1. S. Wie in der Vorlesung betrachten wir eine Gesamtfläche G bestehend aus g Pixeln und eine Teilfläche F bestehend aus f Pixeln. Jetzt geht es um eine dreimal wiederholte rein zufällige Wahl eines Pixels aus G (anders gesagt: um ein Ziehen aus G mit Zurücklegen), beschrieben durch eine auf dem Wertebereich $\{1, \ldots, g\}^3$ uniform verteilte Zufallsvariable $X = (X_1, X_2, X_3)$.

a) Wieviele verschiedene mögliche Ausgänge von X gibt es, bei denen zwei der X_i auf die Menge $\{1, \ldots, f\}$ und eines auf die Menge $\{f + 1, \ldots, g\}$ fallen? Drücken Sie das Ergebnis durch g und durch $p := f/g$ aus.

b) Wie wahrscheinlich ist es, dass genau zwei von den drei Pixeln aus F gewählt werden?

c) Es sei M die zufällige Trefferquote von F . Illustrieren Sie für $p = 0.195$ und $n = 3$ das Ergebnis aus b) 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.

2. Erkunden Sie in der in Aufgabe 1 beschriebenen Situation (wieder für $p = 0.195$) mittels des R-Programms “Monte Carlo Simulation”, wie sich die Genauigkeit der Schätzung verändert, wenn (i) $n = 100$ (ii) $n = 400$ (iii) $n = 1600$ Punkte in die Menge G geworfen werden: Um welchen Faktor (circa) wird jeweils das Histogramm der Schätzwerte schmäler?

3. Q sei das Quadrat $[-1,1] \times [-1,1]$. Wir denken uns Q aus 1000 mal 1000 Pixel bestehend. Es sei D die Menge der Pixel auf der Diagonale von Q, und F die Menge der Pixel, die von der Diagonalen einen Abstand höchstens 0.5 besitzen. Die Zufallsvariable X beschreibe die rein zufällige Wahl eines Pixels aus Q. Berechnen Sie die Wahrscheinlichkeit des Ereignisses

a) $\{X \in D\}$, b) $\{X \in F\}$.

Begnügen Sie sich dabei in b) mit einer Näherungslösung, indem Sie einen passenden Flächeninhalt berechnen.

4 S. 3 Studierende wählen rein zufällig (und ohne irgendeine gegenseitige Absprache) je eine von 6 möglichen Gruppen. Wie wahrscheinlich ist es, dass sie alle in verschiedenen Gruppen landen? Verwenden Sie

(i) die exakte Berechnung

(ii) die Stirling-Approximation

(iii) die in der Vorlesung diskutierte Näherung $\exp(-\frac{n(n-1)}{2r})$

und vergleichen Sie die Ergebnisse.

5. Für die Zyklendarstellung einer Permutation hat sich eine suggestive Schreibweise eingebürgert, die schon an einem Beispiel einsichtig wird: Die Zyklendarstellung der in der Vorlesung betrachteten 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 $1, \ldots, n+1$ aus einer Permutation von $1, \ldots, n$, ausgehend von deren Zyklendarstellung: *Das Element $n+1$ wird jeweils mit W'keit $\frac{1}{n+1}$ auf einen der n Plätze rechts neben $1, 2, \ldots, n$ (innerhalb des jeweiligen Zyklus) gesetzt. Ebenfalls mit W'keit $\frac{1}{n+1}$ wird das Elemet in einen neuen Zyklus (der Länge 1) gesetzt.*

a) Zeichnen Sie (in Form eines Baumes) die 6 Pfade, die, ausgehend vom trivialen Zyklus (1), zu den 6 Permutationen von $1,\ 2,\ 3$ führen.

b) Begründen Sie induktiv, dass zu jeder der $n!$ Permutationen von $1, 2, \ldots, n$ genau ein Pfad (im Sinn von a)) führt.

c) Warum liefert der Algorithmus für jedes $n \in \mathbb{N}$ eine rein zufällige Permutation von $\{1, \ldots, n\}$?

d) Mit welcher Wahrscheinlichkeit liegen in einer rein zufälligen Permutation von $\{1, \ldots, 100\}$
(i) 1, 2 und 3 (ii) 80, 90 und 100
im selben Zyklus?

6. a) Wieviele Möglichkeiten gibt es, aus f Frauen und m Männern ein k-köpfiges Komitee (ohne Reihung) auszuwählen, in dem w Frauen sind? (Dabei ist $w \leq k \leq f + m$.)

b) Beweisen Sie ohne weitere Rechnung, dass

$$\sum_{w=0}^{k} \binom{f}{w} \binom{m}{k-w} = \binom{f+m}{k}.$$

7. S. i) (X_1, X_2, X_3) sei eine uniform verteilte Besetzung von 3 Plätzen mit 10 Objekten. Wie wahrscheinlich ist es, dass jeder Platz mit mindestens zwei Objekten besetzt wird? Skizzieren Sie die Menge der zugehörigen Ausgänge als Teilmenge des in der Vorlesung 2a betrachteten de Finetti-Dreiecks.

(ii) Sei $2r \leq n$. Begründen Sie: Die Anzahl der Besetzungen von r Plätzen mit n Objekten, die auf jeden Platz mindestens zwei Objekte setzen, ist gleich der Anzahl der Besetzungen von r Plätzen mit $n - 2r$ Elementen.[†]

(iii) T sei eine rein zufällige 2-elementige Teilmenge der Menge $\{1, \dots, 20\}$. Wir setzen $V := T \cup \{0, 21\}$. Wie wahrscheinlich ist es, dass je zwei verschiedene Elemente von V mindestens den Abstand 3 haben?

