O'REILLY®

Leicht verständlich
Mit vielen Illustrationen und Beispielen

Mathe-Basics für Data Scientists

Lineare Algebra, Statistik und Wahrscheinlichkeitsrechnung für die Datenanalyse

Thomas Nield

Übersetzung von Frank Langenau

KAPITEL 2

Wahrscheinlichkeit

Was kommt Ihnen in den Sinn, wenn Sie den Begriff *Wahrscheinlichkeit* hören? Vielleicht denken Sie an Beispiele aus dem Bereich des Glücksspiels, wie die Wahrscheinlichkeit, im Lotto zu gewinnen oder ein Paar mit zwei Würfeln zu bekommen. Vielleicht geht es aber auch um die Vorhersage von Aktienkursen, den Ausgang einer politischen Wahl oder die Frage, ob Ihr Flug pünktlich ankommt. Unsere Welt ist voller Ungewissheiten, die wir nur allzu gern messen würden.

Vielleicht ist es das Wort, auf das wir uns konzentrieren sollten: Unsicherheit. Wie messen wir etwas, über das wir uns nicht sicher sind?

Letztlich ist die Wahrscheinlichkeit die theoretische Untersuchung zur Messung der Bestimmtheit, dass ein Ereignis eintreten wird. Sie ist eine Grundlagendisziplin für Statistik, Hypothesentests, maschinelles Lernen und andere Themen in diesem Buch. Viele Menschen halten Wahrscheinlichkeit für selbstverständlich und gehen davon aus, dass sie sie verstehen. Allerdings ist sie vielschichtiger und komplizierter, als die meisten Menschen denken. Während die Theoreme und Ideen der Wahrscheinlichkeit mathematisch fundiert sind, wird es komplexer, wenn wir Daten einführen und uns in die Statistik vorwagen. Wir werden dies in Kapitel 3 zu Statistik und Hypothesentests behandeln.

Zunächst aber beschäftigen wir uns in diesem Kapitel damit, was Wahrscheinlichkeit ist. Dann geht es um mathematische Konzepte der Wahrscheinlichkeitsrechnung, den Satz von Bayes, die Binomialverteilung und die Beta-Verteilung.

Wahrscheinlichkeit verstehen

Die *Wahrscheinlichkeit* gibt an, wie stark wir glauben, dass ein Ereignis eintreten wird, oftmals als Prozentsatz ausgedrückt. Hier sind einige Fragen, die eine Beantwortung durch einen Wahrscheinlichkeitswert rechtfertigen könnten:

- Wie wahrscheinlich ist es, dass ich bei zehn fairen Münzwürfen siebenmal Kopf erhalte?
- Wie hoch stehen meine Chancen, eine Wahl zu gewinnen?
- Wird mein Flug Verspätung haben?
- Wie sicher bin ich, dass ein Produkt defekt ist?

Eine Wahrscheinlichkeit gibt man häufig in Prozent an, wie zum Beispiel: »Mein Flug hat mit 70% Wahrscheinlichkeit Verspätung.« Wir bezeichnen diese Wahrscheinlichkeit als $P(X)$, wobei X das interessierende Ereignis ist. Wenn Sie mit Wahrscheinlichkeiten arbeiten, werden Sie sie jedoch eher als Dezimalzahl (in diesem Fall *0,70*) sehen, die zwischen *0,0* und *1,0* liegen muss:

$$P(X) = .70$$

Die *Likelihood* ist der Wahrscheinlichkeit sehr ähnlich, und man kann beide leicht verwechseln (was auch viele Wörterbücher tun). In der Alltagssprache lassen sich die Begriffe »Wahrscheinlichkeit« und »Likelihood« ohne Weiteres synonym verwenden. Allerdings sollten wir die Unterschiede genau herausarbeiten. Bei Wahrscheinlichkeit (engl. *Probability*) geht es um die quantitative Vorhersage von Ereignissen, die noch nicht eingetreten sind, während Likelihood die Häufigkeit von bereits eingetretenen Ereignissen misst. In der Statistik und beim maschinellen Lernen verwenden wir Likelihood (die Vergangenheit) oft in Form von Daten, um die Wahrscheinlichkeit (die Zukunft) vorherzusagen.

Es ist wichtig zu wissen, dass die Wahrscheinlichkeit, dass ein Ereignis eintritt, streng zwischen 0% und 100% oder 0,0 und 1,0 liegen muss. Logischerweise berechnet man die Wahrscheinlichkeit, dass ein Ereignis *nicht* eintritt, indem man die Wahrscheinlichkeit dieses Ereignisses von 1,0 subtrahiert:

$$P(X) = .70$$
$$P(\text{not } X) = 1 - .70 = .30$$

Dies ist eine weitere Unterscheidung zwischen Probability und Likelihood. Die Wahrscheinlichkeiten aller möglichen sich gegenseitig ausschließenden Ergebnisse für ein Ereignis (d.h., es kann nur ein Ergebnis vorkommen, nicht mehrere) müssen in der Summe 1,0 oder 100% ergeben. Für Likelihoods gilt diese Regel jedoch nicht.

Alternativ kann die Wahrscheinlichkeit als *Chance* $O(X)$ wie 7:3, 7/3 oder $2,\overline{333}$ ausgedrückt werden. Eine Chance $O(X)$ lässt sich in eine proportionale Wahrscheinlichkeit $P(X)$ mit folgender Formel überführen:

$$P(X) = \frac{O(X)}{1 + O(X)}$$

Wenn ich also eine Chance von 7/3 habe, kann ich sie in eine proportionale Wahrscheinlichkeit wie folgt umwandeln:

$$P(X) = \frac{O(X)}{1 + O(X)}$$
$$P(X) = \frac{\frac{7}{3}}{1 + \frac{7}{3}}$$
$$P(X) = .7$$

Umgekehrt kann man eine Wahrscheinlichkeit in eine Chance umwandeln, indem man die Wahrscheinlichkeit, dass das Ereignis eintritt, durch die Wahrscheinlichkeit, dass es nicht eintritt, dividiert:

$$O(X) = \frac{P(X)}{1 - P(X)}$$

$$O(X) = \frac{.70}{1 - .70}$$

$$O(X) = \frac{7}{3}$$

Chancen sind nützlich

Viele Menschen sind zwar vertrauter damit, Wahrscheinlichkeiten als Prozentwerte oder Proportionen auszudrücken, Chancen können aber ebenfalls ein hilfreiches Instrument sein. Wenn ich eine Chance von 2,0 habe, heißt das, dass ich ein Ereignis für doppelt so wahrscheinlich halte, dass es eintritt, als dass es nicht eintritt. Das kann eingängiger sein, als eine Überzeugung mit einem Prozentsatz von $66,\overline{666}$% zu beschreiben. Aus diesem Grund sind Chancen hilfreich, um subjektive Überzeugungen zu quantifizieren, insbesondere im Zusammenhang mit Glücksspielen oder Wetten. Chancen spielen zudem eine Rolle in der bayesschen Statistik (einschließlich des Bayes-Faktors) sowie in der logistischen Regression mit der Log-Odds-Funktion, die wir in Kapitel 6 behandeln.

Wahrscheinlichkeitsrechnung vs. Statistik

Gelegentlich werden die Begriffe *Wahrscheinlichkeit* und *Statistik* synonym verwendet. Doch obwohl es verständlich ist, beide Disziplinen zu vermischen, gibt es doch Unterschiede zwischen ihnen. Die *Wahrscheinlichkeitsrechnung* ist eine rein theoretische Betrachtung der Wahrscheinlichkeit, mit der ein Ereignis eintritt, ohne dass dazu Daten erforderlich wären. Hingegen kann die *Statistik* ohne Daten nicht existieren, und sie stützt sich darauf, um die Wahrscheinlichkeit zu ermitteln und Tools bereitzustellen, mit denen sich die Daten beschreiben lassen.

Denken Sie an die Vorhersage des Ergebnisses, mit einem Würfel eine 4 zu werfen. Nähert man sich dem Problem mit einer reinen Wahrscheinlichkeitsbetrachtung, sagt man einfach, dass es sechs Seiten auf einem Würfel gibt. Wir nehmen an, dass jede Seite gleiche Eigenschaften aufweist, sodass die Wahrscheinlichkeit, eine 4 zu würfeln bei 1/6 oder 16,666 % liegt.

