

Modellierung, Simulation und Optimierungsverfahren

Dr. Colin W. Glass

Institut für Höchstleistungsrechnen

16.04.2014

Modell Definition

Definition

Ein **Modell** ist ein künstlich geschaffenes Objekt, das wesentliche Merkmale, Beziehungen und Funktionen eines zu untersuchenden Objektes (Originals) in vereinfachter Form wiedergibt, nachbildet und damit den Prozess der Informationsgewinnung über dieses Objekt erleichtert [T. I. Ören]

Anmerkung:

Wir interpretieren Objekt sehr allgemein.

Wir interpretieren das erste "und" als "und/oder"

System

Definition

Ein System besteht aus einer Menge von Objekten, die durch Relationen verbunden sind. Es wird von seiner Umgebung durch Systemgrenzen getrennt. Die Umgebung kann nur durch definierte Relationen mit dem System interagieren. Für die Modellierung des Systems werden ausschließlich die definierten Relationen berücksichtigt.

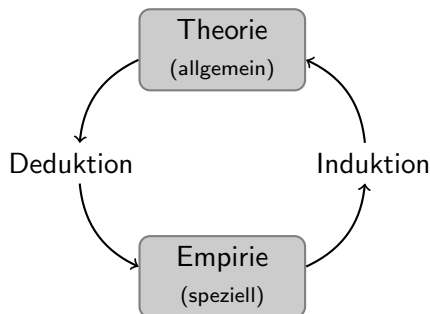
Herangehensweisen bei der Modellbildung

- **deduktiv**

Schluss von einer allgemeingültigen Theorie auf einen speziellen Fall:
Ein Modell wird aus z.B. bekannten mathematischen Theorien abgeleitet.

- **induktiv**

Ableitung einer allgemeinen Erkenntnis aus der Beobachtung:
Ausgehend von z.B. Messdaten wird ein Modell entwickelt.



Die zwei wichtigsten Werkzeuge

- Zahlen: Eigenschaften eines Systems quantifizieren
- Klassen: Entitäten des Systems unterscheiden

Die zwei wichtigsten Werkzeuge

- **Zahlen: Eigenschaften eines Systems quantifizieren - Heute!**
- **Klassen: Entitäten des Systems unterscheiden**

Modellierung Grundlagen

Quelle: genau-dein-obst.de



Delbarestivale®



Jamba



Gravensteiner

Modellierung Grundlagen

Quelle: genau-dein-obst.de



Delbarestivale®



Jamba



Gravensteiner

Zahlen (Anzahl): 1, 2, 3

Modellierung Grundlagen

Quelle: genau-dein-obst.de



Delbarestivale®



Jamba



Gravensteiner

Zahlen (Gramm): 142, 165, 152

Modellierung Grundlagen

Quelle: genau-dein-obst.de



Delbarestivale®



Jamba



Gravensteiner

Klassen (Obst): Apfel, Apfel, Apfel

Modellierung Grundlagen

Quelle: genau-dein-obst.de



Delbarestivale®



Jamba



Gravensteiner

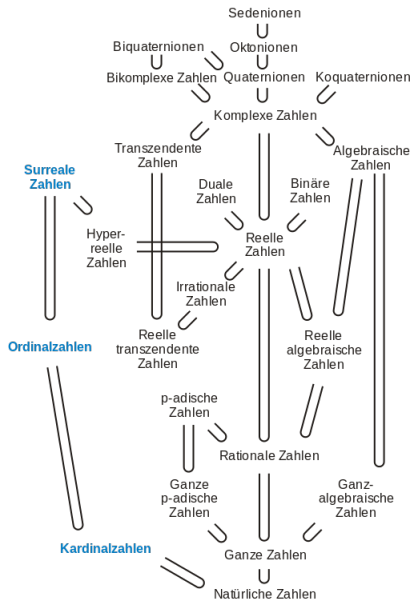
Klassen (Apfelsorte): Delbarestivale, Jamba, Gravensteiner

Zahlen

Definition

Zahlen sind abstrakte mathematische Objekte, also Objekte des menschlichen Denkens, die sich aus Vorstellungen von „Größe“ entwickelt haben: Durch eine Messung wird ein als Größe verstandener Aspekt einer Beobachtung mit einer Zahl in Verbindung gebracht (...). [Wiki]

Ein Vergleich mit der Definition von Modell zeigt: Objekte durch Zahlen zu repräsentieren resultiert in einem Modell.



