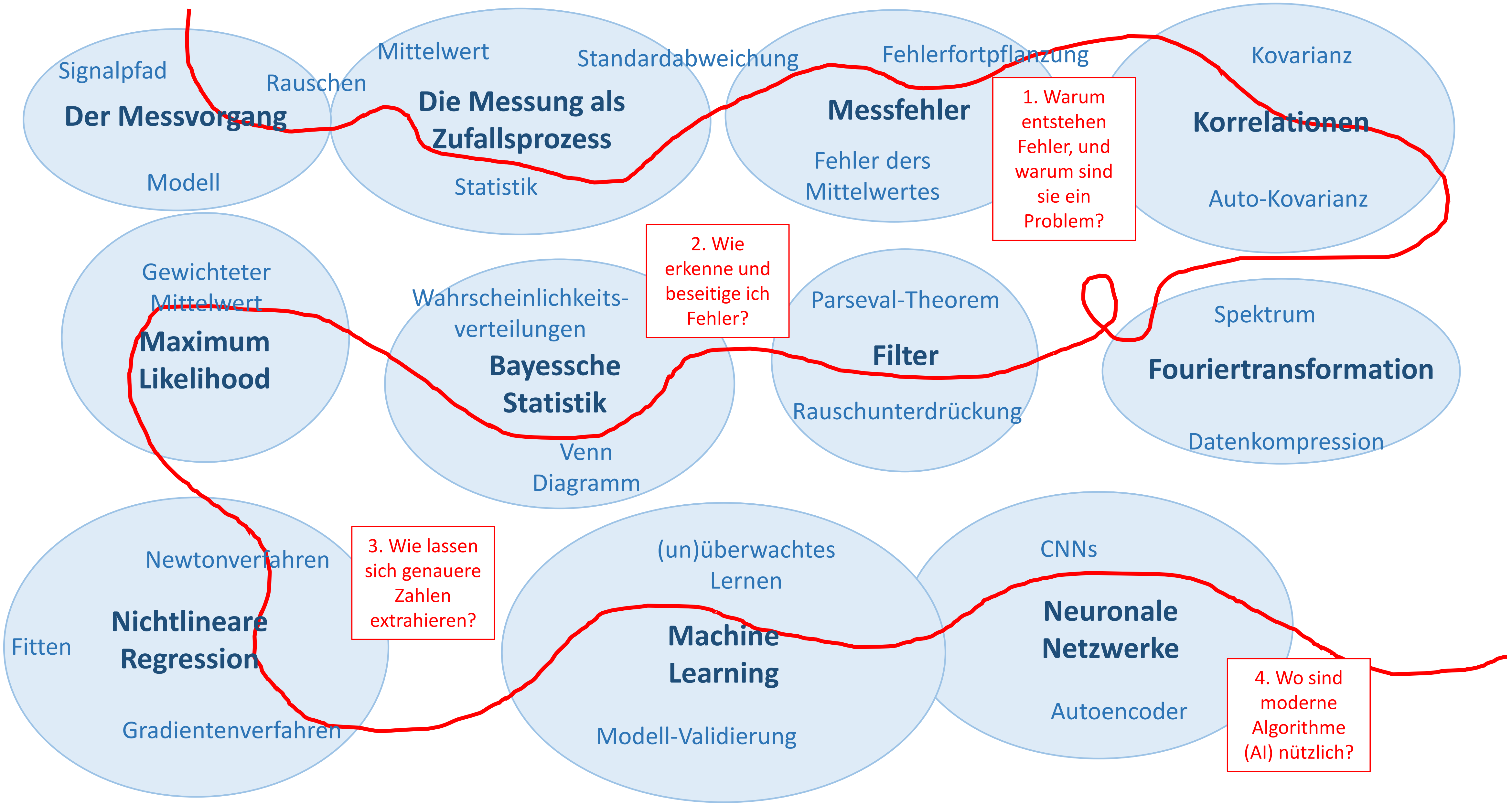
Bayesian vs Frequentist Ansatz

Der rote Faden: wie erhalte ich aussagekräftige Zahlen aus komplexen Daten?



Themen dieser Vorlesung:

- Frequentistischer Wahrscheinlichkeitsbegriff
- Venn Diagramm
- Der Satz von Bayes
- Wahrscheinlichkeitsverteilung und ihre Charakteristika und Momente
- Einführung in die Bayessche Datenanalyse

Repetition zu Messfehlern

Wir interpretieren eine Messung als Zufallsexperiment und das Ergebnis der Messung als Zufallsvariable.

Was bedeutet demnach: $x = (0.52 \pm 0.03)\mathrm{m}$?

Repetition zu Messfehlern - Frequentistischer Wahrscheinlichkeitsbegriff

Unsicherheit einer Messung und Unsicherheit des Mittelwerts.

- Unter welchen Umständen können wir durch wiederholtes Messen die Präzision der Messung verbessern?

- Wie kann die Präzision der Messung eines Parameters verbessert werden, der mit der Samplingzeit von Δ_t gemessen wird?

Frequentistischer Wahrscheinlichkeitsbegriff

Wir interpretieren eine Messung als Zufallsexperiment und das Ergebnis der Messung als Zufallsvariable.

Demnach ist die Wahrscheinlichkeit, dass ein bestimmtes Ereignis A eintritt gegeben durch:

$$P(A) = \frac{k}{n}$$

mit: $n =$ Anzahl der möglichen Ergebnisse des Experiments.
$k =$ Anzahl der Ergebnisse die zu A führen.

Beispiel:
Sechseitiger Würfel, A: Die Augenzahl ist grösser als 4:

$$\rightarrow n = 6, k = 2 \Rightarrow P(A) = \frac{2}{6} = \frac{1}{3}$$

Frequentistischer Wahrscheinlichkeitsbegriff

Wir interpretieren eine Messung als Zufallsexperiment und das Ergebnis der Messung als Zufallsvariable.

Demnach ist die Wahrscheinlichkeit, dass ein bestimmtes Ereignis A eintritt gegeben durch:

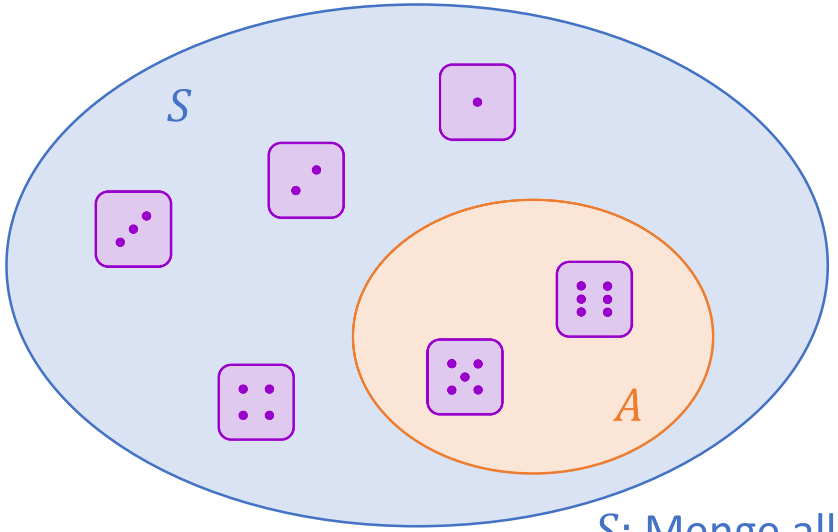
$$P(A) = \frac{k}{n}$$

mit: n = Anzahl der möglichen Ergebnisse des Experiments.
k = Anzahl der Ergebnisse die zu A führen.

