Stochastik (LK Abi16)

B. Waldmüller

28. Juni 2015

Inhaltsverzeichnis

1 Grundlagen

1.1 Vorrede

Stochastik – Mathematik des Zufalls. Das klingt in sich widersprüchlich: In der Mathematik will man Klarheit, aber wo der Zufall regiert, weiß man gerade nicht, was kommt. Wann ein bestimmtes radioaktives Teilchen zerfällt, das wir eben betrachten, wann das Geiger–Müller–Zählrohr im Physikübungsraum den nächsten Klack abgibt, wie hoch der nächste Schaden ist, der der Versicherung gemeldet wird, ob der Würfel beim nächsten Wurf die dringend benötigte Zwei wirft, wer morgen als erster im Kursraum ist, ob Til in der nächsten Stunde seine Hausaufgaben hat – all das weiß man nicht, niemand kann es mit Sicherheit vorhersagen. Aber dennoch kennt die Physik Halbwertszeiten radioaktiver Substanzen, und die Versicherung rechnet aus, wie hoch die Prämie sein muss, damit sie aller Voraussicht nach ihr gegebenes Versprechen, den Schaden zu zahlen, auch einlösen kann. Große Zahlen nehmen dem Zufall das Geheimnis. Über ein einzelnes Teilchen, über die Höhe des nächsten Schadens kann man nicht viel sagen, aber über eine große Zahl radioaktiver Teilchen, über die Gesamtheit aller Versicherten, kann man durchaus Aussagen machen. Welcher Art solche Aussagen sind, sollst du gleich sehen.

1.2 Das empirische Gesetz der großen Zahl

Es waren müßige Adlige vor einigen hundert Jahren, die beim Würfelspiel Beobachtungen machten und sich daraufhin mit Fragen an große Gelehrte wandten. Hier ist ein Beispiel: Ein Würfel wird viermal geworfen. Tritt dabei mindestens eine Sechs auf, gewinnt Spieler A, tritt keine Sechs auf, gewinnt Spieler B. Bei diesem Spiel ist übrigens Spieler A leicht im Vorteil... Kannst du das nachweisen? Kannst du präzise sagen, was das überhaupt heißt, Spieler A sei im Vorteil, und wie sich das praktisch auswirkt?[1]

Wir modernen Menschen haben nicht so viel Zeit und Muße, lange beim Würfeln zuzuschauen. Wir wollen in ein paar Minuten Einsichten gewinnen, für die die Alten beim Würfelspiele Jahre oder gar Jahrzehnte brauchten. Erstaunlicherweise geht das, es gibt ja Computer.[2] Mit MuPADs Hilfe habe ich einige Versuchsreihen mit Würfeln simuliert. Nichts Kompliziertes: ich habe N–mal würfeln und die gewürfelten Sechsen zählen lassen. Die beobachtete Anzahl n der Sechsen nennt man die **absolute** und den Anteil $h = \frac{n}{N}$ der Würfe mit Augenzahl sechs die **relative Häufigkeit** der Sechs. Natürlich liegt die absolute Häufgkeit zwischen 0 und N und die relative Häufigkeit zwischen 0 und 1, beides einschließlich der Grenzen.

Für $N = 1, 5, 25, 625$ und 3025 habe ich also jeweils zwanzig relative Häufigkeiten der Sechs auslosen und graphisch darstellen lassen. Die Ergebnisse siehst du in Abbildung 1 auf Seite 4. Sie illustrieren die folgende Erfahrungstatsache.

1 Satz (Empirisches Gesetz der großen Zahl)
Führt man einen Zufallsversuch sehr oft unabhängig durch, stablilisieren sich die relativen Häufigkeiten der Merkmale.

Was ist mit „Unabhängkeit“ gemeint? Die Ergebnisse verschiedener Würfe eines Würfels sind von einander unabhängig. Entnimmst du jedoch einem Eimer Lose der Reihe nach, hast du ja nach jeder Ziehung eine neue Situation.

[1] Nur aus stilistischen Gründen habe ich da Fragen formuliert. In Wirklichkeit handelt es sich um Aufgaben.

[2] Siehe simul1.mn . Und prüfe dich mal: Sicher verhält sich dein Rechner ab und an anders, als du erwartet hast. Aber kann er echt würfeln, das heißt echt zufällige Ereignisse simulieren? Ich sage es dir im Vertrauen: Nein, kann er nicht.

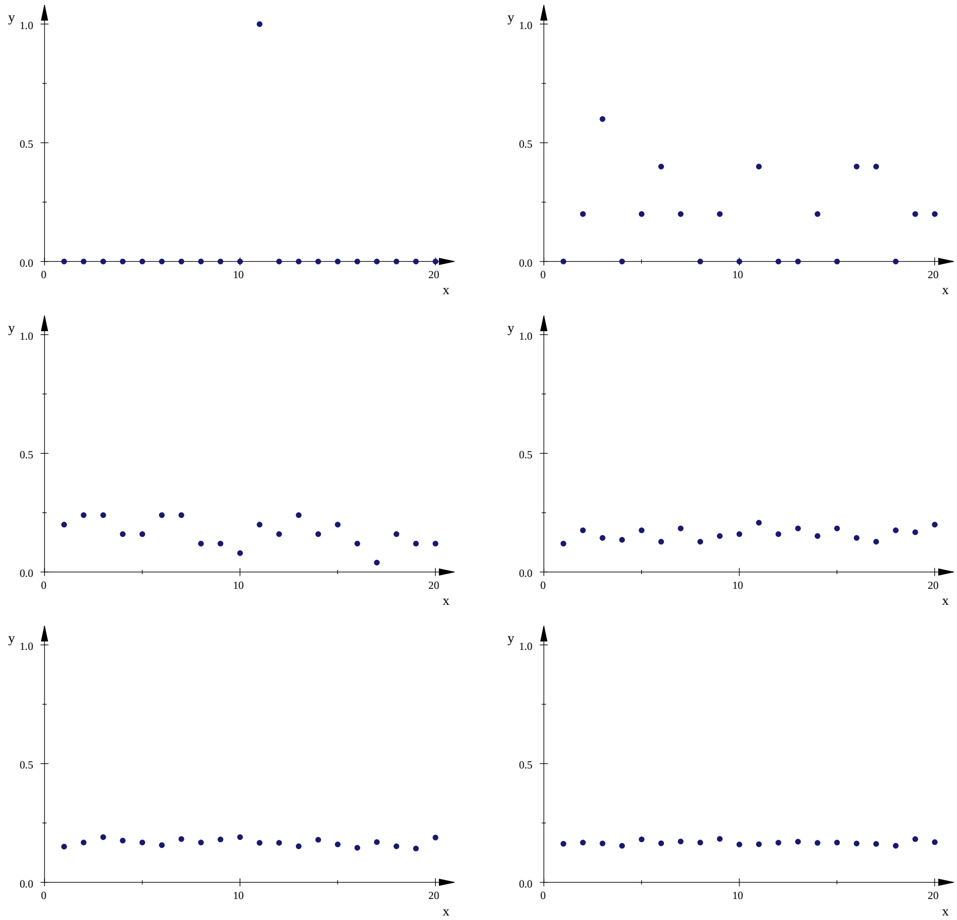


Abbildung 1: Jeweils zwanzig relative Häufigkeiten der Sechs bei N Würfen für $N = 1, 5, 25, 125, 625$ und 3125

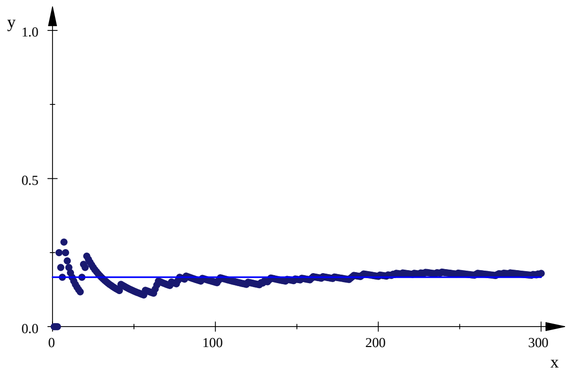


Abbildung 2: Zum Empirischen Gesetz der Großen Zahl: Es wird 300–mal gewürfelt und nach jedem Wurf die relative Häufigkeit der Sechs neu berechnet.

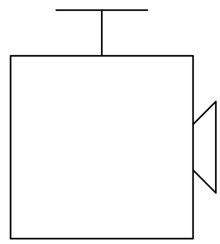
1.3 Ein paar Aufgaben

Du hast ja Vorkenntnisse und gesunden Menschenverstand. Gehe damit an die folgenen Aufgaben. Wahrscheinlichkeitstheorie brauchst du nicht dafür.

1. Wie wahrscheinlich ist es, beim Werfen eines Würfels
   (a) eine 2 oder 3 zu bekommen?
   (b) keine 6 zu bekommen?
   (c) erst eine 2, dann eine 3 zu bekommen, wenn man zweimal wirft?
   (d) eine 2 und eine 3 zu bekommen, wenn man zweimal wirft?
   (e) die erste 6 erst im n–ten Wurf zu bekommen?
   (f) eine 6 zu werfen, obwohl schon der letzte eine 6 war?
   (g) genau eine 6 zu werfen, wenn man dreimal würfelt?
   (h) in sechs Würfen mindestens eine 6 zu werfen?
2. Der Verein der Freunde des Söderblom–Gymnasiums bietet als neuen Service Mitgliedern der Schulgemeinde eine Fahrradversicherung an. Versichert sind tausend Fahrräder. Im letzten Jahr regulierte die Versicherung fünf Schäden: einen zu 2000 Euro und vier zu 100 Euro.
   (a) Wie hoch wird der nächste gemeldete Schaden sein? Was erwartest du? Tim hatte ja eine Berechnungsmethode vorgeschlagen...
   (b) Wie hoch sollte die Jahresprämie sein, die der Förderverein von seinen Versicherungsnehmern erhebt?
3. Ein progressiver Lehrer ermittelt seine Noten, indem er einen Würfel dreimal wirft und die kleinste der drei Augenzahlen als Note gibt.
   (a) Ermittle Klausurnoten für einen Kurs mit achtzehn Schülern – am besten mit einer MuPAD–Simulation, sonst zu Fuß.
   (b) Mit welcher Wahrscheinlichkeit gibt er eine Fünf, mit welcher eine Sechs?
   (c) Wie hoch ist die Durchschnittsnote in seinen Lerngruppen? Stimmt sie mit der Durchschnittsnote deines Beispiels überein?
4. Langjährige Beobachtungen haben ergeben, dass Schüler T. seine Hausaufgaben in 20% der Fälle erledigt. Sein Mathekurs hat drei Termine in der Woche. Mit welcher Wahrscheinlichkeit hat er seine Hausaufgaben in der nächsten Woche immer, zweimal, einmal, nie?

1.4 Kleiner Sprachkurs

Wir haben eine Reihe von Aufgaben gerechnet, die aus diesem Skript hier und welche aus dem Buch, und es ist nun an der Zeit, ein paar formale Dinge festzuklopfen. Es geht immer um einen Vorgang, der zu einem Ergebnis führt, das man nicht von vornherein vorhersagen kann. Stelle dir den Vorgang ruhig so vor, dass du eine Maschine durch einen Knopfdruck dazu bringst, ein Produkt auszustoßen.



Sprechweise: Es wird ein **Zufallsversuch** durchgeführt, und man erhält ein **ausgelostes Ergebnis** des Versuchs. Der Versuch mag das Werfen eines Würfels sein, und das Ergebnis ist die gewürfelte Zahl. Diese (dynamische) Sicht der Dinge soll dir das Verständnis erleichtern, die Wahrscheinlichkeitstheorie braucht keine Zufallsversuche. Da zählt nur die Menge Ω der Ergebnisse, der sogenannte **Stichprobenraum** oder **Ergebnisraum** – beim Würfeln wäre das zum Beispiel

$$\Omega \;=\; \{\,1,2,3,4,5,6\,\}\;\; .$$

Die Elemente von Ω heißen Ergebnisse, sie werden gewöhnlich mit ω bezeichnet.

Ferner braucht man in der Wahrscheinlichkeitstheorie eine Funktion $\mathbb{P}$, die gewissen Teilmengen[3] von Ω Zahlen zwischen 0 und 1 zuordnet. Diese Teilmengen von Ω heißen **Ereignisse**, falls sie nur ein einziges $\omega \in \Omega$ enthalten, **Elementarereignisse**.

Wenn du beim Mensch–ärgere–dich–nicht vor dem Ziel stehst und dir jede Zahl unter fünf nützlich ist, interessierst du dich für das Ereignis

$$A \;=\; \{\,1,2,3,4\,\} \;\subseteq\; \Omega \;\; ,$$

und es tritt ein, wenn beim nächsten Wurf ein $\omega \in A$ gewürfelt wird.

In der Wahrscheinlichkeitstheorie sorgt man durch ein paar Axiome dafür, dass sich diese Funktion $\mathbb{P}$ so verhält wie die „Wahrscheinlichkeiten“, mit denen wir naiv umgehen. Der Vorteil ist, dass man mit dem Konstrukt $(\Omega, \mathbb{P})$ innerhalb der Mathematik ist, also auf der sicheren Seite, wohingegen in der realen Umwelt eigentlich niemand letztlich sauber sagen kann, was Wahrscheinlichkeit ist und ob es so etwas wie Zufall überhaupt gibt. Diese Diskussion halten wir uns vom Halse. Die Anwendbarkeit der Wahrscheinlichkeitstheorie auf reale Phänomene steht jedoch völlig außer Frage; selbst der größte Skeptiker kann nicht leugnen, dass die Sache funktioniert. Diese Anwendbarkeit von Mathematik auf Umweltphänomene,

[3] Ich muss mich hier etwas gewunden ausdrücken; wenn das Ω unendlich groß ist, kann man nicht jeder Teilmenge von Ω eine Wahrscheinlichkeit zuordnen.

der wir ja immer wieder begegnen, ist, bei Licht betrachtet, eine sehr rätselhafte Angelegenheit.

Wir werden den Gebrauch der Begriffe an Beispielen einüben – passe mit auf, dass das auch ausreichend geschieht. Zum Schluss jetzt noch eine Warnung: Die saubere Konstruktion des Ω kann eine sehr aufwendige Angelegenheit sein, die führt man nicht immer korrekt durch. Das spart Arbeit, klar, aber man kann böse auf die Nase fallen, wenn es das Ω und das $\mathbb{P}$ dazu, mit dem man rechnet, gar nicht gibt.

1.5 Laplace–Versuche

Wir wollen einen Würfel dreimal würfeln, das ist unser Zufallsversuch. Wie sieht der Ergebnisraum dazu aus? Wir können die Ergebnisse des Versuchs angeben, indem wir die drei geworfenen Zahlen aufzählen, und zwar in genau der Reihenfolge, in der sie geworfen wurden. Wenn dies die Zahlen $a_1, a_2, a_3 \in \{1, 2, 3, 4, 5, 6\}$ waren, können wir sie als **Wort** $a_1a_2a_3$ notieren – dann müssen wir freilich aufpassen, dass wir das Wort nicht mit dem Produkt der Zahlen verwechseln – oder als 3–Tupel oder Tripel $(a_1|a_2|a_3)$, dann haben wir so etwas gebildet wie den Satz Koordinaten eines Raumpunktes.

Ich nehme mal die Wortschreibweise. Dann ist

$$\Omega \ = \ \{\, a_1a_2a_3 \,|\, a_1, a_2, a_3 \in \{1, 2, 3, 4, 5, 6\}\} \quad .$$

Die **Mächtigkeit** $|\Omega|$, also die Anzahl der Elemente, ist

$$|\Omega| \ = \ 6^3 \ = \ 216 \quad .$$

Klar? Allgemein gilt:

2 Lemma
Die Anzahl der Worte der Länge n aus einem Alphabet mit N Buchstaben ist N^n.

Unser Ω ist ziemlich groß, aber es hat einen entscheidenden Vorteil: es ist völlig symmetrisch. Soll heißen: alle Elemente sind gleichartig. Insbesondere haben alle Elementarereignisse die gleiche Wahrscheinlichkeit

$$\frac{1}{|\Omega|} \ = \ \frac{1}{216} \quad .$$

Für jedes Ereignis $A \subseteq \Omega$ ist dann

$$\mathbb{P}(A) \ = \ \frac{|A|}{|\Omega|} \quad .$$

3 Definition
Ein Zufallsversuch, bei dem jedes Ereignis $A \subseteq \Omega$ die Wahrscheinlichkeit

$$\mathbb{P}(A) \ = \ \frac{|A|}{|\Omega|}$$

hat, heißt **Laplace–Versuch**.

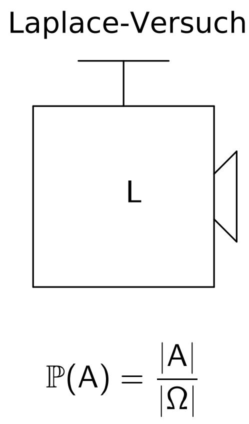


Abbildung 3: Spezieller Zufallsversuch Laplace–Versuch

Wer sich nur dafür interessiert, ob die gewürfelte Zahl eine Sechs war oder nicht, kommt mit einem kleineren Ergebnisraum Ω' aus:

$$\Omega' \;=\; \{\,*\,*\,*,\; 6\,*\,*,\; *6*,\; *\,*\,6,\; 66*,\; 6*6,\; *66,\; 666\,\}$$

Wo ein $*$ steht, kam eben eine andere Zahl als sechs.

Das kleinere Ω hat seinen Preis. Seine Ergebnisse haben nicht mehr alle die gleiche Wahrscheinlichkeit, es handelt sich nicht mehr um einen Laplace–Versuch. Zum Beispiel steht das Ergebnis $6\,*\,* \;\in \Omega'$ für das Ereignis

$$A \;=\; \{\,6ab \,|\, a,b \in \{1,2,3,4,5\}\,\} \;\subseteq\; \Omega$$

des alten Ω, und es ist

$$\mathbb{P}(A) \;=\; \frac{|A|}{|\Omega|} \;=\; \frac{25}{216} \quad .$$

Berechnen wir zum Schluss die Wahrscheinlichkeit dafür, dass mindestens einer der drei Würfe eine Sechs war. Jeder Mensch–ärgere–dich–nicht–Spieler ist brennend an dem Ergebnis interessiert, nicht wahr. Nein, das Ergebnis ist nicht einfach $\frac{1}{2}$, wie manche von euch fix ausgerechnet haben[4]. Man macht es am einfachsten so: Man definiert das Ereignis

$$B \;=\; \{abc \in \Omega \,|\, a,b,c < 6\,\}$$

der Ergebnisse, in denen **keine** Sechs auftrat. Nach dem Lemma ist $|B| = 5^3$, also

$$\mathbb{P}(B) \;=\; \frac{|B|}{|\Omega|} \;=\; \frac{5^3}{6^3} \quad .$$

Die gesuchte Wahrscheinlichkeit ist dann $1 - \mathbb{P}(B) = 1 - \frac{125}{216}$. Das haben wir nun mit dem **Gegenereignis**

$$\overline{B} \;:=\; \{\,\omega \in \Omega \,|\, \omega \notin B\,\}$$

gerechnet, es ist ja stets $\mathbb{P}(\overline{B}) = 1 - \mathbb{P}(B)$. Die Menge $\overline{B}$ heißt das **Komplement** von B in Ω. Du siehst, der Sprachkurs läuft noch eine Weile nebenher.

[4] Denn das bedeutete ja, dass die Wahrscheinlichkeit für mindestens eine Sechs bei achtzehn Würfen 3 wäre. Und immer wenn du Wahrscheinlichkeiten herausbekommst, die größer als 1 sind, hast du Mist gebaut. Nur schütte das Kind nicht mit dem Bade aus: oft helfen Fehler zu fruchtbaren Einsichten. Was mag der eben berechnete Wert $\frac{1}{2}$ bedeuten?

1.6 Zufallsgrößen

Um den armen Til zu schonen, erkläre ich euch diesen Begriff an dem Beispiel des progessiven Lehrers. Wie findet er seine Noten? Er würfelt dreimal und bestimmt dann die kleinste gewürfelte Augenzahl.

Der natürliche Ergebnisraum zum Zufallsversuch „dreimal würfeln“ ist

$$\Omega \;=\; \{\, abc \,|\, a,b,c \in \{1,2,3,4,5,6\}\,\} \;\;.$$

Das Ergebnis $\omega = 255$ führt zur Note 2. Der Mathematiker sieht hier eine Funktion X am Werk, die jedem $\omega \in \Omega$ eine reelle Zahl $X(\omega)$ zuordnet, und er nennt diese Funktion X eine **Zufallsgröße**.

4 Definition
Die Menge Ω sei Ergebnisraum eines Zufallsversuchs. Eine Zufallsgröße X auf Ω ist dann eine Abbildung $X : \Omega \to \mathbb{R}$.

Die Menge der Werte, die X annimmt, bezeichnen wir kurz mit $X(\Omega)$. Damit du dich mit der Terminologie anfreundest, schreibe ich das feierlich formal hin:

$$X(\Omega) \;:=\; \{\, X(\omega) \,|\, \omega \in \Omega\,\} \tag{1}$$

Im Lehrerbeispiel ist $X(\Omega) = \{1,2,3,4,5,6\}$. Wir wollen für jeden dieser Werte die Wahrscheinlichkeit berechnen, mit der X den Wert annimmt. Für die Wahrscheinlichkeit, dass X den Wert $k \in X(\Omega)$ annimmt, schreibt man $\mathbb{P}(X = k)$. Das ist eigentlich nicht zulässig, die Abbildung $\mathbb{P}$ darf man ja nur auf Ereignisse anwenden, also auf Teilmengen von Ω. Man bringt die Sache in Ordnung, indem man sagt, dass $\mathbb{P}(X = k)$ nichts anderes bedeutet als $\mathbb{P}(X^{-1}(k))$ mit

$$X^{-1}(k) \;:=\; \{\, \omega \in \Omega \,|\, X(\omega) = k\,\} \;\;. \tag{2}$$

Klar? Vermutlich eher nicht. Sieh die Sache so: Durch das X wird die Wiese oder der Pfannkuchen, als den wir uns Ω vorstellen, in Bezirke eingeteilt. Zu jedem Wert k, den X annimmt, gehört ein Bezirk, und in dem sind alle die $\omega \in \Omega$, die von X auf eben dieses k abgebildet werden. Und für diesen Bezirk, in Wirklichkeit eine Teilmenge von Ω, hat man das Symbol $X^{-1}(k)$.

Im Lehrerbeispiel ist dann

$$X^{-1}(6) \;=\; \{666\}, \quad \text{also} \quad \mathbb{P}(X = 6) \;=\; \frac{1}{|\Omega|} \;=\; \frac{1}{216} \;\;,$$

das Ω gehört ja zu einem Laplace–Versuch. Die Wahrscheinlichkeit $\mathbb{P}(X = 4)$ ist schon schwieriger zu bestimmen. In statischer Sichtweise zählt man die 3–Tupel, die man mit den Zahlen aus $\{4,5,6\}$ bilden kann, das sind 3^3 Stück, und die, die man mit den Zahlen aus $\{5,6\}$ bilden kann, das sind 2^3. Also gibt es $3^3 - 2^3$ Tripel mit Zahlen aus $\{4,5,6\}$, die wirklich eine 4 enthalten. Die 4 ist auch jeweils die kleinste Zahl in diesen Tripeln, folglich ist

$$\mathbb{P}(X = 4) \;=\; \frac{3^3 - 2^3}{6^3} \;=\; \frac{19}{216} \;\;.$$

In gleicher Weise bestimmt man die übrigen $\mathbb{P}(X = k)$.

