Weiterführende Stochastik

Vorwort

Stochastik hat mich interessiert, seitdem mir dieser Begriff zum ersten Mal begegnet ist. Ich fand es faszinierend, Wahrscheinlichkeiten beim Würfeln, beim Kartenspielen, beim Roulette oder gar beim Lotto berechnen zu können
Aber schon das Rechnen mit bedingten Wahrscheinlichkeiten hat es in sich und führte mich oft in die Irre. Deswegen wollte ich mich irgendwann einmal gründlich mit dieser Thematik auseinandersetzen. Dass ich dazu warten musste, bis ich selbst in Pension gegangen war, wusste ich damals nicht. Aber jetzt hole ich es nach.

Der folgende Text geht nicht über Schulbuchweisheiten hinaus. Das sei vorweggesagt. Das Niveau bewegt sich etwa auf dem eines Leistungskurses in Mathematik. Ich stelle es so dar, wie ich es mir den Stoff selbst erarbeite. Ich bringe ein paar mehr Beispiele und stelle Aufgaben aus unterschiedlicher Sichtweise. Mein erläuternder Text ist da, wo ich es für nötig halte, ausführlich und allgemein verständlich, dort wo ich Verständnis voraussetze, kurz und bündig. Bei den meisten interessierten Schülerinnen und Schülern bin ich in meiner Lehrerzeit damit gut gefahren. Ich hoffe, dass ist auch in diesem Fall so. Bei einem großen Teil meiner Ausführrungen hat mich das wunderbare Lehrbuch von Lambacher Schweizer – Mathematik Qualifikationsphase Leistungskurs/Grundkurs für Nordrhein-Westfalen begleitet. Außerdem gibt es eine Reihe von guten Lehrvideos in YouTube, von denen ich exemplarisch die von Daniel Jung empfehlen kann.

Grundbegriffe

Relative Häufigkeiten

Führe einen Zufallsversuch (z.B. Münzwurf, Würfeln mit 2 Würfeln n- mal durch und notiere die Ergebnisse.
Dann ist die Anzahl k des Ereignisses ω dividiert durch die Anzahl n die relative Häufigkeit des Auftretens des Ereignisses ω

$$h(\omega) = \frac{k}{n}$$

Wahrscheinlichkeiten

Wenn Zufallsversuch beliebig oft durchgeführt werden, so werden sich die relativen Häufigkeiten der einzelnen Ausfälle einem Grenzwert annähern. Diesen Grenzwert bezeichnet man als die *Wahrscheinlichkeit P* eines Ereignisses.
Die Wahrscheinlichkeit P errechnet sich aus:

$$\frac{\textbf{Anzahl günstiger Ausfälle}}{\textbf{Anzahl möglicher Ausfälle}}$$

P ist immer eine Zahl zwischen 0 und 1. Die Summe der Wahrscheinlichkeiten aller möglichen Ausfälle ist immer 1

Die **relative Häufigkeit** ergibt sich aus einem Zufallsversuch oder einer Stichprobe, die n-mal durchgeführt wird. Beispiel: Wenn man 100-mal würfelt, fällt 16- mal eine Sechs. Die relative Häufigkeit beträgt 0,16. Die **Wahrscheinlichkeit** für das Ereignis, dass ein Sechs fällt beträgt allerdings 0,1666… . Die relative Häufigkeit ist ein empirisches Messergebnis. Die Wahrscheinlichkeit ist eine theoretische Größe.

Einstufige Versuche

-Laplace-Versuche: Versuche, in denen die Ausfälle alle die gleiche Wahrscheinlichkeit haben (z.B. Würfeln, Münzen...)
-empirische Versuche. Bei empirischen Versuchen ergeben sich die Wahrscheinlichkeiten aus der relativen Häufigkeit einer langen Versuchsreihe.

Mehrstufige Versuche (z.B. mehrfaches Würfeln, mehrfaches Ziehen einer Karte (mit oder ohne Zurücklegen), Urnenversuche

1. Ziehen mit Zurücklegen

Beispiele für Laplace-Versuche

Werfen einer Münze 3- mal. Wie groß ist die Wahrscheinlichkeit, dass
a) dreimal Wappen fällt
b) zweimal Wappen fällt
c) dass beim Werfen eines Würfels keine Eins fällt

Dazu gehören die folgenden Begriffe, die hier nicht erläutert werden:
<u>Multiplikationsregel</u>
<u>Additionsregel</u>
<u>Gegenwahrscheinlichkeit</u>

Typische Aufgabenstellungen

Die Geburtenstatistiken zeigen, dass die Wahrscheinlichkeit für die Geburt eines Jungen bei etwa 0,514 liegt, die für ein Mädchen bei 0,486. Wie groß ist die Wahrscheinlichkeit bei 2 Kindern für
a) zwei Jungen
b) 1. Kind Junge, 2. Kind Mädchen
c) 1 Junge und ein Mädchen

Eine Familie hat 3 Kinder. Wie groß sind die Wahrscheinlichkeiten für
a) 3 Mädchen
b) mindestens eines der Kinder ist ein Junge

2. Ziehen ohne Zurücklegen

Aus einer Urne mit 5 Kugeln, von denen 2 rot und 3 schwarz sind, werden nacheinander 3 Kugeln (ohne zurücklegen) gezogen.
Zeichne das Baumdiagramm und berechne die Wahrscheinlichkeiten für a) 3 rote Kugeln, b) mindestens 1 rote Kugel

Weiterführende Stochastik

Kombinatorik

Beispiel: Wie viele verschiedene Anordnungen gibt es für die Ziffern 1 bis 5?

Eine Menge, die aus n Elementen besteht, kann mit **n!** Kombinationen angeordnet werden.
5! = 120   Z. B. 12345, 12354, 12434, usw. bis 54321.
Dabei müssen alle fünf Elemente immer in unterschiedlicher Reihenfolge angeordnet werden. Eine Wiederholung von Elementen (11112) ist dabei nicht gestattet.
In diesem Fall spricht man von einer **Permutation**.
Kommen Elemente mehrfach vor, so gilt die Formel:

$\frac{n!}{m1! \cdot m2! \ldots}$ Wie viele Permutationen gibt es für das Wort Mississippi? $\frac{11!}{4!4!2!} = 34650$

Beispiel: Wie viele 3-stellige Zahlen kann man mit den Ziffern 1 bis 5 bilden, wenn Wiederholungen von Ziffern a) nicht erlaubt sind b) erlaubt sind?

zu a) 5*4*3 = 60
zu b) 5^3= 125

Aus einer Menge mit n Elementen lassen sich k Teilmengen (k $\leq n$) bilden.

Man unterscheidet vier Fälle:

a) Jedes Element darf mehrfach auftreten und die Reihenfolge der Elemente ist wesentlich
b) Die Reihenfolge ist wesentlich, aber Wiederholungen sind nicht erlaubt
c) Wiederholungen sind erlaubt, aber die Reihenfolge spielt keine Rolle
d) Die Reihenfolge spielt keine Rolle und Wiederholungen sind nicht erlaubt.
e) Wiederholungen egal, Reihenfolge egal, k Versuche aus n, wobei k auch >n sein kann.

Formeln:

zu a) n^k

zu b) $n \cdot (n-1) \cdot (n-2) \ldots (n-(k-1))$ $oder$ $\frac{n!}{(n-k)!}$

zu c) $\frac{n!}{(n-k)!}$ (Ziehen mit Zurücklegen)

zu d) $\frac{n!}{k!\,(n-k)!}$ (Dies ist die Formel für 6-Richtige im Lotto, Ziehen ohne Zurücklegen)

zu e) $\frac{(n+k-1)!}{k!\,(n-1)!}$

Weitere Aufgaben

1. Wie viele verschiedene Möglichkeiten gibt es, die Buchstaben des Wortes AUTO anzuordnen?
   Permutation: n!
2. Bei einem Rennen starten 12 Rennwagen. Wie viele Kombinationen gibt es für die 3 Erstplatzierten?
   (3 aus 12. Keine Wiederholungen, Reihenfolge ist wesentlich)
   12 x 11 x 10 = 1320
3. Wie viele Möglichkeiten der Ergebnisse gibt es, wenn ein Würfel 3mal geworfen wird?
   n = 6, k = 3
   $\frac{(6+3-1)!}{3!\,(6-1)!} = 56$

4. Es gibt 3 Gerichte zur Auswahl. Es gehen 4 Bestellungen ein. Wie viele Möglichkeiten der Bestellungen gibt es?

$\frac{(3+4-1)!}{4!\,(3-1)!} = 15$ n = 3, k = 4

Der Binomialkoeffizient

Eine Münze wird 5- mal geworfen. Wie groß ist die Wahrscheinlichkeit, dass 3 x Wappen fällt?

c) Lösung mit Baumdiagramm.
   Insgesamt $2^5 = 32$ Pfade
   1. Multiplikationsregel 0,5*0,5*0,5*0,5*0,5 für einen Pfad.
   2. Additionsregel. Wie viele Pfade treffen zu?
      Lösung: Anzahl der Pfade * p(für einen Pfad)

d) Lösung mit Formel n über k
   Wie viele Möglichkeiten gibt es, 3 aus 5 auszuwählen?

5 über 3 = $\frac{5!}{3!(5-3)!} = \frac{120}{6 \bullet 2} = 10$ $\qquad \binom{n}{k} = \frac{n!}{k!(n-k)!}$

D.h. es gibt 10 Pfade im Baumdiagramm. Damit ist
P(3 Wappen)=10/32=0,3125

Anschaulich wird dies anhand eines **Galton-Bretts** verdeutlicht, welches im Prinzip ein Pascal'sches Dreieck nachbildet.

Die Binomialverteilung bei Bernoulli-Experimenten

Unter einer Bernoulli-Kette versteht man eine Reihe von n Versuchen mit den Ergebnissen ‚Treffer' und ‚Niete'.

Binomialverteilung: $P(r)=\binom{n}{r} \cdot p^r \cdot q^{n-r}$

Beispiel: Etwa 20 % aller Menschen sind Linkshänder. Wie groß ist die Wahrscheinlichkeit, dass sich unter 10 Personen 3 Linkshänder befinden?

n = 10, k = 3, p = 0,2, q = 0,8 Lösung: $\binom{10}{3} \cdot 0{,}2^3 \cdot 0{,}8^7 = 0{,}201$

Beispiel: Ein Test besteht aus 5 Fragen mit 4 Antwortmöglichkeiten, von denen eine richtig ist. Wie groß ist die Wahrscheinlichkeit, dass beim zufälligen Ankreuzen genau 3 Antworten richtig sind?

P(3= $\binom{5}{3} \cdot 0{,}25^3 \cdot 0{,}75^2 = \frac{120}{6 \cdot 2} \cdot 0{,}015625 \cdot 0{,}5625 = 0{,}087890625$

Wobei der erste Ausdruck n über r die Anzahl der Möglichkeiten darstellt, 3 richtige Antworten bei 5 Fragen zu haben (hier 10) .

Weiterführende Stochastik

Die Wahrscheinlichkeiten der Binomialverteilung sind in Abhängigkeit von n, r und p in Tabellen nachzulesen bzw. mit einer entsprechenden Funktion im Taschenrechner abzurufen.
Es geht um zwei Berechnungen:

a) Wie groß ist die Wahrscheinlichkeit, dass unter 10 Personen genau 4 Personen mit einer Krankheit infiziert sind, wenn die Infektionswahrscheinlichkeit bei 0,2 liegt.

b) Wie groß ist die Wahrscheinlichkeit, dass unter 10 Personen mindestens 4 infiziert sind, wenn die Infektionswahrscheinlichkeit bei 0,2 liegt.
In diesem Fall handelt es sich um eine kumulierte Wahrscheinlichkeit, deren Werte aus der Tabelle der kumulierten Werte ablesbar sind.
Lösung: Man berechnet oder liest die kumulierte Wahrscheinlichkeit für k = 3.
Der Wert liegt bei 0,8791. D. h. mit 87,91 % iger Wahrscheinlichkeit sind 0,1,2 oder 3 Personen infiziert. Die gesuchte Wahrscheinlichkeit ergibt sich dann aus der Gegenwahrscheinlichkeit
1 - 0,8791 = 0,1209 = 12,09 %.

