Stochastik

LKAbi 2011 (B. Waldmüller)

7. Juli 2010

Inhaltsverzeichnis

1 Grundbegriffe

Stochastik ist die Mathematik des Zufalls, könnte man so sagen. Ob es Zufall überhaupt gibt, ob die Zukunft prinzipiell offen ist oder ob seit Beginn des Zeiten alles vorherbestimmt ist, ist eine große Frage – der Philosophie. Die Frage tangiert durchaus dich persönlich: Ist es Ergebnis deiner freien Entscheidung, für das du die volle Verantwortung trägst, wenn du ohne Hausaufgabe zur Stunde kommst, oder geschieht es in Vollzug unabänderlicher Bestimmung und ist insofern nur tragisch, nicht aber tadelnswert? Es wäre sehr reizvoll, solche Dinge zu bedenken, wir machen das aber nicht, jedenfalls nicht in der Hauptsache; ihr sollt ja Mathematik lernen, und die Mathematik fragt nicht danach, ob es Zufall gibt.

In unserem Stochastikalltag betrachten wir **Zufallsversuche** oder **Zufallsexperimente**, das sind Mechanismen, die Ergebnisse produzieren, die nicht vorhersagbar sind. Das heißt aber keineswegs, dass sie irgendwie beliebig wären. Die Ergebnisse müssen in einem klar anzugebenden **Ergebnisraum** Ω liegen, das ist eine Menge, und es muss völlig klar sein, auf welchem Wege das jeweilige Ergebnis $\omega \in \Omega$ ausgelost wird. Man kann zum Beispiel eine Münze werfen, da ist $\Omega_1 = \{W, Z\}$ geeignet, oder man kann einen Würfel dreimal werfen:

$$\Omega_2 = \{(a, b, c) \,|\, a, b, c \in \{1, 2, \ldots, 6\}\}$$

Wurde im ersten Wurf eine 2 und im zweiten und im dritten Wurf eine 4 geworfen, ist bei der Durchführung $(2, 4, 4)$ das ausgeloste ω.

Ein **Ereignis** A ist eine Teilmenge[1] von Ω, in Formelsprache $A \subseteq \Omega$. Spielt man Mensch–ärgere–dich–nicht, ist durchaus das Ereignis

$$A = \{\{(a, b, 6) \,|\, a, b \in \{1, 2, 3, 4, 5\}\} = \{(a, b, 6) \in \Omega_2 \,|\, a, b < 6\}$$

von Interesse; du kannst sicher die Bedeutung finden.

Ein Ereignis, das nur ein einzelnes Ergebnis enthält, nennt man **Elementarereignis**; es hat also die Gestalt $\{\omega\}$.

1.1 Zufallsgrößen

Es liege nun ein Zufallsversuch mit Ergebnisraum Ω vor. Eine **Zufallsgröße** ist dann eine Abbildung $X : \Omega \to \mathbb{R}$. Die Zufallsgröße X ordnet jedem $\omega \in \Omega$ eine Zahl $X(\omega)$ zu. Beim einleitenden Beispiel des dreifachen Würfelns könnte $X(\omega)$ die Anzahl der gewürfelten Sechsen im Ergebnis ω sein, dann ist $X(\Omega_2) = \{0, 1, 2, 3\}$.

Jede Zufallsgröße X stiftet in natürlicher Weise eine Zerlegung des Ergebnisraumes Ω in die Ereignisse

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

die zu den Werten gehören, die $X(\omega)$ annehmen kann: Jeweils alle ω mit dem gleichen X–Wert $k = X(\omega)$ kommen in einen Topf. Beim Beispiel des dreifachen Würfelns sind zwei der vier Ereignisse

$$X^{-1}(3) = \{(6, 6, 6)\} \quad \text{und} \quad X^{-1}(0) = \{(a, b, c) \in \Omega_2 \,|\, a, b, c < 6\} \quad .$$

Man verwendet dafür übrigens die Bezeichnungen $X = 3$ und $X = 0$; das ist aufreizend schlampig und missverständlich, aber sehr praktisch.

[1] Bei uns sind das erstmal ganz beliebige Teilmengen von Ω, allgemein darf man nicht so großzügig sein.

1.2 Erste Anmerkungen zur Wahrscheinlichkeit

Wir sagen, die Wahrscheinlichkeit, eine sechs zu würfeln, sei $\frac{1}{6}$. Dabei lassen wir uns von der Vorstellung leiten, dass alle sechs Seiten eines mathematischen Würfels gleichberechtigt sind, und verteilen die Gesamtwahrscheinlichkeit 1 zu gleichen Teilen auf die sechs Elementarereignisse $\{1\}, \{2\}, \ldots, \{6\}$. In gleicher Weise ist die Wahrscheinlichkeit, dass eine bestimmte von n Kugeln einer Urne gezogen wird, $\frac{1}{n}$. Die Wahrscheinlichkeit eines Ereignisses $A \subseteq \Omega$ berechnen wir dann nach der Formel

$$P(A) = \frac{|A|}{|\Omega|} \quad , \tag{1}$$

und wir sagen, unser Zufallsversuch sei ein Laplace–Versuch.

Aber was sagt das überhaupt? Wirft man einen ordentlichen Würfel sehr oft und stellt den Anteil der gewürfelten Sechsen an der gesamten Anzahl der Würfe fest, liegt dieser Anteil in der Nähe von $\frac{1}{6}$. Das ist ein **empirischer Befund**, eine Aussage über die reale Welt, und man kann ihn nicht beweisen wie man eine mathematische Aussage beweist. Man kann auch nicht beweisen, dass ein Stein, den man immer wieder aufhob und wieder zu Boden fallen ließ, auch bei den nächsten Versuchen und am nächsten Tag wieder zu Boden fallen wird, aber es ist vernünftig, davon auszugehen, dass das genau so geschehen wird.

Zu einer mathematischen Theorie wurde die Wahrscheinlichkeitsrechnung in den dreißiger Jahren des 20. Jahrhunderts. Vorher hatte man zweihundert Jahre naiv mit Wahrscheinlichkeiten gerechnet und die Erfahrung gemacht, dass die Ergebnisse dieser Rechnungen durchaus etwas über reale Phänomene aussagen. Was so großen Geistern wie Fermat, Laplace und Jakob Bernoulli recht war, sollte uns billig sein; wir treiben also auch erst einmal naive Wahrscheinlichkeitsrechnung mit Laplace–Wahrscheinlichkeiten und Baumdiagrammen. Später denken wir grundsätzlicher nach – vielleicht.

1.3 Aufgaben

Das Aufgabenmaterial im Buch ist sehr dünn. Das ist etwas lästig, aber nicht weiter schlimm. Man kann sich ja welche ausdenken:

1. Wir werfen einen Würfel so lange, bis er eine 6 zeigt. Es sei X die Anzahl der benötigten Würfe. Gib eine Formel für $P(X = k)$, berechne $P(X \leq 10)$ und bestimme ein möglichst kleines N so, dass $P(X > N) \leq 10^{-4}$ ist.

2. Deine kleine Schwester fragt dich, was du in der Aufgabe vorher eigentlich ausgerechnet hast. Gib ihr freundlich Antwort.

3. Dein großer Bruder aus der 13 fragt, wie eigentlich bei der Aufgabe oben das Ω aussieht. Kannst du ihm antworten?

4. Die Klasse 5d plant einen Würfelstand beim MuKu–Fest. Es wird einmal gewürfelt. Bei einer 6 gibt es einen Euro, bei einer 5 und bei einer 4 gibt es 20 Cent, sonst nichts. Welchen Einsatz muss die Klasse mindestens nehmen, damit sie zuversichtlich sein kann, etwas übrig zu behalten?

5. Ein Würfel wird dreimal geworfen.

   (a) Mit welcher Wahrscheinlichkeit hat man mindestens eine 4 dabei?

   (b) Mit welcher Wahrscheinlichkeit hat man mindestens eine 4 und mindestens eine 2?

(c) Es sei X die Anzahl der gewürfelten Einsen. Finde eine Formel für $P(X = k)$.

(d) Ist die Wahrscheinlichkeit, dass alle drei Augenzahlen gerade sind, $\frac{1}{2}$?

6. Es gibt ein ganz stumpfsinniges Würfelspiel, es heißt Pferderennen. Man würfelt mit einem Würfel und setzt die gewürfelte Augenzahl weiter. Gewonnen hat, wer den Parcours als Erster durchlaufen hat. Sagen wir, der Parcours ist 50 Felder lang. Wie lange dauert das Spiel im Mittel?

7. In zwei Klamottenladen darf man seinen Rabatt selbst erwürfeln. Im ersten wirft man zwei gleiche Würfel auf einmal. Die kleinere Zahl gibt die Zehnerziffer, die andere die Einerziffer des Prozentsatzes an. Im zweiten Laden wird ein Würfel zweimal geworfen. Der erste Wurf gibt die Einerziffer, der zweite die Zehnerziffer des Prozentsatzes an.

   (a) Welcher Laden bietet den höheren Rabatt? Antworte nicht voreilig! Ist das nicht Glückssache? Kann man überhaupt eine Antwort geben?

   (b) Simuliere die beiden Rabattsysteme!

8. Simuliere die Aufgabe 4.

1.4 Der Erwartungswert einer Zufallsgröße

Es sei $X : \Omega \to \mathbb{R}$ eine Zufallsgröße. Wenn man den zugrundeliegenden Zufallsversuch N–mal unabhängig wiederholt, also N Werte $x_1, x_2, \ldots, x_N$ von X auslost, wird der Durchschnittswert

$$\bar{x} := \frac{1}{N} \sum_{k=0}^{N} x_k$$

„für große N“ in der Nähe des Erwartungswertes $E(X)$ der Zufallsgröße X liegen:

1 Definition
Für eine Zufallsgröße $X : \Omega \to \mathbb{R}$ definieren wir den Erwartungswert $E(X)$ durch

$$\begin{aligned} E(X) &:= \sum_{k \in X(\Omega)} kP(X = k) \\ &= \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) \quad . \end{aligned}$$

Dir zur Warnung weise ich noch einmal darauf hin: Häufig gehört der Erwartungswert **nicht** zu den Werten, die X annehmen kann, und wenn X den Erwartungswert tatsächlich annehmen kann, ist es keineswegs immer der Wert von X mit der höchsten Wahrscheinlichkeit.

Beispiel Die Zufallsgröße X nehme nur die Werte 1 und 0 an. Es ist üblich, die Wahrscheinlichkeit $P(X = 1)$ mit p und die Wahrscheinlichkeit $P(X = 0)$ mit $q = 1 - p$ zu bezeichnen. Der Erwartungswert von X ist dann

$$E(X) = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = 1 \cdot p + 0 \cdot q = p \quad .$$

Wenn man aus einer Urne mit vier roten und sechs weißen Kugeln sehr oft mit Zurücklegen zieht, wird man je Ziehung etwa $\frac{4}{10}$ rote Kugeln gezogen haben. – Auch dies ist zunächst nur ein empirisch abgesicherter Befund.

1.5 Die Regel $E(X+Y)=E(X)+E(Y)$

Es seien X und Y Zufallsgrößen zum gleichen Zufallsversuch. Wir bilden die neue Zufallsgröße $X+Y:\Omega\to\mathbb{R}$ durch

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

und berechnen den Erwartungswert der neuen Zufallsgröße:

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

Es gilt also

$$\boxed{E(X+Y)=E(X)+E(Y)} \tag{2}$$

für Zufallsgrößen X und Y zum gleichen Zufallsversuch. Die Regel lässt sich natürlich auf Summen von mehr als zwei Zufallsgrößen zum gleichen Zufallsversuch verallgemeinern.

Beispiel. Wir würfeln n–mal und zählen die Anzahl X der Sechsen. Wenn wir für $i=1,2\ldots,n$ die Anzahl der Sechsen im i–ten Wurf mit X_i bezeichnen, ist

$$X = X_1+X_2+X_3+\cdots+X_n=\sum_{i=1}^{n}X_i \quad .$$

Jedes X_i kann nur die Werte 0 und 1 annehmen, und es ist $P(X_i=1)=\frac{1}{6}$ und $P(X_i=0)=\frac{5}{6}$. Es folgt

$$E(X_i)=\frac{1}{6} \quad \text{für } i=1,2,3,\ldots,n \quad ,$$

und daraus ergibt sich nach der gerade bewiesenen Regel

$$E(X)=n\cdot\frac{1}{6} \quad .$$

1.6 Bernoulliketten und binomialverteilte Zufallsgrößen

Ein Zufallsversuch hat im einfachsten nicht uninteressanten Fall nur zwei Ausgänge. Man nennt einen solchen Versuch **Bernoulliversuch** und bezeichnet die Ausgänge mit „Erfolg“ und „Misserfolg“ . Die Wahrscheinlichkeit eines Erfolges heißt üblicherweise p und die Wahrscheinlichkeit eines Misserfolges $q=1-p$. Du kannst zum Beispiel würfeln und eine 6 als Erfolg ansehen. Wirfst du keine 6, ist ein Misserfolg eingetreten. Hier ist $p=\frac{1}{6}$ und $q=\frac{5}{6}$.

Vermutlich findest du Bernoulliversuche nicht besonders aufregend, aber bringe noch einen Augenblick Geduld auf: Wiederholt man einen Bernoulliversuch mit Erfolgswahrscheinlichkeit p n–mal unabhängig, spricht man von einer **Bernoullikette** der Länge n mit Erfolgswahrscheinlichkeit p, und die durch die Anzahl der Erfolge gegebene Zufallsgröße X heißt **binomialverteilt**. Wir haben uns überlegt, dass die Wahrscheinlichkeit $P(X=k)$ durch

$$P(X=k)=\binom{n}{k}p^kq^{n-k} \quad \text{für } k=0,1,\ldots,n \tag{3}$$

gegeben ist; dabei ist

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdot\dots\cdot(n-k+1)}{k(k-1)(k-2)\cdot\dots\cdot 1} = \frac{n!}{k!\cdot(n-k)!} \tag{4}$$

die Anzahl der k–Teilmengen einer n–Menge.

Wir können auch gleich den Erwartungswert der Zufallsgröße X angeben: Es ist $E(X) = np$. Damit sind uns schon Aussagen über zahlreiche Anwendungen möglich; in sehr vielen Situation lassen sich nämlich Abläufe mit Hilfe von Bernoulliketten modellieren.

