

Die Wordle-Entropie

$$p(\text{😊} | \text{🎲}) = 1$$

Ein Unterrichtsentwurf aus der Reihe **Probably Fun – Games to teach Statistics**

von Dr. Kristian Rother (www.academis.eu/probably_fun/)

Nutzbar unter den Bedingungen der Creative Commons Attribution Share-Alike License 4.0

Unterrichtsziel

Die Teilnehmer berechnen die Entropie der Information aus einer Wortliste für Wordle.

Zeit

5 - 90 Minuten

Begriffe

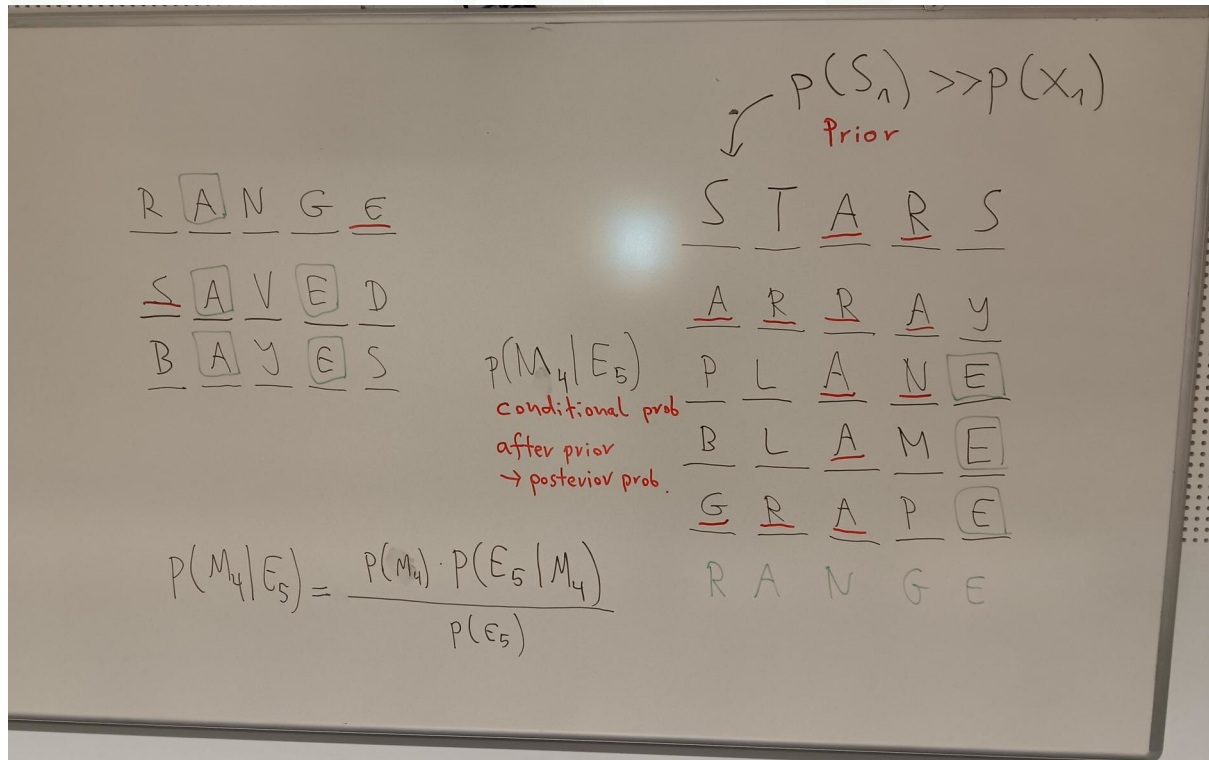
- Shannon-Entropie
- Mutual Information



Nutzbar unter den Bedingungen
der CC-BY-SA 4.0 Lizenz

$$p(\text{😊} | \text{🎲}) = 1$$

Das Spiel: Wordle



Wordles aus dem Unterricht

Wordle ist ein beliebtes Ratespiel irgendwo in der Mitte zwischen Galgenmännchen und MasterMind. Populär wurde es als Handy-App. Die Spieler raten ein Wort mit 5 Buchstaben. Sie erfahren, welche Buchstaben im gesuchten Wort vorhanden sind, und welche sogar an der richtigen Stelle sind. Dann folgt der nächste Versuch.

Christina von der Brettspielakademie brachte die Idee ein, ein Wordle als Energizer im Unterricht einzusetzen. Das funktioniert auch im Statistikunterricht, natürlich mit Wörtern zum Thema, z.B.:

RANGE
PIVOT
BAYES
TABLE

Als Energizer ist Wordle sehr leicht umzusetzen. Du malst 5 Striche an die Tafel und los geht's. Es gibt aber auch statistische Werkzeuge, um Wordles zu beschreiben.



