

Gruppen-Screening Ein Ausflug in die Medizin über die Stochastik, Analysis und CAS

J.-H. Müller
(Rivius-Gymnasium der Stadt Attendorn)

jan.mueller@math.uni-dortmund.de



Zusammenfassung

Am Ende der Sekundarstufe II verfügen Schülerinnen und Schüler im Fach Mathematik über viele Begriffe und Strategien in den Bereichen Analysis und Stochastik. Mit diesem Beitrag soll anhand eines Anwendungsbeispiels gezeigt werden, wie im Rahmen einer integrativen Wiederholung für das Abitur erlerntes Wissen wiederholt, vernetzt und die Benutzung eines CAS sinnvoll in den Unterricht eingebettet werden kann.

Die Idee des Gruppenscreenings

Soll eine Menge von Blutproben auf einen Krankheitserreger hin getestet werden, stellt sich unter anderem die Frage, ob jede Blutprobe einzeln zu testen ist oder ob es effektivere Strategien gibt, wenn man die Verbreitung der Krankheit in der Population mit in Betracht zieht. Wird etwa das Blut von 100 Personen einer Population einzeln untersucht von denen nur eine Person mit einem Krankheitserreger infiziert ist, dann sind maximal 100 Blutproben zu untersuchen. Mischt man jedoch das Blut von jeweils 20 Blutproben und testet anschließend das Blut dieser fünf Mischung auf den Krankheitserreger, dann sind zunächst $100/20 = 5$ Tests erforderlich. Da genau eine dieser fünf Mischungen infiziertes Blut enthält, sind erneut maximal 20 Einzeltests erforderlich. Insgesamt sind also maximal $5 + 20 = 25$ Tests erforderlich, um die infizierte Person herauszufinden. Also insgesamt deutlich weniger Tests als zuvor! Es liegt demzufolge die Frage nahe, ob sich die erforderliche Testanzahl durch eine Variation der Blutprobenanzahl in einer Blutmischung noch weiter minimieren lässt. Eine typische schulnahe Herangehensweise erfolgt über die Infinitesimalrechnung: Wird das Blut (gemäß der Vorgabe auf dem Arbeitsblatt auf Seite 21) von 100 Soldaten gemischt und beschreibt k die Anzahl der Blutproben in einer Blutmischung, dann beschreibt der aufgerundete Wert von $100/k$ die Anzahl

der erforderlichen Tests für alle Blutmischungen und k die maximale Anzahl der erforderlichen Tests für die in einer Blutmischung enthaltenen Blutproben. Der aufgerundete Wert der Folge $t(k) = \frac{100}{k} + k$ beschreibt demnach die maximale Anzahl der insgesamt erforderlichen Tests. Nimmt man k positiv und reellwertig an (was durch die Option des Aufrundens legitimiert werden kann), so erhält man mit den üblichen schulischen Methoden der Analysis (hilfsmittelfrei!) ein globales Minimum von $t(k)$ für $k = 10$. Um eine infizierte Person herauszufinden, sind auf diese Weise für eine Population von 100 Personen also maximal 20 Tests erforderlich.

Verschiedene Modellierungen

Bei einer Prävalenz¹ von 2% wird die Situation komplexer: Betrachtet man wieder eine Population von 100 Personen, dann kann sich das Blut der zwei infizierten Personen in der selben oder in verschiedenen Blutmischungen befinden. In einer solchen Situation kann die Idee des Erwartungswertes hilfreich sein. Hierfür könnte man annehmen, dass jeder Soldat mit einer Wahrscheinlichkeit von 2% infiziert² ist. Die Wahrscheinlichkeit, dass eine Blutmischung mit zwei Blutproben nicht infiziert ist beträgt demnach $0,98^2 \approx 0,9604$. In diesem Fall ist für eine Blutmischung wie zuvor genau ein Test erforderlich. Die Wahrscheinlichkeit, dass eine Blutmischung infiziertes Blut enthält beträgt $1 - 0,98^2$. In diesem Fall sind genau drei Tests erforderlich (einen Test für die Gruppe und zwei Tests für die Einzeluntersuchungen). Auf lange Sicht sind dies bei 100 Soldaten also durchschnittlich

$$100/2 \cdot (1 \cdot 0,98^2 + 3 \cdot (1 - 0,98^2)) \approx 54$$

Tests. Verallgemeinert ergibt sich für k Blutproben in einer Blutmischung, einer Prävalenz von 2% und 100 Soldaten eine zu erwartende Testanzahl $t_1(k)$ von

$$t_1(k) = 100/k \cdot (0,98^k + (k + 1) \cdot (1 - 0,98^k))$$

¹So nennt man den Anteil der Verbreitung einer Krankheit in einer Population

²Man beachte bei dieser Annahme, dass die Wahrscheinlichkeit infiziert zu sein unabhängig ist!