Wenn dir diese Zählmethode zu obskur ist, kann ich dir noch eine andere anbieten. Berechnen wir noch einmal $\mathbb{P}(X = 4)$. Die Note 4 gibt es, wenn mindestens eine vier und sonst nur Zahlen > 4 geworfen wurden. Schauen wir auf die **erste** gewürfelte Vier. Die kann im ersten Wurf erscheinen, das gibt

$$1 \cdot 3 \cdot 3 \;=\; 9$$

Tripel[5], oder im zweiten Wurf, das gibt

$$2 \cdot 1 \cdot 3 \;=\; 6$$

Tripel, oder erst im dritten Wurf, das sind noch einmal

$$2 \cdot 2 \cdot 1 \;=\; 4$$

Tripel. Und $9+6+4$ sind wieder 19.

Wenn man alle $\mathbb{P}(X=k)$ berechnet hat, schreibt man sie in eine Tabelle, und nennt diese die Wahrscheinlichkeitsverteilung von X. Sie sieht hier so aus:

k	1	2	3	4	5	6
$\mathbb{P}(X=k)$	$\frac{91}{216}$	$\frac{61}{216}$	$\frac{37}{216}$	$\frac{19}{216}$	$\frac{7}{216}$	$\frac{1}{216}$

Und natürlich möchte man eine solche Wahrscheinlichkeitsverteilung graphisch darstellen. Dazu zeichnet man ein **Histogramm**. Die $X(\omega)$ markeirt man auf der Rechtsachse oder Merkmalsachse, und dann stellt man auf jedes $k \in X(\Omega)$ ein Rechteck, dessen Flächeninhalt $\mathbb{P}(X=k)$ ist. So sieht das Ding dann aus:

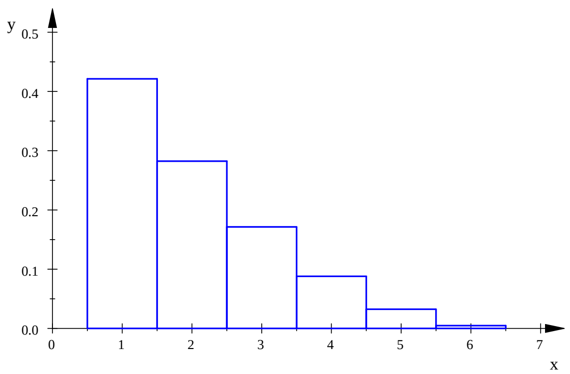


Abbildung 4: Histogramm der Zufallsgröße „minimale Augenzahl bei dreimal Würfeln“

[5] Halt, nimm diese Zahl nur ja nicht einfach mal drei, das ist der Elefantenfehler!

1.7 Der Erwartungswert einer Zufallsgröße

Lost man eine lange Reihe von Ergebnissen $x_1, x_2, x_3, \ldots$ einer Zufallsgröße X aus – dabei ist x_k der X–Wert des Ergebnisses der k–ten Versuchsdurchführung – stabilisiert sich der Mittelwert

$$\frac{1}{n}\sum_{k=1}^{n} x_k$$

dieser X–Werte bei wachsendem n. Würfelt man zum Beispiel sehr oft und bildet dann das arithmetische Mittel der geworfenen Zahlen, bekommt man (bei einem ordentlichen Würfel) immer einen Wert, der nahe bei 3.5 liegt. Dies ist ein empirischer Befund, er gehört eigentlich zum Empirischen Gesetz der großen Zahl. Wir bilden nun eine entsprechende Größe für unsere abstrakten Zufallsgrößen, nämlich den Erwartungswert der Zufallsgröße. Den können wir in unserem innermathematischen Reich der Freiheit einfach definieren. Es sollte sich natürlich beim Gebrauch herausstellen, dass unsere Definition einen brauchbaren Begriff festlegt.

5 Definition
Es sei $X : \Omega \to \mathbb{R}$ eine Zufallsgröße mit endlichem Ω, also mit $|\Omega| < \infty$. Dann definieren wir den **Erwartungswert** *von X durch*

$$\mathbb{E}(X) \;:=\; \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\}) \;=\; \sum_{k\in X(\Omega)} k\cdot\mathbb{P}(X=k) \quad .$$

Die beiden Summen in der Definition sind gleich. Für Beweise ist die erste Summe besser, für praktische Rechnungen meist die zweite, weil sie weniger Summanden hat als die erste Summe.

Aufgaben

1. Es sei X die gewürfelte Augenzahl eines idealen Würfels. Zeichne das Histogramm von X, berechne den Erwartungswert und markiere ihn in deinem Histogramm.

2. Wir würfeln zweimal mit einem idealen Würfel, und es sei X die Summe der gewürfelten Zahlen. Zeichne das Histogramm von X und berechne den Erwartungswert.

3. Die Zufallsgröße X nehme den Wert 1 mit der Wahrscheinlichkeit p und den Wert 0 mit der Wahrscheinlichkeit $q := 1-p$ an. Zeichne das Histogramm von X für $p = \frac{1}{4}$ und berechne für allgemeines p den Erwartungswert.

4. Denke dir einen Zufallsversuch aus, der zu dem X aus der vorigen Aufgabe passt.

5. Beim Spiel „Sechs verliert“ wird mit n Würfeln geworfen. Tritt dabei eine Sechs auf, ist der X–Wert des Ergebnisses Null. Tritt keine Sechs auf, ist der X–Wert die Summe der gewürfelten Augenzahlen. Wir nehmen zunächst $n = 2$. Zeichne das Histogramm von X und berechne den Erwartungswert.

6. Zugunsten des Mensaneubaus wird im Foyer eine Würfelbude eingerichtet. Es wird mit einem Würfel gewürfelt. Bei einer Sechs gibt es einen Euro, jeder Wurf kostet fünfundzwanzig Cent. Es sei X der bei einem Spiel ausgezahlte Betrag und Y der Gewinn des Spielers bei einem Spiel. Berechne die Erwartungswerte von X und von Y.

1.8 Stetig verteilte Zufallsgrössen

Es sei X der Zerfallszeitpunkt eines bestimmten radioaktiven Teilchens, das zum Zeitpunkt 0 noch existiert. Dann kann X jeden Wert ≥ 0 annehmen. Aber die Wahrscheinlichkeit, dass der Zerfall zu einem ganz konkreten vorgegebenen Zeitpunkt eintritt, etwa zur Zeit 2 oder $\sqrt{3}$ oder $\frac{1}{2}\pi$, ist 0.

Für solche Zufallsgrößen braucht man eine **Dichtefunktion**.

6 Definition
Es sei f eine Funktion mit $f(x) \geq 0$ für alle x und

$$\int_{-\infty}^{\infty} f(x)\,dx \;=\; 1 \quad .$$

Dann heißt f Dichtefunktion der Zufallsgröße X mit

$$\mathbb{P}(a \leq X \leq b) \;:=\; \int_a^b f(x)\,dx \quad .$$

Beispiel. Die Zufallsgröße X: „Zerfallszeitpunkt eines gegebenen radioaktiven Teilchens“ hat bei geschickter Wahl der Zeiteinheit die durch

$$f(x) \;=\; \begin{cases} e^{-x} & \text{für } 0 \leq x \\ 0 & \text{für } x < 0 \end{cases}$$

gegebene Dichtefunktion f.

1. Prüfe nach, dass f den Forderungen genügt, die an eine Dichtefunktion gestellt werden.
2. Berechne $\mathbb{P}(2 \leq X \leq 3)$.
3. Bestimmne T so, dass
$$\int_0^T f(x)\,dx \;=\; \frac{1}{2}$$
ist. Diese Zahl T heißt Halbwertszeit.
4. Bestimme den Funktionsterm der **Verteilungsfunktion** F mit
$$F(x) = \mathbb{P}(X \leq x)$$
und zeichne die Graphen von f und von F.

1.9 Die Verteilungsfunktion einer Zufallsgröße

Es sei X eine Zufallsgröße. Die Verteilungsfunktion von X ist dann die durch

$$F(x) \;:=\; \mathbb{P}(X \leq x) \tag{3}$$

definierte Funktion. Für jedes $x \in \mathbb{R}$ gibt sie die Wahrscheinlichkeit dafür an, dass X den Wert x oder einen kleineren annimmt.

Beispiel. Es sei X die Augenzahl beim Würfeln mit einem idealen Würfel und F die Verteilungsfunktion dazu. Dann ist $F(-2) = F(-1) = 0$, aber $F(1) = F(1.3) = \frac{1}{6}$, und es ist $F(x) = 1$ für alle $x \geq 6$. Der Graph von F sieht so aus:

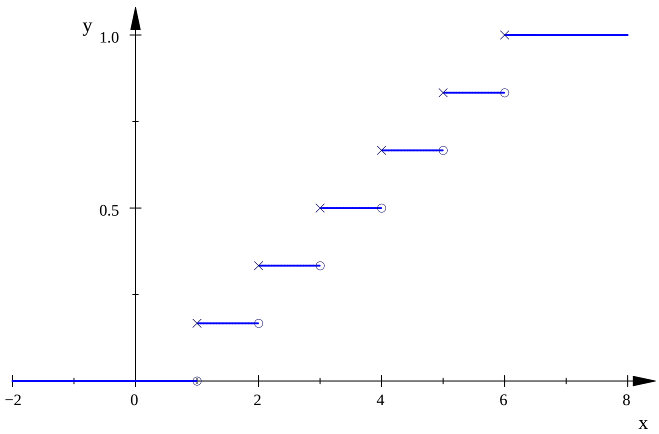


Abbildung 5: Verteilungsfunktion der Augenzahl bei Würfeln

Wir notieren gleich ein paar Eigenschaften der Verteilungsfunktion.

7 Lemma
Es sei X eine Zufallsgröße mit Verteilungsfunktion F. Dann ist F monoton steigend, das heißt für $x_1 < x_2$ ist stets $F(x_1) \leq F(x_2)$, und es gilt

$$F(x) \;\xrightarrow{x \to \infty}\; 1 \quad \textit{und} \quad F(x) \;\xrightarrow{x \to -\infty}\; 0 \ .$$

Falls X stetig verteilt ist mit Dichtefunktion f, gilt

$$F(x) \;=\; \int_{-\infty}^{x} f(t)\,dt \ . \tag{4}$$

Aus dieser Gleichung ergibt sich sofort, dass $F'(x) = f(x)$ ist; F ist ja eine Integralfunktion von f.

Beispiel. Es sei X stetig verteilt mit Dichtefunktion

$$f(x) \;=\; \begin{cases} 0 & \text{für } x < 0 \\ e^{-x} & \text{für } 0 \leq x \end{cases} \ .$$

Die Verteilungsfunktion dazu ist

$$F(x) \;=\; \begin{cases} 0 & \text{für } x < 0 \\ 1 - e^{-x} & \text{für } 0 \leq x \end{cases} \ .$$

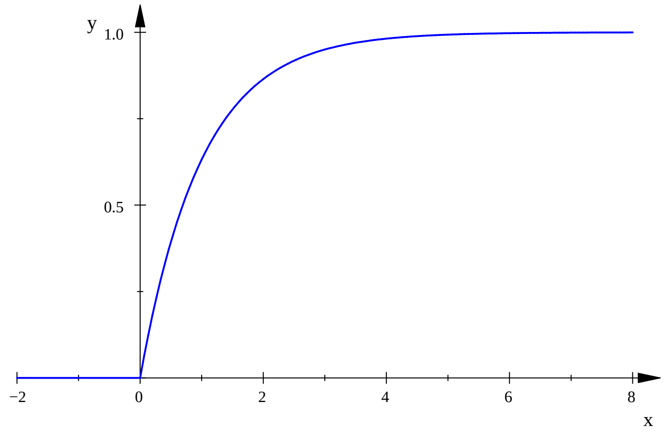


Abbildung 6: Verteilungsfunktion des Zerfallszeitpunktes eines radioaktiven Teilchens

Abbildung 6 zeigt den Graphen von F.

Aufgabe.

1. Der Graph der Dichtefunktion einer Zufallsgröße X ist der Polygonzug ABC mit $A(-1|\frac{1}{2})$, $B(0|\frac{1}{2})$ und $C(2|0)$. Bestimme $F(x)$ und zeichne den Graphen dazu.

2. Die Verteilungsfunktion einer Zufallsgröße X ist

$$F(x) \;:=\; \begin{cases} 0 & \text{für } x < 0 \\ x^3 & \text{für } 0 \leq x \leq 1 \\ 1 & \text{für } 1 < x \end{cases} \;.$$

   Bestimme die Dichtefunktion dazu.

1.10 Transformationen einer Zufallsgröße

Wir haben einen Ergebnisraum Ω und eine Funktion $\mathbb{P}$, die ordentlichen Teilmengen von $A \subseteq \Omega$ ihre Wahrscheinlichkeit $\mathbb{P}(A) \in [0,1]$ zuordnet. Von dem $\mathbb{P}$ verlangt man nur, dass $\mathbb{P}(\Omega) = 1$ ist und dass

$$\mathbb{P}(A \cup B) \;=\; \mathbb{P}(A) + \mathbb{P}(B) \tag{5}$$

ist für alle ordentlichen Teilmengen $A, B \subseteq \Omega$ mit $A \cap B = \emptyset$. Dieses Paar $(\Omega, \mathbb{P})$ ist dann ein Wahrscheinlichkeitsraum[6]; damit sind wir eigentlich begrifflich auf der Höhe der Zeit, bis auf den Umstand, dass ich nicht präzise sage, was eine „ordentliche" Teilmenge von Ω ist[7]. Durch diese Begriffsbildung schuf A. Kolmogorow in den dreißiger Jahren des vorigen Jahrhunderts die Wahrscheinlichkeitstheorie als mathematische Theorie; vorher war die Wahrscheinlichkeitsrechnung eher ein Teilgebiet der Physik, wie die Mechanik.

Aber ich wollte ja von Zufallsgrößen reden, also von Abbildungen X von Ω in $\mathbb{R}$. Damit eine Zufallsgröße sauber definiert ist, muss für jedes $a \in \mathbb{R}$ die Menge

$$\{\, \omega \in \Omega \mid X(\omega) \le a\}$$

eine ordentliche Menge sein, also ein Ereignis. In anderen Worten: Für jedes $a \in \mathbb{R}$ muss es die Wahrscheinlichkeit

$$\mathbb{P}(\, \{\omega \in \Omega \mid X(\omega) \le a \,\}\,)$$

geben, was nicht heißt, dass du in der Lage sein musst, sie auch konkret ausrechnen zu können. Hast du die Bedeutung dieser Wahrscheinlichkeit erkannt? Genau, es ist der Wert der $F(a)$ der Verteilungsfunktion von X.

So, damit du sicheren Stand behältst, zeige ich dir erst an einem einfachen Beispiel, worauf die Sache hinaus soll: Die Zufallsgröße X soll den Wert 1 mit der Wahrscheinlichkeit $p \in [0,1]$ annehmen und den Wert 0 mit der Wahrscheinlichkeit $q := 1 - p$. Den Erwartungswert von X kannst du im Kopf ausrechnen, aber ich zeige dir ein praktisches Schema dafür. Du legst eine Tabelle mit drei Spalten an:

k	$\mathbb{P}(X = k)$	$k \cdot \mathbb{P}(X = k)$
0	q	0
1	p	p
	1	p

Wenn du die Summe der Einträge der zweiten Spalte bildest, muss 1 herauskommen, und die Summe der Einträge der dritten Spalte ist der Erwartungswert $\mathbb{E}(X)$.

Die Zufallsgröße X mag der Betrag sein, den man an einem Stand mit einem Glücksrad an den Spieler auszahlt; der Gewinnsektor hat die Größe $p \cdot 360°$, der andere Sektor hat die Größe $q \cdot 360°$.

Will man nun den Auszahlungsbetrag verdoppeln, hat man eine neue Zufallsgröße Y. Aber diese neue Funktion auf Ω ist eigentlich nur $2X$:

$$Y \;=\; 2X \quad ,$$

[6] probability space

[7] Die ordentlichen Mengen sind die Ereignisse, also Teilmengen von Ω, die eine Wahrscheinlichkeit haben.

was nur heißen soll, dass

$$Y(\omega) \; = \; 2X(\omega) \quad \text{ist für alle } \omega \in \Omega.$$

Was ist mit dem Erwartungswert? Es gehört kein großer Scharfsinn dazu, zu der Vermutung zu kommen, dass

$$\mathbb{E}(Y) \; = \; \mathbb{E}(2X) \; = \; 2\mathbb{E}(X)$$

ist.

Zweitens kann man nach dem zu erwartenden Gewinn des Spielers je Spiel fragen. Beträgt der Einsatz b Euro, hat man die neue Zufallsgröße

$$G \; = \; X - b$$

mit

$$G(\omega) \; = \; X(\omega) - b \quad \text{für alle } \omega \in \Omega.$$

Natürlich ist

$$\mathbb{E}(G) \; = \; \mathbb{E}(X) - b \quad .$$

Obwohl dir das vermutlich völlig klar ist, formulieren wir den Sachverhalt als Lemma und beweisen ihn.

8 Lemma
Es sei $X : \Omega \to \mathbb{R}$ eine Zufallsgröße mit Erwartungswert $\mathbb{E}(X)$. Für $a, b \in \mathbb{R}$ bilden wir die neuen Zufallsgrößen

$$\begin{aligned} Y \; &:= \; aX \quad \textit{mit } Y(\omega) := aX(\omega) \textit{ für alle } \omega \in \Omega \textit{ und} \\ Z \; &:= \; X + b \quad \textit{mit } Z(\omega) := X(\omega) + b \textit{ für alle } \omega \in \Omega. \end{aligned}$$

Dann gilt

$$\mathbb{E}(aX) \; = \; a\mathbb{E}(X) \quad \textit{und} \quad \mathbb{E}(X+b) = \mathbb{E}(X) + b \quad .$$

Beweis. Nach der Definition des Erwartungswertes ist

$$\mathbb{E}(aX) \; = \; \sum_{\omega \in \Omega} aX(\omega)\mathbb{P}(\{\omega\}) \; = \; a \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}) \; = \; a\mathbb{E}(X)$$

und

$$\begin{aligned} \mathbb{E}(X+b) \; &= \; \sum_{\omega \in \Omega} (X(\omega) + b)\, \mathbb{P}(\{\omega\}) \\ &= \; \sum_{\omega \in \Omega} (X(\omega)\mathbb{P}(\{\omega\}) + b\mathbb{P}(\{\omega\})) \\ &= \; \left(\sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}) \right) + \sum_{\omega \in \Omega} b\mathbb{P}(\{\omega\}) \\ &= \; \mathbb{E}(X) + b \cdot 1 \quad , \end{aligned}$$

denn das b in der zweiten Summe können wir ausklammern, und die Summe über alle $\mathbb{P}(\{\omega\})$ ist natürlich 1. $\square$

Warnung. Was ist denn mit $\mathbb{E}(X^2)$? Wo alles so gut lief, denkt man vielleicht, das sei $(\mathbb{E}(X))^2$, aber das stimmt nicht. Bei unserem Beispiel ist X^2 die gleiche Zufallsgröße wie X, also ist $\mathbb{E}(X^2) = p$. Aber $(\mathbb{E}(X))^2 = p^2$.

1.11 Eine physikalische Interpretation des Erwartungswerts

Es sei X eine Zufallsgröße mit Erwartungswert $\mu := \mathbb{E}(X)$. Dann ist

$$\mathbb{E}(X - \mu) \; = \; \mathbb{E}(X) - \mu \; = \; 0 \quad ,$$

das haben wir gerade gelernt. Andererseits ist

$$\mathbb{E}(X - \mu) \; = \; \sum_{k \in X(\Omega)} (k - \mu)\mathbb{P}(X = k) \quad .$$

Schauen wir uns einen Summanden dieser Summe einmal genauer an. Der Wert $k - \mu$ gibt an, wie weit der X–Wert k von μ entfernt ist. Liegt k rechts von μ, ist der Wert positiv, liegt k links von μ, ist der Wert negativ. Der Wert wird nun noch mit der positiven Zahl $\mathbb{P}(X = k)$ multipliziert. Das erinnert an das Hebelgesetz, da rechnet man ja auch „Kraft mal Kraftarm“ , und das Ergebnis ist das Drehmoment, das eine Kraft bewirkt, die im Abstand $k-\mu$ vom Drehpunkt orthogonal zum Hebel angreift.

Wenn du Physik nicht magst, kannst du ja an eine Wippe denken. Insgesamt verhält sich die Sache so, als sei eine unendlich lange – natürlich gewichtslose – Stange mit Längenkala vorhanden, die horizontal liegt und die an der Stelle μ unterstützt wird. An jeder Stelle $k \in X(\Omega)$ wirkt eine Kraft der Größe $\mathbb{P}(X = k)$ auf die Stange , und zwar senkrecht nach unten. Wenn du die Rechtecke des Histogramms auf die Stange stellst, jedes aus einer großen homogenen Sperrholzplatte ausgesägt, hast du genau das Richtige gebastelt. Und es ist nun so, dass die ganze Angelegenheit im labilen Gleichgewicht ist: du weißt nicht, nach welcher Seite die Sache herunterfällt. Hier ist ein Beispiel mit $|X(\Omega)| = 2$:

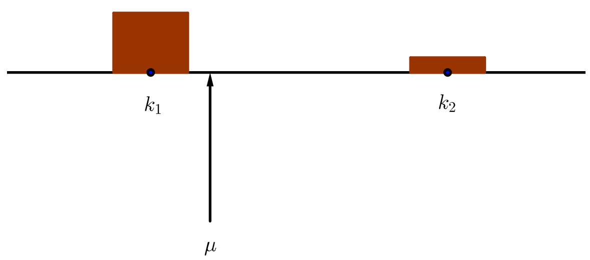


Abbildung 7: Beim Erwartungswert ist das Histogramm im Gleichgewicht

1.12 Die Varianz

Für einen Stand beim Schulfest, der Geld für den Mensaneubau einbringen soll, wird ein einfaches Glücksspiel gesucht. Der Einsatz soll einen Euro betragen, und jedes einzelne Spiel soll im Mittel der betreibenden Klasse einen halben Euro einbringen. Ein möglichst einfaches Spiel sieht so aus: Die Gewinnwahrscheinlichkeit des Spielers ist $p > 0$, und es werden im Gewinnfall $\frac{1}{2p}$ Euro an den Spieler ausgezahlt. Verliert er, bekommt er nichts. – Du kannst leicht ausrechnen, dass $\mathbb{E}(X) = \frac{1}{2}$ ist. Im Extremfall ist $p = 1$, der Spieler zahlt immer einen Euro und bekommt einen halben Euro zurück. Nur wird dann niemand spielen, der einen Funken Verstand hat. Man könnte aber etwa $p = \frac{1}{4}$ setzen, dann werden im Gewinnfall zwei Euro ausgezahlt, das sieht schon reizvoller aus.

Wer risikofreudig ist, nimmt $p = \frac{1}{1000000}$. Zwar muss man dann 500000 Euro an den Spieler zahlen, wenn er gewinnt, aber die Wahrscheinlichkeit, dass von den erwarteten einhundert Besuchern des Standes nicht ein einziger gewinnt, ist ja

$$\left(\frac{999999}{1000000}\right)^{100} \approx 0.9999 \quad ,$$

und das ist praktisch Null.

Hoffen wir einmal, dass die zuständigen Gremien die Genehmigung für diesen Stand versagen, und schauen wir uns an, was uns das Beispiel zeigt. Offensichtlich sagt der Erwartungswert $\mathbb{E}(X)$ allein noch nicht alles über X aus. Wir brauchen auch ein Maß dafür, wie weit die Werte von X von $\mu := \mathbb{E}(X)$ abweichen können. Nun ist

$$\sum_{k \in X(\Omega)} (k - \mu)\mathbb{P}(X = k) \;=\; 0 \quad ,$$

weil die (mit den Wahrscheinlichkeiten gewichteten) Abweichungen der k von μ sich insgesamt ausgleichen. Wir müssen den Abweichungen die Vorzeichen nehmen! Gauß hat vorgeschlagen, dies zu erreichen, indem man die $k - \mu$ quadriert, und dieses Streuungsmaß hat sich in Theorie und Praxis außerordentlich bewährt.

