

Aufgabe 1 (Summenverteilung, bedingter Erwartungswert)

Es seien X_1, X_2, X_3 stochastisch unabhängige exponentialverteilte Zufallsvariablen jeweils mit der Dichte

$$f_{X_i}(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

und es seien $X = X_1 - X_2$ sowie $Y = X_3$. Hinweis: Es darf ohne Nachweis $f_X(x) = \frac{1}{2}e^{-|x|}$ verwendet werden.

- Berechnen Sie die Dichte $f_Z(z)$ von $Z = X + Y$. Hinweis: Unterscheiden Sie die Fälle $z < 0$ und $z \geq 0$.
- Berechnen Sie den bedingten Erwartungswert $E(X + Y|X_3 = x)$ für $x > 0$.

Aufgabe 2 (Parameterschätzung)

X_1, \dots, X_n seien stochastisch unabhängige Zufallsvariablen mit Dichte $f_{X_i}(x) = \frac{1}{2\lambda}e^{-|x|/\lambda}$. Dabei sei $\lambda > 0$.

- Berechnen Sie einen Maximum-Likelihood-Schätzer für λ auf Basis von X_1, \dots, X_n .
Kontrollergebnis: $\lambda_{ML} = \frac{1}{n} \sum_{i=1}^n |X_i|$, hier ist der Rechenweg gesucht, zu dem auch die hinreichende Bedingung für die Lösung des zugehörigen Optimierungsproblems gehört.
- Prüfen Sie, ob der ML-Schätzer aus der vorigen Teilaufgabe erwartungstreu ist.

Aufgabe 3 (Statistische Tests)

An einem bestimmten Tag wurden die Energieerträge von 10 identischen Solarmodulen eines bestimmten etablierten Typs A mit denen von 12 identischen Solarmodulen einer neueren Entwicklungsreihe (Typ B) verglichen. Es ergaben sich die folgenden Tageserträge $X_{A1}, \dots, X_{An_1}, X_{B1}, \dots, X_{Bn_2}$, gemessen in kWh)

Typ A	2,453	2,135	2,466	2,454	2,283	1,892	2,014	2,141	2,199	2,681		
Typ B	2,553	2,240	2,170	2,342	2,340	2,318	2,450	2,222	2,487	2,152	2,355	2,475

Es sollen anhand der Messreihen jeweils zum Niveau 5% die beiden folgenden Nullhypothesen H_0 geprüft werden:

- ob die beiden Solarmodule den gleichen mittleren Ertrag liefern
- ob der Typ B höchstens einen um 0,1 kWh höheren mittleren Ertrag als Typ A liefert.

Dabei darf angenommen werden, dass die Erträge beider Module dieselbe Standardabweichung aufweisen.

- Formulieren Sie die parametrischen Modell-Annahmen und mit Hilfe der Modellparameter die Hypothesenpaare H_0, H_1 zu den beiden obigen Nullhypothesen.
- Führen Sie die Tests zu den Hypothesenpaaren aus a) durch. Eine Tabelle der Quantile der t-Verteilung finden Sie nachfolgend auf Seite 3.

