Übungsaufgaben zur Vorlesung

“Stochastik für die Informatik”, WS 2017/18

Prof. Dr. Anton Wakolbinger

Goethe-Universität Frankfurt

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 viermal wiederholte rein zufällige Wahl eines Pixels aus G (anders gesagt: um ein Ziehen aus G mit Zurücklegen), beschrieben durch eine auf dem Wertebereich $\{1,\ldots,g\}^4$ uniform verteilte Zufallsvariable $X = (X_1, X_2, X_3, X_4)$.

a) Wieviele verschiedene mögliche Ausgänge von X gibt es, bei denen eines der X_i auf die Menge $\{1,\ldots,f\}$ und drei 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 (in der Situation von a) von den vier zufällig aus G gewählten Pixeln genau eines aus F gewählt wird?

c) Es sei M die zufällige Trefferquote von F . Illustrieren Sie für $p = 0.195$ und $n = 4$ 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. Es sei $n \in \mathbb{N}$. Wir nennen $a = (a_0, a_1, \ldots, a_n)$ einen *Standardpfad der Länge* n, falls $a_0 = 0$ und $a_i - a_{i-1} \in \{-1, +1\}$, $i = 1, \ldots, n$. (Wir interpretieren a_i als Position des Pfades zum Zeitpunkt i.) Es sei S die Menge der Standardpfade der Länge 20 und $S_1 := \{a \in S :$ es existiert ein i mit $a_i \geq 1\}$.

a) Begründen Sie die folgende Aussage:

$$\#\{a \in S_1 : a_{20} > 1\} = \#\{a \in S_1 : a_{20} < 1\}.$$

Betrachten Sie dazu den letzten Zeitpunkt $\ell = \ell(a)$*, zu dem der Pfad* $a \in S_1$ *die Position* 1 *besucht, und denken Sie sich Sie das Pfadstück* $(a_\ell, a_{\ell+1}, \ldots, a_{20})$ *an der Höhe* 1 *gespiegelt.*

b) Gilt folgende Aussage:

$$2\#\{a \in S_1 : a_{20} > 1\} = 2\#\{a \in S : a_{20} > 1\} = \#\{a \in S : a_{20} \neq 0\} \quad ?$$

(Hinweis: Nach einer geraden Anzahl von Schritten landet man bei einer geraden Zahl.)

c) Es sei $X = (X_0, X_1, \ldots, X_{20})$ ein rein zufälliger Standardpfad der Länge 20. Drücken Sie die Wahrscheinlichkeit des Ereignisses $\{X$ erreicht jemals einen Wert $\geq 1\}$ durch die Wahrscheinlichkeit des Ereignisses $\{X_{20} = 0\}$ aus.

4 S. Es sei $n \in \mathbb{N}$. Wir nennen $a = (a_1, \ldots, a_n)$ eine 01-*Folge der Länge* n, falls $a_i \in \{0, 1\}$, $i = 1, \ldots, n$. Wir sagen, dass a das Muster 0110 enthält, falls

$$a_i a_{i+1} a_{i+2} a_{i+3} = 0110$$

für ein $i \in \{1, \ldots, n-3\}$.

Für $k \in \mathbb{N}$ sei nun $X = (X_1, \ldots, X_{4k})$ eine rein zufällige 01-Folge der Länge $4k$. Warum gilt folgende Aussage:

$$\mathbf{P}(X \text{ enthält das Muster } 0110) \geq 1 - (1 - 2^{-4})^k?$$

5. n Individuen werden rein zufällig auf $r = 10^6$ Plätze gesetzt.

a) Berechnen Sie die Wahrscheinlichkeit $w_r(n)$ für Kollisionsfreiheit für

(i) $n = 100$ (ii) $n = 1000$

(iii) $n = 10000$.

Sie können dabei die folgende kleine R-Schleife verwenden:

```
r<-10^6
n<-100
w<-function(r,n){
  v<-1
  if (n>1){
    for (i in 1:(n-1)){
      v<-v*(1-i/r)
    }
  }
  return(v)
}
w(r,n)
```

b) Berechnen Sie in a) jeweils den absoluten und den relativen Fehler der Stirling- und der Stirling+Taylor Näherung. (Für eine Näherung u des wahren Wertes w ist dabei $|u - w|$ der absolute und $|(u - w)/w|$ der relative Fehler.)

6. S Es sei S die Menge der 01-Folgen der Länge r. Für $i = 1, \dots, r$ setzen wir $p_i := \frac{i-1}{r}$ und $q_i := 1 - p_i$. Wir betrachten eine Zufallsvariable $Z = (Z_1, \dots, Z_r)$ mit Wertebereich S und

$$\mathbf{P}(Z = (a_1, \dots, a_r)) = \prod_{i \le r : a_i = 1} p_i \prod_{i \le r : a_i = 0} q_i, \quad a = (a_1, \dots, a_r) \in S.$$

a) Zeigen Sie rekursiv für $n \le r$, beginnend mit $n = r - 1$:

$$\mathbf{P}(Z_1 = a_1, \dots, Z_n = a_n) = \prod_{i \le n : a_i = 1} p_i \prod_{i \le n : a_i = 0} q_i.$$

b) Für $a \in S$ sei

$$t(a) := \min\{i \in \{1, \dots, n\} : a_i = 1\}$$

der *Zeitpunkt der ersten Eins* (oder “Zeitpunkt des ersten Erfolges”), mit $t((0, \dots, 0)) := \infty$. Bestimmen Sie $\mathbf{P}(t(Z) > n)$ für $n \le r$.

