Stochastik

Michael Spielmann

Inhaltsverzeichnis

1 Zur Unterscheidung der Begriffe Statistik und Stochastik 4

2 Die Aufgabe der Statistik 4

3 Grundlegendes zur Stochastik 5

3.1 Zufallsexperiment, Ereignis 5

3.2 Grundbegriffe der Mengenlehre 5

4 Begriff der Wahrscheinlichkeit 6

4.1 Probleme der Definition, Geschichte 6

4.2 Wahrscheinlichkeit und relative Häufigkeit 6

4.3 Zusammengesetzte Ereignisse 6

4.4 Mehrstufige Versuche 7

4.5 Eigenschaften der Wahrscheinlichkeit 7

5 Bernoullikette 7

6 Binomialkoeffizienten 8

6.1 Herleitung der Formel 8

6.2 Das Pascaldreieck 9

7 Grundbegriffe der Stochastik 10

8 Anwendung der Binomialverteilung 11

8.1 Hypothesentest 11

8.2 Der gefälschte Würfel 11

8.3 Das Prinzip des Testes 12

8.4 Die Fehlerarten 12

8.5 Zweiseitiger Test 12

8.6 Grundbegriffe des Testens 13

9 Wirkungskurve des Hypothesentests **13**

10 Zum Gebrauch der Tabellen **17**

10.1 Wahrscheinlichkeitsfunktion 17

10.2 Summenfunktion 17

10.3 Bestimmung des kritischen Gebietes zu gegebenem α-Fehler . 19

11 Erwartungswert und Varianz **20**

11.1 beliebige Verteilung 20

11.2 lineare Transformation 21

11.3 Summe von Zufallsvariablen 22

11.4 Produkt von Zufallsvariablen 24

11.5 Binomialverteilung 25

12 Ungleichung von Tschebyscheff **26**

13 Gesetz der großen Zahl **28**

14 Die Poisson-Verteilung **29**

14.1 Geschichtliches 29

14.2 Näherung für die Binomialverteilung 29

14.3 Die Wahrscheinlichkeitsfunktion Poisson-Verteilung 31

14.4 Erwartungswert und Varianz der Poisson-Verteilung 31

15 Die Normalverteilung **33**

15.1 Geschichtliches 33

15.2 Näherung für die Binomialverteilung 33

15.3 Ein Weg zur Funktion $\varphi(x)$ 34

15.4 Die Stirling-Formel und ihre Verwandte 36

15.4.1 StirlingI 36

15.4.2 Stirling II 40

15.5 stetige Verteilung, Wahrscheinlichkeitsdichte 41

15.6 Erwartungswert und Varianz stetiger Zufallsgrößen 42

15.7 Die Gauss-Verteilung 42

16 weitere Verteilungen **44**

17 Normalverteilung in der Praxis **45**

17.1 Qualitätskontrolle 45

17.2 Stetigkeitskorrektur 46
17.3 Toleranzgrenzen 46

18 Unabhängigkeit **50**

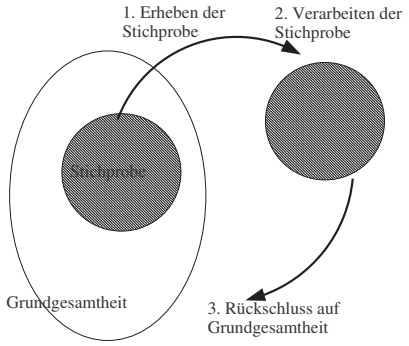
19 bedingte Wahrscheinlichkeit **50**

1 Zur Unterscheidung der Begriffe Statistik und Stochastik

Statistik befasst sich mit dem Sammeln, Beschreiben und Auswerten von Daten. Zu diesem Zweck entnimmt man der Grundgesamtheit Stichproben und bereitet diese auf. Diese Tätigkeiten werden mit dem Begriff deskriptive Statistik umschrieben.
Stochastik (deutsch: Wahrscheinlichkeitsrechnung) ermöglicht Aussagen über die Grundgesamtheit auf der Basis der Stichprobenauswertung. Oder aber sie liefert rein theoretische Erwägungen über das wahrscheinliche Eintreten gewisser Ereignisse.

2 Die Aufgabe der Statistik

In vielen Bereichen der Wissenschaft und Technik will man Aussagen über eine Gesamtheit machen, die zu groß ist, um sie ganz zu erfassen. Man untersucht stattdessen eine Stichprobe. Sie wird geordnet, klassiert, graphisch veranschaulicht, zahlenmäßig ausgewertet. Dazu berechnet man Häufigkeiten, Mittelwerte, Streuungen.
Die absolute Häufigkeit eines Ereignisses gibt an, wie oft das Ereignis tatsächlich eingetreten ist. Die relative Häufigkeit benötigt man als Vergleichszahl bei Serien oder Zufallsexperimenten unterschiedlichen Umfangs. Unter 100 Versuchen trat 55-mal G ein, dann ist die absolute Häufigkeit 55 und die relative Häufigkeit ist $\frac{55}{100} = 55\% = 0,55$. Der Mittelwert soll die Stichprobe als einzelner Zahlenwert repräsentieren. Die Information der gesamten Stichprobe wird dadurch reduziert zugunsten einer kompakten Darstellung. Streuungen, also Abweichungen vom Mittelwert oder Spannweiten, bestimmt man, wenn man Stichproben gleichen Mittelwertes vergleichen will oder über die Struktur der Stichprobe mehr aussagen will als der Mittelwert alleine zulässt.
Mit dem Rückschluss auf die Grundgesamtheit beschäftigt sich die Beur-



teilende Statistik. Sie benötigt zur Begründung die Wahrscheinlichkeitsrechnung.

Wir notieren hier zur Erinnerung die Formeln für Mittelwert und empirische Varianz. Die Streuung oder Standardabweichung ist die Wurzel aus der Varianz.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad v = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad s = \sqrt{v}$$

Dabei war n der Stichprobenumfang; die x_i sind die Merkmalwerte, in der Physik würde man Messwerte sagen. Falls Merkmalwerte mehrfach vorkommen, fasst man sie zu k Klassen mit n_i Werten zusammen und kommt dann zu einer etwas anderen Darstellung der Formeln.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i \qquad v = \frac{1}{n}\sum_{i=1}^{k} (x_i - \bar{x})^2 \cdot n_i$$

3 Grundlegendes zur Stochastik

3.1 Zufallsexperiment, Ereignis

Das Werfen eines Würfels ist ein **Zufallsexperiment**. Man kann nicht mit Bestimmtheit das **Ergebnis** des Wurfes vorhersagen. VIER ist ein mögliches Ergebnis. Verschiedene mögliche Ergebnisse kann man zu einem **Ereignis** wie zum Beispiel GERADE ZAHL zusammenfassen. Das Ereignis GERADE ZAHL ist eingetreten, wenn der Wurf ZWEI oder VIER oder SECHS lieferte. Die Menge aller Ergebnisse eines Zufallsexperimentes heißt **Ergebnismenge**. Ereignisse aus nur einem Element heißen **Elementarereignis**. Die Ergebnismenge selbst heißt auch sicheres Ereignis. Es tritt mit Sicherheit ein, wenn überhaupt gewürfelt wird. Ein unmögliches Ereignis wird durch die leere Menge dargestellt. Es kann nicht eintreten. Zwei einander ausschließende Ereignisse nennt man **Gegenereignisse**. Beim Werfen mit 2 Würfeln besteht ein Ergebnis aus einem Zahlenpaar.

3.2 Grundbegriffe der Mengenlehre

$W = \{1, 2, 3, 4, 5, 6\}$ ist die Menge der Zahlen eines Würfels. Die Zahlen sind Elemente von W; zum Beispiel ist $3 \in W$. Die Menge aller geraden Zahlen in W ist eine Teilmenge von W, $G = \{2, 4, 6\} \subset W$. Die Menge aller gemeinsamen Elemente zweier Mengen nennen wir Schnittmenge. $D = \{3, 6\}$ und $G = \{2, 4, 6\}$ haben die 6 gemeinsam. $D \cap G = \{3, 6\} \cap \{2, 4, 6\} = \{6\}$. Zwei Mengen mit leerer Schnittmenge nennen wir disjunkt. Vereinigen wir die Elemente zweier Mengen zu einer einzigen, so entsteht die Vereinigungsmenge. $D = \{3, 6\}$ und $G = \{2, 4, 6\}$ haben die Vereinigungsmenge $D \cup G = \{3, 6\} \cup \{2, 4, 6\} = \{2, 3, 4, 6\}$. Mehrfach vorkommende Elemente werden nur einmal notiert.

4 Begriff der Wahrscheinlichkeit

4.1 Probleme der Definition, Geschichte

Der Begriff der Wahrscheinlichkeit bedarf einer Definition. Jakob Bernoulli (1654-1705) erklärte ihn als *Grad der Gewissheit*. Umgangssprachlich wird er durch das Wort *Chance* ziemlich gut erfasst.
Wenn wir von Wahrscheinlichkeiten beim Würfeln sprechen, sehen wir von speziellen Gegebenheiten ab, wir abstrahieren, oder idealisieren. Der ideale Würfel ist total symmetrisch und homogen und nichts stört seine Gleichmäßigkeit. So ist garantiert, dass jede Seite mit der gleichen Chance erscheint. Das ist unser Argument für die Zahl $\frac{1}{6}$ als Wahrscheinlichkeit für das Auftreten von z.B. der ZWEI bei einem Wurf. Diese Argumentation geht auf Pierre-Simon Laplace (1749-1829) zurück, der 1812 in seinem Buch *Essai philosophique sur le probabilités* erklärte, die Wahrscheinlichkeit sei der Quotient aus der Zahl der günstigen und der Zahl der möglichen Fälle. Doch ist auch dies keine Definition, sondern eher eine Rechenanweisung. Bis die Wahrscheinlichkeitsrechnung im Sinne einer mathematischen Theorie ausgearbeitet und axiomatisch fundiert dargestellt werden konnte, vergingen noch mehr als 100 Jahre. Andrej Nikolaiewitsch Kolmogorov (1903-1987) entwickelte ein Axiomensystem für die Wahrscheinlichkeitsrechnung, das er 1933 in der Schrift *Grundbegriffe der Wahrscheinlichkeitsrechnung* veröffentlichte. Er benutzt ganz wesentlich die Mengenlehre, und er greift auf die Maß-Theorie, eine Verallgemeinerung der Integralrechnung zurück.

4.2 Wahrscheinlichkeit und relative Häufigkeit

Wir können mit Laplace die Wahrscheinlichkeit durch Symmetrieüberlegung ermitteln. Bei einem Würfel oder einer Münze können wir uns den idealen Fall vollkommener Symmetrie vorstellen; dann sind die sechs Seiten des Würfels gleichwertig, und sie fallen jede mit der Wahrscheinlichkeit $\frac{1}{6}$, und die beiden Seiten der Münze mit der Wahrscheinlichkeit $\frac{1}{2}$. Das klappt aber nicht bei jedem Zufallsexperiment. Werfen wir mit einer Heftzwecke, können wir von Symmetrie nichts voraussetzen. Dann sind wir gezwungen, Näherungswerte zu benutzen. Diese Näherungswerte finden wir mit Hilfe der relativen Häufigkeit. Die Beobachtung, dass sich die relative Häufigkeit bei einer langen Versuchsreihe stabilisiert, ist sehr elementar. Man nennt sie heute **empirisches Gesetz der großen Zahl**. Sie wurde von Bernoulli theoretisch begründet. Er bewies das sogenannte Gesetz der Großen Zahl, welches besagt, dass die Wahrscheinlichkeit, dass die relative Häufigkeit in immer engeren Grenzen liegt, mit steigendem Stichprobenumfang gegen 1 konvergiert.

4.3 Zusammengesetzte Ereignisse

Zieht man eine Karte aus einem Skatspiel und legt Wert auf das Ereignis „rote Karte oder Dame", so spricht man von zusammengesetzten Ereignis-

sen. Die Wahrscheinlichkeit wird prinzipiell durch Addition bestimmt; man muss aber die doppelt gezählten Elemente anschließend wieder subtrahieren. Im Beispiel wäre die Wahrscheinlichkeit $\frac{16}{32} + \frac{4}{32} - \frac{2}{32}$, denn es gibt 16 rote Karten, 4 Damen, aber 2 Damen sind auch rot.

4.4 Mehrstufige Versuche

Zieht man mehrere Karten hintereinander, oder würfelt man mehrmals, so spricht man von mehrstufigen Versuchen. Zur graphischen Darstellung ist ein Baum geeignet. Die Wahrscheinlichkeit längs der Pfade bestimmt man durch Multiplikation. Wie groß ist die Wahrscheinlichkeit, zweimal hintereinander eine drei zu würfeln? Wir können sagen, in $\frac{1}{6}$ aller Fälle würfeln wir eine drei; haben wir sie einmal geworfen, dann würfeln wir in $\frac{1}{6}$ aller dieser Fälle wieder eine drei. $\frac{1}{6}$ von $\frac{1}{6}$ ist aber $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$.

4.5 Eigenschaften der Wahrscheinlichkeit

Eine Wahrscheinlichkeit P ist eine Zahl mit $0 \leq P \leq 1$. Wenn S die Ergebnismenge ist, gilt $P(S) = 1$. Für das unmögliche Ereignis $\emptyset$ ist $P(\emptyset) = 0$.
Gilt $A \cap B = \emptyset$ dann ist $P(A \cup B) = P(A) + P(B)$.
Gilt $A \cap B \neq \emptyset$ dann ist $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
Gilt $A \cup \bar{A} = S$ und $A \cap \bar{A} = \emptyset$, dann ist $P(A) = 1 - P(\bar{A})$.

5 Bernoullikette

Wenn man sich dafür interessiert, ob ein bestimmtes Ereignis eingetreten ist oder nicht, dann rechnet man eigentlich nur mit zwei Ausgängen, wie zum Beispiel Gewinn oder Verlust. Einen Versuch mit nur zwei Ausgängen nennen wir Bernoulli-Versuch; die beiden Ausgänge sind Ereignis und Gegenereignis, die üblicherweise mit zum Beispiel A und $\bar{A}$ beschrieben werden. Diesen Ergebnissen ordnen wir die Wahrscheinlichkeiten p und $1 - p$ zu. Wenn wir einen Bernoulli-Versuch n-mal hintereinander ausführen, entsteht eine Bernoullikette. Bei Bernoulliketten interessiert man sich im allgemeinen für die Wahrscheinlichkeit des k-maligen Auftretens von A bei n Versuchen. Wir bestimmen eine Formel für diese Wahrscheinlichkeit.
Wir drehen ein Glücksrad mit den Feldern *rot* und *nicht rot*, oder A und $\bar{A}$. *rot* soll die Wahrscheinlichkeit $\frac{1}{3}$ haben. Wenn wir zweimal drehen, also $n = 2$ setzen, kann $k = 2, 1, 0$ sein, also *rot* oder A 2-mal, 1-mal, 0-mal auftreten. Die Wahrscheinlichkeit beschreiben wir mit $B_{n,p}(k)$, das bedeutet: Bernoullikette, n-mal ausgeführt, Grundwahrscheinlichkeit p, beobachte k-maliges Auftreten von A.

$$\begin{array}{lll} n=2 & k=2 & B_{2,\frac{1}{3}}(2) = (\frac{1}{3})^2 \\ & k=1 & B_{2,\frac{1}{3}}(1) = 2 \cdot (\frac{1}{3}) \cdot (\frac{2}{3}) \\ & k=0 & B_{2,\frac{1}{3}}(0) = (\frac{2}{3})^2 \end{array}$$

$$\begin{array}{lll} n=3 & k=3 & B_{3,\frac{1}{3}}(3) = (\frac{1}{3})^3 \\ & k=2 & B_{3,\frac{1}{3}}(2) = 3 \cdot (\frac{1}{3})^2 \cdot (\frac{2}{3}) \\ & k=1 & B_{3,\frac{1}{3}}(1) = 3 \cdot (\frac{1}{3}) \cdot (\frac{2}{3})^2 \\ & k=0 & B_{3,\frac{1}{3}}(0) = (\frac{2}{3})^3 \end{array}$$

Das Bauprinzip ist schnell zu durchschauen. Nur die Koeffizienten müssen wir noch näher bestimmen. Sie scheinen den Zahlen des Pascal-Dreiecks zu entsprechen.

$$\begin{array}{ccccccc} & & & 1 & & & \\ & & 1 & & 1 & & \\ & 1 & & 2 & & 1 & \\ 1 & & 3 & & 3 & & 1 \end{array}$$

$$B_{n,p}(k) = \text{Koeffizient} \cdot p^k \cdot (1-p)^{n-k}$$

6 Binomialkoeffizienten

Die Zahlen des Pascal-Dreiecks geben die Anzahl der Wege für das gewünschte Ereignis an. Wir nennen sie *Binomialkoeffizienten* und schreiben $\binom{n}{k}$ und lesen *„n über k“*. Wir berechnen sie nach der Formel $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

6.1 Herleitung der Formel

Als Modell stellen wir uns vor, wir sollten n Plätze mit k mal A und $n-k$ mal $\bar{A}$ belegen. Das entwickeln wir schrittweise.

a) n Plätze belegen mit n unterscheidbaren Elementen $A_1, \ldots, A_n$, die wir abgekürzt schreiben mit $1, 2, \ldots n$.