Weitere Aufgaben

1. Ein Multiple-Choice-Test besteht aus 10 Fragen mit je 4 Auswahlantworten, von denen immer nur eine Antwort richtig ist.
Bestimme die Wahrscheinlichkeit dafür, dass ein zufällig Ratender
a) genau 5 Fragen richtig beantwortet
b) mehr als 5 richtige Antworten hat

2. Beim Abfüllen von Medikamenten in Ampullen soll ein Soll-Wert von 50 ml eingehalten werden. Der Hersteller garantiert, dass mindestens 98 % der Ampullen mindestens 49 ml enthalten.
Es wird eine Stichprobe von 20 Ampullen entnommen. Wie groß ist die Wahrscheinlichkeit, dass
a) mindestens 2 Ampullen weniger als 49 ml enthalten?
b) höchstens 2 Ampullen weniger als 49 ml enthalten?

Zu 1. n = 10, k = 5, p = 0,25, q = 0,75   a) $252 \cdot 0{,}25^5 \cdot 0{,}75^5 = 0{,}584$ = 5,84 %
bei b) sollte man die Tabelle für die kumulierte Binomialverteilung benutzen
für n= 10, k = 5 und p = 0,2 ergibt sich 0,9389. Dieser von 1 abgezogen ergibt 0,0611 also 6,11 %.

zu 2a) wir berechnen, die Wahrscheinlichkeit dafür, dass 0 oder 1 Ampulle weniger als 49 ml enthalten und ziehen diese von 1 ab. n = 20, k = 0 oder 1, p = 0,02, q = 0,98.
Den kumulierten Wert kann man aus der Tabelle ablesen: 0,9401. Damit ist die Wahrscheinlichkeit, dass mindestens 2 Ampullen weniger als 49 ml enthalten, 1-0,9401 = 0,0599 also 6 %.
zu 2b) Dieses ist dann die kumulierte Wahrscheinlichkeit für 0 oder 1 Treffer, also 0,9401 oder 94 %.

Hypergeometrische Verteilung

Beispiel: Aus einem Stapel mit 15 intakten und 5 defekten Batterien werden zufällig 3 Batterien entnommen. Mit welcher Wahrscheinlichkeit befinden sich 2 defekte darunter?

Was ist bei dieser Aufgabe anders? Die Binomialverteilung lässt sich bei Versuchen anwenden, bei denen die Wahrscheinlichkeit p bei jedem Versuch gleichbleibt (Ziehen mit Zurücklegen). Beim **Ziehen ohne Zurücklegen** verändert sich p jedoch mit jedem Zug. Der Unterschied, der sich daraus ergibt, kann bei großem N und kleinen Stichproben n vernachlässigt werden. Wird dieser Umstand allerdings berücksichtigt ergibt sich für die Wahrscheinlichkeit folgende Formel:

$$P(X = k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$$

M = Anzahl der Treffer der Grundgesamtheit N, k = Anzahl der Treffer der Auswahl, n = Auswahl (Stichprobe)

$$P(X = 2) = \frac{\binom{5}{2} \cdot \binom{20-5}{3-2}}{\binom{20}{3}} = \frac{10 \cdot 15}{1140} = 0{,}132 \quad = 13{,}2\ \%$$

Weiterführende Stochastik: Medizinische Tests – Der Fall Gnabry

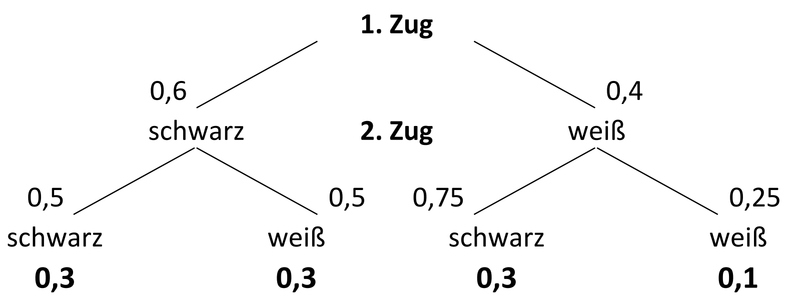
Im Oktober 2020 ging durch die Presse, dass der Fußballnationalspieler Serge Gnabry positiv auf Covid 19 getestet wurde. Daraufhin konnte er nicht an einem Länderspiel teilnehmen und musste in Quarantäne. Einige Tage später ergab sich, dass der Test ein falsches Ergebnis ausgewiesen hatte. Ein weiterer Test ergab, dass Gnabry nicht infiziert war.
Daraufhin war in der Presse vom „Corona-Chaos“ und „Verwirrung“ die Rede.

Es soll untersucht werden, ob es sich hier wirklich um eine chaotische Vorgehensweise gehandelt hat, oder ob solch ein Vorfall nicht sogar zwangsläufig häufig eintreten muss.

Zunächst ein paar theoretische Grundlagen, die erforderlich sind, um das stochastische Vorgehen zu verstehen.
Man benötigt die Begriffe **Vierfeldertafel**, **Baumdiagramm** und die **Formel von Bayes**.

Zwei Ereignisse, die miteinander in Verbindung gebracht werden und zu denen jeweils Wahrscheinlichkeiten für ihr Eintreten bzw. Nichteintreten gehören, können als Baumdiagramm oder in einer Vierfeldertafel dargestellt werden.

Beispiel: In einer Schale liegen 3 schwarze und 2 weiße Kugeln. Es werden nacheinander 2 Kugeln gezogen, wobei die gezogene Kugel nicht zurückgelegt wird. Die Ergebnisse werden in einem **Baumdiagramm** und in einer Vierfeldertafel dargestellt.



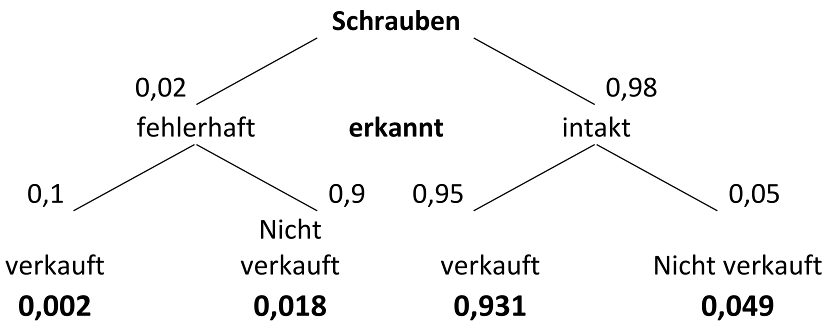
Die Zahlen in der unteren Reihe ergeben sich aus der Multiplikation der Wahrscheinlichkeiten. Die Wahrscheinlichkeit, dass zwei weiße Kugeln gezogen werden liegt beispielsweise bei 0,4 x 0,25 = 0,1 (10 %).
Übertragen in eine Vierfeldertafel ergibt sich folgendes Bild. Die Zahlen in den Zellen ergeben sich aus der Wahrscheinlichkeit, dass sowohl das Spalten- als auch das Zeilenereignis eintritt. Wenn die Zellenwahrscheinlichkeiten das Produkt aus Zeilen- und Spaltensumme sind, dann sind beide Ereignisse (hier: 1. Zug und 2. Zug) voneinander unabhängig. Dies ist hier nicht der Fall, da das Ergebnis des 2. Zuges vom Ergebnis des 1. Zuges abhängt. Die Zahlen der Vierfeldertafel ergeben sich auch aus der unteren Zeile des Baumdiagramms.

Vierfeldertafel

	schwarz	weiß	∑
schwarz	0,3	0,3	0,6
weiß	0,3	0,1	0,4
∑	0,6	0,4	1,0

Beispiel 2.
2 Prozent der Teile einer Produktion sind fehlerhaft. Von den intakten Teilen werden 95 % verkauft. Von den fehlerhaften Teilen gelangen 10 Prozent in den Verkauf.

Darstellung als Baumdiagramm



Die Zahlen unterhalb des Baumdiagramms ergeben sich aus der Multiplikation der Wahrscheinlichkeiten entlang eines Pfades.

Alle diese Und-Verknüpfungen, die sich aus der Multiplikation entlang eines Pfades ergeben werden in der Vierfeldertafel dargestellt.

	fehlerhaft	intakt	∑
Verkauft	0,002	0,931	0,949
Nicht verkauft	0,018	0,049	0,051
∑	0,02	0,98	1

Erläuterung für die Zahl 0,049: 98 Prozent der Teile sind intakt. Die Wahrscheinlichkeit das ein intaktes Teil nicht verkauft wird liegt bei 4,9 %. 0,98 x 0,05 = 0,049.

Mathematisch werden hier zwei Ereignisse mit dem Und-Zeichen verknüpft.
P(intakt ∧ nicht verkauft) = 0,018

Aus der Vierfeldertafel lassen sich nun noch weitere Fragen beantworten:

1. Teil wird verkauft.
2. Schraube ist verkauft und intakt.
3. Ein fehlerhaftes Teil wird verkauft.
4. Ein verkauftes Teil ist intakt.
5. Ein verkauftes Teil ist fehlerhaft.

Zu 1. Die Summenspalte für die 1. Zeile zeigt an, dass 94,9 % aller Teile verkauft werden.
Zu 2. Die Pfadwahrscheinlichkeit ergibt 0,931 = 93,1%.
Zu 3. Die Pfadwahrscheinlich P(fehlerhaft ∧ verkauft) ergibt den Wert 1,8 %
Zu 4. Achtung! Die Ermittlung dieses Wertes ist nicht identisch mit Frage 2. Es handelt sich hier um eine bedingte Wahrscheinlichkeit, deren Wert nicht unmittelbar aus der Tafel abgelesen werden kann:

P(intakt / verkauft) oder $P_{verkauft}$(intakt) Wahrscheinlichkeit, dass eine Schraube intakt ist unter der Bedingung, dass sie verkauft wurde.

Es handelt sich dabei um eine sog. **a posteriori-Wahrscheinlichkeit**, die mit der Formel von Bayes berechnet werden kann.

Formel von Bayes $P_I(A) = \frac{P(A \wedge B)}{P(I)}$

Bedeutet: Die a posteriori-Wahrscheinlichkeit einer Alternative ergibt aus dem Quotienten der Pfadwahrscheinlichkeit (Eintrag in der Tafel bzw. Produkt der Wahrscheinlichkeiten eines Pfades des Baumdiagramms) und der totalen Wahrscheinlichkeit des Indizes (hier sind das die verkauften Teile)

Also $P_{verkauft}(\text{intakt}) = \frac{0{,}931}{0{,}949} = 0{,}98103$ oder 98,1 %

Zu 5. Analog zu Frage 4 $P_{verkauft}(\text{fehlerhaft}) = 0{,}018 : 0{,}949 = 0{,}01897$ oder 1,9 %

Nun zurück zum Fall Gnabry.
Es werden folgende Annahmen getroffen, die ungefähr den realen Werten entsprechen:
Es seien etwa 1 % der Testpersonen infiziert (**Sensitivität**). Ein PCR-Test testet 98 % der Infizierten als positiv und 96 % der nicht-Infizierten als negativ (**Spezifität**).
Es kann folgende Vierfeldertafel erstellt werden:

	Test pos.	Test neg.	∑
Infiziert	0,0098	0,0002	0,0100
Nicht infiziert	0,0396	0,9504	0,9900
∑	0,0494	0,9506	1

Sensitivität: Die Wahrscheinlichkeit, dass das Testergebnis positiv ist, wenn eine Person infiziert ist.
Spezifität: Die Wahrscheinlichkeit, dass das Testergebnis negativ ist, wenn eine Person nicht infiziert ist.

Wenn die Frage beantwortet werden soll, wie wahrscheinlich es ist, dass ein positiv Getesteter nicht infiziert ist, dann haben wir wieder eine bedingte Wahrscheinlichkeit, die mit der Bayes-Formel berechnet werden kann.

P_{positv}(nicht infiziert) = 0,0396 : 0,0494 = 0,8016   oder etwa 80 %

Das bedeutet, dass ein als positiv Getesteter mit der großen Wahrscheinlichkeit von 80 % gar nicht infiziert ist.
Andererseits ist ein positiv Getesteter nur mit 20 % iger Sicherheit tatsächlich infiziert.
Jemand der allerdings negativ getestet wurde kann sich zu 99,98 % sicher sein, dass er tatsächlich nicht infiziert ist.