7. Für gerades $n \in \mathbb{N}$ sei $X = (X_0, X_1, \ldots, X_n)$ ein rein zufälliger Standardpfad der Länge n (vgl. dazu Aufgabe 3 auf Blatt 1). Berechnen Sie die Wahrscheinlichkeit des Ereignisses $\{X$ erreicht zu keinem Zeitpunkt den Wert $1\}$

a) exakt b) näherungsweise mit der Stirlingformel.

8. S 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 $1, \dots, 7$ schreibt man als $(1\,5)(2)(3\,7\,6\,4)$.
Wir beschreiben jetzt ein rekursives Verfahren zur Erzeugung einer zufälligen Permutation von $1, \dots, n+1$ aus einer Permutation von $1, \dots, n$, ausgehend von deren Zyklendarstellung: *Das Element $n+1$ wird jeweils mit W'keit $\frac{1}{n+1}$ auf einen der n Plätze rechts neben $1, 2, \dots, n$ (innerhalb des jeweiligen Zyklus) gesetzt. Ebenfalls mit W'keit $\frac{1}{n+1}$ wird das Element $n+1$ 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 $1, 2, \dots, n$ genau ein Pfad (im Sinn von a)) führt.

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

d) Mit welcher Wahrscheinlichkeit liegen in einer rein zufälligen Permutation von $\{1, \dots, 1000\}$
(i) 1, 2, 3, 4 (ii) 700, 800, 900 und 1000
im selben Zyklus?

9. 60 Karten, von denen 10 die Farbe blau, 20 die Farbe rot und 30 die Farbe grün haben, werden perfekt gemischt und dann eine nach der anderen aufgeschlagen.

(i) Wie wahrscheinlich ist es, dass die erste aufgeschlagene Karte blau ist?

(ii) Wie wahrscheinlich ist es, dass die achte aufgeschlagene Karte blau ist?

(iii) Wie wahrscheinlich ist es, dass die achte und die neunte aufgeschlagene Karte nicht dieselbe Farbe haben?

10. S Es sei $S_{10,3}$ die Menge der Besetzungen[†] von 3 Plätzen mit 10 Objekten.

a) Geben Sie die Anzahl der Elemente von $S_{10,3}$ an, indem Sie
(i) die in der Vorlesung diskutierte Bijektion zwischen $S_{10,3}$ und der Menge aller 01-Folgen der Länge 12 mit genau zwei Nullen verwenden
(ii) das entsprechende de Finetti-Dreieck entlang seiner Zeilen abzählen.

b) Warum hat $\{b \in S_{10,3} : b_j \geq 1 \text{ für alle } j\}$ genau so viele Elemente wie $S_{7,3}$? Veranschaulichen Sie diese Tatsache auch am entsprechenden de-Finetti-Dreieck.

c) 20 Objekte werden gemäß einer uniform verteilten Besetzung auf 5 Plätze gesetzt. Wie wahrscheinlich ist es, dass dabei kein Platz leer bleibt?

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

11. Wir betrachten die in Aufgabe 1 beschriebenen Situation. Zur Erinnerung: Der Pixelanteil (“Flächenanteil”) von A im Quadrat war $p = 0.195$. Die Anzahl der rein zufällig (nacheinander) ins Quadrat gesetzten Pixel (“Punkte”) sei $n = 100$.

a) Mit welcher Wahrscheinlichkeit fallen die ersten 20 Punkte in die Menge A und die restlichen nicht?

b) Mit welcher Wahrscheinlichkeit ist die Anzahl der Punkte, die in A fallen, gleich 20? Vergleichen Sie Ihr Ergebnis mit dem Monte-Carlo-Ergebnis bei 1000 Wiederholungen (der Wahl von jeweils 100 Punkten).

c) Mit welcher Wahrscheinlichkeit fallen
(i) weniger als 10 (ii) mehr als 28 Punkte
in die Menge A? (Hinweis: Der R-Befehl sum(dbinom(k1:k2, n, p)) - mit jeweils passender Wahl der Summationsgrenzen k_1 und k_2 - ist hilfreich. Mehr Information bekommen Sie, wenn Sie z.B. den Befehl *?dbinom* in die R-Konsole eingeben.)

d) Mit welcher Wahrscheinlichkeit kommt der erste Treffer von A erst nach dem 15ten, aber noch vor dem 30ten der 100 “Würfe”?

12. S a) In der Situation von Aufgabe 9 betrachten wir jetzt ein rein zufälliges Ziehen mit Zurücklegen.

i) Wie wahrscheinlich ist es, dass von 10 gezogenen Karten 4 blau, 3 rot und 3 grün sind?

ii) Wie wahrscheinlich ist es, dass von 10 gezogenen Karten 4 blau und die anderen 6 rot oder grün sind?

b) $X = (X_1, X_2, \ldots, X_r)$ sei multinomialverteilt mit Parameter $(n, p_1, \ldots, p_r)$.

Wie ist $X = (X_1 + X_2, X_3, \ldots, X_r)$ verteilt?

13. $\{A, B, C\}$ sei eine Partition (eine “Zerlegung”) von $\{1, \ldots, 20\}$, bestehend aus drei disjunkten Teilmengen A, B, C von $\{1, \ldots, 20\}$ mit 5, 6 und 9 Elementen, und $(X_1, \ldots, X_{20})$ sei eine rein zufällige Permutation von $1, \ldots, 20$.