Eine alternative Modellierung

Die Erfahrung zeigt, dass Unterrichtsgespräche auch andere Ansätze hervorbringen. Das vorliegende Problem lässt sich alternativ aber auch mit dem Ziehen ohne Zurücklegen aus einer Urne modellieren³ (was den Lernenden schließlich anhand vieler Übungsaufgaben zuvor vermittelt wurde und hier sinnvoll genutzt werden kann): Die Blutproben werden durch 100 Kugeln in einer Urne repräsentiert, von denen zwei markiert (infiziert) sind. Die Wahrscheinlichkeit, dass eine Mischung von zwei Blutproben im vorliegenden Fall nicht infiziert ist, kann mit

$$\binom{98}{2} / \binom{100}{2} \approx 0,9602$$

berechnet werden (also minimal unwahrscheinlicher als im ersten Modell). In diesem Fall ist wieder genau ein Test erforderlich. Drei Tests sind demzufolge mit einer Wahrscheinlichkeit von etwa $1 - 0,9602$ zu erwarten. Auch in diesem Modell sind dies auf lange Sicht bei 100 Soldaten wieder durchschnittlich

$$100/2 \cdot (1 \cdot 0,9602 + 3 \cdot (1 - 0,9602)) \approx 54$$

Tests. Verallgemeinert auf k Blutproben in einer Blutmischung, einer Prävalenz von 2% und 100 Soldaten ergibt sich für $t_2(k)$ der Term

$$\frac{100}{k} \cdot \left(1 \cdot \frac{\binom{98}{k}}{\binom{100}{k}} + (k+1) \cdot \left(1 - \frac{\binom{98}{k}}{\binom{100}{k}} \right) \right)$$

Wie kann es weitergehen?

Es ist ein Minimum der Modelle t_1 oder t_2 zu finden. Im ersten Modell kann k als reellwertig angenommen werden und im Unterricht können demnach bekannte Mittel der Differentialrechnung genutzt werden. Wie kann es aber im zweiten Modell weiter gehen? Denn hier ist der Binomialkoeffizient nur für natürliche Werte von k definiert. Es kann zunächst also nur geprüft werden, ob sich die Funktionsterme vereinfachen lassen. Ein CAS bietet hierbei die Möglichkeit experimentell zu arbeiten: Vom Lehrer kann eine Auswahl an Befehlen angeboten werden, anhand derer (oder selbst gefundener weiterer) ein vorliegender Funktionsterm vereinfacht werden soll. Abb. 1 zeigt anhand von Beispielen, dass hierbei die Reihenfolge des Befehlseinsatzes Termumformungen erwartungsgemäß stark beeinflussen kann. Aus Abb. 1 wird ebenfalls ersichtlich, dass sich der vereinfachte Term zu t_2 reellwertig erweiterbar und zudem differenzierbar erweist. Damit hilfsmittelfreie Fähigkeiten ebenfalls geübt

³Mit dieser Modellannahme wird auf die Annahme verzichtet, dass die Wahrscheinlichkeit infiziert zu sein unabhängig ist. Insofern erscheint es realitätsnäher.

⁴Im vorliegenden Fall erscheint z. B. das abhängige (Urnen-) Modell als realitätsnäher, da sich Syphilis durch ungünstige Hygienebedingungen sicher auch innerhalb der Population verbreitet.

⁵Wenn zum Beispiel ein einfaches rechnerisches Modell in etwa gleiche Werte wie ein deutlich komplexeres aber realitätsnäheres Modell liefert.

werden, besteht die Möglichkeit ausgewählte gefundene Termäquivalenzen begründen zu lassen. Dies kann zum Beispiel als Hausaufgabe erfolgen. Der vorherige Einsatz von CAS erweist sich hierbei wiederum als hilfreich, da das Ergebnis der Umformung vorgegeben ist.

```
(%i1) t2:100/k*(binomial(98,k)/binomial(100,k)+(k-1)*(1-binomial(98,k)/binomial(100,k)));
100 * ( 1 - (98 choose k) / (100 choose k) ) * (k+1) + (98 choose k) / (100 choose k)
(%o1) -----
k

(%i2) expand(minfactorial(makefact(t2)));
k^2 - 199 k + 100
(%o2) -----
99 - 99 k + k

(%i3) expand(makefact(minfactorial(t2)));
100 * (100 - k)!
(%o3) -----
k * 99 * (98 - k)! * 100

(%i4) rat(minfactorial(makefact(t2)));
k^3 - 199 k^2 - 9900
(%o4) R/ -----
99 k
```

Abbildung 1: Termumformung mit dem CA-System wxMaxima.

