Zusammenfassung: Datenverarbeitung

Lars Wenning

3. Dezember 2015

Inhaltsverzeichnis

1 Messwert und Messgenauigkeit

- Systematischer Fehler: Fehler bei jeder Messung gleich.
- Statistischer Fehler: Streut um wahren Wert
- Breite der Wahrscheinlichkeitsverteilung ist über Standardabweichung abschätzbar. Diese gibt die Streuung einzelner Messwerte an.
- Mittelwert einer Stichprobe:

$$\langle x \rangle = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{1}$$

- Varianz: Quadratische Abweichung der Messwerte vom Mittelwert:

$$V = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \langle x \rangle)^2 \tag{2}$$

- Standardabweichung $\widehat{=}$ Streuung der Messwerte $\widehat{=}$ Fehler:

$$\sqrt{V} = \sigma \tag{3}$$

- Bei n Messungen liegen 68% $\approx \frac{2}{3}$ aller Messwerte innerhalb von $\pm$ Standardabweichung um den Mittelwert $\langle x \rangle$
- Fehler des Mittelwerts:

$$\sigma_{\langle x \rangle} = \frac{\sigma}{\sqrt{n}} \tag{4}$$

- Fehler der Standardabweichung:

$$\sigma_\sigma = \frac{\sigma}{\sqrt{2(n-1)}} \tag{5}$$

2 Wahrscheinlichkeit

2.1 Kombinatorik

- Wenn ein Ereignis auf n verschiedene gleichwahrscheinliche Arten eintreten kann wovon k die Eigenschaft A haben, dann ist die Wahrscheinlichkeit für das eintreten des Ereignisses:

$$P(A) = \frac{k}{n} \tag{6}$$

- Möglichkeiten r unterscheidbare Objekte in Reihe anzuordnen:

$$N = r! \tag{7}$$

- Die Reihenfolge ist wichtig $\Rightarrow$ Möglichkeiten k Objekte aus n Objekten in Reihe anzuordnen:

$$N = \frac{n!}{(n-k)!} \tag{8}$$

- Die Reihenfolge ist unwichtig $\Rightarrow$ Möglichkeiten k Objekte aus n Objekten in Reihe anzuordnen:

$$N = \binom{n}{k} = \frac{n!}{(n-k)!} \tag{9}$$

2.2 Kombination von Wahrscheinlichkeiten

Es gilt:

$$P(A \underbrace{\vee}_{\text{Disjunktion}} B) = P(A) + P(B) - P(A \underbrace{\wedge}_{\text{Konjunktion}} B) \tag{10}$$

- Disjunktion: Aussage A oder B oder beide.
- Konjunktion: Aussage A und B.

Spezialfälle

- A und B schließen sich aus $\Rightarrow$

$$P(A \wedge B) = 0 \Rightarrow P(A \vee B) = P(A) + P(B) \tag{11}$$

- B ist Negation von A $\Rightarrow$

$$B = \bar{A} \Rightarrow P(A \wedge B) = 0 \Rightarrow P(A \vee B) = P(A \vee \bar{A}) = P(A) + P(\bar{A}) = 1 \tag{12}$$

- A und B schließen sich nicht aus $\Rightarrow$ Definition der Bedingten Wahrscheinlichkeit

$$P(A \wedge B) \geq 0 \Rightarrow P(\underbrace{B|A}_{\text{B bedingt A}}) = \frac{P(A \wedge B)}{P(A)} \tag{13}$$

Bedingt bedeutet hier, dass B auftritt wenn A bereits aufgetreten ist.

Satz von Bayes

$$\begin{aligned} &\textit{Sei } P(A \wedge B) = P(B \wedge A) \\ &\Rightarrow P(A)P(A|B) = P(B)P(A|B) \\ &\Leftrightarrow P(A|B) = P(B|A)\frac{P(A)}{P(B)} \end{aligned} \tag{14}$$