9 Definition
Es sei X eine Zufallsgröße mit Erwartungswert $\mu = \mathbb{E}(X)$. Dann ist die **Varianz** *von X definiert durch*

$$\mathbb{V}(X) \;:=\; \sum_{k \in X(\Omega)} (k - \mu)^2 \mathbb{P}(X = k) \;=\; \sum_{\omega \in \Omega} (X(\omega) - \mu)^2 \mathbb{P}(\{\omega\}) \quad .$$

Anmerkung. Die Varianz, Freunde, ist ein schwieriger Begriff, aber der wird euch noch vertraut werden. Zunächst ist das nur eine Kennzahl, die man nach der Definition ausrechnen kann, und je größer diese Kennzahl ist, desto mehr streuen die X–Werte um μ.

Aufgabe. Berechne die Varianz des Glücksspiels im Text oben (da kommt eine Funktion von p heraus) und die Varianz der Augenzahl beim Würfeln.

1.13 Der Verschiebungssatz

Die Varianz einer Zufallsgröße X mit dem Erwartungswert μ war die Zahl

$$\sum_{\omega\in\Omega}(X(\omega)-\mu)^2\cdot\mathbb{P}(\{\omega\}) \quad .$$

Schau mal genau hin, da steht eine Summe der Form

$$\sum_{\omega\in\Omega}\square(\omega)\cdot\mathbb{P}(\{\omega\}) \quad ,$$

und das ist doch genau die Formel für den Erwartungswert einer Zufallsgröße, nämlich der Zufallsgröße $\square:\Omega\to\mathbb{R}$ mit

$$\square(\omega) \;=\; (X(\omega)-\mu)^2 \quad \text{für alle } \omega\in\Omega.$$

Folglich ist

$$\mathbb{V}(X) \;=\; \mathbb{E}\left((X-\mu)^2\right) \quad .$$

Mit Hilfe unserer Rechenregeln für Erwartungswerte[8] formen wir die rechte Seite um:

$$\begin{aligned}\mathbb{E}\left((X-\mu)^2\right) &= \mathbb{E}(X^2-2X\mu+\mu^2)\\ &= \mathbb{E}(X^2)-2\mathbb{E}(X\mu)+\mu^2\\ &= \mathbb{E}(X^2)-2\mu\mathbb{E}(X)+\mu^2\\ &= \mathbb{E}(X^2)-2\mu^2+\mu^2\\ &= \mathbb{E}(X^2)-\mu^2\end{aligned}$$

Damit haben wir das folgende technische Lemma bewiesen:

10 Lemma (Verschiebungssatz)
Es sei X eine Zufallsgröße mit dem Erwartungswert μ und der Varianz $\mathbb{V}(X)$. Dann gilt

$$\mathbb{V}(X) \;=\; \mathbb{E}(X^2)-\mu^2 \quad .$$

Mit Hilfe dieses Lemmas kann man die Varianz einer Zufallsgröße recht bequem berechnen, man braucht nur die Tabelle zur Berechnung des Erwartungswertes auf Seite 15 um eine Spalte zu ergänzen. Ich führe es an diesem Beispiel vor:

k	$\mathbb{P}(X=k)$	$k\cdot\mathbb{P}(X=k)$	$k\cdot k\cdot\mathbb{P}(X=k)$
0	q	0	0
1	p	$1\cdot p$	$1\cdot 1\cdot p$
	1	p	p

Man musste nur die Einträge der dritten Spalte, also die $k\cdot\mathbb{P}(X=k)$, noch einmal mit dem jeweiligen k aus der ersten Spalte multiplizieren. Die Summe der dritten Spalte war der Erwartungswert $\mu=\mathbb{E}(X)$, die Summe der neuen Spalte ist $\mathbb{E}(X^2)$. Die Varianz ist nach dem Verschiebungssatz $\mathbb{E}(X^2)-\mu^2$, hier

$$\mathbb{V}(X) \;=\; p-p^2 \;=\; p(1-p) \;=\; pq \quad .$$

[8] Hier habe ich stillschweigend die neue Regel $\mathbb{E}(X+Y)=\mathbb{E}(X)+\mathbb{E}(Y)$ benutzt. Den Beweis lasse ich dir als Übungsaufgabe.

1.14 Die Ungleichung von Tschebyschew

Hast du die Defintion der Varianz einer Zufallsgröße X mit Erwartungswert μ vor Augen? Man bildet für jedes $k \in X(\Omega)$ den Ausdruck

$$(k-\mu)^2 \cdot \mathbb{P}(X=k)$$

und summiert über alle $k \in X(\Omega)$. Der Ausdruck hat eine anschauliche Bedeutung: $(k-\mu)^2$ ist der Flächeninhalt des Quadrats mit der Kantenlänge $|k-\mu|$. Als Beitrag des gerade betrachteten X–Werts k wird aber nicht der ganze Inhalt gezählt, sondern nur der Anteil $\mathbb{P}(X=k)$, also die Hälfte oder ein Viertel oder so etwas. Die Summe der Inhalte der ganzen Quadratteile über alle Werte k, die X annimmt, ist halt die Varianz.

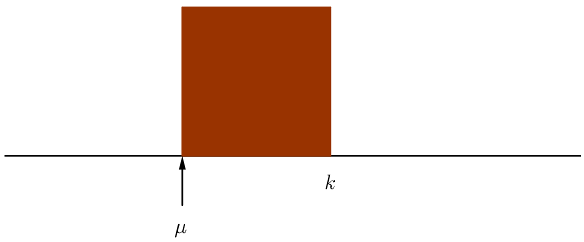


Abbildung 8: Jeder Wert $k \in X(\Omega)$ trägt den Anteil $\mathbb{P}(X=k)$ des Quadrates mit der Kantenlänge $|k-\mu|$ zur Varianz von X bei.

Nun begeben wir uns auf die Spuren des alten Pafnuti Lwowitsch Tschebyschew. Es sei a eine positive Zahl. Wir bilden wieder die Varianz von X, aber wir lassen die Beiträge aller $k \in X(\Omega)$, die echt zwischen $\mu-k$ und $\mu+k$ liegen, einfach weg, und bilden nur noch die Summe der übrigen k, also der k mit

$$|k-\mu| \;\geq\; a \quad .$$

Der Wert dieser Summe ist höchstens so groß wie die Varianz:

$$\sum_{\substack{k\in X(\Omega)\\ |k-\mu|\geq a}} (k-\mu)^2 \cdot \mathbb{P}(X=k) \;\leq\; \mathbb{V}(X)$$

Jedes der $(k-\mu)^2$ in unserer Summe, die jetzt noch da stehen, ist mindestens so groß wie a^2, da ja alle diese k von μ mindestens die Entfernung a haben. Wir ersetzen nun jedes $(k-\mu)^2$ durch a^2. Dadurch wird die Summe noch einmal kleiner, oder sie behält höchstens ihren Wert. Es gilt also

$$\sum_{\substack{k\in X(\Omega)\\ |k-\mu|\geq a}} a^2 \cdot \mathbb{P}(X=k) \;\leq\; \mathbb{V}(X) \quad .$$

Nun sind wir fast am Ziel. Wir klammern das a^2 aus und schauen uns die Summe genau an.

$$a^2 \sum_{\substack{k\in X(\Omega)\\ |k-\mu|\geq a}} \mathbb{P}(X=k) \;\leq \mathbb{V}(X) \quad .$$

Da werden die Wahrscheinlichkeiten aller $k \in X(\Omega)$ aufaddiert, die von μ mindestens den Abstand a haben. Das ergibt natürlich die Wahrscheinlichkeit dafür, dass ein X–Wert k kommt, der von μ mindestens den Abstand a hat:

$$\sum_{\substack{k \in X(\Omega) \\ |k-\mu| \geq a}} \mathbb{P}(X = k) \;=\; \mathbb{P}(|X - \mu| \geq a)$$

Wenn wir dies einsetzen, erhalten wir aus der letzten Ungleichung

$$a^2 \cdot \mathbb{P}(|X - \mu| \geq a) \;\leq\; \mathbb{V}(X) \quad .$$

Wir dividieren noch durch a^2, dann ist sie fertig, die hochberühmte **Ungleichung von Tschebyschew**:

11 Satz (Tschebyschew)
Es sei X Zufallsgröße mit Erwartungswert μ und Varianz $\mathbb{V}(X)$. Dann ist für jedes $a > 0$

$$\mathbb{P}(|X - \mu| \geq a) \;\leq\; \frac{\mathbb{V}(X)}{a^2} \quad .$$

Mit dem Satz können wir für jedes $a > 0$ sagen, wie groß die Wahrscheinlichkeit höchstens ist, dass (beim nächsten Versuch) ein X–Wert k ausgelost wird, der um mindestens a von μ entfernt ist, und das ist eine gute Sache.

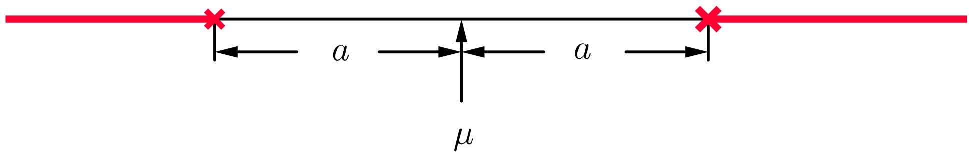


Abbildung 9: Die Wahrscheinlichkeit, dass X einen Wert annimmt, der von μ mindestens a entfernt ist, ist maximal $\frac{\mathbb{V}(X)}{a^2}$. Der Bereich dieser Werte ist der rot markierte Bereich der Merkmalsachse.

1.15 Die Standardabweichung

Ich hatte dir in Aussicht gestellt, dass dir die Bedeutung der Varianz klarer wird, wenn du erst die Ungleichung von Tschebyschew kennst. Es sei also X eine Zufallsgröße mit Erwartungswert μ und Varianz $\mathbb{V}(X)$. Die Tschebyschew–Ungleichung liefert dann eine obere Schranke für die Wahrscheinlichkeit $\mathbb{P}(|X - \mu| \geq a)$, den Erwartungswert um mindestens $a > 0$ zu verfehlen. Diese obere Schranke ist

$$\frac{\mathbb{V}(X)}{a^2} \quad ,$$

man nennt sie auch das **Tschebyschew–Risiko**. Es ist – bei festem X – eine Funktion von a.

Nun weiß jedes Kind, dass eine Wahrscheinlichkeit immer höchstens 1 ist. Falls also das Tschebyschew–Risiko größer als 1 ist, sagt uns die Tschebyschew–Ungleichung nur, was ohnehin jedes Kind weiß. Brauchbar ist sie folglich nur für

$$a \;\geq\; \sqrt{\mathbb{V}(X)} \quad .$$

Diese Größe $\sqrt{\mathbb{V}(X)}$ heißt die Standardabweichung von X, und sie wird gewöhnlich mit dem Buchstaben σ bezeichnet:

$$\sigma \;:=\; \sqrt{\mathbb{V}(X)} \quad \text{heißt } \textbf{Standardabweichung} \text{ von } X. \tag{6}$$

Schauen wir uns nun den Graphen der Funktion $a \mapsto \frac{\mathbb{V}(X)}{a^2}$ an.

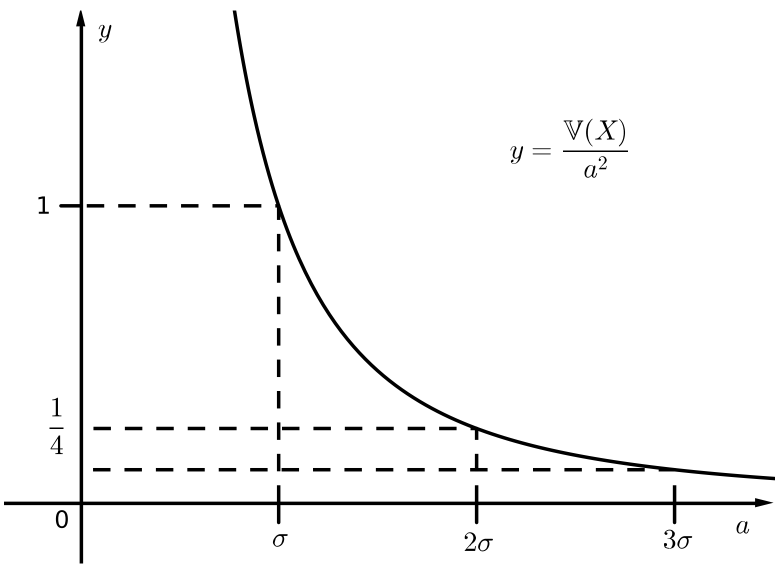


Abbildung 10: Tschebyschew–Risiko $\frac{\mathbb{V}(X)}{a^2}$

Wenn wir von einer Zufallsgröße X die Varianz $\mathbb{V}(X)$ kennen, können wir Folgendes sagen: **Die Wahrscheinlichkeit, den Erwartungswert um mindestens 2σ zu verfehlen, ist maximal $\frac{1}{4}$, die Wahrscheinlichkeit, den Erwartungswert um mindestens 3σ zu verfehlen, ist maximal $\frac{1}{9}$.** Werte von X, die – gemessen in σ – weit vom Erwartungswert weg liegen, können also nur eine kleine Wahrscheinlichkeit haben. Die Varianz ist ein sehr wertvolles Maß für die Streuung der Werte von X, wir werden sie alle Tage brauchen.

1.16 Erwartungswert und Varianz stetig verteilter Zufallsgrößen

12 Definition
Es sei X stetig verteilt mit Dichtefunktion f. Dann definiert man den Erwartungswert von X durch

$$\mathbb{E}(X) \;:=\; \int_{-\infty}^{\infty} x f(x)\,dx$$

und, falls $\mathbb{E}(X) := \mu$ existiert, die Varianz durch

$$\mathbb{V}(X) \;:=\; \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \quad .$$

Die Definitionen sind eigentlich ganz einleuchtend. Die Wahrscheinlichkeit, ein $x \in \mathbb{R}$ genau zu treffen, ist $= 0$, damit kann man nichts tun. Aber die Wahrscheinlichkeit, dass der nächste ausgeloste Wert von X in einem kleinen Intervall der Breite Δx um x liegt, ist etwa $f(x)\Delta x$. Um den Erwartungswert zu bekommen, bildet man eine Summe der Form

$$\sum_k x_k f(x_k)(\Delta x)_k \quad ,$$

und daraus wird für $\Delta x \to 0$ das Integral aus der Definition. Ich hoffe, nun ist dir die Definition halbwegs plausibel. Mehr kann man nicht erreichen, eine Definition kann man nicht „beweisen".

1.17 Klausur am 9. März 2015

1. Da wir mit Stochastik gerade erst angefangen haben, verwenden wir einen Kinderwürfel mit zwei Einsen, zwei Zweien und zwei Dreien. Den wollen wir, solange nichts anderes gesagt ist, zweimal werfen.

   (a) Schreibe einen brauchbaren Ergebnisraum Ω für den Versuch hin. [3]

   (b) Auf dieses Ω setzen wir die Zufallsgrößen U: „Unterschied[9] der beiden Zahlen“ und M: „kleinste der beiden Zahlen“.

      i. Gib für ein $\omega \in \Omega$ deiner Wahl die Zahlen $U(\omega)$ und $M(\omega)$ an. [2]

      ii. Schreibe die folgenden Mengen hin. [6]

      $$U^{-1}(0), \quad M^{-1}(2), \quad U^{-1}(0) \cap M^{-1}(2), \quad U^{-1}(0) \cup M^{-1}(2)$$

      iii. Was für eine Art Ding ist eine Zufallsgröße überhaupt? [2]

      iv. Wie ist $\mathbb{E}(U)$ definiert, und was bedeutet der Wert praktisch? [6]

      v. Jens zahlt an Uwe einen Euro, falls U einen ungeraden Wert annimmt; andernfalls zahlt Uwe an Jens einen Euro. Ist das Spiel fair? [8]

   (c) Jens will den Würfel so lange werfen, bis eine Eins kommt.

      i. Mit welcher Wahrscheinlichkeit braucht er dafür genau k Würfe? [6]

      ii. Die Wahrscheinlichkeit dafür, dass er es nicht schafft, eine 1 zu würfeln, will Jens auf höchstens 10^{-6} drücken, also auf ein Millionstel. Wie oft muss er dann bereit sein zu würfeln? [6+]

      iii. Wie oft muss Jens im Mittel würfeln, bis er seine Eins hat? Schreibe eine Formel hin und äußere eine begründete Vermutung, wie groß der Wert sein sollte, aber versuche nicht, ihn exakt auszurechnen. [0+]

   (d) Jens will seinen Würfel 1200–mal werfen. Wieviele Einsen wird er wohl beobachten, und wie groß ist die Wahrscheinlichkeit höchstens, dass sich die beobachtete Anzahl von diesem Sollwert um mindestens 50 unterscheidet? Du kannst ohne Nachweis $1200 \cdot \frac{1}{3} \cdot \frac{2}{3}$ als Wert der Varianz benutzen. [8]

   (e) Jens will endlich groß sein und einen Sechserwürfel haben. Kann er nicht einen solchen simulieren, indem er die Augenzahlen seiner zwei Dreierwürfel addiert? [6]

   (f) Zeige allgemein, dass für Zufallsgrößen X und Y auf Ω stets

   $$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

   ist. Verwende dazu die Summe über Ω in der Definition von $\mathbb{E}$. [8]

[9] Der Unterschied ist stets ≥ 0.

2. Die Zufallsgröße X sei durch die folgende Tabelle beschrieben:

k	$\mathbb{P}(X = k)$
0	$0,7$
1	$0,1$
2	$0,2$

(a) Berechne den Erwartungswert und die Varianz, zeichne das Histogramm, bestimme die Verteilungsfunktion $F : x \mapsto \mathbb{P}(X \leq x)$ und zeichne ihren Graphen. [24]

(b) Zeichne den Graphen der Funktion $a \mapsto \mathbb{P}(|X - \mu| \geq a)$ und skizziere flüchtig den Graphen des Tschebyschew–Risikos $\frac{\mathbb{V}(X)}{a^2}$ dazu. [8]

3. In der Großen Schulreformkommission wird überlegt, dass Noten ermittelt werden, indem viermal mit einem idealen Würfel gewürfelt wird und die kleinste gewürfelte Augenzahl die Note ergibt – so soll die Abiturientenquote auf OECD–Maß erhöht werden. Wie groß ist dann noch die Wahrscheinlichkeit für eine nicht–ausreichende Leistung? [12]

4. Durch

$$f(x) = \begin{cases} 0 & \text{für } x < 1 \\ \frac{1}{x^2} & \text{für } 1 \leq x \end{cases}$$

ist die Dichtefunktion einer stetig verteilten Zufallsgröße X gegeben. Skizziere den Graphen von f und den der Verteilungsfunktion F dazu, berechne

$$\mathbb{P}(n \leq X \leq n + 1) \quad \text{für } n \in \{1, 2, 3, \dots\}$$

und schaue mal nach dem Erwartungswert – da wartet eine kleine Überraschung auf dich. [24]

1.18 Die Anzahl der k–Teilmengen einer n–Menge

Wir müssen abzählen können, wieviele Pfade vorgegebenen Typs es in einem großen Baumdiagramm gibt, und dabei ist ein besonderes Hilfsmittel nützlich.

13 Definition
Eine Menge mit n Elementen nennen wir kurz eine n–Menge, und wir bezeichnen die Anzahl aller k–Teilmengen einer n–Menge mit dem Symbol

$$\binom{n}{k} \quad ,$$

gelesen „n über k“.

Nun bin ich mit der Tür ins Haus gefallen. Damit du folgen kannst, ein Beispiel: Unser Kurs ist eine 18–Menge, und ein dreiköpfiger Aussschuss zur Planung der Kursfahrt wäre eine 3–Teilmenge davon. Es gibt insgesamt $\binom{18}{3}$ mögliche Ausschüsse.

Natürlich müssen wir ausrechnen können, wieviele das sind. Das gehen wir auf einem kleinen Umweg an. Wir verteilen Ämter in dem Ausschuss. Natürlich brauchen wir einen Vorsitzenden, einen Schriftführer und einen, der für Tee und Kekse zuständig ist. Wieviele Ausschüsse dieser Art es gibt, ist leicht abzuzuzählen: den Vorsitzenden können wir aus achtzehn wählen, den Schriftführer noch aus siebzehn und den Tee– und Kekswart noch aus sechzehn Kandidaten; das macht insgesamt

$$18 \cdot 17 \cdot 16$$

mögliche Ausschüsse.

Nun fragen wir, wieviele dieser Ausschüsse wir aus den gleichen drei Personen bilden können. Das ist leicht: den Vorsitzenden wählen wir aus dreien, den Schriftführer aus zweien, und der letzte ist dann der Tee– und Kekswart. Das macht

$$3 \cdot 2 \cdot 1$$

Ausschüsse aus den gleichen drei Personen. Gut, das heißt aber doch, dass jeweils $3 \cdot 2 \cdot 1$ der $18 \cdot 17 \cdot 16$ geordneten Ausschüsse aus den gleichen drei Personen bestehen. Folglich gibt es insgesamt

$$\frac{18 \cdot 17 \cdot 16}{3 \cdot 2 \cdot 1}$$

ungeordnete Dreierausschüsse, und das sind genau die 3–Teilmengen der 18–Menge.

Diese Überlegung lässt sich leicht auf den allgemeinen Fall übertragen. Es gilt das folgende Lemma.

14 Lemma
Für die Anzahl $\binom{n}{k}$ der k–Teilmengen einer n–Menge gilt

$$\binom{n}{k} = \frac{n(n-1)(n-2) \cdot \ldots \cdot (n-k+1)}{k(k-1)(k-2) \cdot \ldots \cdot 1} \quad .$$

Anmerkung. Es ist $\binom{n}{0} = \binom{n}{n} = 1$, denn es gibt genau eine 0–Teilmenge einer n–Menge, nämlich die leere Menge $\emptyset$, und genau eine n–Teilmenge einer n–Menge, nämlich die Menge selbst. Und für $n < k$ ist $\binom{n}{k} = 0$, denn keine Teilmenge einer n–Menge kann mehr als n Elemente haben.

Die Zahlen $\binom{n}{k}$ heißen **Binomialkoeffizienten**, weil sie in den binomischen Formeln vorkommen. Ich habe dir das Pascalsche Dreieck gezeigt, das einige von euch kannten, daran sieht man das. Ich gebe hier noch ein zusätzliches Argument. Wenn du

$$(a+b)^n \;=\; (a+b)(a+b)\cdot\ldots\cdot(a+b)$$

ausmultiplizierst, ohne zu vereinfachen oder zusammenzufassen, steht da die Summe über alle Elemente der Menge

$$\{p_1p_2p_3\cdot\ldots\cdot p_n \mid p_k \in \{a,b\} \text{ für } k=1,2,3,\ldots,n\} \quad ,$$

und du kannst das so sehen, als stamme der k–te Faktor p_k aus der k–ten Klammer $(a+b)$. Es gibt nur einen Summanden a^n, da musstest du aus jeder Klammer das a nehmen. Der nächste Summand hat die Form $a^{n-1}b$, und der tritt schon mehrfach auf. Du wählst aus **einer** Klammer das b und aus allen anderen das a. Für die b–Klammer hast du aber n Möglichkeiten, der Summand $a^{n-1}b$ tritt n–mal auf. Der nächste Summand ist $a^{n-2}b^2$. Wie oft kommt der vor? Aus **zwei** Klammern wählst du das b, aus den übrigen das a. Wenn du dir die Nummern der b–Klammern merkst, kannst du a–Klammern vergessen. Zu jeder Wahl gehört also genau eine 2–Teilmenge der n–Menge $\{1,2,3,\ldots,n\}$, nämlich die mit den beiden Nummern der b–Klammern; es gibt folglich genau so viele Summanden der Form $a^{n-2}b^2$, wie es 2–Teilmengen einer n–Menge gibt. Und so geht das weiter. Wenn du die Eins bei a^n und bei b^n als $\binom{n}{0}$ bzw. als $\binom{n}{n}$ liest, hast du insgesamt den **binomischen Satz**:

$$(a+b)^n \;=\; a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \ldots \;=\; \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k \tag{7}$$

Mit Hilfe einer Herde von n weißen und einem schwarzen Schaf haben wir noch eingesehen, dass

$$\binom{n}{k} + \binom{n}{k+1} \;=\; \binom{n+1}{k+1} \tag{8}$$

gilt. Die Binomialkoeffizienten passen also exakt zum Bauprinzip des Pascalschen Dreiecks.

