6: Diskrete Wahrscheinlichkeit

- Die Wahrscheinlichkeitsrechnung ist eines der wichtigsten mathematischen Werkzeuge für Informatiker (probabilistische Algorithmen, zuverlässige Systeme, ...).
- Die diskrete Wahrscheinlichkeit ist ein besonders einfacher Fall (im Gegensatz zur kontinuierlichen Wahrscheinlichkeit).

6.1: Grundbegriffe

Definition 93

Ein *diskreter Wahrscheinlichkeitsraum* besteht aus

- der Menge der *Elementarereignisse*, einer abzählbaren Menge Ω, zusammen mit
- der *Wahrscheinlichkeitsverteilung*, einer Funktion

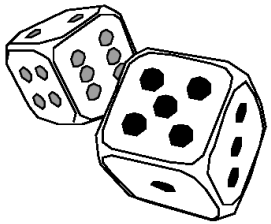
$$\Pr : \Omega \to \mathbb{R}_{\geq 0}, \quad \text{mit} \quad \sum_{x \in \Omega} \Pr(x) = 1.$$

Dabei bezeichnet $\mathbb{R}_{\geq 0}$ die Menge aller reellen Zahlen ≥ 0.

In diesem Kapitel benutzen wir den Bezeichner Ω stets für eine abzählbare Menge von Elementarereignissen und $\Pr$ für die zugehörige Wahrscheinlichkeitsverteilung.

Beispiele (Würfel)

- Fairer Würfel (“Laplace-Würfel”):
  $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\forall x \in \Omega : \Pr[x] = 1/6$.
- Würfel mit Bleigewicht unter der 1:
  $\Omega = \{1, 2, 3, 4, 5, 6\}$,
  $\Pr[2] = \Pr[3] = \Pr[4] = \Pr[5] = 1/6$, $\Pr[6] = 2/9$.
  Was ist $\Pr[\mathbf{1}]$?
  Wie wahrscheinlich ist es, eine gerade Zahl zu würfeln?



- Fairer Würfel mit manipulierten Augenzahlen 2,3,4,5,6,6:
  $\Omega = \{1, 2, 3, 4, 5, 6\}$,
  $\Pr[1] = 0$, $\Pr[2] = \Pr[3] = \Pr[4] = \Pr[5] = 1/6$,
  $\Pr[6] = 1/3$.

Ereignisse und ihre Wahrscheinlichkeit

Definition 94

Die Elemente von Ω sind die *Elementarereignisse*,
alle Teilmengen von Ω sind *Ereignisse*,
und die *Wahrscheinlichkeit* $\Pr[E]$ eines Ereignisses $E \subseteq \Omega$ ist

$$\Pr[E] = \sum_{x \in E} \Pr(x).$$

Das Ereignis $\{\}$ ist das *unmögliche* und Ω selbst ist das *sichere Ereignis*.

Beispiel: Gleichverteilung (“Laplace-Verteilung”)

Sei Ω endlich. Sind alle Elementarereignisse gleich wahrscheinlich, dann ist $\forall x \in \Omega : \Pr[x] = 1/|\Omega|$.

Ferner gilt für alle Ereignisse $E \subseteq \Omega$:

$$\Pr[E] = \frac{|E|}{|\Omega|}.$$

Konkretes Beispiel:
Fairer (“Laplace”-) Würfel
(kennen wir schon).

$\Omega = \{1, 2, 3, 4, 5, 6\}$,
$\forall x \in \Omega : \Pr[x] = 1/6$, $\Pr[\{2, 4, 6\}] = 1/2$.

Eigenschaften der Wahrscheinlichkeit

Satz 95 (Monotonie der Wahrscheinlichkeit)

Sei $A \subseteq B \subseteq \Omega$. Dann gilt

$$\Pr[A] \leq \Pr[B].$$

Satz 96 (Additivität der Wahrscheinlichkeit)

Seien $A, B \subseteq \Omega$. Dann gilt

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B].$$

Folgerungen

Seien A ein Ereignis. Dann gilt:

$$0 \leq \Pr[A] \leq 1.$$

Sind die Ereignisse A und B disjunkt, also $A \cap B = \{\}$, dann gilt

$$\Pr[A \cup B] = \Pr[A] + \Pr[B].$$

Seien $A \subseteq \Omega$. Dann gilt

$$\Pr[\overline{A}] = 1 - \Pr[A].$$

Beispiel: Mehrere Würfe mit einem Würfel

- Man werfe den Laplace-Würfel zweimal.
- Nun ist $\Omega = \{(1,1),(1,2),\ldots,(6,6)\}$
  (Beachte: (i,j) heißt "zuerst i, dann j gewürfelt")
  für $i \neq j$ ist $(i,j) \neq (j,i)$!)
- Man gebe $\Pr[(2,3)]$ und $\Pr[(3,6)]$ an.
- Für (i,j) gebe man $\Pr[i=j]$, $\Pr[i<j]$ und $\Pr[i>j]$ an!