$n=2$	12	2 Möglichkeiten
	21	
$n=3$	123	
	132	
	213	6 Möglichkeiten
	231	
	312	
	321	

Die 4 kann zusätzlich auf 4 Arten einsortiert werden. Daher gibt es $6 \cdot 4 = 24$ oder $2 \cdot 3 \cdot 4 = 4!$ Möglichkeiten.

Auf $n!$ Arten kann man n Plätze mit n unterscheidbaren Elementen belegen.

b) n Plätze belegen mit k gleichen also nicht unterscheidbaren und $n-k$ unterscheidbaren Elementen

$n = 2$ siehe oben

$n = 3$ mit $2 \equiv 3 \equiv 0$

100
100
010
001
010
001

Von den 6 Möglichkeiten sind je 2 identisch.

Es gibt also $\frac{3!}{2} = 3$ Möglichkeiten.

Wir springen zu $n = 7$ und unterscheiden nur 1,2,3,4.

$n = 7$ mit $5 \equiv 6 \equiv 7 \equiv 0$

Bei einer festen Kombination z.B. 1234000 gab es vorher 3! Möglichkeiten, 5, 6, 7, zu belegen; die sind jetzt mit 0 belegt; also sind 3! Möglichkeiten entfallen.

Es gibt also $\frac{7!}{3!}$ Möglichkeiten.

Auf $\frac{n!}{k!}$ Arten kann man n Plätze mit k gleichen und $n-k$ unterscheidbaren Elementen belegen.

c) Jetzt ist sicher klar, wie es weitergeht.

n Plätze belegen mit k gleichen also nicht unterscheidbaren und $n-k$ anderen aber auch nicht unterscheidbaren Elementen

Nun entfallen wieder $(n-k)!$ Möglichkeiten; wir dividieren sie heraus.

$1234\underbrace{567}_{3!} \longrightarrow \underbrace{1234}_{4!}000 \longrightarrow 11110000$

Auf $\frac{n!}{k!(n-k)!}$ Arten kann man n Plätze mit k gleichen und $n-k$ anders gleichen Elementen belegen.

6.2 Das Pascaldreieck

Das Pascaldreieck liefert die sogenannten Binomialkoeffizienten. Sie entstanden als Koeffizienten von Binomen (daher der Name). Was haben sie mit der Anordnung von gewissen A und $\bar{A}$ zu tun?

$(a+b)^3 = a^3+3a^2b+3ab^2+b^3$ hat denselben Aufbau wie die Wahrscheinlichkeit der Bernoulli-Kette. Nur, dass hier a statt p und b statt $1-p$ steht. Im Wahrscheinlichkeitsbaum haben wir längs eines Weges eine bestimmte Abfolge von A und $\bar{A}$ gesehen. Auf die Reihenfolge aber kommt es nicht an. Wir haben also die Wege mit entsprechender Anzahl von A und $\bar{A}$ zusammengefasst. Nichts anderes tun wir bei den Binomen. Beim Ausmultiplizieren wird

jedes Glied der einen mit jedem Glied der anderen Klammer multipliziert. So kommen wir hier zu Anordnungen nicht von A und $\bar{A}$, aber doch von a und b, was natürlich für die Formel gleichgültig ist. Der Binomialkoeffizient gibt also eine bestimmte Anzahl von (a, b)-Kombinationen an, wenn man an Binome denkt, oder $(A, \bar{A})$-Kombinationen, wenn man an Bernoulli denkt, oder $(links, rechts)$-Kombinationen, wenn man nur an den Baum denkt.

7 Grundbegriffe der Stochastik

Zufallsversuch:	ein Versuch mit ungewissem Ausgang
Zufallsexperiment:	anderes Wort für Zufallsversuch
Ergebnis:	spezieller Ausgang eines Zufallsversuches
Ereignis:	Kombination von Ergebnissen eines Zufallsversuches
Elementarereignis:	ein einzelnes Ergebnis; seine Wahrscheinlichkeit ist meist leicht zu bestimmen
Wahrscheinlichkeit:	Maß für die Chance des Eintretens eines Ereignisses
Wahrscheinlichkeits- bestimmung:	1. aus Symmetrieüberlegungen 2. aus relativer Häufigkeit mit Stabilisierung
Wahrscheinlichkeits- baum:	graphische Veranschaulichung eines mehrstufigen Versuches
Häufigkeit, absolute:	Anzahl des Eintretens eines Ergebnisses
Häufigkeit, relative:	Anzahl des Eintretens eines Ergebnisses gemessen an der Anzahl aller möglichen Ergebnisse des Versuches
Stabilisierung:	bei großer Zahl der Versuche erfahrungsgemäße Annäherung der relativen Häufigkeit an einen festen Wert, (sogenanntes empirisches Gesetz der großen Zahl)
Laplace-Experiment:	Zufallsversuch mit gleichwahrscheinlichen Ergebnissen
Bernoulli-Versuch:	Zufallsversuch mit nur 2 Ergebnissen
Bernoulli-Kette:	wiederholtes Ausführen eines gleichen Bernoulli-Versuches
Zufallsvariable:	Zuordnung von Ereignis und Zahl
Bernoulli-Variable:	Zufallsvariable mit den Werten 0 und 1, auch Indikator genannt
Wahrscheinlichkeits- verteilung:	Tabelle oder Zuordnung von Ergebnis und Wahrscheinlichkeit

8 Anwendung der Binomialverteilung

8.1 Hypothesentest

Ein Hypothesentest versucht, Fragen wie die folgenden zu beantworten: Hat sich der Mittelwert einer Produktion von Werkstücken verschoben? Ist der Ausschussanteil wesentlich verschieden vom vereinbarten Prozentsatz? Weicht eine Verteilung von der Normalverteilung ab? Für diese Fragen gibt es verschiedene Methoden. Die zweite Frage können wir mit der Binomialverteilung beantworten.

8.2 Der gefälschte Würfel

Beim letzten Mensch-ärgere-dich-nicht-Spiel hatte ich das Gefühl, viel zu wenige Sechsen zu werfen. Hatte ich einen schlechten Würfel erwischt oder einfach Pech?
Ein Hypothesentest liefert eine Antwort.
Man vereinbart folgende Bedingungen für den Test: zeigt der Würfel bei 30 Würfen weniger als 3 Sechsen, dann hält man ihn für gefälscht.
Wie groß ist die Wahrscheinlichkeit, dass weniger als 3 Sechsen fallen, wenn der Würfel in Ordnung ist? Mit Bernoulli können wir diese Wahrscheinlichkeit berechnen.

$$n = 30; \quad p = \frac{1}{6}; \quad k \leq 2$$

$$P(k \leq 2) = B_{30;\frac{1}{6}}(0) + B_{30;\frac{1}{6}}(1) + B_{30;\frac{1}{6}}(2)$$

$$= 0,00421272 + 0,02527632 + 0,07330133 = 0,10280234$$

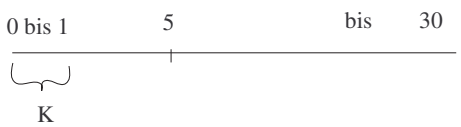
Eigentlich würden wir bei 30 Würfen etwa 5 Sechsen erwarten. Die Wahrscheinlichkeit, dass ein normaler Würfel bei 30 Würfen weniger als 3 Sechsen zeigt, beträgt 10%. Anders formuliert: Nach der obigen Entscheidungsregel beträgt die Wahrscheinlichkeit, dass wir den Würfel verdächtigen, obwohl er in Ordnung ist, 10%. Wenn uns diese Wahrscheinlichkeit zu groß ist, müssen wir den Test verändern; wir verkleinern k.

$$n = 30; \quad p = \frac{1}{6}; \quad k \leq 1$$

$$P(k \leq 1) = B_{30;\frac{1}{6}}(0) + B_{30;\frac{1}{6}}(1)$$

$$= 0,00421272 + 0,02527632 = 0,02950101$$

Die Wahrscheinlichkeit, dass ein normaler Würfel bei 30 Würfen weniger als



2 Sechsen zeigt, beträgt 3%. Dies reicht uns, um den Würfel gegebenenfalls mit geringem Fehler, also großer Sicherheit abzulehnen.

Jetzt führen wir den Test aus. Wir werfen den Würfel 30-mal. Zeigt er weniger als 2-mal Sechs, lehnen wir ihn ab, zeigt er nicht weniger als 2-mal Sechs, können wir ihn nicht ablehnen; es sei denn, wir wollen einen größeren Fehler riskieren.

8.3 Das Prinzip des Testes

Wir formulieren eine Hypothese, mit der wir rechnen können. Es ist allerdings eine Hypothese, deren Ablehnung wir rechtfertigen wollen. Nun gibt es zwei Möglichkeiten, weiterzuarbeiten: wir bestimmen einen Ablehnungsbereich und berechnen den dazugehörigen Fehler; oder: wir geben einen maximalen Fehler vor und bestimmen dazu den passenden Ablehnungsbereich. Üblich ist der zweite Weg. Jetzt führen wir den Test aus und entscheiden, ob die Hypothese abzulehnen ist.

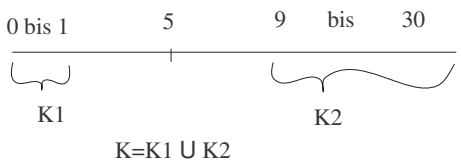
Wir müssen hier auf die Formulierung achten. Mit dem Hypothesentest beweisen wir nichts, wir entscheiden lediglich, ob wir eine Hypothese ablehnen oder nicht. Man darf also auch nicht sagen, der Test habe ergeben, dass der Würfel gefälscht sei. Der Test hat ergeben, dass die Hypothese, der Würfel sei in Ordnung, abgelehnt werden muss.

8.4 Die Fehlerarten

Bei der Entscheidung des Testes sind zwei Fehlentscheidungen möglich.
(a) Die Hypothese ist richtig und wir lehnen sie ab (α-Fehler)
(b) Die Hypothese ist falsch und wir lehnen sie nicht ab (β-Fehler)
Man ist bemüht, beide Fehler so klein wie möglich zu halten. Den α-Fehler können wir berechnen, da wir mit der Wahrscheinlichkeit der Nullhypothese rechnen können. Und diese Wahrscheinlichkeit ist ja dann das wahre p, da die Hypothese richtig ist.
Den β-Fehler können wir nicht ohne weiteres berechnen, da wir die wahre Wahrscheinlichkeit nicht kennen. Die Hypothese ist ja falsch. Wir kommen auf dieses Problem zurück.

8.5 Zweiseitiger Test

Wir haben im Beispiel einen sogenannten einseitigen Test durchgeführt. Der Ablehnungsbereich lag auf nur einer Seite des zu erwartenden Wertes. Das war beim Würfel naheliegend, denn wir wollten Sechsen werfen und lehnen einen Würfel, der viele Sechsen bringt, nicht ab. Streng genommen wäre aber ein solcher Würfel auch nicht gut. Wenn wir einen Würfel auf Gleichwahrscheinlichkeit der Seiten testen wollen, sollten wir zweiseitig testen.



8.6 Grundbegriffe des Testens

Nullhypothese:	Annahme über die Grundwahrscheinlichkeit
Ablehnungsbereich	Ereignis, das zur Ablehung der Hypothese führt
kritisches Gebiet	= Ablehnungsbereich
Verwerfungsbereich	= Ablehnungsbereich
α-Fehler	Wahrscheinlichkeit, eine wahre Hypothese abzulehnen
Signifikanzniveau	Wahrscheinlichkeit, eine wahre Hypothese abzulehnen, = α-Fehler
Sicherheit	Wahrscheinlichkeit, eine wahre Hypothese zu akzeptieren, $S = 1 - \alpha$
β-Fehler	Wahrscheinlichkeit, eine falsche Hypothese zu akzeptieren

9 Wirkungskurve des Hypothesentests

Wir hatten beim Hypothesentest zwei Fehlerarten bemerkt, und wir haben den β-Fehler nicht genau untersuchen können. Das wollen wir jetzt versuchen. Erinnern wir uns an die Schwierigkeit: es gibt zu viele mögliche Wahrscheinlichkeiten p, wenn die Hypothese nicht stimmt.
Untersuchen wir ein einfaches Beispiel.
Wir testen zweiseitig eine Münze mit $H_0: \quad p_0 = \frac{1}{2}$ und $n = 12$. Die Entscheidungsregel lautet: Verwirf die Hypothese, wenn $X \in \{0; 1; 11; 12\}$.

p in Wahrheit	H_0	$P(H_0\ verwerfen)$	$P(H_0\ akzeptieren)$	β-Fehler
0.0	f	1.000	0.000	0.000
0.1	f	0.659	0.341	0.341
0.2	f	0.275	0.725	0.725
0.3	f	0.085	0.915	0.915
0.4	f	0.019	0.981	0.981
0.5	richtig	0.006	0.994	keiner
0.6	f	0.019	0.981	0.981
0.7	f	0.085	0.915	0.915
0.8	f	0.275	0.725	0.725
0.9	f	0.659	0.341	0.341
1.0	f	1.000	0.000	0.000

$$\alpha = 0.006 = 0.6\%$$

Damit die Tabelle besser verstanden wird, sind noch zwei Anmerkungen nötig.

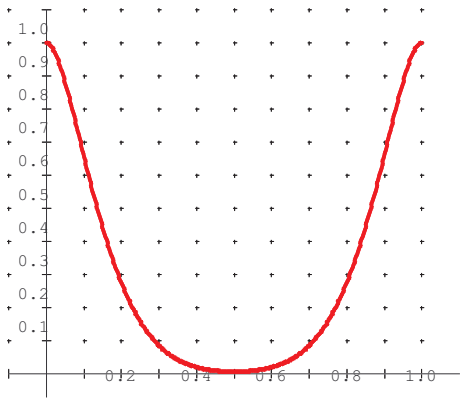
Wir kennen das wahre p nicht, aber irgendeines wird richtig sein. Wenn $p = 0.5$, dann ist die Hypothese richtig, in allen anderen Fällen falsch.

$P(H_0\ verwerfen)$ ist gleich $P(X \in \{0; 1; 11; 12\})$, $P(H_0\ akzeptieren)$

ist die Gegenwahrscheinlichkeit dazu.

Die Fehler sind Wahrscheinlichkeiten für zusammengesetzte Ereignisse: α-Fehler ist die Wahrscheinlichkeit, dass die Hypothese richtig ist und abgelehnt wird, β-Fehler ist die Wahrscheinlichkeit, dass die Hypothese falsch ist und akzeptiert wird. Wo die Hypothese richtig ist, kann also kein β-Fehler existieren.

Jetzt tragen wir die Werte der Ablehnungswahrscheinlichkeit im Koordinatensystem auf, und erhalten etwa folgenden Graphen.