Anhang: Quantile der $t(n)$ -Verteilung

$n =$	$\alpha = .900$.950	.975	.990	.995	.999	.9995
1	3.08	6.31	12.71	31.82	63.66	318.31	636.62
2	1.89	2.92	4.30	6.96	9.92	22.32	31.60
3	1.64	2.35	3.18	4.54	5.84	12.21	12.92
4	1.53	2.13	2.78	3.75	4.60	7.17	8.61
5	1.48	2.02	2.57	3.36	4.03	5.89	6.87
6	1.44	1.94	2.45	3.14	3.71	5.21	5.96
7	1.41	1.89	2.36	3.00	3.50	4.79	5.41
8	1.40	1.86	2.31	2.90	3.36	4.50	5.04
9	1.38	1.83	2.26	2.82	3.25	4.30	4.78
10	1.37	1.81	2.23	2.76	3.17	4.14	4.59
11	1.36	1.80	2.20	2.72	3.11	4.02	4.44
12	1.36	1.78	2.18	2.68	3.05	3.93	4.32
13	1.35	1.77	2.16	2.65	3.01	3.85	4.22
14	1.35	1.76	2.14	2.62	2.98	3.79	4.14
15	1.34	1.75	2.13	2.60	2.95	3.73	4.07
16	1.34	1.75	2.12	2.58	2.92	3.69	4.01
17	1.33	1.74	2.11	2.57	2.90	3.65	3.97
18	1.33	1.73	2.10	2.55	2.88	3.61	3.92
19	1.33	1.73	2.09	2.54	2.86	3.58	3.88
20	1.33	1.72	2.09	2.53	2.85	3.55	3.85
21	1.32	1.72	2.08	2.52	2.83	3.53	3.82
22	1.32	1.72	2.07	2.51	2.82	3.50	3.79
23	1.32	1.71	2.07	2.50	2.81	3.48	3.77
24	1.32	1.71	2.06	2.49	2.80	3.47	3.75
25	1.32	1.71	2.06	2.49	2.79	3.45	3.73
26	1.31	1.71	2.06	2.48	2.78	3.43	3.71
27	1.31	1.70	2.05	2.47	2.77	3.42	3.69
28	1.31	1.70	2.05	2.47	2.76	3.41	3.67
29	1.31	1.70	2.05	2.46	2.76	3.40	3.66
30	1.31	1.70	2.04	2.46	2.75	3.39	3.65

- Die beiden Ertragsdatenreihen sind in zwei R-Vektoren \mathbf{x}_A und \mathbf{x}_B abgelegt. Schreiben Sie zwei R-Befehle, mit denen die beiden Tests aus b) durchgeführt werden.

- d) Sie fragen sich, ob die Annahme gleicher Varianzen von den vorliegenden Daten unterstützt wird, und stellen das Hypothesenpaar $H_0 : \sigma_A^2 = \sigma_B^2$ vs. $H_1 : \sigma_A^2 \neq \sigma_B^2$ auf. Nur eine der drei nachfolgenden Statistiken eignet sich als Teststatistik für diese Situation:

$$(1) V_1 = \frac{\sum_{i=1}^{n_1} (X_{Ai} - \bar{X}_A)^2}{\sum_{j=1}^{n_2} (X_{Bj} - \bar{X}_B)^2} \quad (2) V_2 = \frac{\sum_{i=1}^{n_1} (X_{Ai} - E(X_{Ai}))^2}{\sum_{j=1}^{n_2} (X_{Bj} - E(X_{Bj}))^2} \quad (3) V_3 = \frac{\sum_{i=1}^{n_1} (X_{Ai} - \bar{X}_A)^2}{\sigma_B^2}$$

Wählen Sie begründet diejenige Teststatistik, mit der Sie den Test durchführen würden (auch Begründungen, die zum Ausschluss der ungeeigneten Kandidaten führen, sind möglich). Motivieren Sie einen Ablehnungsbereich für den Test. Eine konkrete rechnerische Durchführung des Tests ist hier nicht verlangt.

Aufgabe 4 (Regression)

Auf einer Immobilie soll eine Photovoltaikanlage installiert werden. Sie verschaffen sich folgende Informationen über die monatliche mittlere Tagesdauer (Zeit von Sonnenaufgang bis Sonnenuntergang in Stunden) und die monatliche Globaleinstrahlung (engl.: „irradiation“, in kWh/m²), für den Standort der Immobilie im Jahr 2020:

Monat	Tagdauer	Einstrahlung	Monat	Tagdauer	Einstrahlung	Monat	Tagdauer	Einstrahlung
1	8,38	20,61	5	15,72	180,58	9	12,60	105,33
2	9,98	38,11	6	16,63	166,86	10	10,60	45,75
3	11,95	99,68	7	16,13	149,36	11	8,80	33,32
4	13,96	166,28	8	14,54	145,81	12	7,85	16,49

