****

Höhere Mathematik für Physiker
Teil III

F. Krause

Kapitel 17
Statistik

****

Inhalt des Kapitels Statistik

Kap. 17 **Statistik**

17.1 **Beschreibende Statistik (I)**

17.1.1 **Das allgemeine Szenenbild**

(1.1.1) Worum geht es in der Statistik? Zunächst einmal immer um **Datensätze**, das sind Abbildungen I$\rightarrow M$ einer endlichen Indexmenge I in eine Menge M von Beobachtungs- oder Messresultaten.
Die genaueren Bezeichnungen:

- Ein "Datensatz (mit Werten in M)" ist eine Abbildung der folgenden Art:

$$\boxed{\text{d=(I,n}\mapsto\text{a}_n\text{=a(n),M)}}$$

  - I **endliche** Menge,
  - M meist $\mathbb{R}$ oder $\mathbb{R}^k$ oder noch andere Mengen.

- (1.1.2) Unterscheide: "Datensätze mit kanonischer Parametrisierung" = "Zeitreihen" und solchen ohne solche. Wir betrachten **hier** Datensätze **ohne** kanonische Parameztrisierung. Dann verliert man insbesondere durch Indexumbenennung keine Sachinformation.

- (1.1.3) **Auszählung eines Datensatzes** (ohne kanonische.Parametrisierung . Geht auch für Datensätze **mit** kan. Param., da man die Interpretation von I vergessen kann):

  Definition: Sei $Bild(d) \subset J$ und Z=$\{Z_1, ..., Z_n\}$ **endliche** Zerlegung von J. $\alpha \in K$ indiziere die Teile der Partition.

- – Bilde die Mengen

$$d_\alpha = \{n \in J | a_n \in Z_\alpha\} \subset J.$$

  – Und deren **Anzahlen**

$$N_\alpha = \natural(d_\alpha) \qquad \alpha \text{ aus K, K endlich.}$$

  Diese liefern einen neuen Datensatz

$$H(d, Z) = (J, \alpha \mapsto N_\alpha, \mathbb{N})$$

  – Schließlich sei N=$\natural(I)$ die Zahl der Daten insgesamt.

  $i \in I$ indiziert die Daten, $\alpha \in K$ die "bins" oder "Wertekörbe" für die Daten.

(1.1.4) Man zählt also aus, wieviele der Daten aus d in den Teil Z_α von J (=Wertekorb) fallen. Die Größe der Z_α ist so zu wählen, daß die Auszählung möglichst aussagekräftig wird: Sind die Z_α zu groß, erkennt man zu wenig Struktur. Sind sie zu klein, wird ihre Anzahl zu groß und es fallen zu wenig Daten in die einzelnen Klassen.
K kann eine reine Zählmenge sein, K=$\{1, 2, ..., N\}$ oder die zugehörigen Intervalle bezeichnen oder auch einen typischen Vertreter der jeweiligen Klasse.

Beispiel: M={1,2,3,4,...,100} mit Datensatz d=(M,n$\mapsto$ $\pi(n), \{0, 1\}$), wobei $\pi(n) = 1$, wenn n Primzahl und =0 sonst. Z={{0},{1}}. Dann ist etwa $d_1 = \{n | n \in M$, n ist Primzahl$\}$ und N(d_1) = 25.

Bilder / Beispiele selbst fertigen!

(1.1.5) Jede Auszählung eines Datensatzes liefert einem zwei neue Abbildungen. Zunächst die eigentliche "Auszählungsabbildung (von d bezüglich Z)" oder "**absolute Häufigkeit"**

$$\boxed{\mathrm{H(d,Z)=(K,}\alpha \mapsto \mathrm{N}_\alpha=\natural(\mathrm{d}_\alpha),\mathbb{R})}$$

und dann die "**relative Häufigkeit"** (es war N=$\natural(Bild(d))$

$$\boxed{\mathrm{h(d,Z)=(K,}\alpha \mapsto \mathrm{h}_\alpha=\tfrac{N_\alpha}{N}=\tfrac{\natural(d_\alpha)}{N},\mathbb{R})}$$

(1.1.6) Ist die Wertemenge von d gleich $\mathbb{R}$ und z$\in \mathbb{R}$,dann gilt für jeden Datenwert d(i) entweder d(i)$\leq z$ oder d(i)>z. In der Regel führt das zu einer zweielementigen Partiton von I mit zugehöriger Auszählung. die angibt, wieviele Datenwerte oberhalb bzw. unterhalb von z liegen. Im Beispiel mit den Primzahlen und z=$\frac{1}{2}$ liegen die 25 Primzahlen mit d(i)=1 oberhalb und der Rest mit d(i)=0 unterhalb von $\frac{1}{2}$.

(1.1.7) Hierzu gibt es eine wichtige Verallgemeinerung. Aber beachten Sie: Die zugehörige Konstruktion ist zunächst nur für sinnvoll, wenn die Urbildmenge I eine Teilmenge von $\mathbb{R}$ ist. Eine Verallgemeinerung auf $\mathbb{R}^k$ ist noch möglich (Beachte: Wertemenge $\mathbb{R}$ bedeutet: Bild(d)$\subset J \subset \mathbb{R}$!)

Definition: Die Verteilungsfunktion des Datensatzes d:$I \to \mathbb{R}$ ist die Abbildung

$$\boxed{\mathrm{P(d)=(J,x}\mapsto \mathrm{P}_d(\mathrm{x})=\tfrac{1}{N}\sum_{d_n \leq x} 1,\mathbb{R})}$$

Der Wert $\mathrm{P}_d(x)$ gibt daher prozentual an, wieviele Elemente einen Wert kleiner oder gleich x haben. (Großer Buchstabe P, da Stammfunktion zu p). In Worten: N$\cdot \mathrm{P}_n(x)$ ist die Anzahl der Daten aus d mit $\mathrm{d}_n \leq x$. Tritt dabei ein und derselbe Wert mehrfach auf, so ist er auch mehrfach zu zählen. Die Verteilungsfunktion ist kanonisch, kommt ohne die Willkür von Zerlegungen aus. Da d nur endlich viele Elemente enthält ist P stets eine Stufenabbildung.

(1.1.8) P(d)=P_d ist eine monoton von 0 auf 1 wachsende Funktion $J \to [0,1]$. Sie ist Verteilungsfunktion P erweist sich als wichtige Ausgangsgröße zur Bestimmung weiterer wahrscheinlichkeitstheoretischer Größen. Insbesondere erwartet man gutes Konvergenzverhalten, etwa mit Wachsen der Datenzahl.

17.1.2 Hauptbeschreibungsgrößen (der deskriptiven Statitik)

Das sind Größen ähnlich dem Funktionaltyp: ***Durch wenige Zahlangaben sollen wesentliche Eigenschaften des Datensatzes möglichst gut charakterisiert*** *werden***.**

Wir beschränken uns auf den Fall von Datensätzen deren Wertemenge $\mathbb{R}$ oder eine Teilmenge davon ist.

(1.2.1) **Charakterisierung des Datensatzes durch eine einzige Zahlangabe**: Das sind die *Mittelwerte.* Gesucht also eine einzige Zahlangabe, die die Lage des Datenschwarmes auf der reellen Zahlengeraden möglichst gut charakterisiert. Hierzu nehmen wir meist das *arithmetische Mittel,* also *die* Zahl $\boxed{\bar{\mathrm{d}}=\frac{1}{N}\Sigma_i d_i}$. Aber andere Charakterisierungen können durchaus sinnvoll sein.

(1.2.2) Ist die Zerlegung z von I so, dass in jeder ihrer Klassen der Datenwert näherungsweise konstant ist und ist x_α ein typischer Vertreter der Klasse Z_α , dann gibt es jeweils eine zugehörige Näherungsformel für die Charakterisierungsgröße $\boxed{\bar{\mathrm{d}}=\frac{1}{N}\Sigma_i d_i}$ über die Auszählungsabbildung, die wir mit angeben. Beachten Sie: $\sum_\alpha$ ist in der Regel kürzer und zugänglicher als $\sum_n$. Überdies leisten diese Formeln später den äußerst wichtigen Übergang zum theoretischen Bereich.

(1.2.3) Definition (arithmetischer) Mittelwert

$$\bar{d} = \frac{1}{N}\Sigma_{n\in I} a_n \approx \Sigma_{\alpha \in K} h_\alpha x_\alpha.$$

Der Mittelwert macht Aussagen über die Lage des Datenschwarmes auf der Zahlengeraden. Er macht noch keinerlei Aussage über dessen **Ausdehnung**. Die nächste Größe tut dies. Erneut geben wir gleich die Näherungsformel über eine Auszählung mit an:

(1.2.4) **Definition:** Varianz und (empirische) Streuung des Datensatzes:

$$Var(d) = \frac{1}{N}\sum_n (d_n - \bar{d})^2 \approx \sum_\alpha h_\alpha (x_\alpha - \bar{d})^2$$

Und

$$\sigma_d = \sqrt{Var(d)} \quad \text{oder} \quad Var(d) = \sigma_d^2$$

(1.2.5) **Zusammenfassung:** Die einfachste Charakterisierung eines numerischen Datensatzes erfolgt durch die Angabe der beidenZalen $\bar{d}$ und σ_d . Der Mittelwert gibt die ungefähre Lage der Daten auf der Achse und die Streung ihre Ausbreitung um diesen Punkt herum. Grob kann man vielfach sagen, daß sich etwa 50% der Daten im Intervall $[\bar{d} - \sigma, d + \sigma]$ befinden werden.

(1.2.6) Man kann die Daten daher wie folgt parametrisieren

$$\boxed{\mathrm{d}_n = \bar{d} + \sigma\delta_n \quad \text{mit} \quad \delta_n = \frac{d_n - \bar{d}}{\sigma}}$$

Die "neuen Koordinaten" δ_n sind typischerweise Zahlen der Größenordnung 1, die um den Mittelwert 0 herum liegen, wobei etwa 50% der Daten zwischen -1 und +1 liegen..

17.2 Mathematische Beschreibung der Häufigkeiten im theoretischen Bereich

17.2.1 Das allgemeine Szenenbild

(2.1.1) Hat man viele Datensätze ein und derselben Art oder einen sehr großen, dann erwarten wir vielfach, dass die Resultate, in einem geeigneten Sinne gegen eine Grenzstruktur konvergieren. Sind nur endlich viele Messwerte möglich (Würfel), dann können wir das einfach durch Ereigniswahrscheinlichkeiten beschreiben. Vgl. Kapitel 1. Aber meist sind unendlich viele unterschiedliche Messwerte zulässig. Nimmt man dann jeweils eine Auszählung vor, so erwarten wir Konvergenz der relativen Häufigkeiten h_α Aber wie hängen die dann von der Wahl der Zerlegung Z ab? Als besonders günstig erweist es sich, die Ankopplung mit Hilfe der (zerlegungsunabhängigen) Verteilungsfunktion.

(2.1.2) Gegeben ein Wahrscheinlichkeitsraum $(\mathbb{R}, \mathfrak{B}, w)$ mit Wahrscheinlichkeitsmaß, also w(E)=1 und $\mathfrak{B}$ die Borel-σ-Algebra .

Dann existieren die folgenden beiden Beschreibungsgrößen für die Mengenfunktion w:

$\mathrm{P}=\mathrm{P}_w$	Verteilungsfunktion	$\mathrm{P}:\mathbb{R} \mapsto \mathbb{R}$	$P(x) = w(]-\infty, x])$
w	Wahrscheinlichkeits intervallfunktion	$[a, b[\mapsto$	$w([a, b[)$

(2.1.3) Ist P ausreichend glatt, so erhält man durch Differenzieren die Dichtefunktion p(x)=P$'(x)$. Also:

$$P(x) = \int_{-\infty}^{x} dtp(t) \quad \text{und } w([a, b[) = \int_a^b dxp(x)$$

Die drei Größen w,p und P nicht verwechseln.

(2.1.4) Beispiel 1): **Gleichverteilung** (b>a):

$$p(x) = \begin{cases} \frac{1}{b-a} & \text{für } \mathrm{a} \le x \le b \\ 0 & \text{sonst} \end{cases}$$

(2.1.5) Beispiel 2) **Normalverteilung**

$$p_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Beachten Sie, daß die Normierungsbedingung $\mathrm{w}(\mathbb{R}) = 1$ jeweils erfüllt ist.

□ Bestimmen Sie in beiden Fällen die zugehörige Verteilungsfunktion.

17.2.1a **Erwartungswerte und Momente**

(2.1.6) Sei jetzt f:$\mathbb{R} \to \mathbb{C}$ oder $f : \mathbb{R} \to \mathbb{V}$ eine L^1-Funktion. Dann bildet man den *Erwartungswert (von f für w)* als Element von $\mathbb{C}$ bzw.V . Diese Größe ist wie folgt definiert (mit einer Vielzahl von Bezeichnungen und eventuellen Konvergenzproblemen):

$$\langle f\rangle = \langle f\rangle_w = \langle f\rangle_p = \int_E dwf = \int_E dxp(x)f(x) = \langle f(x)\rangle = Ef = ...$$

Bei der Schreibweise $\langle f(x)\rangle$ ist x stumme Variable, nicht etwa äußerer Parameter! Erwartungswerte müssen keineswegs immer existieren! Stellen Sie sich daher immer die Einstiegsfrage: Existiert das Integral überhaupt??) . Immer existiert $\langle 1\rangle = 1$.

(2.1.7) Besonders wichtige Erwartungswerte sind die folgenden

Definition: Unter den **n-ten Momente (der Verteilung)** versteht man die folgenden Mittelwerte

$$\boxed{\langle h_n\rangle = \langle x^n\rangle = \int_{\mathbb{R}} \mathrm{dwh}_n = \int_{\mathbb{R}} \mathrm{dxp(x)x}^n = \mu_n}.$$

μ_1 ist der Mittelwert der Verteilung.

☐ Was ist μ_0?

☐ Bestimmen Sie den Mittelwert für die Gleichverteilung und für die Normalverteilung. Dann die höheren Momente dieser beidenVerteilungen. /Machen sie sich mit den Einheiten dieser Größen vertraut./ Zeichnen Sie eine Gleichverteilung samt zugehörigem Mittelwert und Streuung.

(2.1.8) Die **zentralen Momente** (einer Verteilung) sind die folgenden Erwartungswerte:

$$\boxed{\langle (x-\mu_1)^n\rangle = \int_{\mathbb{R}} \mathrm{dxp(x)(x\text{-}}\mu_1)^n. \quad \mathrm{n=0,1,2,..}}$$

$$Var(p) \quad = \quad \sigma^2 = \left\langle (x-\mu_1)^2 \right\rangle = \int_{\mathbb{R}} dxp(x)(x-\mu_1)^2$$

Für n=1 ist das zentrale Moment Null. Für n=2 erhält man die wichtige Varianz und $\sigma = \sqrt{Var}$ ist die Streuung. Beachten sie, daß gilt:

$$\boxed{\sigma^2 = \mu_2 - \mu_1^2.}$$

☐ Verifizieren und rechtfertigen Sie die letzte Gleichung. Was ist vorauszusetzen?

Ist E diskret, gehen alle Integrale in der üblichen Weise in Summen über.

(2.1.9) Verständnisbildend ist auch die näherungsweise Ersetzung der Integrale durch Summen. Sei dazu Z=$\{Z_\alpha\}$ eine Zerlegung von $\mathbb{R}$ wie im deskriptiven Fall beschreiben und f_α ein typischer Funktionswert aus Z_α. Dann folgt:

$$\langle f\rangle = \int_E dxp(x)f(x) \approx \sum_\alpha w(Z_\alpha)f_\alpha.$$

☐ Vergleichen Sie jetzt die Formeln des deskriptiven Bereichs für Mittelwert und Varianz mit den theoretischen Formeln. Was ist, wenn E endlich ist, was ergibt der Vergleich der Näherungsformeln?

☐ Gelten die Definitionen für die Erwartungswerte auch für allgemeinen Grundraum E? Wie sehen die Formeln dann aus?