1.19 Binomialverteilte Zufallsgrößen

Das kleinste Ω, mit dem man noch etwas anstellen kann, hat zwei Elemente. Wir nehmen nun ein solches und bezeichnen die Elemente mit 0 und 1:

$$\Omega \;:=\; \{0,1\}$$

Die 1 heißt gewöhnlich **Erfolg**, und ihre Wahrscheinlichkeit p. Die 0 heißt dann eben **Misserfolg**, und ihre Wahrscheinlichkeit $q := 1 - p$. Hier sind ein paar Beispiele für Zufallsversuche mit einem solchen Ω:

1. Man würfelt, interessiert sich aber nur dafür, ob eine Sechs gewürfelt wurde (Erfolg) oder nicht (Misserfolg). Hier ist $p = \frac{1}{6}$.
2. Man macht eine Umfrage, bei der man nur mit „ja“ oder „nein“ antworten kann: Isst du heute in der Mensa?
3. Man beantwortet eine Frage beim Känguruwettbewerb, richtige Antwort bedeutet Erfolg.
4. Man nimmt ein Produkt aus einer Massenfertigung, zum Beispiel einen Transistor, und prüft, ob er tauglich ist.
5. Man fragt einen Schüler, ob er seine Hausaufgaben hat.
6. Man wirft eine Reißzwecke und schaut, ob sie in der Lage $\perp$ landet.
7. Man beobachtet ein radioaktives Teilchen und schaut, ob es nach einer Stunde noch existiert.

Solche Versuche, bei denen es nur zwei Ergebnisse gibt, heißen **Bernoulliversuche**. Typischerweise wiederholt man sie n–mal **unabhängig**, der Versuch heißt dann **Bernoullikette der Länge n (mit der Erfolgswahrscheinlichkeit p)**. Dann geht man halt den Kurs durch und fragt jeden, ob er seine Hausaufgaben hat oder ob er heute in der Mensa isst, testet einen ganzen Karton Transistoren, wirft hundert Reißzwecken, würfelt 1200–mal oder bearbeitet einen ganzen Zettel Känguruaufgaben. Dagegen ist es **keine** Bernoullikette, wenn man aus einer Urne mit zwei Gewinnen und acht Nieten zweimal ein Los zieht und öffnet, denn die Gewinnwahrscheinlichkeit beim zweiten Ziehen hängt davon ab, was man beim ersten Ziehen hatte. Da sind die Versuche **nicht** unabhängig.

Anmerkung. Bei den praktischen Beispielen darfst du nicht zu kritisch sein. Vielleicht ist es für den Fachdezernenenten aus Detmold eine Bernoullikette, wenn er den Kurs durchgeht und jeden nach seinen Hausaufgaben fragt. Ich käme im Traum nicht darauf, bei jedem von euch das gleiche p zu setzen. Schütte das Kind nicht mit dem Bade aus; trotz gewisser Probleme wendet man Bernoulliketten in der Praxis erfolgreich an.

Gehen wir also von einer Bernoullikette der Länge n mit Erfolgswahrscheinlichkeit p aus. Das passende Ω dazu besteht aus allen Worten der Länge n mit Buchstaben aus dem Alphabet $\{0,1\}$:

$$\Omega \;=\; \{\, a_1a_2a_3\ldots a_n \,|\, a_1,a_2,a_3,\ldots,a_n \in \{0,1\}\,\} \;\;,$$

seine Mächtigkeit ist 2^n. Zu dem Ω passt in natürlicher Weise die Zufallsgröße X : „Anzahl der Erfolge“. Die Werte, die X annehmen kann, sind die Zahlen $0,1,2,\ldots,n$:

$$X(\Omega) \;=\; \{0,1,2,\ldots,n\} \;\;,$$

und erstaunlicherweise kann man sofort die Wahrscheinlichkeit $\mathbb{P}(X = k)$ für (genau) k Erfolge hinschreiben: In einem konkreten Wort $\omega = a_1 a_2 a_3 \dots a_n \in X^{-1}(k)$ stehen genau k Einsen und $n - k$ Nullen. Die Wahrscheinlichkeit, dass es ausgelost wird, ist folglich ein Produkt, in dem k–mal der Faktor p und $(n-k)$–mal der Faktor $q = 1 - p$ steht. Die Faktoren kann man umstellen, also ist

$$\mathbb{P}(\{\omega\}) \ = \ p^k q^{n-k} \quad \text{für jedes } \omega \in X^{-1}(k).$$

Jetzt brauchen wir nur noch die Mächtigkeit von $X^{-1}(k)$. Nun, jedes Wort ist charakterisiert durch die Menge

$$\{i_1, i_2, i_3, \dots, i_k\}$$

der Platznummern, an denen die Einsen stehen. Es gibt also genau so viele dieser Worte in $X^{-1}(k)$, wie es k–Teilmengen der n–Menge $\{1, 2, 3, \dots, n\}$ gibt, nämlich $\binom{n}{k}$. Also gilt

$$\mathbb{P}(X = k) \ = \ \binom{n}{k} p^k q^{n-k} \quad \text{für } k = 0, 1, 2, \dots, n. \tag{9}$$

Eine Zufallsgröße, bei der man die Wahrscheinlichkeiten nach der Formel in Gleichung (9) berechnet, heißt übrigens **binomialverteilt**, genauer $B(n, p)$–verteilt.

Leider, Jungs, hat die Sache einen kleinen Haken, und auf den stoßt ihr sofort, wenn ihr euch vertrauensvoll an euren Taschenrechner wendet, um ein solches $\mathbb{P}(X = k)$ konkret auszurechnen. Zwar hat euer Taschenrechner eine Extrataste nCr, mit der du $\binom{n}{k}$ ausrechnen kannst, das klappt aber nur für ziemlich kleine n oder für k, die nahe bei 0 oder n liegen, sonst geht er schnell in die Knie. Das kommt daher, dass die Potenzen $p^k q^{n-k}$ schnell sehr klein und die Binomialkoeffizienten $\binom{n}{k}$ dazu schnell gigantisch groß werden. Will man über diese Formeln gar noch den Erwartungswert

$$\mathbb{E}(X) \ = \ \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}$$

von X ausrechnen, läuft man in ein numerisches Desaster.

Was den Erwartungswert angeht, ist die Sache leicht zu heilen. Du hast in der Klausur doch schon im Kopf $\mathbb{E}(X)$ ausgerechnet für ein $B(1200, \frac{1}{3})$–verteiltes X. Der Trick ist, die Zufallsgröße X als Summe einfacher Zufallsgrößen zu schreiben. Es sei also X_1 eine $B(1, \frac{1}{3})$–verteilte Zufallsgröße, also die Anzahl der Einsen, die man bekommt, wenn man einen Kinderwürfel einmal wirft. Von X_1 verschaffen wir und 1199 unabhängige Kopien $X_2, X_3, \dots, X_{1200}$. Dabei bedeutet X_k schlicht die Anzahl der Einsen beim k–ten der 1200 Würfe. Die Gesamtzahl X der Einsen ist nun die Summe der X_k:

$$X \ = \ \sum_{k=1}^{1200} X_k$$

In jedem der Würfe bekommt man im Mittel $\frac{1}{3}$ Einsen, oder, falls dir das lieber ist: im Mittel ist jeder dritte Wurf eine Eins. Wenn dir das einleuchtet, wirst du keine Probleme haben zu akzeptieren, dass

$$\mathbb{E}(X) \ = \ \mathbb{E}\left(\sum_{k=1}^{1200} X_k\right) \ = \ \sum_{k=1}^{1200} \mathbb{E}(X_k) \ = \ \sum_{k=1}^{1200} \frac{1}{3} \ = \ 1200 \cdot \frac{1}{3}$$

ist. – Nach diesem Muster sieht man ein, dass $\mathbb{E}(x) = np$ ist für jedes $B(n, p)$–verteilte X.

1.20 Regeln für Varianz und Erwartungswert

Was ich brauchte, Jungs, wäre so etwas wie der folgende Satz.

15 Satz
Es seien X_1 und X_2 zwei von einander unabhängige Zufallsgrößen, die Erwartungswert und Varianz besitzen. Dann gilt für die Summe $X_1 + X_2$ der Zufallsgrößen

$$\mathbb{E}(X_1 + X_2) \;=\; \mathbb{E}(X_1) + \mathbb{E}(X_2) \quad \textit{und} \quad \mathbb{V}(X_1 + X_2) \;=\; \mathbb{V}(X_1) + \mathbb{V}(X_2) \;\;.$$

Ich werde mal nachfragen, ob ihr mit der Aussage des Satzes Probleme habt oder ob ihr die so akzeptiert. Wir brauchen in jedem Fall für die praktische Arbeit Folgerungen aus diesem Satz, nämlich:

16 Satz
Es sei X eine $B(n,p)$–verteilte Zufallsgröße. Dann ist

$$\mathbb{E}(X) \;=\; np \quad \textit{und} \quad \mathbb{V}(X) \;=\; np(1-p) \;\;.$$

Um diese Resultate aus dem Satz 15 zu folgern, braucht man nur den Summentrick am Ende des letzten Kapitels; das ist keine große Sache. Aber der Satz! Wenn ich den mit euch ordentlich beweisen will, muss ich ganz schön was hochfahren. Das geht damit los, dass jede der beiden Zufallsgrößen X_1 und X_2 auf ihrem eigenen Ω definiert ist: $X_1 : \Omega_1 \to \mathbb{R}$ und $X_2 : \Omega_2 \to \mathbb{R}$. Und für die Summe $X_1 + X_2$ muss ich aus Ω_1 und Ω_2 ein neues Ω basteln: Die neue Zufallsgröße $X + Y : \Omega \to \mathbb{R}$ ist auf

$$\Omega \;:=\; \{\omega_1\omega_2 \,|\, \omega_1 \in \Omega_1,\, \omega_2 \in \Omega_2\,\}$$

definiert. Jedes der drei Ω hat sein eigenes $\mathbb{P}$ auf dem passenden System von Teilmengen, und man muss sauber definieren, wie man aus den gegebenen Teilmengensystemen von Ω_1 und Ω_2 ein brauchbares System auf Ω erhält und wie man deren Wahrscheinlichkeiten berechnet; dabei wird dann auch der etwas nebulöse Begriff der Unabhängigkeit klar. Und wenn das alles steht, kann man an die Rechnung gehen … Vielleicht mache ich das an einem konkreten Beispiel mit einer Teilmenge eurer 18–Menge, vielleicht überprüfen wir Satz 16 mit MuPADs Hilfe für einige konkrete Beispiele, mal sehen. Du kannst ja mal darüber nachdenken:

Aufgabe. Nimm $\Omega_1 = \Omega_2 = \{0,1\}$. Sowohl $X_1 : \Omega_1 \to \mathbb{R}$ als auch $X_2 : \Omega_2 \to \mathbb{R}$ mögen den Wert 1 mit der Wahrscheinlichkeit p annehmen (beide sind folglich $B(1,p)$–verteilt). Bilde das passende Ω, lege für die Elementarereignisse in Ω die passenden Wahrscheinlichkeiten fest (das geht nur auf eine Weise!) und lege eine Tabelle an, die folgende Spalten hat:

ω	$(X_1+X_2)(\omega)$	$\mathbb{P}(\{\omega\})$	$(X_1+X_2)(\omega)\mathbb{P}(\{\omega\})$	$(X_1+X_2)(\omega)^2\mathbb{P}(\{\omega\})$

Berechne dann mit Hilfe der Tabelle Erwartungswert und Varianz von $X_1 + X_2$.

Zur Entspannung gibt es noch die Regel, für die Tim einen Beweis und für die Sjard ein plausibles Argument hat: Für eine Zufallsgröße X mit Varianz $\mathbb{V}(X)$ und Zahlen $a, b \in \mathbb{R}$ gilt

$$\mathbb{V}(X+b) \;=\; \mathbb{V}(X) \quad \text{und} \quad \mathbb{V}(aX) \;=\; a^2\mathbb{V}(X) \;\;. \tag{10}$$

Das Histogramm von X wird nur verschoben oder gestreckt, sagt Sjard, und das ist auch so.

1.21 Bedingte Wahrscheinlichkeit und Unabhängigkeit

Wir gehen von einem Ergebnisraum Ω und zwei Teilmengen $A, B \subseteq \Omega$ aus. Hier ist eine Abbildung dazu; der Pfannkuchen steht für Ω; die Menge A ist der Teil von Ω oberhalb der horizontalen Linie, die Menge B ist der Teil von Ω links der vertikalen Linie.

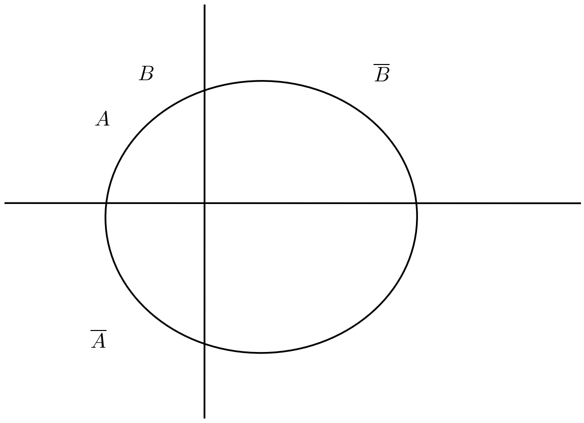


Abbildung 11: Menge Ω mit zwei Teilmengen A und B

Beispiel. Um die Sache konkret zu machen, nehmen wir eine dir wohlbekannte 18–Menge als Ω. Die Teilmenge A besteht aus denen, die regelmäßig Hausaufgaben machen, die Menge B aus denen, die eine gute Note bekommen haben. Dann ist $\mathbb{P}(A)$ die Wahrscheinlichkeit, dass ein zufällig gewählter Schüler regelmäßig seine Hausaufgaben macht. – Wir fragen nun in diesem Beispiel nach $\mathbb{P}(A \cap B)$. Das ist nicht weiter schwierig, man bildet den Anteil von $A \cap B$ an Ω:

$$\mathbb{P}(A \cap B) \;=\; \frac{|A \cap B|}{|\Omega|}$$

Im allgemeinen Fall geht das nicht so einfach, aber mit einem simplen Baumdiagramm bekommt man die Sache in den Griff – siehe Abbildung 12.

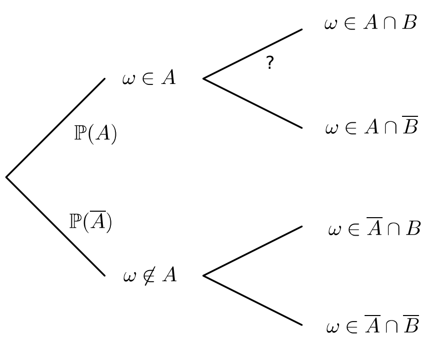


Abbildung 12: Baumdiagramm zur Berechnung von $\mathbb{P}(A \cap B)$

Aber welche Wahrscheinlichkeit gehört an den Zweig, der von $\omega \in A$ nach oben rechts zu $\omega \in \mathbb{P}(A \cap B)$ führt? Die Wahrscheinlichkeit, dass ein Element von B

ausgelost wird, aber es wird nur unter den Elementen von A gewählt; A ist sozusagen ein neues Ω.

Man nennt das, was an den Zweig gehört, die **bedingte Wahrscheinlichkeit** $\mathbb{P}(B|A)$, dass ein Element von B ausgelost wird, wenn nur aus den Elementen von A gewählt wird. Für $\mathbb{P}(A) \neq 0$ definiert man

$$\mathbb{P}(B|A) \ := \ \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \quad . \tag{11}$$

Das passt. Die gesuchte Wahrscheinlichkeit $\mathbb{P}(A \cap B)$ bekommt man nun, indem man die Wahrscheinlichkeiten längs des entsprechenden Pfades multipliziert:

$$\mathbb{P}(A \cap B) \ = \ \mathbb{P}(A) \cdot \mathbb{P}(B|A) \tag{12}$$

Schauen wir noch einmal zurück auf unser Beispiel. $\mathbb{P}(B|A)$ gibt den Anteil derer, die eine gute Note bekommen haben, an denen, die regelmäßig Hausaufgaben machen; dagegen bedeutet $\mathbb{P}(B)$ den Anteil derer in der ganzen 18–Menge, die eine gute Note bekommen haben. Es mag nun sein, dass diese Anteile gleich sind, dann ist A sozusagen ein verkleinertes Abbild von Ω. In diesem Fall wird aus Gleichung (12) diese:

$$\mathbb{P}(A \cap B) \ = \ \mathbb{P}(A) \cdot \mathbb{P}(B) \quad , \tag{13}$$

und man nennt A und B **unabhängig**. Die Gleichung (13) nimmt man als definierende Bedingung für Unabhängigkeit. Ob A und B in unserem Fall unabhängig sind, steht dahin; manchen käme es sicher gelegen.

Mit der Denkmöglichkeit $\mathbb{P}(B|A) = \mathbb{P}(B)$ in meinem Beispiel kann ich einige von euch ein wenig necken; das bedeutete ja, dass Arbeiten nichts bringt. Gesellschaftlich tragen diese bedingten Wahrscheinlichkeiten unter Umständen sehr brisante Botschaften. Nimm als Ω die Menge der Einwohner der Bundesrepublik, als B die Menge der Einwohner von Berlin und als A die Menge der Sozialhilfeempfänger, schaue dir die Einträge in der Vierfeldertafel an und berechne $\mathbb{P}(A|B)$ und $\mathbb{P}(B|A)$. Oder nimm als A die Menge der Frauen und als B die Menge der Aufsichtsräte eines DAX–Unternehmens. Klar, da ist $\mathbb{P}(A|B) < \mathbb{P}(A)$; du brauchst kein besonders aufmerksamer Zeitungsleser zu sein, um das zu wissen. Ein boshafter Mensch könnte sagen, stochastische Unabhängigkeit sei geradezu der Kern des Gerechtigkeitsbegriffs unserer Gesellschaft. Für großes Pathos ist da wenig Raum.

Aber genug davon. Überlege dir, was $\mathbb{P}(B|A)$ ist für $A \subseteq B$ und für $B \subseteq \overline{A}$. Theoriefreie Rechenaufgaben gibt es auf einem anderen Zettel.

1.22 Vom Wesen binomialverteilter Zufallsgrößen

Unser Hauptwerkzeug sind binomialverteilte Zufallsgrößen. Typischerweise geht es um die Anzahl der Erfolge bei einer Bernoullikette der Länge n mit Erfolgswahrscheinlichkeit p, diese Zufallsgröße bezeichnet man kurz als $B(n,p)$–verteilt. Wenn du ein ganz konkretes Beispiel brauchst: Würfle 600–mal und zähle die geworfenen Sechsen. Das entsprechende X ist $B(600, \frac{1}{6}$–verteilt.

Stellen wir zusammen, was wir schon über binomialverteilte Zufallsgrößen wissen: Es ist

$$\mathbb{P}(X = k) \;=\; \binom{n}{k} p^k (1-p)^{n-k} \quad \text{für } k = 0, 1, 2, \ldots, n$$

und es gilt

$$\mathbb{E}(X) = np \quad .$$

Ferner haben wir eine Vorstellung von der Varianz von X – so vorsichtig möchte ich das mal formulieren:

$$\mathbb{V}(X) \;=\; np(1-p)$$

Das Histogramm hat eine typische Buckelform. Das heißt: die Wahrscheinlichkeiten $\mathbb{P}(X = k)$ wachsen mit k etwa bis zum Erwartungswert, dann nehmen sie wieder ab – das rechnen wir heute exakt nach.

Weitere Einsichten sollst du an dem Bilderbuch auf den folgenden Seiten gewinnen. Du siehst an Abbildung 14 auf Seite 35, dass die Histogramme bei festem p mit wachsendem n immer breiter und niedriger werden. Das ist allerdings nicht die ganze Wahrheit. Schau, die Stärke der Streuung wird durch σ gemessen. Nach Tschebyschew ist die gesamte Wahrscheinlichkeit der k, die vom Erwartungswert μ um mindestens fünf σ abweichen, maximal $\frac{1}{25}$. Das σ wächst aber nicht mit n, sondern nur mit $\sqrt{n}$. Deshalb konzentriert sich die Wahrscheinlichkeit bei wachsendem n auf einen (relativ) immer schmaleren Bereich um den Erwartungswert. Das formulieren wir noch genauer.

Die zweite Abbildung zeigt Histogramme für festes $n = 20$ und verschiedene p–Werte. Für $p = \frac{1}{2}$ ist das Histogramm perfekt symmetrisch. Mit wachsendem n werden aber auch die anderen Histogramme tendenziell symmetrischer. Hinter den Formen der Histogramme steckt eine gemeinsame Urform. Sie wird sichtbar, wenn man die Zufallsgrößen so transformiert, dass das Breiterwerden herausgerechnet wird. Dies erreicht man, indem man von X zur zugehörigen **standardisierten** Zufallsgröße

$$Z \;:=\; \frac{X - \mathbb{E}(X)}{\sqrt{\mathbb{V}(X)}} \tag{14}$$

übergeht. Beispiele findest du in Abbildung 15 auf Seite 36. Für die neue Zufallsgröße Z gilt $\mathbb{E}(Z) = 0$ und $\mathbb{V}(Z) = 1$.

Man benutzt die Urform, um damit Näherungswerte der Wahrscheinlichkeiten $\mathbb{P}(X = k)$ zu berechnen; die exakte Formel ist leider zu unhandlich.