Die Daten sind in einem Dataframe `irrdata` abgelegt. Mit R erhalten Sie die folgende Konsolenausgabe:

```
Residuals:
    Min       1Q   Median       3Q      Max
-23.357 -14.987   2.403   5.923  35.842

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -141.551     21.111  -6.705 5.33e-05 ***
Tagdauer      19.483       1.672  11.653 3.85e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.44 on 10 degrees of freedom
Multiple R-squared:  0.9314,    Adjusted R-squared:  0.9245
F-statistic: 135.8 on 1 and 10 DF,  p-value: 3.849e-07
```

- Schreiben Sie einen R-Befehl, welcher diese Konsolenausgabe erzeugt.
- Lesen Sie die Regressionsfunktion zwischen Tagdauer und Einstrahlung aus der Konsolenausgabe ab und erläutern Sie an diesem Fall das Problem der Extrapolation.
- Leiten Sie die Regressionsfunktion zwischen Tagdauer und Einstrahlung mit Hilfe des KQ-Ansatzes aus der angegebenen Datentabelle her. Eventuell dabei benötigte Summen, Quadratsummen und Produktsummen dürfen Sie mit Hilfe Ihres Taschenrechners ermitteln, sie sind aber im Rechenweg anzugeben.
- Lesen Sie das Bestimmtheitsmaß aus der Konsolenausgabe ab, erläutern Sie seine generelle Bedeutung und beurteilen Sie den vorliegenden Wert.
- Die mittlere Tagdauer im Monat August 2022 beträgt 14,60 Stunden. Berechnen Sie ein 95%-Prognoseintervall für die Einstrahlung in diesem Monat. Verwenden Sie hierzu u.a. die Quantiltabelle auf Seite 3 und die obige Konsolenausgabe.

Aufgabe 5 (Multiple Choice)

a) Sind X, Y stochastisch unabhängige und identisch verteilte Zufallsvariablen mit $Y > 0$, so ist X/Y eine diskret verteilte Zufallsvariable.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
b) Ist $X > 0$ eine stetig verteilte Zufallsgröße mit Dichte f_X , so hat $Y = e^X$ ebenfalls eine Dichte, und zwar $h(y) = f_X(\ln(y))$.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
c) Im Einstichprobenmodell mit u.i.v.-Zufallsvariablen X_1, \dots, X_n mit $E(X_i) = 0$ ist $(X_1^2 + \dots + X_n^2)/n$ ein erwartungstreuer Schätzer für $\text{var}(X_1)$.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
d) Ist $\hat{\lambda} = \hat{\lambda}(X_1, \dots, X_n)$ ein ML-Schätzer für einen Parameter λ , so gibt es stets ein $c \in \mathbb{R}$, so dass $c \cdot \hat{\lambda}$ erwartungstreu für λ ist, auch wenn dies für $\hat{\lambda}$ nicht der Fall ist.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
e) Für den Bias B_n eines asymptotisch erwartungstreuen Schätzers T_n eines Parameters θ gilt $\lim_{n \rightarrow \infty} B_n = 0$	<input type="checkbox"/> ja	<input type="checkbox"/> nein