(2.1.10) Sei E=$\mathbb{R}$. Dann bilden μ_1 und σ zusammen ein **systembezogenes Koordinatensystem** mit μ_1 als Ursprung und σ als Einheit. Der Übergang zu den zentralen Momenten entspricht bereits dem Übergang zu dem neuen Ursprung.

(2.1.11) Schließlich benutzt man (neben μ_1 und σ) noch folgende Größen zur Beschreibung der Verteilung

Die Schiefe (der V.)	$S\sigma^3 = \left\langle (x-\mu_1)^3 \right\rangle_p$	$=\kappa_3$
Den Exzess (der V.)	$e\sigma^4 = \left\langle (x-\mu_1)^4 \right\rangle_p - 3\sigma^4$	$=\kappa_4$

Der Buchstabe κ steht für "Kumulante". Auf diese Größen werden wir zurückkommen.

☐ Wie sehen die S und e entsprechenden empirischen Beschreibungsgrößen für Datensätze aus?

(2.1.12) Die Momente müssen keineswegs existieren (L^1-Problem). Als Gegenbeispiel kann man immer die (durchaus wichtige) *Cauchyverteilung* nehmen. Für sie existiert nicht einmal das Mittelwertintegral!! Sie wird über ihre Dichte gegeben:

$$p_{cauchy}(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

Man simuliert sie auf dem Rechner durch eine Datenfolge

$$n \mapsto a_n = \tan(\frac{\pi}{2}z_n) \qquad z_n \text{ Zufallsfolge aus [0,1]}$$

☐ Im Falle riner Gleichverteilung existieren alle Momente. Wieso? Bestimmen Sie die Momente.

☐ **Wichtig:** Sei A$\subset \mathbb{R}$ meßbar und χ_A deren charakteristische Funktion (Integrationstheorie) . Was ist (=welche inhaltliche Bedeutung hat)

$$\int_E dw\chi_A = \int_A dw = \int_A dxp(x) \qquad ??$$

17.2.1b **Die charakteristische Funktion (der Verteilung)**

(2.1.13) Immer exitiert jedoch der folgende Erwartungswert, den man *die charakteristische Funktion (der Verteilung)* nennt.

$$\boxed{\mathrm{k}\mapsto \mathrm{G(k)}=\mathrm{G}_p\mathrm{(k)}=\langle e^{ikx}\rangle = \int_{-\infty}^{\infty}\mathrm{dxp(x)e}^{ikx}}$$

Das ist einerseits ein Erwartungswert, andererseits ein Fouriertransformierte (der Dichte p). Offensichtlich ist G(0)=1.

(2.1.14) Sofern die Taylorentwicklung von G um k=0 existiert und ebenso die Momente, kann man e^{ikx} entwickeln und erwartet die folgende Gleichung

$$G(k) = \sum_{n=0}^{\infty}\frac{\mu_n}{n!}(ik)^n.$$

Die charakteristische Funktion enthält also alle Momente!

(2.1.15) Oder: Im Prinzip legt die Folge der Momente die gesamteVerteilung fest. **Aber man darf die Momente umgekehrt keineswegs beliebig vorgeben!** Denn die inverse Fouriertransformierte der charakteristichen Funktion muß ja die Dichte einer Wahrscheinlichkeitsverteilung ergeben und das kann auf mancherlei Weise mißlingen. Insbesondere muß immer p(x)$\geq$ 0 gelten. Und das wird in der Regel nicht der Fall sein.

(2.1.16) Ein einfaches Gegeneispiel wird durch $\mathrm{G(k)}=\mathrm{e}^{-k^4}$ gegeben, eine Funktion, die als Verallgemeinerung der Normalverteilung naheliegt, deren inverse Fouriertransformierte aber negative Werte annimmt. (Das Integral läßt sich nicht elementar auswerten!)

(2.1.17) Für die Normalverteilung selbst findet man (mit leichter Rechnung, ausführen!)

$$\boxed{G_{\mu,\sigma^2}(k) = e^{\mu(ik)+\frac{1}{2}\sigma^2(ik)^2}}.$$

Allgemein wird man versuchen, möglichst viele auftretetende Größen als Erwartungswert zu schreiben oder durch Erwartungswerte auszudrücken.

(2.1.18) Wird nicht über den gesamten Bereich integriert, wird man mit der mengentheoretischen charakteristischen Funktion arbeiten - $\theta(x) = 1$, für x$\in A$ sont $\theta(x) = 0$.

17.2.1c **Die Kumulanten**

(2.1.19) Jetzt sei p die Dichte und G die charakteristische Funktion einer Verteilung. Beachten Sie, daß die Potenzreihe von G mit 1 beginnt, so daß man die Reihenentwicklung von $\ln(1+\mathrm{x})=\Sigma_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}x^k$ verwenden kann. Man bildet die folgende neue Potenzreihe

$$\boxed{\ln(G(k)) = \ln(1+\overbrace{\sum_{n=1}^{\infty}\frac{\mu_n}{n!}(ik)^n}) = \sum_{n=1}^{\infty}\frac{\kappa_n}{n!}(ik)^n}.$$

Man setzt die Reihe von G(k) in die von ln(1+x) ein. Die Reihe ist absolut konvergent, so dass man umordnen kann. Die durch Umordnung nach Potenzen von k entstehenden Koeffizienten κ_n heißen "Kumulanten". Wir werden ihren Nutzen beim Beweis des zentralen Grenzwertsatzes erkennen. Da die Kumulantenreihe durch Umordnung der Momentenreihe entsteht, kann man die κ_n durch die Momente bzw. die zentralen Momente ausdrücken.

(2.1.20) **Rechnerische Einzelheiten und Beispiel.**

Wir gehe die Definitionen und Schritte nacheinander nochmals durch:

1. Die **Charakteristische Funktion (einer Verteilung)** ist allgemein definiert durch:

$$G(k) = \int dx e^{ikx} p(x).$$

   Mit $x = ik$ folgt beim Entwickeln :

$$G(k) = \sum_{n=0}^{\infty} \frac{\mu_n}{n!} x^n = 1 + \mu_1 x + \frac{1}{2}\mu_2 x^2 + \frac{1}{3!}\mu_3 x^3 + ...$$

2. Die Kumulanten (der Verteilung) sind gegeben durch

$$H(k) = \sum_{n=1}^{\infty} \frac{\kappa_n}{n!} x^n = \kappa_1 x + \frac{1}{2}\kappa_2 x^2 + \frac{1}{3!}\kappa_3 x^3 + ...$$

3. Der Bezug wird hergestellt durch:

$$H = \ln(G) \qquad G = e^H$$

4. Wir werten die Gleichung G=e^H in folgender Form aus.

$$G = e^{\kappa_1 x + \frac{1}{2}\kappa_2 x^2 + \frac{1}{3!}\kappa_3 x^3 + ...} = e^{\kappa_1 x} \cdot e^{\frac{1}{2}\kappa_2 x^2} \cdot e^{\frac{1}{3!}\kappa_3 x^3 + \frac{1}{4!}\kappa_4 x^4 + ...}$$

   (a) Entwickeln der drei Faktoren gibt angemessen angeordnet:

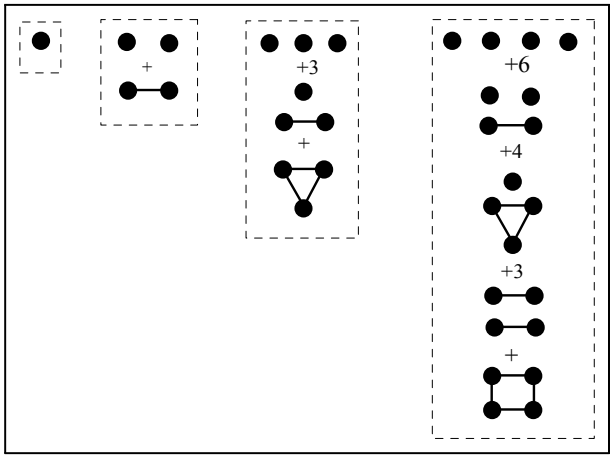
$$\begin{array}{lllll} 1 & +\kappa_1 x & +\frac{1}{2}\kappa_1^2 x^2 & +\frac{1}{3!}\kappa_1^3 x^3 & +\frac{1}{3!}\kappa_1^4 x^4 + ... \\ 1 & & +\left(\frac{1}{2!}\kappa_2 x^2\right) & & +\frac{1}{2!}\left(\frac{1}{2!}\kappa_2 x^2\right)^2 \\ 1 & & & +\frac{1}{3!}\kappa_3 x^3 & +\frac{1}{4!}\kappa_4 x^4 + \end{array}$$

   (b) Wir wollen die zentralen Momente μ durch die Kumulanten κ ausdrücken. Ausmultiplizieren und Sammeln der Beiträge bis zur vierten Ordnung gibt für G:

$$\begin{aligned} & 1 \quad +\kappa_1 x \quad +\tfrac{1}{2}\left(\kappa_1^2 + \kappa_2\right) x^2 \quad +\tfrac{1}{3!}\left(\kappa_1^3 + 3\kappa_1^2\kappa_2 + \kappa_3\right) x^3 \\ & \qquad + \frac{1}{4!}\left(\kappa_1^4 + 6\kappa_1^2\kappa_2 + 4\kappa_1\kappa_3 + 3\kappa_2^2 + \kappa_4\right) x^4 + ... \\ = \quad & 1 + \mu_1 x + \frac{\mu_2}{2!} x^2 + \frac{\mu_3}{3!} x^3 + \frac{\mu_4}{4!} x^4 + ... \end{aligned}$$

   Koeffizientenvergleich liefert die gewünschten Formeln.

   (c) Eine naheliegende **graphische Darstellung und Veranschaulichung** der ersten Beiträge ist:



Beschreibung und Interpretation: Jeder Beitrag μ_n enthält zusammenhängende und unzusammenhängende Komponenten. Jede zusammenhängende Komponente enthält einen "Vertexfaktor" κ_r, wobei r gleich der Zahl der in der Komponente verbundenen Punkte ist. Gehört der Beitrag zu μ_n, so ist die Anzahl aller Punkte gleich n. Schließlich gehört zu jedem Beitrag ein kombinatorischer Faktor.

□ **Auf wieviele Weisen kann man Unzusammenhängendes aus Zusammenhängendem (zu einem gegebenen Grad) kombinieren?** Vgl. die Zykeldarstellung der Permutationen in Kap. 3.

(2.1.22) Natürlich kann man umgekehrt die Kumulanten durch die Momente ausdrücken. Dazu muß man nur H=ln(G) schreiben und die Reihenentwicklung ausführen. Wir haben

$$\begin{aligned} H(k) &= \sum_{n=1}^{\infty} \frac{\kappa_n}{n!} x^n = \kappa_1 x + \frac{1}{2}\kappa_2 x^2 + \frac{1}{3!}\kappa_3 x^3 + \frac{1}{3!}\kappa_4 x^4 + ... \\ G(k) &= \sum_{n=0}^{\infty} \frac{\mu_n}{n!} x^n = 1 + \left(\mu_1 x + \frac{1}{2}\mu_2 x^2 + \frac{1}{3!}\mu_3 x^3 + \frac{1}{4!}\mu_4 x^4 + ... \right) \end{aligned}$$

G=1+X ist in die Entwicklung

$$log(1+X) = X - \frac{1}{2}X^2 + \frac{1}{3}X^3 - \frac{1}{4}X^4 + ...$$

einzusetzen. Es ist

$$\begin{aligned} X^2 &= \left(\mu_1 x + \frac{1}{2}\mu_2 x^2 + \frac{1}{3!}\mu_3 x^3 + ... \right)^2 \\ &= \mu_1^2 x^2 + 2 \cdot \frac{1}{2}\mu_1\mu_2 x^3 + (2\mu_1 \frac{1}{3!}\mu_3 + \frac{1}{4}\mu_2^2) x^4 + ... \end{aligned}$$

usw. Dann gibt Sammeln der einschlägigen Terme:

$\mu_1 x$	$+\frac{1}{2}\mu_2 x^2$	$\frac{1}{3!}\mu_3 x^3$	$+\frac{1}{4!}\mu_4 x^4$	
	$-\frac{1}{2}\cdot\mu_1^2 x^2$	$-\frac{1}{2}\cdot 2\frac{1}{2}\mu_1\mu_2 x^3$	$-\frac{1}{2}\cdot(2\mu_1\frac{1}{3!}\mu_3 + \frac{1}{4}\mu_2^2)x^4$	
		$+\frac{1}{3}\cdot\mu_1^3\mu_1^3 x^3$	$+\frac{1}{3}\cdot\left(\frac{3}{2}\mu_1^2\mu_2\right)$	
			$-\frac{1}{4}\cdot\mu_1^4 x^4$	+...

(2.1.23) Vergleich gibt, wobei die oben mit Hilfe der μ eingeführten Größen s (Schiefe) und e (Exzess) mit verwenden:

$$\begin{aligned} \kappa_1 &= \mu_1 = \mu = Mittelwert \\ \kappa_2 &= \mu_2 - \mu_1^2 = \sigma^2 = Varianz \\ \kappa_3 &= \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 = s\sigma^3 \\ \kappa_4 &= \mu_4 - 4\mu_1\mu_3 - 3\mu_2^2 + 12\mu_1^2\mu_2 - 6\mu_1^4 = e\sigma^4 \end{aligned}$$

Und umgekehrt:

$$\begin{aligned}\mu_1 &= \kappa_1\\ \mu_2 &= \kappa_1^2+\kappa_2\\ \mu_3 &= \kappa_1^3+3\kappa_1^2\kappa_2+\kappa_3\\ \mu_4 &= \kappa_1^4+6\kappa_1^2\kappa_2+4\kappa_1\kappa_3+3\kappa_2^2+\kappa_4\end{aligned}$$

(2.1.24) Berechnen wir als **Beispiel** die Größen für die Normalverteilung. Hier ist G(K) sofort berechenbar zu:

$$G(k)=e^{\mu(ik)+\frac{1}{2}\sigma^2(ik)^2}$$

Alle Kumulanten folgen hieraus unmittelbar. Oder auch: G ergibt sich für die Normalverteilung in der Kumulantenform! Die Momentgrößen lassen sich dann auf mehrere Weisen bestimmen. Sie sind in folgender Tabelle zusammengestellt:

	0	1	2	3	4	5
μ_n	1	μ	$\mu^2+\sigma^2$	$\mu\left(\mu^2+3\sigma^2\right)$	$\mu^4+6\mu^2\sigma^2+3\sigma^4$	$\mu\left(\mu^4+10\mu^2\sigma^2+15\sigma^4\right)$
κ_n	0	μ	σ^2	0	0	0
$\langle(x-\mu)^n\rangle$	1	0	σ^2	0	$3\sigma^4$	

(2.1.25) Allgemein ergibt sich

$$G(k)=e^{\mu(ik)+\frac{1}{2}\sigma^2(ik)^2+\frac{1}{3!}s\sigma^3(ik)^3+\frac{1}{4!}e\sigma^4(ik)^4+....}$$

Man sieht:

> **H liefert uns eine Folge von charakteristischer Beschreibungsgrößen der Verteilung mit abnehmender Bedeutung.** Im Fall der Normalverteilung verschwinden Schiefe und Exzess. Zur Festelgung dieser Verteilung genügen die beiden Größen Mittelwert und Streuung vollständig.

