StofI - WiSe19/20

Rückblick auf die Übungsaufgaben

1. Wie wahrscheinlich ist es, beim dreimaligen (fairen) Würfeln genau zweimal eine Augenzahl größer als 4 zu erzielen?

2.S Wie in der Vorlesung betrachten wir eine Gesamtfläche S bestehend aus g Pixeln und eine TeilflŠche A bestehend aus $f := p \cdot g$ Pixeln. Dabei sei p eine Zahl im Intervall $[0, 1]$, und sowohl g als auch f seien natürliche Zahlen. Die dreimalige rein zufällige Wahl eines Pixels aus S (entsprechend einem "Ziehen mit Zurücklegen"), beschreiben wir durch eine auf dem Wertebereich $\{1, \ldots, g\}^3$ uniform verteilte Zufallsvariable $X = (X_1, X_2, X_3)$.

a) Wieviele Ausgänge von X gibt es?

b) Wieviele Ausgänge von X gibt es, bei denen X_1 auf die Menge $\{1, \ldots, f\}$ und X_2 sowie X_3 auf die Menge $\{f + 1, \ldots, g\}$ fallen? Drücken Sie das Ergebnis durch g und p aus.

c) Wieviele Ausgänge von X gibt es, bei denen genau eines der drei gewählten Pixel auf die Menge $\{1, \ldots, f\}$ fällt?

d) Wie wahrscheinlich ist es, dass von den drei zufällig aus S gewählten Pixeln genau eines aus A gewählt wird?

e) Bestimmen Sie die 4 möglichen Ausgänge der zufälligen Trefferquote M von A, sowie deren Verteilungsgewichte.

f) Illustrieren Sie für $p = 0.195$ und $n = 3$ das Ergebnis aus e) mittels des über den Link auf der StofI-Web-Seite zur Verfügung gestellten R-Programms "Monte Carlo Simulation".* Betrachten Sie dazu ein Histogramm der Schätzwerte aus (z.B.) 1000 Wiederholungen des Zufallsexperiments.

*Das frei verfügbare statistische Programmpaket R bekommen Sie über www.r-project.org, zu finden auch über google $\rightarrow$ R, auf Ihren Rechner.

3. In der Vorlesung haben wir den Anteil p einer Teilfläche A an einer Gesamtfläche ("Quadrat") S mit einem einfachen Monte-Carlo-Verfahren geschätzt: n Punkte wurden rein zufällig in S geworfen und der Anteil M der Treffer von A ermittelt. Die (durch unabhängiges Wiederholen dieses Zufallsexperimentes ermittelte "empirische") Verteilung von M hat uns ein Bild von der Zuverlässigkeit der Schätzung vermittelt. Erkunden Sie (wieder für $p = 0.195$) mittels des R-Programms "Monte Carlo Simulation", wie sich die Genauigkeit der Schätzung verändert, wenn (i) $n = 200$ (ii) $n = 800$ (iii) $n = 3200$ Punkte in die Menge S geworfen werden: Um welchen Faktor (circa) wird jeweils das Histogramm der Schätzwerte schmäler?

4.S. Von 4 Objekten wird jedes rein zufällig auf einen von 16 möglichen Plätzen gesetzt. Berechnen Sie die Wahrscheinlichkeit des Ereignisses “es kommt zu keiner Kollision”

(a) über die exakte Berechnung
(b) über die Stirling-Näherung
(c) über die Approximation durch Linearisierung.

Kommentieren Sie kurz die Güte der in (b) und (c) erzielten Näherung.

5.S. Es sei T der Zeitpunkt der ersten Kollision beim sukzessiven rein zufälligen Besetzen von g Plätzen (mit einem Objekt nach dem anderen), E_n sei das Ereignis "keine Kollision unter den ersten n Objekten",* und $w_n := \mathbf{P}(E_n)$ dessen Wahrscheinlichkeit.

(a) Drücken Sie das Verteilungsgewicht $\rho_n := \mathbf{P}(T = n)$ durch w_{n-1}, n und g aus.

(b) Es sei $g = 100$. Bestimmen Sie ein möglichst kurzes Intervall I, sodass gilt: $\mathbf{P}(T \in I) \geq 0.5$.

*In der gedruckten Version des Übungsblattes war dieses Ereignis mit A_n bezeichnet. In den Folien zu VL1b und im Buch S. 2 ist jedoch A eine Teilmenge von $\{1, \ldots, g\}^n$. Mit der notationelle Änderung in der aktuellen Version ergibt sich die instruktive Beziehung $E_n = \{(X_1, \ldots, X_n) \in A\}$.

6. Beweisen Sie auf den Spuren von Jakob Bernoulli die folgende Identität für $1 \leq k \leq n$:

$$\binom{n}{k} = \sum_{b=k}^{n} \binom{b-1}{k-1}.$$

Denken Sie sich dazu eine k-elementigen Teilmenge B von $\{1, \ldots, n\}$ auf zweistufige Weise gewählt: erst das größte Element b von B, dann die restlichen $k-1$ Elemente. Wieviele Ausgänge dieses zweistufigen Experimentes gibt es?

7. Aus einer rein zufälligen Permutation X von $1, \ldots, n$ kann man "durch zufälliges Einschieben" eine rein zufällige Permutation X' von $1, \ldots, n+1$ gewinnen, indem man erst für $X'(n+1)$ eine rein zufällige Wahl aus $1, \ldots, n+1$ trifft und dann die $X'(i)$, $1 \leq i \leq n$, gewinnt, indem man all diejenigen $X(i)$, die größer oder gleich $X'(n+1)$ sind, um 1 erhöht. Verwenden Sie diese Tatsache, um mit Induktion zu zeigen, dass für jedes $n \in \mathbb{N}$ und jedes $j = 2, \ldots, n$ die Anzahl

$$H(j) := \#\{i < j : X(i) > X(j)\}$$

der *Fehlstände*, an denen j zusammen mit einem kleineren Partner beteiligt ist, uniform auf $\{0, 1, \ldots, j-1\}$ verteilt ist.*

*Der hier vorgeschlagene Lösungsweg ist eine Variante des im Buch, Seite 9, für dieselbe Aufgabenstellung vorgeschlagenen Lösungswegs.

