Stochastik

LK Mathematik Abitur 2013 (B. Waldmüller)

3. Juli 2012

Inhaltsverzeichnis

1 Der Anfang

1.1 Einstimmung

Stochastik ist die mathematische Theorie des Ungewissen. – Diese Formulierung löste möglicherweise in Fachkreisen neben Zustimmung auch wütenden Protest aus, aber so fange ich nun an. Das Ungewisse ist nämlich alltägliche Begleiterscheinung menschlicher Existenz, und der Beitrag der Stochastik ist, dass sie ermöglicht, Ungewissheit zu quantifizieren. Chinesisch? Ich lege dir eine Reihe von Beispielen vor.

- Aus Finanzkreisen ist dieser Tage zu hören, die Wahrscheinlichkeit einer Staatspleite Griechenlands sei auf 40% gestiegen.
- Eine Versicherung verkauft einem vierzigjährigen Mann das Versprechen, den Hinterbliebenen 100.000 Euro zu zahlen, wenn er im Laufe des nächsten Jahres stirbt.
- Ein Wetterdienst im Internet gibt die Niederschlagswahrscheinlichkeit für morgen mit 40% an.
- Die Wahrscheinlichkeit, mit einem Würfel eine fünf zu werfen, ist $\frac{1}{6}$.

Nun, Griechenland geht in absehbarer Zeit pleite oder eben nicht, das sind zwei Möglichkeiten. Dennoch wird man die Wahrscheinlichkeit der Pleite nicht einfach auf $\frac{1}{2}$ setzen. Wenn jemand die Wahrscheinlichkeit der Pleite auf 40% setzt, drückt er damit den Grad der Zuversicht aus, dass die Pleite eintritt, und dieser Grad der Zuversicht mag sich einfach aus seinem Bauchgefühl ergeben. Immerhin ist die Auswirkung des Bauchgefühls der Finanzjongleure, dass die Griechen für ihre Staatsanleihen deutlich erhöhte Zinsen zahlen müssen. Bei der Erstellung der 40%–Prognose war vermutlich Mathematik im Spiel, aber eine mathematische Aussage ist das nicht.

Die Sache mit den Versicherungen ist schon ein wenig unheimlich. Woher wissen die denn, welchen Preis die für das Versprechen nehmen müssen? Es ist doch völlig ungewiss, ob der Mann das Jahr überlebt, und dabei spielen so viele Einflüsse mit, dass auch das ganze Servernetz von Facebook die Datenmenge nicht erfassen geschweige denn verarbeiten könnte; man sagt ja, dass der Flügelschlag eines Schmetterlings in Vietnam einen Tornado im Golf von Mexiko auslösen könne, dem der Mann ja vielleicht zum Opfer fällt. Hören wir einen Vertreter der Versicherungswirtschaft dazu: We do not know, who will die next year, but we know, how many. Da geht die Reise hin: die große Zahl macht die Unsicherheit berechenbar.

Und was mag die Niederschlagswahrscheinlichkeit aussagen? Vielleicht etwas in der Art: Wenn man Rückschau hält auf alle Tage mit vergleichbarer Wetterlage, hat es an etwa 40% dieser Tage Niederschläge gegeben. Deshalb tut man so, als zöge man eine Kugel aus einer Urne mit 40 weißen und 60 schwarzen Kugeln, wenn man auf das morgige Wetter wartet und schaut, ob es Niederschläge gibt. Will man wissen, wie verlässlich Wetterberichte sind, müsste man den Wetterbericht von gestern mit dem Wetter heute vergleichen. Das tut aber niemand, man liest immer den Wetterbericht für morgen.

Wenden wir uns dem letzten Beispiel zu, dem harmlosesten der Liste. Was bedeuten denn die $\frac{1}{6}$? Dass, wenn man sechsmal würfelt, genau eine fünf dabei ist? Nein, das ist falsch; solche Fragen werden wir rechnerisch behandeln. Aber bevor wir anfangen zu rechnen, will ich wissen, worüber wir eigentlich sprechen. Machen wir Aussagen über reale Würfel, wie du sie zu Hause hast, so wie das Fallgesetz der Physik Aussagen über reale fallende Steine macht? Klare Antwort: Nein, wir machen keine Physik. Unsere Aussagen gelten für mathematische Objekte, die nur in unserer Vorstellung existieren. Dass diese Aussagen für reale Würfel halbwegs passen, ist eine empirische Feststellung, also eine Art Physik. Dass mathematische Objekte zur Beschreibung der Wirklichkeit taugen, ist eine praktische Sache, und sie macht die Mathematik zu einem interessanten Werkzeug. Teil der (stofflichen) Wirklichkeit wird die Mathematik dadurch aber keineswegs, das muss man sauber trennen – finde ich jedenfalls.

Als mathematische Theorie ist die Stochastik nicht sonderlich alt. David Hilbert hat bei einem Kongress in Paris im Jahr 1900 als eines von 23 Jahrhundertproblemen die Aufgabe gestellt, eine mathematische Theorie der Wahrscheinlichkeit zu formulieren. Dies gelang dann A. Kolmogoroff, er schlug in den dreißiger Jahren des 20. Jahrhunderts den gordischen Knoten durch und schuf eine axiomatische Theorie der Wahrscheinlichkeit. Mal sehen, vielleicht schauen wir uns sein Axiomensystem einmal an; ich weiß es noch nicht. Jedenfalls: Als Kolmogoroff seine Theorie niederschrieb, hatte man schon ein paar Jahrhunderte lang mit Wahrscheinlichkeiten gerechnet, und zwar, um Glücksspiele zu analysieren.

1.2 Das empirische Gesetz der großen Zahl

Das Handeln von Glücksspielern und Versicherungen beruht auf Erfahrungen mit Zufall, und den Kern dieser Erfahrungen will ich dir an einem Beispiel zeigen. Sagen wir, wir würfeln n-mal, zählen die dabei gewürfelten Fünfen und teilen diese Anzahl durch die Zahl n der Würfe. Die Zahl, die wir dann berechnet haben, heißt die relative Häufigkeit der 5 bei n Würfen. Natürlich kommt nicht jedes Mal, wenn wir das machen, das Gleiche heraus; die relative Häufigkeit ist ein Messwert, der vom Zufall abhängt.

Ich habe nun jeweils 20–mal die relative Häufigkeit der 5 bei n Würfen bestimmt, und zwar für $n = 1, 5, 20, 80, 320$ und $n = 1280$. Die Ergebnisse findest du in Abbildung 1 graphisch dargestellt. Die Gerade hat die Gleichung $y = \frac{1}{6}$.

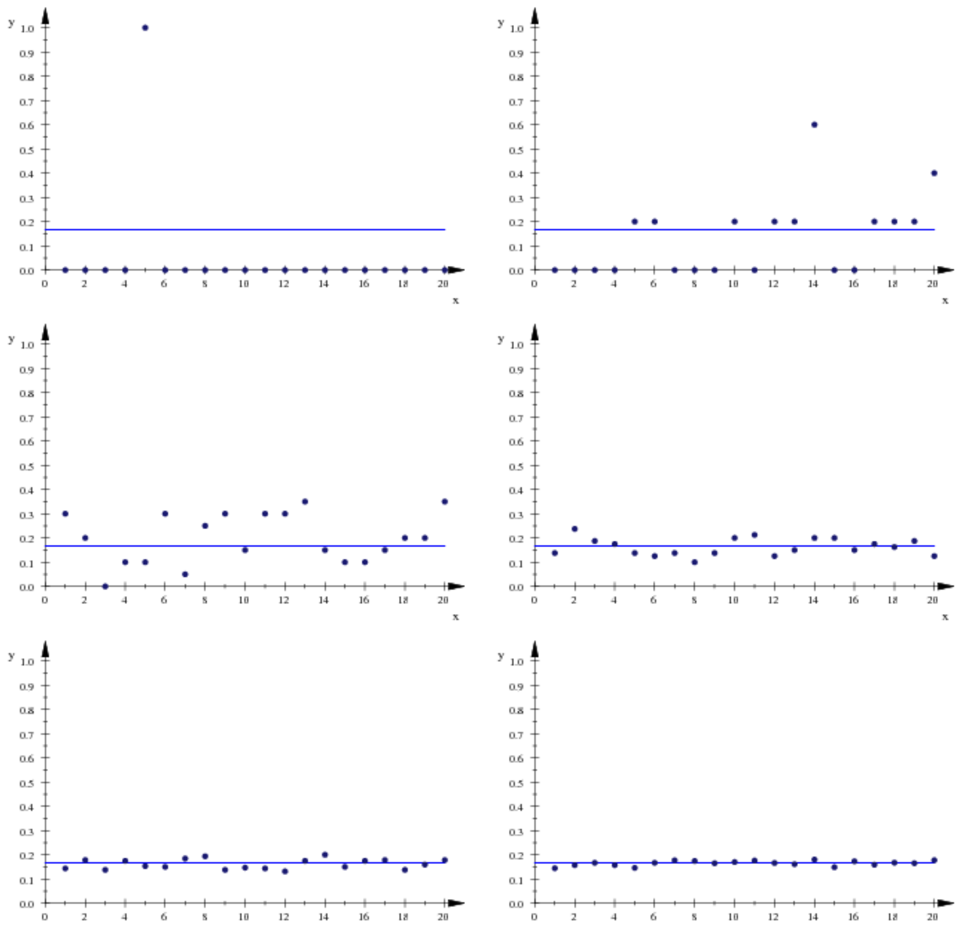


Abbildung 1: Relative Häufigkeit der 5 bei n Würfen für $n = 1, 5, 20, 80, 320, 1280$

Wie du siehst, „stabilisiert sich die relative Häufigkeit der 5 bei n Würfen“ bei wachsendem n. Dass das passiert, ist eine **Erfahrungstatsache**, man spricht vom **empirischen Gesetz der großen Zahl**. Dieses Gesetz ist für die Versicherungen Grundlage ihrer Arbeit.

Die Tendenz der relativen Häufigkeit, bei wachsender Anzahl von Versuchsdurchführungen stabil zu werden, ist erstaunlich stark. Sagen wir, wir machen fünzigmal 100 Würfe mit einem Würfel, stellen jeweils die Anzahl der geworfenen Fünfen fest und berechnen die relative Häufigkeit der 5 nach 100, 200, 300 usw. Würfen. Nach dem empirischen Gesetz der großen Zahl stabilisieren sich die relativen Häufigkeiten, sie werden recht bald nahe bei $\frac{1}{6}$ liegen. Ich mache jetzt Folgendes: Die ersten 100 Würfe führe ich nicht aus, sondern setze die Anzahl der 5 einfach fest. Einmal gehe ich davon aus, dass bei den ersten 100 Würfen überhaupt keine 5 geworfen wurde, und führe die restlichen 49 Wurfserien ordentlich aus, dann gehe ich von 100 Fünfen bei den ersten 100 Würfen aus und mache danach wieder 49 ordentliche Wurfserien. Die Ergebnisse siehst du in der Abbildung 2 links und rechts. Es ist ganz erstaunlich, wie schnell die Störung nicht mehr zu bemerken ist.

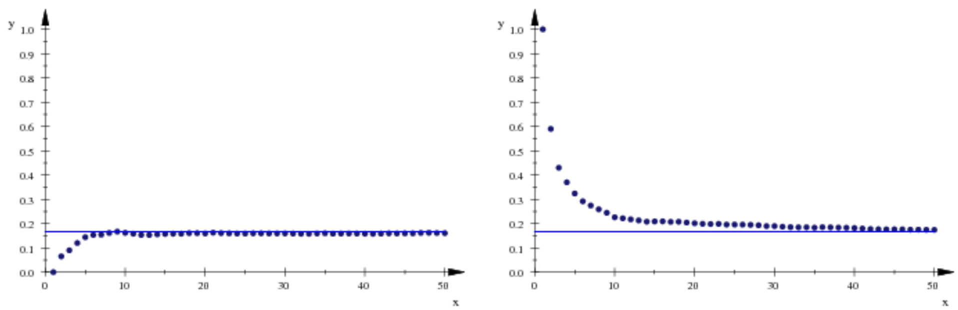


Abbildung 2: Relative Häufigkeit der 5 nach $k \cdot 100$ Würfen. Das Ergebnis der ersten 100 Würfe ist gefälscht, und zwar links auf 0 und rechts auf 100 Fünfen.

2 Grundbegriffe aus der Wahrscheinlichkeitstheorie

2.1 Zufallsversuch, Ergebnis, Ereignis

Man kann

1. einen Würfel einmal werfen,
2. man kann ihn zweimal werfen
3. oder man kann ihn so lange werfen, bis man eine 1 gewürfelt hat.

Das sind drei Beispiele für **Zufallsversuche**. Stelle dir unter einem Zufallsversuch einen Apparat vor, der auf Knopfdruck ein **Ergebnis** ausspuckt. Solche Ergebnisse werden tradionell mit ω bezeichnet. Die Menge aller Ergebnisse, die bei einem Zufallsversuch auftreten können, ist der **Ergebnisraum** Ω des Zufallsversuchs. Sinnvolle Ergebnisräume für die drei Beispiele sind

1. $\Omega = \{ 1, 2, 3, 4, 5, 6 \}$ mit der Mächtigkeit $|\Omega| = 6$.
2. $\Omega = \{ 11, 12, 13, \ldots, 63, 64, 65, 66 \}$ mit der Mächtigkeit $|\Omega| = 36$, dabei bedeutet $\omega = 34$, dass erst eine 3 und dann eine 4 geworfen wurde. Viele bezeichnen dieses ω mit $(3, 4)$. Das ist aufwendiger, hat aber den Vorteil, dass man es als Koordinatenpaar eines Ebenenpunktes auffassen kann, und das ist oft nützlich.
3. $\Omega = \{1; 21, 31, 41, 51, 61, 221, 231, \ldots\}$. Die Elemente dieser Menge haben die Form $\omega = a_1a_2 \ldots a_{n-1}1$ mit einer natürlichen Zahl $n > 1$ und $a_j \in \{2, 3, 4, 5, 6\}$ für $j = 1, 2, \ldots, n-1$. Für $n = 1$ ist $\omega := 1$. Dieses Ω hat (abzählbar) unendlich viele Elemente.

Ereignisse sind beliebige Teilmengen von Ω. Man sagt, ein Ereignis $A \subseteq \Omega$ sei eingetreten, wenn der Zufallsversuch ein Ergebnis $\omega \in A$ ergeben hat, bildlich: wenn der Apparat beim Auslösen ein ω ausgespuckt hat, das in A liegt.

2.2 Über A. Kolmogoroffs Grundlegung der Wahrscheinlichkeitstheorie

Nun kommt die **Wahrscheinlichkeit** ins Spiel. Wie ich euch erzählte, führten Versuche, zu definieren, was Wahrscheinlichkeit ist, genau so wenig zum Ziel wie Euklids Versuche, zu definieren, was ein Punkt ist. Der große David Hilbert beginnt seine Grundlegung der Geometrie so: Ich denke mir ein System von Dingen, die nenne ich Punkte, Geraden und Ebenen... Und dann folgt eine Auflistung von Beziehungen zwischen diesen Dingen: Durch zwei Punkte gibt es genau eine Gerade, und so weiter, und nur diese aufgelisteten Beziehungen dürfen benutzt werden, um Behauptungen zu beweisen. Natürlich stimmt das nicht ganz, Hilbert hat sich ganz sicher Zeichnungen gemacht, um Beweise zu finden, und darin sind Punkte und Geraden, wie in unseren Zeichnungen, aber dass gilt, was er in den Zeichnungen sieht, begründet er penibel nur mit seinen Axiomen – so heißen die formulierten Beziehungen.

Auf die gleiche Weise leistet A. Kolmogoroff die Grundlegung der Wahrscheinlichkeitstheorie. Er hält sich nicht damit auf zu erklären, was Wahrscheinlichkeit ist, sondern er beginnt mit einer Menge Ω und einer Funktion P, die auf Teilmengen von Ω definiert ist und deren Funktionswerte reelle Zahlen sind, und er verlangt, dass diese Funktion P drei Eigenschaften haben soll:

$$P(A) \geq 0 \quad \text{für alle zulässigen } A \subseteq \Omega \tag{1}$$

$$P(\Omega) = 1 \tag{2}$$

$$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B) \tag{3}$$

Ich muss hier etwas unscharf bleiben, welche Teilmengen von Ω zulässig sind – bei uns sind einfach alle Teilmengen von Ω zulässig – und die dritte Bedingung ist auch etwas einfacher als die im Original.

2.3 Wahrscheinlichkeit für Fußgänger

Du hättest wenig Gewinn davon, wenn ich nun versuchte, auf den Axiomen, die A. Kolmogoroff 1933 in einem schmalen Bändchen aus dem Verlag von Julius Springer vorlegte, ein Stück Wahrscheinlichkeitstheorie aufzubauen. Wir gehen von der folgenden einfachen Vorstellung aus:

Die Gesamtwahrscheinlichkeit 1 wird unter den $\omega \in \Omega$ aufgeteilt.

Dann hat schon jedes **Elementarereignis**[1] $\{\omega\}$ seine Wahrscheinlichkeit $P(\{\omega\})$. Für jedes $A \subseteq \Omega$ setzen wir nun

$$P(A) := \sum_{\omega \in A} P(\{\omega\}) \quad , \tag{4}$$

und damit haben wir ein P, das den Forderungen Kolmogoroffs genügt.

Kehren wir zurück auf den sicheren Boden unserer Beispiele. Aus Symmetriegründen werden wir beim ersten Beispiel jedem ω die gleiche Wahrscheinlichkeit geben, es ist also

$$P(\{\omega\}) = \frac{1}{6} \quad \text{für jedes } \omega \in \{1, 2, 3, 4, 5, 6\}.$$

Genauso ist es beim zweiten Beispiel. Dort ist dann

$$P(\{\omega\}) = \frac{1}{36} \quad \text{für jedes } \omega \in \{11, 12, 13, \ldots, 64, 65, 66\}.$$

Bevor ich zum dritten Beispiel komme, will ich dir die $\frac{1}{36}$ des zweiten Beispiels noch auf eine andere Art plausibel machen: Wenn wir 43 würfeln wollen, muss der erste Wurf eine 4 ergeben, das geschieht nur in etwa $\frac{1}{6}$ der Fälle. Und nur in etwa $\frac{1}{6}$ dieses Sechstels wird zusätzlich der zweite Wurf die gewünschte 3 ergeben. Insgesamt erwarten wir die 43 folglich in etwa $\frac{1}{6} \cdot \frac{1}{6}$ aller Fälle. Hier geschieht die Arbeit des Apparats in zwei Teilschritten; die Vorstellung kennst du von den alten Baumdiagrammen.

Kommen wir also zum dritten Beispiel. Auch hier ist natürlich zum Beispiel

$$P(\{23341\}) = \left(\frac{1}{6}\right)^5 \quad ,$$

aber so genau wollen wir es eigentlich nicht wissen. Interessant ist doch, wie wahrscheinlich es ist, dass die erste 1 beim fünften Wurf kommt, und das ist

$$\left(\frac{5}{6}\right)^4 \cdot \frac{1}{6} \quad ,$$

[1] Ein Elementarereignis ist eine Teilmenge von Ω, die nur ein einziges ω enthält

wie du dir mit Hilfe eines Baumdiagramms klarmachen kannst. Allgemein ist die Wahrscheinlichkeit, dass die erste 1 beim k–ten Wurf kommt,

$$\left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6} \quad \text{für } k = 1, 2, 3, \dots$$

2.4 Zufallsgrößen: Erster Kontakt

Eine Zufallsgröße ist eine Abbildung (oder Funktion) $X : \Omega \to \mathbb{R}$. Was das soll, zeige ich dir an zwei Beispielen. Nimm das dritte von oben: Würfeln, bis eine 1 kommt. Eigentlich will man ja nur wissen, wie oft man werfen musste, und nicht, was vor der 1 alles vorkam. An $234331 \in \Omega$ interessiert eigentlich nur die Länge 6 dieses Wortes. Also wird diesem ω die Zahl 6 zugeordnet. Schon hast du die Funktion $X : \Omega \to \mathbb{R}$ des dritten Beispiels:

$$X(a_1 a_2 \dots a_{k-1} 1) = k$$

Nun will man von der Wahrscheinlichkeit sprechen, dass X den Wert k annimmt. Dafür schreibt man frech $P(X = k)$. Aber P wirkt nur auf Teilmengen von Ω. Die Teilmenge, die man hier braucht, ist die Menge aller $\omega \in \Omega$ mit $X(\omega) = k$, und diese Menge bezeichnet man mit $X^{-1}(k)$:

$$X^{-1}(k) := \{\omega \in \Omega \,|\, X(\omega) = k\} \tag{5}$$

Hier erhalten wir

$$P(X = k) = \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6} \quad \text{für } k = 1, 2, 3, \dots \tag{6}$$

Nehmen wir ein zweites Beispiel: Wir würfeln zweimal und addieren die Augenzahlen. Die Summe ist der Wert von X. Zum Beispiel ist $X(34) = 7$. Wir haben uns überlegt, mit welchen Wahrscheinlichkeiten X die möglichen Werte annimmt, und haben diese in einer Tabelle notiert:

k	2	3	4	5	6	7	8	9	10	11	12
$P(X = k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Dies ist die Wertetabelle der Funktion $k \mapsto P(X = k)$, der **Wahrscheinlichkeitsverteilung** von X. Zur graphischen Darstellung von Wahrscheinlichkeitsverteilungen verwendet man **Histogramme**. Abbildung 3 zeigt das Histogramm der Zufallsgröße X des Beispiels. Auf den X–Werten stehen Rechtecke, die Wahrscheinlichkeit der Werte ist jeweils der Flächeninhalt der Rechtecke.

2.5 Technische Übung

Wir wollen einen Würfel N–mal werfen. Ein passender Ergebnisraum zu diesem Zufallsversuch ist

$$\Omega_N := \{a_1 a_2 \dots a_N \,|\, a_1, a_2, \dots, a_N \in \{1, 2, 3, 4, 5, 6\}\} \quad ,$$

oder, in etwas anderer Schreibweise, die Menge der N–Tupel mit Einträgen in $\{1, 2, 3, 4, 5, 6\}$:

$$\{(a_1, a_2, \dots, a_N) \,|\, a_1, a_2, \dots, a_N \in \{1, 2, 3, 4, 5, 6\}\} \quad .$$

Nun kann es sein, dass in den N Würfen keine einzige 1 geworfen wurde. Dann ist das Ereignis

$$A = \{a_1 a_2 \dots a_N \,|\, a_1, a_2, \dots, a_N \in \{2, 3, 4, 5, 6\}\}$$

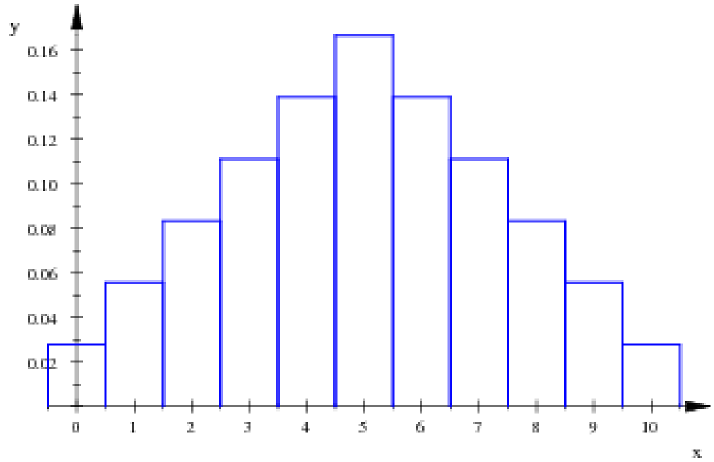


Abbildung 3: Histogramm der Zufallsgröße X: Summe zweier gewürfelter Zahlen

eingetreten. Die Wahrscheinlichkeit dafür ist

$$P(A) = \frac{5^N}{6^N} = \left(\frac{5}{6}\right)^N \quad .$$

Die beiden Terme für $P(A)$ kommen auf ganz verschiedene Weise zu Stande: Der zweite Term ergibt sich aus dem Baumdiagramm, das ist kein Geheimnis. Der erste Term spiegelt eine ganz andere Denkweise wider. Und zwar ist offenbar $|\Omega_N| = 6^N$ und $|A| = 5^N$. Nun haben aus Symmetriegründen alle $\omega \in \Omega_N$ die gleiche Wahrscheinlichkeit, deshalb ist

$$P(A) = \frac{|A|}{|\Omega_N|} = \frac{5^N}{6^N} \quad .$$

Das so genannte **Gegenereignis** oder, in Mengensprache, das **Komplement**

$$\bar{A} := \Omega_N \setminus A := \{\omega \in \Omega_N \,|\, \omega \notin A\}$$

tritt ein, wenn bei den ersten N Würfen mindestens eine 1 erschien. Für die Wahrscheinlichkeiten gilt

$$P(\bar{A}) = 1 - P(A) \quad . \tag{7}$$

Diese Beziehung ist sehr nützlich. Oft ist es viel einfacher, $P(\bar{A})$ zu berechnen und daraus $P(A)$ zu bestimmen, als $P(A)$ selbst direkt zu berechnen. Hier kennen wir auch $P(\bar{A})$, denn wenn in den N Würfen mindestens eine 1 erscheint, muss die erste 1 in einem der Würfe 1 bis N geworfen werden, und die Wahrscheinlichkeit dafür ist nach Gleichung (6)

$$\sum_{k=1}^{N} \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6} \quad .$$

Setzen wir die Terme für $P(A)$ und $P(\bar{A})$ in die Gleichung (7) ein, ergibt das mit der Abkürzung $q = \frac{5}{6}$ und $1 - q = \frac{1}{6}$

$$\sum_{k=1}^{N} q^{k-1}(1-q) = 1 - q^N \quad ,$$

und daraus folgt über

$$\sum_{k=1}^{N} q^k = \frac{1-q^N}{1-q}$$

die **Summenformel für die endliche geometrische Reihe**

$$\sum_{k=0}^{n} q^k = \frac{1-q^{n+1}}{1-q} \quad \text{für } q \neq 1 \ . \tag{8}$$

Für $|q| < 1$ kann man den Grenzwert für n gegen Unendlich bilden, das ergibt die Summenformel für die geometrische Reihe

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \quad \text{für } |q| < 1 \ . \tag{9}$$

Im Unterricht hatten wir die Formel auf andere Weise hergeleitet, hier ist sie mit Mitteln der Stochastik bewiesen.