3 Wahrscheinlichkeitsverteilungen

3.1 Zufalls-variablen und Messdaten

Verteilung von diskreten Zufalls-variablen

- $r \in \mathbb{Z}$
- Wahrscheinlichkeit dass Wert auftritt: $P(r)$
- Normierung: $\sum_{j=1}^{n} P(r_j) = 1$
- Wahrscheinlichkeit, dass Werte zwischen a und b auftreten:

$$P(i_a \leq i \leq i_b) = \sum_{i=i_a}^{i_b} P_i \tag{15}$$

Verteilung von kontinuierlichen Zufalls-variablen

- $r \in \mathbb{R}$
- Wahrscheinlichkeit dass exakter Wert auftritt: $P(r) = 0$
- Normierung: $P(-\infty \leq x \leq \infty) = \int\limits_{-\infty}^{\infty} f(x)dx = 1$
  wobei f(x) die Wahrscheinlichkeits(-dichte)verteilung ist.
- Wahrscheinlichkeit, dass Werte zwischen a und b auftreten:

$$P(a \leq x \leq b) = \int_a^b f(x)dx \tag{16}$$

- $F(x_0) = \int\limits_{-\infty}^{x_0} f(x)dx$ gibt Wahrscheinlichkeit an, dass x in $x \leq x_0$ liegt.
- $f(x) \geq 0$

3.2 Kenngrößen für Wahrscheinlichkeitsverteilungen

- Mittelwert diskret:
$$\langle x\rangle = \sum_{i=1}^{n} x_i P(x_i) \tag{17}$$

- Erwartungswert von einer Funktion h(x) ist definiert als
$$E[h] = \int_{-\infty}^{\infty} h(x)f(x)dx \tag{18}$$
$$\rightarrow E[x^n] = \text{n-tes algebraisches Moment} \tag{19}$$
$$\rightarrow E[(x-\langle x\rangle)^n] = \text{ n-tes zentrales Moment} \tag{20}$$

- Mittelwert kontinuierlich:
$$\langle x\rangle = \int_{-\infty}^{\infty} xf(x)dx \tag{21}$$
entspricht dem ersten algebraischen Moment $\langle x\rangle = E[x]$

- Wahrscheinlichster Wert: Das Maximum von f(x) zeigt den Wahrscheinlichsten Wert an.
$$f_{max} = f(x_{max}) \tag{22}$$

- Median: Zufallswert liegt mit gleicher Wahrscheinlichkeit unterhalb und oberhalb des Median.
$$\int_{-\infty}^{x_{med}} f(x)dx \stackrel{!}{=} \int_{x_{med}}^{\infty} f(x)dx \stackrel{!}{=} 0,5 \tag{23}$$

- Mittelwert $\langle x\rangle$ und wahrscheinlichster Wert x_{max} können weit auseinander liegen.

- Getrimmter Mittelwert: Bei n Messungen werden die kleinsten und und größten Werte ausgelassen.

- Varianz entspricht 2. zentralem Moment

- Root Mean Square(RMS): Wurzel aus dem 2. algebraischen Moment. Auch "Quadratisches Mittel". Mittelwert, bei dem größere Werte einen stärkeren Einfluss haben als kleinere.
$$x_{RMS} = \sqrt{V[x] + \langle x\rangle^2} \tag{24}$$

- Full Width Half Maximum(FWHM): Auch Halbwertbreite
  - Schätzung der Streuung aus Grafischer Verteilung.
  - von f_{max} ausgehend wird zu großen und kleinen x-Werten die Position gesucht an denen die Verteilung auf die Hälfte abgesunken ist.

$$\begin{aligned} f(x_{+\frac{1}{2}}) = f(x_{-\frac{1}{2}}) &= \frac{f_{max}}{2} \\ \Rightarrow FWHM &= |x_{+\frac{1}{2}} - x_{-\frac{1}{2}}| \end{aligned} \tag{25}$$

  - anschaulich die Breite bei halber Höhe.
  - Über Standardabweichung:

$$FWHM = 2\sqrt{2ln(2)}\sigma \tag{26}$$

3.3 Wichtigste Wahrscheinlichkeitsverteilungen

3.3.1 Gleichverteilung (Kontinuierlich)

- Konstanter Funktionswert für alle Werte zwischen $x = a$ und $x = b$

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{sonst} \end{cases} \tag{27}$$

- Normierung durch Faktor $\frac{1}{b-a}$ erfüllt.
- Mittelwert:

$$\langle x \rangle = \mu = \frac{a+b}{2} \tag{28}$$

- Varianz:

$$V[x] = \frac{(b-a)^2}{12} \tag{29}$$

- Standardabweichung:

$$\sigma = \sqrt{V} = \frac{b-a}{\sqrt{12}} \tag{30}$$

3.3.2 Binomialverteilung (diskret)

- p ist Wahrscheinlichkeit, dass E eintritt. Dann ist $(1-p)$ Wahrscheinlichkeit, dass E nicht eintritt.