8.S Wie wahrscheinlich ist es, dass bei einer uniformen Besetzung von drei Plätzen (namens 1,2,3) mit 10 (ununterscheidbaren) Objekten der Platz 1 mit mindestens zwei Objekten besetzt ist? Beantworten Sie die Frage sowohl mittels einer Bijektion zu einem passenden $S_{n,g}$ also auch durch Abzählen (einer passenden Teilmenge) eines de Finetti-Dreiecks.

9.S. Für $n \in \mathbb{N}$ sei $X^{(n)} = (X_1^{(n)}, X_2^{(n)}, X_3^{(n)})$ multinomialverteilt mit den Parametern $3n; \frac{1}{3}, \frac{1}{3}, \frac{1}{3}$.

Berechnen Sie die Näherung von

$$\mathbf{P}\left(X_1^{(n)} = X_2^{(n)} = X_3^{(n)} = n\right)$$

mit der Stirling-Formel.

10. $X = (X_1, X_2, X_3)$ sei multinomialverteilt mit den Parametern $n; p_1, p_2, p_3$. Wie ist die Zufallsvariable $X_1 + X_2$ verteilt?

Argumentieren Sie

(i) intuitiv (über die Summe von Zählvariablen in einem Würfelexperiment), und
(ii) analytisch, indem Sie für $k \leq n$ die folgende Gleichheit nachweisen:

$$\sum_{k_1+k_2=k} \binom{n}{k_1, k_2, n-k} p_1^{k_1} p_2^{k_2} p_3^{n-k} = \binom{n}{k} (p_1 + p_2)^k p_3^{n-k}.$$

11. $W = (W_1, \ldots, W_7)$ beschreibe ein 7-maliges gewöhnliches Würfeln.

Wie wahrscheinlich ist es, dass

(i) mehr als dreimal die 1 gewürfelt wird,
(ii) mehr als dreimal die 6 gewürfelt wird,
(iii) die viertgrößte der 7 gewürfelten Augenzahlen in die Menge $\{2, 3, 4, 5\}$ fällt?*

*Dabei ist die viertgrößte von sieben (nicht notwendigerweise verschiedenen) Zahlen $w_1, \ldots, w_7$ dasjenige (eindeutig bestimmte) Element $v \in \{w_1, \ldots, w_7\}$, für welches für vier der $j \in \{1, \ldots, 7\}$ gilt: $w_j \leq v$ und für drei der $j \in \{1, \ldots, 7\}$ gilt: $w_j \geq v$.

12.S. An einem runden Tisch befinden sich 30 Sitzplätze, nummeriert mit 1,. . . , 30, sodass auch die Plätze 1 und 30 benachbart sind. Die Vertreter von 3 Delegationen (5 aus Land A, 10 aus Land B, 15 aus Land C) werden rein zufällig platziert.

(i) Wie wahrscheinlich ist es, dass Platz 6 und 7 mit Vertretern aus dem selben Land besetzt werden?
(ii) Was ist die erwartete Anzahl von Paaren benachbarter Plätze, auf denen Vertreter aus demselben Land sitzen?
(iii) Was ist die erwartete Anzahl von Paaren benachbarter Plätze, auf denen Vertreter aus verschiedenen Ländern sitzen?

13. Aus einer Population bestehend aus 20 Wiesbadenern und 40 Frankfurtern wurde eine Stichprobe vom Umfang 15 (d.h. eine 15-elementige Teilmenge der Population) herausgegriffen. In dieser befanden sich ein Wiesbadener und 14 Frankfurter.

a) Was ist der Erwartungswert der Anzahl der Wiesbadener in einer *rein zufälligen* Stichprobe vom Umfang 15?

b) Wie wahrscheinlich ist in einer rein zufällig gezogenen Stichprobe eine Anzahl von Wiesbadenern, die vom (in a) berechneten) Erwartungswert mindestens so weit abweicht wie die beobachtete Anzahl 1?

Hinweis: Hier ist der R-Befehl sum(dhyper(. . .)) hilfreich. Finden Sie mittels des Befehls `?dhyper` *heraus, was das mit der Aufgabenstellung zu tun hat, und finden Sie die passenden Summationsgrenzen.*

14. Es sei $(X_1, \ldots, X_7)$ uniform verteilt auf $\{1, \ldots, 10\}^7$. Wir interpretieren X_i als die zufällige Platzwahl des Individuums i, mit 10 möglichen Plätzen. Dabei ist mehrfache Wahl eines Platzes erlaubt, es handelt sich um ein 7-faches $(1/10, \ldots, 1/10)$-Würfeln.)

a) Mit welcher Wahrscheinlichkeit bleiben sowohl Platz 1 als auch Platz 3 leer?

b) Mit welcher Wahrscheinlichkeit bleibt mindestens einer der Plätze 1, 2 oder 3 leer?
Hinweis: Betrachten Sie die Ereignisse $E_i := \{$*Platz* i *wird nicht gewählt*$\}$ *und verwenden Sie die Einschluss-Ausschluss-Formel.*

c) Mit welcher Wahrscheinlichkeit kommen alle drei Plätze 1, 2, 3 zum Zug?

15.S. a) $Z = (Z_1, \ldots, Z_n)$ sei ein n-facher p-Münzwurf. Wie wahrscheinlich ist (für $n \geq 5$) das Ereignis $\{Z_3 = 1, Z_5 = 1\}$?

b) Berechnen Sie $\mathbf{E}[X^2]$ für eine Binom(n, p)-verteilten Zufallsvariable X, indem Sie die Darstellung $X = Z_1 + \cdots + Z_n$ verwenden.

c) Folgern Sie aus b) mit der Linearität des Erwartungswertes, dass für ein Binom(n, p)-verteiltes X gilt:

$$\mathbf{E}[(X - np)^2] = npq \quad \text{und} \quad \mathbf{E}\left[\left(\frac{X}{n} - p\right)^2\right] = \frac{pq}{n}.$$

d) Schätzen Sie die Wahrscheinlichkeit, dass die relative Anzahl der Erfolge beim n-fachen fairen Münzwurf um mehr als 0.01 von 0.5 abweicht,
(i) für $n = 10000$,
(ii) für $n = 1000000$
mittels der Ungleichung von Markov nach oben ab.