2.6 Ein wenig Kombinatorik

Kombinatorik ist die Kunst, geschickt zu zählen. Man kann sich lange damit beschäftigen, aber wir brauchen nicht viel davon.

Unsere Grundmenge ist eine Menge M der Mächtigkeit n. Wir können ruhig

$$M = \{1, 2, 3, \ldots, n\}$$

nehmen; was damit geht, geht mit allen Mengen der Mächtigkeit n. Mit den Elementen von M konstruieren wir verschiedene Gebilde.

Wir beginnen mit der Menge der k–**tupel**

$$\{(a_1, a_2, \ldots, a_k) \,|\, a_1, a_2, \ldots, a_k \in M\}$$

von Elementen aus M. Die kannst du dir vorstellen als die Menge aller k–stelligen Zahlen mit der Ziffernmenge M oder die Menge aller Wörter der Länge k aus dem Alphabet M. Die Mächtigkeit dieser Menge ist offensichtlich n^k.

Nun nehmen wir der Reihe nach k Gegenstände aus M und legen sie so vor uns hin, dass die Reihenfolge, in der sie gezogen wurden, noch zu erkennen ist. Die Anzahl der Möglichkeiten ist

$$n \cdot (n-1) \cdot (n-2) \cdot \ \ldots \ \cdot (n-(k-1)) \ \ .$$

Für $k = n$ ist das $n!$; die Anzahl der Möglichkeiten, die Elemente von M in einer Reihe anzuordnen, oder die Anzahl der **Permutationen** von M ist $n!$. Für $k = 0$ nimmt man überhaupt keinen der Gegenstände, das geht nur auf eine Weise, also ist die gesuchte Anzahl dann 1.

Nun nehmen wir k der Gegenstände heraus, beachten aber die Reihenfolge nicht, in der wir sie gewählt haben. Das heißt, wir fragen nach der Anzahl der k–elementigen Teilmengen von M oder, allgemeiner gesagt, nach der Anzahl der k–Teilmengen einer n–Menge.

Soll die k–Teilmenge $\{1, 2, 3, \ldots, k\}$ von M gezogen werden, geht das, wenn wir die Reihenfolge beachten, auf $k!$ Möglichkeiten. In der Anzahl

$$n \cdot (n-1) \cdot \ \ldots \ \cdot (n-(k-1))$$

der Möglichkeiten, k der n Elemente von M unter Beachtung der Reihenfolge auszuwählen, ergeben jeweils $k!$ Anordnungen die gleiche Teilmenge von M. Die **Anzahl der k–Teilmengen einer n–Menge** ist folglich

$$\frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-(k-1))}{k!} =: \binom{n}{k} \quad . \tag{10}$$

Man liest dies „n über k" und bezeichnet den Term als **Binomialkoeffizienten**.

Erweitert man den Bruch in Gleichung (10) mit $(n-k)!$, kommt man zu der eleganten, für praktische Rechnungen aber wenig geeigneten Darstellung

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} \quad . \tag{11}$$

1 Lemma
Für Binomialkoeffizienten gilt die folgende Regel:

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$$

Beweis. Wir denken uns eine Herde aus n weißen und einem schwarzen Schaf. Die Anzahl der $(k+1)$–Teilmengen dieser $(n+1)$–Menge steht auf der rechten Seite der Gleichung. Es gibt nun zwei Sorten dieser $(k+1)$–Teilmengen: die, die das schwarze Schaf nicht enthalten und die, die es enthalten. Bei der ersten Sorte sind alle $k+1$ Schafe weiß, von dieser Sorte Mengen gibt es natürlich $\binom{n}{k+1}$ Stück. Bei der zweiten Sorte ist das schwarze Schaf gesetzt, man nimmt noch k weiße Schafe hinzu. Von diesen Ergänzungsmengen gibt es $\binom{n}{k}$ Stück. Dies beweist das Lemma. □

Wenn du die Binomialkoeffizienten zu n und zu $n+1$ geeignet untereinanderschreibst, kannst du die der $(n+1)$–Zeile gewinnen, indem du immer benachbarte der n–Zeile addierst; du musst nur zu Beginn und am Ende jeweils eine 1 hinzufügen. Führst du dies ab $n = 0$ durch, entsteht das **Pascalsche Dreieck**. Die kennst du vielleicht von den binomischen Formeln her. In der Tat gilt der folgende Satz; er erklärt auch den Namen Binomialkoenffizient.

2 Satz (Binomischer Satz)

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Die Beweisidee habe ich euch gezeigt, hier spare ich mir die.

3 Über Zufallsgrößen

3.1 Der Erwartungswert

Xaver hat in seiner Schublade sieben braune und drei rote Socken. Er greift blind in die Schublade und zieht einen Socken nach dem anderen, bis er ein gleichfarbiges Paar hat. Unsere Zufallsgröße X bezeichnet die Anzahl der nötigen Ziehungen.

Offensichtlich braucht er mit der Wahrscheinlichkeit

$$p := \frac{7}{10} \cdot \frac{6}{9} + \frac{3}{10} \cdot \frac{2}{9} = \frac{48}{90} = \frac{8}{15}$$

zwei Ziehungen und mit der Gegenwahrscheinlichkeit $q := 1 - p = \frac{7}{15}$ drei Ziehungen, das kannst du mit Hilfe eines Baumdiagramms nachprüfen. Wir stellen uns nun auf den etwas weltfremden Standpunkt, dass Xaver diese Zeremonie jeden Morgen durchführt, und fragen uns, wie oft Xaver im Mittel jeden Morgen ziehen muss.

Schauen wir Xaver N Tage lang zu. Eine Anzahl a Durchführungen wird er zwei Ziehungen benötigen, eine Anzahl $b = N - a$ Durchführungen drei Ziehungen. Insgesamt hat er dann $2a + 3b$ Socken gezogen, das sind je Durchführung

$$\frac{2a + 3b}{N} = 2 \cdot \frac{a}{N} + 3 \cdot \frac{b}{N}$$

gezogene Socken. Nach dem empirischen Gesetz der großen Zahl ist nun

$$\frac{a}{N} \approx p = \frac{8}{15} \quad \text{und} \quad \frac{b}{N} \approx q = \frac{7}{15} \quad ,$$

die Zahl der im Mittel je Durchführung gezogenen Socken ist folglich etwa

$$2p + 3q \quad .$$

Das Sockenbeispiel ist zugegebenermaßen etwas abwegig. Realistischer wäre so etwas: Um Geld für die Mensa zu sammeln, muss jeder Kurs beim Schulfest einen Stand machen. Da könnte man eine Würfelbude einrichten. Geworfen wird mit einem Würfel. Bei einer $1, 2$ oder 3 gibt es nichts, bei einer 4 oder 5 gibt es einen Euro, bei einer 6 gibt es fünf Euro. Wieviel muss man dann im Mittel je Spiel auszahlen? Diese Frage kann man genau wie das Sockenproblem behandeln.

So, das empirische Gesetz der großen Zahl gehört zur physikalischen Welt. In unserer Wahrscheinlichkeitstheorie definieren wir jetzt einfach einen Begriff, der das nachbilden soll, was die Erfahrung lehrt, nämlich den Erwartungswert einer Zufallsgröße.

3 Definition
Es sei $X : \Omega \to \mathbb{R}$ eine Zufallsgröße. Wir definieren den **Erwartungswert** *der Zufallsgröße durch*

$$\mathrm{E}(X) := \sum_{k \in X(\Omega)} kP(X = k) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) \quad .$$

Anmerkung: Häufig verwendet man für $\mathrm{E}(X)$ die Bezeichnung μ.

3.2 Zwei Regeln für Erwartungswerte

Nehmen wir noch einmal das Würfelspiel für das Schulfest: Ein Würfel wird geworfen, bei einer 1,2 oder 3 gibt es nichts, bei einer 4 oder 5 gibt es einen Euro, bei einer 6 fünf Euro. Es sei A der Betrag, den man auszahlen muss. Dann ist

$$\mathrm{E}(A) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{3} + 5 \cdot \frac{1}{6} = \frac{7}{6}$$

der Erwartungswert dieser Zufallsgröße. Nun kann ein fürsorgliches Elternteil bei der Schulkonferenz erreichen, dass die Auszahlungsbeträge halbiert werden müssen. Dann entsteht eine neue Zufallsgröße $X = \frac{1}{2}A$ vermöge

$$X(\omega) := \frac{1}{2}A(\omega) \quad \text{für alle } \omega \in \Omega,$$

und du wirst vermuten, dass $\mathrm{E}(\frac{1}{2}A) = \frac{1}{2}\,\mathrm{E}(A)$ ist.

Wer an dem Spiel teilnimmt, will vielleicht nicht den Erwartungswert des ausgezahlten Betrages wissen, sondern den Erwartungswert des Gewinns, das ist die Differenz aus Auszahlung und Einsatz. Beträgt der Einsatz zwei Euro, ist nun eine neue Zufallsgröße $G = A - 2$ gegeben. Die Definition ist

$$G(\omega) := A(\omega) - 2 \quad \text{fur alle } \omega \in \Omega \ .$$

Auch hier wirst du $\mathrm{E}(G) = \mathrm{E}(A) - 2$ vermuten. In der Tat gilt das folgende Lemma.

4 Lemma
Es sei $X : \Omega \to \mathbb{R}$ eine Zufallsgröße, und es sei $a, b \in \mathbb{R}$. Dann ist

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

Beweis. Es ist

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

denn den konstanten Faktor b kann man aus der Summe ziehen, und die Summe über alle $P(\{\omega\})$ ist natürlich 1. Der Beweis der zweiten Gleichung geht analog. □

3.3 Die Varianz einer Zufallsgröße

Es sei X eine Zufallsgröße auf dem Ergebnisraum Ω mit Erwartungswert $\mu = \mathrm{E}(X)$. Führen wir den zugehörigen Zufallsversuch sehr oft durch, erhalten wir eine Folge

$$\omega_1, \omega_2, \omega_3, \ldots, \omega_N$$

von Elementen von Ω, und der Durchschnittswert

$$\frac{1}{N} \sum_{k}^{N} X(\omega_k)$$

der zugehörigen Werte $x_k = X(\omega_k)$ sollte nahe bei μ liegen[2]. Je nach Zufallsversuch muss das N sehr groß sein; die Werte von X können erheblich streuen, und in der Nähe von μ brauchen überhaupt keine Werte von X zu liegen.

Als Maß dafür, wie stark die Werte von X streuen, benutzt man die Varianz von X.

[2] Wieso eigentlich?

5 Definition
Es sei $X : \Omega \to \mathbb{R}$ eine Zufallsgröße mit Erwartungswert $\mu := \mathrm{E}(X)$. Dann nennt man die Zahl

$$\mathrm{V}(X) := \sum_{k \in X(\Omega)} (k-\mu)^2 P(X=k) \;=\; \sum_{\omega \in \Omega} \left(X(\omega)-\mu\right)^2 P(\{\omega\})$$

die **Varianz** *und deren Wurzel $\sigma := \sqrt{\mathrm{V}(X)}$ die* **Standardabweichung** *von X.*

Wenn du dich nicht zu dumm anstellst, kannst du nun die Varianz und die Standardabweichung einer konkret gegebenen Zufallsgröße ausrechnen; dann hast du zwei Zahlen. Marvins Frage, ob man denn diese Zahlen am Histogramm sehen könne, musste ich leider verneinen. Die Werte, die X annehmen kann, liegen irgendwo auf der Merkmalsachse. Du kannst auf der Merkmalsachse das μ einzeichnen, für jedes $k \in X(\Omega)$ das $k-\mu$ sehen, dir das Quadrat mit dem Inhalt $(k-\mu)^2$ vorstellen und dir den noch mit $P(X=k)$ multipliziert denken. Wenn all diese Produkte aufaddiert sind, ist die Varianz berechnet, aber man sieht nichts mehr. Immerhin mag dir der folgende Satz einen Hinweis geben, dass die Varianz eine sehr nützliche Begriffsbildung ist.

6 Satz (Tschebyscheff–Ungleichung)
Es sei $X : \Omega \to \mathbb{R}$ eine Zufallsgröße mit Erwartungswert μ und Varianz $\mathrm{V}(X)$. Dann gilt

$$P(|X-\mu| \geq a) \;\leq\; \frac{\mathrm{V}(X)}{a^2} \qquad \text{für alle } a > 0.$$

In Worten: Die Wahrscheinlichkeit, dass X einen Wert annimmt, der von μ um mindestens a abweicht, ist höchstens so groß wie das **Tschebyscheff–Risiko** $\mathrm{V}(X)/a^2$.

Beweis. Wenn du den Beweis verstehen willst, solltest du dir die Merkmalsachse hinzeichnen und darin den Bereich von $\mu-a$ bis $\mu+a$ markieren; der Erwartungswert μ liegt in der Mitte dieses Bereichs. Die X–Werte im Inneren des Bereiches färbst du grün, die außerhalb rot. Der Rand selbst gehört zum roten Bereich. In der Formel für die Varianz musst du über alle X–Werte summieren. Du lässt die Beiträge der grünen X–Werte einfach weg, dann erhältst du höchstens noch die Varianz, in aller Regel aber weniger. In den verbleibenden Summanden ersetzt du $(k-\mu)^2$ durch a^2, dann wird wiederum jeder Summand höchstens kleiner. Du klammerst du noch a^2 aus und dividierst durch a^2, schon steht das Ergebnis da. Klar? Also los:

$$\begin{aligned}
\mathrm{V}(X) &= \sum_{k \in X(\Omega)} (k-\mu)^2 P(X=k) \\
&= \sum_{\substack{k \in X(\Omega) \\ |k-\mu|<a}} (k-\mu)^2 P(X=k) \;+\; \sum_{\substack{k \in X(\Omega) \\ |k-\mu|\geq a}} (k-\mu)^2 P(X=k) \\
&\geq \qquad 0 \qquad + \sum_{\substack{k \in X(\Omega) \\ |k-\mu|\geq a}} a^2 P(X=k) \\
&= a^2 \cdot \sum_{\substack{k \in X(\Omega) \\ |k-\mu|\geq a}} P(X=k) \\
&= a^2 P(|X-\mu| \geq a)
\end{aligned}$$

Es gilt also insgesamt

$$\mathrm{V}(X) \geq \quad a^2 P(|X-\mu| \geq a) \quad .$$

Division durch a^2 liefert die gewünschte Ungleichung. $\square$

Wenn du dir die Definition der Varianz genau anschaust, erkennst du, dass

$$\mathrm{V}(X) = \mathrm{E}((X-\mu)^2) \tag{12}$$

ist. Wir formen die rechte Seite um und nutzen dabei aus, was wir über das Rechnen mit Erwartungswerten wissen:

$$\begin{aligned}
\mathrm{V}(X) &= \mathrm{E}((X-\mu)^2) \\
&= \mathrm{E}(X^2 - 2\mu X + \mu^2) \\
&= \mathrm{E}(X^2) - 2\mu\,\mathrm{E}(X) + \mu^2 \\
&= \mathrm{E}(X^2) - \mu^2
\end{aligned}$$

Im Unterricht hatten wir dieses Resultat direkt aus der Definition der Varianz gewonnen, dies hier ist jetzt ein etwas anderer Weg. Es gilt jedenfalls das folgende Lemma.

7 Lemma (Verschiebungssatz der Varianz)
Für die Varianz einer Zufallsgröße X mit Erwartungswert μ gilt

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

3.4 Mehrere Zufallsgrößen auf dem gleichen Wahrscheinlichkeitsraum

Es seien X und Y Zufallsgrößen, die auf dem gleichen Ω definiert sind. Die inhaltsschwere Formulierung „auf dem gleichen Wahrscheinlichkeitsraum“ stellt sicher, dass X und Y nicht nur auf der gleichen Menge definiert sein sollen, sondern dass man auch das gleiche P zum Berechnen von Wahrscheinlichkeiten verwendet. Wir bilden die neue Zufallsgröße

$$S = X + Y \quad \text{mit} \quad S(\omega) = X(\omega) + Y(\omega) \quad \text{für alle } \omega \in \Omega.$$

Dann gilt

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y) \quad . \tag{13}$$

Der Nachweis ist leicht:

$$\begin{aligned}
\mathrm{E}(X+Y) = \sum_{\omega\in\Omega} S(\omega)P(\{\omega\}) &= \sum_{\omega\in\Omega} (X(\omega)+Y(\omega))\,P(\{\omega\}) = \\
&= \sum_{\omega\in\Omega} X(\omega)P(\{\omega\}) + \sum_{\omega\in\Omega} Y(\omega)P(\{\omega\}) = \mathrm{E}(X) + \mathrm{E}(Y)
\end{aligned}$$

Mit dieser einfachen Regel kann man erstaunlich viel anfangen, wie das folgende Anwendungsbeispiel zeigt: Wir würfeln N–mal, es sei X die Anzahl der gewürfelten Sechsen, und wir fragen nach dem Erwartungswert von X. Eine Formel für die Wahrscheinlichkeit, k Sechsen zu würfeln, ist schnell hingeschrieben:

$$P(X=k) = \binom{N}{k} \cdot \left(\frac{1}{6}\right)^k \cdot \left(\frac{5}{6}\right)^{N-k} \qquad \text{für } k = 0,1,2,\ldots,N,$$

aber es ist schon für ziemlich kleine N ein echtes Problem, die Werte der Terme zu berechnen. Hier hilft nun unsere Regel: Wir bilden für $k = 1,2,3,\ldots,N$ die Zufallsgrößen

$$X_k: \quad \text{Anzahl der Sechsen im } k\text{–ten Wurf} \quad ,$$

dann ist

$$X = X_1 + X_2 + X_3 + \cdots + X_N \quad ,$$

und nach unserer Regel ergibt sich

$$\mathrm{E}(X) = \mathrm{E}(X_1) + \mathrm{E}(X_2) + \mathrm{E}(X_3) + \cdots + \mathrm{E}(X_N) = N \cdot \mathrm{E}(X_1) = N \cdot \frac{1}{6} \quad ,$$

und das ist doch ein schönes Ergebnis.

3.5 Durchgerechnetes Beispiel

Wir wollen für die Zufallsgröße X_1 : Anzahl der Sechsen beim einfachen Würfeln zeigen, wie Erwartungswert und Varianz ökonomisch berechnet werden können. Man legt eine Tabelle an:

k	0	1
$P(X_1 = k)$	$\frac{5}{6}$	$\frac{1}{6}$
$kP(X_1 = k)$	0	$\frac{1}{6}$
$k^2P(X_1 = k)$	0	$\frac{1}{6}$

Die Summe der zweiten Zeile liefert den Erwartungswert von X_1

$$\mu = \mathrm{E}(X_1) = 0 + \frac{1}{6}$$

und die Summe der dritten Zeile den Erwartungswert von X_1^2 :

$$\mathrm{E}(X_1^2) = 0 + \frac{1}{6} = \frac{1}{6} \quad .$$

Aus diesen Werten berechnet man die Varianz von X_1

$$\mathrm{V}(X_1) = \mathrm{E}(X_1^2) - \mu^2 = \frac{1}{6} - \frac{1}{36} = \frac{5}{36} \quad .$$

Hier siehst du das Histogramm von X_1, der Erwartungswert ist auch eingezeichnet.

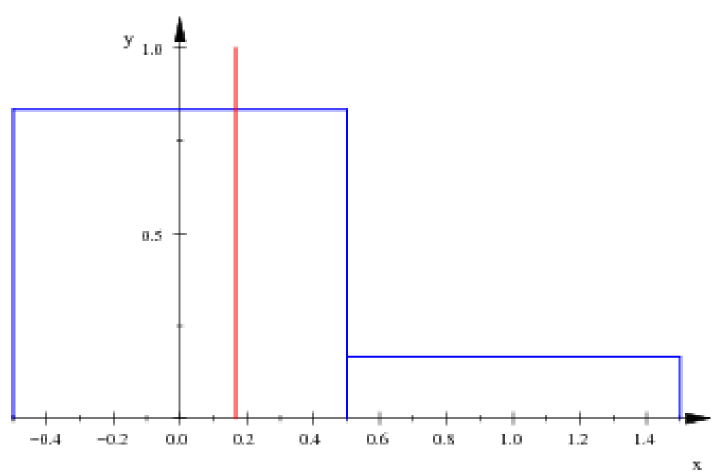


Abbildung 4: Histogramm von X_1 mit Erwartungswert

Was kann man mit diesen Werten anfangen? Wir betrachten die Zufallsgröße X : Anzahl der Sechsen, wenn man 600–mal würfelt, das ist das X aus dem letzten

Unterkapitel mit $N = 600$. Es ist $\mathrm{E}(X) = 600 \cdot \frac{1}{6} = 100$ und, dies nehmen wir einmal vorweg, $\mathrm{V}(X) = 600 \cdot \mathrm{V}(X_1) = \frac{250}{3}$. Müssen wir ernsthaft damit rechnen, dass wir mindestens 150 oder höchstens 50 Sechsen bekommen? Die Tschebyscheff–Ungleichung sagt

$$P(|X - 100| \geq 50) \leq \frac{\mathrm{V}(X)}{50^2} = \frac{1}{30} \quad ;$$

die Wahrscheinlichkeit, vom Erwartungswert um mindestens 50 abzuweichen, ist sicher kleiner als $3\frac{1}{3}\%$. Das Tschebyscheff–Risiko überschätzt die Wahrscheinlichkeit in der Regel, so auch hier, wie du am Histogramm von X erkennen kannst. So ein Histogramm bekommst du mit ein paar MuPAD–Befehlen, aber darin steckt eine enorme Rechenarbeit!

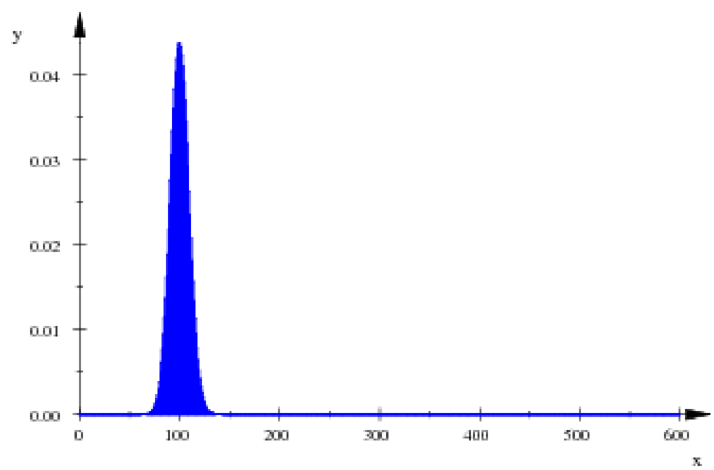


Abbildung 5: Histogramm von X: Anzahl der Sechsen bei 600–mal Würfeln

3.6 Mehr zu Standardabweichung und Varianz

Es sei X eine Zufallsgröße auf Ω mit Erwartungswert μ und Varianz $\mathrm{V}(X)$. Um etwas klarer zu machen, was es mit der Standardabweichung $\sigma = \sqrt{\mathrm{V}(X)}$ von X auf sich hat, schreibe ich die Tschebyscheff–Ungleichung mit $a = r\sigma$ hin für eine positive Zahl r; ich messe die Abweichung a also praktisch in der Einheit σ. Heraus kommt folgendes:

$$P(|X - \mu| \geq r\sigma) \quad \leq \quad \frac{\mathrm{V}(X)}{r^2\sigma^2} \quad = \quad \frac{1}{r^2} \tag{14}$$

Kannst du damit etwas anfangen? Schau: Die Wahrscheinlichkeit, dass X einen Wert annimmt, der von μ mindestens 10σ entfernt ist, ist allerhöchstens 1%. Oder anders: Beim Histogramm von X ist der Inhalt der beiden Schwänze, die aus der 3σ–Umgebung von μ herausschauen, allerhöchstens $\frac{1}{9}$. Und das gilt ganz allgemein für jedes X (das eine Varianz hat).

Dies zum Verständnis, nun noch ein technisches Lemma.