Ein eifriger Statistiker könnte jetzt sagen: »Nein! Wir müssen würfeln, um die Daten zu bekommen. Nur wenn wir 30 Würfe oder mehr machen können – und je mehr Würfe, desto besser –, nur dann bekommen wir genügend Daten, um die Wahrscheinlichkeit zu bestimmen, eine 4 zu würfeln.« Dieser Ansatz mag albern erscheinen, wenn wir davon ausgehen, dass der Würfel fair ist. Aber was ist, wenn

das nicht zutrifft? In diesem Fall ist das Sammeln von Daten die einzige Möglichkeit, die Wahrscheinlichkeit für das Werfen einer 4 zu ermitteln. Über das Testen von Hypothesen sprechen wir in Kapitel 3.

Wahrscheinlichkeitsmathematik

Wenn wir mit einer einzelnen Wahrscheinlichkeit eines Ereignisses $P(X)$ arbeiten, die man als *Randwahrscheinlichkeit* bezeichnet, ist die Idee ziemlich einfach, wie bereits zuvor erwähnt. Doch wenn wir Wahrscheinlichkeiten von verschiedenen Ereignissen zusammenfassen, liegen die Dinge nicht mehr so klar auf der Hand.

Kombinierte Wahrscheinlichkeiten

Angenommen, Sie hätten eine faire Münze und einen fairen sechsseitigen Würfel. Nun möchten Sie wissen, wie groß die Wahrscheinlichkeit ist, mit der Münze Kopf (engl. *Heads*) zu werfen und mit dem Würfel eine 6 zu würfeln. Wir haben es hier mit zwei separaten Wahrscheinlichkeiten und zwei separaten Ereignissen zu tun, doch wir sind an der Wahrscheinlichkeit interessiert, dass beide Ereignisse zusammen eintreten. Dies wird als *kombinierte* oder *multivariate* (bei zwei Wahrscheinlichkeiten als *bivariate*) *Wahrscheinlichkeit* bezeichnet.

Stellen Sie sich eine kombinierte Wahrscheinlichkeit als AND-Operator vor. Ich möchte die Wahrscheinlichkeit ermitteln, dass ich Kopf werfe UND eine 6 würfle. Beide Ereignisse sollen zusammen eintreten. Wie berechnen wir diese Wahrscheinlichkeit?

Eine Münze hat zwei, ein Würfel sechs Seiten. Die Wahrscheinlichkeit für Kopf beträgt also 1/2, und die Wahrscheinlichkeit für eine 6 ist gleich 1/6. Um die Wahrscheinlichkeit zu berechnen, dass beide Ereignisse eintreten (unter der Annahme, dass sie unabhängig sind, mehr dazu später), multipliziert man die beiden Wahrscheinlichkeiten:

$$P(A \text{ AND } B) = P(A) \times P(B)$$

$$P(\text{heads}) = \frac{1}{2}$$

$$P(6) = \frac{1}{6}$$

$$P(\text{heads AND } 6) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12} = .08\overline{333}$$

Recht einfach, aber warum überhaupt? Viele Wahrscheinlichkeitsregeln lassen sich herleiten, indem man alle möglichen Kombinationen von Ereignissen bildet. Dieses Vorgehen stammt aus einem Bereich der diskreten Mathematik, in dem es um Permutationen und Kombinationen geht. In diesem Fall erzeugen Sie jedes mögliche Ergebnis aus Münz- und Würfelwürfen, wobei Sie Kopf (H für *Heads*) und Zahl (T für *Tails*) mit den Zahlen 1 bis 6 paaren. Das uns interessierende Ergebnis – Kopf mit 6 – ist in Sternchen (*) eingeschlossen:

```
H1  H2  H3  H4  H5  *H6*  T1  T2  T3  T4  T5  T6
```

Es gibt 12 mögliche Ergebnisse, wenn wir unsere Münze und unseren Würfel werfen. Als einziges Ergebnis ist für uns *H6* von Interesse, d.h. Kopf werfen und eine 6 würfeln. Da es nur ein Ergebnis gibt, das unsere Bedingung erfüllt, und 12 Ergebnisse möglich sind, beträgt die Wahrscheinlichkeit, Kopf zu werfen und eine 6 zu würfeln, gleich 1/12.

Anstatt alle möglichen Kombinationen zu erzeugen und die uns interessierenden zu zählen, kommen wir schneller zum Ziel, wenn wir die kombinierte Wahrscheinlichkeit per Multiplikation berechnen – mit der *Produktregel* für Wahrscheinlichkeiten:

$$P(A \text{ AND } B) = P(A) \times P(B)$$

$$P(\text{heads AND } 6) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12} = .08\overline{333}$$

Vereinigungswahrscheinlichkeiten

Wir haben über kombinierte Wahrscheinlichkeiten gesprochen, d.h. über die Wahrscheinlichkeit, dass zwei oder mehr Ereignisse gleichzeitig eintreten. Aber mit welcher Wahrscheinlichkeit tritt Ereignis A oder B ein? Bei OR-Verknüpfungen mit Wahrscheinlichkeiten spricht man von einer *Vereinigungswahrscheinlichkeit*.

Sehen wir uns zunächst Ereignisse an, die *sich gegenseitig ausschließen*, d.h. Ereignisse, die nicht gleichzeitig auftreten können. Wenn ich zum Beispiel einen Würfel werfe, kann ich nicht gleichzeitig eine 4 und eine 6 erhalten, sondern nur eines der beiden Ergebnisse. Die Summenwahrscheinlichkeit für diese Fälle lässt sich leicht ermitteln. Ich addiere einfach die Einzelwahrscheinlichkeiten. So berechnet sich die Wahrscheinlichkeit für eine 4 oder eine 6 bei einem Würfelwurf zu 2/6 = 1/3:

$$P(4) = \frac{1}{6}$$

$$P(6) = \frac{1}{6}$$

$$P(4 \text{ OR } 6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Doch wie sieht es aus mit Ereignissen, die *sich nicht gegenseitig ausschließen*, die also gleichzeitig auftreten können? Kehren wir zum Beispiel mit den Münz- und Würfelwürfen zurück. Wie hoch ist die Wahrscheinlichkeit, Kopf ODER eine 6 zu erhalten? Bevor Sie in Versuchung geraten, diese Wahrscheinlichkeiten zu addieren, wollen wir noch einmal alle möglichen Ergebnisse bilden und diejenigen markieren, die uns interessieren:

```
*H1*  *H2*  *H3*  *H4*  *H5*  *H6*  T1  T2  T3  T4  T5  *T6*
```

Wir sind jetzt an allen Kopf-Ergebnissen sowie an den 6-Ergebnissen interessiert. Wenn wir die 7 interessierenden zu den 12 möglichen Ergebnissen ins Verhältnis setzen, erhalten wir mit 7/12 eine korrekte Wahrscheinlichkeit von $0{,}58\overline{333}$.

Aber was passiert, wenn wir die Wahrscheinlichkeiten von Kopf und 6 addieren? Wir erhalten eine andere (und falsche!) Antwort von $0{,}\overline{666}$:

$$P(heads) = \frac{1}{2}$$
$$P(6) = \frac{1}{6}$$
$$P(\text{heads OR } 6) = \frac{1}{2} + \frac{1}{6} = \frac{4}{6} = .\overline{666}$$

Weshalb ist das so? Sehen Sie sich die Kombinationen von Münzwurf und Würfelergebnis noch einmal genau an. Finden Sie etwas Verdächtiges? Wenn wir die Wahrscheinlichkeiten addieren, zählen wir die Wahrscheinlichkeit, eine 6 zu erhalten, sowohl in *H6* als auch in *T6*, also doppelt! Sollte Ihnen das nicht einleuchten, versuchen Sie, die Wahrscheinlichkeit für Kopf oder einen Würfelwurf von 1 bis 5 zu ermitteln:

$$P(heads) = \frac{1}{2}$$
$$P(1 \text{ bis } 5) = \frac{5}{6}$$
$$P(\text{heads OR 1 bis } 5) = \frac{1}{2} + \frac{5}{6} = \frac{8}{6} = 1.\overline{333}$$

Wir erhalten eine Wahrscheinlichkeit von 133,333 %, was definitiv nicht korrekt ist, da eine Wahrscheinlichkeit nicht mehr als 100 % oder 1,0 betragen darf. Das Problem ist wieder, dass wir die Ergebnisse doppelt zählen.