Beispiel:
Sechseitiger Würfel, A: Die Augenzahl ist grösser als 4:

$$\rightarrow n = 6, k = 2 \Rightarrow P(A) = \frac{2}{6} = \frac{1}{3}$$

Wir können das in einem sogenannten Venn Diagramm darstellen.



S: Menge aller möglichen Ergebnisse

Venn Diagramm

$$P(A) = \frac{n_A}{n_S}$$

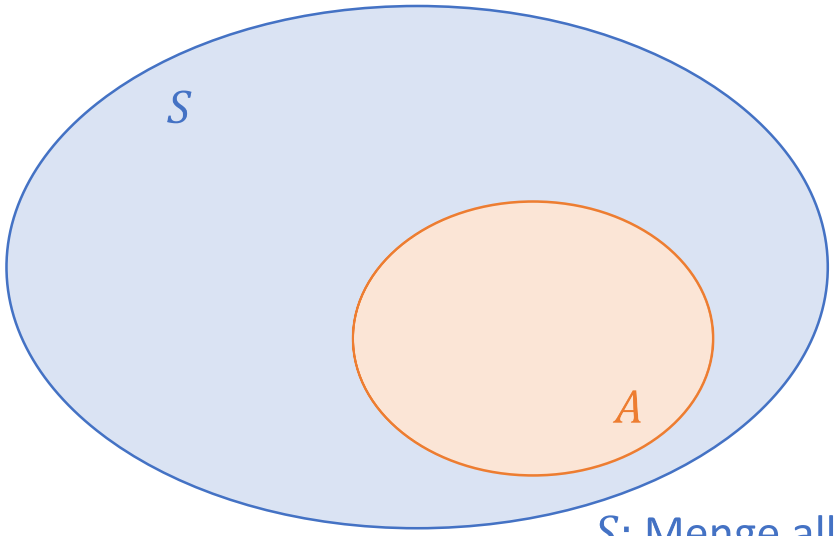
n_S: Anzahl der Ergebnisse in S
n_A: Anzahl der Ergebnisse in A

Es gilt:

$A \subset S$: $\quad 0 \leq P(A) \leq 1$

Und:

$$1 = P(S) = P(A \cup \bar{A}) = P(A) + P(\bar{A})$$



S: Menge aller möglichen Ergebnisse

Venn Diagramm

$$P(A) = \frac{n_A}{n_S}$$

n_S: Anzahl der Ergebnisse in S
n_A: Anzahl der Ergebnisse in A

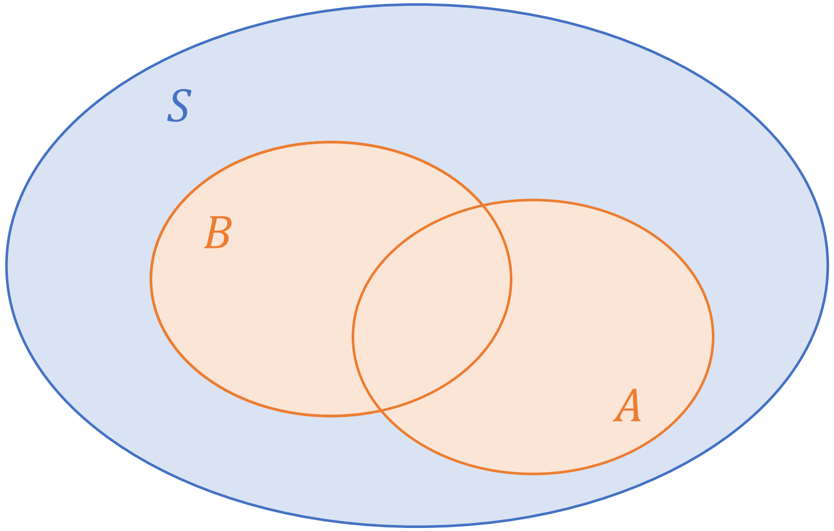
Es gilt:

$A \subset S$: $\quad 0 \leq P(A) \leq 1$

Und:

$$1 = P(S) = P(A \cup \bar{A}) = P(A) + P(\bar{A})$$

Wir können das für mehrere Ereignisse erweitern.

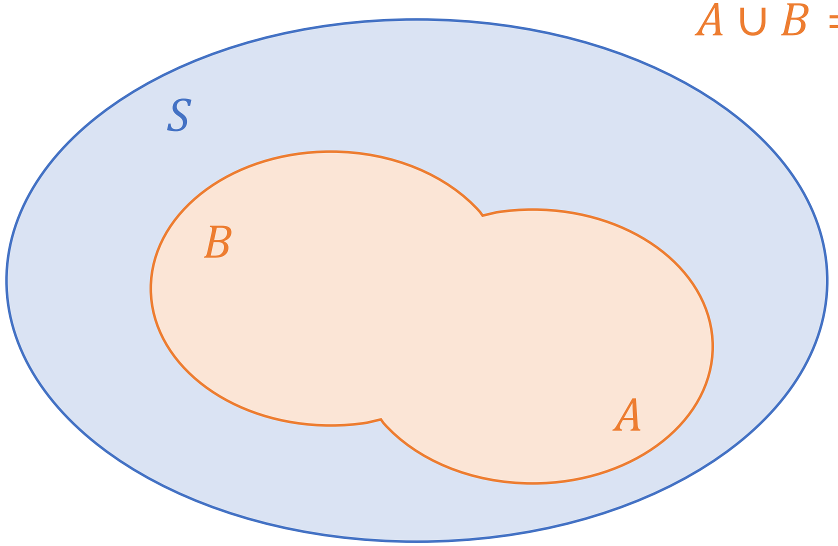


Venn Diagramm

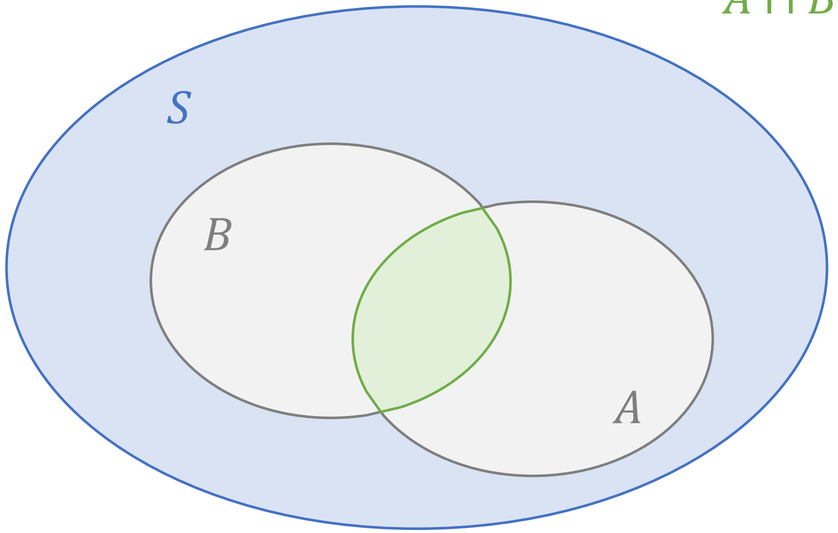
Die Wahrscheinlichkeit für ein Ergebnis in A oder B ist dann:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$A \cup B = B \cup A$