Quelle: Wikipedia

Variablen und Konstanten

- Variablen: Veränderlich, diskret oder stetig
 - Freie Variable: Wert kann innerhalb einer Definitionsmenge frei gewählt werden.
 - Zufallsvariable: Wert hängt vom Ereignis eines Zufallsexperiments ab.
 - Abhängige Variable: Wert ist fest an den Wert anderer Variablen gebunden.
 - Parameter: an und für sich frei wählbar, aber innerhalb eines Kontexts konstant.
- Konstanten: Unveränderlich
 - Mathematisch: eine Zahl, die in der Mathematik von besonderem Interesse ist, z.B. π
 - Physikalisch: ein Messwert der Physik, deren Größenwert sich weder räumlich noch zeitlich verändert, z.B. Lichtgeschwindigkeit im Vakuum.

Zufallsvariable und Verteilungen

Definition

Eine Zufallsvariable X ist ein Zufallsexperiment mit möglichen Werten in $\mathbb{R}$, bzw. in einer Teilmenge von $\mathbb{R}$. Deren Wert ist im voraus nicht bekannt, sondern hängt vom Ergebnis des Zufallsexperiments ab. Wenn das Ergebnis ω herauskommt, nimmt die Zufallsvariable den Wert $X(\omega)$ an.

Eine Zufallsvariable X legt eine Wahrscheinlichkeit Q auf $\mathbb{R}$ fest, die sogenannte Verteilung von X . [Peter Bühlmann]

$$Q(B) = P[\{\omega; X(\omega) \in B\}] = P[X \in B] \quad (B \subseteq \mathbb{R})$$

- Beispiele: Würfeln, Markov Chains, Börse etc.

Wahrscheinlichkeitsverteilung

Eine Wahrscheinlichkeitsverteilung beschreibt, wie sich die Wahrscheinlichkeiten auf die möglichen Zufallsergebnisse (Zustände) verteilen.

Die Wahrscheinlichkeitsverteilung ist das theoretische Pendant zur Häufigkeitsverteilung, die aus empirischen Daten erstellt wird.

Kennzahlen von Verteilungen

Eine Verteilung einer Zufallsvariable kann durch Kennzahlen beschrieben werden. Üblicherweise durch den Erwartungswert und die Varianz:

$$\begin{aligned} E[X] &= \mu && \text{(Erwartungswert, auch } \mu_x \text{ genannt)} \\ E[(X - \mu)^2] &= \mu_2 = \sigma^2 && \text{(Varianz, } \sigma: \text{ Standardabweichung)} \end{aligned}$$

Erwartungswert $\hat{=}$ Lage der Zufallsvariable

Varianz $\hat{=}$ Streuung der Zufallsvariable (2. zentrales Moment)

Wahrscheinlichkeitsfunktion

Eine Zufallsvariable X ist diskret, wenn die Menge W der möglichen Werten von X abzählbar ist, z.B. $W = \mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$

Die Verteilung ist gegeben durch die Wahrscheinlichkeitsfunktion:

$$p(x_i) = P[X = x_i] \text{ für } x_i \in W$$

$$E[X] = \mu = \sum_{x_i \in W} x_i p(x_i) \quad \text{Erwartungswert}$$

$$E[(X - \mu)^2] = \sigma^2 = \sum_{x_i \in W} (x_i - \mu)^2 p(x_i) \quad \text{Varianz}$$