16.S. Ähnlich wie in der ersten Stunde denken wir uns eine Quadratfläche S in g gleich große Teilquadrate zerlegt, wobei g eine große Quadratzahl ist. Die Teilquadrate seien mit $j = 1, \ldots, g$ indiziert. Wir wählen wiederholt rein zufällig Punkte aus S und notieren die Indizes der Teilquadrate, in die sie fallen, als $X_1, X_2, \ldots$. Mit A bezeichnen wir die Vereinigung der Teilquadrate 1,2,3,4.

a) Wie ist die Anzahl derjenigen Punkte aus den ersten $\lfloor g/2 \rfloor$ Punkten verteilt, die in A fällt?

b) Wie ist die Wartezeit bis zum ersten Treffer von A verteilt?

c) Berechnen Sie eine Näherung für die Wahrscheinlichkeit dafür, dass keiner der ersten $\lfloor g/2 \rfloor$ Punkte in A fällt

i) mit der Poissonappromation (für die Verteilung der Anzahl der Treffer)

ii) mit der Exponentialapproximation (für die Verteilung der Wartezeit bis zum ersten Treffer).

17. *Unabhängig oder nicht?*

a) *Für zwei diskrete Zufallsvariable X_1, X_2 ist die Unabhängigkeit äquivalent dazu, dass die Zeilen der Matrix ihrer gemeinsamen Verteilungsgewichte zueinander proportional sind.* Rekapitulieren (und vervollständigen) Sie den Beweis dieses Kriteriums anhand der Hinweise auf der neuen Folie am Ende von Abschnitt 1 der Vorlesung 4b.

b) Wir betrachten vier Beispiele, bei den ersten beiden ist $S_1 = \{1, 2\}$, $S_2 = \{b, c\}$ bei den letzten beiden ist $S_1 = \{1, 2, 3\}$, $S_2 = \{b, c, d\}$. Die Matrizen der gemeinsamen Verteilungsgewichte von (X_1, X_2) sind

i)

	b	c
1	0.1	0.3
2	0.15	0.45

ii)

	b	c
1	0.1	0.3
2	0.2	0.4

iii)

	b	c	d
1	6γ	7γ	10γ
2	12γ	14γ	20γ
3	18γ	21γ	30γ

iv)

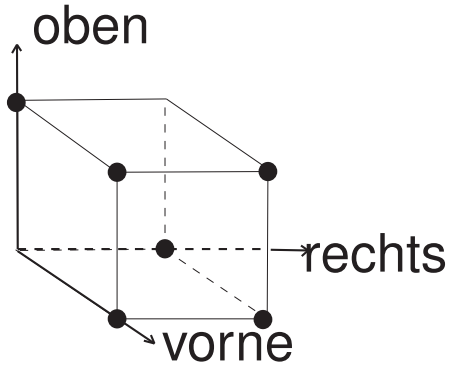
	b	c	d
1	6γ	7γ	10γ
2	13γ	14γ	20γ
3	17γ	21γ	30γ

Was ist in (iii) bzw. (iv) der Wert von γ? In welchen der Beispiele sind X_1 und X_2 unabhängig, in welchen nicht?

18. ***Paarweise Unabhängigkeit.***

$X = (X_1, X_2, X_3)$ mit Wertebereich $\{o, u\} \times \{\ell, r\} \times \{h, v\}$ beschreibe eine rein zufällige Wahl aus den 6 in der Skizze markierten Ecken des Würfels (die beiden nicht markierten Ecken bleiben tabu). Vier der markierten Ecken sind "vorne", drei sind "rechts" und drei sind "oben". Wir betrachten die Ereignisse

$E_1 := \{X_1 = o\} = \{X$ landet oben$\}$,
$E_2 := \{X_2 = r\} = \{X$ landet rechts$\}$,
$E_3 := \{X_3 = v\} = \{X$ landet vorne$\}$.



a) Bestimmen Sie die 2×2-Matrix der Verteilungsgewichte von (X_1, X_2). Sind E_1 und E_2 unabhängig? Sind sie positiv korelliert?

b) Gilt $\mathbf{P}(E_1 \cap E_2 \cap E_3) = \mathbf{P}(E_1) \cdot \mathbf{P}(E_2) \cdot \mathbf{P}(E_3)$?

19. S ***Die Standardabweichung der zufälligen Trefferquote.*** In der Stunde Eins haben wir den Anteil p einer Teilfläche F an einer Gesamtfläche G dadurch geschätzt, dass wir n Punkte rein zufällig in G geworfen und als Schätzer $\widehat{p}$ die relative Treffzahl von F genommen haben. Berechnen Sie die Standardabweichung von $\widehat{p}$, wenn der tatsächliche Wert von p gleich $1/5$ ist. Was ergibt sich für (i) $n = 100$, (ii) $n = 400$, (iii) $n = 1600$?

20. S ***Die Varianz des Stichprobenmittels beim Ziehen ohne Zurücklegen.***

In einer Population von 90 Individuen haben 45 Individuen die Größe 10, 15 Individuen die Größe 5 und 30 Individuen die Größe 20. Es sei X die Größe eines rein zufällig aus der Population gewählten Individuums.

a) Berechnen Sie (i) den Erwartungswert, (ii) die Varianz, (iii) die Standardabweichung von X.

b) Wir ziehen rein zufällig und ohne Zurücklegen aus der Population und bezeichen mit X_i die Größe des i-ten gezogenen Individuums.

α) Warum hängt (für $1 \leq i \neq j \leq 90$) die Kovarianz **Cov**$[X_i, X_j]$ nicht von i und j ab?

β) Berechnen Sie die Kovarianz von X_1 und X_2 aus der Identität

$$0 = \textbf{Var}(X_1 + \ldots + X_{90}).$$

γ) Berechnen Sie die Varianz des Stichprobenmittels $\frac{1}{30}(X_1 + \cdots + X_{30})$.

c) Schätzen Sie mittels der Ungleichung von Chebyshev die Wahrscheinlichkeit dafür ab, dass das Stichprobenmittel $\frac{1}{30}(X_1 + \cdots + X_{30})$ um mehr als 2 von μ abweicht.