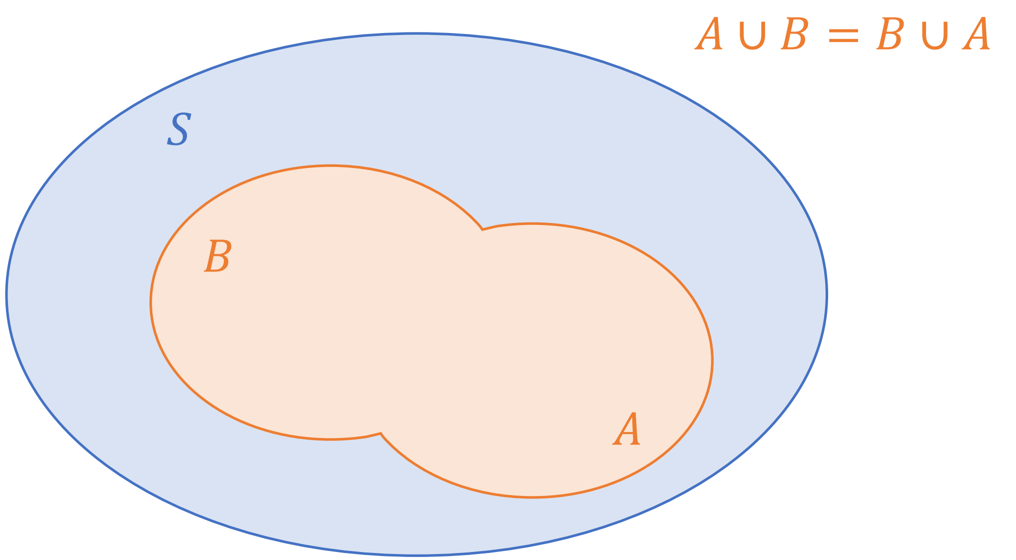
$A \cap B = B \cap A$



Venn Diagramm

Die Wahrscheinlichkeit für ein Ergebnis in A oder B ist dann:

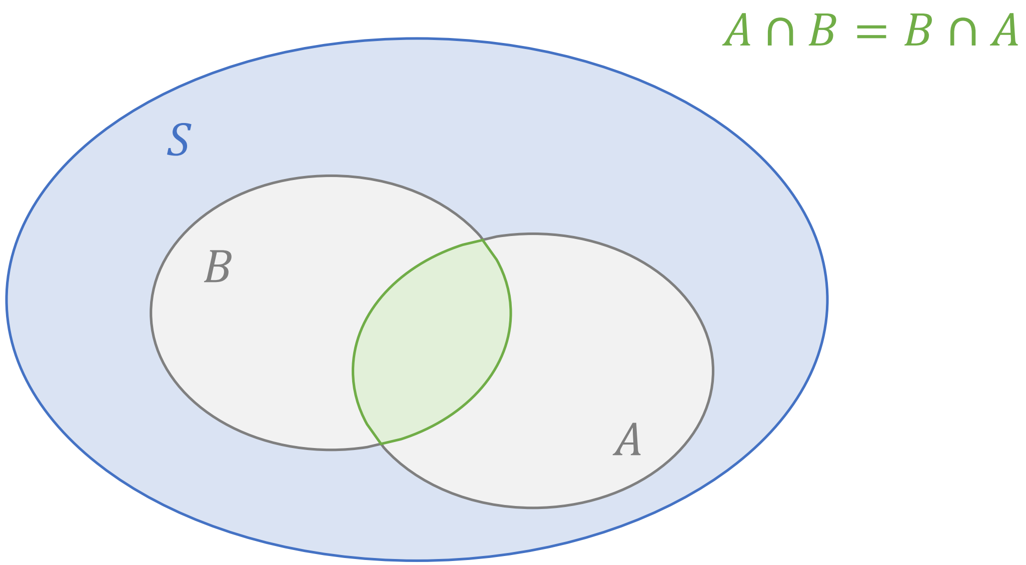
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Die Wahrscheinlichkeit dass A eintritt, wenn B eintritt, also die Wahrscheinlichkeit für A gegeben B:

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

oder: $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$



Achtung: Es muss kein kausaler Zusammenhang zwischen A und B herrschen!

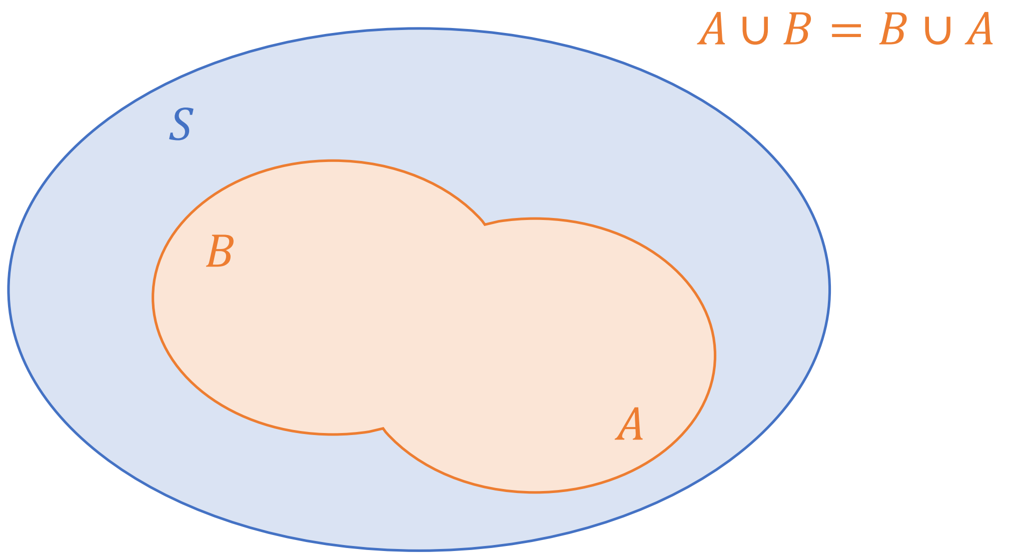
Wenn A und B unabhängig sind dann ist:

$P(A|B) = P(A)$ und $P(A \cap B) = P(A)P(B)$

Der Satz von Bayes

Die Wahrscheinlichkeit für ein Ergebnis in A oder B ist dann:

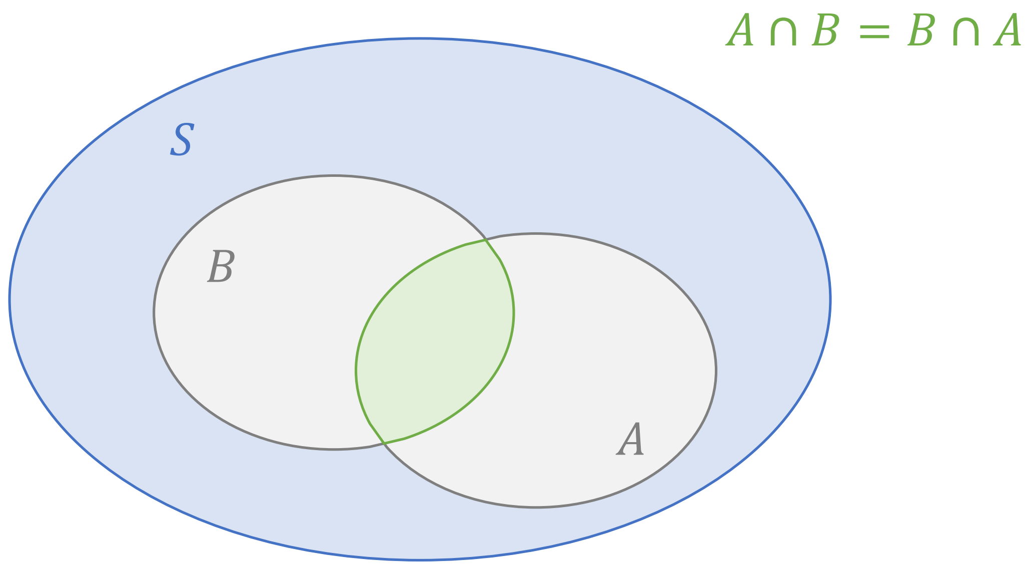
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Die Wahrscheinlichkeit dass A eintritt, wenn B eintritt, also die Wahrscheinlichkeit für A gegeben B:

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

oder: $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$



Hieraus folgt:

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

Satz von Bayes

Wahrscheinlichkeitsverteilungen

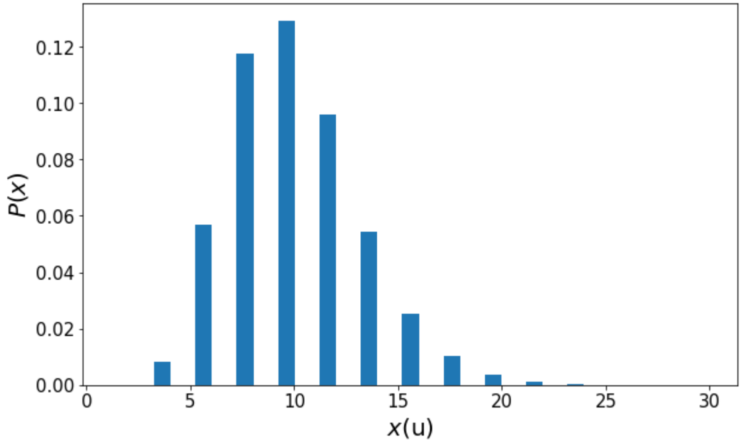
Wenn es viele mögliche Ergebnisse eines Experiments gibt, schreiben wir diese als A_i. Wenn jedes mögliche Ergebnis von den A_i abgedeckt wird, dann ist $P(A_i)$ eine Wahrscheinlichkeitsverteilung mit:

$$P(A_i) \geq 0$$

und

$$\sum_i P(A_i) = 1$$

Für einen diskreten Satz von Ereignissen A_i nennen wir das auch eine Wahrscheinlichkeitsmassefunktion, Probability Mass Function – PMF



Wahrscheinlichkeitsverteilungen

Wenn es viele mögliche Ergebnisse eines Experiments gibt, schreiben wir diese als A_i. Wenn jedes mögliche Ergebnis von den A_i abgedeckt wird, dann ist $P(A_i)$ eine Wahrscheinlichkeitsverteilung mit:

$$P(A_i) \geq 0$$

und

$$\sum_i P(A_i) = 1$$

Für einen diskreten Satz von Ereignissen A_i nennen wir das auch eine Wahrscheinlichkeitsmassefunktion, Probability Mass Function – PMF

Falls das Ergebnis des Experiment eine kontinuierliche Zufallsvariable x ist, dann können wir schreiben:

$$f(x) \geq 0$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Dies nennen wir eine Wahrscheinlichkeitsdichtefunktion, Probability Density Function – PDF

Wahrscheinlichkeitsverteilungen

Falls das Ergebnis des Experiment eine kontinuierliche Zufallsvariable x ist, dann können wir schreiben:

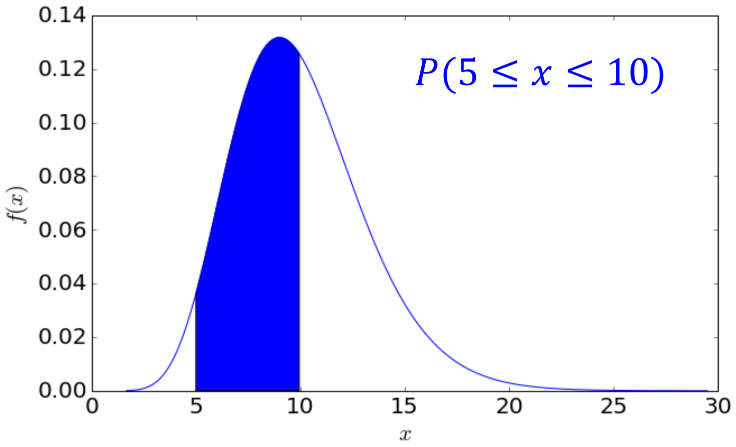
$$f(x) \geq 0$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Dies nennen wir eine Wahrscheinlichkeitsdichtefunktion, Probability Density Function – PDF

Die Wahrscheinlichkeit, dass das Ergebnis x in einem Intervall zwischen a und b liegt ist:

$$P(a \leq x \leq b) = \int_a^b f(x)dx$$



Achtung: $f(x)$ ist nicht die Wahrscheinlichkeit dass x auftritt!

Beispiel: Die Dirac-δ-Funktion erfüllt alle Eigenschaften einer PDF:

$$f(x) = \delta(x) \equiv \begin{cases} \infty, & x = 0 \\ 0, & x \neq 0 \end{cases}$$

Wahrscheinlichkeitsverteilungen

Falls das Ergebnis des Experiment eine kontinuierliche Zufallsvariable x ist, dann können wir schreiben:

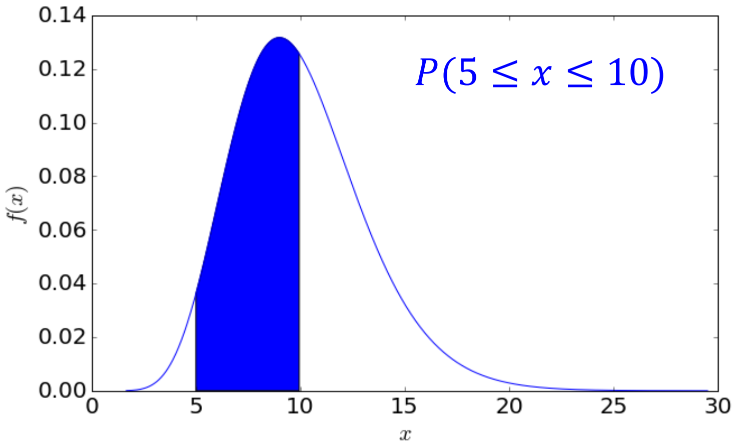
$$f(x) \geq 0$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Dies nennen wir eine Wahrscheinlichkeitsdichtefunktion, Probability Density Function – PDF

Die Wahrscheinlichkeit, dass das Ergebnis x in einem Intervall zwischen a und b liegt ist:

$$P(a \leq x \leq b) = \int_a^b f(x)dx$$



Achtung: $f(x)$ ist nicht die Wahrscheinlichkeit dass x auftritt!

Beispiel: Die Dirac-δ-Funktion erfüllt alle Eigenschaften einer PDF:

$$f(x) = \delta(x) \equiv \begin{cases} \infty, & x = 0 \\ 0, & x \neq 0 \end{cases}$$

Von nun an benutze ich die Abkürzung PDF und nehme kontinuierliche Wahrscheinlichkeitsverteilungen an.

Charakteristiken von Wahrscheinlichkeitsverteilungen

Generell wird eine PDF durch ihre funktionale Form $f(x)$ beschrieben.

Manchmal ist es Nützlich oder Ausreichend die PDF durch ihre wichtigsten Eigenschaften zu charakterisieren.