- Nun werfe man den "fairen" 2,3,4,5,6,6-Würfel zweimal.
  Man beachte: $\Pr[(1,1)] = \cdots = \Pr[(1,6)] = 0$ und
  $\Pr[(2,1)] = \cdots = \Pr[(6,1)] = 0$.
- Man gebe $\Pr[(2,3)]$ und $\Pr[(3,6)]$ an.
- Für (i,j) gebe man $\Pr[i=j]$, $\Pr[i<j]$ und $\Pr[i>j]$ an!

Zwei unabhängige Ereignisse

Definition 97

Zwei Ereignisse $A, B \subseteq \Omega$ heißen *(stochastisch) unabhängig*, wenn

$$\Pr[A \cap B] = \Pr[A] * \Pr[B]$$

gilt. Andernfalls heißen sie *(stochastisch) abhängig*.

Beispiel: Laplace-Würfel.

$E_1 = \{2, 3\}$
$E_2 = \{3, 5\}$
$E_3 = \{2, 3, 4\}$.
$E_4 = \{1, 2, 3\}$.

- Es gibt 6 Paare von Ereignissen (E_1, E_2), $(E_1, E_3) \ldots, (E_3, E_4)$.
- Welche dieser Paare sind unabhängig, welche nicht?

6.2: Bedingte Wahrscheinlichkeit

Zwei Münzen:

- eine faire mit “Kopf” und “Zahl” und
- eine unfaire mit “Kopf” auf beiden Seiten.

Wir wählen zufällig eine der beiden Münzen und werfen sie.

Unser Wahrscheinlichkeitsraum:

- $\Omega = \{$ “Kopf”, “Zahl” $\}$
- Pr[“Kopf”] $= 3/4$; Pr[“Zahl”] $= 1/4$.

Zweifaches Werfen der Münze

Nun werfen wir die zufällige Münze zweimal, also

$$\Omega = \{\text{“Kopf”}, \text{“Zahl”}\}^2.$$

Sei X das Ereignis, dass wir im ersten Wurf eine Zahl werfen, Y das Ereignis, dass wir im zweiten Wurf eine Zahl werfen.

Offenbar ist $\Pr[X] = \Pr[Y] = 1/4$.

Aber: wenn X bereits eingetreten ist (oder $\overline{X}$), dann “ändert” sich die Wahrscheinlichkeit des Ereignisses Y auf einmal ...

Wie kann man das formal beschreiben?

Das Knabenchor-Experiment

Felix klappert die Familien der Nachbarschaft ab um Nachwuchs für seinen Knabenchor anzuwerben. Er hat eine Liste mit allen Familien mit mindestens einem Jungen.

Die Wahrscheinlichkeit, dass ein Kind ein Junge ist, ist (ungefähr) 1/2. Deshalb erwart Felix, dass in Hälfte der Familien mit genau zwei Kindern sogar zwei potentielle Kandidaten für den Knabenchor sind.

Hat Felix recht mit dieser Erwartung?

Felix behauptet, bei “fast allen Familien mit zwei Kindern ist das andere ein Mädchen”. Tatsächlich stellte sich heraus, dass von den Famlien mit zwei Kindern nur zwei Jungen hatten, die anderen acht hatten einen Jungen und ein Mädchen.

Hat Felix bei seiner Werbung für den Knabenchor besonderes Pech gehabt, oder ist das Ergebnis statistisch normal?

Antwort 1: Geringe Stichprobengröße

Unser Wahrscheinlichkeitsraum:

- $\Omega = \{\text{Junge}, \text{Mädchen}\}$
- $\Pr[\text{Junge}] = \Pr[\text{Mädchen}]$

Die Wahrscheinlichkeit, dass von 10 zufällig gewählten Kindern, die unabhängig voneinander mit jeweils der Wahrscheinlichkeit$^{(*)}$ 1/2 Junge bzw. Mädchen sein könnten, nur zwei Jungen sind, ist

$$\frac{10 * 9}{2} * 2^{-10} = 45 * 2^{-10} \approx 4.4\% \approx \frac{1}{23}.$$

(Warum?)

So erscheint das Ergebnis von Felix zwar nicht gerade als Senstation, aber doch als kleiner statistischer Ausreißer.

