Aufgabe 1 (Summenverteilung, bedingter Erwartungswert)

Es seien X, Y stochastisch unabhängige Zufallsvariablen.

1. X besitze eine stetige Verteilung mit Dichte $f(x) = \lambda^2 \cdot x \cdot e^{-\lambda x} \cdot \mathbf{1}_{]0;\infty[}(x)$,
2. Y besitze eine stetige Verteilung mit Dichte $g(y) = \mu \cdot e^{-\mu y} \cdot \mathbf{1}_{]0;\infty[}(y)$,

Dabei seien $\lambda, \mu > 0$ mit $\lambda \neq \mu$.

a) Berechnen Sie eine Dichte von $Z = X + Y$ und stellen Sie diese in der Form

$$f_Z(z) = K \cdot \left((Az + B)e^{-\lambda z} + (Cz + D)e^{-\mu z}\right)$$

mit geeigneten, gegebenenfalls von λ, μ abhängigen Werten $A, B, C, D, K \in \mathbb{R}$ dar.

b) Berechnen Sie den bedingten Erwartungswert $E(X + Y|X = x)$ für $x > 0$.
Hinweis: Sie müssen hier nicht die bedingte Dichte ausrechnen.

Aufgabe 2 (Parameterschätzung)

Bei einer verbogenen Münze (mit den Seiten „Kopf“,„Wappen“) sei $p \in]0; 1[$ die unbekannte Wahrscheinlichkeit, dass bei einem einzelnen Wurf „Kopf“ fällt. Sie wiederholen 50-mal folgendes Einzelexperiment: Sie werfen eine verbogene Münze so lange, bis zum **zweiten** Mal Kopf fällt und zählen, wie oft bis zu diesem Zeitpunkt „Wappen“ fällt. Die Wahrscheinlichkeit, dass beim i-ten dieser unabhängigen Einzelexperimente x_i-mal Wappen fällt, ist

$$f_{X_i}(x_i) = (x_i + 1) \cdot p^2 \cdot (1 - p)^{x_i}, \quad x_i \in \mathbb{N}_0$$

a) Es sei angenommen, dass in den 50 Einzelexperimenten insgesamt 85-mal Wappen fällt. Berechnen Sie eine Maximum-Likelihood-Schätzung für p. Hier ist der vollständige Rechenweg gefordert. Gehen Sie auch auf hinreichende Bedingungen ein.

b) Berechnen Sie nun einen Schätzer für p nach der Momentenmethode. Auch hier ist der vollständige Rechenweg gesucht, allerdings dürfen Sie ohne gesonderte Rechnung ausnutzen, dass $\sum_{k=0}^{\infty} k(k+1)x^{k-1} = \frac{2}{(1-x)^3}$ für alle $x \in]0; 1[$ gilt.

Aufgabe 3 (Regression, Tests)

Die satirische Website `www.der-postillon.com` „berichtet“ im Jahr 2017: „Wachsende Altersarmut: Immer mehr Rentner werden von Enten gefüttert“. Zum „Nachweis“ führt der Artikel Daten zu Empfängern von Grundsicherung (x, in Tausend) und dokumentierten Fällen von Rentnerfütterung durch Enten (y) an:

	Jahr	x	y		Jahr	x	y		Jahr	x	y		Jahr	x	y
1	2003	22	260	5	2007	112	292	9	2011	141	437	13	2015	250	535
2	2004	32	293	6	2008	125	410	10	2012	161	464	14	2016	260	525
3	2005	72	343	7	2009	119	400	11	2013	200	500				
4	2006	122	371	8	2010	136	412	12	2014	241	511				

a) Berechnen Sie auf Grundlage der Datentabelle mit der KQ-Methode eine Regressionsfunktion der Form $y = ax + b$ sowie das zugehörige Bestimmtheitsmaß. Geben Sie alle relevanten Zwischenschritte (Summen bzw. Matrizen etc.) an. Die Angabe der Endergebnisses – z.B. anhand eines R-Befehls – ist in dieser Teilaufgabe nicht ausreichend.

b) Die Daten seien in einem Data-Frame `redata` mit den Attributen `jahr,x,y` abgelegt. Schreiben Sie einen kurzen R-Befehl, mit dem die folgende Ausgabe erzeugt werden kann:

```
Residuals:
   Min     1Q Median     3Q    Max
-83.97 -13.15   9.11  18.29  31.03

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 247.4408    18.1812   13.61 1.18e-08 ***
x             1.1475     0.1136   10.10 3.22e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.63 on 12 degrees of freedom
Multiple R-squared:  0.8947,    Adjusted R-squared:  0.8859
F-statistic:   102 on 1 and 12 DF,  p-value: 3.224e-07
```

c) Erläutern Sie für die in der vorigen Ausgabe auftretenden p-values, welcher Test jeweils durchgeführt wird und welche Bedeutung der p-value hat.