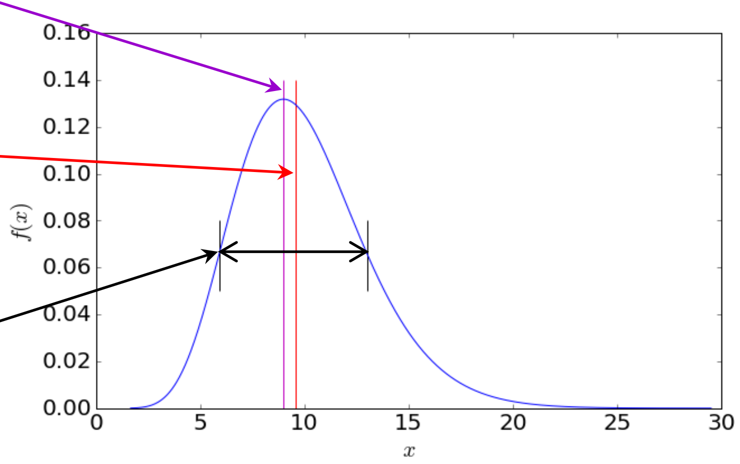
Mode Der Wert von x bei dem die PDF ihr Maximum erreicht.
Für eine PDF mit einem einzigen Maximum: x_{mode}: $\left.\frac{df(x)}{dx}\right|_{x_{\mathrm{mode}}} = 0$

Median Die Mitte der PDF:

$$x_{\mathrm{median}}: \frac{1}{2} = \int_{-\infty}^{x_{\mathrm{median}}} f(x)dx$$

FWHM Die Halbwertsbreite (Full Width at Half Maximum) der PDF:

$$f(a) = f(b) = \frac{1}{2} f(x_{\mathrm{mode}}); \quad \mathrm{FWHM} = b - a$$



Charakteristiken von Wahrscheinlichkeitsverteilungen

Generell wird eine PDF durch ihre funktionale Form $f(x)$ beschrieben.

Manchmal ist es Nützlich oder Ausreichend die PDF durch ihre wichtigsten Eigenschaften zu charakterisieren.

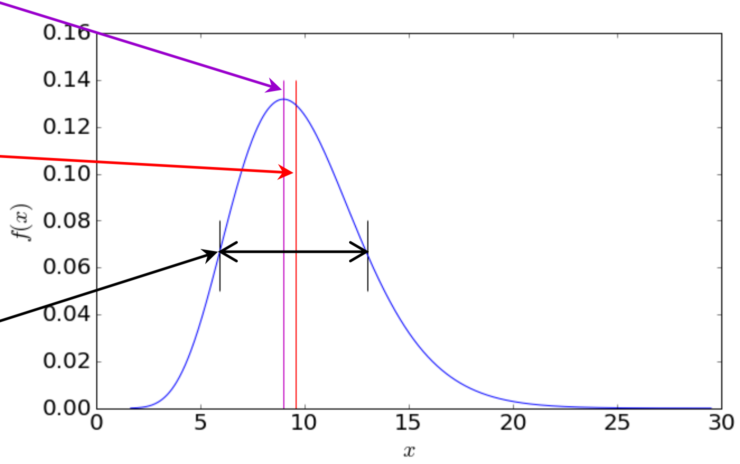
Mode Der Wert von x bei dem die PDF ihr Maximum erreicht.
Für eine PDF mit einem einzigen Maximum: $x_{\text{mode}}\colon \left.\frac{df(x)}{dx}\right|_{x_{\text{mode}}} = 0$

Median Die Mitte der PDF:

$$x_{\text{median}}\colon \frac{1}{2} = \int_{-\infty}^{x_{\text{median}}} f(x)dx$$

FWHM Die Halbwertsbreite (Full Width at Half Maximum) der PDF:

$$f(a) = f(b) = \frac{1}{2} f(x_{\text{mode}}); \quad \text{FWHM} = b - a$$



Im Falle einer Normalverteilung $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$: $x_{\text{mode}} = \mu$, setzen wir o.B.d.A $\mu = 0$

$$f(\mu = 0) = \frac{1}{\sigma\sqrt{2\pi}}, f(x') = \frac{1}{2} f(\mu = 0) = \frac{1}{2\sigma\sqrt{2\pi}} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x'^2}{2\sigma^2}\right) \to \log 2 = \frac{x'^2}{2\sigma^2} \to x' = \sigma\sqrt{2\log 2} \to \text{FWHM}_{\text{Gauss}} = 2\sigma\sqrt{2\log 2}$$

Charakteristiken von Wahrscheinlich

Generell wird eine PDF durch ihre funktionale Form $f(x)$ b

Manchmal ist es Nützlich oder Ausreichend die PDF durch i

Mode Der Wert von x bei dem die PDF ihr Maximum e

Für eine PDF mit einem einzigen Maximum: x_{mo}

Median Die Mitte der PDF:

$$x_{\mathrm{median}}: \frac{1}{2} = \int_{-\mathrm{c}}^{x}$$

FWHM Die Halbwertsbreite (Full Width at Half Maximur

$$f(a) = f(b) = \frac{1}{2} f(x_{\mathrm{moc}}$$

Im Falle einer Normalverteilung $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right.$

$f(\mu = 0) = \frac{1}{\sigma\sqrt{2\pi}}, f(x') = \frac{1}{2} f(\mu = 0) = \frac{1}{2\sigma\sqrt{2\pi}} = \frac{1}{\sigma\sqrt{2\pi}} \exp$

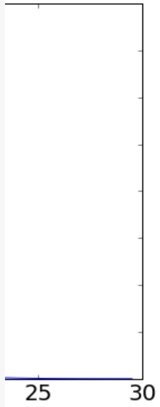
MK write a python code that loads a dataset (x, y, with y a function of x) from a file and calculates the full width at half max of the y

Here is an example of Python code that loads data from a file, and calculates the full width at half max (FWHM) of the y values:

Copy code

```
import numpy as np

# Load data from file
data = np.loadtxt('data.txt')
x = data[:,0]
y = data[:,1]

# Find the maximum value of y
ymax = max(y)

# Find the index of the maximum value
ymax_index = np.argmax(y)

# Find the half maximum value
half_max = ymax / 2

# Find the indexes of the points where y is greater than or equal to half_max
above_half_max_indexes = np.where(y >= half_max)[0]

# Find the first and last index of the points where y is greater than or equal
to half_max
left_index = above_half_max_indexes[0]
right_index = above_half_max_indexes[-1]

# Calculate the full width at half max
fwhm = x[right_index] - x[left_index]

print("FWHM:", fwhm)
```

Regenerate response



g 2

Momente einer Wahrscheinlichkeitsverteilung

Darüber hinaus sind für viele PDFs ihre Momente definiert, das m-te Moment:

$$M_m = \int_{-\infty}^{\infty} x^m f(x) dx$$

Beispiel:

Das 0-te Moment: $M_0 = \int_{-\infty}^{\infty} f(x) dx = 1$

Das 1-te Moment: $M_1 = \int_{-\infty}^{\infty} x f(x) dx$ Der Mittelwert oder der Erwartungswert.

Bemerkung: nicht jede PDF hat einen Mittelwert (zB die Cauchy- oder Lorentzverteilung: $f(x) = \frac{1}{\pi} \frac{\gamma}{(x-x_0)^2+\gamma^2}$)

Darüber hinaus ist das m-te zentrale Moment einer PDF definiert als:

$$\widetilde{M}_m = \langle (x - M_1)^m \rangle$$

Damit ist die Varianz einer PDF gegeben durch das zweite zentrale Moment: $\boxed{\sigma^2 = \langle (x - M_1)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 = M_2 - M_1^2}$

Das dritte quantifiziert die „Schiefe“ und das vierte die „Wölbung“ der PDF.

Momente einer Wahrscheinlichkeitsverteilung

Darüber hinaus sind für viele PDFs ihre Momente definiert, das m-te Moment:

$$M_m = \int_{-\infty}^{\infty} x^m f(x) dx$$

Beispiel:

Das 0-te Moment: $M_0 = \int_{-\infty}^{\infty} f(x) dx = 1$

Das 1-te Moment: $M_1 = \int_{-\infty}^{\infty} x f(x) dx$ Der Mittelwert oder der Erwartungswert.