- E trete bei den ersten r aber nicht bei den letzten $n-r$ Versuchen auf. $\Rightarrow P = p^r(1-p)^{n-r}$ ist die Gesamtwahrscheinlichkeit. Die r treten ohne feste Reihenfolge auf. Dann ist die Wahrscheinlichkeit, dass E mit der Wahrscheinlichkeit p in n Versuchen auftritt:

$$P = \binom{n}{r} p^r (1-p)^{n-r} \tag{31}$$

- Die Binomialverteilung beschreibt also die Anzahl der Erfolge in einer Serie von gleichartigen und unabhängigen Versuchen, die jeweils genau zwei mögliche Ergebnisse haben („Erfolg“ oder „Misserfolg“). Solche Versuchsserien werden auch Bernoulli-Prozesse genannt.

- Mittelwert:

$$\langle r \rangle = \mu = np \tag{32}$$

- Varianz:

$$V[x] = np(1-p) \tag{33}$$

- Standardabweichung:

$$\sigma = \sqrt{np(1-p)} \tag{34}$$

Bilder/399px-Galton-Brett.png

Abbildung 1: Galtonbrett

Bilder/692px-Pascal's_triangle;_binomial_distributi

Abbildung 2: Pascal's triangle binomial distribution. The area of all bars in a row is always 1.

3.3.3 Poisson Verteilung (diskret)

- Gibt wie Binomialverteilung die Wahrscheinlichkeit P für das r-fache Aufteten in n Versuchen an, nur dass n groß und das Eintreten eines einzelnen Ereignisses sehr unwahrscheinlich ist. ($p \ll 1$)

- Grenzfall für Binomialverteilung bei dem die Varianz ungefähr dem Mittelwert entspricht.

$$\begin{aligned} V &\approx \mu = np \\ \Leftrightarrow p &= \frac{\mu}{n} \text{ in Binomialverteilung einsetzen} \\ n &\to \infty \\ \Rightarrow P(r) &= e^{-\mu}\frac{\mu^r}{r!} \end{aligned} \tag{35}$$

- Mittelwert:
$$\langle r \rangle = \mu = np \tag{36}$$

- Varianz:
$$V[x] = \mu = np \tag{37}$$

- Standardabweichung:
$$\sigma = \sqrt{np} \tag{38}$$

3.3.4 Gaussverteilung Normalverteilung (Kontinuierlich)

- Eine stetige Zufallsvariable x mit der Wahrscheinlichkeitsdichte f(x) gegeben durch:
$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{39}$$
heißt Normalverteilt mit den Parametern mit dem Erwartungswert μ und der Standardabweichung σ.

- Vorfaktor ergibt sich aus der Normierungsbedingung für Wahrscheinlichkeitsdichten.

- Zufallsgrößen mit Normalverteilung benutzt man zur Beschreibung zufälliger Vorgänge wie: zufällige Messfehler, zufällige Abweichungen vom Sollmaß bei der Fertigung von Werkstücken, Beschreibung der brownschen Molekularbewegung.

- $$f_{max} = \frac{1}{\sqrt{2\pi\sigma}} \tag{40}$$

- Varianz:
$$V[x] = \sigma^2 \tag{41}$$

- Im Fall $\mu = 0$ und $\sigma^2 = 1$ wird diese Verteilung Standardnormalverteilung genannt.

- bei Aufgaben hilft die Substitution $Z = \frac{x-\mu}{\sigma}$. Dadurch wird die Verteilung in eine Standardnormalverteilung transformiert.

- f ist eine Gaußsche Glockenkurve, deren Höhe und Breite von σ abhängt. Sie ist achsensymmetrisch zur Geraden der Gleichung $x = \mu$

- um zu überprüfen ob es sich um eine Gaussverteilung handelt, kann beispielsweise der Chi-Quadrat-Test angewandt werden.