8 Lemma
Es sei $X : \Omega \to \mathbb{R}$ eine Zufallsgröße mit Erwartungswert μ und Varianz $\mathrm{V}(X)$. Dann ist

$$\mathrm{V}(aX) = a^2\,\mathrm{V}(X) \quad \textit{für alle } a \in \mathbb{R} \textit{ und}$$
$$\mathrm{V}(X + b) = \mathrm{V}(X) \quad \textit{für alle } b \in \mathbb{R}\ .$$

Beweis. Wir wenden die Definition der Varianz auf die Zufallsgröße $aX : \omega \mapsto aX(\omega)$ an – dass letztere den Erwartungswert $a\mu$ hat, wissen wir ja schon.

$$\begin{aligned}\mathrm{V}(aX) &= \sum_{\omega \in \Omega} (aX(\omega) - a\mu)^2 \cdot P(\{\omega\}) \\ &= a^2 \sum_{\omega \in \Omega} (X(\omega) - \mu)^2 \cdot P(\{\omega\}) \\ &= a^2\,\mathrm{V}(X)\end{aligned}$$

Damit ist die erste Eigenschaft bewiesen.

Um einzusehen, dass die Varianz auch die zweite Eigenschaft hat, brauchen wir ebenfalls nur die Definition hinzuschreiben:

$$\mathrm{V}(X+b) = \sum_{\omega \in \Omega} (X(\omega) + b - (\mu + b))^2 \cdot P(\{\omega\})$$

Das b geht sofort weg, es bleibt $\mathrm{V}(X)$. □

3.7 Noch ein durchgerechnetes Beispiel

Sechs von euch sitzen mit dem Rücken zur Wand, achtzehn mit dem Rücken zu einem Fenster. Es werden fünf von euch zufällig gewählt. Mit welcher Wahrscheinlichkeit sitzen (genau) zwei davon mit dem Rücken zur Wand?

Damit ich mich leichter ausdrücken kann betrachte ich anstelle des Kurses eine Urne mit achtzehn weißen und sechs roten Kugeln. Dieser Urne werden fünf Kugeln entnommen, und es sei X die Anzahl der roten Kugeln darunter. Gefragt ist dann nach $P(X = 2)$.

Einer von euch sagte, er verstehe die Aufgaben nicht. Mein Rat ist: stellt euch den Ablauf des Experimentes genau vor, lasst es vor euren Augen ablaufen. Hier gibt es dann zwei mögliche Wege. Man greift in die Urne und nimmt fünf Kugeln auf einmal heraus oder man nimmt eine Kugel nach der anderen, bis man fünf hat.

Die zweite Vorstellung führt zu einem Baumdiagramm. Wir erhalten

$$P(X = 2) = \frac{6}{24} \cdot \frac{5}{23} \cdot \frac{18}{22} \cdot \frac{17}{21} \cdot \frac{16}{20} \cdot \binom{5}{2} \quad .$$

Hier wurden zuerst zwei rote, dann drei weiße Kugeln gezogen. Verändert man die Reihenfolge, verändern sich die einzelnen Brüche, nicht aber der Wert des Produktes, deshalb haben alle Pfade, die zu genau zwei gezogenen roten Kugeln gehören, die gleiche Wahrscheinlichkeit. Der Binomialkoeffizient am Ende bringt die Anzahl der Pfade ins Spiel.

Wer alle Kugeln auf einmal herausnimmt, muss anders denken. Er wählt dann eine 5–Teilmenge einer 24–Menge, und von denen gibt es $\binom{24}{5}$ Stück, die alle gleich wahrscheinlich sind. Um eine solche mit genau zwei roten Kugeln zu erstellen, haben wir $\binom{5}{2} \cdot \binom{18}{3}$ Möglichkeiten, denn wir können jede 2–Teilmenge roter Kugeln mit jeder 3–Teilmenge weißer Kugeln kombinieren. Wer so denkt, erhält

$$P(X = 2) = \frac{\binom{5}{2} \cdot \binom{18}{3}}{\binom{24}{5}} \quad .$$

Natürlich haben beide Terme den gleichen Wert, das ist völlig klar.

Der zweite Term lässt sich besonders leicht in MuPAD eingeben. Man bekommt dann bequem[3] die Wahrscheinlichkeitsverteilung, den Erwartungswert $\mu = \frac{5}{4}$ und die Varianz $\mathrm{V}(X) = \frac{285}{368}$.

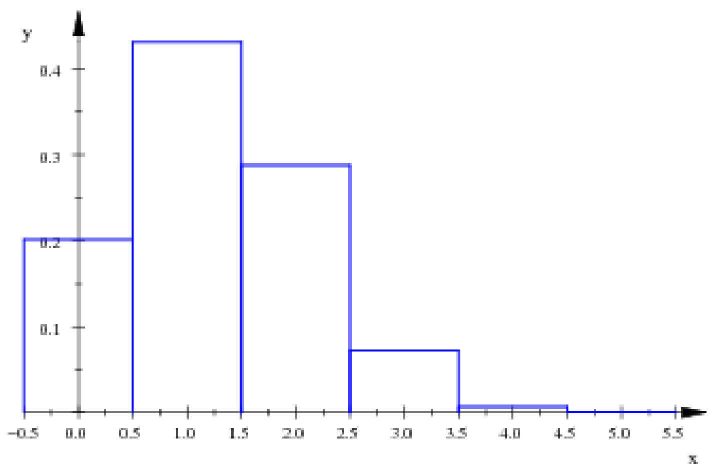


Abbildung 6: Histogramm von X: Anzahl der roten Kugeln unter fünf

Offensichtlich ist die Wahrscheinlichkeit, dass die erste gezogene Kugel rot ist, gerade $\frac{1}{4}$. Mit Hilfe eines Baumdiagramms habt ihr ausgerechnet, dass auch die Wahrscheinlichkeit, dass die zweite gezogene Kugel rot ist, $\frac{1}{4}$ ist, und einige haben das gleiche Ergebnis auch wieder für die Wahrscheinlichkeit erhalten, dass die dritte gezogene Kugel rot ist. Das ist schon etwas erstaunlich, aber es stimmt sogar noch für die vierte und für die fünfte gezogene Kugel. Das kann man mit vollständiger Induktion beweisen, aber das machen wir von mir aus nicht, sondern wir behelfen uns mit der folgenden Betrachtung. Stelle dir vor, die fünf Kugeln werden der Reihe nach in Fächer mit den Nummern eins bis fünf gelegt. Betrachte nun ein Fach, meinetwegen das Fach mit der Nummer vier. Bevor der Ziehvorgang begonnen hat, hat jede der 24 Kugeln die gleiche Chance, in das Fach vier zu kommen. Deshalb ist die Wahrscheinlichkeit, dass eine rote Kugel in das Fach vier kommt, gerade $\frac{1}{4}$.

Bezeichnen wir für $i = 1, 2, \ldots, 5$ mit X_i wieder die Anzahl der roten Kugeln bei der i–ten Ziehung (oder die Anzahl der roten Kugeln im i–ten Fach), gilt $\mathrm{E}(X_i) = \mathrm{E}(X_i^2) = \frac{1}{4}$ und $\mathrm{V}(X_i) = \frac{1}{4} - \frac{1}{16} = \frac{3}{16}$. Aus

$$X = \sum_{i=1}^{5} X_i$$

folgt nach unserer Summenregel für Erwartungswerte $E(X) = \frac{5}{4}$. Jedoch ist

$$5 \cdot \mathrm{V}(X_i) = 5 \cdot \frac{3}{16} \neq \frac{285}{368} = \mathrm{V}(X) \quad ,$$

es kann also keine Regel der Art $\mathrm{V}(X + Y) = \mathrm{V}(X) + \mathrm{V}(Y)$ für beliebige Zufallsgrößen X, Y auf dem gleichen Wahrscheinlichkeitsraum geben; diese Regel gilt nur unter gewissen Voraussetzungen, die X und Y erfüllen müssen.

[3] Siehe stochastik.mn

3.8 Das schwache Gesetz der großen Zahl

Wenn wir einen wirklichen Würfel N–mal werfen und die Anzahl H der Einsen zählen, wird die relative Häufigkeit $\frac{H}{N}$ der Eins für genügend große N sehr nah bei $\frac{1}{6}$ liegen – das sagt das empirische Gesetz der großen Zahl. Wir werfen aber keine wirklichen Würfel, sondern rechnen in einem „Wahrscheinlichkeitsraum (Ω, P)", und da haben wir $P(\{1\}) = \frac{1}{6}$ gesetzt, weil uns das wegen der Symmetrie des Würfels vernünftig erschien, und wir nennen diesen Wert Wahrscheinlichkeit. Was passiert eigentlich, wenn wir unseren gedachten Würfel N–mal werfen, die Anzahl X der Einsen zählen und durch N teilen? Dies wollen wir nun untersuchen.

Wir werfen also unseren gedachten Würfel N–mal. Es sei X_i die Anzahl der Einsen im i–ten Wurf für $i = 1, 2, \ldots, N$. Für die Gesamtzahl X der Einsen gilt dann

$$X = \sum_{i=1}^{N} X_i \quad .$$

Wir haben früher schon ausgerechnet, dass $\mathrm{E}(X_i) = \frac{1}{6}$ und $\mathrm{V}(X_i) = \frac{5}{36}$ ist, und wir berechnen damit nach unseren Regeln[4]

$$\mathrm{E}\left(\frac{1}{N}X\right) = \frac{1}{N}\,\mathrm{E}\left(\sum_{i=1}^{N} X_i\right) = \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}(X_i) = \frac{1}{N}\cdot N\cdot\frac{1}{6} = \frac{1}{6} \quad \text{und}$$

$$\mathrm{V}\left(\frac{1}{N}X\right) = \mathrm{V}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{V}(X_i) = \frac{1}{N^2}\cdot N\cdot\frac{5}{36} = \frac{5}{36N} \quad .$$

Diese Daten setzen wir in die Ungleichung von Tschebyscheff ein. Für jedes noch so kleine $a > 0$ gilt dann

$$P\left(\left|\frac{1}{N}X - \frac{1}{6}\right| \geq a\right) \leq \frac{5}{36a^2N} \xrightarrow{N\to\infty} 0 \quad . \tag{15}$$

In Worten: Die Wahrscheinlichkeit dafür, dass $\frac{1}{N}X$ vom Erwartungswert $\frac{1}{6}$ um mindestens a abweicht, strebt für jedes noch so kleine $a > 0$ gegen 0 für $N \to \infty$. Dies ist die Aussage von **Bernoullis schwachem Gesetz der großen Zahl**. Ich habe sie nur für das Beispiel formuliert, das Gesetz gilt aber allgemein für Zufallsgrößen X, die eine Varianz haben. Es bestätigt, dass die Konstrukte der Wahrscheinlichkeitstheorie so gebildet sind, dass man erwarten kann, dass Aussagen der Wahrscheinlichkeitstheorie auf reale Zufallsvorgänge anwendbar sind. Und so soll es ja auch sein!

3.9 Eine Anwendung

Ein Würfel werde N–mal geworfen. Wie eben sei X_k die Anzahl der Sechsen im k–ten Wurf für $k = 1, 2, 3, \ldots, N$ und X die Anzahl der insgesamt geworfenen Sechsen. Wir hatten gesehen, dass

$$P\left(\left|\frac{1}{N}X - \frac{1}{6}\right| \geq \frac{1}{12}\right) \quad \leq \quad \frac{5\cdot 144}{36N} = \frac{20}{N} \xrightarrow{N\to\infty} 0$$

gilt. Für große N sollte also der beobachtete Wert der relativen Häufigkeit der Sechs zwischen $\frac{1}{12}$ und $\frac{3}{12}$ liegen. Wer nun meint, für genügend große N läge die relative Häufigkeit der Sechs **mit Sicherheit** in diesem Bereich, irrt. Man kann N so groß

[4] Dass wir hier die Summenregel für die Varianz benutzen dürfen, werden wir noch beweisen.

machen, wie man nur immer will, die Wahrscheinlichkeit, dass dennoch überhaupt keine Sechs geworfen wird, ist

$$P(X=0) = \left(\frac{5}{6}\right)^N > 0 \quad .$$

Man kann sich höchstens wünschen, dass das Risiko, dass die relative Häufigkeit von $\frac{1}{6}$ um mindestens $\frac{1}{12}$ abweicht, höchstens so groß wird wie eine kleine Schranke, die man vorgibt, zum Beispiel 5%.

Für die Wahrscheinlichkeit selbst, die auf der linken Seite der Ungleichung steht, kann man zwar einen Term hinschreiben, aber den Wert des Terms kann man schon für recht überschaubare N nicht berechnen, und es gibt keinerlei Hoffnung, dass man die Ungleichung nach N auflösen könnte. Deshalb macht man N so groß, dass das Tschebyscheffrisiko 5% ist; die Wahrscheinlichkeit selbst ist dann garantiert höchstens 5%. Aus

$$\frac{20}{N} = \frac{5}{100}$$

berechnen wir $N = 400$. In Wirklichkeit kommt man mit deutlich weniger Würfen aus, aber uns fehlen noch die Mittel, ein besseres N zu bestimmen. Da ist das berechnete $N = 400$ schon ein Gewinn.

3.10 Frequentistische Deutung des Erwartungswertes

Wir beginnen mit einer technischen Aufgabe: Wir werfen einen Würfel. Es sei X die gewürfelte Augenzahl bei einem Wurf, und es sei Y die mittlere Augenzahl je Wurf bei N Würfen. Berechne den Erwartungswert und die Varianz von X und daraus den Erwartungswert und die Varianz von Y [die Summenregel für die Varianz darfst du dabei benutzen]. Zeige, dass die Wahrscheinlichkeit, dass Y vom Erwartungswert um mindestens a abweicht, für jedes $a > 0$ gegen Null strebt für n gegen Unendlich.

Was die Aufgabe verlangt, ist mit überschaubarer Mühe erledigt. Es ist $\mathrm{E}(X) = \frac{7}{2}$ und $\mathrm{V}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \frac{35}{12}$. Wie üblich bezeichnen wir die im i–ten Wurf gewürfelte Augenzahl mit X_i für $i = 1, 2, 3, \ldots, N$, dann ist

$$Y = \frac{1}{N}\sum_{i=1}^{N} X_i \quad , \quad \mathrm{E}(Y) = \frac{7}{2} \quad \text{und} \quad \mathrm{V}(Y) = \frac{35}{12N} \quad .$$

Für $N \to \infty$ strebt somit $\mathrm{V}(Y)$ gegen Null. Nach der Tschebyscheffungleichung strebt dann für jedes noch so kleine positive a die Wahrscheinlichkeit, dass Y den Erwartungswert um mindestens a verfehlt, gegen Null für $N \to \infty$. Das heißt, **wenn man den Zufallsversuch sehr häufig durchführt, kann man davon ausgehen, dass das arithmetische Mittel $Y = \frac{1}{N}\sum_{i=1}^{N} X_i$ der Einzelwerte nahe bei $\mathrm{E}(X)$ liegt**. Entsprechendes gilt für die Wahrscheinlichkeit $P(A)$ eines Ereignisses $A \subseteq \Omega$. Führt man den Zufallsversuch oft durch, wird der Anteil der Durchführungen, bei denen ein $\omega \in A$ ausgelost wurde, in der Nähe von $P(A)$ liegen. Unsere theoretischen Zufallsgrößen halten, was das empirische Gesetz der großen Zahl für real durchgeführte Zufallsversuche verspricht.

4 Bedingte Wahrscheinlichkeit und Unabhängigkeit

Wir betrachten einen „Wahrscheinlichkeitsraum“ (Ω, P) und zwei Ereignisse $A, B \subseteq \Omega$. Es mag hilfreich sein, die Situation in einem Mengenbild darzustellen (siehe Abbildung 7, A liegt oberhalb der horizontalen Linie, B links der vertikalen Linie). Lost man ein $\omega \in \Omega$ aus, wird es mit der Wahrscheinlichkeit $P(B)$ zu B gehören.

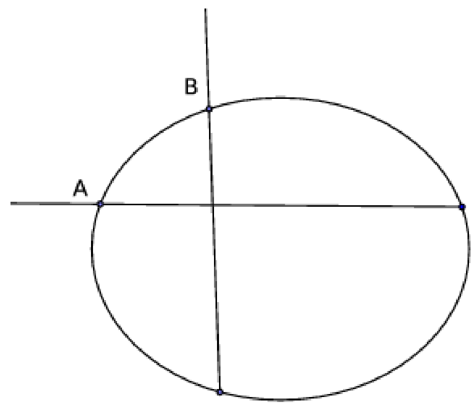


Abbildung 7: Ergebnisraum Ω mit Ereignissen $A, B \subseteq \Omega$

Nun kann man nach der Wahrscheinlichkeit fragen, dass B eintritt, wenn man zur Auslosung nur die Elemente von A zulässt. Die entsprechende Wahrscheinlichkeit wird **bedingte Wahrscheinlichkeit** genannt und mit den Symbolen

$$P_A(B) \quad \text{und} \quad P(B|A)$$

bezeichnet. Praktisch ist das ein neuer Zufallsversuch mit einem neuen Ω, eben A, und man überlegt sich, dass

$$P_A(B) := \frac{P(A \cap B)}{P(A)} \tag{16}$$

zu setzen ist. Das geht natürlich nur für $A \neq \emptyset$.

Man wird die Ereignisse A und B **unabhängig** nennen, wenn $P_A(B) = P(B)$ ist. Man kann sich das so vorstellen, dass dann der Anteil der Elemente von B in A genau so groß ist wie im gesamten Ergebnisraum. Man definiert den Begriff ein wenig anders, damit er auch noch sinnvoll ist, wenn $P(A) = 0$ ist:

9 Definition
Die Ereignisse A und B eines Ergebnisraums Ω heißen **unabhängig**, *wenn*

$$P(A \cap B) = P(A) \cdot P(B)$$

ist.

Schauen wir uns ein Beispiel an. Aus einer Urne mit drei roten und sieben weißen Kugeln werden zwei gezogen. Dann ist $\Omega = \{rr, rw, wr, ww\}$ ein geeigneter Ergebnisraum. Es sei $A = \{rr, rw\}$ und $B = \{rr, wr\}$; in Worten formuliert „erste Kugel ist rot“ bzw. „zweite Kugel ist rot“. Wie wir wissen, ist

$$P(A) = P(B) = \frac{3}{10} \quad ,$$

ganz gleich, ob wir die erste gezogene Kugel zurücklegen oder nicht. Für $P(A \cap B) = P(\{rr\})$ sieht das aber anders aus. Legen wir die gezogene Kugel zurück, wird

$$P(A \cap B) = \frac{3}{10} \cdot \frac{3}{10} = P(A) \cdot P(B)$$

sein, die Ereignisse sind folglich unabhängig. Ziehen wir dagegen ohne Zurücklegen, ist

$$P(A \cap B) = \frac{3}{10} \cdot \frac{2}{9} \neq P(A) \cdot P(B) \quad ,$$

hier sind A und B abhängig. Für die bedingte Wahrscheinlichkeit erhalten wir in diesem Fall

$$P_A(B) = \frac{2}{9} \quad .$$

Was du beim Baumdiagramm an die Pfadabschnitte geschrieben hast, waren alles bedingte Wahrscheinlichkeiten, und die hast du automatisch richtig gebildet.

Wir könnten nun als Ω die Menge der Bundesbürger, als A die Menge der Einwohner Berlins und als B die Menge der Sozialhilfeempfänger nehmen. In solchen Kontexten spielt der Begriff der Unabhängigkeit im Alltag eine durchaus brisante Rolle. Natürlich werden da keine Mengen gebildet und es wird kein ω ausgelost; das Phänomen der Unabhängigkeit bleibt aber.

Übrigens sagt sich das leicht, Einwohner von Berlin. Es ist durchaus schwierig, im Einzelfall zu entscheiden, wer dazugehört und wer nicht. Aber mit diesem Problem wollen wir uns nicht befassen.

5 Erste Klausur am 28. März 2012

1. **Einstiegsaufgabe**
   Bei einem Spiel wird der Betrag X ausgezahlt. Es ist $P(X = 0) = \frac{1}{2}$, $P(X = 1) = \frac{1}{3}$ und $P(X = 2) = \frac{1}{6}$. Zeichne das Histogramm von X, berechne Erwartungswert und Varianz von X und erläutere kurz, was der Erwartungswert eigentlich bedeutet. [20]

2. **Kursstochastik**
   In unserem Kurs sind vierundzwanzig Schüler, und von denen kommen sechs aus Hille. Gib durchsichtige Terme an, Werte brauchst du nicht auszurechnen.

   (a) Wieviele verschiedene Vierergruppen kann man insgesamt bilden? [3]

   (b) Wieviele verschiedene Vierergruppen kann man bilden, in denen zwei Hiller und zwei andere sind? [5]

   (c) Es werden vier Leute aus dem Kurs ausgelost. [10]

      i. Mit welcher Wahrscheinlichkeit ist Marvin dabei?

      ii. Mit welcher Wahrscheinlichkeit gehört Lucas nicht zu der Gruppe?

   (d) Die Stufenleiterin ruft einen nach dem anderen der Schüler des Kurses in ihre Sprechstunde. Wie groß ist die Wahrscheinlichkeit, dass der vierte Schüler der erste ist, der nicht aus Hille kommt? [5]

   (e) Nun soll ein Kurssprecher gewählt werden. Wenn der Gewählte nicht aus Hille kommt, legt der Bürgermeister der Gemeinde Widerspruch ein, und dann muss die Wahl wiederholt werden. Es sei X die Anzahl der nötigen Wahlgänge. Gib eine Formel für $P(X = k)$ an. [8]

   (f) Mit welcher Wahrscheinlichkeit sind höchstens zehn Wahlgänge nötig? [8]

(g) Wenn es nur nach dem Zufall geht, ist die Wahrscheinlichkeit dafür, dass Kevin, Nick und Eike als Letzte zum Unterricht erscheinen,

$$p = \frac{3! \cdot 21!}{24!} \quad .$$

i. Begründe, dass das stimmt, und berechne den Wert von p. [8]

ii. Wir schauen uns sechzig Tage mit Mathematikunterricht an. Begründe, dass der Erwartungswert μ der Anzahl der Tage, an denen die drei als Letzte zum Unterricht erscheinen, $60p$ ist. [8]

iii. Wie groß ist die Wahrscheinlichkeit höchstens, dass die beobachtete Anzahl von Tagen, an denen die drei als Letzte kommen, vom Erwartungswert μ um mindestens 40 abweicht? Hier ist ein Wert verlangt, rechne mit $\mathrm{V}(X) = 60p(1-p)$. [10]

iv. Die Beobachtung ergab, dass die drei an 45 von den sechzig Terminen als Letzte kamen. Wie verträgt sich das mit den Ergebnissen der Rechnung? [4]

(h) Weil sie eine verschworene Gemeinschaft sind und weil es eine Stochastikklausur ist, verabreden die Schüler, dass jeder bei der Rückgabe der Klausur einfach das nächste Heft vom Stapel nimmt. [16]

i. Mit welcher Wahrscheinlichkeit bekommt Burak sein eigenes Heft?

ii. Was ist der Erwartungswert der Anzahl der Hefte, die bei ihrem Besitzer ankommen? Vorsicht, mit dem richtigen Trick hat man die Antwort in wenigen Zeilen, ohne ist sie praktisch nicht lösbar.

3. **Kleiner Sprachtest**

(a) Es sei $\Omega = \{-1, 0, 1\}$, und jedes $\omega \in \Omega$ besitze die gleiche Wahrscheinlichkeit. Die Zufallsgröße X auf Ω ist definiert durch $X(\omega) = \omega^2$.

i. Schreibe die Wahrscheinlichkeitsverteilung von X hin. [6]

ii. Was ist $X^{-1}(1)$? [4]

(b) Zeichne für die Zufallsgröße X der Einstiegsaufgabe zu Beginn der Klausur den Graphen der Funktion $x \mapsto P(X \leq x)$. [8]

(c) Es seien X und Y Zufallsgrößen auf dem gleichen Wahrscheinlichkeitsraum (Ω, P). Was ist dann [6]

$$\sum_{\omega \in \Omega} X(\omega) \cdot Y(\omega) \cdot P(\{\omega\}) \quad ?$$

4. **Taylorpolynom**

Ein Berliner Institut denkt über Aufgaben für die nächste Lernstandserhebung der Stufe 8 nach: eine Seitenfläche eines großen Würfels ist um 20% größer als eine Seitenfläche eines kleinen Würfels. Um wieviel ist das Volumen des großen Würfels größer als das des kleinen Würfels? – Du weißt ja mehr als ein Achter. Stelle das Taylorpolynom $p(x)$ vom Grad 2 der Funktion

$$f(x) = (1+x)^{\frac{3}{2}}$$

auf, berechne damit $p(0.2)$ und vergleiche den Wert mit dem exakten Wert von $f(0.2)$. Die Taylorreihe ist recht unbequem hinzuschreiben, die brauchst du nicht. Aber wenn du Zeit und Lust hast, kannst du mit dem Ergebnis dem Achter die Antwort verraten. [24+]

6 Mehr über Zufallsgrößen

6.1 Die Varianz einer Summe

Es seien X und Y Zufallsgrößen auf dem gleichen Ω. Wir wünschen uns auch für die Varianz $\mathrm{V}(X+Y)$ der Summe von X und Y eine schöne Regel, wir haben aber gesehen, dass eine solche Regel nicht für beliebige X und Y gilt. Nun versuchen wir, die Varianz der Summe zu berechnen; wir werden ja sehen, wie weit wir damit kommen.