$^{(*)}$ In Wirklichkeit ist die Wahrscheinlichkeit, dass ein Neugeborenes ein Junge ist, etwas größer als 1/2.

Antwort 2: Wahrscheinlich ist “das andere” ein Mädchen!

Wir wiederholen das Experiment in einer Computersimulation mit 100 000 Familien mit jeweils zwei Kindern. Es gibt vier mögliche Kombinationen von Kindern, und, wie man auch mathematisch herleiten könnte, hat jede hat die Wahrscheinlichkeit 1/4:

1.	2.	Anzahl
Junge	Junge	25 070
Junge	Mädchen	25 117
Mädchen	Junge	24 803
Mädchen	Mädchen	25 010

Dabei ist das Geschlecht des zweiten Kindes unabhängig$^{(**)}$ von dem des ersten.

$^{(**)}$ In Wirklichkeit gibt es auch eineiige Zwillinge, bei denen beide zwangsläufig das gleiche Geschlecht haben.

Zu Antwort 2

1.	2.	Anzahl
Junge	Junge	25 070
Junge	Mädchen	25 117
Mädchen	Junge	24 803
Mädchen	Mädchen	25 010

- Auf Felix Liste sind keine Familien mit (nur) zwei Mädchen.
- Statistisch zu erwarten: Zwei Drittel aller Familien mit zwei Kindern auf der Liste haben einen Jungen und ein Mädchen, ein Drittel zwei Jungen.
- Das sind etwa 3 von 10.
- Die 2 von 10, die Felix tatsächlich antraf, sind tatsächlich *“statistisch normal” und grade kein Ausreißer!*

Bedinge Wahrscheinlichkeit

Definition 98

Seien A und B zwei Ereignisse mit $\Pr[B] > 0$. Die *bedingte Wahrscheinlichkeit* $\Pr[A|B]$, dass das Ereignis A unter der Bedingung B eintritt, ist

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]}.$$

Beispiel: Das Experiment mit der fairen und der unfairen Münze:

Wie groß sind $\Pr[X|Y]$ und $\Pr[\overline{X}|Y]$?

Satz 99

Zwei Ereignisse $A, B \subseteq \Omega$ sind unabhängig, genau dann, wenn $\Pr[A|B] = \Pr[A]$.

Eigenschaften der bedingten Wahrscheinlichkeit

Satz 100 (Multiplikationssatz für Wahrscheinlichkeiten)

$$\Pr[A \cap B] = \Pr[B] * \Pr[A|B].$$

Satz 101 (Bayes)

Seien $A, B \subset \Omega$ *Ereignisse mit* $\Pr[A] > 0 < \Pr[B]$. *Dann gilt:*

$$\Pr[A|B] = \frac{\Pr[A]}{\Pr[B]} * \Pr[B|A].$$

Anwendung: Ein statistischer Test

- Gegeben sei ein Test auf eine bestimmte Krankheit.
- Es seien 0.1 % aller Personen erkrankt.
- Falsch-Negatives Resultat (eine kranke Person wird fälschlich als “gesund” eingestuft) mit der Wahrscheinlichkeit 0.2 %.
- Falsch-Positives Resultat (eine gesunde Person wird fälschlich als “krank” eingestuft) mit der Wahrscheinlichkeit 0.3 %.

- Mariette lässt sich testen. Das Ergebnis ist “krank”. Wie wahrscheinlich ist Mariette tatsächlich krank?

Anwendung: Rechtschreibkorrektur

Neulich bei Google:

Wie macht Google das?

Können wir ein Programm schreiben, das das auch kann?

Idee für eine Rechtschreibkorrektur

- Eingegebenes Wort **W**
- Gesucht: Korrekturvorschlag **C** mit maximaler $\Pr[\mathbf{C}|\mathbf{W}]$
  ... leider ist $\Pr[\mathbf{C}|\mathbf{W}]$ unbekannt ...
- Satz von Bayes:

$$\Pr[\mathbf{C}|\mathbf{W}] = \frac{\Pr[\mathbf{W}|\mathbf{C}] * \Pr[\mathbf{C}]}{\Pr[\mathbf{W}]}$$

- Gesucht: **C**, so dass $\Pr[\mathbf{W}|\mathbf{C}] * \Pr[\mathbf{C}]$ maximal ist
- **Fehlermodell:** $\Pr[\mathbf{W}|\mathbf{C}]$
- **Sprachmodell:** $\Pr[\mathbf{C}]$

Das Sprachmodell – ein Training für die Fehlerkorrektur

- Liste von Dokumenten in der Zielsprache!
- Für alle auftretenden Wörter $\mathbf{C}$: Zähle, wie oft $\mathbf{C}$ insgesamt auftritt.
- Eleminiere Wörter $\mathbf{C}$, die extrem selten oder nur in wenigen Dokumenten auftretren (keine Fehler eintrainieren).
- Schätzwert für $\Pr[\mathbf{C}]$: relative Häufigkeit
- (Der tatsächliche Wert $\Pr[\mathbf{C}]$ ist gar nicht wichtig – wir müssen für $\mathbf{C} \neq \mathbf{C}'$ nur abschätzen können, ob $\Pr[\mathbf{C}] > \Pr[\mathbf{C}']$ gilt, oder nicht.)

