

Die exakte Wissenschaft des Pi-mal-Daumens

Wenn man etwas nicht so ganz genau weiß, muss man schätzen. Falls man aber eine präzise Schätzung haben möchte, sollte man es mit der richtigen Mathematik probieren.

» [spektrum.de/artikel/2099886](https://www.spektrum.de/artikel/2099886)

Kürzlich hatte ich Glück. Es gab keine Weihnachtskekse mehr, und die von den Feiertagen übrig gebliebene Schokolade war ebenfalls schon aufgegessen. Doch in der hinteren Ecke eines Küchenschrank habe ich noch eine Packung bunter Schokolinsen gefunden. Der erste Griff in die Tüte lieferte eine grüne, zwei rote und sieben blaue Süßigkeiten. Eigentlich war mir das egal, denn der Farbstoff ist geschmacklos. Trotzdem habe ich mich über die blaue Übermacht in meiner Stichprobe gewundert und mich gefragt, wie wahrscheinlich so ein Ereignis ist. Und ich habe mich entschieden, dem Ganzen mathematisch nachzugehen. Dafür kann man diese Formel verwenden:

$$\mathbb{P}(k, \lambda, n, N) = \frac{\binom{\lambda N}{k} \binom{N(1-\lambda)}{n-k}}{\binom{N}{n}}$$

Für das genannte Beispiel bezeichnet N die Gesamtzahl der Schokolinsen in der Packung. Der Anteil der blauen Linsen wird durch λ beschrieben, die Anzahl der bei der Stichprobe herausgegriffenen Süßigkeiten durch n und die darin gefundenen blauen Exemplare durch k . Setzt man die entsprechenden Werte ein, ergibt sich daraus die Wahrscheinlichkeit \mathbb{P} , dass man unter den n Linsen exakt k blaue findet.

Das Problem daran: Der Anteil der blauen Schokostückchen λ ist unbekannt, dabei ist es genau dieser Wert, der mich interessiert. Das macht aber nichts, denn dafür kann man die »Maximum-Likelihood-Methode« verwenden. Vereinfacht gesagt probiert man verschiedene Werte für λ aus (bei einer festen Zahl N) und sieht nach, wann die Chance, unter n Linsen genau k mit der Farbe blau zu finden, maximal wird.

Im Fall der Schokolinsen hätte ich natürlich alle nach Farben sortieren und zählen können – so groß war die Tüte nicht. Doch es gibt durchaus Situationen, in denen man auf Schätzungen angewiesen ist. Ein historisch interessantes Beispiel ist das »German tank problem«: Im Zweiten Weltkrieg wollten die Alliierten wissen, wie viele Panzer die Wehrmacht baute. Diese Information hatten sie natürlich nicht, dafür aber zumindest ein paar erbeutete Panzer. Die dort verwendeten Bauteile waren mit Seriennummern versehen. So war es möglich, die Anzahl der produzierten Teile zu schätzen. Hat man etwa vier Panzer mit den Seriennummern 20, 31, 44 und 61, kann man sich fragen: Für welche Gesamtzahl an Panzern ist es am wahrscheinlichsten, bei einer Stichprobe von vier Stück keine Nummer zu finden, die größer als 61 ist?

In der Realität war die Sache natürlich komplexer: Unterschiedliche Bauteile der Panzer hatten verschiedene Seriennummern, nicht alle Modelle wurden fortlaufend nummeriert und so weiter. Am Ende waren die mathematischen Schätzungen jedoch überraschend gut.

Schätzmethoden wie das Maximum-Likelihood-Prinzip werden aber nicht nur für Geheimdienstoperationen im Krieg verwendet. Man benutzt sie in der Biologie, um Verwandtschaftsbeziehungen zwischen Lebewesen zu bestimmen, bei Datenanalysen und Vorhersagen in Sozial- und Wirtschaftswissenschaften oder bei Algorithmen zum maschinellen Lernen.

Bei meinen Schokolinsen habe ich am Ende auf eine Maximum-Likelihood-Schätzung verzichtet und sie einfach alle aufgegessen. Die Farbe hat ja keinen Einfluss auf den Geschmack – ganz anders als bei Gummibärchen, bei denen die Weißen definitiv besser schmecken als die anderen. Keine noch so komplexe Formel wird mich je vom Gegenteil überzeugen können.



Florian Freistetter
ist Astronom, Autor und Wissenschaftskabarettist bei den »Science Busters«.