1.7 Der Einfluss der Länge einer Bernoullikette

Wenn man aus einer Urne mit vier weißen und sechs roten Kugeln N–mal mit Zurücklegen zieht und die Anzahl der gezogenen weißen Kugeln betrachtet, hat man es mit einer binomialverteilten Zufallsgröße X mit Erfolgswahrscheinlichkeit $p = 0.4$ zu tun. Im Mittel sollte man je Durchführung $E(X) = pN$ weiße Kugeln erhalten. Das ist schön und gut, aber was können wir einen konkreten ausgelosten Wert von X sagen? Wie stark darf der vom Erwartungswert abweichen, ohne dass uns die Sache verdächtig vorkommen sollte?

Wir schauen uns zunächst Histogramme für einige N–Werte an – siehe Abbildung 1 auf Seite 7. Lasse dich nicht täuschen, die Histogramme werden immer breiter und flacher; das siehst du an den Werten auf der Hochachse. Dennoch wird der Bereich von Werten, die nur eine sehr kleine Wahrscheinlichkeit haben, relativ immer größer. Diesem Phänomen gehen wir im nächsten Abschnitt nach.

Alle Histogramme zeigen Wahrscheinlichkeiten, die etwa bis zum Erwartungswert ansteigen und danach wieder kleiner werden. Wie wir im Unterricht bewiesen haben, gilt dies allgemein für binomialverteilte Zufallsgrößen: Es gilt nämlich

$$P(X = k) < P(X = k+1) \Longleftrightarrow k < np - q \quad . \tag{5}$$

1.8 Varianz und Tschebyschew–Ungleichung

Die Form der Histogramme in Abbildung 1 passt zu unserer Vorstellung, dass die ausgelosten Werte einer binomialverteilten Zufallsgröße X um den Erwartungswert schwanken sollten. Dies schauen wir uns nun genauer an. Ein Maß dafür, wie stark die Werte einer Zufallsgröße streuen, ist die Varianz der Zufallsgröße.

2 Definition
Die **Varianz** $V(X)$ *einer Zufallsgröße* X *mit dem Erwartungswert* $\mu = E(X)$ *ist definiert durch*

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

Die Größe $\sigma := \sqrt{V(X)}$ *heißt* **Standardabweichung**.

Ich habe versucht, euch die Varianz etwas anschaulich zu machen; dies lasse ich hier aus. Ferner haben wir die Ungleichung von Tschebyschew hergeleitet. Sie macht eine Aussage darüber, wie groß die Wahrscheinlichkeit sein kann, dass die Werte einer Zufallsgröße X sich vom Erwartungswert $\mu = E(X)$ um mindestens $a > 0$ unterscheiden. Man gewinnt sie durch zwei sehr rabiate Abschätzungen. Zunächst lässt man in der Summe, mit der man $V(X)$ berechnet, alle k mit $|k - \mu| < a$ weg. dadurch wird die Summe in der Regel kleiner, jedenfalls nicht größer. Bei den

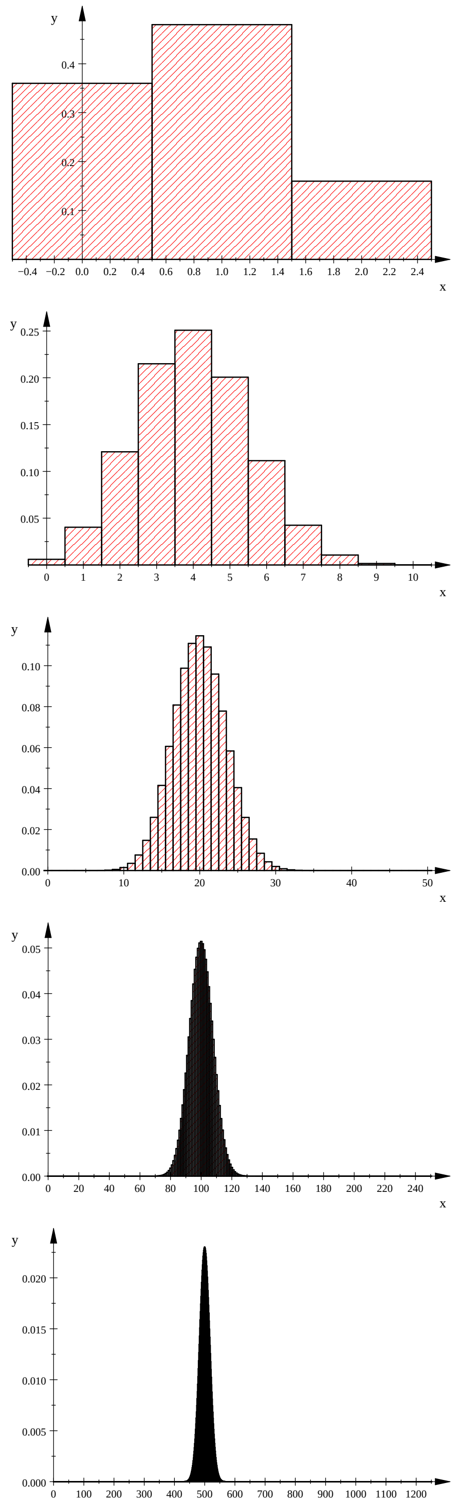


Abbildung 1: Histogramme binomialverteilter Zufallsgrößen mit $p = 0.4$ für die N–Werte $2, 10, 50, 125$ und $N = 1250$

verbleibenden k ersetzt man $(k-\mu)^2$ durch a^2. Dadurch werden die Summanden in der Regel kleiner, jedenfalls aber nicht größer. Insgesamt erhält man:

$$\begin{aligned}
V(X) &= \sum_{k\in X(\Omega)} (k-\mu)^2 P(X=k) \\
&\geq \sum_{k\in X(\Omega),|k-\mu|\geq a} (k-\mu)^2 P(X=k) \\
&\geq \sum_{k\in X(\Omega),|k-\mu|\geq a} a^2 P(X=k) \\
&= a^2 \sum_{k\in X(\Omega),|k-\mu|\geq a} P(X=k) \\
&= a^2 P(|X-\mu|\geq a)
\end{aligned}$$

Dividiert man noch durch a^2, gewinnt man daraus die **Tschebyschew–Ungleichung**:

$$\boxed{P(|X-\mu|\geq a) \leq \frac{V(X)}{a^2}} \tag{6}$$

1.9 Regeln für Varianz und Erwartungswert

Für den Erwartungswert einer Zufallsgröße gelten einige Rechenregeln. Man prüft sie leicht anhand der Definition des Erwartungswertes nach.

3 Lemma
Es sei $X:\Omega\to\mathbb{R}$ eine Zufallsgröße, und es seien a und b reelle Zahlen. Dann gilt

$$E(aX) = aE(X) \quad \text{und} \quad E(X+b) = E(x)+b \quad .$$

Mache dir gründlich klar, dass $X+b$ eine neue Zufallsgröße auf Ω ist, und zwar ist $(X+b)(\omega) = X(\omega)+b$.

Entsprechende Regeln gelten für die Varianz:

4 Lemma
Es sei $X:\Omega\to\mathbb{R}$ eine Zufallsgröße, und es seien a und b reelle Zahlen. Dann gilt

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

Wir brauchen die Varianz binomialverteilter Zufallsgrößen. Man bekommt sie mit dem gleichen Kunstgriff, den wir bei der Berechnung des Erwartungswertes angewandt haben: Wiederholt man einen Bernoulliversuch mit der Erfolgswahrscheinlichkeit p n–mal und bezeichnet man die Anzahl der Erfolge bei der k–ten Durchführung mit X_k, dann gilt für die Anzahl X der Erfolge

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

Für jedes $k = 1,2,\ldots,n$ ist $V(X_k) = pq$, wie man leicht nachrechnet. Mit Hilfe des nachfolgenden Satzes können wir nun die Varianz einer binomialverteilten Zufallsgröße X mit den Parametern n und p angeben:

$$\boxed{V(X) = npq} \tag{7}$$

5 Satz
Es seien X und Y unabhängige Zufallsgrößen auf dem Ergebnisraum Ω, es sei also

$$P_{X=x}(Y=y) = P(Y=y) \quad \text{für alle } x \in X(\Omega) \text{ und für alle } y \in Y(\Omega) \ .$$

Dann gilt:

$$\boxed{V(X+Y) = V(X) + V(Y)}$$

Der Beweis dieses Satzes ist für uns gut erreichbar, aber seine Durchführung frisst Zeit, die ich lieber darauf verwende, deine Vertrautheit mit Zufallsgrößen und ihren Eigenschaften zu verbessern. Ich lege dir dringend ans Herz, immer genau zu durchdenken, wie der Zufallsversuch abläuft, über den du etwas sagen willst, die Zufallsgröße sorgfältig zu definieren, die du verwendest, und ihre Natur zu klären. Und dann: denke in Histogrammen. Viele Routinen der Beurteilenden Statistik werden dann durchsichtig und einfach. Mache Gebrauch von den MuPAD–Befehlen, die ich dir zum Zeichnen von Histogrammen und zur Berechnung von Wahrscheinlichkeiten formuliert habe. Furcht vor Stochastik ist gänzlich unbegründet.

Eine **Anmerkung** zur Unabhängigkeit von Zufallsgrößen: Es darf keinen Einfluß auf die Wahrscheinlichkeit haben, dass Y einen Wert y annimmt, welcher Wert der Zufallsgröße ausgelost wurde. Da wir unseren Bernoulliversuch unabhängig wiederholen, hat das Ergebnis einer Durchführung keinen Einfluß auf die Ergebnisse der anderen Durchführungen; wir durften den Satz auf die Summe der X_k anwenden.

Aufgabe Es sei X eine Zufallsgröße. Dann ist $X+X = 2X$. Berechne die Varianz von $X+X$ und die von $2X$ nach den Regeln; du wirst sehen, dass die Ergebnisse verschieden sind. Die Summenregel für die Varianz darf man nur bei unabhängigen Summen verwenden, und X ist keineswegs unabhängig von sich selbst.

1.10 Das schwache Gesetz der großen Zahl

Es sei X die Anzahl der Erfolge bei einer Bernoullikette der Länge n mit Erfolgswahrscheinlichkeit p. Dann ist, wie wir wissen, $E(X) = np$ und $V(X) = npq$. Wir betrachten die Zufallsgröße $\frac{1}{n}X$, also den Anteil der Versuche, die „Erfolg“ ergaben, an der Gesamtzahl der Versuche oder die relative Häufigkeit von „Erfolg“ . Nun ist nach unseren Regeln

$$E(\frac{1}{n}X) = \frac{1}{n}E(X) = \frac{1}{n}np = p$$

und

$$V(\frac{1}{n}X) = \frac{1}{n^2}V(X) = \frac{1}{n^2}npq = \frac{pq}{n} \quad .$$

Für jedes $a > 0$ ergibt nun die Tschebyschew–Ungleichung

$$P\left(|\frac{1}{n}X - p| \geq a\right) \leq \frac{pq}{na^2} \xrightarrow{n\to\infty} 0 \quad .$$

Siehst du, was das heißt? Für jedes noch so kleine $a > 0$ wird die Wahrscheinlichkeit, dass $\frac{1}{n}X$ von p um mindestens a abweicht, beliebig klein, wenn man nur das n groß genug macht. Die ist das **schwache Gesetz der großen Zahl**.

Der Deutlichkeit halber bringe ich ein **Beispiel**. Wir wollen n–mal würfeln und die Anzahl X der Sechsen zählen. Für einen realen Würfel sagt das empirische Gesetz der großen Zahl, dass der Anteil der Sechsen an den gewürfelten Zahlen nahe bei $\frac{1}{6}$ liegt. Unsere Zufallsgröße X aber ist kein reales Objekt, sondern ein Wesen aus der Wahrscheinlichkeitstheorie, von dem man behaupten kann, dass es nur in den Köpfen der Mathematiker existiere, ohne dass man jemand diese Behauptung schlüssig widerlegen könnte. Und dass dieses Wesen sich einen Deut um die Empirie schert, ist alles andere als sicher.

Wir fragen uns, wie oft man würfeln muss, damit $\frac{1}{n}X$ mit einer Wahrscheinlichkeit von höchstens $\frac{1}{1000}$ um $\frac{1}{100}$ oder mehr vom Erwartungswert $\frac{1}{6}$ abweicht. Wir machen das Tschebyschew–Risiko kleiner als $\frac{1}{1000}$, dann ist die Wahrscheinlichkeit selbst erst recht kleiner als dieser Wert. Dies führt auf den Ansatz

$$\frac{V(\frac{1}{n}X)}{\left(\frac{1}{100}\right)^2} = \frac{10^4 pq}{n} \leq \frac{1}{1000} \quad ,$$

und dies formen wir um zu

$$n \geq 10^7 pq = \frac{5 \cdot 10^7}{36} \quad .$$

Ergebnis: Wenn wir 1388889–mal oder öfter würfeln, werden wir mit einer Wahrscheinlichkeit von höchstens $\frac{1}{1000}$ beobachten, dass der Anteil der Sechsen um $\frac{1}{100}$ oder mehr von $\frac{1}{6}$ abweicht.

Das n, das wir hier ausgerechnet haben, ist viel zu groß, aber das stört uns im Moment noch nicht. Wir stellen zufrieden fest, dass sich unser $\frac{1}{n}X$ so verhält, wie es das empirische Gesetz der großen Zahl für reales Würfeln aussagt. Das macht uns zuversichtlich, dass die Wahrscheinlichkeitstheorie Modelle liefert, mit denen wir brauchbare Aussagen über reale Vorgänge machen können.

Anmerkung Das schwache Gesetz der großen Zahl sagt nicht, dass die Abweichung auch kleiner als die vorgegebene Schranke bleibt, wenn wir n weiter vergrößern. Das $\frac{1}{n}X$ kann noch erheblich schwanken. Aber das ist eine schwierige Kiste.

2 Klausur Nr. 3 am 11. März 2010

Vergiss nicht, dein X zu deklarieren und seine Eigenschaften zu formulieren, bevor du losrechnest!

1. In dieser Aufgabe benutzen wir eine Urne mit drei weißen und fünf schwarzen Kugeln. Aus dieser Urne ziehen wir mehrfach, die gezogenen Kugeln legen wir nicht zurück.

   (a) Mit welcher Wahrscheinlichkeit

      i. ist die erste gezogene Kugel weiß?
      ii. ist die erste Kugel schwarz und die zweite weiß?
      iii. erhält man bei drei Ziehungen nur schwarze Kugeln?
      iv. ist die zweite gezogene Kugel weiß?
      v. hat man bei drei Ziehungen genau zwei schwarze Kugeln?

   (b) Man zieht so lange, bis man eine schwarze Kugel gezogen hat. Es sei X die Anzahl der nötigen Ziehungen. Berechne $P(X = k)$ für die Werte, die X annehmen kann, zeichne ein Histogramm und berechne den Erwartungswert von X.

