

Zusammenfassung Datenanalyse

Monte-Carlo Methode

- Erzeugung von "Pseudozufallszahlen" mit Hilfe von Computeralgorithmen; Simulation von Prozessen statistischer Natur
- **Hit-And-Miss-Methode:** Erzeugt Zufallszahlen in Intervall $a \leq x \leq b$ gemäss Verteilung $p(x)$. Sei p_{max} maximalwert, den $p(x)$ über das Intervall annimmt: Es werden statistisch unabhängige Zufallszahlenpaare (x_i, y_i) erzeugt, wobei x_i gleichverteilt zwischen a und b , y_i gleichverteilt zwischen 0 und p_{max} . Akzeptiere x_i , wenn $y_i \leq p(x_i)$. Die akzeptierten x_i sind gemäss $p(x)$ verteilt. [Funktioniert für beliebige $p(x)$]
- **Invertieren der Verteilungsfunktion:** Das Ziel hierbei ist, Zufallszahlen x gemäss der Wahrscheinlichkeitsdichte $p(x)$ zu erzeugen. Zunächst werden Zufallszahlen u_i zwischen 0 und 1 erzeugt. Diese bilden sozusagen die Wahrscheinlichkeiten, die wir für die Verteilung erhalten. Ist $P(x)$ die (umkehrbare) kumulative Verteilungsfunktion für x^1 , so gilt: $x_i = P^{-1}(u_i)$ ist gemäss $p(x)$ verteilt. Blöd gesagt wird von der Wahrscheinlichkeit auf die kumulative Wahrscheinlichkeitsverteilung und daraus dann auf die Wahrscheinlichkeitsdichte geschlossen (vgl. Abb. 0.2)

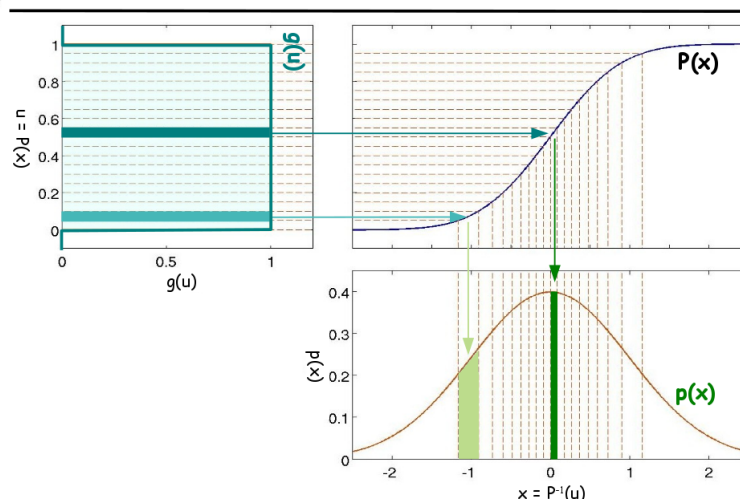


Abbildung 0.1: Invertieren der Verteilungsfunktion

Faltung zweier Verteilungen

- Faltung = "Überlagerung" zweier Wahrscheinlichkeitsverteilungen.
- x folgt Wahrscheinlichkeitsverteilung $f(x)$, y folgt Wahrscheinlichkeitsverteilung $g(y)$; Suche Wahrscheinlichkeitsverteilung $h(z)$ für $z = x + y$
- Notation: $h(z) = f(x) \otimes g(y)$ oder $h(z) = f(x) \cdot g(y)$
- Wahrscheinlichkeitsverteilungen $f(k_x)$, $g(k_y)$ diskreter Zufallsvariablen k_x , k_y :

$$h(k_z) = \sum_{k_x=0}^{k_z} f(k_x) \cdot g(k_z - k_x), \text{ wobei } k_z = k_x + k_y \Rightarrow k_y = k_z - k_x$$

$${}^1P(x) = \int_{-\infty}^x p(x') dx'$$

- Faltung einer Poissonverteilung mit Erwartungswerten μ_x, μ_z :

$$P(k_z) = \frac{(\mu_z)^{k_z}}{k_z!} \cdot e^{-\mu_z}$$

- Wahrscheinlichkeitsverteilungen $f(x), g(y)$ kontinuierlicher Zufallsvariablen x, y :

$$h(z) = \int_{-\infty}^{\infty} f(x) \cdot g(z-x) dx$$

- Monte-Carlo-Faltung: Addition statistisch unabhängig erzeugter Zufallszahlen (Einfach alle "Ergebnisse" im Histogramm zusammenaddieren)