(2.1.26) **Beispiel: Dreiecksverteilung.**
Ein weiteres gut rechenbares Beispiel ist:

$$f_{m,a}(x)=\begin{cases}\frac{1}{a}\frac{x+a}{m+a} & \text{für } -a\le x\le m\\ \frac{1}{a}\frac{x-a}{m-a} & \text{für } m\le x\le a\end{cases}$$

oder einheitenfrei

$$f_e(u)=\begin{cases}\frac{u+1}{e+1} & \text{für } -1\le u\le e\\ \frac{u-1}{e-1} & \text{für } e\le u\le 1\end{cases}\quad .\qquad (\text{-1}\le e\le 1)$$

Bestimmung des **Mittelwertes**:

$$\mu_1=\int_{-1}^{e}\frac{u+1}{e+1}udu+\int_{e}^{1}\frac{u-1}{e-1}udu=\frac{1}{3}e^2+\frac{1}{6}e-\frac{1}{6}-\frac{1}{6}\frac{1+2e^3-3e^2}{e-1}=\frac{1}{3}e$$

$$\boxed{\mu_1=\tfrac{1}{3}\mathrm{e}}$$

Bestimmung der **Varianz:**

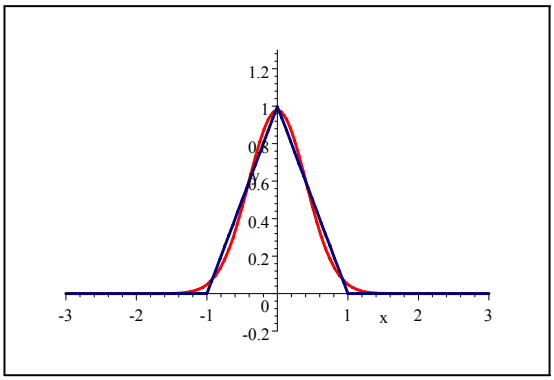
$$\begin{aligned}\mu_2 &= \int_{-1}^{e}\frac{u+1}{e+1}u^2du+\int_{e}^{1}\frac{u-1}{e-1}u^2du=\frac{1}{4}e^3+\frac{1}{12}e^2-\frac{1}{12}e+\frac{1}{12}-\frac{1}{12}\frac{1+3e^4-4e^3}{e-1}\\ &= \frac{1}{6}e^2+\frac{1}{6}\qquad \text{Also}\qquad \sigma^2=\frac{1}{6}e^2+\frac{1}{6}-\frac{1}{9}e^2=\frac{1}{18}e^2+\frac{1}{6}\end{aligned}$$

$$\begin{aligned}\mu_2&=\tfrac{1}{6}e^2+\tfrac{1}{6}\\ \sigma^2&=\tfrac{1}{18}e^2+\tfrac{1}{6}=\tfrac{1}{18}(e^2+3)\end{aligned}$$

Vergleich mit der Normalverteilung mit demselben Mittel und derselben Varianz. Sei (e=0)

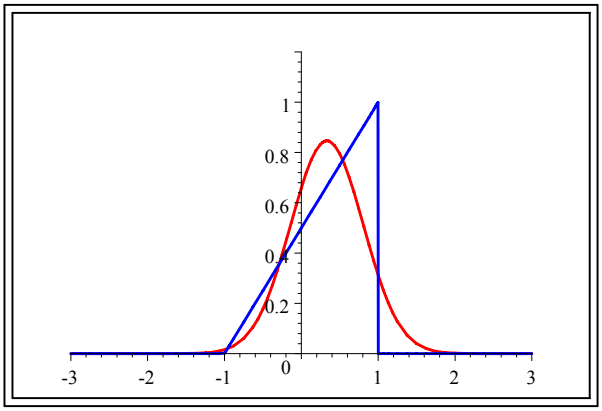
$$f(x) = \begin{cases} 0 & x < -1 \\ (x+1) & -1 \le x < 0 \\ (1-x) & 0 \le x < 1 \\ 0 & 1 \le x \end{cases}$$

Also $\mu_1 = 0$ und $\sigma^2 = \frac{1}{6}$. Die Normalverteilung $\frac{1}{\sqrt{2\pi\frac{1}{6}}} e^{-\frac{6x^2}{2}}$ mit derselben Streuung sieht wie folgt aus:



Und nun der extrem schiefe Fall e=1 und die Normalverteilung h mit denselben Beschreibungsgrößen.

$$g(x) = \begin{cases} 0 & x \le -1 \\ \frac{1}{2}(x+1) & -1 < x \le 1 \\ 0 & 1 < x \end{cases} \qquad h(x) = \sqrt{\frac{9}{4\pi}} e^{-\frac{9(x-\frac{1}{3})^2}{4}}$$



(2.1.27) Jetzt berechnen wir noch die beiden anderen Beschreibungsgrößen für die Dreiecksverteilung. Die **Schiefe.**

$$\left\langle (x-\mu_1)^3 \right\rangle = \int_{-1}^{e} \frac{u+1}{e+1}(u-\frac{e}{3})^3 du + \int_{e}^{1} \frac{u-1}{e-1}(u-\frac{e}{3})^3 du = ... \frac{1}{135}\left(e^2-9\right)e$$

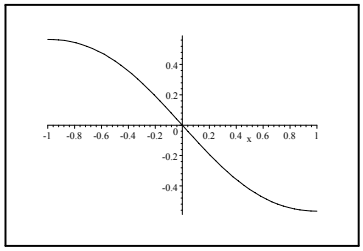
Also:

$$s = \frac{\kappa_3}{\sigma^3} = \frac{\frac{1}{135}\left(e^2-9\right)e}{\sqrt{\left(\frac{1}{18}e^2+\frac{1}{6}\right)^3}} = \frac{4}{5}\left(e^2-9\right)\frac{e}{\left(e^2+3\right)\sqrt{\left(2e^2+6\right)}}$$

$$\left\langle (x-\mu_1)^3 \right\rangle = \tfrac{1}{135}\left(e^2-9\right)e$$

$$s = \tfrac{4}{5}\frac{e\left(e^2-9\right)}{\left(e^2+3\right)\sqrt{\left(2e^2+6\right)}} \qquad (-1 \le e \le 1)$$

Graphisch gibt das für die Schiefe:



Und nun der **Exzess:**

$$\begin{aligned}\left\langle (x-\mu_1)^4 \right\rangle &= \int_{-1}^{e} \frac{u+1}{e+1}(u-\frac{e}{3})^4 du + \int_{e}^{1} \frac{u-1}{e-1}(u-\frac{e}{3})^4 du = ... \\ &= \frac{1}{135}e^4 + \frac{2}{45}e^2 + \frac{1}{15} = \frac{1}{135}\left(e^2+3\right)^2\end{aligned}$$

Also

$$\begin{aligned}e\sigma^4 &= \left\langle (x-\mu_1)^4 \right\rangle - 3\sigma^4 = \frac{1}{135}\left(e^2+3\right)^2 - 3\left(\frac{1}{18}e^2+\frac{1}{6}\right)^2 \\ &= -\frac{1}{540}e^4 - \frac{1}{90}e^2 - \frac{1}{60} \qquad \text{Und:} \quad \boxed{\frac{e\sigma^4}{\sigma^4} = \frac{-\frac{1}{540}\left(e^2+3\right)^2}{\left(\frac{1}{18}e^2+\frac{1}{6}\right)^2} = -\frac{3}{5}}\end{aligned}$$

Und

$$\begin{aligned}&\left\langle (x-\mu_1)^4 \right\rangle = \tfrac{1}{135}\left(e^2+3\right)^2 \\ &e = -\tfrac{3}{5}\end{aligned}$$

Und schließlich die charakteristische Funktion:

$$\begin{aligned}G(k) &= \int_{-1}^{r} \frac{u+1}{r+1}e^{iku}du + \int_{r}^{1} \frac{u-1}{r-1}e^{iku}du = \\ &\frac{kr\sin kr + \cos kr + k\sin kr - \cos k}{k^2r+k^2} + \frac{\cos k + k\sin kr - kr\sin kr - \cos kr}{k^2r-k^2} + \\ &i\left(\frac{-kr\cos kr + \sin kr - k\cos kr + \sin k}{k^2r+k^2} + \frac{\sin k - k\cos kr + kr\cos kr - \sin kr}{k^2r-k^2}\right) \\ &= -2\frac{\cos kr - \cos k + i\sin kr - i\left(\sin k\right)r}{k^2\left(r^2-1\right)}\end{aligned}$$

$$G(k) = -2\frac{\cos(kr) - \cos k + i\sin(kr) - i\left(\sin k\right)r}{k^2\left(r^2-1\right)}$$

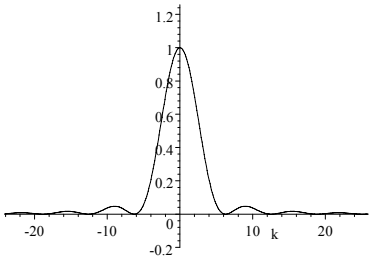
$$\boxed{\mathrm{G(k)}=-2\frac{e^{ikr}-\cos k - ir\sin k}{k^2(r^2-1)}}$$

Test für k=0: $-2\frac{(1+ikr+\frac{1}{2}(ikr)^2-1+\frac{1}{2}k^2-irk}{k^2(r^2-1)} = \frac{k^2r^2-k^2}{k^2(r^2-1)} = 1$ Stimmt!

G(k) ist eine ganze holomorphe Funktion!. Das folgt daraus, daß der Träger der Verteilung selbst kompakt ist!

Für r=0 folgt

$G_{r=0}(k) = 2\frac{1-\cos k}{k^2}$



Die Entwicklung von G nach den Momenten:

$$\begin{aligned} -2\frac{e^{ikr}-\cos k-ir\sin k}{k^2\left(r^2-1\right)} &= 1+\frac{1}{3}irk-\frac{1}{12}(r^2+1)k^2 \\ &\quad -\frac{1}{60}i\left(r^2+1\right)rk^3+\frac{1}{360}\left(r^2-r+1\right)\left(r^2+r+1\right)k^4 \\ &\quad +\frac{1}{2520}i\left(r^2-r+1\right)\left(r^2+r+1\right)rk^5+.... \end{aligned}$$

Und die Kumulantenentwicklung:

$$\begin{aligned} H(k) &= \left(\frac{1}{3}ir\right)k+\left(-\frac{1}{36}r^2-\frac{1}{12}\right)k^2+\left(-\frac{1}{810}i\left(r^2-9\right)r\right)k^3 \\ &\quad +\left(-\frac{1}{12\,960}\left(r^2+3\right)^2\right)k^4+\left(-\frac{1}{68\,040}i\left(r^2+3\right)(r-3)\,r\,(r+3)\right)k^5 \\ &\quad +\left(\frac{1}{12\,247\,200}\left(r^6-153r^4+351r^2-135\right)\right)k^6+.... \end{aligned}$$

Und in der richtigen Form:

$$\boxed{\begin{array}{l} H(k)=\left(\frac{1}{3}r\right)(ik)+\frac{1}{2}\left(\frac{1}{18}(r^2+3)\right)(ik)^2+\frac{1}{6}\left(\frac{r(r^2-3)}{135}\right)(ik)^3 \\ +\frac{1}{24}\left(\frac{-\left(r^2+3\right)^2}{540}\right)(ik)^4+\frac{1}{120}\left(-\frac{\left(r^2+3\right)(r-3)r(r+3)}{567}\right)(ik)^5+... \end{array}}$$

17.3 Der volle theoretische Bereich und seine Kopplung an die Datenwelt

17.3.1 Der Ereignisraum

*Was entspricht den **Datensätzen** im theoretischen Bereich? Offenbar nicht unsere bisherigen Wahrscheinlichkeitsräume $(\mathbb{R}, \mathcal{B}, w)$. Die gehören zu den Auszählungen und den damit verbundenen relativen Häufigkeiten. Wir erwarten, daß die **Verteilungsfunktion** des theoretischen Bereichs der **Verteilung** der relativen Häufigkeiten entspricht. Die jetzt einzuführende Größe schließt diese Lücke, wie wir sehen werden.*

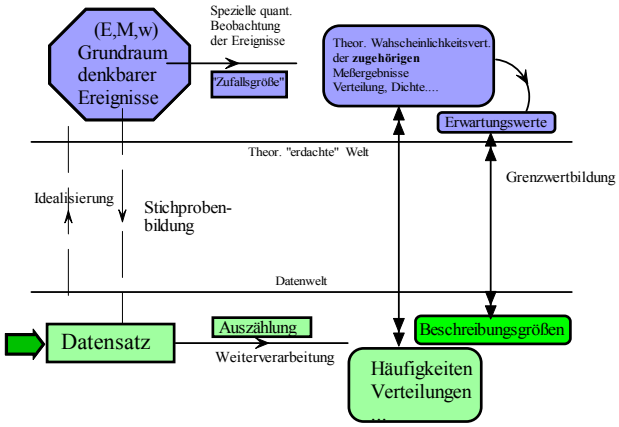
(3.1.1) Wir starten mit einem allgemeineren Wahrscheinlichkeitsraum $(E, \mathfrak{M}, w)$, bei dem der Grundraum, der "Ereignisraum" in der Regel nicht $\mathbb{R}$ ist. Dieses E soll so etwas sein, wie der "Raum aller möglichen Ereignisse eines bestimmten Typs. Wir wollen nicht versuchen, das genauer zu formalisieren (d.h. die Elemente im Sinne mathematischer Präzisierung festzulegen), sondern uns darauf beschränken, anzugeben, was dieser Wahrscheinlichkeitsraum $(E, \mathfrak{M}, w)$ leisten soll. (Das erspart es dem Leser keineswegs, weiter darüber nachzudenken und speziell konkrete Fälle zu durchdenken. Aber es ist günstig, zunächst zu sehen, was dieser Raum leisten soll.)

- Zum einen soll er zugehörige Datensätze (des konkreten Bereiches) produzieren. D.h. sein w legt - eventuell mit Beobachtungsfehlern - die Häufigkeitsverteilung der Daten dieser Datensätze fest. Das gibt die Beziehung zum konkreten Bereich. Die so entstehenden Datensätze sind sind das, was man üblicherweise als "Stichproben" bezeichnet.
- An jedem Ereignis lassen sich viele Beobachtungen und Messungen mit quantitativem Ergebnis vornehmen. Über die nachfolgende Konstruktion der Zufallsgröße produziert er - der Ereignisraum - für jeden solchen Meßtyp die zugeörigen theoretischen Häufigkeitsverteilungen, die man dann mit entsprechenden Datensatzauszählungen vergleicht .

(3.1.2) Damit besitzen wir eine erste ganz grobe Skizze über die Zusammenhänge der beiden Bereiche.

Nochmals zum Grundraum $(E, \mathfrak{M}, w)$, dem "Ereignisraum". Man soll mit ihm weniger konkret mathematisch, als konzeptionell arbeiten. Er dient dazu, die Gedanken zu ordnen und voranzubringen. Und das hängt wenig von einer genauen Elementfestlegung (der Menge) ab. Vielfach ist es nützlich, sich E als eine Art unausschöpfliches Reservoir vorzustellen, aus dem man nach Belieben Datensätze oder Stichproben ziehen kann, die dann (unter gewissen Vorsichtsmaßnahmen) gemäß w verteilt sind.

(3.1.4) Die Zusammenhänge als Diagramm:



17.31a **Unabhängige Daten**

(3.1.5) Häufig ist E ein Produktraum. Sagen wir $E = F \times G = \{(f,g), | f \in F, g \in G\}$. Die Ereigniselemente bestehen hier aus geordneten Paaren von Einzelereignissen. So kann man beispielsweise dieselbe Messung zweifach ausführen und die beiden Resultate zu einer neuen zweikomponentigen Größe zusammenfassen. Natürlich ist F=G möglich.

(3.1.6) Formal haben wir es in einer solchen Situation mit drei Wahrscheinlichkeitsräumen zu tun. Meist interessiert, welche Beziehung zwischen den drei Wahrscheinlichkeiten besteht. Die Maßtheorie bietet als naheliegenden Kandidaten das Produktmaß $\mu \otimes \nu$ an. Damit erhebt sich die Frage, wie dieses Maß im Rahmen der Wahrscheinlichkeitstheorie zu interpretieren ist.