Diskret: Geometrische Verteilung

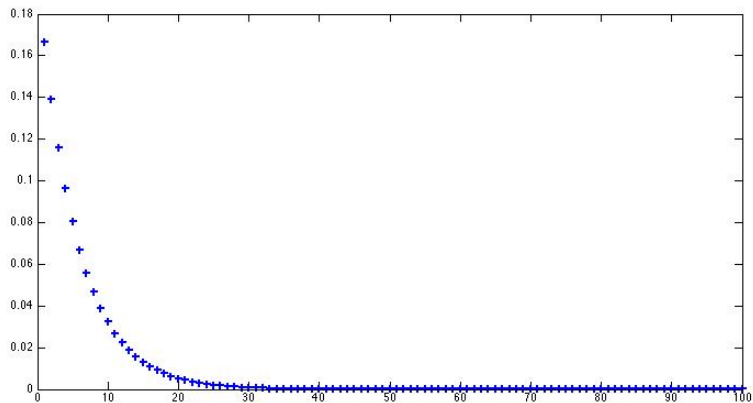
Die Anzahl Wiederholungen bis ein Ereignis mit Wahrscheinlichkeit p eintritt ist geometrisch verteilt. Z.B. wie oft würfeln bis eine 6 fällt.

$$p(x) = p(1 - p)^{x-1}, \quad W = \{1, 2, 3, \dots\}$$

$$\begin{aligned} E[X] &= \mu = 1/p \\ E[(X - \mu)^2] &= \sigma^2 = \frac{1-p}{p^2} \end{aligned}$$

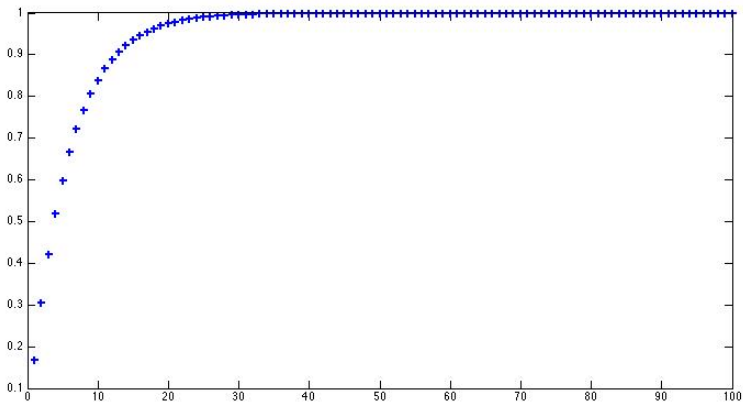
P.s: Es gibt eine zweite Definition der Geometrischen Verteilung.

Diskret: Geometrische Verteilung



$$p = 1/6 \quad \mu = 6 \quad \sigma^2 = 30$$

Diskret: Kumulative Geometrische Verteilung



Diskret: Binomialverteilung

Die Anzahl Erfolge bei n unabhängigen Wiederholungen eines Experiments mit Erfolgswahrscheinlichkeit p ist binomial verteilt. Z.B. wie viele 6 bei 100 mal würfeln.

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad W = \{0, 1, 2, \dots, n\}$$

$$\begin{aligned} E[X] &= \mu = n \cdot p \\ E[(X - \mu)^2] &= \sigma^2 = n \cdot p(1-p) \end{aligned}$$

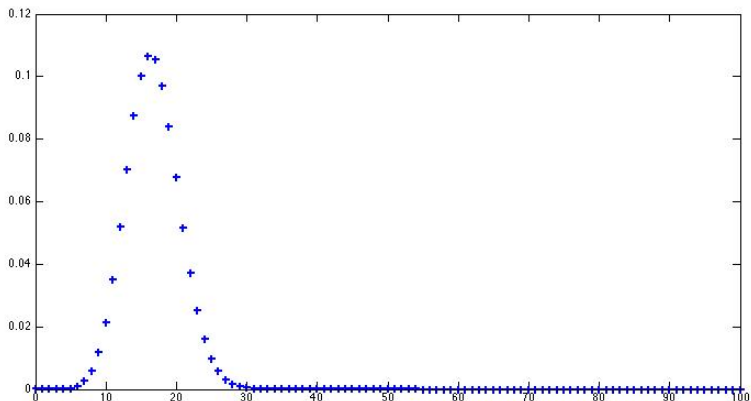
Diskret: Binomialverteilung

$\binom{n}{x}$ ist der Binomialkoeffizient: er quantifiziert die Anzahl Arten, x Objekte aus einer Menge von n Objekten auszuwählen (ohne Beachtung der Reihenfolge, ohne Zurücklegen).