[†]Zur Erinnerung: Bei einer *Besetzung* (als Synonym für ein r-Tupel von Besetzungszahlen) unterscheiden wir nicht, mit welchen Objekten der jeweilige Platz besetzt ist.

8. S. a) In einer Urne befinden sich 10 rote, 20 blaue und 30 grüne Kugeln. Man zieht n-mal rein zufällig mit Zurücklegen und notiert $Z_i = 1$ (bzw. $Z_i = 2$ bzw. $Z_i = 3$) falls die i-te gezogene Kugel rot (bzw. blau bzw. grün) ist. Prüfen Sie nach, dass $(Z_1, \dots, Z_n)$ ein n-faches (p_1, p_2, p_3)-Würfeln (mit passendem (p_1, p_2, p_3)) im Sinn der in der Vorlesung gegebenen Definition ist.
b) Überprüfen Sie anhand der Definition (und illustrieren Sie anhand des Beispiels in Teil a)): Wenn man bei einem n-fachen $(p_1, \dots, p_r)$-Würfeln $(Z_1, \dots, Z_n)$ zwei Ausgänge zu einem zusammenfasst in dem Sinn, dass man $Z_i' = r - 1$ setzt falls $Z_i \in \{r-1, r\}$, und $Z_i' = Z_i$ sonst, dann ist $(Z_1', \dots, Z_n')$ ein n-faches $(p_1, \dots, p_{r-2}, p_{r-1} + p_r)$-Würfeln.
c) $(X_1, \dots, X_r)$ sei multinomial $(n; p_1, \dots, p_r)$-verteilt. Wie ist dann $(X_1 + X_2, X_3, \dots, X_r)$ verteilt?

9. S a) Für natürliche Zahlen r und n sei $(X_1, \ldots, X_n)$ uniform verteilt auf $\{1, \ldots, r\}^n$. Berechnen Sie die erwartete Anzahl der Kollisionen, d.h. den Erwartungswert der Anzahl der Paare (i, j) mit $1 \leq i < j \leq n$ und $X_i = X_j$. Fällt Ihnen ein Zusammenhang mit einer der in Vorlesung 1b hergeleiteten Näherungen für die “Wahrscheinlichkeit von null Kollisionen” auf?

b) Es seien p_1, p_2 und p_3 drei nichtnegative Zahlen mit $p_1 + p_2 + p_3 = 1$. In einem vollständigen Graphen mit n Knoten werden die Kanten per (p_1, p_2, p_3)-Würfeln blau, rot oder grün gefärbt. Berechnen Sie die erwartete Anzahl der Dreicke mit drei gleichfarbigen Seiten.

10. X sei binomialverteilt zu den Parametern n und p. Berechnen Sie $\mathbf{E}[X^2]$, indem Sie X als Summe von Zählvariablen schreiben.

11. $X_1, \ldots, X_{10}$ seien die Resultate eines rein zufälligen Ziehens ohne Zurcklegen aus der Menge $\{1, \ldots, 100\}$. Berechnen Sie den Erwartungswert der Zufallsvariablen $M_n := \frac{1}{10} \sum_{i=1}^{10} X_i^2$.

Zur Erinnerung: $\sum_{k=1}^{m} k^2 = \frac{1}{6} m(m+1)(2m+1)$.

12. S 32 Karten, bestehend aus je 8 derselben Farbe, werden in rein zufälliger Reihenfolge aufgeschlagen. Für jede der 32 Karten bekommen Sie einen Euro, wenn sie dieselbe Farbe hat wie die unmittelbar zuvor oder die unmittelbar danach aufgeschlagene Karte.[‡] Berechnen Sie die erwartete Gesamtauszahlung. (Man beachte: Die erste aufgeschlagene Karte hat keinen Vorgänger und die letzte aufgeschlagene Karte hat keinen Nachfolger.)

[‡]Oder ist hier - wie auch sonst in dem in der Stochastik üblichen Sprachgebrauch - im nicht ausschließenden Sinn verstanden. Haben z.B. die ersten 4 aufgeschlagenen Karten die Farben *rot, rot, rot, blau*, dann bekommen Sie aus den ersten 3 Aufschlägen insgesamt 3 Euro. Haben die ersten 4 aufgeschlagenen Karten die Farben *rot, blau, rot, rot*, dann bekommen Sie aus den ersten 3 Aufschlägen insgesamt 1 Euro.

13. In Aufgabe 10 haben wir gezeigt, dass für eine Bin(n, p)-verteilte Zufallsvariable X_n gilt: $\mathbf{E}[X_n^2] = npq + (np)^2$.
a) Folgern Sie daraus mit der Linearität des Erwartungswertes, dass

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

b) Beweisen Sie mit der Ungleichung von Markov: Für jedes $\varepsilon > 0$ ist

$$\mathbf{P}\left(\left(\frac{X_n}{n} - p\right)^2 \geq \varepsilon^2\right) \leq \frac{1}{\varepsilon^2}\frac{pq}{n}.$$

c) Von b) ist es nur mehr ein kleiner Schritt zum *Gesetz der großen Zahlen von Jakob Bernoulli*:
Sei $Z_1, Z_2, \ldots$ *ein fortgesetzter* p*-Münzwurf und* $\varepsilon > 0$*. Dann ist* $\mathbf{P}(|\frac{1}{n}(Z_1 + \cdots + Z_n) - p| \geq \varepsilon)$, $n = 1, 2, \ldots$ *eine gegen* 0 *konvergente Folge, d.h. die Wahrscheinlichkeit, dass die relative Anzahl der Erfolge um mehr als* ε *von der Erfolgswahrscheinlichkeit* p *abweicht, wird verschwindend klein im Grenzwert* $n \to \infty$*.* Führen Sie diesen Schritt durch.