(i) Mit welcher Wahrscheinlichkeit fallen X_3 und X_4 in verschiedene Partitionselemente, d.h. nicht gemeinsam in eine der Mengen A, B oder C?

(ii) 20 Städte, von denen 5 im Land A, 6 im Land B und 9 im Land C liegen, werden in rein zufälliger Reihenfolge besucht. Jeder Grenzübergang kostet k Euro. Berechnen Sie den Erwartungswert der gesamten Grenzübertrittskosten.

14. 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 den Erwartungswert des Quadrats einer Binom(n, p)-verteilten Zufallsvariablen X, indem Sie X als Summe von Zählvariablen schreiben.

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

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

15. S Aus einer Population bestehend aus 40 Hessen und 60 Bayern wurde eine Stichprobe vom Umfang 20 (d.h. ein 20-elementige Teilmenge der Population) herausgegriffen. In dieser befanden sich 5 Hessen und 15 Bayern.

a) Was ist der Erwartungswert der Anzahl der Hessen in einer *rein zufälligen* Stichprobe vom Umfang 20?

b) Wie wahrscheinlich ist in einer rein zufällig gezogenen Stichprobe eine Anzahl von Hessen, die vom (in a) berechneten) Erwartungswert mindestens so weit abweicht wie die beobachtete Anzahl 5?

Hinweis: Ganz ähnlich wie in Aufgabe 11 ist hier der R-Befehl sum(dhyper(k1:k2,r,b, hilfreich. Finden Sie heraus, was diese Gewichte mit der Aufgabenstellung zu tun haben, und was die Summationsgrenzen sind.

16. Für natürliche Zahlen r und n sei $(X_1, \ldots, X_n)$ uniform verteilt auf $\{1, \ldots, r\}^n$.

a) Wie wahrscheinlich ist es, dass (für $n \geq 5$) die beiden Komponenten X_3 und X_5 gleich ausfallen?

b) Berechnen Sie die erwartete Anzahl der Kollisionen, d.h. den Erwartungswert der Anzahl der Paare (i, j) mit $1 \leq i < j \leq n$, für die X_i und X_j gleich ausfallen.

c) Es sei λ_n der in b) berechnete Wert. Vergleichen Sie $e^{-\lambda_n}$ mit der ersten in Vorlesung 1b hergeleiteten Näherung für die “Wahrscheinlichkeit von null Kollisionen”.

17. a) X_n sei Binom(n,p)-verteilt. Beweisen Sie mit der Ungleichung von Markov, angewandt auf ein passend gewähltes Y: Für jedes $\delta > 0$ ist

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

b) Folgern Sie aus a) das *Gesetz der großen Zahlen von Jakob Bernoulli*:
Sei $Z_1, Z_2, \ldots$ ein fortgesetzter p-Münzwurf und $\varepsilon > 0$. Dann ist

$$\mathbf{P}(|\frac{1}{n}(Z_1 + \cdots + Z_n) - p| \geq \varepsilon), \quad n = 1, 2, \ldots$$

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 $n \to \infty$.

18. S. $X^{(n)}$ sei eine rein zufällige Permutation von $1, \ldots, n$.

a) Berechnen Sie die erwartete Anzahl der Fixpunkte von $X^{(n)}$.

b) Für $0 \leq k \leq n$ sei $f_k^{(n)}$ die Anzahl der Permutationen von $1, , n$ mit genau k Fixpunkten. Finden Sie eine Beziehung zwischen $f_k^{(n)}$ und $f_0^{(n-k)}$.

c) Es sei $p_k^{(n)}$ die Wahrscheinlichkeit des Ereignisses, dass $X^{(n)}$ genau k Fixpunkte hat. Zeigen Sie, dass für jedes $k \in \mathbb{N}_0$ die Folge $p_k^{(n)}$ für $n \to \infty$ gegen das Poissongewicht von k zum Parameter 1 konvergiert.

19. S. Für $r \in \mathbb{N}$, sei $Z_1, Z_2, \dots$ ein fortgesetztes r-Würfeln mit den gleich wahrscheinlichen Ausgängen $1, \dots, r$.

a) Was ist die erwartete Anzahl der Würfe, bis erstmals alle r Ausgänge erzielt worden sind?

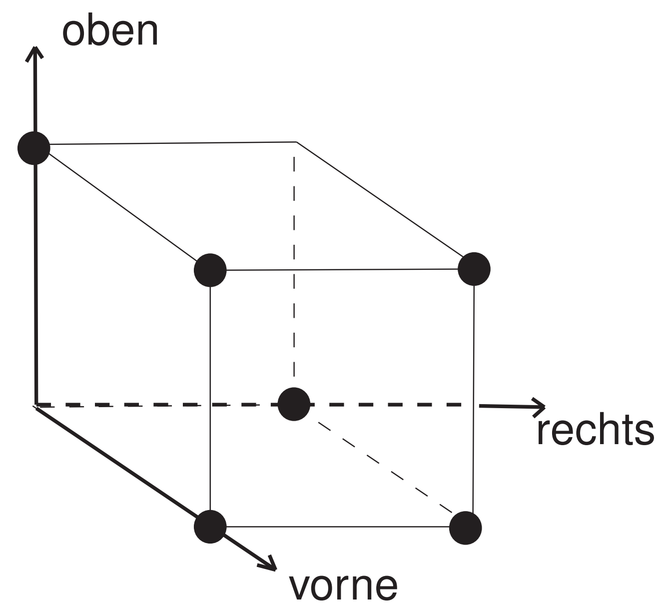
b) Es sei $r = 1000$; man würfelt in jeder Sekunde einmal. Was ist die Wahrscheinlichkeit, dass man bis zum Ausgang 1 länger als eine Stunde warten muss? Verwenden Sie die Exponentialapproximation.

