Tutorium:Statistische Methoden in der Sprachverarbeitung

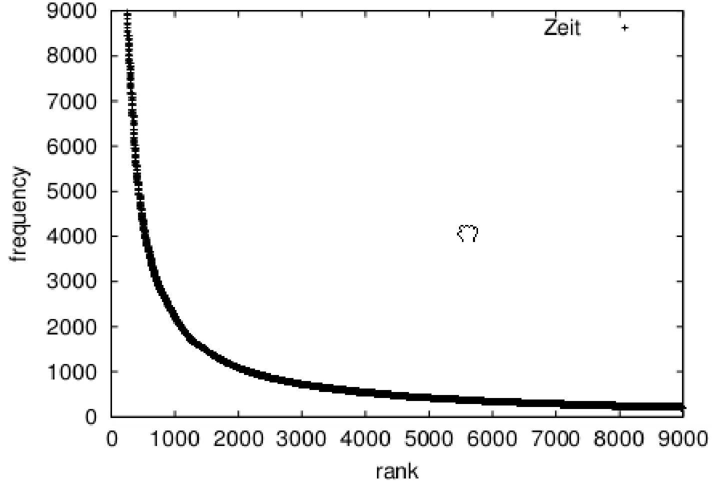
Organisatorisches

- Website: https://tutorium-statistische-methode-ws2021.github.io/
- Tutorin: Suteera Seeha (s.seeha@campus.lmu.de)
- Fragensammlung: https://docs.google.com/document/d/1hSTtDnCD4haLUXybEkNCLI_tYMLDx-FZaZDUp34aP9U/edit?usp=sharing
- Wunschliste: https://docs.google.com/document/d/1KWWZQjQr_h4n8rptCKLz67phYf0XHoNiM44-MTAWvig/edit?usp=sharing

- Inhalt: Vorlesungsstoff wiederholen, Frage beantworten zur Vorlesungsstoff und Übung

Zipf's Gesetz:

$$f \sim \frac{1}{r} \quad (f \cdot r \approx K)$$



- sagt über die Verteilung von Wörter in einem Korpus
- sagt über die Beziehung zw. Wortrang und Worthäufigkeit
- Die Worthäufigkeit ist also proportional zum Kehrwert des Wortranges
- Wenige Wörter sind sehr häufig
- Die meisten Wörter sind sehr selten

Zipf's Gesetz

Wort	f	r	$f \cdot r$
the	3332	1	3332
and	2972	2	5944
a	1775	3	5235
he	877	10	8770
but	410	20	8400
be	294	30	8820
two	104	100	10400
turned	51	200	10200
you'll	30	300	9000
family	8	1000	8000
brushed	4	2000	8000
sins	2	3000	6000
Could	2	4000	8000
Applausive	1	8000	8000

Zipf's Gesetz:

$$f \sim \frac{1}{r} \quad (f \cdot r \approx K)$$

$\Rightarrow$ Wenige Wörter sind sehr häufig

$\Rightarrow$ Die meisten Wörter sind sehr selten.

Beispiel von Klausuraufgabe

Aufgabe 1) Was sagt das **Zipf'sche Gesetz** aus? (2 Punkte)

Aufgabe 2) Was ist eine **Zufallsvariable**? Was haben Zufallsvariablen mit der Berechnung eines Notendurchschnittes zu tun? (Denken Sie an die Noten sehr gut, gut, befriedigend, ausreichend, mangelhaft, ungenügend.) (2 Punkte)

Aufgabe 3) Wie wird der **Erwartungswert** für die Zahl der Augen eines fairen Würfels berechnet? (2 Punkte)

Aufgabe 4) Wie lautet das **Theorem von Bayes**? Nennen Sie ein Beispiel, wo wir dieses Theorem benutzt haben. (2 Punkte)

Zufallsexperiment

In der Statistik geht es um die Wahrscheinlichkeit von Ereignissen:

Beispiel: Wie wahrscheinlich ist es, sechs Richtige im Lotto zu haben?