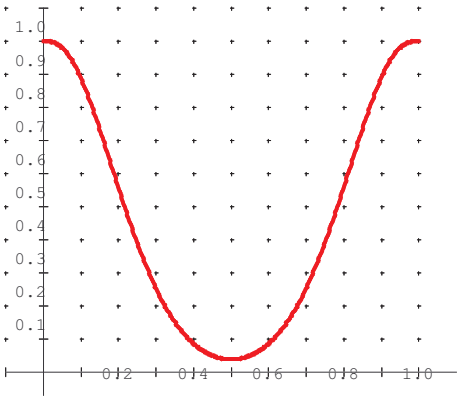


Diese Kurve wird im allgemeinen Wirkungskurve des Tests genannt. Die Kurve der Annahmewahrscheinlichkeiten wird Operations-Charakteristik genannt. Die β-Fehler erkennen wir als Differenzen $1 - P(H_0\ verwerfen)$. Die Fehler hängen vom Ablehnungsbereich ab.

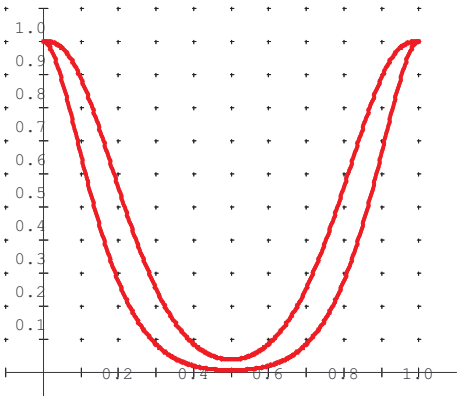
Ob dies nun ein guter oder nicht so guter Test ist, erkennt man, wenn man ihn mit einem zweiten Test vergleicht. $H_0: \quad p_0 = \frac{1}{2}$ und $n = 12$. Die Entscheidungsregel lautet: Verwirf die Hypothese, wenn $X \in \{0; 1; 2; 10; 11; 12\}$.

p in Wahrheit	H_0	$P(H_0\ verwerfen)$	β-Fehler
0.0	f	1.000	0.000
0.1	f	0.889	0.111
0.2	f	0.559	0.441
0.3	f	0.253	0.747
0.4	f	0.085	0.915
0.5	richtig	0.038	keiner
0.6	f	0.085	0.915
0.7	f	0.253	0.747
0.8	f	0.559	0.441
0.9	f	0.889	0.111
1.0	f	1.000	0.000

$$\alpha = 0.038 = 3.8\%$$

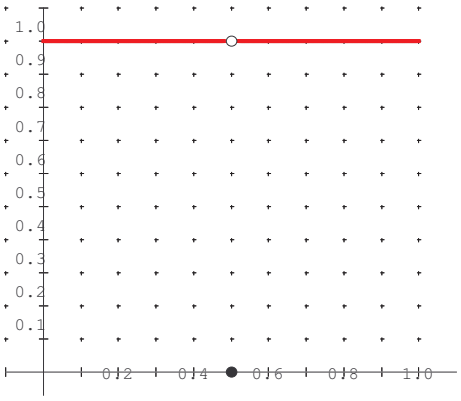


Wir zeichnen beide Kurven in ein gemeinsames System.



Bei Test1 wird der β-Fehler nicht so schnell klein wie bei Test2; dafür hat Test1 einen kleineren α-Fehler.

Wie sähe denn nun ein idealer Test aus? Er dürfte gar keinen Fehler zulassen; dann müsste also $\alpha = 0$ und $\beta = 0$ sein.
Daran gemessen ist offenbar ein Test besser, wenn der β-Fehler schnell klein



wird; gut ist der Test, wenn dann auch noch der α-Fehler klein ist.

Wir haben einen Graphen gezeichnet, dann müsste doch dazu ein Funktionsterm existieren. Viele verschiedene p bereiten dann kein Problem, wenn wir p als Variable auffassen und die davon abhängigen Wahrscheinlichkeiten als Funktion. Tatsächlich finden wir als Funktion der Ablehnungswahrscheinlichkeit ein Polynom:

$$M_{12;\{0;1;11;12\}}(p) = (1-p)^{12} + 12p(1-p)^{11} + 12p^{11}(1-p) + p^{12}$$

ist das Polynom zu $K = \{0; 1; 11; 12\}$,

$$M_{12;\{0;1;2;10;11;12\}}(p) = (1-p)^{12}+12p(1-p)^{11}+66p^2(1-p)^{10}+66p^{10}(1-p)^2+12p^{11}(1-p)+p^{12}$$

ist das Polynom zu $K = \{0; 1; 2; 10; 11; 12\}$.
Allgemein kann man bei Stichprobenumfang n und Verwerfungsbereich K schreiben

$$M_{n;K}(p) = \sum_{k \in K} \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

Das sieht abschreckend aus; ist es auch, wenn n groß und K umfangreich ist. Aber dafür gibt es Computer.

10 Zum Gebrauch der Tabellen

10.1 Wahrscheinlichkeitsfunktion

Die Tabellen der Binomialverteilung und der summierten Binomialverteilung sind nicht ganz leicht zu handhaben. Die Benutzung soll hier erklärt werden. Für die Wahrscheinlichkeitsfunktion $B_{n;p}(k)$ hatten wir herausgefunden, dass

$$B_{n;p}(k) = B_{n;1-p}(n-k)$$

Das bedeutet, dass die Wahrscheinlichkeit für k Erfolge bei Grundwahrscheinlichkeit p gleich ist der Wahrscheinlichkeit für $n-k$ Miss-Erfolge, die natürlich die Grundwahrscheinlichkeit $1-p$ haben. Wir können daher die Tabellen auf p-Werte von 0 bis 0.5 beschränken, da sie automatisch die anderen Wahrscheinlichkeiten von 0.5 bis 1 abdecken. Man muss nur die Tabelle rückwärts lesen. Daher ergibt sich folgender Aufbau:

$n=4$	k	$p=0.30$
	0	0.2401
	1	0.4116
	2	0.2646
	3	0.0756
	4	0.0081

$n=4$	k	$p=0.70$
	0	0.0081
	1	0.0756
	2	0.2646
	3	0.4116
	4	0.2401

oder

$n=4$	k	$p=0.70$
	4	0.2401
	3	0.4116
	2	0.2646
	1	0.0756
	0	0.0081

Man vereinigt die beiden Tabellen zu

$n=4$	k	$p=0.30$	
	0	0.2401	4
	1	0.4116	3
	2	0.2646	2
	3	0.0756	1
	4	0.0081	0
		$p=0.70$	k

10.2 Summenfunktion

Jetzt kommen wir zur Summenfunktion, wobei wir beachten, dass

$$F_{n;p}(k) = F_{n;p}(X \leq k)$$

$n=4$	k	$p=0.30$
	0	0.2401
	1	0.6517
	2	0.9163
	3	0.9919
	4	1.0000

Den Wert 1.0000 für $k=4$ lassen wir später weg, er kostet nur Platz.

Es gilt

$$\begin{aligned} F_{n;1-p}(k) &= B_{n;1-p}(0) + \cdots + B_{n;1-p}(k) \\ &= B_{n;p}(n-0) + \cdots + B_{n;p}(n-k) \\ &= 1 - (B_{n;p}(0) + \cdots + B_{n;p}(n-k-1)) \end{aligned}$$

denn die Ereignisse $\{0; 1; 2; \cdots; n-k-1\}$ und $\{n-k; n-k+1; \cdots; n-1; n\}$ sind Gegenereignisse.
Und wir erhalten

$$F_{n;1-p}(k) = 1 - F_{n;p}(n-k-1)$$

Sehen wir uns jetzt an, welche Bedeutung das für die Tabelle hat.

		$F_{n;p}(k)$			$F_{n;1-p}(k)$
$n = 4$	k	$p = 0.30$	$n = 4$	k	$p = 0.70$
	0	0.2401		0	0.0081
	1	0.6517		1	0.0837
	2	0.9163		2	0.3483
	3	0.9919		3	0.7599
	4	1.0000		4	1.0000

und umgekehrt sortiert

		$F_{n;p}(k)$			$F_{n;1-p}(k)$
$n = 4$	k	$p = 0.30$	$n = 4$	k	$p = 0.70$
	0	0.2401		4	1.0000
	1	0.6517		3	0.7599
	2	0.9163		2	0.3483
	3	0.9919		1	0.0837
	4	1.0000		0	0.0081

und jetzt noch geschickt nebeneinander geschrieben, dann sehen wir, wie die Tabelle zu lesen ist

		$F_{n;p}(k)$			$F_{n;1-p}(k)$
$n = 4$	k	$p = 0.30$	$n = 4$	k	$p = 0.70$
				4	1.0000
	0	0.2401		3	0.7599
	1	0.6517		2	0.3483
	2	0.9163		1	0.0837
	3	0.9919		0	0.0081
	4	1.0000			

Will ich zum Beispiel $F_{4;0.70}(3)$ bestimmen, lese ich also bei $F_{4;0.30}(4-3-1) = F_{4;0.30}(0)$ ab und berechne die Differenz zu 1.

10.3 Bestimmung des kritischen Gebietes zu gegebenem α-Fehler

Im allgemeinen wird α vorgegeben und dazu werden die Grenzen des Kritischen Gebietes, des Ablehnungsbereiches bestimmt. Betrachten wir einen zweiseitigen Test; der einseitige wird genauso behandelt, nur dass eben eine einzige Grenze bestimmt wird. Sei $\alpha = 0.10$, dann teilt man in zwei Bereiche gleicher Wahrscheinlichkeit. Es ist

$$K = \{0; \cdots ; g_l\} \cup \{g_r; \cdots ; n\}$$

mit $P(\{0; \cdots ; g_l\}) \leq \alpha/2$ und $P(\{g_r; \cdots ; n\}) \leq \alpha/2$. Wir suchen in der Tabelle die Grenzen g_l und g_r.

Beispiel: $n = 20$ $p = 0.30$ $\alpha = 0.10$
Suche den Wert, der gerade eben kleiner als 0.05 ist:
das ist 0.0355; dazu gehört $k = 2$; das ist unser g_l.
Für die Grenze g_r müssen wir mit Gegenereignissen rechnen, da die Summenfunktion von 0 bis k summiert.

$$P(\{g_r; \cdots ; n\}) = 1 - P(\{0; \cdots ; g_r - 1\}) \leq \alpha/2 \Rightarrow P(\{0; \cdots ; g_r - 1\}) \geq 1 - \alpha/2$$

Das heisst für unser Beispiel:
Suche den Wert, der gerade eben größer als 0.95 ist:
das ist 0.9520; dazu gehört $k = 9$; das ist unser $g_r - 1$; also ist $g_r = 10$.
Wir finden $K = \{0 \cdots 2; 10 \cdots 20\}$.

Beispiel: $n = 20$ $p = 0.70$ $\alpha = 0.10$
Wir lesen die Tabelle mit Eingang von unten und rechnen mit $1 - P$.
Suche den Wert, der gerade eben kleiner als 0.05 ist:
das ist $1 - 0.9520 = 0.0480$; dazu gehört $k = 10$; das ist unser g_l.
Für die Grenze g_r müssen wir mit Gegenereignissen rechnen, da die Summenfunktion von 0 bis k summiert.
Das heisst für unser Beispiel:
Suche den Wert, der gerade eben größer als 0.95 ist: das ist $1 - 0.0355 = 0.9645$; dazu gehört $k = 17$; das ist unser $g_r - 1$; also ist $g_r = 18$.
Wir finden $K = \{0 \cdots 10; 18 \cdots 20\}$.

11 Erwartungswert und Varianz

Jetzt sollten wir noch einige weitere wichtige Grundbegriffe der Stochastik kennenlernen. **Erwartungswert** und **Varianz** können wir übertragen aus der eindimensionalen Statistik. Der Begriff **Unabhängigkeit** tauchte bei der Untersuchung zweidimensionaler Stichproben auf. Er fiel bereits bei der Bernoullikette, wurde aber noch nicht so richtig erklärt. Wir benötigen den Begriff und seine rechnerische Beschreibung bei einem der folgenden Beweise. Wir werden Unabhängigkeit aber erst später genauer untersuchen.
Zunächst betrachten wir eine einzelne Zufallsgröße und erklären, was wir unter Erwartungswert und Varianz verstehen.

11.1 beliebige Verteilung

Bei Häufigkeitsverteilungen haben wir das arithmetische Mittel als Kennzahl berechnet. Diesem entspricht bei Wahrscheinlichkeitsverteilungen der Erwartungswert. Betrachten wir als Beispiel ein Würfelspiel mit Geldeinsatz, bei dem einige Ereignisse zur Auszahlung eines Betrages, andere zum Verlust des Einsatzes führen. Man zahlt 6 Euro ein, bekommt bei 1 bis 5 jeweils 1 bis 5 Euro und bei 6 den dreifachen Einsatz. Für wen lohnt sich das Spiel?

Augenzahl	1	2	3	4	5	6
ausgezahlt	1	2	3	4	5	18
Gewinn	-5	-4	-3	-2	-1	12
W´keit	1/6	1/6	1/6	1/6	1/6	1/6
zu erwarten	-5/6	-4/6	-3/6	-2/6	-1/6	12/6

Insgesamt kann man erwarten,

$$E = -5 \cdot \frac{1}{6} - 4 \cdot \frac{1}{6} - 3 \cdot \frac{1}{6} - 2 \cdot \frac{1}{6} - 1 \cdot \frac{1}{6} + 12 \cdot \frac{1}{6} = -15 \cdot \frac{1}{6} + 12 \cdot \frac{1}{6}$$
$$= -\frac{1}{2}\ (Euro)$$

ausgezahlt zu bekommen. Man macht also auf Dauer Verlust. Das Spiel lohnt sich nur für die „Bank“.

Definition 11.1 *Ist X eine Zufallsvariable mit den Werten $x_1, \ldots, x_n$ und den Wahrscheinlichkeiten $P(X = x_i) = p_i$, so heißt der Term*

$$E(X) = x_1 \cdot p_1 + \cdots + x_n \cdot p_n$$

Erwartungswert von X.

Statt $E(X)$ schreiben wir oft μ_x oder auch einfach μ, wenn der Bezug klar ist. Wir können $E(X)$ als Summe schreiben mit

$$E(X) = \sum_{i=1}^{n} x_i \cdot p_i$$

In der deskriptiven Statistik haben wir als Steuungsmaß die mittlere Summe der quadrierten Abweichungen benutzt.

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

oder bei klassierter Stichprobe mit k Klassen und den Besetzungszahlen n_i

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot n_i = \sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot \frac{n_i}{n}$$

Die Terme $\frac{n_i}{n}$ sind relative Häufigkeiten, sie entsprechen den Wahrscheinlichkeiten. Wir kommen also zu der Definition der Varianz:

Definition 11.2 *Ist X eine Zufallsvariable mit den Werten $x_1, \ldots, x_n$ und dem Erwartungswert μ, so heißt der Term*

$$V(X) = (x_1 - \mu)^2 \cdot p_1 + \cdots + (x_n - \mu)^2 \cdot p_n$$

Varianz von X.

Statt $V(X)$ schreiben wir oft σ_x^2 oder auch einfach σ^2, wenn der Bezug klar ist. σ nennen wir auch **Standardabweichung**. Wir können $V(X)$ als Summe schreiben mit

$$V(X) = \sum_{i=1}^{n}(x_i - \mu)^2 \cdot p_i$$

Satz 11.1 *Die Varianz ist der Erwartungswert der quadrierten Abweichungen.*

$$V(X) = E((X - \mu)^2)$$

11.2 lineare Transformation

Satz 11.2

$$E(aX + b) = aE(X) + b$$

Beweis*:*

$$E(X) = \sum_{i=1}^{n} x_i \cdot p_i$$

$$E(aX+b) = \sum_{i=1}^{n}(ax_i+b)\cdot p_i = \sum_{i=1}^{n} ax_i \cdot p_i + \sum_{i=1}^{n} b \cdot p_i = a\sum_{i=1}^{n} x_i \cdot p_i + b\sum_{i=1}^{n} \cdot p_i$$
$$= aE(X) + b$$

Satz 11.3

$$V(aX+b) = a^2 V(X)$$

Beweis:

$$E(X) = \mu \Longrightarrow E(aX+b) = a\mu + b$$

$$V(X) = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot p_i = V(aX+b) = \sum_{i=1}^{n} ((ax_i + b) - (a\mu + b))^2 \cdot p_i$$

$$= \sum_{i=1}^{n} ((ax_i - a\mu) + b - b)^2 \cdot p_i$$

$$= \sum_{i=1}^{n} a^2 (x_i - \mu)^2 \cdot p_i = a^2 \sum_{i=1}^{n} (x_i - \mu)^2 \cdot p_i$$

$$= a^2 V(X)$$

11.3 Summe von Zufallsvariablen

Sobald wir zwei Zufallsvariable zusammenfassen zum Beispiel durch Addition oder Multiplikation, müssen wir überlegen, ob die Zufallvariablen unabhängig voneinander sind.