20. Es sei Q ein Quadrat mit 100×100 Pixeln und T ein Teilquadrat von Q mit 10×10 Pixeln. $(X_1, X_2, \ldots,)$ sei eine fortgesetzte rein zufällige Wahl von Pixeln aus Q. (Die mehrfache Wahl eines Pixels ist damit nicht ausgeschlossen.) Wie wahrscheinlich ist es, dass unter den ersten 200 gewählten Pixeln genau 5 Treffer von T sind? Rechnen Sie

(i) "exakt" (z.B. mit einem entsprechenden R-Befehl) und

(ii) mit der Poissonapproximation.

21. ***Paarweise Unabhängigkeit von Ereignissen.*** X sei 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 $E_1 := \{X \text{ landet vorne}\}$, $E_2 := \{X \text{ landet rechts}\}$, $E_3 := \{X \text{ landet oben}\}$.



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

b) Sind die drei Ereignisse E_1, E_2, E_3 unabhängig? Argumentieren Sie hier sowohl rechnerisch als auch anschaulich.

22. S. ***Paare und Summen von Zählvariablen.*** In der Situation von Aufgabe 13 sei Z_i die Indikatorvariable des Ereignisses, dass X_i und X_{i+1} in verschiedene Partionselemente fallen ($i = 1, \dots, 19$).

a) Bestimmen Sie, gerundet auf 4 Nachkommastellen, die Matrix der gemeinsamen Verteilungsgewichte von Z_3 und Z_4.

b) Welche der drei folgenden Aussagen trifft zu: Z_3 und Z_4 sind

(i) unkorreliert (ii) positiv korreliert (iii) negativ korreliert ?
c) Berechnen Sie die Varianz von $Z_1 + \cdots + Z_{19}$.

d) Zusatzfrage für Extra-Punkte: Von n Kugeln sind n_1 schwarz, n_2 grün und n_3 weiß gefärbt (mit $n = n_1 + n_2 + n_3$). Die Kugeln werden rein zufällig ohne Zurücklegen gezogen. Wir betrachten das Ereignis $E_1 :=$ "die Farbe der ersten und der zweiten gezogenen Kugel sind gleich" und das Ereignis $E_2 :=$ "die Farbe der zweiten und der dritten gezogenen Kugel sind gleich". Finden Sie ein Beispiel für n, n_1, n_2 und n_3, in dem E_1 und E_2 negativ korreliert sind, und ein weiteres, in dem E_1 und E_2 positiv korreliert sind. *(Eine Frage an den Hausverstand: Wenn es deutlich viel mehr schwarze Kugeln gibt als grüne und weiße und Ihnen jemand verrät, dass E_1 eingetreten ist: würden Sie dann vielleicht eher auf das Eintreten von E_2 wetten als ohne die besagte Information, und wenn ja, aus welchem intuitivem Grund?)*

23. S. ***Varianz des Stichprobenmittels.*** a) Es sei n eine natürliche Zahl, und Y eine auf $\{1, \ldots, n\}$ uniform verteilte Zufallsvariable. Berechnen Sie
(i) den Erwartungswert μ,
(ii) die Varianz σ^2,
(iii) die Standardabweichung von Y.

(Die Formel $\sum_{a=1}^{n} a^2 = \frac{1}{6}n(n+1)(2n+1)$ *ist dabei hilfreich.)*

b) Wir ziehen rein zufällig und ohne Zurücklegen aus der Menge $\{1, \ldots, 20\}$ und bezeichnen mit X_i das beim Zug Nr. i gezogene Element.

α) Warum hängt (für $1 \leq i \neq j \leq 20$) die Kovarianz $\mathbf{Cov}[X_i, X_j]$ nicht von i und j ab?

β) Berechnen Sie die Kovarianz von X_1 und X_2 aus der Identität
$0 = \mathbf{Var}(X_1 + \ldots + X_{20})$.

γ) Berechnen Sie die Varianz des Stichprobenmittels $\frac{1}{10}(X_1 + \cdots + X_{10})$.

δ) Vergleichen Sie das Ergebnis aus γ) mit der Varianz des entsprechenden Stichprobenmittels beim Ziehen mit Zurücklegen.

24. *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) $n = 1600$?

25. X sei uniform verteilt auf $[-1, 1]$, $Y := X^2$.

Sind X und Y

(i) unkorreliert

(ii) unabhängig ?

Hinweis zu (ii): Betrachten Sie die Ereignisse

$\{|X| < 1/2\}$ und $\{Y > 1/4\}$.

26. Es sei U eine Unif($[0,1]$)-verteilte Zufallsvariable.

a) Finden Sie eine Abbildung
$h=(h_1,h_2):[0,1]\to\{0,1\}^2$ so, dass $X=h(U)$ so verteilt ist wie
(i) ein zweimaliger fairer Münzwurf
(ii) ein zweimaliger Münzwurf zum Parameter $p=1/3$.

b) Finden Sie eine Abbildung $h:[0,1]\to\mathbb{R}_+$ so, dass $X=h(U)$
(i) binomialverteilt ist zu den Parametern $n=2$ und $p=1/4$
(ii) exponentialverteilt ist zum Parameter $\lambda=1/4$.