Ist n eine nichtnegative ganze Zahl und $n \geq x$, so gilt:

$$\binom{n}{x} = \frac{n!}{x! \cdot (n-x)!}$$

Diskret: Binomialverteilung



$$n = 100 \quad p = 1/6 \quad \mu = 16.67 \quad \sigma^2 = 13.89$$

Wahrscheinlichkeitsdichte

Für eine stetige Zufallsvariable X ist $P[X = x] = 0$ für jedes x . Deshalb wird hier die Dichte betrachtet:

$$f(x) = \frac{P[x \leq X \leq x + dx]}{dx}$$

$$E[X] = \mu = \int_{-\infty}^{\infty} xf(x)dx \quad \text{Erwartungswert}$$

$$E[(X - \mu)^2] = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx \quad \text{Varianz}$$

Stetig: Uniforme Verteilung

Die uniforme Verteilung wird gerne bei fehlenden Kenntnissen des Sachverhalts verwendet. Kommt auch tatsächlich vor, z.B. für Rundungsfehler (je nach dem wo die Zahlen her kommen, die gerundet werden).

$$f(x) = \frac{1}{(b-a)} \quad \text{für } a \leq x \leq b, \quad W = [a, b]$$

$$\begin{aligned} E[X] &= \mu = \frac{a+b}{2} \\ E[(X - \mu)^2] &= \sigma^2 = \frac{(b-a)^2}{12} \end{aligned}$$

Stetig: Normalverteilung

Wichtig - Zentraler Grenzwertsatz besagt: eine Summe von unabhängigen, identisch verteilten Zufallsvariablen mit endlicher Varianz ist im Grenzwert normalverteilt.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp - \frac{(x - \mu)^2}{2\sigma^2}, \quad W = \mathbb{R}$$

μ = Erwartungswert

σ = Standardabweichung

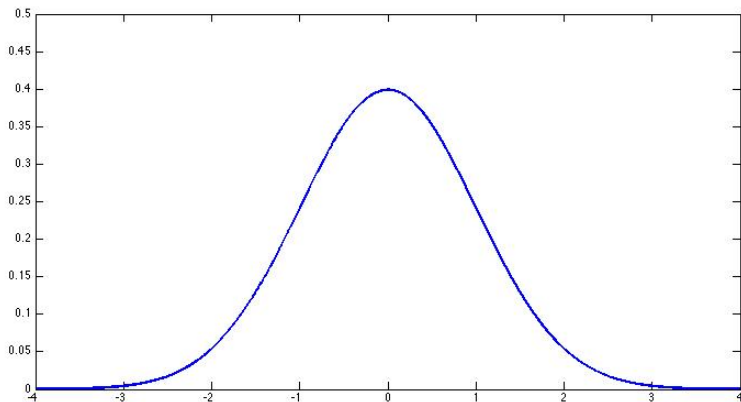
P.s: Abstand von μ vs. Wahrscheinlichkeit:

Innerhalb σ $\rightarrow$ 68,27 % aller Werte

Innerhalb 2σ $\rightarrow$ 95,45 % aller Werte

Innerhalb 3σ $\rightarrow$ 99,73 % aller Werte

Stetig: Normalverteilung



$$\mu = 0 \quad \sigma^2 = 1$$

Weitere Kennzahlen von Verteilungen

Zwei weitere praktische Kennzahlen: Skewness (Schiefe) und Kurtosis (Wölbung) basieren auf dem 3. und 4. zentralen Moment, normiert auf die Varianz (standardized moment).