- Faltung zweier Gaussverteilungen: $\mu_z = \mu_x + \mu_y, \sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}$

Verteilungen zweier Variablen

- Zweidimensionale Gaussverteilung:

Einfacher Fall: $\mu_x = \mu_y = 0, \sigma_x = \sigma_y = 0$: $p(x, y) = p(x) \cdot p(y) = \frac{1}{2\pi} \cdot e^{-\frac{x^2+y^2}{2}}$

Kurven konstanter Wahrscheinlichkeitsdichte: $x^2 + y^2 = \text{const} = k^2$

Wahrscheinlichkeit für Wertepaar mit $x^2 + y^2 < k^2$: $P(x^2 + y^2 < k^2) = 1 - e^{-\frac{k^2}{2}}$

Allgemeiner Fall: x und y korreliert², $\mu_x \neq \mu_y \neq 0, \sigma_x \neq \sigma_y \neq 0$:

$$p(x, y) = \frac{1}{2\pi\sqrt{\det(V)}} \cdot \exp\left[-\frac{1}{2} \left((x - \mu_x, y - \mu_y) V^{-1} \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix} \right)\right]$$

$$= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \cdot \exp\left[\frac{1}{2\pi(1-\rho_{xy}^2)} \left(\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho_{xy} \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right)\right]$$

- N-Dimensionale Gaussverteilung

Wahrscheinlichkeitsdichte³ $p(\vec{x}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det(V)}} \cdot e^{-\frac{1}{2} \cdot (\vec{x} - \vec{\mu}_x)^T \cdot V^{-1} \cdot (\vec{x} - \vec{\mu}_x)}$

Stichproben und Schätzfunktionen

- $f(x|a)$ hängt von Parameter(n) a ab; Experiment ergibt endliche Anzahl gemessener Werte x_i (= "Stichprobe"). Die Werte der Parameter a sollen nun aus dieser Stichprobe geschätzt werden.

- Schätzwerte sind Zufallsvariablen! Diverse Stichproben ergeben unterschiedliche Schätzwerte \hat{a} .

- Die Verteilung der Schätzwerte \hat{a} hat den Erwartungswert $\langle \hat{a} \rangle$ und Varianz $V(\hat{a})$.

- Die Schätzfunktion ist ein Algorithmus zur Berechnung von Schätzwerten.

- Ansprüche an Schätzfunktionen:

Konstistenz: Schätzwert \hat{a} soll mit zunehmendem Umfang der Stichprobe dem wahren Wert a entgegenstreben.

Erwartungstreue: Erwartungswert $\langle \hat{a} \rangle$ des Schätzwertes soll unabhängig vom Umfang der Stichprobe gleich dem wahren Wert des Parameters a sein \Rightarrow Schätzwert soll "im Mittel" über viele Stichproben a ergeben.

Effizienz: Varianz $V(\hat{a})$ des Schätzwertes soll für gegebenen Umfang der Stichprobe möglichst klein sein.

Robustheit: \hat{a} soll möglichst unempfindlich gegen einzelne Ausreisser sein.

Der Anspruch auf Effizienz und Robustheit haben einen Interessenkonflikt, denn eine kleine Varianz ist bei Ausreissern nicht korrekt bzw. schliesst diese aus; Unempfindlichkeit auf Ausreisser hingegen würde eine grosse Varianz bedingen.

²Mit: $V(x, y) = \begin{pmatrix} \sigma_x^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{xy}\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$; $\rho_{xy} = \frac{\text{cov}(x, y)}{\sigma_x\sigma_y}$

³Wobei $\vec{x} = (x_1, x_2, \dots, x_N)$; $\vec{\mu} = (\mu_1, \mu_2, \dots, \mu_N)$ und $V_{ij} = \langle (x_i - \mu_i) \cdot (x_j - \mu_j) \rangle$

- Aus diesen Ansprüchen resultieren zwei Definitionen für die Standardabweichung:

$$\sigma = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (x_i - \bar{x})^2} \text{ ist die Standardabweichung der Stichprobe selbst}$$

$$s = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^N (x_i - \bar{x})^2} \text{ ist die erwartungstreue Schätzfunktion für die Standardabweichung der Wahrscheinlichkeitsverteilung.}$$

Dieser Unterschied ist insbesondere dann wichtig, wenn N nicht gross ist. Deshalb muss immer angegeben werden, welche Definition benutzt wird.