(3.1.7) Gehen wir auf die Definition des Produktmaßes zurück. Zu F gehöre die Wahrscheinlichkeit w und zu G die Wahrscheinlichkeit v. Dann gilt für allgemeine Rechtecke

$$(w \otimes v)(A \times B) = w(A)v(B)$$

Das beinhaltet aber gerade die **statistische Unabhängigkeit** der beiden Ereignisse, wie aus der elementaren Wahrscheinlichkeitsrechnung her bekannt ist. Inhaltlich: das Resultat der zweiten Messung wird durch das der ersten nicht beeinflußt oder eingeschränkt, ist unabhängig vom ersten Resultat.

(3.1.8) Wir wissen bereits, dass allgemein für die Verteilung und ihre eventuelle Dichte gilt

$$\Phi(x,y) = \int \int_{\substack{u \leq x \\ v \leq y}} d\mu(u) d\nu(v) \quad \text{und} \quad \varphi(x,y) = \frac{\partial^2 \Phi}{\partial x \partial y}(x,y).$$

Im Falle der Unabhängigkeit folgt daraus sofort, daß auch die Verteilungsfunktion und die die Dichte faktorisieren:

$$\boxed{\begin{array}{l} \Phi(x,y) = \Phi_1(x)\Phi_2(y) \\ \varphi(x,y) = \varphi_1(x)\varphi_2(y) \end{array}}$$

Auch die Umkehrung dieser Aussage gilt.

(3.1.9) Sind die Ereignisse nicht unabhängig, dann wird die Verteilung durch ein anderes Maß als $\mu \otimes \nu$ beschrieben.

Vgl....

17.3.1b **Zufallsgrößen (=stochastische Variable)**

(3.1.10) Wir verbleiben jetzt im theoretischen Bereich. Sei also $(E, \mathfrak{M}, w)$ unser Wahrscheinlichkeitsraum. Wir betrachten eine Abbildung $X : E \to \mathbb{R}$. Das ist so etwas wie ein **Skalarfeld möglicher Meßergebnisse auf E**, ist also interpretierbar als eine bestimmte Meßoperation, die man an den Ereignissen von E vornehmen kann, wobei jedem Ereignis e$\in E$ eine Zahl X(e) zugeschrieben wird, interpretiert als Ergebnis einer idealen Messung am Ereignis e. Welche Information liefert uns dann w über X? Nun, ist $A \subset E$, dann ist $w(A)$ die Wahrscheinlichkeit, ein Ergebnis in A zu finden.

> Vermutung: Welche Beziehung besteht zu X? Ist $B \subset \mathbb{R}$, dann können wir $\overset{-1}{\underline{X}}(B) \subset E$ und folglich $w(\overset{-1}{\underline{X}}(B))$ bilden. Das ist eine interessante Größe. Sie gibt die Wahrscheinlichkeit dafür, unter X ein Meßergebnis aus dem Bereich B vorzufinden.

(3.1.11) Können wir das als **Wahrscheinlichkeitsmaß auf $\mathbb{R}$ interpretieren**? Offenbar muß dazu für jede Borelmenge $B \in \mathbb{R}$ die Menge $\overset{-1}{\underline{X}}(B)$ in der σ-Algebra $\mathfrak{M}$ von E liegen. Ist das der Fall, dann definiert

$$\boxed{\Phi_X(\mathrm{x}) = \mathrm{w}(\overset{-1}{\underline{X}}(]\text{-}\infty, \mathrm{x}]))}$$

tatsächlich die Verteilungsfunktion Φ_X einer Wahrscheinlichkeit W_X auf $\mathbb{R}$. Nicht auf E, wo das ursprüngliche w wirkt. Man definiert:

> (3.1.12) **Definition:** $(E, \mathfrak{M}, w)$ Wahrscheinlichkeitsraum. Eine Abbildung $X : E \to \mathbb{R}$ heißt "stochastische Variable" oder "Zufallsgröße" auf E, wenn für jede Borelmenge $B \subset \mathbb{R}$ gilt: $w(\overset{-1}{\underline{X}}(B)) \in \mathfrak{M}$. D.h., wenn die Abbildung X im Sinne der allgemeinen Maßtheorie meßbar ist.

Dann definiert $\Phi_X(x) = w(\overset{-1}{\underline{X}}(]-\infty, x])$ eine reelle Verteilungsfunktion, die man "Verteilungsfunktion der stochastischen Variablen X" nennt. Die zugehörige Wahrscheinlichkeit W_X auf $\mathbb{R}$ erfüllt für alle Borelmengen B von $\mathbb{R}$ die Beziehung

$$W_x(B) = w(\overset{-1}{\underline{X}}(B)).$$

□ Was wird man unter einer "vektorwertigen Zufallsvariablen" verstehen? Als Vektorraum kommt vornehmlich $\mathbb{R}^n$ in Betracht, insbesondere $\mathbb{R}^2$.

Orientierungsbeispiele:....(Endlicher Bereich mit Stichproben / Computermodell)

(3.1.13) Die Maß- und Integrationstheorie liefert unmittelbar die folgenden Beziehungen, die es ermöglichen, Integrale zum eigentlich theoretisch interessierenden (schwer zugänglichen) E in solche über den Meßbereich $\mathbb{R}$ einer bestimmten Meßvorschrift X umzuwandeln. (dW und $d\Phi$ bezeichnen also hier Integrale über $\mathbb{R}$) :

$\int_A dw(e) = \int_B dW_X(x) \qquad \text{mit} \quad A = \overset{-1}{\underline{X}}(B) \subset E$
$\langle f \circ X \rangle_w = \int_E dw(e) f(X(e)) = \int_{\mathbb{R}} d\Phi_X(x) f(x) = \langle f \rangle_{W_X}$

Aus der vielfach unhandlichen Mengenfunktion w auf der unzugänglichen σ-Algebra $\mathfrak{M}$ wird die besser zugängliche Verteilungsfunktion Φ_X für die Meßergebnisse von X. Und ist die Verteilung glatt, dann kann man sogar zur Dichte $\varphi_X(x) = \Phi'(x)$ übergehen. Nochmals zur Verdeutlichung: E beschreibt die "Ereignisse". Diese sind idealisiert, benötigen aber in der Regel zur Erfassung und Beschreibung nicht nur eine, sondern mehrere bis viele Zahlangaben. Die beschreibenden Zahlangaben erhält man jeweils über eine Zufallsgröße.

17.3.1c **Die mit den Zufallsgrößen verbundene mathematische Struktur**

(3.1.14) Zufallsgrößen $E \to \mathbb{R}$ sind Abbildungen. Durch Wertemengenübertragung übernehmen sie die algebraische Struktur von $\mathbb{R}$. d.h. hier eine Ringstruktur. Die mathematische Frage der Übertragbarkeit der Meßbarkeitseigenschaft klammern wir aus. **Praktisch bedeutet das, daß man Formelausdrücke aus Zufallsvariablen bilden kann!** Sind X,Y,Z Zufallsgrößen auf E, dann ist auch X+2Y-3XZ eine Zufallsgröße, die eine zugehörige kombinierte Beobachtung repräsentiert. Die Schreibweise ist allerdings nicht ganz eindeutig, weil man neben der Wertemengenübertragung auch noch das Abbildungsprodukt hat. So läßt sich X+X auf zwei **hier zu unterscheidende Weisen** als Abbildung interpretieren.

$$\begin{aligned} (E, e &\mapsto X(e) + X(e) = 2X(e), \mathbb{R}) \\ &\qquad \text{Wertemengenübertragung} \\ (E \times E, (e_1, e_2) &\mapsto X(e_1) + X(e_2), \mathbb{R}) \\ &\qquad \text{Abbildungsprodukt } X \times 1 + 1 \times X \end{aligned}$$

(3.1.15) Beide Interpretationen sind für die Statistik wichtig. Im Falle des Produktraumes ist noch sorgsam auf das zugehörige Maß zu achten. Liegt das Produktmaß vor im Sinne unabhängiger Daten oder ein anderes Maß in den **zwei** Variablen?

Sind schließlich E und F zwei verschiedene Ereignisräume mit zugehörigen Zufallsvariablen X und Y, dann ist auch $(E \times F, (e, f) \to X(e) + Y(f), \mathbb{R})$ eine Zufallsvariable, die man meist erneut mit X+Y bezeichnet.

Damit es nicht zu Verwechslungen kommt, werden wir im Zweifelsfall immer das zugehörige Abbildungstripel angeben.

(3.1.16) Zu jeder so gebildeten Zufallsgröße S gehört nach unserem Konzept eine zugehörige Wahrscheinlichkeitsfunktion W_s mit Verteilung Φ_S. Es entsteht das Problem, etwa die Verteilung von W_S aus den Verteilungen der Bestandteile zu bilden. **Das ist häufig deutlich komplizierter als die Bildung der stochastischen Variablen selbst. Andererseits benötigt man diese Verteilungen.**

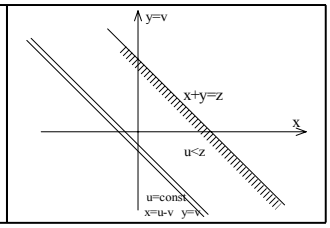
(3.1.17) Wir diskutieren das Problem am **Beispiel der Summenbildung.**

Es seien X und Y Zufallsgrößen zum Ereignisraum E mit Dichten φ und ψ. Welche Dichte hat dann die Zufallsgröße

$$S = X + Y = (E \times E, (e, f) \mapsto X(e) + Y(f), \mathbb{R}) \quad ?$$

Der Einfachheit halber nehmen wir an, daß die Produktwahrscheinlichkeit W_S zu (X, Y) glatt ist und durch eine zweidimensionale Dichte $\rho_S(x, y)$ beschrieben wird. Dann gilt:

$$\begin{aligned}\Phi_S(z) &= w(\{(e,f)|X(e)+Y(f)\le z\})\\ &= W_S(\{(x,y)|x+y\le z\})\\ &= \int\int_{x+y\le z} dxdy\rho_S(x,y)\\ &= \int_{-\infty}^{z} du\left(\int_{-\infty}^{\infty} dv\rho_S(u-v,v)\right)\end{aligned}$$



(3.1.18) Damit haben wir die gesuchte Dichte zu Φ_S gefunden! Im letzten Schritt haben wir die Substitution u=x+y, v=y ausgeführt, wobei u die gesuchte Summengröße ergibt. Ergebnis:

Satz: Die Verteilung der Zufallsgröße S=X+Y ergibt sich wie folgt:

$$\boxed{\Phi_S(\mathrm{z})=\int_{-\infty}^{z}\mathrm{du}\left(\int_{-\infty}^{\infty} dv\rho_S(u-v,v)\right)}$$

$$\boxed{\varphi_S(\mathrm{z})=\int_{-\infty}^{\infty}\mathrm{dv}\rho_S(\mathrm{z\text{-}v,v})}$$

Sind X und Y statistisch unabhängig, d.h. gilt $\rho_S(x, y) = p(x)q(y)$, dann folgt

$$\boxed{\varphi_S(\mathrm{z})=\int_{-\infty}^{\infty}\mathrm{dvp(z\text{-}v)q(v)}}$$

(3.1.19) Durch Fouriertransformation folgt die charakteristische Funktion schließlich zu

$$\boxed{\mathrm{G}_S(\mathrm{k})=\mathrm{G}_X(\mathrm{k})\mathrm{G}_Y(\mathrm{k})}.$$

Das ist ein überaus wichtiges Resultat!

☐ Beweisen Sie den folgenden Sachverhalt:

Satz: Sind X_1 und X_2 beides normalverteilte Zufallsvariablen mit Mittelwerten μ_1 bzw. μ_2 und Varianzen σ_1^2 bzw. σ_2^2, **dann** ist $S = X_1 + X_2$ erneut normalverteilt mit Mittelwert $\mu_1+ \mu_2$ und Varianz $\sigma_1^2 + \sigma_2^2$.

Der Beweis ist leicht, wenn Sie frühere Resultate geschickt verwenden! Und beachten Sie, daß das Resultat die Ausnahme ist: **Typischerweise hat die Summe zweier Zufallsgrößen einen anderen Verteilungstyp als die Summanden.**

☐ X_1 und X_2 seien gleichverteilt und unabhängig. Wie sieht die Verteilung von $X_1 + X_2$ aus? Und analog von $X_1 + X_2 + X_3$? Wie läßt sich das geometrisch interpretieren? Kann man mit der charakteristischen Funktion arbeiten?

17.3.2 Die Weiterverarbeitung von Verteilungen und Datensätzen

(3.2.1) Sei X eine Zufallsgröße mit Wahrscheinlichkeitsverteilung Q und Wahrscheinlichkeit w=w_Q. Weiter sei f eine reelle Funktion. Dann ist $f \circ X$ auch eine Zufallsgröße nach deren Verteilung P_f bzw. Dichte p_f wir jetzt fragen. Im Datenbereich heißt das, daß wir den neuen Datensatz $f \circ d$ bilden und dessen Häufigkeitsverteilung suchen.

Sei dazu $I_y =]-\infty, y]$. Dann folgt:

$$P_f(y) = w(\{t|f(t)\le y\}) = w(\{t|f(t)\in I_y)\}) = w(\{t|t\in \underline{f}^{-1}(I_y)\}) = \int_{\underline{f}^{-1}(I_y)} dw_Q$$

Jetzt nehmen wir an, daß BildX=I ein Intervall sei und daß $f : I \to J$ bijektiv und monoton wachsend sei. J sei auch ein Intervall. Dann gilt "$(f(t) \leq y \text{ mit } y \in J) \Leftrightarrow \left(t \leq \overset{-1}{f}(y)\right)$". Und das heißt

$$P_f(y) = w(\{t|f(t) \leq y\}) = w(\{t|t \leq \overset{-1}{f}(y)\}) = Q(\overset{-1}{f}(y)).$$

Für y außerhalb von J ist P_f durch 0 oder 1 zu ergänzen. Das ist die gesuchte Gleichung.

Allgemeinere Fälle lassen sich entsprechend behandeln.

□ Sei f:I→ J bijektiv und monoton fallend. Was gilt dann? Welche Komplikation tritt auf?

(3.2.2) Weiter nehmen wir an, daß f bijektiv, monoton wachsend und glatt ist. Dann können wir differenzieren und finden für die Dichten (q die gegebene von X und p die gesuchte von Y=f∘X):

$$\boxed{p(y) = f^{-1\prime}(y)q(f^{-1}(y))}$$

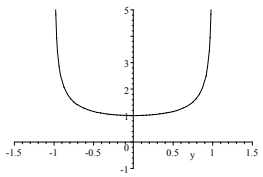
Ist f nicht wachsend, sondern fallend, ist der Betrag der Ableitung zu nehmen.

□ Das ist gut wie folgt zu merken:

$$p(y)dy = q(x)dx$$

Ist f nur nicht injektiv, nur stückweise glatt, wird man das wie folgt abändern: p(y)dy=$\Sigma q(x_i)dx_i$. Wie ist das zu verstehen?

(3.2.3) Beispiel: $X : E \to I = [-\frac{\pi}{2}, \frac{\pi}{2}]$ gleichverteilt und f(x)=sin(x). D.h. $q_X(x)=\frac{1}{\pi}$ auf I. Und J=[-1,1]. Das gibt für die gesuchte Dichte



$$p(y)dy = \frac{1}{\pi}\frac{dy}{\sqrt{1-y^2}}$$
$$P_f(y) = \frac{1}{\pi}\left(asn(y) + \frac{\pi}{2}\right)$$

Ein zugehöriger Datensatz läßt sich problemlos mit dem Computer simulieren und auszählen

□ Behandeln Sie entsprechend f=tan.

