und
$$F_b(a,b) = 2\cdot n\cdot b - 2\cdot\sum_{i=1}^{n} u_i + 2\cdot a\cdot\sum_{i=1}^{n} t_i\,.$$
Division durch zwei ergibt die beiden Gleichungen für das Verschwinden dieser partiellen Ableitungen:
$$a\cdot\sum_{i=1}^{n} t_i^2 + b\cdot\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} t_i\cdot u_i$$
und
$$a\cdot\sum_{i=1}^{n} t_i + b\cdot n = \sum_{i=1}^{n} u_i\,.$$
Um $b$ zu eliminieren, multiplizieren wir die zweite Gleichung mit $\frac{1}{n}\sum_{i=1}^{n} t_i$
$$a\cdot\frac{1}{n}\left(\sum_{i=1}^{n} t_i\right)^2 + b\cdot\sum_{i=1}^{n} t_i = \frac{1}{n}\left(\sum_{i=1}^{n} t_i\right)\cdot\left(\sum_{i=1}^{n} u_i\right)$$
und subtrahieren dies von der ersten Gleichung:
$$a\cdot\left(\sum_{i=1}^{n} t_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} t_i\right)^2\right) = \sum_{i=1}^{n} t_i\cdot u_i - \frac{1}{n}\left(\sum_{i=1}^{n} t_i\right)\cdot\left(\sum_{i=1}^{n} u_i\right),$$
d.h.
$$a = \frac{\displaystyle\sum_{i=1}^{n} t_i\cdot u_i - \frac{1}{n}\left(\sum_{i=1}^{n} t_i\right)\cdot\left(\sum_{i=1}^{n} u_i\right)}{\displaystyle\sum_{i=1}^{n} t_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} t_i\right)^2}\,.$$
Entsprechend läßt sich $a$ eliminieren, wenn wir die erste Gleichung mit $\sum t_i$ und die zweite mit $\sum t_i^2$ multiplizieren, und wir erhalten
$$b = \frac{\displaystyle\left(\sum_{i=1}^{n} t_i\right)\cdot\left(\sum_{i=1}^{n} t_i\cdot u_i\right) - \left(\sum_{i=1}^{n} t_i^2\right)\cdot\left(\sum_{i=1}^{n} u_i\right)}{\displaystyle\left(\sum_{i=1}^{n} t_i\right)^2 - n\sum_{i=1}^{n} t_i^2}\,.$$



Als Beispiel betrachten wir den Zusammenhang zwischen dem Grad der Korruption in einem Land und dessen Bruttosozialprodukt pro Einwohner. Jedes Jahr veröffentlicht die Organisation *Transparancy International* ihren *corruption perceptions index (CPI),* in dem jedem Land eine Zahl zwischen null und zehn zugeordnet wird, je nachdem, wie stark Geschäftsleute, Risikospezialisten und die Bevölkerung die Korruption im betreffenden Land einschätzen: Ein Index von zehn bedeutet, daß es praktische keine Korruption gibt, während bei null nichts ohne Korruption geht. Der neueste Index stammt von Ende 1999 und ist unter

`http://www.transparency.org/documents/cpi/index.html`

zu finden.

Dieser Index wir hier verglichen mit dem Bruttosozialprodukt per Einwohner, wie es die Weltbank für 1997 festgestellt hat. Dies sind die neuesten verfügbaren Zahlen; sie sind beispielsweise auf dem Server des Statistischen Bundesamtes unter

`http://www.statistik-bund.de/basis/d/ausl/ausllae5.htm`

zu finden. In der folgenden Tabelle sind alle Staaten aufgelistet, für die beide Zahlen vorliegen; das Bruttosozialprodukt pro Einwohner in US-$ ist kursiv gedruckt, der Korruptionsindex fett:

| | | |
|---|---|---|
| **Ägypten** | *1200* | **3,3** |
| **Albanien** | *760* | **2,3** |
| **Argentinien** | *8950* | **3,0** |
| **Armenien** | *560* | **2,5** |
| **Aserbaidschan** | *510* | **1,7** |
| **Australien** | *20650* | **8,7** |
| **Belgien** | *26730* | **5,3** |
| **Bolivien** | *970* | **2,5** |
| **Botsuana** | *3310* | **6,1** |
| **Brasilien** | *4790* | **4,1** |
| **Chile** | *4820* | **6,9** |
| **China (ohne Hongkong & Taiwan)** | *860* | **3,4** |
| **Costa Rica** | *2680* | **5,1** |
| **Côte d'Ivoire** | *710* | **2,6** |

| | | |
|---|---|---|
| **Dänemark** | *34890* | **10,0** |
| **Deutschland** | *28280* | **8,0** |
| **Ecuador** | *1570* | **2,4** |
| **El Salvador** | *1810* | **3,9** |
| **Estland** | *3360* | **5,7** |
| **Finnland** | *24790* | **9,8** |
| **Frankreich** | *26300* | **6,6** |
| **Ghana** | *390* | **3,3** |
| **Griechenland** | *11640* | **4,9** |
| **Großbritannien und Nordirland** | *20870* | **8,6** |
| **Guatemala** | *1580* | **3,2** |
| **Honduras** | *740* | **1,8** |
| **Hongkong** | *25200* | **7,7** |
| **Indien** | *370* | **2,9** |
| **Indonesien** | *1110* | **1,7** |
| **Irland** | *17790* | **7,7** |
| **Island** | *26580* | **9,2** |
| **Israel** | *16180* | **6,8** |
| **Italien** | *20170* | **4,7** |
| **Jamaika** | *1550* | **3,8** |
| **Japan** | *38160* | **6,0** |
| **Jordanien** | *1520* | **4,4** |
| **Kamerun** | *620* | **1,5** |
| **Kanada** | *19640* | **9,2** |
| **Kasachstan** | *1350* | **2,3** |
| **Kenia** | *340* | **2,0** |
| **Kirgisistan** | *480* | **2,2** |
| **Kolumbien** | *2180* | **2,9** |
| **Korea, Republik** | *10550* | **3,8** |
| **Kroatien** | *4060* | **2,7** |
| **Lettland** | *2430* | **3,4** |
| **Litauen** | *2260* | **3,8** |
| **Luxemburg** *(BSP von 1996)* | *45360* | **8,8** |
| **Malawi** | *210* | **4,1** |
| **Malaysia** | *4530* | **5,1** |
| **Marokko** | *1260* | **4,1** |