Zufallsexperiment: Experiment (Versuch) mit mehreren möglichen Ausgängen (*Wurf von zwei Würfeln*)

Ergebnis: Resultat eines Experimentes (*3 Augen auf Würfel 1 und 4 Augen auf Würfel 2*)

Ergebnisraum: Menge aller möglichen Ergebnisse

Ereignis: Teilmenge des Ergebnisraumes (*7 Augen auf zwei Würfeln*)

Elementarereignis: anderes Wort für Ergebnis

Stichprobe: Folge von Ergebnissen bei einem wiederholten Experiment

- **Wahrscheinlichkeit:**
    - We use probability to measure the chance or likelihood of an event (or events) occurring in the future.
    - Wie wahrscheinlich ist es, beim Wurf eines Würfels Kopf zu bekommen
    - Wie wahrscheinlich ist es, dass ein bestimmtes Wort vorkommt gegeben die vorherigen Wörter.
    - Wie wahrscheinlich ist es, dass das Wort “gehen” den POS-Tag “VERB” oder “NOUN” hat
- **Zufallsexperiment:** Experiment / observation / process that has more than one possible outcomes
    - roll a die: 1,2,3,4,5,6
    - flip a coin: head, tail
    - take an exam: grade 1,2,3,4,5,6
    - observe the weather tomorrow: sunny, rainy, cloudy, snowy
    - observe words and their frequencies in a corpus: “the”: 5 times, “dog”:1 times, “eat”: 3 times, etc.
- **Ergebnis/Elementarereignis (outcome):** an outcome of the experiment
    - pick a ball from a box containing 3 red balls and 2 blue balls : outcome are red and blue ball
- **Ergebnisraum (sample space):** a set of all possible outcomes of an experiment
    - roll a die: {1,2,3,4,5,6}
    - flip a coin: {head, tail}
    - take an exam: grade {1,2,3,4,5,6}
    - observe the weather tomorrow: {sunny, rainy, cloudy, snowy}
    - pick a ball from a box containing 3 red balls and 2 blue balls : {red ball, blue ball}
- **Ereignis(event):** a subset of the sample space (Ergebnisraum). We usually intested in finding the probability of events
    - number on the die is even: { 2,4,6}
    - number on the die greater than 3: {4,5,6 }
    - number on the die is equal to 5: {5} , number on the die is **not** equal to 5: {1,2,3,4,6}
- **Stichprobe (sample):** Folge von Ergebnisse bei einem wiederholten Experiment
    - roll a die 10 times and observe the number: 2,1,1,3,4,6,6,6,6,5
- **Note:**
    - **Ergebnis, Ergebnisraum, Ereignis** can be defined without doing any actual experiments
    - To get a **Stichprobe,** you need to do actual experiments and collect the outcomes. Doing experiments does not affect the sample space in any way.

Definition

A random experiment is a mechanism that produces a definite outcome that cannot be predicted with certainty. **The sample space** associated with a random experiment is the set of all possible outcomes. **An event** is a subset of the sample space.

ref: https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_Introductory_Statistics_(Shafer_and_Zhang)/03%3A_Basic_Concepts_of_Probability/3.01%3A_Sample_Spaces%2C_Events%2C_and_Their_Probabilities

Statistical Experiment

All statistical experiments have three things in common:

- The experiment can have more than one possible outcome.
- Each possible outcome can be specified in advance.
- The outcome of the experiment depends on chance.

ref: https://stattrek.com/statistics/dictionary.aspx?definition=statistical_experiment

To compute the probablity of an event, we **sum** probablities of outcomes that define the event. Probablity distribution **p** is a function that maps each outcomes to a value between 0 and 1(the probablity).

The sum of the probabilities of all outcomes must be equal to 1.

Wahrscheinlichkeitsverteilung

Wahrscheinlichkeitsverteilung: Funktion, die jedem Ergebnis o einen Wert zwischen 0 und 1 zuweist, so dass

$$\sum_o p(o) = 1$$

Die Wahrscheinlichkeit eines Ereignisses ist die Summe der Wahrscheinlichkeiten der entsprechenden Ergebnisse.

