

# Stochastik

LK Mathematik Abitur 2014 (B. Waldmüller)

8. Juli 2013

# Inhaltsverzeichnis

<b>1</b>	<b>Das empirische Gesetz der großen Zahl</b>	<b>1</b>
<b>2</b>	<b>Kleiner Sprachführer</b>	<b>3</b>
<b>3</b>	<b>Wahrscheinlichkeitsfunktion und Zählverfahren</b>	<b>4</b>
<b>4</b>	<b>Unabhängigkeit und bedingte Wahrscheinlichkeit</b>	<b>6</b>
<b>5</b>	<b>Zufallsgrößen</b>	<b>8</b>
5.1	Die von $X$ gestiftete Zerlegung von $\Omega$ . . . . .	8
5.2	Der Erwartungswert $E(X)$ einer Zufallsgröße $X$ . . . . .	9
5.3	Rechenregeln für Erwartungswerte . . . . .	9
5.4	Standardbeispiele . . . . .	10
5.5	Die Varianz einer Zufallsgröße $X$ . . . . .	11
5.6	Das Histogramm einer binomialverteilten Zufallsgröße . . . . .	13
5.7	Mehr Regeln für Erwartungswert und Varianz . . . . .	13
<b>6</b>	<b>Klausur Nr. 3 am 21. März 2013</b>	<b>16</b>
<b>7</b>	<b>Die lokale Näherungsformel von de Moivre und Laplace</b>	<b>18</b>
7.1	Einstieg . . . . .	18
7.2	Standardisierung einer Zufallsgröße . . . . .	18
7.3	Die Gaußsche Glockenkurve . . . . .	20
7.4	Lokale Näherungsformel von de Moivre und Laplace . . . . .	21
7.5	Die Verteilungsfunktion einer Zufallsgröße . . . . .	21
7.6	Integrale Näherungsformel von de Moivre und Laplace . . . . .	23
<b>8</b>	<b>Stetige Zufallsgrößen</b>	<b>24</b>
8.1	Beispiele und Grundbegriffe . . . . .	24
8.2	Über Erwartungswert und Varianz stetiger Zufallsgrößen . . . . .	25
8.3	Regeln für Erwartungswert und Varianz . . . . .	26
8.4	Die Normalverteilung . . . . .	26
8.5	Eine Anwendungsaufgabe . . . . .	28
<b>9</b>	<b>Zwei Grenzwertsätze</b>	<b>29</b>
9.1	Bernoullis Schwaches Gesetz der großen Zahl . . . . .	29
9.2	Der Zentrale Grenzwertsatz . . . . .	30
9.3	Ein Anwendungsbeispiel . . . . .	32
<b>10</b>	<b>Klausur Nr. 4 am 16. Mai 2013</b>	<b>34</b>
<b>11</b>	<b>Beurteilende Statistik</b>	<b>36</b>
11.1	Einstiegsbeispiel . . . . .	36
11.2	Zweites Beispiel . . . . .	37
11.3	Der Apparat eines Hypothesentests . . . . .	39
<b>12</b>	<b>Klausur Nr. 4 (Nachschreibausgabe)</b>	<b>40</b>
<b>13</b>	<b>Schätzen einer unbekanntem Wahrscheinlichkeit</b>	<b>42</b>
13.1	Das Konfidenzintervall . . . . .	42
13.2	Berechnung des Konfidenzintervalls . . . . .	45
13.3	Was bedeutet die Irrtumswahrscheinlichkeit $\alpha$ ? . . . . .	46

# 1 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  $n$  Würfeln. 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  $n$  Würfeln 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  $y = \frac{1}{6}$ .

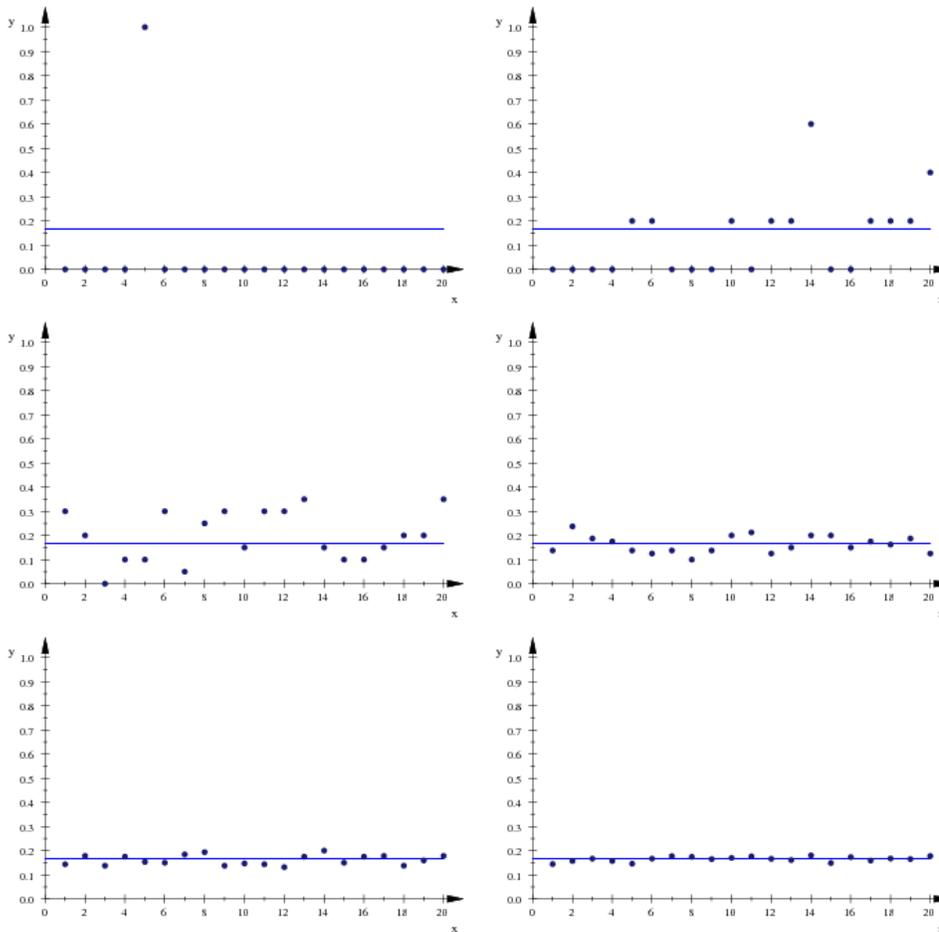


Abbildung 1: Relative Häufigkeit der 5 bei  $n$  Würfeln für  $n = 1, 5, 20, 80, 320, 1280$

Wie du siehst, „stabilisiert sich die relative Häufigkeit der 5 bei  $n$  Würfeln“ 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. Würfeln. 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 Würfeln überhaupt keine 5 geworfen wurde, und führe die restlichen 49 Wurfserien ordentlich aus, dann gehe ich von 100 Fünfen bei den ersten 100 Würfeln 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.

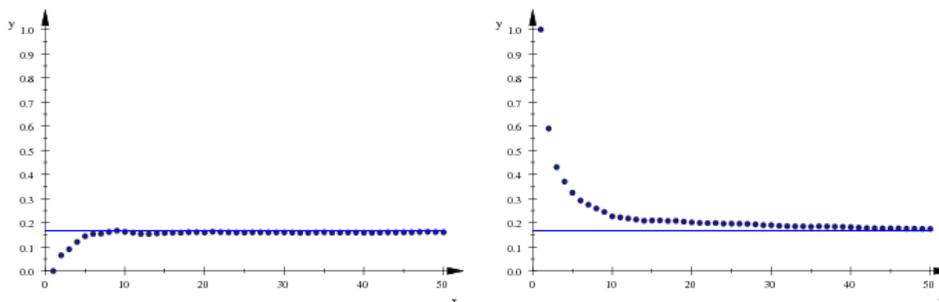


Abbildung 2: Relative Häufigkeit der 5 nach  $k \cdot 100$  Würfeln. Das Ergebnis der ersten 100 Würfe ist gefälscht, und zwar links auf 0 und rechts auf 100 Fünfen.

## 2 Kleiner Sprachführer

In der abstrakten Wahrscheinlichkeitstheorie kommt man mit sehr wenigen Begriffen aus, und man versucht nicht einmal zu beschreiben, was Wahrscheinlichkeit ihrem Wesen nach ist; da gibt es dann eine Funktion  $P$ , die auf einem System von Mengen definiert ist, die eine Handvoll Eigenschaften hat. Dieser moderne Ansatz bringt den Schlag durch den gordischen Knoten; der alte Euklid mühte sich zu sagen, was ein Punkt seiner Geometrie sein soll, aber das führt im Grunde zu nichts. Der moderne Geometer sagt nicht, was ein Punkt ist, sondern er sagt, er denke sich ein System von Dingen, die er Punkte, Geraden und Ebenen nennt. Jeder kann sich darunter vorstellen, was er mag; was zählt, sind ausformulierte Beziehungen zwischen diesen Objekten, zum Beispiel, dass zwei Punkte genau eine Gerade festlegen. Dieser axiomatische Ansatz entstand in reiner Form Anfang des 20. Jahrhunderts, der berühmte David Hilbert in Göttingen war ein wichtiger Wegbereiter. Man darf die Beliebigkeit der Interpretation der Begriffe aber nicht zu buchstäblich nehmen; wer wirklich schöpferisch Geometrie treiben will, macht auch Zeichnungen und hat Bilder im Kopf.

In den dreißiger Jahren des 20. Jahrhunderts schuf Andrej Kolmogorov ein Axiomensystem, auf dem die Wahrscheinlichkeitstheorie als mathematische Theorie aufgebaut ist. Obwohl es sich um ein sehr einfaches System handelt, ist der Weg, aus Axiomen Sätze zu beweisen und so eine Theorie aufzubauen, für uns nicht gangbar; dir ginge sehr bald die Puste aus. Wir brauchen wenigstens die idealen Zufallsversuche, die wir in unseren Köpfen durchführen, wie wir ideale Punkte und Geraden brauchen, um Geometrie zu treiben. Unter einem **Zufallsversuch** stellst du dir am besten einen Apparat vor, der auf Knopfdruck ein **Ergebnis**  $\omega$  ausspuckt.

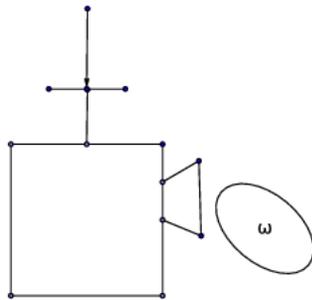


Abbildung 3: Zufallsversuch: Apparat, den man durch Knopfdruck auslösen kann

Die Menge aller möglichen Ergebnisse ist der **Ergebnisraum**  $\Omega$ . Bei jeder Durchführung liefert der Zufallsversuch genau ein  $\omega \in \Omega$ , das muss garantiert sein. **Ereignisse** sind gewisse Teilmengen von  $\Omega$ . Wir sagen, das Ereignis  $A \subseteq \Omega$  sei (bei einer Durchführung des Versuchs) eingetreten, wenn ein  $\omega \in A$  ausgespuckt<sup>1</sup> wurde. Dies geschieht mit einer gewissen **Wahrscheinlichkeit**  $P(A)$ , einer Zahl, die zwischen 0 und 1 liegt. Es ist stets  $P(\Omega) = 1$  und  $P(\{\}) = 0$ .

Nehmen wir ein Beispiel. Eine (ideale) Münze werde zweimal geworfen. Dann können wir mit

$$\Omega := \{KK, KW, WK, WW\}$$

arbeiten; dabei bedeutet  $KW$ , dass erst Kopf und dann Wappen erschien. Aus Symmetriegründen hat jedes dieser vier Ergebnisse die gleiche Chance, ausgelost zu

<sup>1</sup>Eine seriöse Vokabel dafür wäre: ausgelost.

werden, deshalb geben wir jedem der vier **Elementarereignisse**

$$\{KK\}, \{KW\}, \{WK\}, \{WW\}$$

die Wahrscheinlichkeit  $\frac{1}{4}$ :

$$P(\{KK\}) = P(\{KW\}) = P(\{WK\}) = P(\{WW\}) = \frac{1}{4}$$

Wenn wir uns nur dafür interessieren, wie oft Kopf geworfen wurde, haben wir es mit einer **Zufallsgröße**  $X$  zu tun. Das ist eine Abbildung (oder Funktion)  $\Omega \rightarrow \mathbb{R}$ . Hier ist

$$X(KK) = 2, \quad X(KW) = X(WK) = 1 \quad \text{und} \quad X(WW) = 0 \quad .$$

Die Wahrscheinlichkeit  $P(X = 1)$ , dass  $X$  den Wert 1 annimmt, ist gegeben durch

$$P(X = 1) := P(X^{-1}(1)) = P(\{KW, WK\}) \quad ,$$

dabei ist

$$X^{-1}(1) := \{\omega \in \Omega \mid X(\omega) = 1\} \quad .$$

Die **Wahrscheinlichkeitsfunktion** von  $X$  schreibt man zum Beispiel in Form einer Wertetabelle hin:

$k$	0	1	2
$P(X = k)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

– So, jetzt sind erst einmal ein paar Sprachübungen fällig, aber mit dem Begriffsapparat können wir schon etwas anstellen.

### 3 Wahrscheinlichkeitsfunktion und Zählverfahren

Eine Wahrscheinlichkeit  $P(A)$  wollten wir von Ereignissen  $A$  bilden, also von Teilmengen von  $\Omega$ . Fürs erste befassen wir uns nur mit endlichen  $\Omega$ , in Zeichen: es ist  $|\Omega| < \infty$ . Dann können wir die Sache so ins Lot bringen, dass wir die Gesamtwahrscheinlichkeit 1 unter den  $\omega \in \Omega$  aufteilen; jedes  $\omega \in \Omega$  bekommt seinen Anteil<sup>2</sup>  $p_\omega$ . Dabei können wir im Prinzip ganz nach Belieben vorgehen; die Zahlen  $p_\omega$  müssen nur  $\geq 0$  sein und ihre Summe muss 1 ergeben:

$$\sum_{\omega \in \Omega} p_\omega = 1 \tag{1}$$

Für jedes  $A \subseteq \Omega$  ist dann die Wahrscheinlichkeit festgelegt:

$$P(A) = \sum_{\omega \in A} p_\omega \tag{2}$$

Vermutlich ist dir das nicht ganz geheuer, dass du die  $p_\omega$  beliebig festsetzen kannst. Gewöhnlich geht das auch nicht, man hat ja einen Zufallsversuch im Kopf, und die  $p_\omega$  ergeben sich aus den Eigenheiten dieses Versuchs. Im Idealfall ist der

---

<sup>2</sup>Schau, jedes  $\omega$  bekommt sein  $p_\omega \in \mathbb{R}$ . Was ist das? Klar, eine Funktion  $p : \omega \mapsto p_\omega$  mit Definitionsbereich  $\Omega$  und Wertebereich  $\mathbb{R}$ , also eine reellwertige Funktion. Die Bezeichnungen sind nicht einheitlich, ich nenne sie jetzt Wahrscheinlichkeitsfunktion, im Buch heißt sie Wahrscheinlichkeitsverteilung. Darüber können wir aber noch einmal reden.

Versuch und das  $\Omega$  so konstruiert, dass alle  $p_\omega$  gleich sind. In dem Fall ist dann für jedes  $A \subseteq \Omega$

$$P(A) = \frac{|A|}{|\Omega|} , \quad (3)$$

und wir sprechen von einem Laplaceversuch. Um die Wahrscheinlichkeit zu bestimmen, muss man „nur noch“ die Anzahl  $|A|$  der Elemente von  $A$  zählen, und dafür lernst du jetzt ein paar Hilfsmittel. Sie gehören zur Kombinatorik, der mathematischen Theorie des geschickten Zählens.

Es sei  $M$  eine Menge mit  $m$  Elementen, es gelte also  $|M| = m$ . Diese Menge  $M$  stellt den Vorrat an Ziffern oder Buchstaben dar. Dann gelten die folgenden Aussagen:

1. Es gibt  $m^n$  Wörter der Länge  $n$  mit Buchstaben aus  $M$ . Bei Wörtern kommt es auf die Reihenfolge der Buchstaben an, und jeder Buchstabe darf beliebig oft verwandt werden.
2. Es gibt  $m \cdot (m - 1) \cdot (m - 2) \cdot \dots \cdot (m - k + 1)$  Wörter der Länge  $k$ , bei denen alle Buchstaben verschieden sind.
3. Es gibt

$$\binom{m}{k} := \frac{m \cdot (m - 1) \cdot (m - 2) \cdot \dots \cdot (m - k + 1)}{1 \cdot 2 \cdot 3 \cdot \dots \cdot k} = \frac{m!}{k!(m - k)!}$$

$k$ -elementige Teilmengen von  $M$ . Man sagt auch kurz:  $k$ -Teilmengen einer  $m$ -Menge. Dabei ist  $k!$  eine Abkürzung für  $1 \cdot 2 \cdot 3 \cdot \dots \cdot k$ , und man setzt  $0! := 1$ .