Das Fehlermodell – eine einfache Heuristik

Google hat vermutlich ein sehr ausgeklügeltes Fehlermodell.

Aber korrekte Wörter sind so selten, dass die folgende Regel bereits brauchbare Ergebnisse liefert.

Zwei Wörter **W** und **C** haben *Editier-Distanz*

- 0, falls **C** $=$ **W** gilt,
- 1, falls man **W** in **C** umwandeln kann, durch eine der folgenden Operationen:
  - Einfügen eines Buchstabens
  - Entfernen eines Buchstabens
  - Ersetzen eines Buchstabens durch einen anderen
  - Vertauschen zweier benachbarter Buchstaben
- i, falls man mindestens i dieser Operationen braucht

Eine einfache, aber schon leistungsfähige Rechtschreibkorrektur

Seien W gegeben und das Sprachmodell in Form einer nach Häufigkeit sortierten Wortliste eintrainiert.

1. Suche das kleinste i, für das es (mindestens) ein **C** in der Wortliste gibt, mit der Editier-Distanz i zwischen **C** und **W**.

   (Wird i zu groß, brich ab!)
2. Gib das Wort **C** aus, das von allen Wörtern mit der Editier-Distanz i die maximale Wahrscheinlichkeit $\Pr[\mathbf{C}]$ hat.

Siehe ⟨`http://norvig.com/spell-correct.html`⟩.

Bayes'sche Statistik

- Die Anwendung des Satzes von Bayes ist ein sehr verbreiteter Trick, für die Auswertung von Daten.
- Z.B. arbeiten die meisten SPAM-Filter auf dieser Basis.
- Firmen nutzen den Satz von Bayes, um Dinge über ihre Nutzer herauszufinden, die die Nutzer nicht gewollt preisgegeben haben.
- Z.B. kann man, durch die Analyse der Verbindungen in sozialen Netzen, erkennen (die begründete Vermutung aufstellen), ob eine Person homo- oder heterosexuell ist, auch wenn man die Information nicht explizit im Netz findet.
- Wetten, Google kann erkennen, ob Sie Rechts- oder Linkshänder sind?

6.3: Perfekte Verschlüsselung

- $M \in \{0,1\}^n$: vertrauliche Nachricht
- $C = E_K(M)$: Chiffretext (Verschlüsselung von M unter einem geheimen Schlüssel K)
- "Angreifer" kennt C, aber weder M noch K
- Legaler Empfänger kennt K, erfährt C und berechnet $M = D_K(C)$ durch Entschlüsselung von C unter K

Idee

Ein Verschlüsselungssystem heißt *perfekt*, falls der "Angreifer" aus dem Chiffretext "nichts über M erfährt, was er nicht sowieso schon weiß".

Von der Idee zur formalen Definition

Definition 102

Ein Verschlüsselungssystem heißt *perfekt*, falls für den "Angreifer" und für alle Nachrichten $M \in \{0,1\}^n$ und alle Chiffretexte C gilt:

$$\Pr[\text{“}M\text{”} | \text{“}C\text{”}] = \Pr[\text{“}M\text{”}].$$

(Achtung: Etwas missbräuchliche Notation! Für $M \in \{0,1\}^n$ bezeichnet "M" das Ereignis, dass der Sender diese Nachricht verschickt. Ebenso für Chiffretexte "C" und Schlüssel "K".)

Die Vernam-Chiffre ("One-Time Pad", Vernam, 1917)

- $M \in \{0,1\}^n$: vertrauliche Nachricht
- Schlüssel $K \in \{0,1\}^n$, gleichverteilt
- Chiffretext $C \in \{0,1\}^n$: $C = M \oplus K$.
- Entschlüsselung: $M = C \oplus K$.

Satz 103 (Shannon, 1949)

Die Vernam-Chiffre ist perfekt.

Perfekte Verschlüsselung ist in der Praxis eine seltene Ausnahme

Satz 104

Bei jedem perfekten Verschlüsselungssystem muss der Schlüssel mindestens so lang sein, wie die Nachricht.

