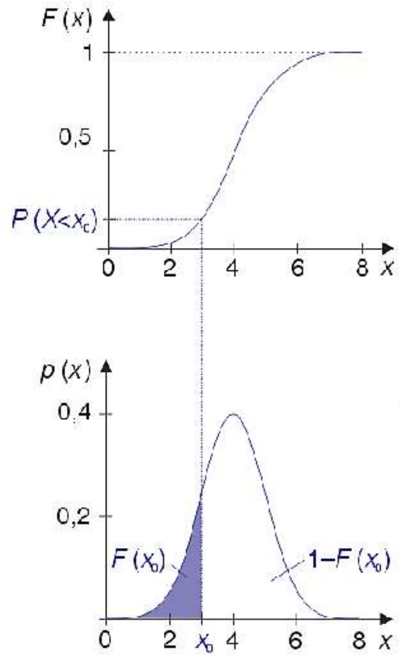
2 Signalstatistik

- Abhängigkeit von 2 Variablen heißt, dass auch Verteilungen zweidimensional (Glocken, Quader, ...)
- 2D-Verteilungsfunktion:

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(u,v)\, du\, dv$$

- 2D-Wahrscheinlichkeitsdichtefunktion:

$$p(x,y) = \frac{\partial^2 F(x,y)}{\partial x\, \partial y}$$



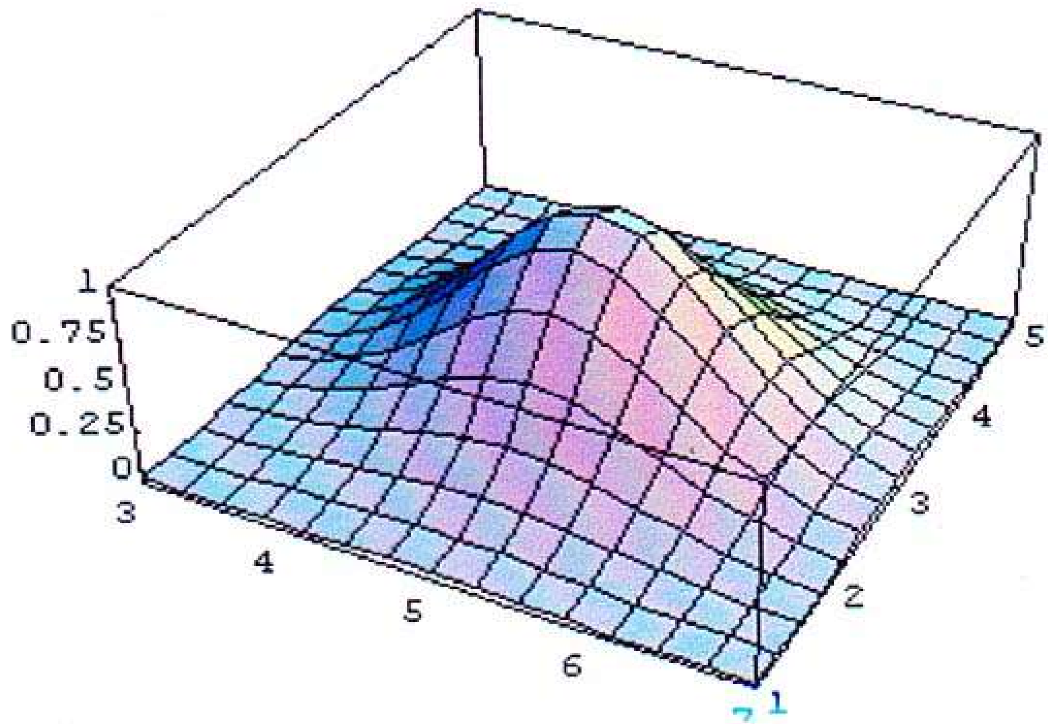
bei Unabhängigkeit der beiden Variablen untereinander gilt mit:

$$
\begin{aligned}
F(x, y) &= \mathrm{P}(X \leq x, Y \leq y) \\
F(x, \infty) &= \mathrm{P}(X \leq x) = F_x(x) \\
F(\infty, y) &= \mathrm{P}(Y \leq y) = F_y(y)
\end{aligned}
$$

die folgende Beziehung, die auch als Prüfung auf Unabhängigkeit verwendet werden kann:

$$
\begin{aligned}
F(x, y) &= F_x(x) \cdot F_y(y) \\
p(x, y) &= p_x(x) \cdot p_y(y)
\end{aligned}
$$