| | | |
|---|---|---|
| **Mauritius** | *3870* | **4,8** |
| **Mazedonien** | *1100* | **3,3** |
| **Mexiko** | *3700* | **3,4** |
| **Mongolei** | *390* | **4,3** |
| **Mosambik** | *140* | **3,5** |
| **Namibia** | *2110* | **5,3** |
| **Neuseeland** | *15830* | **9,4** |
| **Nicaragua** | *410* | **3,1** |
| **Niederlande** | *25830* | **9,0** |
| **Nigeria** | *280* | **1,6** |
| **Norwegen** | *36100* | **8,9** |
| **Österreich** | *27920* | **7,6** |
| **Pakistan** | *500* | **2,2** |
| **Paraguay** | *2000* | **2,0** |
| **Peru** | *2610* | **4,5** |
| **Philippinen** | *1200* | **3,6** |
| **Portugal** | *11010* | **6,7** |
| **Sambia** | *370* | **3,5** |
| **Schweden** | *26210* | **9,4** |
| **Schweiz** | *43060* | **8,9** |
| **Senegal** | *540* | **3,4** |
| **Singapur** | *32810* | **9,1** |
| **Slowenien** | *9840* | **6,0** |
| **Spanien** | *14490* | **6,6** |
| **Südafrika** | *3210* | **5,0** |
| **Tansania, Vereinigte Republik** | *210* | **1,9** |
| **Thailand** | *2740* | **3,2** |
| **Tunesien** | *2110* | **5,0** |
| **Türkei** | *3130* | **3,6** |
| **Uganda** | *330* | **2,2** |
| **Uruguay** | *6130* | **4,4** |
| **Usbekistan** | *1020* | **1,8** |
| **Venezuela** | *3480* | **2,0** |
| **Vereinigte Staaten** | *29080* | **7,5** |
| **Vietnam** | *310* | **2,6** |

Der erste Augenschein zeigt, daß korruptionsärmere Länder oftmals reicher sind: Das praktisch korruptionsfreie Dänemark hat ein Bruttosozialprodukt von 34 890 \$ pro Einwohner, das deutlich korruptere Deutschland nur 28 280 \$ und ein stark korruptes Land wie Aserbaidschan nur 510 \$. Es lohnt sich also, eine Ausgleichsgerade zu berechnen; die obigen Formeln führen auf

$$\mathrm{CPI} = 3{,}177461442 + 0{,}1692187403 \cdot 10^{-3} \cdot \mathrm{BSP}\,.$$

Abbildung 64 zeigt diese Gerade zusammen mit den Datenpunkten.

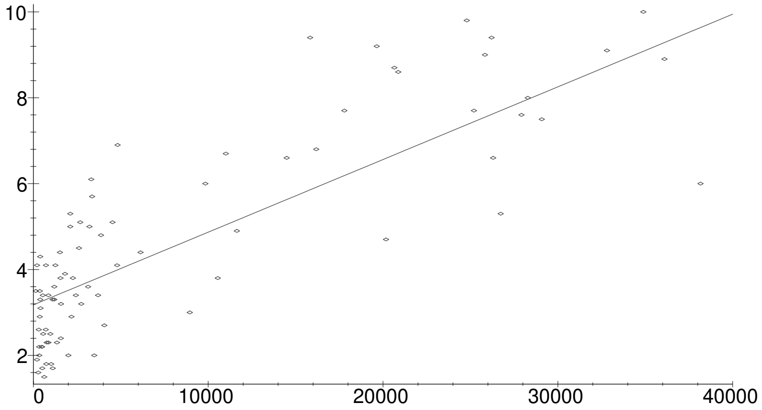


Abb. 64: Korruption über Bruttosozialprodukt für 86 Staaten

Natürlich liegen nicht alle Datenpunkte auf der Ausgleichsgeraden: Beispielsweise sind Japan, Belgien und Italien deutlich korrupter, als ihre Wirtschaftskraft vermuten läßt, wohingegen Dänemark, Finnland und Neuseeland deutlich ärmer sind, als man aufgrund der geringen Korruption erwarten sollte. Trotzdem beschreibt die Ausgleichsgerade doch einigermaßen genau einen Trend, und zumindest Deutschland liegt ziemlich genau auf der Gerade. (Die Daten stammen allerdings aus einer Zeit, als das Wort *Bimbes* noch nicht so bekannt war wie heute.)

Ähnlich sieht es aus, wenn man das Klausurergebnis dieses Semesters mit den bearbeiteten Übungsaufgaben in Verbindung bringt: Hier ergibt



sich die Gerade

$$\text{Klausurpunkte} = 1{,}097325998 + 0{,}5192298020 \cdot \text{Übungsprozente}\,,$$

die in Abbildung 65 zusammen mit den Datenpunkten eingezeichnet ist.

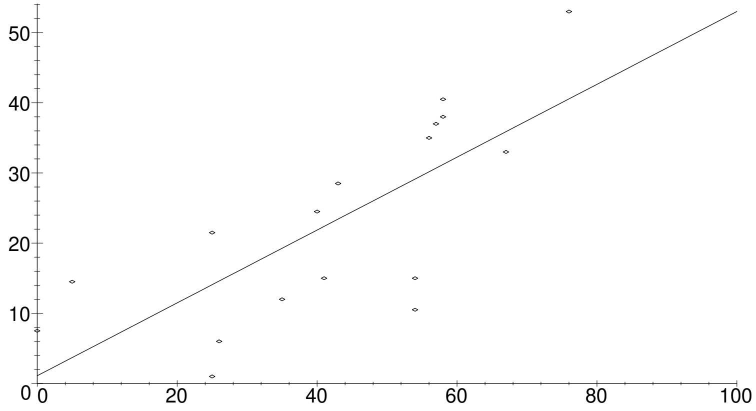


Abb. 65: Zusammenhang zwischen Klausur und Übungen in diesem Semester

Wieder sieht man, daß im großen und ganzen höhere Punktezahlen bei den Übungsaufgaben einem besseren Klausurergebnis entsprechen, aber es gibt auch recht deutliche Ausreißer, deren Interpretation dem Leser überlassen sei.