Werden mehrere Nachrichten verschlüsselt, muss der Schlüssel mindestens so lang sein, wie die Längen aller Nachrichten zusammen.

(Ohne Beweis)

6.4: Das Geburtstagsparadoxon

Wie wahrscheinlich ist es, dass **von 22 Spielern** bzw. **von 23 Leuten** (Spieler + Schiedsrichter) auf einem Fußballfeld **zwei am gleichen Tag Geburtstag haben**

22 Spieler: etwa 47.6 %
23 Leute: etwa 50.7 %

Überrascht?

Wie rechnet man das aus?

- $n \leq 365$ Bälle, die jeweils zufällig in einen von 365 Körben geworfen werden.
- Die Wahrscheinlichkeit p_n, dass von n Bällen in jedem Korb höchstens ein Ball ist:
  - $p_1 = (365/365) = 1$
  - $p_2 = (364/365)$
  - $p_3 = (364/365) * (363/365)$
  - $p_4 = (364/365) * (363/365) * (362/365)$
  - ...

$$p_n = \prod_{0 \leq i < n} \frac{365 - i}{365}.$$

- Die Wahrscheinlichkeit, dass in mindestens einem Korb mehr als ein Ball liegt, ist natürlich $1 - p_n$, siehe Tabelle.

Anz.	$1 - p_n$
1	0.0000
2	0.0027
3	0.0082
4	0.0164
⋮	⋮
20	0.4114
21	0.4437
22	0.4757
23	**0.5073**
24	0.5383
25	0.5687
26	0.5982
27	0.6269
28	0.6545

Verallgemeinerung des Geburtstagsproblems

- $\mathbf{k}$ Körbe und (wie bisher) $\mathbf{n} \leq \mathbf{k}$ Bälle.
- Die Wahrscheinlichkeit, dass wenn alle $\mathbf{n} - \mathbf{1}$ Bälle bisher in verschiedenen Körben gelandet sind, auch der $\mathbf{n}$-te in einem noch leeren Korb landet:

$$\mathbf{p_n} = \prod_{0 \leq i < \mathbf{n}} \frac{\mathbf{k} - i}{\mathbf{k}}.$$

- Für "große" $\mathbf{k}$ erwartet man eine Kollision (zwei Bälle, ein Korb) bei

$$\mathbf{n} \approx \sqrt{\frac{\pi}{2}} * \sqrt{\mathbf{k}}$$

(ohne Beweis, siehe aber nächste Folie). Es ist $\sqrt{\frac{\pi}{2}} \approx 1.25$.

- Für $\mathbf{n} = \mathbf{c} * \sqrt{\frac{\pi}{2}} * \sqrt{\mathbf{k}}$, ist die erwartete Anzahl an Kollisionen $\mathbf{c^2}$.

Informelle Begründung

- Zwei Bälle landen genau

  mit der Wahrscheinlichkeit $\mathbf{1/K}$ im gleichen Korb.

- Bei insgesamt $\mathbf{n}$ Bällen gibt es

$$\frac{\mathbf{n}(\mathbf{n}-1)}{2} \approx \frac{\mathbf{n}^2}{2} \text{ Paare von Bällen,}$$

  die im gleichen Korb landen können (ohne Reihenfolge).

- Summe der Einzelwahrscheinlichkeiten:

$$\approx \frac{\mathbf{n}^2}{\mathbf{2K}}.$$

	1	2	3	$\cdots$	$\mathbf{n-1}$	$\mathbf{n}$
1	–	$\mathbf{1/K}$	$\mathbf{1/K}$		$\mathbf{1/K}$	$\mathbf{1/K}$
2	–	–	$\mathbf{1/K}$		$\mathbf{1/K}$	$\mathbf{1/K}$
3	–	–	–		$\mathbf{1/K}$	$\mathbf{1/K}$
$\vdots$						
$\mathbf{n-1}$	–	–	–		–	$\mathbf{1/K}$
$\mathbf{n}$	–	–	–		–	–

Anwendung: Diskrete Logarithmen

- Die Effizienz vieler probabilistischer Algorithmen hängt eng mit der Anzahl an “zufälligen Kollisionen” zusammen.
- Uns schon bekannt ist das **Problem des Diskreten Logarithmus:** ($\rightarrow$ Diffie-Hellman Schlüsselaustausch): Seien ein erzeugendes Element g einer (zyklischen) Gruppe der Ordnung q und $x = g^{\mathbf{y}}$ gegeben, gesucht ist $\mathbf{y}$.
- Es gibt Algorithmen zur Berechnung des Diskreten Logarithmus, die mit

$$\sqrt{\tfrac{\pi}{2}} * \sqrt{\mathbf{q}} \quad \text{Rechenschritten}$$

  auskommen (Berechne viele $g^i * x^j$, $\rightarrow$ Tafel).
- **Warum verlangen wir für den D.-H. Schlüsselaustausch $q > 2^{160}$?** Wir wollen Angreifer zwingen, mindestens $\mathbf{2^{80}}$ Rechenschritte durchzuführen – in der Erwartung, dass sie diese Rechenleistung nicht aufbringen können.