Wenn Sie lange genug nachdenken, werden Sie feststellen, dass die logische Methode zur Beseitigung von Doppelzählungen in einer Summenwahrscheinlichkeit darin besteht, die kombinierte Wahrscheinlichkeit zu subtrahieren. Diese sogenannte Summenregel für Wahrscheinlichkeiten stellt sicher, dass jedes kombinierte Ereignis nur einmal gezählt wird:

$$P(A \text{ OR } B) = P(A) + P(B) - P(A \text{ AND } B)$$
$$P(A \text{ OR } B) = P(A) + P(B) - P(A) \times P(B)$$

Um zu unserem Beispiel zurückzukehren und die Wahrscheinlichkeit von Kopf oder 6 zu berechnen, müssen wir die kombinierte Wahrscheinlichkeit, Kopf oder 6 zu erhalten, von der Summenwahrscheinlichkeit subtrahieren:

$$P(heads) = \frac{1}{2}$$
$$P(6) = \frac{1}{6}$$
$$P(A \text{ OR } B) = P(A) + P(B) - P(A) \times P(B)$$
$$P(\text{heads OR } 6) = \frac{1}{2} + \frac{1}{6} - \left(\frac{1}{2} \times \frac{1}{6}\right) = .58\overline{333}$$

Diese Formel gilt auch für sich gegenseitig ausschließende Ereignisse. Wenn sich die Ereignisse gegenseitig ausschließen, d. h. nur ein Ergebnis A oder B möglich ist, nicht aber beide, ist die kombinierte Wahrscheinlichkeit *P*(A AND B) gleich 0, sodass sie sich selbst aus der Formel entfernt. Dann bleibt nur noch die Summierung der Ereignisse, wie wir es zuvor schon durchgeführt haben.

Zusammenfassend lässt sich sagen, dass bei einer Vereinigungswahrscheinlichkeit zwischen zwei oder mehreren Ereignissen, die sich nicht gegenseitig ausschließen, die kombinierte Wahrscheinlichkeit abgezogen werden muss, damit Wahrscheinlichkeiten nicht doppelt gezählt werden.

Bedingte Wahrscheinlichkeit und der Satz von Bayes

Ein Thema der Wahrscheinlichkeitsrechnung, das häufig Verwirrung stiftet, ist das Konzept der bedingten Wahrscheinlichkeit, d. h. der Wahrscheinlichkeit eines Ereignisses A, das eintritt, sofern Ereignis B eingetreten ist. Normalerweise drückt man das als *P*(A GIVEN B) oder *P*(A|B) aus.

Angenommen, eine Studie behauptet, dass 85 % der Krebspatienten Kaffee getrunken haben. Wie reagieren Sie auf diese Behauptung? Beunruhigt Sie das und fühlen Sie sich veranlasst, auf Ihr Lieblingsgetränk am Morgen zu verzichten? Definieren wir dies zunächst als bedingte Wahrscheinlichkeit *P*(Coffee given Cancer) oder *P*(Coffee|Cancer). Das stellt die Wahrscheinlichkeit dar, dass Menschen, die an Krebs erkrankt sind, Kaffee trinken.

Vergleichen wir dies in den Vereinigten Staaten mit dem Prozentsatz der Menschen, bei denen Krebs diagnostiziert wurde (0,5 % laut *cancer.gov*), und dem Prozentsatz der Menschen, die Kaffee trinken (65 % laut *statista.com*):

$$P(\text{Cancer}) = .005$$
$$P(\text{Coffee}) = .65$$
$$P(\text{Coffee}|\text{Cancer}) = .85$$

Hmmmm ... lassen Sie diese Zahlen für einen Moment auf sich wirken und fragen Sie sich, ob Kaffee hier wirklich das Problem ist. Bedenken Sie auch, dass nur 0,5 % der Bevölkerung zu irgendeinem Zeitpunkt an Krebs erkrankt sind. Allerdings trinken 65 % der Bevölkerung regelmäßig Kaffee. Wenn Kaffee zu Krebs beiträgt, müssten wir dann nicht viel höhere Krebszahlen als 0,5 % haben? Müssten es nicht eher 65 % sein?

Dies ist das Heimtückische an proportionalen Zahlen. Zunächst können sie ohne jeglichen Kontext signifikant erscheinen, und die Schlagzeilen der Medien können dies sicherlich für Klicks ausnutzen: »Neue Studie deckt auf, dass 85 % der Krebspatienten Kaffee trinken«, könnte es heißen. Natürlich ist dies albern, denn wir haben eine häufige Eigenschaft (Kaffee trinken) mit einer ungewöhnlichen Eigenschaft (an Krebs erkrankt sein) in Verbindung gebracht.

Der Grund, warum Menschen so leicht durch bedingte Wahrscheinlichkeiten verwirrt werden können, liegt darin, dass die Richtung der Bedingung eine Rolle spielt und die beiden Bedingungen irgendwie als gleichwertig verschmelzen. Die »Wahrscheinlichkeit, an Krebs zu erkranken, wenn man Kaffeetrinker ist«, ist etwas anderes als die »Wahrscheinlichkeit, Kaffeetrinker zu sein, wenn man Krebs hat«. Einfacher ausgedrückt: Nur wenige Kaffeetrinker haben Krebs, aber viele Krebspatienten trinken Kaffee.

Wenn wir untersuchen wollen, ob Kaffee zu Krebs beiträgt, sind wir eigentlich an der ersten bedingten Wahrscheinlichkeit interessiert: der Wahrscheinlichkeit, dass jemand Krebs hat, wenn er Kaffeetrinker ist.

$$P(\text{Coffee}|\text{Cancer}) = .85$$
$$P(\text{Cancer}|\text{Coffee}) = ?$$

Wie können wir die Bedingung umkehren? Es gibt mit dem *Satz von Bayes* eine leistungsfähige kleine Formel, mit der sich bedingte Wahrscheinlichkeiten umkehren lassen.

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

Wenn wir die Informationen, die wir bereits haben, in diese Formel einsetzen, können wir sie nach der Wahrscheinlichkeit auflösen, mit der jemand Krebs hat, wenn er Kaffee trinkt:

$$\mathrm{P}(\mathrm{A}|\mathrm{B}) = \frac{\mathrm{P}(\mathrm{B}|\mathrm{A}) * \mathrm{P}(\mathrm{A})}{\mathrm{P}(\mathrm{B})}$$
$$\mathrm{P}(\text{Cancer}|\text{Coffee}) = \frac{\mathrm{P}(\text{Coffee}|\text{Cancer}) * \mathrm{P}(\text{Cancer})}{\mathrm{P}(\text{Coffee})}$$
$$\mathrm{P}(\text{Cancer}|\text{Coffee}) = \frac{.85 * .005}{.65} = .0065$$

Beispiel 2-1 zeigt, wie Sie das in Python berechnen können.

Beispiel 2-1: Den Satz von Bayes in Python verwenden

```
p_coffee_drinker = .65
p_cancer = .005
p_coffee_drinker_given_cancer = .85

p_cancer_given_coffee_drinker = p_coffee_drinker_given_cancer *
    p_cancer / p_coffee_drinker

# gibt 0.006538461538461539 aus
print(p_cancer_given_coffee_drinker)
```

Die Wahrscheinlichkeit, dass jemand Krebs hat, wenn er Kaffeetrinker ist, beträgt also nur 0,65 %! Dieser Wert unterscheidet sich stark von der mit 85 % angegebenen Wahrscheinlichkeit, dass jemand Kaffeetrinker ist, wenn er Krebs hat. Verste-

hen Sie nun, warum die Richtung der Bedingung wichtig ist? Der Satz von Bayes ist aus diesem Grund hilfreich. Man kann ihn auch heranziehen, um mehrere bedingte Wahrscheinlichkeiten zu verketten und unsere Überzeugungen basierend auf neuen Informationen zu aktualisieren.

Was macht einen »Kaffeetrinker« aus?

Natürlich hätte ich hier auch andere Variablen berücksichtigen können, insbesondere die Frage, was einen als »Kaffeetrinker« qualifiziert. Wenn jemand einmal im Monat Kaffee trinkt, ein anderer aber täglich, sollte ich dann beide als »Kaffeetrinker« bezeichnen? Was ist mit der Person, die vor einem Monat mit dem Kaffeetrinken begonnen hat, im Gegensatz zu der, die seit 20 Jahren Kaffee trinkt? Wie oft und wie lange müssen Menschen Kaffee trinken, bevor sie die Schwelle zum »Kaffeetrinker« in dieser Krebsstudie erreichen?