   (c) Dein kleiner Bruder aus der 10 fragt dich, was der Wert $E(X)$ bedeutet, den du eben ausgerechnet hast. Schreibe auf, was du ihm sagst.

   (d) Nehmen wir an, es sei zweimal gezogen worden und dabei sei genau eine schwarze Kugel gezogen worden. Mit welcher Wahrscheinlichkeit war die zweite gezogene Kugel schwarz?

(e) Erkläre, was die Wahrscheinlichkeit, die du gerade berechnet hast, eigentlich bedeutet.

(f) Wir ziehen viermal. Klaus behauptet, der Erwartungswert der Anzahl der bei der k–ten Ziehung gezogenen schwarzen Kugeln sei $\frac{5}{8}$ für jedes k, $k = 1, 2, 3, 4$.

i. Stimmst du ihm zu?

ii. Wie groß ist der Erwartungswert der Anzahl der insgesamt gezogenen schwarzen Kugeln, wenn Klaus Recht hat?

2. Langjährige Beobachtungen des Schülers Erik ergaben, dass Erik seine Hausaufgaben nur in 30% der Fälle erledigt – zum Glück haben wir solche ja nicht in unserem Kurs! Es sei X die Anzahl unter 10 zufällig gewählten Stunden, in denen Erik mit gemachten Hausaufgaben erscheint.

(a) Nimm an, X sei binomialverteilt, und berechne $P(X < 2)$, $P(X = 4)$ und $P(X \geq 5)$.

(b) Siehst du Probleme, X als binomialverteilt anzunehmen?

3. „Die Anzahl der Gasteilchen in unserem Hörsaal liegt so in der Größenordnung von 10^{28}“, doziert der Physikprofessor. „In jeder Hälfte des Hörsaals sollten sich also $\frac{1}{2} \cdot 10^{28}$ Teilchen aufhalten.“
Es sei X die Anzahl der Teilchen, die sich zu einem zufällig gewählten Zeitpunkt in der linken Hälfte des Hörsaals aufhalten. Sage, was die Wahrscheinlichkeit

$$P(|X - \frac{1}{2} \cdot 10^{28}| \geq 10^{20})$$

bedeutet, und gib eine obere Schranke für diese Wahrscheinlichkeit an. Berechne, wieviel Prozent 10^{20} von $\frac{1}{2} \cdot 10^{28}$ ist, und gib Auskunft, was das alles für die Studenten im Hörsaal bedeutet.

4. Die Zufallsgröße X nehme den Wert 1 und den Wert -1 jeweils mit der Wahrscheinlichkeit $\frac{1}{2}$ an.

(a) Berechne $E(X)$, $E(X^2)$ und $E(X)^2$.

(b) Wir haben so schöne Regeln für das Rechnen mit Erwartungswerten, aber keine für $E(XY)$. Wenn du dir das Ergebnis der letzten Teilaufgabe anschaust, weißt du, warum nicht.

5. Eine Schule veranstaltet ein Schulfest, um Geld für den Bau einer Mensa einzunehmen. Der pfiffige Felix schlägt seiner Klasse vor, einen Stand einzurichten, bei dem eine Kugel aus einer Urne gezogen wird. Der Einsatz beträgt einen Euro. Wird eine der 9999 schwarzen Kugeln in der Urne gezogen, geht der Einsatz an die Klasse. Wird die eine weiße Kugel gezogen, bekommt der Spieler 10000 Euro. Felix behauptet, dass das Spiel auf lange Sicht weder den Spieler noch die Klasse benachteiligt. Dennoch könne die Klasse mit 99%–iger Sicherheit 100 Euro Gewinn machen, wenn 100 Spieler an dem Stand spielen.

(a) Sagt Felix die Wahrheit?

(b) Empfiehlst du der Klasse, den Stand einzurichten?

3 Die Näherungsformeln von de Moivre und Laplace

3.1 Hinführung

Wir werden es in nächster Zeit vor allem mit binomialverteilten Zufallsgrößen zu tun haben. Zwar kennen wir im Prinzip für binomialverteiltes X mit den Parametern n und p die Wahrscheinlichkeit

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

aber im konkreten Fall kann es sehr mühsam sein, die Wahrscheinlichkeiten zu berechnen, die wir brauchen. Für manche Wertesätze n, p haben wir Tabellen, allzuweit kommt man auch damit nicht. Deshalb lernst du nun ein neues Hilfsmittel.

Einer von euch hatte die Ahnung, dass hinter den Histogrammen binomialverteilter Zufallsgrößen immer die gleiche Kurve steckt – freilich nur, wenn n nicht zu klein und p nicht zu extrem ist. Wenn du dir zum Beispiel das Histogramm zu $n = 30$ und $p = \frac{1}{2}$ in Abbildung 2 anschaust, magst du den Eindruck gewinnen, dass an der Ahnung etwas ist.

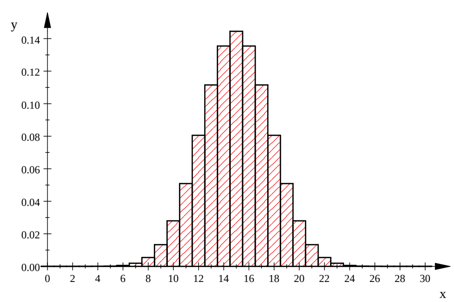


Abbildung 2: Histogramm des $B(30, \frac{1}{2})$–verteilten X

Freilich hat dir MuPAD durch die Wahl der Einheiten auf den Achsen dabei geholfen, normalerweise sieht das Histogramm so wie in Abbildung 3 gezeigt, und daran sieht man nicht viel.

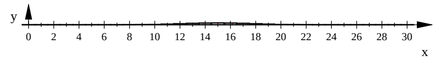


Abbildung 3: Histogramm des $B(30, \frac{1}{2})$–verteilten X mit einheitlichem Achsenmaßstab

Man muss von X zu rX übergehen für ein geeignetes $r < 1$, wenn man ein vernünftiges Histogramm bekommen will. Es hat sich herausgestellt, dass $r = \frac{1}{\sigma}$ eine gute Wahl ist, man misst dann auf der Merkmalsachse praktisch in der Einheit σ. Zusätzlich verschiebt man das Histogramm so, dass der Erwartungswert 0 ist:

6 Definition
Es sei X eine Zufallsgröße mit dem Erwartungswert μ und der Standardabweichung σ. Dann heißt die Zufallsgröße

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

die zu X gehörige standardisierte Zufallsgröße.

Die Aussage des folgenden Lemmas beweist du leicht mit Hilfe unserer Regeln.

7 Lemma
Es sei X eine Zufallsgröße mit Erwartungswert μ und Standardabweichung σ. Dann ist

$$E\left(\frac{X-\mu}{\sigma}\right) = 0 \quad \textit{und} \quad V\left(\frac{X-\mu}{\sigma}\right) = 1 \quad .$$

Die Abbildung 4 zeigt die Histogramme unseres $B(30, \frac{1}{2})$–verteilten X und der zugehörigen standardisierten Zufallsgröße.

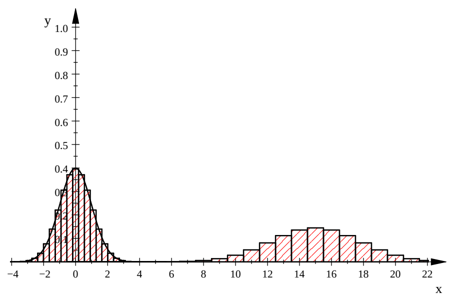


Abbildung 4: Histogramm von X mit $n = 30$ und $p = \frac{1}{2}$ und zugehöriger standardisierte Zufallsgröße

Wir schauen uns das Histogramm der standardisierten Zufallsgröße Z noch einmal aus der Nähe an:

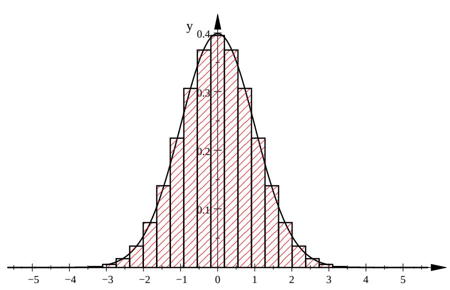


Abbildung 5: Histogramm der standardisierten Zufallsgröße Z

Die eingezeichnete Kurve ist der Graph der Gauß–Funktion φ mit

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad , \tag{8}$$

und die ist des Pudels Kern, den wir gesucht haben.

Überlegen wir uns nun, wie man mit Hilfe des Gaußschen φ Wahrscheinlichkeiten berechnet: Es ist

$$P(X = k) = P\left(Z = \frac{k-\mu}{\sigma}\right)$$

der Inhalt des Kästchens zu k im Histogramm von X und des Kästchens zu $\frac{k-\mu}{\sigma}$ im Histogramm von Z. Die Höhe des Kästchens im Histogramm von Z ist $\varphi(\frac{k-\mu}{\sigma})$, die Breite des Kästchens ist $\frac{1}{\sigma}$. Es folgt

$$P(X = k) \approx \frac{1}{\sigma}\varphi\left(\frac{k-\mu}{\sigma}\right) \quad . \tag{9}$$

Für $P(X \leq k)$ erhalten wir, wenn wir bedenken, dass das Kästchen zu k bei $k + \frac{1}{2}$ endet

$$P(X \leq k) = P\left(Z \leq \frac{k-\mu}{\sigma}\right) \approx \int_{-\infty}^{\frac{k+\frac{1}{2}-\mu}{\sigma}} \varphi(x)\,dx =: \Phi\left(\frac{k+\frac{1}{2}-\mu}{\sigma}\right) \quad . \tag{10}$$

Für $\Phi(x) = \int_{-\infty}^{x} \varphi(t)\,dt$ gibt es Tabellen, geschlossen integrieren kann man φ nämlich nicht. Auch φ selbst ist tabelliert.

3.2 Hinweise zur Anwendung der Näherungsformeln

Die Formeln in den Gleichungen 9 und 10 heißen die lokale und die integrale Näherungsformel von de Moivre und Laplace. Der Beweis ist sehr technisch, er hielte uns lange auf. Du darfst die Näherungen nur benutzen, wenn npq nicht zu klein ist. Nach einem alten Volksbrauch sieht man $npq \geq 9$ als ausreichend an. Einen Beweis dafür, dass das in Ordnung ist, habe ich noch nirgends gesehen; da folgen wir einfach der Tradition.

Die σ–Regeln im Buch, die zum Beispiel sagen, dass $P(|X - \mu| \leq 2\sigma \approx 0.955$ ist, beruhen auf diesen Näherungsformeln; man muss einfach $\Phi(2) - \Phi(-2)$ bilden, um die angegebene Wahrscheinlichkeit näherungsweise zu erhalten. Damit ist aber auch klar, dass man diese Sigmaregeln nur nehmen darf, wenn $V(X)$ groß genug ist.

Schauen wir uns zur Warnung die Aufgabe 5 der Klausur an. Es sei X der Betrag in Euro, den die Klasse bei einem Spiel an den Spieler zahlt. Dann ist

$$E(X) = -1 \cdot \frac{9999}{10000} + 9999 \cdot \frac{1}{10000} = 0 \quad ,$$

das Spiel ist fair. Wenn 100 Spieler spielen, nimmt die Klasse mit der Wahrscheinlichkeit

$$\left(\frac{9999}{10000}\right)^{100} \approx 0.99$$

100 Euro ein, wie Felix gesagt hat. Dennoch wäre es völlig unverantwortlich, der Klasse die Einrichtung des Standes zu empfehlen. Aber nun zu unserem Gegenstand: Einige von euch haben die Einnahme der Klasse in 100 Spielen als Zufallsgröße betrachtet. Nennen wir diese Zufallsgröße Y. Sie kann die Werte

$$100, -9900, -19900, \ldots$$

Euro annehmen. Es ist

$$Y = 100 - 10000Z \quad ,$$

dabei bedeutet Z die Anzahl der in hundert Ziehungen gezogenen weißen Kugeln. Offensichtlich ist Z $B(100, \frac{1}{10000})$–verteilt, mit dieser Information und unseren Regeln können wir den Erwartungswert und die Varianz von Y ausrechnen: Es ist

$$E(Y) = 100 - 10000E(Z) = 100 - 10000 \cdot 100 \cdot \frac{1}{10000} = 0 \quad \text{und}$$

$$V(Y) = 10000^2 \cdot 100 \cdot \frac{1}{10000} \cdot \frac{9999}{10000} \approx 10^6 \quad .$$

Der einzige Wert von Y, der in der $k\sigma$–Umgebung des Erwartungswertes liegt, ist 100 für $k = 1, 2, 3, \ldots 9$, der nächste Wert -9900 ist etwa 9.9σ vom Erwartungswert entfernt. Entsprechend ist

$$P(|Y - E(Y)| < k\sigma) = P(Y = 100) = 0.9999^{100} \approx 0.9900 \text{ für } k = 1, 2, 3, \ldots, 9,$$

die σ–Regeln versagen völlig.

4 Zum Testen von Hypothesen

4.1 Einführung

Wie kann man bei einem Würfel nachprüfen, ob de Wahrscheinlichkeit für eine Zwei wirklich $\frac{1}{6}$ ist? Wie stellt man fest, ob ein neues Medikament wirksamer ist als das alte, und wie überprüft man, ob der Ausschussanteil in einer Serienfertigung nicht größer ist als zulässig? Bei solchen Fragen kommt mathematische Statistik zur Anwendung. Sie bewahrt einen nicht zuverlässig vor Fehlentscheidungen, aber sie liefert Entscheidungsregeln, und sie gibt jeweils die Risiken an, mit der man eine falsche Entscheidung trifft.

Beginnen wir mit dem Würfel. Wir wollen ihn meinetwegen 300–mal werfen und die Anzahl X der geworfenen Zweien feststellen.

Der Erwartungswert für die Anzahl der Zweien bei $n = 300$ Würfen ist bei einem idealen Würfel 50. Wenn der beobachtete Wert X unseres Würfels zu weit von 50 abweicht, wollen wir den Würfel als gefälscht zurückweisen.

Natürlich kann ein idealer Würfel bei 300 Würfen auch 300 Zweien liefern, also werden wir unter Umständen auch einen idealen Würfel zurückweisen. Wenn wir kein Risiko eingehen wollen, brauchen wir keine Statistik. Immerhin verlangt man, dass das Risiko α , den Würfel zu Unrecht abzulehnen, klein ist; häufig nimmt man $\alpha = 0,05$.