f)	Die Anteilswerthypothese $H_0 : p = p_0$ wird abgelehnt, wenn der hypothetische Wert p_0 außerhalb der Realisation des zweiseitigen Konfidenzintervalles für p liegt.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
g)	Gehört zu einem Datensatz und einem statistischen Test der p -Wert 0,06, so wird die Nullhypothese zum Signifikanzniveau 5% nicht verworfen.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
h)	Beim Chi-Quadrat-Anpassungstest mit festem Datensatz kann durch eine Erhöhung der Anzahl der Kategorien der Fehler zweiter Art minimiert werden.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
i)	Die Dauer einer Aktivität in einem Simulationsmodell ist bei ihrem Beginn unbekannt.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
j)	Der Beginn einer Aktivität wird in der Regel durch ein Ereignis ausgelöst.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
k)	Die Simulation Clock ist eine Variable, die zählt, wie lange die Ausführung einer Simulationstrajektorie auf der genutzten CPU gedauert hat.	<input type="checkbox"/> ja	<input type="checkbox"/> nein
l)	Der gewählte Ansatz zum Fortschreiben der Simulation Clock kann die für eine Simulationstrajektorie benötigte Rechenzeit beeinflussen.	<input type="checkbox"/> ja	<input type="checkbox"/> nein

Aufgabe 6 (Zufallszahlen)

- Nennen Sie drei gewünschte Eigenschaften eines arithmetischen Generators für (Pseudo-)Zufallszahlen.
- Es sind die folgenden Informationen über die Parameter eines linearen Kongruenzgenerators geben: modulus $m = 5$, multiplier $a = 3$, increment $c = 0$, Startwert $X_0 = 10$. Formulieren Sie den Generator mathematisch und beantworten Sie, um was für eine Art von linearem Kongruenzgenerator es sich handelt. Generieren Sie schliesslich die ersten 3 Pseudozufallszahlen in $[0,1)$, die der Generator liefert.
- Nennen Sie einen Vorteil und einen Nachteil der Polarmethode zur Erzeugung normalverteilter Pseudozufallszahlen im Vergleich zum Verfahren von Box und Müller.
- Beschreiben Sie kurz die Schritte der Polarmethode zur Erzeugung normalverteilter Pseudozufallszahlen.

Aufgabe 7 (Warteschlangen)

Sie diskutieren mit dem Auftraggeber einer Simulationsstudie über das zu simulierende System. Der Auftraggeber möchte, dass Sie zunächst eine bestimmte Arbeitsstation aus seiner Fabrikhalle analysieren. Die Arbeitsstation kann als Wartesystem mit einer Bedienstation aufgefasst werden, in der sich eine Maschine befindet. Die Maschine dient dazu, beim Wareneingang an den Toren der Fabrikhalle, die angelieferten Bauteile optisch auf Fehler zu überprüfen. Eine Überprüfung dauert immer genau 30 Sekunden. Angelieferte Bauteile werden direkt vor der Fabrikhalle zwischengelagert und von dort aus zur Maschine gebracht, wenn diese frei ist. Um die angelieferten Bauteile im Zwischenlagerbereich vor der Witterung zu schützen, denkt der Auftraggeber darüber nach, vor der Fabrikhalle eine kostspielige Überdachung aus Glas anzubringen. Ihn interessiert vor allem, wie gross das Glasdach mindestens sein muss, um alle Bauteile schützen zu können. Sie können dabei davon ausgehen, dass alle Bauteile die gleiche Grösse haben. Die Fabrikhalle ist die ganze Woche über non-stop in Betrieb. Die Bauteile werden von verschiedenen Zulieferern in unregelmässigen Abständen zur Fabrikhalle gebracht. Der Auftraggeber sagt, dass die Ankunft einer neuen Lieferung von vielen Faktoren abhängt, etwa vom Arbeitsaufkommen des Zulieferers und vom Verkehrsaufkommen. Das einzige, was er sicher weiss, ist, dass neue Lieferungen sehr oft in den frühen Morgenstunden eintreffen, niemals in den Mittagsstunden und ab und zu mal in den Nachtstunden.