Für eine PMF: $M_1^{PMF} = \sum_{n=1}^{N} x_n P(x_n)$

Bemerkung: nicht jede PDF hat einen Mittelwert (zB die Cauchy- oder Lorentzverteilung: $f(x) = \frac{1}{\pi} \frac{\gamma}{(x-x_0)^2+\gamma^2}$)

Darüber hinaus ist das m-te zentrale Moment einer PDF definiert als:

$$\widetilde{M}_m = \langle (x - M_1)^m \rangle$$

Damit ist die Varianz einer PDF gegeben durch das zweite zentrale Moment: $\boxed{\sigma^2 = \langle (x - M_1)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 = M_2 - M_1^2}$

Das dritte quantifiziert die „Schiefe“ und das vierte die „Wölbung“ der PDF.

Momente einer Wahrscheinlichkeitsverteilung

Es ist hilfreich die Momenterzeugende Funktion zu definieren:

$$M(\zeta) = \int_{-\infty}^{\infty} \exp(\zeta x)\, f(x) dx$$

Um dies etwas besser zu verstehen betrachten wir die Taylor Reihe von $\exp(\zeta x)$:

$$\begin{aligned} M(\zeta) &= \int_{-\infty}^{\infty} \left(1 + \zeta x + \frac{\zeta^2 x^2}{2!} + \cdots \right) f(x) dx \\ &= M_0 + \zeta M_1 + \frac{\zeta^2}{2!} M_2 + \cdots \end{aligned}$$

Für eine normierte Wahrscheinlichkeitsverteilung gilt: $M_0 = 1$ und höhere Momente $m \geq 1$ können iterativ aus der Ableitung der Momenterzeugenden Funktion berechnet werden:

$$\boxed{M_m = \left. \frac{\partial^m M(\zeta)}{\partial \zeta^m} \right|_{\zeta=0}}$$

Charakteristische Funktion einer Wahrscheinlichkeitsverteilung

Darüber hinaus ist es hilfreich die Charakteristische Funktion einer Wahrscheinlichkeitsverteilung $f(x)$ zu definieren:

$$\phi(\nu) = \int_{-\infty}^{\infty} \exp(i\nu x)\, f(x) dx$$

Dies ist die inverse Fourier Transformierte der Wahrscheinlichkeitsverteilung. Die Fourier Transformation haben wir in der Vorlesung zur Spektralanalyse kennengelernt.

Wir können die Exponentialfunktion wieder erweitern:

$$\phi(\nu) = \int_{-\infty}^{\infty} \left(1 + i\nu \mathrm{x} + \frac{1}{2!}(i\nu)^2 x^2 + \frac{1}{3!}(i\nu)^3 x^3 + \cdots \right) f(x) dx =$$
$$= 1 + i\nu M_1 + \frac{1}{2!}(i\nu)^2 M_2 + \frac{1}{3!}(i\nu)^3 M_3 + \cdots$$

Charakteristische Funktion einer Wahrscheinlichkeitsverteilung

Darüber hinaus ist es hilfreich die Charakteristische Funktion einer Wahrscheinlichkeitsverteilung $f(x)$ zu definieren:

$$\phi(\nu) = \int_{-\infty}^{\infty} \exp(i\nu x)\, f(x) dx$$

Dies ist die inverse Fourier Transformierte der Wahrscheinlichkeitsverteilung. Die Fourier Transformation haben wir in der Vorlesung zur Spektralanalyse kennengelernt.

Ausserdem können wir über die Fourier Transformation aus $\phi(\nu)$ die Wahrscheinlichkeitsverteilung $f(x)$ berechnen:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\nu x)\, \phi(\nu) d\nu$$

Charakteristische Funktion einer Wahrscheinlichkeitsverteilung

Darüber hinaus ist es hilfreich die Charakteristische Funktion einer Wahrscheinlichkeitsverteilung $f(x)$ zu definieren:

$$\phi(\nu) = \int_{-\infty}^{\infty} \exp(i\nu x)\, f(x) dx$$

Dies ist die inverse Fourier Transformierte der Wahrscheinlichkeitsverteilung. Die Fourier Transformation haben wir in der Vorlesung zur Spektralanalyse kennengelernt.

Die Kombination dieser beiden Ergebnisse ergibt:

$$\boxed{f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\nu x) \left(1 + i\nu M_1 + \frac{1}{2!}(i\nu)^2 M_2 + \frac{1}{3!}(i\nu)^3 M_3 + \cdots\right) d\nu}$$

Ein wichtiges und bemerkenswertes Ergebnis!
Wir können aus den Momenten (die wir zB aus der Messung bestimmen können) die Wahrscheinlichkeitsverteilung abschätzen.

Zusammenfassung

- Eine Wahrscheinlichkeitsdichtefunktion oder Wahrscheinlichkeitsverteilung, ist eine positive, normierte Funktion, die die Verteilung der Wahrscheinlichkeit einer Zufallsvariablen beschreibt.

- Falls die Zufallsvariable nur diskrete Werte annehmen kann nennen wir die Verteilungsfunktion eine Probability Mass Function – PMF

- Falls die Zufallsvariable kontinuierliche Werte annehmen kann nennen wie die Verteilungsfunktion eine Probability Density Function – PDF

- Die Momente einer Wahrscheinlichkeitsverteilung können mit der Momenterzeugenden Funktion berechnet werden

- Über die Charakteristische Funktion ist die Wahrscheinlichkeitsverteilung vollständig über ihre Momente definiert.

Bayessche Datenanalyse

Soweit erhalten wir die PDF aus wiederholten Messungen und die Wahrscheinlichkeit des Ergebnisses ist dann gegeben durch dessen gemessene Frequenz.
Das ist der sogenannte frequentistische Ansatz der Datenanalyse.

Aber was bedeutet hier eigentlich Wahrscheinlichkeit? Und Wahrscheinlichkeitsverteilung?

Ein gemessener Parameter hat ja einen tatsächlichen (unbekannten) Wert.

Der Bayessche Ansatz der Datenanalyse fasst Wahrscheinlichkeiten nicht als Frequenzen auf sondern als Plausibilität: Wie sehr denken wir das etwas wahr ist, basierend auf der Datenlage.

Somit beruht die Bayessche Datenanalyse auf bedingten Wahrscheinlichkeiten und insbesondere auf dem Satz von Bayes:

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

Bayessche Datenanalyse

Die Bayessche Datenanalyse beruht auf bedingten Wahrscheinlichkeiten und insbesondere auf dem Satz von Bayes:

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

Um dies auf ein Experiment anzuwenden definieren wir:

$A \rightarrow D$	Die gemessenen Daten des Experiments
$B \rightarrow B_1, \dots, B_n$	Ein endlicher Satz von Ergebnissen die wie mit Hilfe der Daten unterscheiden möchten
I	Unser gesamtes zusätzliches Wissen (wird oft vernachlässigt, so auch von mir)

$$P(B_i|D,I) = \frac{P(D|B_i,I)P(B_i|I)}{P(D|I)}$$

Wenn D impliziert, dass eines der B_i eintritt, dann ist die Wahrscheinlichkeit dass D eintritt:

$$P(D) = \sum_k P(D|B_k)P(B_k)$$

Bayessche Datenanalyse

Die Bayessche Datenanalyse auf bedingten Wahrscheinlichkeiten uns insbesondere auf dem Satz von Bayes:

$$P(B_i|D) = \frac{P(D|B_i)P(B_i)}{P(D)}$$

Wenn D impliziert, dass eines der B_i eintritt, dann ist die Wahrscheinlichkeit dass D eintritt:

$$P(D) = \sum_k P(D|B_k)P(B_k)$$

$P(B_i)$ ist die Wahrscheinlichkeit, dass B_i eintritt basierend auf unserem ursprünglichen (*A-priori*) Wissen (vor der Durchführung des Experiments). Es ist die sogenannte: *Prior probability distribution,* oder *A-priori-Verteilung,* kurz *Prior*.