$$\begin{aligned}
 E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] &= \frac{\mu_3}{\sigma^3} &=: \gamma_1 & \quad (\text{Skewness (Schiefe)}) \\
 E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] &= \frac{\mu_4}{\sigma^4} &=: \beta_2 & \\
 E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3 &= \frac{\mu_4}{\sigma^4} - 3 &= \beta_2 - 3 =: \gamma_2 & \quad (\text{Kurtosis (Wölbung)})
 \end{aligned}$$

Weitere Kennzahlen von Verteilungen

Skewness $\hat{=}$ Maß für die Asymmetrie der Verteilung

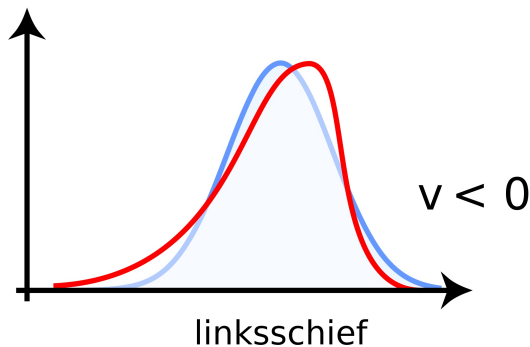
Kurtosis $\hat{=}$ Maß für die "Steilheit" bzw. Maß für die "Endlastigkeit"
(verschiedene Interpretationen)

Skewness und (excess) Kurtosis sind für Normalverteilungen = 0.

Es gibt nicht nur verschiedene Interpretationen von Kennzahlen, es gibt auch verschiedene Kennzahlen. . . , z.B. die Kumulanten. Diese werden hier nicht betrachtet.

Weitere Kennzahlen von Verteilungen

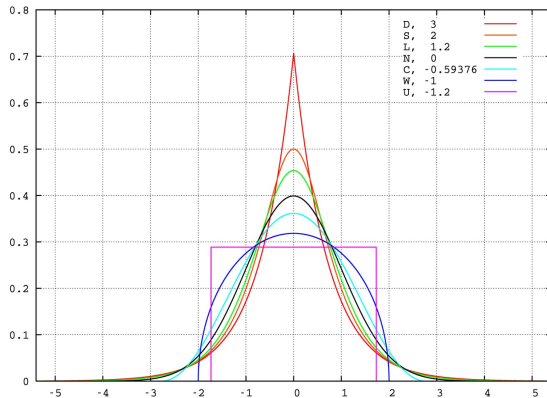
Skewness (Schiefe)



[Chrischi]

Weitere Kennzahlen von Verteilungen

Kurtosis (Wölbung)



[Wiki]

Berechnung von Kennzahlen

Für Wahrscheinlichkeitsverteilungen:

$p(x)$ (diskret) bzw. $f(x)$ (stetig) in die entsprechende Summe bzw. Integral (Erwartungswerte $E[\dots]$) einsetzen, lösen.

Für Häufigkeitsverteilungen:

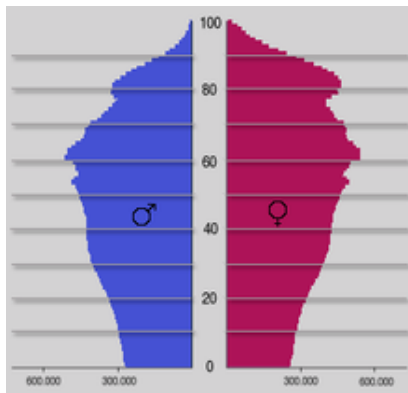
$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{Arithmetisches}) \text{ Mittel} \hat{=} \mu$$

$$\hat{\sigma}^2 = S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{korrigierte Stichprobenvarianz}$$

Häufigkeitsverteilung (Induktiv)

Gemessenen Verteilung einer Zufallsvariable. Statistiken beruhen auf Häufigkeitsverteilungen.

Beispiel: Alter von Menschen.



[Wiki]

Beispiel: Risiko

- Lustiges Brettspiel
- Zwei Armeen treffen auf einander, es geht darum ein Land zu befreien ;)
- Wer gewinnt?
- Verfahren:
 - Angreifer 3 Würfel, schlechtesten verwerfen
 - Verteidiger 2 Würfel
 - Zahlen sortieren, pärenchenweise vergleichen
 - Pro Pärchen: Angreifer gewinnt wenn $>$ Verteidiger
 - ... und natürlich die ganzen Spezialfälle für weniger wie 3 Angreifer, bzw. weniger wie 2 Verteidiger
- Kompliziert!

