.

Die Anzahl der gesichteten Meteoroiden in den letzten 10 Tagen sind in der folgenden Tabelle gegeben.

$i$ (Tag)	1	2	3	4	5	6	7	8	9	10
$x_i$ (in Millionen)	0.79	0.10	0.63	0.13	0.02	0.69	0.91	0.71	0.19	1.11

Für die gegebenen Daten gilt  $\sum_{i=1}^{10} x_i = 5.28$  und  $\sum_{i=1}^{10} x_i^2 \approx 4.125$  und  $\sum_{i=1}^{10} x_i^3 \approx 3.561$ .

- (1 Punkt) Bestimmen Sie  $c$  in Abhängigkeit von  $k$ , so dass  $f(x)$  eine Wahrscheinlichkeitsdichte ist. Begründen Sie Ihr Vorgehen.
- (2 Punkte) Angenommen  $k > 0$  ist bekannt. Geben Sie die Likelihood-Funktion und die Log-Likelihood-Funktion von  $\lambda$  in allgemeiner Form an. Definieren Sie dafür  $n \in \mathbb{N}$  passende i.i.d. Zufallsvariablen.
- (2 Punkte) Berechnen Sie den Maximum-Likelihood-Schätzer von  $\lambda$  in allgemeiner Form. Berechnen Sie auch den numerischen Wert mit den gegebenen Zahlen und  $k = 3$ .

Wir nehmen nun an, dass  $k = c = 1$ . Sei  $X$  eine Zufallsvariable mit Wahrscheinlichkeitsdichte  $f(x|(k = 1, \lambda))$ .

- (2 Punkte) Zeigen Sie, dass  $\mathbb{E}[X] = \lambda$  und  $\text{Var}(X) = \lambda^2$  gilt.

**Tipp:** Sie können ohne Beweis verwenden, dass  $\Gamma(n) = (n-1)!$  für die Gammafunktion  $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$ , für  $z > 0$ , gilt. Dies ist allerdings nicht unbedingt notwendig.

- (1 Punkt) Verwenden Sie **d)** und treffen Sie geeignete Annahmen, um die Verteilung der insgesamt gesichteten Meteoroiden (in Millionen) während der nächsten 50 Tage zu approximieren. Geben Sie die Verteilung und alle Parameter sowie eine Begründung an.
- (2 Punkte) Bestimmen Sie ein Intervall symmetrisch um den Erwartungswert, so dass die Anzahl insgesamt gesichteter Meteoroiden (in Millionen) innerhalb der nächsten 50 Tage mit Wahrscheinlichkeit 99% in diesem Intervall liegt. Geben Sie das Intervall zuerst in Abhängigkeit von  $\lambda$  und danach für den Wert  $\lambda = 0.7$  an.

**Siehe nächstes Blatt!**

4. (10 Punkte) Zwei Typen von Segelfliegern werden darauf getestet, wie lange sie maximal in der Luft bleiben können. In der folgenden Tabelle sind die Zeiten (in Stunden) von jeweils 10 Testflügen von beiden Modellen sowie deren Differenzen eingetragen.

$x_i$ (Typ 1)	6.67	13.48	9.03	9.95	13.96	12.62	6.58	11.0	10.4	8.78
$y_i$ (Typ 2)	3.69	15.29	5.24	12.98	7.38	5.53	3.55	10.97	6.45	1.04
$d_i = x_i - y_i$	2.98	-1.81	3.79	-3.03	6.58	7.09	3.03	0.03	3.95	7.74

Für diese Daten kann verwendet werden, dass  $\bar{x}_n = 10.247$ ,  $\bar{y}_n = 7.212$ ,  $\bar{d}_n = 3.035$  und  $\sum_{i=1}^n (x_i - \bar{x}_n)^2 = 60.423$ ,  $\sum_{i=1}^n (y_i - \bar{y}_n)^2 = 183.882$ ,  $\sum_{i=1}^n (d_i - \bar{d}_n)^2 = 121.846$ , wobei  $n = 10$ .

Wir wollen testen, ob sich die Flugzeiten der beiden Modelle signifikant unterscheiden. Dazu nehmen wir an, dass die Daten normalverteilt sind und alle Varianzen identisch sind und führen Tests mit dem Signifikanzniveau  $\alpha = 0.05$  durch.

- a) (1 Punkt) Begründen Sie, warum ein 2-Stichproben  $t$ -Test verwendet werden sollte.
- b) (1.5 Punkte) Geben Sie das Modell für die Daten an. Definieren Sie die Null- und Alternativhypothese und geben Sie die Teststatistik mit allen dazugehörigen Parametern an. Beschreiben Sie die Verteilung der Teststatistik unter der Nullhypothese.
- c) (1.5 Punkte) Berechnen Sie den  $p$ -Wert und formulieren Sie das Testergebnis.