Zur Verdeutlichung ein paar Beispiele:  $M$  ist der Kurs, dann ist  $|M| = 24$ , und die Elemente seid ihr. Sagen wir, an jedem der nächsten fünf Tage muss einer zum Sekretariat, nach Anweisungen für den Kurs fragen. Es muss eine Liste erstellt werden, auf der für jeden Tag der verantwortliche Schüler notiert ist. Es ist möglich, dass ein Schüler mehrfach dran ist, im Extremfall muss ein einziger Schüler fünfmal gehen. Auf die Reihenfolge kommt es auch an, es geht also um Wörter der Länge fünf mit Wiederholung. Davon gibt es  $24^5$ .

Nächstes Beispiel: Zum Schülerrat sollen Kurssprecher und Stellvertreter gehen. Keiner kann beide Ämter ausfüllen, es geht also um die Anzahl der Worte der Länge zwei ohne Wiederholung. Davon gibt es  $24 \cdot 23$  Stück.

Letztes Beispiel: Es sollen acht Schüler zur Reinigung der Freizeitzone abgestellt werden. Da die Reihenfolge keine Rolle spielt, geht es um die Anzahl der 8-Teilmengen eurer 24-Menge, und davon gibt es  $\binom{24}{8}$  Stück. Dein Taschenrechner rechnet dir die Zahl aus: Tippe  $24 nCr 8$  ein.

## 4 Unabhängigkeit und bedingte Wahrscheinlichkeit

In unserem Kurs sitzen (bzw. liegen, Tom) fünf Schüler mit dem Rücken zum Fenster. Wählt man also einen Schüler des Kurses zufällig, ist die Wahrscheinlichkeit  $\frac{5}{24}$ , dass ein Schüler ausgelost wird, der mit dem Rücken zum Fenster sitzt. Ferner hatten in der letzten Klausur sieben Schüler eine eins, und drei von denen sitzen mit dem Rücken zum Fenster.

Wir bezeichnen die Menge der Schüler, die mit dem Rücken zum Fenster sitzen, mit  $A$ , und die Menge derer, die in der letzten Klausur eine eins hatten, mit  $B$ . Dann haben wir zwei Teilmengen  $A, B$  der Menge  $\Omega$  aller Schüler des Kurses. Überlege dir ruhig kurz, was die Mengen  $A \cup B$ ,  $A \cap B$ ,  $\bar{A}$  und  $\bar{A} \cap B$  bedeuten und was die Wahrscheinlichkeiten dieser Ereignisse sind.

Eigentlich ist es müßig, welche von euch auszulosen. Dennoch kann es von Interesse sein, nach Anteilen zu fragen. Zum Beispiel ist die Wahrscheinlichkeit  $P(B)$ , dass ein zufällig gewählter Schüler in der letzten Klausur eine eins hatte,

$$P(B) = \frac{|B|}{|\Omega|} = \frac{7}{24} .$$

Wählt man aber einen der Schüler der Menge  $A$  zufällig und fragt nach der Wahrscheinlichkeit, dass dieser eine eins in der letzten Klausur hatte, ergibt sich dafür

$$P(B|A) = \frac{3}{5} ,$$

dies ist die **bedingte Wahrscheinlichkeit**  $P(B|A)$ , dass  $B$  eintritt unter der Bedingung, dass  $A$  eingetreten ist<sup>3</sup>; sie weicht hier von der nackten Wahrscheinlichkeit  $P(B) = \frac{7}{24}$  ab. In diesem Fall sagt man, die Ereignisse  $A$  und  $B$  seien (stochastisch) **abhängig**. Viele ignorieren den Zusatz „stochastisch“, aber sie tun nicht gut daran. Stochastische Abhängigkeit bedeutet keineswegs automatisch, dass ein kausaler Zusammenhang besteht. Natürlich können wir in die pädagogische Diskussion mit Verweis auf unseren empirischen Befund die These einbringen, dass es schlau macht, am Fenster zu sitzen; das wird dort aber erst akzeptiert werden, wenn man eine größere statistische Erhebung durchgeführt hat – hoffe ich zumindest. In jedem Fall hat man häufig nicht mehr als eben einen solchen statistischen Befund.

So, ich will nicht abschweifen, sondern dir ordentlich erklären, was stochastische Unabhängigkeit bedeutet: Wenn du die Wahrscheinlichkeit  $P(A \cap B)$  berechnen willst, kannst du erst danach fragen, ob  $A$  eingetreten ist. Dies ist mit der Wahrscheinlichkeit  $P(A)$  der Fall. Wenn nun ein  $\omega \in A$  ausgelost wurde, kannst du als nächstes fragen, ob das ausgeloste  $\omega$  auch noch zu  $B$  gehört. Die Wahrscheinlichkeit dafür ist die bedingte Wahrscheinlichkeit  $P(B|A)$ . Insgesamt erhältst du<sup>4</sup>

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

Analog gilt natürlich  $P(A \cap B) = P(B) \cdot P(A|B)$ , insgesamt folglich

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B) . \quad (4)$$

Falls  $B$  von  $A$  unabhängig und  $P(B) \neq 0$  ist, folgt aus dieser Gleichung nach Einsetzen und Division durch  $P(B)$ , dass  $P(A|B) = P(A)$  ist, dass also auch  $A$  von  $B$  unabhängig ist. Man definiert Unabhängigkeit von vornherein so:

<sup>3</sup>Diese Formulierung ist etwas unglücklich, es muss überhaupt keine zeitliche Reihenfolge vorliegen.

<sup>4</sup>Vielleicht hilft es dir, wenn du einen Baum zeichnest.

### 1 Definition

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

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

ist.

Über dieser technischen Definition solltest du aber nie die anschauliche Bedeutung der Unabhängigkeit vergessen, die darzustellen ich anfangs versucht habe.

Zum Schluss trage ich eine ordentliche Definition der bedingten Wahrscheinlichkeit nach:

### 2 Definition

Es sei  $A \subseteq \Omega$  ein Ereignis mit  $P(A) \neq 0$ . Dann definiert man für jedes Ereignis  $B \subseteq \Omega$  die **bedingte Wahrscheinlichkeit**  $P(B|A)$  durch

$$P(B|A) := \frac{P(A \cap B)}{P(A)} .$$

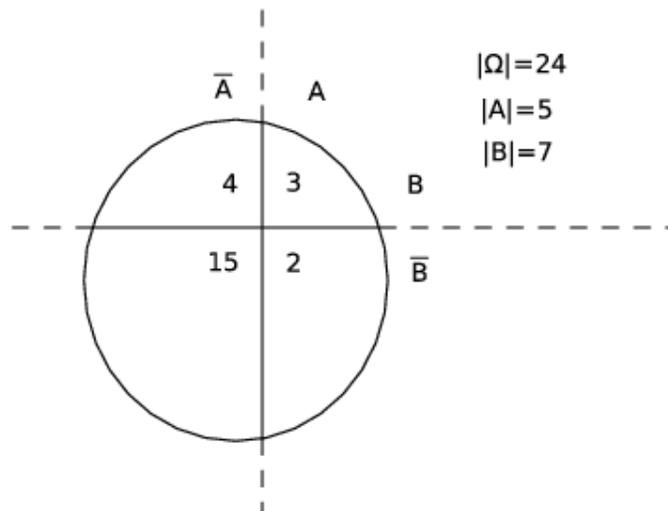


Abbildung 4: Mengendiagramm zum Beispiel, zur Vierfeldertafel ist kein weiter Weg mehr

## 5 Zufallsgrößen

### 5.1 Die von $X$ gestiftete Zerlegung von $\Omega$

Eine Zufallsgröße  $X$  (auf  $\Omega$ ) ist eine Funktion, die jedem  $\omega \in \Omega$  eine Zahl  $X(\omega)$  zuordnet; das ist dir schon kurz begegnet. Die Menge aller Werte, die  $X$  annehmen kann, ist

$$X(\Omega) := \{ X(\omega) \mid \omega \in \Omega \} .$$

Für jeden Wert  $k \in X(\Omega)$  ist

$$X^{-1}(k) := \{ \omega \in \Omega \mid X(\omega) = k \}$$

die Menge der Urbilder von  $k$ . Für die Wahrscheinlichkeit

$$P(X^{-1}(k)) ,$$

dass  $X$  bei Durchführung des Versuchs den Wert  $k$  annimmt, schreibt man kurz  $P(X = k)$ .

Schauen wir uns ein Beispiel an. Wir werfen einen Würfel, schauen uns die gewürfelte Zahl an, und kippen den Würfel dann über eine zufällig gewählte Kante um. Jedes Ergebnis des Versuchs ist ein Wort  $ab$ , dabei ist  $a$  eine Zahl von 1 bis 6 und  $b$  ist eine Zahl von 1 bis 6, aber weder  $a$  noch  $7 - a$ .

Wir können die Elemente des Ergebnisraums  $\Omega$  dieses Versuchs in folgender Weise sinnvoll anordnen:

	12	13	14	15	
21		23	24		26
31	32			35	36
41	42			45	46
51		53	54		56
	62	63	64	65	

Nun definieren wir uns Zufallsgrößen auf diesem  $\Omega$ :

$X$  : erste gewürfelte Zahl

$Y$  : zweite gewürfelte Zahl

$S$  : Summe der gewürfelten Zahlen

$Z$  : Zahl, auf der der Würfel beim ersten Wurf lag

Jede dieser Zufallsgrößen stiftet auf  $\Omega$  eine Struktur, genauer: eine Zerlegung in Teilmengen. Und zwar steckt man jeweils alle die  $\omega$  in eine Teilmenge, die von der Zufallsgröße auf denselben Wert abgebildet werden. In unserem Beispiel haben jeweils alle  $\omega$  einer Zeile denselben  $X$ -Wert, alle  $\omega$  einer Spalte denselben  $Y$ -Wert; es ist etwa

$$X^{-1}(2) = \{21, 23, 24, 26\} \quad \text{und}$$

$$Y^{-1}(4) = \{14, 24, 54, 64\} .$$

Es folgt

$$P(X = 2) = P(X^{-1}(2)) = \frac{4}{24} \quad \text{und} \quad P(Y = 4) = P(Y^{-1}(4)) = \frac{4}{24} ,$$

$$P(X = 2, Y = 4) = P(\{24\}) = \frac{1}{24} \quad \text{und} \quad P(X = 2 \mid Y = 4) = \frac{4}{24} ,$$

$$P(S = 3) = \frac{2}{24} \quad \text{und} \quad P(X = 2 \mid S \leq 6) = \frac{3}{24} .$$

## 5.2 Der Erwartungswert $E(X)$ einer Zufallsgröße $X$

Es sei  $X : \Omega \mapsto \mathbb{R}$  eine Zufallsgröße. Führen wir den zugrundeliegenden Zufallsversuch  $N$ -mal aus, erhalten wir  $N$  ausgeloste Werte

$$x_1, x_2, x_3, \dots, x_N$$

von  $X$ ; man nennt sie Realisierungen, aber das brauchst du dir nicht zu merken. Es ist ein empirischer Befund, dass sich der Mittelwert

$$\bar{x} := \frac{1}{N} \sum_{k=1}^N x_k \quad (5)$$

dieser Werte bei wachsendem  $N$  stabilisiert, sich also bei einem festen Wert einpendelt.

In der Theorie haben wir ja nur unsere ausgedachten  $\Omega$ ,  $P$  und  $X$ , und hier definieren wir uns einfach eine Größe  $E(X)$ , mit der wir das Verhalten realer Zufallsgrößen nachbilden wollen:

### 3 Definition

Es sei  $X : \Omega \mapsto \mathbb{R}$  eine Zufallsgröße. Dann ist der **Erwartungswert** von  $X$  definiert durch

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

**Beispiel.** Wir betrachten einen Zufallsversuch, der nur zwei Ergebnisse hat. Du kannst zum Beispiel einen Würfel werfen und schauen, ob du eine sechs geworfen hast – das nennst du Erfolg – oder nicht, das ist dann eben ein Misserfolg. Ein solcher Zufallsversuch, der nur zwei Ergebnisse hat, heißt **Bernoulliversuch**; für sein  $\Omega$  ist  $|\Omega| = 2$ . Die Wahrscheinlichkeit für Erfolg wird traditionell mit  $p$  bezeichnet, die für Misserfolg ist dann  $1 - p$ . Die Zufallsgröße

$$X : \text{Anzahl der Erfolge}$$

kann dann nur die Werte 0 und 1 annehmen. Berechne  $E(X)$ .

## 5.3 Rechenregeln für Erwartungswerte

Es seien  $X$  und  $Y$  Zufallsgrößen auf dem gleichen  $\Omega$ . Dann ist durch

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

eine neue Zufallsgröße  $S : \Omega \rightarrow \mathbb{R}$  definiert, die wir die Summe von  $X$  und  $Y$  nennen. Manchmal schreibt man das  $S$  nicht einmal hin und spricht einfach von der Summe  $X + Y$  der Zufallsgrößen  $X$  und  $Y$ . Dieses  $X + Y$  hat schöne Eigenschaften:

### 4 Lemma

Es seien  $X$  und  $Y$  Zufallsgrößen auf dem gleichen  $\Omega$ . Dann gilt

$$E(X + Y) = E(X) + E(Y) .$$

**Beweis.** Es ist

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

damit ist das Lemma bewiesen.  $\square$

### 5 Lemma

Es sei  $X : \Omega \rightarrow \mathbb{R}$  eine Zufallsgröße und  $a \in \mathbb{R}$ . Dann ist der Erwartungswert der durch

$$(aX)(\omega) := aX(\omega) \quad \text{für alle } \omega \in \Omega$$

definierten Zufallsgröße  $aX$  gegeben durch

$$E(aX) = a E(X) \quad .$$

Den **Beweis** kannst du selbst.

## 5.4 Standardbeispiele

Wir gehen zunächst davon aus, dass die betrachtete Zufallsgröße  $X$  nur endlich viele Werte annimmt:  $|X(\Omega)| < \infty$ .

Im einfachsten Fall haben alle Werte, die  $X$  annehmen kann, die gleiche Wahrscheinlichkeit. In diesem Fall ist

$$P(X = k) = \frac{1}{|X(\Omega)|} \quad \text{für alle } k \in X(\Omega),$$

und die Zufallsgröße  $X$  heißt **gleichverteilt** oder **uniform verteilt**. Zum Beispiel ist die gewürfelte Augenzahl beim einfachen Wurf eines Würfels gleichverteilt; es ist  $X(\Omega) = \{1, 2, 3, 4, 5, 6\}$ ,  $|X(\Omega)| = 6$  und  $P(X = k) = \frac{1}{6}$  für alle  $k \in X(\Omega)$ .

Von herausragender Bedeutung sind Zufallsgrößen der folgenden Bauart: Man wirft einen Würfel  $n$ -mal, und  $Y$  ist die Anzahl der gewürfelten Sechsen. Oder man zieht aus einer Urne mit weißen und roten Kugeln  $n$ -mal **und legt die gezogene Kugel jedesmal zurück**. Wenn  $p$  der Anteil der roten Kugeln ist und  $X$  die Anzahl der gezogenen roten Kugeln bedeutet<sup>5</sup>, ist

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

Mit Zufallsgrößen dieser Sorte musst du gut vertraut sein.

### 6 Definition

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

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

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

Wenn bei unserem Urnenbeispiel  $X_i$  die Anzahl der roten Kugeln bedeutet, die bei der  $i$ -ten Ziehung gezogen wurde, kann  $X_i$  jeweils nur die Werte 0 und 1 annehmen, und es ist

$$E(X_i) = 1 \cdot p + 0 \cdot (1-p) = p \quad .$$

Da

$$X = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i \quad (6)$$

ist, können wir nach unserer Summenregel für Erwartungswerte sofort  $E(X)$  angeben! Es gilt der folgende Satz:

<sup>5</sup>Die Anzahl  $Y$  der gewürfelten Sechsen ist der Sonderfall von  $X$  mit  $p = \frac{1}{6}$ .

### 7 Satz

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

$$E(X) = np \quad .$$

Zieht man aus einer Urne mit  $N$  Kugeln, von denen  $m$  rot sind,  $n$ -mal **ohne Zurücklegen**<sup>6</sup>, ist die Anzahl  $X$  der gezogenen roten Kugeln nicht mehr binomialverteilt. In diesem Fall gilt, wie wir uns an Beispielen überlegt haben,

$$P(X = k) = \frac{\binom{m}{k} \cdot \binom{N-m}{n-k}}{\binom{N}{n}} \quad \text{für } k = 0, 1, 2, \dots, n \quad ,$$

und  $X$  heißt **hypergeometrisch verteilt**. Wie wir uns – wieder mit Hilfe der Summenregel – überlegt haben, muss für hypergeometrisch verteiltes  $X$

$$E(X) = n \cdot \frac{m}{N}$$

gelten.

Zum Schluss noch ein komplizierteres Beispiel: Wir ziehen aus der Urne mit einem Anteil von  $p$  roten Kugeln so lange mit Zurücklegen, bis wir eine rote Kugel gezogen haben<sup>7</sup>. Es sei  $X$  die Anzahl der nötigen Ziehungen. Hier kann  $X$  nicht mehr nur endlich viele Werte annehmen, und es ist

$$P(X = k) = (1-p)^{k-1}p \quad \text{für } k = 1, 2, 3, \dots \quad .$$

Diese Zufallsgröße heißt **geometrisch verteilt**, aber das musst du nicht unbedingt wissen.