Beispiel:
Wahrscheinlichkeit, dass die Zahl der Augen beim Wurf eines Würfels gerade ist

$$p(e) = p(2) + p(4) + p(6) = \frac{3}{6} = \frac{1}{2}$$

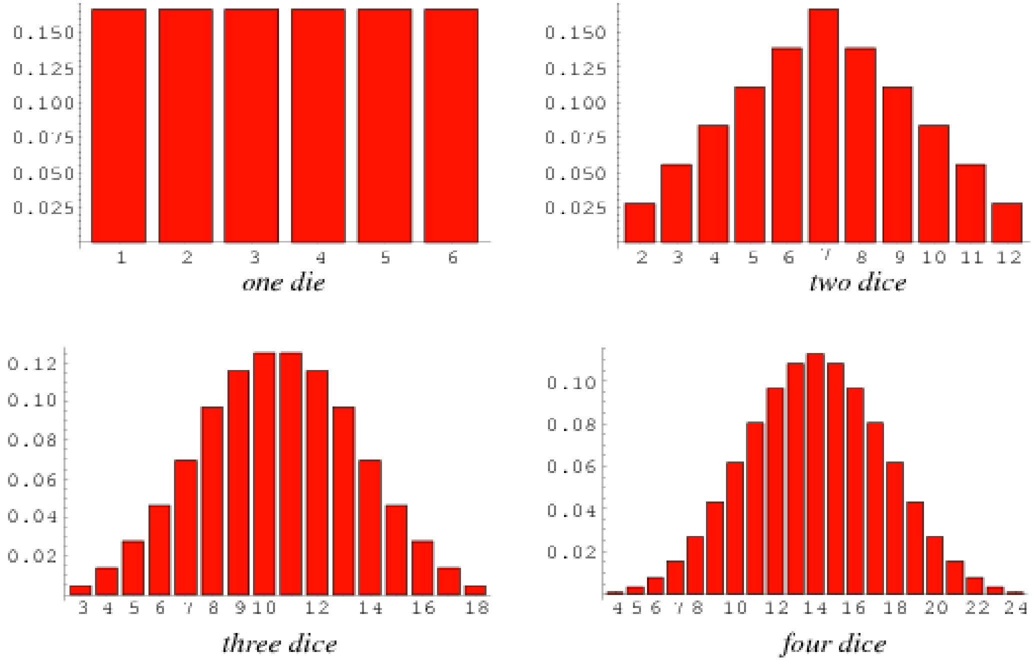
p of an outcome in case of rolling a fair die is easy to compute. If we know that each outcome is likely to occur, then p(outcome) is "1 / number of all outcomes" (In this example, 1/6). We called this uniform distribution.

If the die has been modified so that the side with 1 dot is heavier than other sides, p(1) increases while p(other side) decreases.
Or if we add another 6 side to the die, p(6) will increase:
p(6) = **2/7**
p(1) = 1/7
p(2) = 1/7
p(3) = 1/7
p(4) = 1/7
p(5) = 1/7

In other more complicated problems, we might not be able to compute the probablity easily like this, so we will have to use other methods to estimate the probability.

Different types of probablity distributions

- Roll one die and count the number of dots.
- All outcomes have the same probablity.



- Roll two die and count the number of dots
- outcome 7 has the highest probability.

Übung:

1. Experiment: Werfen einen Würfel und betrachte die Augenzahl
    a. Ergebnisraum: {1,2,3,4,5,6}
    b. Ereignisse aus dem Ergebnisraum: {2,5}, {1}, {2,4,6}, {1,2,3,4,5,6}, {4,5,6}
    c. Nenne ein Beispiel von einer Stichprobe:
        - 10 mal das Experiment wiederholen ergibt z.B. Stichprobe: 1,1,1,2,3,4,5,6,3,3

2. Experiment: Werfen 2 Würfeln und betrachte die Augenzahl von beiden
    a. Ergebnisraum: {(1,1),(1,2),(1,3),(1,4)(1,5),(1,6), ….}
    b. Nenne 2 Ereignisse aus dem Ergebnisraum:
        - {(1,2), (1,3)}
        - Summe ist gleich 7 : {(1,6), (2,5), (3,4)}
    c. Ein Beispiel von einer Stichprobe:
        - 3 mal das Experiment wiederholen ergibt z.B. Stichprobe: (1,6), (2,5), (3,4)