14. S Wir betrachten dieselbe Situation wie in Aufgabe 9a) und interpretieren X_i als die zufällige Platzwahl des Individuums i, mit r möglichen Plätzen für das Individuum i, $i = 1, \ldots, n$. (Zur Erinnerung: Mehrfache Wahl eines Platzes ist erlaubt, es handelt sich um ein n-faches $(1/r, \ldots, 1/r)$-Würfeln.) Es sei $n = 8$ und $r = 10$.

a) Mit welcher Wahrscheinlichkeit wird weder der Platz 2 noch der Platz 3 gewählt?

b) Mit welcher Wahrscheinlichkeit wird mindestens einer der Plätze 1, 2 oder 3 nicht gewählt?

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. Wieder betrachten wir dieselbe Situation wie in den Aufgaben 9a und 14, diesmal mit $n = 3r$. Berechnen Sie für $r \to \infty$ den Grenzwert der Wahrscheinlichkeit, dass die Plätze $1, 2, 3$

a) allesamt leer ausgehen

b) ingesamt genau viermal gewählt werden.

16. S a) Mit welcher Wahrscheinlichkeit kommt beim p-Münzwurf mit $p = 1/4$ der erste Erfolg später als zum Zeitpunkt 10, aber nicht später als zum Zeitpunkt 20?
b) Wir betrachten jetzt einen Münzwurf, bei dem pro Millisekunde ein Versuch kommt. Die Erfolgswahrscheinlichkeit sei so eingestellt, dass die erwartete Anzahl der Erfolge in 20 Sekunden gleich 5 ist. Berechnen Sie *näherungsweise* die Wahrscheinlichkeit dafür, dass der erste Erfolg später als 10 Sekunden, aber nicht später als 20 Sekunden kommt.

17 S. a) S_1, S_2 und S_3 seien endliche Mengen, und (X_1, X_2, X_3) sei uniform verteilt auf $S_1 \times S_2 \times S_3$. Sind dann X_1, X_2, X_3 unabhängig? Geben Sie eine Begründung oder ein Gegenargument.

b) (X_1, X_2) sei uniform verteilt auf
$\{(-1,0),(0,1),(0,-1),(1,0)\}$.
Sind X_1 und X_2 unabhängig?
Sind X_1 und X_2 unkorreliert?

18. a) Die Verteilung des zufälligen Paares (X_1, X_2) mit Werten in $S_1 \times S_2$ lässt sich angeben durch die Matrix der gemeinsamen Verteilungsgewichte $\rho(a_1, a_2)$, $a_1 \in S_1$, $a_2 \in S_2$. Wir betrachten vier Beispiele, bei den ersten beiden ist $S_1 = \{1, 2\}$, $S_2 = \{b, c\}$ bei den letzten beiden ist $S_1 = \{1, 2, 3\}$, $S_2 = \{b, c, d\}$.

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

jeweils mit $\gamma := \frac{1}{138}$. In welchen Fällen sind X_1 und X_2 unabhängig, und in welchen nicht? Bei (iii) und (iv) sei Ihnen dabei erlaubt, das folgende Kriterium zu verwenden: *X_1 und X_2 sind genau dann unabhängig, wenn die Zeilen der Matrix der gemeinsamen Verteilungsgewichte zueinander proportional sind*.

b) (als Kür für diejenigen, die der Sache auf den Grund gehen wollen): Knöpfen Sie sich Übung 25b aus dem WS 15/16 samt den dort gegegeben ausführlichen Lösungshinweisen vor, und leiten Sie damit das obige Kriterium her. (Auf die alte Lehrveranstaltungsseite werden Sie geführt, indem Sie in der Web-Adresse der aktuellen Seite die Jahreszahlen verändern.)

19. S In einer Population von 100 Individuen haben 60 Individuen die Größe 10, 30 die Größe 5 und 10 die Größe 15. Es sei X die Größe eines rein zufällig aus der Population gewählten Individuums.
a) Berechnen Sie
(i) den Erwartungswert μ, (ii) die Varianz σ^2, (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 $1 \leq i \neq j \leq 100$) 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 $0 =$ **Var**$(X_1 + \ldots + X_{100})$.
γ) Berechnen Sie die Varianz des Stichprobenmittels $\frac{1}{10}(X_1 + \cdots + X_{10})$.

20. Schätzen Sie in der Situation von Aufgabe 19 mittels der Ungleichung von Chebyshev die Wahrscheinlichkeit dafür ab, dass das Stichprobenmittel $\frac{1}{10}(X_1 + \cdots + X_{10})$ um mehr als 2 von μ abweicht.

21 S. U sei uniform auf $[0,3]$ verteilt, und X sei Exp(3)-verteilt.

Berechnen Sie

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

von

a) U^4 b) $6X+2$.