Es ist also so, dass lediglich negative Testergebnisse sehr verlässlich sind. Positive Testergebnisse müssen immer durch mindestens einen zweiten Test abgesichert werden.

In der Praxis werden bei einem PCR-Test aber mehrere Gen-Orte getestet, sodass die Fehlerwahrscheinlichkeit von 2 % deutlich gesenkt werden kann.
Dabei werden die errechneten a posteriori-Wahrscheinlichkeiten zu den neuen a priori-Wahrscheinlichkeiten und eine analoge Rechnung wird durchgeführt.

	M1 ja	M1 nein	∑
M2 ja	0,1943	0,0040	0,1983
M2 nein	0,0321	0,7696	0,8016
∑	0,2264	0,7735	1,0000

0,0321 / 0,2264 = 0,1418 oder 14,18 %
0.1943 / 0,2264 = 0, 8582 oder 85,82 %

Diese ergibt bei einem zweistufigen Test eine Wahrscheinlichkeit von 14,18 %, dass der positiv getestete nicht infiziert ist und eine Wahrscheinlichkeit von 85,82 %, dass das positive Testergebnis stimmt.
Schnelltests, wie sie oft bei Massentestungen (Flughafen) eingesetzt werden, haben in der Regel eine geringere Sicherheit, so dass die Fehlerrate beim ersten Test noch höher ist. Allerdings behaupten einige Hersteller, dass ein Schnelltest eine Spezifität von 99,68 % und eine Sensitivität von 96,52 % erreichen kann.
In diesem Fall werden so gut wie alle Infizierten erkannt. Allerdings ist immer noch die Wahrscheinlichkeit sehr hoch, dass jemand irrtümlich positiv getestet wird.

Fazit: Sichere Ergebnisse liefern nur mehrstufige Tests. Wer positiv getestet ist, muss sich auf jeden Fall noch ein 2. Mal testen lassen, um sicher sein zu können. Je geringer der Anteil der Infizierten ist, desto höher ist die Wahrscheinlichkeit für ein falsches positives Testergebnis.

Geht es auch (fast) ohne Mathematik?

Zugegeben, das Ganze war recht kompliziert und umständlich. Stochastik ist aber manchmal so. Viele Dinge muss man erst mehrfach durcharbeiten, um sie richtig zu verstehen. Allzu leicht formuliert man Bedingungen oder Wahrscheinlichkeiten nicht exakt und produziert damit falsche Ergebnisse. Im Fall Gnabry kommt man allerdings auch mit dem sog. gesunden Menschenverstand zum Ziel.
Wenn 1 % der Untersuchten infiziert ist, sind dass 100 von 10 000 Personen.
Davon werden durch den Test 98 (98 %) als positiv erkannt. Von den 9900 nicht infizierten werden 9504 (96 %) als solche erkannt. Also werden die restlichen 396 falsch positiv getestet. Damit haben wir 495 Personen, die positiv getestet sind, von denen aber nur 99 tatsächlich infiziert sind. D.h. 80 Prozent der positiv getesteten sind falsch positiv getestet.

Testen von Hypothesen – Sigma-Regeln und Binomialverteilung

Der Graph einer Binomialverteilung hat eine **Glockenform.** Mit wachsendem n wird der Graph breiter und flacher. Der höchste Punkt des Graphen ist der Erwartungswert.

Erwartungswert: $\mu = n \cdot p$

Standardabweichung: $\sigma = \sqrt{n \cdot p \cdot (1-p)}$

Für das Beispiel oben ist der Erwartungswert 5 * 0,25 = 1,25 und die Standardabweichung 0,97.

Die Standardabweichung ist der Abstand des Wendepunktes der Glockenkurve zum Maximum.

Sigma-Regeln

Die Bedeutung der Standardabweichung ergibt sich dadurch, dass bei der Binomialverteilung im Intervall $[\mu-\ \sigma;\ \mu+\ \sigma]$ immer 68 % aller Treffer liegen. Im Intervall vom Dreifachen der Standardabweichung um den Mittelwert liegen 99,7 % aller Treffer.
Tabelle für Sigma-Regeln

Intervall	Trefferwahr-scheinlichkeit	Intervall	Trefferwahr-scheinlichkeit
$[\mu-\ \sigma;\ \mu+\ \sigma]$	68,3 %	$[\mu-\ 1{,}64\sigma;\ \mu+\ 1{,}64\sigma]$	90 %
$[\mu-\ 2\sigma;\ \mu+\ 2\sigma]$	95,4 %	$[\mu-\ 1{,}96\sigma;\ \mu+\ 1{,}96\sigma]$	95 %
$[\mu-\ 3\sigma;\ \mu+\ 3\sigma]$	99,7 %	$[\mu-\ 2{,}58\sigma;\ \mu+\ 2{,}58\sigma]$	99 %

Man kann auch sagen, im 1-Sigma-Bereich liegen 68,3 % aller Fälle usw.

Signifikanztest – Testen von Hypothesen bei binomialverteilten Zufallsgrößen

Wenn man bei Bernoulli-Versuchen die Trefferwahrscheinlichkeit p nicht kennt, aber eine Vermutung (**Hypothese)** dazu hat, kann man diese mithilfe einer **Stichprobe** testen.

Sehr häufig verwendet man dafür das $1{,}96\sigma$ Intervall als ausreichende Näherung. Das heißt, wenn in der Stichprobe die Trefferzahl innerhalb des Intervalls liegt, kann die Hypothese angenommen werden. Die Irrtumswahrscheinlichkeit liegt dann nur bei 5 %.

Beispiel:
Eine Losbude wirbt damit, dass jedes 4. Los gewinnt. Eine Person kauft 50 Lose, von denen 8 Gewinnlose sind. Kann die Angabe des Losverkäufers stimmen, wenn man ein Signifikanzniveau von 5 % zugrunde legt?

Wir wollen die Hypothese testen, dass die Trefferwahrscheinlichkeit für ein Gewinnlos bei 25 % liegt(**Nullhypothese**). Dazu legen wir als Stichprobengröße den Wert von 50 gekauften Losen fest. Wenn die Stichprobenergebnisse innerhalb eines bestimmten Bereiches liegen, kann die Hypothese mit einer Irrtumswahrscheinlichkeit angenommen werden. Andernfalls wird sie verworfen. Die Grenzen des

Annahmebereichs können mit der Sigma-Regel bestimmt werden oder werden aus der Tabelle der kumulierten Werte der Binomialverteilung abgelesen.

Lösung:
Ein Signifikanzniveau von 5 % bedeutet, dass die maximale Irrtumswahrscheinlichkeit bei 5 % liegen soll. D. h. 2,5 % können unterhalbe der Intervallgrenze liegen und 2,5 % darüber. Deshalb verwenden wir die Zelle mit der Trefferwahrscheinlichkeit von 95 % in der Tabelle.

Es ist μ = 12,5 (50 x 0,25) und σ= 3,06

Für das 1,96 Intervall ergibt sich [6,502; 18,502] bzw. [7; 19].
Damit kann man die Hypothese bei 8 Gewinnlosen annehmen, da der Wert innerhalb des Annahmebereichs liegt.

Alternativ kann man die Werte aus der **Tabelle der kumulierten Werte der Binomialverteilung** ablesen für **n = 50 und p = 0,25.**

k	P
6	0,0193
7	0,0452
18	0,9712
19	0,986

Daraus ist ersichtlich, dass die 2,5 % Grenze der Treffer unterhalb von 7 und oberhalb von 18 liegt.

Einseitiger Signifikanztest

Man unterscheidet zwischen einem **linksseitigen Signifikanztest: $p < p_0$** und einem **rechtsseitigen Signifikanztest, wenn $p > p_0$** ist.

Beispiel: Bisher befürworten 70 % der Bevölkerung die Sommerzeit. Gegner behaupten aber, dass sich der Anteil verändert hat und der Anteil der Befürworter deutlich abgenommen hat. Das soll nun mit einem Stichprobenumfang von 100 und einem Signifikanzniveau von 5 % untersucht werden .

D. h. die Nullhypothese H_0 liegt bei $p \geq 0{,}7$, dass die Sommerzeit befürwortet wird. Die Hypothese wird verworfen, wenn die Anzahl der Befürworter unterhalb des Annahmebereiches liegt.
Da hier lediglich untersucht werden soll, ob sich die Wahrscheinlichkeit signifikant nach unten verändert hat, liegt hier ein **linksseitiger Signifikanztest** vor.
Die Hypothese H_1 besagt, dass $p < 0{,}7$ ist, weil H_0 verworfen wird, wenn es weniger als 70 % Befürworter gibt.
Da die meisten Tabellen nur bis p = 0,5 ausgewiesen sind, suchen wir nach der Gegenwahrscheinlichkeit, dass mehr als 30 % die Sommerzeit mit einer Wahrscheinlichkeit von mindestens 95 % ablehnen (Tabelle der summierten binomialen Wahrscheinlichkeiten: n = 100, p = 0,3, siehe Anhang) . Damit formulieren wir die Aufgabe zu einem **rechtseitigen Signifikanztest** um.

Wenn mehr als 37 Personen die Sommerzeit ablehnen, dann beträgt die Wahrscheinlichkeit, dass die Hypothese „70 % befürworten die Sommerzeit“ falsch ist mehr als 95 %. Wenn es weniger sind, sinkt die Signifikanz.
Die Nullhypothese wird also verworfen, wenn die Stichprobe ergibt, dass weniger als 63 Personen die Sommerzeit befürworten (100-37).

Man kann dieses Ergebnis sowohl aus dem oberen Teil der Tabelle ablesen, wenn p >0,05 ist, als auch aus dem unteren Teil, wenn k mit einer Wahrscheinlichkeit von >0,95 größer als 37 ist.

k	P		k	P
3	0,0288		36	0,9201
22	0,0479		37	0,9470
23	0,0755		38	0,9660

Eine Hypothese, von der man annimmt, sie durch Beobachtung ablehnen zu können, heißt **Nullhypothese** H_0.

Wenn die Nullhypothese verworfen wird, obwohl sie richtig ist, spricht man von einem **Fehler 1. Art**.
Wenn die Nullhypothese akzeptiert wird, obwohl sie falsch ist, spricht man von einem **Fehler 2. Art**.
Je größer die Stichprobe, desto geringer ist die Wahrscheinlichkeit für einen Fehler.

Jemand behauptet, ein Würfel sei gezinkt. Die Wahrscheinlichkeit eine 6 zu würfeln sei geringer als 1/6. Als Gegenbeweis soll der Würfel 10-mal geworfen werden. Wenn dabei keine oder nur eine 6 fällt, dann sei die Behauptung war.

Die Nullhypothese besagt also, dass p = 1/6 sei. Die Addition der Wahrscheinlichkeiten für keine und eine Sechs bei 10 Versuchen ergibt 0,4845. D.h. dass man mit ca. 48 % Wahrscheinlichkeit damit rechnen muss, dass keine oder nur eine Sechs gewürfelt wird, auch wenn der Würfel nicht gezinkt ist. Das ist kein aussagefähiger Test.

Die Stichprobe wird erweitert auf 100 Versuche. Die Hypothese wird abgelehnt (Würfel ist gezinkt), wenn sich zeigt, dass weniger als 10 Sechsen gewürfelt werden.

Die Tabelle ergibt den Wert 0,0213 als kumulierten Wert für 0 bis 9 Sechsen. Damit kann man mit einer signifikanten Irrtumswahrscheinlichkeit von 2,13 % schon eher den Verdacht haben, das mit dem Würfel etwas nicht stimmt.

Umgekehrt kann man die Irrtumswahrscheinlichkeit auch im Voraus angeben und dann berechnen oder nachsehen, wo die Intervallgrenzen für den Annahme- oder Ablehnungsbereich liegen.

Beispiel:
Ein neuer Impfstoff wird mit einer Wirksamkeit von 95 % beworben.
Ein unabhängiger Test mit 10 000 Personen soll die Behauptung mit einer Sicherheit von 99 % überprüfen.
Bei wie vielen Probanden muss der Impfstoff mindestens wirksam sein, wenn die Behauptung mit der angegebenen Sicherheit gelten soll?