Dies stellt die Frage, wie man die Qualität der Ausgleichsgerade messen kann. Die Antwort kennen wir im wesentlichen schon aus **[HM1]**, wo wir Kovarianz und Korrelation zweier Zufallsvariablen definiert haben; der besseren Verständlichkeit wegen sei der Korrelationskoeffizient hier aber noch einmal direkt eingeführt.

Wir argumentieren wie bei der Herleitung des GAUSSschen Fehlerfortpflanzungsgesetzes: Seien im folgenden $\bar t$ der Mittelwert der $t_i$ und $\bar u$ der Mittelwert der $u_i$. Falls die $u_i$ und die $t_i$ aus voneinander unabhängigen Prozessen stammen, ist der Erwartungswert von

$$\sum_{i=1}^{n} (t_i - \bar t) \cdot (u_i - \bar u)$$

null, da die Summen der $t_i - \bar t$ und der $u_i - \bar u$ jeweils null sind. Falls aber im anderen Extrem $u_i = a \cdot t_i + b$ ist für alle $i$, so ist $\bar u = a \cdot \bar t + b$ und

$$\sum_{i=1}^{n}(t_i - \bar t)\cdot(u_i - \bar u) = \sum_{i=1}^{n}(t_i - \bar t)\cdot(a\cdot t_i - a\cdot \bar t) = a\cdot\sum_{i=1}^{n}(t_i - \bar t)^2$$

eine (fast immer) positive Zahl, die eng mit der Standardabweichung der Verteilung der $t_i$ zusammenhängt. Um sie weiter auszurechnen, beachten wir, daß

$$\sum_{i=1}^{n}(u_i - \bar u)^2 = a^2\cdot\sum_{i=1}^{n}(t_i - \bar t)^2$$

ist und damit

$$\sum_{i=1}^{n}(t_i - \bar t)^2\cdot\sum_{i=1}^{n}(u_i - \bar u)^2 = a^2\cdot\left(\sum_{i=1}^{n}(t_i - \bar t)^2\right)^2$$

oder

$$\sqrt{\sum_{i=1}^{n}(t_i - \bar t)^2\cdot\sum_{i=1}^{n}(u_i - \bar u)^2} = a\cdot\sum_{i=1}^{n}(t_i - \bar t)^2 .$$

Damit ist also der Quotient

$$\rho = \frac{\displaystyle\sum_{i=1}^{n}(t_i - \bar t)\cdot(u_i - \bar u)}{\sqrt{\displaystyle\sum_{i=1}^{n}(t_i - \bar t)^2\cdot\sum_{i=1}^{n}(u_i - \bar u)^2}}$$

je nach Vorzeichen von $a$ gleich $\pm 1$, wenn alle $u_i = a \cdot t_i + b$ sind, und $r = 0$, wenn die $u_i$ unabhängig von den $t_i$ sind. Für beliebige Wertepaare $(t_i, u_i)$ sagt uns die CAUCHY-SCHWARZsche Ungleichung aus **[HM1]**, Kap. I, §7*b),* daß stets

$$-1 \le \rho \le 1$$

ist. Dieses $\rho$ nennen wir den *Korrelationskoeffizienten;* sein Betrag gibt an, wie gut die Beziehung zwischen den $u_i$ und den $t_i$ durch eine Gerade beschrieben werden kann, während sein Vorzeichen angibt, ob diese Gerade ansteigt (die Daten sind *positiv korreliert*) oder sinkt (*negative Korrelation*).

Bei den obigen Zahlenbeispielen ergeben sich die Korrelationskoeffizienten zu

$$\rho \approx 0{,}8317249988$$

für den Zusammenhang zwischen Korruption und Bruttosozialprodukt und

$$\rho \approx 0{,}7465571442$$

bei der Klausur. In beiden Fällen ist die Korrelation also recht gut.

### f) Anhang: Mehr zum Laplaceschen Fehlermodell und zur Maximum Likelihood Methode

Leider ist in dieser Vorlesung nicht genug Zeit, um auch nur annähernd die notwendigen Grundlagen für einen Beweis des zentralen Grenzwertsatzes bereitszustellen. Wir haben aber immerhin schon am Beispiel des LAPLACEschen Fehlermodell graphisch gesehen, daß für die Verteilungen mit $u_i = \pm\varepsilon$, jeweils mit Wahrscheinlichkeit $\frac{1}{2}$, die Verteilung der Summen gegen eine Normalverteilung konvergiert.

Zumindest in diesem einfachen Fall können wir dies auch rechnerisch einsehen und auf diesem Weg insbesondere auch die Gleichung der Glockenkurve herleiten. Dazu müssen wir uns zunächst überlegen, was $\varepsilon$ tun soll, wenn $n$ gegen unendlich geht.

Wir kennen zwei statistische Kennzahlen zur Beschreibung der Fehlerverteilung: Das arithmetische Mittel und die Varianz. Das arithmetische Mittel ist (z.B. aus Symmetriegründen) null, bleibt also noch die Varianz.