Definition 11.3 *Zwei Zufallsvariable X und Y heißen unabhängig, wenn gilt*

$$P(X \cap Y) = P(X) \cdot P(Y)$$

Definition 11.4 *Die Summe $X + Y$ der Zufallsvariablen X und Y ordnet den Paaren $(x_i; y_j)$ die Summen $x_i + y_j$ zu.*

Definition 11.5 *Die Wahrscheinlichkeitsverteilung der Summe $Z = X + Y$ der Zufallsvariablen X und Y ordnet den Summen $z_k = x_i + y_j$ die Wahrscheinlichkeiten $P(x_i; y_j)$ zu.*

Beispiel:

X		
x_i	2	3
p_i	0,3	0,7

$E(X) = 2 \cdot 0,3 + 3 \cdot 0,7 = 2,7$

Y			
y_j	0	1	2
p_j	0,1	0,5	0,4

$E(Y) = 0 \cdot 0,1 + 1 \cdot 0,5 + 2 \cdot 0,4 = 1,3$

$$E(X) + E(Y) = 2,7 + 1,3 = 4$$

Die beiden Wahrscheinlichkeitsverteilungen kombinieren wir. Wenn die Variablen voneinander unabhängig sind, können wir die Wahrscheinlichkeiten multiplizieren.

$z_k = x_i + y_j$	0	1	2 (y_j)	Summe
2	0,03	0,15	0,12	0,30
(x_i) 3	0,07	0,35	0,28	0,70
Summe	0,10	0,50	0,40	1,00

Jetzt bilden wir für die Zufallsvariable Z die Summen $x_i + y_j$ für alle i, j.

$z_k = x_i + y_j$	0	1	2 (y_j)
2	2	3	4
(x_i) 3	3	4	5

Und nun können wir die Wahrscheinlichkeitsverteilung von Z aufschreiben.

$z_k = x_i + y_j$	2	3	4	5
	0,03	0,22	0,47	0,28

$E(Z) = 2 \cdot 0,03 + 3 \cdot 0,22 + 4 \cdot 0,47 + 5 \cdot 0,28 = 4 = E(X) + E(Y)$
Dass das kein Zufall ist, beweisen wir jetzt.

Satz 11.4 *E(X+Y)=E(X)+E(Y)*
Beweis*:*
Beim Beweis benutzen wir die Tatsache

$$\sum_{alle\ j} P(x_i; y_j) = P(x_i)$$

Wir bilden alle Kombinationen von Summen; dabei laufe i von 1 bis n, und j laufe von 1 bis m.

$$\begin{aligned}
E(X+Y) &= \sum_{i,j} (x_i + y_j) P(x_i; y_j) \\
&= (x_1 + y_1) P(x_1; y_1) + \cdots + (x_1 + y_m) P(x_1; y_m) \\
&\cdots \\
&+ (x_n + y_1) P(x_n; y_1) + \cdots + (x_n + y_m) P(x_n; y_m) \\
&= x_1 \cdot P(x_1; y_1) + \cdots + x_1 \cdot P(x_1; y_m) \\
&\cdots \\
&+ x_n \cdot P(x_n; y_1) + \cdots + x_n \cdot P(x_n; y_m) \\
&+ y_1 \cdot P(x_1; y_1) + \cdots + y_1 \cdot P(x_n; y_1) \\
&\cdots \\
&+ y_m \cdot P(x_1; y_m) + \cdots + y_m \cdot P(x_n; y_m) \\
&= \sum x_i \cdot P(x_i) + \sum y_j \cdot P(y_j) \\
&= E(X) + E(Y)
\end{aligned}$$

Der entsprechende Satz für Varianzen ist nicht ganz so einfach. Wenn wir den Term der Varianz umformen, erscheint im gemischten Glied das Produkt XY, welches wir also auch noch untersuchen müssen. Aber alles zu seiner Zeit.
Wir hatten bereits gesehen, dass

$$V(X) = E((X-\mu)^2)$$

Diesen Term formen wir um, wobei wir beachten, dass man Konstante ausklammert und dass der Erwartungswert einer Konstanten natürlich diese Konstante ist. Wir erhalten

$$V(X) = E(X^2 - 2X\mu + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu\mu + \mu^2 = E(X^2) - \mu^2$$

Angewandt auf $V(X+Y)$ erhalten wir nach den bekannten Sätzen über den Erwartungswert

$$\begin{aligned} V(X+Y) &= E((X+Y)^2) - (\mu_x + \mu_y)^2 = E(X^2 + 2XY + Y^2) - (\mu_x^2 + 2\mu_x\mu_y + \mu_y^2) \\ &= E(X^2) + 2E(XY) + E(Y^2) - \mu_x^2 - 2\mu_x\mu_y - \mu_y^2 \\ &= \underbrace{E(X^2) - \mu_x^2} + \underbrace{E(Y^2) - \mu_y^2} + 2E(XY) - 2\mu_x\mu_y \\ &= V(X) + V(Y) + 2(E(XY) - \mu_x\mu_y) \end{aligned}$$

Wenn $E(XY) - \mu_x\mu_y = 0$ ist, können wir die Varianz zerlegen in eine Summe von Varianzen; dies ist aber nur der Fall, wenn X und Y unabhängig sind.
Der Term $E(XY) - \mu_x\mu_y$ ist uns übrigens begegnet bei der Regressionsanalyse in der Darstellung $\frac{1}{n}S_{xy} = \frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}$. Wir haben ihn **Kovarianz** genannt und zur Bestimmung statistischer Abhängigkeit benutzt.

11.4 Produkt von Zufallsvariablen

Satz 11.5 *Wenn die Zufallsvariablen X und Y unabhängig sind, gilt*

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

Beweis*:*

$$E(X \cdot Y) = \sum_{i,j} x_i y_j \cdot P(x_i; y_j)$$

$$= x_1 y_1 \cdot P(x_1; y_1) + \cdots + x_1 y_m \cdot P(x_1; y_m)$$

$$\cdots$$

$$+ x_n y_1 \cdot P(x_n; y_1) + \cdots + x_n y_m \cdot P(x_n; y_m)$$

$$= x_1 \cdot P(x_1) \cdot [y_1 \cdot P(y_1) + \cdots + y_m \cdot P(y_m)]$$

$$\dots$$
$$+x_n \cdot P(x_n) \cdot [y_1 \cdot P(y_1) + \dots + y_m \cdot P(y_m)]$$

$$= \sum x_i \cdot P(x_i) \cdot \sum y_j \cdot P(y_j)$$
$$= E(X) \cdot E(Y)$$

Die Varianz eines Produktes lassen wir weg. Wir können aber zeigen, dass für unabhängige Variable gilt

Satz 11.6
$$V(XY) = E(X^2Y^2) - E^2(X)E^2(Y)$$

11.5 Binomialverteilung

Jetzt können wir Erwartungswert und Varianz eines Bernoulli-Versuches und dann einer Bernoulli-Kette bestimmen. Die Verteilung des Bernoulli-Versuches ist

X	1	0
$P(X)$	p	$1-p$

$$E(X) = 1 \cdot p + 0 \cdot (1-p) = p$$

$(X - E(X))^2$	$(1-p)^2$	$(0-p)^2$
$P(X)$	p	$1-p$

$$V(X) = (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) = p - 2p^2 + p^3 + p^2 - p^3$$
$$= p - p^2 = p \cdot (1-p)$$

Um den Erwartungswert des n-stufigen Versuches zu bestimmen, brauchen wir den Satz über den Erwartungswert der Summe. Die Zufallsvariable der Bernoullikette mit Länge n ist eine Summe aus n Zufallsvariablen des einzelnen Bernoulliversuches. Es ist leicht zu zeigen, dass

$$E(\sum X_i) = n \cdot E(X) = np$$

Wir finden dann

Satz 11.7 *Ist X die Zufallsvariable einer Bernoullikette, also $B_{n;p}$-verteilt, so gilt mit $q = 1 - p$*

$$E(X) = np, \qquad V(X) = npq, \qquad \sigma_x = \sqrt{npq}$$

12 Ungleichung von Tschebyscheff

Bei der Analyse eindimensionaler Stichproben hatten wir Streuungsintervalle untersucht, also gefragt, welche Anteile der Stichprobe in den Intervallen

$$[\mu - \sigma; \mu + \sigma],\ [\mu - 2\sigma; \mu + 2\sigma],\ [\mu - 3\sigma; \mu + 3\sigma]$$

liegen. Damals hatten wir gesagt, dass bei normaler Verteilung im 1σ-Intervall etwa 68% der Werte liegen, im 2σ-Intervall etwa 95%, und im 3σ-Intervall etwa 99%. Diese Frage wollen wir auf Wahrscheinlichkeitsverteilungen übertragen. Die Ungleichung von Tschebyscheff ermöglicht eine Abschätzung dieser Anteile bei Wahrscheinlichkeitsverteilungen.

Satz 12.1 (Ungleichung von Tschebyscheff) *Für die Wahrscheinlichkeit, dass der Wert der Zufallsvariable X um mindestens c vom Erwartungswert μ abweicht, gilt mit $V(X) = \sigma^2$*

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

Beweis*:*
Wir teilen die Werte, die X annehmen kann, in zwei getrennte Bereiche.

$$A = \{x \| x - \mu| < c\} \quad \textit{und} \quad B = \{x \| x - \mu| \geq c\}$$

Jetzt bestimmen wir die Varianz.

$$V(X) = \sum_{alle\ i} (x_i - \mu)^2 P(x_i)$$

und trennen nach den beiden Bereichen

$$= \underbrace{\sum_{x_i \in A} (x_i - \mu)^2 P(x_i)}_{weglassen,\ uninteressant} + \underbrace{\sum_{x_i \in B} (x_i - \mu)^2 P(x_i)}_{interessiert\ uns}$$

$$\geq \sum_{x_i \in B} (x_i - \mu)^2 \cdot P(x_i)$$

Da diese $x_i \in B$, kann man $(x - \mu)^2$ abschätzen mit $|x - \mu|^2 \geq c^2$.

$$\geq \sum_{x_i \in B} c^2 \cdot P(x_i)$$

$$= c^2 \cdot \sum_{x_i \in B} P(x_i)$$

Nun sind nur noch die $x_i \in B$ übrig; für die gilt aber gerade, dass sie um mindestens c von μ abweichen. Und $P(x_i)$ ist die Wahrscheinlichkeit für dieses Ereignis.

$$= c^2 \cdot P(|X - \mu| \geq c)$$

$$\Longrightarrow \sigma^2 \geq c^2 \cdot P(x_i \in B)$$

$$\Longrightarrow P(x_i \in B) \leq \frac{\sigma^2}{c^2}$$

$$\Longrightarrow P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

Diese Abschätzung ist sehr grob, da sie für jede beliebige Verteilung gilt. Zum Beispiel kann man sehen, dass die Wahrscheinlichkeit, einen Wert außerhalb des 2σ-Intervalls zu finden, bei höchstens 25% liegt, außerhalb des 3σ-Intervalls bei höchstens 11%. Wenn man mehr über die spezielle Verteilung weiß, kann man die Wahrscheinlichkeiten genauer einschätzen.
Kommen wir auf den Würfel zurück.
Mit $p = \frac{1}{6}$ und $n = 30$ gilt $E(X) = 30 \cdot \frac{1}{6} = 5$, was wir auch erwartet hatten.
Weiter gilt $V(X) = 30 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{25}{6} \approx 4,2$ also $\sigma \approx 2,04$.
Wir erhalten nach Tschebyscheff die oben errechneten Werte. Mit der Binomialverteilung jedoch

$$P(|X - \mu| \geq 2\sigma) = P(|X - 5| \geq 4,08) = P(0; 10; \ldots; 30) \approx 2,4\%$$

und

$$P(|X - \mu| \geq 3\sigma) = P(|X - 5| \geq 6,12) = P(12; \ldots; 30) \approx 0,2\%$$

Bei einer symmetrischen Verteilung wie der des Münzwurfes erhalten wir mit $p = 0,5$ und $n = 50 \quad E(X) = 25$ und $V(X) = 12,5$, also $\sigma \approx 3,54$.

$$P(|X - \mu| \geq 2\sigma) = P(|X - 25| \geq 7,08) = P(0; \ldots; 17; 33; \ldots; 50) \approx 3,3\%$$

und

$$P(|X - \mu| \geq 3\sigma) = P(|X - 25| \geq 10,6) = P(0; \ldots; 14; 36; \ldots; 50) \approx 0,26\%$$

Hier sei eine ergänzende Frage erlaubt: Hätte man beim Beweis der Ungleichung nicht auch den Bereich B weglassen können um so eine Abschätzung für
$P(|X - \mu| < c)$ zu erhalten? Wenn man die Richtung der Abschätzungen beachtet, sieht man, dass das nicht ginge.

13 Gesetz der großen Zahl

Jetzt endlich sind wir so weit fortgeschritten, dass wir uns auch theoretisch mit dem Gesetz der Stabilisierung relativer Häufigkeiten befassen können. Wir haben gesagt, wenn man nur hinreichend oft einen Würfel wirft, dann wird es immer wahrscheinlicher, dass h in der Nähe von $\frac{1}{6}$ herauskommt. Das wollen wir rechnerisch erfassen beziehungsweise formal beschreiben. Zunächst bestimmen wir Erwartungswert und Varianz der relativen Häufigkeit. Dann schätzen wir mit Tschebyscheff die Intervalle ab.
Sei also h die Zufallsvariable; man kann sie mittels Bernoullikette als Mittelwert der Summe der X_i, also der Einzelversuche, auffassen. Diese Summen hatten wir als Zufallsvariable X beschrieben.

$$h = \frac{1}{n}\sum X_i = \frac{1}{n} \cdot X$$

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

Das wundert uns nicht sehr; der Erwartungswert der relativen Häufigkeit sollte natürlich die Wahrscheinlichkeit sein.

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

Diese Zeile sagt uns, dass die Abweichung mit steigendem n geringer wird. Und das bestätigt unsere Idee der Stabilisierung.
Nun aber zu der Aussage des Gesetzes der großen Zahl.

Satz 13.1 (Gesetz der großen Zahl) *Die Wahrscheinlichkeit, dass bei n unabhängigen Versuchen die relative Häufigkeit eines Ereignisses von der Wahrscheinlichkeit dieses Ereignisses um weniger als $c > 0$ abweicht, strebt für $n \to \infty$ gegen 1.*

$$\lim_{n\to\infty} P(|h - p| < c) = 1$$

Beweis*: Beim Beweis des Satzes haben wir ein kleines Problem: die Formulierung entspricht wegen $< c$ nicht Tschebyscheff. Wir verändern die Ungleichung von Tschebyscheff.*

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

$$\Longrightarrow 1 - P(|X - \mu| \geq c) \geq 1 - \frac{\sigma^2}{c^2}$$

$$\Longrightarrow P(|X - \mu| < c) \geq 1 - \frac{\sigma^2}{c^2}$$

Die Zufallsvariable heißt jetzt speziell h, ihr Erwartungswert ist $\mu = p$, ihre Varianz $\sigma^2 = \frac{pq}{n}$; und so wird die Ungleichung zu

$$P(|h - p| < c) \geq 1 - \frac{\sigma^2}{c^2} = 1 - \frac{pq}{nc^2}$$

Hier ist zu erkennen, dass mit wachsendem n auch die Wahrscheinlichkeit zunimmt; sie strebt gegen 1.