21. S. ***Münzwurffolgen übersetzt ins Einheitsintervall***

(Z_1, Z_2, Z_3) sei ein dreifacher p-Münzwurf mit $p = 1/4$. Die $[0, 1]$-wertige Zufallsvariable Y sei definiert durch $Y := \frac{1}{2}Z_1 + \left(\frac{1}{2}\right)^2 Z_2 + \left(\frac{1}{2}\right)^3 Z_3$.

a) Bestimmen Sie die Verteilungsgewichte von Y.

b) Bestimmen Sie den Wert $F(b)$ der Verteilungsfunktion von Y für
(i) $b = 1/2$, (ii) $b = 3/8$, (iii) $b = 9/16$, (iv) $b = 2$.

c) Skizzieren Sie die Funktion F.

d) Besitzt die Verteilung von Y eine Dichte?

22. S. ***Rechnen mit kontinuierlich verteilten Zufallsvariablen***

U sei uniform verteilt auf $[0, 1]$ und X sei standard-exponentialverteilt. Berechnen Sie

(i) den Erwartungswert
(ii) die Varianz
(iii) die Verteilungsfunktion
(iv) die Dichte

von

a) $3 + U^{1/3}$
b) $4X + 5$.

23. *Verteilungstransformationen hin und zurück*

a) X sei exponentialverteilt zum Parameter 2. Wie ist $Y := e^{-2X}$ verteilt?

b) Z sei standard-normalverteilt und Φ sei die Verteilungsfunktion von $\mathrm{N}(0,1)$. Wie ist $\Phi(Z)$ verteilt?

c) Φ sei wie in b), $\Phi^{-1} : (0,1) \to \mathbb{R}$ sei die Umkehrfunktion von Φ, und U sei uniform verteilt auf $(0,1)$. Wie ist $\Phi^{-1}(U)$ verteilt?

d) Geben Sie eine Funktion $h : (0,1) \to \mathbb{R}$ an, sodass für uniform auf $(0,1)$ verteiltes U die Zufallsvariable $h(U)$ normalverteilt ist mit Erwartungswert 10 und Varianz 4.

24. ***Wieviele Standardabweichungen?***

a) Wieviele Standardabweichungen muss man links und rechts vom Erwartungswert legen, damit eine normalverteilte Zufallsvariable mit W'kt 0.9 in das so entstehende Intervall fällt? Anders gesagt: Für welches c gilt $\mathbf{P}(X \in (\mu - c\sigma, \mu + c\sigma)) = 0.9$ für ein N(μ, σ^2)-verteiltes X?
Hinweis: Der R-Befehl `qnorm` *ist hilfreich. Warum reicht es, die Frage für* $\mu = 0$ *und* $\sigma = 1$ *zu beantworten?*

b) Finden Sie einen um den Erwartungswert symmetrischen Bereich, in den eine binomialverteilte Zufallsvariable mit $n = 1000$ und $p = 0.2$ mit Wahrscheinlichkeit annähernd 0.9 fällt.

25. S. (X_1, X_2) ist das Koordinatenpaar eines auf $S \subset \mathbb{R}^2$ uniform verteilten Punktes.

Sind X_1 und X_2

(i) unkorreliert (ii) unabhängig

für

a) $S := [-1, 1] \times [0, 1]$,

b) $S := ([-1, 0] \times [-1, 0]) \cup ([0, 1] \times [0, 1])$,

c) $S := \{(a_1, a_2) : a_1^2 + a_2^2 \leq 1, a_1 \geq 0\}$?

26. a) Berechnen Sie, gerundet auf 2 Nachkommastellen, die Zahl c jeweils so, dass eine normalverteilte Zufallsvariable Y mit Wahrscheinlichkeit (i) 0.01 (ii) 0.001
außerhalb ihrer $c\sigma_Y$-Grenzen fällt. (Zur Erinnerung: In R bekommt man die Verteilungsfunktion der Standardnormalverteilung als $\Phi(b) = \texttt{pnorm}(b)$ und die zugehörige *Quantilfunktion* als $\Phi^{-1}(p) = \texttt{qnorm}(p)$.)

b) Y sei eine reellwertige Zufallsvariable mit Erwartungswert μ, und d sei eine positive Zahl.
Sind die beiden Ereignisse "Y hat von seinem Erwartungswert einen Abstand $\leq d$" und "Das zufällige Intervall $I := [Y - d, Y + d]$ überdeckt die Zahl μ" gleich?

c) Y sei eine normalverteilte Zufallsvariable mit Varianz 16 und unbekanntem Erwartungswert μ. Geben Sie ein zufälliges Intervall I and, welches μ mit Wahrscheinlichhkeit (i) 0.99 (ii) 0.999
überdeckt.

27 S. a) Es sei B eine Bin$(n, 0.9)$-verteilte Zufallsvariable. Was ist das maximale n, für welches die Wahrscheinlichkeit, dass B den Wert 100 überschreitet, noch unter 0.025 bleibt? Beantworten Sie die Frage

i) unter Verwendung der R-Befehle `pbinom` und/oder `qbinom`
ii) durch Auffinden des maximalen n, für welches
$n \cdot 0.9 + 1.96\sqrt{n \cdot 0.9 \cdot 0.1} < 100 + \frac{1}{2}$
gilt.
iii) Wie kommt das Rezept in ii) mittels der Normalapproximation der Binomialverteilung zustande? Eine Skizze ist hilfreich.

b) Es sei bekannt, dass jede einzelne bis zum Tag x angenommene Buchung eines Fluges mit Wahrscheinlichkeit 0.1 nach dem Tag x storniert wird. Wieviele Buchungen dürfen für diesen Flug bis zum Tag x höchstens angenommen werden, wenn bei 100 Plätzen im Flugzeug alle gebuchten Passagiere mit einer Wahrscheinlichkeit von mindestens 0.975 Platz finden sollen?