Zunächst nutzen wir unsere Regeln für die Varianz und für Erwartungswerte aus.

$$\begin{aligned}
\mathrm{V}(X+Y) &= \mathrm{E}\left((X+Y)^2\right) \quad - \quad (\mathrm{E}(X+Y))^2 \\
&= \mathrm{E}\left(X^2+2XY+Y^2\right) \quad - \quad \left(\mathrm{E}(X)^2+2\,\mathrm{E}(X)\,\mathrm{E}(Y)+\mathrm{E}(Y)^2\right) \\
&= \mathrm{E}(X^2)+2\,\mathrm{E}(XY)+\mathrm{E}(Y^2) \quad - \quad \mathrm{E}(X)^2-2\,\mathrm{E}(X)\,\mathrm{E}(Y)-\mathrm{E}(Y)^2 \\
&= \mathrm{V}(X)+\mathrm{V}(Y) \quad + \quad 2\,\mathrm{E}(XY)-2\,\mathrm{E}(X)\,\mathrm{E}(Y)
\end{aligned}$$

Die gesuchte Summenregel für die Varianz gilt folglich genau dann, wenn der Erwartungswert $\mathrm{E}(XY)$ des Produkts gleich dem Produkt der Erwartungswerte ist.

Nehmen wir uns also $\mathrm{E}(XY)$ vor. Es hilft nichts, da müssen wir auf die Definition zurück:

$$\mathrm{E}(XY) = \sum_{\omega\in\Omega} X(\omega)Y(\omega)P(\{\omega\}) \tag{17}$$

Die Zufallsgrößen X und Y stiften jeweils eine Ordnung in Ω. Zum Einen wird Ω zerlegt in $|X(\Omega)|$ Teilmengen $X^{-1}(k)$ der $\omega\in\Omega$ mit $X(\omega)=k$, dabei läuft k durch $X(\Omega)$. Zum Anderen wird Ω zerlegt in $|Y(\Omega)|$ Teilmengen $Y^{-1}(j)$ der $\omega\in\Omega$ mit $Y(\omega)=j$, dabei läuft j durch $Y(\Omega)$. Abbildung 8 zeigt diese Ordnung für $|X(\Omega)|=|Y(\Omega)|=3$.

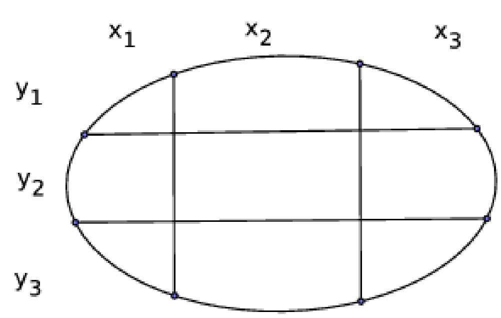


Abbildung 8: Von X und Y gestiftete Einteilung von Ω für $|X(\Omega)|=|Y(\Omega)|=3$

Es sei nun $k\in X(\Omega)$ und $j\in Y(\Omega)$. In der Summe in Gleichung (17) fassen wir die Beiträge aller ω zusammen, für die $X(\omega)=k$ und $Y(\omega)=j$ ist; das sind alle ω eines der Kästchen in Abbildung 8. Diese ω bilden die Menge

$$X^{-1}(k)\cap Y^{-1}(j) \quad ,$$

und der Beitrag dieser ω zur Summe ist

$$kjP\left(X^{-1}(k)\cap Y^{-1}(j)\right) \quad .$$

Du kennst dich mit bedingten Wahrscheinlichkeiten aus. Es gilt

$$P\left(X^{-1}(k)\cap Y^{-1}(j)\right) = P\left(X^{-1}(k)\right)\cdot P_{X^{-1}(k)}\left(Y^{-1}(j)\right) \quad .$$

In Worten: Die Wahrscheinlichkeit, ein ω der Schnittmenge $X^{-1}(k)\cap Y^{-1}(j)$ zu bekommen, ist das Produkt der Wahrscheinlichkeit $P\left(X^{-1}(k)\right)$, ein Element von $X^{-1}(k)$ zu bekommen, mit der Wahrscheinlichkeit $P_{X^{-1}(k)}\left(Y^{-1}(j)\right)$, ein Element von $Y^{-1}(j)$ zu bekommen, wenn nur unter den Elementen von $X^{-1}(k)$ ausgelost wird. Nun sind wir an der entscheidenden Stelle. Wenn die Ereignisse $X^{-1}(k)$ und $Y^{-1}(j)$ **unabhängig** sind, ist

$$\begin{aligned} P_{X^{-1}(k)}\left(Y^{-1}(j)\right) &= P\left(Y^{-1}(j)\right) \quad \text{und} \\ P\left(X^{-1}(k)\cap Y^{-1}(j)\right) &= P\left(X^{-1}(k)\right)\cdot P\left(Y^{-1}(j)\right) \quad , \end{aligned}$$

und dann bringen wir unsere Rechnung zum gewünschten Ende:

$$\begin{aligned} \mathrm{E}(XY) &= \sum_{\omega\in\Omega} X(\omega)Y(\omega)P(\{\omega\}) \\ &= \sum_{k\in X(\Omega)}\sum_{j\in Y(\Omega)} kjP\left(X^{-1}(k)\cap Y^{-1}(j)\right) \\ &= \sum_{k\in X(\Omega)}\sum_{j\in Y(\Omega)} kjP\left(X^{-1}(k)\right)\cdot P\left(Y^{-1}(j)\right) \\ &= \sum_{k\in X(\Omega)}\left(kP\left(X^{-1}(k)\right)\sum_{j\in Y(\Omega)} jP\left(Y^{-1}(j)\right)\right) \\ &= \sum_{j\in Y(\Omega)} jP\left(Y^{-1}(j)\right)\cdot\sum_{k\in X(\Omega)} kP\left(X^{-1}(k)\right) \\ &= \mathrm{E}(Y)\cdot\mathrm{E}(X) \end{aligned}$$

Du musst dich nun nicht damit belasten, für jeden Wertesatz $k\in X(\Omega)$ und $j\in Y(\Omega)$ die Unabhängigkeit der Ereignisse $X^{-1}(k)$ und $Y^{-1}(j)$ zu prüfen. Wenn sich X und Y nicht beeinflussen, sich zum Beispiel aus der unabhängigen Wiederholung eines Zufallsversuchs ergeben, kannst du getrost die Summenregel für die Varianz verwenden, die ich hier etwas nachlässig formuliere:

10 Lemma
Es seien $X, Y:\Omega\to\mathbb{R}$ unabhängig. Dann gilt $\mathrm{V}(X+Y)=\mathrm{V}(X)+\mathrm{V}(Y)$.

Du wirst sehen, dass dieses Lemma für uns sehr nützlich ist.

6.2 Die Binomialverteilung

Wir würfeln und interessieren uns nur dafür, ob eine Sechs geworfen wurde oder nicht. Führen wir den Versuch n–mal durch und bezeichnen wir die Anzahl der geworfenen Sechsen mit X, gilt bekanntlich

$$P(X=k)=\binom{n}{k}\cdot\left(\frac{1}{6}\right)^k\cdot\left(\frac{5}{6}\right)^{n-k} \qquad \text{für } k=0,1,2,\ldots,n \ .$$

Eine solche Situation tritt recht häufig auf, und deshalb prägen wir eine Reihe von Begriffen.

11 Definition
Eine Zufallsgröße X, die die Werte $k = 0, 1, 2, 3, \dots, n$ mit der Wahrscheinlichkeit

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

annimmt, heißt **binomialverteilt**, *genauer $B(n,p)$–verteilt.*

12 Definition
Ein Zufallsversuch, der nur zwei Ergebnisse hat, heißt **Bernoulli–Versuch**. *Eines der Ergebnisse nennt man* **Erfolg**, *und man bezeichnet seine Wahrscheinlichkeit mit p. Das andere heißt Misserfolg, seine Wahrscheinlichkeit $1-p$ wird oft mit q bezeichnet. Wiederholt man den Versuch n–mal so, dass die Erfolgswahrscheinlichkeit dabei immer gleich bleibt, nennt man dies eine* **Bernoullikette der Länge** *n* **mit Erfolgswahrscheinlichkeit** *p.*

Die Zufallsgröße „Anzahl der Erfolge bei einer Bernoullikette der Länge n mit der Erfolgswahrscheinlichkeit p “ ist $B(n,p)$–verteilt. Solche Zufallsgrößen treten in Anwendungen häufig auf, deshalb notieren wir wichtige Eigenschaften. Zunächst geben wir den Erwartungswert und die Varianz an:

13 Satz
Es sei X binomialverteilt mit den Parametern n und p. Dann ist

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

Man beweist den Satz mit dem üblichen Summentrick, indem man X als Summe $X_1 + X_2 + \cdots + X_n$ darstellt.

Zum Verständnis von Anwendungssiationen müssen wir eine Vorstellung davon haben, wie das Histogramm einer binomialverteilten Zufallsgröße aussieht. Es hat, grob gesprochen, Buckelform, es ist eingipflig. Die Kuppe des Buckels liegt in der Nähe des Erwartungswertes. Beide Aussagen ergeben sich aus diesem technischen Lemma.

14 Lemma
Es sei X eine $B(n,p)$–verteilte Zufallsgröße. Dann gilt

$$P(X = k) \, < \, P(X = k+1) \quad \Longleftrightarrow \quad k \, < \, np - (1-p)$$

für $k = 0, 1, 2, \dots, n-1$.

Wir haben das Lemma bewiesen, indem wir die beiden Terme ausgeschrieben haben und eine Reihe naheliegender Vereinfachungen vorgenommen haben.

6.3 Standardisierte Zufallsgrößen

15 Definition
Es sei X eine Zufallsgröße mit Erwartungswert μ und Standardabweichung σ. Dann heißt

$$X^* := \frac{X - \mu}{\sigma}$$

die zu X gehörige standardisierte Zufallsgröße.

16 Lemma
Es sei X^* die zu X gehörige standardisierte Zufallsgröße. Dann ist $\mathrm{E}(X^*) = 0$ und $\mathrm{V}(X^*) = 1$.

Beweis. Wende die Regeln an!

Warum standardisiert man Zufallsgrößen? Wahrscheinlichkeiten $P(X = k)$ sind für binomialverteilte X schon bei ziemlich kleinen n nur mühsam auszurechnen. In Anwendungen braucht man aber oft ziemlich große n. Zum Glück sind die Histogramme von binomialverteilten Zufallsgrößen mit nicht zu kleiner Varianz Transformationen der gleichen Urkurve. Das heißt, wenn man sie standardisiert, bekommt immer das gleiche Histogramm, und damit kann man dann die benötigten Wahrscheinlichkeiten berechnen.

Beispiel. Es sei X $B(50, \frac{1}{5})$–verteilt. Dann nimmt X die Werte $k = 0, 1, 2, \dots, 50$ an. Wie wir wissen, ist $\mathrm{E}(X) = 50 \cdot \frac{1}{5} = 10$ und $\sigma = \sqrt{50 \cdot \frac{1}{5} \cdot \frac{4}{5}} = \sqrt{8}$. Die zugehörige standardisierte Zufallsgröße X^* nimmt folglich die Werte

$$-\frac{10}{\sqrt{8}}, -\frac{9}{\sqrt{8}}, \dots, \frac{40}{\sqrt{8}}$$

an, und es ist

$$P\left(X^* = \frac{k-\mu}{\sigma}\right) \;=\; P(X = k) \;=\; \binom{50}{k}\left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{50-k} \quad .$$

Beim Zeichnen des Histogramms von X^* muss man beachten, dass die Wahrscheinlichkeit eines Wertes der Inhalt des Kästchens ist, das auf dem Wert steht. Da die Kästchen im Histogramm von X^* die Breite $\frac{1}{\sigma} = \frac{1}{\sqrt{8}}$ haben, hat das Kästchen auf $\frac{k-\mu}{\sigma}$ die Höhe $\sigma P(X = k)$.

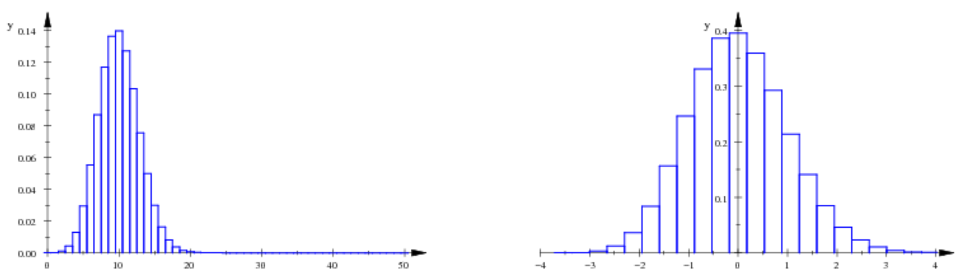


Abbildung 9: Histogramme von X (links) und von X^* (rechts, mit dem Graphen von Finns Kurve) für $B(50, \frac{1}{5})$–verteiltes X

6.4 Einschub – für Marvin

Wenn man zu der Überzeugung gelangt ist, dass es eine Funktion f gibt, deren Graph die Urkurve ist, die zu den Histogrammen der standardisierten Zufallsgrößen X^* der $B(n, p)$–verteilten X passt, kann man versuchen, f zu bestimmen.

Abbildung 10 zeigt links das Histogramm des $B(50, \frac{1}{5})$–verteilten X, das Kästchen zu einem Wert k ist ausgefüllt. Rechts ist Finns Kurve und das Kästchen auf dem Wert

$$x = \frac{k-\mu}{\sigma}$$

zu k. Beide Kästchen haben den Inhalt $P(X = k)$. Das rechte Kästchen hat die Breite $\frac{1}{\sigma}$ und die Höhe $f(x)$, also gilt

$$P(X = k) = \frac{1}{\sigma} f(x) \quad . \tag{18}$$

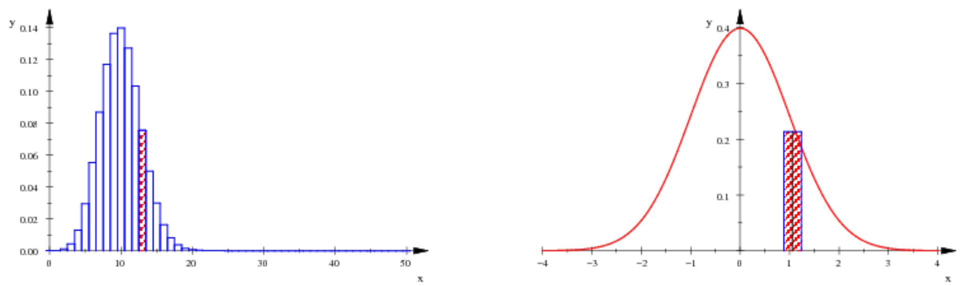


Abbildung 10: Säulen auf dem Wert k im Histogramm von X (links) und auf dem zugehörigen $x = \frac{k-\mu}{\sigma}$

Wir können die Parameter n und p von X ruhig verstellen, allerdings soll der Wert von x dabei nicht verändert werden; es muss also stets

$$k = x\sigma + \mu \tag{19}$$

gelten.

Wir werden nun die Ableitung $f'(x)$ von $f(x)$ bestimmen. Zu $k+1$ gehört der Wert

$$\frac{k+1-\mu}{\sigma} = x + \frac{1}{\sigma} \quad .$$

Wir bilden

$$\begin{aligned} \frac{f\left(x+\frac{1}{\sigma}\right) - f(x)}{\frac{1}{\sigma}} &= \sigma\left(f\left(x+\frac{1}{\sigma}\right) - f(x)\right) = \\ &= \sigma^2\left(P(X=k+1) - P(X=k)\right) \quad . \end{aligned} \tag{20}$$

Nun ist

$$\begin{aligned} P(X=k+1) - P(X=k) &= \binom{n}{k+1} p^{k+1}(1-p)^{n-k-1} - \binom{n}{k} p^k(1-p)^{n-k} = \\ &= \binom{n}{k} \frac{n-k}{k+1} p^{k+1}(1-p)^{n-k-1} - \binom{n}{k} p^k(1-p)^{n-k} = \\ &= \binom{n}{k} p^k(1-p)^{n-k}\left(\frac{n-k}{k+1}\cdot\frac{p}{1-p} - 1\right) = \\ &= \binom{n}{k} p^k(1-p)^{n-k}\cdot\frac{np-kp-k-1+kp+p}{(k+1)(1-p)} = \\ &= \binom{n}{k} p^k(1-p)^{n-k}\frac{-x\sigma-1+p}{(k+1)(1-p)} \quad , \end{aligned} \tag{21}$$

denn $np - k = -x\sigma$. Wir setzen dieses Ergebnis in Gleichung (20) ein und erhalten

$$\frac{f\left(x+\frac{1}{\sigma}\right) - f(x)}{\frac{1}{\sigma}} = \sigma^2\binom{n}{k} p^k(1-p)^{n-k}\frac{-x\sigma-1+p}{(k+1)(1-p)} \quad .$$

Nun ist $\sigma P(X=k) = f(x)$. Damit erhalten wir

$$\begin{aligned} \frac{f\left(x+\frac{1}{\sigma}\right) - f(x)}{\frac{1}{\sigma}} &= \sigma f(x)\frac{-x\sigma-1+p}{(k+1)(1-p)} = \\ &= f(x)\frac{-xnp(1-p)-\sigma+p\sigma}{(k+1)(1-p)} = f(x)\frac{-xp(1-p)-\frac{\sigma}{n}+\frac{p\sigma}{n}}{(\frac{k}{n}+\frac{1}{n})(1-p)} \quad . \end{aligned}$$

Schauen wir uns den letzten Term an. Wenn n gegen Unendlich läuft, strebt $\frac{\sigma}{n}$ gegen 0, denn in σ steht ja nur $\sqrt{n}$. Für das $\frac{k}{n}$ im Nenner gilt

$$\frac{k}{n} = \frac{np + x\sigma}{n} = p + \frac{x\sigma}{n} \xrightarrow{n\to\infty} p \quad .$$

Insgesamt erhalten wir

$$\frac{f\left(x + \frac{1}{\sigma}\right) - f(x)}{\frac{1}{\sigma}} \xrightarrow{n\to\infty} f(x)\frac{-xp(1-p)}{p(1-p)} = -xf(x) \quad ,$$

also schließlich

$$f'(x) = -xf(x) \quad . \tag{22}$$

So, das ist es! Dies ist eine **Differentialgleichung** für $f(x)$, und wir können sie sogar lösen! Wir müssen nur durch $f(x)$ dividieren und einmal scharf hinschauen:

$$\begin{aligned}
\frac{f'(x)}{f(x)} &= -x \\
(\ln(f(x)))' &= -x \\
\ln(f(x)) &= -\frac{1}{2}x^2 + c \quad \text{für eine feste Zahl } c \in \mathbb{R} \\
f(x) &= e^{-\frac{1}{2}x^2 + c} = Ce^{-\frac{1}{2}x^2} \quad \text{für eine feste Zahl } C > 0
\end{aligned}$$

Nun muss nur noch der richtige Wert von C ausgerechnet werden. Das hat Finn erledigt: Finns Kurve ist der Graph der Gaußfunktion

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \quad . \tag{23}$$

Die Rechnung soll den Term der Gaußfunktion plausibel machen, ein strenger Beweis ist das nicht. Vor allem fehlt der Nachweis, dass überhaupt eine Urkurve existiert. Ein solcher Nachweis führte aber wirklich zu weit.

6.5 Ein paar Bilder

In Abbildung 11 siehst du Histogramme $B(200, p)$–verteilter Zufallsgrößen für die p–Werte $\frac{1}{20}$, $\frac{2}{10}$, $\frac{5}{10}$, $\frac{8}{10}$ und $\frac{19}{20}$. Die Abbildung 12 zeigt die Histogramme der zugehörigen standardisierten Zufallsgrößen, jeweils mit Finns Kurve. Du siehst, dass Finns Kurve gut passt. Dass du etwas achtgeben musst, siehst du an den Histogrammen in Abbildung 13. Wenn die Varianz zu klein ist, solltest du nicht mit Finns Kurve arbeiten. Nimm sie nur, wenn du nicht zu Fuß rechnen kannst.

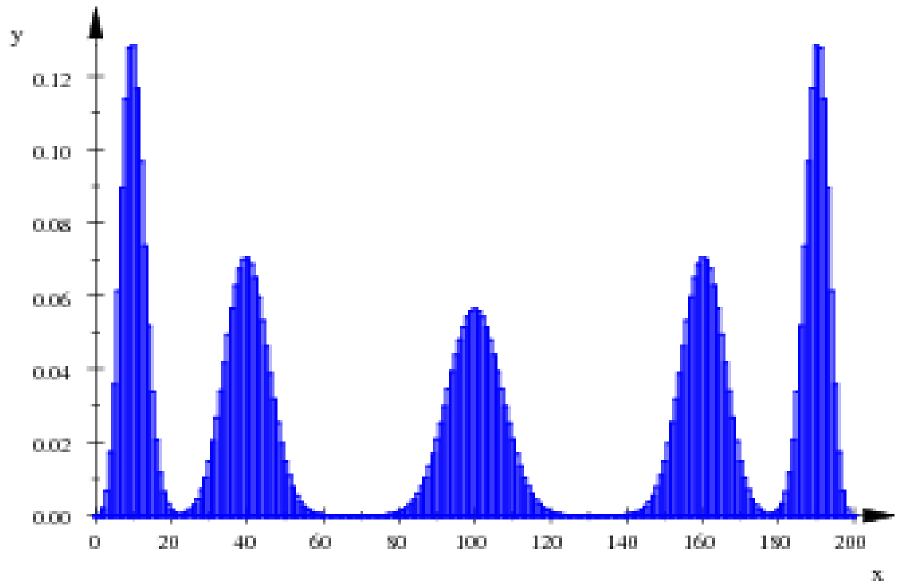


Abbildung 11: Histogramme $B(200, p)$–verteilter X für $p = \frac{1}{20}, \frac{2}{10}, \frac{5}{10}, \frac{8}{10}, \frac{19}{20}$

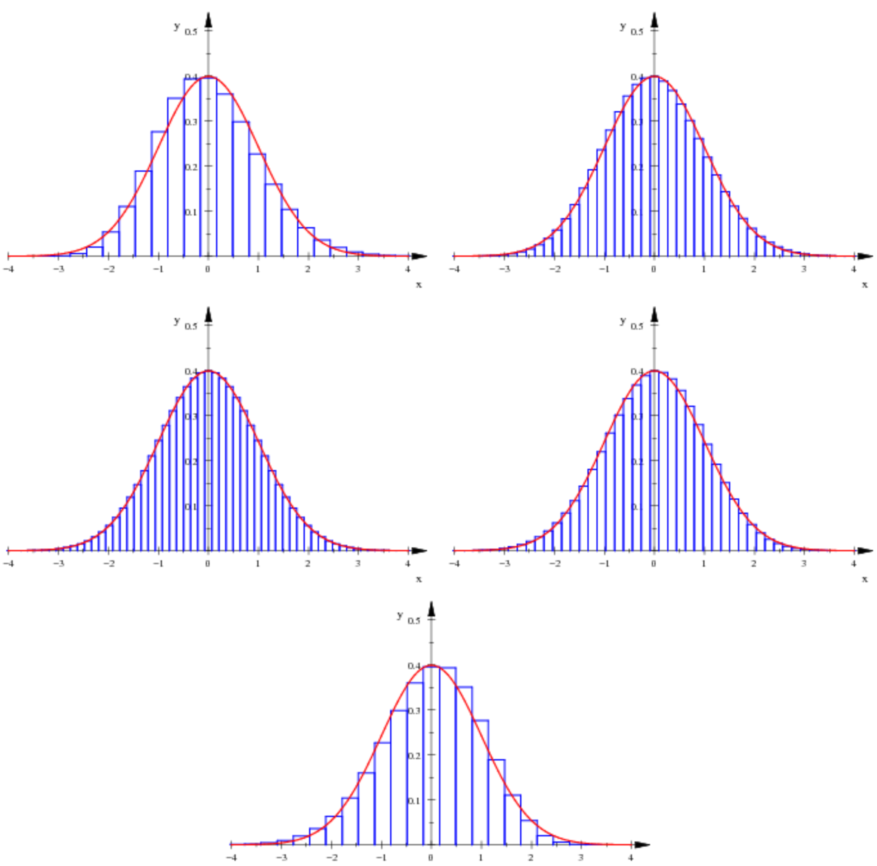


Abbildung 12: Histogramme der standardisierten Zufallsgrößen X^* zu $B(200, p)$–verteilten X für $p = \frac{1}{20}, \frac{2}{10}, \frac{5}{10}, \frac{8}{10}, \frac{19}{20}$, jeweils mit Finns Kurve