Abbildung 3: Dichtefunktion der Normalverteilung

Bilder/500px-Normal_Distribution_PDF.png

3.4 Zweidimensionale Wahrscheinlichkeitsdichten

- Zufallsprozess hängt von mehreren Zufallsvariablen ab
- Einziger wichtiger Unterschied zu eindimensionalen Dichten ist, dass die Variablen nun in Abhängigkeit voneinander Streuen können(Korrelationen).
- Wahrscheinlichkeit, dass Zufallsvariablen (x, y) zwischen $a \leq x \leq b, c \leq y \leq d$ liegen:

$$P = \int_a^b \int_c^d f(x,y)dxdy \tag{42}$$

- Normierung:

$$P = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy = 1 \tag{43}$$

- Mittelwerte:

$$\langle x \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x,y)dxdy \tag{44}$$

$$\langle y \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x,y)dxdy \tag{45}$$

- Varianzen:

$$V[x] = \sigma_x^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \langle x \rangle)^2 f(x,y)dxdy \tag{46}$$

$$V[y] = \sigma_y^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \langle y \rangle)^2 f(x,y)dxdy \tag{47}$$

- Kovarianz beschreibt Korrelationen zwischen x und y:

$$\sigma_{xy} = \sigma_{yx} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \langle x \rangle)^2 (y - \langle y \rangle)^2 f(x,y)dxdy \tag{48}$$

- Kovarianz Matrix

$$\hat{V} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix} \tag{49}$$

- alternative Berechnung für den Korrelationskoeffizient:

$$\sigma_{xy} = \rho\sqrt{\sigma_x^2 \sigma_y^2} mit - 1 \leq \rho \leq 1 \tag{50}$$

$$\rho = \begin{cases} 1 & \text{vollständig korreliert} \\ 0 & \text{unkorreliert} \\ -1 & \text{vollständig antikorreliert} \end{cases} \tag{51}$$

4 Kombination von Zufallsvariablen

4.1 Charakteristische Funktionen zur Berechnung von kombinierten Wahrscheinlichkeitsdichtefunktionen

- Analogon zu Fourier:

$$\Phi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)dx \ mit \ \Phi(0) = 1 \tag{52}$$

- Rücktransformation:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \Phi(t)dt \tag{53}$$

- Beispiel: Charakteristische Funktion der Gaussfunktion:

$$\Phi(t) = e^{-\frac{t^2}{2\sigma_t^2}}$$

4.2 Faltung von Wahrscheinlichkeitsdichten

Seien x und y Zufallsvariablen $\Rightarrow z = x + y$ ist wieder Zufallsvariable.
$\Rightarrow f_z(z)$ ist Kombination von $f_y(y)$ und $f_x(x)$ folgendermaßen:

$$\begin{aligned} f_z(z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_x(x) f_y(y) \cdot \underbrace{\delta(z - (x + y))}_{Dirac-Delta} dxdy \\ &= \int_{-\infty}^{\infty} f_y(y) f_x(z - y)dy \end{aligned} \tag{54}$$

4.3 Zentraler Grenzwertsatz (ZGW)

"Die Kombination von vielen Zufallsvariablen aus der gleichen Verteilung ergeben immer eine Gaussverteilung"

`Bilder/2000px-CLTBinomConvergence.png`

Abbildung 4: Konvergenz der Binomial- gegen die Normalverteilung

4.4 Kenngrößen aus Messdaten

- Mittelwert mit Hilfe des ZGW:

$$
\begin{aligned}
\text{Sei } \hat{v} &= \frac{1}{n}\sum_{i=1}^{n} v_i \\
\overset{ZGW}{\Rightarrow} \omega &= \sum_{i=1}^{n} v_i \\
\Rightarrow \hat{v} &= \frac{\omega}{n} \ mit \ \langle \hat{v} \rangle = \frac{\langle \hat{\omega} \rangle}{n}
\end{aligned} \tag{55}
$$

- Fehler des Mittelwerts
  - Varianz:

$$V[\hat{v}] = V[\frac{\omega}{n}] = \frac{1}{n^2}V[\omega] = \frac{n}{n^2}V[v_i] = \frac{V[v_i]}{n} \tag{56}$$

  - Fehler des Mittelwerts ist damit:

$$\sigma_{\hat{v}} = \frac{\omega}{\sqrt{n}} \tag{57}$$

- Bestimmung der Standardabweichung

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(v_i - \hat{v})^2 \tag{58}$$

  Faktor $\frac{1}{n-1}$ kommt daher, dass $\hat{v}$ bereits eine Schätzung ist. Wird der wahre Mittelwert verwendet, dann wird der Faktor durch $\frac{1}{n}$ ersetzt.