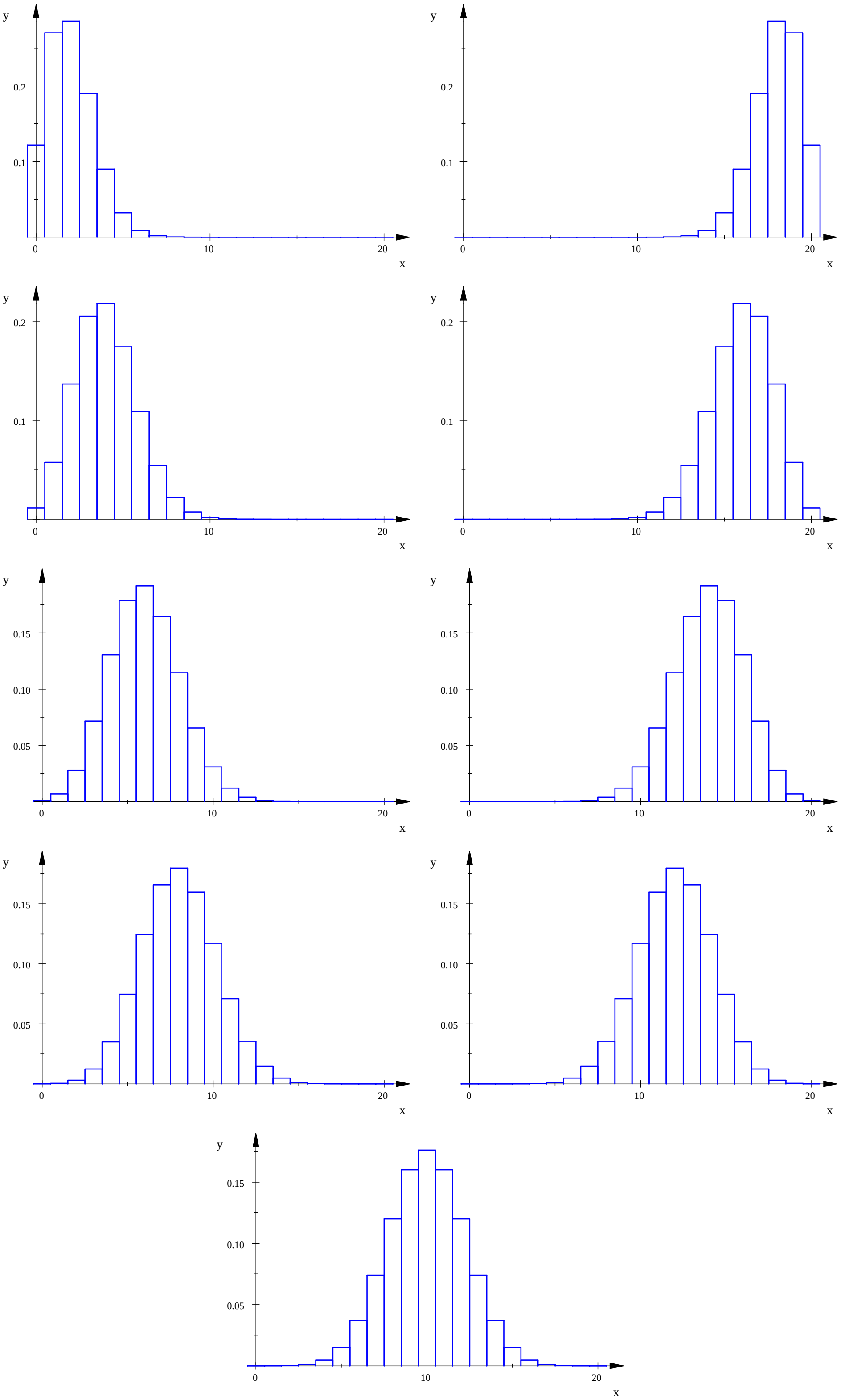


Abbildung 13: $B(20, p)$–verteiltes X, $p = \frac{1}{10}, \frac{2}{10}, \ldots, \frac{9}{10}$

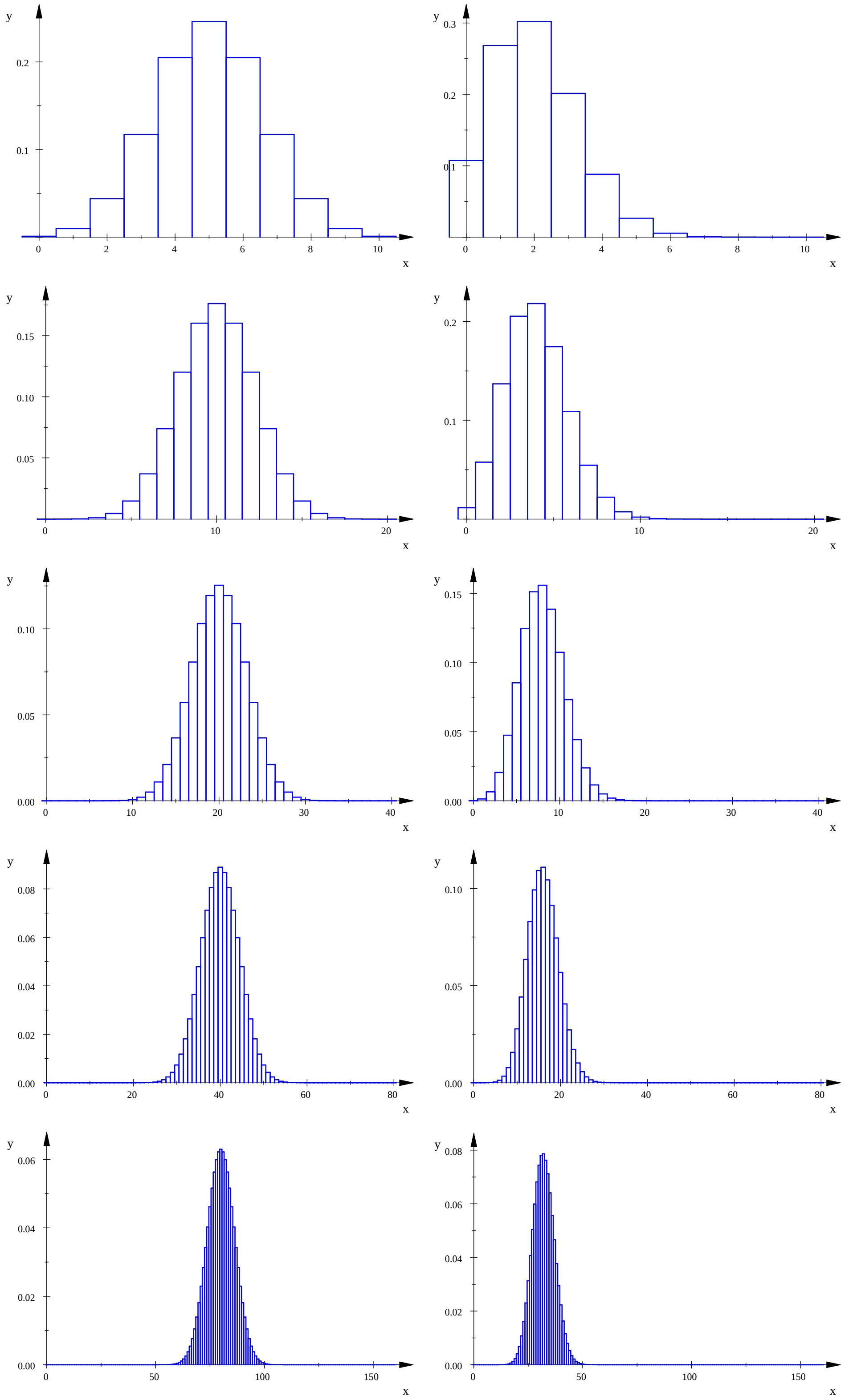


Abbildung 14: $B(n,p)$–verteiltes X, $p = \frac{1}{2}$ (links) und $p = \frac{1}{10}$ (rechts) für $n = 10, 20, 40, 80, 160$

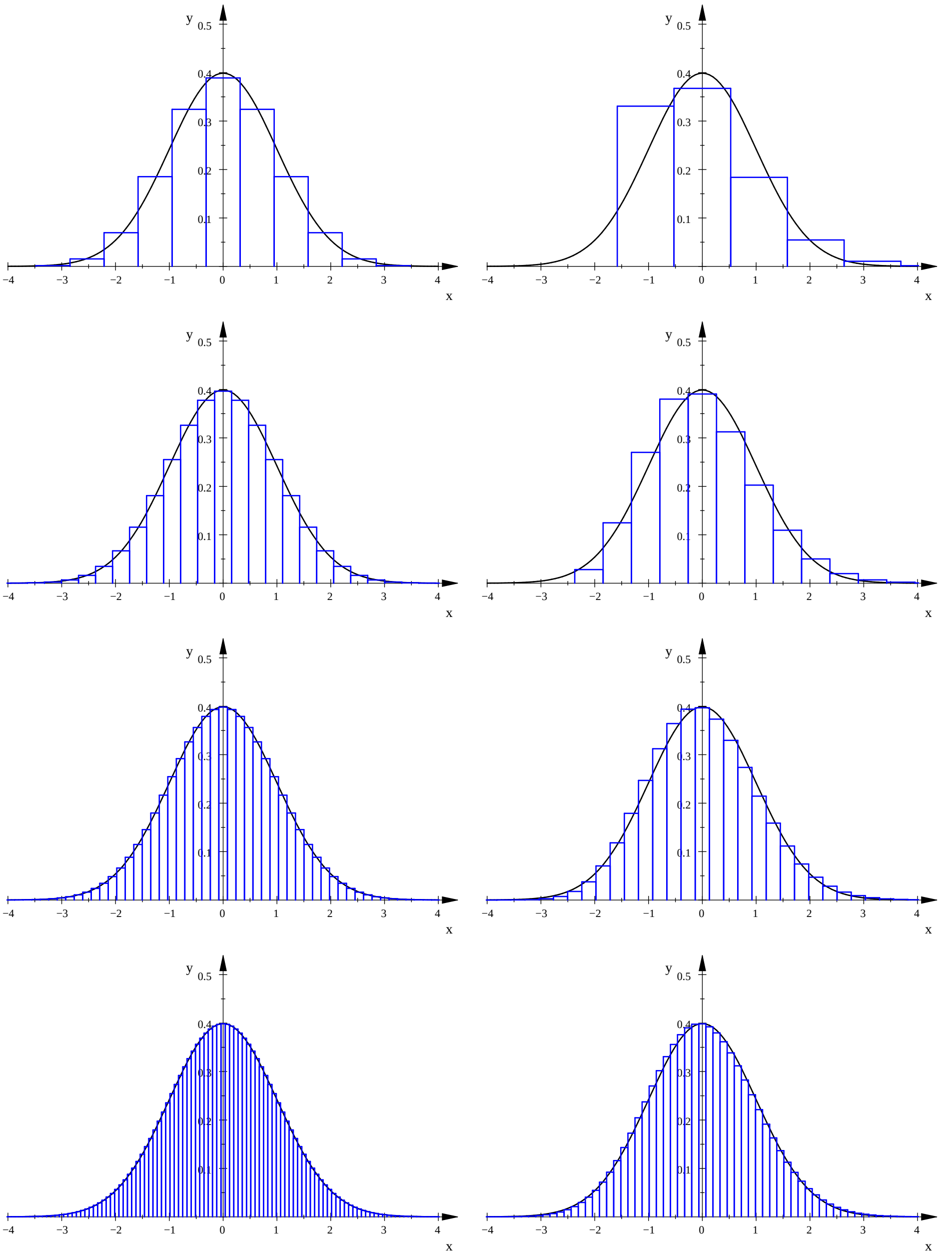


Abbildung 15: Standardisiertes $B(n,p)$–verteiltes X, $p = \frac{1}{2}$ (links) und $p = \frac{1}{10}$ (rechts) für $n = 10, 40, 160$ und 640 mit der Gaußkurve $y = \varphi(x)$

1.23 Ein Beispiel

Achtzehn Kunden beim Erlebniseinkauf – gleich werden sie nach und nach an der Kasse eintrudeln. M. hat noch ein wenig Zeit, seine Gedanken schweifen zu lassen. Vierzig Prozent der Kunden sollen eine Kundenkarte haben, behaupten die Marktforscher der Edekazentrale, das wären $0.4 \cdot 18 = 7.2$ Kunden. Geht ja gar nicht, klar, das ist auch nur der Erwartungswert der Zufallsgröße X: „Anzahl der Kunden mit Kundenkarte unter achtzehn zufällig gewählten“. In Wirklichkeit kann X jeden der Werte aus

$$X(\Omega) \;=\; \{0, 1, 2, \ldots, 18\}$$

annehmen, und die Wahrscheinlichkeit für den Wert k ist, wenn man die Daten der Marktforscher zu Grunde legt,

$$\mathbb{P}(X = k) \;=\; \binom{18}{k} \cdot 0.4^k \cdot 0.6^{18-k} \quad .$$

Die maximale Wahrscheinlichkeit hat das k im Intervall

$$[\,\mu - q\,,\ \mu + p\,] \;=\; [\,6.6\,,\ 7.6\,] \quad ,$$

also $k = 7$. M. hat das Histogramm klar vor Augen:

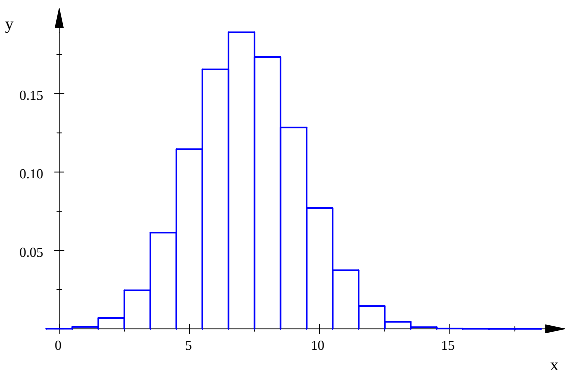


Abbildung 16: Histogramm des $B(18, 0.4)$–verteilten X

Bei dem kleinen $n = 18$ kann man alles noch bequem mit dem Taschenrechner ausrechnen, sonst müsste man die Zufallsgröße standardisieren. Wie ging das noch? Zuerst verschiebt man das Histogramm um μ nach links. Das ist keine große Sache, das Histogramm der neuen Zufallsgröße $X - \mu$ sieht genau so aus wie das alte[10]. Anschließend geht man noch von $X - \mu$ zu

$$Z \;:=\; \frac{X - \mu}{\sigma}$$

über, das ist dann die standardisierte Zufallsgröße zu X. Die Werte, die Z annimmt, sind die in

$$Z(\Omega) \;=\; \left\{ \frac{-\mu}{\sigma}, \frac{1-\mu}{\sigma}, \ldots, \frac{18-\mu}{\sigma} \right\} \quad .$$

Die Kästchenbreite im Histogramm von Z ist nicht mehr 1, sondern $\frac{1}{\sigma}$. Damit die Wahrscheinlichkeit immer noch der Flächeninhalt des Kästchens ist, muss die

[10] Siehe Abbildung 17 auf Seite 38

Höhe des Kästchens auf k nun $\sigma\mathbb{P}(X = k)$ sein. Das Histogramm von Z passt dann in etwa zur Gaußkurve

$$y = \varphi(x) \;=\; \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} \quad , \tag{15}$$

allerdings noch nicht besonders gut, wie Abbildung 18 zeigt.

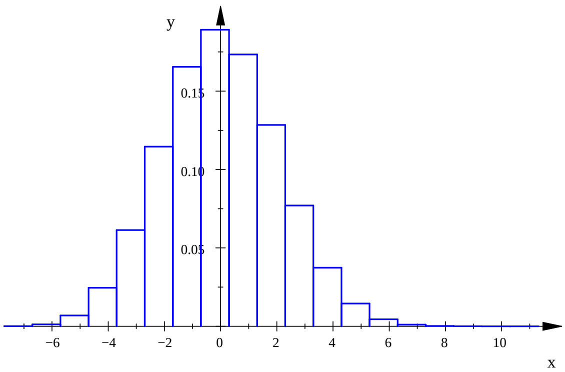


Abbildung 17: Histogramm von $X - \mu$

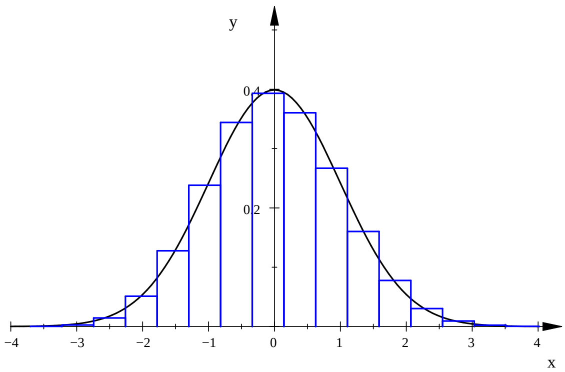


Abbildung 18: Histogramm von $Z = \frac{X-\mu}{\sigma}$

Nun sind die achtzehn Kunden bedient, und es hatten zwölf von ihnen eine Kundenkarte. M. überlegt, ob das wohl im Rahmen zufälliger Schwankungen liegt oder ob man bei der Edekazentrale wegen einer Bonuszahlung vorstellig werden sollte, weil in Espelkamp so Hervorragendes in Sachen Kundenbindung geleistet wird. Ein Maß dafür wäre die Wahrscheinlichkeit $\mathbb{P}(X \geq 12)$, M. rechnet die gleich aus. Und wenn sie klein genug ist, versucht er es vielleicht mit dem Bonus.

1.24 Die Näherungsformeln von de Moivre und Laplace

Viele Fragestellungen aus den Anwendungen laufen darauf hinaus, dass man für ein $B(n,p)$–verteiltes X Wahrscheinlichkeiten berechnen muss. Konkret sucht man

1. $\mathbb{P}(X = k)$ oder
2. $\mathbb{P}(X \leq k)$ oder
3. ein möglichst kleines k so, dass $\mathbb{P}(X \leq k)$ mindestens z.B. 90% ist, oder
4. ein möglichst kleines a, für das $\mathbb{P}(|X - \mu| \leq a)$ mindestens z.B. 90% ist.

Zwar haben wir eine exakte Formel für $\mathbb{P}(X = k)$, aber schon für sehr moderate n kann man die Formel nicht mehr numerisch auswerten. Hier benutzt man eine Beobachtung, die in Abbildung 15 auf Seite 36 an einer Reihe von Beispielen nachzuvollziehen ist: Das Histogramm der standardisierten Zufallsgröße $Z = \frac{X-\mu}{\sigma}$ zu X wird ganz gut durch die Gaußsche Glockenkurve angenähert. Diese hat die Gleichung

$$y \;=\; \varphi(z) \;:=\; \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}z^2} \quad . \tag{16}$$

Mit Hilfe dieser Funktion bekommt man leicht die **lokale Näherungsformel von de Moivre und Laplace**

$$\mathbb{P}(X = k) \;\approx\; \frac{1}{\sigma} \cdot \varphi\left(\frac{k-\mu}{\sigma}\right) \;=\; \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\cdot\left(\frac{k-\mu}{\sigma}\right)^2} \quad . \tag{17}$$

Denn diese Wahrscheinlichkeit ist der Flächeninhalt des Rechtecks auf k im Histogramm von X und gleichzeitig der Inhalt des Rechtecks auf $z = \frac{k-\mu}{\sigma}$ im Histogramm von Z. Die Höhe des Rechtecks im Histogramm von Z ist ungefähr $\varphi\left(\frac{k-\mu}{\sigma}\right)$, aber die Breite ist nur $\frac{1}{\sigma}$, so kommt der Term zustande.

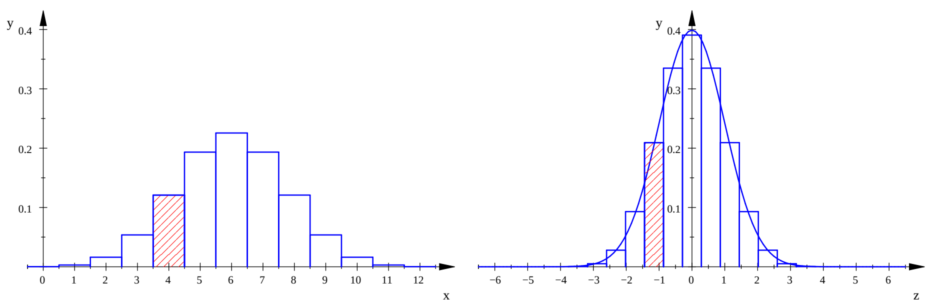


Abbildung 19: Histogramme der $B(12, \frac{1}{2})$–verteilten Zufallsgröße X (links) und der zugehörigen standardisierten Zufallsgröße Z (rechts). Die Rechtecke auf $k = 4$ und auf dem zugehörigen Wert $\frac{4-\mu}{\sigma}$ sind hervorgehoben, ihr Inhalt ist $\mathbb{P}(X = 4)$.

Auch die Wahrscheinlichkeit $\mathbb{P}(X \leq k)$ tritt als Flächeninhalt im Histogramm von Z auf. Die erste Säule steht auf $\frac{0-\mu}{\sigma}$, die letzte auf $\frac{k-\mu}{\sigma}$. Der Inhalt der Figur ist näherungsweise der Wert des Integrals $\int_a^b \varphi(z)\,dz$. Die Grenzen des Integrals sind

$$a \;=\; \frac{0-\mu-\frac{1}{2}}{\sigma} \quad \text{und} \quad b \;=\; \frac{k-\mu+\frac{1}{2}}{\sigma}$$

– der Summand $\frac{1}{2}$ verhindert dabei, dass ein halbes Rechteck abgeschnitten wird. Leider hat φ keine Stammfunktion, die man elementar hinschreiben könnte, deshalb verwendet man hier die Funktion

$$\Phi(z) \;:=\; \int_{-\infty}^{z} \varphi(t)\,dt \quad , \tag{18}$$

deren Werte tabelliert sind. Damit bekommt man die **integrale Näherungsformel von de Moivre und Laplace**

$$\mathbb{P}(X \le k) \;\approx\; \int_{-\infty}^{\frac{k-\mu+\frac{1}{2}}{\sigma}} \varphi(t)\,dt \;=\; \Phi\left(\frac{k-\mu+\frac{1}{2}}{\sigma}\right) \quad . \tag{19}$$

Anmerkungen

1. Der Kram sieht schwierig aus, er ist es aber nicht. Du musst zu den Wahrscheinlichkeiten **immer** die Flächen in den Histogrammen sehen, deren Inhalt die Wahrscheinlichkeiten sind, und zwar sowohl im Histogramm von X als auch in dem von Z.

2. Natürlich ist

$$\int_{-\infty}^{\infty} \varphi(z)\,dz \;=\; 1 \quad .$$

Das Histogramm von Z hat auch den Flächeninhalt 1, es erstreckt sich aber nur über einen endlichen Bereich der z–Achse. Dennoch wird das Histogramm von Z näherungsweise durch die Glockenkurve dargestellt. Das klappt, weil

$$\int_{-3}^{3} \varphi(z)\,dz \;\approx\; 0.997$$

ist: fast der gesamte Flächeninhalt 1 gehört zu dem Flächenstück unter dem Graphen von φ zwischen $z=-3$ und $z=3$. Die untere Grenze a liegt weit unterhalb von -3, deshalb liest du in der integralen Näherungsformel auch als untere Grenze $-\infty$. Der Wert von $\Phi\left(-\frac{\mu}{\sigma}\right)$ ist stets praktisch $=0$, wenn man die Näherungsformeln anwenden kann.

3. Die Grenze $-\infty$ macht Probleme, wenn du Werte von Φ mit der Integraltaste deines Taschenrechners ausrechnest. Da nimmst du besser als eine Grenze 0 und nutzt die Symmetrie der Glockenkurve aus, oder du benutzt die eingebaute Funktion Φ deines Taschenrechners.

Anwendungsbeispiel. Wir wollen nun für $B(2000, \frac{3}{5})$–verteiltes X die zu Beginn dieses Kapitels auf Seite 39 aufgeführten Werte berechnen. Es ist

$$\mu \;=\; \mathbb{E}(X) \;=\; 2000 \cdot \frac{3}{5} = 1200 \quad \text{und} \quad \sigma \;=\; \sqrt{\mathbb{V}(X)} \;=\; \sqrt{480} \quad .$$

1. Fragen wir nach $\mathbb{P}(X=1190)$.

$$\mathbb{P}(X=1190) \;\approx\; \frac{1}{\sqrt{480}} \cdot \varphi\left(\frac{1190-1200}{\sqrt{480}}\right) \;\approx\; 0.0164078$$

Gibt man den Term für den exakten Wert in MuPAD ein, bekommt man einen Bruch, der nicht in eine Zeile passt, und dafür den Näherungswert 0.01637476165.