2D-NV



pD-NV

mit Σ als Kovarianzmatrix

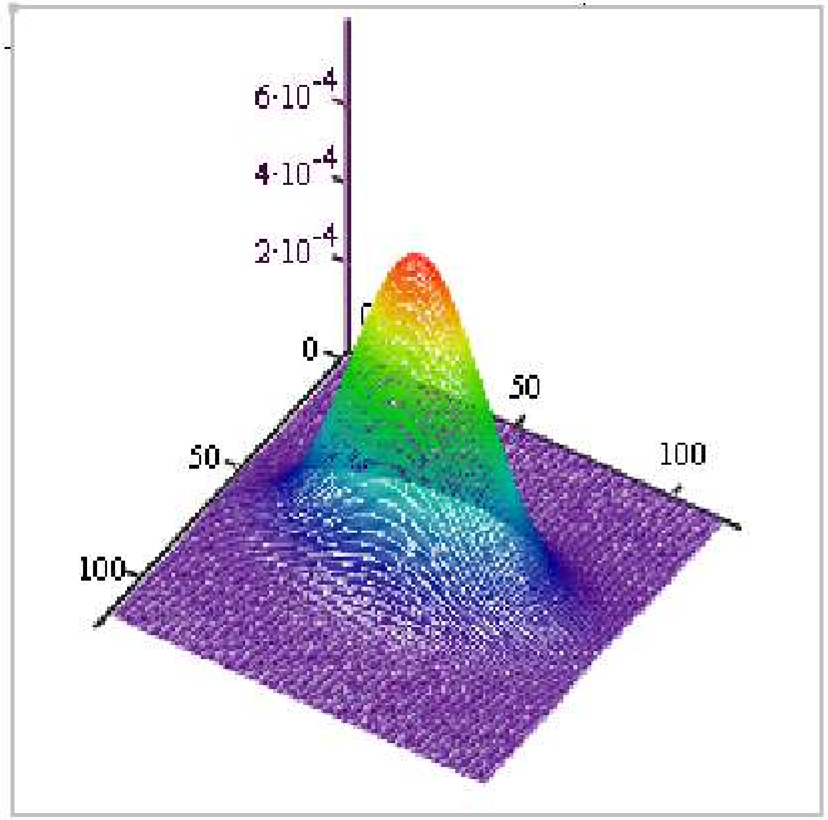
$$f_X(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left\{ -\tfrac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

$$f1(x1,y1) := \frac{1}{2\cdot\pi\cdot s1\cdot s2}\cdot e^{\left[\frac{-1}{2}\cdot\left[\frac{(x1-a)^2}{s1^2}+\frac{(y1-b)^2}{s2^2}\right]\right]}$$