27. S U sei uniform auf $[0,2]$ verteilt, und X sei Exp(3)-verteilt. Berechnen Sie

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

(iv) die Dichte

von

a) U^5 b) $4X-3$.

Hinweis zu (i) und (ii): Bei a) ist es geraten, mit der Transformationsformel für Erwartungswerte zu rechnen; bei b) helfen grundlegende Eigenschaften von Erwartungswert und Varianz.

28. S a) Wieviele Versuche muss man in einem Bernoulli-Experiment mit Erfolgswahrscheinlichkeit $p = 0.8$ mindestens machen, damit die Wahrscheinlichkeit, mindestens 30 Erfolge zu erzielen, nicht geringer ist als 0.975? Rechnen Sie mit der Normalapproximation der Binomialverteilung. (Eine Skizze ist hilfreich.)

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:

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

Vergleichen Sie in der Situation von a) die naive Approximation von $\mathbf{P}(X \geq 30)$ durch $\mathbf{P}(Y \geq 30)$ mit der genaueren Approximation von $\mathbf{P}(X \geq 30)$ mittels $(*)$.

c) Lösen Sie a) auch noch (ohne Normalapproximation) mittels des R-Befehls `dbinom`.

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

(i) die Verteilungsfunktion (ii) die Dichte

von

a) $\sqrt{2X}$ b) $\dfrac{1-U}{U}$.

30. a) *Der (infinitesimale) Flächeninhalt eines Kreisringes mit Radius* r *und Breite* dr *ist* $2\pi r\, dr$*.* Geben Sie hierfür eine anschauliche Begründung mittles einer Skizze.

b) $Z = (Z_1, Z_2)$ sei standard-normalverteilt auf $\mathbb{R}^2$, und $|Z| := \sqrt{Z_1^2 + Z_2^2}$. Verwenden Sie Teil a) dieser Aufgabe zur Bestimmung der Dichte von $|Z|$.

c) Vergleichen Sie das Ergebnis aus b) mit dem aus Aufgabe 29 a).

31. $U_1, U_2, \ldots$ seien unabhängig und uniform verteilt auf $[0,1]$.
Für $i = 1, 2, \ldots$ setzen wir
$X_i := U_i^5$, $i = 1, 2, \ldots,$ und
$Y := X_1 + \ldots + X_{100}$.

a) Bestimmen Sie unter Verwendung des Zentralen Grenzwertsatzes ein um $\mathbf{E}[Y]$ symmetrisches Intervall J, für das gilt:

$$\mathbf{P}(Y \in J) \approx 0.95.$$

(*Hinweis: Das Ergebnis aus Aufgabe 27 a) ist hilfreich.)*

b) Lösen Sie die zu a) analoge Aufgabe auch noch für
$M := \frac{1}{100}(X_1 + \ldots + X_{100})$ anstelle von Y.

32. S. Aus den Zahlen $1, \dots, 100$ werden 49 Zahlen ohne Zurücklegen gezogen und der arithmetische Mittelwert M aus den gezogenen Zahlen (der "Stichprobenmittelwert") gebildet. Es sei Ihnen verraten, dass trotz der in den Zügen enthaltenen Abhängigkeiten die Zufallsvariable M annähernd normalverteilt ist. Bestimmen Sie mit dieser Information ein um den Erwartungswert von M symmetrisches Intervall J, für das gilt:

$$\mathbf{P}(M \in J) \approx 0.95.$$

(*Hinweis: Aufgabe 23 b γ) ist hilfreilch, mit entsprechend angepassten Zahlen.*)

33. S. a) Y sei normalverteilt mit Erwartungswert μ_Y und Varianz σ_Y^2. Wie ist b zu wählen, damit das zufällige Intervall $I := [Y - b, Y + b]$ die Zahl μ_Y mit Wahrscheinlichkeit ≈ 0.95 überdeckt? Begründen Sie Ihre Antwort, indem Sie das Ereignis $\{\mu_Y \in I\}$ passend umformen.

b) In einer Stichprobe des Umfangs 100 aus einer großen Population ergab sich der Stichprobenmittelwert $m = 7.8$ und die Stichprobenvarianz $s^2 = 4.0$. Bestimmen Sie daraus die Realisierung eines zufälligen Intervalls, welches den Populationsmittelwert mit Wahrscheinlichkeit ≈ 0.95 überdeckt.

34. Z_1 und Z_2 seien unkorreliert und jeweils mit Erwartungswert 0 und Varianz 1; es sei $X := 3 + 3Z_1 + 4Z_2$ und $Y := 5 + 5Z_1 - 3Z_2$. Berechnen Sie

(a) diejenige affin lineare Funktion $h = h(x)$, für die der erwartete quadratische Prognosefehler $\mathbf{E}[(Y - h(X))^2]$ unter allen affin linearen Funktionen minimal wird

(b) den mit der Lösung aus (a) resultierenden erwarteten quadratischen Prognosefehler.