14 Die Poisson-Verteilung

14.1 Geschichtliches

Simeon-Denis Poisson (1781-1840) wurde als empfindsames und kränkelndes Kind geboren; seine Eltern waren froh, ihn über die ersten kritischen Jahre der Kindheit mit Unterstützung einer Amme am Leben erhalten zu haben. Seine Feinmotorik war leicht gestört, so konnte er nicht gut zeichnen, aber er war hoch begabt in mathematischen Fragen. Er wurde von Laplace und Lagrange unterrichtet (sie blieben lebenslang befreundet) und er beschäftigte sich mit Differentialgleichungen, theoretischer Mechanik und Fragen der Erdrotation, der Schwingungslehre, mit Magnetismus und Elektizität. Er verfasste 300 bis 400 mathematische Aufsätze. 1837 schrieb er ein Buch über die Wahrscheinlichkeitsrechnung, in dem er den Begriff **Gesetz der großen Zahl** prägte, und in dem er die nach ihm benannte Verteilung entwickelte. Er fand zwar nicht die angemessene Resonanz in seiner näheren mathematischen Umgebung, aber Tschebyscheff entwickelte die Ideen in Russland weiter.

14.2 Näherung für die Binomialverteilung

Wenn man die Werte der Binomialverteilung für große n bestimmen will, wird die Rechnung im allgemeinen unhandlich. Wir verdanken Poisson eine Näherungsformel. Er fand heraus, dass die Verteilungen mit konstantem Erwartungswert, also für große n und passend kleine p fast identisch sind, wenn man n hinreichend groß wählt. Und so fand er eine Näherungsfunktion für die Binomialverteilung.

$$B_{n;p}(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

$$= \frac{n \cdot (n-1) \cdots (n-k+1)}{k!} \cdot p^k \cdot (1-p)^{n-k}$$

Wegen $\mu = n \cdot p$ können wir ersetzen $p = \mu/n$.

$$= \frac{n \cdot (n-1) \cdots (n-k+1)}{k!} \cdot (\frac{\mu}{n})^k \cdot (1 - \frac{\mu}{n})^{n-k}$$

Wir tauschen die beiden Nenner aus.

$$= \frac{n \cdot (n-1) \cdots (n-k+1)}{n^k} \cdot \frac{\mu^k}{k!} \cdot (1 - \frac{\mu}{n})^{n-k}$$

Wir zerlegen den Quotienten.

$$= \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} \cdot \frac{\mu^k}{k!} \cdot (1 - \frac{\mu}{n})^{n-k}$$

Jetzt lassen wir n groß werden. Dann werden die Quotienten nahezu 1 sein.

$$\approx \frac{\mu^k}{k!} \cdot (1 - \frac{\mu}{n})^{n-k}$$

Was machen wir mit dem Term $(1-\frac{\mu}{n})^{n-k}$? Wir können die Potenz nach den Exponenten zerlegen, dann wird der Teil mit Exponent k auch 1; nur der Teil mit Exponent n ist unklar: Basis und Exponent verändern sich gleichzeitig. Wenn wir uns an die Differentialrechnung erinnern und dort speziell an die Gleichung $y' = y$ und die Ableitung der Funktion $f(x) = a^x$, dann wissen wir, dass

$$\lim_{n\to\infty}(1+1/n)^n = e$$

ist. Man kann nachweisen, dass

$$\lim_{n\to\infty}(1-1/n)^n = e^{-1}$$

und allgemein

$$\lim_{n\to\infty}(1+a/n)^n = e^a$$

Wir können weiter abschätzen.

$$\cdots \approx \frac{\mu^k}{k!}\cdot(1-\frac{\mu}{n})^{n-k}$$

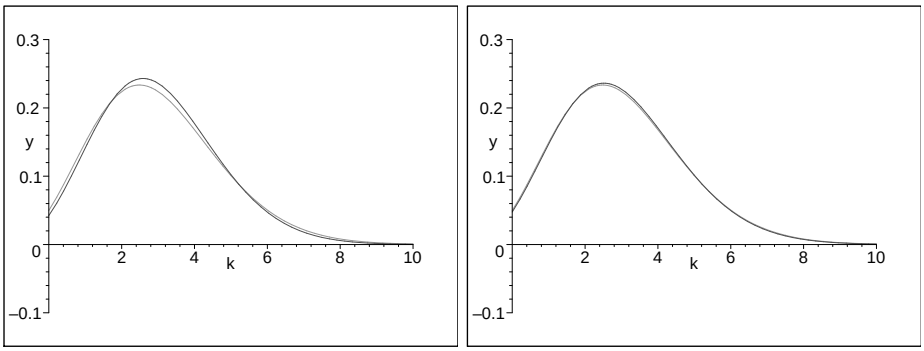
$$= \frac{\mu^k}{k!}\cdot(1-\frac{\mu}{n})^{n}\cdot\underbrace{(1-\frac{\mu}{n})^{-k}}_{\to 1}$$

$$\approx \frac{\mu^k}{k!}\cdot\underbrace{(1-\frac{\mu}{n})^{n}}_{\to e^{-\mu}}$$

$$\approx \frac{\mu^k}{k!}\cdot e^{-\mu}$$

Wir erhalten mit hinreichend großen n nach Poisson

$$B_{n;p}(k) \approx \frac{\mu^k}{k!}\cdot e^{-\mu}$$



Die Graphen von $B_{30;0.1}(k)$ und der Näherung unterscheiden sich noch; bei $B_{100;0.03}(k)$ sieht man kaum noch eine Differenz.

14.3 Die Wahrscheinlichkeitsfunktion Poisson-Verteilung

Mit Poisson haben wir eine Ersatzfunktion für die Binomialverteilung erarbeitet. Sie unterscheidet sich von der Binomialverteilung darin, dass sie nicht mehr von 3 Variablen n, p, k abhängt, sondern nur noch von den beiden Variablen μ und k. Darüberhinaus ist es eine Funktion, deren Definitionsbereich nach oben nicht beschränkt ist; wir können beliebig große k einsetzen; bei Bernoulli war k auf den Bereich 0 bis n beschränkt. Die von Poisson entwickelte Funktion ist sogar eine Wahrscheinlichkeitsfunktion, daher nennen wir sie Verteilung. Wir können zeigen, dass die Summe aller Funktionswerte 1 ergibt.

$$\sum_{k=0}^{\infty} \frac{\mu^k}{k!} \cdot e^{-\mu}$$

Alles, was nicht von k abhängt, klammern wir aus.

$$= e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!}$$

$$= e^{-\mu} \underbrace{(1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \cdots)}_{Taylor-Reihe\ von\ e^{\mu}}$$

$$= e^{-\mu} \cdot e^{\mu} = 1$$

14.4 Erwartungswert und Varianz der Poisson-Verteilung

Jetzt haben wir das Rüstzeug beisammen, mit dem wir Erwartungswert und Varianz der Poisson-Verteilung bestimmen können.

$$E(k) = \sum_{k=0}^{\infty} k \cdot \frac{\mu^k}{k!} \cdot e^{-\mu}$$

Auf $k = 0$ können wir verzichten, da der Summand NULL wird.

$$E(k) = \sum_{k=1}^{\infty} k \cdot \frac{\mu^k}{k!} \cdot e^{-\mu}$$

Wir klammern $e^{-\mu}$ aus. Der Term ähnelt der Taylor-Reihe. Das k stört; wir kürzen.

$$E(k) = e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^k}{(k-1)!}$$

Exponent und Fakultät stimmen nicht überein. Wir klammern ein μ aus; dann können wir die Taylor-Reihe ersetzen und erhalten

$$E(k) = e^{-\mu} \cdot \mu \cdot \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} = e^{-\mu} \cdot \mu \cdot e^{\mu} = \mu$$

Nun bestimmen wir die Varianz.

$$V(k) = \sum_{k=0}^{\infty}(k-\mu)^2 \cdot \frac{\mu^k}{k!} \cdot e^{-\mu}$$

Der Term scheint keine neuen Ideen zu verlangen. Wir klammern $e^{-\mu}$ aus. Dann ist da noch das Binom, das wir auflösen können. Dies führt zu drei Summanden; beim letzten klammern wir μ^2 aus; das gemischte Glied enthält $\mu \cdot k$, da klammern wir μ aus und benutzen den Erwartungswert. Aber das erste Quadrat mit k^2 macht Probleme. Einmal kann man k kürzen; den Trick kennen wir schon; da bleibt aber immer noch ein k übrig. Besser wäre es, wenn wir auch noch $k-1$ kürzen könnten.
Also muss ein weiterer Trick her.

$$k^2 = k \cdot (k-1) + k$$

$$V(k) = \sum_{k=0}^{\infty}(k^2 - 2\mu k + \mu^2) \cdot \frac{\mu^k}{k!} \cdot e^{-\mu}$$

$$= e^{-\mu}(\sum_{k=0}^{\infty} k^2 \cdot \frac{\mu^k}{k!} - \sum_{k=0}^{\infty} 2\mu k \cdot \frac{\mu^k}{k!} + \sum_{k=0}^{\infty} \mu^2 \cdot \frac{\mu^k}{k!})$$

$$= e^{-\mu}(\sum_{k=0}^{\infty} k^2 \cdot \frac{\mu^k}{k!} - 2\mu \underbrace{\sum_{k=0}^{\infty} k \cdot \frac{\mu^k}{k!}}_{\mu \cdot e^{\mu}} + \mu^2 \underbrace{\sum_{k=0}^{\infty} \cdot \frac{\mu^k}{k!}}_{e^{\mu}})$$

$$= e^{-\mu}(\sum_{k=0}^{\infty} k^2 \cdot \frac{\mu^k}{k!} - 2\mu \cdot \mu \cdot e^{\mu} + \mu^2 \cdot e^{\mu})$$

$$= e^{-\mu}(\sum_{k=0}^{\infty}(k(k-1)+k) \cdot \frac{\mu^k}{k!} - 2\mu^2 \cdot e^{\mu} + \mu^2 \cdot e^{\mu})$$

$$= e^{-\mu}(\sum_{k=0}^{\infty} \underbrace{(k(k-1)+k)}_{aufteilen} \cdot \frac{\mu^k}{k!} - \mu^2 \cdot e^{\mu})$$

und nun mit den bekannten Tricks weiter

$$= e^{-\mu}(\sum_{k=2}^{\infty} k(k-1) \cdot \frac{\mu^k}{k!} + \sum_{k=1}^{\infty} k \cdot \frac{\mu^k}{k!} - \mu^2 \cdot e^{\mu})$$

$$= e^{-\mu}(\mu^2 \sum_{k=2}^{\infty} \frac{\mu^{k-2}}{(k-2)!} + \mu \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} - \mu^2 \cdot e^{\mu})$$

$$= e^{-\mu}(\mu^2 \cdot e^{\mu} + \mu \cdot e^{\mu} - \mu^2 \cdot e^{\mu})$$

$$= \mu$$

Erwartungswert und Varianz sind gleich μ. Das muss uns nicht wundern. Erwartungswert bei Bernoulli war $\mu = np$ und die Varianz $\sigma^2 = npq$; nun ist aber bei Poisson p klein, nahe Null, und somit q groß, also nahe 1; also ist $npq \approx np$.

15 Die Normalverteilung

15.1 Geschichtliches

Die Werte der Binomialverteilung sind für große n nicht leicht zu berechnen. Schon mit Poisson hatten wir eine gute Näherung gefunden, die allerdings auf kleine p beschränkt ist. Abraham De Moivre (1667-1754) und Pierre-Simon Laplace (1749-1827) haben eine Näherung entwickelt, die später von Karl-Friedrich Gauss (1777-1855) genauer untersucht wurde. De Moivre emigrierte als französischer Protestant 1685 nach England; er wurde 1710 in die Royal Society berufen, und hatte dort die Aufgabe, den Streit zwischen Newton und Leibniz um die Vorherrschaft der Erfindung der Analysis zu entscheiden. Er war mit Newton gut befreundet. 1730 veröffentlichte er die später fälschlich nach Stirling bezeichnete Formel zur Berechnung von Fakultäten.

$$n! \approx \sqrt{2\pi n} \cdot (n/e)^n$$

Sie wurde bei der Näherung für die Binomialverteilung benötigt.
De Moivre entwickelte entscheidend die Analytische Geometrie und die Wahrscheinlichkeitsrechnung weiter. Mithilfe der komplexen Zahlen verband er Trigonometrie und Analysis. Laplace studierte in Paris; seine ersten Veröffentlichungen behandelten die Untersuchung von Extrema und Differentialgleichungen. Er war Prüfer bei der Königlichen Artillerieschule und hatte Napoleon Bonaparte als Examenskandidaten zu prüfen. 1812 veröffentlichte er seine Wahrscheinlichkeitstheorie. Er gilt als einer der hervorragendsten Naturwissenschaftler aller Zeiten: seine Untersuchungen über Kapillardruck, Mond- und Planetenbewegung, Bewegung von Flüssigkeiten, Theorie der Wärme waren richtungweisend.

15.2 Näherung für die Binomialverteilung

Lässt man bei der Binomialverteilung n immer größer werden, so beobachtet man, dass die Kurve immer mehr einer Glockenkurve ähnelt. Erwartungswert und Varianz allerdings verändern sich. Diese Abhängigkeit von den beiden Kennzahlen einer Verteilung kann man vermeiden, wenn man die Verteilung normiert, das heißt, mittels der Transformation $\frac{x-\mu}{\sigma}$ auf eine standardisierte Form bringt. Der Erwartungswert wird damit zu 0 und die Varianz zu 1. De Moivre und Laplace zeigten, dass sich die Binomialverteilung der heute so genannten Gauss-Kurve annähert. De Moivre zeigte es 1734 für den Fall $p = \frac{1}{2}$, Laplace verallgemeinerte es 1812 auf beliebige p.

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

Dass diese Annäherung mathematisch korrekt ist, belegt der **Zentrale Grenzwertsatz**, der besagt, dass die Summe einer großen Zahl von Zufallsvariablen durch die Funktion φ beschrieben werden kann. Den Beweis dafür erbringen wir hier nicht.