6.5: Zufallsvariablen und Erwartungswerte

Eine *Zufallsvariable* (“random variable”) ist weder zufällig, noch eine Variable! Sie ist eine Funktion, die jedem Ereignis eines Ereignisraumes einen numerischen Wert zuordnet:

Definition 105

Sei $(\Omega, \Pr)$ ein diskreter Wahrscheinlichkeitsraum.
Eine *diskrete Zufallsvariable* ist eine Funktion

$$X : \Omega \to R.$$

Dabei ist R eine abzählbare Teilmenge von $\mathbb{R}$.

Jedes $a \in R$ hat eine bestimmte Wahrscheinlichkeit:

$$\Pr[X = a] = \sum_{\omega \in \Omega, X(\omega) = a} \Pr(\omega).$$

Wozu braucht man Zufallsvariablen?

Entwicklung der Wahrscheinlichkeitsrechnung: Verstehen der Gewinnerwartung von Glücksspielen! Es kommt nicht darauf an, welche Karten man zieht, sondern **ob** man mit diesen Karten gewinnt oder verliert, bzw. **wieviel** man gewinnt oder verliert:

Definition 106

Der *Erwartungswert* einer diskreten Zufallsvariablen X ist

$$E(X) = \sum_{\omega \in \Omega} X(\omega) * \Pr[\omega].$$

Satz 107

$$E(X) = \sum_{a \in R} a * \Pr[X = a].$$

Beispiele

- Ist der Wertebereich $R = \{a_1, \ldots, a_n\}$ von X endlich, und ist jedes Ergebnis a_i gleich wahrscheinlich, dann ist

$$E(X) = \frac{a_1 + a_2 + \cdots + a_n}{n}$$

  das arithmetische Mittel der a_i.
- Sind, z.B., $R = \{1, 2, 3, 4, 5, 6\}$ und X die Augenzahl eines (fairen) Würfels, dann ist $E(X) = 3.5$.
- Bei einem Spiel mit einem fairen Würfel gewinnen wir 4 Euro, wenn wir eine 6 würfeln. Sonst verlieren wir 1 Euro. D.h. $X(1, 2, 3, 4, 5) = -1$, $X(6) = 4$ und

$$E(X) = \sum_{a \in R} a * \Pr[X = a] = ((-1) * (5/6)) + (4 * (1/6)) = -1/6.$$

Die Linearität des Erwartungswertes

Satz 108

Seien X und Y Zufallsvariablen und $a, b \in \mathbb{R}$. Dann gilt

$$E(aX + bY) = aE(X) + bE(Y).$$

Anwendung: Der Massen-Gentest

- Bei der Fahndung nach einem Raubkopierer gab es einen Durchbruch. Die Polizei konnte einen gebrauchten Kaugummi des Täters sicherstellen. Nun ruft die Polizei zu einem freiwilligen Gentest auf.
- Der Gentest ist höchst zuverlässig: Nur mit der Wahrscheinlichkeit 1/1 Million muss ein Unschuldiger befürchten, dass sein Test "positiv" ausfällt, d.h., dass er, obwohl unschuldig, mit einer Anklage als mutmaßlicher Raubkopierer rechnen muss.

- Einer der ersten Freiwilligen ist Josef K. Sein Testergebnis ist "positiv".
- Kann die Polizei den teuren Massentest nun abblasen, weil sie den mutmaßlichen Täter bereits gefunden hat?

6.6: Drei wichtige Verteilungen
1. Die Bernoulli-Verteilung mit Parameter p

- Jede Zufallvariable X hat nur zwei mögliche Werte 0 und 1; $p := \Pr[X = 1]$, $E(X) = p$
- Beispiel:
  Würfeln mit einem Würfel; der Wert sei 1 wenn wir eine Sechs würfeln und 0 sonst.
  Ist der Würfel fair, dann gilt $p = 1/6$.

2. Die Binomialverteilung $B(n, p)$

- Seien $X_1, \ldots, X_n$ n unabhängige Bernoulli-verteilte Zufallsvariablen mit Parammeter p.
- Dann ist die Summe $Y = X_1 + X_2, + \cdots + X_n$ eine binomialverteilte Zufallsvariable.
- Erwartungswert: $E(Y) = nE(X)$.
- Beispiel $B(n, 1/6)$: Würfele einen fairen Würfel n-mal und zähle die Sechsen!