35. Wir betrachten ein simples Modell für die Körpergrößen (X, Y) eines zufällig aus der Population gewählten Vater-Sohn-Paares: Z_0, Z_1, Z_2 seien unabhängig und standard-normalverteilt, $X := 170 + 10Z_0 + 10Z_1$, $Y := 170 + 10Z_0 + 10Z_2$. Berechnen Sie den Wert der besten affin linearen Prognose

a) von Y auf Grundlage der Beobachtung $X = 185$,

b) von X auf Grundlage der Beobachtung $Y = 180$.

36 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	0.4	0.2	0.4
c	0.3	0.2	0.5
d	0.5	0.4	0.1

(i) Sind X_1 und X_2 unabhängig?

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

(iii) 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.

37. 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.

38. S a) X sei Bernoulli($1/3$)-verteilt. Gegeben $\{X = 0\}$ sei Y Exp(3)-verteilt, und gegeben $\{X = 1\}$ sei $-Y$ Exp(5)-verteilt. Berechnen Sie $\mathbf{E}[Y]$ und $\mathbf{Var}[Y]$.

b) Y sei uniform verteilt auf $[0, 1] \cup [10, 20]$. Berechnen Sie $\mathbf{E}[Y]$ und $\mathbf{Var}[Y]$.

Verwenden Sie sowohl in a) als auch in b) den Satz von der Zerlegung der Varianz. Was war nochmal die Varianz einer auf dem Intervall $[0, 1]$ *uniform verteilten ZV'en, und wie verhält sich die Varianz unter Verschiebung und Streckung?*

39. S a) Y sei N(3,1)-verteilt und N sei N(2,1)-verteilt; Y und N seien unabhängig.
(i) Wie ist $Y + N$ verteilt?
(ii) Finden Sie die bedingte Dichte von Y gegeben $\{Y + N = 5\}$.
Hinweis: Für die gemeinsame Dichtefunktion gilt
$f_{Y,Y+N}(a,5) = f_{Y,N}(a,5-a) = \frac{1}{2\pi} e^{-\frac{(a-3)^2}{2}} e^{-\frac{(5-a-2)^2}{2}} = \cdots$. *Gefragt ist nach* $\frac{f_{Y,Y+N}(a,5)}{f_{Y+N}(5)}\, da.$
(iii) Bestimmen Sie den bedingten Erwartungswert und die bedingte Varianz von Y, gegeben $\{Y + N = 5\}$.
(iv) Bestimmen Sie die (im Sinn des erwarteten quadratischen Fehlers) beste *lineare* Prognose von Y auf der Basis von $X := Y + N$. Welchen Wert nimmt sie an, wenn X auf den Ausgang $b = 5$ fällt?

b) Finden Sie ein einfaches Beispiel von zwei (diskreten) Zufallsvariablen X, Y, für die die beste *lineare* Prognose von Y auf der Basis von X verschieden ist von der besten Prognose von Y auf der Basis von X.

(Hier und auch im folgenden Teil c) ist das Gütekriterium jeweils die Kleinheit des erwarteten quadratischen Fehlers.)

c) (für Zusatzpunkte) Finden Sie ein Beispiel von zwei nicht unabhängigen Zufallsvariablen X, Y mit einer gemeinsamen Dichte, für die die beste *lineare* Prognose von Y auf der Basis von X und die beste Prognose von Y auf der Basis von X übereinstimmen.

40. 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.
Wie kann das sein? Berechnen Sie aus den Angaben, ein wie großer Prozentsatz der Männer (Frauen) im Fachbereich B angetreten ist, und formulieren Sie mit Blick darauf eine auch für alle Ihre Freunde verständliche Erklärung des Paradoxons. ‡

‡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

41. (Eine Modifikation von Aufgabe 3.15 aus L. Dümbgen, Stochastik für Informatiker, Springer 2003). Jede der 5 Kanten im skizzierten Netwerk sei (unabhängig von den anderen) mit Wahrscheinlichkeit p intakt.

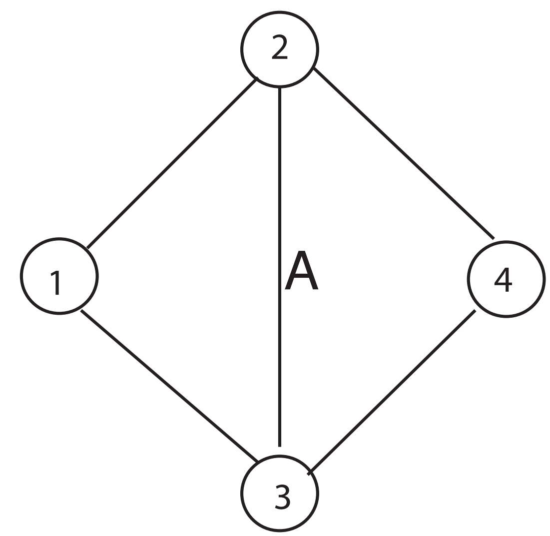
a) Mit welcher Wahrscheinlichkeit besteht eine Verbindung ("fließt Strom") vom Knoten 1 zum Knoten 4,

(i) gegeben die Kante A (zwischen Knoten 2 und Knoten 3) ist intakt

(ii) gegeben die Kante A ist defekt (d.h. nicht intakt)?

b) Mit welcher Wahrscheinlichkeit besteht eine Verbindung ("fließt Strom") zwischen Knoten 1 und Knoten 4?