3. Experiment: Beobachte das Wetter :
    a. Ergebnisraum: {sunny, rainy, cloudy}
    b. Nenne 1 Ereignisse aus dem Ergebnisraum: {sunny}
    c. Ein Beispiel von einer Stichprobe:
        - 4 mal das Experiment wiederholen ergibt z.B. Stichprobe: sunny, rainy, cloudy, cloudy

4. Experiment: Gegeben einen Sack mit 3 roten Bällen, 4 schwarzen, nimmt einen Ball daraus
    a. Ergebnisraum: {rot, blau}
    b. Nennen 1 Ereignisse aus dem Ergebnisraum: {rot}
    c. Ein Beispiel von einer Stichprobe:
        - 4 mal das Experiment wiederholen ergibt z.B. Stichprobe: rot, blau, blau, blau

Bedingte Wahrscheinlichkeit

Bedingte Wahrscheinlichkeit: Wahrscheinlichkeit eines Ereignisses A, wenn das Ereignis B bekannt ist:

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

$\cap$ ist der Schnittmengenoperator.

Übung 3 (aus der Vorlesungfolien)
Wahrscheinlichkeit, dass die Augenzahl eines Würfels gerade ist, wenn die Augenzahl größer als 3 ist

Lösung: **2/3**

Übung 1
Beim Wurf eines Würfel, was ist die Wk dass die Augenzahl ist 4, wenn wir wissen, die Augenzahl ist gerade.

Experiment: Werfen einen Würfel
Ergebnisraum: {1,2,3,4,5,6}
Ereignisse: A={4}, B={2,4,6}
Wir suchen P(A | B), also wir müssen P(A ∩ B) und P(B) ausrechnen
P(A ∩ B)= P({4}) = p(4) =1/6 (angenommen der Würfel ist fair)
P(B) = P({2,4,6}) = p(2)+p(4)+p(6) = 3 *1/6
P(A ∩ B) / P(B) = (1/6) / (3 /6) = **1/3**
- Das Vorwissen, dass die Augenzahl gerade ist, erhöht die Wk von Augenzahl 4

Übung 2: Was ist die Wk, dass die Augenzahl größer als 3, wenn wir wissen, die Augenzahl ist 5.
Ereignis A: Augenzahl größer als 3 = {4,5,6}
Ereignis B: Augenzahl ist 5 = {5}
Wir suchen P(A|B)
A ∩ B = {5} und B = {5}
P(A ∩ B) / P(B) = P({5}) / P({5} = (1/6) / (1/6) = **1**
- Wenn die Augenzahl ist 5, es ist 100% sicher größer als 3

Zufallsvariablen

Zufallsvariable: Funktion, welche jedem Ergebnis eine reelle Zahl zuweist.

Beispiel: Abbildung der Noten *sehr gut, gut, befriedigend, ausreichend, mangelhaft, ungenügend* auf die Zahlen 1, 2, 3, 4, 5, 6.

Wahrscheinlichkeit eines Wertes x der Zufallsvariablen X:

$$P(X = x) = p(x) = P(A_x)$$

A_x: Menge der Ergebnisse, für welche die Zufallsvariable X den Wert x liefert.

$p(x)$ ist eine abkürzende Schreibweise, die oft verwendet wird, wenn klar ist, zu welcher Zufallsvariablen der Wert x gehört.

Eine Zufallsvariable, die nur die Werte 0 und 1 liefert, nennt man **Bernoulli-Experiment**.

A randome variable is a function that assigns each outcome of an experiment to a real number. This is a way to represent the outcomes with numeric values so that it is easy to compute.