(3.2.4) Die hergeleitete Formel bietet sich zur Behandlung der folgenden Problemtypen an:

- p und q gegeben. bestimme f (über die entstehende Differentialgleichung)
- p und f gegeben. Bestimme q. (Wie im Beispiel mit f=sin)

(3.2.5) Noch ein Beispiel. Es sei f(x)=ax und a>0. Dann ist $f \circ X = aX$. Unsere Bedingungen sind erfüllt und wir finden für die Dichten

$$q_{aX}(y) = \frac{1}{a}q_X(\frac{1}{a}y).$$

Durch Integration folgt für die charakteristische Funktion wegen ky=(ak)(y/a):

$$\begin{aligned} G_{aX}(k) &= \int_{-\infty}^{\infty} dy e^{iky}\frac{1}{a}q_X(\frac{1}{a}y) = \int_{-\infty}^{\infty} dx e^{i(ak)x}q_X(x) = G_X(ak) \\ \text{Also} \quad &: \quad G_{aX}(k) = \sum \frac{\mu_n}{n!}(iak)^n = \sum \frac{(a^n\mu_n)}{n!}(ik)^n \\ \text{Und ebenso} \quad H_{aX}(k) &= \ln G_{aX}(k) = ... = \sum \frac{(a^n\kappa_n)}{n!}(ik)^n \end{aligned}$$

Das zeigt, wie sich die Momente und die Kumulanten bei einer Skalierung transformieren! **Zu merken über** $G_{aX}(k) = G_X(ak)$

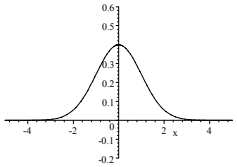
(3.2.6) Damit können wir zusammenfassend ein wichtiges Resultat formulieren:

> Sei $a_1, ..a_n \in \mathbb{R}$ und $X_1, ...X_n$ unabhängige Zufallsgrößen mit charakteristischen Funktionen $G_1, ...G_n$. Dann hat die (linear kombinierte) Zufallsgröße $S = a_1X_1 + ...a_nX_n$ die folgende charakteristische Funktion
>
> $$G_S(k) = G_1(a_1k) \cdot G_2(a_2k) \cdot .... \cdot G_n(a_nk)$$

□ Interpretieren Sie auch $(X, Y) \mapsto X + Y$ im Sinne einer Weiterverarbeitung.

17.3.2 Ergänzung: **Die Standardnormalverteilung**

(3.2.7) Die Normlverteilung mit Mittelwert 0 und Varianz 1 ist die "Standardnormalverteilung", deren Werte man üblicherweise in den Tabellen findet. Sei φ die zugehörige Dichte, Φ die Verteilung und X zugehörige stochastische Variable. Also



$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$
$$\Phi(x) = \int_{-\infty}^{x} \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$
$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\frac{x}{\sqrt{2}}} du e^{-u^2}$$

Dann führt die Weitertransformation $\sigma X + \mu$ zur allgemeinen Normalverteilung. Unsere Resultate liefern für die Dichte erwartungsgemäß

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} = \frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma})$$

Und über Integration folgt

$$\Phi_{\mu,\sigma^2}(x) = \Phi(\frac{x-\mu}{\sigma}).$$

□ Der übliche Rechenkalkül für die δ-Funktion enthält die folgende Regel:

$$\delta(f(x)) = \sum \frac{1}{|f'(x_0)|}\delta(x - x_0)$$

Summiert wird über die Nullstellen x_0 von f(x), die einfach sein sollen. Begründen Sie diese Regel mit den Resultaten dieses Abschnittes.

17.3.3 **Der zentrale Grenzwertsatz**

(3.3.1) Jetzt betrachten wir n unabhängige Zufallsgrößen $X_1, X_2, ..., X_n$ mit Mittelwerten $\mu_1, ...\mu_n$. Die übrigen Kumulanten sollen alle übereinstimmen. **D.h. es liegt immer dieselbe Verteilung vor, nur unterschiedlich verschoben.** Damit bilden wir die Linearkombination

$$S = \frac{1}{\sqrt{n}}(X_1 + X_2 + ...X_n)$$

(3.3.2) Wir bestimmen S näherungsweise, indem wir das Verhalten der Kumulanten von S bestimmen. Unser Resultat besagt: Die k-ten Kumulanten sind zu addieren und werden dabei jeweils mit einem Faktor $(\frac{1}{\sqrt{n}})^k$ verziert!

Für den Mittelwert (k=1) folgt sofort:

$$\kappa_{1S} = \frac{1}{\sqrt{n}}(\mu_1 + ... + \mu_n) = \sqrt{n}\bar{\mu}.$$

Dabei ist $\bar{\mu}$ das arithmetische Mittel der Einzelmittelwerte.

Für die nächste Kumulante folgt

$$\kappa_{2S} = \left(\frac{1}{\sqrt{n}}\right)^2 (\sigma^2 + \sigma^2 + ... + \sigma^2) = \sigma^2.$$

D.h. S hat dieselbe Varianz wie alle Einzelgrößen! Für die nächste Kumulante folgt

$$\kappa_{3S} = \left(\frac{1}{\sqrt{n}}\right)^3 (\kappa_3 + .. + \kappa_3) = \frac{1}{\sqrt{n}}\kappa_3$$

Und das geht mit n nach Null. Für die höheren Kumulanten gilt das ebenso, ja sie gehen zunehmend stärker nach Null.

(3.3.3) **Ergebnis**:

Wir sehen: Mit zunehmendem n nähert sich die charakteristische Funktion von S der charakteristischen Funktion einer Normalverteilung mit Mittelwert $\sqrt{n} \cdot \bar{\mu}$ und Varianz σ^2 . **Und damit nähert sich die Verteilung selbst der entsprechenden Normalverteilung!**

(3.3.4) Das ist eine einfache Form des zentralen Grenzwertsatzes. Bemerkenswert dabei ist, wie schnell diese Annäherung vielfach geht, d.h. daß teilweise bereits für n=3 oder n=4 eine gute Annäherung an die Normalverteilung zu beobachten ist.

(3.3.5) Wir betrachten jetzt statt S die Zufallsvariable

$$\bar{X} = \frac{1}{n}(X_1 + .. + X_n).$$

Hierfür finden wir entsprechend

$$\begin{aligned} \kappa_{1\bar{X}} &= \frac{1}{n}(\mu_1 + .. + \mu_n) = \bar{\mu} \\ \kappa_{2\bar{X}} &= \frac{1}{n}\sigma^2 \end{aligned}$$

Die höheren Kumulanten gehen erneut nach Null. Nun entspricht $\kappa_{1\bar{X}}$ aber in der Datenwelt gerade der Varianz des Datenmittels.

(3.3.6) **Die "1/$\sqrt{n}$-Regel**": Streuen die Daten mit σ, dann streut das arithmetische Mittel nur mit $\sigma/\sqrt{n}$, wenn n die Anzahl der Daten ist.

Konkret bedeutet das: Will man die Genauigkeit des Datenmittels um eine Stelle verbessern (Faktor 1/10), dann muß man die Zahl der Daten um einen Faktor 100 erhöhen!

17.3.4 Statistisch abhängige Größen. Die Kovarianz

(3.4.1) Wir haben bereits gesehen, wie der Begriff der "statistischen Unabhängigkeit" im theoretischen Bereich zu formalisieren ist. Man geht aus von einem Ereignisraum $(E \times F, \mathfrak{M}, w)$ mit Produktstruktur. Notfalls muß man entsprechend idealisieren. Dazu eine Zufallsgröße $Z : E \times F \to \mathbb{R}$ mit zweidimensionaler Verteilung $\Phi_Z(x, y)$ und eventuell einer Dichte $\varphi_Z(x, y)$. Die Frage: **Kann man mit Hilfe von Information über die eine Komponente Voraussagen über die andere machen?** Wir wissen: Sind die Komponenten unabhängig, dann geht das nicht. Jetzt soll das Problem etwas weiter verfolgt werden.

(3.4.2) Zunächst bilden wir die beiden "Komponenten" von Z, die "**Randverteilungen**" durch

$$\begin{aligned} X &= X_Z = (E, e \mapsto X(e), \mathbb{R}) \quad \text{mit } \Phi_X(x) = \int_{-\infty}^{\infty} dy \Phi_Z(x, y) \\ Y &= Y_Z = (E, f \mapsto Y(f), \mathbb{R}) \quad \text{mit } \Phi_Y(y) = \int_{-\infty}^{\infty} dx \Phi_Z(x, y) \end{aligned}$$

Das sind tatsächlich Verteilungsfunktionen. Sie entstehen durch Mittelung über die jeweils andere Komponente. Entsprechende Gleichungen gelten für eventuelle Dichten. Sind die beiden Größen statistisch unabhängig, dann gilt

$$\Phi_Z(x, y) = \Phi_X(x)\Phi_Y(y)$$

Offenbar gilt generell:

$$\langle X \rangle = \langle X \rangle_{\Phi_X} = \langle X \times 1 \rangle_{\Phi_Z} \qquad \text{und} \quad \langle Y \rangle = \langle Y \rangle_{\Phi_Y} = \langle 1 \times Y \rangle_{\Phi_Z} .$$

Ein anderer Extremfall einer Verteilung ist $Y = f \circ X$. D.h. Y ist Weiterverarbeitung von X im besprochenen Sinn. Also $Z = X \times f(X)$. Dann gilt für die Dichten

$$\varphi_Z(x, y) = \varphi_X(x)\delta(y - f(x))$$

(3.4.3) Beachten Sie, daß man dann die Verteilung und Dichte von Y mit Hilfe der früher diskutierten Regel $\delta(F(x)) = \sum_i \frac{1}{|F'(x_{0i})|}\delta(x - x_{0i})$ für die δ-Funktion bestimmen kann. Dabei wird über die Nullstellen von F summiert, die einfach sein müssen.

☐ $\delta(x^2 - 4) = ...$

☐ Wie sieht das zugehörige Φ_Z aus?

(3.4.4) **Defintion**: In (naheliegender) Verallgemeinerung zur Varianz bildet man die "**Kovarianz** (vonZ)" genannte Größe:

$$\begin{aligned} Cov(X,Y) &= \langle (X - \langle X\rangle)\cdot(Y - \langle Y\rangle)\rangle_Z \\ &= \int_{E\times F} dxdy\varphi_Z(x,y)(x - \langle X\rangle)(y - \langle Y\rangle) \end{aligned}$$

(3.4.5) Mit der Linearität des Erwartungswertes folgt über die Bilinearität des Zahlproduktes ((2.1.8) verallgemeinernd)

$$Cov(X,Y) = \langle XY\rangle_Z - \langle X\rangle_X \langle Y\rangle_Y .$$

Wie bei der Varianz heben sich beim bilinearen Rechnen zwei Terme fort. Es verbleibt die angegebene Differenz. Aus dieser Gleichung folgt: **Sind X und Y statistisch unabhängig, dann ist die Kovarianz Null**. Aber Achtung: Es kann vorkommen, daß X und Y nicht unabhängig sind und daß die Kovarianz trotzdem verschwindet.

(3.4.6) Der andere Extremfall ist der, daß beide Faktoren übereinstimmen oder daß Y=aX+b einfach eine lineare Transformation von X ist. Dann folgt sofort (aX+b=f$\circ X$ mit f(y)=ay+b und $\varphi_Z(x,y) = \varphi(x)\delta(y - (ax + b))$):

$$Cov(X, aX + b) = a\langle XX\rangle_X = aVar(X).$$

Ebenso wie bei den üblichen Momenten gibt es Verteilungen, für die die Kovarianz nicht existiert. D.h. das zugehörige Integral divergiert.

(3.4.7) Die Kovarianz scheint daher geeignet, ein **erstes** orientierendes Maß zur Beschreibung von Daten**abhängigkeit** zu liefern.

(3.4.8) Allerdings ist diese Größe "Kovarianz" noch nicht systembezogen. Die vorgegebenen Koordinaten gehen in die Kovarianz ein, noch nicht die systembezogenen, in den Einheiten der beiden Streuungen σ_x und σ_y gemessenen. Der Übergang zu systembezogenen Einheiten ist leicht unde ergibt die folgende Größe, die vollständig analog zur Skalarproduktbildung ist:

$$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

Dies ist der "**Korrelationskoeffizient**", eine reine Zahl. Wie bei Cauchy-Schwarz kann sie nur Werte zwischen -1 und +1 annehmen. Sind X und Y unabhängig, dann gilt r=0. Die Umkehrung ist falsch. Aus r=0 darf man nicht auf Unabhängigkeit der Daten schließen. Sind dagegen X und Y identisch oder ist Y identisch aX+b mit a$\neq$ 0, dann ist r=1 für a>0 und r=-1 für a<0. Aber auch hier ist es so, daß der zweite Faktor durch den ersten völlig festgelegt sein kann und r ist trotzdem nicht $\pm$1.

Nochmals: r liefert Information über die Abhängigkeit von Datensätzen, die aus einem gemeinsam Datensatz Z mit Randverteilungen X und Y entstanden sind. Wir werden sehen, daß r darüber Aussagen macht, wie gut sich zugehörige Datenschwärme durch eine optimale Gerade beschreiben lassen.

(3.4.9) In der Datenwelt lautet die Formel für die "Kovarianz zweier Datensätze", die der Kovarianz entspricht, wobei $\langle d\rangle$ und $\langle e\rangle$ für die "wahren" oder idealen Mittel stehen sollen, nicht aber für eine Schätzung etwa über die arithmetischen Mittel

$$\begin{aligned} Cov(d,e) &= \frac{1}{N}\sum_i (d_i - \langle d\rangle)(e_i - \langle e\rangle) \\ &= \sum_{\alpha,\beta} h_{\alpha\beta}(x_a - \langle d\rangle)(y_\beta - \langle e\rangle) \end{aligned}$$

Wegen $\langle X + Y\rangle = \langle X\rangle + \langle Y\rangle$ gilt (Eigentlich $X\times 1 + 1\times Y$ mit $\varphi_Z(x,y)$

$$\begin{aligned} & (X + Y - \langle X + Y\rangle)^2 \\ = \; & (X - \langle X\rangle)^2 + (Y - \langle Y\rangle)^2 + 2(X - \langle X\rangle)(Y - \langle Y\rangle) \end{aligned}$$

Bildet man von beiden Seiten dieser Gleichung den Erwartungswert, so folgt

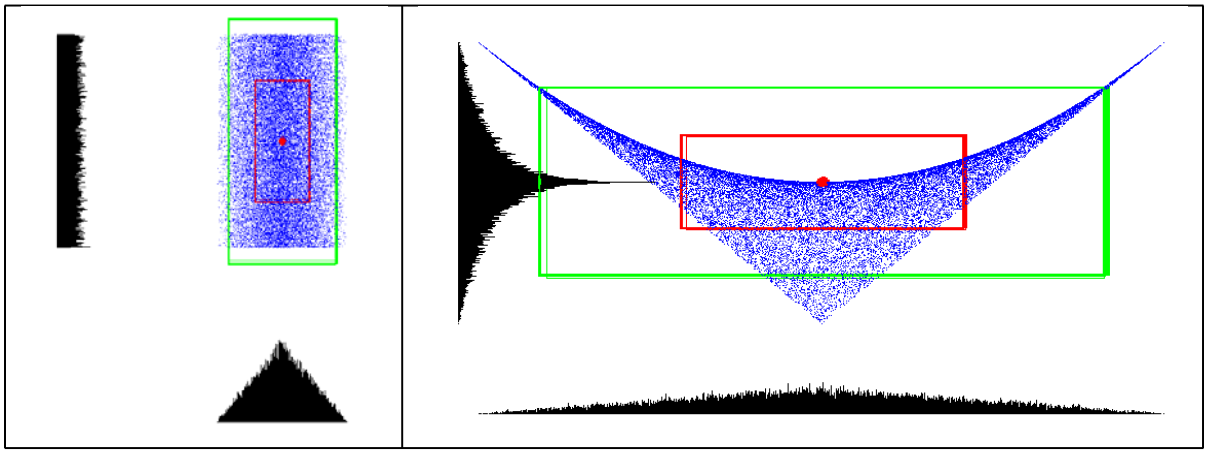
$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y).$$

Das erinnert natürlich an den Cosinussatz. Verschwindet die Kovarianz, etwa weil X und Y unabhängig sind, dann addieren sich die Varianzen:

$$Var(X+Y) = Var(X) + Var(Y).$$

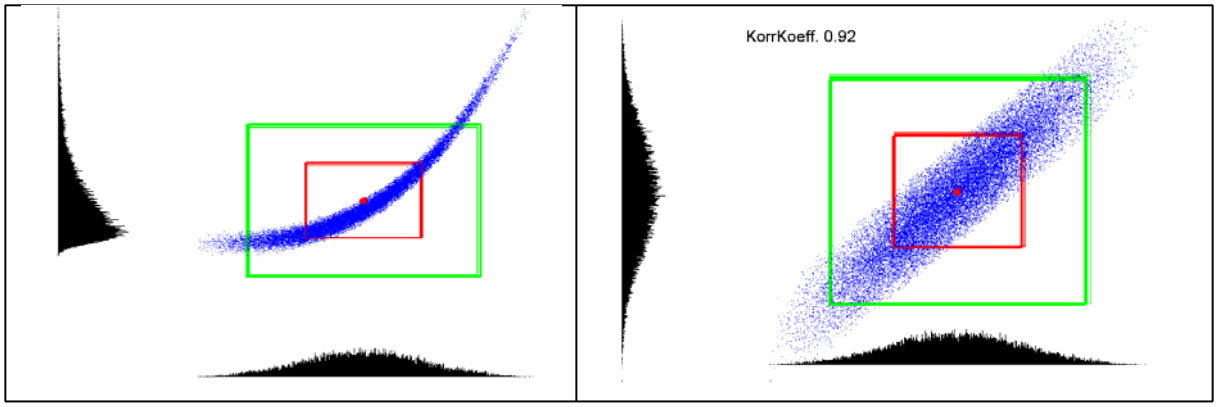
Ein wichtiges Resultat, das dem Pythagoras entspricht.