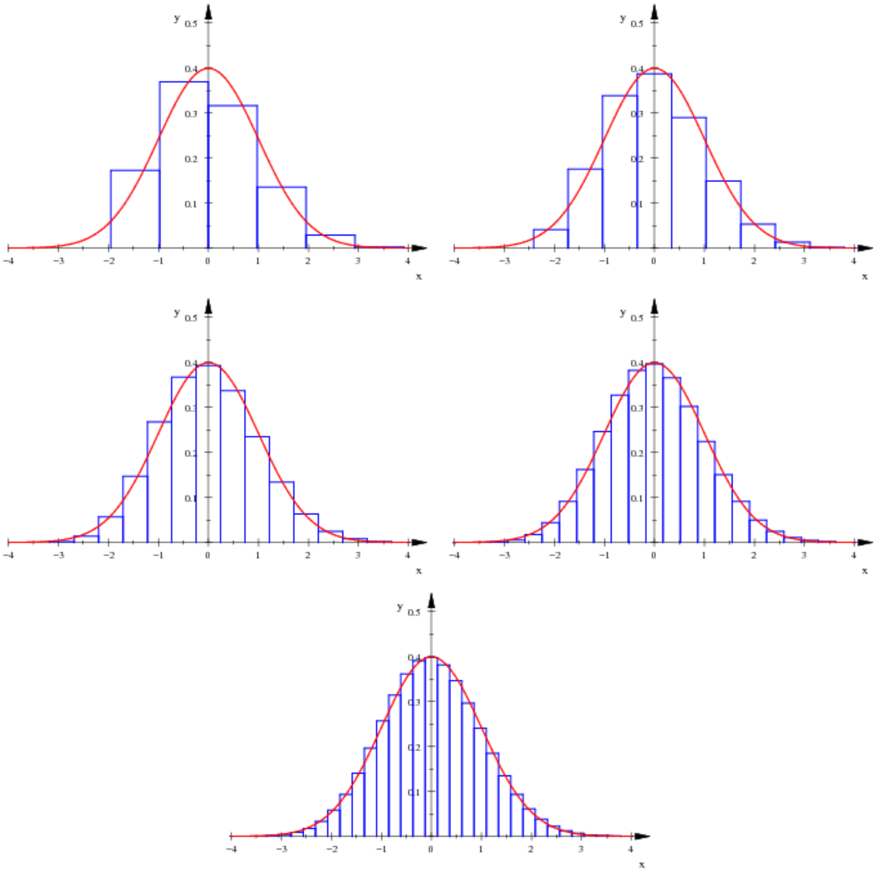


Abbildung 13: Histogramme der standardisierten Zufallsgrößen X^* zu $B(n, \frac{3}{10})$–verteilten X für $n = 5, 10, 20, 40, 80$, jeweils mit Finns Kurve

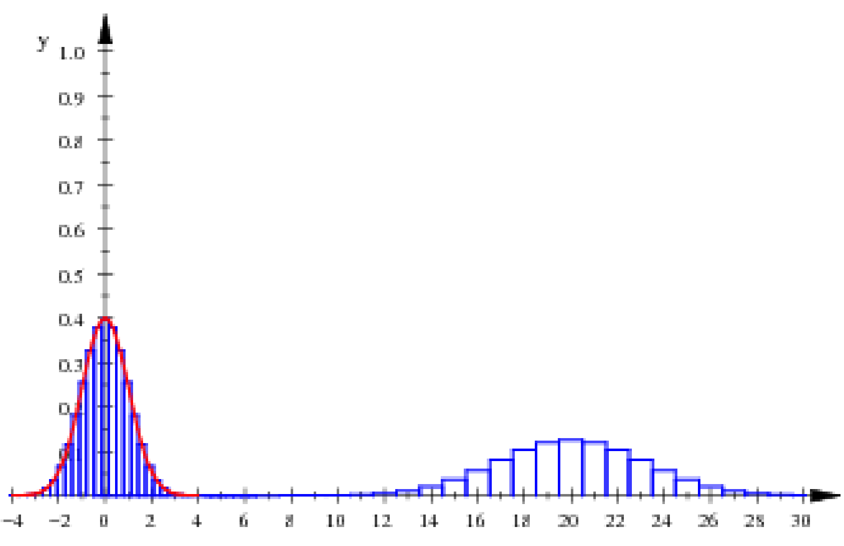


Abbildung 14: $B(40, \frac{1}{2})$–verteiltes X und zugehöriges X^* in einem Schaubild

6.6 Die Näherungsformeln von de Moivre und Laplace

Es sei X eine $B(n,p)$–verteilte Zufallsgröße. Wahrscheinlichkeiten $P(X=k)$ und erst recht welche der Form $P(X \le k)$ kann man nur für sehr kleine Werte von n und k ausrechnen, viel kleinere, als in den Anwendungen vorkommen. Hier hilft man sich mit Näherungsformeln, und das geht so: Zunächst standardisiert man X, man geht also über zu

$$X^* = \frac{X-\mu}{\sigma} \quad .$$

Aus dem Rechteck auf dem Wert k im Histogramm von X mit der Breite 1 und der Höhe $P(X=k)$ wird das Rechteck auf $\frac{k-\mu}{\sigma}$ mit der Breite $\frac{1}{\sigma}$ im Histogramm von X^*. Die Höhe des Rechtecks muss folglich $\sigma \cdot P(X=k)$ sein. Da das Histogramm von X^* gut zu Finns Kurve mit der Gleichung

$$y = \varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} \tag{24}$$

passt, gilt folglich

$$\sigma P(X=k) \approx \varphi\left(\frac{k-\mu}{\sigma}\right) \quad ,$$

also

$$P(X=k) = P\left(X^* = \frac{k-\mu}{\sigma}\right) \approx \frac{1}{\sigma}\varphi\left(\frac{k-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\cdot\sigma} e^{-\frac{1}{2}\left(\frac{k-\mu}{\sigma}\right)^2} \tag{25}$$

– dies ist die **lokale Näherungsformel von de Moivre und Laplace**.

Völlig analog erhält man

$$P(X \le k) = P\left(X^* \le \frac{k-\mu}{\sigma}\right) \approx \int_{-\infty}^{\frac{k+\frac{1}{2}-\mu}{\sigma}} \varphi(x)\,dx = \Phi\left(\frac{k+\frac{1}{2}-\mu}{\sigma}\right) \tag{26}$$

mit

$$\Phi(x) := \int_{-\infty}^{x} \varphi(t)\,dt \quad . \tag{27}$$

Gleichung (26) ist die **integrale Näherungsformel von de Moivre und Laplace**. Das $\frac{1}{2}$ bei der oberen Grenze des Integrals verhindert, dass man im Histogramm von X^* die Hälfte des Kästchens auf $\frac{k-\mu}{\sigma}$ weglässt.

Das Integral über φ kann man nicht geschlossen berechnen, da ist man auf Tabellen für Φ angewiesen. Ansonsten ist die Anwendung der Formeln eigentlich nicht schwierig. Wichtig ist, dass du stets die Flächen in den Histogrammen vor Augen hast, die zu den gesuchten Wahrscheinlichkeiten gehören, dann kommst du gut zurecht. Insbesondere solltest du gut verstanden haben, was bei der Standardisierung passiert; vertiefe dich ruhig noch einmal in die Abbildung 14 auf Seite 31.

6.7 Die σ–Regeln

Es sei X eine binomialverteilte Zufallsgröße. Wenn die Varianz von X groß genug ist, können wir das Histogramm von X^* getrost durch Finns Kurve ersetzen, also durch den Graphen von φ. Nun ist – du solltest den Graphen von φ immer vor Augen haben –

$$\int_{-1}^{1} \varphi(x)\,dx = \Phi(1) - \Phi(-1) = \Phi(1) - (1-\Phi(1)) = 2\Phi(1) - 1 \approx 2\cdot 0.8413 - 1 = 0.6826 \ ,$$

und das heißt doch

$$P(-1 < X^* < 1) = P(|X^* - 0| < 1) \approx 0.6826 \quad .$$

Wir machen daraus eine Aussage über X selbst. Dazu überlegen wir uns, welcher Wert x der Merkmalsachse beim Standardisieren auf die 1 kam:

$$\frac{x-\mu}{\sigma} = 1 \quad \Longleftrightarrow \quad x = \mu + \sigma \cdot 1$$

Also gilt

$$P(|X-\mu| < \sigma) \;=\; P(\mu - \sigma < X < \mu + \sigma) \;\approx\; 0.6826 \quad .$$

Man kann folglich sagen: Die Wahrscheinlichkeit, dass X einen Wert in der σ–Umgebung von μ annimmt, ist etwa 68%. Man sollte die Zahl nicht zu genau nehmen, da X ja nur ganzzahlige Werte annimmt, aber bei richtigem Gebrauch hat man eine sehr nützliche Faustregel. Ich formuliere sie als Lemma:

17 Lemma (σ–Regel)
Es sei X eine binomialverteilte Zufallsgröße. Dann ist

$$\begin{aligned} P(|X-\mu| < \sigma) &\approx 0.68 \\ P(|X-\mu| < 2\sigma) &\approx 0.95 \\ P(|X-\mu| < 3\sigma) &\approx 0.997 \end{aligned}$$

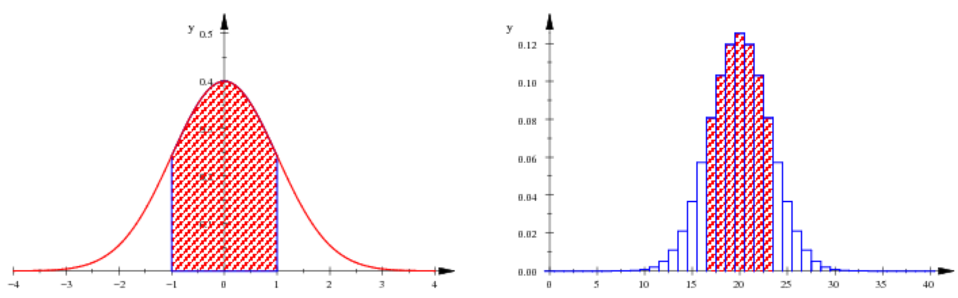


Abbildung 15: Fläche unter Finns Kurve zwischen -1 und 1 (links) und die σ–Umgebung von μ im Histogramm des $B(40, \frac{1}{2})$–verteilten X (rechts). Beide Flächen haben jeweils etwa den Inhalt 0.68 .

Beispiel. Ein Würfel wird 1200–mal geworfen, und es sei X die Anzahl der geworfenen Sechsen. Dann ist X binomialverteilt, $\mu = 1200 \cdot \frac{1}{6} = 200$, $\sigma = \sqrt{200 \cdot \frac{5}{6}} \approx 12.9$. Führt man den Versuch durch, wird die Anzahl der gewürfelten Sechsen mit der Wahrscheinlichkeit 0.68 zwischen 187 und 213 liegen und mit der Wahrscheinlichkeit 0.95 zwischen 174 und 226, und nur sehr selten wird man mehr als 239 oder weniger als 161 Sechsen beobachten. Die tatsächlich beobachteten Werte liegen also in einem recht kleinen Teil der Merkmalsachse.

6.8 Technische Übung

Hinweis: Zeichne dir immer Skizzen der Histogramme!

1. Berechne $\int_2^3 \varphi(x)\,dx$.

2. Für welches x ist $\Phi(x) = 0.8$? Was bedeutet das anschaulich?

3. Es sei X $B(2400, \frac{1}{6})$–verteilt. Berechne

   (a) $P(X = \mu)$.

   (b) $P(X \leq 350)$.

   (c) $P(|X - \mu| \geq 80)$ (vergleiche den Wert auch mit dem entsprechenden Tschebyscheffrisiko) .

   (d) a so, dass $P(|X - \mu| < a) \approx 0.7$ ist.

   (e) x so, dass $P(X < x) \approx 0.01$ ist.

4. Es für jedes Paar (n, p) jeweils eine möglichst kleine Umgebung des Erwartungswertes so gesucht, dass die $B(n, p)$–verteilte Zufallsgröße X mindestens mit der Wahrscheinlichkeit 0.95 einen Wert in dieser Umgebung annimmt. Verwende dazu die Tabelle, Finns Funktion und den Befehl Ualpha aus der Datei simulieren.mn . Vertragen sich die Ergebnisse? Werte: $n = 200$ und $n = 20$, $p = 0.4$ und $p = 0.1$.

5. Es sei X eine $B(n, p)$–verteilte Zufallsgröße. Wie schauen uns an, wie die Standardabweichung σ von n und von p abhängt. Bei festem p ist σ proportional zu $\sqrt{n}$; wenn man n vervierfacht, verdoppelt sich σ, das ist klar. Wie σ bei festem n von p abhängt, muss man sich genauer anschauen. Wegen $\sigma = \sqrt{np(1-p)} = \sqrt{n}\sqrt{p(1-p)}$ kommt es auf die Funktion $p \mapsto \sqrt{p(1-p)}$ an. Wie sieht die aus? Was kommt heraus, wenn p nahe bei 0 oder nahe bei 1 liegt? Für welches p ist der Wert der Funktion maximal, und wie groß ist der maximale Funktionswert?

6. Zum Schluß noch eine interessantere Aufgabe: Acht Jäger haben eine Rotte von zwölf Wildschweinen gestellt. Jeder Jäger nimmt ein Wildschwein aufs Korn, und auf ein Zeichen schießen alle gleichzeitig. Wildschweine sind schlau, jeder kann nur einmal schießen, und die Jäger können keine Absprachen treffen, wer auf welches Wildschwein schießt. Dafür sind die Jäger gute Schützen; jedes Wildschwein, auf das ein Jäger schießt, ist hin. Wieviele der zwölf Wildschweine werden das Treffen überleben?

7 Schätzen einer unbekannten Wahrscheinlichkeit

7.1 Das Konfidenzintervall

Du findest nahezu täglich Meldungen, wieviel Prozent Wählerstimmen eine Partei bekäme, wenn heute Wahl wäre, oder welcher Anteil der Leute für ein absolutes Rauchverbot in Gaststätten oder für ein generelles Tempolimit auf den Autobahnen ist. Woher wissen die eigentlich diese Zahlen? Die befragen nicht alle Leute – hat dich denn schon einmal jemand befragt? – denn das wäre kaum durchführbar und sehr teuer, sondern man „zieht eine Stichprobe", das heißt, man wählt eine Anzahl von Leuten aus und befragt die, und das Ergebnis der Befragung in der Stichprobe rechnet man hoch auf die Grundgesamtheit.

Man muss natürlich etwas Acht geben. Befragt man nur Leute aus einem Stadtteil in Berlin, der mehrheitlich von Sozialhilfeempfängern bewohnt wird, wird man andere Antworten bekommen als in einem Münchner Vorort mit hoher Millionärsdichte. Demoskopen versuchen, eine „repräsentative" Befragung durchzuführen; ihre Stichprobe soll die Verhältnisse in der Gesamtbevölkerung gut widerspiegeln.

Nun, solche praktischen Fragen überlassen wir den Demoskopen. Wir gehen davon aus, dass eine sehr große Grundgesamtheit vorliegt, in der ein unbekannter Anteil p ein bestimmtes Merkmal hat, zum Beispiel am nächsten Wahltag die Piraten wählen zu wollen. Wir wählen n Elemente der Grundgesamtheit zufällig aus und stellen fest, wieviele von diesen n das Merkmal haben. Diese beobachtete Anzahl k ist dann eine Realisierung einer $B(n, p)$–verteilten Zufallsgröße X . Ist dir klar, was das heißt? Noch einmal: **Eine solche Befragung durchführen ist nichts Anderes als einen Wert einer $B(n, p)$–verteilten Zufallsgröße X auslosen**.

Mit solchen Dingen kennen wir uns ja aus. Der gesuchte Anteil p ist der Erwartungswert der Zufallsgröße $\bar{X} = \frac{1}{n}X$, und mit wachsendem n stabilisiert sich der beobachtete Wert von $\bar{X}$ um das gesuchte p. Für die tatsächliche Befragung ergibt sich das aus dem empirischen Gesetz der großen Zahl, für das abstrakte X aus Bernoullis schwachem Gesetz der großen Zahl; das wissen wir alles. Der springende Punkt ist, dass wir nur einen einzigen Wert k für ein festes n bekommen, eben das Ergebnis der Befragung, und der Schätzwert $\hat{p}$, der sich daraus ergibt, ist

$$\hat{p} = \frac{k}{n} \quad ,$$

da sind sich wohl vom Grundschulkind über uns bis zum Professor für Mathematische Statistik alle einig.

Und nun? Ist das schon das Ende der Geschichte? Natürlich nicht. Wir sollten schon mehr sagen können als das Grundschulkind. Du weißt, wie das Histogramm einer $B(n, p)$–verteilten Zufallsgröße X aussieht, und du kennst die Wahrscheinlichkeit

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

dafür, dass der beobachtete Wert k auftritt. Diese Wahrscheinlichkeit ist für jedes p mit $0 < p < 1$ positiv, das heißt, auf Grund der Beobachtung k können wir kein einziges dieser p ausschließen – eine trostlose Konsequenz. Was tun? Stochastik treiben, meine Lieben, heißt zocken. Wer kein Risiko eingehen will, eine falsche Auskunft zu geben, dem hilft die Stochastik herzlich wenig. Was den Stochastiker auszeichnet, ist, dass er dieses Risiko quantifiziert. Er nennt es **Irrtumswahrscheinlichkeit** und bezeichnet es gewöhnlich mit dem Buchstaben α. Ein häufig verwendeter Wert ist $\alpha = 5\% = \frac{1}{20}$. Und so läuft die Sache: Man geht davon aus, dass das beobachtete k nicht zu weit vom Erwartungswert np von X entfernt ist. Genauer: Man bildet eine Umgebung des Erwartungswertes, die so groß ist, dass ein ausgeloster Wert von X mit mindestens der Wahrscheinlichkeit $1-\alpha$ in dieser Umgebung liegt. Natürlich macht man diese Umgebung nur so groß, wie sie unbedingt sein muss. Ihren Radius bestimmt man mit Finns Funktion zu

$$\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sigma \approx 1.96\sigma \quad ,$$

sie geht also von $np-1.96\sqrt{np(1-p)}$ bis $np+1.96\sqrt{np(1-p)}$. Nun bestimmt man das Intervall all der Werte p, für die das beobachtete k in dieser Umgebung liegt: Dieses Intervall heißt **Konfidenzintervall** zur Irrtumswahrscheinlichkeit α, und das Intervall ist die Lösung des Problems.

Der Politiker ist in der Regel nicht besonders glücklich mit diesem Ergebnis, er wollte ja eine klare Antwort, also eine scharfe Zahl, also das wahre p, aber das kann ihm niemand sagen. Jede neue Befragung wird ein neues k und damit ein neues Intervall ergeben, und man muss davon ausgehen, dass im Mittel 5% dieser Intervalle das wahre p nicht einmal überdecken. Natürlich ist es einfacher, nur eine Zahl zur Kenntnis zu nehmen, aber es ist Selbstbetrug. Wenn es den Kunden glücklich macht, sollte man ihm halt seine Illusionen lassen; du allerdings sollst verstehen, wie die Dinge wirklich liegen.

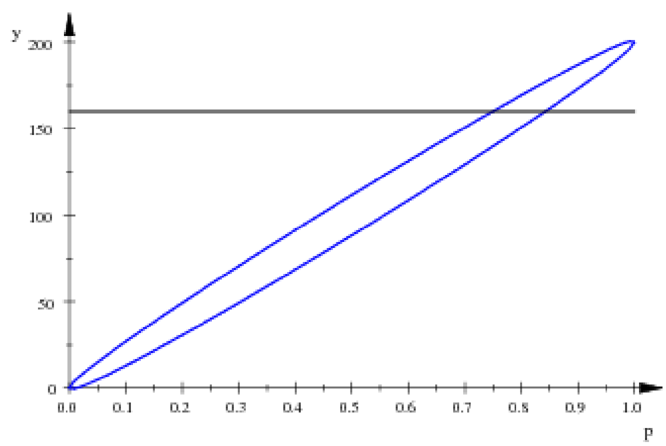


Abbildung 16: Hier siehst du die Kurven $y = 200p \pm \Phi^{-1}(0.95)\sqrt{200p(1-p)} = \mu \pm \Phi^{-1}(0.05)\cdot\sigma$ angetragen. Bei jedem p liegt zwischen den Kurven genau der Bereich mit den Werten des $B(200,p)$–verteilten X, die zusammen etwa die Wahrscheinlichkeit 90% haben. Die waagerechte Linie beim beobachteten Ausgang 160 schneidet die beiden Kurven genau bei den Grenzen des Konfidenzintervalls zur Irrtumswahrscheinlichkeit 0.1.

Aufgabe zum Schätzwert $\hat{p}$. Wir haben die Realisierung k einer $B(n,p)$–verteilten Zufallsgröße X beobachtet. Der Wert von p ist unbekannt und soll geschätzt werden. Es ist plausibel, als Schätzwert $\hat{p}$ für p den Wert zu nehmen, für den die Wahrscheinlichkeit

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} =: f(p) \quad ,$$

dass X den beobachteten Wert k annimmt, möglichst groß wird (Maximum–Likelihood–Methode). Man sieht also diese Wahrscheinlichkeit als Funktion von p an. Rechne aus, für welchen Wert $p \in [0;1]$ das $f(p)$ maximal wird.

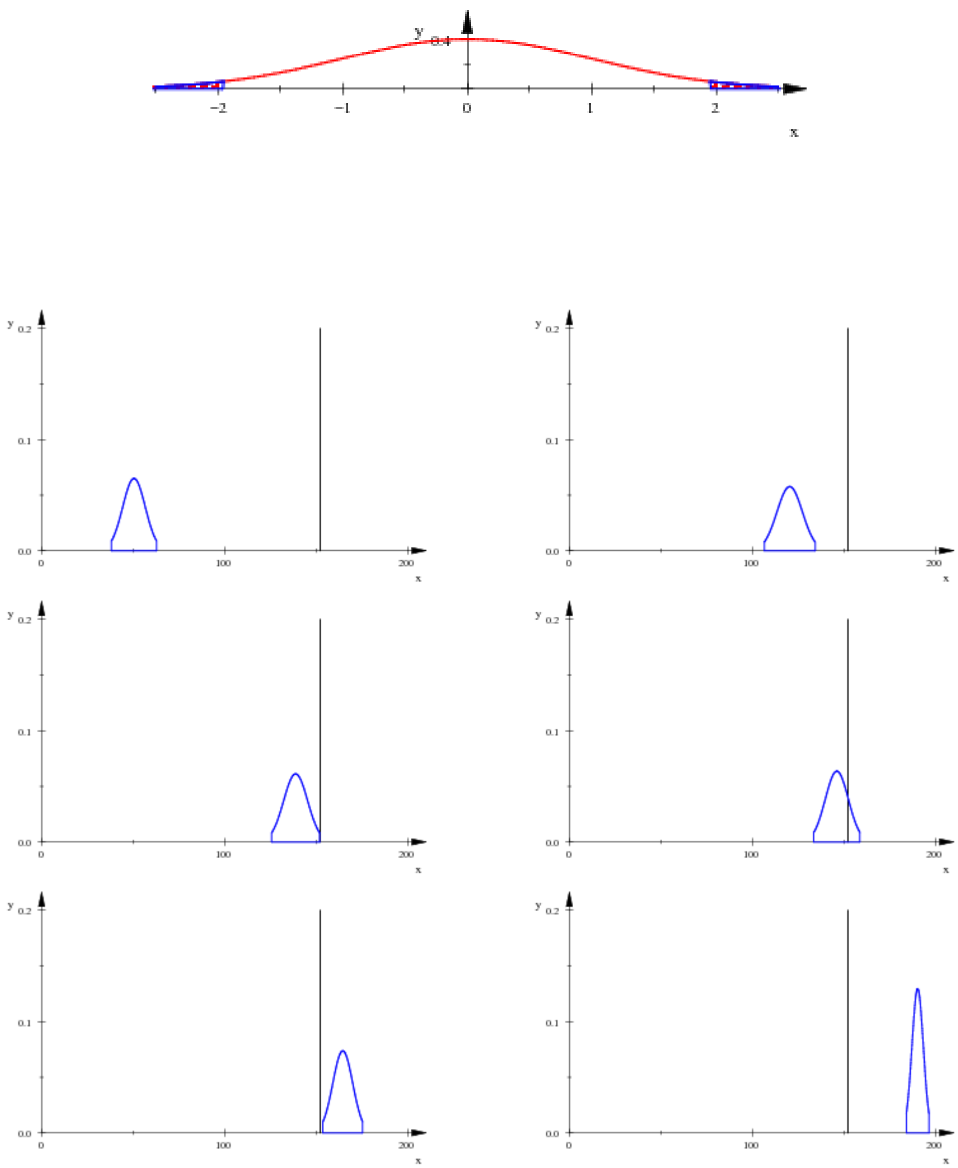


Abbildung 17: Der oberste Graph gehört zu Finns Funktion. Der Inhalt der (vollständigen) Schwänze links und rechts ist jeweils 0.025, die schneiden wir weg. Das Mittelstück hat dann den Flächeninhalt 0.95. Dieses gestutzte Mittelstück destandardisieren wir und erhalten den Mittelteil von Histogrammen zu $n = 200$ mit einigen Werten von p; das sind die übrigen Graphen. Du siehst die Merkmalsachse mit $n = 200$, der schwarze Strich markiert das beobachtete k. Am besten stellst du dir vor, dass p von 0 bis 1 wächst. Die p–Werte der ersten beiden sind zu klein, das p des dritten ist das kleinste p des Konfidenzintervalls, das p des vierten liegt im Innern des Konfidenzintervalls, das p des fünften schon wieder außerhalb. Am sechsten erkennt man schön, wie sich die Form des Histogramms mit p ändert.