Beispiel: Risiko - Algorithmus

```

function Armies_Left = Beispiel_Risiko(Armeen_a, Armeen_b)
while (Armeen_a>0.5)&& (Armeen_b>0.5)
    if Armeen_a>2.5
        Wurf_A = sort(randi(6,3,1));
        Wurf_A (1) =[];
    elseif Armeen_a >1.5
        Wurf_A = sort(randi(6,2,1));
    else
        Wurf_A = sort(randi(6,1,1));
    end

    if Armeen_b>1.5
        Wurf_B = sort(randi(6,2,1));
    else
        Wurf_B = sort(randi(6,1,1));
    end

    if length(Wurf_A)>length(Wurf_B)
        Wurf_A (1) =[];
    elseif length(Wurf_B)>length(Wurf_A)
        Wurf_B (1) =[];
    end

    Success_A = sum(Wurf_A > Wurf_B);

    Armeen_b = Armeen_b -Success_A;
    Armeen_a = Armeen_a -length(Wurf_A) +Success_A;

end

if Armeen_a > Armeen_b
    Armies_Left = Armeen_a;
else
    Armies_Left = -Armeen_b;
end

```

P.s: Dieses Programm ist nicht nach best-practice entstanden ...

Einschub: Monte Carlo

Monte Carlo ist ein Verfahren aus der Stochastik, bei welchem die (einer Zufallsvariable zugrundeliegende) Wahrscheinlichkeitsdichte bzw. -verteilung durch eine Häufigkeitsverteilung angenähert wird.

Die Häufigkeitsverteilung wird durch häufiges durchführen des entsprechenden Zufallsexperiments erzeugt. Dies kann ein reales oder virtuelles Zufallsexperiment sein. Als Grundlage ist vor allem das Gesetz der großen Zahlen zu sehen.

Monte Carlo kommt insbesondere dann zum Einsatz, wenn die theoretische Wahrscheinlichkeit sehr schwierig zu bestimmen ist.

Beispiel: Risiko - Monte Carlo!

Szenario:

- immer 100 vs. 100 bis einer gewinnt
- 10000 Wiederholungen
- Positive Zahlen: Angreifer gewinnt und hat noch Leute
- Negative Zahlen: Verteidiger gewinnt und hat noch Leute

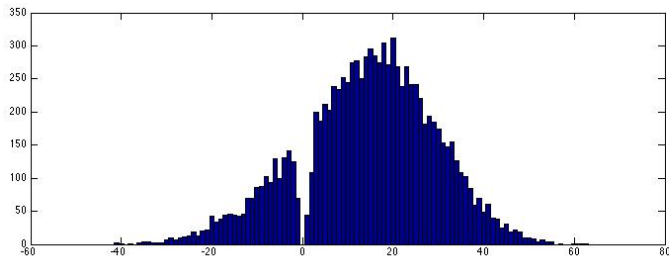
Häufigkeitsverteilung ... wer gewinnt?

Beispiel: Risiko - Monte Carlo!

Szenario:

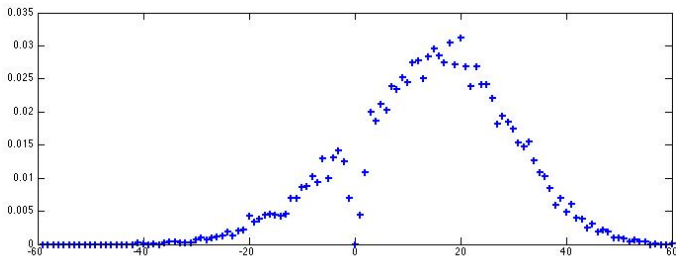
- immer 100 vs. 100 bis einer gewinnt
- 10000 Wiederholungen
- Positive Zahlen: Angreifer gewinnt und hat noch Leute
- Negative Zahlen: Verteidiger gewinnt und hat noch Leute

Häufigkeitsverteilung ... wer gewinnt?