- Definieren Sie eine Kennzahl, durch deren Berechnung das Anliegen des Auftraggebers am besten geklärt werden kann.
- Als Sie dem Auftraggeber den Preis für das Durchführen einer Simulationsstudie zur Berechnung der Kennzahl nennen, kommt dieser ins Schwitzen. Er fleht Sie an, ein günstigeres Angebot zu machen und fragt, ob eine Simulationsstudie hier wirklich notwendig ist. Er erinnere sich aus Studienzeiten noch an das M|M|1 Wartesystem und die Warteschlangentheorie, mit der man doch ganz schnell Lösungen berechnen kann. Nennen Sie drei Gründe, warum der vom Auftraggeber vorgeschlagene Ansatz für das vorliegende System nicht gewählt werden sollte.
- Sie haben den Auftraggeber davon überzeugt, dass eine Simulationsstudie durchgeführt werden muss. Die Kennzahlen sind definiert, das Simulationsmodell (konzeptionell und präzisiert) steht und ist, soweit es bisher geht, validiert. Sie haben sich dafür entschieden, das Softwaremodell mit einer Simulationsumgebung zu erstellen. Welcher Schritt ist noch erforderlich, bevor Sie mit der ersten Umsetzung des Softwaremodells beginnen können?

Aufgabe 8 (Arena)

Frau A. Weiler betreibt ein kleines Einzelhandelsgeschäft in der Innenstadt. Sie bittet Sie darum, Ihr bei der Simulation des Geschehens im Geschäft zu helfen. Sie hofft, durch die Simulation einen besseren Eindruck von der Auslastung zu erhalten und so mögliche Ressourcen-Engpässe frühzeitig identifizieren zu können. Sie beschreibt die erwarteten Abläufe wie folgt:

- Etwa alle 3 Minuten trifft ein neuer Kunde bei uns ein.
- An unserer Eingangstür steht ein Mitarbeiter, der den aktuellen Impfstatus der Kunden überprüft. Dazu benötigt er zwischen 1 und 2 Minuten.
- Etwa 3 % bestehen die Kontrolle nicht und werden fortgeschickt.
- Hat der Kunde die Kontrolle bestanden, nimmt er sich einen Einkaufskorb und tätigt seinen Einkauf. Wir haben aktuell 70 Körbe. Falls alle Einkaufskörbe in Benutzung sind, wartet der Kunde auf den nächsten freien Korb. Für den Einkauf benötigen die Kunden wahrscheinlich 9 Minuten, minimal 3 Minuten und maximal 27 Minuten.
- Hat der Kunde den Einkauf beendet sucht er sich für das Bezahlen die Kasse mit der kürzeren Warteschlange aus.
- An Kasse Nr. 1 sitzt Frieda. Sie benötigt für das Kassieren zwischen 30 Sekunden und 5 Minuten.
- An Kasse Nr. 2 sitzt Otto. Er benötigt für das Kassieren zwischen 2 und 4 Minuten.
- Nach dem Bezahlen gibt der Kunde den Korb wieder zurück.
- Anschließend verlässt er das Geschäft.

Skizzieren Sie ein entsprechendes ARENA-Modell zur Simulation des Einzelhandelsgeschäfts. Benennen Sie alle Module aussagekräftig und notieren Sie zu treffende Einstellungen in den Flowchart- und Daten-Modulen. Mögliche Module und die im Falle der Verwendung eines Moduls jeweils geforderten Einstellungen sind:

- **Flowchart Module**
 - Create: Zwischenankunftszeit | Anzahl Entitäten pro Ankunft
 - Dispose
 - Process: Seize-Delay-Release Typ | (Bearbeitungs-/ Delay-Zeit) | (Ressource(n))
 - Decide: by-condition + Bedingung/ by-chance + Wahrscheinlichkeiten
 - Assign: Art | Name | Wert
 - Record: Count (Name)/ Time Interval (Name, Attributsname)
- **Data Module**
 - Ressource: Ressource | Anzahl

Bitte beachten Sie, dass Sie keine Angaben zum Run-Setup machen müssen und dass Sie davon ausgehen können, dass als Standard-Zeiteinheit Minuten eingestellt ist. Dokumentieren Sie (falls nötig) getroffene Annahmen.