15.3 Ein Weg zur Funktion $\varphi(x)$

Es ist $B(k) = \frac{n!}{k!(n-k)!}p^k q^{n-k}$ die Wahrscheinlichkeit nach Bernoulli; anschaulich klar ist, dass sie beim Erwartungswert maximal ist.
Dann ist $k = np$ und $n - k = nq$

$$B_{max} = \frac{n!}{(np)!(nq)!}p^{np}q^{nq}$$

Wir formen B_{max} in Zähler und Nenner mittels der Stirling-Formel

$$n! \approx \sqrt{2\pi n} \cdot (n/e)^n$$

um zu

$$B_{max} = \frac{n^n e^{-n}\sqrt{2\pi n}}{(np)^{np} \cdot e^{-np}\sqrt{2\pi np} \cdot (nq)^{nq} \cdot e^{-nq}\sqrt{2\pi nq}} \cdot p^{np}q^{nq}$$

$$= \frac{n^n}{(np)^{np}(nq)^{nq}} \frac{e^{-n}}{e^{-np}e^{-nq}} \sqrt{\frac{2\pi n}{2\pi np \cdot 2\pi nq}} p^{np}q^{nq}$$

$$= \frac{n^n}{(np)^{np}(nq)^{nq}} \frac{e^{-n}}{e^{-np}e^{-nq}} \frac{1}{\sqrt{2\pi n \cdot pq}} p^{np}q^{nq}$$

Wegen $p + q = 1$ kann man kürzen

$$= \frac{n^n}{(np)^{np}(nq)^{nq}} \frac{1}{\sqrt{2\pi n \cdot pq}} p^{np}q^{nq}$$

und den vorderen Bruch geschickt zerlegen

$$= \frac{n^n}{n^{np}p^{np}n^{nq}q^{nq}} \frac{1}{\sqrt{2\pi n \cdot pq}} p^{np}q^{nq}$$

und wieder kürzen

$$= \frac{1}{\sqrt{2\pi \cdot npq}}$$

So weit, so gut. Wie kommen wir jetzt zu der gesuchten Exponentialfunktion? Wir zählen einmal die Variable k nicht von 0 bis n, sondern vom Maximum aus vorwärts und rückwärts. Dann wird $k = np + x$ mit passendem x, und $n - k = n - np - x = n(1 - p) - x = nq - x$. Im Grunde führen wir hier eine Transformation durch, die die Funktion um den Erwartungswert nach links schiebt. Das sieht man sofort, wenn man die neue Variable $x = k - np = k - \mu$ betrachtet.

$$B(x) = \frac{n!}{(np + x)!(nq - x)!}p^{np+x}q^{nq-x}$$

Wenn wir jetzt eine Formel für $(n + a)!$ hätten! Da ist sie:

$$(n + a)! \approx n!n^a e^{\frac{a^2+a}{2n}}$$

Diese Formel nennen wir einfach einmal StirlingII. Wir sehen sie uns später genauer an. Jetzt formen wir noch einmal kräftig um.

$$B(x) = \frac{n^n e^{-n}\sqrt{2\pi n} \cdot p^{np+x} q^{nq-x}}{(np)!(np)^x e^{\frac{x^2+x}{2np}} (nq)!(nq)^x e^{\frac{x^2-x}{2nq}}}$$

Im Nenner haben wir StirlingII benutzt, im Zähler stirlingI; das passiert dann im Nenner auch; und weil wir oben so fleißig umgeformt haben, sehen wir jetzt ganz schnell, dass der Term

$$= \frac{n^n e^{-n}\sqrt{2\pi n} \cdot p^{np+x} q^{nq-x}}{(np)!(np)^x e^{\frac{x^2+x}{2np}} (nq)!(nq)^x e^{\frac{x^2-x}{2nq}}}$$

verkürzt wird zu

$$= \frac{e^{-\frac{x^2+x}{2np} - \frac{x^2-x}{2nq}}}{\sqrt{2\pi npq}}$$

$$= \frac{1}{\sqrt{2\pi npq}} \cdot e^{-\frac{1}{2} \cdot \frac{qx^2+qx+px^2-px}{npq}}$$

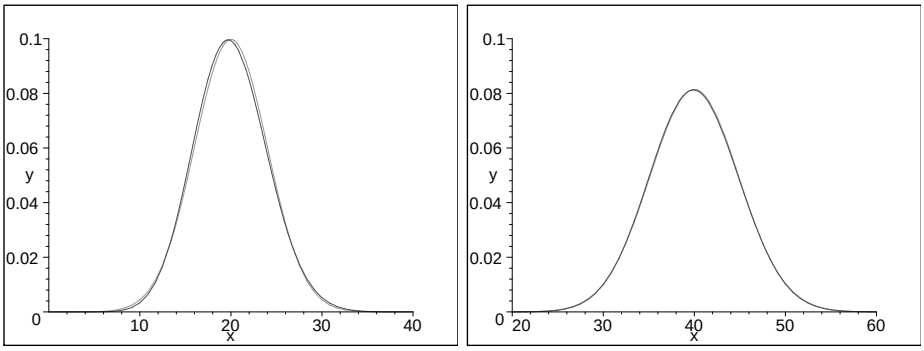
$$= \frac{1}{\sqrt{2\pi npq}} e^{-\frac{1}{2} \cdot \frac{x^2+(q-p)x}{npq}}$$

Der Zähler im Exponenten ist zu kompliziert, wir vereinfachen ihn. Für eine Abschätzung gibt es zwei Argumente: (a) für ganze x ist ein linearer Term im Vergleich zu einem quadratischen zu vernachlässigen; (b) der lineare Term enthält $q - p = 1 - 2p$; kleine p oder große p (oder kleine q) interessieren uns nicht; die hat Poisson bearbeitet; für mittlere p ist aber $1 - 2p \approx 0$.

$$\approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x^2}{2npq}}$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}$$

Um die Qualität der Annäherung zu erfassen, zeichnen wir die Graphen von $B_{100;0.2}(k)$ und $B_{100;0.4}(k)$ im Vergleich zur Exponentialfunktion. Für $B_{100;0.2}(k)$ erhalten wir $\mu = 20$ und $\sigma^2 = 16$, für $B_{100;0.4}(k)$ erhalten wir $\mu = 40$ und $\sigma^2 = 24$.



Die Graphen von $B_{100;0.2}(k)$ und der Näherung unterscheiden sich noch; bei $B_{100;0.4}(k)$ sieht man kaum noch eine Differenz.

15.4 Die Stirling-Formel und ihre Verwandte

15.4.1 StirlingI

Wir wollen zeigen, dass

$$n! \approx \sqrt{2\pi n} \cdot (n/e)^n$$

Diese Formel ist nach James Stirling (1692-1770) benannt. Möglicherweise war sie aber schon Abraham de Moivre (1667-1754) vorher bekannt.

Um die nachfolgenden Ausführungen zu verstehen, benötigen wir eine gewisse Kenntnis der Taylor-Reihe, der geometrischen Reihe, Verständnis für Abschätzungen, Monotonie und Konvergenz von Zahlenfolgen und eine Vorstellung der Idee der Intervallschachtelung, und natürlich ein bißchen Rechnen mit Logarithmen. Und bei einem zweiten Nachweis, kurz bevor wir fertig sind, benötigen wir noch einmal Integrale von Potenzen trigonometrischer Funktionen. Die Mathematik geht merkwürdige Wege.

Wir sehen uns **den ersten Weg** an.

Zum Beweis der Stirling-Formel betrachten wir die Doppelungleichung

$$n^n \cdot e^{-n}\sqrt{2\pi n} < n! < n^n \cdot e^{-n}\sqrt{2\pi n} \cdot e^{\frac{1}{12n}} \tag{1}$$

die übrigens relativ scharf ist. Für $n = 1000$ beträgt der relative mittlere Fehler etwa $0,004^o/_{oo}$.

Wir helfen uns mit einer Taylorreihe.
Am besten abzuschätzen sind Reihen mit nur positiven Gliedern. Wenn man ganz viel ausprobiert, kommt man darauf, dass mit $x = \frac{1}{2n+1}$ gilt

$$\frac{1+x}{1-x} = \frac{1+\frac{1}{2n+1}}{1-\frac{1}{2n+1}} = 1 + \frac{1}{n}$$

Nach Taylor gilt

$$\ln\frac{1+x}{1-x} = 2 \cdot (x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots)$$

Das sollte man nachrechnen.

Mit $x = \frac{1}{2n+1}$ haben wir dann nur noch positive Summanden; wenn davon einige entfallen, wird der Term kleiner.

$$\ln(1+\frac{1}{n}) = 2 \cdot \left(\frac{1}{2n+1} + \frac{1}{3(2n+1)^3} + \frac{1}{5(2n+1)^5} \cdots\right) > 2 \cdot \left(\frac{1}{2n+1}\right)$$

Daher ist

$$1 < \frac{1}{2}(2n+1) \cdot \ln(1+\frac{1}{n})$$

Auf $\ln(1+\frac{1}{n})$ wenden wir nocheinmal die Taylor-Reihe an, multiplizieren aus, verkleinern die Nenner, sehen die geometrische Reihe:

$$= \frac{1}{2}(2n+1)\cdot 2\cdot\left(\frac{1}{2n+1}+\frac{1}{3(2n+1)^3}+\frac{1}{5(2n+1)^5}\cdots\right)$$

$$= 1+\frac{1}{3(2n+1)^2}+\frac{1}{5(2n+1)^4}+\cdots$$

$$< 1+\frac{1}{3}\underbrace{\left(\frac{1}{(2n+1)^2}+\frac{1}{(2n+1)^4}+\cdots\right)}_{\textit{geom. Reihe}} = 1+\frac{1}{3}\cdot\frac{1}{(2n+1)^2-1}$$

$$= 1+\frac{1}{3}\cdot\frac{1}{4n^2+4n} = 1+\frac{1}{12n(n+1)}$$

also

$$1 < \frac{1}{2}(2n+1)\ln(1+\frac{1}{n}) < 1+\frac{1}{12n(n+1)}$$

und delogarithmiert

$$e < (1+\frac{1}{n})^{n+\frac{1}{2}} < e\cdot e^{\frac{1}{12n(n+1)}} \tag{2}$$

Die linke Seite stellt übrigens ein interessantes Ergebnis dar. Wenn man die Konvergenz der Folge $(1+1/n)^n$ gegen e nachweist, stößt man auf $(1+1/n)^n < e$.

Wir betrachten jetzt noch einmal die anfängliche Doppelungleichung. Wir weisen nach, dass die Folge

$$a_n = \frac{n!}{n^n\cdot e^{-n}\sqrt{2\pi n}}$$

monoton fällt und gegen 1 konvergiert. $a_n > 1$ entspricht dem linken Teil der Ungleichung. Der wäre dann erledigt. Ebenso weisen wir nach, dass die Folge

$$b_n = \frac{n!}{n^n\cdot e^{-n}\sqrt{2\pi n}\cdot e^{\frac{1}{12n}}}$$

monoton steigt und gegen 1 konvergiert. Dann wäre der rechte Teil der Ungleichung erledigt.
Wir weisen also jetzt folgendes nach: a_n steigt monoton, b_n fällt monoton, $a_n > b_n$, dann wissen wir, dass a_n und b_n einen Grenzwert besitzen; $a_n - b_n$ konvergiert gegen null, dann hätten wir die Intervallschachtelung bewiesen, das heißt, beide gehen gegen denselben Grenzwert. Und dann zeigen wir noch, dass der Grenzwert von a_n gleich 1 ist, was für b_n ja dann auch gilt.

Monotonie von a_n:

$$a_n = \frac{n!}{n^n\cdot e^{-n}\sqrt{2\pi n}}$$

$$a_{n+1} = \frac{(n+1)!}{(n+1)^{n+1}\cdot e^{-(n+1)}\sqrt{2\pi(n+1)}}$$

und nach (2) linke Seite

$$\frac{a_n}{a_{n+1}} = \frac{\left(1+\frac{1}{n}\right)^{n+\frac{1}{2}}}{e} > 1$$

also ist a_n monoton fallend.

Monotonie von b_n:

$$b_n = \frac{n!}{n^n \cdot e^{-n}\sqrt{2\pi n} \cdot e^{\frac{1}{12n}}}$$

$$b_{n+1} = \frac{(n+1)!}{(n+1)^{n+1} \cdot e^{-(n+1)}\sqrt{2\pi(n+1)} \cdot e^{\frac{1}{12(n+1)}}}$$

und nach (2) rechte Seite

$$\frac{b_n}{b_{n+1}} = \frac{(1+\frac{1}{n})^{n+\frac{1}{2}}}{e \cdot e^{\frac{1}{12n(n+1)}}} < 1$$

also ist b_n monoton wachsend. Und es ist leicht zu sehen, dass $a_n > b_n$, daher sind die Folgen konvergent.

Konvergenz von $a_n - b_n$:

$$a_n - b_n = \frac{n!}{n^n \cdot e^{-n}\sqrt{2\pi n}} - \frac{n!}{n^n \cdot e^{-n}\sqrt{2\pi n} \cdot e^{\frac{1}{12n}}}$$

$$= a_n(1 - e^{-\frac{1}{12n}})$$

a_n konvergiert, das sahen wir oben; der Klammerterm geht gegen null; also geht $a_n - b_n$ gegen null.

Jetzt ist man fast fertig, denn man stellt fest: $a_n > 1 > b_n$ und $a_n - b_n \to 0$ bedeutet, dass a_n und b_n beide gegen 1 gehen. Und die Sache ist erledigt.

Wir hatten die **Möglichkeit eines zweiten Weges** angedeutet. Wir nutzen zum Nachweis von $a_n \to 1$ nur die linke Seite der Doppelungleichung 1, also auch nur die Folge a_n. Wir hatten herausgefunden, dass a_n monoton fällt. Da sie nur positive Glieder hat, ist sie beschränkt und muss also konvergieren.

Der Grenzwert sei g. Auf ziemlich raffinierte Art berechnen wir diesen Grenzwert. Wir bestimmen nämlich a_n^2 und a_{2n} und deren Quotient; dessen Grenzwert ist auch g, aber leichter zu bestimmen.

$$\lim_{n\to\infty} a_n = g \quad \Rightarrow \lim_{n\to\infty} a_n^2 = g^2 \quad \wedge \quad \lim_{n\to\infty} a_{2n} = g$$

$$\lim_{n\to\infty} \frac{a_n^2}{a_{2n}} = g^2/g = g$$

$$\frac{a_n^2}{a_{2n}} = \frac{\frac{(n!)^2}{n^{2n}e^{-2n}\cdot 2\pi n}}{\frac{(2n)!}{(2n)^{2n}e^{-2n}\cdot\sqrt{2\pi 2n}}} = \frac{(n!)^2 \cdot 2^{2n}}{(2n)!\sqrt{n\pi}}$$

Dieser Term hat den Grenzwert 1. Warum?
Unserer Ankündigung folgend müßten jetzt die Integrale dran sein.
Setzt man $S_n = \int \sin^n x dx$ dann gilt

$$S_n = -\frac{1}{n}\sin^{n-1} x \cdot \cos x + \frac{n-1}{n} S_{n-2}$$

Das kann man mit partieller Integration leicht zeigen. Daraus ergibt sich dann
für gerade n

$$\int_0^{\frac{\pi}{2}} \sin^n x dx = \frac{n-1}{n} \cdot \frac{n-3}{n-2} \cdot \dots \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{\pi}{2} \tag{3}$$

für ungerade n

$$\int_0^{\frac{\pi}{2}} \sin^n x dx = \frac{n-1}{n} \cdot \frac{n-3}{n-2} \cdot \dots \cdot \frac{4}{5} \cdot \frac{2}{3} \tag{4}$$

Nun ist aber

$$\int_0^{\frac{\pi}{2}} \sin^{n+1} x dx < \int_0^{\frac{\pi}{2}} \sin^n x dx < \int_0^{\frac{\pi}{2}} \sin^{n-1} x dx$$

und daher mit geradem n, also $n = 2m$ nach (3) und (4)

$$\frac{2 \cdot 4 \cdot 6 \cdots (2m)}{3 \cdot 5 \cdot 7 \cdots (2m+1)} < \frac{\pi}{2} \cdot \frac{1 \cdot 3 \cdot 5 \cdots (2m-1)}{2 \cdot 4 \cdot 6 \cdots (2m)} < \frac{2 \cdot 4 \cdot 6 \cdots (2m-2)}{3 \cdot 5 \cdot 7 \cdots (2m-1)}$$

oder

$$\frac{2^2 \cdot 4^2 \cdot 6^2 \cdots (2m-2)^2 \cdot (2m)^2}{3^2 \cdot 5^2 \cdot 7^2 \cdots (2m-1)^2 \cdot (2m+1)} < \frac{\pi}{2} < \frac{2^2 \cdot 4^2 \cdot 6^2 \cdots (2m-2)^2 \cdot 2m}{3^2 \cdot 5^2 \cdot 7^2 \cdots (2m-1)^2}$$

Der Quotient aus oberer und unterer Schranke für $\frac{\pi}{2}$ ist $\frac{2m}{2m+1}$; mit $m \to \infty$ strebt er gegen 1. Die Grenzwerte von linkem und rechtem Term sind also gleich. Daher ist

$$\frac{\pi}{2} = \lim_{m\to\infty} \frac{2^2 \cdot 4^2 \cdot 6^2 \cdots (2m-2)^2 \cdot (2m)^2}{3^2 \cdot 5^2 \cdot 7^2 \cdots (2m-1)^2 \cdot (2m+1)} \tag{5}$$

oder

$$\frac{\pi}{2} = \lim_{m\to\infty} \frac{2^2 \cdot 4^2 \cdot 6^2 \cdots (2m-2)^2 \cdot 2m}{3^2 \cdot 5^2 \cdot 7^2 \cdots (2m-1)^2} \tag{6}$$

Die Gleichungen (5) und (6) findet man in der Literatur als sogenanntes Wallis-Produkt. Es wurde benannt nach John Wallis (1616-1703). Die Gleichung (6) können wir umformen zu

$$\sqrt{\pi} = \lim_{m\to\infty} \frac{2 \cdot 4 \cdot 6 \cdots (2m-2) \cdot 2m}{3 \cdot 5 \cdot 7 \cdots (2m-1) \cdot \sqrt{m}} = \lim_{m\to\infty} \frac{2^{2m} \cdot (m!)^2}{(2m)! \cdot \sqrt{m}}$$

Somit ist

$$\lim_{n\to\infty} \frac{a_n^2}{a_{2n}} = \lim_{n\to\infty} \frac{2^{2n}\cdot(n!)^2}{(2n)!\cdot\sqrt{n}}/\sqrt{\pi}$$
$$= \sqrt{\pi}/\sqrt{\pi} = 1$$

Das war´s.