7.2 Berechnung des Konfidenzintervalls

Es wurde ein Wert k einer $B(n,p)$–verteilten Zufallsgröße X beobachtet, und nun soll der unbekannte Wert p geschätzt werden. Der, der den Schätzwert haben möchte, muss eine Irrtumswahrscheinlichkeit α vorgeben, die er in Kauf zu nehmen bereit ist. Wir berechnen uns zu diesem α den Wert

$$a := \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad , \tag{28}$$

dann nimmt X etwa mit der Wahrscheinlichkeit $1-\alpha$ einen Wert zwischen $\mu - a\sigma$ und $\mu + a\sigma$ an. Wir gehen davon aus, dass das beobachtete k in diesem Bereich liegt. Dann ist μ höchstens $a\sigma$ von k entfernt:

$$|k - \mu| \leq a\sigma$$

Wir dürfen diese Ungleichung quadrieren, weil beide Seiten nichtnegativ sind. Wenn wir noch μ bzw. σ durch np bzw. $\sqrt{np(1-p)}$ ersetzen, erhalten wir

$$(k - np)^2 \leq a^2 np(1-p) \quad .$$

Umformen ergibt, wenn ich richtig gerechnet habe,

$$(n^2 + a^2 n)p^2 - (2kn + a^2 n)p + k^2 \leq 0 \quad .$$

Auf der linken Seite steht der Term einer nach oben offenen Parabel, die Lösungsmenge der Ungleichung ist also das von den Nullstellen begrenzte Intervall.

Hat das Intervall die Form $[p_l, p_r]$, ist p_l der kleinste p–Wert, für den k noch in der $a\sigma$–Umgebung des Erwartungswertes liegt, und p_r ist der größte (siehe Abbildung 18). Die σ–Werte unterscheiden sich etwas, und zwar um so mehr, je näher $\hat{p} := \frac{k}{n}$

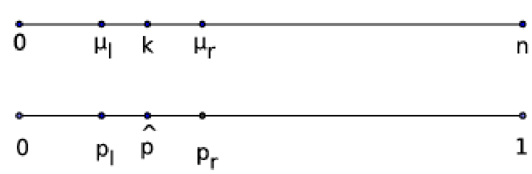


Abbildung 18: k, $\mu_l = np_l$ und $\mu_r = np_r$

bei 0 oder bei 1 ist. In der Regel kann man sich das Leben leicht machen und die Änderung ignorieren. Man geht dann von $\hat{p}$ einfach um

$$\frac{a\sqrt{n\hat{p}(1-\hat{p})}}{n} = a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

nach rechts und nach links und hat schon das **Näherungskonfidenzintervall**

$$\left[\hat{p} - a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \;,\; \hat{p} + a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] \quad . \tag{29}$$

Beispiel. Um herauszufinden, welcher Anteil der Bevölkerung Espelkamps einen Vatertagsausflug macht, befragt ein Institut die 200 Schüler der Stufe Q1 des Söderblom–Gymnasiums. Es antworteten 160 Schüler mit ja. Obwohl die Umfrage offensichtlich nicht repräsentativ ist, kann man natürlich ein Konfidenzintervall zur Irrtumswahrscheinlichkeit $\alpha = \frac{1}{10}$ ausrechnen; die Theorie schützt nicht vor törichten Handlungen. Das exakte Konfidenzintervall berechnet MuPAD zu $[0.74961, 0.84238]$. Um das Näherungskonfidenzintervall zu bekommen, berechnen wir

$$a = \Phi^{-1}(0.95) \approx 1.645\ ,\ \hat{p} = \frac{160}{200},\ \sqrt{\frac{\hat{p}(1-\hat{p})}{200}} \approx 0.028284\quad ,$$

und das führt zu dem Intervall $[0.7534, 0.8465]$.

7.3 Was bedeutet die Irrtumswahrscheinlichkeit α?

Gerade haben wir ein Konfidenzintervall zur Irrtumswahrscheinlichkeit $\alpha = \frac{1}{10}$ für p ausgerechnet und $[0.75, 0.85]$ gefunden. Am liebsten sagte man nun, dass das gesuchte p mit 90%–iger Sicherheit in diesem Intervall läge, aber das ist völliger Quatsch. Das hieße ja, wenn man den Versuch oft durchführte, läge das gesuchte p in etwa 90% der Fälle in diesem Intervall. In Wirklichkeit aber bekäme man bei jeder Durchführung ein neues Intervall, und auf lange Sicht überdeckten etwa 90% dieser Intervalle das wahre p. Ergebnis des Zufallsversuchs „Stichprobe ziehen und zum Ergebnis das 90%–Konfidenzintervall ausrechnen“ ist **ein Intervall** und nicht etwa ein p. Man kann also nur das gefundene Intervall anschauen und zur Kenntnis nehmen, dass es mit einem Verfahren bestimmt wurde, das im Mittel in 90% der Fälle brauchbare Intervalle liefert; mehr ist nicht drin.[5]

Ich spitze die Sache noch etwas zu. Bei der Lernstandserhebung der Stufe 8 verwendet die auswertende Landesbehörde einen Test mit der Irrtumswahrscheinlichkeit 20%. Nehmen wir an, dank der hervorragenden Arbeit der Landesregierung erfüllten alle achten Klassen in ganz Nordrhein–Westfalen die gestellten Anforderungen. Dann würden trotzdem im Mittel 20% der teilnehmenden Klassen beanstandet – zu Unrecht, wohlgemerkt. Schau, da lernst du was über Statistik!

[5]Sieh dir dazu die Abbildung 19 auf Seite 40 an und meditiere ein paar Minuten darüber!

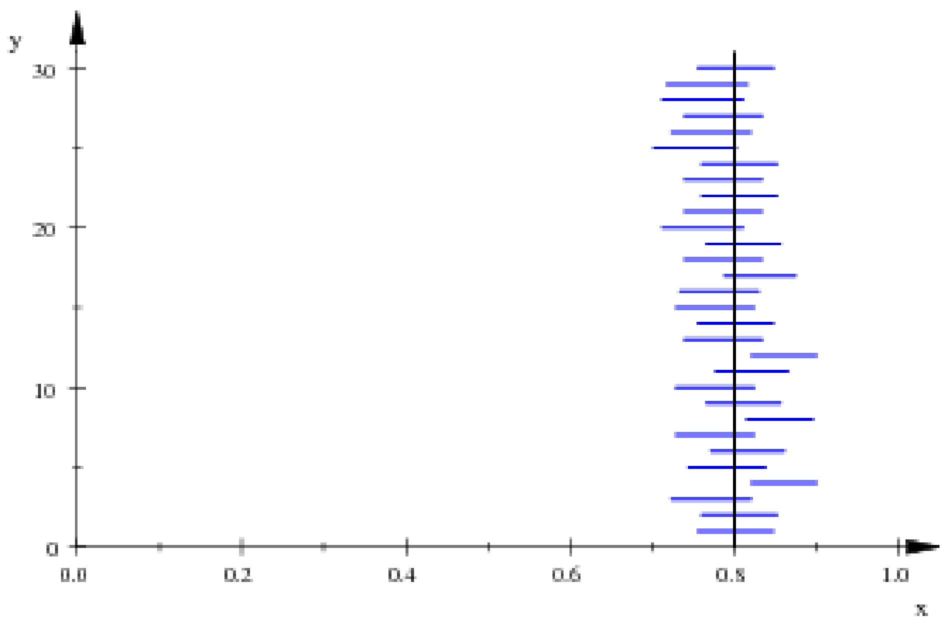


Abbildung 19: Zur Bedeutung von α beim Konfidenzintervall: Hier wurden dreißig Werte der $B(200, \frac{4}{5})$–verteilten Zufallsgröße X ausgelost und zu jedem Wert das Konfidenzintervall zur Irrtumswahrscheinlichkeit $\alpha = 0.1$ berechnet. In dem Bild sind die dreißig Konfidenzintervalle dargestellt, jedes als Strecke parallel zur p–Achse. Bei $p = \frac{4}{5}$ ist eine senkrechte Strecke eingezeichnet. Man erkennt leicht, dass zwar die meisten, aber nicht alle Intervalle das zu schätzende p überdecken. Nochmal: Die Irrtunmswahrscheinlichkeit sagt nur etwas über die Güte des Verfahrens und nichts über das konkrete Intervall!

8 Praxis

8.1 Erinnerung an den Kern der integralen Näherungsformel

Mein Hinweis, dass ihr den Graphen von Finns Funktion φ stets vor Augen haben müsst und dass euch klar sein muss, was sie für binomialverteilte Zufallsgrößen bedeutet, liebe Leute, war nicht nur so dahingesagt. Wenn ihr das außer Acht lasst, dürft ihr euch nicht wundern, wenn ihr im Dunkeln tappt. Hier ist noch einmal der Kern der Sache in hoffentlich klaren Worten.

1. Wir beginnen mit Finns Funktion. Häufig spielt eine Irrtumswahrscheinlichkeit α eine Rolle. Zu α bilden wir

$$a := \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad .$$

Dann ist der Inhalt der Fläche unter dem Graphen von Finns Funktion zwischen $-a$ und a gerade $1 - \alpha$, und die Schwänze haben jeweils den Inhalt $\frac{\alpha}{2}$. Die Situation ist für $\alpha = \frac{1}{10}$ in Abbildung 20 dargestellt.

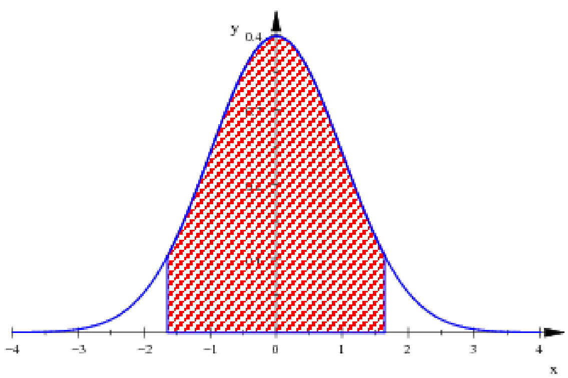


Abbildung 20: Achsensymmetrisches Flächenstück unter Finns Kurve mit Inhalt $1 - \alpha$

2. Es sei X eine $B(n,p)$–verteilte Zufallsgröße mit Erwartungswert μ und Standardabweichung σ. Wie du weißt, wird das Histogramm von

$$X^* = \frac{X - \mu}{\sigma}$$

für nicht zu kleine $V(X)$ gut durch Finns Kurve angenähert. Wir drehen den Spieß einmal herum. Wenn wir Finns Kurve in x–Richtung mit dem Faktor σ und in y–Richtung mit dem Faktor $\frac{1}{\sigma}$ strecken und anschließend um μ nach rechts verschieben, nähert diese neue Kurve mit der Gleichung

$$y = \frac{1}{\sigma} \cdot \varphi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

das Histogramm von X ganz gut an – siehe Abbildung 21.

3. Zentrale Folgerung: Die Werte k von X, die in der $\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \sigma$–Umgebung von μ liegen, haben insgesamt etwa die Wahrscheinlichkeit $1 - \alpha$:

$$P\left(|X - \mu| \leq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sigma\right) \approx 1 - \alpha \qquad (30)$$

Dabei haben wir auf halbe Kästchen keine Rücksicht genommen.

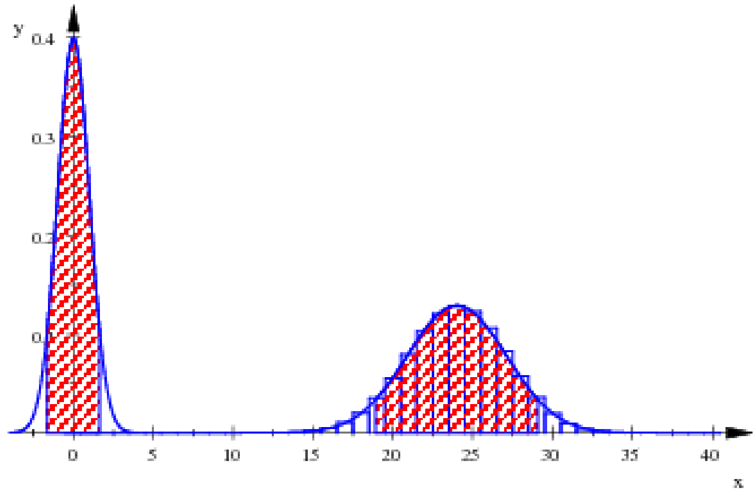


Abbildung 21: Finns Kurve (links) geeignet transformiert passt zum Histogramm des binomialverteilten X (rechts)

4. Ein Beispiel dazu: Wir suchen den Radius einer Umgebung des Erwartungswertes μ der $B(40, \frac{3}{5})$–verteilten Zufallsgröße X, in der ein ausgeloster Wert von X mit der Wahrscheinlichkeit 90% liegt. Der Radius dieser Umgebung ist – ohne Rücksicht auf halbe Kästchen –

$$\Phi^{-1}(0.95)\sigma \approx 1.645\sqrt{40 \cdot \frac{3}{5} \cdot \frac{2}{5}} \approx 5.1 \quad .$$

Die gesuchte Umgebung geht also von $24 - 5.1$ bis $24 + 5.1$. Wir können sagen: Etwa mit der Wahrscheinlichkeit 90% nimmt X einen Wert 19 bis 29 an. In der Tat ist $P(19 \leq X \leq 29) \approx 0.926$.

8.2 Skalen

Es sei X eine $B(n, p)$–verteilte Zufallsgröße. Du kannst an eine Umfrage denken; es werden n Leute aus einer Grundgesamtheit gefragt, ob sie für die Sommerzeit sind, wobei der Anteil der Befürworter in der Grundgesamtheit p ist. Will man den Anteil p auf der Grundlage des beobachteten Ergebnisses der Befragung schätzen, betrachtet man auch die Zufallsgröße $\bar{X} = \frac{1}{n}X$; sie gibt die relative Häufigkeit der Befürworter in der Stichprobe an.

Und schon haben wir drei Skalen, die du gleichzeitig im Blick haben solltest; sie sind in Abbildung 22 dargestellt. Die obere Skala ist die x–Achse von Finns Kurve, die mittlere die Merkmalsachse von X und die untere die Merkmalsachse von $\bar{X}$. Zu einem Wert a auf der Skala von Finns Kurve gehört in natürlicher Weise der Wert $\mu + a\sigma$ auf der Merkmalsachse von X und der Wert $p + a\frac{\sigma}{n}$ auf der Achse von $\bar{X}$, denn es gilt

$$\int_{-a}^{a} \varphi(x)\, dx \;\approx\; P(\mu - a\sigma \leq X \leq \mu + a\sigma) \;=\; P\left(p - a \cdot \frac{\sigma}{n} \leq \bar{X} \leq p + a \cdot \frac{\sigma}{n}\right) \quad .$$

Schauen wir uns ein **Beispiel** an: Es sei $p = \frac{3}{5}$. Wie groß muss n mindestens sein, damit die beobachtete relative Häufigkeit mit der Wahrscheinlichkeit 90% um höchstens 0.01 von p abweicht? Nun, die 0.01 gehören zur unteren Skala, also der

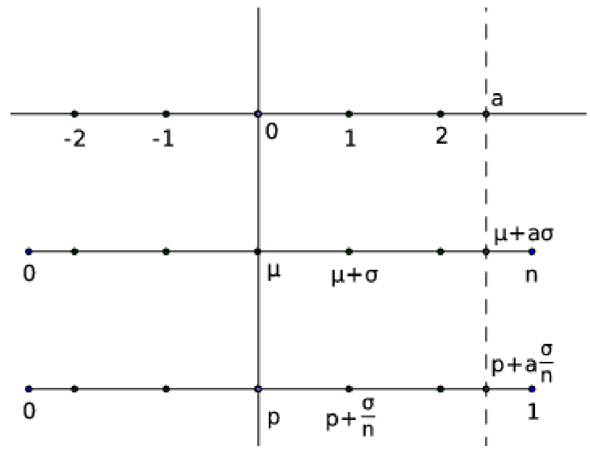


Abbildung 22: Skalen von φ, X und $\bar{X}$

von $\bar{X}$. Auf der mittleren Skala ist die zulässige Abweichung $0.01n$. Wegen der 90% ist auf der Skala von Finns Kurve

$$a = \Phi^{-1}(0.95) \approx 1.645 \quad .$$

Daraus ergibt sich der Ansatz $0.01n = a\sigma$. Man rechnet ein wenig:

$$\begin{aligned} 0.01n &= 1.645\sqrt{n \cdot \frac{3}{5} \cdot \frac{2}{5}} \\ 0.0001n^2 &= 1.645^2 \cdot n \cdot \frac{6}{25} \end{aligned}$$

und findet dann $n \geq 6493.3$. Man sollte also mindestens 6494 Leute befragen. Das ist ziemlich viel. Wohl die allermeisten Aussagen, die auf der Basis von Umfrageergebnissen gemacht werden, stützen sich auf bedeutend kürzere Stichproben; und das verrät Einiges über die Belastbarkeit dieser Aussagen.

8.3 Beispiel für die Berechnung von n bei bekanntem p

Eine beliebig gewählte Schraube aus einer Serienfertigung sei mit der Wahrscheinlichkeit $p = \frac{1}{20}$ unbrauchbar. Wie viele Schrauben sollte man kaufen, wenn man unbedingt 100 brauchbare Schrauben haben muss?

Das ist die Aufgabe. So können wir sie nicht lösen: Selbst bei einer Million gekaufter Schrauben besteht eine positive Wahrscheinlichkeit, dass weniger als 100 davon brauchbar sind. Man muss ein Risiko quantifizieren, das man bereit ist, einzugehen. Sagen wir also, die Schrauben sollen mit 99%–iger Sicherheit reichen. Dann ist die Anzahl n der zu kaufenden Schrauben so zu bestimmen, dass für die $B(n, \frac{19}{20})$–verteilte Zufallsgröße X

$$P(100 \geq X) \geq 0.99 \quad \text{bzw.} \quad P(X < 100) \leq 0.01$$

ist; hier ist X die Anzahl der Brauchbaren unter den n Schrauben.

Mit den üblichen Bezeichnungen heißt das, dass

$$\Phi\left(X^* < \frac{100-\mu}{\sigma}\right) \leq 0.01 \quad ,$$

also

$$100 - \mu - \frac{1}{2} \leq \Phi^{-1}(0.01) \cdot \sigma$$

sein muss. Wir ersetzen das $\leq$ durch $=$ und quadrieren. Das ergibt die quadratische Gleichung

$$(99.5 - np)^2 = \left(\Phi^{-1}(0.01)\right)^2 np(1-p) \quad ,$$

für n. MuPAD gibt die Lösungen 99.4 und 110.34 als Näherungswerte der Lösungen der Gleichung aus, und wir entnehmen dem, dass man 111 Schrauben kaufen sollte. Zur Kontrolle bilden wir für $B(111, \frac{19}{20})$–verteiltes X

$$P(X \geq 100) \approx 0.9903 \quad ,$$

das passt.

Anmerkung: Du folgst hoffentlich meinem Rat, dich immer an den Histogrammen zu orientieren und auch sonst deinen Verstand zu benutzen. Sonst musst du die quadratische Ungleichung selbst lösen, und das ist recht lästig.

8.4 Berechnung von n bei unbekanntem p

Der Anteil p der Leute in der Gesamtbevölkerung, die für die Abschaffung der Hausaufgaben sind, soll auf einen Prozentpunkt genau bestimmt werden. Wie viele Leute muss man befragen?

Wir brauchen eine Irrtumswahrscheinlichkeit, da nehmen wir 5%. Die Anzahl der Leute in einer Stichprobe der Länge n, die für die Abschaffung der Hausaufgaben sind, ist eine $B(n,p)$–verteilte Zufallsgröße X, und der Radius der Umgebung des Erwartungswertes, in dem ein ausgeloster Wert von X mit etwa der Wahrscheinlichkeit 95% liegt, ist $\Phi^{-1}(0.975) \cdot \sigma$ – in solchen Dingen bist du nun fit, da wird kein großes Theater mehr drum gemacht. Damit können wir einen Ansatz hinschreiben, wir brauchen nur die 0.01 von der Skala der relativen Häufigkeiten auf die Merkmalsskala von X umzurechnen:

$$\Phi^{-1}(0.975)\sigma \leq 0.01n$$

Ausschreiben, quadrieren und nach n auflösen ergibt

$$n \geq \frac{\Phi^{-1}(0.975)^2}{0.01^2} p(1-p) \quad .$$

Im ungünstigsten Fall wissen wir über p überhaupt nichts. Dann können wir uns daran halten, dass

$$p(1-p) \leq \frac{1}{4} \tag{31}$$

ist[6], und dann ist wegen

$$1.96^2 \cdot 100^2 \cdot \frac{1}{4} \approx 9603.6$$

$n = 9604$ garantiert ausreichend. Es mag aber sein, dass man Vorinformationen über p hat, vielleicht durch eine kleine Vorumfrage, durch alte Umfragergebnisse oder infolge der Überlegung, dass die selbständige Auseinandersetzung mit dem Gegenstand bei der Erledigung von Hausaufgaben für das Verständnis für jeden erkennbar so wichtig ist, dass p sicherlich mindestens 99% beträgt. In dem Fall kann man $p(1-p)$ durch $0.99 \cdot 0.01$ ersetzen und erhält $n = 381$ als Stichprobenumfang. Und solltest du p in der Nähe von einem Prozent schätzen, kommt dasselbe heraus.

[6] siehe die Aufgabe unten

Aufgabe zu $p(1-p)$. Wie groß kann $p(1-p)$ werden? Du kannst das als gewöhnliche Extremwertaufgabe behandeln, du kannst aber auch nachrechnen, dass jeder Punkt der durch $y = \sqrt{p(1-p)}$ gegebenen Kurve von $(\frac{1}{2}, 0)$ die Entfernung $\frac{1}{2}$ hat: die Kurve ist ein Halbkreis! Vergleiche Abbildung 23.

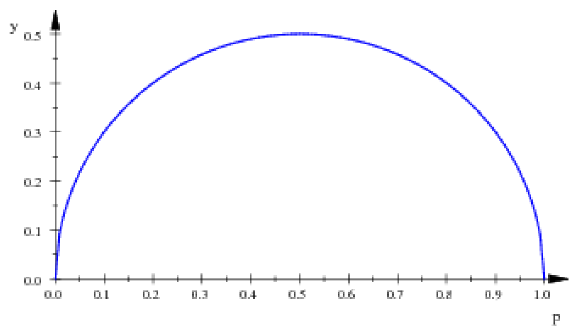


Abbildung 23: Der Graph von $p \mapsto \sqrt{p(1-p)}$ ist ein Halbkreis!

8.5 Ein Testproblem

Ein Großhändler hat mit dem Schraubenhersteller eine Vereinbarung getroffen: Finden sich in einer Stichprobe von 100 Schrauben mehr als 7 unbrauchbare, wird die Lieferung zurückgeschickt.[7] Mit welcher Wahrscheinlichkeit wird eine Lieferung mit 5% unbrauchbaren Schrauben zurückgeschickt, mit welcher Wahrscheinlichkeit wird eine Lieferung mit 12% unbrauchbaren Schrauben akzeptiert?

Wir bezeichnen die $B(100, p)$–verteilte Zufallsgröße mit X_p. Die Wahrscheinlichkeit, dass eine Lieferung mit 5% unbrauchbaren Schrauben zurückgeschickt wird, ist

$$P(X_{0.05} > 7) = 1 - P(X_{0.05} \leq 7) \approx 1 - 0.872 = 0.128 \quad ,$$

und die Wahrscheinlichkeit, dass eine Lieferung mit 12% unbrauchbaren Schrauben akzeptiert wird, ist

$$P(X_{0.12} \leq 7) \approx 0.0761 \quad .$$

[7] Eine solche Vorschrift heißt Entscheidungsregel.

9 Zweite Klausur am 6. Juni 2012

1. **Idealer Würfel**
   Ein idealer Würfel wird 2400–mal geworfen. Es sei X die Anzahl der geworfenen Fünfen.

   (a) Wieviele Fünfen wird man im Mittel erhalten? [3]

   (b) Mit welcher Wahrscheinlichkeit wird man bei einer Durchführung des Versuches genau diese Anzahl von Fünfen beobachten? [10]

   (c) Mit welcher Wahrscheinlichkeit wird man höchstens 410 Fünfen beobachten? [6]

   (d) Gib einen Bereich an, in dem die beobachtete Anzahl von Fünfen aller Voraussicht nach liegen wird. [6]

   (e) Wie oft muss man mindestens würfeln, wenn man mit mindestens 90% Wahrscheinlichkeit mindestens 100 Fünfen haben will? Schreibe einen Ansatz hin und bearbeite ihn, bis nur noch eine Ungleichung mit konkreten Zahlen für die gesuchte Anzahl n gelöst werden müsste. Berechne den Wert selbst **nicht**. [14]

   (f) Jens schreibt die Zahlen von 1 bis 6 auf einen Zettel und wirft den idealen Würfel fünfmal. Nach jedem Wurf streicht er die gewürfelte Zahl auf dem Zettel durch. Mit welcher Wahrscheinlichkeit p wird die 2 auf dem Zettel alle fünf Würfe überstehen? Wieviele Zahlen werden am Ende des Spiels im Mittel noch auf dem Zettel sein? [8]

2. **Schätzen**
   Der Gesundheitsminister will wissen, wie hoch der Anteil der Leute ist, die zu Organspenden bereit sind.