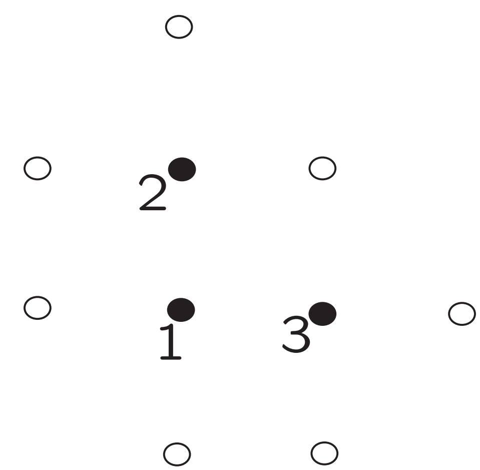
c) Es sei $p = 0.1$. Mit welcher Wahrscheinlichkeit ist die Kante A defekt, gegeben es fließt Strom zwischen Knoten 1 und Knoten 4?



42. Für $n \geq 0$ sei $S_n := \{a_0 a_1 \ldots a_n : a_i \in \{0, \ldots, n\}$ paarweise verschieden$\}$ und X_n eine Zufallsvariable mit Wertebereich S_n, wobei die Folge $(X_0, X_1, \ldots)$ so aufgebaut wird:
Gegeben $X_n = a_0 a_1 \ldots a_n$ entsteht X_{n+1} so, dass $n+1$ rein zufällig in einen der insgesamt $n+2$ möglichen Slots "links von a_0, zwischen a_0 und a_1, ..., rechts von a_n" platziert wird.
a) Zeigen Sie mit vollständiger Induktion: X_n ist uniform verteilt auf S_n.
b) Für $a = a_0 a_1 \ldots a_n \in S_n$ sei $L(a)$ die Anzahl der $j \in \{1, \ldots, n\}$, für die j links von 0 steht, und $R(a)$ die Anzahl der $j \in \{1, \ldots, n\}$, für die j rechts von 0 steht (z. B. ist $L(350142) = 2$).
Begründen Sie: Gegeben $\{L(X_1) = k_1, \ldots, L(X_n) = k_n\}$ ist die Wahrscheinlichkeit des Ereignisses $\{L(X_{n+1}) = k_n + 1\}$ gleich $\frac{k_n+1}{n+2}$.
c) Folgern Sie aus b): Für jedes $m \in \mathbb{N}$ ist $((L(X_n), R(X_n)))_{n=1,\ldots,m}$ so verteilt wie die Farbanzahlen der ersten m Zugänge in einer Pólya-Urne (mit anfänglich einer weißen und einer blauen Kugel in der Urne).

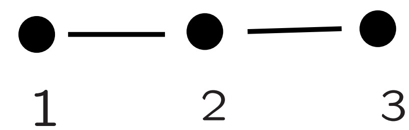
43. S Wir betrachten die gewöhnliche Irrfahrt in $\mathbb{Z}^2$: von jedem Punkt geht man jedesmal einen Schritt der Größe 1, unabhängig von der Vorgeschichte mit W'kt 1/4 nach Osten, Norden, Westen oder Süden.

a) Berechnen Sie die Wahrscheinlichkeit dafür, dass man ausgehend von dem in der nebenstehenden Skizze mit 3 bezeichneten Punkt die Menge $\{1, 2, 3\}$ nach Norden oder Osten verlässt.



b) Berechnen Sie die erwartete Anzahl der Schritte bis zum erstmaligen Verlassen der Menge $\{1, 2, 3\}$ bei Start in 1.

44. S a) Es sei $X_0, X_1, \ldots$ eine gewöhnliche Irrfahrt auf dem nebenstehend skizzierten Graphen mit den Knoten 1,2,3, und $Y_n := \mathbf{1}_{\{1,2\}}(X_n)$.



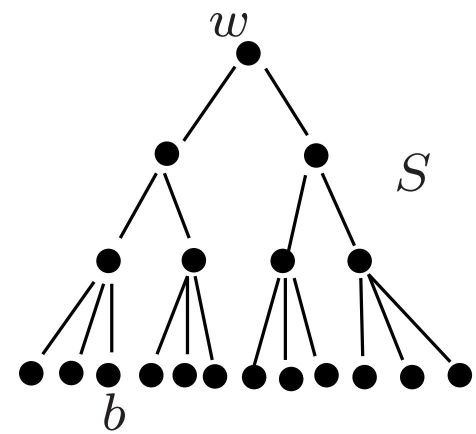
Berechnen Sie

(i) $\mathbf{P}(Y_3 = 0 \,|\, Y_0 = 0, Y_1 = 1, Y_2 = 1)$

(ii) $\mathbf{P}(Y_3 = 0 \,|\, Y_0 = 1, Y_1 = 0, Y_2 = 1)$

Ist (Y_n) eine Markovkette?

44 S b) Wieder sei $X_0, X_1, \ldots$ eine gewöhnliche Irrfahrt, diesmal auf dem nebenstehend skizzierten Baum mit der Knotenmenge S. Für jeden Knoten k sei $h(k)$ die *Tiefe* von k (also z.B. $h(w) = 0$, $h(b) = 3$).