Wir geben jetzt einige Simulationen von zweidimensionalen Punkteverteilungen. Der Datensatz enthält jeweils etwa 20.000 Elemente und ist als (blauer) Punkteschwarm in der x-y-Ebene aufgetragen, so dass die Punktedichte die Grenzdichte näherungsweise veranschulicht. Dazu (schwarz) die beiden Randverteilungen (summiert über den jeweils rechts bzw darüber leigenden Streifen). Weiter werden Approximationen der ersten Beschreibungsgrößen angegeben. Jeweils nach 1000 weiteren Punkten werden diese Beschreibungsgrößen für alle Vorgängerdaten gezeichnet, so dass man die Konvergenzeigenschaften sehen kann. Als roter Punkt die Approximation des Mittelwertes. Der rote inner Kasten gibt den 1-σ-Bereich der beiden Randverteilungen, der grüne den 2-σ-Bereich.

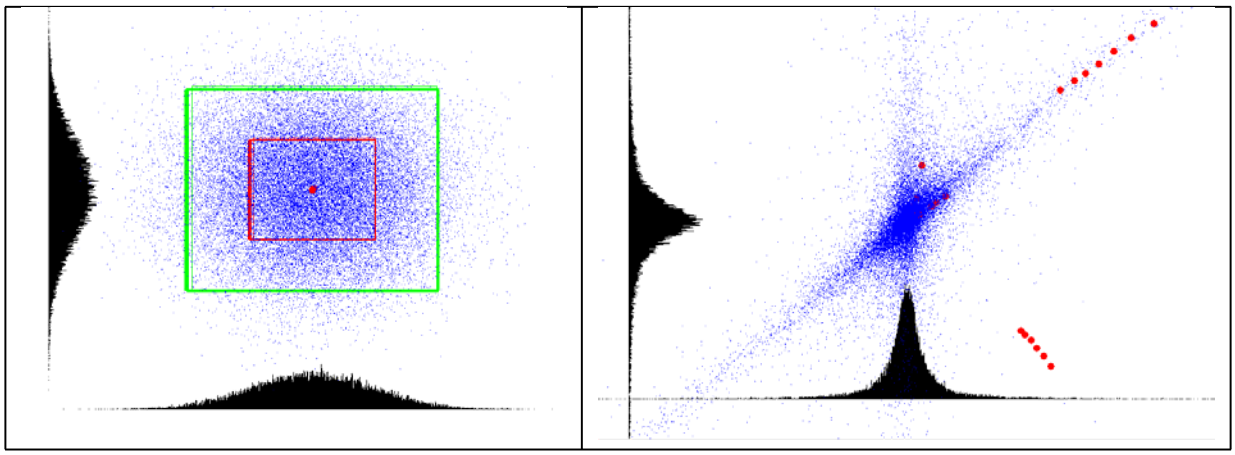


Es sei R eine gleichverteilte Zufallsgröße mit Werten zwischen -1 und 1. Dann gehört das linke Bild zu $R_1 + R_2$ und R_3. Es wird jeweils zuerst die horizontale x, dann die vertikale y-Koordinate angeben. Das rechte Bild dagegen gehört zu $R_1 + R_2$ und $R_1 \cdot R_2$.

☐ Verwenden Sie die Methode der Lagrangschen Multiplikatoren, um den oberen parabelförmigen Rand im rechten Bild zu verstehen und genauer zu bestimmen. Beide Verteilungen -auch die rechte - haben übrigens Korrelation Null!



Zwei Verteilungen mit positivem Korrelationskoeffizienten von r≥0.9. Beachten Sie, dass man mit Hilfe von r allein nicht zwischen den beiden sehr unterschiedlichen Abhängigkeiten unterscheiden kann. Beachten Sie auch unsere Faustregel zur Streuung: Etwa50% der Daten innerhalb des 1-σ-Bereiches.



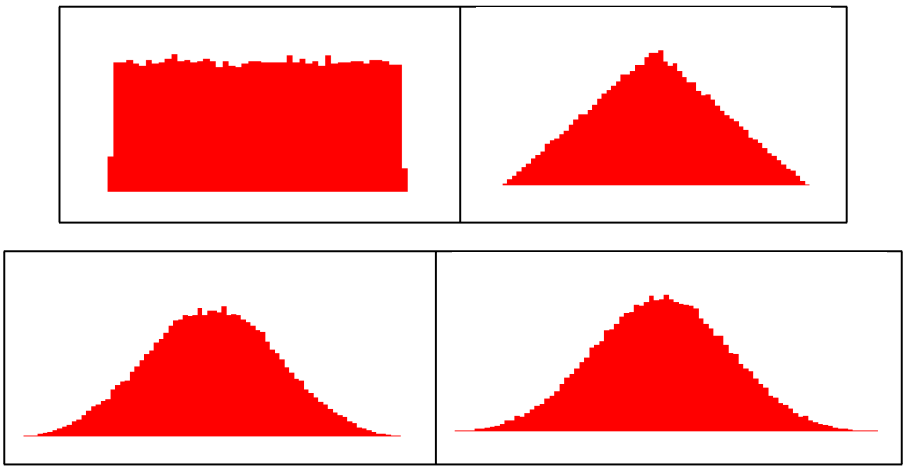
Im linken Bild sind beide Randverteilungen Normalverteilungen. Im rechten dagegen ist X=tan($\frac{\pi}{2}R$) und Y=X+tan($\frac{\pi}{R}R$) wobei R gleichverteilt zwischen -1 und +1 ist. Das ergibt eine Cauchyverteilung, von der wir gesehen haben, dass die zugehörigen Momente nicht existieren. Man sieht das am Schwerpunkt. Während dieser im linken Bild bereits nach 1000 Daten im Bild völlig stabil ist, spring er rechts herum. Immer wieder einmal kommt es zu einem so großen Datenwert, dass der Mittelwert springt. Dass gilt erst recht für die Streuung. diese ist hier von Anfang n so groß, dass sie außerhalb des Bildschirmes liegt und dort entsprechend springt. Erneut zeigt das linke Bild auch für die Streuung rasche Konvergenz.

☐ Wie wurden die Normalverteilungen im linken Bild erzeugt? Hierzu die folgende Frage:

Die beiden Zufallsgrößen R_1 und R_2 seien über J=[0,1] beide gleichverteilt. Wir bilden daraus die Vektorgröße $\vec{X} = (X_1, X_2)$. Weiter sei $\vec{f}$=(J×$J, \vec{x} \mapsto \vec{f}(\vec{x}), \mathbb{R}^2$) eine glatte injektive Abbildung, so daß eine glatte inverse Abbildung existiert. Wir bilden den neuen Zufallsvektor $\vec{Y} = \vec{f}(\vec{X})$.

- a) Bestimmen Sie die Dichte, mit der $\vec{Y}$ verteilt ist. (Substitutionsregel für $\mathbb{R}^2$ nutzen!)
- b) Wählen Sie $y_1 = \sqrt{-2\ln(x_1)}\cos(2\pi x_2)$ und $y_2 = \sqrt{-2\ln(x_1)}\sin(2\pi x_2)$. Wie sind y_1 und y_2 dann verteilt?

Bilder zum Grenzwertsatz: Es sei R_i gleichverteilt von -1 bis +1. Der zugehörige Mittelwert ist Null. Wir bilden $X_1 = R_1$, $X_2 = \frac{1}{\sqrt{2}}(R_1 + R_2)$ und $X_3 = \frac{1}{\sqrt{3}}(R_1 + R_2 + R_3)$ und analog X_4. Dann ziehen wir je eine Stichprobe von 100.000 Elementen und nehmen eine Auszählung in 100 Bins vor. Die Bilderserie zeigt die entstehenden Verteilungen. Die Streuung bleibt bei dieser Bildung in etwa konstant. Beachten Sie die Datenfluktuation besonders auch im ersten Bild.



17.3.4 **Endliche Ereignisräume**

(3.4.1) Endliche Ereignisräume bilden einen einfachen Spezialfall der allgemeinen Theorie. Wir erläutern das an einem wichtigen Beispiel. Sei E={wahr,falsch} ein zweielementiger Ereignisraum mit suggestiver Bezeichnung der beiden Elemente und mit Wahrscheinlichkeiten w(wahr)=p und w(falsch)=q und p+q=1. Dazu gibt es eine ausgezeichnete Zufallsgröße B mit B(wahr)=1 und B(falsch)=0. Dies B (B für Bernoulli) enthält die gesamte Infomation für den theoretischen Bereich. Die zugehörige Verteilungsfunktion ist ein Stufenfunktion. Die Dichte besteht aus zwei δ-Funktionen:

$$\sigma_B(x) = p\delta(x-1) + q\delta(x)$$

(3.4.2) Die charakteristische Funktion und deren Logarithmus folgen sofort zu

$$\begin{aligned} G_B(k) &= pe^{ik} + q \\ H_B(k) &= \ln(pe^{ik} + q) = \ln(1 + p(e^{ik} - 1)) \end{aligned}$$

Entwicklung gibt die ersten Kumulanten:

$$\begin{aligned} \ln(1 + p(e^{ik} - 1)) &= p \cdot (ik) + p(1-p) \cdot \frac{(ik)^2}{2!} \\ &\quad + (p-1)(2p-1)\, p \cdot \frac{(ik)^3}{3!} \\ &\quad - p(p-1)\left(6p^2 - 6p + 1\right) \cdot \frac{(ik)^4}{4!} \end{aligned}$$

(3.4.3) Natürlich lassen sich die Kumulanten und Momente auch direkt ausrechnen. Zu der Bernoulliverteilung gehört daher ein Mittelwert von p und eine Streuung $\sqrt{pq}$. Das ist ein wichtiges zu merkendes Resultat.

□ Wie groß ist die Schiefe?

(3.4.4) Jetzt seien $B_1, B_2, ...B_n$ n unabhängige Bernoulligrößen, die alle zu demselben p gehören sollen. Wir bilden $B = B_1 + B_2 + ... + B_n$. Da jedes B_i den Wert 0 oder 1 annehmen kann, kann B alle Werte zwischen 0 und n annehmen und höchstens diese. B hat die folgende Interpretion: Aus einer Urnenziehung erhält man ein bestimmtes Ereignis mit Wahrscheinlichkeit p. **Wie groß ist dann die Wahrscheinlichkeit, bei n Ziehungen gerade r mal dieses Ereignis vorzufinden**?

(3.4.5) Die gesuchte Verteilung für B finden wir leicht: Die charakteristische Funktion ist

$$G_B(k) = (pe^{ik} + q)^n = \sum_{r=0}^{n} \binom{n}{r} e^{ikr} p^r q^{n-r}$$

Wie sieht die Dichte aus, die diese Verteilung hat? Es folgt offensichtlich (über die inverse Fouriertransformation):

$$\varphi_X(x) = \sum_{r=0}^{n} \binom{n}{r} \delta(x-r) p^r q^{n-r}$$

Und das ergibt sofort die gewünschten Wahrscheinlichkeiten für die (n+1) diskreten Werte von r.

(3.4.6) Wie steht es mit dem Mittelwert und der Varianz? Wegen

$$\ln G_B(k) = \ln(pe^{ik} + q)^n = n \ln(pe^{ik} + q)$$

folgt mit den Resultaten für die Bernoulliverteilung

$$E(B) = \langle B \rangle = np \quad \text{und} \quad Var(B) = npq.$$

Das liefert uns die folgende nützliche Anwendung:

17.3.4a **Die Fluktuationsschätzung**

(3.4.7) Zunächst die **Fragestellung**: Angenommen ein bestimmter Typ von Datensatz kann nur N diskrete Werte annehmen. Ein Beispiel ist eine Auszählung in N Klassen. Dann hat man eine absolute Häufigkeit von H_r für die r-te Klasse. Die relative Häufigkeit ist $h_r = H_r/n$, wenn n die Gesamtzahl der Daten ist. **Wie genau ist dieser Wert, der einem neben n zur Verfügung steht?**

Wenn man die Datennahme wiederholt, erhält man ja einen etwas anderen Wert. Erst bei sehr großen Datensätzen ist eine Konvergenz gegen den wahren Wert zu erwarten. Dazu bilden wir folgendes Modell:

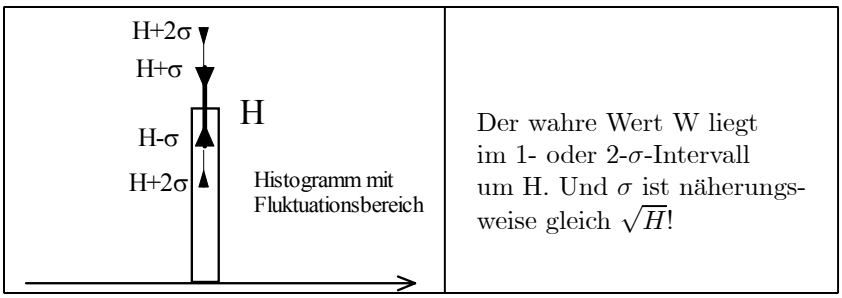
Beim Auszählen fällt ein Datenelement mit einer Wahrscheinlichkeit p in die r-te Klasse und mit einer Wahrscheinlichkeit q=1-p in eine der anderen Klassen. Sind die Datenelemente unabhängig, dann ergibt das gerade eine Binomialverteilung (n=Datenzahl, p Wahrscheinlichkeit für die Klasse). Die Streuung dieser Verteilung (für die absoluten Häufigkeiten) ist $\sigma = \sqrt{npq}$. Und das ist ein grobes Maß der auftretenden Datenfluktuation. Für die Größe der Fluktuationen der relativen Häufigkeiten ist dann $\frac{1}{n}\sigma = \sqrt{\frac{pq}{n}}$ zu erwarten. Natürlich kennt man p nicht exakt, aber man hat für p eine Schätzung in Form von $h_r = H_r/n$. Meist ist p auch noch sehr klein. Dann kann man q durch 1 ersetzen und erhält: $\sigma \approx \sqrt{n\frac{H_r}{n}1} = \sqrt{H_r}$. Insbesondere geht n in diese (grobe) Schätzung nicht mehr ein.

D.h. hat man 100 Elemente in einer Klasse, so ist eine typische statistische Unsicherheit von ±10 zu erwarten. Will man sicher gehen sogar eher von ±20.