Maximum-Likelihood-Methode

- Gegeben: Wahrscheinlichkeitsverteilung $f(x|a)$, Stichprobe aus N Werten x_i . Suche Schätzfunktion \hat{a} für Parameter a .

- Daraus bildet man die **Likelihood-Funktion** $L(a)$:

$$L(a) = f(x_1|a) \cdot f(x_2|a) \cdot \dots \cdot f(x_N|a) = \prod_{i=1}^N f(x_i|a)$$

Die Likelihood-Funktion gibt die Wahrscheinlichkeit dafür an, die gegebene Stichprobe zu erhalten und ist eine Funktion des Parameters a .

- Wähle als Schätzwerte \hat{a} diejenigen Werte, die $L(a)$ maximieren. (Die Wahrscheinlichkeit, diese Stichprobe zu erhalten, wird maximiert.)

- Rechentechnisch vorteilhaft: Suche Maximum der Log-Likelihood-Funktion

$$\ln L(a) = \ln \left(\prod_{i=1}^N f(x_i|a) \right) = \sum_{i=1}^N \ln (f(x_i|a))$$

Die Bestimmung des Maximums von \hat{a} erfolgt analytisch mit $\frac{d \ln L(\hat{a})}{d \hat{a}} = 0$ oder numerisch durch Abtasten.

- Unsicherheit auf Maximum-Likelihood-Schätzwert:

Grosser Stichprobenumfang ($N \rightarrow \infty$): Log-Likelihood Funktion ist parabelförmig, Unsicherheit $\sigma_{\hat{a}}$ wird aus der Breite der Parabel bestimmt.

$$\ln L(a) - \ln L(\hat{a}) = -\frac{(a-\hat{a})^2}{2\sigma_{\hat{a}}^2}; \text{ Mit}$$

$$\ln L(\hat{a} \pm \sigma_{\hat{a}}) = \ln L(\hat{a}) - 0.5$$

$$\ln L(\hat{a} \pm 2\sigma_{\hat{a}}) = \ln L(\hat{a}) - 2.0$$

$$\ln L(\hat{a} \pm 3\sigma_{\hat{a}}) = \ln L(\hat{a}) - 4.5$$

Stichprobenumfang nicht gross genug: Log-Likelihood-Funktion im Allgemeinen nicht parabelförmig. Die Unsicherheit auf \hat{a} wird definiert durch:

$$\ln L(\hat{a} \pm n \cdot \sigma_{\hat{a}}) := \ln L(\hat{a}) - n^2/2$$

Das ergibt unter Umständen asymmetrische Fehler, z.B. $\hat{a} = 3.5_{-1.1}^{+1.7}$

- Für n Dimensionen

$$\hat{a} = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n) \text{ wird berechnet aus } \frac{\partial \ln L(\hat{a})}{\partial \hat{a}_i} = 0$$

$$\text{Fehlermatrix für die } \hat{a}_i: V_{ij}^{-1} = \text{cov}^{-1}(a_i, a_j) = \left[-\frac{\partial^2 \ln L}{\partial a_i \partial a_j} \right]_{a_i=\hat{a}_i, a_j=\hat{a}_j}$$

- Anpassung einer Geraden $\hat{a} \cdot x = \hat{b}$:

$$\hat{a} = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}$$

$$\hat{b} = \frac{\overline{x^2 y} - \bar{x}\overline{xy}}{\overline{x^2} - \bar{x}^2}$$

$$\text{Kovarianzmatrix } \text{cov}(\hat{a}, \hat{b}) = \frac{\sigma^2}{N \cdot (\overline{x^2} - \bar{x}^2)} \cdot \begin{pmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

- Die Maximum-Likelihood-Methode ist eine häufig benutzte Methode, um Schätzfunktionen herzuleiten. Die Schätzfunktionen sind konsistent, meist effizient und für grossen Stichprobenumfang

erwartungstreu. Die Schätzwerte sind für grosse N gaussverteilt um den wahren Wert. Die Daten brauchen nicht histogrammiert zu werden, was einen Informationsverlust vermeidet. Zudem kann die Methode leicht auf mehrere Parameter erweitert werden. Jedoch sind die Schätzfunktionen bei kleinen N oft nicht erwartungstreu und die Resultate lassen keinen Rückschluss auf die Qualität der Anpassung zu.