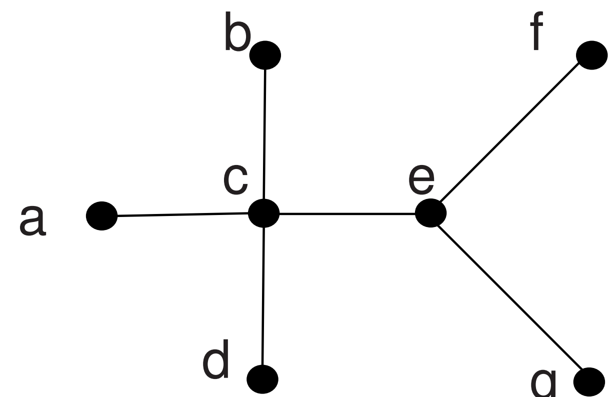


i) Ist $Y_n := h(X_n)$, $n = 0, 1, 2, \ldots$ eine Markovkette?

ii) Wieviele Schritte benötigt die Irrfahrt bei Start in b in Erwartung bis zum ersten Treffen von w?

45. S a) Auf $S = \{1, 2, 3, 4\}$ betrachten wir die Markovkette mit der durch
$P(1, 2) = 1,\ P(2, 1) = P(2, 3) = 1/2, P(3, 2) = 1/3,$
$P(3, 4) = 2/3, P(4, 3) = 1$
gegebenen Übergangsmatrix. Berechnen Sie die Gleichgewichtsverteilung.
(*Hinweis: Suchen Sie nach einem reversiblen Gleichgewicht.*)

b) Wir betrachten eine Markovkette auf dem skizzierten Graphen mit den folgenden Übergangsgewichten:
$P(a,c) = P(b,c) = P(d,c) = P(f,e) = P(g,e) = 1$,
$P(c,e) = 1/2$,
$P(c,a) = P(c,b) = P(c,d) = 1/6$,
$P(e,c) = P(e,f) = P(e,g) = 1/3$.
Bestimmen Sie die Gewichte der Gleichgewichtsverteilung. (*Hinweis: Machen Sie sich Symmetrien zunutze und verwenden Sie a).*)



c) Bestimmen Sie die Übergangsmatrix und die Gleichgewichtsverteilung der gewöhnlichen Irrfahrt auf dem skizzierten Graphen.

46. S In zwei großen Populationen ist der Mittelwert der Größe der Individuen jeweils μ_1 bzw. μ_2. Aus den beiden Populationen wurden Stichproben der Größe 100 bzw 200 erhoben. Die Stichprobenmittelwerte waren 8.0 und 8.1 und die Stichprobenstandardabweichungen waren 1.1 und 1.0.
a) Finden Sie die Realisierung eines approximativen (i) 95%- (ii) 99%-Konfidenzintervalls für die Differenz $\mu_1 - \mu_2$. (Hinweis: Der R-Befehl `qnorm(p)` liefert $\Phi^{-1}(p)$.)
b) Wie wahrscheinlich ist eine betragsmäßig mindestens so große Differenz der Stichprobenmittelwerte wie die beobachtete, wenn die beiden Populationsmittelwerte gleich sind? Verwenden Sie die approximative Normalverteiltheit der Differenz der Stichprobenmittelwerte.[§] (Hinweis: Der R-Befehl `pnorm(b)` liefert $\Phi(b)$.)

[§] Im Jargon der Statistik: *Zu welchem p-Wert lässt sich die Hypothese "$\mu_1 = \mu_2$" aufgrund des z-Tests ablehnen?*

47. In einer großen Population gibt es einen Anteil p von Typ A-Individuen. In einer Stichprobe der Größe 60 waren 9 Typ A-Individuen.
a) Bestimmen Sie ein approximatives 95%-Konfidenzintervall für p.
b) Wie wahrscheinlich ist eine betragsmäßig so große Abweichung des Anteils in der Stichprobe vom Anteil in der Population, wenn p in Wahrheit 0.2 ist? Rechnen Sie mit approximativer Normalität.¶

¶Anders gesagt: *Zu welchem p-Wert lässt sich die Hypothese“$p = 0.2$” aufgrund des z-Tests ablehnen?*

48. a) 6 Objekte werden (ohne Mehrfachbelegungen) auf die Plätze $\{1, \dots, 18\}$ gesetzt.
(i) Was ist unter der Hypothese der reinen Zufälligkeit der Erwartungswert μ der Anzahl X der Objekte, die auf der Platzmenge $\{1, \dots, 10\}$ landen?
(ii) Es ist nur eines der sechs Objekte auf $\{1, \dots, 10\}$ gelandet. Welche Ausgänge von X sind mindestens so "extrem", d.h. mindestens so weit entfernt von μ wie der beobachtete Ausgang 1? Wie wahrscheinlich ist es unter der Hypothese der reinen Zufälligkeit, einen mindestens so extremen Ausgang wie den beobachteten zu bekommen?[‖]

[‖]Anders gesagt: *Zu welchem p-Wert können Sie unter Verwendung von Fishers exaktem Test die Hypothese der reinen Zufälligkeit verwerfen?*

48 b) (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. Wie wahrscheinlich ist unter der Hypothese der reinen Zufälligkeit ein Ausgang der "Rangsumme", der mindestens so nahe zu einem der beiden extremen Ausgänge $1 + 2 + 3$ bzw. $28 + 29 + 30$ liegt wie die beobachtete Stichprobensumme $7 + 2 + 1$?**

**Anders gesagt: *Zu welchem p-Wert können Sie unter Verwendung des Wilcoxon-Rangsummentests die Hypothese der reinen Zufälligkeit (zugunsten einer "Tendenz hin zum Rand", d.h. zu den Plätzen mit den kleinen bzw zu den Plätzen mit den großen Rängen) verwerfen?*