Nutzbar unter den Bedingungen
der CC-BY-SA 4.0 Lizenz

$$p(\text{😊} | \text{🎲}) = 1$$

Unterrichtsablauf

Um tiefer in das Wordle-Thema einzutauchen wollen, lässt sich etwas Informationstheorie anwenden. Dreh- und Angelpunkt ist ein guter Datensatz wie die SOWPODS-Liste, eine offizielle Liste aller erlaubten Scrabble-Wörter auf Englisch. Um mit der Wörterliste zu arbeiten, solltest du alle Wörter mit 5 Buchstaben herausfiltern und in eine fünfspaltige Tabelle überführen. Im Englischen enthält die Liste immerhin 12478 Wörter, was für eine Tabellenkalkulation aber gut zu schaffen ist. Auch eine Programmiersprache kann hier zum Einsatz kommen; die Berechnung ist nicht sehr schwierig.

Um optimale Wörter zum Raten zu finden, kann man zunächst einmal nach den häufigsten Buchstaben für jede Position suchen. Interessanter wird es, die **Shannon-Entropie** jeder Buchstabenposition zu berechnen. Hier die Gleichung:

$$H(X) = - \sum p(x) \log_2 p(x)$$

Es lohnt sich, die Entropie für ein paar einfache Beispiele an der Tafel zu bestimmen. Auch die Einheit (bit) sollte besprochen werden, da auch ein Informatiker nicht ganzzahligen Bits nicht unbedingt gleich etwas abgewinnen kann.

Im nächsten Schritt wird die Liste gefiltert: *“Wenn wir wissen, dass in Position 1 ein A ist, wie verändert sich die Entropie?”* Der pragmatische Coding-Junkie wird vermutlich eine brute force Heuristik losballern wollen, die alle Buchstaben durchprobiert und eine Kombination mit der kleinsten Entropie sucht. Das ist sicher zielführend, auch wenn die optimale Lösung etwas komplizierter ist (siehe Link).

Zur weiteren Vertiefung lohnt sich die Berechnung der **Mutual Information**: *“Wie viele Bit Information verrät uns Position 1 über Position 2?”* Auf technischer Ebene lässt sich damit eine Matrix aller 5 Positionen gegeneinander berechnen. Da die Mutual Information mit der Kullback-Leibler Divergenz eng verwandt ist, ist eine weitere Vertiefung in Richtung Machine Learning möglich. Hier noch einmal die Formel:

$$I(X; Y) = \sum_x \sum_y p(x, y) \log_2 \left(\frac{p(x, y)}{p(x) p(y)} \right)$$

Entropy und Mutual Information unterscheiden sich in unterschiedlichen Sprachen übrigens sehr. Die Untersuchung hat dann aber schon Projektcharakter.



Nutzbar unter den Bedingungen
der CC-BY-SA 4.0 Lizenz

$$p(\text{😊} | \text{🎲}) = 1$$

Hier ist ein einfacher Unterrichtsablauf:

Schritt	Aktivität	Zeit
1.	Spielt ein Wordle mit einem Begriff, der bereits im Unterricht dran war	5'
2.	Lade den SOWPODS-Datensatz in eine Tabelle und extrahiere Wörter mit 5 Buchstaben	10'
3.	Berechne relative Häufigkeiten für jeden Buchstaben und jede Position	15'
4.	Berechne die Shannon-Entropie für jede Position	15'
5.	Wende eine Bedingung an (z.B. "A in Position 1") und berechne die Entropien erneut für die verbliebenen Wörter	15'
6.	Berechnet die Mutual Information von zwei Spalten	15'
7.	Diskutiert mögliche Gewinnstrategien für Wordle	10'

Reflexionsfragen für den Unterricht

- Was ist gute Gewinnstrategie für Wordle?

Links

- SOWPODS-Liste der erlaubten Scrabble-Wörter
www.freescrabledictionary.com/sowpods/download/sowpods.txt
- Wordle auf Wikipedia
en.wikipedia.org/wiki/Wordle
- Using Information Entropy to solve Wordle
tomrocksmaths.com/wp-content/uploads/2023/07/using-information-entropy-to-e28098solve-wordle.pdf



Nutzbar unter den Bedingungen
der CC-BY-SA 4.0 Lizenz

$$p(\text{😊} | \text{🎲}) = 1$$