Es handelt sich hier um einen **linksseitigen** Hypothesentest, da die Hypothese (Wirksamkeit mindestens 95 %) nur abgelehnt wird, wenn das Ergebnis der Stichprobe signifikant kleiner als 95 % ist.
Wenn jemand behauptet, die Aussage der Wirksamkeit von 95 % sei richtig, dann genügt auch eine geringfügig geringere Anzahl, um diese Behauptung mit einer bestimmten Sicherheit zu bestätigen. Liegt die Anzahl dann allerdings unter diesem Niveau, kann die Aussage nicht aufrechterhalten werden.

Da es kaum eine kumulierte Tabelle der Binomialverteilung gibt, mit eine Stichprobengröße von 10 000, kann man auch wieder mit der Sigma-Regel rechnen, da es für die Signifikanz von 95 % ein Sigma-Intervall gibt.

p = 0,95, n = 10 000, Sigma-Wert = 2,58. Dann ist μ = 10000 * 0,95 = 9500 und σ = 21,79.
9500 – 2 * 2,58 * 21,79 = 9387,5. (2-fach, da es sich um ein einseitiges Intervall handelt). Nur, wenn es weniger als 9388 wirksame Impfungen gibt, kann die Hypothese (95 %) nicht aufrechterhalten werden.

Wenn nur eine Stichprobenauswertung vorliegt anhand derer man die relative Häufigkeit für ein Ereignis ermittelt, dann kann die Wahrscheinlichkeit für das Ereignis nur geschätzt werden. Diese kann dann innerhalb eines Intervalls liegen, welches als **Vertrauensintervall** bezeichnet wird. Dieses kann mithilfe einer Formel berechnet werden (Lambacher S. 382).
Das Vertrauensintervall zu einer relativen Häufigkeit h enthält alle Wahrscheinlichkeiten p, die mit der beobachteten relativen Häufigkeit h mit einer vorgegebenen Sicherheitswahrscheinlichkeit (Vertrauensniveau) β vereinbar sind.
Die Herleitung für die Berechnung des Vertrauensintervalls ist nicht gerade einfach. Die beiden Intervallgrenzen ergeben sich aus den Lösungen einer quadratischen Gleichung bei der Erwartungswert, Standardabweichung und der jeweilige Sigmaintervallwert verwendet werden.

Die Grenzen des Vertrauensintervalls bestimmt man durch Auflösen der Gleichung
h = $p \pm k\sqrt{\frac{p(1-p)}{n}}$ nach p, wobei k der **Sigmaintervallwert** für das gewünschte Vertrauensniveau ist. Die relative Häufigkeit h ergibt sich aus der Trefferzahl der Stichprobe dividiert durch die Anzahl der Stichprobenelemente (r/n).

Beispiel: Bei einer Stichprobe n = 100 und einer Trefferzahl von r = 61 liegen die Ergebnisse mit 99 % iger Wahrscheinlichkeit im Bereich von

0,61 = $p \pm 2{,}58\sqrt{\frac{p(1-p)}{n}}$ ergibt aufgelöst nach p ergibt ein Intervall [0,481;0,725].
Das bedeutet, dass alle Wahrscheinlichkeiten zwischen 48,1 % und 72,5 % zu 99 % mit der Stichprobenwahrscheinlichkeit von 61 % vereinbar sind. Das Vertrauensintervall liegt demnach zwischen 48,1 % und 72,5 %.

Bei einer Stichprobe von 1000 liegt das **Vertrauensintervall** zwischen 57 % und 64,9 %.
Je größer die Stichprobe, desto kleiner ist das Vertrauensintervall.

Etwas einfacher ist die Berechnung mit der **Näherungsformel:**

Linke Grenze $= h - k\sqrt{\frac{h(1-h)}{n}}$ Rechte Grenze$= h + k\sqrt{\frac{h(1-h)}{n}}$

Wobei h die relative Häufigkeit p/n ist.

Beispiel:
Ein neuer Impfstoff wird getestet. Zunächst wird er an 1000 mit einem Virus infizierten Personen getestet. Davon genesen aufgrund der Impfung 810 Personen. Das Vertrauensintervall soll eine Zuverlässigkeit von 99 % haben.
a) Bestimme das Vertrauensintervall bei n = 1000.
b) Bestimme das Vertrauensintervall, wenn in einer zweiten Studie 10 000 Personen getestet werden und mit 8 100 Genesenen die Quote genauso hoch ist wie beim ersten Mal.

n = 1000, h = 0,81, k = 2,58
Zu a) Mit der Näherungsformel und dem Sigma Wert von 2,58 erhalten wir für a) ein Vertrauensintervall von 77,8 % bis 84,2 %.
Die genaue Rechnung mit der Auflösung nach p ergibt einen Bereich von 77,6 bis 83,99 %.
Zu b) ergibt sich ein Intervall von 80,0% bis 82,0 % mit der Näherungsformel.
Interpretation: Bei Anwendung der 1. Studie kann man zu 99% darauf vertrauen, dass die Zuverlässigkeit zwischen 77,8 % und 84,2 % liegt.
Bei Anwendung der 2. Studie kann man zu 99 % darauf vertrauen, dass die Zuverlässigkeit zwischen 80 % und 82 % liegt.

Zusammenfassung:

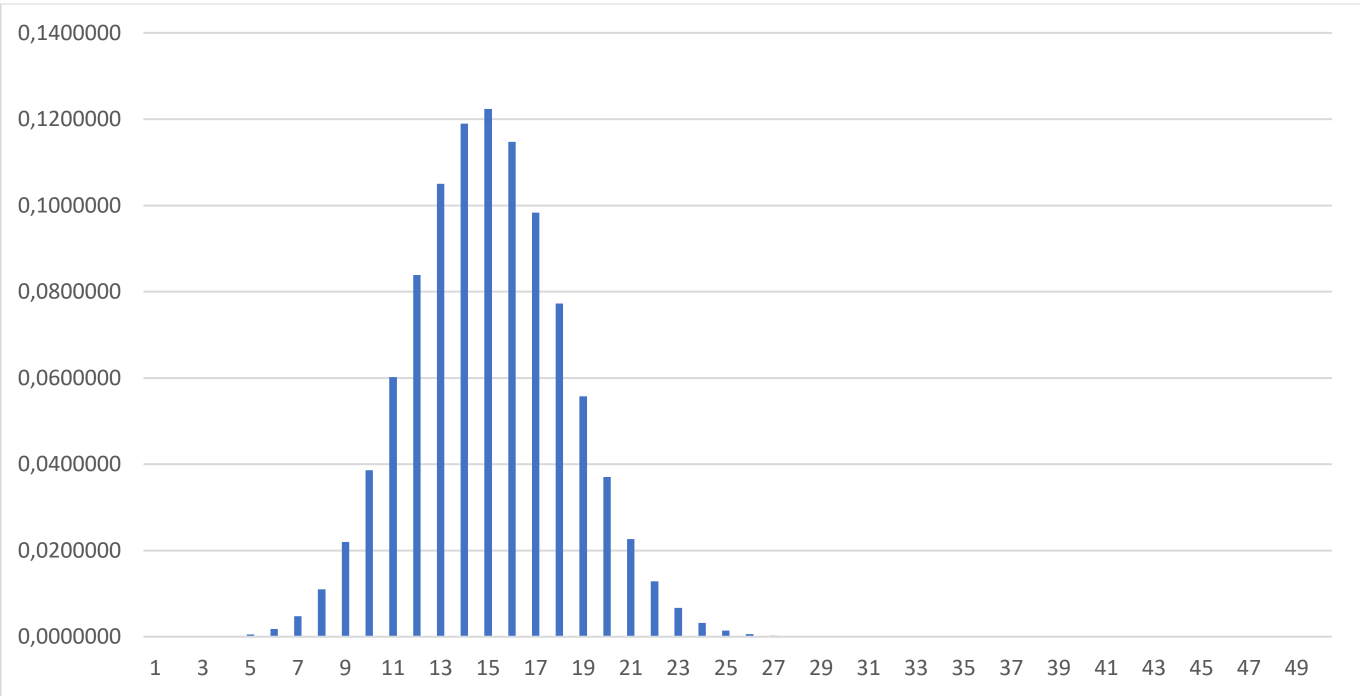
Hypothesentests werden durchgeführt, wenn man mit den Ergebnissen einer Stichprobe eine Behauptung überprüfen möchte.

Nullhypothese H_0: mindestens 20 % aller Kinder im Alter von 10 Jahren sind Nichtschwimmer.
Alternativhypothese H_1: weniger als 20% aller Kinder im Alter von 10 Jahren sind Nichtschwimmer.

Eine Stichprobe von n = 100 ergibt, dass 17 von 100 Kindern nicht schwimmen können. Kann die Behauptung, dass mindestens 20 % aller 10-jährigen nicht schwimmen können aufrecht erhalten werden (linksseitiger Hypothesentest)?
Um diese Frage zu beantworten, muss man festlegen, mit welcher Wahrscheinlichkeit die Antwort auf die Behauptung sicher sein soll. Dies wird als **Signifikanzniveau** bezeichnet. Ein Signifikanzniveau von 5% bedeutet, dass mit einer Wahrscheinlichkeit von 5 % die Nullhypothese fälschlicherweise abgelehnt wird. Umgekehrt ergibt sich, dass die Nullhypothese mit einer Wahrscheinlichkeit von 95 % aufrecht erhalten werden kann, wenn sich ergibt, dass mehr als 24 Kinder nicht schwimmen können.
Je größer die Stichprobe, desto sicherer ist die Aussage. D. h. bei 17 von 100 kann die Abweichung von 20% noch eher zufällig sein, als wenn bei einer Stichprobe von 1000 170 Kinder nicht schwimmen können.
Die Tabelle der kumulierten Binomialverteilung ergibt bei n = 100 und p = 0,2 bei einem k von 13 den Wert

0,0469 und bei k = 14 den Wert 0,0804. Das bedeutet, dass ein Wert von 13 oder weniger bei einem Signifikanzniveau von 4,69 % zu einer Ablehnung der Hypothese führt, da das Signifikanzniveau unterschritten ist (kleiner gleich 13 %). Bei Ergebnissen von 14 und mehr Nichtschwimmern, kann die Nullhypothese noch aufrecht erhalten werden (Trefferwahrscheinlichkeit über 95 %). Würde man die Stichprobe auf 200 erhöhen, so führen Ergebnisse von 30 und weniger zu einer Ablehnung (kleiner gleich 15 %).
Bei einem Signifikanzniveau von 1 %, könnte die Nullhypothese noch bei mehr als 10 festgestellten Nichtschwimmern aufrecht erhalten werden, da die Trefferwahrscheinlichkeit, dass die Nullhypothese aufrecht erhalten werden kann 99 % beträgt.
Bei großen Stichproben kann der Annahmebereich mit Hilfe der Sigma-Regeln bestimmt werden, sofern die Standardabweichung größer als 3 ist.

Graphische Darstellung:
Im Schaubild wird die Verteilung der Ergebnisse einer Stichprobe mit n = 50 und p = 0,3 dargestellt. Der Erwartungswert liegt bei 15 (50 x 0,3). 50 % aller Ergebnisse liegen im Bereich 0 – 15, 50 % liegen im Bereich 15 -50. Ergibt bei einer Stichprobe ein Ergebnis von k = 11, so ergibt sich aus den kumulierten Werten, dass 13,9 % der Ergebnisse im Bereich von 0 bis 11 liegen.
Im Bereich von 13 bis 17 liegen 55,93 % aller Ergebnisse.
Im Bereich von mehr als 20 liegen noch 4,78 % aller Ergebnisse.