## 5.5 Die Varianz einer Zufallsgröße $X$

Wir kamen heute zu der Erkenntnis, dass es nicht reicht, den Erwartungswert einer Zufallsgröße zu kennen, sondern dass auch von Bedeutung ist, wie stark die Werte von  $X$  um den Erwartungswert streuen. Bei einer realen Zufallsgröße erwarten wir intuitiv, dass es bei starker Streuung sehr lange dauert, bis sich der empirische Mittelwert stabilisiert, und das ist auch in der Tat so.

Wir wünschen uns also ein Maß für die Stärke der Streuung der Werte einer Zufallsgröße  $X$ . So, wie wir den Erwartungswert abstrakt definiert haben, setzen wir auch ein Streuungsmaß fest:

### 8 Definition

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

$$\text{Var}(X) := \sum_{k \in X(\Omega)} (k - \mu)^2 P(X = k) \quad .$$

Die Varianz von  $X$  ist eine Zahl. Ist sie groß, streuen die Werte von  $X$  stark, ist sie klein, streuen sie wenig. Dabei ist sowohl die Größe der Abweichung eines Wertes  $k$  von  $\mu$  von Bedeutung als auch die Wahrscheinlichkeit, dass dieser Wert ausgelost

<sup>6</sup>Man kann der Urne auch  $n$  Kugeln auf einmal entnehmen, das läuft auf das Gleiche hinaus.

<sup>7</sup>Du kennst dieses Beispiel in der Form, dass jemand würfelt, bis er eine sechs geworfen hat.

wird. Den Beitrag eines Wertes  $k \in X(\Omega)$  zur Varianz verdeutlicht Abbildung 5: Du siehst die Merkmalsachse mit dem Erwartungswert  $\mu$  und einem Wert  $k \in X(\Omega)$  und das Quadrat mit der Kantenlänge  $|k - \mu|$ . Vom Inhalt dieses Quadrates geht der Anteil  $P(X = k)$  in die Varianz ein; das mag der Inhalt des markierten Teilstücks sein,  $P(X = k)$  wäre etwa  $\frac{1}{6}$ . Wenn es Werte gibt, die weit von  $\mu$  weg liegen und deren Wahrscheinlichkeit groß ist, machen diese die Varianz groß.

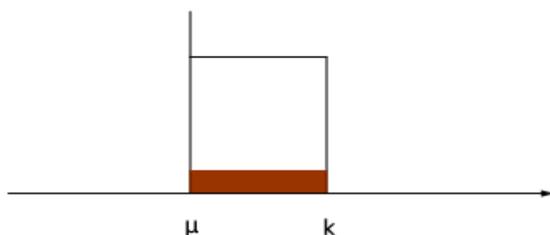


Abbildung 5: Beitrag eines  $k \in X(\Omega)$  zur Varianz  $\text{Var}(X)$

Man kann auch andere Streuungsmaße definieren. Dass ein Begriff gut gewählt ist, zeigt sich beim praktischen Gebrauch. Dass die Varianz etwas taugt, sollst du sogleich sehen.

Wir beginnen mit der Berechnungsvorschrift für die Varianz:

$$\text{Var}(X) = \sum_{k \in X(\Omega)} (k - \mu)^2 P(X = k) .$$

Stelle dir die Merkmalsachse gut vor: der Erwartungswert  $\mu$  und die Werte  $k \in X(\Omega)$  seien markiert. Wähle nun eine positive Zahl  $a$ . Die  $k$ -Werte mit  $|k - \mu| < a$  färben wir grün, die  $k$  mit  $|k - \mu| \geq a$  färben wir rot. Dann sind die  $k$  in der Nähe von  $\mu$  grün und die, die mindestens  $a$  von  $\mu$  entfernt sind, rot. Jedes  $k$  liefert einen Beitrag zur Varianz, und wir schreiben die Beiträge der verschieden gefärbten  $k$  getrennt auf:

$$\text{Var}(X) = \sum_{|k - \mu| < a} (k - \mu)^2 P(X = k) + \sum_{|k - \mu| \geq a} (k - \mu)^2 P(X = k)$$

Die erste Summe ist der Beitrag der grünen  $k$ , die zweite der Beitrag der roten. Nun kommt ein kühner Eingriff, Alexanders Durchschlagen des Gordischen Knotens vergleichbar. Die Summen enthalten nur positive Summanden; lassen wir die erste Summe weg, bleibt die rechte Seite gleich oder wird kleiner:

$$\text{Var}(X) \geq \sum_{|k - \mu| \geq a} (k - \mu)^2 P(X = k)$$

Letzter Schritt: Die Werte  $(k - \mu)^2$  der roten  $k$  sind mindestens gleich  $a^2$ . Ersetzen wir alle  $(k - \mu)^2$  in der Summe auf der rechten Seite durch  $a^2$ , bleibt die rechte Seite gleich oder wird höchstens kleiner. Es gilt also erst recht

$$\text{Var}(X) \geq \sum_{|k - \mu| \geq a} a^2 P(X = k) = a^2 \sum_{|k - \mu| \geq a} P(X = k) = a^2 P(|X - \mu| \geq a) .$$

Dividieren wir diese Ungleichung durch  $a^2$ , erhalten wir die berühmte **Ungleichung von Tschebyschew**:

$$P(|X - \mu| \geq a) \leq \frac{\text{Var}(X)}{a^2} \quad (7)$$

Sie gilt für jede Zufallsgröße  $X$ , die eine Varianz hat.

## 5.6 Das Histogramm einer binomialverteilten Zufallsgröße

Der Erwartungswert einer  $B(n, p)$ -verteilten Zufallsgröße  $X$  ist  $E(X) = np$ . Bei binomialverteilten Zufallsgrößen ist es tatsächlich auch so, wie man es gern hat: die Werte in der Nähe des Erwartungswertes haben die höchste Wahrscheinlichkeit. Dies sieht man, wenn man die Ungleichung

$$P(X = k) < P(X = k + 1)$$

ausschreibt, vereinfacht und schließlich nach  $k$  auflöst. Das kostet einige Mühe, die Stefan aber willig auf sich nahm. Er erhielt schließlich

$$P(X = k) < P(X = k + 1) \Leftrightarrow k < np + p - 1 .$$

Um zu sehen, was das bedeutet, zeichnest du dir die Merkmalsachse hin und markierst irgendwo zwischen 0 und  $n$  den Erwartungswert  $E(X) = np$ . Die Zahl  $np + p - 1$  liegt links vom Erwartungswert  $np$ , aber nur weniger als 1. Markiere sie in deiner Zeichnung. Für die  $k$ -Werte, die links von der Markierung liegen, ist die rechte Nachbarsäule des Histogramms, die auf  $k + 1$  steht, höher als die Säule auf  $k$ . Die Säulen des Histogramms werden also höher bis zu der, die auf dem ersten  $k$  ab der Markierung steht. Diese ist die höchste Säule, danach werden sie wieder niedriger.<sup>8</sup> Jedenfalls haben die Histogramme binomialverteilter Zufallsgrößen immer Buckelform<sup>9</sup>, und der Buckel liegt in der Nähe des Erwartungswertes. Wenn du mit binomialverteilten Zufallsgrößen zu tun hast, solltest du die Form des Histogramms stets vor Augen haben.

Das Histogrammbilderbuch in der Abbildung 6 auf Seite 14 soll dir Hinweise geben, wie sich die Form des Histogramms ändert, wenn du bei festem  $p$  das  $n$  veränderst: schau dir dazu die Spalten an. Die Histogramme in einer Zeile haben verschiedene  $p$ -Werte, aber den gleichen  $n$ -Wert. Ich habe nur Beispiele mit  $p \leq \frac{1}{2}$  aufgeführt, Histogramme für  $p = \frac{9}{10}$  und für  $p = \frac{3}{10}$  erhältst du durch Spiegeln der abgebildeten Histogramme an geeigneten Achsen.

## 5.7 Mehr Regeln für Erwartungswert und Varianz

Wir werfen einen Würfel und bilden die Zufallsgrößen  $X$  und  $Y$ . Dabei ist  $X$  die gewürfelte Zahl und  $Y$  die Zahl auf der Fläche, auf der der Würfel liegt. Bilden wir jetzt die Zufallsgröße  $S = X + Y$ , passiert etwas Seltsames:  $S$  nimmt nur den einzigen Wert 7 an. Vielleicht magst du dieses  $S$  nicht als Zufallsgröße ansehen, weil dem Zufall kein Raum bleibt; dem Mathematiker sind aber lästige Sonderfälle überaus verhasst, und deshalb lässt er auch konstante Zufallsgrößen zu, insbesondere auch unser  $S$ .

Wir berechnen Erwartungswert und Varianz von  $S$ . Das ist kein Problem, wenn du dich streng an die Definition hältst: Es ist

$$E(S) = \sum_{k \in S(\Omega)} k P(S = k) = 7 \cdot P(S = 7) = 7 \cdot 1 = 7$$

<sup>8</sup>Von Sonderfällen sehe ich hier einmal ab. Es kann zum Beispiel sein, dass zwei benachbarte Säulen gleich hoch sind .

<sup>9</sup>Fachterminus: die Verteilung ist eingipflig.

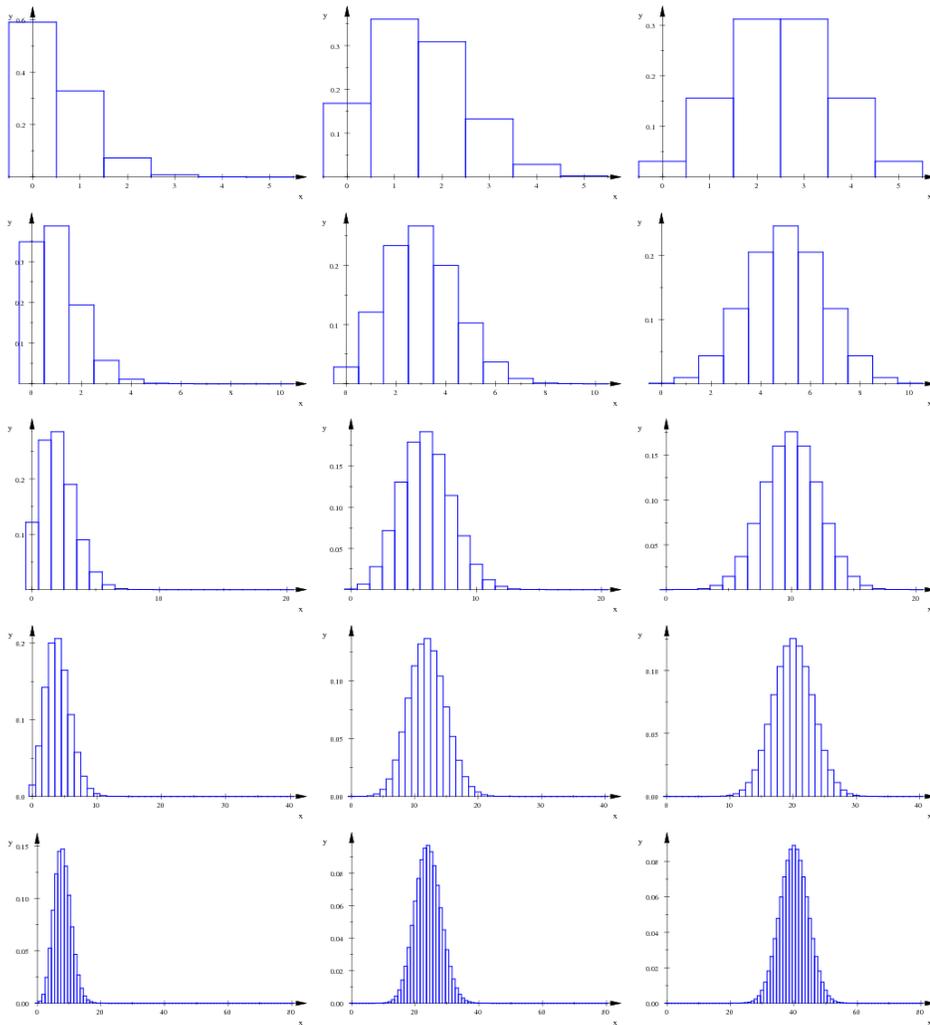


Abbildung 6: **Bilderbuch:** Hier siehst du Histogramme  $B(n, p)$ -verteilter Zufallsgrößen. In der linken Spalte ist  $p = \frac{1}{10}$ , in der mittleren ist  $p = \frac{3}{10}$  und in der rechten Spalte ist  $p = \frac{1}{2}$ . Die  $n$ -Werte betragen jeweils 5, 10, 20, 40 und 80.

und

$$\text{Var}(S) = \sum_{k \in S(\Omega)} (k - 7)^2 P(S = k) = 0 \quad .$$

Wir wissen schon<sup>10</sup>, dass  $E(aX) = aE(X)$  ist. Tragen wir eine Regel für den Erwartungswert nach:

### 9 Lemma

Es sei  $X$  eine Zufallsgröße und  $b \in \mathbb{R}$ . Dann ist

$$E(X + b) = E(X) + b \quad .$$

Das kannst du leicht direkt beweisen, du kannst auch  $b$  als konstante Zufallsgröße auf dem  $\Omega$  des  $X$  ansehen und die Summenregel benutzen, wie du magst.

<sup>10</sup>Siehe Lemma 5 auf Seite 10

Wenden wir uns nun der Varianz zu.

### 10 Satz

Es sei  $X$  eine Zufallsgröße auf  $\Omega$  mit Erwartungswert  $\mu = E(X)$  und  $a, b \in \mathbb{R}$ . Dann gelten die folgenden Aussagen:

$$\text{Var}(X + b) = \text{Var}(X) \quad (8)$$

$$\text{Var}(aX) = a^2 \text{Var}(X) \quad (9)$$

$$\text{Var}(X) = E((X - \mu)^2) \quad (10)$$

$$\text{Var}(X) = E(X^2) - \mu^2 \quad (11)$$

Die ersten beiden Aussagen rechnet du mit Hilfe der Definition nach, die dritte siehst du ein, wenn du die Definition scharf anschaut, und die letzte bekommst du aus der dritten, wenn du die binomischen Formeln und die Regeln für das Rechnen mit Erwartungswerten anwendest.

Nun bleibt uns noch ein schwierigeres Problem:

### 11 Satz

Es seien  $X$  und  $Y$  **unabhängige** Zufallsgrößen auf dem gleichen  $\Omega$ . Dann ist

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \quad .$$

Die Zufallsgrößen  $X$  und  $Y$  heißen unabhängig, wenn für jedes  $k \in X(\Omega)$  die Wahrscheinlichkeit  $P(X = k)$  nicht davon abhängt, welchen Wert  $Y$  annimmt:

$$P(X = k | Y = j) = P(X = k) \quad \text{für alle } j \in Y(\Omega).$$

Ich könnte die Sache aus so ausdrücken: Für alle  $k \in X(\Omega)$  und alle  $j \in Y(\Omega)$  muss

$$P(X^{-1}(k) \cap Y^{-1}(j)) = P(X^{-1}(k)) \cdot P(Y^{-1}(j))$$

gelten, aber ich glaube nicht, dass ich dir damit einen Gefallen tue. Schau dir jedenfalls noch einmal an, was es mit der Unabhängigkeit von Ereignissen auf sich hatte.

Ein Beweis des Satzes stand an der Tafel; ich versuche nicht, ihn hier nachzuerzählen. Es kommt mir auf wichtige Anwendungen des Satzes an: damit kann man nämlich die Varianz einer binomialverteilten Zufallsgröße allgemein angeben!

### 12 Satz

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

$$\text{Var}(X) = np(1 - p) \quad .$$

Zum Beweis stellt man  $X$  als Summe  $X_1 + X_2 + \dots + X_n$  unabhängiger  $B(1, p)$ -verteilter Zufallsgrößen  $X_1, X_2, \dots, X_n$  dar, wie wir es schon bei der Berechnung von  $E(X) = np$  gemacht haben. Dann muss man nur noch

$$\text{Var}(X_1) = (1 - p)^2 p + (0 - p)^2 (1 - p) = p(1 - p) \cdot (p + 1 - p) = p(1 - p)$$

ausrechnen, und schon folgt die Behauptung des Satzes aus der Summenregel für die Varianz.  $\square$

## 6 Klausur Nr. 3 am 21. März 2013

### 1. Einstiegsaufgabe

Aus einer Urne mit sechs weißen und vier roten Kugeln wird zweimal gezogen, und die gezogene Kugel wird **nicht** zurückgelegt. Es sei  $X$  die Anzahl der gezogenen weißen Kugeln.