- Binned Maximum-Likelihood: (bin-zentren x_i , bin-Breiten Δx_i , n_i Einträge pro bin:

Erwartete anzahl Einträge im bin i : $\nu_i = n \cdot \int_{x_i - \Delta x_i/2}^{x_i + \Delta x_i/2} f(x|a) dx \approx n \cdot \Delta x_i \cdot f(x_i|a)$

W'keit, n_i Einträge zu beobachten: $p_i(a) = \frac{(\nu_i(a))^{n_i}}{n_i!} \cdot e^{-\nu_i(a)}$

Log-Likelihood Funktion: $\ln L(a) = \sum_{i=1}^N (n_i \cdot \ln \nu_i(a) - \nu_i(a)) + const.$

- Erweiterte Maximum-Likelihood

Die Anzahl Ereignisse bei Experimenten ist meist nicht fest vorgegeben. Dies muss bei der Definition der Likelihood-Funktion berücksichtigt werden. Wir benutzen den Ansatz - die Gesamtzahl n betrachteter Ereignisse als poissonverteilt angenommen - den Erwartungswert ν dieser Poissonverteilung als zusätzlichen Parameter in der Likelihood-Funktion zu betrachten. Daraus folgt die erweiterte Likelihood- und Log-Likelihood-Funktion:

$$L(a, \nu(a)) = \frac{\nu^n(a)}{n!} e^{-\nu(a)} \cdot \prod_{i=1}^n f(x_i|a)$$

$$\ln L(a, \nu(a)) = -\nu(a) + \sum_{i=1}^n \ln [\nu(a) \cdot f(x_i|a)]$$

- Signal und Untergrund:

Seien die Form der Signalverteilung $s(x)$ und die Form der Untergrundverteilung $b(x)$ bekannt. Gesucht werden Schätzwerte für n_s und n_b . Nach der Standard-Max.-Lik. würde man die Gesamtzahl der Ereignisse $n_{tot} = n_s + n_b$ als fest annehmen. Daraus folgt:

$$\ln L(n_s) = \sum_{i=0}^{n_{tot}} \ln \left(\frac{n_s}{n_{tot}} \cdot s(x_i) + \frac{n_{tot} - n_s}{n_{tot}} \cdot b(x_i) \right)$$

Mit der erweiterten Max.-Lik. nehmen wir die Gesamtzahl der Ereignisse als poissonverteilt an:

$$\ln L(n_s, n_b) = -(n_s + n_b) + \sum_{i=0}^{n_{tot}} \ln (n_s \cdot s(x_i) + n_b \cdot b(x_i))$$

Methode der Kleinsten Quadrate

Auch "χ²" - Methode

- Die Methode kleinsten Quadrate ist eine häufig benutzte Methode zur Herleitung von Schätzfunktionen. Im Gegensatz zur Maximum-Likelihood Methode besteht hier die Stichprobe hier aus N Wertepaaren (x_i, y_i) , wobei wir Messunsicherheiten σ_i auf den y_i haben und diejenigen auf x_i vernachlässigbar klein sind. Wir wollen nun eine Funktion $f(x|a)$, welche uns für alle x -Werte den erwarteten y -Wert voraussagt. $f(x|a)$ hängt von einem oder mehreren Parametern a ab, deren Werte aus der Stichprobe geschätzt werden sollen.

- Wir definieren die "Chi-Quadrat"-Funktion:

$$\chi^2(a) = \sum_{i=1}^N \left(\frac{y_i - f(x_i|a)}{\sigma_i} \right)^2$$

Die $\chi^2(a)$ -Funktion misst die Diskrepanz zwischen erwarteten und beobachteten y -Werten und ist eine Funktion der Parameter a .