22. a) Die Zufallsvariable U sei uniform auf $[0, 1]$ verteilt. Finden Sie eine Abbildung $h : [0, 1] \to \mathbb{R}$, sodass $h(U)$
i) Bin (3, 1/4)-verteilt
ii) Exp(3)-verteilt
ist. Skizzieren Sie jeweils die Lösung.
Hinweis zu i): Zerlegen Sie das Einheitsintervall in vier Teilintervalle passender Länge. Zu ii): Das letzte Beispiel aus Vorlesung 6a ist hilfreich.

b) Z sei standard-normalverteilt. Berechnen Sie den Erwartungswert von e^Z. *Hinweis: Erweitern Sie den mittels der Dichte von Z ausgedrückten Erwartungswert $\mathbf{E}[e^Z]$ mit $e^{-1/2}e^{1/2}$ und verwenden Sie die Identität $x^2 - 2x + 1 = (x-1)^2$.*

23. a) Mit welcher Wahrscheinlichkeit fällt eine Exp(1/4)-verteilte Zufallsvariable in das Intervall $[10, 20]$? Stellen Sie eine Beziehung zur Aufgabe 16 b) her.

b) Mit welcher Wahrscheinlichkeit haben die Komponenten X_1 und X_2 einer auf $[0, 1] \times [0, 1]$ uniform verteilten Zufallsvariablen $X = (X_1, X_2)$ den Abstand $\geq \frac{1}{2}$?

24. S a) X sei normalverteilt mit Erwartungswert 5 und Varianz 16. Bestimmen Sie Zahlen c und d so, dass $Y := d + cX$ mit Wahrscheinlichkeit 0.95 in das Intervall $[9, 11]$ fällt. Rechnen Sie dabei mit der Näherung $\mathbf{P}(|Z| \leq 2) \approx 0.95$ für standard-normalverteiltes Z.

b) Für $n \in \mathbb{N}$, $p \in (0, 1)$ und $q := 1 - p$ sei X Bin(n, p)-verteilt und Y N(np, npq)-verteilt. In der Vorlesung haben wir zumindest ansatzweise begründet, warum für großes npq gilt:

$$P(X = k) \approx \mathbf{P}(k - \frac{1}{2} \leq Y \leq k + \frac{1}{2}).$$

In diesen Sinn ist die Verteilung von X durch die von Y approximierbar. Die Zahl n sei nun gerade mal so groß, dass $n \cdot 0.9 + 2\sqrt{n \cdot 0.9 \cdot 0.1} \geq 100 + \frac{1}{2}$. Begründen Sie, warum für dieses n und ein Bin$(n, 0.9)$-verteiltes X gilt: $\mathbf{P}(X > 100) \approx 0.025$. Eine Skizze ist hilfreich!

c) Es sei bekannt, dass jede einzelne bis zum Tag x angenommene Buchung eines Flug-es 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?

25. S *(Serien- und Parallelschaltung)* Für $i = 1, 2, 3$ sei X_i exponentialverteilt zum Parameter $2i$ und die X_i seien unabhängig.

a) Berechnen Sie die Wahrscheinlichkeit dass

(i) jedes der X_i größer als 2 ausfällt

(ii) mindestens eines der X_i größer als 2 ausfällt.

b) Berechnen Sie die Dichte von

(i) $\min(X_1, X_2, X_3)$

(ii) $\max(X_1, X_2, X_3)$.

26. ***(Rotationssymmetrie der 2-dimensioanlen Standardnormalverteilung)*** Es sei (X, Y) standard-normalverteilt auf $\mathbb{R}^2$. Weiter sei $R := \sqrt{X^2 + Y^2}$ der zu (X, Y) gehörige Radius und Θ der zu (X, Y) gehörige (im Bogenmaß gemessene) Winkel.

a) Begründen Sie auf den Spuren der Vorlesung 7a, dass
(i) R^2 exponentialverteilt ist zum Parameter $1/2$,
(ii) Θ uniform verteilt ist auf $[0, 2\pi)$.

b) In der Vorlesung haben wir rekapituliert, dass das Kreissegment G (in der (x, y)-Ebene) zwischen den Radien r_1 und r_2 und zwischen den (im Bogenmaß gemessenen) Winkeln θ_1 und θ_2 dem Rechteck $H = [r_1, r_2] \times [\theta_1, \theta_2]$ im "Radius-Winkel-Streifen" $\mathbb{R}_+ \times [0, 2\pi)$ entspricht (in dem Sinn, dass $(x, y) \in G \iff (r, \theta) \in H$ für $x = r \cos \theta$, $y = r \sin \theta$, vgl. Folie 19 aus VL 7a).

Weisen Sie damit die (schon aus der Rotationssymmetrie recht anschauliche einzusehende) Unabhängigkeit der beiden Zufallsvariablen R und Θ auch formal nach.

27. S *(Dreieck und Glocke)* X_1 und X_2 seien unabhängig und uniform auf $[0, 1]$ verteilt. Wir setzen $Y := X_1 + X_2$.

a) Nun ist ja 2 noch recht weit entfernt von ∞, trotzdem (oder gerade deshalb) wollen wir hier Y mit seiner Normalapproximation im Sinn des Zentralen Grenzwertsatzes vergleichen. Es sei dazu N eine normalverteilte Zufallsvariable mit $\mu_N = \mu_Y$ und $\sigma_N = \sigma_Y$. Berechnen Sie $\mathbf{P}(|Y - \mu_Y| > 2\sigma_Y)$ und vergleichen Sie das Ergebnis mit der Zahl $\mathbf{P}(|N - \mu_N| > 2\sigma_N)$. *(Hinweis: Weil* (X_1, X_2) *so verteilt ist wie* $(X_1, 1-X_2)$, *gilt* $\mathbf{P}(|X_1+X_2-1| > 2\sigma_Y) = \mathbf{P}(|X_1 - X_2| > 2\sigma_Y)$*; hier hilft dann Aufgabe 23 b) weiter.)*

b) (als Kür) Skizzieren Sie die Verteilungs- und die Dichtefunktion von Y.