(3.4.8) **Die Fluktuationsschätzung**: Von n Daten eines Datensatzes sollen H_r in die r-te Klasse fallen. Die theoretische Wahrscheinlichkeit dafür, daß ein Datenelement in die r-te Klasse fällt, sei p, der wahre Erwartungswert für diese Klasse k_r.. Dann streuen die Daten bei n-facher Datennahme mit $\sigma = \sqrt{npq}$ um den Wert k_r. Für die auftretenden Größen liefert der konkrete Datensatz Schätzwerte: $h_r = H_r/n$ für k_r und ebenso für p. Ist p nahe bei Null, so darf man q=1 setzen und erhält: Der "wahre" Datenzahlwert nk_r für die r-te Klasse sollte in etwa 50% der Fälle im Intervall

$$H_r - \sqrt{H_r} \leq nk_r = K_r \leq H_r + \sqrt{H_r}$$

liegen. Division durch n gibt die Schätzung für die relativen Häufigkeiten



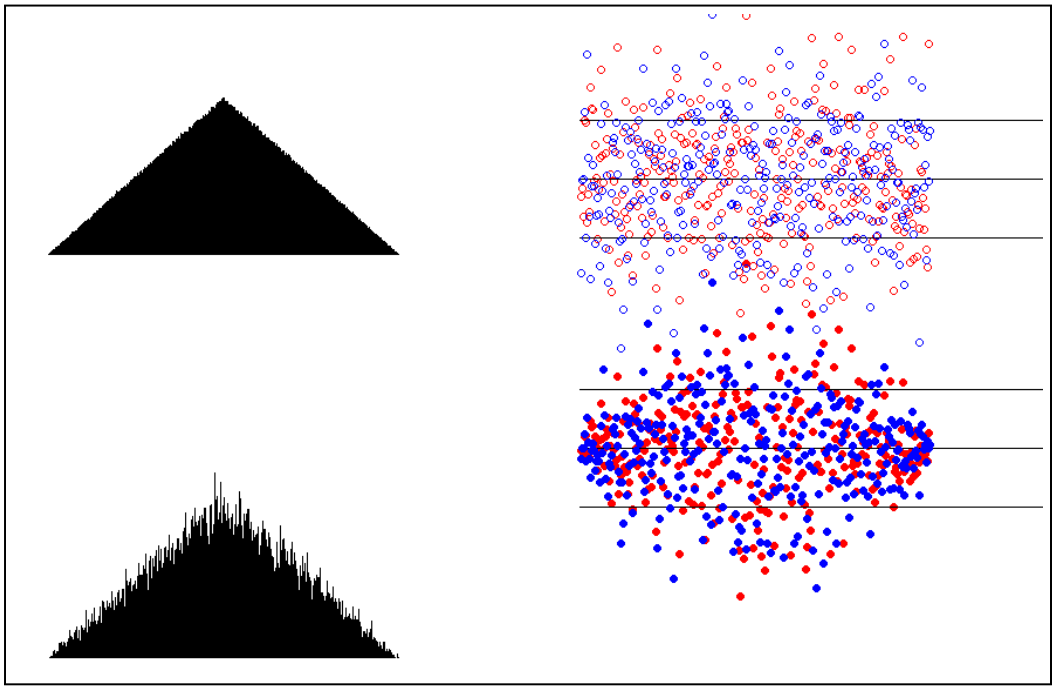
Testbeispiel für den Computer: X und Y gleichverteilt.Dann ist X+Y dreiecksverteilt mit p_r. Die jeweils ereichten Binwerte seien H_r , die vorhergesagten np_r. Wir bilden

$$\frac{H_r - np_r}{\sqrt{np_rq_r}} = \varepsilon_r \quad \text{oder} \qquad \frac{H_r - np_r}{\sqrt{np_r}} = \varepsilon_r$$

Diese Größe sollte unabhängig von der tatsächlichen Größe von H_r im Bereich von ±1 bzw. ±2 liegen.

Illustration der Fluktuation: Wir betrachten eine Variable X mit Dreiecksverteilung. $X=R_1+R_2$ wie üblich. Dazu ziehen wir eine Stichprobe von einmal 20.000 Daten (links unten) und eine von 2.000.000 Daten (links oben) mit jeweils 300 bins. Aufgetragen sind die relativen Häufigkeiten. Die Fluktuationen sollten um einen Faktor $10=\sqrt{100}$ geringer werden. Rechts unten sind die relativen Fluktuationen aufgetragen $\frac{(Ni-w_iN)}{N_i}$

über dem zugehörigen bin. (N_i Datenzahl im i-ten bin, $w_i N$ zugehörige ideale Vorhersage). Rot für den ersten und blau für den zweiten Datensatz. Die blauen Daten sind um den Faktor 10 vergrößert. Man findet die vorhergesagte Deckung. Weiter sieht man deutlich, dass die Fluktuationen zu den Rändern hin abnehmen. Rechts oben - offene Kreise - dagegen ist jeweils durch $\sqrt{Nw_i}$ geteilt. Alle Abweichungen sind jetzt von der gleichen Größenordnung und diese stimmt mit dem vorhergesagen 1-σ-Niveaus überein.



17.4 **Datensätze (II)**

17.4.1 **Die n-1 Regel für die Streuungsschätzung**

(4.1.1) Welche Konsequenzen ergeben sich aus den bisherigen Überlegungen für die Analyse der Datensätze? Nun, ein erstes Problem ist die "Schätzung von Beschreibungsgrößen". D.h.: Man hat einen Datensatz, von dem man weiß oder glaubt, daß er zu einer Stichprobe für eine Zufallsgröße X des theoretischen Bereichs (E,$\mathfrak{M}, w$) gehört. Man möchte jetzt mit Hilfe der Daten etwa die Erwartungswerte E(X) und Var(X) vorhersagen. Wie geht man vor?

(4.1.2) Im Falle des Mittelwertes wird man das arithmetische Mittel des Datensatzes nehmen. Das ist sinnvoll, solange das Moment E(X) existiert. Was aber ist mit der Varianz? Dort hat man das Problem, daß man in der Regel den **wahren** Erwartungswert $\mu = \langle X \rangle$ von X nicht kennt. Wir haben es mit n unabhängigen Wiederholungen X_i der stochastischen Variablen X zu tun. Man kennt nur den Schätzwert $\bar{d}$ in Form des arithmetischen Datenmittels. Dieses gehört zur stochastischenVariablen $\bar{X} = \frac{1}{n}(X_1 + ... + X_n)$. Jetzt rechnen wie folgt

$$X_i - \bar{X} = (X_i - \mu) + (\mu - \bar{X}).$$

Quadrieren und Bilden des Erwartungswertes (bezüglich der Verteilung von X) gibt:

$$E((X_i - \bar{X})^2) = E((X_i - \mu)^2) + E((\bar{X} - \mu)^2) - 2Cov(X_i, \bar{X})$$

Die (gemeinsame) Varianz aller X_i ist σ^2. Wir haben bereits gesehen, daß $E((\bar{X} - \mu)^2) = \frac{1}{n}\sigma^2$ gilt. Für die Kovarianz schließlich folgt über die Unabhängigkeit von X_i und X_j für i≠j:

$$\begin{aligned} -2Cov(X - \mu, \mu - \bar{X}) &= -\frac{2}{n}\sum_{j=1}^{n} Cov(X_i, X_j) \\ &= -\frac{2}{n}Cov(X_i, X_i) = -\frac{2}{n}\sigma^2 \end{aligned}$$

Insgeamt:

$$E((X_i - \bar{X})^2) = \sigma^2 + \frac{1}{n}\sigma^2 - \frac{2}{n}\sigma^2 = \frac{n-1}{n}\sigma^2.$$

(4.1.3) Das das gesuchte Resultat:

Die "n-1-Regel": Die aus den Daten bildbare Größe

$$\frac{1}{n}\sum(d_i - \bar{d})^2,$$

die zu der Zufallsgröße X-$\bar{X}$ gehört, nähert sich nicht der gesuchtenVarianz σ^2, sondern nur $\frac{n-1}{n}\sigma^2$. Zur Schätzung der Varianz muß man daher die Größe

$$s^2 = \frac{1}{n-1}\sum(d_i - \bar{d})^2 \quad \text{verwenden.}$$

(4.1.4) Wir sehen: Mit n=1 ist es vernünftigerweise nicht möglich, eine Schätzung für die Streuung zu gewinnen. Dividiert mn durch n, so wird die Schätzung zu klein. Das liegt daran, daß das Datenmittel systematisch näher an den Einzeldaten liegt, als das wahre Mittel. Für kleine n, etwa n=2,3,4 ist der Effekt beträchtlich.

□ Wählen Sie eine Gleichverteilung zwischen 0 und 1 und bestimmen Sie Mittelwert und Streuung. Nehmen Sie dann 100 Computerdatensätze mit 2 (bzw. 3) Elementen

17.4.2 **Lineare Regression**

(4.2.1) Gegeben ein Datensatz $i \mapsto (x_i, y_i)$, also anschaulich ein Datenschwarm der Ebene. Kann man diese Punkte durch eine Gerade optimal approximieren? Das ist in gewissen Fällen naheliegend, wenn man die graphische Darstellung inspiziert. Wir wollen eine solche optimale Gerade jetzt rechnereisch bestimmen.

(4.2.2) Dazu verwenden wir folgende Idee: Wir bilden die Zufallsgröße mit 2 äußeren Parametern Z_{mb} =Y-mX-b. Dabei seien X und Y die beiden Randgrößen, die zum Datensatz gehören. Ist jetzt Y=mX+b eine Verarbeitung des Datensatzes e, also Y(e)=mX(e)+b, dann ist Z=0, hat insbesondere Erwartungswert und Varianz Null. Und der Datenschwarm liegt exakt auf der Geraden y=mx+b in der Ebene. Liegt der Schwarm nicht auf der Geraden, dann sollte die Abweichungsgröße $\left\langle (Y - mX - b)^2 \right\rangle$ zumindest möglichst klein werden. Also:. **Für welche Wahl von m und b wird dieser Erwartungswert möglichst klein**

(4.2.3) Die rechnerische Ausführung:

$$\begin{aligned}(Y - mX - b)^2 &= Y^2 - 2Y(mX + b) + (mX + b)^2 \\ & \quad Y^2 - 2mYX - 2bY + m^2X^2 + 2mbX + b^2\end{aligned}$$

Bildung des Erwartungswertes:

$$\begin{aligned}& \left\langle (Y - mX - b)^2 \right\rangle \\ = \; & \left\langle Y^2 \right\rangle - 2m \left\langle YX \right\rangle - 2b \left\langle Y \right\rangle + m^2 \left\langle X^2 \right\rangle + 2mb \left\langle X \right\rangle + b^2\end{aligned}$$

Das ist jetzt ein Skalarfeld in (m,b). Ein Minimum muß existieren, da der Erwartungswert eines Quadrates vorliegt. Für einen Extremwert muß der Gradient verschwinden. Wir setzen die beiden partiellen Ableitungen (nach m und b) Null und finden:

$$\begin{aligned}-2 \left\langle YX \right\rangle + 2m \left\langle X^2 \right\rangle + 2b \left\langle X \right\rangle &= 0 \\ -2 \left\langle Y \right\rangle + 2m \left\langle X \right\rangle + 2b &= 0.\end{aligned}$$

Die zweite Gleichung besagt, daß die optimale Gerade durch den Schwerpunkt des Schwarmes $(\langle X \rangle, \langle Y \rangle)$ gehen muß. Dann folgt für die erste Gleichung

$$\begin{aligned}-2 \left\langle YX \right\rangle + 2m \left\langle X^2 \right\rangle + 2(\left\langle Y \right\rangle - m \left\langle X \right\rangle) \left\langle X \right\rangle &= 0 \\ m(\left\langle X^2 \right\rangle - \left\langle X \right\rangle^2) &= \left\langle YX \right\rangle - \left\langle Y \right\rangle \left\langle X \right\rangle\end{aligned}$$

Das gibt für m

$$m = \frac{\langle YX\rangle - \langle Y\rangle\langle X\rangle}{\langle X^2\rangle - \langle X\rangle^2} = \frac{Kov(X,Y)}{Var(X)} =$$
$$= \left(\frac{Kov(X,Y)}{\sqrt{VarX}\sqrt{Vary}}\right)\sqrt{\frac{Vary}{VarX}} = r\frac{\sigma_Y}{\sigma_X}.$$

Damit ist die (im beschriebnenen Sinne) optimale Gerade bestimmt.

(4.2.4) **Lineare Regression:** Zu einer zweidimensionale Zufallsgröße Z mit Projektionen X und Y gehört die optimal approximierende Geradengröße $Y_g = mX + b$. Dabei ist

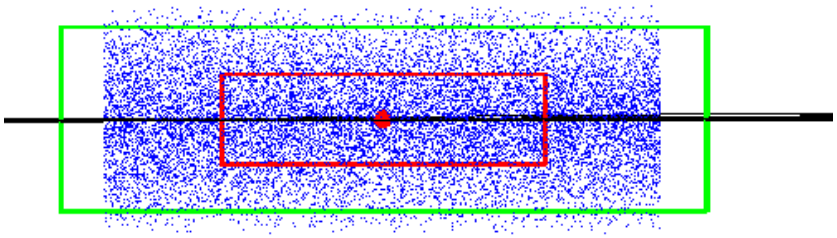
$$m = r\frac{\sigma_Y}{\sigma_X} \quad \text{und} \quad \langle Y\rangle = m\langle x\rangle + b.$$

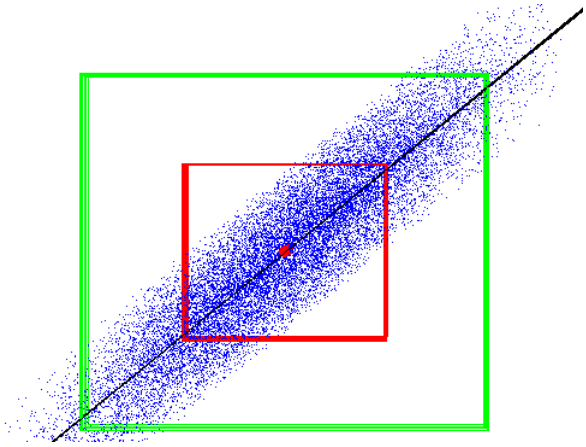
Hat man einen zugehörigen Datensatz i$\mapsto$ $d_i = (x_i, y_i)$, so sind die zugehörigen Schätzwerte einzusetzen.

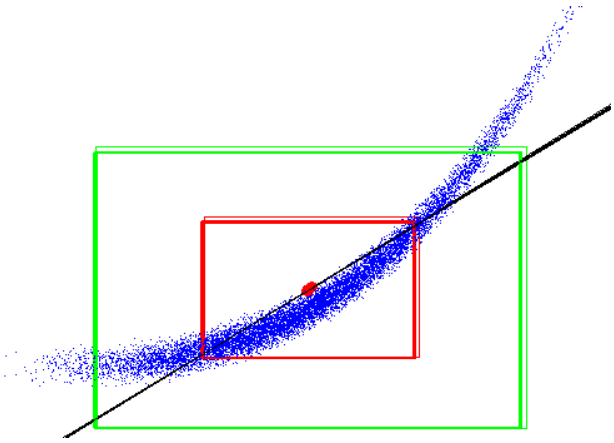
☐ Was ergibt die oben eingeführte Größe Q=$\langle (Y - mX - b)^2\rangle$ für die optimale Wahl von m und b? Zeigen Sie, daß dann Q=(1-r^2)σ_Y^2 folgt und interptretieren Sie das in dem Sinne, daß r die Anpaßbarkeit der Daten an eine Gerade beschreibt.

☐ Was ist, wenn r=0 gilt? Was besagt $\sigma_X = 0$?