Folgende Hypothesen können beispielsweise getestet werden:

1. Linksseitiger Test: Es wird behauptet, dass mindestens 30 % aller Jungen gerne Fußball spielen. Eine Stichprobe von 50 Personen soll das mit einem Signifikanzniveau von 5 % bestätigen. Die Stichprobe ergibt, dass 12 Jungen gerne Fußball spielen. Kann die Behauptung aufrecht erhalten werden?
   In der Tabelle findet man bei k = 9 den Wert 0,0402 und bei k = 10 den Wert 0,0789. D.h., dass etwas 4 % aller Stichproben einen Wert von weniger als 10 ergeben. Die Nullhypothese kann demnach bei diesem Ergebnis noch aufrecht erhalten werden. Das Stichprobe ist hier noch zu klein, um ein wirklich aussagefähiges Ergebnis liefern zu können.
2. Rechtseitiger Test: Es wird behauptet. Dass höchstens 30 % aller 10-jährigen Schwimmen können. Ein Test von 100 10-jährigen ergibt, dass 41 Schwimmen können. Das Ergebnis soll ein Signifikanz-niveau von 1 % haben. Es ergibt sich aus der Tabelle bei 40 der Wert 0,9875 und bei 41 der Wert 0,9928. Damit ist das Ergebnis von 41 knapp außerhalb der 99 % Grenze, sodass die Hypothese nicht aufrecht erhalten werden kann.
3. Beidseitiger Test: Ein Glücksrad soll in genau 10 % aller Versuche die Zahl 7 anzeigen. In welchem Bereich darf die Anzahl der Treffer bei einem Signifikanzniveau von 5 % sein, wenn a) 100 Versuche gemacht werden, b) bei 1000 Versuchen?
   Ein Signifikanzniveau von 5 % bei einem beidseitigen Test bedeutet, dass nach oben und nach unten jeweils eine Abweichung von maximal 2,5 % zulässig ist. Aus der Tabelle ergibt sich, dass dieser Bereich zwischen 5 und 15 Treffern liegt.
   Bei 1000 Versuchen wendet man die Sigma-Regel an, wenn die Standardabweichung einen Wert von größer 3 ergibt. Diese liegt in diesem Fall bei ca. 9,5. Dann gilt die Regel: (100 – 1,96x9,5; 100 + 1,96x9,5). Das ergibt, dass ein Intervall von 82 bis 118 noch innerhalb der Trefferwahrscheinlichkeit von 95 % liegt.

Dichtefunktion

Eine Funktion f, aus der man die Wahrscheinlichkeiten durch Integration bestimmen kann, heißt Dichtefunktion oder Wahrscheinlichkeitsdichte.

Was bedeutet das?
Bei einem Zufallsversuch weichen die Ergebnisse vom Sollmaß ab. Dieses Sollmaß kann auch ein Mittelwert oder ein erwarteter Wert sein.

Beispiele:
1. Bei einer Produktion soll ein Teil von exakt 1,72 mm hergestellt werden. Tatsächlich weichen die Teile aber immer wieder geringfügig von diesem Wert ab.
2. Die Zufallszahlen, die der TR erzeugt, liegen zwischen 0 und 1. Sie streuen um den Mittelwert von 0,5.
3. Es regnet gleichmäßig auf einen kreisrunden Tisch mit dem Radius r. Wie groß ist die Wahrscheinlichkeit dafür, dass ein Tropfen einen bestimmten Abstand vom Mittelpunkt hat.

Die Funktionswerte f(x) sind keine Wahrscheinlichkeiten, da man die Wahrscheinlichkeit an einem Punkt nicht messen kann (P = 0). Es muss immer ein Intervall angegeben werden (P(r <=x<=s))um die Dichte (Wahrscheinlichkeit) berechnen zu können. Die Fläche unter der Kurve f bzw. das Flächenstück unter der Kurve f im Intervall [r; s] gibt die Wahrscheinlichkeit an.
Damit ist klar, welche **Eigenschaften** die Dichtefunktion haben muss:

$$f(x) \geq 0 \qquad \text{und} \qquad \int_r^s f(x)dx = 1$$

Oft geht es darum, diese Dichtefunktion zu bestimmen. Dazu muss allerdings die Grundfunktion bekannt sein. Es ist dann der Parameter **k** gesucht, der die Steigung oder die Form des Graphen von f beschreibt.
Allgemein: Dichtefunktion: k· f(x)

Beispiel 1:
Regentropfen fallen gleichmäßig auf einen runden Tisch mit einem Radius von 50 cm. Je weiter man sich vom Mittelpunkt zum Rand entfernt, desto häufiger ist die Anzahl der Regentropfen. Stellt man diese relativen Häufigkeiten in Abhängigkeit vom Radius in einem Achsenkreuz dar, so erhält man eine Gerade durch den Ursprung. Man kann nun folgenden Aufgaben stellen.
a) Wie ist der Faktor k zu wählen, damit es sich um die Dichtefunktion handelt?
b) Wie groß ist die Wahrscheinlichkeit, dass ein Regentropfen in einem Kreisring von 10 bis 15 cm um den Mittelpunkt fällt?
c) Welche Entfernung eines Regentropfens vom Mittelpunkt kann man im Durchschnitt erwarten? D. h. wo liegt die größte Wahrscheinlichkeit. Wir suchen den Erwartungswert.

<u>Lösung:</u>
zu a) k ist so zu bestimmen, dass der gesamte Flächeninhalt unter der Kurve gleich 1 ist.

$\int_0^{50} k \cdot x\,dx = 1$ -> $[0{,}5k \cdot x^2]_0^{50}$=1 -> (1250k) = 1 -> k = 0,0008

Die Dichtefunktion heißt: f(x) = 0,0008x

Die Probe ergibt: $\int_0^{50} 0{,}0008\,x\,dx = 1$

zu b)

$\int_{10}^{15} 0{,}0008\,x\,dx =$ $\lfloor 0{,}0004\,x^2 \rfloor_{10}^{15} =$ 0,09 – 0,04 = 0,05 oder 5 %.

zu c) der mittlere Abstand eines Regentropfens vom Zentrum wird durch den **Erwartungswert** dargestellt.
Zur Berechnung muss die Dichtefunktion noch mit x multipliziert werden. Warum? Wenn man x Zufallsversuche durchführt, deren Ergebnisse mit einer Wahrscheinlichkeitsverteilung auftreten, dann erhält man den Mittelwert durch die Summe der Wahrscheinlichkeiten für ein einzelnes Ereignis. Beispiel: Bei einem Würfel mit drei Zahlen (1,2,3) wird die 1 mit p=0,2 geworfen, die 2 mit p=0,5 und die 3 mit p=03. Bildet man die Summe 1*0,2+2*0,5+3*03 erhält man 2,1. Das ist für diesen Zufallsversuch der Mittelwert. Für die Dichtefunktion gilt:

Erwartungswert der Dichtefunktion: $\mu = \int_r^s x \cdot f(x)dx$

hier: $\int_0^{50} x \cdot 0{,}0008x\,dx$ = $\left[\frac{1}{3750}x^3\right]_0^{50}$ = 33,33 Im Durchschnitt fallen die Tropfen 33,33 cm vom Zentrum entfernt.

Ebenso wie der Erwartungswert ist auch die **Standardabweichung** eine oft verwendete Kenngröße. Sie berechnet, wie sehr die Werte der Zufallsgröße x um den Erwartungswert schwanken.

Standardabweichung der Dichtefunktion: $\sigma = \sqrt{\int_r^s (x-\mu)^2 \cdot f(x)dx}$

Beide Werte sind theoretische Größen und sagen lediglich voraus, wie der empirische Mittelwert und die empirische Abweichung sich bei sehr großen Versuchsreihen entwickeln würden.

Beispiel:
Bei einem Zufallsversuch ergibt sich eine Wahrscheinlichkeitsdichte von $f(x) = 3(x-1)^2$ im Intervall [0;5].
a) Suchen Sie die obere Grenze des Intervalls, wenn f eine Dichtefunktion ist.
b) Berechnen Sie die Wahrscheinlichkeit für x <= 0,5.
c) Berechnen Sie Erwartungswert und Standardabweichung.

zu a)

$\int_0^s 3(x-1)^2\,dx = 1$ -> $\int_0^s 3x^2 - 6x + 3 = 1$ -> $\lfloor x^3 - 3x^2 + 3x \rfloor_0^s$

$S^3 - 3s^2 + 3s = 1$ Gl. 3. Grades mit dem TR lösen: s = 1 (rechte Grenze liegt bei 1)

zu b)

$\int_0^{0,5} 3x^2 - 6x + 3\,dx$ = $\lfloor x^3 - 3x^2 + 3x \rfloor_0^{0,5}$ = 0,125 – 0,75 + 1,5 = 0,875 oder 87,5 %

zu c)

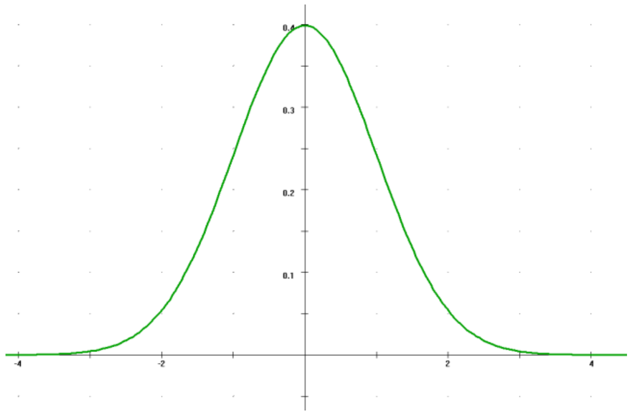
$\mu = \int_0^1 x(3x^2 - 6x + 3)\,dx$ = $\int_0^1 3x^3 \;\; - 6x^2 + 3x\,dx$ = = $\lfloor 0{,}75x^4 - 2x^3 + 1{,}5x^2 \rfloor_0^1$ =

0,75 – 2 + 1,5 = 0,25

$\sigma = \sqrt{\int_0^1 (x - 0{,}25)^2 \cdot (3x^2 - 6x + 3)dx}$ Das Ausrechnen überlassen wir einem Computersystem

es ergibt sich $\sigma = 0{,}2$

Die Gauß'sche Glockenfunktion als Dichtefunktion (Normalverteilung)



Carl Friedrich Gauß hat herausgefunden, dass viele zufällige Vorgänge ähnliche Abweichungen vom Mittelwert haben, die sich durch den Verlauf einer Glockenkurve gut darstellen lassen. Abweichungen in der engen Umgebung des Mittelwerts sind häufiger, größere Abweichungen werden immer seltener je größer die Abweichung ist. Wenn der Mittelwert bei 0 liegt und die Standardabweichung bei 1 liegt, dann ergibt sich obiges Bild der Glockenkurve. Bei anderen Mittelwerten ist sie nach links oder rechts verschoben, bei anderen Standardabweichungen verläuft sie flacher oder steiler.
Genau wie bei der Sigma-Regel kann man Intervalle bestimmen, in denen ein gewisser Anteil aller Messwerte zu finden sind. So sind z.B. im Intervall +/- σ 68,27 % aller Messwerte zu finden, im Intervall +/- 2σ 95,45 % aller Messwerte usw. (die Tabelle der Sigma-Werte gilt auch hier).
Die Formel der Dichtefunktion nach Gauß lautet:

f(x)= $\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)2}{2\sigma^2}}$ für $\mu = 0 \; und \; \sigma = 1$ gilt dann f(x) = $\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$

Um zu beweisen, dass die Funktion tatsächlich eine Dichtefunktion ist, muss das Integral im Intervall [-∞; +-∞] den Wert 1 haben. Näherungsweise kann man aber auch das Intervall im Bereich [-10; + 10] bestimmen, da dieser Wert schon sehr nahe an 1 sein muss.
Mithilfe des TR ermittelt man tatsächlich ein Ergebnis, welches mit 0,999… nahezu den Wert 1 hat.

Die Wahrscheinlichkeiten, die sich aus der Integration der Gauß-Funktion ergeben sind zum Glück in Tabellen festgehalten. Zu erwähnen ist noch, dass man **Hochpunk**t und **Wendepunkt** der Funktion bestimmen kann. Der Hochpunkt ergibt sich aus dem Erwartungswert μ, die Wendepunkte ergeben sich aus der Addition der Standardabweichung W($\mu \pm \sigma$).

Standardisierung: Aus der Tabelle entnimmt man Werte, die für eine Normalverteilung mit dem Erwartungswert 0 und der Standardabweichung 1 gelten. Bei anderen Erwartungswerten und Standardabweichungen (Kurve ist verschoben, flacher oder steiler) müssen die Grenzen **standardisiert** werden. Dies geschieht mit der Umrechnungsformel $\frac{r,s-\mu}{\sigma}$, wobei r und s die untere bzw. obere Grenze ist.