Wir brauchen nun ein paar Fachbegriffe. Es sei p die (unbekannte) Wahrscheinlichkeit, mit der der Würfel eine Zwei wirft. Wir gehen von der so genannten **Nullhypothese**

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

über diese Wahrscheinlichkeit p aus. Man verwirft die Nullhypothese, wenn das beobachtete Versuchsergebnis in einem Ablehnungsbereich R liegt. Den Ablehnungsbereich wählt man so, dass die Wahrscheinlichkeit, die Nullhypothese zu Unrecht zu verwerfen, höchstens so groß ist wie eine vorgegebene Irrtumswahrscheinlichkeit α:

$$P_{H_0}(X \in R) \leq \alpha$$

Um R zu finden, bestimmen wir ein $c > 0$ so, dass

$$\Phi(c) - \Phi(-c) = \int_{-c}^{c} \varphi(x)\, dx = 0.95$$

ist. Mache dir den Sinn dieses Ansatzes anhand des Graphen von φ klar! Lösung ist

$$c = \Phi^{-1}(0.975) \approx 1.96 \quad .$$

Wenn wir für den Moment abgeschnittene Kästchen außer Acht lassen, können wir sagen, dass

$$P(|X - 300p| \leq c\sigma) \approx 0.95$$

ist. Für $p = \frac{1}{6}$ ist $c\sigma \approx 12.65$ und $50 + c\sigma \approx 62.65$. Das Kästchen der 62 geht nur bis 62.5; die Wahrscheinlichkeit

$$P_{H_0}(38 \leq X \leq 62)$$

sollte also etwas kleiner sein als 0.95. Wenn man es genau nimmt, sollte der Ablehnungsbereich

$$R = \{0, 1, \ldots, 36\} \cup \{64, 65, \ldots, 300\}$$

sein. Ob man in der Praxis so genau ist, weiß ich nicht.

Die Testvorschrift sieht nun folgendermaßen aus: Wirf den Würfel 300–mal und stelle die Anzahl X der geworfenen Zweien fest. Ist der beobachtete Wert von X kleiner als 47 oder größer als 63, lehnst du den Würfel ab.

Formal gesprochen: Für $X \in R$ verwirfst du die Nullhypothese.

Dir muss von vornherein eines klar sein: Wenn du sehr viele ideale Würfel nach dieser Vorschrift testest, wirst du auf lange Sicht etwa 5% als falsch ablehnen!

Nun haben wir auf das Risiko geachtet, die Nullhypothese zu Unrecht zu verwerfen, das heißt, einen Fehler erster Art zu machen. Es ist aber auch zu fragen, mit welcher Wahrscheinlichkeit der Test einen falschen Würfel erkennt. Es sei X_p eine $B(300, p)$–verteilte Zufallsgröße. Wir bilden die Funktion

$$g(p) = P(X_p \in R) = P(X_p \leq 46) + P(X_p \geq 64) \quad ,$$

sie heißt die **Gütefunktion** des Tests.

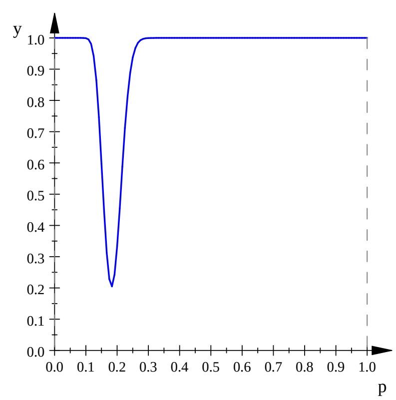


Abbildung 6: Gütefunktion des Würfeltests

Der Funktionswert $g(p)$ der Gütefunktion gibt an, mit welcher Wahrscheinlichkeit ein Würfel abgelehnt wird, der mit der Wahrscheinlichkeit p eine Zwei wirft. Würfel mit $p \geq 0.3$ und welche mit $p \leq 0.1$ werden recht zuverlässig erkannt, ein Würfel mit $p = 0.2$ besteht den Test mit der Wahrscheinlichkeit 67.4%.

Die Wahrscheinlichkeit, dass man die Nullhypothese nicht verwirft, obwohl sie falsch ist, heißt **Risiko zweiter Art**, der zugehörige Fehler Fehler zweiter Art. Ich erwähne dies schon einmal, wir werden die Begriffe später noch genauer behandeln.

4.2 Beispiel für die Konstruktion eines Tests

Bald ist Landtagswahl. Versetzen wir uns in die Lage einer Partei, deren Stimmenanteil um die 5% herum liegt. Zwischen den beiden Möglichkeiten $p < 0.05$ und

$p \geq 0.05$ liegen für die Partei Welten. Nun soll ein Test Auskunft geben, ob der Einzug in den Landtag gefährdet ist oder nicht.

Der Statistiker, der den Test konstruieren soll, will erst einmal das **Signifikanzniveau** wissen. Es wird gewöhnlich mit α bezeichnet. Der Test wird so konstruiert, dass das Risiko erster Art, dass nämlich die Nullhypothese zu Unrecht abgelehnt wird, höchstens α ist. Üblich sind Werte wie $\alpha = 0.1$ oder $\alpha = 0.05$ oder $\alpha = 0.01$. Ich wähle hier $\alpha = 0.01$, weil der Wert 0.05 schon vorkommt.

Man kann natürlich die Nullhypothese immer annehmen, dann ist das Risiko erster Art 0. Aber das ist nicht sehr sinnvoll, weil ja nie erkannt werden kann, wenn die Nullhypothese falsch ist; deshalb macht man das nicht.

Es soll also wirklich ein Test gemacht werden. Dann will der Statistiker wissen, wieviele Leute befragt werden sollen. Du vermutest richtig, dass der Test um so aussagekräftiger wird, je mehr Leute man fragt. Jede Befragung kostet aber Geld, deshalb muss man einen Kompromiss machen. Im Unterricht hat einer von euch $n = 1000$ für die Länge der Stichprobe vorgeschlagen, wir nehmen also $n = 1000$.

So, schauen wir uns die beiden Hypothesen an, die zu unterscheiden sind. Welche soll zur Nullhypothese gemacht werden? Wenn die Partei zu Unrecht annimmt, dass ihr Einzug in den Landtag gesichert ist, kommt sie nicht hinein, und das wäre eine ziemliche Katastrophe. Nimmt sie zu Unrecht an, dass der Einzug in den Landtag nicht gesichert ist, unternimmt sie teure Anstrengungen, die unnötig sind. Wir halten den ersten Irrtum für schlimmer und nehmen als Nullhypothese, dass $p < 0.05$ ist. Der erste Irrtum ist dann das Risiko erster Art, und das wird durch das Signifikanzniveau gedeckelt.

Nun haben wir alles zusammen: Wir testen die Hypothese $H_0 : p < 0.05$ gegen die Alternative $H_1 : p \geq 0.05$ mit Hilfe einer Stichprobe der Länge $n = 1000$ bei einem Signifikanzniveau $\alpha = 0.01$. Wir werden H_0 ablehnen, wenn wir unter den Befragten viele Wähler finden. Der Verwerfungsbereich wird also die Gestalt $R = \{u, u+1, \ldots, 1000\}$ haben, und wir müssen u so bestimmen, dass

$$P_{H_0}(X \in R) \leq \alpha = 0.01$$

ist; dabei bedeutet X die Anzahl der Wähler der Partei unter den 1000 Befragten. Wenn die Alternative H_1 zutrifft, sollte der Test das auch registrieren, deshalb wird man u möglichst groß machen.

Es sei nun X_p $B(1000, p)$–verteilt. Die Wahrscheinlichkeit $P(X_p \in R)$ wird um so größer, je größer p ist. Deshalb ist

$$P_{H_0}(X \in R) \leq P(X_{0.05} \in R) = P(X_{0.05} \geq u) \quad .$$

Mit Hilfe der Tabelle auf der Seite 180 unseres Buches, rechte Spalte, finden wir

$$c := \Phi^{-1}(0.99) \approx 2.326$$

– den Sinn dieses Ansatzes musst du dir mit Hilfe des Histogramms von $X_{0.05}$ klarmachen. Zu diesem c gehört der Wert

$$\mu + c\sigma = 1000 \cdot 0.05 + 2.326 \cdot \sqrt{1000 \cdot 0.05 \cdot 0.95} \approx 66.03$$

der Merkmalsachse von $X_{0.05}$. Wir wählen also $u = 67$. Der Test sieht nun so aus: Finden sich unter den 1000 Befragten mindestens 67 Wähler der Partei, wird die Hypothese H_0 verworfen. Die Partei kann dann davon ausgehen, dass ihr Einzug in den Landtag gesichert ist; jedenfalls sagt das der Test. Sicherheit gibt es erst am Wahlabend.

Der Schatzmeister der Partei ist über das große u nicht sehr glücklich. Er muss ja auch dann viel Geld ausgeben, wenn der Wähleranteil der Partei bei den 5.4% liegt, die er nach seinem Bauchgefühl vermuten mag. Können wir ihm mehr Informationen geben?

Um klarer zu sehen, bilden wir die Gütefunktion

$$g : p \mapsto P(X_p \geq 67)$$

des Tests. Die Abbildung 7 zeigt den Graphen der Gütefunktion einmal für $p \in [0; 1]$ und einmal für $p \in [0.05, 0.10]$.

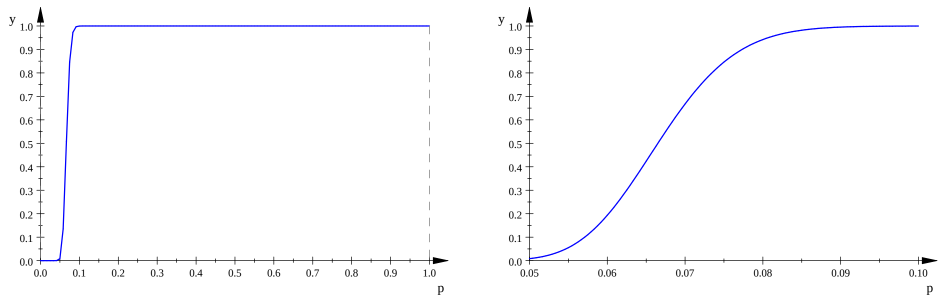


Abbildung 7: Graph der Gütefunktion des Tests der Partei

Den Schatzmeister trügt sein Gefühl keineswegs, es ist $g(0.054) \approx 0.040$. Das Risiko 2. Art, dass man nämlich bei H_0 bleibt, obwohl $p = 0.054$ ist, ist somit etwa 96%. Um den Test zu verbessern, könnte man das n vergrößern oder das Signifikanzniveau auf $\alpha = 0.05$ setzen. Rechne die Aufgabe mit $\alpha = 0.05$ noch einmal durch!

4.3 Zur Bestimmung des Verwerfungsbereichs R

Im vorigen Abschnitt war die Hypothese $H_0 : p < 0.05$ gegen die Alternative $H_1 : p \geq 0.05$ zu testen. Wir schauen uns noch einmal genau an, wie der Verwerfungsbereich R bestimmt wurde. Die Wahrscheinlichkeit, dass die Testgröße X einen Wert im Verwerfungsbereich annimmt, obwohl H_0 gilt, ist maximal

$$P(X_{0.05} \in R) \quad ,$$

und diese Wahrscheinlichkeit darf das gegebene $\alpha = 0.01$ nicht überschreiten. Wir haben da ein $c = \Phi^{-1}(0.99)$ bestimmt, für dieses c ist also

$$\Phi(c) = \int_{-\infty}^{c} \varphi(z)\, dz = 0.99 \quad \text{und entsprechend} \quad \int_{c}^{\infty} \varphi(z)\, dz = 0.01 \quad .$$

Du musst jetzt klar vor Augen haben, was all das am Graphen von φ bedeutet, und dir muss klar sein, dass der Graph von φ eine gute Näherung des Histogramms der standardisierten Zufallsgröße

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

zu X ist.

Schauen wir nun auf das Histogramm von $X_{0.05}$ selbst. Zu $c \approx 2.236$ gehört der Wert

$$g = c\sigma + \mu = 2.236\sqrt{1000 \cdot 0.05 \cdot 0.95} + 1000 \cdot 0.05 \approx 66.265$$

auf der Merkmalsachse von $X_{0.05}$. Der Graph von

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

nähert das Histogramm von $X_{0.05}$ gut an, denn die Klötzchen des Histogramms haben die Breite 1, und damit ist die Höhe des Klötzchens auf k gerade die Wahrscheinlichkeit $P(X = k) \approx f(k)$. Schaue dir dazu die Abbildung 8 an.

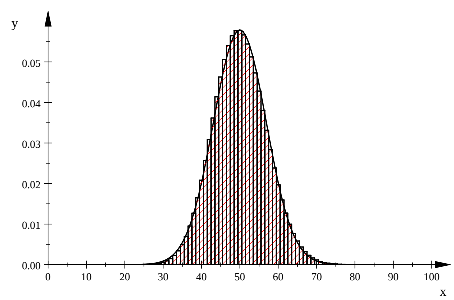


Abbildung 8: Histogramm von $X_{0.05}$ und Graph von f für $0 \leq x \leq 100$

Die Abbildung 9 zeigt uns die Graphen in der Nähe des Wertes g. Wir erkennen daran zwei Dinge:

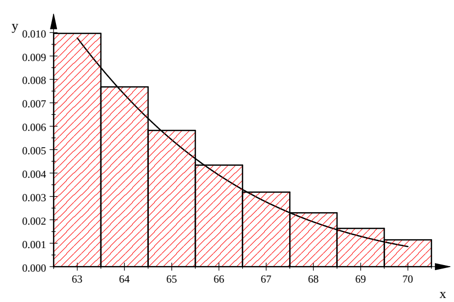


Abbildung 9: Histogramm von $X_{0.05}$ und Graph von f für $63 \leq x \leq 70$

- Der Wert $g \approx 66.265$ teilt das Kästchen auf dem x–Wert 66. Das Integral über f von g bis Unendlich ergibt den Wert 0.01. Nehmen wir den Wert $x = 66$ noch zu R hinzu, müssen wir damit rechnen, dass die Irrtumswahrscheinlichkeit den zulässigen Wert 0.01 überschreitet. Deshalb dürfen wir R erst bei 67 beginnen lassen: $R = \{67, 68, \ldots, 1000\}$
- Die Güte der Näherung ist im betrachteten Bereich schon ziemlich mies, deshalb sollte man sich nicht zu viele Gedanken machen. Als ambitionierter Schüler solltest du dir allerdings der Problematik bewusst sein.

5 Schätzen: das Konfidenzintervall

5.1 Einführung und Berechnung des Konfidenzintervalls

Eine Umfrage unter 200 zufällig ausgewählten Wahlberechtigten ergab, dass 80 von ihnen die Spaßpartei wählen wollen. Wie groß mag der Anteil der Anhänger der Spaßpartei unter allen 10 Millionen Wahlberechtigten sein?