d) Geben Sie für das von Ihnen in a) berechnete Bestimmtheitsmaß (alternativ: für den entsprechenden Wert in der R-Ausgabe) an, welche allgemeine Bedeutung es hat und wie der konkrete Wert im Beispiel zu bewerten ist.

e) Für das Jahr 2017 wird mit 280000 Empfängern von Grundsicherung gerechnet. Geben Sie ein 95%-Prognoseintervall für die dokumentierten Fälle von Rentnerfütterung durch Enten in 2017 an (mit Rechenweg).

f) Schreiben Sie ein kurzes R-Skript, mit welchem die standardisierten Residuen aus dem Regressionsansatz berechnet werden. Berechnen Sie mit einer geeigneten Tabelle die Kolmogoroff-Smirnoff-Statistik für die standardisierten Residuen (d.h. der Rechenweg ist anzugeben). Prüfen Sie mit dem Kolmogoroff-Smirnoff-Test die Hypothese standardnormalverteilter standardisierter Residuen. Hinweis: Sie dürfen hier auf R zurückgreifen, müssen die Rechnung in Ihrer Antwort zu dieser Teilaufgabe aber dokumentieren.

Aufgabe 4 (Multiple Choice)

Die nachfolgenden Aussagen sind entweder wahr (Antwort „ja“) oder falsch bzw. nicht in jedem Fall richtig (Antwort „nein“). Geben Sie Sie jeweils die richtige Antwort an:

a)	Ist Y eine stetig verteilte Zufallsgröße und $X_i = \mathbf{1}_{[i;i+1]}(X)$, so ist $Z = X_1 + \cdots + X_n$ eine diskret verteilte Zufallsgröße.	ja	nein
b)	Ist $X > 0$ eine stetig verteilte Zufallsgröße mit Dichte f_X, so hat $Y = \ln(X)$ ebenfalls eine Dichte, und zwar $h(y) = f_X(e^y) \cdot e^y$.	ja	nein
c)	Im Einstichprobenmodell mit u.i.v.-Zufallsvariablen $X_1, \ldots, X_n$ ist $\sqrt{\frac{X_1^2+\cdots+X_n^2}{n}}$ stets ein erwartungstreuer Schätzer für $E(X_1)$.	ja	nein
d)	Bei einer asymptotisch erwartungstreuen Schätzstatistik ist das zugehörige Schätzerrisiko gleich der Schätzervarianz.	ja	nein
e)	Bei zwei für θ erwartungstreuen Schätzern T_1, T_2 gilt stets: Entweder T_1 ist MSE-effizienter als T_2 oder T_2 ist MSE-effizienter als T_1.	ja	nein
f)	Ist $V_n = (T_n - \theta)/A_n$ standardnormalverteilt, so gibt es zu vorgegebenem $\alpha \in]0;1]$ ein $t \in \mathbb{R}$, so dass θ vom Konfidenzintervall $[T_n - t \cdot A_n; T_n + t \cdot A_n]$ mit Wahrscheinlichkeit $1 - \alpha$ überdeckt wird.	ja	nein
g)	Gehört zu einem Datensatz und einem statistischen Test der p-Wert $0,04$, so wird die Nullhypothese zum Signifikanzniveau 5% abgelehnt.	ja	nein
h)	Beim Einstichproben-Chi-Quadrat-Anpassungstest besitzt die Prüfgröße unter der Nullhypothese eine Chi-Quadrat-Verteilung.	ja	nein
i)	Eine Trajektorie eines diskreten, ereignisbasierten Simulationsmodells kann als Realisierung eines Markov-Prozesses aufgefasst werden.	ja	nein
j)	Im Rahmen einer Simulationsstudie sollte das präzisierte Simulationsmodell vom Auftraggeber der Studie formuliert werden.	ja	nein
k)	Die Verfahren aus dem Bereich Simulation Optimization sind vor allem bei Vorliegen einer unendlichen Anzahl von Alternativen erfolgsversprechend.	ja	nein
l)	Bei zweistufigen Simulation Optimization Verfahren kann es durch Überschätzen der Varianzen in Stufe 1 zu einer unnötig hohen Anzahl von Simulationstrajektorien in Stufe 2 kommen.	ja	nein

Aufgabe 5 (Zufallszahlen)

a) Nennen Sie zwei Ansätze zum Erzeugen echter Zufallszahlen für eine Computersimulation.

b) Definieren Sie den linearen Kongruenzgenerator formal und benennen Sie alle Elemente der Definition.

c) Wie gehen Sie vor, um eine Realisierung der gleichverteilten Zufallsvariablen $X_1 \sim U(0,1)$ und $X_2 \sim U(5,10)$ zu erzeugen? Erläutern Sie jeweils Ihre Vorgehensweisen, und berechnen Sie eine Realisierung von X_2 unter der Annahme, dass Sie $X_1 = 0.2$ beobachtet haben.