28. X und Y seien zwei auf $\mathbb{R}$ standard-normalverteilte Zufallsvariable, und V sei der zufällige Punkt in $\mathbb{R}^2$ mit den Koordinaten X und Y.

a) Bestimmen Sie den Erwartungswert von $T := X^2 + Y^2$.

b) Sind (für $b > 0$) die beiden Ereignisse $\{T > b\}$ und "V fällt nicht in die Kreisscheibe $K := \{(x, y) : x^2 + y^2 \leq b\}$" gleich?

c) Wir nehmen nun zusätzlich an, dass X und Y unabhängig sind. Bestimmen Sie die Wahrscheinlichkeit $\mathbf{P}(T > b)$. Dabei dürfen Sie folgende Gleichheit verwenden:

$$\int_K e^{-(x^2+y^2)/2} dx\, dy = \int_0^{\sqrt{b}} e^{-r^2/2}\, 2\pi r\, dr.$$

d) Wie ist das Quadrat des Abstands vom Ursprung eines auf $\mathbb{R}^2$ standard-normalverteilten Punktes verteilt?

29 S. a) $x_1, \ldots, x_n$ seien reelle Zahlen, und $m := \frac{1}{n}(x_1 + \cdots + x_n)$. Begründen Sie die Identität

$$(*) \qquad \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - m^2$$

ohne weitere Rechung aus unserer "hilfreichen Formel für die Varianz" durch Angabe eines passenden Zufallsexperiments.*

*Die Geometrie hinter der Identität $(*)$ kann man so einsehen: $(x_1 - m, \ldots, x_n - m)$ und $(m, \cdots, m)$ sind zwei Vektoren im $\mathbb{R}^n$, deren Skalarprodukt Null ist (überprüfen Sie das!). Multipliziert man $(*)$ mit n, dann wird es zur Formel von Pythagoras für die (aufeinander ortogonal stehenden) Vektoren $(x_1 - m, \ldots, x_n - m)$ und $(m, \cdots, m)$.

b) $X_1, \ldots, X_n$ seien unabhängige reellwertige Zufallsvariable mit Erwartungswert μ und endlicher Varianz σ^2. Wie in Vorlesung 7b setzen wir $M := \frac{1}{n}(X_1 + \cdots + X_n)$ und $\hat{\sigma}^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i - M)^2$. Verwenden Sie (unter dem Motto "Was einer Gleichheit von Termen für Variablen recht ist, ist derseben Gleichheit für Zufallsvariablen billig") die Identität $(*)$ jetzt für X_i anstelle von x_i, um zu zeigen:

$$\mathbf{E}[\hat{\sigma}^2] = \frac{n-1}{n}\sigma^2.$$

Drücken Sie dazu $\mathbf{E}[X_i^2]$ und $\mathbf{E}[M^2]$ jeweils unter Verwendung der "hilfreichen Formel für die Varianz" durch μ und σ^2 aus.

30. Wir betrachten die Situation von Aufgabe 20. Es sei μ der in A20a) berechnete Populationsmittelwert, σ^2 die in A20a) berechnete Populationsvarianz, und $X_1, \ldots, X_{30}$ die zufälligen Werte, die bei einem 30-maligen Ziehen ohne Zurücklegen entstehen (vgl. A20b)).
a) Warum sind die X_i nicht unabhängig?

b) Wir setzen $M := \frac{1}{30}(X_1 + \cdots + X_{30})$. Es sei Ihnen verraten, dass trotz der fehlenden Unabhängigkeit der X_i auch hier die asymptotische Normalität greift, vgl. dazu die Ausblicke in VL 7a Folie 89 und VL 7b Folie 16. Deshalb dürfen Sie im Rest der Aufgabe mit der Normalapproximation rechnen, Sie sollten dabei aber die Standardabweichung von M verwenden, die wir in Aufgabe 20 berechnet haben.

(i) Für welche Zahl δ ist die Wahrscheinlichkeit, dass M um mehr als δ von μ abweicht, ungefähr gleich 0.05?

(ii) Geben Sie ein um M zentriertes Intervall I an, sodass $\mathbf{P}(\mu \in I) \approx 0.95$.*

c) Überprüfen Sie, mit welcher Wahrscheinlichkeit das in b) ii) berechnete Intervall dem Populationsmittelwert μ überdeckt, indem Sie den Rechner viele Stichproben der Größe 30 ziehen lassen. Verwenden Sie dazu das unter dem Link A30.R auf der Lehrveranstaltungs-Webseite bereitgestellte R-Programm.

*Ein zufälliges Intervall I mit der Eigenschaft $\mathbf{P}(\mu \in I) \approx 0.95$ heißt *Konfidenzintervall für μ, approximativ zum Niveau* 0.95.

31. $X_1, \ldots, X_k$ seien unabhängig und Exp(1)-verteilt (mit $k \in \mathbb{N}$), und $Y := X_1 + \ldots + X_k$.

Berechnen Sie für $-\infty < t < 1$

(i) $\mathbf{E}[e^{tX_1}]$, (ii) $\mathbf{E}[e^{tY}]$.

32 S. X_1, X_2 und Y seien reellwertige Zufallsvariable. Die Standardabweichungen von X_1 und X_2 seien gleich.

a) Das Wievielfache von σ_{X_1} ist $\sigma_{X_1+X_2}$, wenn
(i) X_1 und X_2 unkorreliert sind,
(ii) X_1 und X_2 Korrelationskoeffizient -0.5 haben?

b) Die Standardabweichung von Y sei doppelt so groß wie die von X_1+X_2 und die Standardabweichungen von X_1 und X_2 seien gleich. Der Korrelationskoeffizient von X_1 und X_2 sei -0.5. Außerdem sei der Korrelationskoeffizient von Y und X_1 gleich dem Korrelationskoeffizienten von Y und X_2. Die beste affin lineare Prognose von Y auf der Basis von $X_1 + X_2$ sei $\hat{Y} = (X_1 + X_2) + 5$. Berechnen Sie den Korrelationskoeffizienten von Y und X_1.