- Im Gegensatz zur Maximum-Likelihood-Methode wählt man hier als Schätzwerte \hat{a} nicht diejenigen Werte der Parameter a , die χ^2 maximieren, sondern diejenigen, die sie minimieren. (Merkhilfe: Methode der *kleinsten* Quadrate vs *Maximum*-Likelihood-Methode)

- Beispiele:

$$f(x) = a \cdot x, \sigma_i = \sigma \text{ für alle gleich Gross} \quad \Rightarrow \quad \chi^2(a) = \sum_{i=1}^N \left(\frac{y_i - a \cdot x_i}{\sigma} \right)^2 \quad \Rightarrow \quad \hat{a} = \frac{\overline{xy}}{\overline{x^2}}$$

$$V(\hat{a}) = \sum_{i=1}^N \left(\frac{\partial \hat{a}}{\partial y_i} \right)^2 \sigma^2 = \frac{\sigma^2}{N \cdot \overline{x^2}}^4$$

$$f(x) = a \cdot x + b:$$

Das Ergebnis ist identisch zur Maximum-Likelihood Methode: $\hat{a} = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}$, $\hat{b} = \frac{\overline{x^2\bar{y}} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}$

$$\text{cov}(\hat{a}, \hat{b}) = \frac{\sigma^2}{N \cdot (\overline{x^2} - \bar{x}^2)} \cdot \begin{pmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

- Die Methode der kleinsten Quadrate und die Maximum-Likelihood-Methode sind identisch, wenn die Messfehler auf den y_i gaussverteilt sind.

- Anpassung der Geraden $y = a \cdot x + b$, allgemeiner Fall (Messunsicherheiten σ_i unterschiedlich gross)
Ersetze in allen Formeln einfache Mittelwerte durch gewichtete Mittelwerte:

$$\bar{x} = \frac{\sum_{i=1}^N x_i / \sigma_i^2}{1 / \sigma_i^2}; \bar{y} = \frac{\sum_{i=1}^N y_i / \sigma_i^2}{1 / \sigma_i^2}; \overline{x^2} = \frac{\sum_{i=1}^N x_i^2 / \sigma_i^2}{1 / \sigma_i^2}; \overline{xy} = \frac{\sum_{i=1}^N x_i \cdot y_i / \sigma_i^2}{1 / \sigma_i^2};$$

Ersetze in der Kovarianzmatrix σ^2 durch gewichteten Mittelwert der σ_i^2 :

$$\overline{\sigma^2} = \frac{N}{\sum_{i=1}^N N / \sigma_i^2}$$

- Extrapolation der Geraden zu einem Wert X : $Y(X) = \hat{b} + \hat{a} \cdot X$
Varianz auf Y durch Fehlerfortpflanzung: $V(Y) = V(\hat{b}) + X^2 \cdot V(\hat{a}) + 2X \cdot \text{cov}(\hat{a}, \hat{b})$
Nach einer Koordinatentransformation $x \rightarrow x' = x - \bar{x}$ verschwindet der Kovarianzterm!

- χ^2 ist ein quantitatives Mass für die Güte der Anpassung an die Daten. Erwartet werden im Mittel $\chi^2 = 1$ pro Freiheitsgrad n_f , wobei n_f die Anzahl Messpaare minus die Anzahl zu bestimmender Parameter ist.

$\chi^2 / n_f \gg 1$: Abweichungen der gemessenen von den vorhergesagten Werten viel grösser als aufgrund der angegebenen Messunsicherheiten erwartet. Wurden die Messfehler unterschätzt oder ist die angenommene Form der Verteilung $f(x|a)$ falsch?

$\chi^2 / n_f \ll 1$: Übereinstimmung der gemessenen und vorhergesagten Werte viel besser als aufgrund der angegebenen Messunsicherheiten erwartet. Wurden die Messfehler überschätzt? Aber: grosse und kleine χ^2 kommen auch aufgrund zufälliger Fluktuationen vor.

- χ^2 -Verteilung für gaussverteilte Messfehler auf den y_i :

$$p(\chi^2 | n_f) = \frac{2^{-n_f/2}}{\Gamma(n_f/2)} \cdot \chi^{n_f-2} \cdot e^{-\chi^2/2}$$