Binomialverteilungen und die **Gauß'sche Verteilungsfunktion** verlaufen ähnlich und können oft gegenseitig ersetzt werden. Oftmals lassen sich die Werte für die Gauß-Funktion besser aus der den Tabellen ablesen, sodass viele binomialverteilte Aufgaben mit der Normalverteilung gelöst werden.

Rechnen mit der Normalverteilung

Beispiel :
Berechnen Sie die Integrale: a) $\int_{-0,2}^{0,5} \varphi_{0,1}(x)dx$ b) $\int_{0,5}^{\infty} \varphi_{1,2}(x)dx$

Das Zeichen φ (phi) ist das Symbol für die Glockenfunktion. Die Indices bezeichnen die Werte für den Erwartungswert und die Standardabweichung.
Im Fall a) handelt es sich um die Standardnormalverteilung, weil $\mu = 0$ und $\sigma = 1$ ist.
Aus der Tabelle entnimmt man die Werte für x = 0,5 und den Wert für x = -0,2 mit 0,6915 und 0,4207. Die Differenz ergibt 0,2708.

Zu b) In diesem Fall handelt es sich um eine Normalverteilung, die im Gegensatz zur Standardform verschoben ist und flacher verläuft (breitere Streuung wegen größerer Standardabweichung). In diesem Fall werden die obere bzw. untere Grenze (r oder s) umgerechnet und dann ebenfalls die Tabelle angewendet.

Die Umrechnung erfolgt mit der Formel $\frac{r,s-\mu}{\sigma}$. Hier: r = $\frac{0,5-1}{2} = -0,25$,
damit 1 – 0,4013 = 0,5987.

Für stetige Werte erfolgt eine **Stetigkeitskorrektur**. D.h. dass man von dem unteren Wert 0,5 abzieht und beim oberen Wert 0,5 hinzuzählt.

<u>Dazu zwei Textaufgaben:</u>

1) Auf eine Fähre passen 250 Fahrzeuge. Im Durchschnitt befindet sich in 30 % aller Fälle nur eine Person. Wie groß ist die Wahrscheinlichkeit, dass sich in weniger als 80 Fahrzeugen nur der Fahrer befindet?

Im Prinzip ist diese Aufgabe durch eine Binomialverteilung lösbar. Allerdings können wir sie auch durch Annäherung mit einer Normalverteilung lösen.
Zunächst muss man die linke und die rechte Grenze des Integrals bestimmen. Die linke Grenze kann man ignorieren, da sie bei -∞ liegt. Bei der rechten Grenze muss man die Stetigkeitskorrektur anwenden. Da es weniger als 80 Fahrzeuge sind, liegt der Wert bei 79 also 79,5.
Dann muss dieser Wert auf die Standardnormalverteilung umgerechnet werden. Dazu benötigt man den Erwartungswert und die Standardabweichung.

$\mu = n \cdot p$ und $\sigma = \sqrt{n \cdot p(1-p)}$, also $\mu = 75$ und $\sigma = 7{,}25$. Dann gilt

$\int_{-\infty}^{79,5} \varphi_{75;7,25}(x)dx$. Die Umrechnung der oberen Grenze ergibt $\frac{79,5-75}{7,25} = 0{,}621$.
Das ergibt: $\Phi(0{,}621) - \Phi(-\infty) = 0{,}7324 - 0 = 0{,}73 = 73\ \%$

2) Der Benzinverbrauch eines Autos ist normalverteilt mit $\mu = 8{,}5$ und $\sigma = 0{,}8$.
a) Berechnen Sie die Wahrscheinlichkeit dafür, dass der Verbrauch über 10 Liter liegt.
b) In welchem Intervall mit dem Mittelpunkt μ liegt der Verbrauch mit der Wahrscheinlichkeit von 80 %?

Zu a) $\int_{10}^{\infty} \varphi_{8,5,0,8}(x)dx$ Umrechnung r = (10-8,5)/0,8 = 1,875
Der Tabellenwert für x = 1,875 ergibt (1-0,9693) = 0,0307 oder 3,07 %

zu b) Aus der Tabelle den Wert 0,9 ablesen. Dazu gehört der x-Wert 1,28. D.h. die 80 % ige Wahrscheinlichkeit bewegt sich bei der Normalverteilung im Intervall [-1,28; 1,28]. Dieser Wert muss allerdings wieder mit obiger Formel umgerechnet werden. = $\frac{r-8,5}{0,8} = 1{,}28$. Das ergibt für r den Wert 9,524. Dann ist der untere Wert 8,5-(9,524-8,5) = 7,476. Das Intervall liegt zwischen 7,476 und 9,524.

Binomialverteilte Zufallsgrößen stimmen mit der Normalverteilung gut überein. Wie schon erwähnt, gelten die Sigma-Regeln auch hier. Die benötigten Größen sind n, p, $\mu = n \cdot p$ und $\sigma = \sqrt{n \cdot p(1-p)}$.

Beispiel:
Die Anzahl der Rosinen in einem Rosinenbrötchen kann durch eine Binomialverteilung mit $\mu = 16{,}2$ und σ = 3,5 beschrieben werden. Wie groß ist die Wahrscheinlichkeit, dass ein zufällig gewähltes Brötchen a) genau 15 Rosinen enthält, b) zwischen 14 und 18 Rosinen enthält ?

zu a) Weil die Wahrscheinlichkeitsverteilung stetig ist, wird durch eine **Stetigkeitskorrerktur** auf das Intervall [14,5; 15,5] korrigiert.
$\int_{14,5}^{15,5} \varphi_{16,2;3,5}(x)dx$ = 0,4207 – 0,3156 = 0,105 = 10,5 %

Zu b)
$\int_{13,5}^{18,5} \varphi_{16,2;3,5}(x)dx$ = 0,7454 – 0,2206 = 0,5248 = 52,5 %.

Bestimmung des Stichprobenumfangs

Beispiel 1:
Für die Hochrechnung des Wahlergebnisses für eine Partei soll ermittelt werden, wie groß die Stichprobe n der befragten Personen sein muss, wenn das Ergebnis mit 96 % Sicherheit nur um höchstens 0,5 % vom Wahlergebnis abweichen soll.

Bei dieser Aufgabenstellung werden verschiedene Größen verwendet, die genau unterschieden werden müssen:
n = die gesuchte Stichprobengröße
r = die Anzahl der ‚Treffer' in der Stichprobe

h = die relative Häufigkeit r/n für das durch die Stichprobe ermittelte Wahlergebnis
p = das unbekannte, tatsächliche Wahlergebnis der Partei
α = die Sicherheitswahrscheinlichkeit für das Ergebnis (hier 0,96)
d = die gewünschte Genauigkeit (Abweichung) (hier 0,005)
Mit Φ (phi) bezeichnet man das Integral der Gauß'schen Dichtefunktion

Letztlich müssen wir für die Lösung in der Tabelle nachsehen, für welchen x-Wert die gesuchte Fläche erreicht oder übertroffen wird. Dieser Wert muss dann mit einer Formel umgerechnet werden, um n zu erhalten.

Dafür gibt es die Formel: $2 \cdot \Phi\left(\frac{\sqrt{n} \cdot d}{\sqrt{p(1-p)}}\right) - 1 = \alpha$

Eingesetzt $2 \cdot \Phi\left(\frac{\sqrt{n} \cdot 0{,}005}{\sqrt{p(1-p)}}\right) - 1 \geq 0{,}96$ daraus ergibt sich $\Phi\left(\frac{\sqrt{n} \cdot 0{,}005}{\sqrt{p(1-p)}}\right) \geq 0{,}98$

Die Flächensumme von mindestens 0,98 finden wir bei dem **x-Wert 2,06**.
Wenn p unbekannt ist, nimmt man den maximal möglichen Wert für den Ausdruck p(1-p). Dieser kann maximal den Wert 0,25 annehmen, da p(1-p) = p-p² maximal 0,25 ist, wenn 0<=p<=1.
Dann gilt im ungünstigsten Fall $\frac{\sqrt{n} \cdot 0{,}005}{0{,}5} \geq 2{,}06$ $\quad \sqrt{n} \cdot 0{,}005 \geq 1{,}03$ $\quad \underline{n \geq 42\,436}$.

In Kurzform: Wir suchen in der Tabelle den X-Wert für die gesuchte Genauigkeitswahrscheinlichkeit (nicht von -∞ bis a, sondern von -a bis a. D. h. bei α = 90 % lesen wir den Wert von 0,95 ab).
Dann bilden wir die Gleichung $2\sqrt{n} \cdot d \geq x$, setzen die gewünschte Genauigkeit ein und lösen nach n auf.

Formel für die Größe des Stichprobenumfangs: $n \geq \frac{s^2}{d^2} \cdot p(1-p)$

Dabei ist s entweder die Zahl, die zu einem Sigma-Vertrauensintervall gehört oder der errechnete x-Wert aus der Tabelle der Dichtefunktion. Wenn p nicht bekannt ist, ersetzt man p(1-p) durch den maximalen Wert 0,25.

Herleitung:
1. $Gleichung\ |X - n \cdot p| \leq s \cdot \sigma$ dabei ist X die absolute Anzahl der Treffer, s ist entweder der Sigma-Wert für eine bestimmte Wahrscheinlichkeit (z.B. 1,96 für 95 %) oder der errechnete X-Wert der Dichtefunktion, der in der Tabelle eine bestimmte Wahrscheinlichkeit (Fläche) ergibt.
Umformung: $\left|\frac{x}{n} - p\right| \leq s \cdot \frac{\sigma}{n}$ Gleichung 1 dividiert durch n
2. $Gleichung\ |\frac{x}{n} - p| \leq d$ die Abweichung soll kleiner sein als der Prozentsatz d
Beide Gleichungen kombinieren: s - $\frac{\sigma}{n}$ ≤ d wobei σ die **Standardabweichung** ist. $\boldsymbol{\sigma = \sqrt{n \cdot p(1-p)}}$
s $\cdot \frac{\sqrt{n \cdot p(1-p)}}{n} \leq d$ Quadrieren
s² $\cdot \frac{n \cdot p(1-p)}{n^2} \leq d^2$ Kürzen und mit n multiplizieren
s² · p(1-p) ≤ d² · n p(1-P) ist maximal 0,25

n ≥ $\frac{s^2}{d^2}$ · 0,25

Beispiel 2:
Wenn die gewünschte Genauigkeit einem bekannten Sigma-Wert, z. B. 95 %, entspricht, kann man eine alternative Rechnung durchführen.

|X - n·p| ≤ 1,96 · σ
X ist die absolute Häufigkeit der Treffer in der Stichprobe.
Wenn man obige Gleichung durch n dividiert erhält man: $\left|\frac{X}{n} - p\right| \leq 1{,}96\ \frac{\sigma}{n}$
D.h. die Abweichung der relativen Häufigkeit von dem tatsächlichen Ergebnis soll kleiner als d (gewünschte Abweichung) sein.

$\left|\frac{X}{n} - p\right| \leq d$ dann gilt aber auch $1{,}96\ \frac{\sigma}{n} \leq d$ in der Aufgabe $1{,}96\ \frac{\sigma}{n} \leq 0{,}005$
ersetzt man σ, erhält man $1{,}96\ \frac{\sqrt{n \cdot p(1-p)}}{n}$n $\leq 0{,}005$, dann $1{,}96^2\ \frac{p(1-p)}{n} \leq 0{,}005^2$
Da p(1-p) maximal 0,25 ist, ist <u>n = 38 416</u>. N ist hier etwas kleiner, da die Genauigkeit nur 95 % und nicht 96 % wie oben sein soll.