Beide Modelle bieten also die Möglichkeit mit den Methoden der Differentialrechnung untersucht zu werden. Aber für welches Modell sollten sich die Schülerinnen und Schüler entscheiden? Genauer: Anhand welcher Kriterien sollte entschieden werden? Neben einer inhaltlichen Analyse⁴ (Welches Modell beschreibt die Realität „besser“?) kann ebenso eine „Kosten-Nutzen“-Analyse⁵ oder die Frage nach einer Verallgemeinerung des Problems hilfreich sein. Eine im Mathematikunterricht zunächst naheliegende Methode ist der graphische Vergleich der konkurrierenden Modelle (auch hierfür kann ein CAS nützlich sein!). Abb. 2 zeigt, dass die beiden Modelle t_1 und t_2 im Bereich der optimalen Blutprobenanzahl k zwischen 5 und 10 kaum variieren und ein Optimum bei etwa 7 Blutproben pro Blutmischung vorliegt.

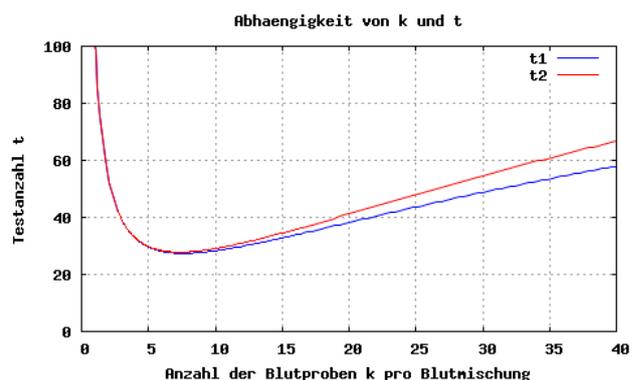


Abbildung 2: Modellvergleich mit wxMaxima.

Auch zu Übungszwecken könnte diese Einsicht nun rechnerisch von den Schülerinnen und Schülern bestätigt werden.

Den Forscherdrang fördern

Das vorliegende Problem bietet im Zusammenhang mit der Frage „Welches Modell ist besser?“ viele Möglichkeiten für selbständiges Weiterforschen oder (falls noch Zeit ist) zur Bearbeitung im Unterricht. Offen sind nach wie vor Fragen, inwiefern sich gefundene Modelle für eine Verallgemeinerung der Prävalenz eignen. Im sehr lesenswerten Beitrag [1] wird zum Beispiel gezeigt, dass sich anhand von Modell t_1 das Optimum der Gruppengröße k unabhängig von der vorliegenden Populationsgröße lediglich durch die Prävalenz p mittels

$$k = \frac{1}{\sqrt{p}}$$

annähern lässt. Neben der vorgestellten Art der Gruppenbildung sind natürlich auch noch andere Strategien denkbar, die in [2] diskutiert werden. Die beiden in diesem Beitrag vorgestellten Modelle stehen selbstverständlich ebenfalls nur für eine weitere Vielzahl an Ideen, die Schülerinnen und Schüler im Mathematikunterricht diskutieren können. Probieren Sie es einmal aus!

Literatur

- [1] Henn, H.-W., Humenberger, H. Gruppenscreening – ein Paradebeispiel für Vernetzungsmöglichkeiten im Mathematikunterricht. *Biehler, R., J. Engel u. J. Meyer (Hrsg.): Neue Medien und innermathematische Vernetzungen in der Stochastik. Anregungen zum Stochastikunterricht*, Franzbecker, Hildesheim, 2:19–32, 2004. Im Internet verfügbar unter: homepage.univie.ac.at/hans.humenberger/Aufsaeetze/0090.pdf (Zugriff am 15.2.2015)
- [2] Henn, H.W., Jock, W. Gruppenscreening. *ISTRON (Band 6): Materialien für einen realitätsbezogenen Mathematikunterricht*, Franzbecker, Hildesheim, 123–137, 2000.
- [3] Dorfman, R. The detection of defective members of large populations. *The Annals of Mathematical Statistics*, 14:436–440, 1943.

mathemas ordinate  www.ordinate.de

 0431 23745-00/  -01, info@ordinate.de → Software for mathematical people !

 **Mathematische Software u. Consulting, MathType, Optica, ExtendSim, KaleidaGraph, Intel-Software, Fortran, NSBasic, @Risk, Chemistry, Satellitensteuerung u.a.** $\infty + \mu < \heartsuit$

mathemas ordinate, Dipl. Math. Carsten Herrmann, M. Sc.
Königsbergerstr. 97, 24161 Altenholz

Fast 30 Jahre Erfahrung mit Software-Distribution !

$$\int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$