- Given outcomes: sehr gut, gut, befriedigend, ausreichend, mangelhaft, ungenügend
- How can we define a random variable?
    - We ask a question, what are we interested in these outcomes?
- Defining a random variable is like asking a question from the given outcomes.
    - X = What grade do I get (in numerical value)?, Values x are e.g. 1, 2, 3, 4, 5, 6
        - you can also use other numbers to represent the grades, but it makes sense to pick values that represent the meaning well.
    - Y = Do I pass the test? Values x are e.g. 1 (stands for yes), 0 (stands for no)
        - you can also use other numbers, but conventionally we use 1 and 0
        - This kind of random variable is called Bernoulli random variable
    - Z = Do I get grade “sehr gut”? Values x are e.g. 1 (stands for yes), 0 (stands for no)

Wahrscheinlichkeit eines Wertes x der Zufallsvariablen X:

$$P(X = x) = p(x) = P(A_x)$$

A_x: Menge der Ergebnisse, für welche die Zufallsvariable X den Wert x liefert.

$p(x)$ ist eine abkürzende Schreibweise, die oft verwendet wird, wenn klar ist, zu welcher Zufallsvariablen der Wert x gehört.

Übung: Write the probabilty of each value of X using 3 notations for the following random variables

1. X = What grade do I get (in numerical value)

p(X = 1) = p(1) = p({1})
p(X = 2) = p(2) = p({2})
p(X = 3) = p(3) = p({3})
p(X = 4) = p(4) = p({4})
p(X = 5) = p(5) = p({5})
p(X = 6) = p(6) = p({6})

2. Y = Do I pass the test with value 1 for yes, 0 for no
p(Y = 1) = p(1) = p({1,2,3,4})
p(Y = 0) = p(0) = p(5,6})

3. Z = grade is “sehr gut”
p(Y = 1) = p(1) = p({1})
p(Y = 0) = p(0) = p({2,3,4,5,6})

4. W = grade is better than “ausreichend”
p(Y = 1) = p(1) = p({1,2,3})
p(Y = 0) = p(0) = p({4,5,6})

Übung

- Given a bag containing several word bigrams
  - black tea, the dog, white tea, black tea, drink milk, like to
- Experiment: we take one bigram out of the bag and observe each word
  - What ist the sample space (Ergebnisraum) ?
    - {black tea, the dog, white tea, drink milk, like to}
- We want to compute the probablity that we pick “black tea” and the probabilty that we pick other bigrams, how can we define a random variable? And what are the values that it can take?
  - B = pick “black tea”
  - with value 1 for picking “black tea” and 0 for other cases
- How to denote the probabilty of picking “black tea” and the probabilty of picking other bigrams?
  - P(B = 1) or p(1) or p({black tea})
  - P(B = 0) or p(0) or p({the dog, white tea, drink milk, like to})

Erwartungswert

Der **Erwartungswert** ist der Mittelwert einer Zufallsvariablen:

$$E(X) = \sum_x p(x)x$$

Beispiel: erwartete Augenzahl bei einem fairen Würfel: 1/6*(1+2+3+4+5+6)=3,5

Erwartungswert einer Funktion f:

$$E(f(X)) = \sum_x p(x)f(x)$$

Beispiel: erwartete quadrierte Augenzahl: 1/6*(1+4+9+16+25+36) = 91/6

Übung:
https://de.serlo.org/mathe/stochastik/zufallsgr%C3%B6%C3%9Fen/aufgaben-erwartungswert

Erklärung:
https://www.toppr.com/guides/business-economics-cs/mathematics-of-finance-and-elementary-probability/expected-values/

Gemeinsame Verteilungen

Die **gemeinsame Verteilung** zweier Zufallsvariablen X und Y:

$$p(x, y) = P(X{=}x, Y{=}y) = P(A_x \cap A_y)$$

Übung (Aus der Vorlesung): Berechne die gemeinsame Verteilung von Zufallsvariablen X und Y

X = Augenzahl > 3
Y = Augenzahl ist gerade

Gemeinsame Verteilungen

Die **gemeinsame Verteilung** zweier Zufallsvariablen X und Y:

$$p(x, y) = P(X{=}x, Y{=}y) = P(A_x \cap A_y)$$

Übung: Berechne die gemeinsame Verteilung von Zufallsvariablen X und Y

X = Augenzahl > 3
Y = Augenzahl ist gerade

Lösung:
Wir wollen p(x,y) für alle x von X und y von Y berechnen
Welche Werte kann X haben?
- X ist eine Bernoulli-ZV, es hat also Werte 1 und 0
- 1 steht für Ergebnisse mit “Augenzahl > 3”, 0 steht für Ergebnisse mit “ Augenzahl **nicht** > 3”