Aufgabe 6 (Warteschlangen)

a) Formulieren Sie das Gesetz von Little mit präziser Benennung von Variablen und Parametern.

b) Zur Bekämpfung der Pandemie und zur Aufbesserung Ihres Monatseinkommens haben Sie sich für den Betrieb einer Corona-Teststation entschieden. Durch intensives Training sind Sie mittlerweile in der Lage relativ viele Kunden pro Stunde zu testen. Sie haben in den ersten Wochen des Betriebes der Station folgende drei Beobachtungen gemacht: (1) Im Durchschnitt stehen über die Öffnungszeit hinweg zu jedem Zeitpunkt 3 Personen in der Warteschlange vor Ihrer Station. (2) Einige der Personen, die sich Ihrer Station mit der Absicht nähern, sich testen zu lassen, entscheiden sich beim Anblick der Warteschlange aus Angst vor der Menschenmenge dazu umzudrehen und eine andere (weniger frequentierte) Teststation aufzusuchen. Sie schätzen den Anteil dieser Personen auf ca. 25% der sich ihrer Station mit Testabsicht nähernden Personen. (3) Obwohl der Test an sich durch Ihr hartes Training sehr schnell von statten geht, nutzen die Getesteten die Gelegenheit, um sich bei Ihnen über die Wartezeit in der Schlange vor der Teststation zu beklagen. Sie entnehmen den Rückmeldungen, dass die Getesteten im Durchschnitt ca. 20 Minuten in der Warteschlange vor Ihrer Teststation verbracht haben. Nehmen Sie an, dass das geschilderte Wartesystem sich im Gleichgewicht befindet. Wieviele Kunden stellen sich pro Stunde im Durchschnitt tatsächlich in die Warteschlange vor Ihrer Teststation?

Aufgabe 7 (ARENA)

a) Herr T. Ice plant die Eröffnung einer neuen Eissporthalle am Stadtrand. Während der Planung, bittet er Sie darum, Ihn bei der Simulation der Abläufe zu unterstützen. Er hofft durch die Simulation einen besseren Eindruck über die Auslastung seiner Ressourcen zu erhalten. Er beschreibt die erwarteten Abläufe wie folgt:

   - Wir erwarten, dass etwa alle 5 Minuten ein neuer Gast an der Eissporthalle eintrifft.
   - Jeder Gast muss sich bei uns ein Paar Schlittschuhe leihen. Der Gast kann sich selbst an einem Regal mit Schlittschuhen bedienen und benötigt für das nehmen der Schlittschuhe ca. 30 Sekunden. Wir planen aktuell 120 Paar Schlittschuhe bereitzustellen. Zur Vereinfachung können Sie davon ausgehen, dass alle Gäste die gleiche Schuhgröße haben. Falls ein Gast eintrifft und keine Schlittschuhe verfügbar sind, geht der Gast unverrichteter Dinge.
   - Ansonsten nimmt sich der Gast ein Schlittschuhpaar und geht zu einer der beiden Kassen. Der Gast stellt sich dabei stets an der Kasse mit der kürzeren Warteschlange an.
   - Der Mitarbeiter an den Kasse 1 benötigt zwischen 1 und 2 Minuten für das Kassieren, wohingegen der Mitarbeiter an Kasse 2 zwischen 30 Sekunden und 3 Minuten benötigt.
   - Anschließend nutzt der Gast die Schlittschuhbahn zwischen 45 Minuten und 2 Stunden.
   - Nach dem Schlittschuhfahren gibt der Gast seine Schlittschuhe am Serviceschalter zurück. Der Mitarbeiter am Serviceschalter desinfiziert die Schlittschuhe und stellt sie zurück ins Regal. Dafür benötigt er ungefähr 1 Minute.

   Skizzieren Sie ein entsprechendes ARENA-Modell zur Simulation der Eissporthalle. Benennen Sie alle Module aussagekräftig und notieren Sie zu treffende Einstellungen in den Flowchart- und Data-Modulen.

   - **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
     * Batch: Anzahl
     * Separate: Dublicate Original / Split Batch
     * Assign: Art | Name | Wert
     * Record: Count / Time Interval
   - **Data Module**
     * Ressource: Ressource | Anzahl

   Bitte beachten Sie, dass Sie **nicht** alle der hier aufgeführten Flowchart-Module in Ihrem Modell verwenden müssen und hier nicht aufgeführte Module nicht beschrieben werden müssen. Sie müssen desweiteren keine Angaben zum Run-Setup machen und können davon ausgehen, dass als Standard-Zeiteinheit Minuten eingestellt ist. Dokumentieren Sie (falls nötig) getroffene Annahmen.

b) Erläutern Sie kurz den Unterschied zwischen den Attributen und Variablen in ARENA.