Wer sicher sein will, kann nur sagen, dass es unter den 10 Millionen mindestens 80 Anhänger und mindestens 120 Nichtwähler der Spaßpartei gibt. Diese Aussage ist garantiert wahr, aber völlig unbrauchbar.

Um zu einer interessanten Aussage zu kommen, geht man davon aus, dass das Ergebnis der Umfrage einigermaßen typisch ist, genauer, dass es in der 95%-Umgebung des Erwartungswertes liegt. Bei einer $B(200, p)$–verteilten Zufallsgröße X_p, und als solche können wir die Anzahl der Wähler der Spaßpartei in der Umfrage ansehen, heißt das, dass der beobachtete Wert 80 in der $\Phi^{-1}(0.975)\cdot\sqrt{200p(1-p)}$–Umgebung des Erwartungswertes $\mu = 200p$ liegen soll. Man bestimmt nun alle p, für die dies erfüllt ist. Sie bilden ein Intervall, man nennt es das 95%–Konfidenzintervall zum beobachteten Wert x. Es enthält den Wert 80/200, und man geht davon aus, dass das berechnete Intervall das gesuchte wahre p überdeckt.

Die Grenzen des Intervalls erfüllen die Gleichung[2]

$$np \pm c\sqrt{np(1-p)} = x \quad , \tag{11}$$

dabei ist x der beobachtete Wert, $n = 200$ und $c = \Phi^{-1}(0.975)$. Jetzt ist eigentlich nur noch eine quadratische Gleichung zu lösen, aber auf dieser Aufgabe liegt scheinbar ein alter Fluch: nur recht selten führt die Rechnung per Hand zum richtigen Ergebnis. Ich führe die Rechnung hier vor, damit alles seine gute Ordnung habe. Belaste dich nicht damit; wenn du den Ansatz eingesehen hast, kannst du ruhig zu Gleichung 12 gehen.

Also nun die Rechnung: Wir bringen np auf die rechte Seite und quadrieren – dabei verschwindet das $\pm$ vor der Wurzel.

$$c^2np(1-p) = (x-np)^2$$

Diese quadratische Gleichung für p lösen wir wie üblich.

$$\begin{aligned} c^2np - c^2np^2 &= x^2 - 2npx + n^2p^2 \\ 0 &= (n^2 + c^2n)p^2 - (c^2n + 2nx)p + x^2 \\ 0 &= p^2 - \frac{c^2n + 2nx}{n^2 + c^2n}p + \frac{x^2}{n^2 + c^2n} \end{aligned}$$

Ein n können wir im Vorfaktor von p noch herauskürzen:

$$p^2 - \frac{c^2 + 2x}{n + c^2}p + \frac{x^2}{n^2 + c^2n} = 0$$

[2] Eigentlich muss ja noch das ganze Klötzchen auf x im Intervall $[\mu - c\sigma; \mu + c\sigma]$ liegen. Deshalb hat Jonathan im Grunde recht, wenn er das maximale zulässige p mit dem Ansatz

$$np - c\sqrt{np(1-p)} = x - \frac{1}{2}$$

und das minimale p mit dem Ansatz

$$np + c\sqrt{np(1-p)} = x + \frac{1}{2}$$

berechnen will.

Nun können wir die Lösung hinschreiben:

$$
\begin{aligned}
p &= \frac{c^2+2x}{2(n+c^2)} \pm \sqrt{\left(\frac{c^2+2x}{2(n+c^2)}\right)^2 - \frac{x^2}{n(n+c^2)}} \\
&= \frac{c^2+2x}{2(n+c^2)} \pm \sqrt{\frac{(c^2+2x)^2}{4(n+c^2)^2} - \frac{4x^2(n+c^2)}{4n(n+c^2)^2}} \\
&= \frac{c^2+2x \pm \sqrt{c^4+4c^2x+4x^2-4x^2-\frac{4c^2x^2}{n}}}{2(n+c^2)} \\
&= \frac{c^2+2x \pm c\sqrt{c^2+4x-\frac{4x^2}{n}}}{2(n+c^2)}
\end{aligned}
$$

Damit haben wir unser Ergebnis! Das Konfidenzintervall hat die Grenzen:

$$\boxed{p = \frac{c^2+2x \pm c\sqrt{c^2+4x-\frac{4x^2}{n}}}{2(n+c^2)}} \tag{12}$$

Die Umfragewerte der Spaßpartei liefern das Intervall $[0.3346, 0.4692]$ für den Wähleranteil in der Gesamtbevölkerung; Jonathans Intervall wäre $[0.3369, 0.4667]$.

Es ist nicht anzunehmen, dass der Vorstand der Spaßpartei mit dem Ergebnis zufrieden wäre; der Durchmesser des Intervalls ist viel zu groß. Will man ihn kleiner haben, muss man mehr Leute befragen, also das n vergrößern. Sagen wir, der Vorstand will den Wähleranteil auf $\pm 1\%$ wissen, er verlangt also ein Konfidenzintervall, dessen Durchmesser höchstens $d = 0.02$ ist. Bezeichnen wir die Grenzen des Konfidenzintervalls mit p_u und p_o, ist der Durchmesser b des entsprechenden Intervalls auf der Merkmalsachse gegeben durch

$$b = c\sqrt{np_u(1-p_u)} + c\sqrt{np_o(1-p_o)} = c\sqrt{n}\left(\sqrt{p_u(1-p_u)} + \sqrt{p_o(1-p_o)}\right) \quad .$$

Wie wir wissen, ist durch $y = \sqrt{p(1-p)}$ ein Halbkreis mit Mittelpunkt $(\frac{1}{2}|0)$ und Radius $\frac{1}{2}$ gegeben. Folglich ist stets

$$\sqrt{p(1-p)} \leq \frac{1}{2} \quad , \tag{13}$$

und daraus ergibt sich

$$b = c\sqrt{n}\left(\sqrt{p_u(1-p_u)} + \sqrt{p_o(1-p_o)}\right) \leq c\sqrt{n}\left(\frac{1}{2}+\frac{1}{2}\right) = c\sqrt{n} \quad .$$

Wenn wir n so wählen, dass $c\sqrt{n} \leq d = 0.02$ ist, ist erst recht der Durchmesser b des Konfidenzintervalls kleiner oder gleich 0.02. Dies liefert für n die Bedingung

$$n \geq \frac{c^2}{d^2} = \frac{c^2}{0.02^2} \approx \frac{1.96^2}{0.02^2} = 9604 \quad .$$

Das ist eine sehr große Zahl, Umfragen dieses Umfangs sind meines Wissens nicht üblich. Wenn du nun vermutest, dass viele Werte in den zahlreichen Zeitungsmeldungen, dass soundsoviel Prozent der Bevölkerung für oder gegen dieses oder jenes sei, nicht sehr verlässlich sind, werde ich deiner Skepsis nicht entgegentreten. Die Zahlen als „frei erfunden“ zu bezeichnen, wäre wohl zu viel gesagt; für bare Münze sollte man sie aber nicht nehmen.

Aufgabe Die PQR–Partei krebste bei den letzten Wahlen so um die 7% Wähleranteil. Nun steht die nächste Wahl vor der Tür. Bei einer aktuellen Umfrage fanden sich unter 400 Wahlberechtigten 30 Wähler der PQR–Partei. Werte die Umfrage aus und trage dem Parteivorstand dein Ergebnis vor. Der Vorstand erwartet auch eine Empfehlung, ob eine weitere Umfrage in Auftrag gegeben werden soll und welchen Umfang sie gegebenenfalls haben sollte.

5.2 Vertiefende Betrachtung des Konfidenzintervalls

Schauen wir uns noch einmal das Einführungsbeispiel an: Unter 200 Befragten fanden sich 80 Anhänger der Spaßpartei, und daraus ergab sich als 95%–Konfidenzintervall für den Anteil p der Anhänger der Spaßpartei in der wahlberechtigten Bevölkerung das Intervall $[0.3369, 0.4667]$. Man ist versucht zu sagen, mit der Wahrscheinlichkeit 95% liege p in diesem Intervall, aber das ist grober Blödsinn. Natürlich kann man sehr oft ein Konfidenzintervall bestimmen: man muss dann jeweils eine Umfrage machen und zum Umfrageergebnis das Konfidenzintervall bestimmen. Das gibt so viele Konfidenzintervalle, wie man Umfragen gemacht hat, und etwa 95% dieser Intervalle überdeckten das wahre p, das der Umfrage zugrunde liegt. Nicht das p ist Ergebnis eines Zufallsversuchs, sondern das Intervall.

Sebastian hat eine Simulation geschrieben, die zu $B(n,p)$–verteiltem X sehr oft einen Wert auslost, zu diesem Wert das 95%–Konfidenzintervall ausrechnet und feststellt, ob das Intervall das wahre p überdeckt, und in der Tat lag der Anteil der guten Intervalle etwa bei 95%. Mehr kann man eigentlich nicht sagen: Das konkrete vorliegende Konfidenzintervall ist mit einer Maschine bestimmt, die im Mittel 95% gute Intervalle liefert. Es enthält gerade die p–Werte, bei denen der beobachtete Wert noch in der 95%–Umgebung von p liegt, und die Annahme, dass bei dem Versuch ein Wert aus der 95%–Umgebung des Erwartungswertes ausgelost wurde, ist eben nur in 95% der Fälle wahr. Das klingt kompliziert, aber einfacher geht es nicht, wenn man redlich bleiben will.

6 Stetige Zufallsgrößen

6.1 Einführendes Beispiel: radioaktiver Zerfall

Wir denken uns eine sehr große Anzahl N Teilchen einer radioaktiven Substanz. Wie du weißt, ist nach einer gewissen Zeit, der vom Stoff abhängigen Halbwertszeit, die Hälfte der Teilchen zerfallen. Wenn wir die Zeit t in Halbwertzeiten messen, beträgt die Anzahl der Teilchen zum Zeitpunkt $t \geq 0$

$$a(t) = N \cdot 2^{-t} \quad . \tag{14}$$

Der Graph der Funktion a ist in Abbildung 10 dargestellt.

Wir markieren eines der Teilchen und bezeichnen den Zeitpunkt, zu dem es zerfällt, mit T. Meines Wissens kann niemand den Zerfallszeitpunkt zuverlässig vorhersagen, wir haben es mit einer Zufallsgröße zu tun. Unser T kann jeden Wert $t \geq 0$ annehmen, und für jedes t ist, wie wir uns überlegt haben $P(T = t) = 0$. Das ist seltsam; dass $P(T = t) = 0$ ist, bedeutet ja nicht, dass der Zerfall nicht zu eben diesem Zeitpunkt t erfolgen kann. Die Zufallsgröße T ist von einer für uns neuen Art, man nennt sie stetig.

Zum Zeitpunkt t sind schon $N - N \cdot 2^{-t} = N(1 - 2^{-t})$ Teilchen zerfallen. Die Wahrscheinlichkeit $P(T \leq t)$, dass unser markiertes Teilchen eines davon ist, ist demnach

$$F(t) := P(T \leq t) = \frac{N(1-2^{-t})}{N} = 1 - 2^{-t} \quad ,$$

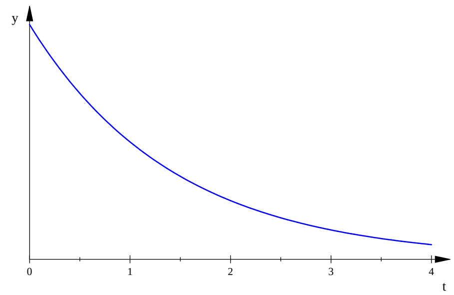


Abbildung 10: Anzahl $a(t) = N \cdot 2^{-t}$ radioaktiver Teilchen nach t Halbwertzeiten

die Abbildung 11 zeigt den Graphen.

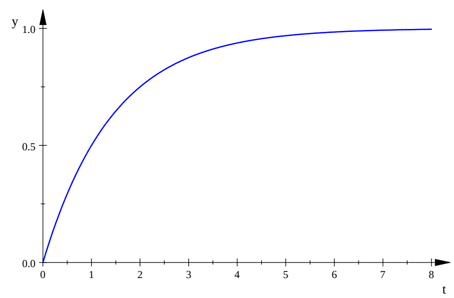


Abbildung 11: Graph der Verteilungsfunktion $F(t) = P(T \leq t) = 1 - 2^{-t}$ von T

Die Wahrscheinlichkeit, dass unser Teilchen im Zeitintervall $[a, b]$ zerfällt, ist dann[3]

$$P(a \leq T \leq b) = P(T \leq b) - P(T \leq a) = F(b) - F(a) \quad . \tag{15}$$

Wir wollen nach dem Erwartungswert von T fragen, also nach mittleren Lebensdauer eines Teilchens. Unsere alte Setzung $\sum xP(X = x)$ versagt hier, weil $P(T = t) = 0$ ist und weil wir die reellen Zahlen nicht der Reihe nach aufzählen können. Nun ist

$$P(a \leq T \leq b) = F(b) - F(a) = F'(z)(b - a)$$

für ein $z \in [a, b]$, das ist dir ja noch vertraut; die Funktion $f(t) := F'(t)$ heißt Dichtefunktion von T.

Der Beitrag eines sehr kleinen Intervalls $[a, b]$ zum Erwartungswert ist dann gegeben durch

$$zf(z)(b - a)$$

für ein $z \in [a, b]$. Der Beitrag des Intervalls $[0, b]$ ist dann von der Art

$$\sum_{k=1}^{n} t_k f(t_k) \Delta t \quad ,$$

dabei wurde $[0, b]$ in n Teile eingeteilt und t_k geeignet im k–ten Teilintervall gewählt. Die Summe ist offensichtlich eine Riemannsche Summe für

$$\int_0^b t f(t)\, dt \quad .$$

[3] Eigentlich müsste es $P(T \leq b) - P(T < a)$ heißen, aber $P(T = a)$ ist ja $= 0$.

Nach diesen Überlegungen wird dir die folgende Definition einleuchten:

8 Definition
Es sei T eine stetige Zufallsgröße mit Verteilungsfunktion $F(t) = P(T \leq t)$ und Dichtefunktion $f(t) = F'(t)$. Dann sind Erwartungswert und Varianz von T definiert durch

$$E(T) := \int_{-\infty}^{\infty} t f(t)\, dt \quad \text{und} \quad V(T) := \int_{-\infty}^{\infty} (t - E(T))^2 f(t)\, dt \quad ,$$

falls diese Integrale existieren.

Für unser Teilchen haben wir $E(T) = \frac{1}{\ln(2)}$ ausgerechnet, es lebt also im Mittel 1.44 Halbwertzeiten.