Zu deren Berechnung müssen wir die Fehlerquadrate über die $2^n$ möglichen Verhaltensweisen der „Dämonen“ summieren, d.h. wir müssen

$$\sigma^2 = \frac{1}{2^n}\sum(\varepsilon_1 + \ldots + \varepsilon_n)^2$$

berechnen, wobei sich die Summation über alle Tupel

$$(\varepsilon_1, \cdots, \varepsilon_n) \quad \text{mit} \quad \varepsilon_i = \pm\varepsilon$$

erstreckt. Beim Ausmultiplizieren heben sich alle gemischten Terme der Form $\varepsilon_i\varepsilon_j$ gegenseitig weg, denn das Tupel, bei dem nur an der $i$-ten Stelle das Vorzeichen das Vorzeichen geändert wurde, liefert einen Summanden $-\varepsilon_i\varepsilon_j$. Also bleiben nur die Quadrate; diese sind alle gleich $\varepsilon^2$, und es sind pro Summand $n$ Stück. Da die Anzahl der Summanden gleich dem Nenner des Vorfaktors ist, berechnet sich die Varianz daher zu

$$\sigma^2 = n\varepsilon^2$$

Somit müssen wir für $n \to \infty$ den Einfluß $\varepsilon$ jedes einzelnen „Dämonen“ so gegen null gehen lassen, daß $n\varepsilon^2$ konstant bleibt, d.h. wir setzen

$$\varepsilon = \frac{\sigma}{\sqrt{n}}$$

für eine geeignet zu wählende Konstante $\sigma > 0$, die Standardabweichung.

Wir wollen nun für $n \to \infty$ berechnen, mit welcher Wahrscheinlichkeit ein bestimmter Fehler $u$ auftritt. Für festgehaltenes $n$ und

$$u = (n-2k)\varepsilon = \frac{(n-2k)\sigma}{\sqrt{n}}$$

ist diese Wahrscheinlichkeit gleich

$$\binom{n}{k} 2^{-n-1} \frac{\sqrt{n}}{\sigma} = \frac{n!}{k!(n-k)!} 2^{-n-1} \frac{\sqrt{n}}{\sigma} ;$$

wir müssen also diesen Ausdrucks für $n \to \infty$ und konstantes $u$ betrachten.

Was uns interessiert ist allerdings nicht der Grenzwert dieses Ausdrucks: Mit $n$ geht schließlich auch die Anzahl der zu betrachtenden Fälle gegen unendlich, so daß die Wahrscheinlichkeit für jeden einzelnen dieser Fälle notgedrungen gehen Null gehen muß: Die Summe aller Wahrscheinlichkeiten kann schließlich nicht größer als eins werden.



Wenn wir $n$ variieren lassen, ist es ohnehin sinnlos, einen genauen Wert des Fehlers zu betrachten: Für jedes $n$ gibt es nur $n+1$ mögliche Werte, und bei den meisten größeren Werten von $n$ werden diese Zahlen – abgesehen von der Null – nicht auftreten. Wenn wir etwas von $n$ unabhängiges definieren möchten, dürfen wir also nicht den genauen *Wert* des Fehlers festlegen, sondern müssen ein *Fehlerintervall* betrachten.

Nun wollen wir aber natürlich als Ergebnis keine Funktion, die von einem Intervall abhängt, sondern eine gewöhnliche Funktion von $u$. Um eine solche zu bekommen, betrachten wir nicht die Wahrscheinlichkeit, sondern die *Wahrscheinlichkeitsdichte:* Für eine kontinuierlich variierende zufällige Größe definieren wir die Wahrscheinlichkeitsdichte $\varphi(u_0)$ im Punkt $u_0$ als

$$\varphi(u_0) = \lim_{\varepsilon \to 0} \frac{\text{Wahrscheinlichkeit für } u_0 - \varepsilon \le u \le u_0 - \varepsilon}{2\varepsilon} ,$$

d.h. also als Wahrscheinlichkeit dividiert durch die Intervallbreite. Falls diese Dichte existiert, folgt mehr oder weniger sofort aus der Definition des RIEMANN-Integrals, daß

$$(\text{Wahrscheinlichkeit für } a \le u \le b) = \int_a^b \varphi(u)\,du$$

ist. Unsere Ziel ist also, diese Wahrscheinlichkeitsdichte $\varphi$ für $n \to \infty$ zu berechnen.

Für festes $n$ haben die möglichen Fehlerwerte den Abstand $2\varepsilon$, wir betrachten daher Intervalle der Länge $2\varepsilon$ mit den Werten $(n-2k)\varepsilon$ als Mittelpunkten; diese überdecken den möglichen Fehlerbereich lückenlos, und die Wahrscheinlichkeit für einen Fehler in diesem Intervall ist $\binom{n}{k}2^{-n}$.

Wir interessieren uns daher für den Grenzwert von

$$\frac{\binom{n}{k}2^{-n}}{2\varepsilon} = \frac{\binom{n}{k}2^{-n}}{2\sigma/\sqrt{n}} = \binom{n}{k} 2^{-n-1}\sqrt{n}$$

für $n \to \infty$.

Das Problem bei der Berechnung dieses Grenzwerts ist der Binomialkoeffizient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}\,;$$

um diesen abzuschätzen brauchen wir einen handhabbaren Ausdruck für $n!$.

Dazu betrachten wir

$$\ln n! = \sum_{k=1}^{n} \ln k$$

und berechnen dies nach einer Methode von EULER: Bezeichnet man wie üblich mit $[x]$ die größte ganze Zahl kleiner oder gleich $x$ und mit

$$\{x\} \underset{\text{def}}{=} x - [x]$$

den gebrochenen Anteil von $x$, so zeigt partielle Integration, daß für jede differenzierbare Funktion $f$ und jede natürliche Zahl $k$ gilt

$$\begin{aligned}\int_k^{k+1} \left(\{x\} - \tfrac12\right) f'(x)\,dx &= \left(x - k - \tfrac12\right) f(x)\,\Big|_k^{k+1} - \int_k^{k+1} f(x)\,dx \\ &= \frac{f(k+1) + f(k)}{2} - \int_k^{k+1} f(x)\,dx\,.\end{aligned}$$