15.4.2 Stirling II

Unser Ziel ist die Abschätzung

$$(n+a)! \approx n! n^a e^{\frac{a^2+a}{2n}}$$

Wir logarithmieren die Ungleichung Stirling I

$$n! \approx \sqrt{2\pi n}\cdot(n/e)^n$$

$$\ln n! \approx \ln(\sqrt{2\pi n}\cdot(n/e)^n)$$
$$= \ln\sqrt{2\pi} + \frac{1}{2}\ln n + n\ln n - n\ln e$$

Jetzt betrachten wir nicht $(n+a)!$ alleine, sondern nehmen das $n!$ von der anderen Seite herüber und logarithmieren ebenfalls.

$$\ln\frac{(n+a)!}{n!}$$

Wenn wir diesen Tern ausgedrückt haben, können wir delogarithmieren und $n!$ auf die andere Seite bringen und sind fertig.

$$\ln\frac{(n+a)!}{n!} = \ln(n+a)! - \ln n!$$

und wir setzen den logarithmierten Stirling-Term ein.

$$= \frac{1}{2}\ln(n+a) + \ln\sqrt{2\pi} + (n+a)\ln(n+a) - (n+a)\ln e) - (\frac{1}{2}\ln n + \ln\sqrt{2\pi} + n\ln n - n\ln e)$$

$$= \frac{1}{2}\ln(n+a) + (n+a)\ln(n+a) - a - \frac{1}{2}\ln n - n\ln n$$

Ein kleiner Trick:

$$= \frac{1}{2}\ln\frac{n(n+a)}{n} + (n+a)\cdot\ln\frac{n(n+a)}{n} - a - \frac{1}{2}\ln n - n\ln n$$

$$= \underbrace{\frac{1}{2}\ln n} + \frac{1}{2}\ln 1 + \frac{a}{n} + \underline{(n+a)\cdot\ln n} + (n+a)\cdot\ln(1+\frac{a}{n}) - a - \underbrace{\frac{1}{2}\ln n} - \underline{n\ln n}$$

$$= \frac{1}{2}\ln(1+\frac{a}{n}) + a\cdot\ln n + (n+a)\cdot\ln(1+\frac{a}{n}) - a$$

$$= a\ln n - a + (n+a+\frac{1}{2})\cdot ln(1+\frac{a}{n})$$

Nun ist aber nach Taylor

$$ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots$$

$$= a\ln n - a + (n + a + \frac{1}{2}) \cdot \left(\frac{a}{n} - \frac{1}{2}(\frac{a}{n})^2 + \frac{1}{3}(\frac{a}{n})^3 - \cdots\right)$$

$$= a\ln n - a + a + \frac{a^2+a}{2n} \cdots$$

und wir erhalten

$$\ln \frac{(n+a)!}{n!} \approx a \cdot \ln n + \frac{a^2+a}{2n}$$

und durch potenzieren

$$\frac{(n+a)!}{n!} \approx n^a \cdot e^{\frac{a^2+a}{2n}}$$

also

$$(n+a)! \approx n! \cdot n^a \cdot e^{\frac{a^2+a}{2n}}$$

15.5 stetige Verteilung, Wahrscheinlichkeitsdichte

Mit der Funktion φ liegt eine stetige Funktion vor, die alle Zwischenwerte für x zulässt. Die Wahrscheinlichkeit, genau eine der Reellen Zahlen zwischen $-\infty$ und ∞ zu treffen, ist aber gleich 0. Daher kann eine sinnvolle Frage nach Wahrscheinlichkeiten in solchen Fällen nur mit Intervallen verbunden sein.
Andererseits ist die Zahl n bei Bernoulli so groß, dass auch hier die einzelnen Wahrscheinlichkeiten sehr klein sind; auch dies spricht für die Untersuchung von Intervallwahrscheinlichkeiten.
Eine stetige Wahrscheinlichkeitsfunktion nennen wir **Wahrscheinlichkeits-Dichte**, ihre Summenkurve nennen wir **Wahrscheinlichkeits-Verteilung**. Die Wahrscheinlichkeiten werden mittels Integralen berechnet.

Definition 15.1 (Dichte) *f heißt Wahrscheinlichkeitsdichte der stetigen Zufallsgröße x, wenn*

$$f(x) \geq 0 \quad \wedge \quad \int_{-\infty}^{\infty} f(x)dx = 1$$

Definition 15.2 (Verteilung) *F heißt Wahrscheinlichkeitsverteilung der stetigen Zufallsgröße x, wenn*

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

15.6 Erwartungswert und Varianz stetiger Zufallsgrößen

Erwartungswert und Varianz definieren wir in Entsprechung zu den Kennzahlen der Statistik.

Definition 15.3 (Erwartungswert) *E oder μ heißt Erwartungswert der stetigen Zufallsgröße x, wenn*

$$E(x) = \mu_x = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

Definition 15.4 (Varianz) *V oder σ^2 heißt Varianz der stetigen Zufallsgröße x, wenn*

$$V(x) = \sigma_x^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)dx$$

15.7 Die Gauss-Verteilung

Will man auf die Umformungen von Moivre-Laplace und Stirling verzichten, kann man auf folgende Weise die Gauss-Funktion anschaulich klarmachen.
Die Glockenform der Verteilung kann durch unterschiedliche Funktionen beschrieben werden. Den Erfindern der Wahrscheinlichkeitstheorie erschien eine Exponentialfunktion der Form

$$g(x) = c \cdot e^{-kx^2}$$

angemessen. Wir bestimmen die Parameter c und k aus den Bedingungen, dass die Funktion den Erwartungswert 0 und die Varianz 1 haben soll.

$$\int_{-\infty}^{\infty} x \cdot c \cdot e^{-kx^2} dx = \lim_{t\to\infty} [-\frac{1}{2k} ce^{-kx^2}]_{-t}^{+t} = 0$$

Der Erwartungswert ist also offenbar automatisch 0 geworden. Das ist gut so, liefert aber keine Information über c oder k.

$$\int_{-\infty}^{\infty} x^2 \cdot ce^{-kx^2} dx$$

Wir integrieren partiell.

$$= \int_{-\infty}^{\infty} x \cdot x \cdot ce^{-kx^2} dx$$

$$= \underbrace{\lim_{t\to\infty} [x \cdot (-\frac{1}{2k} ce^{-kx^2})]_{-t}^{+t}}_{=\,0} + \underbrace{\int_{-\infty}^{\infty} \frac{1}{2k} \cdot ce^{-kx^2} dx}_{nicht\ elementar}$$

$$= \frac{1}{2k} \int_{-\infty}^{\infty} ce^{-kx^2} dx$$

Das Integral über die Dichte soll 1 sein. Wir erhalten daher zunächst

$$= \frac{1}{2k}$$

und weil die Varianz 1 sein soll

$$k = \frac{1}{2}$$

Ohne Rechnung geben wir jetzt das nicht elementare Integral an.

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}$$

Daraus erhalten wir c.

$$\int_{-\infty}^{\infty} \frac{1}{2k} \cdot ce^{-kx^2} dx = c \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = c \cdot \sqrt{2\pi} = 1$$

$$c = \frac{1}{\sqrt{2\pi}}$$

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} = \varphi(x)$$

Damit haben wir die Dichte der standardisierten Gauss-Verteilung gefunden.

Definition 15.5 (Normalverteilung) *Die zur Gauss-Funktion gehörende Verteilung*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt$$

heißt ***Gauss-Verteilung*** *oder* ***Normalverteilung****.*
Da sie Erwartungswert 0 und Varianz 1 hat, heißt sie auch Standard-Normalverteilung oder N(0;1)-Verteilung.

Viele Zufallsversuche können mit der Normalverteilung beschrieben werden, sie haben jedoch den Erwartungswert $\mu \neq 0$ und die Varianz $\sigma^2 \neq 1$. Diese Verteilungen werden transformiert mit

$$z = \frac{x - \mu}{\sigma} \quad bzw. \quad u = \frac{t - \mu}{\sigma}$$

Durch die Transformation mit $\frac{1}{\sigma}$ wurde das Integral, also die Fläche verändert. Wir erhalten

$$du = \frac{1}{\sigma} dt$$

$$\Phi(z) = \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{z} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2}dt$$

$$= \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{z} e^{-\frac{1}{2}u^2}\sigma du$$

$$= \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{z} e^{-\frac{1}{2}u^2}du$$

Die Dichte der $N(\mu;\sigma)$-Verteilung ist

$$\varphi_{\mu;\sigma}(x) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

16 weitere Verteilungen

Wenn ein Merkmal binomialverteilt ist, und wir haben die passende Tabelle, dann können wir die Wahrscheinlichkeiten relativ exakt bestimmen. Haben wir die Tabelle nicht zur Hand, benötigen wir eine Ersatzfunktion. Für kleine Wahrscheinlichkeiten in Verbindung mit großen Stichprobenumfängen ist die Poissonverteilung eine gute Näherung. Für mittlere Wahrscheinlichkeiten und große Stichprobenumfänge ist die Normalverteilung eine gute Näherung.

Nun gibt es aber auch Merkmale, die nicht binomialverteilt sind, da die Übergänge zwischen den Merkmalwerten praktisch stetig sind. Zum Beispiel Körpergrößen: die Kleidungsbranche möchte die Verteilung kennen, damit sie entsprechende Fabrikationsanteile berechnen kann. Zum Beispiel Schraubenlängen: die Fabrik möchte die Verteilung kennen, damit sie die Qualitätskontrolle mathematisch begründen kann. Zum Beispiel die Lebensdauer elektronischer Bauteile: hier möchte nicht zuletzt der Händler die Verteilung kennen, damit er den Bedarf an Ersatzteilen oder überhaupt den Reparaturbedarf planen kann. Die beiden ersten Beispiele sind normalverteilt; neuere Untersuchungen scheinen allerdings zu zeigen, dass das bei den Körpergrößen nicht so ganz stimmt. Das letzte Beispiel kann mit der **Exponentialverteilung** beschrieben werden. Diese Verteilungen werden allgemein untersucht, ihr Erwartungswert und ihre Varianz bestimmt, um sie auf verschiedene Gegebenheiten anwenden zu können.

Es sind aber noch ganz andere Verteilungen denkbar. Bei der Binomialverteilung haben wir Unabhängigkeit vorausgesetzt, sozusagen mit Zurücklegen gezogen. Was geschieht, wenn wir ohne Zurücklegen ziehen? Wir erhalten dann die **hypergeometrische Verteilung**. Wenn wir den ersten Erfolg einer Versuchsreihe als Variable betrachten, erhalten wir die **geometrische Verteilung**. Diese untersuchen wir hier aber nicht.

17 Normalverteilung in der Praxis

17.1 Qualitätskontrolle

Man hat eine Stichprobe entnommen, aus der Folgerungen gezogen werden sollen. Es kann eine Qualitätskontrolle sein, die Aufschluss über den Ausschussanteil einer Fertigung gibt. Der Erwartungswert ist entweder schon bekannt; er ist dann ein Sollwert, der bei der Fertigung eingehalten werden soll; oder er ist aus der Probe zu bestimmen; dann nehmen wir den Mittelwert als Ersatz. Die Streuung gibt Auskunft über Toleranzbereiche.

Da die Stichprobe zwar als normalverteilt angesehen werden kann, aber meist nicht standardisiert ist, transformieren wir mit $z = \frac{x-\mu}{\sigma}$ und können dann Wahrscheinlichkeiten mit der Standardnormalverteilung bestimmen.

Als Beispiel betrachten wir die Fabrikation von Bolzen der Länge 3 cm; das wäre der Sollwert. Eine Probe von $n = 100$ Stück ergab:

x cm Länge	<2.6	2.6	2.8	3	3.2	3.4	>3.4
bei y Schrauben	0	15	50	28	6	1	0

Wir erhalten mit den üblichen Rechenverfahren

einen Mittelwert von $\overline{x} = 2.856$ und eine Streuung von $s = 0.165$.

Wieviel Ausschuss ist zu erwarten, wenn der Toleranzbereich $\pm$ eine Standardabweichung vom Soll beträgt, und die Maschine richtig eingestellt ist?

Es ist $\mu = 3$ und $\sigma = 0.165$, der Toleranzbereich $2.835 < X < 3.165$.

$$P(2.835 < X < 3.165) = \Phi(\frac{3.165-3}{0.165}) - \Phi(\frac{2.835-3}{0.165})$$

$$= 2\Phi(1) - 1 = 2 \cdot 0.8413 - 1 = 0.6827$$

68% werden also akzeptiert, 32% erfüllen nicht die Bedingungen.

Offenbar ist die Maschine ja nicht justiert. Wieviel Ausschuss ist zu erwarten, wenn wir mit dem Toleranzbereich wie oben $2.835 < X < 3.165$ rechnen? Jetzt ist $\mu = 2.856$ und $\sigma = 0.165$.

$$P(2.835 < X < 3.165) = \Phi(\frac{3.165-2.856}{0.165}) - \Phi(\frac{2.835-2.856}{0.165})$$

$$= \Phi(1.818) - \Phi(-0.127) = 0.96546 - 0.44947 = 0.51599$$

Es sind 49% als Ausschuss zu bezeichnen.

49% Ausschuss ist zu viel; in welchem Toleranzbereich würden die Bolzen ausgeliefert, wenn man bei so ungenauer Fertigung 5% Ausschuss für vertretbar hält?

$$\Phi(1.96) = 0.975 \Rightarrow 1.96 = \frac{x_o - 3}{0.165} \Rightarrow x_o = 3.323;\ x_u = 2.676$$

Der Toleranzbereich wäre jetzt $2.676 < X < 3.323$.

Den Mittelwert haben wir jetzt justiert, die Streuung aber bleibt. Eine einfache Methode der Fertigungskontrolle nutzt sogenannte Kontrollkarten. Man berechnet Kontrollgrenzen; in unserem Beispiel wäre die Kontrollgrenze $3.0 \pm 1.96 \cdot 0.165 = 3.0 \pm 0.32$. Nun greifen wir ein Stück aus der Fertigung und messen; verläßt es diesen Bereich, sucht man nach Störungen. Im Grunde wird hier ein Test mit dem Signifikanzniveau 5% durchgeführt.

17.2 Stetigkeitskorrektur

Die Normalverteilung wird oft als Ersatz für die Binomialverteilung genommen. Die Binomialverteilung ist eine diskrete Funktion, sie ist nur für natürliche Zahlen $0, 1, 2, \cdots$ definiert. Die Normalverteilung hingegen ist für alle Zwischenwerte definiert. Der Wert $B_{300;0.7}(20)$ zum Beispiel kann mit der Funktion φ näherungsweise bestimmt werden, zumeist liegen aber nur die Werte von Φ tabelliert vor. Um so den Wert für $k = 20$ zu erhalten, benötigen wir ein Intervall. Im Histogramm der Binomialverteilung hat jedes Rechteck die Breite 1. Das Intervall ist daher $k \pm 0.5$. Wir nutzen die sogenannte globale Näherung mit Stetigkeitskorrektur

$$B_{n;p}(k) \approx \Phi(\frac{k - \mu + 0.5}{\sigma}) - \Phi(\frac{k - \mu - 0.5}{\sigma})$$

Im allgemeinen geht man davon aus, dass $\sigma > 3$ sein sollte, damit die Normalverteilung ein guter Ersatz für die Binomialverteilung ist. Wie groß sollte σ sein, damit man auf die Stetigkeitskorrektur verzichten kann? Die Tabellen sind bis zur zweiten Nachkommastelle berechnet. Für $n = 500$ und $p = 0.6$ erhält man $\sigma = \sqrt{120} \approx 11$. Ist $\sigma > 10$, so ist die Korrektur $\frac{0.5}{\sigma} < 0.05$.