33. S. ***Diskret und kontinuierlich - diesmal gemischt.*** Die Verteilungsfunktion F der reellwertigen Zufallsvariablen Y habe die folgenden Eigenschaften (i)-(iii):
(i) $F(0) = 0$, $F(5) = 0.7$, $F(8) = 0.8, F(10) = 1$.
(ii) Auf jedem der Intervalle $(0, 5)$, $(5, 8)$ und $(8, 10)$ hat F konstante Steigung.
(iii) F hat an der Stelle 5 einen Sprung der Höhe 0.3, ansonsten hat F keine Sprünge.

a) Skizzieren Sie F.
b) Wir fassen jetzt Y als Ergebnis der zweiten Stufe eines zweistufigen Zufallsexperimentes auf, in dem wir in der ersten Stufe entscheiden, ob die einelementige Menge $\{5\}$ zum Zug kommt oder eines der drei angegebenen Intervalle. Beschreiben Sie die Verteilungsgewichte der ersten Stufe, sowie für jeden der 4 Ausgänge der ersten Stufe die Übergangsverteilung.
c) Berechnen Sie den Erwartungswert von Y.

34. ***Box-Muller Verfahren.*** Mir wurde von den Tutoren berichtet, dass einige von Ihnen gerne ein ähnlich hübsches Verfahren zur Erzeugung einer standard-normalverteilten Zufallsvariablen sehen würden, wie wir das für eine Exp(1)-verteilte Zufallsvariable kennengelernt haben (nämlich $T := -\ln U$, mit U uniform auf $[0, 1]$). Über diese Neugierde habe ich mich gefreut.

Was halten Sie von folgendem Verfahren: U und V seien unabhängige auf $[0, 1]$ uniform verteilte Zufallsvariable. Erzeuge einen zufälligen Punkt $\vec{Z} = (Z_1, Z_2)$ in $\mathbb{R}^2$ so: Als Abstand R des Punktes $\vec{Z}$ vom Ursprung nimm $\sqrt{-2 \ln U}$, und als seinen *Winkel* Θ (im Bogenmaß) nimm $2\pi V$ (damit ist $\vec{Z}$ rotationssymmetrisch verteilt). Stellen Sie sich vor, Ihr Opa oder Ihre Chefin ist Hobby-Stochastiker*in und hat Aufgabe 28 schon verstanden. Wie können Sie ihn/sie überzeugen, dass

(i) Z_1 und Z_2 unabhängig und $N(0, 1)$-verteilt sind,

(ii) die Zufallsvariable $\sqrt{-2 \ln U} \cos(2\pi V)$ standard-normalverteilt ist?

Bei Ihrer Überzeugungsarbeit dürfen Sie sich auf folgendes Argument stützen: *Jede rotationssymmetrische Verteilung auf $\mathbb{R}^2$ eintsteht auf zweistufige Weise so: Wähle in der ersten Stufe einen zufälligen Radius, d.h. Abstand vom Ursprung. Gegeben den Ausgang der ersten Stufe, wähle den Winkel (im Bogenmaß) uniform verteilt auf $[0, 2\pi)$. Insbesondere ist also jede rotationssymmetrische Verteilung auf $\mathbb{R}^2$ durch die Verteilung des zufälligen Radius bestimmt.*

35. S. ***Zweistufigkeit hin und zurück.*** Das zufällige Paar (X_1, X_2) mit Werten in $\{1,2,3\}\times\{b,c,d\}$ komme durch ein zweistufiges Experiment zustande, wobei die Verteilung von X_1 uniform sei und die Übergangswahrscheinlichkeiten $P(a_1,.)$, $a_1 \in \{1,2,3\}$, durch die rechts angegebene Matrix bestimmt sind.

	b	c	d
1	0	0.5	0.5
2	0.3	0.2	0.5
3	0.6	0.3	0.1

(i) Finden Sie die Matrix der gemeinsamen Verteilungsgewichte von (X_1, X_2) und die Verteilungsgewichte von X_2.

(ii) Finden Sie Übergangswahrscheinlichkeiten $Q(a_2,.),\ a_2 \in \{b,c,d\}$ so, dass das zufällige Paar (X_2, X_1) als zweistufiges Zufallsexperiment, jetzt mit X_2 als erster Stufe, entsteht.

(iii) Berechnen Sie den bedingten Erwartungswert von X_1 gegeben $X_2 = b$.

36. ***Erwartete Suchtiefe.*** 15 Namen sind in 5 Listen einsortiert. Die Längen $Z_1, \ldots, Z_5$ der Listen sind identisch verteilt und haben Varianz 16. Die Suchtiefen der in Liste j einsortierten Namen sind $0, 1, \ldots, Z_j - 1$.

a) Finden Sie $\mathbf{E}[Z_j]$.

b) Aus den 15 Namen wird rein zufällig einer gewählt. Berechnen Sie den Erwartungswert seiner Suchtiefe.

37. S. X sei uniform verteilt auf $[-1, 1]$, Z sei $\mathsf{N}(0, 1)$-verteilt und unabhängig von X. Es sei $Y := X^3 + \sigma Z$, mit $\sigma > 0$.

a) Finden Sie die im Sinn des erwarteten quadratischen Fehlers

i) beste ii) beste affin lineare

Prognose von Y auf der Basis von X,
d.h. jeweis eine Funktion $e : [-1, 1] \to \mathbb{R}$, für die $\mathbf{E}[(Y - e(X))^2]$ minimal wird, wobei in ii) die Form $e(x) = \beta_0 + \beta_1 x$ verlangt wird.

b) Was ist jeweils der erwartete quadratische Prognosefehler $\mathbf{E}[(Y - e(X))^2]$ und der erwartete quadratische Bias? (Siehe dazu Folie 10 in V9b.)

38. Aus der Vereinigung von drei disjunkten Populationen, deren Größen im Verhältnis $1:2:3$ stehen, wird rein zufällig ein Individuum J gewählt und eine reelles Merkmal $h(J)$ beobachtet. In den drei Populationen ist das Merkmal der Individuen jeweils uniform verteilt auf einem Intervall, und zwar in der kleinsten Population (d.h. der mit den wenigsten Individuen) auf dem Intervall $[20, 50]$, in der größten Population auf $[30, 40]$, und in der zweitgrößten Population auf $[30, 80]$.

a) Was ist (i) der bedingte Erwartungswert, (ii) die bedingte Varianz,

von $h(J)$, gegeben dass J aus der Population mit den wenigsten Individuen gewählt wurde?

b) Berechnen Sie $\mathbf{Var}[h(J)]$, also die Varianz des Merkmals in der Gesamtpopulation.