Summation von $k = 1$ bis $n-1$ ergibt

$$\int_1^n \left(\{x\} - \tfrac12\right) f'(x)\,dx = \frac{f(1)}{2} + \sum_{k=2}^{n-1} f(k) + \frac{f(n)}{2} - \int_1^n f(x)\,dx$$

oder

$$\sum_{k=1}^{n} f(k) = \int_1^n f(x)\,dx + \frac{f(1) + f(n)}{2} + \int_1^n \left(\{x\} - \tfrac12\right) f'(x)\,dx\,.$$



Speziell für $f(x) = \ln x$ erhalten wir

$$\begin{aligned}\ln n! &= \int_1^n \ln x\,dx + \frac{\ln n}{2} + \int_1^n \frac{\{x\} - \frac12}{x}\,dx \\ &= x(\ln x - 1)\,\Big|_1^n + \frac{\ln n}{2} + \int_1^n \frac{\{x\} - \frac12}{x}\,dx \\ &= n(\ln n - 1) + 1 + \frac{\ln n}{2} + \int_1^n \frac{\{x\} - \frac12}{x}\,dx\,.\end{aligned}$$

In dieser Formel stört noch das rechte Integral; dieses können wir wie folgt abschätzen: Für eine natürliche Zahl $k$ ist

$$\begin{aligned}\int_k^{k+1} \frac{\{x\} - \frac12}{x}\,dx &= \int_{-\frac12}^{\frac12} \frac{x}{k + \frac12 + x}\,dx \\ &= \int_0^{\frac12} \left(\frac{x}{k + \frac12 + x} - \frac{x}{k + \frac12 - x}\right) dx \\ &= \int_0^{\frac12} \frac{-2x^2}{\left(k + \frac12\right)^2 - x^2}\,dx\,.\end{aligned}$$

Im Intervall von 0 bis $\frac12$ ist der Integrand monoton fallend, d.h.

$$0 \ge \frac{-2x^2}{\left(k + \frac12\right)^2 - x^2} \ge \frac{-\frac12}{\left(k + \frac12\right)^2 - \frac14} = \frac{-2}{(2k+1)^2 - 1} \ge -\frac{1}{4k^2}\,,$$

und damit ist

$$0 \ge \int_k^{k+1} \frac{\{x\} - \frac12}{x}\,dx = \int_0^{\frac12} \frac{-2x^2}{\left(k + \frac12\right)^2 - x^2}\,dx \ge -\frac{1}{8k^2}\,,$$

denn wir können das Integral abschätzen durch das Produkt aus der Länge des Integrationsintervalls und dem Minimum des Integranden. Summation von $k = 1$ bis $n - 1$ schließlich gibt die Abschätzung

$$0 \ge \int_1^n \frac{\{x\} - \frac12}{x}\,dx \ge -\sum_{k=1}^{n-1} \frac{1}{4k^2}$$

für das störende Integral aus der obigen Formel.

Wie wohl jeder schon einmal in einer Analysis I Übungsaufgabe zeigen mußte, konvergiert die rechtsstehende Summe (egal ob mit oder ohne acht im Nenner) für $n \to \infty$, und daher konveriert auch das uneigentliche Integral

$$\int_1^\infty \frac{\{x\} - \frac12}{x}\,dx$$

gegen einen uns bislang noch unbekannten Wert $I$. Also ist

$$\ln n! = n(\ln n - 1) + \frac{\ln n}{2} + C + o(1) \quad \text{mit} \quad C = I + 1$$

oder

$$n! \approx e^C \cdot n^n e^{-n} \sqrt{n}\,.$$

Mit dieser Formel können wir nun daran gehen, den Binomialkoeffizienten $\binom{n}{k}$ abzuschätzen:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \approx \frac{e^C n^n e^{-n} \sqrt{n}}{e^C k^k e^{-k} \sqrt{k}\, e^C (n-k)^{n-k} e^{-(n-k)} \sqrt{(n-k)}}$$
$$= \frac{1}{e^C} \frac{n^n}{k^k\,(n-k)^{n-k}} \sqrt{\frac{n}{k\cdot(n-k)}}\,.$$

Ausgedrückt durch $u = (n - 2k)\varepsilon = (n - 2k)\sigma/\sqrt{n}$ ist

$$k = \frac{n}{2} - \frac{u\sqrt{n}}{2\sigma}\,,$$

und setzen wir zur Vereinfachung der Schreibweise

$$m = \frac{n}{2}, \quad v = \frac{u}{2\sigma} \quad \text{und} \quad \ell = \sqrt{2m}\cdot v\,,$$



so ergeben sich die Formeln

$$k = m - \sqrt{2m}\cdot v = m - \ell \quad \text{und} \quad n - k = m + \sqrt{2m}\cdot v = m + \ell\,.$$

Setzen wir dies alles in die Formel für die Wahrscheinlichkeitsdichte ein, erhalten wir

$$\begin{aligned}
&\binom{n}{k} 2^{-n-1} \frac{\sqrt{n}}{\sigma} \\
&\approx \frac{1}{e^C} \frac{n^n}{k^k\,(n-k)^{n-k}} \sqrt{\frac{n}{k\cdot(n-k)}}\, 2^{-n-1} \frac{\sqrt{n}}{\sigma} \\
&= \frac{1}{e^C\sigma} \frac{n^n \cdot 2^{-n}}{k^k\,(n-k)^{n-k}} \frac{n\cdot 2^{-1}}{\sqrt{k\cdot(n-k)}} \\
&= \frac{1}{e^C\sigma} \frac{(2m)^{2m}\cdot 2^{-2m}}{(m-\ell)^{m-\ell}(m+\ell)^{m+\ell}} \frac{2m\cdot 2^{-1}}{\sqrt{(m-\ell)(m+\ell)}} \\
&= \frac{1}{e^C\sigma} \frac{m^{2m}}{(m-\ell)^m (m+\ell)^m} \left(\frac{m-\ell}{m+\ell}\right)^\ell \frac{m}{\sqrt{m^2-\ell^2}} \\
&= \frac{1}{e^C\sigma} \frac{m^{2m}}{(m^2-\ell^2)^m} \left(\frac{m-\ell}{m+\ell}\right)^\ell \frac{m}{\sqrt{m^2-\ell^2}} \\
&= \frac{1}{e^C\sigma} \frac{1}{\left(1-\frac{\ell^2}{m^2}\right)^m} \left(\frac{1-\ell/m}{1+\ell/m}\right)^\ell \frac{1}{\sqrt{1-\ell^2/m^2}} \\
&= \frac{1}{e^C\sigma} \frac{1}{\left(1-\frac{2v^2}{m}\right)^m} \left(\frac{1-\sqrt{2/m}\cdot v}{1+\sqrt{2/m}\cdot v}\right)^{\sqrt{2m}\cdot v} \frac{1}{\sqrt{1-2v^2/m}}\,.
\end{aligned}$$