Welche Werte kann Y haben?
- analog zu X

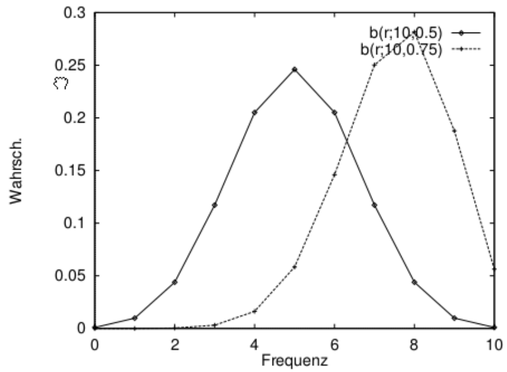
Wir suchen also
P(X=1, Y=1) = P({4,6}) = p(4) + p(6) = 2/6
P(X=1, Y=0) = P({5}) = p(5) = 1/6
P(X=0, Y=1) = P({2}) = p(2) = 1/6
P(X=0, Y=0) = P({1,3}) = p(1) + p(3) = 2/6

Binomialverteilung

Eine **Binomialverteilung** ergibt sich, wenn ein **Bernoulli-Experiment** (Ergebnisse 0 und 1, Bsp. Münzwurf: X(Kopf)=1, X(Zahl)=0) n Mal wiederholt wird. Die Binomialverteilung ist die Wahrscheinlichkeit, dabei genau r viele "1"-Ereignisse zu bekommen, wenn die Wahrscheinlichkeit des "1"-Ereignisses p ist:

$$b(r; n, p) = \binom{n}{r} p^r (1-p)^{n-r}$$

$$\binom{n}{r} = \frac{n!}{(n-r)!\, r!}$$



Beispiel: Die mittlere Kurve gibt an, wie wahrscheinlich es bei 10 Münzwürfen ist, 0-mal, 1-mal, 2-mal etc. Kopf zu erhalten.

Übung:

-Was können wir mit Binomialverteilung berechnen?

- Wenn wir eine Bernoulli Experiment haben und möchte wissen, wenn wir das Experiement n mal wiederholen, wie wahrscheinlich ist es, dass wir 1-Ereignisse r mal bekommen (ohne echte Experimente durchzuführen)

-Was brauchen wir um diese Binomial-Wahrscheinlichkeit zu berechnen?

- n : Anzahl der Experimente
- r : Anzahl von 1-Ereignisse, für die wir uns interessiert
- p: Wk von 1-Ereignisse aus der Bernoulli-Experiment

-Welche Werte kann eine Binomialverteilung nehmen?

- r = 0,1,2,…,n
- Es berechnet also p(0), p(1), …, p(n)
- e.g. p(0) bedeutet die Wk, dass wir 1-Ereignisse null mal bekommen in n Experimente

Übung:
In a Bernoulli-Experiment, you pick a bigram from a bag containing 1000 bigrams, and ask if you get “black tea” or not. (with replacement--put the bigram back before doing the next experiment)

Since you formulate this as a Bernoulli-Experiment, the sample space is {black tea, not black tea}. Here we see other words besides “black tea” as “not black tea”, while in a general experiment, we might want to consider each actual word as an outcome. You are also given that P(black tea) = 0.2 and P(not black tea) = 0.7

You want to know if you repeat the experiment 1000 times, what is the probabilty of getting “black tea” 300 times?

Lösung:
- We can use Binomialverteilung to calculate this prob
- r: 300
- n: 1000
- p: 0.2
- Put this to the formula

$$b(r; n, p) = \binom{n}{r} p^r (1-p)^{n-r}$$

$$\binom{n}{r} = \frac{n!}{(n-r)!\, r!}$$

Done