Die Dichtefunktion f hat die schöne Eigenschaft, dass

$$P(a \leq T \leq b) = \int_a^b f(t)\, dt \tag{16}$$

ist.

6.2 Normalverteilte Zufallsgrößen

9 Definition
Eine stetig verteilte Zufallsgröße mit der Dichtefunktion

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

heißt normalverteilt mit dem Erwartungswert $\mu \in \mathbb{R}$ und der Standardabweichung $\sigma > 0$. Die Verteilungsfunktion von X ist

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad .$$

Natürlich muss man nachrechnen, dass das auch alles in sich stimmig ist, was ich in dieser Definition notiert habe, aber das ist kein großes Problem; man braucht etwas Integralrechnung.

Die Normalverteilung ist unglaublich wichtig, du wirst noch mehr dazu erfahren. Im Prinzip kennst du sie schon: Das Histogramm einer binomialverteilten Zufallsgröße X mit den Parametern n und p wird für nicht zu kleine npq gut durch die Dichtefunktion f mit $\mu = np$ und $\sigma = \sqrt{np(1-p)}$ angenähert; das ist der lokale Grenzwertsatz von de Moivre und Laplace.

7 Der Zentrale Grenzwertsatz

7.1 Der Satz

Dieser Satz ist, so ein Kenner, das schönste Ergebnis der Wahrscheinlichkeitstheorie.[4] Nun, seht selbst.

[4] Johann Pfanzagl, Elementare Wahrscheinlichkeitsrechnung. Berlin – New York 1988, Seite 183.

10 Satz (Zentraler Grenzwertsatz)
Es sei X eine Zufallsgröße mit Erwartungswert $\mu = E(X)$ und Varianz $V(X)$, und es seien $X_1, X_2, \ldots, X_n$ unabhängige Kopien von X. Wir setzen

$$S := X_1 + X_2 + \cdots + X_n = \sum_{k=1}^{n} X_k \quad .$$

Dann gilt

$$P(S \leq s) \approx \Phi\left(\frac{s - E(S)}{\sqrt{V(S)}}\right) \quad .$$

Wir schauen uns einige Beispiele an.

Beispiel 1 Es sei X die Anzahl der Erfolge bei einem einfachen Bernoulliversuch, also $P(X = 1) = p$ und $P(X = 0) = 1 - p$. Wir führen diesen Versuch n–mal unabhängig durch, und für $k = 1, 2, \ldots, n$ sei X_k die Anzahl der Erfolge bei der k–ten Durchführung. Die Zufallsgröße $S := X_1 + X_2 + \cdots + X_n$ gibt dann die Anzahl der Erfolge insgesamt an. Die Summenregeln für Erwartungswert und Varianz ergeben $E(S) = np$ und $V(S) = np(1 - p)$. Nach dem Zentralen Grenzwertsatz ist nun

$$P(S \leq s) \approx \Phi\left(\frac{s - np}{\sqrt{np(1-p)}}\right) \quad .$$

Du erkennst unschwer die integrale Näherungsformel von de Moivre und Laplace. Für $s = k \in \mathbb{Z}$ verwenden wir sie in der Form

$$P(S \leq k) \approx \Phi\left(\frac{k - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right) \quad ,$$

das liefert einen besseren Näherungswert als der Zentrale Grenzwertsatz.

Beispiel 2 Was passiert, wenn man einen Würfel n–mal wirft und die Summe S der Augenzahlen bildet, haben wir uns mit Hilfe von MuPAD angeschaut. Es zeigte sich, dass das Histogramm von S recht schnell die Form einer Glockenkurve annahm, also die Form der Dichtefunktion einer normalverteilten Zufallsgröße, und dann ist es kein Wunder, dass man $P(S \leq s)$ mit Hilfe von Φ ausrechnen kann. Dies funktionierte auch, wenn man statt eines Würfels einen Legostein warf.

7.2 Konkretes Beispiel: Legoklotz

Auf der Seite 48 unseres Buches ist ein besonderer Würfel beschrieben: die Flächen eines Legoklotzes sind mit den Zahlen von 1 bis 6 bezeichnet. Wir stellen uns vor, dass die Wahrscheinlichkeitsfunktion der Zufallsgröße X: „mit dem Klotz geworfene Augenzahl" durch diese Matrix gegeben ist:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{54}{500} & \frac{7}{500} & \frac{235}{500} & \frac{143}{500} & \frac{7}{500} & \frac{54}{500} \end{pmatrix} \quad .$$

Abbildung 12 (oben) zeigt das Histogramm, es sieht recht unregelmäßig aus.

Wir bilden nun für n unabhängige Kopien von X die Zufallsgröße

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

Praktisch wirft man den Legoklotz n–mal und addiert die geworfenen Augenzahlen, die Summe ist der Wert von S. Die Abbildung 12 zeigt Histogramme der standardisierten Zufallsgrößen zu S_2, zu S_4, zu S_8 und zu S_{16}, jeweils mit dem Graphen der Gaußfunktion φ. Man sieht, wie sie sich die Histogramme der Gaußkurve annähern, wie es der Zentrale Grenzwertsatz sagt.

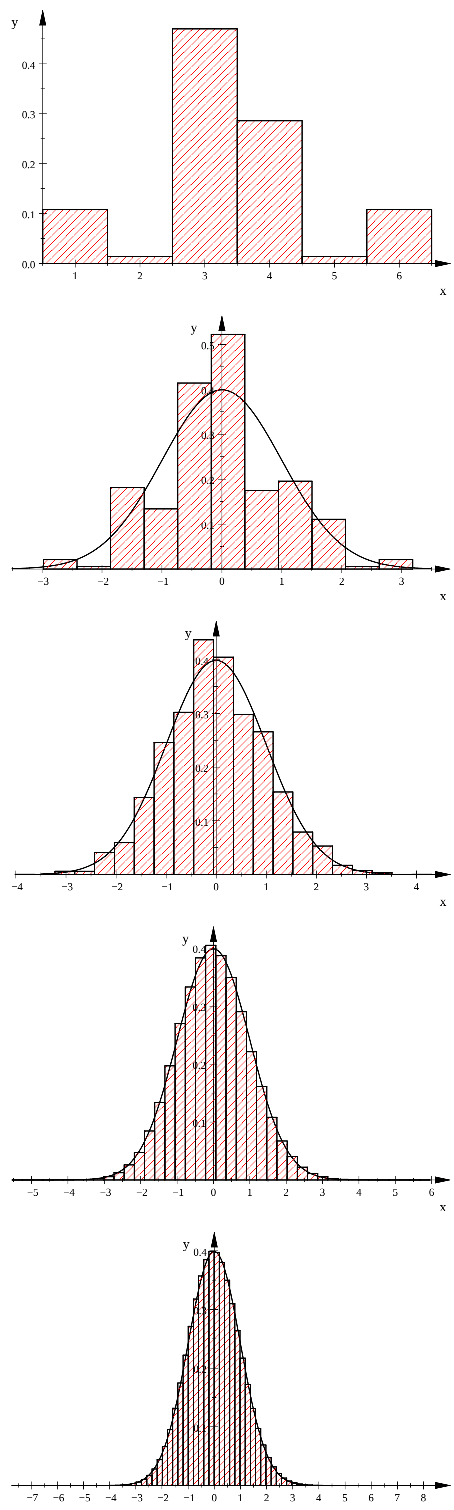


Abbildung 12: Histogramme von X und der standardisierten Zufallsgrößen zu S_2, S_4, S_8 und S_{16}

8 Klausur Nr. 4 am 19. Mai 2010

1. Es sei X_p eine $B(1000, p)$–verteilte Zufallsgröße.

   (a) Berechne $P(X_{0,3} \leq 295)$, $P(X_{0,3} > 310)$, $P(X_{0,3} = 300)$ und schließlich $P(X_{0,001} = 0)$ (pass auf!).

   (b) Du hast gerade die Näherungsformeln von de Moivre und Laplace verwandt. Erkläre mit Hilfe einer Skizze und mit ein paar Worten, wie das läuft (die Idee, nicht wie man die Zahlen einsetzt!). Warum hast du beim letzten Term nicht mit der Näherung gearbeitet?

2. Am Gymnasium in Birkenkamp soll eine Mensa gebaut werden. Nun ist die Frage, wie viele Schüler denn wohl da essen werden, wenn sie fertig ist. Man könnte eine Befragung der gesamten Schülerschaft durchführen, aber das wäre zu einfach. Statt dessen soll im Sinne des Schulprogramms ein Mathematikkurs eine Umfrage planen und auswerten, Projekte sind ja immer gut. Es soll jeder 10. der 1200 Schüler befragt werden.

   (a) Es gilt zunächst, die zu befragenden Schüler auszuwählen. Dem Kurs liegen folgende Vorschläge aus dem Mensabauausschuss vor:

   - die ganze Stufe 5,
   - spontane Interviews am Busbahnhof, bis man 120 beisammen hat,
   - jeder 10. in der alphabetischen Liste aller Schüler.

   Was sollte man machen?

   (b) Endlich ist die Befragung durchgeführt: 14 der befragten 120 Schüler gaben an, in der Mensa essen zu wollen. Schüler J.-U. nimmt seinen Taschenrechner und bildet

   $$\frac{1}{120}\left(14 \pm 1,96 \cdot \sqrt{120 \cdot \frac{14}{120} \cdot \frac{106}{120}}\right) \quad .$$

   Was macht der denn da?? Schreib' was Gescheites!

   (c) S. hat eine vornehme MuPAD–Routine, die spuckt ihm das Konfidenzintervall aus: $[0.07399734968, 0.1814560128]$. Na bitte. Berechne J.–U.s Werte, vergleiche und kommentiere.

   (d) Schüler $*$, der immer nur die Hälfte mitkriegt, stößt dich mit dem Ellbogen an und fragt, was das Konfidenzintervall überhaupt ist. Gib ihm Auskunft. Der Umfang soll im Rahmen dessen liegen, was du dem Nachbarn während des Unterrichts sagen kannst.

   (e) Der nachdenkliche und etwas grüblerische E. hat einen Einwand zu der ganzen Sache, und den trägt er jetzt vor: Wenn es in der gesamten Schülerschaft 100 Mensagänger gibt, dann ist doch $P(X = 101) = P(X = 102) = \cdots = P(X = 120) = 0$. Das binomialverteilte X, mit dem der Apparat arbeitet, kann doch hier eigentlich nicht passen. M. hört ihm gut zu, denkt nach und berechnet $P(X > 100)$ für $B(120, \frac{100}{1200})$–verteiltes X. Was bekommt M. heraus, und was hat das mit E.s Einwand zu tun? Kommentiere.

   (f) Formuliere das Ergebnis der Untersuchung für den Mensabauausschuss. Ein heikler Punkt dabei ist die Bedeutung des gewählten Signifikanzniveaus. Gib dir Mühe, so zu formulieren, dass du nicht verbreitete Missverständnisse förderst.

3. Ein Kräuterweiblein hat eine neue Salbe angerührt, die besser sein soll als ihre alte Salbe. Nun ist sie im Zweifel, ob sie auf die neue Salbe umsteigen soll, denn dann kann sie einen beträchtlichen Teil der im letzten Jahr gesammelten Kräuter nicht mehr verwenden. Die alte Salbe half in 80% der Fälle; die neue Salbe brachte 63 von den 70 Patienten Linderung, bei denen sie ausprobiert wurde. Konstruiere einen ordentlichen Test und gib dem Weiblein anhand dieses Tests eine Empfehlung, was sie tun soll. Natürlich wird das Kräuterweiblein nur deine Empfehlung anhören, aber ich lese deine Durchführung. Schreibe das also ordentlich auf.

4. Es sei T der Zeitpunkt, zu dem eine Seifenblase zerplatzt, die zur Zeit $t = 0$ erzeugt wurde. Die Dichtefunktion von T ist:

$$f(t) = \begin{cases} 0 & \text{für } t \leq 0 \\ \frac{1}{2} - \frac{1}{8}t & \text{für } 0 < t \leq 4 \\ 0 & \text{für } 4 < t \end{cases}$$

   Gib $P(T = 1)$ an, berechne $P(T > 2)$ und den Erwartungswert $E(T)$.

5. Hinz schlägt Kunz ein Spiel vor. Die beiden würfeln abwechselnd, Kunz fängt an. Bei einer 6 gewinnt Kunz. Da Hinz als zweiter Spieler im Nachteil ist, will er gewinnen, wenn er eine 1 oder 2 wirft.

   (a) Berechne die Wahrscheinlichkeit, dass das Spiel den folgenden Verlauf nimmt: Kunz gewinnt das Spiel mit seinem zweiten Wurf.

   (b) Mit welcher Wahrscheinlichkeit gewinnt Hinz, mit welcher Kunz?

   (c) „Wenn ich 100–mal würfle und die geworfenen Augenzahlen addiere, bekomme ich im Mittel die Augensumme 350“, sagt Kunz zu Hinz. „Kann ich denn zuversichtlich sein, dass ich mindestens die Augensumme 330 bekomme?“ Hinz antwortet: „Du kannst die Wahrscheinlichkeit, dass die Augensumme kleiner als 330 ist, mit dem Zentralen Grenzwertsatz ausrechnen. Die Varianz der Augenzahl bei einem Wurf ist $\frac{35}{12}$.“
   Gib eine Formel für die $\frac{35}{12}$ an, den Wert kannst du aber glauben, ebenso wie die 350 in Kunzens Aussage. Berechne dann die Wahrscheinlichkeit, die Kunz wissen will. Wenn du die Gleichung des Zentralen Grenzwertsatzes nicht mehr zur Hand hast: Für $S = X_1 + X_2 + \cdots + X_n$ ist

$$P(S \leq s) \approx \Phi\left(\frac{s - E(S)}{\sqrt{V(S)}}\right) \quad .$$

   Die Bedeutung der Variablen musst du allerdings wissen.

9 Poissonverteilte Zufallsgrößen

Wer in dieser Jahreszeit seinen Hund auf dem Gelände des Ludwig–Steil–Hofes spazieren führt, muss damit rechnen, dass der Hund Zecken mit nach Hause bringt, und zwar, sagen wir, im Mittel zwei Stück je Spaziergang. Manchmal sind es mehr, manchmal weniger, im Mittel eben zwei Stück. Wenn die Anzahl der Zecken vom Zufall abhängt, haben wir es mit einer Zufallsgröße X zu tun, und wir können nach der Wahrscheinlichkeitsfunktion $k \mapsto P(X = k)$ fragen.