**Bemerkung:** Falls Ihr Taschenrechner die nötigen Verteilungen nicht berechnen kann, genügt es, wenn Sie ein Intervall angeben, in dem sich der  $p$ -Wert befindet. Verwenden Sie dafür die Tabelle. Sie müssen die Werte der Tabelle dafür **nicht** interpolieren.

Wir nehmen nun an, dass die Tests der beiden Typen jeweils gleichzeitig unter den selben Bedingungen durchgeführt wurden.

- d) (1 Punkt) Begründen Sie, warum ein 1-Stichproben  $t$ -Test durchgeführt werden sollte. Was ist der Unterschied zu a) und wie erwarten Sie, dass sich das Ergebnis verändert?
- e) (1.5 Punkte) Geben Sie das Modell für die Daten an. Definieren Sie die Null- und Alternativhypothese und geben Sie die Teststatistik mit allen dazugehörigen Parametern an. Beschreiben Sie die Verteilung der Teststatistik unter der Nullhypothese.
- f) (1.5 Punkte) Berechnen Sie den Verwerfungsbereich für die Teststatistik und formulieren Sie das Testergebnis.
- g) (2 Punkte) Angenommen der Segelflieger vom Typ 1 kann im Schnitt 2 Stunden länger in der Luft bleiben. Berechnen Sie die Macht des Tests aus Aufgabenteil f).

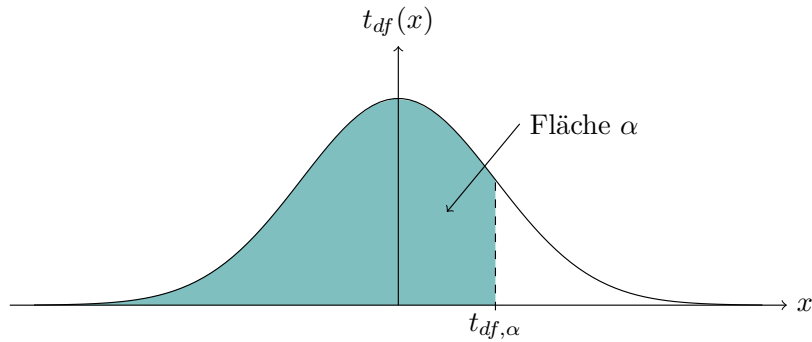
**Bitte wenden!**

**Siehe nächstes Blatt!**





Tabelle der  $t$ -Verteilung:



Lesebeispiel Tabelle:  $t_{9, 0.975} = 2.262$

$df \setminus \alpha$	0.60	0.70	0.80	0.90	0.95	0.975	0.99	0.995
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
90	0.254	0.526	0.846	1.291	1.662	1.987	2.368	2.632
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
$\infty$	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576

## Prüfung - Stochastik (401-0603-00L)

### Antwortblatt Aufgabe 1

Nachname: ..... Vorname: ..... Stud.-Nr: .....

Bitte kreuzen Sie die richtige Antwort an. Es ist jeweils genau eine Antwort richtig.

*Beachten Sie: Jede richtige Antwort gibt einen Punkt. Es gibt keinen Abzug für falsche Antworten. Wir empfehlen daher alle Fragen zu beantworten.*

- |     |                             |                              |                               |
|-----|-----------------------------|------------------------------|-------------------------------|
| a)  | i) <input type="checkbox"/> | ii) <input type="checkbox"/> | iii) <input type="checkbox"/> |
| b)  | i) <input type="checkbox"/> | ii) <input type="checkbox"/> | iii) <input type="checkbox"/> |
| c)  | i) <input type="checkbox"/> | ii) <input type="checkbox"/> | iii) <input type="checkbox"/> |
| d)  | i) <input type="checkbox"/> | ii) <input type="checkbox"/> | iii) <input type="checkbox"/> |
| e)  | i) <input type="checkbox"/> | ii) <input type="checkbox"/> | iii) <input type="checkbox"/> |
| f)  | i) <input type="checkbox"/> | ii) <input type="checkbox"/> | iii) <input type="checkbox"/> |
| g)  | i) <input type="checkbox"/> | ii) <input type="checkbox"/> | iii) <input type="checkbox"/> |
| h1) | i) <input type="checkbox"/> | ii) <input type="checkbox"/> | iii) <input type="checkbox"/> |
| h2) | i) <input type="checkbox"/> | ii) <input type="checkbox"/> | iii) <input type="checkbox"/> |
| h3) | i) <input type="checkbox"/> | ii) <input type="checkbox"/> | iii) <input type="checkbox"/> |