Nun können wir langsam daran denken, $n$ (und damit auch $m$) gegen unendlich gehen zu lassen; wir verwenden dazu die aus der Analysis I und wahrscheinlich auch aus der Schule bekannte Beziehung

$$e^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n\,.$$

Danach ist insbesondere

$$\lim_{m\to\infty} \left(1 - \frac{2v^2}{m}\right)^m = e^{-2v^2}$$

und

$$\lim_{m\to\infty}\left(1\pm\sqrt{\frac{2}{m}}\cdot v\right)^{\sqrt{2m}\cdot v} = \lim_{m\to\infty}\left(1\pm\frac{\sqrt{2}\cdot v}{\sqrt{m}}\right)^{\sqrt{m}\cdot\sqrt{2}\cdot v}$$

$$= \lim_{q\to\infty}\left(1\pm\frac{\sqrt{2}\cdot v}{q}\right)^{q\cdot\sqrt{2}\cdot v} = \left(e^{\pm\sqrt{2}\cdot v}\right)^{\sqrt{2}\cdot v} = e^{\pm 2v^2},$$

denn es bleibt sich natürlich gleich, ob $m$ oder $q = \sqrt{m}$ gegen unendlich geht. Da der Term $v^2/m$ gegen null geht, erhalten wir somit als Grenzwert des gesamten obigen Ausdrucks

$$\frac{1}{e^C\sigma}\cdot\frac{1}{e^{-2v^2}}\cdot\frac{e^{-2v^2}}{e^{+2v^2}}\cdot 1 = \frac{1}{e^C\sigma}e^{-2v^2}.$$

Beachten wir nun noch, daß $v = u/2\sigma$ war, erhalten wir

$$\frac{1}{e^C\sigma}e^{-u^2/2\sigma^2}.$$

Damit sind wir fast am Ziel; das einzige, was noch fehlt, ist die Konstante $C$. Diese können wir bestimmen, indem wir ausnutzen, daß jeder Fehler mit Wahrscheinlichkeit eins zwischen $-\infty$ und $\infty$ liegt, d.h.

$$\frac{1}{e^C\sigma}\int_{-\infty}^{\infty} e^{-u^2/2\sigma^2}\,du = 1.$$

Aus §6*c)* wissen wir, daß

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$$

ist; mit der Substitution $x = u/\sqrt{2}\sigma$ folgt, daß dann

$$\int_{-\infty}^{\infty} e^{-u^2/2\sigma^2}\,du = \sqrt{2\pi}\sigma$$

ist und

$$e^C = \sqrt{2\pi} \quad\text{oder}\quad C = \tfrac{1}{2}\ln(2\pi).$$



Damit haben wir die Wahrscheinlichkeitsdichte endlich vollständig berechnet; das Endergebnis ist

$$\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-u^2}{2\sigma^2}}.$$

Auch die Formel für $n!$ können wir nach der Bestimmung von $C$ nun vollständig hinschreiben:

$$\ln n! = n(\ln n - 1) + \ln\sqrt{2\pi n} + o(1) \quad\text{oder}\quad n! \approx \left(\frac{n}{e}\right)^n\sqrt{2\pi n}.$$

Beides bezeichnet man als STIRLINGsche Formel.

Der schottische Mathematiker JAMES STIRLING (1692–1770) war Anhänger des gestürzten Königs Jakob II Stuart und hatte deshalb große politische Probleme bei seinem Studium; unter anderem wurde er deshalb von der Universität Oxford ausgeschlossen. 1717–1722 lebte er in Venedig und hatte auch gute Kontakte zu NICOLAUS BERNOULLI an der Universität von Padua; außerdem brachte er aus Venedig die Produktionsgeheimnisse der dortigen Glasbläser mit. Ab 1724 arbeitete er zehn Jahre lang als Mathematiklehrer in London, wo er viel mit NEWTON zusammentraf; 1735 wurde er Direktor einer schottischen Bergbaugesellschaft. In seine Londoner Zeit fällt die Veröffentlichung seines bedeutendsten Werks *Methodus Differentialis sive Tractatus de Summatione er Interpolatione Serierum Infinitarum* im Jahre 1730, das die obige Formel als Beispiel zwei zu Proposition 28 enthält. Ebenfalls ziemlich bekannt wurde seine 1735 veröffentlichte Arbeit über die Gestalt der Erde.

Oft interessiert nicht so sehr die Verteilung der Fehler, sondern die der Meßwerte selbst. Ist $\hat{x}$ der korrekte Wert und $x_i$ der $i$-te Meßwert dafür, der gemäß $x_i = \hat{x} + u_i$ mit dem Fehler $u_i$ behaftet ist, so können wir mit $x = \hat{x} + u$ die obigen Wahrscheinlichkeitsdichte auch als

$$\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\hat{x})^2/2\sigma^2}$$

schreiben.