Wir schauen uns in unserer Werkzeugkiste um: unser bestes Werkzeug sind Bernoulliketten, und die helfen uns auch hier weiter. Ein Bernoulliversuch darf nur zwei Ausgänge haben, hier dann „Zecke“ oder „keine Zecke“, das ist natürlich ein

Problem. Aber es lässt sich lösen: Wir unterteilen den ganzen Spaziergang in n gleichwertige Teilstücke und fragen, ob der Hund auf dem i–ten Teilstück eine Zecke eingefangen hat – das ist der Erfolg, $X_i = 1$ – oder nicht, das ist der Misserfolg. Wie üblich ist dann $X = X_1 + X_2 + \cdots + X_n$ binomialverteilt.

Du siehst natürlich sofort das Problem: unser X ist die Anzahl der Teilstücke des Weges mit Zeckenangriff. Ob der Hund sich auf dem i–ten Teilstück des Spaziergangs eine, zwei oder noch mehr Zecken eingefangen hat, unterscheidet unser Modell nicht. Aber das kriegen wir in den Griff, indem wir n so groß machen, dass wir diesen Fall außer Acht lassen können.

Wenn die Erfolgswahrscheinlichkeit p ist, bekommen wir im Mittel np Teilstücke mit Zeckenangriff; folglich muss $\lambda = np$ oder $p = \frac{\lambda}{n}$ sein, damit der Hund im Mittel λ Zecken einfängt.

Es sei also n groß. Wir bilden $P(X = k)$ und lassen n schließlich sogar, zu Christians Erheiterung, wieder gegen Unendlich laufen, um zur echten Poissonverteilung zu kommen:

$$\begin{aligned}
P(X=k) &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{n\cdot(n-1)\cdot\cdots\cdot(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k \cdot \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!}\cdot\frac{n}{n}\cdot\frac{n-1}{n}\cdot\cdots\cdot\frac{n-k+1}{n}\cdot\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!}\cdot 1\cdot\left(1-\frac{1}{n}\right)\cdot\cdots\cdot\left(1-\frac{k-1}{n}\right)\cdot\left(1-\frac{\lambda}{n}\right)^{n-k} \quad .
\end{aligned}$$

Schauen wir uns an, wie sich die Faktoren verhalten, wenn n gegen Unendlich läuft. Der erste Faktor ist fest. Jeder der k Faktoren der Form $1 - \frac{i}{n}$ strebt gegen 1, die können wir vergessen. Natürlich strebt auch $1 - \frac{\lambda}{n}$ gegen 1, aber gleichzeitig wächst die Anzahl dieser Faktoren, und es ist überhaupt nicht leicht zu sagen, wer sich durchsetzt. Als Grenzwert für n gegen Unendlich könnte zwischen 0 und 1 so ziemlich alles herauskommen. – Nun, wie Dominik und Henrich euch gelegentlich vorrechnen werden, gilt

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

und das hättest du vermutlich nicht erraten. Aber wir sind jetzt am Ziel:

11 Definition
Es sei $\lambda \geq 0$. Eine Zufallsgröße X heißt Poissonverteilt mit Parameter λ, wenn gilt

$$P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda} \quad \textit{für } k = 0,1,2,3,\ldots .$$

Dass die Summe über alle $P(X = k)$ auch genau 1 ergibt, dass $E(X) = \lambda$ und ebenfalls $V(X) = \lambda$ ist, wird euch Jens gelegentlich vorrechnen. Und mit wievielen Zecken der Hund rechnen muss, siehst du in Abbildung 13 auf Seite 30.

10 Simulation radioaktiven Zerfalls

Wir betreiben Stochastik als mathematische Theorie, mit Experimenten geben wir uns praktisch nicht ab. Immerhin lassen wir gelegentlich den Rechner Werte von

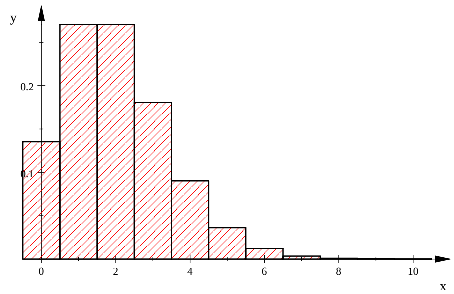


Abbildung 13: Zeckenbefall – Poissonverteilung mit Parameter $\lambda = 2$

Zufallsversuchen auslosen. Das macht ein ausgewiesener Experte für uns, aber du solltest dich nicht zurücklehnen, Sebastian machen lassen und ein paar Dateien abspeichern, sondern dich selbst auch ein wenig kundig machen.

Der MuPAD–Befehl

```
random()()
```

führt zur Ausgabe einer zwölfziffrigen Zahl, einer Pseudozufallszahl. MuPAD rechnet diese Zahlen aus, von Zufall kann überhaupt keine Rede sein, aber eine Folge dieser Zahlen sieht regellos aus. Braucht man eine Zahl zwischen 1 und 6, will man also Würfeln simulieren, benutzt man

```
random(1..6)() ,
```

und braucht man eine in einem Intervall $[a, b]$ zufällig gewählte reelle Zahl, hilft

```
a+random()()*(b-a)/10^12 .
```

Es sei nun T stetig verteilte Zufallsgröße mit Verteilungsfunktion

$$P(T \leq t) = F(t) = \begin{cases} 0 & \text{für } t \leq 0 \\ 1 - e^{-t} & \text{für } x > 0 \end{cases}$$

– es handelt sich um den Zerfallszeitpunkt eines radioaktiven Teilchens[5]. Die Dichtefunktion von T ist $f(t) = F'(t) = e^{-t}$ für $t > 0$. Der Graph von f sieht überhaupt nicht wie eine Gaußkurve aus, nach dem Zentralen Grenzwertsatz soll aber eine Summe unabhängiger Kopien von T annähernd normalverteilt sein. Sebastian hatte den Auftrag, dies mit Hilfe einer Simulation zu verdeutlichen.

Um den Auftrag auszuführen, musste er mehrere Kunststücke vollbringen. Zunächst einmal musste er Werte von T auslosen, aber die liegen irgendwo auf der reellen Achse rechts von 0. Dies machte er so, dass er eine Zufallszahl $z \in [0; 1]$ ausloste[6] und $t = F^{-1}(z)$ als ausgelosten t–Wert nahm; dabei ist F^{-1} die Umkehrfunktion von F. Um nun einen Wert s von

$$S = \sum_{k=1}^{n} T_k$$

zu bekommen, musste er n T–Werte auslosen und addieren. Von diesen S–Werten brauchte er sehr viele. Dann musste er die ausgelosten S–Werte in angemessener

[5]Siehe Seite 22

[6]Eine zufällig gewählte Zahl in $[0; 1]$ bekommt man am bequemsten mit dem Befehl frandom(), den hat Marcel entdeckt.

Weise graphisch darstellen und in die Darstellung die annähernde Gaußkurve einzeichnen, deren Parameter aber auch erst noch berechnet werden mussten. All dies hat er sehr schön erledigt, seine Datei simgwshist.mn hast du ja erhalten. Die folgenden beiden Bilder sind mit seiner Datei erstellt.

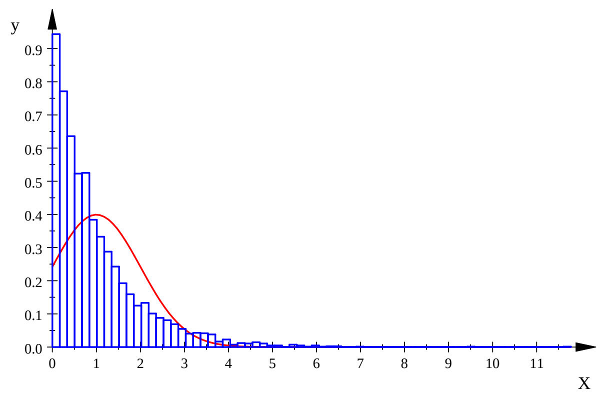


Abbildung 14: 5000 ausgeloste T–Werte

Im ersten Bild sind 5000 ausgeloste T–Werte graphisch dargestellt. Das Histogramm sieht natürlich nicht aus wie eine Gaußkurve, sondern wie die Dichtefunktion einer exponentialverteilten Zufallsgröße. So soll es ja auch sein. Im nächsten Bild dann hat Sebastian 5000 Werte von S_{50} berechnet und verarbeitet. Dazu mussten 5000–mal jeweils 50 T–Werte ausgelost und aufaddiert werden. Dieses Histogramm wird schon ganz gut durch die eingezeichnete Gaußkurve angenähert, wie es der Zentrale Grenzwertsatz in Aussicht stellt. Wenn du Sebastians Programm mit wachsenden n–Werten laufen lässt, siehst du sehr schön, wie die Histogramme ihrer Gaußkurve immer ähnlicher werden. Mache das ruhig, wenn es wieder regnet.

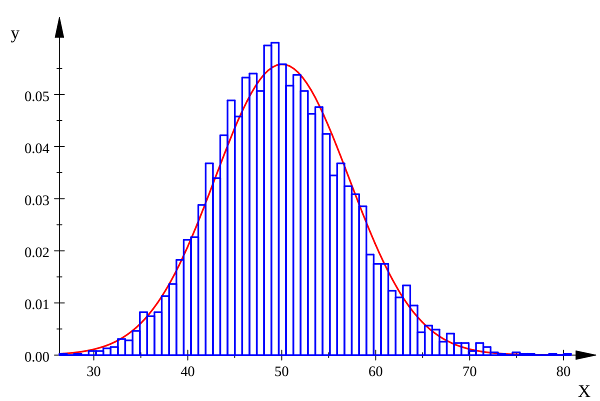


Abbildung 15: 5000 ausgeloste Werte von $S_{50} = \sum_{k=1}^{50} T_k$

11 Klausur Nr. 4 (Nachschreibversion) vom 22. Juni 2010

1. Es sei X eine stetige Zufallsgröße mit Dichtefunktion f. Dabei ist $f(x) = \frac{3}{2}x^2$ für $-1 \leq x \leq 1$ und $f(x) = 0$ sonst.

(a) Zeichne den Graphen von f und begründe, dass $E(X) = 0$ ist.

(b) Bestimme die Verteilungsfunktion F.

(c) Berechne die Varianz von X. [Zwischenergebnis: $V(X) = \frac{3}{5}$]

(d) Es sei $S = X_1 + X_2 + \cdots + X_{100}$ Summe unabhängiger Kopien von X. Bestimme $E(S)$ und $V(S)$.

(e) Berechne $P(-10 \leq S \leq 10)$.

2. Es sei T der Zerfallszeitpunkt eines zufällig gewählten Teilchens einer radioaktiven Substanz. Die Dichtefunktion von T ist die Funktion f mit

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

Berechne den Erwartungswert von T, $P(T > 2)$ und die Halbwertszeit der Substanz.

3. Wir besuchen die bekannte Fabrik für Plastikflaschen. Eine als erste Wahl verkaufte Partie darf höchstens 5% Ausschuss enthalten; enthält sie mehr als 5% Ausschuss, darf sie nur noch als zweite Wahl mit Preisabschlag verkauft werden. Wird eine gute Partie mit Preisabschlag verkauft, entsteht dem Hersteller ein Verlust. Wird eine schlechte Partie als erste Wahl verkauft, riskiert der Hersteller das Vertrauen seiner Kunden, deshalb darf das nur selten vorkommen. Die Entscheidung, ob eine Lieferung erste oder zweite Wahl ist, soll mit Hilfe eines Tests mit $\alpha = 0,025$ getroffen werden. Es werden 100 Flaschen untersucht. Konstruiere einen geeigneten Test und berechne die Wahrscheinlichkeiten p_1, dass eine Lieferung mit 2% Ausschuss als zweite Wahl eingestuft wird, und p_2, dass eine Lieferung mit 10% Ausschuss als erste Wahl eingestuft wird.

4. Die Auszählung der bis zum 18. Juni in unserem Kurs angefallenen Fehlstunden ergab in etwa dieses Ergebnis:

gefehlte Doppelstunden	0	1	2	3	4	5	6
Anzahl der Schüler	6	12	2	2	1	3	2

Mag ja sein, dass die Anzahl der Doppelstunden, die ein Schüler des Kurses fehlt, nur vom Zufall abhängt! Wir vergleichen die erhobenen Daten mit dem, was zu erwarten wäre, wenn die Anzahl der gefehlten Doppelstunden eines zufällig gewählten Schülers des Kurses eine poissonverteilte Zufallsgröße X mit Parameter $\lambda = \frac{53}{28}$ wäre[7].

(a) Wie kommt man auf den angegebenen Wert von λ ?

(b) Berechne $P(X = k)$ für $k = 0, 1, 2, \ldots, 6$.

(c) Zeichne das Histogramm von X für den Bereich $0 \leq k \leq 6$.

(d) Wie groß ist $P(X > 6)$, und warum frage ich danach?

(e) Stelle die Tabellenwerte graphisch so dar, dass ein Vergleich mit dem Histogramm von X möglich ist, und tue deine Meinung kund, ob die Annahme einer Poissonverteilung vernünftig ist. Harte Fakten könntest du erst liefern, wenn Jonathan sein Referat schon gehalten hätte. Im Prinzip kannst du hier nur ein wenig herumschwadronieren, also tue dir keinen Zwang an. Es sollte allerdings gefällig sein, was du aufschreibst.

[7] Im Buch wird dieser Wert mit μ bezeichnet.

5. Beschreibe den Zufallsversuch, der zum Bertrandschen Paradoxon führt, und erläutere, worin das Problem bei diesem Versuch besteht.

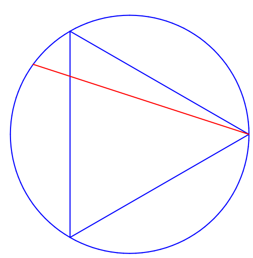
6. Es soll ein Punkt der Einheitskreisscheibe zufällig gewählt werden Kunz regelt das so: Er betrachtet das Rechteck mit der Breite 2π und der Höhe 1. Dann wählt er zufällig eine Zahl $t \in [0; 2\pi]$ und zufällig eine Zahl $r \in [0, 1]$, so erhält er einen zufällig gewählten Punkt (t, r) des Rechtecks. Sein ausgeloster Punkt der Kreisscheibe ist dann der Punkt $Q(r\cos(t); r\sin(t))$. Zeichne Kreis und Rechteck, markiere die folgenden Punktmengen im Rechteck und die zugehörigen Punktmengen im Einheitskreis und gib jeweils die Wahrscheinlickeit an, dass ein Punkt diesr Menge ausgelost wird. Stelle sicher, dass dein Taschenrechner auf Bogenmaß steht!