Es gilt $\Phi(0.60) = 0.7257$ und $\Phi(0.65) = 0.7422$, das macht eine Differenz von 2 Prozentpunkten. Bei $x = 1.5$ macht die Differenz kaum noch einen Unterschied.

17.3 Toleranzgrenzen

Mit der Ungleichung von Tschebyscheff für binomialverteilte Größen können wir Aufgaben lösen, die für die Fertigungskontrolle interessant sind. Nehmen wir an, der Ausschussanteil einer Fertigung sei $q = 0.1 = 10\%$, dann ist in einer Stichprobe von 1000 Stück eine mittlere Zahl von 900 brauchbaren Stücken zu erwarten. In welchem Bereich schwankt diese Zahl? Wir fragen nach der Wahrscheinlichkeit, dass von den 1000 herausgegriffenen Stücken zwischen 880 und 920 gute sind.
Es ist $n = 1000,\ p = 0.9,\ q = 0.1,\ \mu = 900,\ \sigma = 9.49,\ c = \frac{20}{1000} = 0.02.$

$$P(|h - p| < c) \geq 1 - \frac{pq}{nc^2}$$

$$P(|h - 0.9| < 0.02) \geq 1 - \frac{0.9 \cdot 0.1}{1000 \cdot 0.02^2} = 0.775$$

Mit einer Wahrscheinlichkeit von 78% liegen die getesteten Stücke in den gegebenen Schranken.

Mit Hilfe der Normalverteilung hätten wir etwas andere Werte erhalten.

$$\Phi(\frac{920-900}{9.49}) - \Phi(\frac{880-900}{9.49}) = \Phi(2.107) - \Phi(-2.107)$$
$$= 2 \cdot 0.98249 - 1 = 0.96498$$

Naja, Tschebyscheff liefert nun einmal eine ziemlich grobe Abschätzung. Berechnen wir den Bereich mit Bernoulli, erhalten wir 96.474%. Gauss liegt etwas höher, Tschebyscheff weit niedriger. Das bedeutet nicht, dass Gauss am genauesten ist: er liefert auch nur eine Abschätzung für Bernoulli.

Eine zweite Fragestellung ist aber zu den Toleranzgrenzen denkbar. Wie groß ist n zu wählen, damit die Wahrscheinlichkeit für die oben angenommene Abweichung größer als 95% ist?
Es ist $P = 0.95,\ p = 0.9,\ q = 0.1$; wir bleiben bei $c = 0.02$.
Mit Tschebyscheff erhalten wir

$$1 - \frac{0.9 \cdot 0.1}{n \cdot 0.02^2} > 0.95$$
$$\frac{0.9 \cdot 0.1}{n \cdot 0.02^2} < 0.05$$
$$\frac{0.9 \cdot 0.1}{0.05 \cdot 0.02^2} < n$$
$$n > 4500$$

Allgemein erhalten wir

$$n > \frac{pq}{c^2(1-P)}$$

Groß P und klein p könnte man verwechseln, daher ersetzt man oft P durch S; S wird auch **statistische Sicherheit** genannt.

Die Rechnung mit Normalverteilung gestaltet sich folgendermaßen.

Wir halten zum besseren Verständnis zunächst p, X, c und S variabel.

$$P(|h-p| < c) \geq S$$

$$|h-p| < c \Leftrightarrow |\frac{X}{n} - p| < c \Leftrightarrow |X - np| < nc \Leftrightarrow |\frac{X-np}{\sigma}| < \frac{nc}{\sigma}$$
$$\Leftrightarrow -\frac{nc}{\sigma} < \frac{X-\mu}{\sigma} < \frac{nc}{\sigma}$$

Es wird also

$$P(|h-p| < c) = P(-\frac{nc}{\sigma} < \frac{X-\mu}{\sigma} < \frac{nc}{\sigma})$$

Der Einfachheit halber setzen wir $\frac{nc}{\sigma} = z$. Dann ergibt sich mit Gauss

$$\Phi(z) - \Phi(-z) \geq S \Leftrightarrow 2\Phi(z) - 1 \geq S \Leftrightarrow \Phi(z) \geq (1+S) : 2$$

Jetzt nutzen wir die Zahlen aus der Aufgabe. $S = 0.95$, $p = 0.9$, $q = 0.1$, $c = 0.02$. Aus der Tabelle entnehmen wir $z \geq 1.96$.

$$\frac{0.02n}{\sqrt{n \cdot 0.9 \cdot 0.1}} \geq 1.96 \Rightarrow 0.02 \cdot \sqrt{n} \geq 1.96 \cdot \sqrt{0.9 \cdot 0.1}$$

$$\Rightarrow 0.0004n \geq 1.96^2 \cdot 0.09 \Rightarrow n \geq 864.36 > 864$$

Man muss also 865 Stücke untersuchen, um mit 95% Sicherheit in der relativen Häufigkeit eine maximale Abweichung von 0.02 von der Wahrscheinlichkeit $p = 0.9$ zu erreichen; das heißt, dass $0.88 < h < 0.92$ ist.

Das hätten wir auch ohne den Umweg über X rechnen können. Wir erinnern uns, dass $E(h) = p$ und $V(h) = \sigma_h^2 = \frac{pq}{n}$ ist.

$$|h - p| < c \Leftrightarrow -c < h - p < c \Leftrightarrow \frac{-c}{\sigma_h} < \frac{h-p}{\sigma_h} < \frac{c}{\sigma_h}$$

Wir setzen $\frac{c}{\sigma_h} = z$, also $\frac{c}{\sqrt{\frac{pq}{n}}} = \frac{c\sqrt{n}}{\sqrt{pq}} = z$. Zu $S = 95\%$ erhalten wir $z = 1.96$, das kennen wir schon.

$$\frac{0.02\sqrt{n}}{\sqrt{0.9 \cdot 0.1}} \geq 1.96 \Rightarrow 0.02 \cdot \sqrt{n} \geq 1.96 \cdot \sqrt{0.9 \cdot 0.1}$$

$$\Rightarrow 0.0004n \geq 1.96^2 \cdot 0.09 \Rightarrow n \geq 864.36 > 864$$

Interessant ist die Frage nach diesem Intervall, wenn p unbekannt ist. Das Produkt pq kommt in der Varianz vor, man sollte also $p \cdot q$ abschätzen.

Es gilt $p(1-p) \leq \frac{1}{4}$, was mit der Quadratfunktion leicht nachzuweisen ist. Wir setzen die oben ausgeführte Rechnung mit unbekanntem p fort.

$$P(|h - p| < c) = P(-\frac{nc}{\sigma} < \frac{X - \mu}{\sigma} < \frac{nc}{\sigma})$$

$$\frac{0.02n}{\sqrt{n \cdot p \cdot (1-p)}} \geq 1.96 \Rightarrow 0.02 \cdot \sqrt{n} \geq 1.96 \cdot p \cdot (1-p)$$

$$\Rightarrow 0.0004n \geq 1.96^2 \cdot p \cdot (1-p)$$

$$n \geq 9604 \cdot p \cdot (1-p) \Rightarrow n \geq 9604 \cdot \frac{1}{4} \Rightarrow n \geq 2401$$

Bei dieser Abschätzung darf man nicht den Fehler machen, $pq \leq 0.25$ an die Ungleichung anzuhängen; das würde zu einem unsinnigen Schluss führen.

$$n \geq 9604 \cdot p \cdot (1-p) \leq 9604 \cdot 0.25$$

Die Argumentation läuft folgendermaßen: Es soll gelten $n \geq 9604 \cdot p \cdot (1-p)$; wenn man nun ein größeres n wählt, bleibt die Ungleichung erhalten; man setzt also die Abschätzung $pq \leq 0.25$ zwischen n und $9604 \cdot p \cdot (1-p)$ und erhält

$$n \geq 9604 \cdot p \cdot (1-p) \Rightarrow n \geq 9604 \cdot 0.25 = 2401$$

Bei fünf Variablen P, n, c, h und p sind natürlich weitere Fragen sinnvoll. In welchem Bereich schwankt das h, wenn wir als Wahrscheinlichkeit 95% vorgeben? Wir fragen also nach dem c.
Ausgehend von Tschebyscheff erhalten wir

$$1 - \frac{p(1-p)}{nc^2} = P = S \Rightarrow c = \sqrt{\frac{p(1-p)}{n(1-S)}}$$

In unserem Beispiel sei $P = S = 0.95,\ p = 0.9,\ q = 0.1,\ n = 1000.$

$$c = \sqrt{\frac{0.1 \cdot 0.9}{1000 \cdot (1 - 0.95)}} = 0.0424$$

Wir erhalten den Bereich 1000 ± 42; darin liegen die Ergebnisse mit 95% Wahrscheinlichkeit.

Mit Gauss errechnen wir zunächst die Grenze in der Standardverteilung, transformieren, und erhalten dann das c.

$$0.95 = 2\Phi(z) - 1 \Rightarrow \Phi(z) = 0.975 \Rightarrow z = 1.96$$

$$\Rightarrow x = 1.96 \cdot \sigma + \mu = 1.96 \cdot 9.49 + 900 = 918.6$$

also $c = \frac{19}{1000} = 0.019$.
Wir erhalten den Bereich 1000 ± 19; darin liegen die Ergebnisse mit 95% Wahrscheinlichkeit.

Nun ist es oft so, dass man aus einer Stichprobe eine relative Häufigkeit h gewinnen konnte, aber den wahren Wert der Wahrscheinlichkeit p nicht kennt. Man möchte aber wissen, wie weit p um h schwankt. Mit Gauss bestimmen wir zu $P = S$ die Grenze z und berechnen anschließend p. In unserem Beispiel sei $P = 0.95$, also $z = 1.96$, p unbekannt, $n = 1000$ und h ein bekannter Wert; hier sei $h = 0.89$; wir rechnen zunächst mit variablem h. Aus vorherigen Kapiteln wissen wir, dass $\mu_h = p$, $\sigma_h = \sqrt{\frac{pq}{n}}$.

$P(|\frac{h-p}{\sigma_h}| < z) = 0.95$ mit $z = 1.96$ ergibt

$$(\frac{h-p}{\sqrt{\frac{pq}{n}}})^2 < z^2 \Rightarrow z^2 \cdot p \cdot (1-p) > n \cdot (h-p)^2$$

$$1.96^2 \cdot p \cdot (1-p) > 1000 \cdot (0.89 - p)^2$$

Die quadratische Ungleichung hat die Lösungen $0.8691 < p < 0.90793$, wozu $p = 0.9$ passen würde.

18 Unabhängigkeit

Nun ist es an der Zeit, dem Thema **Unabhängigkeit** nachzugehen. Als wir die Bernoullikette untersucht haben, wurde vorausgesetzt, dass von Stufe zu Stufe die Wahrscheinlichkeit unverändert bleibt. Das dazu passende Urnenmodell wurde durch ,,Ziehen mit Zurücklegen" beschrieben. Das Erscheinen einer FÜNF beim siebten Würfelwurf beeinflusst nicht die Warscheinlichkeit beim achten Wurf. Die Ereignisse sind unabhängig. Legt man nicht zurück, ändert sich die Wahrscheinlichkeit, weil sich der Anteil an der Population ändert. Die Ereignisse sind abhängig.

Meinungsumfragen oder auch Fertigungskontrollen entsprechen meist dem Ziehen ohne Zurücklegen; eine Person wird ja nicht zweimal befragt, oder ein Bauteil zweimal kontrolliert. Aber wegen des großen Stichprobenumfangs kann man den dadurch entstehenden Fehler vernachlässigen.

Das sind einfach zu durchschauende Beispiele für Abhängigkeit oder Unabhängigkeit. In der Regressionsanalyse haben wir Merkmale auf Abhängigkeit untersucht und gesehen, dass diese Abhängigkeit nicht leicht zu durchschauen ist.

Definition 18.1 *Zwei Ereignisse A und B heißen stochastisch unabhängig, wenn gilt $P(A \cap B) = P(A) \cdot P(B)$.*

Betrachten wir die Merkmale Geschlecht und sportliche Aktivität.

	Männer	Frauen	Summe
Sport	48	12	60
kein Sport	16	24	40
Summe	64	36	100

Gilt $P(Mann \cap Sport) = P(Mann) \cdot P(Sport)$?
$P(Mann \cap Sport) = 0.48$; $P(Mann) \cdot P(Sport) = 0.64 \cdot 0.6 = 0.384 \neq 0.48$.
Die Ereignisse sind nicht unabhängig.

19 bedingte Wahrscheinlichkeit

Hängt die Wahrscheinlichkeit eines Ereignisses von der Wahrscheinlichkeit eines anderen Ereinisses ab, so sprechen wir von bedingter Wahrscheinlichkeit. $P(A|B)$ bedeutet die Wahrscheinlichkeit von A, unter der Bedingung, dass B eingetreten ist.

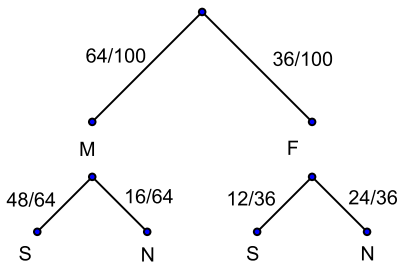
Jetzt können wir Unabhängigkeit neu - aber gleichwertig - definieren.

Definition 19.1 *Zwei Ereignisse A und B heißen stochastisch unabhängig, wenn gilt $P(A|B) = P(A)$.*

Offenbar gilt allgemein für die bedingte Wahrscheinlichkeit die sogenannte Bayes-Formel

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

Beziehen wir uns auf das oben angeführte Beispiel und fragen: wir treffen einen Sport treibenden Menschen; wie wahrscheinlich ist es ein Mann?



Man erkennt an diesem Zahlenbeispiel sehr gut, dass $\frac{48}{64}$ eine bedingte Wahrscheinlichkeit ist. Erstens steht sie für ein Ereignis in der zweiten Stufe des Baumes, hat also als Bedingung M. Das haben wir bei den Bäumen anfangs so nicht wahrgenommen. Zweitens sieht man am Nenner 64, dass sich der Bruch auf die Anzahl der Männer beschränkt; und das ist klar, weil die Bedingung M gilt.

In der Schreibweise der bedingten Wahrscheinlichkeit könnte man also schreiben $P(S|M) = \frac{48}{64}$, $P(N|M) = \frac{16}{64}$, $P(S|F) = \frac{12}{36}$, $P(N|F) = \frac{24}{36}$.

Wenn jetzt der Begriff der bedingten Wahrtscheinlichkeit noch nicht ganz verstanden wurde, kann man sich vielleicht mit folgenden Formulierungen helfen.
$P(S|M)$ ist der Anteil der Sportler an den Männern.

$$P(S|M) = P(\text{wenn Mann, dann Sportler})$$

$P(M|S)$ ist der Anteil der Männer an den Sportlern.

$$P(M|S) = P(\text{wenn Sportler, dann Mann})$$

$P(M \cap S)$ ist der Anteil der sporttreibenden Männer an allen Personen, Männern **und** Frauen.

Gesucht ist $P(M|S)$.

Es gilt $P(M \cap S) = P(M|S) \cdot P(S) = P(S|M) \cdot P(M)$. Eines der beiden Produkte kennen wir.

Wir nutzen die Formel für die bedingte Wahrscheinlichkeit, oder wir verfolgen im Baum die entsprechenden Wege.

$$P(M|S) = \frac{P(M \cap S)}{P(S)} = \frac{\frac{64}{100} \cdot \frac{48}{64}}{\frac{64}{100} \cdot \frac{48}{64} + \frac{36}{100} \cdot \frac{12}{36}} = \frac{48}{60}$$

Eine sinnvolle Fragestellung ist in diesem Zusammenhang folgende: Eine Alarmanlage gibt bei Einbruch Alarm mit einer Wahrscheinlichkeit 0.99; obwohl kein Einbruch stattfindet, gibt sie mit 0.5% Alarm. In dieser Gegend

findet in 1000 Nächten 1 Einbruch statt. Die Anlage gibt Alarm; wie groß ist die Wahrscheinlichleit, dass wirklich eingebrochen wird?

$$P(E|A) = \frac{P(E \cap A)}{P(A)} = \frac{0.001 \cdot 0.99}{0.001 \cdot 0.99 + 0.999 \cdot 0.005} = 0.1654$$