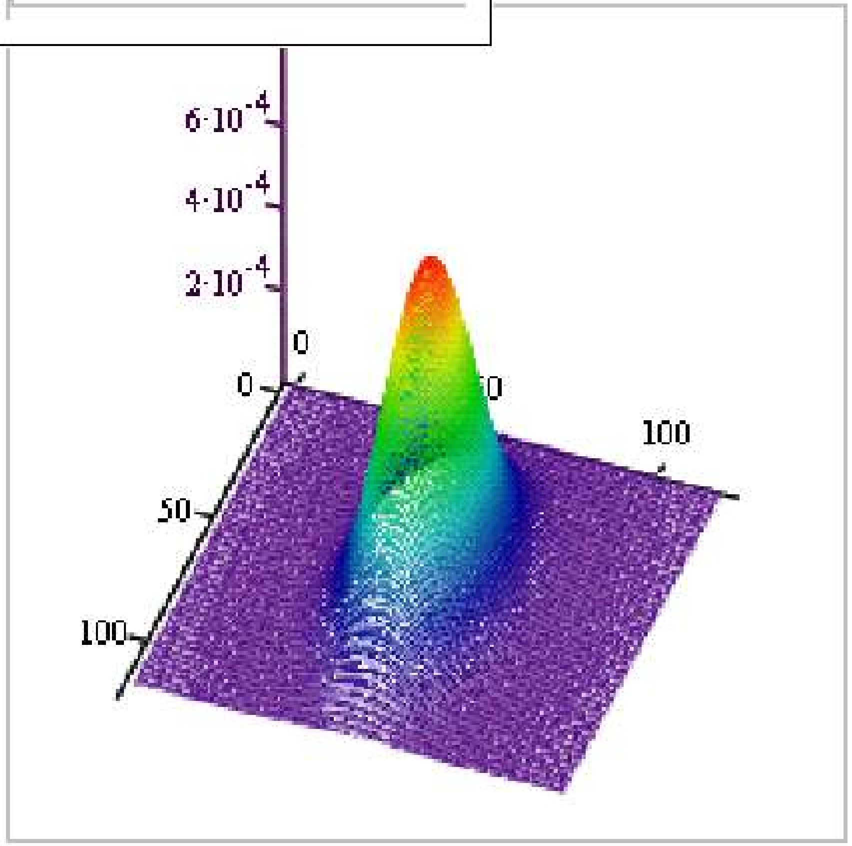
$a := 60 \quad b := 60$

$s1 := 10 \quad s2 := 20$



M1

$$f2(x2,y2) := \frac{1}{2\cdot\pi\cdot s3\cdot s4}\cdot e^{\left[\frac{-1}{2}\cdot\left[\frac{(x2-c)^2}{s3^2}+\frac{(y2-d)^2}{s4^2}\right]\right]}$$



M2

$c := 60 \quad d := 60$

$s3 := 20 \quad s4 := 10$

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{5.1}$$

Where $*$ is the convolution operation in x and y, and $G(x, y, \sigma)$ is a variable-scale Gaussian defined as Govindarajulu and Kumar Reddy (2012):

$$G(x, y, \sigma) = \frac{1}{2\Pi\sigma^2} e^{-(x^2+y^2)//2\sigma^2} \tag{5.2}$$

Beispiel aus Ahrens/Läuter:
Mehrdimensionale Varianzanalyse, S. 82

Messung des Geburtsgewichts und der Körperlänge von Neugeborenen an der Universitätskinderklinik der Charité Berlin

Stichprobe: 20 Kinder eines Kollektivs

$$gg := \begin{pmatrix} 3300 \\ 3250 \\ 3900 \\ 3800 \\ 4460 \\ 3650 \\ 3350 \\ 3650 \\ 2950 \\ 2950 \\ 3250 \\ 3370 \\ 2740 \\ 2950 \\ 3790 \\ 3420 \\ 3550 \\ 3450 \\ 4000 \\ 4400 \end{pmatrix} \qquad kl := \begin{pmatrix} 50 \\ 50 \\ 53 \\ 51 \\ 54 \\ 53 \\ 52 \\ 50 \\ 53 \\ 48 \\ 51 \\ 50 \\ 48 \\ 49 \\ 53 \\ 51 \\ 52 \\ 51 \\ 54 \\ 57 \end{pmatrix}$$

Messung des Geburtsgewichts und der Körperlänge
von Neugeborenen an der Kinderklinik der Charité
[Ahrens/Läuter: Mehrdimensionale Varianzanalyse, S. 82]

Umfang der Stichprobe = 20
Geburtsgewicht gg in Gramm
Körperlänge kl in Zentimeter"

$gg^T =$

	0	1	2	3	4	5	6	7	8	9
0	3300	3250	3900	3800	4460	3650	3350	3650	2950	2950

....