28. ***(Viele Versuche, kleine Erfolgswahrscheinlichkeit)*** 20000 Punkte werden unabhängig und uniform verteilt ins Quadrat $[0, 100] \times [0, 100]$ gesetzt. Berechnen Sie mit der Poisson-Näherung die Wahrscheinlichkeit, dass die Kreisscheibe mit Radius 1 um den Mittelpunkt des Quadrats

i) leer ausgeht

ii) genau 4 Punkte abbekommt.

29. S a) Z sei standard-normalverteilt. Bestimmen Sie $c > 0$ so, dass für das Intervall $I = [-c, c]$ gilt: $\mathbf{P}(Z \in I) = 0.80$. (Hinweis: Der R-Befehl `qnorm(p)` liefert $\Phi^{-1}(p)$. Mehr Informationen darüber bekommen Sie, wenn Sie `?qnorm` in die R-Konsole eingeben.)

b) Y sei $N(0, \sigma^2)$-verteilt. Bestimmen Sie $c > 0$ so, dass für das Intervall $I = [-c, c]$ gilt: $\mathbf{P}(Y \in I) = 0.80$.

c) Wie groß muss der Stichprobenumfang mindestens sein, damit das Intervall mit den Grenzen $M \pm 2$ den Populationsmittelwert mit Wahrscheinlichkeit

(i) 0.99

(ii) 0.9

(iii) 0.8

überdeckt? Dabei sei M der (zufällige) Stichprobenmittwelwert, und es sei bekannt, dass die Populationsvarianz den Wert 100 beträgt. Rechnen Sie mit der Normalapproximation.

30. Wir betrachten die Situation von Aufgabe 19. Es sei μ der in 19a) berechnete Populationsmittelwert, σ^2 die in 19a) berechnete Populationsvarianz, und $X_1, \dots, X_{50}$ die zufälligen Werte, die bei einem 50-maligen Ziehen ohne Zurücklegen entstehen (vgl. 19b)).
a) Warum sind die X_i nicht unabhängig?
b) Wir setzen $M := \frac{1}{50}(X_1 + \cdots + X_{50})$. Es sei Ihnen verraten, dass trotz der fehlenden Unabhängigkeit der X_i auch hier die asymptotische Normalität greift, vgl. dazu den Ausblick auf der vorletzten Folie von VL 7b). Deshalb dürfen Sie im Rest der Aufgabe mit der Normalapproximation rechnen, Sie sollten dabei aber die Standardabweichung von M verwenden, die wir in Aufgabe 19 berechnet haben (jetzt mit $n = 50$ anstelle des dortigen $n = 10$).
(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 50 ziehen lassen. Verwenden Sie dazu das unter dem Link A30.R von Benjamin Straub 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 S a) Z_1, Z_2 und Z_3 seien unabhängig und $N(0,1)$-verteilt. Es sei $X := 2Z_1 + 3Z_2$ und $Y := 6Z_2 + Z_3$. Berechnen Sie diejenigen Zahlen β_0 und β_1, für die der erwartete quadratische Abstand zwischen den beiden Zufallsvariablen $\beta_0 + \beta_1 X$ und Y minimal wird.

b) X und Y seien wie in a). Berechnen Sie die Regressionsgerade für $Y + 3$ auf der Basis von $X + 2$.

32. a) (X, Y) sei uniform verteilt auf der dreielementigen Menge $\{(0,0),(1,0),(3,2)\}(\subset \mathbb{R}^2)$. Berechnen Sie die Regressionsgerade von Y auf der Basis von X.

b) Es sei $x_1 = 0$, $x_2 = 1$, $x_3 = 3$ sowie $y_1 = y_2 = 0$, $y_3 = 2$. Für welche Zahlen β_0 und β_1 wird $\sum_{i=1}^{3} (y_i - \beta_0 - \beta_1 x_i)^2$ minimal?

33. S. Das zufällige Paar (X_1, X_2) mit Werten in $\{b, c, d\} \times \{1, 2, 3\}$ komme durch ein zweistufiges Experiment zustande, wobei $\mathbf{P}(X_1 = b) = 2\mathbf{P}(X_1 = c) = 2\mathbf{P}(X_1 = d)$ gelte und die Übergangswahrscheinlichkeiten $P(a_1, .)$, $a_1 \in \{b, c, d\}$, durch die rechts angegebene Matrix bestimmt sind.

	1	2	3
b	0	0.6	0.4
c	0.3	0.2	0.5
d	0.6	0.3	0.1

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

(ii) Berechnen Sie den bedingten Erwartungswert von X_2 gegeben $X_1 = c$.

(iii) Finden Sie die *(im Sinn des erwarteten quadratischen Abstandes) beste Prognose von* X_2 auf der Basis von X_1, d.h. diejenige Zufallsvariable der Form $h(X_1)$, für die $\mathbf{E}[(X_2 - h(X_1))^2]$ minimal (über alle möglichen Funktionen h) wird.