Als *Normalverteilung mit Mittelwert $a$ und Standardabweichung $\sigma$* bezeichnen wir daher die Verteilung mit Wahrscheinlichkeitsdichte

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Diese Wahrscheinlichkeitsdichte hängt offensichtlich nur von der *normierten* Variablen

$$z = \frac{x-a}{\sigma}$$

ab; diese hat Mittelwert null und Standardabweichung eins. Daher gibt es für die Normalverteilung nicht – wie für viele andere statistische Verteilungen – je nach Parameterwerten verschiedene Tabellen, sondern man findet in allen Tabellenwerken nur die Normalverteilung mit Mittelwert null und Standardabweichung eins, man findet also die Wahrscheinlichkeitsdichte

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

und deren Integral

$$F(z) = \int_{-\infty}^{z} f(u)\,du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}}\,du\,.$$

Dieses Integral läßt sich nicht weiter vereinfachen, da sich die Stammfunktion von $e^{-u^2/2}$ nicht durch elementare Funktionen ausdrücken läßt. Für die Bestimmung von $F(z)$ ist man daher auf Tabellen oder Computerprogramme angewiesen; eine graphische Darstellung von $F(z)$ ist in Abbildung 66 zu sehen. Mit dieser Funktion läßt sich die Wahrscheinlichkeit dafür, daß

$$c \le z = \frac{x-a}{\sigma} \le d$$

ist berechnen als $F(d) - F(c)$, und damit läßt sich auch leicht die Wahrscheinlichkeit berechnen, daß $x$ selbst zwischen zwei gegebenen Schranken liegt.

Wie Abbildung 66 zeigt, liegt $F(-2)$ sehr nahe bei null und $F(2)$ sehr nahe bei eins. In der Tat ist die Wahrscheinlichkeit dafür, daß ein Wert $z$ Betrag größer zwei hat, kleiner als 5%; genauer gilt für eine beliebige Normalverteilung, daß der Wert der Variablen mit folgenden Wahrscheinlichkeiten um höchstens $i\sigma$ vom Mittelwert abweicht:

| $i =$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Wahrscheinlichkeit: | 0,683 | 0,954 | 0,9973 | 0,99994 |

Damit liegen also etwa zwei Drittel aller Fehler zwischen $-\sigma$ und $\sigma$, 95% liegen zwischen $-2\sigma$ und $2\sigma$ und 99,7% zwischen $-3\sigma$ und $3\sigma$; die Wahrscheinlichkeit dafür, daß der Fehler größer als $3\sigma$ ist, beträgt



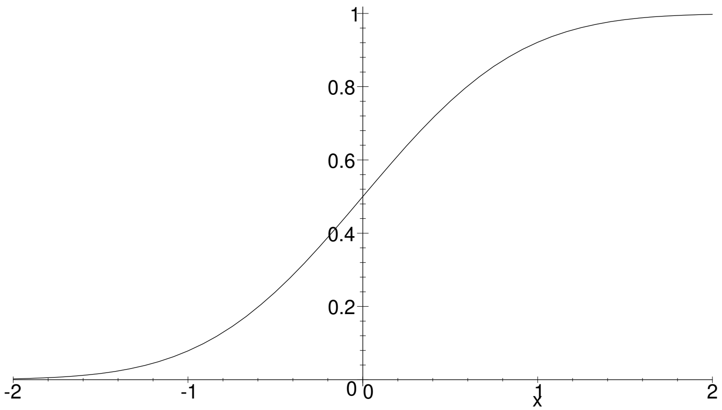


Abb. 66: Das Integral $F(z)$ über die „Glockenkurve"

also nur etwa 0,27%. Da Ereignisse mit einer so geringen Wahrscheinlichkeit seltener als in einem von 300 Fällen auftreten, betrachtet man Fehler, die außerhalb des $3\sigma$-Bereichs liegen, oft als „Ausreißer", d.h. als grobe Meßfehler, die bei der Bestimmung des Ergebnisses nicht berücksichtigt werden. Sehr vorsichtige Leute reden allerdings erst ab einer Abweichung von $4\sigma$ von Ausreißern.

Für Leser, die ihren Computer selbst programmieren und keine spezielle Statistiksoftware haben, sei hier eine Näherungsformel für $F(z)$ angegeben: Mit einem Fehler von höchstens $7{,}5 \cdot 10^{-8}$ ist

$$F(z) = 1 - \varphi(z) \cdot \left(a_1 t + a_2 t^2 + a_3 a^3 + a_4 t^4 + a_5 t^5\right)$$

mit $t = \dfrac{1}{1+pz}$ und $\varphi(z) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ sowie

$$\begin{array}{lll} a_1 = \phantom{-}0{,}319\,381\,530 & a_2 = -0{,}356\,563\,782 & a_3 = 1{,}781\,477\,973 \\ a_4 = -1{,}821\,255\,978 & a_5 = \phantom{-}1{,}330\,274\,429 & p = 0{,}231\,641\,9 \quad . \end{array}$$

Beim Rechnen mit dem Taschenrechner kann man sich auch mit einer vereinfachten Version begnügen, bei der $a_4 = a_5 = 0$ ist und

$$a_1 = 0{,}436\,1836 \quad a_2 = -0{,}120\,1676 \quad a_3 = 0{,}937\,2980 \quad p = 0{,}332\,67\,;$$

hier kann der Fehler bis zu $10^{-5}$ betragen.