Dies sind wichtige Fragen, die zu bedenken sind, und sie zeigen, warum Daten selten die ganze Geschichte erzählen. Wenn Ihnen jemand eine Tabelle mit Patienten vorlegt, auf der einfach nur angekreuzt ist, ob sie Kaffeetrinker sind, muss dieser Schwellenwert definiert werden! Oder wir brauchen eine aussagekräftigere Metrik wie »Anzahl der in den letzten drei Jahren konsumierten Kaffeegetränke«. Ich habe dieses Gedankenexperiment einfach gehalten und nicht definiert, wie sich jemand als »Kaffeetrinker« qualifiziert, aber seien Sie sich darüber im Klaren, dass es in der Praxis immer eine gute Idee ist, Daten kritisch zu bewerten. In Kapitel 3 gehe ich ausführlicher darauf ein.

Wenn Sie sich eingehender mit der Intuition hinter dem Satz von Bayes befassen möchten, lesen Sie Anhang A. Fürs Erste genügt es, zu wissen, dass er uns hilft, eine bedingte Wahrscheinlichkeit umzudrehen. Als Nächstes sprechen wir darüber, wie bedingte Wahrscheinlichkeiten mit kombinierten und Vereinigungsoperationen zusammenwirken.

Naive Bayes

Der Satz von Bayes spielt eine zentrale Rolle in einem verbreiteten Algorithmus für maschinelles Lernen, Naive Bayes genannt. Joel Grus behandelt ihn in seinem Buch *Einführung in Data Science* (O'Reilly).

Kombinierte und vereinigte bedingte Wahrscheinlichkeiten

Sehen wir uns noch einmal die kombinierten Wahrscheinlichkeiten an und wie sie mit bedingten Wahrscheinlichkeiten interagieren. Ich möchte die Wahrscheinlichkeit ermitteln, dass jemand ein Kaffeetrinker ist UND Krebs hat. Soll ich $P(\text{Coffee})$ und $P(\text{Cancer})$ multiplizieren? Oder sollte ich $P(\text{Coffee}|\text{Cancer})$ anstelle von $P(\text{Coffee})$ verwenden, sofern dies verfügbar ist? Welche Werte soll ich verwenden?

Option 1:
$P(\text{Coffee}) \times P(\text{Cancer}) = .65 \times .005 = .00325$

Option 2:
$P(\text{Coffee}|\text{Cancer}) \times P(\text{Cancer}) = .85 \times .005 = .00425$

Wenn wir bereits festgestellt haben, dass unsere Wahrscheinlichkeit nur für Menschen mit Krebs gilt, ist es dann nicht sinnvoll, $P(\text{Coffee}|\text{Cancer})$ anstelle von $P(\text{Coffee})$ zu verwenden? Das eine ist spezifischer und bezieht sich auf eine Bedingung, die bereits bekannt ist. Wir sollten also $P(\text{Coffee}|\text{Cancer})$ nehmen, da $P(\text{Cancer})$ bereits Teil unserer kombinierten Wahrscheinlichkeit ist. Das bedeutet, dass die Wahrscheinlichkeit, dass jemand Krebs hat und Kaffeetrinker ist, bei 0,425 % liegt:

$$P(\text{Coffee and Cancer}) = P(\text{Coffee}|\text{Cancer}) \times P(\text{Cancer}) = .85 \times .005 = .00425$$

Diese kombinierte Wahrscheinlichkeit gilt auch in umgekehrter Richtung. Ich kann die Wahrscheinlichkeit, dass jemand Kaffeetrinker ist und Krebs hat, ermitteln, indem ich $P(\text{Cancer}|\text{Coffee})$ und $P(\text{Coffee})$ multipliziere. Wie Sie sich überzeugen können, komme ich auf das gleiche Ergebnis:

$$P(\text{Cancer}|\text{Coffee}) \times P(\text{Coffee}) = .0065 \times .65 = .00425$$

Hätten wir keine bedingten Wahrscheinlichkeiten zur Verfügung, könnten wir bestenfalls $P(\text{Coffee Drinker})$ und $P(\text{Cancer})$ wie folgt multiplizieren:

$$P(\text{Coffee Drinker}) \times P(\text{Cancer}) = .65 \times .005 = .00325$$

Überlegen Sie nun Folgendes: Wenn Ereignis A keinen Einfluss auf Ereignis B hat, was bedeutet das dann für die bedingte Wahrscheinlichkeit $P(\text{B}|\text{A})$? Es bedeutet, dass $P(\text{B}|\text{A}) = P(\text{B})$ ist, also das Eintreten von Ereignis A keinen Einfluss darauf hat, wie wahrscheinlich Ereignis B auftritt. Demzufolge können wir unsere Formel für die kombinierte Wahrscheinlichkeit wie folgt aktualisieren – und zwar unabhängig davon, ob die beiden Ereignisse voneinander abhängig sind:

$$P(A \text{ AND } B) = P(B) \times P(A|B)$$

Und schließlich wollen wir uns noch mit vereinigten und bedingten Wahrscheinlichkeiten befassen. Wenn ich die Wahrscheinlichkeit berechnen will, ob A oder B eintritt, aber A die Wahrscheinlichkeit von B beeinflussen kann, aktualisieren wir unsere Summenregel so:

$$P(A \text{ OR } B) = P(A) + P(B) - P(A|B) \times P(B)$$

Zur Erinnerung: Dies gilt auch für sich gegenseitig ausschließende Ereignisse. Die Summenregel $P(\text{A}|\text{B}) \times P(\text{B})$ würde 0 ergeben, wenn die Ereignisse A und B nicht gleichzeitig eintreten können.

Binomialverteilung

Im weiteren Verlauf dieses Kapitels lernen Sie zwei Wahrscheinlichkeitsverteilungen kennen: die Binomial- und die Beta-Verteilung. Auch wenn wir diese Verteilungen für den Rest des Buchs nicht mehr verwenden werden, sind sie doch nützliche Hilfsmittel und grundlegend für das Verständnis, wie Ereignisse bei einer bestimmten Anzahl von Versuchen auftreten. Zudem bilden sie eine gute Überleitung zum Verständnis von Wahrscheinlichkeitsverteilungen, die wir in Kapitel 3 intensiv nutzen werden. Schauen wir uns einen Anwendungsfall an, der in einem praktischen Szenario auftreten könnte.

Angenommen, Sie arbeiteten an einem neuen Strahltriebwerk und haben zehn Tests durchgeführt. Die Ergebnisse ergaben acht Erfolge und zwei Fehlschläge:

✓ ✓ ✓ ✓ ✓ ✘ ✓ ✘ ✓ ✓

Ursprünglich hatten Sie gehofft, eine Erfolgsquote von 90 % zu erreichen. Aber auf der Grundlage der vorliegenden Daten kommen Sie zu dem Schluss, dass Ihre Tests mit nur 80 % Erfolg gescheitert sind. Da jeder Test zeitaufwendig und teuer ist, entscheiden Sie, das Design überarbeiten zu lassen.

Allerdings besteht eine Ihrer Ingenieurinnen darauf, weitere Tests durchzuführen: »Der einzige Weg, um sicherzugehen, ist, mehr Tests durchzuführen«, argumentiert sie. »Was ist, wenn mehr Tests einen Erfolg von 90 % oder mehr bringen? Denn wenn man eine Münze zehnmal wirft und achtmal Kopf erhält, heißt das ja nicht, dass die Münze zu 80 % korrekt liegt.«

Sie denken kurz über das Argument der Ingenieurin nach und stellen fest, dass sie nicht ganz unrecht hat. Selbst ein fairer Münzwurf hat nicht immer ein gleichmäßig aufgeteiltes Ergebnis, schon gleich gar nicht bei nur zehn Würfen. Am wahrscheinlichsten ist es, dass Sie fünfmal Kopf bekommen, aber Sie können auch drei-, vier-, sechs- oder siebenmal Kopf bekommen. Es ist sogar möglich, dass Sie zehnmal Kopf werfen, auch wenn dies ziemlich unwahrscheinlich ist. Wie kann man also die Wahrscheinlichkeit für einen 80%igen Erfolg bestimmen, wenn die zugrunde liegende Wahrscheinlichkeit 90 % beträgt?