2. Für $\mathbb{P}(X \le 1190)$ erhalten wir nach der integralen Näherungsformel

$$\mathbb{P}(X \le 1190) \;\approx \Phi\left(\frac{1190-1200+\frac{1}{2}}{\sqrt{480}}\right) \;\approx\; 0.3322845 \quad .$$

Für den exakten Wert gibt mir MuPAD auf meinem Rechner nach fünf Sekunden Rechenzeit wieder einen langen Bruch und dafür den dezimalen Näherungswert 0.3318413738 . Ohne Stetigkeitkorrekturliefert die integrale Näherungsformel übrigens das Ergebnis 0.3240384341 – da hat man schon eine merkliche Abweichung vom dezimalen Näherungswert des exakten Wertes.

3. Ein möglichst kleines k so, dass $\mathbb{P}(X \leq k) \geq 0.9$ ist, findet man zu Fuß nicht. Mit der integralen Näherungsformel ist das kein Problem. Zunächst bestimmt man z so, dass $\Phi(z) = 0.9$ ist:

$$\Phi(z) \;=\; 0.9 \quad\Leftrightarrow\quad z \;=\; \Phi^{-1}(0.9) \;\approx\; 1.281551566 \;\;,$$

dann macht man die Standardisierung rückgängig:

$$x \;=\; \mu + \Phi^{-1}(0.9) \cdot \sigma \;\approx\; 1228.077$$

Daraus ergibt sich, dass $k = 1228$ geeignet ist. Vorsicht, bei $x \approx 1228.7$ hätten wir $k = 1229$ nehmen müssen!

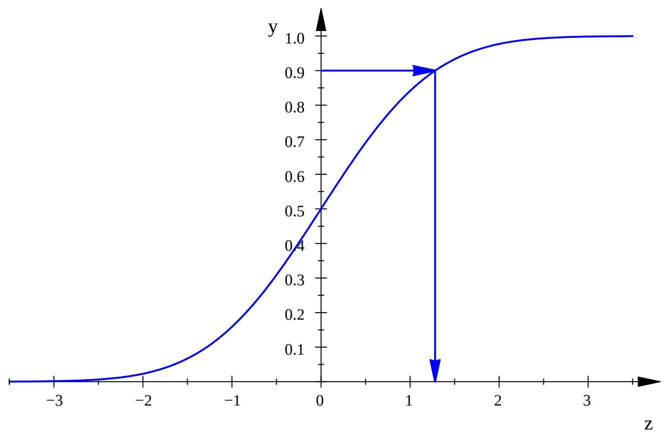


Abbildung 20: Man bestimmt $z = \Phi^{-1}(0.9)$, indem man Φ rückwärts liest.

4. Ein möglichst kleines a so, dass $\mathbb{P}(|X - \mu| \leq a) \geq 0.9$ ist, findet man auf diesem Wege: Zunächst bestimmt man z so, dass

$$\int_{-z}^{z} \varphi(t)\, dt \;=\; 0.9$$

ist. Das ergibt $z = \Phi^{-1}(0.95) \approx 1.6449$. Das gesuchte a ist dann

$$a \;=\; \phi^{-1}(0.95) \cdot \sigma \;\approx\; 36.037 \;\;.$$

Aufgaben

1. Mache dir das alles gut anhand der Histogramme und der Graphen von φ und Φ klar.

2. Berechne für $B(1600, \frac{3}{4})$–verteiltes X

$$\mathbb{P}(X = \mathbb{E}(X)) \;, \qquad \mathbb{P}(X = 1220) \quad \text{und} \quad \mathbb{P}(1205 \leq X \leq 1241)$$

und bestimme ein möglichst kleines a so, dass $\mathbb{P}(|X - \mu| \leq a)$ mindestens 80% ist.

3. Berechne $\int_{-a}^{a} \varphi(z)\, dz$ für $a \in \{1, 2, 3\}$ und überlege dir, was das Ergebnis für X bedeutet.

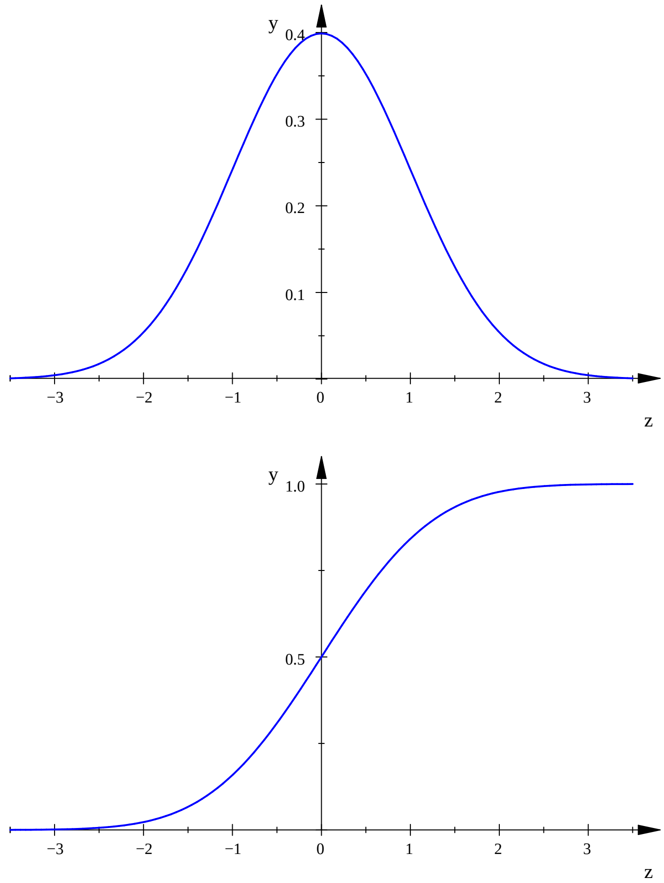


Abbildung 21: Graphen von φ und Φ

1.25 Normalverteilte Zufallsgrößen

Bei der Anwendung der Näherungsformeln von de Moivre und Laplace begegneten wir solchen Ausdrücken:

$$\mathbb{P}(Z \leq z) \;=\; \int_{-\infty}^{z} \varphi(t)\, dt$$

Dabei war das Z eine Zufallsgröße, die aus einem binomialverteilten X durch Standardisierung entstanden war, klar. Die Sache funktioniert aber auch, wenn wir Z als stetig verteilte Zufallsgröße und φ als ihre Dichtefunktion ansehen. Die so konstruierte Zufallsgröße Z heißt standardnormalverteilt. Der Name verrät es schon: normalverteilte Zufallsgrößen treten in den Anwendungen sehr häufig auf.

17 Definition
Es sei $\sigma > 0$ und $\mu \in \mathbb{R}$. Eine stetig verteilte Zufallsgröße X mit der Dichtefunktion[11]

$$f(x) \;=\; \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\left(\frac{x-\mu}{\sigma}\right)^2} \tag{20}$$

heißt **normalverteilt** *mit den Parametern μ und σ, kurz $N(\mu, \sigma^2)$–verteilt. Eine $N(0,1)$–verteilte Zufallsgröße heißt* **standardnormalverteilt**.

[11] Gib gut acht, das σ im Nenner der Vorfaktors steht **nicht** in der Wurzel.

Wenn man eine solche Definition hinschreibt, muss man natürlich sicherstellen, dass sie sinnvoll ist, also dass f auch wirklich als Dichtefunktion geeignet ist.

Aufgabe. Weise nach, dass die in Gleichung (20) definierte Funktion f den Forderungen genügt, die an eine Dichtefunktion gestellt werden.

Die Namen der Parameter sind mit Bedacht gewählt:

18 Lemma
Es sei X eine $N(\mu, \sigma^2)$–verteilte Zufallsgröße. Dann ist

$$\mathbb{E}(X) \;=\; \mu \qquad \textit{und} \qquad \mathbb{V}(X) = \sigma^2 \quad .$$

Der Beweis ist eine gute Integrationsübung. Ich weiß noch nicht, ob wir uns der jetzt unterziehen sollen.

Im Buch findest du eine ganze Reihe von Aufgaben. Der Eindruck täuscht nicht, normalverteilte Zufallsgrößen sind allgegenwärtig. Ich werde vorsichtig versuchen, ein wenig zu erklären, wie das kommt. Aber erst sollst du den Umgang mit normalverteilten Zufallsgrößen etwas trainieren.

2 Anwendungen

Etwas Wahrscheinlichkeitstheorie kennen wir nun, jetzt wollen wir sehen, wie man diese Theorie verwenden kann, in der Praxis vernünftige Entscheidungen zu treffen. Man spricht da von Beurteilender Statistik; in der Meditation unseres M. an seinem Arbeitsplatz klang schon ein wenig davon an. So ganz scharf zweigeteilt, wie du jetzt vielleicht denkst, ist unser Stochastikkurs aber nicht. Auch bisher kamen schon Anwendungen vor, und wir werden weiterhin unsere Theorie ausbauen.

2.1 Prognoseintervalle

Wieviele Sechsen wird es geben, wenn man 5000–mal würfelt? Wieviele von 5000 befragten Kunden eines Supermarktes haben eine Kundenkarte? Wieviele Schwarzfahrer sind unter 5000 Reisenden bei der Eurobahn? Wieviele von 5000 gesetzten Tulpenzwiebeln werden aufgehen, wenn bei jeder einzelnen die Wahrscheinlichkeit 93% beträgt, dass sie keimt?

Der Übervorsichtige, der Gürtel und Hosenträger gleichzeitig verwendet, wird jeweils sagen, zwischen Null und 5000. Damit hat er völlig recht, aber mit der Antwort kann man nichts anfangen. Ein Anderer prophezeit vielleicht, dass 4650 Tulpen aufgehen werden. Das ist eine präzise Vorhersage, aber in der großen Mehrzahl der Fälle wird sie nicht eintreffen.

Wer gar kein Risiko eingehen will, sich zu irren, kann sich nur der unfruchtbaren Haltung des Übervorsichtigen anschließen, und das wollen wir nicht. Wir wollen eine brauchbare Vorhersage machen, obwohl sie vielleicht nicht eintrifft. Was uns von gewöhnlichen Kaffeesatzlesern unterscheidet, ist, dass wir die Unsicherheit unserer Aussage quantifizieren können: wir nennen die Wahrscheinlichkeit dafür, dass unsere Vorhersage nicht eintritt. Einen Auftraggeber werden wir fragen, welche Unsicherheit er bereit ist, in Kauf zu nehmen, und ihm dann eine Auskunft geben, die zu seiner Risikofreude passt. Je präziser die Vorhersage ist, desto größer ist das Risiko, dass sie nicht eintrifft.

Möchte der Beauftragte für die Jubiläumsfeierlichkeiten im nächsten Jahr wissen, wieviele der 5000 Tulpen mit 99%-tiger Sicherheit aufgehen, können wir ihm zwei Antworten anbieten. Wir bilden

$$\Phi^{-1}(0.995) \;\approx\; 2.5758 \qquad \text{und} \qquad \Phi^{-1}(0.99) \;\approx\; 2.3263 \;\;,$$

berechnen mit $\sigma = \sqrt{5000 \cdot 0.93 \cdot 0.07} \approx 18.0416$ die Werte

$$\Phi^{-1}(0.995) \cdot \sigma \;\approx 46.47 \qquad \text{und} \qquad \Phi^{-1}(0.99) \cdot \sigma \;\approx\; 41.97$$

und sagen unserem Auftraggeber, dass mit 99%–iger Sicherheit zwischen 4604 und 4696 Tulpenzwiebeln beziehungsweise mindestens 4608 Tulpenziebeln keimen werden.

Aufgabe. In welchem theoretischen Modell wurden die angegebenen Zahlen berechnet? Verstehst du die Frage? Was ist das X, mit welchem Recht verwendet man es, und wie kommt man von X auf die Zahlen?

So, du hast nun eine Vorstellung davon, was ein Prognoseintervall ist. Es ist an der Zeit, eine ordentliche Definition zu geben.

19 Definition
Es sei X eine Zufallsgröße mit Erwartungswert μ und $w \in [0;1]$. Das **Prognoseintervall** *zur Sicherheitswahrscheinlichkeit w ist das kleinste Intervall $[\mu - a; \mu + a]$, für das*

$$\mathbb{P}(|X - \mu| \leq a) \;\geq\; w$$

ist.

Aufgabe. Bestimme das Prognoseintervall für die Anzahl der Sechsen, wenn man 5000–mal würfelt, zur Sicherheitswahrscheinlichkeit $w = 80\%$ und formuliere damit ein allgemeinverständliches Gutachten für den Auftraggeber, den Bundesverband der Mensch–ärgere–dich–nicht–Spieler.

2.2 Klausur am 4. Mai 2015

1. **Binomialverteilte Zufallsgrößen**

   (a) Beschreibe kurz, wie das Histogramm einer $B(n,p)$–verteilten Zufallsgröße typischerweise aussieht. [6]

   (b) Es sei X eine $B(n,p)$–verteilte Zufallsgröße. Berechne

   i. $\mathbb{P}(X = 3)$ für $n = 8$, $p = \dfrac{2}{3}$; [4]

   ii. $\mathbb{P}(100 \leq X \leq 150)$ für $n = 2400$, $p = \dfrac{1}{20}$; [6]

   iii. für $n = 1200$ und $p = \dfrac{1}{10}$ das 90%–Prognoseintervall. [6]

   iv. Ist 200 für $n = 2400$ und $p = \dfrac{1}{20}$ ein ungewöhnliches Ereignis? [4]

2. **Wachzeiten**
   Ab und an schreckt Schüler A. im Unterricht auf, und dann bleibt er eine Zeit X [in Sekunden] wach, bevor er wieder wegduselt. Beobachtungen legen nahe, dass X in guter Näherung normalverteilt ist mit $\mu = 300$ und $\sigma = 60$. Behalte den Graphen der Dichtefunktion von X im Auge, wenn du die folgenden Teilaufgaben bearbeitest.

   (a) Mit welcher Wahrscheinlichkeit dauert As nächste Wachphase mindestens 400 Sekunden? [5]

   (b) Berechne ein möglichst kleines Intervall um μ, in dem 80% der Wachphasenlängen liegen. [5]

   (c) Berechne das obere und das untere Quartil von X. [6]

   (d) Mit welcher Wahrscheinlichkeit dauert die nächste Wachphase ganz exakt 300 Sekunden? [3]

   (e) Mit welcher Wahrscheinlichkeit liegt die Dauer der nächsten Wachphase in einem kleinen Intervall der Breite Δt, das μ enthält? [Hinweis: da ist nicht viel zu rechnen] [0+]

3. **Tunichtgute aufspüren**
   Ein Institut bietet dem SG einen Test an. Einen Tunichtgut erkennt der Test mit 98%–iger Sicherheit. Leider besteht ein geringes Risiko von 1%, dass ein braver Schüler irrtümlich als Tunichtgut verbellt wird. Für die Schulleitung wäre natürlich gut, sie kennte alle Tunichtgute.

   (a) Wieviele Schüler wird der Test als Tunichtgute melden, wenn es am SG n Tunichtgute gibt und der Test auf alle 1250 Schüler angewandt wird? Was ist das konkret für $n = 5$? [10]

   (b) Mit welcher Wahrscheinlichkeit ist ein Schüler, der vom Test als Tunichtgut gemeldet wird, wirklich einer? Was ergibt sich konkret für $n = 5$? [14]

4. **Radioaktiver Zerfall**
Zu einem Zeitpunkt $t = 0$ sehen wir ein Teilchen einer radioaktiven Substanz. Meines Wissens kann niemand vorhersagen, wann das Teilchen zerfällt. In der Physik verwendet man mit gutem Erfolg das Modell, dass der Zerfallszeitpunkt des Teilchens eine Zufallsgröße T ist. Die Dichtefunktion von T ist

$$f(t) \;=\; \begin{cases} 0 & \text{für } t \leq 0 \\ ca^{-t} & \text{für } 0 < t \end{cases}$$

mit einer Zahl $a > 1$, die materialabhängig ist, und einer geeigneten Konstanten c.

(a) Berechne c so, dass f eine Dichtefunktion ist – zur Not für $a = 2$. [12]

(b) Für den Rest der Aufgabe nehmen wir $a = e$. Dann ist $c = 1$, also $f(t) = e^{-t}$ für $0 < t$, und die Rechnungen werden einfacher. Skizziere zu den Wahrscheinlichkeiten auch jeweils das Flächenstück im Graphen von f. Skizzen: [10]

   i. Mit welcher Wahrscheinlichkeit zerfällt das Teilchen bis zum Zeitpunkt $t = 3$? [4]

   ii. Mit welcher Wahrscheinlichkeit existiert das Teilchen noch nach 3 Zeiteinheiten? [3]

   iii. Mit welcher Wahrscheinlichkeit zerfällt es in der Zeit von 3 bis 6?[3]

   iv. Erkläre, was $\mathbb{P}(T \leq 6 \,|\, T > 3)$ bedeutet, und berechne den Wert.[10]

   v. Berechne $\mathbb{E}(T)$. [12]

   vi. Bestimme $\tilde{t}$ so, dass $\mathbb{P}(T \leq \tilde{t}) \;=\; \dfrac{1}{2}$ ist, und schreibe auf, wie diese Kennzahl von T heißt. [10]

(c) Das $\tilde{t}$ aus der letzten Teilaufgabe ist in der Physik die Halbwertszeit der Substanz. Schon seltsam, dass es so etwas gibt, wo doch der Zerfallszeitpunkt eines jeden Teilchens zufällig sein soll. Nun, schauen wir uns eine kleine Menge der Substanz an, sie möge zur Zeit $t = 0$ aus $N = 10^{22}$ Teilchen bestehen. Die Zufallsgröße X sei die Anzahl dieser Teilchen, die bis zum Zeitpunkt $\tilde{t}$ zerfallen.

   i. Erläutere, dass X eine $B\left(10^{22}\ ,\ \frac{1}{2}\right)$–verteilte Zufallsgröße ist, und berechne μ und σ von X. [4+]

   ii. Gib eine möglichst kleine Zahl a an, für die $\mathbb{P}(|X - \mu| \geq a) \leq 0,0001$ ist. Mir reicht schon, wenn das Tschebyschew–Risiko $\leq 0,0001$ ist. [8+]

   iii. Der Wert $\frac{a}{\frac{1}{2}N}$ ist eine sehr kleine Zahl, von daher ist der Begriff Halbwertszeit doch plausibel. Erkläre! [0+]

2.3 Schätzen: Konfidenzintervalle

Du hast für eine $B(n,p)$–verteilte Zufallsgröße X zu gegebenen Sicherheitswahrscheinlichkeiten Prognoseintervalle berechnet. Beispiel: Mit wie vielen Sechsen kann man rechnen, wenn man einen idealen Würfel 1200–mal wirft? Natürlich muss man eine Sicherheitswahrscheinlichkeit festlegen, meinetwegen 95%, dann kann man das Intervall ausrechnen. Und wenn man das 1200–mal Würfeln sehr oft durchführt, wird der beobachtete Wert im Mittel in 95% der Durchführungen im angegebenen Intervall liegen; das ist eine klare Sache.

Man kennt hier die Erfolgswahrscheinlichkeit p von X exakt und sagt etwas über den Wert k, den die Zufallsgröße bei der nächsten Durchführung annimmt. Häufig aber ist es so, dass man zwar davon ausgehen kann, dass man es mit einer binomialverteilten Zufallsgröße zu tun hat, nur kennt man die Erfolgswahrscheinlichkeit p der Zufallsgröße nicht. Zum Beispiel möchte man wissen, mit welcher Wahrscheinlichkeit eine geworfene Reißzwecke in der Lage $\perp$ landet, oder, und das interessiert viel mehr Leute, wie hoch der Anteil der FDP–Wähler unter den Wahlberechtigten ist.

Sjard hat das gestellte Problem fix gelöst: Man wirft eine Reißzwecke 100–mal. Landet sie 34–mal in der Lage $\perp$, sollte $p = \frac{34}{100} = 0.34$ sein. Oder? Nun, da gibt es zwei Haken bei der Sache. Wirft man die Reißzwecke wieder 100–mal, wird man vermutlich nicht wieder genau 34–mal $\perp$ beobachten. Und wenn man bei 1000 Versuchen 340–mal $\perp$ beobachtet, oder bei 100000 Versuchen 34000–mal $\perp$, wird man jedesmal $p = 0,34$ sagen, aber es ist völlig klar, dass diese Werte nicht alle gleich vertrauenswürdig sind. Deshalb gibt man in der Statistik nur dann eine einzige Zahl als Schätzwert für das gesuchte p an, wenn man dem Fragesteller keine komplexere Antwort zumuten möchte. Dem Kundigen nennt man ein ganzes Intervall, das Konfidenzintervall für p. Je größer die Anzahl n der Versuche ist, desto schmaler wird das Intervall. Andererseits geht man immer das Risiko ein, dass das Intervall das gesuchte p nicht überdeckt, und je kleiner dieses Risiko ist, desto breiter wird das Intervall.

Genug der Vorüberlegungen. Man führt den Zufallsversuch n–mal aus und bestimmt die Anzahl k der Erfolge. Als „Schätzer“ $\hat{p}$ für p nimmt man natürlich[12]

$$\hat{p} \;=\; \frac{k}{n} \quad .$$

Um ein Konfidenzintervall für p zum beobachteten Wert k zu bekommen, muss man festlegen, welches Risiko man eingehen will, sich zu irren. Üblich sind zum Beispiel 5%. Man nimmt dann an, dass der beobachtete Wert k im 95%–Prognoseintervall von X liegt. Die Menge aller p, für das der Fall ist, bildet ein Intervall. Beim kleinsten p des Intervalls liegt k gerade an der rechten Grenze des zugehörigen Prognoseintervalls, beim größten p des Intervalls liegt k an der linken Grenze. Ein wenig formaler ausgedrückt: Bezeichnen wir die $B(n,p)$–verteilte Zufallsgröße mit X_p, ist das 95%–Konfidenzintervall zum beobachteten Wert k die Menge

$$\{\, p \in [0;1] \mid k \text{ liegt im 95\%–Prognoseintervall von } X_p \,\} \quad . \tag{21}$$

Ich will dir die Sache etwas anschaulicher machen. Die Histogramme der X_p haben Buckelform. Der Gipfel liegt in der Nähe des Erwartungswertes np von X_p, das weißt du alles. Lässt du das p von $0,1$ bis $0,9$ laufen, rutscht aber nicht nur das Histogramm von links nach rechts, sondern es ändert seine Form. Je näher p an $\frac{1}{2}$ liegt, desto breiter und flacher ist das Histogramm.

[12] Beachte, dass dieses $\hat{p}$ ist selbst eine Zufallsgröße ist.