39. S (Frei nach dem Eingangsbeispiel im 2. Vortrag der Ringvorlesung 2018/19)

https://www.mathe-uni-ffm.de/ringvorlesung/algorithmen-maschinelles-lernen-quantencomputing

Jemand führt einen Münzwurf vor. Aus gewissen Gründen kommt nur in Frage, dass er entweder die ganze Zeit eine faire 01-Münze verwendet, oder eine mit $p = 0.9$. Bevor er beginnt, schätzen Sie die Wahrscheinlichkeit, dass er eine faire Münze verwendet, mit 0.8 ein.

Wie aktualisieren Sie die Wahrscheinlichkeit, dass es sich um eine faire Münze handelt, nachdem

(i) beim ersten Wurf eine Eins
(ii) bei den ersten drei Würfen jeweils eine Eins

geworfen wurde?

40. a) U_1 und U_2 seien unabhängige, auf $[0, 1]$ uniform verteilte Zufallsvariable. Bestimmen Sie

(i) die Verteilungsfunktion (ii) die Dichte

von $Y := \max(U_1, U_2)$.

b) X_1, X_2 und X_3 seien unabhängig und Exp(1)-verteilt. Begründen Sie, warum die Zufallsvariable $X_1 + X_2$ gegeben $\{X_1 + X_2 + X_3 = 1\}$ so verteilt ist wie Y aus a). Argumentieren Sie
(i) anschaulich über einen Münzwurf mit kleiner Erfolgswahrscheinlichkeit durch Betrachten der Zeitpunkte der ersten drei Erfolge (vgl. V10a Abschnitte 2 und 3)
sowie
(ii) durch eine Rechnung analog zu der im Beispiel in V10 Abschnitt 4.

Zusatzaufgabe. U_0, U_1, U_2, U_3 seien unabhängig und uniform auf $[0, 1]$ verteilt.

(a) Für $i = 0, 1, 2, 3$ setzen wir $R(i) := \sum_{j \in \{0,1,2,3\} \setminus \{i\}} I_{\{U_j < U_i\}}$ Begründen Sie, warum $(R(0), R(1), R(2), R(3))$ eine rein zufällige Permutation von 0,1,2,3 ist.

b) Bestimmen Sie $\mathbf{P}(U_1 < U_0, U_2 < U_0)$ und
$\mathbf{P}(U_1 < U_0, U_2 < U_0, U_3 \geq U_0)$.

c) Wir definieren $Z_1 := I_{\{U_1 < U_0\}}, Z_2 := I_{\{U_2 < U_0\}}, Z_3 := I_{\{U_3 < U_0\}}$.
(i) Berechnen Sie $\mathbf{P}(Z_2 = 1 | Z_1 = 1)$ und $\mathbf{P}(Z_3 = 0 | Z_1 = 1, Z_2 = 1)$.
(ii) Tragen Sie die Gewichte der Übergangsverteilungen des durch (Z_1, Z_2, Z_3) beschriebenen 3-stufigen Experiments in einen Baum der Tiefe 3 ein.

41. Für die Zyklendarstellung einer Permutation hat sich eine suggestive Schreibweise eingebürgert, die an dem folgenden Beispiel einsichtig wird: Die Zyklendarstellung der in Vorlesung 2a betrachteten (und dort auf Folie 14 veranschaulichten) Permutation $5, 2, 7, 3, 1, 4, 6$ von $1, \ldots, 7$ schreibt man als $(1\,5)(2)(3\,7\,6\,4)$.
Wir beschreiben jetzt ein rekursives Verfahren zur Erzeugung einer zufälligen Permutation von $1, \ldots, i$ aus einer Permutation von $1, \ldots, i-1$, ausgehend von deren Zyklendarstellung: *Das Element i wird jeweils mit W'keit $\frac{1}{i}$ auf einen der $i-1$ Plätze rechts neben $1, 2, \ldots, i-1$ (innerhalb des jeweiligen Zyklus) gesetzt. Ebenfalls mit W'keit $\frac{1}{i}$ wird das Element i in einen neuen Zyklus (der Länge 1) gesetzt.*

Dieses Verfahren fassen wir als mehrstufiges Zufallsexperiment namens $(X_1, \ldots, X_n)$ auf, bei dem der Wertebereich von X_i die Permutationen von $1, \ldots, i$ sind.

a) Beweisen Sie induktiv (mittels der Multiplikationsregel), dass das oben beschriebene Verfahren auf eine rein zufällige Permutation von $1, \ldots, n$ führt.

b) Mit welcher Wahrscheinlichkeit liegen in einer rein zufälligen Permutation von $1, \ldots, 100$ die Zahlen $1, 2, 3$ im selben Zyklus?

c) Mit welcher Wahrscheinlichkeit liegen in einer rein zufälligen Permutation von $1, \ldots, 100$ die Zahlen $50, 60, 75$ im selben Zyklus?

42. S Wir betrachten die gewöhnliche Irrfahrt auf dem skizzierten Graphen.

a) Berechnen Sie

i) die Wahrscheinlichkeit, bei Start in c den Zustand f vor dem Zustand b zu treffen,

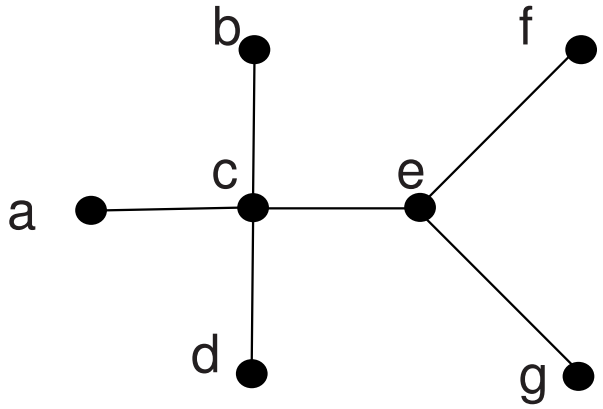
ii) die Wahrscheinlichkeit, bei Start in g nach drei Schritten in e zu sein,

iii) die Wahrscheinlichkeit, bei Start in e nach drei Schritten in g zu sein.

b) Sei π die Gleichgewichtsverteilung dieser Irrfahrt. Berechnen Sie

i) $\mathbf{P}_\pi(X_3 = e)$,

ii) $\mathbf{P}_\pi(X_0 = e | X_3 = g)$.