Weitere Beispiele:
1. Wie oft muss man eine Münze werfen, damit die relative Häufigkeit für „Kopf“ mit 99,5 % Wahrscheinlichkeit zwischen 0,47 und 0,53 liegt?
2. Eine Partei möchte vor der Wahl eine Umfrage über den zu erwartenden Wahlerfolg starten. Das Ergebnis der Umfrage soll mit 99 % Sicherheit um nicht mehr als 1 % vom tatsächlichen Ergebnis abweichen. Wie viele Personen müssen befragt werden?
3. Ein Pharmaunternehmen möchte einen Impfstoff auf den Markt bringen, der eine 90 % Wirksamkeit ausweisen soll.
   a) Wie viele Personen müssen an der Impfstudie teilnehmen, wenn eine 99,7 % Sicherheit bestehen soll, dass die Wirksamkeit um nicht mehr als 2 % nach unten abweicht?
   b) Bei einer Studie mit 5000 Personen ergibt sich eine Wirksamkeit von 88,1 %. Muss die ausgewiesene Wirksamkeit korrigiert werden?
4. Für eine Untersuchung werden zufällig 4900 Personen ausgewählt. Wie groß ist die Wahrscheinlich-keit, dass davon mehr als 2600 davon Frauen sind?
5. Eine Fluggesellschaft rechnet damit, dass 5 % der Passagiere ihren gebuchten Flug nicht antreten. Damit keine Plätze leer bleiben, werden die Maschinen überbucht. Bei einer Maschine mit 260 Plätzen werden 268 Tickets verkauft.
   a) Wie groß ist die Wahrscheinlichkeit, dass die Plätze nicht ausreichen?
   b) Wie viele Tickets dürfen maximal mit einer Sicherheit von 99,5 % verkauft werden, wenn 4 % der Passagiere den Flug nicht antreten?

Zu 1) Die Abweichung beträgt 0,03. P = 0,5. Dann gilt die Formel:

$$2 \cdot \Phi\left(\frac{\sqrt{n}\ \cdot 0{,}03}{\sqrt{0{,}5(1-0{,}5)}}\right) - 1 = 0{,}995 \ \rightarrow\ \Phi\left(\frac{\sqrt{n}\ \cdot 0{,}03}{\sqrt{0{,}5(1-0{,}5)}}\right) \geq 0{,}9975$$

Für 0,9975 gilt der Tabellenwert 2,81. $\left(\frac{\sqrt{n} \cdot 0{,}03}{\sqrt{0{,}5(1-0{,}5)}}\right) = 2{,}81,$ dann <u>n = 2193.</u>

Zu 2) $\left(\frac{\sqrt{n} \cdot 0{,}01}{\sqrt{0{,}5(1-0{,}5)}}\right) \geq 0{,}99$ $\quad \sqrt{n}\ \cdot 0{,}01 = 2{,}33\ \cdot 0{,}5$ $\quad$ <u>n = 13572</u>

Zu 3) Wir wenden die Formel mit dem Vertrauensintervall von 97,7 % an.

$n \geq \frac{3^2}{0{,}02^2} \cdot 0{,}9 \cdot 0{,}1$ -> <u>n = 2025</u>

Zu 4) P(X > 2600) oder P(X ≥ 2600), P(2601 ≤ ∞). Damit ist die untere Grenze 2601.
Der Erwartungswert ist 4900 · 0,5 = 2450, die Standardabweichung ist 35.
Daraus ergibt sich $\frac{2601-2450}{35} = 4{,}28$.

$\Phi(\infty) = 1$ und $\Phi(4{,}28) = 1$ (dieser Wert ist nicht mehr tabelliert). Damit ist P = 1 – 1 = 0.
Die Wahrscheinlichkeit ist also so gering, dass sie rechnerisch den Wert 0 ergibt.

Zu 5a) Größen: n = 268, p = 5 %, das Problem besteht dann, wenn mehr als 260 Fluggäste erscheinen.
P(x ≥ 261) = 1 – P(x ≤ 260) oder $1 - \Phi\left(\frac{k-\mu+0{,}5}{\sigma}\right) = 1 - \Phi\left(\frac{260-268 \cdot 0{,}95+0{,}5}{\sqrt{268 \cdot 0{,}95 \cdot 0{,}05}}\right)$

$1 - \Phi(1{,}65) = 0{,}0495 = 4{,}95\ \%$

Zum Vergleich: Es werden 275 Plätze verkauft. Dann steigt die Wahrscheinlichkeit, dass die Plätze nicht ausreichen auf 58,32 % . $1 - \Phi\left(\frac{260-275 \cdot 0{,}95+0{,}5}{\sqrt{275 \cdot 0{,}95 \cdot 0{,}05}}\right)$ = 1 – 0,4168.

Zu 5b) 1 – P(X ≤ 260) < 0,005 oder P(X ≤ 260) > 0,995 heißt: Die Wahrscheinlichkeit, dass insgesamt weniger als 261 Passagiere kommen, soll größer als 99,5 % sein.
n = ? p = 96% (für Antritt der Reise). P(X ≥ 261) = 1 – P(X ≤ 260) < 0,005
$\Phi\left(\frac{260-n \cdot 0{,}96+0{,}5}{\sqrt{n \cdot 0{,}96 \cdot 0{,}04}}\right) > 0{,}995$ Wir suchen den Tabellenwert für die Dichte 0,995 -> X = 2,58

$\Phi\left(\frac{260-n \cdot 0{,}96+0{,}5}{\sqrt{n \cdot 0{,}96 \cdot 0{,}04}}\right) > 2{,}58$ oder $\left(\frac{260-n \cdot 0{,}96+0{,}5}{\sqrt{n \cdot 0{,}96 \cdot 0{,}04}}\right) > 2{,}58$ 260 - n·0,96 + 0,5 > 2,58 $\cdot\sqrt{n \cdot 0{,}96 \cdot 0{,}04}$

260,5- n·0,96 > 2,58 $\cdot\sqrt{n \cdot 0{,}0384}$ | $()^2$ 67860,25 – 500,16n + 0,9216n² > 6,6564 · 0,0384 n
0,9216n² - 500,42n + 67860,25 = 0 -> n1 = 280,2 n2 = 262,74
In diesem Fall ist wohl die zweite Lösung (262,74) die richtige Lösung, denn 280 liegt ja deutlich über dem Erwartungswert von ca. 10 Ausfällen.
Antwort: Wenn mit 99,5 % Sicherheit bei einer Ausfallquote von 4 % keine Überbuchungen stattfinden sollen, dann dürfen maximal 262 Plätze verkauft werden.
<u>Übrigens:</u> Würde man die Sicherheit auf 90 % senken, dann könnten 267 Plätze verkauft werden. Die kleine Änderung von 500,42 n auf 500,2229 n ist dafür verantwortlich. Es werden 5 Plätze mehr verkauft, mit dem Risiko, dass bei jedem 10. Flug überbucht wird. Dann hat man aber die Einnahmen von 50 Flügen, die dann den Entschädigungen für die überbuchten Fluggäste gegenübergestellt werden.

Testen von Hypothesen bei normalverteilten Zufallsgrößen

Wenn die Standardabweichung und der Erwartungswert bekannt sind, kann man auf dem 5 % Signifikanzniveau testen:
a) Man wählt den Stichprobenumfang n und berechnet den Mittelwert
b) Man berechnet den Annahmebereich mit A = $\left[\mu - 1.96\frac{\sigma}{\sqrt{n}};\ \mu + 1{,}96\frac{\sigma}{\sqrt{n}}\right]$
c) Man verwirft die Hypothese, wenn der Mittelwert außerhalbe des Annahmebereichs liegt.

Einfache Rechnungen

a) Der Erwartungswert bei der Körpergröße von Studenten liegt bei 1,72 cm. Die Standardabweichung liegt bei 8. Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Student zwischen 1,70 und 1,74 cm groß ist?

Ansatz: P(1,70 ≤ X ≤ 1,74) $\Phi(1{,}74)$ - $\Phi(1{,}70)$
Standardisierung: $\Phi\left(\frac{1{,}74-1{,}72}{8}\right)$ - $\Phi\left(\frac{1{,}70-1{,}72}{8}\right)$ -> $\Phi(0{,}025)$ - $\Phi(-0{,}025)$
0,51 – 0,49 = 0,02 Die Wahrscheinlichkeit liegt bei 2 %

Schwieriger (Testen von Hypothesen, Erwartungswert oder Standardabweichung unbekannt)

b) Es werden 20 Studenten ausgewählt. Die mittlere Körpergröße dieser Studenten liegt bei 1,74 cm. In welchem Bereich muss der Erwartungswert liegen, wenn die Standardabweichung mit 8 bekannt ist. Die Schätzung soll im Bereich des 90 % Vertrauensniveau liegen.

Aus der Sigma-Tabelle entnehmen für ein 90 % Vertrauensintervall den Wert
$[\mu - 1{,}64\sigma;\ \mu + 1{,}64\sigma]$, wobei wir den Erwartungswert durch den Mittelwert ersetzen.

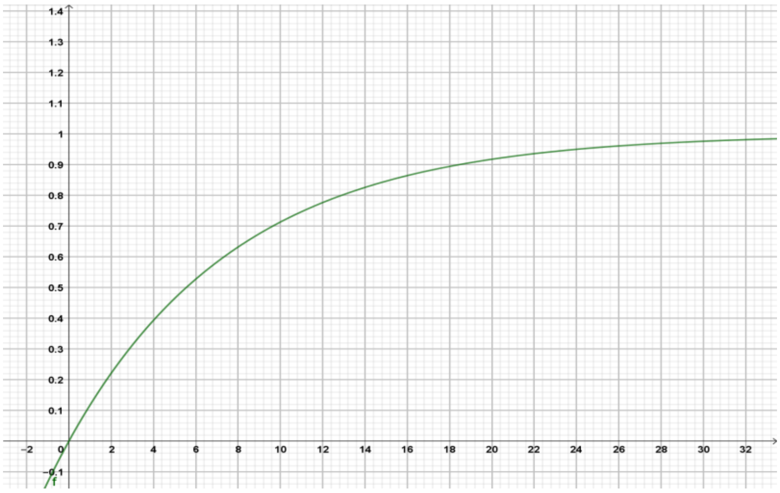
Dann ergibt sich eine Abweichung von $174 \pm \frac{1{,}64 \cdot 8}{\sqrt{20}}$ = 174 $\pm$ 2,93.
Der Erwartungswert liegt mit 90 % Wahrscheinlichkeit im Bereich [1,7107; 176,92]

Die Exponentialverteilung

Beispiel 1:
In einem großen Unternehmen wird die Lebensdauer der Rechner ermittelt. Der Anteil defekter Rechner steigt im Zeitablauf. Die mittlere Lebensdauer liegt bei 8 Jahren.
a) Wie groß ist die Wahrscheinlichkeit, dass ein Rechner die mittlere Lebensdauer erreicht?
b) Nach welcher Zeit sind 50 % aller Rechner defekt?

Um was für eine Art von Verteilung handelt es sich?
Wenn auf der x-Achse die unabhängige Größe (Zeit in Jahren) abgetragen wird, dann wird auf der y-Achse der Anteil der defekten Rechner abgetragen. Der Anteil liegt zwischen 0 und 1 und entspricht gleichzeitig der Wahrscheinlichkeit, dass Rechner nach x-Jahren defekt sind. Die Kurve wird sich also mit abnehmendem Wachstum dem Wert 1 nähern.
Man bezeichnet diesen Verlauf **als beschränktes Wachstum** mit der allgemeinen Formel

$F(x) = 1 - e^{-\lambda x}$, dabei ist λ der sogenannte Wachstums- oder Zerfallsfaktor



In diesem Fall ergibt sich die Gleichung

$$P(X \leq x) = F(x) = 1 - e^{-0{,}125x} \quad \text{wobei } \lambda = \frac{1}{\mu} = \frac{1}{8} \text{ oder } \mu = \frac{1}{\lambda}$$

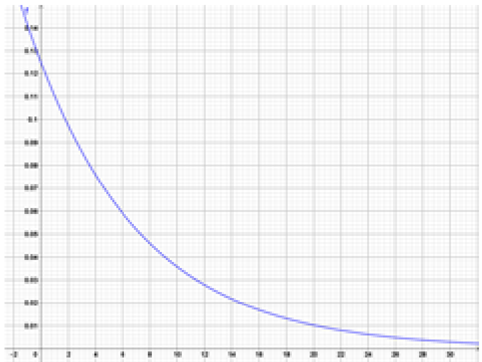
Die Lösungen zu den Aufgaben ergeben sich durch Einsetzen in die Funktionsgleichung.
Zu a) x = 8 also P(X >=8) = F(8) = 0,6321. D. h. nach 8 Jahren sind 63,21 % der Rechner defekt. Deshalb liegt die Wahrscheinlichkeit dafür, dass ein Rechner diese Lebensdauer erreicht bei
100-63,21 % = **36,79 %.**
Zu b) Ansatz: $1 - e^{-0{,}125x} = 0{,}5$ durch Logarithmieren ergibt sich **x = 5,545.**

Die genannte Funktion ist eine Dichtefunktion, da man aus ihr unmittelbar die Wahrscheinlichkeiten ablesen kann. Bildet man die Ableitung, so erhält man nach Anwenden der Kettenregel

$F(x) = 1 - e^{-0{,}125x}$ -> f(x)= $0{,}125e^{-0{,}125x}$

Der Verlauf dieses Graphen gibt dann an, wie groß der Anteil (oder die Wahrscheinlichkeit) der Rechner ist, genau nach x Jahren defekt zu sein.