  Zu Anfang des Semesters hatten wir *empirisch* untersucht, wie stark die Streuung von Y ist:

Stichprobengröße	Interval	Anteil (in %)
600	$100 \pm 10\,\%$,	> 75
6000	$1000 \pm 10\,\%$,	≈ 100
60000	$10000 \pm 1\,\%$,	> 70
600000	$100000 \pm 1\,\%$,	≈ 100

Anwendung: Wie viele Sechsen mit n Würfen?

- Die sogenannte *Standardabweichung* ist ein Maß für die Streuung einer Zufallsvariablen.
- Die Standardabweichung einer binomialverteilten Zufallsvariablen $Y = X_1 + X_2, + \cdots + X_n$ ist

$$\sigma_Y = \sqrt{n * E(X) * (1 - E(X))}.$$

- Beispiel: $B(n, 6)$ für verschiedene n:

n	600	6000	60000	600000	
$E(Y)$	100	1000	10000	100000	
σ_Y	9	29	91	289	(gerundet)

Von der Standardabweichung zur Vorhersage der Abweichung

Ohne mathematische Herleitung bzw. Begründung

Die folgende Tabelle gibt näherungsweise an, wie wahrscheinlich es ist, dass eine *binomialverteilte Zufallsvariable* Y innerhalb (bzw. außerhalb) eines gegebenen Intervalls liegt:

Wahrscheinlichkeit $\Pr[Y \in I]$	Intervall I
68,3	$E(Y) \pm \sigma_Y$
95,4	$E(Y) \pm 2\sigma_Y$
99,7	$E(Y) \pm 3\sigma_Y$

Achtung: Jede Zufallsvariable hat eine Standardabweichung, aber die obige Näherung gilt nur für bestimmte Verteilungen, unter anderem eben für die Binomialverteilung!

Anwendung: Umfrageergebnisse

- Genau 1/6 der Wähler wählen die Partei P, also $\approx 16.7\,\%$
- Bei einer Umfrage werden 1000 Wähler befragt.
- Sei Y die Anzahl der Befragten, die P wählen. Es gilt:

$$E(Y) = \frac{1000}{6} \approx 167, \quad \sigma_Y \approx \sqrt{\frac{1000}{6} * \frac{5}{6}} \approx 12.$$

$E(Y) \pm \sigma_Y$: Bei etwa jeder dritten Umfrage ist die Prognose für P liegt die Prognose außerhalb des Bereichs von 15.5 bis 17.9 Prozent.

$E(Y) \pm 2\sigma_Y$: Bei etwa jeder 23.-ten Umfrage ($\approx 4.4\,\%$) liegt die Prognose außerhalb des Bereichs von 14.3 bis 19.1 Prozent.

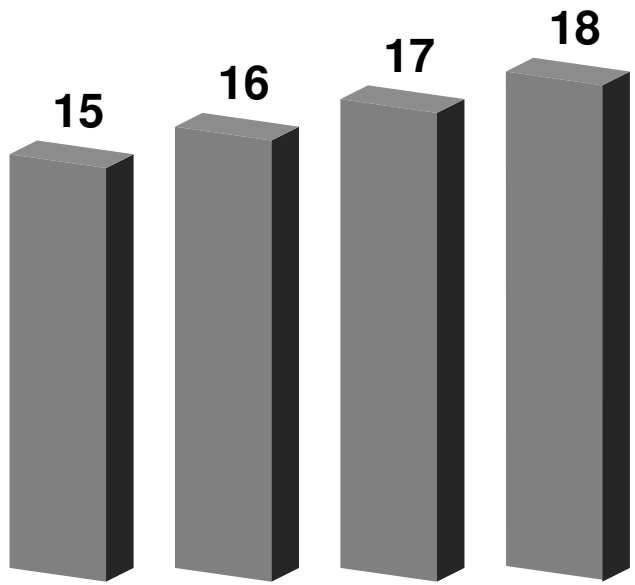
Die Simulation von Umfragen

In einer Simulation verbessert sich P kontinuierlich von Umfrage zu Umfrage, von anfangs 15.4 % auf 17.7 % am Ende.

Auf ganze Zahlen gerundet scheint die Partei jeden Monat einen Prozentpunkt hinzugewinnen.