Wobei die Γ -Funktion die Erweiterung der Fakultät $n!$ auf reelle Zahlen ist: $\Gamma(x + 1) := x \cdot \Gamma(x)$ mit $\Gamma(1) := 1$

$$\langle \chi^2 \rangle = n_f; V(\chi^2) = 2 \cdot n_f$$

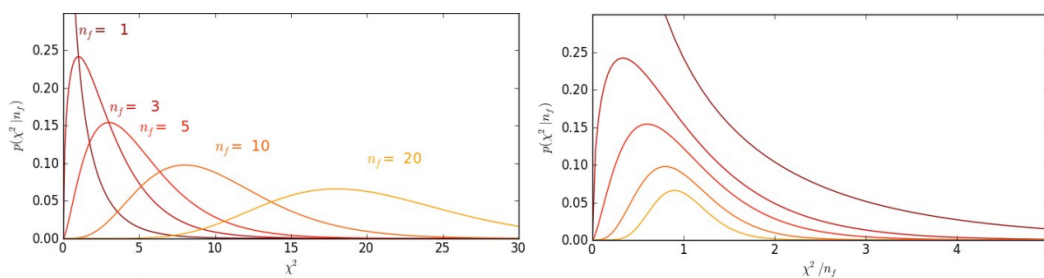


Abbildung 0.2: χ^2 -Verteilungen

⁴Dies folgt aus der Fehlerfortpflanzung...

Formelsammlung

Faltung einer Funktion

Faltung einer Funktion	$h(z)$	$=$	$f(x) \otimes g(y)$
oder	$h(z)$	$=$	$f(x) \cdot g(y)$
W'keitsvert. $f(k_x), g(k_y), k_x, k_y$ diskret:	$h(k_z)$	$=$	$\sum_{k_x=0}^{k_z} f(k_x) \cdot g(kz - k_x)$
Poissonverteilung	$P(k_z)$	$=$	$\frac{(\mu_z)^{k_z}}{k_z!} \cdot e^{-\mu_z}$
kontinuierliche Zufallsvariablen x, y :	$h(z)$	$=$	$\int_{-\infty}^{\infty} f(x) \cdot g(z - x) dx$
zwei Gaussverteilungen:	μ_z	$=$	$\mu_x + \mu_y, \sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}$

Verteilungen zweier Variablen

Zweidimensionale Gaussverteilung:

Einfacher Fall:

$\mu_x = \mu_y = 0, \sigma_x = \sigma_y = 0$:	$p(x, y)$	$=$	$p(x) \cdot p(y) = \frac{1}{2\pi} \cdot e^{-\frac{x^2+y^2}{2}}$
W'keit für Wertepaar mit $x^2 + y^2 < k^2$:	$P(x^2 + y^2 < k^2)$	$=$	$1 - e^{-\frac{k^2}{2}}$

Allgemeiner Fall:

x und y korreliert	$p(x, y)$	$=$	$\frac{1}{2\pi\sqrt{\det(V)}} \cdot \exp\left[-\frac{1}{2}\left((x - \mu_x, y - \mu_y)V^{-1}\begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix}\right)\right]$
$\mu_x \neq \mu_y \neq 0, \sigma_x \neq \sigma_y \neq 0$		$=$	$\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \cdot e^{A(x,y)}$
Mit	$A(x, y)$	$=$	$\left[\frac{1}{2\pi(1-\rho_{xy}^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho_{xy}\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right)\right]$

Wobei	$V(x, y)$	$=$	$\begin{pmatrix} \sigma_x^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{xy}\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$
	ρ_{xy}	$=$	$\frac{\text{cov}(x,y)}{\sigma_x\sigma_y}$

N-Dimensionale Gaussverteilung:

W'keitsdichte	$p(\vec{x})$	$=$	$\frac{1}{(2\pi)^{N/2}\sqrt{\det(V)}} \cdot e^{-\frac{1}{2}\cdot(\vec{x}-\vec{\mu}_x)^T \cdot V^{-1} \cdot (\vec{x}-\vec{\mu}_x)}$
---------------	--------------	-----	--

Wobei	\vec{x}	$=$	(x_1, x_2, \dots, x_N)
	$\vec{\mu}$	$=$	$(\mu_1, \mu_2, \dots, \mu_N)$
	V_{ij}	$=$	$\langle (x_i - \mu_i) \cdot (x_j - \mu_j) \rangle$

Schätzfunktionen und Maximum-Likelihood

Standardabweichung der Stichprobe	σ	$=$	$\sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (x_i - \bar{x})^2}$
-----------------------------------	----------	-----	---

Schätzfunktion für Standardabweichung	s	$=$	$\sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^N (x_i - \bar{x})^2}$
---------------------------------------	-----	-----	---

Likelihood-Funktion	$L(a)$	$=$	$f(x_1 a) \cdot f(x_2 a) \cdot \dots \cdot f(x_N a) = \prod_{i=1}^N f(x_i a)$
---------------------	--------	-----	---