Die neuen Daten erscheinen in $P(D|B_i)$, der Wahrscheinlichkeitsverteilung, D zu messen, wenn B_i eintritt. Dies ist im Prinzip eine Likelihood Funktion wie wir sie später genauer kennen lernen werden.

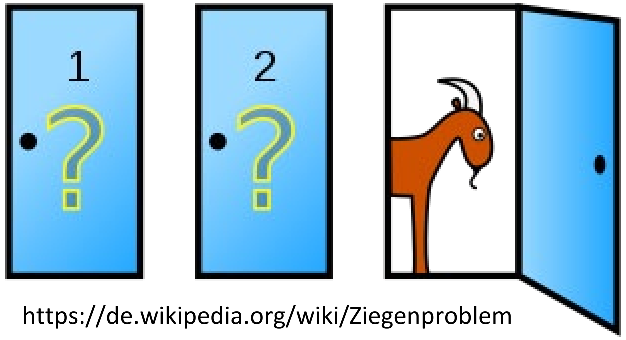
$P(D)$ garantiert die Normierung, so dass $P(B_i|D)$ eine PDF bzw. PMF ist.

Das Ergebnis, $P(B_i|D)$ ist die A-posteriori-Wahrscheinlichkeitsverteilung

Für eine kontinuierliche PDF schreiben wir:

$$f_1(B|D) = \frac{L(D|B)f_0(B)}{\int L(D|B)\, f_0(B)dB}$$

Bayessche Datenanalyse



https://de.wikipedia.org/wiki/Ziegenproblem

Beispiel:
Das „Ziegenproblem": Hinter einem von drei Toren ist ein Preis, hinter den anderen eine Ziege.

$B_i,\ \ i = 1, 2, 3$: Gewinn hinter Tor 1, 2 oder 3, also ist: $P(B_i) = \frac{1}{3}$

Bayessche Datenanalyse

Beispiel:
Nehmen wir an es gibt eine Infektionskrankheit, mit einer Inzidenzrate von 100. Das bedeutet, dass von 100000 Menschen, 100 infiziert sind, oder 0.1%.

Um festzustellen ob jemand infiziert ist existiert ein Test.

Nehmen wir des weiteren an, dass der Test sehr gut ist und in 99% der Fälle das korrekte Ergebnis liefert:

$$P(\text{positive}|\text{infected}) = 0.99$$
$$P(\text{negative}|\text{infected}) = 0.01$$

$$P(\text{positive}|\text{healthy}) = 0.01$$
$$P(\text{negative}|\text{healthy}) = 0.99$$

Bayessche Datenanalyse

Beispiel:
Nehmen wir an es gibt eine Infektionskrankheit, mit einer Inzidenzrate von 100. Das bedeutet, dass von 100000 Menschen, 100 infiziert sind, oder 0.1%.

Um festzustellen ob jemand infiziert ist existiert ein Test.

Nehmen wir des weiteren an, dass der Test sehr gut ist und in 99% der Fälle das korrekte Ergebnis liefert:

$$P(\text{positive}|\text{infected}) = 0.99$$
$$P(\text{negative}|\text{infected}) = 0.01$$

$$P(\text{positive}|\text{healthy}) = 0.01$$
$$P(\text{negative}|\text{healthy}) = 0.99$$

Person X lässt sich testen und der Test ist positiv.

Was ist die Wahrscheinlichkeit, dass X tatsächlich infiziert ist?

Man währe versucht zu sagen 99%, basierend auf der Genauigkeit des Tests.

Bayessche Datenanalyse

Beispiel:
Nehmen wir an es gibt eine Infektionskrankheit, mit einer Inzidenzrate von 100. Das bedeutet, dass von 100000 Menschen, 100 infiziert sind, oder 0.1%.

Um festzustellen ob jemand infiziert ist existiert ein Test.

Die ursprüngliche Wahrscheinlichkeit (vor dem Test), dass eine Person infiziert ist:

$$P(\text{infected}) = 0.001$$
$$P(\text{healty}) = 0.999$$

Nehmen wir des weiteren an, dass der Test sehr gut ist und in 99% der Fälle das korrekte Ergebnis liefert:

$$P(\text{positive}|\text{infected}) = 0.99$$
$$P(\text{negative}|\text{infected}) = 0.01$$

$$P(\text{positive}|\text{healthy}) = 0.01$$
$$P(\text{negative}|\text{healthy}) = 0.99$$

Person X lässt sich testen und der Test ist positiv.

Was ist die Wahrscheinlichkeit, dass X tatsächlich infiziert ist?

Man währe versucht zu sagen 99%, basierend auf der Genauigkeit des Tests.

Bayessche Datenanalyse

Beispiel:
Nehmen wir an es gibt eine Infektionskrankheit, mit einer Inzidenzrate von 100. Das bedeutet, dass von 100000 Menschen, 100 infiziert sind, oder 0.1%.

Um festzustellen ob jemand infiziert ist existiert ein Test.

Die ursprüngliche Wahrscheinlichkeit (vor dem Test), dass eine Person infiziert ist:

$$P(\text{infected}) = 0.001$$
$$P(\text{healty}) = 0.999$$

Nehmen wir des weiteren an, dass der Test sehr gut ist und in 99% der Fälle das korrekte Ergebnis liefert:

$$P(\text{positive}|\text{infected}) = 0.99$$
$$P(\text{negative}|\text{infected}) = 0.01$$

$$P(\text{positive}|\text{healthy}) = 0.01$$
$$P(\text{negative}|\text{healthy}) = 0.99$$

Person X lässt sich testen und der Test ist positiv.

Was ist die Wahrscheinlichkeit, dass X tatsächlich infiziert ist?

Man währe versucht zu sagen 99%, basierend auf der Genauigkeit des Tests.

Nach dem Satz von Bayes:

$$P(\text{infected}|\text{positive}) = \frac{P(\text{positive}|\text{infected})P(\text{infected})}{P(\text{positive}|\text{infected})P(\text{infected}) + P(\text{positive}|\text{healthy})P(\text{healty})} = 0.1$$

Die Wahrscheinlichkeit, dass X infiziert ist, ist „nur“ 10%.

Bayessche Datenanalyse

Beispiel Fortsetzung:
Person X lässt sich erneut testen und der Test ist wieder positiv.