(iv) Finden Sie Übergangswahrscheinlichkeiten $Q(a_2, .)$, $a_2 \in \{1, 2, 3\}$ so, dass das zufällige Paar (X_2, X_1) als zweistufiges Zufallsexperiment (jetzt mit X_2 als erster Stufe) entsteht.

34. 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.

35 S X sei Exp(2)-verteilt, und gegeben $X = a$ sei Y normalverteilt mit Erwartungswert a und Standardabweichung $\sqrt{a}$. Berechnen Sie **Var**$(X + Y)$ über die Zelegung der Varianz nach der ersten Stufe. (Sie dürfen dabei verwenden, dass die in der Vorlesung hergeleitete Formel auch für kontinuierlich verteilte Zufallsvariable gilt.)

36. Y sei uniform verteilt auf $S := \{1, 2, \ldots 5\} \cup \{21, 22, \ldots 30\}$. Berechnen Sie $\mathbf{Var}[Y]$, indem Sie Y als zweite Stufe in einem zweistufigen Zufallsexperiment auffassen, dessen erste Stufe eine Bernoulli$(1/3)$-Zufallsvariable X ist. Was sind hier

(i) die Varianzen innerhalb der beiden Gruppen

(ii) die Varianz zwischen den Gruppen?

37. a) X und Y seien unabhängig und Pois(α) bzw. Pois(β)-verteilt. Begründen Sie, dass $X+Y$ Pois($\alpha+\beta$) verteilt ist. *Hinweis: Für alle $k \in \mathbb{N}_0$ ist*

$$\sum_{i+j=k} \frac{\alpha^i}{i!} e^{-\alpha} \frac{\beta^j}{j!} e^{-\beta} = \frac{(\alpha+\beta)^k}{k!} e^{-(\alpha+\beta)}$$ *- warum gilt das?*

b) Berechnen Sie die Gewichte der bedingten Verteilung von X gegeben $X+Y=n$, $n \in \mathbb{N}$.

c) Berechnen Sie den Erwartungswert und die Varianz von $Z = \sum_{i=1}^{Y} X_i$. Dabei sind $X_1, X_2, \ldots$ unabhängige Kopien von X, die auch von Y unabhängig sind.

38. Im Fachbereich A einer Universität war die Durchfallquote bei der Aufnahmsprüfung 0.4 bei den Männern und 0.3 bei den Frauen, im Fachbereich B war sie 0.8 bei den Männern und 0.7 bei den Frauen. (In beiden Fachbereichen haben Frauen damit jeweils besser abgeschnitten als Männer.) Paradoxerweise war die Durchfallsquote bei den Männern insgesamt (mit den beiden Fachbereichen zusammengenommen) niedriger, nämlich 0.5, im Gegensatz zu 0.6 bei den Frauen.
a) Wie kann das sein? Berechnen Sie aus den Angaben, ein wie großer Prozentsatz der Männer (Frauen) im Fachbereich B angetreten ist.¶

b) Wie groß war der Anteil der Frauen an den “Erfolgreichen”, wenn
(i) insgesamt gleich viele Männer und Frauen angetreten sind,
(ii) doppelt so viele Frauen wie Männen angetreten sind?

c) Ein wie großer Anteil der “Erfolgreichen” entfiel auf den Fachbereich A, wenn
(i) in beiden Fachbereichen gleich viele Personen angetreten sind,
(ii) im Fachbereich A doppelt so viele Personen angetreten sind wie im Fachbereich B ?

¶Die Zahlen hier sind erfunden. Eine ähnliche Geschichte hat sich allerdings tatsächlich einmal an der Universität Berkeley zugetragen und führte sogar zu einer Diskriminierungsklage, siehe https://de.wikipedia.org/wiki/Simpson-Paradoxon

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

a) Bestimmen Sie $\mathbf{P}(U_1 < U_0, U_2 < U_0)$ und $\mathbf{P}(U_1 < U_0, U_2 < U_0, U_3 \geq U_0)$. (*Tipp: Arbeiten Sie mit rein zufälligen Permutationen, siehe Folie 14 in VL 2a.*)

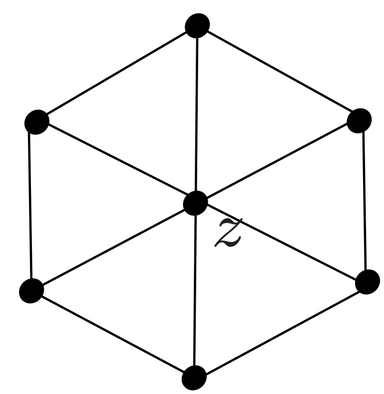
b) Wir definieren $Z_1 := I_{\{U_1 < U_0\}}, Z_2 := I_{\{U_2 < U_0\}}, Z_3 := I_{\{U_3 < U_0\}}$.

(i) Berechnen Sie $\mathbf{P}(Z_2 = 1 | Z_1 = 1)$ und $\mathbf{P}(Z_3 = 1 | Z_1 = 1, Z_2 = 0)$.

(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.

40. S a) $(X_n)_{n=0,1,2,\ldots}$ sei eine (p,q)-Irrfahrt auf $\mathbb{Z}$ mit $p = 3/4$ (d.h. man addiert zur aktuellen Position in jedem Schritt unabhängig $+1$ mit W'keit p und -1 mit W'keit $q = 1 - p$) und Start in 0. Wie groß muss n mindestens sein, damit X_n mit Wahrscheinlichkeit 0.975 größer als 100 ist? Verwenden Sie die Normalapproximation.

b) **40. S** b) $(Y_n)_{n=0,1,2,\ldots}$ sei eine gewöhnliche Irrfahrt auf dem skizzierten Graphen (d.h. man geht in jedem Schritt unabhängig zu einem rein zufällig gewählten Nachbarn der aktuellen Position). Der Startpunkt sei $Y_0 = z$. Finden Sie die Verteilung von Y_2.