Ein Instrument, das hier infrage kommt, ist die *Binomialverteilung*, die ein Maß dafür liefert, wie wahrscheinlich k Erfolge auftreten, wenn n Versuche mit einer Wahrscheinlichkeit von p durchgeführt werden.

Abbildung 2-1 zeigt, wie eine Binomialverteilung grafisch aussieht.

Hier sehen Sie die Wahrscheinlichkeit von k Erfolgen für jeden Balken von insgesamt zehn Versuchen. Diese Binomialverteilung geht von einer Wahrscheinlichkeit p von 90 % aus, d. h., es besteht eine Chance von 0,90 (oder 90 %) dafür, dass ein Erfolg auftritt. Unter diesen Voraussetzungen beträgt die Wahrscheinlichkeit 0,1937, dass wir acht Erfolge bei zehn Versuchen erhalten. Die Wahrscheinlichkeit, nur einen Erfolg bei zehn Versuchen zu erhalten, ist mit 0,000000008999 äußerst gering, weshalb der Balken im Diagramm nicht einmal sichtbar ist.

Wir können auch die Wahrscheinlichkeit für acht oder weniger Erfolge berechnen, indem wir die Balken für acht oder weniger Erfolge addieren. Dies würde uns eine Wahrscheinlichkeit von 0,2639 für acht oder weniger Erfolge liefern.

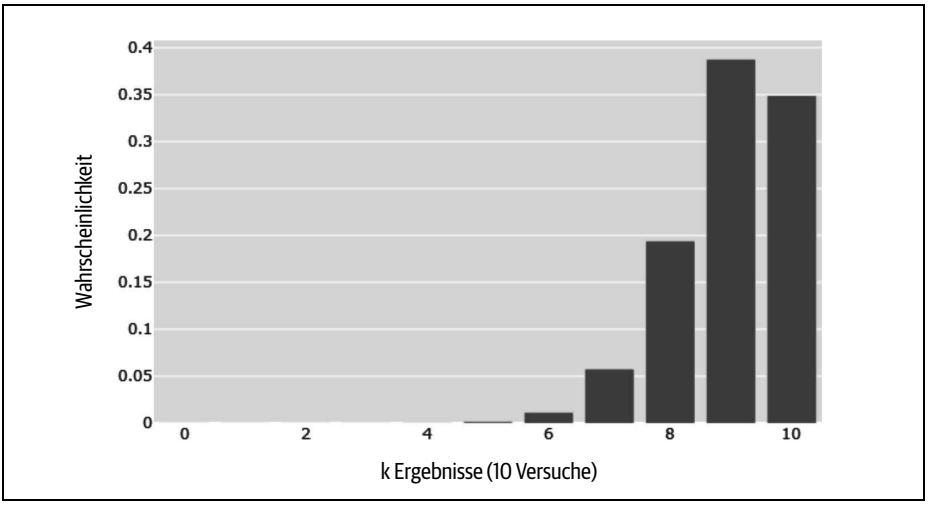

Abbildung 2-1: Eine Binomialverteilung

Wie implementieren wir nun die Binomialverteilung? Zum einen können wir es relativ einfach von Grund auf tun (wie in Anhang A beschrieben), oder wir greifen auf Bibliotheken wie SciPy zurück. Beispiel 2-2 zeigt, wie wir die SciPy-Funktion `binom.pmf()` verwenden (wobei `pmf` für *probability mass function* steht), um alle elf Wahrscheinlichkeiten für unsere Binomialverteilung von 0 bis 10 Erfolge auszugeben.

Beispiel 2-2: SciPy für die Binomialverteilung verwenden

```
from scipy.stats import binom

n = 10
p = 0.9

for k in range(n + 1):
    probability = binom.pmf(k, n, p)
    print("{0} - {1}".format(k, probability))

# OUTPUT:

# 0 - 9.99999999999996e-11
# 1 - 8.999999999999996e-09
# 2 - 3.644999999999996e-07
# 3 - 8.748000000000003e-06
# 4 - 0.0001377809999999999
# 5 - 0.0014880347999999988
# 6 - 0.011160260999999996
# 7 - 0.05739562800000001
# 8 - 0.19371024449999993
# 9 - 0.38742048900000037
# 10 - 0.34867844010000004
```

Wie aus dem Code hervorgeht, spezifizieren wir mit `n` die Anzahl der Versuche, mit `p` die Erfolgswahrscheinlichkeit und mit `k` die Anzahl der Erfolge, für die wir die Wahrscheinlichkeit ermitteln möchten. Wir iterieren über jeder Anzahl von Erfolgen k mit der entsprechenden Wahrscheinlichkeit, so viele Erfolge zu sehen. Wie die Ausgabe zeigt, ist 9 die wahrscheinlichste Anzahl von Erfolgen.

Aber wenn wir die Wahrscheinlichkeit für acht oder weniger Erfolge aufsummieren, erhalten wir 0,2639. Das bedeutet, dass wir mit einer Chance von 26,39% acht oder weniger Erfolge sehen, wenn die zugrunde liegende Erfolgsquote 90% beträgt. Vielleicht hat die Ingenieurin also recht: Eine Chance von 26,39% ist nicht nichts und durchaus möglich.

Allerdings haben wir hier in unserem Modell eine Annahme getroffen, die wir als Nächstes mit der Beta-Verteilung diskutieren werden.

Binomialverteilung von Grund auf

In Anhang A erfahren Sie, wie Sie die Binomialverteilung von Grund auf ohne SciPy erstellen können.

Beta-Verteilung

Wovon bin ich bei meinem Triebwerktestmodell mit der Binomialverteilung ausgegangen? Gibt es einen Parameter, den ich als `true` angenommen habe und um den ich dann mein gesamtes Modell aufgebaut habe? Denken Sie genau nach und lesen Sie weiter.

Bei meiner Binomialverteilung könnte es problematisch sein, dass ich die zugrunde liegende Erfolgsquote mit 90% *angenommen* habe. Das soll nicht heißen, dass mein Modell wertlos ist. Ich habe lediglich gezeigt, dass bei einer zugrunde liegenden Erfolgsquote von 90% eine Chance von 26,39% besteht, dass ich bei zehn Versuchen acht oder weniger Erfolge erzielen würde. Die Ingenieurin hatte also nicht unrecht, als sie meinte, dass die zugrunde liegende Erfolgsquote bei 90% liegen könnte.

Aber drehen wir die Frage um: Was wäre, wenn es neben den 90% noch andere zugrunde liegende Erfolgsquoten gäbe, die 8/10 Erfolge liefern würden? Könnten wir 8/10 Erfolge mit einer zugrunde liegenden Erfolgsquote von 80% sehen? 70%? 30%? Wenn wir die 8/10 Erfolge festlegen, können wir dann die Wahrscheinlichkeiten der Wahrscheinlichkeiten untersuchen?

Anstatt zahllose Binomialverteilungen zu erstellen, um diese Frage zu beantworten, können wir mit der *Beta-Verteilung* auf ein Werkzeug zurückgreifen, mit dem wir die Likelihood der verschiedenen zugrunde liegenden Wahrscheinlichkeiten für das Eintreten eines Ereignisses bei *Alpha*-Erfolgen oder *Beta*-Misserfolgen sehen. Abbildung 2-2 zeigt ein Diagramm der Beta-Verteilung bei acht Erfolgen und zwei Misserfolgen.

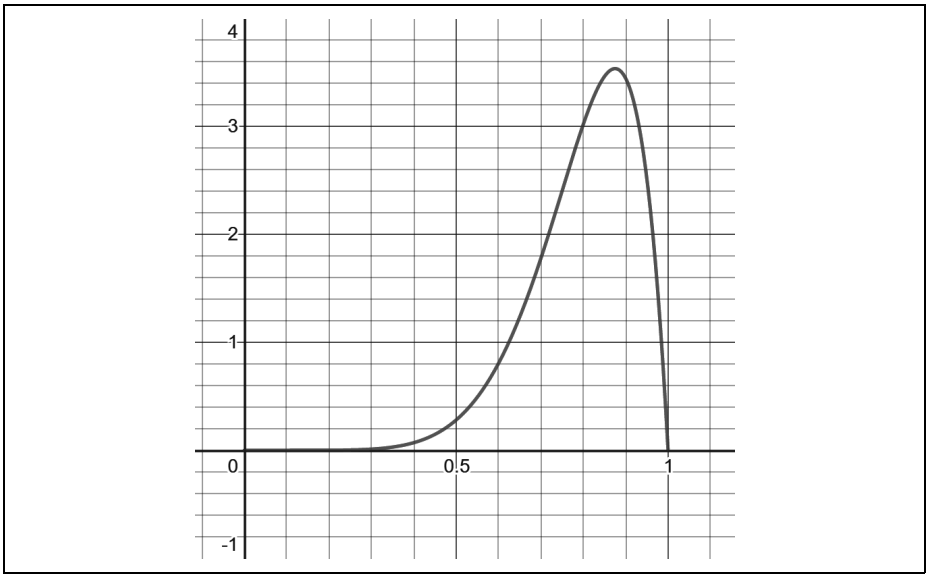

Abbildung 2-2: Beta-Verteilung

Beta-Verteilung auf Desmos
Wenn Sie mit der Beta-Verteilung spielen möchten, finden Sie unter *https://oreil.ly/pN4Ep* einen Desmos-Graphen.