Log-Likelihood-Funktion	$\ln L(a)$	$=$	$\sum_{i=1}^N \ln(f(x_i a))$
-------------------------	------------	-----	------------------------------

Unsicherheit auf Log-Likelihood:	$\ln L(a) - \ln L(\hat{a})$	$=$	$-\frac{(a-\hat{a})^2}{2\sigma_{\hat{a}}^2}$
----------------------------------	-----------------------------	-----	--

(Für $N \rightarrow \infty$) Mit	$\ln L(\hat{a} \pm \sigma_{\hat{a}})$	$=$	$\ln L(\hat{a}) - 0.5$
-----------------------------------	---------------------------------------	-----	------------------------

	$\ln L(\hat{a} \pm 2\sigma_{\hat{a}})$	$=$	$\ln L(\hat{a}) - 2.0$
--	--	-----	------------------------

	$\ln L(\hat{a} \pm 3\sigma_{\hat{a}})$	$=$	$\ln L(\hat{a}) - 4.5$
--	--	-----	------------------------

Ansonsten:	$\ln L(\hat{a} \pm n \cdot \sigma_{\hat{a}})$	$:=$	$\ln L(\hat{a}) - n^2/2$
------------	---	------	--------------------------

Fehlermatrix bei n Dimensionen
Anpassung einer Geraden $y = a \cdot x + b$

$$V_{ij}^{-1} = cov^{-1}(a_i, a_j) = \left[-\frac{\partial^2 \ln L}{\partial a_i \partial a_j} \right]_{a_i = \hat{a}_i, a_j = \hat{a}_j}$$

$$\hat{a} = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}$$

$$\hat{b} = \frac{\overline{x^2 y} - \bar{x}\overline{xy}}{\overline{x^2} - \bar{x}^2}$$

$$cov(\hat{a}, \hat{b}) = \frac{\sigma^2}{N \cdot (\overline{x^2} - \bar{x}^2)} \cdot \begin{pmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

Kovarianzmatrix

Binned Maximum-Likelihood: (bin-Zentren x_i ,
bin-Breiten Δx_i , n_i Einträge pro bin)

Erwartete Anzahl Einträge im bin i :

$$\nu_i = n \cdot \int_{x_i - \Delta x/2}^{x_i + \Delta x/2} f(x|a) dx \approx n \cdot \Delta x_i \cdot f(x_i|a)$$

W'keit, n_i Einträge zu beobachten:

$$p_i(a) = \frac{(\nu_i(a))^{n_i}}{n_i!} \cdot e^{-\nu_i(a)}$$

Log-Likelihood Funktion:

$$\ln L(a) = \sum_{i=1}^N (n_i \cdot \ln \nu_i(a) - \nu_i(a)) + const.$$

Erweiterte Maximum-Likelihood-Funktion:

$$L(a, \nu(a)) = \frac{\nu^n(a)}{n!} e^{-\nu(a)} \cdot \prod_{i=1}^n f(x_i|a)$$

$$\ln L(a, \nu(a)) = -\nu(a) + \sum_{i=1}^n \ln [\nu(a) \cdot f(x_i|a)]$$

Signal und Untergrund

Standard Max. Lik.

$$\ln L(n_s) = \sum_{i=0}^{n_{tot}} \ln \left(\frac{n_s}{n_{tot}} \cdot s(x_i) + \frac{n_{tot} - n_s}{n_{tot}} \cdot b(x_i) \right)$$

Erweiterte Max.-Lik.

$$\ln L(n_s, n_b) = -(n_s + n_b) + \sum_{i=0}^{n_{tot}} \ln (n_s \cdot s(x_i) + n_b \cdot b(x_i))$$

Methode kleinster Quadrate

$$\chi^2(a): \quad \chi^2(a) = \sum_{i=1}^N \left(\frac{y_i - f(x_i|a)}{\sigma_i} \right)^2$$

$$f(x) = a \cdot x, \quad \sigma_i = \sigma \text{ für alle gleich Gross} \quad \hat{a} = \frac{\overline{xy}}{\overline{x^2}}$$

$$V(\hat{a}) = \sum_{i=1}^N \left(\frac{\partial \hat{a}}{\partial y_i} \right)^2 \sigma^2 = \frac{\sigma^2}{N \cdot \overline{x^2}}$$