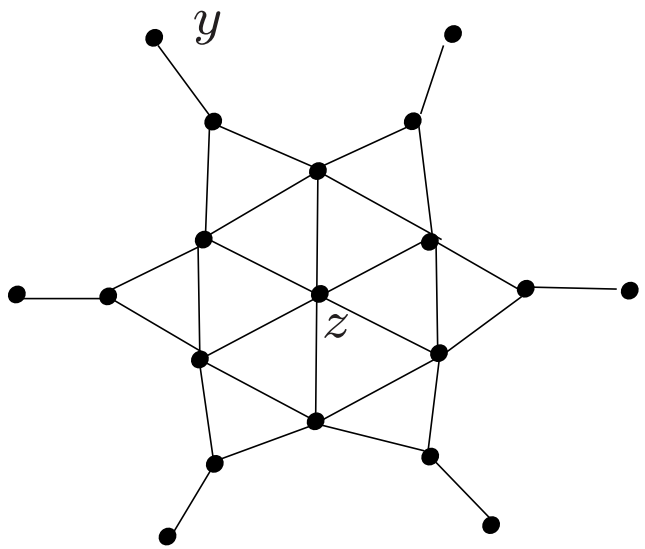


43. S Wir betrachten die gewöhnliche Irrfahrt auf dem skizzierten Graphen. Berechnen Sie

a) die erwartete Treffzeit des "zentralen" Knotens z bei Start im Knoten y,

b) die Gleichgewichtsverteilung der Irrfahrt.

Hinweis zu a): Wieso reicht es hier, den "Abstand" vom Zentrum zu betrachten, sodass man am Ende kein großes Gleichungssystem lösen muss?



44. Die Markovkette X auf $\{0, 1, 2, 3\}$ habe die folgenden Übergangsgewichte:
$P(0, 1) = 1$, $P(1, 0) = 1/5$, $P(1, 1) = 2/5$, $P(1, 2) = 2/5$,
$P(2, 1) = 2/3$, $P(2, 3) = 1/3$, $P(3, 2) = 1$.

Berechnen Sie die erwartete Treffzeit des Zustands 3 bei Start im Zustand 0.

45. S a) Es geht darum, in einer großen Population den Anteil p der Individuen mit einem bestimmten Merkmal mit einem approximativen 99%-Konfidenzniveau auf $\pm 2\%$ genau zu schätzen. Wie groß muss dafür der Stichprobenumfang n mindestens sein? Finden Sie jeweils eine möglichst gute untere Schranke für n,
(i) die vom Anteilschätzer $\widehat{p}$ abhängt,
(ii) die nicht von $\widehat{p}$ abhängt.
b) Bei 28% der Individuen einer großen Population wird ein bestimmtes Merkmal festgestellt. Wie groß muss die Stichprobengröße n mindestens sein, damit die Hypothese "der Anteil des Merkmals in der Population ist 30 %" mit einem p-Wert 0.01 abgelehnt werden kann?

Verwenden Sie in beiden Aufgabenteilen die asymptotische Normalität des Anteilschätzers $\widehat{p}$.

46. S Aus zwei reellwertigen Stichproben $x_1, x_2, \ldots, x_{100}$ und $y_1, y_2, \ldots, y_{200}$, die aus zwei großen Populationen $\mathcal{P}_1$ und $\mathcal{P}_2$ rein zufällig gezogen wurden, ergaben sich die Stichprobenmittelwerte $\bar{x} = 10$ und $\bar{y} = 11$ sowie die Stichprobenstandardabweichungen $s_x = 2.5$ und $s_y = 3.0$.

a) Was ist die geschätzte Standardabweichung
(i) des zufälligen Stichprobenmittelwertes M_X
(ii) des zufälligen Stichprobenmittelwertes M_Y
(iii) der Differenz $M_X - M_Y$?

b) Geben Sie ein approximatives 99%-Konfidenzintervall für die Differenz der beiden Populationsmittelwerte an.

c) Zu welchem p-Wert können Sie die Hypothese ablehnen, dass die beiden Populationsmittelwerte gleich sind?

47. Wir betrachten die in Aufgabe 13 beschriebene Situation. Zu welchem p-Wert lässt sich die Hypothese, das das Ziehen der Stichprobe rein zufällig erfolgte, ablehnen unter Verwendung

(i) von Fishers exaktem Test

(ii) der asymptotischen Normalität des Anteilschätzers einschließlich der Varianzkorrektur für endliche Populationen?

48. $\mathcal{P}$ sei eine Menge bestehend aus 40 reellen Werten, und T sei eine zufällige vierelementige Teilmenge von $\mathcal{P}$. Sie bekommen gesagt, dass T aus dem kleinsten, zweitkleinsten, viert- und achtkleinsten Element von $\mathcal{P}$ besteht. Zu welchem p-Wert können Sie unter Verwendung des Wilcoxon-Rangsummentests die Hypothese der reinen Zufälligkeit der Wahl von T (zugunsten einer Tendenz “hin zum Rand von $\mathcal{P}$”) verwerfen?

Zusatzaufgabe.* $X_1, \ldots, X_{10}$ seien reellwertige unabhängig und identisch verteilte Zufallsvariable. Ihre Verteilung ρ besitze eine Dichte. Berechnen Sie dier Wahrscheinlichkeit, mit der das Intervall $[X_{(2)}, X_{(9)}]$ den Median von ρ überdeckt.

*Ein schriftliches Bearbeiten dieser Aufgabe ist optional. Es wird wie auch die anderen S-Aufgaben auf die Bonuspunkte angerechnet.