Die x-Achse stellt alle zugrunde liegenden Erfolgsraten von 0,0 bis 1,0 (0% bis 100%) dar und die y-Achse die Likelihood dieser Wahrscheinlichkeit bei acht Erfolgen und zwei Misserfolgen. Mit anderen Worten, die Beta-Verteilung erlaubt uns, die Wahrscheinlichkeiten von Wahrscheinlichkeiten bei 8/10 Erfolgen zu sehen. Betrachten Sie sie als eine Meta-Wahrscheinlichkeit und nehmen Sie sich Zeit, um diese Idee zu begreifen!

Darüber hinaus ist die Beta-Verteilung eine stetige Funktion, bildet also eine kontinuierliche Kurve von Dezimalwerten (im Unterschied zu den ordentlichen und diskreten Ganzzahlen in der Binomialverteilung). Das macht die Mathematik mit der Beta-Verteilung etwas schwieriger, da ein bestimmter Dichtewert auf der y-Achse keine Wahrscheinlichkeit ist. Stattdessen ermitteln wir Wahrscheinlichkeiten anhand der Flächen unter der Kurve.

Die Beta-Verteilung ist eine Art *Wahrscheinlichkeitsverteilung*, d.h., die Fläche unter der gesamten Kurve ist 1,0 oder 100%. Um eine Wahrscheinlichkeit zu ermitteln, müssen wir die Fläche innerhalb eines Bereichs bestimmen. Wenn wir zum Beispiel die Wahrscheinlichkeit bewerten wollen, dass 8/10 Erfolge eine Erfolgsquote von 90% oder mehr ergeben, müssen wir die Fläche zwischen 0,9 und 1,0 finden, die 0,225 beträgt. In Abbildung 2-3 ist sie schattiert dargestellt.

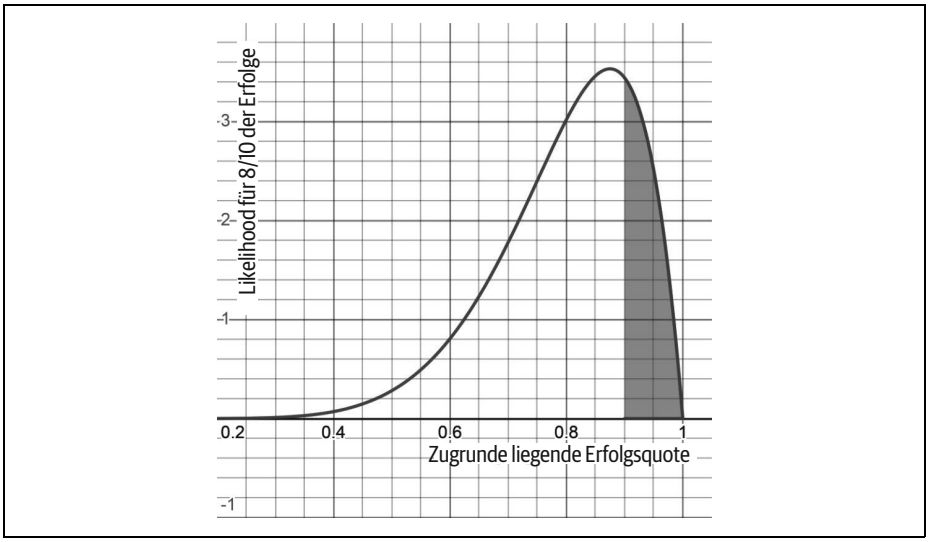

Abbildung 2-3: Die Fläche zwischen 90% und 100%, die 22,5% ausmacht

Wie die Binomialverteilung können wir die Beta-Verteilung mithilfe von SciPy implementieren. Jede stetige Wahrscheinlichkeitsverteilung hat eine *kumulative Verteilungsfunktion* (*Cumulative Distribution Function*, CDF), die die Fläche bis zu einem bestimmten x-Wert berechnet. Angenommen, ich möchte die Fläche bis zu 90% (0,0 bis 0,90) berechnen, wie sie in Abbildung 2-4 schattiert ist.

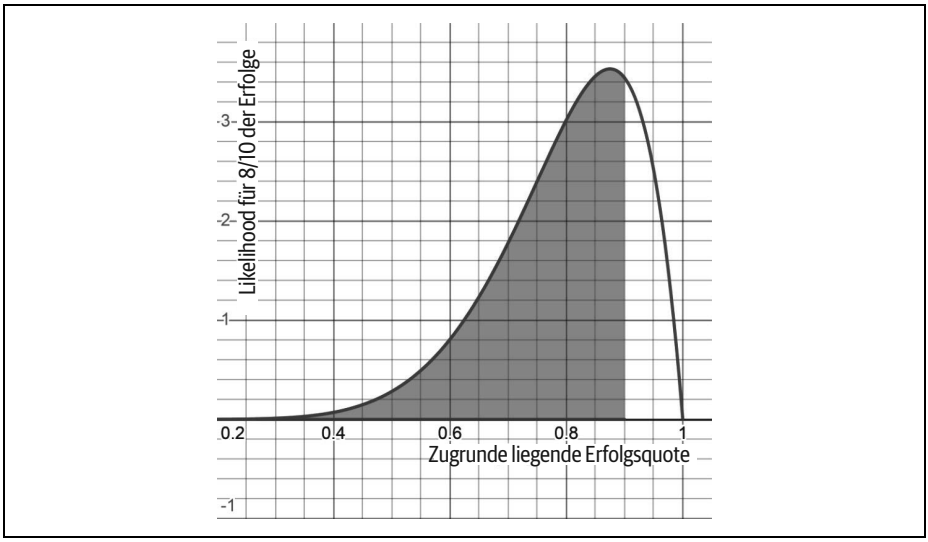

Abbildung 2-4: Die Fläche bis zu 90% (0,0 bis 0,90) berechnen

Es ist ziemlich einfach, SciPy mit der Funktion `beta.cdf()` zu verwenden. Als Parameter muss ich lediglich den `x`-Wert, die Anzahl der Erfolge `a` und Anzahl der Misserfolge `b` angeben, wie Beispiel 2-3 zeigt.

Beispiel 2-3: Beta-Verteilung mithilfe von SciPy

```
from scipy.stats import beta

a = 8
b = 2

p = beta.cdf(.90, a, b)

# 0.7748409780000001
print(p)
```

Entsprechend unserer Berechnung besteht also eine Chance von 77,48% für die zugrunde liegende Erfolgswahrscheinlichkeit von 90% oder weniger. Wie berechnen wir die Erfolgswahrscheinlichkeit für 90% oder mehr, wie in Abbildung 2-5 schattiert dargestellt?