   (a) Im Jahr 2008 waren 2680 von 4000 befragten Personen zu einer Organspende bereit. Zum Ärger des Ministers wollte ein Statistiker, der daraus den Anteil der Organspender berechnen sollte, eine Irrtumswahrscheinlichkeit wissen. Der Minister will sich ja gar nicht irren! Wozu braucht man denn eine Irrtumswahrscheinlichkeit? [4]

   (b) Auf der nächsten Seite siehst du eine Grafik, die zu den vorliegenden Daten passen sollte – allerdings ist der Bearbeiter vom Umfrageergebnis 268 unter 400 ausgegangen. Welches Konfidenzintervall ergibt sich aus der Zeichnung? Trage Hilfslinien ein, so dass zu erkennen ist, wie du dein Ergebnis gefunden hast. [6]

   (c) Kann das eben zeichnerisch bestimmte Intervall zu den Originaldaten passen? [4]

   (d) Der Minister hat auf Rat seiner Fachleute die Irrtumswahrscheinlichkeit 5% gewählt. Schreibe eine Gleichung hin, aus der man die Grenzen des exakten Konfidenzintervalls berechnen kann. Löse die Gleichung nicht. [8]

   (e) Der Minister schaut sich das vom Statistiker berechnete Intervall für den Anteil p an und wundert sich, dass der beobachtete Schätzwert $\hat{p}$ nicht in der Mitte des Intervalls liegt. Gib du Auskunft, ob $\hat{p}$ näher an der linken oder näher an der rechten Grenze des Intervalls liegt, und gib einen Hinweis, warum das so ist. Du sollst das Intervall dafür aber keinesfalls ausrechnen! [8]

   (f) Berechne die Grenzen des Näherungskonfidenzintervalls zu den gegebenen Daten. [8]

(g) Der Minister schaut sich dein Ergebnis an. Aha, in diesem Intervall liegt der gesuchte Anteil p mit der Wahrscheinlichkeit 95%, ja? Schön wär's. Erkläre ihm, was die 95% wirklich bedeuten. [6]

(h) Eine neue Umfrage im Jahr 2011 ergab 74% potentielle Organspender. Leider konnte ich nicht in Erfahrung bringen, wieviele Leute befragt wurden, und ohne diese Angabe bedeutet der Wert wenig. Nehmen wir an, dass der im Jahr 2008 bestimmte Anteil das wahre p ist und dass man jetzt nur 100 Leute befragt hat. Kann man dann sagen, dass sich der Anteil der Organspender erhöht hat? [8]

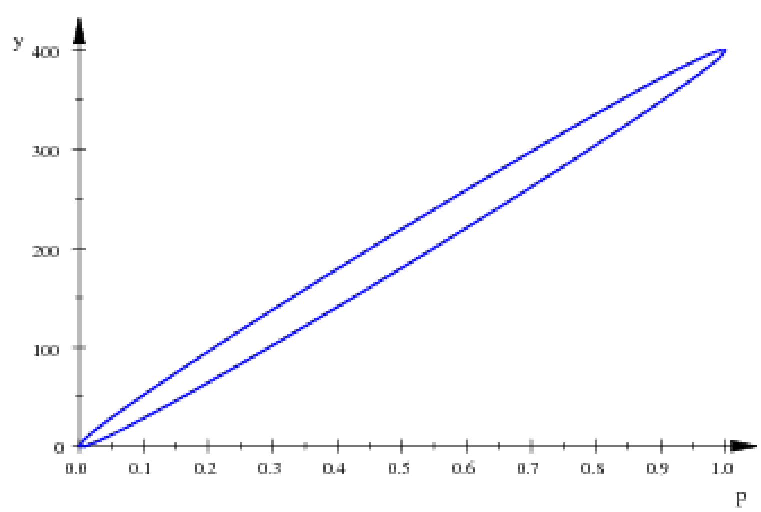
(i) Wieviele Leute muss man befragen, wenn das Näherungskonfidenzintervall höchstens den Radius 0.02 haben soll? [10]

3. **Noten**

Acht der vierundzwanzig Schüler des Kurses hatten eine drei auf dem Zeugnis. Von den neun Schülern mit dem zweiten Leistungsfach Biologie hatten vier eine drei.

(a) Sind die Merkmale „eine drei auf dem Zeugnis haben“ und „Biologie als zweites Leistungsfach haben“ stochastisch unabhängig? [6]

(b) Die Stufenleiterin sieht in ihrem Verwaltungsprogramm bei einem Schüler des Kurses die Zeugnisnote drei in Mathematik. Mit welcher Wahrscheinlichkeit hat der Schüler Biologie als zweites Leistungsfach? [6]



10 Stetig verteilte Zufallsgrößen

10.1 Erstes Beispiel: Eine gleichverteilte Zufallsgröße

Es sei X eine zufällig gewählte Zahl aus dem Intervall $[0;2]$. Wir können uns die Sache so vorstellen, dass wir aus diesem Stück der Zahlengeraden einen Kreis machen und einen Zeiger drehen, wie beim Glücksrad, und die Zahl ist ausgelost, bei der der Zeiger stehenbleibt. Allerdings soll der Zeiger genau eine reelle Zahl auslosen, und da siehst du schon, dass die Sache nur in unserer Vorstellung perfekt läuft; jede Realisierung ist unvollkommen.

Vielleicht stößt du dich daran, dass die Zahlen 0 und 2 beim Verkleben zu einer wurden, aber das ist nicht schlimm: jede Zahl aus $[0;2]$ wird notwendig mit der Wahrscheinlichkeit 0 ausgelost, auch die, die wirklich drankommt. Die Zahl 3 hingegen wird ebenfalls mit der Wahrscheinlichkeit 0 ausgelost, sie kann aber nicht drankommen. In der Stochastik sagt man übrigens von einem Ereignis, das die Wahrscheinlichkeit 0 hat, es sei fast unmöglich.

Aber zurück zu unserem X. Wir losen einmal 60 Werte von X aus und stellen sie auf der Zahlengeraden dar (siehe Abbildung 24), sie liegen regellos verstreut, wie es ja auch sein soll.

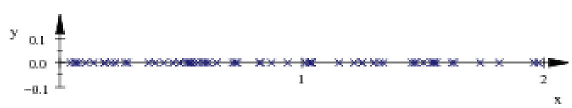


Abbildung 24: Sechzig ausgeloste Werte von X

Offensichtlich gilt für alle Zahlen a, b mit $0 \leq a \leq b \leq 2$

$$P(a \leq X \leq b) = \frac{b-a}{2} \quad .$$

Teilen wir das Intervall in zwanzig gleiche Teile ein, fällt ein ausgeloster Wert von X mit der Wahrscheinlichkeit $\frac{1}{20}$ in ein bestimmtes Teilstück. Wir führen den Versuch sehr oft durch und zeichnen ein empirisches Histogramm: auf jedem Teilstück steht ein Kästchen, dessen Inhalt die beobachtete relative Häufigkeit der Werte angibt, die in dieses Teilstück gefallen sind (Abbildung 25).

Wenn wir $n = 500$ Werte von X auslosen, diese der Größe nach ordnen, beim k–ten Wert x_k den Punkt $(x_k, \frac{k}{n})$ zeichnen und diese Punkte der Reihe nach verbinden, erhalten wir so etwas, wie es in Abbildung 26 dargestellt ist. Es handelt sich um eine empirische Verteilungsfunktion von X, also näherungsweise um den Graphen der Funktion F mit

$$F(x) = P(X \leq x) \quad ;$$

idealerweise kommt ein Stück der Geraden $y = \frac{1}{2}x$ heraus.

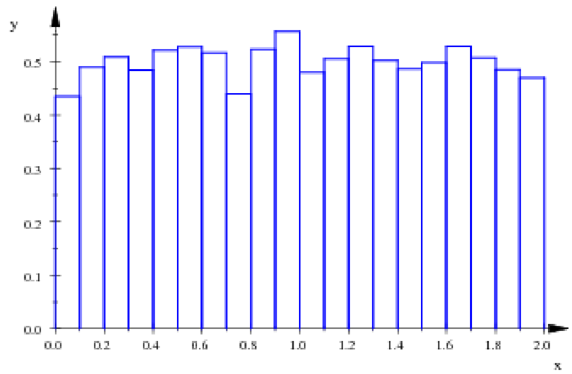


Abbildung 25: Empirisches Histogramm von X

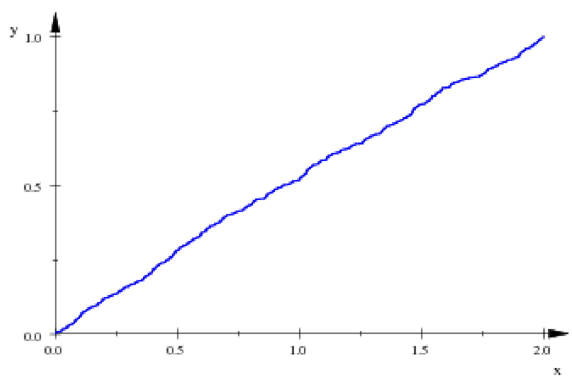


Abbildung 26: Empirische Verteilungsfunktion von X

10.2 Eine Summe gleichverteilter Zufallsgrößen

Wir verschaffen uns $m = 30$ Kopien $X_1, X_2, \ldots, X_m$ von X und bilden die Zufallsgröße

$$Y = X_1 + X_2 + \cdots + X_m = \sum_{k=1}^{m} X_k \quad ;$$

praktisch gesprochen: wir losen m Werte von X aus und addieren sie. Die Werte von Y liegen dann im Intervall $[0, 2m]$. Die den Abbildungen 24, 25 und 26 entsprechenden Abbildungen für Y siehst du in Abbildung 27 auf der Seite 50.

Beim empirischen Histogramm von Y in Abbildung 27 ist ein deformiertes Exemplar von Finns Kurve mit eingezeichnet, es passt ganz gut. Das heißt aber doch, dass für alle a, b mit $0 \leq a \leq b \leq 60$

$$P(a \leq Y \leq b) \approx \int_a^b f(x)\, dx \tag{32}$$

ist; dabei ist f die deformierte Gaußkurve. Für Y gilt eine ähnliche Aussage wie der integrale Näherungssatz von de Moivre und Laplace für binomialverteilte Zufallsgrößen!

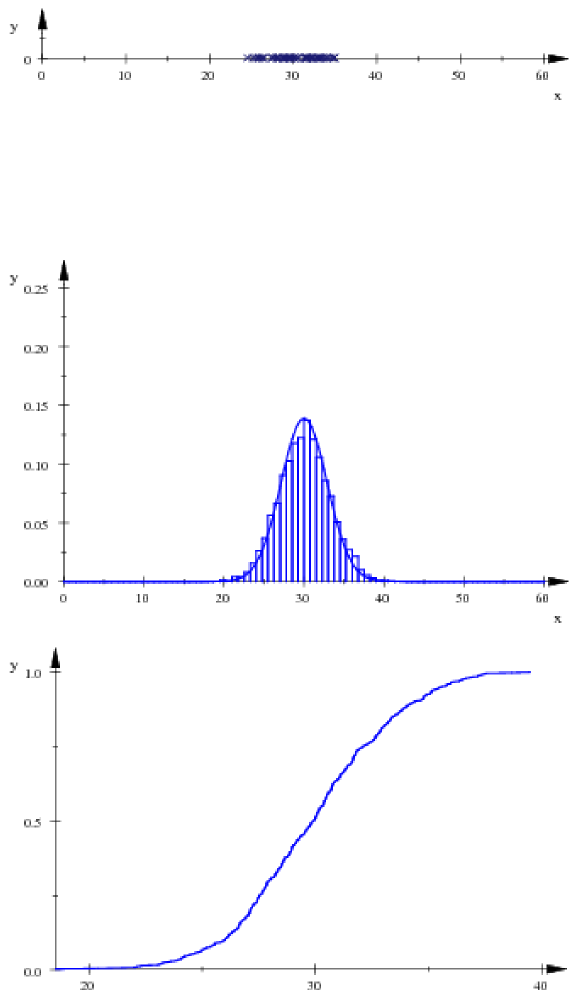


Abbildung 27: Ausgeloste Werte, empirisches Histogramm und empirische Verteilungsfunktion von Y

10.3 Deformationen von Finns Kurve

Auf welche Weise darf man die Gaußkurve $y = \varphi(x)$ deformieren? Natürlich darf man sie in x–Richtung verschieben, aber das reicht nicht, wir müssen auch die Merkmalsachse strecken. Abbildung 28 zeigt die Gaußkurve $y = \varphi(x)$ und die Transformierte $y = \varphi\left(\frac{x}{\sigma}\right)$ für $\sigma = 2$ Diese Transformation macht die Kurve für $\sigma > 1$ breiter; an der Stelle x wird ja der Funktionswert von $\frac{x}{\sigma}$ angetragen, und $\frac{x}{\sigma}$ liegt näher an 0 als x selbst. Damit der Flächeninhalt bleibt, müssen die y–Werte mit $\frac{1}{\sigma}$ multipliziert werden. Nun sind wir am Ziel: Eine brauchbare Deformierte von Finns Kurve $y = \varphi(x)$ hat eine Gleichung der Form

$$f(x) \;=\; \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) \;=\; \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{für } \sigma > 0 \text{ und } \mu \in \mathbb{R}. \tag{33}$$

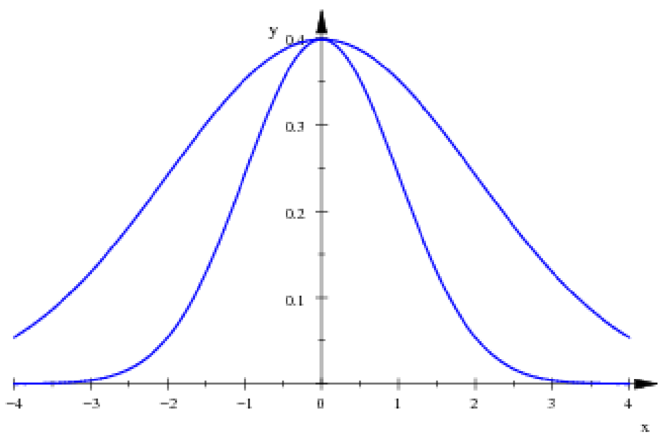


Abbildung 28: Die Kurven $y = \varphi(x)$ und $y = \varphi\left(\frac{x}{2}\right)$

Für $\mu = 30$ und $\sigma = \sqrt{10}$ bekommen wir die passende Kurve zu unserem empirischen Histogramm! Und mit Hilfe dieser Kurve können wir die Wahrscheinlichkeit, dass die Zufallsgröße Y einen Wert in einem Intervall $[a, b]$ annimmt, auf folgende Weise exakt angeben. Es ist

$$\begin{aligned} P(a \le Y \le b) &= \int_a^b f(x)\,dx = \int_a^b \frac{1}{\sqrt{2\pi}\sqrt{10}} \cdot e^{-\frac{1}{2}\left(\frac{x-30}{\sqrt{10}}\right)^2} dx = \\ &\qquad = \int_{\frac{a-30}{\sqrt{10}}}^{\frac{b-30}{\sqrt{10}}} \varphi(x)\,dx = \Phi\left(\frac{b-30}{\sqrt{10}}\right) - \Phi\left(\frac{a-30}{\sqrt{10}}\right) \quad . \end{aligned}$$

Diese Funktion f heißt **Dichtefunktion** von Y.

10.4 Die allgemeinen Begriffe

Nach diesen einführenden Beispielen will ich euch die allgemeinen Begriffe vorsetzen.

18 Definition
Es sei $f : \mathbb{R} \to \mathbb{R}$ eine Funktion mit folgenden Eigenschaften:

1. *$f(x) \geq 0$ für alle $x \in \mathbb{R}$*
2. *Für alle $a, b \in R$, $a \leq b$, existiert das Integral $\int_a^b f(x)\,dx$*
3. *$\int_{-\infty}^{\infty} f(x)\,dx = 1$*

Dann ist durch

$$P(a < X \leq b) := \int_a^b f(x)\,dx \quad \text{für alle } a, b \in \mathbb{R},\ a \leq b$$

eine **stetig verteilte Zufallsgröße** *X definiert. Die Funktion f heißt* **Dichtefunktion** *von X und die Funktion F mit*

$$F(x) := P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

heißt **Verteilungsfunktion** *von X. Erwartungswert und Varianz von X sind definiert durch*

$$\mu = \mathrm{E}(X) := \int_{-\infty}^{\infty} x f(x)\,dx \quad \text{und} \quad \mathrm{V}(X) := \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \quad ,$$

falls diese Integrale existieren.

Da man unseren Beweis der Tschebyscheffungleichung wörtlich auf die neuen Zufallsgrößen übertragen kann, haben μ und σ die gewohnten Bedeutungen. Auch die Rechenregeln für Erwartungswert und Varianz gelten für die neuen Zufallsgrößen, aber da schauen wir nicht genauer hin, das würde doch etwas aufwendig.

Aufgaben

1. Bestimme für die Zufallsgröße X, also die Gleichverteilung auf $[0, 2]$, Dichtefunktion und Verteilungsfunktion und berechne Erwartungswert und Varianz.
2. Berechne mit Hilfe der Regeln Erwartungswert und Varianz der Zufallsgröße $Y = X_1 + X_2 + \cdots + X_n$, also der Summe von n unabhängigen Kopien von X.
3. Es sei
$$f(x) = \begin{cases} 0 & \text{für } x \leq 0 \\ e^{-x} & \text{für } 0 < x \end{cases} \quad .$$
Die Zufallsgröße X mit der Dichtefunktion f heißt **exponentialverteilt**, X könnte zum Beispiel der Zerfallszeitpunkt eines radioaktiven Teilchens sein. Berechne die Verteilungsfunktion F sowie Erwartungswert und Varianz von X.

10.5 Normalverteilte Zufallsgrößen

Offensichtlich sind die Funktionen, deren Graphen wir als Ergebnis zulässiger Deformationen von Finns Kurve erhalten haben, als Dichtefunktionen geeignet.

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

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

heißt **normalverteilt**, *genauer $N(\mu, \sigma^2)$–verteilt.*

20 Lemma
Es sei X eine $N(\mu, \sigma^2)$–verteilte Zufallsgröße. Dann gelten die folgenden Aussagen.

1. $P(a < X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$
2. $\mathrm{E}(X) = \mu$
3. $\mathrm{V}(X) = \sigma$

Zum Beweis benötigt man nur Integrationsregeln und etwas Durchhaltevermögen.

10.6 Der Zentrale Grenzwertsatz

Dass wir bei unserem Einführungsbeispiel Y wieder auf Finns Kurve gestoßen sind, ist kein Zufall. Es gibt einen Satz, eigentlich eine ganze Sammlung von Sätzen, die man unter dem Namen „Zentraler Grenzwertsatz“ führt. Beweise dieser Sätze sind sehr schwer, und sie brauchen viel Theorie. Ein Mathematikstudent, der eine Einführungsvorlesung in die Wahrscheinlichkeitstheorie hört, sieht darin keinen Beweis des Satzes; der kommt erst in weiterführenden Vorlesungen. Aber die Aussage des Satzes ist für uns erreichbar: Gegeben seien unabhängige Zufallsgrößen

$$X_1, X_2, X_3, \ldots \quad ,$$

und die Größen $\mu_i := \mathrm{E}(X_i)$ und $\sigma_i := \sqrt{\mathrm{V}(X_i)}$ mögen für alle i existieren. Wir setzen für $n \in \mathbb{N}$

$$m_n := \mu_1 + \mu_2 + \cdots + \mu_n \quad \text{und} \quad s_n := \sigma_1 + \sigma_2 + \cdots + \sigma_n \quad .$$

Unter sehr weitherzigen Voraussetzungen sagt der Zentrale Grenzwertsatz dann, dass die Summe $X_1 + X_2 + \cdots + X_n$ für genügend große n in guter Näherung $N(m_n, s_n^2)$–verteilt ist. Griffig formuliert: Die Summe unabhängiger Zufallsgrößen ist approximativ normalverteilt. Bei Kugellagerkugeln aus einer Serienfertigung ist die Abweichung vom Sollwert Ergebnis vieler unabhängiger zufallsbedingter kleiner Größen, deshalb greift der Zentrale Grenzwertsatz, und das ist der Grund, dass man hier wie in zahllosen anderen Fällen einer Normalverteilung begegnet.

Einen Sonderfall schauen wir uns noch an: Es sei X die Anzahl der Erfolge bei einer Bernoullikette der Länge n mit Erfolgswahrscheinlichkeit p. Dann können wir X als Summe

$$X = X_1 + X_2 + \cdots + X_n$$

auffassen, wobei die X_i unabhängige Kopien der Anzahl der Erfolge bei einem einfachen Bernoulliversuch mit der Erfolgswahrscheinlichkeit p sind – diesen Ansatz haben wir ja häufig verwandt. Nun ist $\mathrm{E}(X_i) = p$ und $\mathrm{V}(X_i) = p(1-p)$, folglich ist $\mathrm{E}(X) = np$ und $\mathrm{V}(X) = np(1-p)$, das kennst du alles. Aber nun sagt der Zentrale Grenzwertsatz, dass X approximativ $B(np, np(1-p))$–verteilt ist. Das weißt du schon, das ist nämlich die Aussage der Näherungsformeln von de Moivre und Laplace, und die sind folglich ein Spezialfall der Zentralen Grenzwertsatzes! Das mag dir ein Hinweis sein, dass ein Beweis des Zentralen Grenzwertsatzes kein Pappenstiel ist; den haben Leute wie A. Kolmogoroff mit erheblichen Anstrengungen zuwege gebracht.

10.7 Beispiel

Die Zufallsgröße X : „Zerfallszeitpunkt eines radioaktiven Teilchens" ist exponentialverteilt, das heißt, X hat die Dichtefunktion f und die Verteilungsfunktion F mit

$$f(x) = \begin{cases} 0 & \text{für } x \le 0 \\ e^{-x} & \text{für } 0 < x \end{cases} \quad \text{und} \quad F(x) = \begin{cases} 0 & \text{für } x \le 0 \\ 1 - e^{-x} & \text{für } 0 < x \end{cases} .$$

Die Graphen kennst du gut, du findest sie noch einmal in Abbildung 29. Physiker,

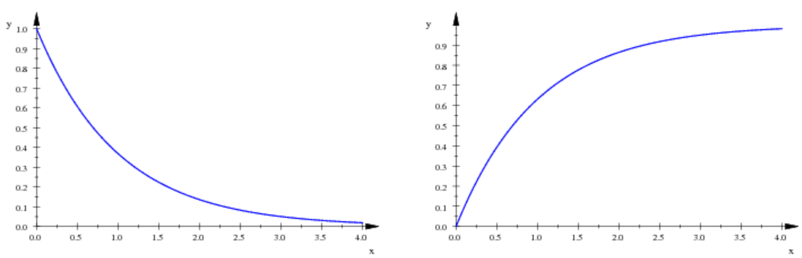


Abbildung 29: Dichtefunktion f und Verteilungsfunktion F des exponentialverteilten X

die reale Teilchen betrachten, brauchen noch einen einstellbaren Parameter, aber das soll uns nicht kümmern.

Wir möchten gern Werte von X auslosen. Das geht erstaunlich einfach: Es sei G gleichverteilt auf $[0;1]$. Wir bilden die Zufallsgröße $Z := F^{-1}(G)$. Das ist es schon! Schau, es sei $z \in \mathbb{R}$. Genau dann wird ein Wert $Z \le z$ ausgelost, wenn ein $G \le g := F(z)$ kommt, und die Wahrscheinlichkeit dafür ist eben $g = F(z)$. Das heißt, dass Z die Verteilungsfunktion F hat, und dann brauchen wir Z von X nicht zu unterscheiden; die Verteilungsfunktion bestimmt die Zufallsgröße vollständig.

Wenn wir $N = 1000$ Werte von X auslosen, werden die alle verschieden sein. Am besten veranschaulichen wir uns die Werte mit Hilfe der empirischen Verteilungsfunktion. Dazu ordnen wir die Werte der Größe nach und bilden den Polygonzug, der die Punkte $(x_k, \frac{k}{N})$, $k = 1, 2, 3, \ldots, N$, der Reihe nach verbindet. Das Ergebnis findest du in Abbildung 30, die schwarze Kurve ist die theoretische Kurve, also der Graph von F. Das passt ganz gut.

Der Zentrale Grenzwertsatz sagt nun, dass die Summe unabhängiger Kopien von X approximativ normalverteilt ist. Um dies sichtbar zu machen, haben wir die Summe

$$Y = X_1 + X_2 + \cdots + X_{30}$$

dreißig unbahängiger Kopien von X gebildet, $N = 1000$ Werte von Y ausgelost und wieder die empirische Verteilungsfunktion gebildet. Es kam die vertraute Form des Graphen von Φ heraus, natürlich deformiert. Du findest die Kurve in Abbildung 31. Alle Abbildungen wurden mit dem zweiten Teil von gleichverteilung.mn erstellt, dort findest du auch weitere Abbildungen und die Möglichkeit, selbst zu experimentieren.