Beispiel: Risiko - Häufigkeitsverteilung

Berechnet: Mittelwert = 14.48 $\hat{\sigma} = 14.9231$



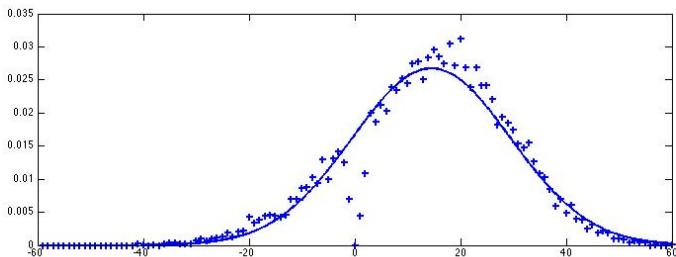
Fitting (Ausgleichsrechnung)

Fitting ist eine Optimierungsmethode, um für eine Reihe von Messdaten die unbekannten Parameter ihres Modells oder die Parameter einer vorgegebenen Funktion zu bestimmen (zu schätzen).

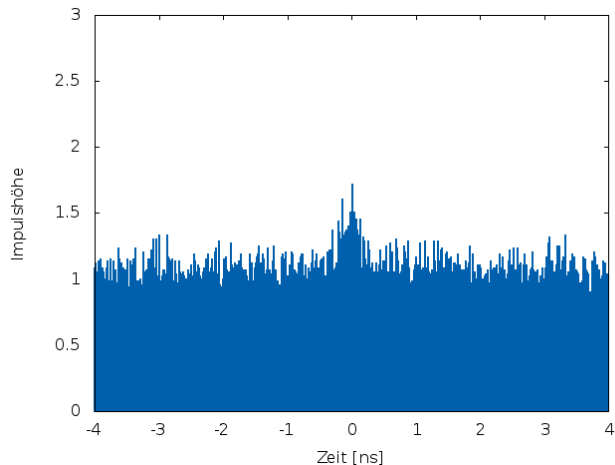
Beispiel: Risiko - Häufigkeitsverteilung

Berechnet: Mittelwert = 14.48 $\hat{\sigma} = 14.9231$

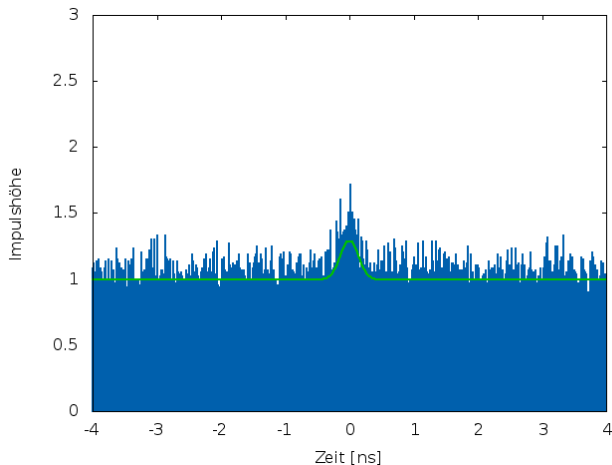
Normalverteilung:



Beispiel: Photonenstatistik



Beispiel: Photonenstatistik



$$f(x) = 1 + \frac{a^2}{2\pi\sigma^2} \cdot \exp\left(-\frac{x^2}{\sigma^2}\right) \quad \text{Quadrierte Normalverteilung}$$

Mehrere Zufallsvariablen

Gemeinsame Verteilungen von Zufallsvariablen werden schnell kompliziert. Z.B.: die Verteilung von zwei stetigen Zufallsvariablen kann über die gemeinsame Dichte beschrieben und die Wahrscheinlichkeit durch Integration bestimmt werden.

$$P[x \leq X \leq x + dx, y \leq Y \leq y + dy] = f_{X,Y}(x, y) dx dy$$

Die Dichte und die gemeinsame Verteilung von abhängigen Zufallsvariablen sind i.A. sehr kompliziert, deshalb wird oft auf eine Kennzahl zurückgegriffen: die Kovarianz bzw. Korrelation. Wir werden uns darauf beschränken.