***Die Maximum Likelihood Methode:*** GAUSS gab im Laufe seines Lebens mehrere Begründungen für die Methode der kleinsten Quadrate (die er bei sowohl bei seinen astronomischen Arbeiten wie auch bei der von ihm geleiteten Vermessung des Königreichs Hannover zwischen 1818 und 1832 ständig benutzte); die unter dem Gesichtspunkt einer in sich geschlossenen Fehlertheorie interessanteste beruht auf dem LAPLACEschen Fehlermodell:

Danach sollte der Wert $u_i$ für die korrekten Parameterwerte $a, b, \dots$ aus einer Normalverteilung mit Mittelwert $f(a, b, \dots; t_i)$ kommen, deren Standardabweichung $\sigma_i$ von der Genauigkeit abhängt, mit der $u_i$ bestimmt werden kann. Die Wahrscheinlichkeit dafür, daß $u_i$ zwischen zwei Werten $a$ und $b$ liegt, ist damit

$$\int_a^b e^{-\left(u - f(a,b,\dots;t_i)\right)^2/2\sigma_i} .$$

Von der Wahrscheinlichkeit, daß $u_i$ gleich einem Wert $c$ ist, können wir natürlich nicht reden, da diese nach obiger Formel ein Integral von $c$ nach $c$ wäre, also Null. Aber die Wahrscheinlichkeit dafür, daß $u_i$ in einem kleinen Intervall der Länge $\varepsilon_i$ um einen Wert $c_i$ liegt, ist ungefähr proportional zum $\varepsilon_i$-fachen Wert des Integranden an der Stelle $c_i$, also zu

$$\varepsilon_i \cdot e^{-\left(c_i - f(a,b,\dots;t_i)\right)^2/2\sigma_i} .$$

Entsprechend ist

$$\varepsilon_j \cdot e^{-\left(c_j - f(a,b,\dots;t_j)\right)^2/2\sigma_j}$$

ungefähr gleich der Wahrscheinlichkeit dafür, daß $u_j$ in einem Intervall der Breite $\varepsilon_j$ um $c_j$ liegt.

Wenn wir wie üblich davon ausgehen, daß keine systematischen Fehler auftreten, sind die Fehler von $u_i$ und $u_j$ voneinander unabhängig, die Wahrscheinlichkeit dafür, daß $(u_i, u_j)$ in einem Rechteck mit Seiten $\varepsilon_i$ und $\varepsilon_j$ um $(c_i, c_j)$ liegt, ist also proportional zum Produkt der beiden obigen Einzelwahrscheinlichkeiten, d.h. zu

$$\varepsilon_i\varepsilon_j\, e^{-\left(c_i - f(a,b,\dots;t_i)\right)^2/2\sigma_i - \left(c_j - f(a,b,\dots;t_j)\right)^2/2\sigma_j} .$$



Entsprechend kann auch die Wahrscheinlichkeit dafür berechnet werden, daß der Punkt $(u_1, \dots, u_n)$ in einem kleinen gegebenen Quader mit Kantenlängen $\varepsilon_1, \dots, \varepsilon_n$ liegt; sie ergibt sich zu

$$L(a, b, \dots) \cdot \prod_{i=1}^n \varepsilon_i$$

mit

$$L(a, b, \dots) \underset{\text{def}}{=} e^{\sum_{i=1}^n \left(u_i - f(a,b,\dots;t_i)\right)^2/2\sigma_i} .$$

Diese Größe ist selbst keine Wahrscheinlichkeit, sondern der Quotient aus einer Wahrscheinlichkeit und einem Volumen; man spricht daher von einer *Wahrscheinlichkeitsdichte.*

Wenn wir diese Wahrscheinlichkeitsdichte als Funktion von $a, b, \dots$ betrachten, macht sie eine Aussage über die Güte der Parameter: Schließlich wird man einem Modell, das dem beobachteten Ausgang eines Experiments eine hohe Wahrscheinlichkeit zuweist, eher glauben als einem alternativen Modell, das die beobachteten Daten zu Ausreißern erklärt. Aus diesem Grund kann die Funktion $L$ auch als Maß dafür betrachtet werden, wie „wahrscheinlich" in irgendeinem umgangssprachlichen (und schwer präzisierbaren) Sinne die Parameter $a, b, \dots$ sind.

Im englischen gibt es zwei Wörter für Wahrscheinlichkeit: Das romanische Wort *probability* und das germanische Wort *likelihood.* Für den mathematisch exakten Wahrscheinlichkeitsbegriff verwendet man probability, für „Wahrscheinlichkeit" im Sinne der Funktion $L$ likelihood. Da es im deutschen kein zweites Wort für Wahrscheinlichkeit gibt, spricht man hier in Anlehnung an das Englische von einer *Likelihoodfunktion.*

Die Maximum Likelihood Methode besteht nun genau in dem, was ihr Name besagt: *Man wähle die Parameter $a, b, \dots$ so, daß die Likelihoodfunktion maximal wird.*

Da $L(a, b, \dots)$ durch eine Exponentialfunktion beschrieben wird, wird die Likelihoodfunktion genau dann maximal, wenn ihr Exponent maximal wird. Dieser Exponent ist eine negative Zahl, wird also genau dann maximal, wenn sein Betrag *minimal* wird, das heißt, wenn die Quadratsumme

$$\sum_{i=1}^{n} \frac{\big(u_i - f(a,b,\ldots;t_i)\big)^2}{2\sigma_i}$$

minimal wird.

In vielen Fällen wird die Zuverlässigkeit der einzelnen Paare $(t_i, u_i)$ miteinander vergleichbar sein, so daß alle $\sigma_i$ gleich sind; in diesem Fall kann man die $\sigma_i$ ignorieren und einfach die Quadratsumme

$$\sum_{i=1}^{n} \big(u_i - f(a,b,\ldots;t_i)\big)^2$$

minimieren, d.h. wir kommen wieder zur klassischen Methode der kleinsten Quadrate. Es gibt aber auch Anwendungen, wie etwa oben beim überexponentiellen Bevölkerungswachstum, bei denen die Verschiedenheit des $\sigma_i$ sehr wesentlich ist: Sicherlich wird man etwa der auf Volkszählungen beruhenden Weltbevölkerungszahl, die die Vereinten Nationen für 1995 veröffentlichten, mehr Vertrauen entgegenbringen als der Schätzung eines Historikers für die Weltbevölkerung des Jahres Null, und selbst bei ein und derselben Meßreihe im Labor kommt es gelegentlich vor, daß (beispielsweise aufgrund unterschiedlicher Genauigkeit eines Meßinstruments in verschiedenen Bereichen) manche Daten zuverlässiger sind als andere.



E N D E

S C H Ö N E F E R I E N !