- (a) Gib  $X(\Omega)$  an, berechne die Werte  $P(X = k)$ ,  $E(X)$  und  $\text{Var}(X)$  und zeichne das Histogramm von  $X$ . [22]
- (b) Zeichne eine Skizze von  $\Omega$  mit der von  $X$  gestifteten Zerlegung. [4]
- (c) Was ist die praktische Bedeutung der Zahl  $E(X)$ , die du eben ausgerechnet hast? [6]

### 2. Würfelaufgaben

- (a) Der kleine Hans will unbedingt eine Sechs würfeln. Wie oft muss er bereit sein zu würfeln, wenn er mit 99,9% Wahrscheinlichkeit erfolgreich sein will? [10]
- (b) Gleich beginnt er. Wie wahrscheinlich ist es, dass die erste Sechs beim einunddreißigsten Wurf kommt? [4]
- (c) Nun hat Hans dreißigmal gewürfelt, und es ist noch keine Sechs gekommen. Wie wahrscheinlich ist es, dass es beim einunddreißigsten Wurf klappt? [3]
- (d) Es sei  $X$  die Anzahl der Sechsen, wenn man einmal würfelt. Jan besorgt sich sechs unabhängige Kopien  $X_1, X_2, \dots, X_6$  von  $X$  und bildet  $S = X_1 + X_2 + \dots + X_6$ . Gerd bildet einfach  $Z = 6X$ . Vergleiche Erwartungswert und Varianz von  $S$  und von  $Z$ . Hinweis:  $E(X) = \frac{1}{6}$  und  $\text{Var}(X) = \frac{5}{36}$  darfst du ohne Beweis benutzen. [12]
- (e) Ein Würfel wird 7200-mal geworfen, und es sei  $X$  die Anzahl der geworfenen Sechsen. Erkläre, was die Wahrscheinlichkeit  $P(|X - 1200| \geq 100)$  bedeutet, gib einen exakten Term für diese Wahrscheinlichkeit und eine obere Schranke für ihren Wert an. [18]
- (f) Ich könnte eure Mitarbeitsnoten ja so bestimmen: ich würfle zweimal und gebe die kleinste gewürfelte Zahl als Note. Mit welcher Wahrscheinlichkeit bekommst du eine zwei, mit welcher eine fünf? [8]

### 3. Vierfeldertafel [12]

In der 9c sind elf Jungen und dreizehn Mädchen. Von den Jungen machen alle bis auf zwei beim Känguruwettbewerb mit, bei den Mädchen machen alle bis auf zwei nicht mit.

- (a) Die Tür des Raums der 9c geht auf. Mit welcher Wahrscheinlichkeit kommt ein Mädchen heraus?
- (b) Auf dem Tisch liegt ein Känguruantwortzettel aus der 9c. Mit welcher Wahrscheinlichkeit ist er von einem Jungen?
- (c) Sind die beiden Merkmale eines Schülers der 9c, ein Junge zu sein und beim Känguruwettbewerb mitzumachen, stochastisch unabhängig?

### 4. Sprachübungen

- (a) Wann ist  $P(A \cup B) = P(A) + P(B)$ ? Erkläre die Aussage auch anhand eines Diagramms. [6]
- (b) Was ist  $P(B | A)$ ? Erkläre auch. [8]

(c)  $\sum_{\omega \in \Omega} P(\{\omega\}) = ?$  [3]

5. **Binomialverteilung.** Es sei  $X$  eine  $B(n, p)$ -verteilte Zufallsgröße.

(a) Gib unseren Term für  $P(X = k)$  an und erkläre die Bestandteile. [8]

(b) Beschreibe, wie das Histogramm von  $X$  aussieht. [4]

(c) Löse die Ungleichung  $\binom{n}{k} < \binom{n}{k+1}$  nach  $k$  auf und erkläre, was das Ergebnis bedeutet. [15]

(d) Bestimme für  $n = 1000$  und  $p = \frac{1}{10}$  die Werte von  $P(X \leq 120)$ , von  $P(X = 100)$  und von  $P(X > 95)$ . [10]

6. **Grundsätzliches**

(a) Was bedeutet es, wenn man sagt, eine geworfene Reißzwecke lande mit der Wahrscheinlichkeit  $\frac{1}{3}$  in der Position  $\perp$ ? Wie bestimmt man diese Zahl eigentlich? [6]

(b) Wie kommt man darauf zu sagen, die Wahrscheinlichkeit, beim Würfeln eine Sechs zu werfen, sei  $\frac{1}{6}$ ? Kann man das beweisen, wie den Innenwinkelsatz? [6+]

## 7 Die lokale Näherungsformel von de Moivre und Laplace

### 7.1 Einstieg

Du weißt, dass das Histogramm einer binomialverteilten Zufallsgröße Buckelform hat. Zwar wird der Buckel immer breiter und flacher, wenn du bei festem  $p$  das  $n$  vergrößerst, aber die Form des Buckels bleibt in einem Maße erhalten, dass man auf den Gedanken kommen kann, all diese Buckel seien in etwa verzerrte Versionen eines Urbuckels.

Dem ist in der Tat so, und wir wollen versuchen, diese Urform herauszuarbeiten. Dazu muss man das Breiterwerden herausrechnen, und das geschieht am einfachsten, indem man die Merkmalsachse mit einem Faktor  $< 1$  streckt.

Einen Hinweis auf den passenden Streckfaktor gibt die Ungleichung von Tschebyschew

$$P(|X - \mu| \geq a) \leq \frac{\text{Var}(X)}{a^2} .$$

Misst man den Radius  $a$  der Umgebung von  $\mu$  in der Einheit  $1\sigma$ , wird  $a = r\sigma$  sein für ein  $r > 0$ , und die Ungleichung erhält die Gestalt

$$P(|X - \mu| \geq r\sigma) \leq \frac{\text{Var}(X)}{r^2\sigma^2} = \frac{1}{r^2} ,$$

das heißt, das Tschebyschewrisiko auf der rechten Seite hängt nur noch von  $r$  ab. Man kann deshalb hoffen, dass  $\frac{1}{\sigma}$  ein geeigneter Streckfaktor ist.

### 7.2 Standardisierung einer Zufallsgröße

Es sei also  $X$  eine Zufallsgröße mit Erwartungswert  $\mu$  und Standardabweichung  $\sigma = \sqrt{\text{Var}(X)}$ . Wir gehen erst einmal von  $X$  zu der neuen Zufallsgröße  $Y := X - \mu$  über. Dann ist

$$E(Y) = E(X - \mu) = E(X) - \mu = 0 \quad \text{und} \quad \text{Var}(Y) = \text{Var}(X - \mu) = \text{Var}(X) .$$

Die Varianz hat sich nicht geändert, und das ist kein Wunder: das Histogramm von  $Y$  bekommt man, indem man das Histogramm von  $X$  um  $\mu$  nach links verschiebt.

Nun strecken wir die Merkmalsachse mit dem Faktor  $\frac{1}{\sigma}$ . Dazu gehen wir von  $Y$  zu der neuen Zufallsgröße

$$Z := \frac{1}{\sigma} Y$$

über. Nun ist

$$E(Z) = E\left(\frac{1}{\sigma} Y\right) = \frac{1}{\sigma} E(Y) = 0 \quad \text{und} \quad \text{Var}(Z) = \text{Var}\left(\frac{1}{\sigma} Y\right) = \frac{1}{\sigma^2} \text{Var}(Y) = 1 .$$

Diese neue Zufallsgröße  $Z$  heißt die **standardisierte Zufallsgröße** zu  $X$ . Für sie gilt, wie wir gerade gezeigt haben, das folgende Lemma.

#### 13 Lemma

Es sei  $X$  eine Zufallsgröße mit Erwartungswert  $\mu$  und Standardabweichung  $\sigma$ . Für die zu  $X$  gehörige **standardisierte Zufallsgröße**

$$Z := \frac{X - \mu}{\sigma} \quad \text{gilt dann} \quad E(Z) = 0 \quad \text{und} \quad \text{Var}(Z) = 1 .$$

Wir schauen uns ein konkretes Beispiel an. Unser  $X$  sei  $B(200, \frac{3}{5})$ -verteilt. Dann ist  $\mu = E(X) = 200 \cdot \frac{3}{5} = 120$  und  $\text{Var}(X) = 200 \cdot \frac{3}{5} \cdot \frac{2}{5} = 48$  und  $\sigma = \sqrt{48} = 4\sqrt{3}$ . Abbildung 7 zeigt Histogramme von  $X$ , von  $Y = X - \mu$  und von  $Z = \frac{X - \mu}{\sigma}$ .

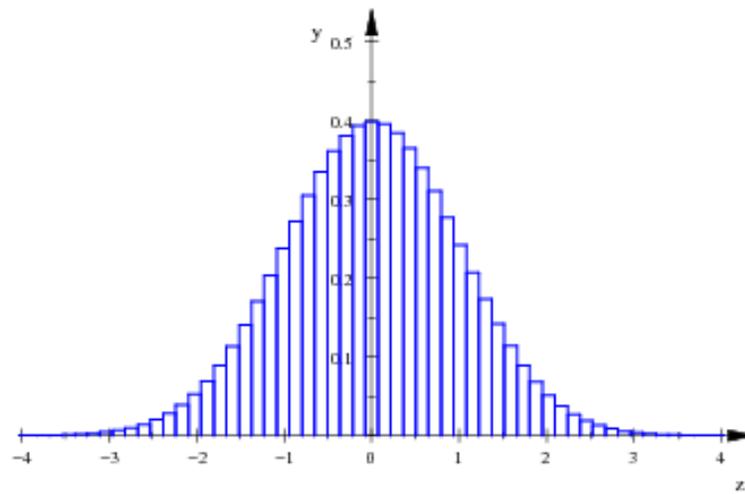
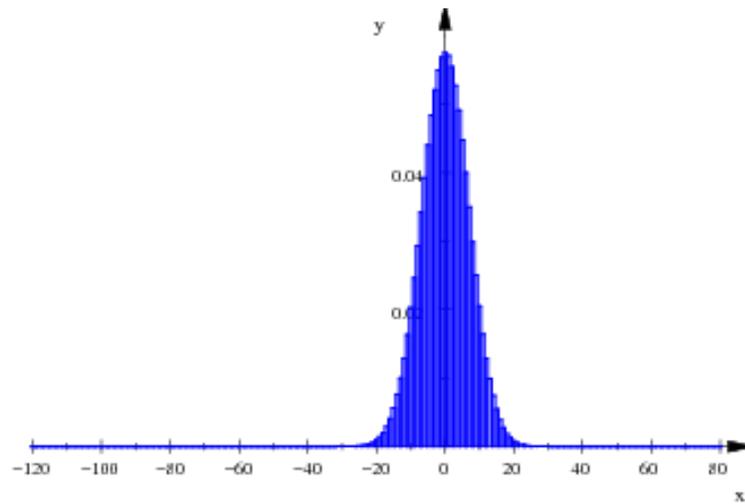
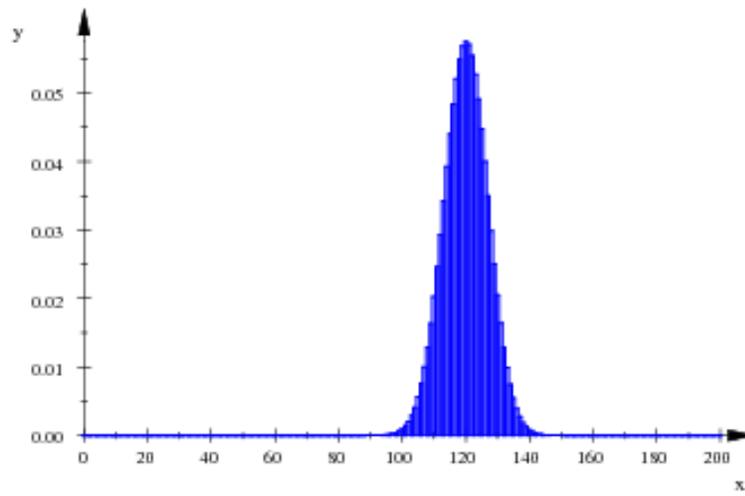


Abbildung 7: Histogramme der  $B(200, \frac{3}{5})$ -verteilten Zufallsgröße  $X$  (oben), von  $Y = X - \mu$  (Mitte) und der zugehörigen standardisierten Zufallsgröße  $Z = \frac{X - \mu}{\sigma}$  (unten)

### 7.3 Die Gaußsche Glockenkurve

An der letzten Kurve in Abbildung 7 kannst du die Form des Urbuckels schon gut ablesen; es handelt sich um die hochberühmte Gaußsche Glockenkurve, sie ist Graph der Funktion

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} . \quad (12)$$

Abbildung 8 zeigt diese Kurve allein und in das Histogramm der standardisierten Zufallsgröße  $Z$  zu unserem  $B(200, \frac{3}{5})$ -verteilten  $X$  aus dem letzten Unterkapitel eingezeichnet.

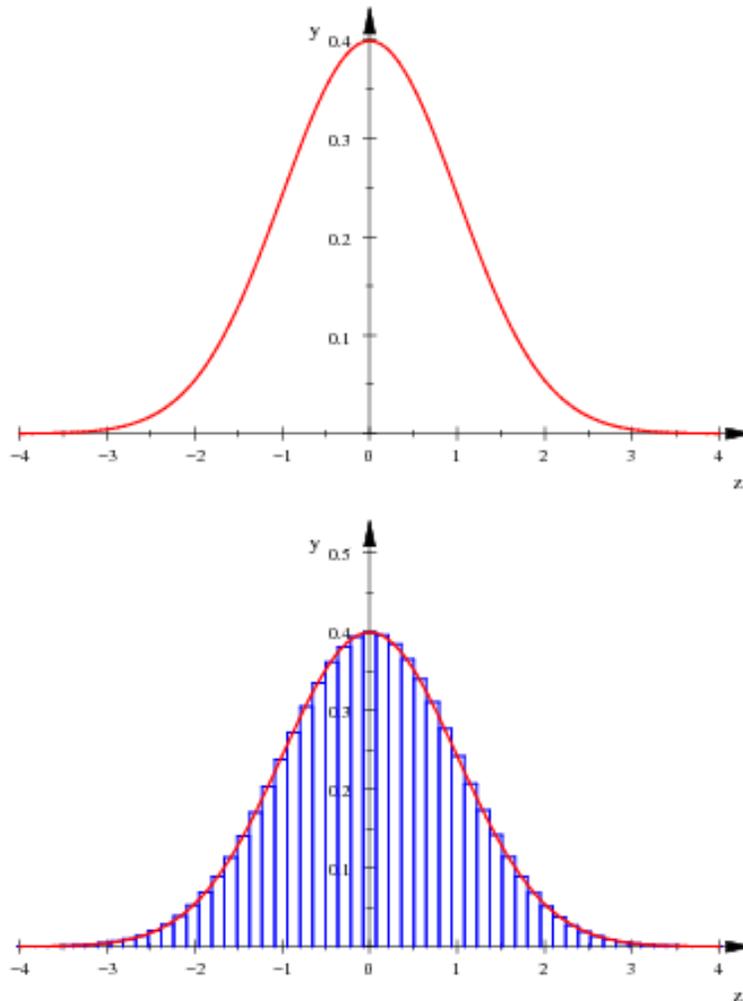


Abbildung 8: Gaußsche Glockenkurve allein und im Histogramm von  $Z$  aus Abbildung 7 unten

## 7.4 Lokale Näherungsformel von de Moivre und Laplace

Ich zeige dir nun an dem schon vertrauten Beispiel des  $B(200, \frac{3}{5})$ -verteilten  $X$ , wie man mit Hilfe der Gaußfunktion  $\varphi$  einen Näherungswert für  $P(X = k)$  bekommt. Wir müssen uns nur genau anschauen, was beim Standardisieren passiert. Schau dir die Abbildung 9 an:

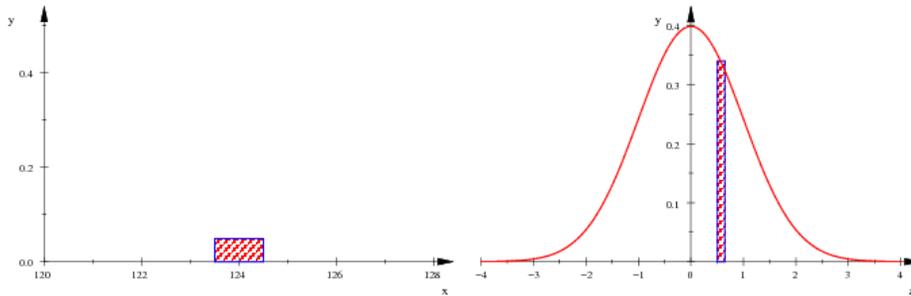


Abbildung 9: Säule beim Wert  $k$  im Histogramm von  $X$  (links) und die Säule im Histogramm der zugehörigen standardisierten Zufallsgröße  $Z$  an der Stelle  $\frac{k-\mu}{\sigma}$

Im Histogramm von  $X$  steht an der Stelle  $k = 124$  eine Säule der Breite 1 und der Höhe  $P(X = k)$ . Beim Standardisieren geht  $k$  in den Wert

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

über. Im Histogramm von  $Z$  steht an der Stelle  $z$  eine Säule mit dem Flächeninhalt  $P(Z = z)$ . Aber  $Z$  nimmt genau dann den Wert  $z$  an, wenn  $X$  den Wert  $k$  annimmt, es ist also  $P(Z = z) = P(X = k)$ . Die Säulen im Histogramm von  $Z$  haben aber nur die Breite  $\frac{1}{\sigma}$ , folglich ist die Höhe der Säule an der Stelle  $z$  nun  $\sigma P(X = k)$ , sonst stimmt der Flächeninhalt nicht. Es folgt

$$\sigma P(X = k) \approx \varphi(z) \quad ,$$

und daraus ergibt sich die **lokale Näherungsformel von de Moivre und Laplace**

$$P(X = k) \approx \frac{1}{\sigma} \varphi\left(\frac{k - \mu}{\sigma}\right) \quad \text{für } B(n, p)\text{-verteiltes } X. \quad (13)$$

Du darfst sie anwenden, wenn du  $P(X = k)$  nicht geschlossen ausrechnen kannst; dann kannst du davon ausgehen, dass der Fehler nicht zu groß ist. Manche benutzen als Faustformel die Bedingung, dass  $np(1-p)$  mindestens 9 betragen soll, aber ich kenne keinen Beweis für diese Kennzahl 9. – Einen Eindruck von der Güte der Näherung mag Abbildung 11 auf Seite 22 geben.