Unsere Ursprüngliche Information beinhaltet nun das Ergebnis des ersten Tests:

$$P(\text{infected}) = 0.1$$
$$P(\text{healty}) = 0.9$$

Damit wird

$$P(\text{infected}|\text{positive}) = \frac{P(\text{positive}|\text{infected})P(\text{infected})}{P(\text{positive}|\text{infected})P(\text{infected}) + P(\text{positive}|\text{healthy})P(\text{healty})} = 0.92$$

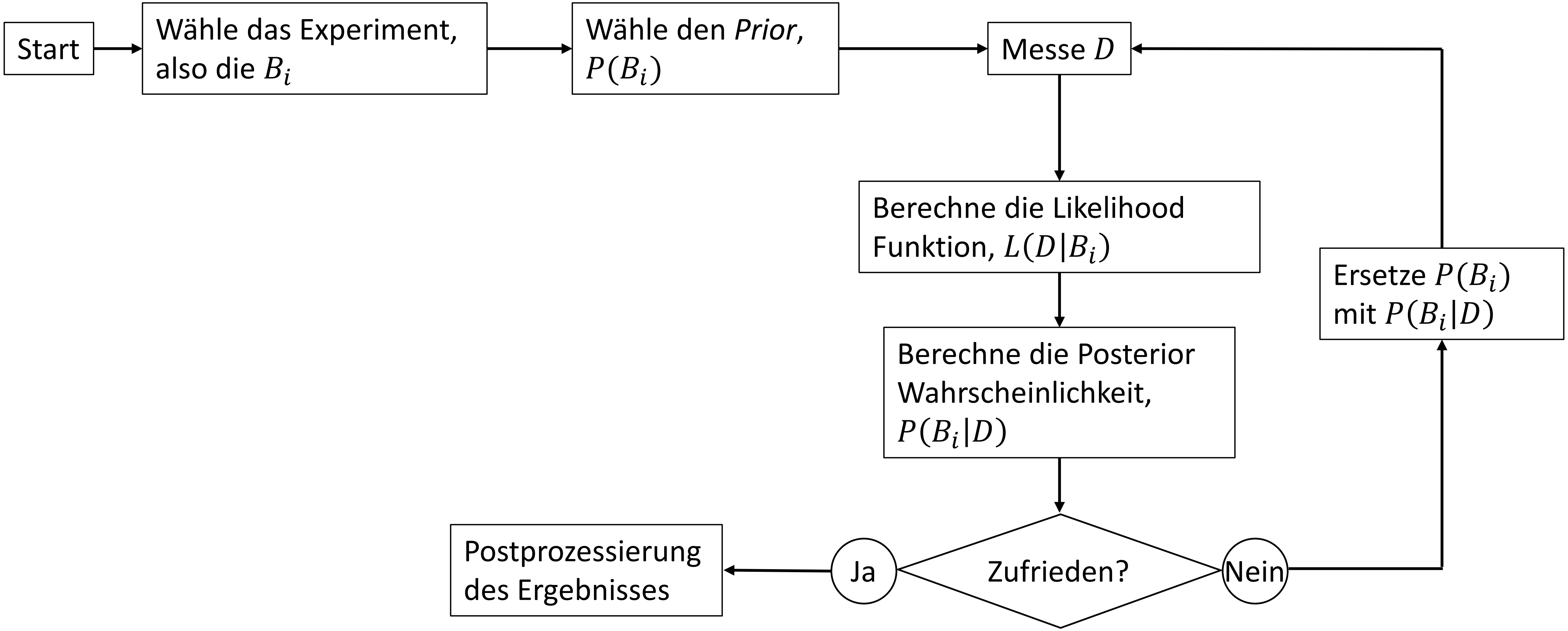
Die Wahrscheinlichkeit, dass X infiziert ist nun also 92%.

Der Ansatz der Bayesschen Datenanalyse lässt uns sehr elegant unser Ursprüngliches Wissen in die Datenanalyse einfliessen. So können wir auch anhand einer einzelnen Messung eine Aussage machen.
Mit Aussagen basierend auf der Frequenz von Ereignissen müssten wir eine grosse Statistik „erfinden“ um zum gleichen Schluss zu kommen.

Des weiteren können wir unsere Analyse um weitere Messergebnisse erweitern, bis wir mit der Präzision des Ergebnisses zufrieden sind.

Bayessche Datenanalyse

Bayessche Datenanalyse *„on the fly“*



Schätzen von Parametern

Um den Wert eines Parameters a zu schätzen verwenden wir den Ausdruck:

$$f_1(a|D) = \frac{L(D|a) f_0(a)}{\int L(D|a)\, f_0(a) da}$$

Der *Prior* $f_0(a)$ repräsentiert unser gesamtes Wissen über a bevor wir die Daten D messen.

$L(D|a)$ ist die Likelihood von a gegeben D.

Das Ergebnis $f_1(a|D)$ ist die Posterior-Wahrscheinlichkeitsverteilung von a gegeben D. (Beachte: $f_0 \neq f_1$)

Wenn die Daten aus mehreren Messungen $x: x_1, \dots, x_n$, mit ihren Unsicherheiten $\sigma: \sigma_1, \dots, \sigma_n$ bestehen benutzen wir die Vektorschreibweise:

$$f_1(a|\boldsymbol{x}, \boldsymbol{\sigma}) = \frac{L(\boldsymbol{x}, \boldsymbol{\sigma}|a) f_0(a)}{\int L(\boldsymbol{x}, \boldsymbol{\sigma}|a)\, f_0(a) da}$$

Das Ergebnis ist die PDF von a. Typischerweise wollen wir die PDF durch charakteristische Parameter wie den Mittelwert und die Standardabweichung beschreiben:

$$\hat{a} = \langle a \rangle = \int a f_1(a|D) da$$

$$\sigma_{\hat{a}}^2 = \langle a^2 \rangle - \langle a \rangle^2, \text{ mit } \langle a^2 \rangle = \int a^2 f_1(a|D) da$$

Siehe Beispiel Python Notebook:
„Bayesian Analysis of a Coin“

Schätzen von Parametern

Beispiel:

Wir wollen einen Parameter d messen. Aus vorhergegangenen Messungen haben wir die Abschätzung: $\mu_0 \pm \sigma_0$.

Wir nehmen einen gaussschen *Prior* an: $f_0(d) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{1}{2}\frac{(d-\mu_0)^2}{\sigma_0^2}\right)$

Wir messen $d_1 \pm \sigma_{d_1}$, damit wird die Likelihood Funktion:

$$L\left(d_1, \sigma_{d_1} \middle| d\right) = \frac{1}{\sqrt{2\pi}\sigma_{d_1}} \exp\left(-\frac{1}{2}\frac{(d_1 - d)^2}{\sigma_{d_1}^2}\right)$$

Die nicht normierte Posterior PDF ist dann:

$$f_1\left(d \middle| d_1, \sigma_{d_1}\right) \sim L\left(d_1, \sigma_{d_1} \middle| d\right) f_0(d) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\frac{(d_1 - \mu_1)^2}{\sigma_1^2}\right)$$

Mit: $\mu_1 = \frac{\sigma_{d_1}^2 \mu_0 + \sigma_0^2 d_1}{\sigma_0^2 + \sigma_{d_1}^2}$, $\sigma_1^2 = \frac{\sigma_0^2 \sigma_{d_1}^2}{\sigma_0^2 + \sigma_{d_1}^2}$

Auch wenn das Ergebnis keine PDF ist da es nicht normiert ist, können wir als Ergebnis den Mittelwert und Standardabweichung von f_1 extrahieren.

Zusammenfassung

- Bayessche Datenenalyse ist eine elegante Methode insbesondere aus kleinen Datensätzen sinnvolle Informationen zu extrahieren.

- Bayessche Datenanalyse erlaubt uns auf elegante Weise unser ursprüngliches Wissen in die Analyse mit einfliessen zu lassen.

- Das Ergebnis der Bayesschen Datenanalyse hängt potentiell sehr stark von der Wahl der *Priors* ab.

- Generell (für grosse Datensätze) sollte der frequentistische und Bayes-Ansatz die gleichen Ergebnisse liefern.

- Wir sollten nicht naiv oder ohne nachzudenken eine Datenanalysemethode anwenden, sondern immer überlegen was wir wissen, was wir messen und was wir erreichen wollen. Dann sollten wir die geeignete Methode wählen.