$\text{mittelwert}(gg) = 3509$

$\text{var}(gg) = 202999$ $\quad$ $\text{stdev}(gg) = 450.554$

$kl^T =$

	0	1	2	3	4	5	6	7	8	9
0	50	50	53	51	54	53	52	50	53	48

....

$\text{mittelwert}(kl) = 51.5$

$\text{var}(kl) = 4.65$ $\quad$ $\text{stdev}(kl) = 2.156$

Messung des Geburtsgewichts und der Körperlänge
von Neugeborenen an der Frauenklinik in Chemnitz

Umfang der Stichprobe = 27
Geburtsgewicht gg in Gramm
Körperlänge kl in Zentimeter"

$gg^T =$

	0	1	2	3	4	5	6	7	8	9
0	2350	3570	3400	2590	3490	3510	3290	4115	2850	3290

mittelwert(gg) = 3264.444

var(gg) = 282606.173 stdev(gg) = 531.607

$kl^T =$

	0	1	2	3	4	5	6	7	8	9
0	45	50	49	46	50	52	49	52	46	48

mittelwert(kl) = 49.074

var(kl) = 5.032 stdev(kl) = 2.243

Messung des Geburtsgewichts und der Körperlänge
von Neugeborenen an der Kinderklinik der Charité (2005)

Umfang der Stichprobe = 48
Geburtsgewicht gg in Gramm
Körperlänge kl in Zentimeter"

$gg^T =$

	0	1	2	3	4	5	6	7	8	9
0	3695	3850	3565	3875	3470	3645	4050	3710	3585	3355

....

mittelwert(gg) = 3367.396

var(gg) = 199291.656 stdev (gg) = 446.421

$kl^T =$

	0	1	2	3	4	5	6	7	8	9
0	52	53	51	54	50	53	56	52	53	51

....

mittelwert(kl) = 50.125

var(kl) = 7.193 stdev (kl) = 2.682

2 Signalstatistik

Kovarianz: Erwartungswert des Produkts der Abweichungen der Zufallsvariablen von ihren Mittelwerten

$$\begin{aligned} cov\,(X,Y) &= E\,[(X-E\,(X))\cdot(Y-E\,(Y))] \\ &= E\left[\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)\right] \\ &= \overline{XY}-\overline{X}\;\overline{Y}-\overline{X}\;\overline{Y}+\overline{X}\;\overline{Y} \\ &= \overline{XY}-\overline{X}\;\overline{Y} \quad \text{mit} \quad E\,(X)=\overline{X} \quad \text{und} \quad E\,(Y)=\overline{Y} \end{aligned}$$

Kovarianz sagt etwas darüber aus, ob sich zwei (oder mehrere) Prozesse
in etwa um dasselbe Mittel herum abspielen und ob die Abweichungen davon ähnlich sind

Kovarianz erfasst Grad der (linearen) Abhängigkeit zwischen X und Y

offensichtlich gilt für $Y = X$:

$$cov\,(X,X)=\overline{X\;X}-\overline{X}\;\overline{X}=E\left(X^2\right)-{m_1}^2=D^2\,(X)=var(X)$$

Messung des Geburtsgewichts und der Körperlänge
von Neugeborenen an der Kinderklinik der Charité
[Ahrens/Läuter: Mehrdimensionale Varianzanalyse, S. 82]

Umfang der Stichprobe = 20
Geburtsgewicht gg in Gramm
Körperlänge kl in Zentimeter"

$gg^T =$

	0	1	2	3	4	5	6	7	8	9
0	3300	3250	3900	3800	4460	3650	3350	3650	2950	2950

....

$\text{mittelwert}(gg) = 3509$

$\text{var}(gg) = 202999 \qquad \text{stdev}(gg) = 450.554$

$kl^T =$

	0	1	2	3	4	5	6	7	8	9
0	50	50	53	51	54	53	52	50	53	48

....