## 7.5 Die Verteilungsfunktion einer Zufallsgröße

### 14 Definition

Die Verteilungsfunktion  $F$  von  $X$  ist definiert durch

$$F(x) := P(X \leq x) \quad .$$

Die Verteilungsfunktion  $F$  enthält die volle Information über die Zufallsgröße. Offensichtlich ist  $F$  monoton steigend. Bei den Zufallsgrößen, mit denen wir bisher zu tun hatten, sind die Verteilungsfunktionen Treppenfunktionen; an jeder Stelle

$k$ , die  $X$  annehmen kann, ist ein Sprung der Größe  $P(X = k)$ . Abbildung 10 zeigt den Graphen der Verteilungsfunktion des  $B(200, \frac{3}{5})$ -verteilten  $X$ , mit dem wir uns jetzt länger beschäftigt haben. Man erkennt in der Abbildung gerade noch, dass es sich um den Graphen einer Treppenfunktion handelt.

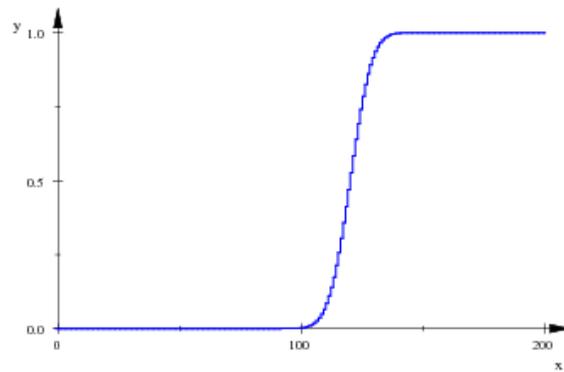


Abbildung 10: Graph der Verteilungsfunktion des  $B(200, \frac{3}{5})$ -verteilten  $X$

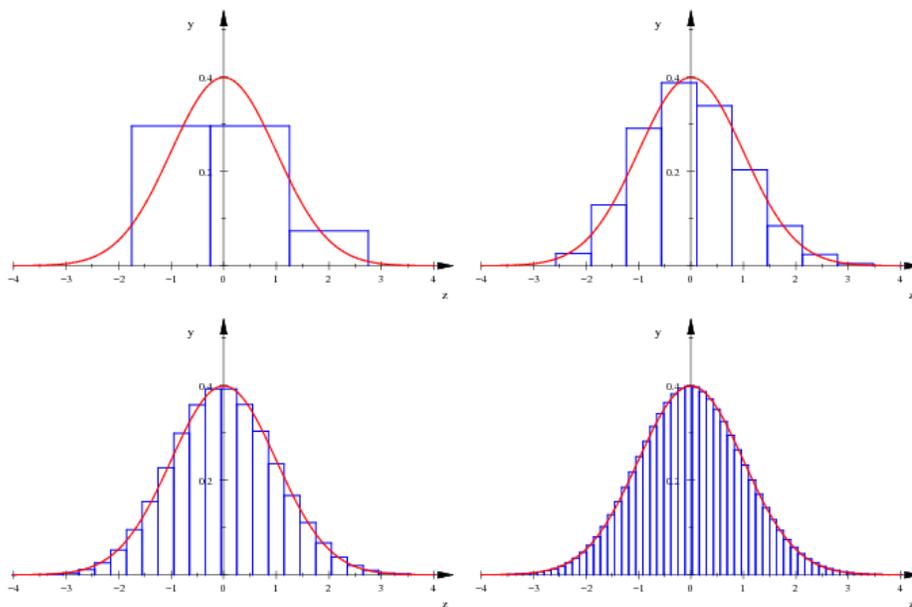


Abbildung 11: Graphen der Gaußschen Glockenkurve und Histogramme der standardisierten Zufallsgrößen zu  $B(n, p)$ -verteiltem  $X_n$  für  $p = \frac{1}{3}$  und  $n \in \{2, 10, 50, 250\}$

## 7.6 Integrale Näherungsformel von de Moivre und Laplace

Die integrale Näherungsformel von de Moivre und Laplace liefert einen Näherungswert für die Wahrscheinlichkeit  $P(X \leq k)$ , dass ein ausgeloster Wert der  $B(n, p)$ -verteilten Zufallsgröße  $X$  höchstens  $k$  ist. Ich demonstriere sie an unserem Beispiel  $n = 200$  und  $p = \frac{3}{5}$ . Wir fragen nach  $P(X \leq 125)$ . Im Histogramm von  $X$  ist die gesuchte Wahrscheinlichkeit der Inhalt einer Fläche, in Abbildung 12 ist die Fläche schraffiert. Beim Standardisieren geht die Fläche in das schraffierte Flächenstück im Histogramm der standardisierten Zufallsgröße  $Z = \frac{X - \mu}{\sigma}$  über, und die beiden Flächenstücke haben den gleichen Inhalt.

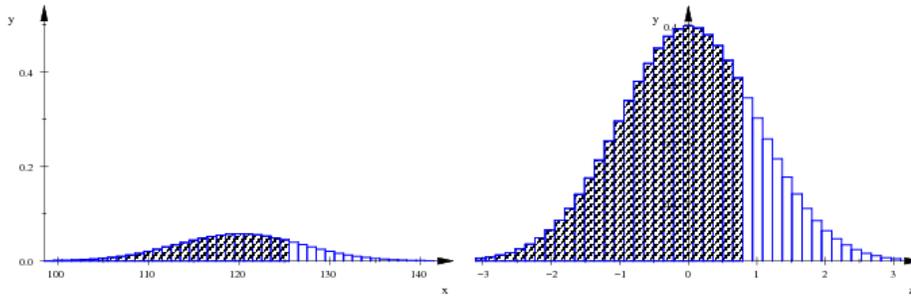


Abbildung 12: Zur Bestimmung von  $P(X \leq k) = P\left(Z \leq \frac{k - \mu}{\sigma}\right)$

Den Inhalt des Flächenstücks im Histogramm von  $Z$  bestimmt man näherungsweise mit einem Integral über die Gaußfunktion  $\varphi$ ; dazu verwendet man die Integralfunktion

$$\Phi(z) := \int_{-\infty}^z \varphi(t) dt \quad . \quad (14)$$

Man muss nur noch aufpassen, dass man vom letzten Kästchen im Histogramm von  $Z$ , das auf  $\frac{k - \mu}{\sigma}$  steht, nicht die Hälfte wegschneidet, dann hat man schon die **integrale Näherungsformel von de Moivre und Laplace**:

$$P(X \leq k) \approx \int_{-\infty}^{\frac{k - \mu}{\sigma} + \frac{1}{2\sigma}} \varphi(x) dx = \Phi\left(\frac{k - \mu}{\sigma} + \frac{1}{2\sigma}\right) \quad \text{für } B(n, p)\text{-verteiltes } X \quad (15)$$

Werte von  $\Phi$  muss man in Tabellen nachschlagen, das Integral kann man nicht analytisch ausrechnen.

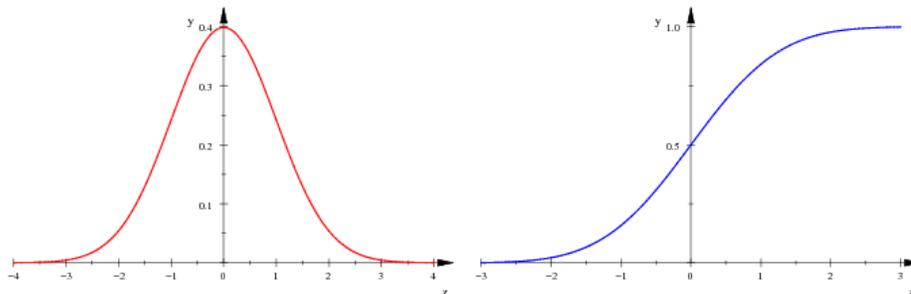


Abbildung 13: Graphen von  $\varphi$  und  $\Phi$

## 8 Stetige Zufallsgrößen

### 8.1 Beispiele und Grundbegriffe

Wir drehen ein ideales Glücksrad am Einheitskreis, wir wiegen eine zufällig gewählte Packung Zucker oder wir messen den Zerfallszeitpunkt eines bestimmten radioaktiven Teilchens. Die Ergebnisse dieser Versuche hängen vom Zufall ab, wir haben es also mit drei Zufallsgrößen  $X$ ,  $Y$  und  $Z$  zu tun. Neu an diesen Zufallsgrößen ist, dass sie jede Zahl eines bestimmten Intervalls annehmen können. Klar, wenn man würfelt und die Anzahl  $W$  der Würfe bis zur ersten Sechs zählt, hat man auch eine Zufallsgröße mit unendlich vielen Werten; sie nimmt die positive ganze Zahl  $k$  mit der Wahrscheinlichkeit

$$P(W = k) = p(1 - p)^{k-1} \quad (16)$$

an, und sie heißt **geometrisch verteilt**. Die kennen wir schon, aber die Werte, die sie annehmen kann, kann man der Reihe nach aufzählen. Man sagt deshalb, ihr Wertebereich  $W(\Omega)$  sei **abzählbar unendlich**. Die Zahlen eines Intervalls kann man nicht der Reihe nach aufzählen, dafür sind es zu viele: ein Intervall  $[a, b]$  mit  $a < b$  enthält **überabzählbar unendlich** viele Zahlen.<sup>11</sup>

Beim Glücksrad hat jede Zahl  $x \in [0, 2\pi[$  die gleiche Wahrscheinlichkeit, ausgelost zu werden. Diese Wahrscheinlichkeit muss dann Null sein, sonst kämen wir nicht mit der Eins für die Wahrscheinlichkeit  $P(\Omega)$  aus. Aber jede Drehung des Glücksrads beschert uns eine dieser Zahlen! Deshalb nennt man in der Stochastik ein Ereignis mit der Wahrscheinlichkeit Null **fast unmöglich** und eines mit der Wahrscheinlichkeit Eins **fast sicher**. Schau, die Wahrscheinlichkeit  $P(X > 0)$ , dass das Glücksrad nicht bei Null stehen bleibt, ist Eins. Dennoch würdest du dich nicht auf einen Stuhl setzen, der dich mit Hochspannung umbringt, wenn das Glücksrad die Null auslost, und du tust gut daran, obwohl natürlich  $P(X \neq 0) = 1$  ist.

Für die Wahrscheinlichkeit bei den neuen Zufallsgrößen braucht man ein anderes Konzept:

#### 15 Definition

Es sei  $f : \mathbb{R} \rightarrow \mathbb{R}^{\geq 0}$  eine Funktion mit  $\int_{-\infty}^{\infty} f(t) dt = 1$ . Die Zufallsgröße  $X$  heißt **stetig mit Dichtefunktion  $f$** , wenn

$$P(a < X \leq b) = \int_a^b f(t) dt$$

ist. Die Funktion  $F$  mit

$$F(x) = \int_{-\infty}^x f(t) dt$$

heißt **Verteilungsfunktion** von  $X$ .

Dichtefunktion der Glücksradzufallsgröße  $X$  von oben ist

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

Der Zerfallszeitpunkt  $Z$  eines bestimmten radioaktiven Teilchens hat bei geeigneter Wahl der Zeiteinheit die Dichtefunktion

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

<sup>11</sup>Das sind Entdeckungen Georg Cantors aus der zweiten Hälfte des 19. Jahrhunderts.

## 8.2 Über Erwartungswert und Varianz stetiger Zufallsgrößen

Bei einer (diskreten) Zufallsgröße  $X$ , die nur die Werte  $x_1, x_2, \dots, x_n$  annehmen kann, ist der Erwartungswert

$$E(X) = \sum_{k=1}^n x_k P(X = x_k) .$$

Für stetiges  $X$  ist diese Definition nicht brauchbar, denn es ist ja

$$P(X = x) = 0 \quad \text{für alle } x \in \mathbb{R}.$$

Wir können aber ein kleines Intervall der Breite  $\Delta x$  betrachten, das  $x$  enthält, und als Beitrag des Intervalls zum Erwartungswert den Term

$$x f(x) \Delta x$$

heranziehen, denn die Wahrscheinlichkeit, dass  $X$  einen Wert in dem kleinen Intervall annimmt, ist ungefähr  $f(x) \Delta x$ .<sup>12</sup>

Den Beitrag eines größeren Abschnitts  $[a, b]$  der Merkmalsachse können wir nun so erfassen: Wir teilen den Abschnitt in  $n$  Teile der Länge  $\Delta x = \frac{b-a}{n}$  ein und wählen in jedem Teilstück einen  $x$ -Wert. Nennen wir den  $x$ -Wert im  $k$ -ten Teilstück  $x_k$ , erhalten wir als Beitrag des Abschnitts  $[a, b]$  die Summe

$$\sum_{k=1}^n x_k f(x_k) \Delta x .$$

Dies ist offensichtlich eine Riemannsche Summe des Integrals

$$\int_a^b x f(x) dx ,$$

folglich strebt die Summe für  $n \rightarrow \infty$  gegen dieses Integral. Nun müssen wir  $a$  gegen  $-\infty$  und  $b$  gegen  $\infty$  schieben, dann haben wir einen sinnvollen Term für  $E(X)$ . Für die Varianz läuft das auf die gleiche Weise.

### 16 Definition

Es sei  $X$  eine stetige Zufallsgröße mit Dichtefunktion  $f$ . Dann sind Erwartungswert und Varianz von  $X$  definiert durch

$$\begin{aligned} \mu = E(X) &:= \int_{-\infty}^{\infty} x f(x) dx \quad \text{und} \\ \text{Var}(X) &:= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx , \end{aligned}$$

falls diese Integrale existieren.

Der Vorbehalt, dass die Integrale existieren sollen, ist durchaus praktischer Natur. Es kann passieren, dass der Erwartungswert existiert, die Varianz aber nicht. Wenn der Erwartungswert nicht existiert, braucht man nach der Varianz natürlich nicht mehr zu fragen.

**Beispiel.** Zeige, dass die Funktion

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

als Dichtefunktion einer Zufallsgröße  $X$  geeignet ist, aber kein Erwartungswert von  $X$  existiert.

<sup>12</sup>Mache dir dies sorgfältig anhand einer Skizze klar!

### 8.3 Regeln für Erwartungswert und Varianz

Wenn du dir die Beweise unserer Regeln für Erwartungswert und Varianz anschaust, siehst du, dass du sie problemlos auf stetige Zufallsgrößen übertragen kannst. Eigentlich musst du nur die Summen durch Integrale ersetzen, wie du es in dem Text vor der Definition des Erwartungswerts auf Seite 25 gesehen hast. Du hast also die gewohnten Regeln zur Verfügung:

$$\begin{aligned}E(aX + b) &= a E(X) + b \\E(X + Y) &= E(X) + E(Y) \\ \text{Var}(aX) &= a^2 \text{Var}(X) \\ \text{Var}(X + b) &= \text{Var}(X) \\ \text{Var}(X) &= E(X^2) - E(X)^2 \\ \text{Var}(X + Y) &= \text{Var}(X) + \text{Var}(Y) \quad \text{für unabhängige } X \text{ und } Y,\end{aligned}$$

und es gilt sogar die Ungleichung von Tschebyschew. Du solltest allerdings nicht penibel nachfragen, was die Unabhängigkeit von Zufallsgrößen formal genau bedeutet; das ist eine ziemlich unangenehme Angelegenheit.

### 8.4 Die Normalverteilung

Von der Gaußfunktion  $\varphi$  wissen wir längst, dass  $\varphi(z) \geq 0$  ist für alle  $z \in \mathbb{R}$  und dass  $\int_{-\infty}^{\infty} \varphi(z) dz = 1$  ist. Das heißt, dass  $\varphi$  als Dichtefunktion geeignet ist.

#### 17 Definition

Eine stetige Zufallsgröße mit Dichtefunktion  $\varphi$  heißt *standardnormalverteilt*, in Zeichen:  $N(0, 1)$ -verteilt.

Wir können die Gaußkurve in Richtung der Merkmalsachse mit einem Faktor  $\sigma > 0$  strecken und die gestreckte Gaußkurve um  $\mu$  nach rechts verschieben<sup>13</sup>. Aus dem Punkt  $(z, \varphi(z))$  der Gaußkurve wird dabei der Punkt  $(\sigma z + \mu, \varphi(z))$ . Das Integral über die Funktion zur verschobenen Gaußkurve ist nun nicht mehr 1, sondern  $\sigma$ , weil wir das Flächenstück in  $z$ -Richtung mit  $\sigma$  gestreckt haben. Um wieder eine Dichtefunktion zu bekommen, strecken wir in  $y$ -Richtung mit  $\frac{1}{\sigma}$ , dann haben wir den Graphen einer Dichtefunktion  $f$ , und es ist

$$f(x) = \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2} . \quad (17)$$

#### 18 Definition

Eine stetige Zufallsgröße, deren Dichtefunktion die Funktion  $f$  aus Gleichung 17 ist, heißt *normalverteilt*, genauer  $N(\mu, \sigma^2)$ -verteilt.