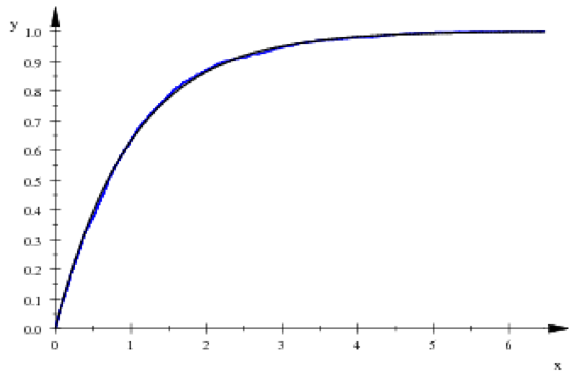


Abbildung 30: Empirische und theoretische Verteilungsfunktion von X

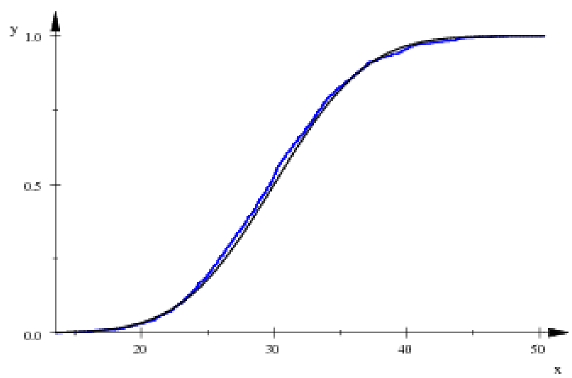


Abbildung 31: Empirische und theoretische (schwarz) Verteilungsfunktion der Summe von 30 unabhängigen exponentialverteilten Zufallsgrößen

10.8 Eine merkwürdige Sache

Es sei $x > 0$. Die Wahrscheinlichkeit, dass unser radioaktives Teilchen bis zum Zeitpunkt x zerfallen sein wird, ist

$$P(X \leq x) = F(x) = 1 - e^{-x} \quad .$$

Nun sei auch $a > 0$. Wir berechnen die bedingte Wahrscheinlichkeit, dass das Teilchen bis zum Zeitpunkt $a + x$ zerfallen sein wird, unter der Bedingung, dass es zum Zeitpunkt a noch existiert, also

$$P_{X>a}(X \leq a + x) \quad .$$

Wenn du strikt anwendest, was du über bedingte Wahrscheinlichkeiten gelernt hast, bekommst du

$$\begin{aligned} P_{X>a}(X \leq a + x) = \frac{P(a < X \leq a + x)}{P(X > a)} &= \frac{F(a + x) - F(a)}{1 - F(a)} \\ &= \frac{1 - e^{-(a+x)} - (1 - e^{-a})}{e^{-a}} = 1 - e^{-x} = F(x) \quad , \end{aligned}$$

du kannst aber auch anschaulich argumentieren: Die gesuchte bedingte Wahrscheinlichkeit ist der Anteil des Flächenstücks unter dem Graphen der Dichtefunktion f zwischen a und $a + x$ am gesamten Flächenstück rechts von a (siehe Abbildung

32). Das Merkwürdige dabei ist, dass die Wahrscheinlichkeit, dass das Teilchen innerhalb der kommenden Zeitspanne x zerfällt, wenn es erst den Zeitpunkt a erlebt hat, gleich der Wahrscheinlichkeit $F(x)$ ist, dass es in der Zeitspanne der Länge x zerfällt, die jetzt beginnt. Radioaktive Teilchen altern nicht! Ganz gleich, ob das betrachtete Teilchen schon seit tausenden von Jahren existiert oder ob es vor zehn Sekunden entstanden ist, die Wahrscheinlichkeit, dass es innerhalb der nächsten Stunde zerfällt, ist immer die gleiche! Das sieht bei Lebewesen gänzlich anders aus. Der letzte Grund ist, mathematisch gesehen, eine besondere Eigenschaft der Exponentialfunktion: Ausgehend von der gleichen Ausgangskurve kann man sie auf die gleiche Endkurve bringen, indem man sie in y–Richtung streckt oder in x–Richtung verschiebt.

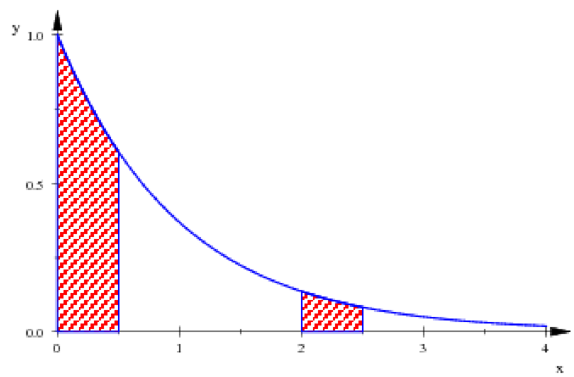


Abbildung 32: Radioaktive Teilchen altern nicht!

11 Hypothesentests

11.1 Die Entscheidungsregel und ihre Gütefunktion

Johann arbeitet in der Qualitätskontrolle der Firma Hammer und Nagel. Seine Aufgabe ist es, jeder eingehenden Lieferung Schrauben 100 zufällig gewählte Schrauben zu entnehmen und festzustellen, wie viele von ihnen schadhaft sind. Falls er dabei mehr als fünf schadhafte Schrauben findet, geht die Lieferung zurück.

Das, was Johann da täglich durchführt, ist ein Test, und seine Arbeitsanweisung ist die **Entscheidungsregel** des Tests. Beschreiben wir den Vorgang mit unseren Begriffen aus der Wahrscheinlichkeitstheorie: Die Anzahl der schadhaften Schrauben, die Johann in einer Stichprobe findet, ist eine Zufallsgröße X, und diese Zufallsgröße ist $B(100, p)$–verteilt; dabei ist p der Anteil der schadhaften Schrauben in der Lieferung, die er gerade untersucht. Liegt der beobachtete Wert von X im **kritischen Bereich** $W_1 := \{6, 7, 8, \ldots, 100\}$, wird die Lieferung beanstandet.

Damit wir eine Vorstellung bekommen, wie der Test wirkt, betrachten wir seine **Gütefunktion**

$$p \mapsto P(W_1) = P(X > 5) = \sum_{j=6}^{100} \binom{100}{j} p^j (1-p)^{100-j} \quad ,$$

ihr Graph ist in Abbildung 33 zu sehen. Die Gütefunktion gibt dir für jeden Anteil

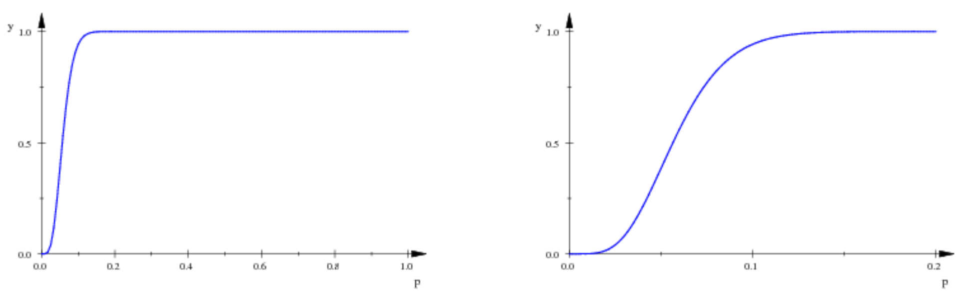


Abbildung 33: Graph der Gütefunktion $p \mapsto P(W_1)$

p schadhafter Schrauben in der Lieferung die Wahrscheinlichkeit an, dass der Test anschlägt und die Lieferung abgelehnt wird. Ist $p \leq 2\%$, wird die Lieferung praktisch immer angenommen, ist $p \geq 15\%$, wird sie praktisch immer abgelehnt.

11.2 Fehler erster und zweiter Art

Johanns Chef kennt die Vereinbarung, die die Firma Hammer und Nagel mit dem Schraubenkonzern getroffen hat, und die sieht so aus: Eine Lieferung darf maximal 5% schadhafter Schrauben enthalten. Johann kann nun zwei Fehler machen (obwohl er eigentlich alles richtig macht!): er kann eine ordentliche Lieferung ablehnen, das ist der **Fehler erster Art**, und er kann eine schlechte Lieferung durchwinken, das ist der **Fehler zweiter Art**. Die Wahrscheinlichkeit für einen Fehler erster Art ist das Risiko erster Art, es ist hier maximal

$$\alpha = P(X_{0.05} > 5) \approx 0.384 \quad .$$

Die Wahrscheinlichkeit für einen Fehler zweiter Art ist das Risiko zweiter Art, es ist kleiner als

$$\beta = P(X_{0.05} \leq 5) = 1 - P(X_{0.05} > 5) \approx 0.616 \quad .$$

Man hätte natürlich am liebsten, dass der Test **jede** ordentliche Lieferung akzeptiert und **jede** schlechte Lieferung ablehnt, wie der Graph in Abbildung 34 zeigt, aber das ist nicht möglich. Man kann nur die Flanke der Gütefunktion steiler machen, indem man den Stichprobenumfang vergrößert.

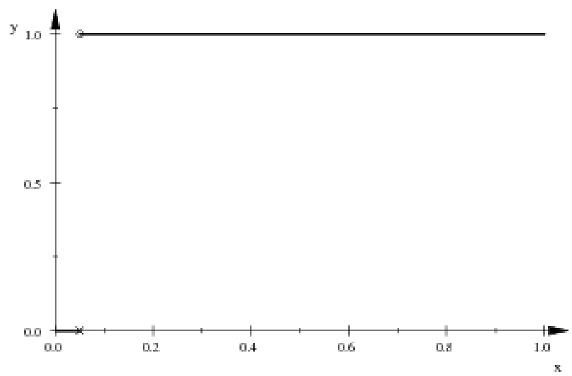


Abbildung 34: Gütefunktion eines idealen (einseitigen) Tests

11.3 Noch ein Beispiel

Johanns Frau Ute arbeitet in einer Manufaktur für Spielwürfel. Jeder Würfel wird getestet, ob die Wahrscheinlichkeit für eine Sechs auch wirklich exakt $\frac{1}{6}$ ist, bevor er zum Kunden geht. Dazu würfelt sie 120–mal und zählt die Sechsen. Sind es weniger als 15 oder mehr als 25, wird der Würfel nicht verkauft.

Hier ist die betrachtete Zufallsgröße $B(120, p)$–verteilt, dabei ist p die Wahrscheinlichkeit für eine Sechs. Der kritische Bereich ist

$$W_1 = \{1, 2, 3, \dots, 14\} \cup \{26, 27, 28, \dots, 120\} \quad ,$$

und man nennt diesen Test einen **zweiseitigen** Test; Johanns Test war dementsprechend ein **einseitiger** Test. Die Gütefunktion hat die Vorschrift

$$p \mapsto P(W_1) = 1 - \sum_{j=15}^{25} \binom{120}{j} p^j (1-p)^{120-j} \quad ,$$

ihr Graph ist in Abbildung 35 zu sehen.

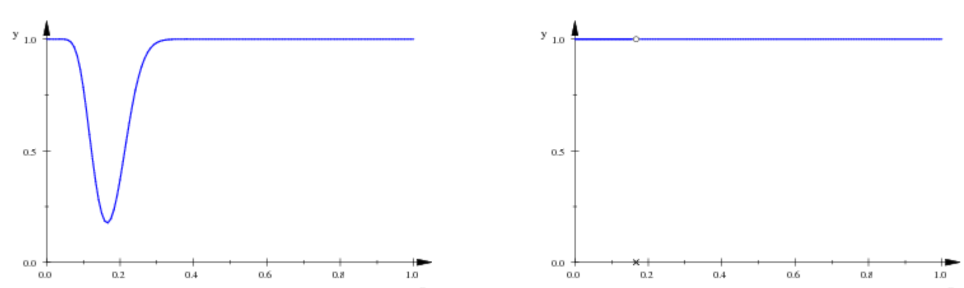


Abbildung 35: Gütefunktion eines realen (links) und eines idealen zweiseitigen Tests

Der Test ist offensichtlich völlig nutzlos. Beinahe jeder fünfte ideale Würfel wird ausgesondert, und nur sehr schlechte Würfel werden einigermaßen zuverlässig erkannt.

Aufgabe. Der Test der Würfel soll künftig so erfolgen, dass ein Automat 1200–mal würfelt und die Anzahl der Sechsen feststellt. Ziel ist, dass ein idealer Würfel nur noch mit maximal 2% Wahrscheinlichkeit aussortiert wird.

1. Formuliere eine Entscheidungsregel, die das leistet.

2. Zeichne die Gütefunktion zu deiner Entscheidungsregel.

3. Es sollen keine Würfel in den Handel kommen, bei denen die Wahrscheinlichkeit für eine Sechs um $\frac{1}{100}$ oder noch mehr zu klein ist, weil das zu Sammelklagen amerikanischer Eltern führen könnte. Wie groß ist das Risiko der Firma, dass ein solcher Würfel durch die Kontrolle kommt?

11.4 Nullhypothese und Alternativhypothese

Ein (ausgedachtes) Beispiel: Zum Kummer der Branche lag der Anteil der Deutschen, die meinten, dass Investmentbanken gut für Wirtschaft und Gesellschaft sind, bei maximal 10%. Deshalb hat die Branche mit erheblichem Mitteleinsatz versucht, ihr Image zu verbessern, etwa durch Sponsoring, Präsentation der Wetterkarte im Fernsehen, Pressearbeit und ähnliche Dinge. Der Koordinator dieser Anstrenungen ist davon überzeugt, dass der Anteil p der Deutschen, die gut über Investmentbanken denken, nun über 10% liegt. Ein Test soll Klarheit bringen.

In Hypothesentestchinesisch ist die Aussage $p > 10\%$, deren Gültigkeit gezeigt werden soll, die **Alternativhypothese**

$$H_1 : p > 10\% \quad .$$

Man bildet zu H_1 die **Nullhypothese**

$$H_0 : p \leq 10\% \quad .$$

Ziel ist, die Nullhypothese mit Hilfe einer Stichprobe abzulehnen; dann ist H_1 etabliert. Auf den ersten Blick scheint das etwas verquer zu sein; wenn du den Mechanismus einige Male angewandt hast, leuchtet er dir aber hoffentlich ein.

Wie in der Stochastik üblich und nötig, wird nun eine Irrtumswahrscheinlichkeit α festgelegt, die Nullhypothese H_0 zu Unrecht abzulehnen. Will man dieses Risiko nicht eingehen, wird man die Nullhypothese nie ablehnen können, und dann kann man sich den ganzen Kram sparen. Häufig wählt man $\alpha = 5\%$, und das sei auch hier der Fall. Umfragen kosten Geld; nach einem Blick in die Portokasse wurde als Stichprobenumfang $n = 600$ festgelegt. Aus diesen Daten ergibt sich nun die Entscheidungsregel des Tests: Man wird H_0 verwerfen, wenn die Anzahl X der Befragten, die Investmentbanken für gut und nützlich halten, mindestens einen noch zu bestimmenden Wert g annimmt. Natürlich soll g möglichst klein sein, man will H_0 ja ablehnen. Aber die Wahrscheinlichkeit, H_0 zu Unrecht abzulehnen, darf höchstens α sein:

$$P_{H_0}(X \geq g) \leq \alpha$$

Nun ist $P(X \geq k)$ um so größer, je größer p ist. Man muss g also so bestimmen, dass

$$P_{p=0.1}(X \geq g) \leq \alpha$$

ist. Damit kennen wir uns aus. Der Ansatz

$$P_{p=0.1}(X \geq g) = 1 - P_{p=0.1}(X < g) \approx 1 - \Phi\left(\frac{g - \frac{1}{2} - 60}{\sqrt{600 \cdot 0.1 \cdot 0.9}}\right) \leq \alpha$$

führt zu

$$g \geq 60.5 + \Phi^{-1}(0.95) \cdot \sqrt{54} \approx 72.6 \quad ,$$

und das ergibt folgende Entscheidungsregel: Es wird H_0 abgelehnt, wenn X einen Wert im kritischen Bereich

$$W_1 = \{73, 74, 75, \ldots, 600\}$$

annimmt. Wenn dies, wie erhofft, passiert, kann der Koordinator stolz verkünden, dass der Test ein signifikantes Ergebnis erbracht hat und die Behauptung mit der Sicherheitswahrscheinlichkeit 95% (bzw. der Irrtumswahrscheinlichkeit 5% bzw. auf dem Signifikanzniveau 5%) nachgewiesen ist.

Die **Bedeutung dieser Formulierung** ist dir hoffentlich klar: Wie beim Konfidenzintervall ist die Irrtumswahrscheinlichkeit eine **Kenngröße des Verfahrens** und nicht des konkreten Testergebnisses. Wenn das Institut sehr viele Tests mit der Irrtumswahrscheinlichkeit 5% durchführt[8], wird die Nullhypothese in etwa 5% der Fälle, in denen sie richtig ist, dennoch abgelehnt werden.

Und wenn X einen Wert in

$$W_0 := \{0, 1, 2, \ldots, 72\}$$

annimmt? Dann wird H_0 eben nicht verworfen, was aber nicht heißt, dass damit die Gültigkeit von H_0 nachgewiesen wäre. In diesem Fall hat der Test einfach nichts gebracht. Damit du siehst, wie vertrackt die Situation ist, spinne ich den Faden noch etwas weiter: Man hätte ja auch $\alpha = 1\%$ als Irrtumswahrscheinlichkeit nehmen können; das hätte zu

$$g \geq 60.5 + \Phi^{-1}(0.99) \cdot \sqrt{54} \approx 77.6$$

geführt, also zu $g = 78$ als kleinster Zahl im kritischen Bereich. Das Umfrageergebnis 75 führt bei $\alpha = 5\%$ zur Verwerfung von H_0 und folglich zur Annahme von $H_1 : p > 10\%$. Bei $\alpha = 1\%$ jedoch wird $H_0 : p \leq 10\%$ beibehalten. Kurios, nicht?

Zum Schluss findest du hier noch eine Tabelle, in der die Fehler erster und zweiter Art im Zusammenhang mit den Namen der Hypothesen aufgelistet sind.

Die Hypothese H_0	wird abgelehnt	wird nicht abgelehnt
stimmt	Fehler erster Art	richtige Entscheidung
stimmt nicht	richtige Entscheidung	Fehler zweiter Art

Die Wahrscheinlichkeit für einen Fehler erster Art darf eine vorgegebene Schranke α nicht überschreiten. Die Wahrscheinlichkeit für einen Fehler zweiter Art wird gewöhnlich mit β bezeichnet.

[8] und diese Tests unabhängig von einander sind

12 Poissonverteilte Zufallsgrößen

12.1 Vorbereitung: Über $\left(1+\frac{1}{n}\right)^n$

Wenn man in einer Zeiteinheit 3% Zinsen bekommt, ist der Wachstumsfaktor 1.03. Bekommt man sogar 100% Zinsen, ist der Wachstumsfaktor 2. Wenn man für den halben Zeitraum nur 50% Zinsen bekommt, steht man sich wegen des Zinseszinseffektes natürlich besser. Der Wachstumsfaktor für den gesamten Zeitraum ist dann nämlich

$$\left(1+\frac{1}{2}\right)^2 = \frac{9}{4} = 2.25 \quad .$$

Noch besser steht man sich, wenn man für $\frac{1}{3}$ des Zeitraums auch nur $\frac{1}{3}$ von 100% Zinsen bekommt; das ergibt den Wachstumsfaktor

$$\left(1+\frac{1}{3}\right)^3 = \frac{64}{27} \approx 2.37 \quad .$$

Fährt man in dieser Weise fort, erhält man für $\frac{1}{n}$ des Zeitraums $\frac{1}{n}$ von 100% Zinsen, und das macht den Wachstumsfaktor

$$\left(1+\frac{1}{n}\right)^n \quad .$$

Was wird daraus für n gegen Unendlich? Eine theoretische Antwort wirst du nicht finden, aber vielleicht hilft dir dein Taschenrechner (oder MuPAD).

Es gilt der folgende bemerkenswerte Satz, den wir leider nicht beweisen können:

21 Satz

$$\left(1+\frac{1}{n}\right)^n \xrightarrow{n\to\infty} e$$

Wenn man diesen Satz erst einmal hat, findet man leicht weitere erstaunliche Resultate:

$$\left(1-\frac{1}{n}\right)^n = \left(\frac{n-1}{n}\right)^n = \frac{1}{\left(\frac{n}{n-1}\right)^n} = \frac{1}{\left(1+\frac{1}{n-1}\right)^n} \xrightarrow{n\to\infty} \frac{1}{e} = e^{-1} \quad ,$$

denn für jedes feste $k \in \mathbb{Z}$ gilt

$$\left(1+\frac{1}{n}\right)^{n+k} = \left(1+\frac{1}{n}\right)^n \cdot \left(1+\frac{1}{n}\right)^k \xrightarrow{n\to\infty} e\cdot 1 = e \quad .$$

Und für $x > 0$ erhalten wir

$$\left(1+\frac{x}{n}\right)^n = \left(1+\frac{1}{\frac{n}{x}}\right)^n = \left(\left(1+\frac{1}{\frac{n}{x}}\right)^{\frac{n}{x}}\right)^x \xrightarrow{n\to\infty} e^x \quad .$$

Die Gleichung gilt auch für $x < 0$, man muss dann nur etwas anders umformen.

12.2 Typisches Beispiel

Vielleicht kennst du Lackdraht: dünner Kupferdraht wird durch einen Behälter mit Klarlack gezogen. Wenn die Lackschicht getrocknet ist, hat man einen isolierten Draht. Der Draht, den man nimmt, um Spulen zu wickeln, ist auf diese Weise isoliert.

Wo etwas gemacht wird, treten Fehler auf; bei der Herstellung von Lackdraht sind das schadhafte Stellen in der Isolierschicht. Sagen wir, bei einem Kilometer Lackdraht sind das im Mittel drei Fehler. Wir betrachten ein Stück der Länge 100 Meter. Mit welcher Wahrscheinlichkeit ist das Stück fehlerfrei?

Die Anzahl der schadhaften Stellen auf diesem Stück ist eine Zufallsgröße, wir bezeichnen sie mit X. Den Erwartungswert von X kennen wir schon, er ist $\frac{3}{10}$. In unserem Kontext bezeichnet man diesen Wert mit λ. Um die Wahrscheinlichkeitsverteilung von X zu bekommen, basteln wir uns ein Modell, und zwar teilen wir das Stück Draht in n Teile gleicher Länge auf. Auf ein Teilstück kommen dann im Mittel $\frac{\lambda}{n}$ Fehler. Wenn n sehr groß ist, können wir davon ausgehen, dass kein Teilstück zwei oder mehr Fehler enthält, und dann brauchen wir nur noch zu fragen, ob das betrachtete Teilstück fehlerhaft ist oder nicht. Das ist für jedes Teilstück ein Bernoulliversuch, und wenn wir alle Teilstücke der Reihe nach anschauen, haben wir es mit einer Bernoullikette der Länge n zu tun. Damit der richtige Erwartungswert λ herauskommt, nehmen wir als Erfolgswahrscheinlichkeit $p = \frac{\lambda}{n}$. Die Wahrscheinlichkeit für k Treffer auf dem Drahtstück ist dann

$$\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \quad ,$$

das weißt du ja. Wenn wir an dem Ausdruck etwas herumschrauben, erhalten wir daraus

$$\begin{aligned}\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} &= \frac{\lambda^k}{k!}\cdot\frac{n(n-1)(n-2)\cdot\dots\cdot(n-k+1)}{n^k}\cdot\left(1-\frac{\lambda}{n}\right)^{n-k}\\ &= \frac{\lambda^k}{k!}\cdot\frac{n}{n}\cdot\frac{n-1}{n}\cdot\frac{n-2}{n}\cdot\dots\cdot\frac{n-k+1}{n}\cdot\left(1-\frac{\lambda}{n}\right)^{n-k}\\ &\xrightarrow{n\to\infty} \frac{\lambda^k}{k!}\cdot 1\cdot 1\cdot 1\cdot\dots\cdot 1\cdot e^{-\lambda} = \frac{\lambda^k}{k!}e^{-\lambda} \quad .\end{aligned}$$

Und damit haben wir es. Unser Modell liefert für die betrachtete Zufallsgröße X die Wahrscheinlichkeitsverteilung

$$P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda} \quad \text{für } k = 0, 1, 2, \dots . \tag{34}$$

Eine Zufallsgröße mit dieser Wahrscheinlichkeitsverteilung heißt **poissonverteilt**.

Beispiel. Bei leichtem Regen hält jeder Schüler des Kurses einen Augenblick die flache Hand aus dem Fenster. Im Mittel habe jede Hand drei Regentropfen abbekommen. Mit welcher Wahrscheinlichkeit fielen auf Karolines Hand $0, 1, 2, 3, \dots$ Regentropfen?

In unserem dicken grünen Stochastikbuch stehen noch einige historische Beispiele. Spätestens nach denen haben wir das Ende unserer Reise in die Mathematik des Zufalls erreicht. War doch ganz nett, oder?