   (a) Der Punkt mit $r = \frac{1}{2}$ und $t = 1$.
   (b) Ein Punkt mit $t = \frac{1}{2}\pi$, $r \in [0; 1]$ beliebig.
   (c) Ein Punkt mit $r \in [0; \frac{1}{2}]$, $t \in [0, 2\pi]$ beliebig.
   (d) Ein Punkt mit $r \in [0, 1]$ beliebig, $t \in [\pi, \frac{3}{2}\pi]$ beliebig.

   Meinst du, dass diese Konstruktion einen zufällig gewählten Punkt der Kreisscheibe liefert?

12 Das Bertrandsche Paradoxon

Gegeben sei der Einheitskreis mit einbeschriebenem gleichseitigen Dreieck. Gesucht ist die Wahrscheinlichkeit, dass eine zufällig gewählte Sehne des Kreises länger ist als die Seite des Dreiecks.



Diese Aufgabe wurde von Joseph Bertrand gestellt, einem französischen Mathematiker des 19. Jahrhunderts. Ihr habt schnell eine Lösung gefunden: Man kann den Kreis mit dem Dreieck immer so drehen, dass ein Endpunkt der Sehne ein Eckpunkt des Dreiecks ist. Die beiden anderen Punkte teilen den Kreisbogen in drei Stücke gleicher Länge. Liegt der andere Endpunkt der Sehne auf dem Bogen zwischen den beiden anderen Punkten, ist die Sehne länger als die Dreieckseite, sonst ist sie kürzer oder genau so lang wie die Dreiecksseite. Demnach wäre die gesuchte Wahrscheinlichkeit $\frac{1}{3}$, die Aufgabe wäre gelöst.

Wäre dies die ganze Wahrheit, hätte die Aufgabe nicht diesen Bekanntheitsgrad erreicht. Die zentrale Einsicht, die das Problem vermittelt, ist, dass man mit großer Vorsicht zu Werke gehen muss, wenn man von einem „zufällig gewählten“ Objekt spricht. Gemeint ist damit immer eine Art Gleichverteilung. Es soll ein $\omega \in \Omega$ so gewählt werden, dass für jedes Ereignis $A \subseteq \Omega$

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

gilt. Für endliches Ω haben wir von einem Laplaceversuch gesprochen. Das Ω wäre hier die Menge aller Sehnen des Kreises, und das ist eine recht unübersichtliche Angelegenheit.

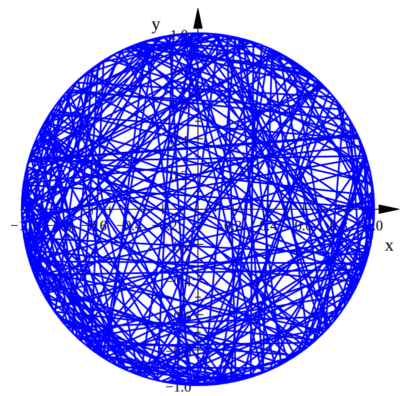


Abbildung 16: 300 mit Marcels Programm „zufällig gewählten“ Sehnen im Einheitskreis

Man muss die Menge aller Sehnen irgendwie so parametrisieren, dass man die Auswahl organisieren kann. Wir diskutieren nun einige Möglichkeiten, wie diese Parametrisierung geschehen kann.

Modell 1 Wir wählen zwei Punkte A und B der Kreislinie zufällig. Die Verbindungsstrecke $\overline{AB}$ ist dann die ausgeloste Sehne. Technisch geht das so, dass wir zwei Zahlen u und v in $[0, 2\pi]$ zufällig wählen und $A = (\cos(u)|\sin(u))$ und $B = (\cos(v)|\sin(v))$ setzen. Zu jeder Sehne gehört also genau ein Punkt (u, v) eines Quadrates mit der Kantenlänge 2π, du denkst am besten an das Quadrat mit den Eckpunkten $(0|0)$, $(2\pi|0)$, $(2\pi|2\pi)$ und $(0|2\pi)$ im uv–System. Abbildung 17 zeigt das Quadrat (verkleinert) mit einem Punkt sowie den Kreis mit der Sehne, die zu dem Punkt gehört.

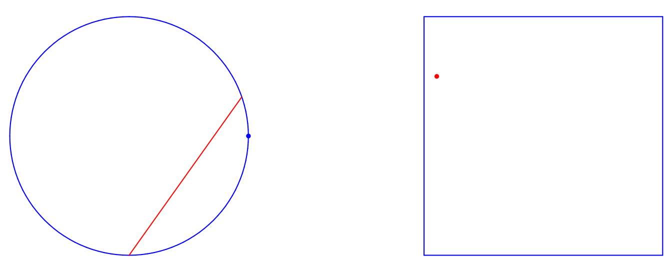


Abbildung 17: Parametrisierung nach Modell 1

Man überlegt sich, dass zu den langen Sehnen genau die Punkte in dem schraffierten Gebiet gehören (siehe Abbildung 18); dieses Gebiet enthält gerade $\frac{1}{3}$ der Fläche des Quadrats. Die Wahrscheinlichkeit, eine lange Sehne auszulosen, ist folglich $\frac{1}{3}$. Wir erhalten den gleichen Wert für die Wahrscheinlichkeit wie beim ersten Ansatz, bei dem wir einen Kreispunkt als Sehnenanfang festhielten und nur den zweiten Kreispunkt als Sehnenendpunkt zufällig wählten.

Modell 2 Jede Sehne ist schon durch ihren Mittelpunkt eindeutig festgelegt: Es sei $C \neq M(0|0)$ ein Punkt der Einheitskreisscheibe. Es gibt eine und nur eine Sehne, die C als Mittelpunkt hat. Man findet sie, indem man in C die Senkrechte auf der Strecke $\overline{MC}$ errichtet. Sieht man von den Durchmessern ab, kann man

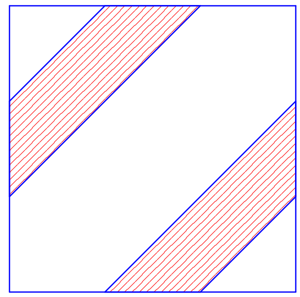


Abbildung 18: Menge der Punkte zu langen Sehnen in Modell 1

die Menge der Sehnen des Einheitskreises durch die Menge der Punkte $\neq M$ der Einheitskreisscheibe parametrisieren. Wir losen nun einen Punkt $C \neq M$ der Einheitskreisscheibe aus und betrachten die Sehne als ausgelost, deren Mittelpunkt C ist. Zu langen Sehnen gehören die Punkte im Innern des Kreises um M mit dem Radius $\frac{1}{2}$. Die Wahrscheinlichkeit, dass eine solche lange Sehne ausgelost wird, ist folglich

$$\frac{\pi \left(\frac{1}{2}\right)^2}{\pi \cdot 1^2} = \frac{1}{4} \quad ,$$

und das verträgt sich nicht mit dem Ergebnis aus Modell 1. Aber das mag ja daran liegen, dass wir die Durchmesser weggelassen haben?!

Modell 3 Wir wählen einen festen Radius des Einheitskreises, zum Beispiel die Verbindungsstrecke der Punkte M und $P(0|1)$. Jede Sehne können wir so drehen, dass ihr Mittelpunkt auf diesem festen Radius zu liegen kommt, deshalb könnten wir doch stellvertretend für alle Sehnen nur die betrachten, deren Mittelpunkt auf dem festen Radius liegt, und aus denen eine zufällig wählen. Wir wählen also einen Punkt C auf $\overline{MP}$ zufällig und betrachten die Sehne als ausgelost, die C als Mittelpunkt hat. Die Wahrscheinlichkeit, eine lange Sehne auszulosen, ist dann $\frac{1}{2}$.

Fazit: Wir müssen die Redeweise von der zufällig gewählten Sehne aufgeben. Ein Zufallsversuch ist erst dann gegeben, wenn präzise beschrieben ist, wie er durchgeführt wird.

Aufgabe Wir wählen einen Punkt $P \neq (0|0)$ der Einheitskreisscheibe, indem wir Zahlen R mit $0 < R \leq 1$ und T mit $0 \leq T < 2\pi$ zufällig wählen und $P = (R\cos(T)|R\sin(T))$ setzen. Dann ist P keineswegs zufällig gewählt, denn für eine Teilmenge A der Einheitskreisscheibe ist im Allgemeinen eben nicht $P(A) = \frac{|A|}{\pi}$. Mache dir dies noch einmal klar, indem du als A die Kreisscheibe A_r um den Nullpunkt mit dem Radius r wählst.

Gesucht ist nun eine eine zweidimensionale Dichtefunktion $f : (x, y) \mapsto f(x, y)$. Für eine vernünftige Teilmenge A der Einheitskreisscheibe ist dann

$$P(A) = \frac{1}{\pi} \int_A f(x, y)\, dx\, dy \quad .$$

Es ist einleuchtend, dass f rotationssymmetrisch sein sollte; alle Punkte auf einem jeden Kreis um den Nullpunkt sollten von f auf dieselbe Zahl abgebildet werden. Es gibt eine solche Funktion f, und du kannst sie finden! Hinweis: Betrachte die Mengen A_r, mit denen du dich oben schon befasst hast.

13 Der χ^2–Test

Wir werfen einen idealen Würfel 24–mal und zählen, wie oft eine 1, eine 2 usw. geworfen wurde. Dann haben wir es mit sechs Zufallsgrößen $H_1, H_2, \ldots, H_6$ zu tun; dabei bezeichnet H_k, wie oft die Augenzahl k geworfen wurde. Natürlich ist $E(H_k) = 4$, das heißt, wenn wir den Versuch sehr oft durchführen, werden wir im Mittel je Durchführung zum Beispiel vier Einsen beobachten. Betrachtet man aber nicht, was im Mittel auf lange Sicht passiert, sondern schaut man sich eine konkrete Durchführung an, bekommt man einen Wertesatz $(h_1, h_2, h_3, h_4, h_5, h_6)$, und die einzelnen h_k können von 4 deutlich abweichen. Es ist lediglich klar, dass $h_1 + h_2 + \cdots + h_6 = 24$ ist.

Um die Abweichung quantitativ zu erfassen, bildet man folgende Messgröße:

$$\sum_{k=1}^{6} \frac{(h_k - 4)^2}{4}$$

Je größer diese Kennzahl ist, desto größer ist die Abweichung. – Für diese Kennzahl gebe ich zunächst eine allgemeinere Beschreibung:

12 Definition
Es sei $X : \Omega \to \mathbb{R}$ eine Zufallsgröße mit Wertevorrat $X(\Omega) = \{x_1, x_2, \ldots, x_n\}$. Wir losen N Werte von X aus, und es bezeichne H_k, wie oft der Wert x_k beobachtet wurde. Dann ist H_k eine Zufallsgröße mit Erwartungswert $E(H_k) = N \cdot P(X = x_k)$. Wir definieren zu X und N die Zufallsgröße

$$Y := \sum_{k=1}^{n} \frac{\left(H_k - E(H_k)\right)^2}{E(H_k)} \quad . \tag{17}$$

Kehren wir zurück zu unserem Würfelbeispiel. Jede konkrete Durchführung des Versuchs liefert eine konkrete Zahl. Marcel hat den Versuch in einer Simulation 100–mal durchführen lassen und damit einhundert dieser Kennzahlen gewonnen und in einem Histogramm dargestellt.[8]

So weit, so gut. Nun lebte vor etwa hundert Jahren ein Karl Pearson, und er hat eine Familie von Verteilungen entdeckt, die man χ^2–Verteilungen nennt. Es sind stetige Verteilungen, und ihre Dichten sehen so aus:

$$f_m(x) = \begin{cases} \frac{1}{2^{\frac{1}{2}m}\Gamma(\frac{1}{2}m)} x^{\frac{m-2}{2}} e^{-\frac{1}{2}x} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases} \tag{18}$$

Dabei ist Γ die berühmt–berüchtigte Gammafunktion, die du ehrfürchtig betrachten und dann ignorieren kannst. Der Parameter m wird Anzahl der Freiheitsgrade genannt. Nun kommt der Clou der Sache: Pearson hat bewiesen, dass

$$P(Y \le y) \approx F_m(y) = \int_{-\infty}^{y} f_m(x)\,dx \tag{19}$$

ist, wenn man $m = n-1$ wählt. Hier wird die Wortwahl nachvollziehbar: Die Summe der H_k muss stets N ergeben, also sind in gewissem Sinne nur $n-1$ der H_k frei.[9]

In der Praxis läuft die Sache nun wie folgt ab. Der Anwender hat einen Wertesatz und vermutet, dass es sich um einen Satz von N ausgelosten Werten einer Zufallsgröße X handelt. Beim Würfelbeispiel wäre X gleichverteilt, in der Aufgabe mit der

[8] Siehe die Datei chi.mn .
[9] Ist X poissonverteilt, muss man sogar $m = n - 2$ setzen.

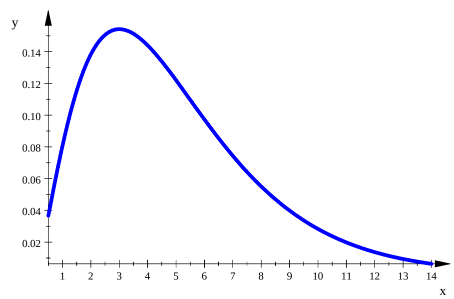


Abbildung 19: Dichte der χ^2–Verteilung mit $m = 5$

Verkehrszählung aus dem Buch sollte X poissonverteilt sein. Der Anwender berechnet nun den Wert y der Zufallsgröße Y aus Gleichung 17 und bestimmt den Wert des Parameters m (im Würfelbeispiel ist $m = 5$, beim Verkehrszählungsbeispiel ist $m = 9 - 2 = 7$). Mit Hilfe einer Tabelle stellt er nun fest, ob y im rechten 5%–Schwanz der entsprechenden χ^2–Verteilung liegt. Dann ist die Kenngröße nämlich auffällig groß, und die Hypothese, der Wertesatz stamme von der vermuteten Zufallsgröße X, wird verworfen. Liegt y dagegen in dem Bereich von 0 bis $F_m^{-1}(0,95)$, geht man davon aus, dass es sich um zufällige Abweichungen handelt, und behält die Hypothese, die Werte stammten von X, bei. Bei der ganzen Sache handelt es sich also um einen einseitigen Test mit Nullhypothese X und $\alpha = 0,05$. Man macht einen einseitigen Test, weil eben große Werte von Y verdächtig sind.

Das war in etwa das, was Jonathan vorgetragen hat. Hier enden nun meine Aufzeichnungen zur Stochastik; ich hoffe, dass sie dir nützlich waren oder noch werden. – Im nächsten Halbjahr beginnen wir ein ganz anderes Thema.