Wie man Wahrscheinlichkeiten bei normalverteilten Zufallsgrößen berechnen kann, zeigt das folgende Lemma:

#### 19 Lemma

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

$$P(a < X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right) .$$

---

<sup>13</sup>Siehe Abbildung 14 auf Seite 27.

**Beweis.** Es ist

$$P(a < X \leq b) = \int_a^b f(x) dx = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx .$$

Wir transformieren dieses Integral durch die Substitution

$$z = \frac{x - \mu}{\sigma} ,$$

das ergibt

$$\frac{dz}{dx} = z' = \frac{1}{\sigma} , \quad \text{also} \quad dz = \frac{1}{\sigma} dx .$$

Wir beachten noch, dass  $z$  von  $\frac{a-\mu}{\sigma}$  bis  $\frac{b-\mu}{\sigma}$  läuft, wenn  $x$  von  $a$  bis  $b$  läuft, und setzen alles ein. Wir erhalten

$$P(a < X \leq b) = \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \varphi(z) dz = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) ,$$

wie behauptet. □

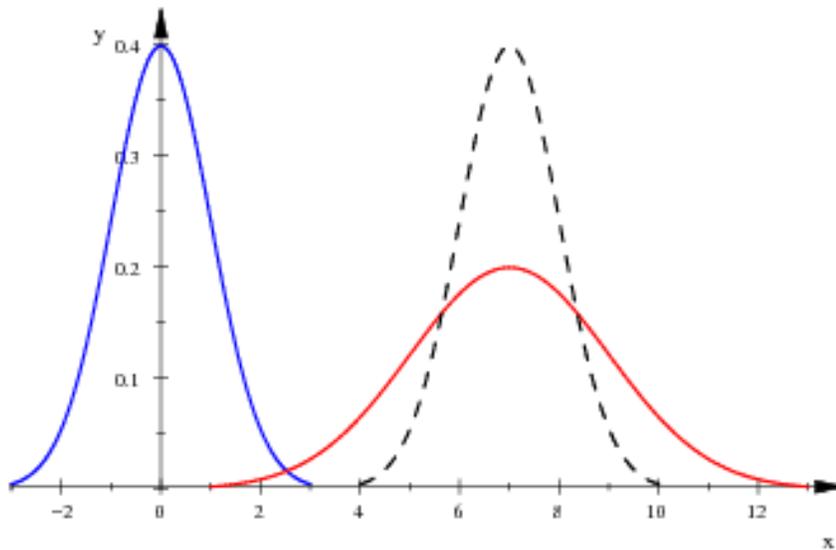


Abbildung 14: Gaußkurve (blau) und mit  $\mu = 7$  und  $\sigma = 2$  transformierte Kurve (rot), also die Dichtefunktionen einer  $N(0, 1)$ - und einer  $N(7, 2^2)$ -verteilten Zufallsgröße. Die schwarze gestrichelte Kurve ist die um  $\mu$  nach rechts verschobene Gaußkurve.

## 8.5 Eine Anwendungsaufgabe

Im Buch<sup>14</sup> findest du eine Aufgabe, die in einer Zuckerfabrik spielt. Es werden Zuckerpackungen abgefüllt, jede soll 1 kg Zucker enthalten. Das stimmt natürlich nie ganz exakt. Wählt man ein Zuckerpaket zufällig aus der laufenden Produktion, kann man seine Masse als Zufallsgröße  $X$  ansehen. Mit Hilfe der Daten  $\bar{x} = 1006\text{ g}$  und  $s = 5\text{ g}$  am Rand des Buches berechnest du die Wahrscheinlichkeit

$$P(X < 1000) \approx \Phi\left(\frac{1000 - 1006}{5}\right) \approx 0.1539 \text{ ,}$$

dass das Zuckerpaket weniger als 1 kg Zucker enthält, und daraus sollst du schließen, dass etwa 15.4% der Fertigung Mindergewicht aufweisen.

So weit, so gut. Ich wünsche mir allerdings, dass du jetzt nicht dein Heft zufrieden zuklappst, sondern dass du nachfragst. Wenigstens zwei Fragen liegen doch auf der Hand:

1. Was hat die Normalverteilung mit Zuckerpaketen zu tun?
2. Woher kommen die Werte  $\bar{x}$  und  $s$ , die in der Aufgabe angegeben sind?

Mehr dazu erfährst du im nächsten Kapitel.

---

<sup>14</sup>Seite 145 Aufgabe 15

## 9 Zwei Grenzwertsätze

### 9.1 Bernoullis Schwaches Gesetz der großen Zahl

Den Wert  $\bar{x}$  in der Anwendungsaufgabe auf Seite 28 ermittelt die Zuckerfabrik, indem sie zufällig ausgewählte Pakete auswiegt und den Mittelwert der Messwerte bildet; der ist schon das  $\bar{x}$ . Solange sich nichts Grundsätzliches ändert, also niemand die Maschinen verstellt oder so etwas, gibt es nur zufallsbedingte Schwankungen der Messwerte, und dann stabilisiert sich der Mittelwert bei wachsendem Stichprobenumfang – das sagt uns das Empirische Gesetz der großen Zahl.

So weit, so gut. Nun ist unser Job nicht die Qualitätskontrolle in einer Fertigung, bei uns geht es um Wahrscheinlichkeitstheorie, also um reine Mathematik. Eine Zufallsgröße ist ein abstraktes Konstrukt, gebildet in der Gedankenwelt der Mathematiker, und ob die materielle Welt in diese Bereiche ausstrahlt, ist keineswegs klar. Deshalb wird sich der Mathematiker eine seiner Zufallsgrößen  $X$  hernehmen, sich davon beliebig viele „unabhängige Kopien“

$$X_1, X_2, X_3, \dots$$

beschaffen und die Mittelwerte

$$S_n := \frac{1}{n} \sum_{k=1}^n X_k$$

bilden. Diese  $S_n$  sind selbst Zufallsgrößen, das sollte dir klar sein! Man lost  $n$  Werte unserer Zufallsgröße  $X$  aus, indem man den zugrundeliegenden Zufallsversuch  $n$ -mal unabhängig ausführt und jeweils den  $X$ -Wert des Ergebnisses bildet. Der  $X$ -Wert der  $k$ -ten Durchführung ist der Wert von  $X_k$ , so sind die  $X_k$  definiert. Die  $n$  Einzelwerte addiert man und dividiert die Summe durch  $n$ , das Ergebnis ist ein Wert von  $S_n$ . Jede Wiederholung liefert einen neuen zufälligen Wert von  $S_n$ . Durchdenke das gründlich an einem Beispiel,  $X$  sei dabei meinetwegen die Augenzahl beim Werfen eines Würfels.

Wenn dir die Angelegenheit hinreichend klar ist, können wir die theoretische Untersuchung weitertreiben. Ist  $E(X) = \mu$  und  $\text{Var}(X) = \sigma^2$ , erhalten wir nach unseren Regeln

$$E(S_n) = E\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} E\left(\sum_{k=1}^n X_k\right) = \frac{1}{n} \sum_{k=1}^n E(X_k) = \frac{1}{n} \cdot n\mu = \mu$$

und, in gleicher Weise,

$$\text{Var}(S_n) = \text{Var}\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n} .$$

Jetzt müssen wir den Sack nur noch zubinden. Nach der Ungleichung von Tschebyschew gilt für jede noch so kleine Zahl  $\epsilon > 0$

$$P(|S_n - \mu| \geq \epsilon) \leq \frac{\text{Var}(S_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n \rightarrow \infty} 0 .$$

Dies bedeutet Folgendes: Wir können eine noch so kleine Umgebung um den Erwartungswert  $\mu$  vorgeben. Die Wahrscheinlichkeit, dass ein beobachteter Wert von  $S_n$  außerhalb dieser Umgebung liegt, geht stets gegen Null für  $n$  gegen Unendlich. Anders formuliert: Mittelwerte von Werten einer Zufallsgröße<sup>15</sup> streuen weniger als die Einzelwerte selbst. – Wir formulieren das Ergebnis in präziser Form als Satz:

<sup>15</sup>mit endlicher Varianz

**20 Satz (Schwaches Gesetz der großen Zahl von J. Bernoulli)**

Es sei  $X$  eine Zufallsgröße mit Erwartungswert  $\mu$  und Varianz  $\text{Var}(X)$  und es seien  $X_1, X_2, X_3, \dots$  unabhängige Kopien von  $X$ . Dann gilt

$$\mathbb{P} \left( \left| \frac{1}{n} \sum_{k=1}^n X_k - \mu \right| \geq \epsilon \right) \xrightarrow{n \rightarrow \infty} 0$$

für jedes noch so kleine  $\epsilon > 0$ .

Im Grunde rechtfertigt Bernoullis schwaches Gesetz der großen Zahl nachträglich unsere Definition des Erwartungswerts auf Seite 9!

**9.2 Der Zentrale Grenzwertsatz**

Es sei  $X$  eine Zufallsgröße mit Erwartungswert  $\mu$  und Standardabweichung  $\sigma$ . Wir verschaffen uns wieder unabhängige Kopien

$$X_1, X_2, X_3, \dots$$

von  $X$  und bilden die Summe

$$Y_n := \sum_{k=1}^n X_k .$$

Der Zentrale Grenzwertsatz sagt dann, dass die  $Y_n$  näherungsweise normalverteilt sind:

**21 Satz (Zentraler Grenzwertsatz von Lindeberg und Levy)**

Es sei  $X$  eine Zufallsgröße mit Erwartungswert  $\mu$  und Varianz  $\sigma^2$ , und es seien  $X_1, X_2, X_3, \dots$  unabhängige Kopien von  $X$ . Für jedes  $n \in \mathbb{N}$  definieren wir die Zufallsgröße

$$Y_n := \sum_{k=1}^n X_k$$

und die zugehörige standardisierte Zufallsgröße

$$Z_n := \frac{Y_n - \mathbb{E}(Y_n)}{\sqrt{\text{Var}(Y_n)}} = \frac{Y_n - n\mu}{\sqrt{n}\sigma} .$$

Dann gilt

$$\mathbb{P}(Z_n \leq z) \xrightarrow{n \rightarrow \infty} \Phi(z) \quad \text{für alle } z \in \mathbb{R}.$$

**Bemerkung 1.** Diesen Satz können wir nicht beweisen, aber er ist von großer praktischer Bedeutung, und manche schätzen ihn als schönstes Ergebnis der Wahrscheinlichkeitstheorie ein. Damit dir die Aussage klar wird, solltest du bemerken, dass die Summe  $Y_n$  der ersten  $n$  Kopien  $X_k$  den Erwartungswert  $n\mu$ , die Varianz  $n \text{Var}(X) = n\sigma^2$  und folglich die Standardabweichung  $\sqrt{n}\sigma$  hat. Du siehst dann, dass eine Wahrscheinlichkeit der standardisierten Zufallsgröße  $Z_n$  näherungsweise mit Hilfe von  $\Phi$  berechnet werden kann: Wenn  $\mathbb{P}(Z_n \leq z)$  gegen  $\Phi(z)$  strebt, muss ja

$$\mathbb{P}(Z_n \leq z) \approx \Phi(z)$$

sein für genügend große  $n$ .

**Bemerkung 2.** Schau hin: Falls  $X$  eine  $B(1, p)$ -verteilte Zufallsgröße ist, ist  $Y_n$  binomialverteilt mit den Parametern  $n$  und  $p$ , und aus unserem Zentralen Grenzwertsatz wird die integrale Näherungsformel von de Moivre und Laplace!

**Bemerkung 3.** Es reicht, wenn die  $X_1, X_2, X_3, \dots$  lediglich unabhängig sind, sie brauchen nicht unabhängige Kopien einer Vorlage  $X$  zu sein. Die  $X_k$  müssen Erwartungswerte  $\mu_k$  und Varianzen  $\sigma_k^2$  haben. Die entsprechend gebildeten Summen

$$Y_n = \sum_{k=1}^n X_k$$

haben dann den Erwartungswert

$$E(Y_n) = \sum_{k=1}^n \mu_k$$

und die Standardabweichung

$$\sqrt{\text{Var}(Y_n)} = \sqrt{\sum_{k=1}^n \sigma_k^2} ,$$

und es gilt unter sehr schwachen Voraussetzungen

$$P\left(\frac{Y_n - E(Y_n)}{\sqrt{\text{Var}(Y_n)}} \leq z\right) \xrightarrow{n \rightarrow \infty} \Phi(z) . \quad (18)$$

Viele Anwender berechnen Wahrscheinlichkeiten bei Summen von unabhängigen Zufallsgrößen für genügend große  $n$  näherungsweise mit Hilfe von  $\Phi$ , und dies rechtfertigt der Zentrale Grenzwertsatz in der Form von Gleichung (18).

**Bemerkung 4.** Die Abweichung der Füllmenge einer Zuckerpackung vom Sollwert ergibt sich als Summe vieler kleiner zufälliger Einzelfehler, die idealerweise unabhängig sind. So kommt es, dass die Füllmenge wenigstens näherungsweise normalverteilt ist.

### 9.3 Ein Anwendungsbeispiel

Mit Hilfe eines Glücksrades werden Zahlen ausgelost. Dabei hat die Null die Wahrscheinlichkeit  $\frac{1}{2}$ , die Eins die Wahrscheinlichkeit  $\frac{1}{6}$  und die Zwei die Wahrscheinlichkeit  $\frac{1}{3}$ . Wir haben es also mit einer Zufallsgröße  $X$  zu tun mit  $X(\Omega) = \{0, 1, 2\}$ ,  $P(X = 0) = \frac{1}{2}$ ,  $P(X = 1) = \frac{1}{6}$  und  $P(X = 2) = \frac{1}{3}$ . Es ist nicht schwer,  $E(X) = \frac{5}{6}$  und  $\text{Var}(X) = \frac{29}{36}$  auszurechnen.

Wir verschaffen uns wieder unabhängige Kopien  $X_1, X_2, X_3, \dots$  von  $X$  und bilden die Summen

$$Y_n := \sum_{k=1}^n X_k \quad .$$

Dann ist

$$E(Y_n) = n E(Y) = n \cdot \frac{5}{6} \quad \text{und} \quad \text{Var}(Y_n) = n \text{Var}(X) = n \cdot \frac{29}{36} \quad .$$

Nach dem Zentralen Grenzwertsatz sollte  $Y_n$  für genügend große  $n$  näherungsweise normalverteilt sein. Abbildung 15 auf Seite 33 zeigt einige Histogramme der zugehörigen standardisierten Zufallsgrößen

$$Z_n = \frac{Y_n - n \cdot \frac{5}{6}}{\frac{1}{6} \sqrt{n \cdot 29}} \quad ,$$

die Gaußsche Glockenkurve ist gleich mit eingezeichnet. Offensichtlich kann man spätestens ab  $n = 20$  die Glockenkurve verwenden, um Wahrscheinlichkeiten auszurechnen. Es wäre zum Beispiel

$$P(Y_{20} \leq 18) \approx \Phi \left( \frac{18 + \frac{1}{2} - 20 \cdot \frac{5}{6}}{\sqrt{20 \cdot \frac{29}{36}}} \right) \approx \Phi(0.4567501392) \approx 0.6761 \quad .$$

Der exakte Wert dieser Wahrscheinlichkeit ist

$$P(Y_{20} \leq 18) = \frac{275559354604243}{406239826673664} \approx 0.6783 \quad ,$$

unser Näherungswert ist also hinreichend genau.<sup>16</sup>

---

<sup>16</sup>Erzeugung der Bilder und Berechnung der Werte erfolgte mit zgws.mn .