- Fehler der Standardabweichung: Da σ aus Zufallszahlen berechnet wurde.

$$\Sigma_\sigma = \frac{\sigma}{\sqrt{2(n-1)}} \tag{59}$$

- Gewichteter Mittelwert:
  Messwerte mit kleinen Fehlern sollen stärker berücksichtigt werden.

$$\omega_i = \frac{1}{\sigma_i^2} \tag{60}$$

  - Mittelwert:

$$\langle x \rangle = \frac{\sum_{i=1}^{n} \omega_i x_i}{\sum_{i=1}^{n} \omega_i} \tag{61}$$

  - Varianz:

$$V[\langle x \rangle] = \frac{1}{\sum_{i=1}^{n} \omega_i} \tag{62}$$

  - Standardabweichung:

$$\sigma_{\langle x \rangle} = \sqrt{\frac{1}{\sum_{i=1}^{n} \omega_i}} \tag{63}$$

  - Gewichtetes Mittel hat kleineren Fehler bei unterschiedlichen Messgenauigkeiten.

5 Messfehler und Fehlerfortpflanzung

5.1 Transformationen von Wahrscheinlichkeitsdichten

- Transformation so dass Wahrscheinlichkeiten erhalten bleiben. $\Rightarrow$ Die Flächen unter den Dichtefunktionen sind gleich.

$$f_y(y) = f_x(x) \cdot |\frac{dx}{dy}|$$

5.2 Fehlerfortpflanzung

- Herleitung durch Taylor-Entwicklung der Transformationsformel:

$$y(x) = y(\langle x \rangle) + \frac{dy}{dx} |_{x=\langle x \rangle} (x - \langle x \rangle) + \frac{1}{2}\frac{d^2y}{dx^2} |_{x=\langle x \rangle} (x - \langle x \rangle)^2 + ...$$

- Erwartungswert der neuen Wahrscheinlichkeitsdichte:

$$E[x] = \underbrace{E[y(\langle x \rangle)]}_{y(\langle x \rangle)} + \frac{dy}{dx} |_{x=\langle x \rangle} \underbrace{E[(x - \langle x \rangle)]}_{=0} + \frac{1}{2}\frac{d^2y}{dx^2} |_{x=\langle x \rangle} \underbrace{E[(x - \langle x \rangle)^2]}_{V[x]=\sigma_y^2}$$

  umformen nach $\sqrt{V} = \sigma_y$ liefert:

- Standardabweichung der neuen Wahrscheinlichkeitsdichte:

$$\underbrace{\sigma_y = \frac{dy}{dx} |_{x=\langle x \rangle} \sigma_x}_{\textbf{Gesetz der Fehlerfortpflanzung}} \tag{64}$$

- Relative Fehler sind beispielsweise:

$$\frac{\sigma_x}{x} \tag{65}$$

5.3 Fehlerfortpflanzung mit vielen Variablen

- Transformation:

$$\vec{y} = \hat{B} \cdot \vec{x}, \; B_{ik} = \frac{\partial y_i}{\partial x_k} = const$$

- Erwartungswert

$$\langle \vec{y} \rangle = \vec{y}$$

- Kovarianzmatrix

$$\begin{aligned} \hat{V}[\vec{y}] &= E[(\vec{y} - \langle \vec{y} \rangle)^2] \\ &= E[\hat{B} \cdot (\vec{x} - \langle \vec{x} \rangle) \cdot (\vec{x} - \langle \vec{x} \rangle)^T \hat{B}^T] \\ \Rightarrow \quad & \underbrace{\hat{V}[\vec{y}] = \hat{B}\hat{V}[\vec{x}]\hat{B}^T}_{\textit{Gesetz der Fehlerfortpflanzung für viele Variable}} \end{aligned} \tag{66}$$