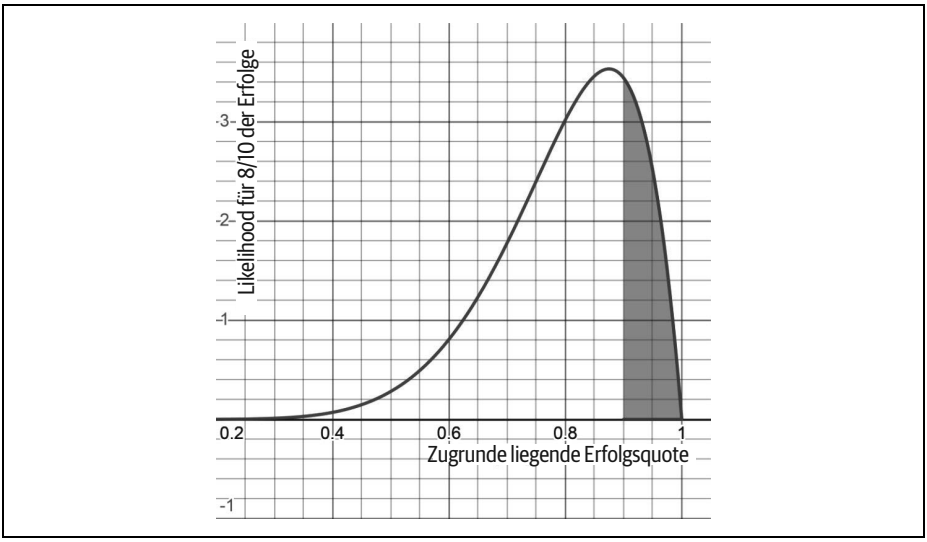

Abbildung 2-5: Die Erfolgswahrscheinlichkeit beträgt 90% oder mehr.

Unsere CDF berechnet nur die Fläche links von unserer Grenze, nicht rechts davon. Denken Sie an die Wahrscheinlichkeitsregeln, und bei einer Wahrscheinlichkeitsverteilung ist die Gesamtfläche unter der Kurve 1,0. Wenn wir die gegenteilige Wahrscheinlichkeit eines Ereignisses ermitteln wollen (größer als 0,90 im Unterschied zu kleiner als 0,90), subtrahieren wir einfach die Wahrscheinlichkeit, kleiner als 0,90 zu sein, von 1,0, und die restliche Wahrscheinlichkeit wird größer als 0,90 sein. Abbildung 2-6 veranschaulicht diese Subtraktion.

Beispiel 2-4 zeigt, wie wir diese Subtraktion in Python berechnen.

Beispiel 2-4: Subtrahieren, um einen rechten Bereich in einer Beta-Verteilung zu erhalten

```
from scipy.stats import beta

a = 8
```

```
b = 2

p = 1.0 - beta.cdf(.90, a, b)

# 0.22515902199999993
print(p)
```

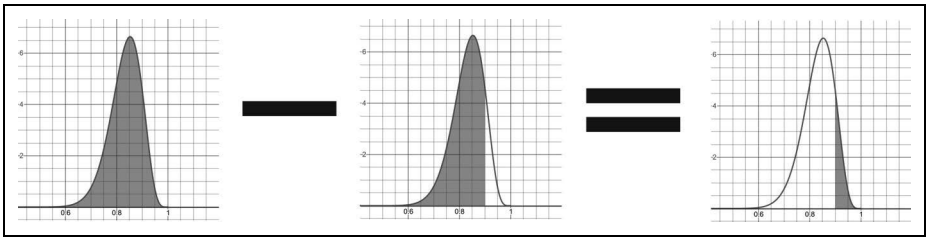

Abbildung 2-6: Die Wahrscheinlichkeit ermitteln, dass der Erfolg größer als 90% ist

Das bedeutet, dass bei 8/10 erfolgreichen Motortests nur eine Chance von 22,5% besteht, dass die zugrunde liegende Erfolgsrate 90% oder größer ist. Es besteht aber eine Wahrscheinlichkeit von 77,5%, dass die Chance kleiner als 90% ist. Die Chancen sehen also nicht gut aus, dass unsere Tests erfolgreich waren, aber wir könnten darauf setzen, dass eine Rate von 22,5% bei weiteren Tests erreicht wird, wenn wir Glück haben. Sollte unser CFO weitere 26 Tests bewilligen, die 30 Erfolge und 6 Misserfolge ergeben, würde unsere Beta-Verteilung wie in Abbildung 2-7 aussehen.

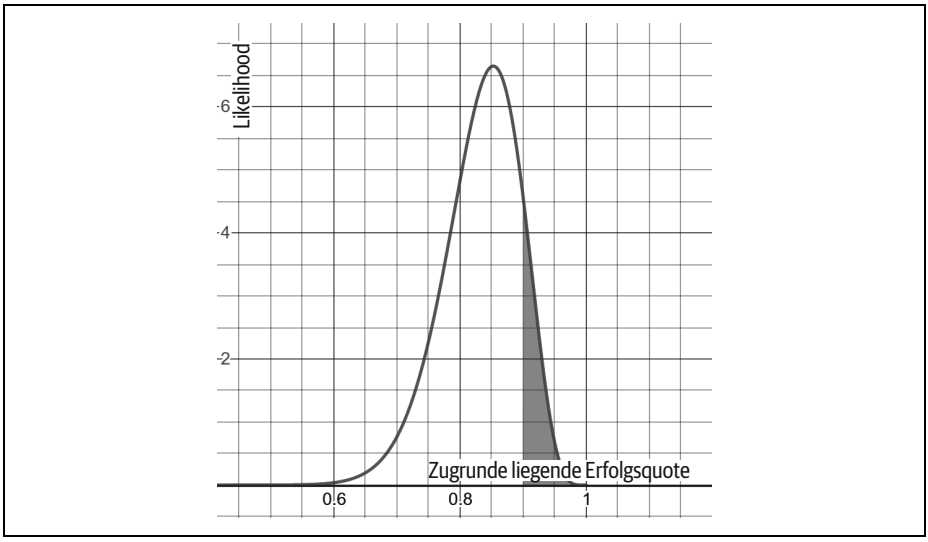

Abbildung 2-7: Beta-Verteilung nach 30 Erfolgen und 6 Misserfolgen

Unsere Verteilung ist nun enger geworden. Damit steigt die Zuversicht, dass die zugrunde liegende Erfolgsquote in einem kleineren Bereich liegt. Leider ist die Wahrscheinlichkeit, dass wir unser Minimum bei 90% erreichen, von 22,5% auf 13,16% gesunken, wie Abbildung 2-5 zeigt.

Beispiel 2-5: Eine Beta-Verteilung mit mehr Versuchen

```
from scipy.stats import beta
a = 30
b = 6
p = 1.0 - beta.cdf(.90, a, b)
# 0.13163577484183708
print(p)
```

An diesem Punkt wäre es vielleicht eine gute Idee, die Tests vorübergehend abzubrechen, es sei denn, Sie spielen weiterhin gegen diese 13,16%-Chance und hoffen, dass sich die Spitze nach rechts bewegt.

Zu guter Letzt: Wie berechnen wir einen Bereich in der Mitte? Wie sieht es aus, wenn Sie die Wahrscheinlichkeit ermitteln möchten, dass meine Erfolgsrate zwischen 80% und 90% liegt, wie Abbildung 2-8 zeigt?

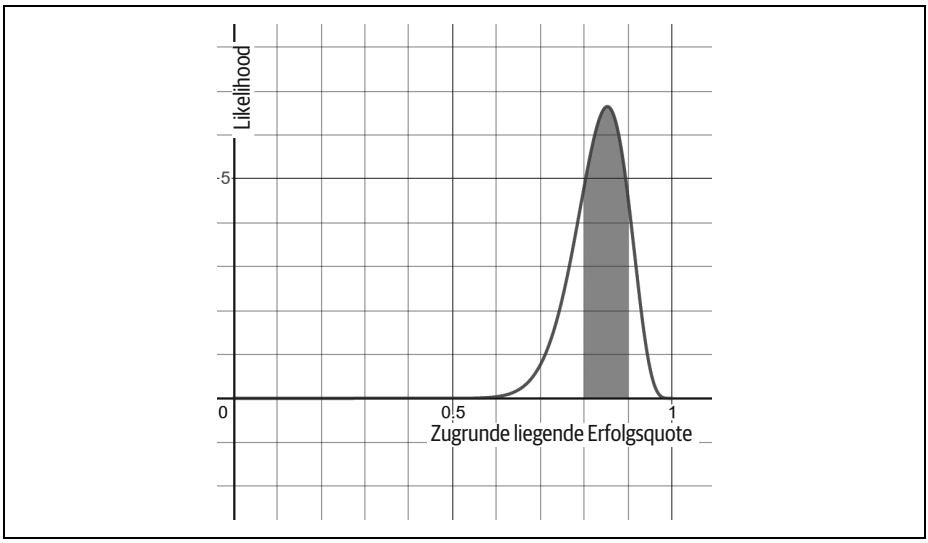

Abbildung 2-8: Wahrscheinlichkeit, dass die zugrunde liegende Erfolgsrate zwischen 80% und 90% liegt

Denken Sie genau darüber nach, wie Sie dies angehen. Was wäre, wenn wir die Fläche hinter 0,80 von der Fläche hinter 0,90 subtrahieren, wie Abbildung 2-9 zeigt?