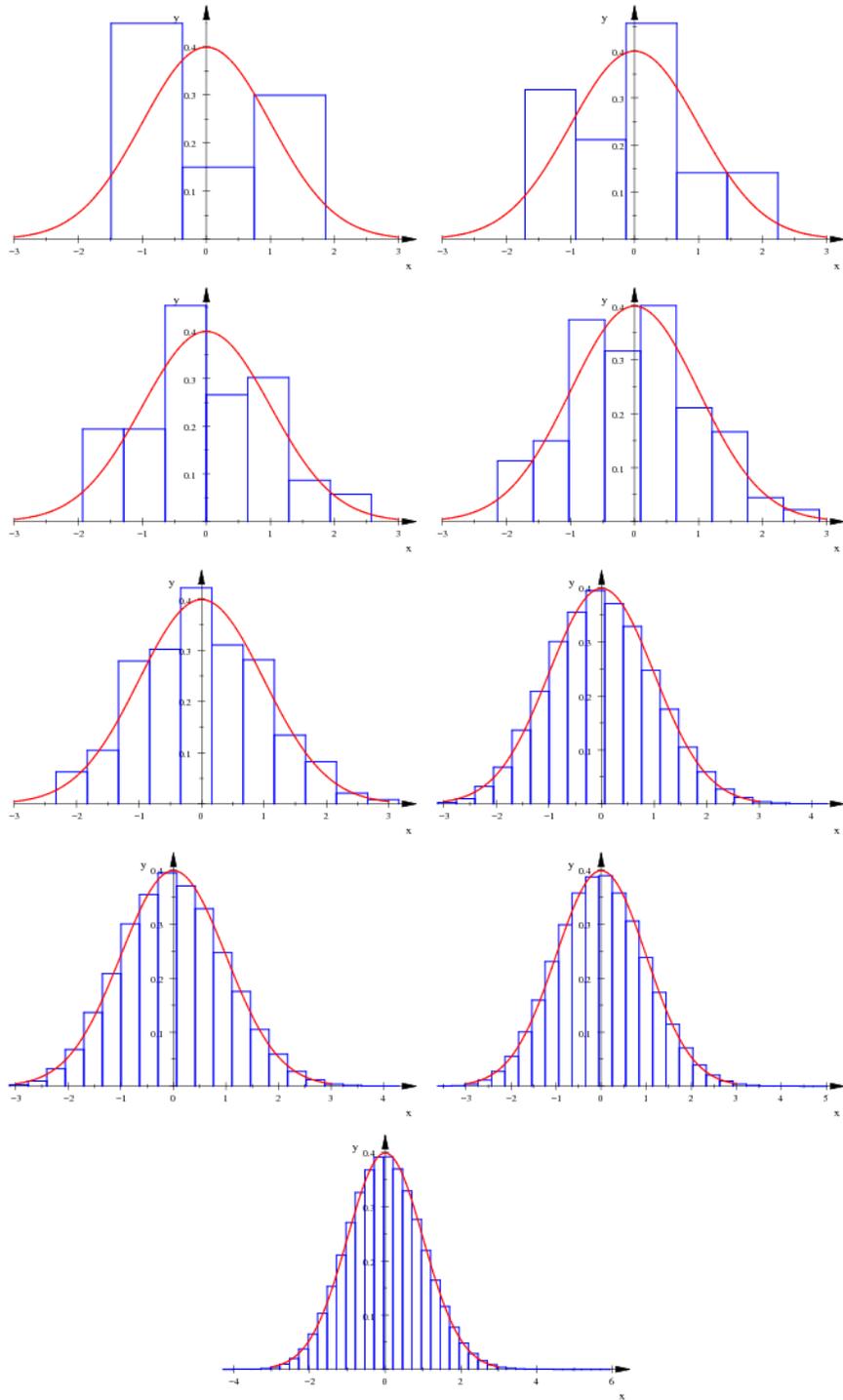


Abbildung 15: Histogramme der standardisierten Zufallsgrößen zu  $Y_n$  für  $n = 1, 2, 3, 4, 5, 10, 15$  und  $n = 20$ , jeweils mit Gaußscher Glockenkurve

## 10 Klausur Nr. 4 am 16. Mai 2013

### 1. Rechenaufgaben

[26]

Es sei  $X$  binomialverteilt mit  $n = 400$  und  $p = \frac{1}{10}$ . Berechne

- (a)  $P(X = 0)$
- (b)  $P(X = 1)$
- (c)  $P(X = E(X))$
- (d)  $P(41 \leq X \leq 46)$
- (e) eine möglichst kleine Zahl  $a$  so, dass  $P(X \leq a) \geq \frac{4}{5}$  ist;
- (f) eine möglichst kleine Zahl  $a$  so, dass sich  $X$  vom Erwartungswert höchstens mit der Wahrscheinlichkeit  $\frac{1}{10}$  um  $a$  oder mehr unterscheidet.

### 2. Würfeln

Ein gewöhnlicher Würfel wird geworfen, und  $X$  ist die gewürfelte Augenzahl.

- (a) Berechne  $E(X)$  und schreibe eine Formel für  $\text{Var}(X)$  hin. [6]
- (b) Zeichne den Graphen der Funktion  $F$  mit  $F(x) = P(X \leq x)$ . Wie nennt man diese Funktion bezüglich  $X$ ? [5]
- (c) Der kleine Kurt will 2400-mal würfeln. Wie groß wird die Augensumme sein
  - i. mindestens?
  - ii. höchstens?
  - iii. im Mittel? [5]
- (d) Mit welcher Wahrscheinlichkeit verfehlt Kurts Augensumme bei 2400 Würfeln den Erwartungswert um mindestens 150? Was sagt die Tschebyschew-Ungleichung, was der Zentrale Grenzwertsatz? Und rechne nicht gleich los, bilde erst die nötigen Objekte. Du kannst  $\text{Var}(X) = \frac{35}{12}$  für das  $X$  von oben ohne eigene Rechnung verwenden. [10]
- (e) Kurt stellte fest, dass bei seinen 2400 Würfeln 502 Sechsen auftraten. Ist das ungewöhnlich? [6]

### 3. $\sigma$ -Regeln

Jan kommt empört zu dir und beschwert sich: „Da wollte ich mal die  $1\sigma$ -Regel ausprobieren und habe mir die allereinfachste Zufallsgröße dafür hergenommen, nämlich die Anzahl von Kopf beim einfachen Münzwurf. Und was passiert? Der Wert liegt mit der Wahrscheinlichkeit 1 in der  $1\sigma$ -Umgebung des Erwartungswerts und nicht mit der, die im Buch steht. Es ist doch ein Kreuz mit der Stochastik!“

- (a) Hat Jan richtig gerechnet? [8]
- (b) Was sagt die  $1\sigma$ -Regel denn überhaupt, und woher kommt sie? [4+]
- (c) Nimm Stellung zu Jans Problem. [0+]

### 4. Nagelfabrik

[12]

In einer Nagelfabrik werden Nägel der Länge 40mm hergestellt. Im Messprotokoll der Maschine steht  $\bar{x} = 40.2\text{mm}$  und  $s = 0.15\text{mm}$ .

- (a) Welcher Anteil der Fertigung hat weniger als die verlangten 40mm Länge?
- (b) Wie muss  $\bar{x}$  eingestellt werden, damit nur 2% der Fertigung Minderlänge aufweist?

### 5. Radioaktiver Zerfall

Wir betrachten ein Teilchen eines bestimmten radioaktiven Stoffs. Der Zeitpunkt, zu dem das Teilchen zerfällt, ist eine Zufallsgröße  $T$ . Bei geeigneter Wahl der Einheiten ist die Dichtefunktion von  $T$  die Funktion  $f$  mit

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

- (a) Zeichne eine Skizze des Graphen von  $f$ . [3]
- (b) Für welches  $t$  ist  $P(T \leq t) = \frac{1}{2}$ ? [8]
- (c) Berechne  $P(T > t)$  für  $t \geq 0$ . [Kontrollergebnis:  $e^{-t}$ ] [4]
- (d) Es seien  $s, t > 0$ . Berechne  $P(T > s)$ ,  $P(T > t + s)$  und die bedingte Wahrscheinlichkeit  $P(T > t + s | T > t)$  und erlaüttere am Graphen, was diese Werte anschaulich bedeuten. [10]

## 11 Beurteilende Statistik

### 11.1 Einstiegsbeispiel

Ein Biologe fängt in der Nordsee 53 Exemplare einer Krebsart mit unaussprechlichem Namen, und er stellt fest, dass 30 davon männliche Tiere sind.<sup>17</sup> Kann die hohe Anzahl männlicher Tiere in der Stichprobe Zufall sein oder sollte man eher annehmen, dass das Geschlechterverhältnis bei dieser Krebsart nicht 1:1 ist?

Joschka entwarf ein sauberes Testproblem mit vollständigem Apparat. Er nimmt an, aus der sehr großen Grundgesamtheit  $\Omega$  aller Krebse dieser Art werden 53 Exemplare zufällig gewählt, und die Zufallsgröße  $X$  ist die Anzahl der männlichen Tiere in der Stichprobe. Den Anteil der männlichen Tiere sieht er als  $\frac{1}{2}$  an, das heißt, er bildet die Nullhypothese

$$H_0 : p = \frac{1}{2}$$

und fragt, ob das beobachtete Stichprobenergebnis  $x = 30$  mit dieser Annahme verträglich ist. Dazu berechnet er den  $p$ -Wert dieses Ergebnisses, das ist die Wahrscheinlichkeit

$$P(X \geq 30) \quad ,$$

das das beobachtete Ergebnis eintritt oder ein noch extremeres. Hier können wir  $X$  ohne Weiteres als binomialverteilt annehmen, und der  $p$ -Wert ist

$$\begin{aligned} P(X \geq 30) &= \sum_{k=30}^{50} \binom{50}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{50-k} \approx \\ &\approx 1 - \Phi\left(\frac{29 + \frac{1}{2} - 53 \cdot \frac{1}{2}}{\sqrt{53 \cdot \frac{1}{2} \cdot \frac{1}{2}}}\right) \approx 1 - \Phi(0.824) \approx 1 - 0.795 = 0.205 \quad . \end{aligned}$$

Das heißt, ein Ergebnis wie das beobachtete erhält man etwa bei jeder fünften Untersuchung, wenn das Geschlechterverhältnis ausgeglichen ist. Niemand wird bei dem Ergebnis die Nullhypothese ablehnen.

Das heißt keineswegs, dass der Test die Nullhypothese bestätigt oder gar bewiesen hätte; man kann die Nullhypothese nicht ablehnen, Schluss. So ist die Logik eines Hypothesentests. Wir kauen das noch ein paarmal durch, aber du solltest es dir möglichst gleich korrekt einprägen.

---

<sup>17</sup>Aus Kersting und Wakolbinger: Elementare Stochastik. Birkhäuser Verlag, Basel 2008

## 11.2 Zweites Beispiel

Beim Känguruwettbewerb muss jeder Teilnehmer der Q1 bei dreißig Fragen jeweils die richtige von fünf Antworten herausfinden. Wer nun nicht ernsthaft teilnimmt, sondern bei jeder Frage aufs Geratewohl eine Antwort ankreuzt, verschafft sich für zwei Euro zwei Freistunden, und ein kleines Spiel erhält er obendrein. Solche Tunichtgute will der strenge Stufenleiter natürlich entlarven. Um den Kreis der Verdächtigen einzugrenzen, sollen erst einmal die Teilnehmer aussortiert werden, die ernsthaft gearbeitet haben, und die restlichen Teilnehmer dann gründlicher unter die Lupe genommen werden – man kennt ja die üblichen Verdächtigen. Ab wievielen richtigen Antworten soll der Teilnehmer als ernsthaft angesehen werden?

Der Stufenleiter beschließt, einen Teilnehmer bei mindestens zehn richtigen Antworten als ernsthaften Teilnehmer anzusehen. Behandle die folgenden Fragen.

1. Mit welcher Wahrscheinlichkeit besteht ein Tunichtgut den Test?
2. Ein schwacher ernsthafter Teilnehmer beantwortet jede Frage mit der Wahrscheinlichkeit  $\frac{2}{5}$  richtig. Mit welcher Wahrscheinlichkeit bleibt er in dem Verdacht, Tunichtgut zu sein?
3. Übersetze alles in Testchinesisch: Was ist  $H_0$ , was ist  $X(\Omega)$ , was ist  $V$ ? Welche Sorte Zufallsgröße ist  $X$ ?
4. Bestimme einen Term der Gütefunktion

$$g : p \mapsto P(X \in V)$$

des Tests und lasse dir den Graphen der Funktion zeichnen. Beurteile anhand des Graphen, wie gut der Test arbeitet.

5. Ändere den Test so ab, dass ein Tunichtgut maximal mit der Wahrscheinlichkeit 5% unter die Guten gerechnet wird, und erstelle die Gütefunktion des neuen Tests. Wie geht der neue Test mit dem schwachen ernsthaften Teilnehmer um? Ab welchem  $p$  wird ein Teilnehmer mit ziemlicher Sicherheit unter die Guten gerechnet?
6. Welche Fachbezeichnung hat die Schranke 5% der vorigen Teilaufgabe? Welche Fehler können bei der Anwendung des Tests begangen werden?
7. Das Känguruzentrum in Adelaide gibt an, dass auch ein nur mässig talentierter Q1-Schüler jede Frage mindestens mit der Wahrscheinlichkeit  $\frac{2}{5}$  richtig beantwortet. Konstruiere einen Test, der Tunichtgute ertappt, ehrliche Teilnehmer aber höchstens mit 1% Wahrscheinlichkeit zum Tunichtgut erklärt, und berechne die Wahrscheinlichkeit, dass ein Tunichtgut nicht erkannt wird.

**Ergebnisse:** Das  $X$  ist  $B(30, p)$ -verteilt, und  $p$  ist die Wahrscheinlichkeit des Teilnehmers, eine Frage richtig zu beantworten. Natürlich ist

$$X(\Omega) = \{0, 1, 2, \dots, 30\} \quad \text{und} \quad V = \{10, 11, \dots, 30\},$$

und die Gütefunktion hat den Term

$$g(p) = P(X_p \geq 10) = 1 - P(X_p \leq 9) \quad .$$

Den Graphen zeigt Abbildung 16 auf Seite 38.

Ein ordentlicher Q1-Teilnehmer beantwortet jede Frage mindestens mit der Wahrscheinlichkeit  $p \geq \frac{2}{5}$  richtig. Wir gehen also nun von

$$H_0 : p \geq \frac{2}{5}$$

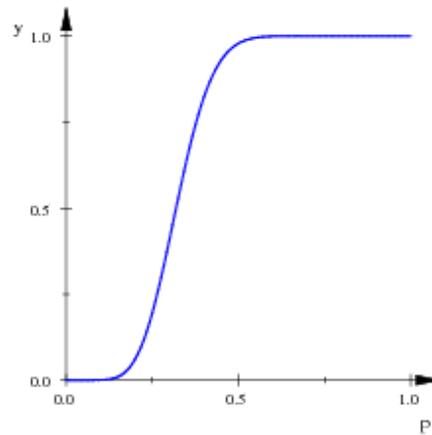


Abbildung 16: Gütefunktion zu  $V = \{10, 11, \dots, 30\}$

aus. Der Verwerfungsbereich des Tests muss die Form  $V = \{0, 1, \dots, k\}$  haben, und es muss

$$P\left(X_p \in V \mid p \geq \frac{2}{5}\right) \leq 0.01$$

sein. Nun ist

$$P\left(X_p \leq k \mid p \geq \frac{2}{5}\right) \leq P\left(X_{\frac{2}{5}} \leq k\right) .$$

Wegen<sup>18</sup>  $P(X_{\frac{2}{5}} \leq 5) \approx 0.00566$  und  $P(X_{\frac{2}{5}} \leq 6) \approx 0.0172$  müssen wir  $k = 5$  nehmen. Folglich ist der maximale Verwerfungsbereich  $V = \{0, 1, 2, 3, 4, 5\}$ , und der Graph der Gütefunktion  $p \mapsto P(X_p \leq 5)$  sieht so aus:

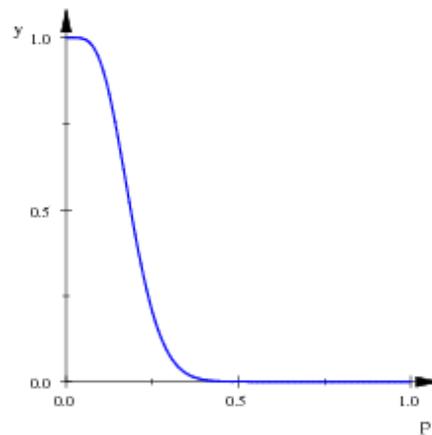


Abbildung 17: Gütefunktion zu  $V = \{0, 1, \dots, 5\}$

<sup>18</sup>Die Wahrscheinlichkeit  $P(X \leq k)$  für  $B(n, p)$ -verteiltes  $X$  bekommst du mit MuPAD durch `stats::binomialCDF(n,p)(k)`. Das C steht für „kumuliert“, und der Befehl ist aus dem Paket `stats`. Die Rechnungen und die Befehle für die Bilder findest du in der Datei `guetefkt.mn`.

### 11.3 Der Apparat eines Hypothesentests

Im Kern haben wir es immer mit einer  $B(n, p)$ -verteilten Zufallsgröße  $X$  zu tun. Der Wert von  $n$  ist bekannt, der Wert von  $p$  jedoch nicht, und mit Hilfe des Tests möchte man etwas über  $p$  herausfinden. Dazu stellt man eine Hypothese  $H_0$  auf, die sogenannte **Nullhypothese**. Dabei handelt es sich stets um eine Aussage über  $p$ , zum Beispiel

$$H_0 : p = \frac{1}{6} \quad \text{oder} \quad H_0 : p \leq \frac{1}{3} .$$

Im Erfolgsfall führt der Test zur Ablehnung von  $H_0$ , über andere Verläufe redet man nicht viel. Das muss man wissen, sonst formuliert man keine brauchbare Nullhypothese. Hat man zum Beispiel den Verdacht, dass ein Würfel zu viele Sechsen wirft oder dass der Ausschussanteil einer Lieferung größer ist als der vereinbarte Anteil  $\frac{1}{6}$ , stellt man die Hypothese

$$H_0 : p \leq \frac{1}{6}$$

auf und hofft darauf, dass die beobachtete Anzahl von Sechsen bei  $n$  Würfeln oder die beobachtete Anzahl unbrauchbarer Stücke unter  $n$  geprüften hoch ist. Dazu legt man einen **Verwerfungsbereich**  $V$  fest, das ist die Menge der Werte von  $X$ , bei deren Eintreten man  $H_0$  ablehnen will. Die Größe von  $V$  hängt von der Wahrscheinlichkeit ab,  $H_0$  zu Unrecht abzulehnen, die man tolerieren will. In der Regel darf diese Irrtumswahrscheinlichkeit eine vorgegebene Schranke  $\alpha$  nicht überschreiten:

$$P(X \in V | H_0) \leq \alpha \tag{19}$$

Wir haben gelernt, zu  $H_0$  und  $n$  ein möglichst großes  $V$  so zu bestimmen, dass diese Ungleichung erfüllt ist.