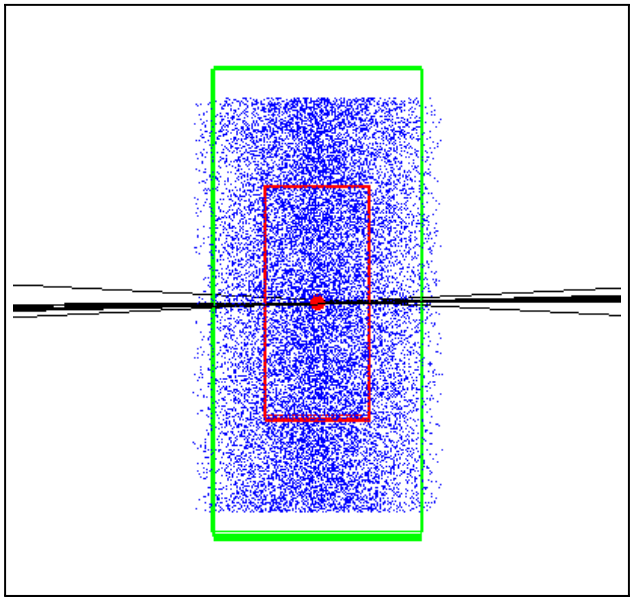
Die nachfolgenden Bilder zeigen für einige Fälle die über einen zugehörigen Datensatz bestimmte Regressionsgerade. Erfolgt die Zeichnung in 1000er Schritten. Man sieht, dass auch hier die Schätzung ziemlich stabil ist.







Das sieht jeweils vernünftig aus. Das nächste Beispiel zeigt, dass die Methode auch für r$\neq$ 0 problematisch sein kann! Das liegt hier an der ungleichen Behandlung der beiden Achsen. Die benutzte Gleichungsform y=mx+b beschreibt keine Parallen zur y-Achse!



17.4.3 **Modellanpassung / Die Chi-Quadrat-Verteilung**

(4.3.1) Wir betrachten die folgende Situation: Wir haben eine Auszählung H_r eines (skalaren) Datensatzes und eine theoretische Vorhersage w_r für die zugehörigen Wahrscheinlichkeiten. Ist n die Datenzahl, dann ist nw_r die theoretische Vorhersage für die (absolute) Häufigkeit H_r. Beide Größen unterscheiden sich voneinander. Bisher hat uns der Unterschied $H_r - nw_r$ für ein **festes r** interessiert. Nach dem Fluktuationssatz gilt $H_r - nw_r \approx \pm\sqrt{nw_r}$. Oder:

$$\frac{H_r - nw_r}{\sqrt{nw_r}} \approx \pm 1 \quad \text{für jedes r.}$$

(4.3.2) Jetzt fragen wir danach, ob die Gesamtheit aller Unterschiede mit der theoretischen Vorhersage verträglich ist. Dazu bilden wir die folgende Größe

$$V_{n-1}^2 = \sum_r \frac{(H_r - nw_r)^2}{nw_r}.$$

Jeder Summand trägt infolge des Quadrierens einen Wert von etwa +1 bei. **Zu kleine oder zu große Werte werden selten sein.** Da $\sum H_r = n$ gelten muß, sind nur n-1 der Fluktuationen unabhängig.

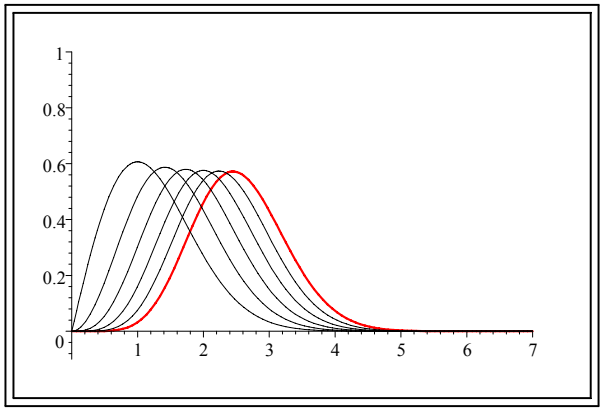
(4.3.3) Man kann erneut die Verteilungsfunktion von V^2 bestimmen, wenn man annimmt, daß die H_r erzeugenden Daten zufällig von den theoretischen Werten abweichen. Diese Verteilungen heißen $\chi^2 - Verteilung$ (mit n-1) Freiheitsgraden. Die Verteilungen sind in Form ihrer Verteilungsfunktion tabelliert, speziell für solche Werte von V^2, die sehr selten auftreten: Etwa in weniger als 1% oder 5% aller Fälle. Liefern die Daten einen Wert in diesem (seltenen) Bereich, wird man an der Vorhersage zweifeln. Liefern sie einen Wert im zulässigen Bereich, wird man die Daten als vereinbar mit der Vorhersage interpretieren.

Beachten Sie, daß man **sowohl zu große als auch zu kleine Werte von V^2 auszuschließen hat.**

Wir geben die zugehörigen Verteilungsdichten noch an. Mit $\chi^2 = V_n^2$ folgt

$$p_n(\chi)d\chi = d\chi \cdot 2\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\chi^{n-1}e^{-\frac{\chi^2}{2}} = (2\chi d\chi)\,\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\left(\chi^2\right)^{\frac{n-2}{2}}e^{-\frac{1}{2}\chi^2}$$

Beachten Sie, dass $d\chi^2 = 2\chi d\chi$ gilt. Als Funktion von χ erhält man für n=2,3,..,7 die folgenden Kurven, wobei n=7 rot gezeichnet ist:



17.4.4 Fehlerfortpflanzung

(4.4.1) Das zugehörige Szenenbild sieht wie folgt aus:

- Man hat einen Datensatz d mit Mittelwert m und Streuung σ. Dabei ist m typisch Schätzwert für eine physikalische Größe M. Sagen wir Mittelwert von Messungen einer Schwingungsdauer, deren "wahrer Wert" gesucht ist. m ist Schätzwert dieser Größe. U.U. auch ein anderer Schätzwert als das arithmetische Mittel.

- Die Verteilung von m (etwa bei vielfacher Meßwiederholung) ist eine andere als die für die Daten. Im Falle des Mittels streut m nach dem $1/\sqrt{n}$-Gesetz mit $\sigma_0 = \frac{\sigma}{\sqrt{n}}$. Sei σ_0 generell die Streuung von m. σ_0 ist typischerweise viel kleiner als die Streuung σ der Daten! Ja, soll so klein sein, daß Tangentenapproximation zulässig ist.

- Die zugehörige theoretische Verteilung von m nennen wir Ψ. Zugehörige Erwartungswerte werden mit $\langle..\rangle_\Psi$ bezeichnet. Insbesondere gilt $\langle x\rangle_\Psi = m$ und VarΨ=σ_0^2. Wir erwarten also, daß der wahre Wert von M im 1-σ_0- oder 2-σ_0-Intervall um m liegt. σ_0 ist ein Maß für die Unsicherheit der Vorhersage und kann statistischer oder systematischer Art sein.

- **Und jetzt interessiert für die weitere Arbeit nicht der (wahre) Wert $\mathbf{m}_w$ von M selbst, sondern ein fortgerechneter Wert f($\mathbf{m}_w$), wobei f eine glatte Funktion ist.** Typischerweise möchte man den Wert von f(m_w) vorhersagen. Dann ist f(m) zunächst einmal der richtige Schätzwert von m. Aber wie groß ist die Unsicherheit?

- Wir lösen das Problem mit der Tangentenapproximation! Sei $m_w = m + \epsilon\sigma_0$, wobei ε eine Zahl in der Größenordnung zwischen -2 und 2 ist. Es folgt:

$$f(m_w) = f(m + \epsilon\sigma_0) = f(m) + f'(m)\epsilon\sigma_0 + ...$$

  Der typische Fehler ist daher von der Größe $|f'(m)\epsilon|\,\sigma_0$.

- Meist ist es so, daß die gesuchte Größe von mehreren Variablen $m_1, ...m_k$ samt zugehörigen Meßfehlern abhängt. Dann rechnet man wie folgt

$$\begin{aligned} & f(m_1 + \epsilon_1\sigma_{10}, ..., m_k + \epsilon_k\sigma_{k0}) \\ = \quad & f(m_1, .., m_k) + \sum \varepsilon_j \frac{\partial f}{\partial m_j}(..)\sigma_{j0} + ... \end{aligned}$$

  Jetzt beschreibt die Summe über j die typische Abweichung zwischen Vorhersagewert und wahrem Wert. Die tatsächlicheGröße hängt natürlich von der zugehörigen Gesamtverteilung Ψ ab.

  (4.4.2) Wir unterscheiden **zwei Extremfälle**:

- In dem einen Fall ist es möglich, im Sinne von nicht unwahrscheinlich,daß sich die Vorzeichen der ε durch irgendeinen Mechanismus verschwören, derart daß alle Summanden dasselbe Vorzeichen erhalten. Diese Verschwörung produziert dann so etwas wie den *größtmöglichen Fehler*. In diesem ungünstigen Fall ergibt sich näherungsweise

$$\text{Maximaler Fehlerbetrag} = \sum_j \left| \frac{\partial f}{\partial m_j}(...) \right| \sigma_{0j}.$$

  In so einem Fall sind die verschiedenen ε natürlich nicht voneinander unabhängig.

- In der Regel werden die Messungen und die ε voneinander unabhängig sein. Dann müssen wir über viele unabhängige Datennahmen mitteln und erwarten für die Erwartungswerte

$$\left\langle \varepsilon_j^2 \right\rangle_\Psi = 1 \quad \text{und} \quad \left\langle \varepsilon_i \varepsilon_j \right\rangle_\Psi = 0 \quad \text{für i} \neq j.$$

  Der Fehler ergibt sich jetzt über die Varianz der f-Daten um den Mittelwert in der Ψ-Verteilung. Das gibt mit unserer Tangentenapproximation:

$$\begin{aligned} & \left\langle (f(m_1 + \varepsilon_1\sigma_{10}, ..., m_k + \varepsilon_k\sigma_{k0}) - f(m_1, ..m_k))^2 \right\rangle_\Psi \\ = & \sum_{ij} \left\langle \varepsilon_i \varepsilon_j \right\rangle_\Psi \frac{\partial f}{\partial m_i} \frac{\partial f}{\partial m_j} \sigma_{i0}\sigma_{j0} \\ = & \sum_j 1 \cdot \frac{\partial f}{\partial m_j} \frac{\partial f}{\partial m_j} \sigma_{j0}\sigma_{j0} = \sum_j \left(\frac{\partial f}{\partial m_j} \right)^2 \sigma_{j0}^2 \end{aligned}$$

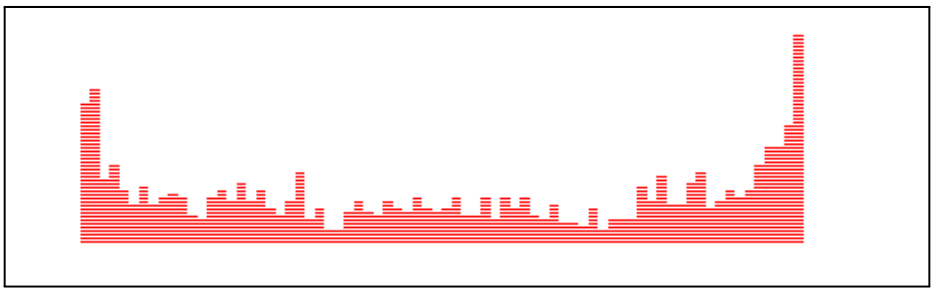
  Damit ist die zur Beschreibung des Fehlers benötigte Varianz für den zweiten Extremfall bestimmt!

(4.4.3) Das Ergebnis ist das **Gaußsche Fehlerfortpflanzungsgesetz**:

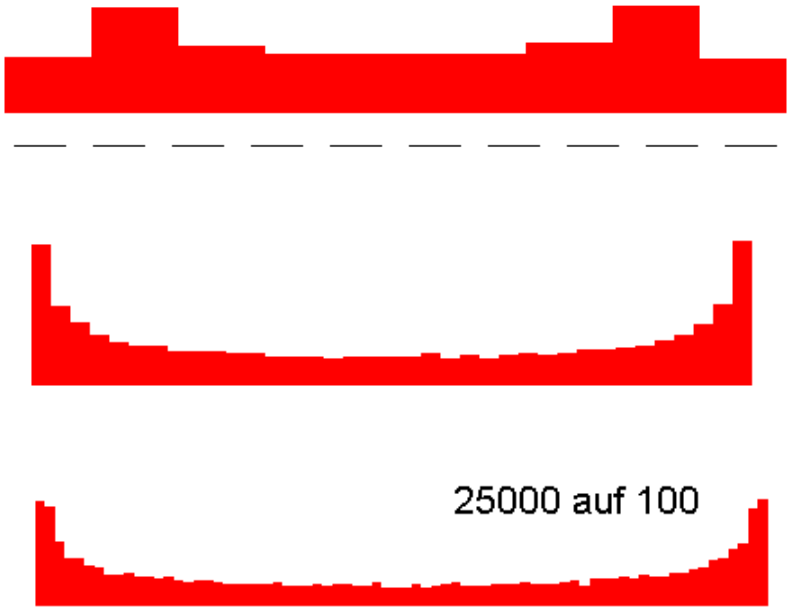
Die Größen $m_i + \varepsilon\sigma_{i0}$ seien voneinander unabhängig verteilt mit $\left\langle \varepsilon_i \varepsilon_j \right\rangle_\Psi = \delta_{ij}$. Weiter sei f:$(x_1, ...x_k) \mapsto f(x_1, ..x_k)$ ein glattes Skalarfeld. **Dann** ist der Fehler für f$(m_1, ..., m_k)$ typischerweise

$$\boxed{\sqrt{\sum_j \left(\frac{\partial f}{\partial x_j}(m_1, ..., m_k) \right)^2 \sigma_{j0}^2}}$$

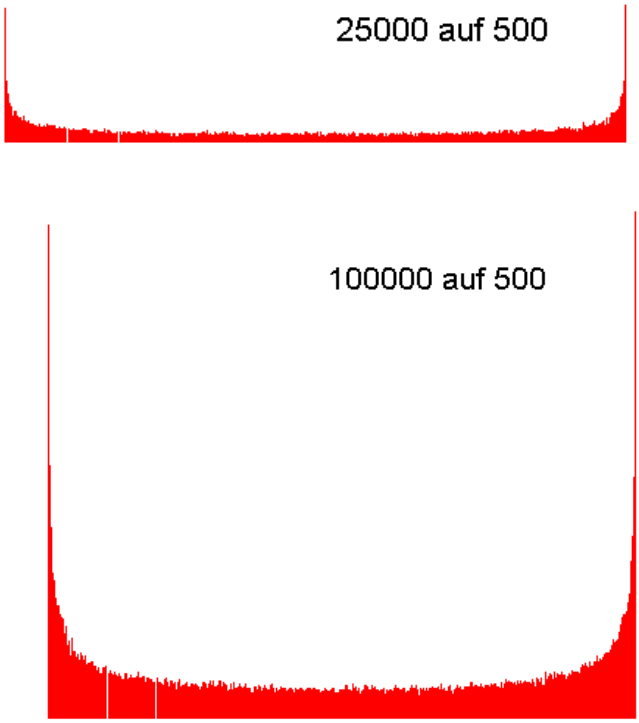
Ergänzung zu Auszählung: Sei R gleichverteilt zwischen -1 und +1. Damit bilden wir die neue Zufallsgröße X=sin(πR) deren theoretische Verteilung wir bestimmt haben. Wir wählen einen Datensatz mit 2000 Elementen und eine Zerlegung in 80 bins, die etwas über das Werteintervall hinausreichen. Dann ergibt sich beispielsweise folgendes Bild mit recht großen Datenfluktuationen:



Jetzt wählen wir Datensätz mit 25000 Elementen und nacheinander 10, 50 und 100 Häuser:



Und jetzt noch 500 Häuser, einmal mit 25000 und einmal mit 100000 Daten



Denken Sie daran, die ideale theoretische Dichte ist proportional zu $\frac{1}{\sqrt{1-x^2}}$.*Die* Konvergenz der Histogramme dagegen ist sicher nicht gleichmäßig. Für die zugehörige Verteilungsfunktion dagegen erwarten wir gleichmäßige Konvergenz.