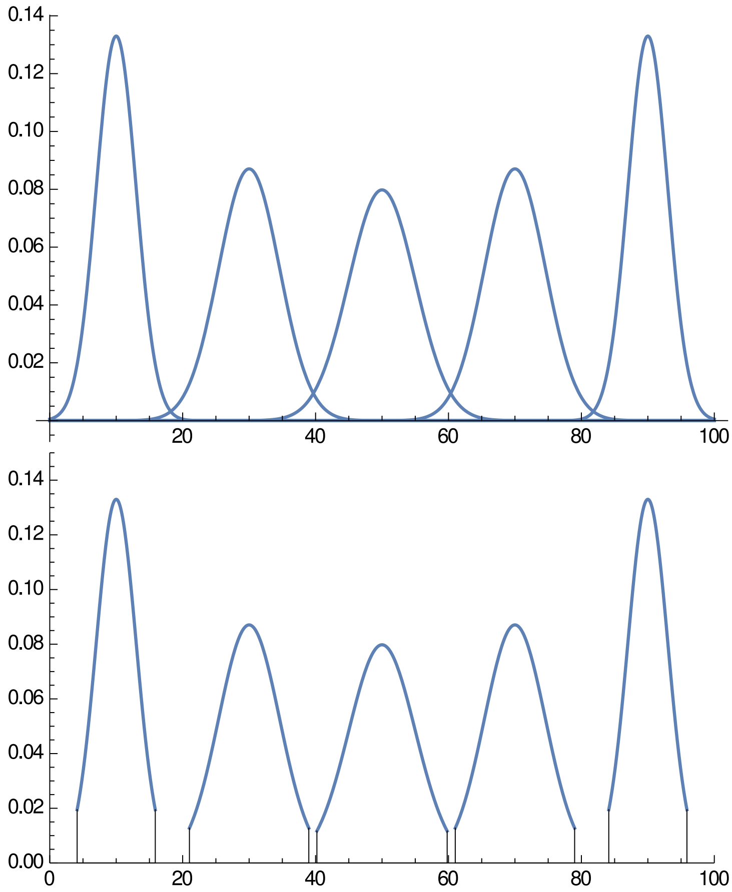


Abbildung 22: Histogramme $B(100, p)$–verteilter Zufallsgrößen für einige p, näherungsweise dargestellt durch die Dichten der entsprechenden normalverteilten Zufallsgrößen, und die auf die 95%–Prognoseintervalle gestutzten Histogramme dazu

Nun hat das 95%–Prognoseintervall von X_p den Radius

$$\Phi^{-1}(0.975)\sigma \;\approx\; 1.96\sqrt{np(1-p)} \quad ,$$

und es reicht von $\mu - 1.96\sigma$ bis $\mu + 1.96\sigma$. Für das kleinste p des Konfidenzintervalls ist folglich

$$k \;=\; \mu + 1.96\sigma \;=\; np + 1.96\sqrt{np(1-p)} \quad ,$$

für das größte p dagegen

$$k \;=\; \mu - 1.96\sigma \;=\; np - 1.96\sqrt{np(1-p)} \quad .$$

Das sind zwei Gleichungen für p. Um sie nach p aufzulösen, bringt man jeweils np auf die andere Seite und quadriert. In beiden Fällen kommt man zu derselben Gleichung, nämlich zu

$$(k - np)^2 \;=\; 1.96^2 \cdot np(1-p) \quad . \tag{22}$$

Das ist bloß eine quadratische Gleichung für p. Sie hat zwei Lösungen, und die sind die Grenzen des **(exakten)** 95%–**Konfidenzintervalls**. Schneller bekommt man das **Näherungskonfidenzintervall**. Dazu bildet man einfach das 95%–Prognoseintervall von $X_{\frac{k}{n}}$ und dividiert dessen Grenzen durch n.

Beispiel. Für den beobachteten Wert $k = 34$ und $n = 100$ ist der Radius des 95%–Prognoseintervalls $1.96 \cdot \sqrt{100 \cdot 0.34 \cdot 0.66} \approx 9.28$. Das Intervall ist folglich $[24.71; 43.28]$,[13] und das Näherungskonfidenzintervall deshalb

$$[0.2471; 0.4328] \quad .$$

Für das exakte 95%–Konfidenzintervall bekommt man nach dem Lösen der quadratischen Gleichung (22) das Intervall

$$[0.254613795; 0.4372245434] \quad .$$

Die Grenzen des Näherungskonfidenzintervalls sind nicht einfach gerundete Werte der exakten Grenzen, die Intervalle sind unglücklicherweise gegeneinander verschoben – die Gründe haben wir besprochen.

2.4 Was die Sicherheitswahrscheinlichkeit für das konkrete Konfidenzintervall bedeutet

Was war das noch mit dem Konfidenzintervall? Wir haben es mit einer Bernoullikette der Länge n mit der unbekannten Erfolgswahrscheinlichkeit p zu tun, und wir wollen das p schätzen. Dazu lassen wir den Versuch einmal laufen und erhalten eine Anzahl k von Erfolgen; dieses k ist dann ein ausgeloster Wert der $B(n, p)$–verteilten Zufallsgröße X : „Anzahl der Erfolge bei unserer Bernoullikette“. Zu einer gewählten oder von einem Auftraggeber vorgegebenen Sicherheitswahrscheinlichkeit von z.B. 95% berechnen wir dann das 95%–Konfidenzintervall $[p_1, p_2]$ zum beobachteten Wert k. Im Beispiel am Ende des letzten Unterkaptitels war $n = 100$, $k = 34$, $p_1 \approx 0.2546$ und $p_2 \approx 0.4373$.[14]

Wir wollen nun untersuchen, was die Sicherheitswahrscheinlichkeit 95% für das konkrete berechnete Intervall $[p_1, p_2]$ bedeutet. Man ist versucht zu sagen, dass das gesuchte p mit der Wahrscheinlichkeit 95% in dem berechneten Intervall liegt, und dieser Versuchung sind auch schon viele erlegen, aber die Formulierung ist Blödsinn, und du sollst sie nicht verwenden. Warum ist sie Blödsinn? Da musst du fragen, was es bedeutet, dass ein Ereignis die Wahrscheinlichkeit 95% hat: Wenn man den Zufallsversuch zu dem Ereignis „sehr oft“ durchführt, wird man im Mittel in etwa 95% der Fälle beobachten, dass das Ereignis eingetreten ist. So weit, so gut. Aber was genau ist der Zufallsversuch? Der sollte nicht im Lösen der quadratischen Gleichung bestehen, du bist ja Lk–Schüler. Sondern: Du musst die Bernoullikette sehr oft laufen lassen. Dann bekommst du jedesmal einen Wert k und jedesmal das 95%–Konfidenzintervall zu diesem Wert k. Heißt: Du bekommst so viele Intervalle, wie du die Bernoullikette hast laufen lassen, und im Mittel werden 95% dieser Intervall das wahre p enthalten – oder überdecken, wie man auch sagt. Die 95% sind eine Kenngröße des **Prozesses**, mit dem du das konkrete Intervall wie zum Beispiel $[0.2546; 0.4373]$ bestimmt hast. Zu dem konkreten Intervall kann man nur sagen, dass es mit einem Prozess bestimmt wurde, der im Mittel in 95% der Fälle gute Intervalle liefert.

[13] Auf ganze Zahlen achtet man hier nicht.

[14] Du wunderst dich vielleicht darüber, wie ich gerundet habe. In der Tat sollte man die linke Grenze stets abrunden und die rechte stets aufrunden, damit das Intervall keinesfalls zu klein ist.

Klar? Ich gebe dir noch ein Beispiel. Wenn du einen Würfel wirfst, wirst du mit der Wahrscheinlichkeit $\frac{5}{6}$ keine Sechs bekommen. Nun würfelst du konkret und erhältst, sagen wir, eine Zwei. Dann wirst du doch nicht sagen, mit der Wahrscheinlichkeit $\frac{5}{6}$ sei die Zwei keine Sechs. Siehst du, und deshalb solltest du auch die falsche Sprechweise meiden, das wahre p liege mit der Wahrscheinlichkeit 95% im berechneten Intervall. Sage nur, das Intervall sei das 95%–Konfidenzintervall zum beobachteten Versuchergebnis, und hoffe, dass der Auftraggeber keine Nachfragen stellt, sonst musst du ihm nämlich erklären, was ich dir eben aufgeschrieben habe – und was du jetzt hoffentlich auch verstanden hast.

Sicherheitshalber bekommst du noch eine kleine Simulation. Wir wollen die Erfolgswahrscheinlichkeit p einer $B(100;\ 0.4)$–verteilten Zufallsgröße X schätzen. Da wir das wahre p kennen, können wir sehen, was unsere Konfidenzintervalle taugen. Wir losen $N = 50$ Werte der Zufallsgröße aus und berechnen zu jeder das 95%–Konfidenzintervall. Die N Intervalle sind in Abbildung 23 als waagerechte Balken dargestellt, die rote Linie markiert das $p = 0.4$. Schneidet die rote Linie einen Balken, überdeckt das zugehörige Konfidenzintervall unser p, sonst nicht.

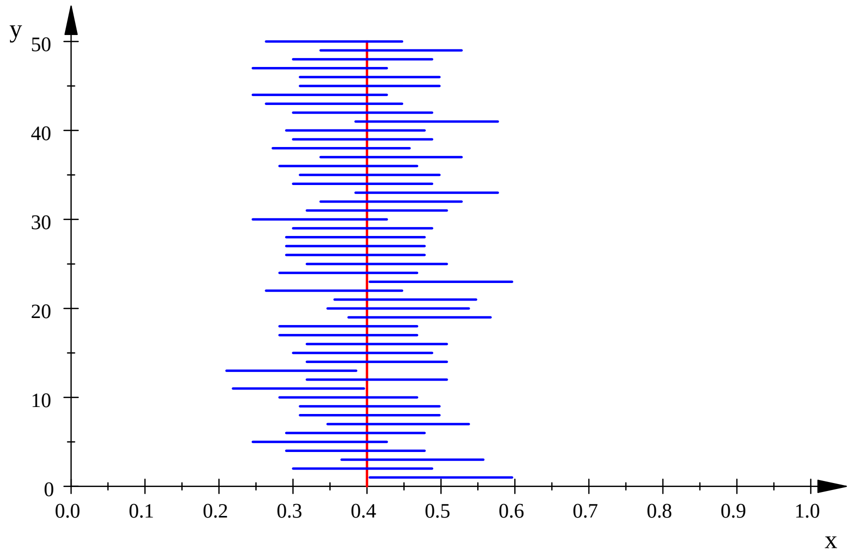


Abbildung 23: Fünfzig Konfidenzintervalle für das p einer $B(100; 0.4)$–verteilten Zufallsgröße

2.5 Wie σ von p abhängt bei $B(n,p)$–verteiltem X

Es sei X eine $B(n,p)$–verteilte Zufallsgröße. Dann ist bekanntlich

$$\sigma \;=\; \sqrt{np(1-p)} \;=\; \sqrt{n}\cdot\sqrt{p(1-p))} \;\; .$$

Wir wollen untersuchen, wie σ von p anhängt, deshalb lassen wir den Faktor $\sqrt{n}$ außen vor; er bewirkt eine Streckung des Graphen von $p \mapsto \sqrt{p(1-p)}$ in y–Richtung und ist unproblematisch.

Wir lassen uns die Kurve $y = \sqrt{p(1-p)}$ zeichnen (siehe Abbildung 24), das Ergebnis sieht erstaunlicherweise aus wie ein Halbkreis. Der Mittelpunkt sollte $(\frac{1}{2}|0)$ und der Radius sollte $r = \frac{1}{2}$ sein. Passt das? Nun, der Kreis um den Nullpunkt hat die Gleichung

$$x^2 + y^2 \;=\; \frac{1}{4} \;\; .$$

Verschieben wir den um $\frac{1}{2}$ nach rechts, erhalten wir ein Gebilde mit der Gleichung

$$\left(x - \frac{1}{2}\right)^2 + y^2 \;=\; \frac{1}{4} \;\; .$$

Wir lösen die Klammern auf und vereinfachen:

$$\begin{aligned} x^2 - x + y^2 &= 0 \\ y^2 &= x - x^2 \;=\; x(1-x) \end{aligned}$$

– tatsächlich, es stimmt, wie wir sehen, ohne Ableitungen zu verwenden: der Graph von $p \mapsto \sqrt{p(1-p)}$ ist ein Halbkreis.

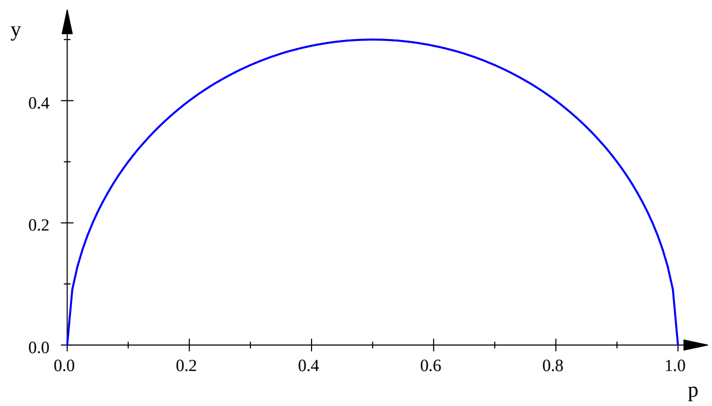


Abbildung 24: Der Graph von $p \mapsto \sqrt{p(1-p)}$

2.6 Bestimmung der Stichprobenlänge bei vorgegebener Genauigkeit

Bei einer $B(n,p)$–verteilten Zufallsgröße X soll ein 95%–Konfidenzintervall für p bestimmt werden. Das n soll so gewählt werden, dass der Radius des Konfidenzintervalls eine vorgegebene Größe nicht überschreitet. Zum Beispiel möchte Bürgermeister Vieker vor der Wahl wissen, wie hoch sein Wähleranteil sein wird, und zwar mit einer Genauigkeit von $\pm 1\%$.

Wir schauen uns nur das Näherungskonfidenzintervall an, sonst wird die Sache zu kompliziert. Sein Radius ist

$$r \;=\; 1.96\frac{\sqrt{n\hat{p}(1-\hat{p})}}{n} \;=\; 1.96\cdot\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \quad .$$

Wenn wir gar nichts über das gesuchte p wissen, können wir lediglich darauf zurückgreifen, dass

$$\sqrt{p(1-p)} \;\leq\; \frac{1}{2}$$

ist; das ist uns aus der Untersuchung der Abhängigkeit des σ von p aus dem letzten Unterkapitel bekannt. In diesem Fall gilt

$$r \;=\; \frac{1.96\cdot\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \;\leq\; \frac{1.96}{2\sqrt{n}} \quad .$$

Wir können $r \leq a$ garantieren, wenn wir sicherstellen, dass

$$\frac{1.96}{2\sqrt{n}} \;\leq\; a$$

ist, und dies formen wir leicht um zu

$$n \;\geq\; \left(\frac{1.96}{2a}\right)^2 \quad . \tag{23}$$

Dem Bürgermeister Vieker müssen wir dann zu seinem Entsetzen 9604 als erforderliche Stichprobenlänge nennen, eine aus mehreren Gründen indiskutable Zahl.

Geht es mit einer kleineren Stichprobenlänge? Wenn man das zu schätzende p in der Nähe von 0.1 vermutet, wird man $\sqrt{\hat{p}(1-\hat{p})}$ nicht durch $\frac{1}{2}$ abschätzen, sondern darin $\hat{p}$ durch 0.01 ersetzen. Dann erhält man

$$n \approx \left(\frac{1.96}{0.01}\right)^2 \cdot 0.1(1-0.1) \;\approx\; 3457 \quad .$$

Das ist immer noch sehr viel. In der Regel sind die Stichprobenlängen bei Umfragen viel kleiner, folglich die Konfidenzintervalle viel breiter als $2\cdot 0.01$. Nach meiner Einschätzung haben viele Umfrageergebnisse, die als aufregende Neuigkeiten verbreitet werden, lediglich einen gewissen Unterhaltungswert. Aber das muss ja auch nichts Schlechtes sein.

2.7 Tests: Entscheidungsregel und Gütefunktion

Wir kommen nun zu Hypothesentests, **dem** abiturrelevanten Aufgabentyp aus der Stochastik. Ich will euch mit Hilfe zweier Beispiele an den Testapparat heranführen.

A Der Hersteller von Tintenpatronen hat mit seinem Großkunden verabredet, dass höchstens 10% der Patronen zu wenig Tinte enthalten dürfen. Mit Hilfe einer Stichprobe der Länge $n = 100$ soll entschieden werden, ob eine Lieferung diese Zusage erfüllt.

B Bei einem idealen Würfel sollte die Wahrscheinlichkeit einer Sechs genau $\frac{1}{6}$ sein. Bevor ein solcher Würfel die Spielwarenmanufaktur verlässt, wird er geprüft, indem er 600–mal geworfen wird und die Anzahl der Sechsen ermittelt wird.

Der Prüfer, der die Tintenpatronenlieferung begutachten soll, bekommt eine **Entscheidungsregel:** „Lehne die Lieferung ab, wenn in der Stichprobe mehr als g Patronen mit Minderfüllung sind.“ Nun ist die Anzahl der Patronen mit Minderfüllung in der Stichprobe eine Zufallsgröße. Wenn p der Anteil der Patronen mit Minderfüllung in der Fertigung ist, hat jede zufällig gewählte Patrone mit der Wahrscheinlichkeit p Minderfüllung, und die Anzahl der Patronen mit Minderfüllung in der Stichprobe können wir als eine $B(100, p)$–verteilte Zufallgröße X_p ansehen. Die Lieferung wird dann mit der Wahrscheinlichkeit $\mathbb{P}(X_p > g)$ abgelehnt, und dieser Wert hängt ab von p und von g.

Damit du siehst, wie die Sache läuft, lasse ich den Graphen der Funktion f_g mit

$$f_g(p) \;=\; \mathbb{P}(X_p > g)$$

einmal für einige Werte von g zeichnen. Hier ist der für $g = 10$ (Abbildung 25).

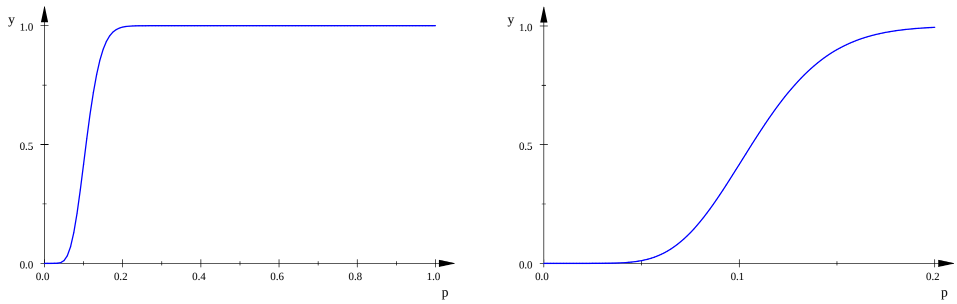


Abbildung 25: Graphen von $p \mapsto \mathbb{P}(X_p > 10)$ für $0 \le p \le 1$ und für $0 \le p \le 0.2$

Was siehst du daran? Eine Lieferung, die gerade noch in Ordnung ist, wird etwa mit der Wahrscheinlichkeit $\frac{1}{2}$ abgelehnt. Damit eine Lieferung ziemlich sicher angenommen wird, darf p eigentlich maximal 5% sein, und erst ab $p = 15\%$ wird die Lieferung mit ziemlicher Sicherheit abgelehnt.

Man möchte natürlich gern, dass eine Lieferung mit $p \le 10\%$ mit Sicherheit angenommen und dass eine Lieferung mit $p > 10\%$ mit Sicherheit abgelehnt wird. Der Graph der Funktion, die die Ablehnungswahrscheinlichkeit beschreibt, sollte im Idealfall aussehen wie der in Abbildung 26, aber das lässt sich nicht realisieren.

In der Praxis muss man Kompromisse machen. Je nach Marktmacht oder mit Blick auf den Schaden, der entsteht, wenn eine falsche Entscheidung getroffen wird,

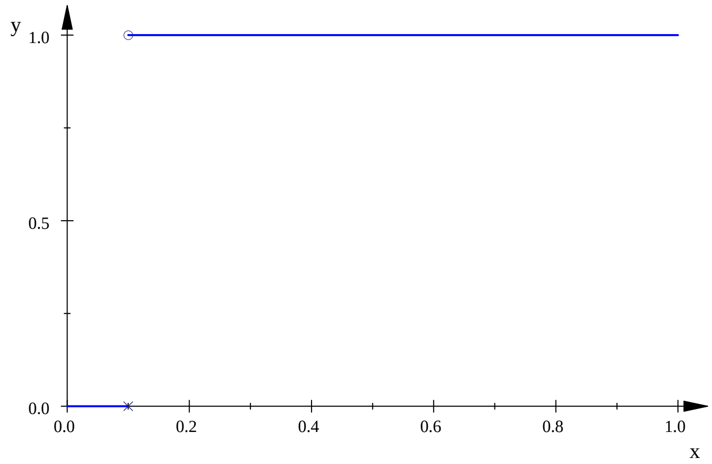


Abbildung 26: Graph von $p \mapsto \mathbb{P}(X_p > 10)$ für einen idealen Test

konstruiert man den Test so, dass die Lieferung in der Regel angenommen wird und nur abgelehnt wird, wenn sie deutlich zu viele Patronen mit Minderfüllung enthält, oder dass sie nur angenommen wird, wenn sie mit ziemlicher Sicherheit nur einen geringen Anteil schlechter Patronen enthält. Im ersten Fall nimmt man ein größeres g, im zweiten Fall ein kleineres. Abbildung 27 veranschaulicht, wie sich das auswirkt.

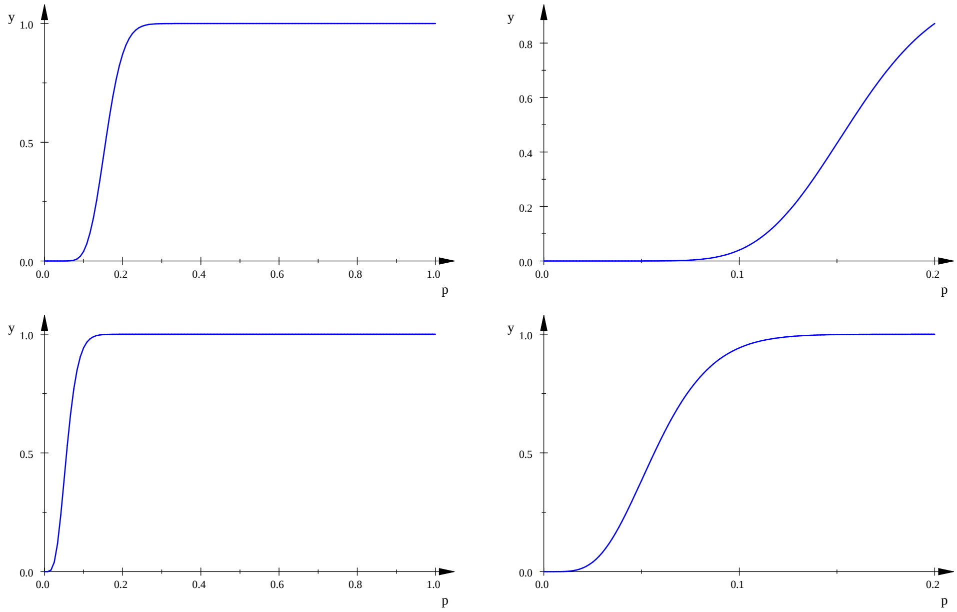


Abbildung 27: Graphen von $p \mapsto \mathbb{P}(X_p > 15)$ (oben) und von $p \mapsto \mathbb{P}(X_p > 5)$ (unten)

Bei dem **Würfelbeispiel** sieht die Sache anders aus. Man wird den Würfel sowohl ablehnen, wenn er zu viele als auch wenn er zu wenige Sechsen liefert. Probieren wir es mit der Entscheidungsregel: „Lehne den Würfel ab, wenn bei den 600 Würfen weniger als 85 oder mehr als 115 Sechsen auftreten.“ Bezeichnen wir die $B(600, p)$–verteilte Zufallsgröße mit X_p, erfolgt dann die Ablehnung eines Würfels, der mit der Wahrscheinlichkeit p eine Sechs liefert, mit der Wahrscheinlichkeit

$$f(p) \ := \ \mathbb{P}(X_p < 85) \ + \ \mathbb{P}(X_p > 115) \ = \ 1 - \mathbb{P}(85 \leq X_p \leq 115) \ .$$

Den Graphen von f_p siehst du in Abbildung 28.