$\text{mittelwert}(kl) = 51.5$

$\text{var}(kl) = 4.65 \qquad \text{stdev}(kl) = 2.156$

$\text{kvar}(gg, kl) = 781.5$

$$S := \begin{pmatrix} \text{var}(gg) & \text{kvar}(gg, kl) \\ \text{kvar}(gg, kl) & \text{var}(kl) \end{pmatrix} \qquad S = \begin{pmatrix} 202999 & 781.5 \\ 781.5 & 4.65 \end{pmatrix}$$

Messung des Geburtsgewichts und der Körperlänge
von Neugeborenen an der Kinderklinik der Charité (2005)

Umfang der Stichprobe = 48
Geburtsgewicht gg in Gramm
Körperlänge kl in Zentimeter"

$gg^T =$

	0	1	2	3	4	5	6	7	8	9
0	3695	3850	3565	3875	3470	3645	4050	3710	3585	3355

....

$\text{mittelwert}(gg) = 3367.396$

$\text{var}(gg) = 199291.656$ $\text{stdev}(gg) = 446.421$

$kl^T =$

	0	1	2	3	4	5	6	7	8	9
0	52	53	51	54	50	53	56	52	53	51

....

$\text{mittelwert}(kl) = 50.125$

$\text{var}(kl) = 7.193$ $\text{stdev}(kl) = 2.682$

$$S := \begin{pmatrix} \text{var}(gg) & \text{kvar}(gg, kl) \\ \text{kvar}(gg, kl) & \text{var}(kl) \end{pmatrix} \qquad S = \begin{pmatrix} 199291.656 & 1033.659 \\ 1033.659 & 7.193 \end{pmatrix}$$

$\text{kvar}(gg, kl) = 1033.659$

Messung des Geburtsgewichts und der Körperlänge
von Neugeborenen an der Kinderklinik in Chemnitz
Geburten vom 8.11.2010 bis 24.11.2010

Umfang der Stichprobe = 27
Geburtsgewicht gg in Gramm
Körperlänge kl in Zentimeter"

$gg^T =$

	0	1	2	3	4	5	6	7	8	9
0	2350	3570	3400	2590	3490	3510	3290	4115	2850	3290

$\text{mittelwert}(gg) = 3264.444$

$\text{var}(gg) = 282606.173$ $\text{stdev}(gg) = 531.607$

$kl^T =$

	0	1	2	3	4	5	6	7	8	9
0	45	50	49	46	50	52	49	52	46	48

$\text{mittelwert}(kl) = 49.074$

$\text{var}(kl) = 5.032$ $\text{stdev}(kl) = 2.243$

$\text{kvar}(gg, kl) = 1038.745$

$$S := \begin{pmatrix} \text{var}(gg) & \text{kvar}(gg,kl) \\ \text{kvar}(gg,kl) & \text{var}(kl) \end{pmatrix} \qquad S = \begin{pmatrix} 282606.173 & 1038.745 \\ 1038.745 & 5.032 \end{pmatrix}$$

Kovarianzmatrix für mehr als zwei Zufallsgrößen

$$\mathbf{COV}\left(X_1, X_2, \ldots, X_n\right) = \begin{bmatrix} cov\left(X_1, X_1\right) & cov\left(X_1, X_2\right) & \cdots & cov\left(X_1, X_n\right) \\ cov\left(X_2, X_1\right) & cov\left(X_2, X_2\right) & \cdots & cov\left(X_2, X_n\right) \\ \vdots & \vdots & \ddots & \vdots \\ cov\left(X_n, X_1\right) & cov\left(X_n, X_2\right) & \cdots & cov\left(X_n, X_n\right) \end{bmatrix}$$

Korrelation:

Erwartungswert des Produkts Zufallsvariablen

Korrelationsmatrix für mehr als zwei Zufallsgrößen

$$\mathbf{COR}(X_1, X_2, \ldots, X_n) = \begin{bmatrix} cor(X_1, X_1) & cor(X_1, X_2) & \cdots & cor(X_1, X_n) \\ cor(X_2, X_1) & cor(X_2, X_2) & \cdots & cor(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ cor(X_n, X_1) & cor(X_n, X_2) & \cdots & cor(X_n, X_n) \end{bmatrix}$$

Beispiel Benzinverbrauch

Gallonen pro Meile

Beispiel Neugeborene:

Pfund und Fuß

Beispiel Leuchtdichte:

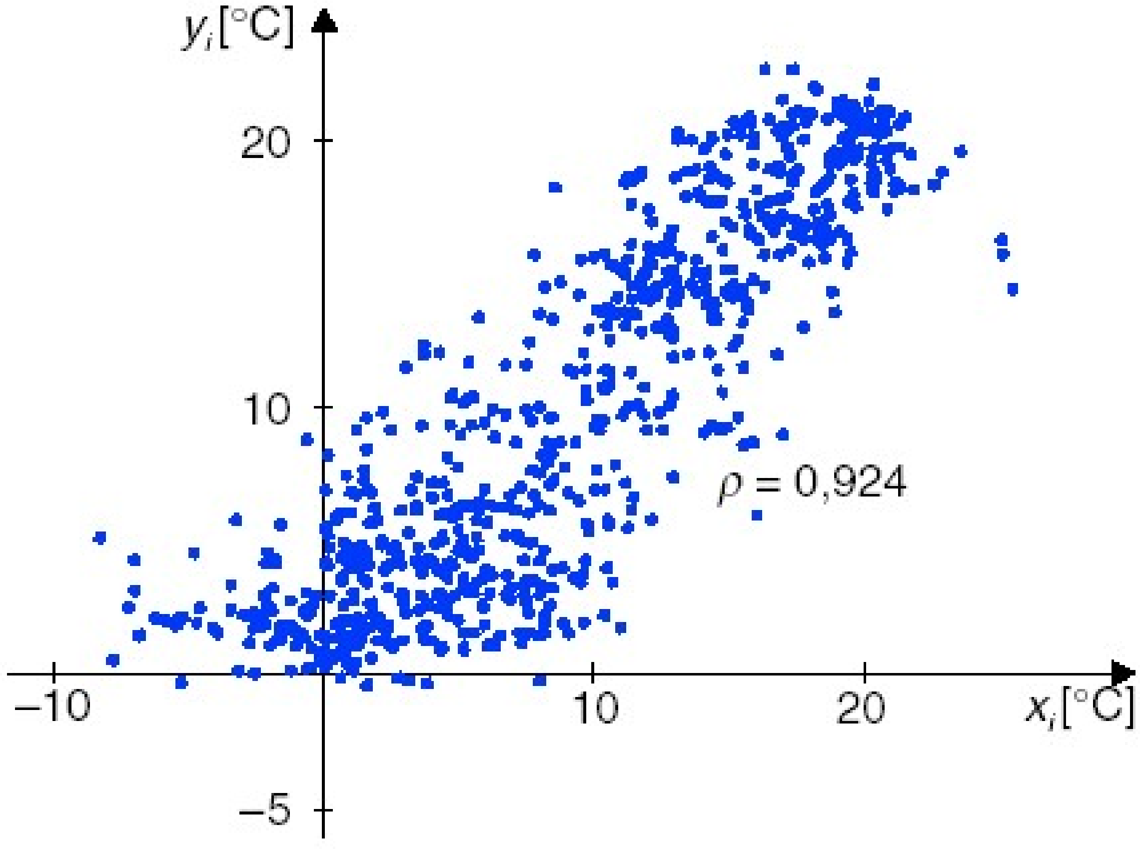
Candela pro Quadratmeter

ursprünglich: ft.la.