Kovarianz und Korrelation

Die Kovarianz und die Korrelation zwischen zwei Zufallsvariablen X und Y sind definiert über:

$$\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)] \text{ (Kovarianz)}$$

$$\text{Corr}(X, Y) = \rho_{xy} = \text{Cov}(X, Y) / (\sigma_x \sigma_y) \text{ (Korrelation)}$$

Folglich misst die Korrelation sowohl Stärke als auch Richtung der linearen (!) Abhängigkeit zwischen X und Y .

Kausalität

Wenn die Veränderung einer Variable, Veränderung einer anderen bedingt, spricht man von einem kausalen Zusammenhang.

Kausalität bezeichnet die Beziehung zwischen Ursache und Wirkung, betrifft also die Abfolge aufeinander bezogener Ereignisse und Zustände. Die Kausalität (ein kausales Ereignis) hat eine feste zeitliche Richtung, die immer von der Ursache ausgeht, auf die die Wirkung folgt. [Wiki]

Kausalität lässt sich nicht zuverlässig induktiv bestimmen!!!

Cum hoc ergo propter hoc: mit diesem, also deswegen

Der Fehlschluss von Korrelation auf Kausalität wird auch als Cum hoc ergo propter hoc bezeichnet.

Beispiel:

- Schlafen gehen ohne die Schuhe auszuziehen, ist korreliert mit Kopfschmerzen am nächsten Morgen.

Cum hoc ergo propter hoc: In Schuhen schlafen verursacht Kopfschmerzen.

Vermutlich kommt jedoch die Korrelation daher, dass "in Schuhen schlafen" und "Kopfscherzen" eine gemeinsame Ursache haben: übermässiger Alkoholkonsum.

Metriken

Wir betrachten hier Metrik als eine Größe, welche eine relevante Eigenschaft eines Systems wiedergibt und zuverlässig quantifiziert.

In sehr vielen Disziplinen ist die Formulierung einer guten Metrik schwierig und sehr zentral.

Metriken

Beispiele für knifflige Metriken:

- Gini-Koeffizient
- Inflation
- "presidential job approval ratings"
- Big Mac Index
- GFlops (Linpack)
- ...

Metriken

Gini-Koeffizient:

Ein Maß für die Ungleichverteilung einer Größe. Der Gini-koeffizient wird gerne für die Einkommens- bzw. Vermögensverteilung in einer Bevölkerung verwendet.

$$G = \frac{2\sum_{i=1}^n iy_i}{n\sum_{i=1}^n y_i} - \frac{n+1}{n}$$

Metriken

Inflation:

Ein Maß für die Abwertung einer Währung über die Zeit. Wird oft mittels eines Warenkorbs bestimmt (consumer price index).

Metriken

"presidential job approval ratings":

Ein Maß für die Unterstützung des amerikanischen Präsidenten seitens des Volks. Basiert auf Stichproben-Umfragen.

Metriken

GFlops (Linpack):

Maß für die Güte eines Supercomputers. Es wird Linpack - ein Solver, also ein Algorithmus der lineare Gleichungssysteme löst - auf der ganzen Maschine laufen gelassen. Die Anzahl Flops (Floating point operations per second) ist die Metrik.

Metriken

Big Mac Index:

Maß für die Kaufkraft verschiedener Währungen gemessen anhand von genau einem Gut: dem Big Mac.

Literatur: Stochastik (und Statistik)

"Mini-Skript zur Vorlesung Stochastik" Peter Bühlmann, ETH Zürich.

Google: "Miniskript Stochastik Peter Bühlmann"

Erster Treffer - ausser Google hat eure Interessen anders modelliert wie meine ;)

Super Skript! ... und weniger wie 30 Seiten!
Verteilungen, Wahrscheinlichkeitsfunktion etc.