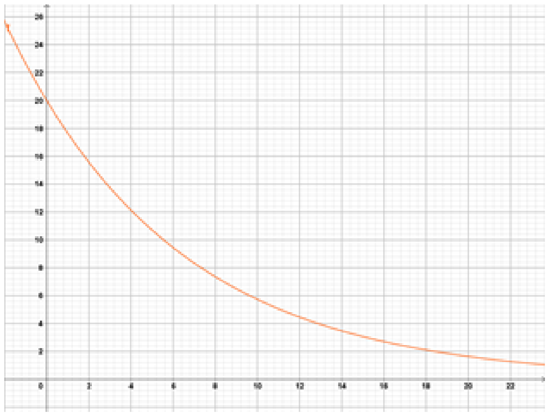
Die Fläche unter der Kurve gibt dann den Anteil aller Rechner an, die bis zu einem Zeitpunkt defekt sind.



Beispiel 2:
Ein Taxiunternehmen hat 20 Fahrzeuge. Die mittlere Laufzeit eines Fahrzeuges liegt bei 8 Jahren.
a) Wie groß ist die Wahrscheinlichkeit, dass ein Taxi nach 8 Jahren noch in Betrieb ist.
b) Nach welcher Zeit sind 75 % aller Fahrzeuge nicht mehr einsatzbereit?

Da der Erwartungswert bekannt ist, liegt die Zerfallskonstante bei 1/8 = 0,125. Bei einem Anfangswert von 20 lautet die Exponentialfunktion $f(x) = 20e^{-0{,}125x}$

Der Graph zeigt die Entwicklung des Bestandes an Taxis. Für die Wahrscheinlichkeiten oder die Berechnung der Anteile benötigt man die Dichtefunktion als Stammfunktion.
Bei der Integration verschwindet der Anfangswert, sodass die Dichtefunktion keine mengenmäßigen Angaben macht, sondern die Wahrscheinlichkeiten angibt.

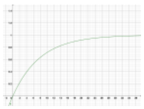


$F(x) = 1 - e^{-0,125x} \quad = \int_0^\infty 20e^{-0,125x}$

Das Schaubild zeigt die Dichtefunktion, die darstellt, wie groß der Anteil der Fahrzeuge ist, die nach x Jahren nicht mehr einsatzbereit ist.

Zu 2a) $1 - e^{-0,125 \cdot 8} = 0,6321$
Die Gegenwahrscheinlichkeit ist dann **36,79 %**

Zu 2b) $1 - e^{-0,125 \cdot x} = 0,75$



$0,25 = e^{-0,125 \cdot x} \quad | \ln$
$\ln 0,25 = -0,125\, x$
x = 11,09 Jahre

Beispiel 3:
Die mittlere Reichweite eines Motors liegt bei 250 000 km.
a) Wie groß ist die Wahrscheinlichkeit, dass ein Motor mindestens 250 000 km lang hält?
b) Nach welcher Laufleistung sind 50 % aller Motoren defekt?

Zu 3a) $\Lambda = \frac{1}{250000}$, dann heißt die Dichtefunktion $F(x) = 1 - e - {}^{-\frac{1}{250000}x}$
Für x 250000 eingesetzt ergibt 0,632 bzw. 36,8 % da die Dichtefunktion angibt, wie groß der Anteil der defekten Motoren ist.

Zu 3b) $1 - e - {}^{-\frac{1}{250000}x} = 0,5$ ergibt 173286 km.

Schwierigkeiten: Die drei Beispiele sollen zeigen, dass es darauf ankommt, die Ausgangsfunktion und die Dichtefunktion richtig zu interpretieren und zu wissen, wann es darauf ankommt, die Gegenwahrscheinlichkeit zu ermitteln. Interessant ist es auch, dass bei der Integration zur Ermittlung der Dichtefunktion der Anfangswert c verschwindet. Dies ergibt sich aus der partiellen Integration, was aber in diesem Skript nicht behandelt werden soll.

Geometrische Verteilungen

Geometrische Verteilungen von Zufallsgrößen erhält man, wenn gilt:

$q < 1$ und $p = 1 - q$ $\quad P(Z \le t) = 1 - q^t$ mit dem Erwartungswert $\mu = \frac{1}{p}$

Beispiel:

Beim Würfelspiel „Scheiße“ wird mit einem Würfel geworfen und die Augenzahlen zusammengezählt. Wenn allerdings eine 1 geworfen wird, wird die geworfene Augensumme verworfen und der nächste Spieler ist an der Reihe. Wer zuerst insgesamt 100 Punkte zusammen hat, gewinnt.
a) Mit welcher Wahrscheinlichkeit fällt die 1 nach dem 10. Wurf?
b) Mit welcher Wahrscheinlichkeit fällt in den ersten 5 Würfen eine 1?

Zu a) Hier ist $P(X = 10) = \frac{5}{6}^{9} \cdot \frac{1}{6} = 0{,}032 \;=\; 3{,}2\%$

Zu b) Hier gilt die Dichtefunktion $P(X \le 5) = 1 - \frac{5}{6}^{5} \;=\; 0{,}598 \;=\; 60\,\%$

Ergänzende Aufgaben

1. Bei der Auszählung einer Wahl erhalten die Grünen einen Stimmenanteil von 25 %. Es sind bisher 70 % der Stimmen ausgezählt worden.

   Wie groß ist die Wahrscheinlichkeit, dass das Endergebnis um nicht mehr als 2 % von diesem vorläufigen Ergebnis abweicht, wenn es für die Aussage eine Sicherheitswahrscheinlichkeit von 95 % geben soll?

   Bei dieser Aufgabe handelt es sich um eine sogenannte **Hochrechnung**, bei der anhand eines Stichprobenergebnisses auf das Endergebnis geschlossen werden soll. Auch wenn das Wahlergebnis der Grünen einer Normalverteilung unterliegen soll, so fehlen hier doch die Zahlen für die Größe der Stichprobe. Damit kann man die Standardabweichung nicht berechnen.
   In der Praxis werden solche Hochrechnungen entweder durch Regressionsrechnungen oder anhand von Erfahrungswerten einzelner Wahlbezirke in früheren Wahlen getroffen.

Anhang: Auszug aus der Tabelle der kumulierten Binomialverteilung für eins Stichprobe n = 100

Summierte Binomialverteilung (n = 100)

$$P(X \le k) = \sum_{i=0}^{k} \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$$

n	k	p 0,01	0,05	0,1	$\frac{1}{6}$	0,2	0,25	0,3	$\frac{1}{3}$	0,4	0,5	*k*	*n*
	0	0,3660	0,0059	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	*99*	
	1	0,7358	0,0371	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	*98*	
	2	0,9206	0,1183	0,0019	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	*97*	
	3	0,9816	0,2578	0,0078	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	*96*	
	4	0,9966	0,4360	0,0237	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	*95*	
	5	0,9995	0,6160	0,0576	0,0004	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	*94*	
	6	0,9999	0,7660	0,1172	0,0013	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	*93*	
	7	1,0000	0,8720	0,2061	0,0038	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	*92*	
	8		0,9369	0,3209	0,0095	0,0009	0,0000	0,0000	0,0000	0,0000	0,0000	*91*	
	9		0,9718	0,4513	0,0213	0,0023	0,0000	0,0000	0,0000	0,0000	0,0000	*90*	
	10		0,9885	0,5832	0,0427	0,0057	0,0001	0,0000	0,0000	0,0000	0,0000	*89*	
	11		0,9957	0,7030	0,0777	0,0126	0,0004	0,0000	0,0000	0,0000	0,0000	*88*	
	12		0,9985	0,8018	0,1297	0,0253	0,0010	0,0000	0,0000	0,0000	0,0000	*87*	
	13		0,9995	0,8761	0,2000	0,0469	0,0025	0,0001	0,0000	0,0000	0,0000	*86*	
	14		0,9999	0,9274	0,2874	0,0804	0,0054	0,0002	0,0000	0,0000	0,0000	*85*	
	15		1,0000	0,9601	0,3877	0,1285	0,0111	0,0004	0,0000	0,0000	0,0000	*84*	
	16			0,9794	0,4942	0,1923	0,0211	0,0010	0,0001	0,0000	0,0000	*83*	
	17			0,9900	0,5994	0,2712	0,0376	0,0022	0,0002	0,0000	0,0000	*82*	
	18			0,9954	0,6965	0,3621	0,0630	0,0045	0,0005	0,0000	0,0000	*81*	
	19			0,9980	0,7803	0,4602	0,0995	0,0089	0,0011	0,0000	0,0000	*80*	
	20			0,9992	0,8481	0,5595	0,1488	0,0165	0,0024	0,0000	0,0000	*79*	
	21			0,9997	0,8998	0,6540	0,2114	0,0288	0,0048	0,0000	0,0000	*78*	
	22			0,9999	0,9369	0,7389	0,2864	0,0479	0,0091	0,0001	0,0000	*77*	
	23			1,0000	0,9621	0,8109	0,3711	0,0755	0,0164	0,0003	0,0000	*76*	
	24				0,9783	0,8686	0,4617	0,1136	0,0281	0,0006	0,0000	*75*	
	25				0,9881	0,9125	0,5535	0,1631	0,0458	0,0012	0,0000	*74*	
	26				0,9938	0,9442	0,6417	0,2244	0,0715	0,0024	0,0000	*73*	
	27				0,9969	0,9658	0,7224	0,2964	0,1066	0,0046	0,0000	*72*	
	28				0,9985	0,9800	0,7925	0,3768	0,1524	0,0084	0,0000	*71*	
	29				0,9993	0,9888	0,8505	0,4623	0,2093	0,0148	0,0000	*70*	
	30				0,9997	0,9939	0,8962	0,5491	0,2766	0,0248	0,0000	*69*	
	31				0,9999	0,9969	0,9307	0,6331	0,3525	0,0398	0,0001	*68*	
	32				1,0000	0,9984	0,9554	0,7107	0,4344	0,0615	0,0002	*67*	
	33					0,9993	0,9724	0,7793	0,5188	0,0913	0,0004	*66*	
100	34					0,9997	0,9836	0,8371	0,6019	0,1303	0,0009	*65*	*100*
	35					0,9999	0,9906	0,8839	0,6803	0,1795	0,0018	*64*	
	36					0,9999	0,9948	0,9201	0,7511	0,2386	0,0033	*63*	
	37					1,0000	0,9973	0,9470	0,8123	0,3068	0,0060	*62*	
	38						0,9986	0,9660	0,8630	0,3822	0,0105	*61*	
	39						0,9993	0,9790	0,9034	0,4621	0,0176	*60*	
	40						0,9997	0,9875	0,9341	0,5433	0,0284	*59*	
	41						0,9999	0,9928	0,9566	0,6225	0,0443	*58*	
	42						0,9999	0,9960	0,9724	0,6967	0,0666	*57*	
	43						1,0000	0,9979	0,9831	0,7635	0,0967	*56*	
	44							0,9989	0,9900	0,8211	0,1356	*55*	
	45							0,9995	0,9943	0,8689	0,1841	*54*	
	46							0,9997	0,9969	0,9070	0,2421	*53*	
	47							0,9999	0,9983	0,9362	0,3086	*52*	
	48							0,9999	0,9991	0,9577	0,3822	*51*	
	49							1,0000	0,9996	0,9729	0,4602	*50*	