Wie gut der Test arbeitet, sieht man an der **Gütefunktion**

$$g : p \mapsto P(X_p \in V) \quad \text{für } B(n, p)\text{-verteiltes } X . \tag{20}$$

Sie gibt für jedes  $p \in [0, 1]$  die Wahrscheinlichkeit an, dass eine Lieferung mit dem Ausschussanteil  $p$  abgelehnt wird. Ist  $H_0$  wahr, ist  $g(p)$  eben die Wahrscheinlichkeit, dass die Lieferung zu Unrecht abgelehnt wird (**Fehler erster Art**). Ist  $H_0$  falsch, ist  $1 - g(p)$  die Wahrscheinlichkeit dafür, dass eine Lieferung mit dem (zu hohen) Ausschussanteil  $p$  dennoch angenommen wird (**Fehler zweiter Art**).

Häufig muss man zu vorgegebenem  $\alpha$  ein zur Situation passendes  $H_0$  formulieren und  $V$  dazu bestimmen, oft noch zusätzlich ausrechnen, mit welcher Wahrscheinlichkeit eine Lieferung mit einem zu hohen Ausschussanteil  $p$  noch angenommen wird. Sucht man  $V$  zu  $H_0 : p \leq \frac{1}{6}$  bei gegebenem  $\alpha$ , rechnet man mit dem Extremfall  $p = \frac{1}{6}$ , denn das  $V$  muss bei der Konstellation rechts auf der Merkmalsachse liegen, und dann ist  $P(X_p \in V)$  umso größer, je größer  $p$  ist.

Du hast nun einige Beispielaufgaben gerechnet und im Buch sind Tests auch ganz ordentlich behandelt, deshalb solltest du nun mit den Tests auf vertrautem Fuße stehen. Stochastikaufgaben im Abitur enthalten immer etwas Wahrscheinlichkeitsrechnung und als Krönung etwas zu Hypothesentests. Wenn du dir Lücken gönnen willst, sollten die nicht gerade an der Stelle sein, das könnte sonst unangenehme Folgen haben. Und heute geschieht vielerorts „Wahrheitsfindung“ mit Hilfe von Tests; da schadet es nichts, wenn du etwas davon verstehst, auch ganz unabhängig vom Abitur.

## 12 Klausur Nr. 4 (Nachschreibausgabe)

### 1. Einmal längs durch die Stochastik

Ein Glücksrad ist in drei Sektoren eingeteilt, die Winkel sind  $\alpha_1 = 180^\circ$ ,  $\alpha_2 = 120^\circ$  und  $\alpha_3 = 60^\circ$ . Mit Hilfe dieses Glücksrads werden die Werte  $x_1 = 0$ ,  $x_2 = 1$  und  $x_3 = 2$  einer Zufallsgröße  $X$  ausgelost.

- Zeichne das Glücksrad.
- Zeichne das Histogramm von  $X$  und den Graphen der Verteilungsfunktion.
- Berechne Erwartungswert und Varianz von  $X$ .
- Das Glücksrad wird achtzigmal gedreht, und  $S$  ist die Summe der ausgelosten Zahlen. Berechne  $P(S = 0)$  und gib an, welche Werte  $S$  annehmen kann.
- Wir bleiben bei dem  $S$  aus der vorigen Teilaufgabe. Damit du die Macht der Theorie, die du gelernt hast, richtig einschätzt: du hast überhaupt keine Chance,  $P(S = k)$  anzugeben für einen beliebigen Wert  $k$  von  $S$ . Aber mit Hilfe unserer Regeln kannst du  $E(S)$  und  $\text{Var}(S)$  bestimmen und dann mit Hilfe des Zentralen Grenzwertsatzes Näherungswerte für  $P(S = 130)$  und für  $P(S \leq 140)$  berechnen. Führe das so durch, dass man den Einsatz der Theorie sehen kann, schreibe also nicht nur Ergebnisse hin.

### 2. Technisches

Es sei  $X$  eine  $B(n, \frac{1}{5})$ -verteilte Zufallsgröße. Der Wert von  $n$  ist in jeder Teilaufgabe angegeben.

- Gib ein Beispiel für eine solche Zufallsgröße.
- Berechne  $P(X = n)$  für  $n = 10$ .
- Berechne  $P(X = E(X))$  und  $P(X \leq 95)$  für  $n = 500$ .
- Bestimme für  $n = 500$  ein  $k$  so, dass  $P(X \leq k) \approx 0.99$  ist, und formuliere das Ergebnis der Aufgabe in allgemeinverständlicher Sprache.
- Es sei jetzt  $n = 100$  und  $p$  variabel. Wie groß wird

$$P(X \geq 49 \mid p \leq 0.4)$$

höchstens?

### 3. Testaufgabe

Eine Großbäckerei stellt an zwei Standorten Laugenbrezeln her. Eine schwäbische Laugenbrezel sollte nach alter Sitte mindestens 62 Gramm haben, sonst gilt sie als untergewichtig.<sup>19</sup>

- An der Brezelmaschine im Standort Pringsheim steht<sup>20</sup> heute  $\mu = 63.4 [g]$  und  $\sigma = 0.65 [g]$ . Gib Auskunft, was diese Werte bedeuten, und berechne daraus den Anteil der untergewichtigen Brezeln in Pringsheim an diesem Tag.
- Mit einem Großkunden wurde vereinbart, dass höchstens 10% der Brezeln einer Lieferung untergewichtig sein dürfen. Finden sich in einer Stichprobe von fünfzig Brezeln mehr als sieben untergewichtige Brezeln, wird die ganze Lieferung abgelehnt.

<sup>19</sup>In einer neuzeitlichen Klausur kommt jetzt eine Internetadresse als Quelle, aber ich habe mir die Daten einfach ausgedacht.

<sup>20</sup>Eigentlich steht da  $\bar{x}$  und  $s$ , aber damit bist du nicht so vertraut, deshalb bleiben wir bei  $\mu$  und  $\sigma$ .

- i. Welches Risiko hat die Bäckerei, dass eine Lieferung abgelehnt wird, obwohl sie in Ordnung ist?
- ii. Welches Risiko hat der Kunde, dass eine Lieferung vom anderen Standort mit 20% untergewichtigen Brezeln angenommen wird?
- iii. Verwende die errechneten Daten und skizziere grob den Graphen der Gütefunktion des Tests.

#### 4. Gütefunktionen

Hier siehst du drei Graphen von Gütefunktionen. Sie gehören zu drei Tests zum gleichen  $H_0$  und zum gleichen Signifikanzniveau, aber mit verschiedenen Stichprobenlängen. Du sollst der Zeichnung einige Informationen entnehmen. Gib nicht nur die Daten an, es muss zu sehen sein, wie du sie gewonnen hast.

- (a) Wie groß ist das Signifikanzniveau ungefähr?
- (b) Gib  $H_0$  so gut an, wie du es der Zeichnung entnehmen kannst.
- (c) Welche Kurve gehört zur größten Stichprobenlänge, welche zur kleinsten? Begründe deine Antwort!

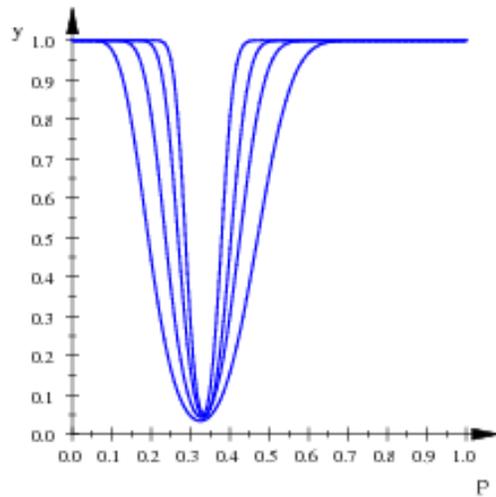


Abbildung 18: Drei Graphen von Gütefunktionen

## 13 Schätzen einer unbekannt Wahrscheinlichkeit

Hinweis: Dieser Abschnitt ist aus dem Skript eurer Vorgänger geklaut.

### 13.1 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

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

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  $\alpha$ . 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 der Gaußfunktion 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  $\alpha$ , 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.

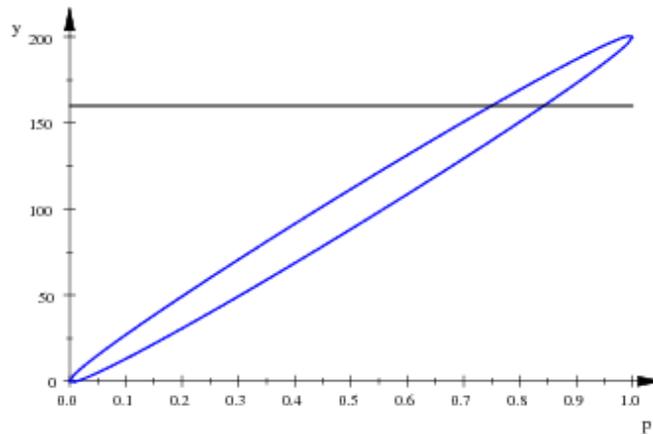


Abbildung 19: Hier siehst du die Kurven  $y = 200p \pm \Phi^{-1}(0.95)\sqrt{200p(1-p)} = \mu \pm \Phi^{-1}(0.05) \cdot \sigma$  angetragen. Bei jedem  $p$  liegt zwischen den Kurven genau der Bereich mit den Werten des  $B(200, p)$ -verteilten  $X$ , die zusammen etwa die Wahrscheinlichkeit 90% haben. 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.

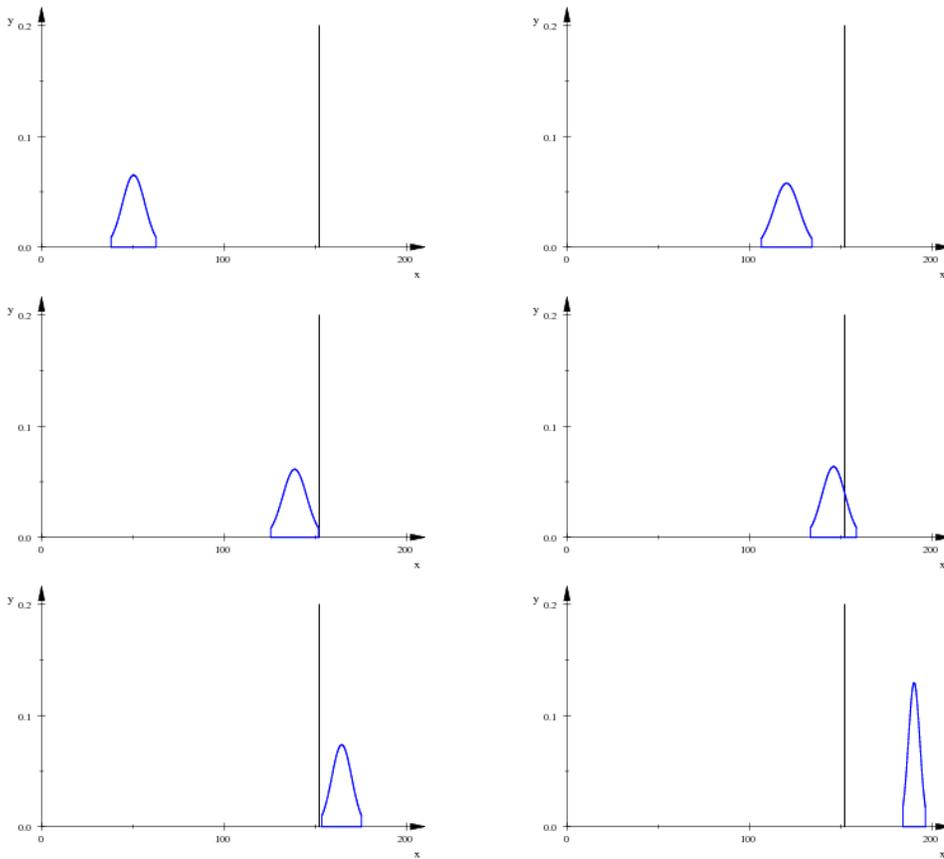
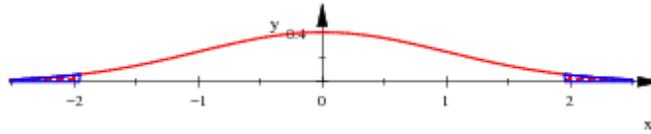


Abbildung 20: Der oberste Graph gehört zur Gaußfunktion  $\varphi$ . Der Inhalt der (vollständigen) Schwänze links und rechts ist jeweils 0.025, die schneiden wir weg. Das Mittelstück hat dann den Flächeninhalt 0.95. Dieses gestutzte Mittelstück destandardisieren wir und erhalten den Mittelteil von Histogrammen 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.

### 13.2 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  $\alpha$  vorgeben, die er in Kauf zu nehmen bereit ist. Wir berechnen uns zu diesem  $\alpha$  den Wert

$$a := \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) , \quad (21)$$

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  $\mu$  höchstens  $a\sigma$  von  $k$  entfernt:

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

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

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

Umformen ergibt, wenn ich richtig gerechnet habe,

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

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 21). Die  $\sigma$ -Werte unterscheiden sich etwas, und zwar um so mehr, je näher  $\hat{p} := \frac{k}{n}$

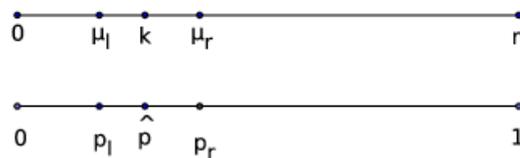


Abbildung 21:  $k$ ,  $\mu_l = np_l$  und  $\mu_r = np_r$

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**

$$\left[ \hat{p} - a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} , \hat{p} + a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] . \quad (22)$$

**Beispiel.** Um herauszufinden, welcher Anteil der Bevölkerung Espelkamps einen Vaternatagsausflug 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

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

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

### 13.3 Was bedeutet die Irrtumswahrscheinlichkeit $\alpha$ ?

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.<sup>21</sup>

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!

---

<sup>21</sup>Sieh dir dazu die Abbildung 22 auf Seite 47 an und meditiere ein paar Minuten darüber!

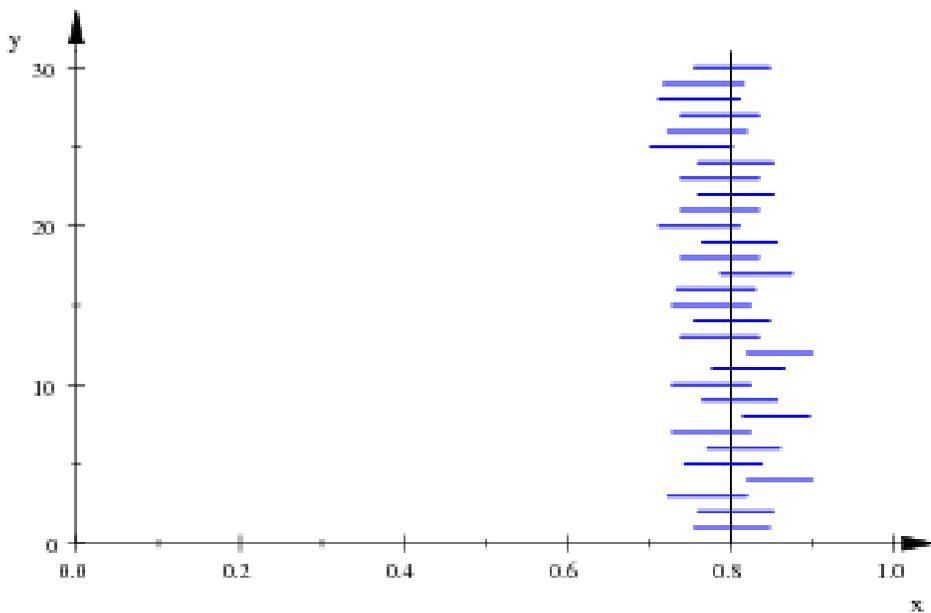


Abbildung 22: Zur Bedeutung von  $\alpha$  beim Konfidenzintervall: Hier wurden dreißig Werte der  $B(200, \frac{4}{5})$ -verteilten Zufallsgröße  $X$  ausgelost und zu jedem Wert das Konfidenzintervall zur Irrtumswahrscheinlichkeit  $\alpha = 0.1$  berechnet. In dem Bild sind die dreißig Konfidenzintervalle 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 Irrtumswahrscheinlichkeit sagt nur etwas über die Güte des Verfahrens und nichts über das konkrete Intervall!