41. Wir betrachten eine gewöhnliche Irrfahrt auf dem skizzierten Graphen. Berechnen Sie

i) die Wahrscheinlichkeit, bei Start in a nach drei Schritten in g zu sein,

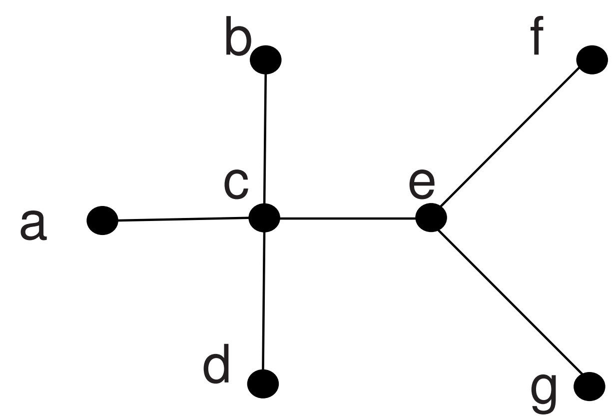
ii) die Wahrscheinlichkeit, bei Start in e nach drei Schritten in g zu sein,

iii) die erwartete Anzahl von Schritten bei Start in e bis zum Treffen von g,

iv) die Gleichgewichtsverteilung π auf $S := \{a, b, c, d, e, f, g\}$,

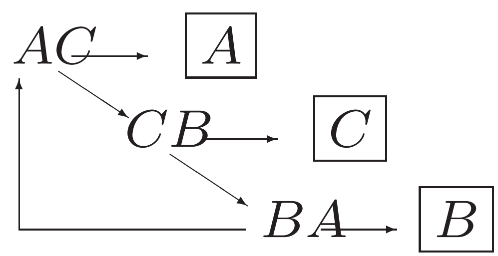
v) $\mathbf{P}_\pi(X_3 = e)$,

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



42. Drei Spieler A, B, C spielen ein Turnier. Alle Spieler sind gleich stark. In jeder Partie nehmen zwei Spieler teil. A und B beginnen. Der aussetzende Spieler spielt gegen den Gewinner der vorherigen Partie. Das Turnier endet, wenn ein Spieler zwei Partien hintereinander gewinnt. Berechnen Sie die Chancen von C, A und B, das Turnier zu gewinnen.

Hinweis : Fassen Sie den Fall ins Auge, daß A die erste Partie gewinnt, und betrachten Sie den Graphen



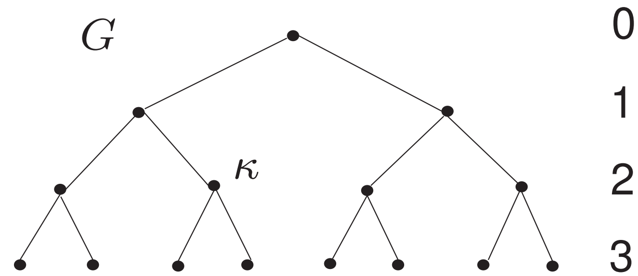
43. S Berechnen Sie den Erwartungswert der Anzahl der Würfe, bis in einer fairen Münzwurffolge erstmals das Muster

a) `010`

b) `111`

vollendet ist. (Zur Erklärung: Bei der Realisierung `110010` hat es 6 Würfe gebraucht, bis erstmals das Muster `010` vollendet ist.)

44. S Im abgebildeten Graphen G gibt es 2^i Knoten in der Tiefe i, $i = 0, 1, 2, 3$, darunter eine Wurzel (in Tiefe 0) und 8 Blätter (in Tiefe 3). Wir betrachten die gewöhnliche Irrfahrt auf G.



a) Finden Sie die Gewichte der einzelnen Knoten in der Gleichgewichtsverteilung.

b) Mit welcher Wahrscheinlichkeit macht man ausgehend von Tiefe 2 den nächsten Schritt nach Tiefe 3?

c) Mit welcher Wahrscheinlichkeit trifft man ausgehend vom Knoten κ (siehe Bild) die Wurzel eher als die Menge der Blätter. Anders gefragt: Mit welcher Wahrscheinlichkeit kommt man bei Start in κ in die Tiefe 0, bevor man in die Tiefe 3 kommt?

d) Was ist die erwartete Anzahl der Schritte, die man bei Start in κ macht, bis man die Wurzel trifft?

45. S In einer Population gibt es Befürworter und Gegner eine bestimmten politischen These. Im Kanton A wurde eine Stichprobe des Umfangs 100 befragt, im Kanton B eine des Umfangs 200. Die Anteile der Befürworter in den beiden Stichproben waren 40% im Kanton A und 41% im Kanton B.

(i) Geben Sie ein 95% Konfidenzintervall für die Differenz der Anteile der Befürworter in den beiden Kantonen an.

(ii) Zu welchem p-Wert lässt sich die Hypothese der Gleichheit der beiden Populationsmittelwerte ablehnen? (Hinweis: Der R-Befehl `pnorm(x)` liefert $\Phi(x)$.)