5.4 Fehlerfortpflanzung bei zusammengesetzten Messgrößen

$$Sei: y = y(x_1, x_2)$$

$$Dann\ gilt: \sigma_y^2 = \frac{\partial^2 y}{\partial x_1^2}\sigma_1^2 + 2 \cdot \frac{\partial y}{\partial x_1}\frac{\partial y}{x_2} + \frac{\partial^2 y}{\partial x_2^2}\sigma_2^2 \tag{67}$$

$$\textit{für den unkorrelierten Fall gilt: } \sigma_y = \sum_{i=1}^{n=2} \frac{\partial y}{\partial x_i}\sigma_i \tag{68}$$

Beispiele:

- 1. Addition und Subtraktion

$$y = x_1 \pm x_2$$
$$\Rightarrow \sigma_y = \sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}$$

- 2. Multiplikation und Division

$$y = x_1 \cdot x_2 \textbf{ oder } y = \frac{x_1}{x_2}$$
$$\Rightarrow (\frac{\sigma_y}{y})^2 = (\frac{\sigma_{x_1}}{x_1})^2 + (\frac{\sigma_{x_2}}{x_2})^2$$

6 Parameterschätzung aus Daten

6.1 Maximum Likelihood Methode

- Likelihood Funktion: sei a der Schätzwert und die x_i die Messwerte.

$$\mathcal{L} = f(x_1|a) \cdot f(x_2|a) \cdots f(x_n|a) = \prod_{i=1}^{n} f(x_i|a) \tag{69}$$

  $\mathcal{L}$ wird für den besten Schätzwert maximal.

- Bedingte Wahrscheinlichkeit:
  $\mathcal{L}$ gibt ein Maß für die bedingte Wahrscheinlichkeit $P(Daten|a)$ an bei gegebener Wahl von a diese Messwerte zu erhalten.

$$P(Daten|a) = \mathcal{L} \cdot const.$$

  gewünscht ist in praktischen zusammenhängen allerdings $P(a|Daten)$.

$$\begin{aligned} &Satz\ von\ Bayes\ \Rightarrow \\ &P(a|Daten) = P(Daten|a) \cdot \underbrace{\frac{P(a)}{P(Daten)}}_{im\ Allgemeinen\ konstant} \\ &\Rightarrow P(a|Daten) = \mathcal{L} \cdot const \end{aligned} \tag{70}$$

- Minimum der negativen logarithmischen Likelihood Funktion:

$$\ln \mathcal{L} = \ln \prod_{i=1}^{n} f(x_i|a) = \sum_{i=1}^{n} \ln f(x_i|a) notag \tag{71}$$

  - Konvention zur Vereinfachung der χ^2-Methode:

$$F(a) = -2 \ln \mathcal{L} \tag{72}$$

6.2 Standardabweichung der Logarithmischen Likelihood Funktion

grobe Herleitung:

Taylor Entwicklung von F(a) bis in 2te Ordnung dann

$$\begin{aligned} &\to \lim_{n\to\infty} \\ &\to Gauss \\ &\to F(a) = F(\langle a\rangle) + (\frac{a - \langle a\rangle}{\sigma})^2 \\ &\to \Delta F = F(a) - F(\langle a\rangle) = (\frac{a - \langle a\rangle}{\sigma})^2 \\ &\Rightarrow \Delta F(\langle a\rangle + n\sigma) = n^2\ (Parabel) \end{aligned}$$

Ändert sich F um n^2 ist die Entfernung zum nächsten Schätzwert $n \cdot \sigma$

6.3 Methode der kleinsten Quadrate (Least Square)

- Anpassung von n Messwerten mit Parametern a_j.
- Residuum: Differenz zwischen gemessenem Wert und Modellwert.

$$\rho_i(x_i) := y_i(x_i) - y_{Modell}(x_i) \tag{73}$$

- das Modell liefert die beste Anpassung, wenn :

$$h := \sum_{i=1}^{n} \rho_i^2$$

- Modellfunktion:

$$y_{Modell} = a_1 f_1(x) + (\cdots) + a_n f_n(x) = \sum_{j=1}^{m} a_j f_j$$

- Fehler der Parameter

$$\sigma^2 = \frac{1}{n-m} \sum_{i=1}^{n} (y_i - y_{Modell})^2 \tag{74}$$

  dabei ist m die Anzahl der Parameter a_j

7 Statistische Testverfahren

7.1 Test einer Hypothese

- Wenn das Messergebnis weniger als $2{\cdot}\sigma$ vom theoretischen/vorhergesagtem Wert entfernt liegt wird er nach Konvention als verträglich bezeichnet.
- Messung und Vorhersage müssen kompatibel sein.
- Konfidenzgrenze: Die Konfidenzgrenzen geben an, mit welcher Wahrscheinlichkeit der wahre Mittelwert einer Grundgesamtheit in einem gewissen Bereich liegt.
- Konfidenzintervall: Das Konfidenzintervall ist der Bereich, der bei unendlicher Wiederholung eines Zufallsexperiments mit einer gewissen Häufigkeit (dem Konfidenzniveau) die wahre Lage des Parameters einschließt. Ein häufig verwendetes Konfidenzniveau ist 95%.
- Konfidenzniveau: Das Konfidenzniveau gibt an, mit welcher Wahrscheinlichkeit die Lageschätzung eines statistischen Parameters (zum Beispiel eines Mittelwertes) aus einer Stichprobenerhebung auch für die Grundgesamtheit zutreffend ist.

$$CN = 1 - \alpha_{unterhalb} - \alpha_{oberhalb} \tag{75}$$

- t-Test:
  Student'sche t-Verteilung für Verfahren, bei denen die Genauigkeit des Mttelwerts aus den Daten selbst bestimmt wird.
  - Testgröße
  $$t = \frac{\langle y \rangle - y_t}{\sigma_{\langle y \rangle}} \tag{76}$$
  - Gamma Funktion
  $$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \quad \textit{für} \;\; \Gamma(n+1) = n! \tag{77}$$
  - t-Verteilung
  $$f_n(t) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} (1 + \frac{t^2}{2})^{-\frac{n+1}{2}} \tag{78}$$
  - Vorgehensweise:
    Zur Überprüfung ob n Messdaten mit m anderen Messdaten verträglich sind werden zunächst die jeweiligen Mittelwerte und die Fehler auf die Mittelwerte berechnet. Dann werden die Freiheitsgrade nach $f_n = n - 1 \; bzw. \; f_m = m - 1$ bestimmt. Die Testgröße t kann nun ausgerechnet werden. t sollte einer t-Verteilung mit $k = n + m - 2$ Freiheitsgraden folgen. Sind beide Experimente als Referenz wählbar muss nur $|t|$ betrachtet werden. Die Wahrscheinlichkeit P um einen Wert größer oder gleich $|t|$ zu kann nach
    $$P = 2 \cdot \int_{|t|}^\infty f_n(t) dt$$
    berechnet werden. Wenn nun das berechnete Konfidenzniveau $CN = 1-P$ innerhalb des gewählten Konfidenzniveau liegt, ($z.B. \; berechnet \; CN = 93,1\% \leq CN_{Gauss} = 95\%$) sind die Werte untereinander verträglich.

7.2 χ^2-Test

- Gauss Verteilte Messfehler
- Prüfung ob Messdaten mit theoretischer Vorhersage kompatibel sind.
- χ^2-Verteilung: benötigt Standardnormalverteilung. n unabhängige Zufallsvariablen z_i mit Mittelwert $\mu = 0$ und Standardabweichung $\sigma = 1$.

$$\chi^2 := \sum_{i=1}^{n} z_i^2 \tag{79}$$

$$\stackrel{Faltung}{\rightarrow} f_n(\chi^2) = \frac{(\frac{\chi^2}{2})^{\frac{n-2}{2}} \cdot exp^{-\frac{\chi^2}{2}}}{2\Gamma(\frac{n}{2})} \tag{80}$$

- Erwartungswert:

$$\langle \chi^2 \rangle = n \tag{81}$$

- Standardabweichung:

$$\sigma = \sqrt{2n} \tag{82}$$

- Maximum:

$$(\chi^2)_{max} = n - 2 \tag{83}$$

- Residuen für χ^2 als Testgröße bei n Messwerten.

$$\rho_i = \frac{y_i(x_i) - f(x_i)}{\sigma_{y,i}(x_i)} \tag{84}$$

falls die theoretische Beschreibung f(x) korrekt ist folgen die Residuen selbst auch der Gauss-verteilung mit Mittelwert $\langle \rho \rangle = 0$ *und* $\sigma_\rho = 1$

$$\Rightarrow \chi^2 = \sum_{i=1}^{n} \rho_i \tag{85}$$

Bilder/gm_compdistrib_tusche.png