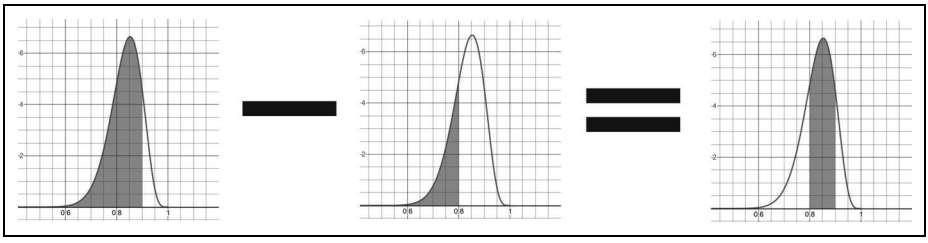

Abbildung 2-9: Die Fläche zwischen 0,80 und 0,90 ermitteln

Würden wir damit die Fläche zwischen 0,80 und 0,90 erhalten? Ja, und es würde sich eine Fläche von 0,3386 oder 33,86 % Wahrscheinlichkeit ergeben. Beispiel 2-6 zeigt, wie sich dies in Python berechnen lässt.

Beispiel 2-6: Mittlere Fläche der Beta-Verteilung mithilfe von SciPy

```
from scipy.stats import beta

a = 8
b = 2

p = beta.cdf(.90, a, b) - beta.cdf(.80, a, b)

# 0.33863336200000016
print(p)
```

Die Beta-Verteilung ist ein faszinierendes Instrument, mit dem sich auf der Grundlage einer begrenzten Anzahl von Beobachtungen die Wahrscheinlichkeit berechnen lässt, dass ein Ereignis auftritt bzw. dass es nicht auftritt. Damit haben wir die Möglichkeit, über Wahrscheinlichkeiten von Wahrscheinlichkeiten nachzudenken. Zudem können wir sie aktualisieren, wenn wir neue Daten bekommen, und darüber hinaus für das Testen von Hypothesen heranziehen. Allerdings befassen wir uns ausführlicher mit der Normalverteilung und der t-Verteilung für diesen Zweck in Kapitel 3.

Beta-Verteilung von Grund auf neu
Um mehr darüber zu lernen, wie sich die Beta-Verteilung von Anfang an implementieren lässt, sollten Sie sich Anhang A zuwenden.

Zum Schluss

Dieses Kapitel hat eine Menge Grundlagenthemen behandelt! Wir beschäftigen uns hier nicht nur mit den grundlegenden Elementen der Wahrscheinlichkeit, ihren logischen Operatoren und dem Satz von Bayes, sondern führen auch Wahrscheinlichkeitsverteilungen ein, einschließlich der binomialen und der Beta-Verteilungen. Im nächsten Kapitel befassen wir uns mit einer der bekannteren Wahrscheinlichkeiten, der Normalverteilung, und ihrer Bedeutung für Hypothesentests.

Wenn Sie mehr über bayessche Wahrscheinlichkeit und Statistik lernen wollen, empfiehlt sich *Bayesian Statistics the Fun Way* von Will Kurt (No Starch Press). Außerdem gibt es interaktive Katacoda-Szenarios, die auf der O'Reilly-Plattform verfügbar sind (*https://oreil.ly/OFbai*).

Übungen

1. Es besteht eine 30%ige Chance, dass es heute regnet, und eine 40%ige Chance, dass Ihre Regenschirmbestellung rechtzeitig eintrifft.

   Sie wollen heute unbedingt im Regen spazieren gehen, können es aber ohne den Schirm nicht tun!

   Wie hoch ist die Wahrscheinlichkeit, dass es regnet UND dass Ihr Regenschirm eintrifft?

2. Es gibt eine 30%ige Chance, dass es heute regnet, und eine 40%ige Chance, dass Ihre Regenschirmbestellung rechtzeitig eintrifft.

   Besorgungen können Sie nur machen, wenn es nicht regnet oder Ihre Regenschirmbestellung eintrifft.

   Wie groß ist die Wahrscheinlichkeit, dass es nicht regnet ODER Ihr Regenschirm eintrifft?

3. Heute besteht eine 30%ige Chance, dass es regnet, und eine 40%ige Chance, dass Ihre Regenschirmbestellung rechtzeitig eintrifft.

   Allerdings haben Sie herausgefunden, dass bei Regen nur eine 20%ige Chance besteht, dass Ihr Regenschirm rechtzeitig eintrifft.

   Wie groß ist die Wahrscheinlichkeit, dass es regnet UND Ihr Regenschirm rechtzeitig eintrifft?

4. Einen Flug von Las Vegas nach Dallas haben 137 Passagiere gebucht. Allerdings startet der Flug in Las Vegas an einem Sonntagmorgen, und Sie schätzen, dass ein Passagier mit 40%iger Wahrscheinlichkeit nicht auftaucht.

   Sie versuchen herauszufinden, wie viele Sitze überbucht sind, damit das Flugzeug nicht leer fliegt.

   Wie wahrscheinlich ist es, dass mindestens 50 Passagiere nicht eintreffen?

5. Eine Münze haben Sie 19-mal geworfen und dabei 15-mal Kopf und 4-mal Zahl erhalten.

   Denken Sie, dass diese Münze mit einigermaßen guter Wahrscheinlichkeit fair ist? Warum oder warum nicht?

Die Antworten finden Sie in Anhang B.

Inhalt

O'REILLY®

Mathe-Basics für Data Scientists

Um als Data Scientist erfolgreich zu sein, müssen Sie über ein solides mathematisches Grundwissen verfügen. Dieses Buch bietet einen leicht verständlichen Überblick über die Mathematik, die Sie in der Data Science benötigen. Thomas Nield führt Sie Schritt für Schritt durch Bereiche wie Infinitesimalrechnung, Wahrscheinlichkeit, lineare Algebra, Statistik und Hypothesentests und zeigt Ihnen, wie diese Mathe-Basics beispielsweise in der linearen und logistischen Regression und in neuronalen Netzen eingesetzt werden. Zusätzlich erhalten Sie Einblicke in den aktuellen Stand der Data Science und erfahren, wie Sie dieses Wissen für Ihre Karriere als Data Scientist nutzen.

- Verwenden Sie Python-Code und Bibliotheken wie SymPy, NumPy und scikit-learn, um grundlegende mathematische Konzepte wie Infinitesimalrechnung, lineare Algebra, Statistik und maschinelles Lernen zu erkunden
- Verstehen Sie Techniken wie lineare und logistische Regression und neuronale Netze durch gut nachvollziehbare Erklärungen und ein Minimum an mathematischer Terminologie
- Wenden Sie deskriptive Statistik und Hypothesentests auf einen Datensatz an, um p-Werte und statistische Signifikanz zu interpretieren
- Manipulieren Sie Vektoren und Matrizen und führen Sie Matrixzerlegung durch
- Vertiefen Sie Ihre Kenntnisse in Infinitesimal- und Wahrscheinlichkeitsrechnung, Statistik und linearer Algebra und wenden Sie sie auf Regressionsmodelle einschließlich neuronaler Netze an
- Erfahren Sie, wie Sie Ihre Kenntnisse und Fähigkeiten in der Datenanalyse optimieren und gängige Fehler vermeiden, um auf dem Data-Science-Arbeitsmarkt zu überzeugen

»Eine Ressource mit vielen klaren, praktischen Beispielen für die Grundlagen, die man benötigt, um Daten zu verstehen und mit ihnen zu arbeiten.«

— **Vicki Boykis**
Senior Machine Learning Engineer bei Tumblr

»Eine solide Basis für das Verständnis der mathematischen Instrumente der Data Science.«

— **Mike X Cohen**
sincXpress

Thomas Nield ist Gründer der Nield Consulting Group sowie Dozent an der University of Southern California und bei O'Reilly Media. Es macht ihm Freude, technische Inhalte insbesondere all jenen, die sich zunächst von ihnen abgeschreckt fühlen, verständlich zu erklären. Thomas Nield unterrichtet regelmäßig Kurse zu Datenanalyse, Machine Learning, mathematischer Optimierung und praktischer künstlicher Intelligenz.

9 783960 092155

www.oreilly.de

Euro 39,90 (D)
ISBN 978-3-96009-215-5