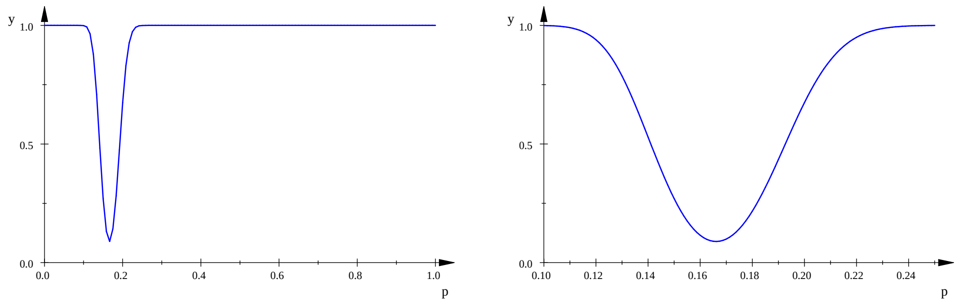


Abbildung 28: Graphen von $p \mapsto \mathbb{P}(X_p < 85) + \mathbb{P}(X_p > 115)$

Fragen

1. Was genau bedeutet das $f_g(p)$, das in Abbildung 25 dargestellt ist?
2. Was müsste man tun, wenn der Graph von $f_g(p)$ in Abbildung 25 dem idealen Graphen in Abbildung 26 ähnlicher werden soll?
3. Wie müsste man g wählen, damit eine Lieferung mit dem Anteil 10% von Patronen mit Minderfüllung maximal mit der Wahrscheinlichkeit 20% abgelehnt wird? Dabei soll g natürlich nicht unnötig groß gewählt werden.

2.8 Testchinesisch

Für den Anteil p der Tintenpatronen mit Minderfüllung in Beispiel A auf Seite 54 gibt es zwei Möglichkeiten: Es ist $p \leq \frac{1}{10}$ oder es ist $p > \frac{1}{10}$. Auf Testchinesisch heißen diese beiden Aussagen **Hypothesen**. Die eine heißt **Nullhypothese** und wird mit H_0 bezeichnet, die andere heißt **Alternative** H_1. Damit du hier die richtige Zuordnung treffen kannst, musst du erst einmal wissen, wie ein Hypothesentest arbeitet: **Man geht davon aus, dass die Nullhypothese wahr ist, und schaut, ob man diese Annahme auf Grund des Stichprobenergebnisses verwerfen kann.** Ist dies der Fall, hat der Test etwas bewirkt. Kann man die Nullhypothese nicht verwerfen, bleibt man bei der Annahme, sie sei wahr, was aber keineswegs heißt, dass der Test ihre Wahrheit nachgewiesen oder auch nur bestätigt hätte.

Ich demonstriere den Apparat einmal an Beispiel A. Als Nullhypothese wähle ich die Annahme, dass die Lieferung in Ordnung ist:

$$H_0 : \quad p \leq \frac{1}{10}$$

Der **Verwerfungsbereich** oder **kritische Bereich**, der die Stichprobenergebnisse enthält, die zur Ablehnung von H_0 führen, ist dann von der Art

$$\{\{\, k \in \{0, 1, \ldots, 100\} \mid k > g \,\}$$

für eine ganze Zahl g. Die Wahrscheinlichkeit, dass wir ein solches k beobachten, ist $\mathbb{P}(X_p > g)$.

Der Test kann nun dazu führen, dass die Nullhypothese verworfen wird, obwohl sie wahr ist; das ist der **Fehler erster Art**. Die Wahrscheinlichkeit eines Fehlers erster Art ist das **Risiko erster Art**, es wird gewöhnlich durch eine vorgegebene obere Schranke α gedeckelt. Üblich sind Werte wie 10%, 5% oder 1% für dieses α. Man wählt das g nun möglichst groß, denn der Test soll ja wirken[15], aber nur so groß dass

$$\mathbb{P}(X_p > g \,|\, H_0) \;\leq\; \alpha$$

ist. Wir nehmen mal $\alpha = 5\%$. Nun ist $\mathbb{P}(X_p > g)$ um so größer, je größer p ist, also gilt

$$\mathbb{P}(X_p > g \,|\, H_0) \;\leq\; \mathbb{P}(X_{0.1} > g) \quad .$$

Mit Hilfe der integralen Näherungsformel (oder mit einer Tabelle) finden wir

$$\mathbb{P}(X_{0.1} > 15) \quad \approx 0.0399 \quad \text{und} \quad \mathbb{P}(X_{0.1} > 14) \;\approx\; 0.0726 \quad ,$$

folglich ist $g = 15$ zu wählen.

Unsere Entscheidungsregel lautet also: Weise die Lieferung zurück, wenn du in der Stichprobe der Länge 100 mehr als 15 Tintenpatronen mit Minderfüllung findest. Wie der Test arbeitet, siehst du am besten an seiner **Gütefunktion**

$$f : p \mapsto \mathbb{P}(X_p > 15) \quad .$$

Den Graphen der Gütefunktion findest du in Abbildung 27 oben auf der Seite 55.

Für den Lieferanten ist unser Test natürlich recht komfortabel. Bringt er eine Lieferung, in der 15% der Tintenpatronen Minderfüllung haben, wird diese Lieferung noch mit der Wahrscheinlichkeit

$$\mathbb{P}(X_{0.15} \leq 15) \;\approx\; 0.568$$

angenommen. Der Fehler, dass die Nullhypothese nicht verworfen wird, obwohl sie falsch ist, heißt übrigens **Fehler zweiter Art**.

Nehmen wir nun einmal an, das Stichprobenergebnis sei 15. Dann wird H_0 nicht verworfen. Aber niemand, der bei klarem Verstand ist, wird ein Stichprobenergebnis mit der relativen Häufigkeit 0.15 als Beleg dafür nehmen, dass das wahre p maximal 0.1 ist. Also noch einmal: Der Test führt zur Ablehnung der Nullhypothese oder er hat (erstmal) nichts gebracht.

Jetzt weißt du Bescheid, und wir können das Testen üben. Achte mit darauf, dass wir genügend Beispiele behandeln.

[15] er soll „mächtig sein“, sagt man in Fachkreisen

2.9 Einschub: Bonusaufgabe

Vor dem Spieler liegen zwei Umschläge mit Geld. Der Spieler weiß, dass ein Umschlag das Doppelte des Betrages enthält, der in dem anderen ist. Einen Umschlag darf der Spieler auswählen und öffnen. Dann darf er entscheiden, ob er das Geld behalten will oder ob er lieber den anderen Umschlag wählt. Was soll der Spieler tun?

1. Eine naive Rechnung geht so: In dem geöffneten Umschlag sei der Betrag x. Dann ist in dem anderen Umschlag der Betrag $2x$ oder $\frac{1}{2}x$, jeweils mit der Wahrscheinlichkeit $\frac{1}{2}$. Wechselt er, wird er im Mittel den Betrag

$$\frac{1}{2} \cdot 2x + \frac{1}{2} \cdot \frac{1}{2}x \;=\; \frac{5}{4}x$$

   bekommen. Deshalb sollte er wechseln!? Und wenn er zweimal wechselt, bekommt er am Ende noch mehr??

2. Oliver S. hat eine MuPAD–Prozedur geschrieben, die die Umschläge zufällig füllt. Den Betrag x in dem einen Umschlag bestimmt er so: Er lost mit dem Befehl „frandom()“ eine Zahl $z \in [0;1]$ aus. Die Zufallsgröße Z dazu ist gleichverteilt, das heißt, sie ist stetig verteilt mit der Dichtefunktion

$$g(z) = \begin{cases} 0 & \text{für } z < 0 \\ 1 & \text{für } 0 \leq z \leq 1 \\ 0 & \text{für } 1 < z \end{cases} \quad .$$

   Zu dem ausgelosten z berechnet er

$$x \;=\; -\ln(1-z) \quad ,$$

   und das ist der Geldbetrag.

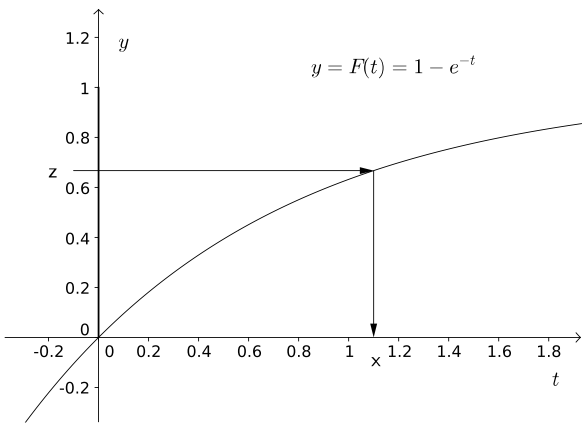


Abbildung 29: Zur Auslosung von x

   Der Hintergrund dazu sieht so aus: Du kennst die Zufallsgröße T: „Zerfallszeitpunkt eines radioaktiven Teilchens“ mit der Dichtefunktion

$$f(t) \;=\; \begin{cases} 0 & \text{für } t < 0 \\ e^{-t} & \text{für } 0 \leq t \end{cases} \quad .$$

Die Verteilungsfunktion von T ist für $0 \le t$

$$F(t) \;=\; \int_0^t e^{-u}\,du \;=\; -e^{-u}\,|_{u=0}^{u=t} \;=\; 1 - e^{-t} \quad \text{für } 0 \le t.$$

Wir berechnen nun die Verteilungsfunktion der Zufallsgröße X. Für $0 < x$ ist

$$\mathbb{P}(X \le x) \;=\; \mathbb{P}(Z \le 1 - e^{-x}) \;=\; 1 - e^{-x} \quad ,$$

folglich ist die Dichtefunktion von X für $x > 0$ gegeben durch

$$\big(1 - e^{-x}\big)' \;=\; e^{-x} \quad ,$$

wie gewünscht.

3. Oliver lost nun auf die beschriebene Weise einen (gerundeten) Geldbetrag x aus, füllt einen Umschlag mit x und den anderen mit $2x$. Dabei entsteht das Problem, dass man unter Umständen erkennen kann, dass man nicht den Umschlag mit dem Inhalt $2x$ geöffnet hat, weil die Endziffer der Zahl ungerade ist (Oliver G.). Das soll natürlich nicht möglich sein.

4. Wir haben nun eine MuPAD–Routine, die die Umschläge füllt. Man kann wählen, welchen Umschlag man öffnen möchte, sieht den Betrag in diesem Umschlag und hat dann einige Sekunden Zeit zu entscheiden, ob man bei dem Umschlag bleiben will oder wechseln will.

5. Nach einigen Versuchen kam die Idee auf, eine geeignete Zahl S festzulegen und den Umschlag immer dann (und nur dann) zu wechseln, wenn der geöffnete Umschlag weniger als S enthält.

6. Nun ist das Folgende zu tun: Man bestimme – durch Simulation oder durch Rechnung – den Erwartungswert der Auszahlung, wenn man

   (a) immer den Umschlag 1 nimmt (d.h. beim gewählten Umschlag bleibt),

   (b) immer den Umschlag 2 nimmt (d.h., stets wechselt),

   (c) dann und nur dann wechselt, wenn im Umschlag 1 weniger als S ist.

   Bei der dritten Strategie ist der Erwartungswert eine Funktion von S, und man kann nach dem besten S fragen.

2.10 Tests auf dem Prüfstand

Nehmen wir die Aufgabe 18 aus dem Buch auf der Seite 174, in der es heißt, dass in Reihenuntersuchungen zur Zahngesundheit bei durchschnittlich 20% der Jugendlichen Zahnschäden festgestellt wurden und nun nach einer Aufklärungskampagne durch einen Test festgestellt werden soll, ob der Anteil der Jugendlichen mit Zahnschäden gesunken ist. Dazu soll eine Stichprobe der Länge 800 gezogen werden, und es soll auf dem Signifikanzniveau 5% getestet werden.

Als Hypothesen über den unbekannten Anteil p der Jugendlichen mit Zahnschäden bieten sich $p < 20\%$ und $p \geq 20\%$ an. Aber welche davon soll H_0 werden? Wir wählen $H_0: \ p < 20\%$, und du rechnest gleich los:

Aufgabe 1. Entwerfe den Test zu unserem H_0. Mache das ordentlich: Markiere auf der Merkmalsachse, wo die Erwartungswerte der $B(800, p)$–verteilten X_p liegen, die zu H_0 gehören, und wo der Verwerfungsbereich V liegen muss. Berechne dann diesen Verwerfungsbereich V und formuliere die Entscheidungsregel. Mit welcher Wahrscheinlichkeit wird H_0 verworfen, wenn $p = 22\%$ ist? Wie sieht die Gütefunktion deines Tests aus?[16] Denke dich in den Inhaber der Werbeagentur und in den Kämmerer des Landkreises hinein: der eine bekommt im Erfolgsfall das Geld, der andere steht für die Ausgaben gerade. Wer von denen kann mit dem Test gut leben, der eine, der andere, keiner oder alle beide?

Na gut, aber es könnte ja auch $H_0: \ p \geq 20\%$ gewählt werden.

Aufgabe 2. Mache die Aufgabe 1 neu für das alternative H_0.

Aufgabe 3. Was ist dein Fazit: welches H_0 ist angemessen?

Aufgabe 4. Was hältst du von einem zweiseitigen Test? Wie lauten dann H_0 und H_1? Wäre der hier sinnvoll?

Werfen wir nun einen Blick auf die Daten. Das alte p in der Grundgesamtheit war $p = 20\%$, und es ist herauszufinden, ob es sich verändert hat. Was heißt denn 20%? So ungefähr, auf 10% gerundet? Auf 1% gerundet? Exakt $20\% = \frac{1}{5}$ als reelle Zahl? Wie wurde das p gewonnen, durch eine Stichprobe? Dann kennt man nur ein Konfidenzintervall. Wie breit ist das? Wann kann man von einem Erfolg der Werbekampagne sprechen? Wenn man $p = \frac{1}{5}$ zu Grunde legt, ganz exakt, und sich der Wert dann auf $\frac{1}{5} - 10^{-6}$ geändert hat?

Du siehst schon, je nachdem, wie man an die Aufgabe herangeht, ist der Schwachsinngehalt der Lösung mindestens 64 Volumenprozent. Wir wollen es nun besser machen.

Aufgabe 5. Agenturinhaber und Kämmerer verabreden, dass die Kampagne als erfolgreich angesehen und die Zahlung dafür fällig wird, wenn in der Stichprobe der Länge 800 maximal 16% Jugendliche mit Zahnschäden gefunden werden. Berechne das Risiko des Kämmerers, zahlen zu müssen, obwohl alles beim Alten geblieben ist, und das Risiko der Agentur, kein Geld zu bekommen, obwohl $p = 15\%$ ist.

2.11 Noch ein Testproblem

Unsere Q2 hat 179 Schüler, von denen haben 42 Sport als Leistungsfach. Jetzt, am Tag vor den Prüfungen im 1. bis 3. Fach, haben 13 Schüler noch keine 100 Punkte zusammen, deshalb müssen sie um ihr Abitur fürchten. Sieben der 13 haben Sport

[16] Den Graphen kannst du unter Verwendung der berechneten Daten grob skizzieren oder dir vom Rechner zeichnen lassen – wenn du mit MuPAD umgehen kannst. Der Rechnenaufwand ist bei $n = 800$ erheblich. Du kannst allerdings die Näherungsformel nehmen, dann geht vielleicht sogar Geogebra.

als Leistungsfach. Na ja: Ist das ein Beleg für die Vorbehalte, die im Untergrund grummeln, dass die Dumpfbackendichte unter den Sportlern signifikant höher ist als in der gesamten Stufe? Oder fallen diese Zahlen noch in den Bereich zufälliger Schwankungen? Das sollst du nun testen!

Unser Instrumentarium passt nicht so gut auf diese Situation, bei dem kleinen $|\Omega| = 179$ sollte man nicht mit der Binomialverteilung arbeiten. Man macht das so: Man füllt eine Urne mit 179 Kugeln, davon sind 13 weiß und 166 rot. Dann zieht man „auf einen Griff“ 42 Kugeln aus der Urne. Testgröße ist die Zufallsgröße X: „Anzahl der weißen Kugeln in der gezogenen 42er–Menge“. Nun berechnet man den p–Wert des beobachteten Wertes $X = 7$, also die Wahrscheinlichkeit $\mathbb{P}(X \geq 7)$, dass X den Wert 7 oder einen noch extremeren annimmt. Ist diese Wahrscheinlichkeit zu klein, kippt man die Nullhypothese, dass nur Zufall im Spiel ist.

Und wie groß ist $\mathbb{P}(X = k)$? Es gibt $\binom{179}{42}$ 42er–Teilmengen von Schülern der Stufe bzw. Kugeln der Urne – das ist lediglich eine Frage der Pietät – eine davon wird gezogen. Man sieht diese Mengen als gleichwahrscheinlich an und bestimmt die Anzahl der 42er–Teilmengen, die genau k weiße Kugeln enthalten: Um eine solche Menge zu bekommen, muss man k der 13 weißen Kugeln nehmen, da gibt es $\binom{13}{k}$ Möglichkeiten, und $42-k$ rote Kugeln dazutun. Von diesen Ergänzungsmengen gibt es $\binom{166}{42-k}$ Stück. Folglich gibt es

$$\binom{13}{k} \cdot \binom{166}{42-k}$$

42er–Teilmengen der Menge der Kugeln der Urne, die genau k weiße Kugeln enthalten, und daraus ergibt sich[17]

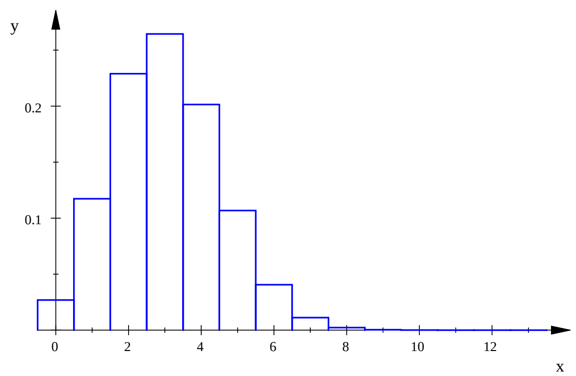
$$\mathbb{P}(X = k) \;=\; \frac{\binom{13}{k} \cdot \binom{166}{42-k}}{\binom{179}{42}} \quad , \quad \text{allgemein} \quad \mathbb{P}(X = k) \;=\; \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}} \quad .$$

Jetzt muss man bloß noch rechnen, allerdings z.B. mit MuPAD; deinem Taschenrechner geht vermutlich die Puste aus.

Das Ergebnis der Rechnung ist

$$\mathbb{P}(X \geq 7) \;=\; \frac{60312505795807}{4402176279479290} \;\approx\; 0.01370061124 \quad ,$$

und hier ist auch noch das Histogramm der Zufallsgröße X.



[17] Eine Zufallsgröße X, bei der man die Wahrscheinlichkeiten nach dieser Formel berechnet, heißt **hypergeometrisch verteilt** mit den Parametern $N, M, n \in \mathbb{N}$; $0 < M \leq N$, $n \leq N$.

3 Rückblick

Schauen wir kurz zurück auf unseren Stochastikkurs, auch mit Blick auf die Abiturklausur. Du hast elementare Wahrscheinlichkeitsrechnung gelernt: Zufallsversuche, Baumdiagramme, Vierfeldertafel, bedingte Wahrscheinlichkeit, Unabhängigkeit. Dann Zufallsgrößen $X : \Omega \to \mathbb{R}$. Am wichtigsten sind für uns binomialverteilte Zufallsgrößen, es sind auch praktisch die einzigen, die in Abiturklausuren vorkommen. Freilich solltest du nicht über den Formeln $\mu = np$ und $\sigma = \sqrt{np(1-p)}$ für Erwartungswert und Standardabweichung binomialverteilter Zufallsgrößen die allgemeinen Definitionen und unsere Theorie dieser Begriffe vergessen, das wäre schade. Als wichtige Anwendungen kennst du Schätzen (Konfidenzintervalle) und Hypothesentests, wobei lediglich Tests in der Abiturklausur vorkommen.

Stetig verteilte Zufallsgrößen, die wir auch behandelt haben, spielen in der Abiturklausur lediglich da eine Rolle, wo Wahrscheinlichkeiten bei binomialverteilten Zufallsgrößen mit Hilfe der Näherungsformeln von de Moivre und Laplace berechnet werden oder wo Sigmaregeln verwandt werden.

So weit das Praktische. Ich möchte ein paar grundsätzliche Bemerkungen zum Phänomen „Zufall“ anfügen. Dass die Wahrscheinlichkeitstheorie, die als abstrakte mathematische Theorie ja gar nicht nach dem Wesen des Zufalls in der Realität fragt, sehr erfolgreich auf reale Phänomene angewandt wird, weißt du. Aber was ist Zufall? Gibt es überhaupt Dinge, die „zufällig“geschehen, oder ist nicht vielmehr alles Teil einer langen Kette von Ursachen und Wirkungen? „Nur die Quantenphysik kennt echten Zufall“, heißt es in einer Veröffentlichung des Max–Planck–Instituts für die Physik des Lichts[18], und der Zugriff auf echten Zufall wird als etwas sehr Wertvolles herausgestellt. Natürlich ist das Öffentlichkeitsarbeit, aber es stimmt schon: Bei Lotterien und Glücksspielen soll es fair zugehen, und wenn ein Schlüssel für ein Verschlüsselungsverfahren ausgelost wird, muss der die buchstäbliche Nadel im Heuhaufen sein. Weiß jemand, wie hier „Zufall“ generiert wird, ist er klar im Vorteil. Generiert wird Zufall in der Regel im Computer, und das geschieht nach strengen Rechenregeln. Die Ergebnisse sind scheinbar regellos, das ist der Zufall dabei.

Wer die Ausgangsdaten und den Bewegungsablauf kennt, kann im Prinzip ausrechnen, welche Zahl ein Würfel anzeigen wird, oder er kann den Lauf einer Billardkugel ausrechnen. Aber eben nur im Prinzip: Nach zwanzig Kollisionen wirkt sich beim Billard schon die Massenanziehungskraft aus, die die Spieler auf die Kugeln ausüben – eine Kraft, die so klein ist, dass man sie kaum messen kann – Vorausberechnungen auf lange Sicht sind praktisch unmöglich.[19] Die Abläufe sind deterministisch, aber nicht vorhersagbar.

Von Quantenphysik verstehe ich so gut wie nichts, deshalb greife ich auf ein anderes Beispiel zurück. Der Zerfallszeitpunkt eines radioaktiven Teilchens geschieht zufällig, sagte ich euch, niemand kann den Zeitpunkt vorhersagen. Zur Zeit, muss ich hinzufügen. Es mag doch sein, dass die Physiker neue Gesetze entdecken, die den Zerfallszeitpunkt festlegen, dann geschähe der Zerfallsprozess zwangsläufig und nicht mehr zufällig, wobei durchaus sein könnte, dass niemand den Zerfallszeitpunkt praktisch vorausberechnen kann. Wenn das Ausmaß unseres Nichtwissens groß genug ist, sollte man meiner Meinung nach ruhig von Zufall reden und Stochastik – die Mathematik des Zufalls – anwenden, so gut es geht. Außerdem ist es Grund

[18] http://www.forschen-mit-Licht.mpg.de/68981/geheimcode-im-laserblitz am 26. Juni 2015

[19] Siehe zum Beispiel Volker Enß, Das Pendel im Gezeitenhub – alles vorherbestimmt und doch nicht vorhersagbar. Ein Artikel auf der Seite des Instituts für Reine und Angewandte Mathematik der RWTH Aachen, mit Google zu finden.

genug, sich mit Mathematik zu beschäftigen, wenn sie schön ist. Dass man damit Geld verdienen kann, ist natürlich auch nicht unwichtig, das gebe ich gern zu.

Wenn du noch etwas Interessantes zur Stochastik suchst, greife zu dem schönen Buch von Olle Häggström, Streifzüge durch die Wahrscheinlichkeitstheorie. Du findest es in der Schulbücherei. Im Kurs fassen wir allerdings nun ein neues Gebiet ins Auge: Lineare Algebra und Geometrie. Freu dich drauf.