(iii) Um welchen gemeinsamen Faktor müssen - bei unveränderten Anteilen in den Stichproben - die Stichprobenumfänge größer sein, damit die Hypothese der Gleichheit der beiden Populationsmittelwerte mit einem p-Wert von 0.05 abgeleht werden könnte?

46. Aus zwei Populationen wurden Stichproben des Umfangs 30 bzw. 50 entnommen und die Werte eines reellen Merkmals x_i und y_j gemessen. Die Stichprobenmittelwerte und-standardabweichungen waren $m_x = 10$, $m_y = 14$, $s_x = 5$, $s_y = 8$.

(i) Bestimmen Sie ein 95% Konfidenzintervall für die Differenz der beiden Populationsmittelwerte.

(ii) Zu welchem p-Wert lässt sich die Hypothese der Gleichheit der beiden Populationsmittelwerte ablehnen?

47. S a) (i) Wie wahrscheinlich ist es, bei einem fairen Münzwurf der Länge 18 weniger als 4 oder mehr als 15 Erfolge zu haben?

(ii) $X_1, \ldots, X_{18}$ seien unabhängig und identisch verteilt auf $\mathbb{R}$, mit $\mathbf{P}(X_1 = a) = 0$ für alle $a \in \mathbb{R}$. Berechnen Sie die Wahrscheinlichkeit, mit der das Intervall $[X_{(4)}, X_{(15)}]$ den Median der Verteilung von X_1 überdeckt.

47. S b) 6 Objekte werden (ohne Mehrfachbelegungen) auf die Plätze $\{1, ..., 18\}$ gesetzt.

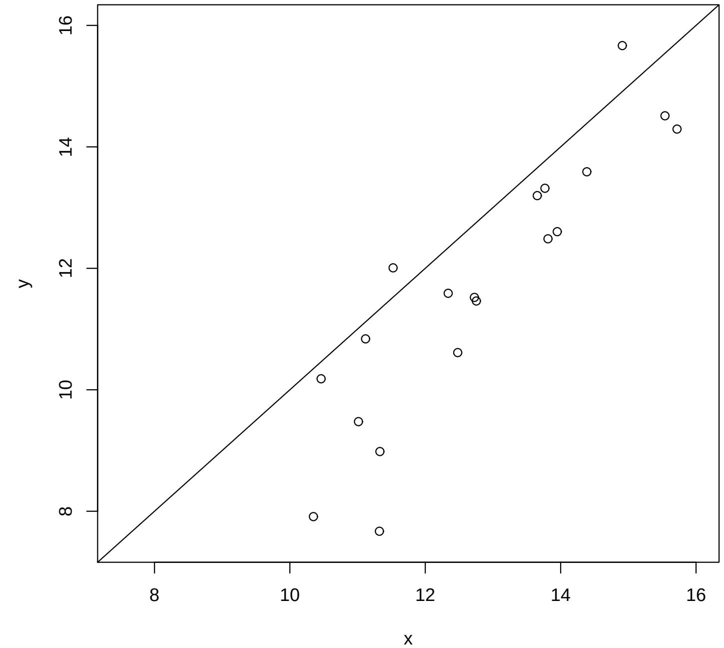
(i) Was ist der Erwartungswert der Anzahl der Objekte, die auf der Platzmenge $M := \{1, ..., 10\}$ landen?

(ii) Es ist nur eines der sechs Objekte auf M gelandet. Zu welchem p-Wert können Sie unter Verwendung von Fishers exaktem Test die Hypothese der reinen Zufälligkeit verwerfen?

47. S c) (i) Für wieviele dreielementige Mengen aus $\{1, 2, \dots, 30\}$ ist die Summe ihrer Elemente ≤ 10?

(ii) 3 Objekte wurden (ohne Mehrfachbelegungen) auf die Plätze $\{1, \dots, 30\}$ gesetzt, sie fielen auf die Plätze $7, 2$ und 1. Zu welchem p-Wert können Sie unter Verwendung des Wilcoxon-Rangsummentests die Hypothese der reinen Zufälligkeit (zugunsten einer “Tendenz an die Ränder”, d.h. zu den Plätzen mit den kleinen bzw zu den Plätzen mit den großen Rängen) verwerfen?

48. 20 Probanden wurden einer extensiven Bewegungstherapie unterzogen, verbunden mit langen Radtouren an mehreren Wochenenden. Gemessen wurde bei jedem der Probanden $i = 1, \ldots, 20$ ein bestimmter Blutfettwert, mit dem Ergebnis x_i vor und y_i nach der Therapie. Beobachtet wurden $\bar{x} = 12.97$, $\bar{y} = 11.92$, $s_x = 1.78$, $s_y = 2.40$, $s_{x-y} = 1.07$. Es ging darum, die systematische Komponente der Änderung des Wertes vor und nach der Therapie zu schätzen, und eine Aussage ber die Signifikanz der beobachteten Änderungen zu treffen.



Dazu wurde (in Anlehnung an das Vorgehen in Aufgabe 46) das Intervall

$$I := \left[\bar{x} - \bar{y} - 2\sqrt{s_x^2/20 + s_y^2/20}, \bar{x} - \bar{y} + 2\sqrt{s_x^2/20 + s_y^2/20}\right] \qquad \text{angegeben.}$$

Es stellte sich heraus, dass dieses Intervall die Zahl 0 enthielt, also glaubte man keinen signifikanten Effekt feststellen zu dürfen. Hätten Sie die Versuchsauswertung besser gekonnt?