Das Ergebnis ist kein politischer “Trend”, sondern nur “Zufall”: Der tatsächliche “Wähler”-Anteil von P war konstant 16,7 %.

```
>>> dice.count_six(1000)
154
>>> dice.count_six(1000)
156
>>> dice.count_six(1000)
170
>>> dice.count_six(1000)
177
```



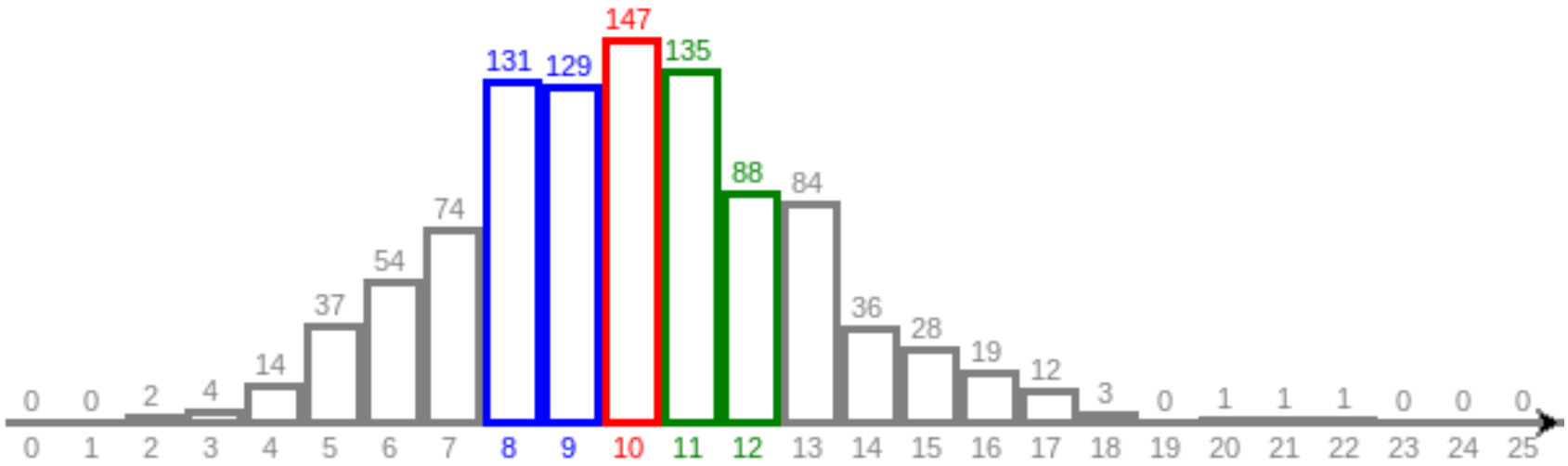
Eine echte Umfrage

Die Welt vom 10.01.2015, Hervorhebung von sl
Kann man aus den Zahlen einen Stimmungswandel ablesen?

Allerdings hat sich die Stimmung nach dem Anschlag von Paris ein wenig gewandelt. Am 8. Januar befragte Infratest Dimap die Bürger dazu nämlich ein zweites Mal. Während vor dem Anschlag **21 Prozent** ein großes oder sehr großes Verständnis für Pegida und **76 Prozent** wenig oder keins geäußert hatten, lagen diese Werte am 8. Januar bei **22 bzw. 72 Prozent**.

Idealtypische Binomialverteilung

Zur Erinnerung: Die Ergebnisse von 1000-mal 60 Würfel-Würfen



3. Die geometrische Verteilung mit Parameter p

- Wiederhole ein Bernoulli-Experiment so lange, bis das Ergebnis 1 ist.
- Die Anzahl $Z \in \mathbb{N}$ ist eine geometrisch verteilte Zufallsvariable.
- Beispiel: Würfele so lange, bis das Ereignis “Sechs” eintritt. Zähle die Anzahl der Würfe.

Satz 109

Sei Z eine geometrisch verteilte Zufallsvariable, basierend auf einer Bernoulli-Verteilung $B(1,p)$. Dann gilt

$$E(Z) = \frac{1}{p}.$$

(Fallunterscheidung: $E(Z) = p * 1 + (1-p) * (1 + E(Z)) = p + (1-p) + (1-p) * E(Z) = 1 + (1-p) * E(Z)$)

Anwendung: Primzahlsuche

Bekannter Algorithmus, um (zufällige) Primzahlen zu finden:

Wiederhole:

1. wähle eine Zufallszahl r (in einem vorgegebenen Intervall)
2. teste, ob r prim ist oder zusammengesetzt

bis r eine Primzahl ist.

Die Anzahl Z der Schleifendurchläufe ist eine geometrisch verteilte Zufallsvariable.

Ist Z mit der Wahrscheinlichkeit p prim, dann ist $E(Z) = 1/p$.