foot-lambert = $(1/\pi)$ Kerzen pro Quadratfuß

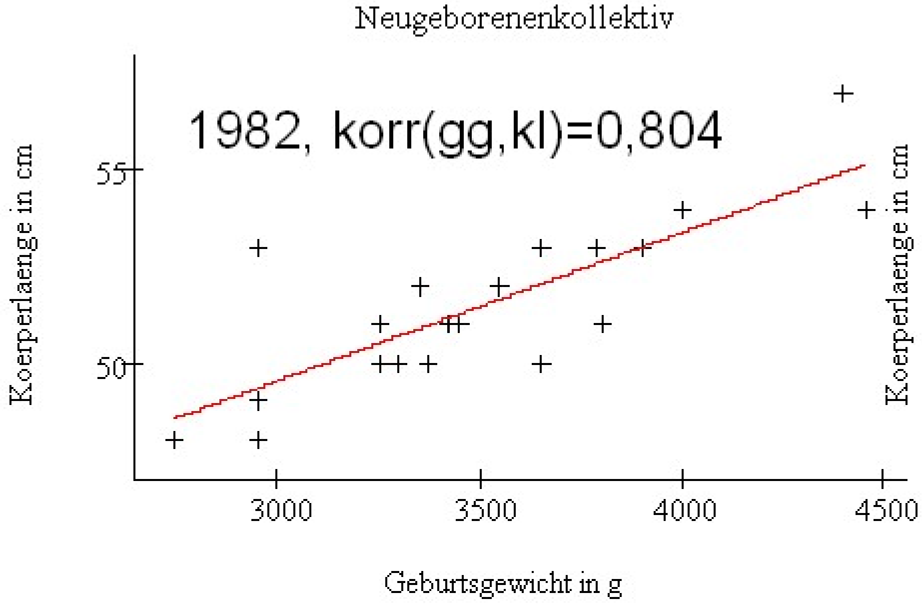
2. Beispiel:

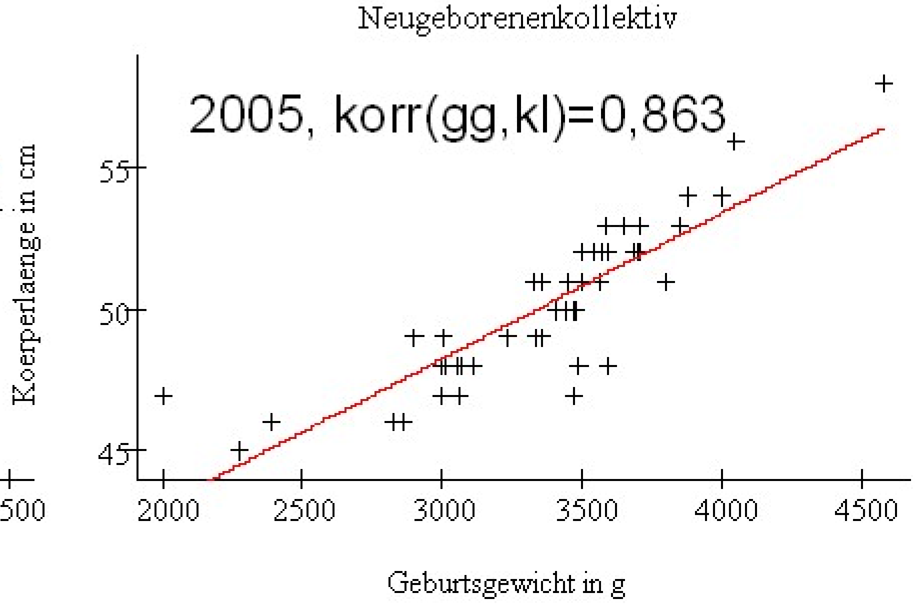
- Messreihe der mittleren Lufttemperaturen x_i in Zinnowitz (Insel Usedom)
- Messreihe der Wassertemperaturen y_i der Ostsee, gemessen in Koserow (Insel Usedom)
- Zeitraum Januar 2001 bis Dezember 2002 (730 Tage)
- empirischer Korrelationskoeffizient $\rho = 0{,}924$

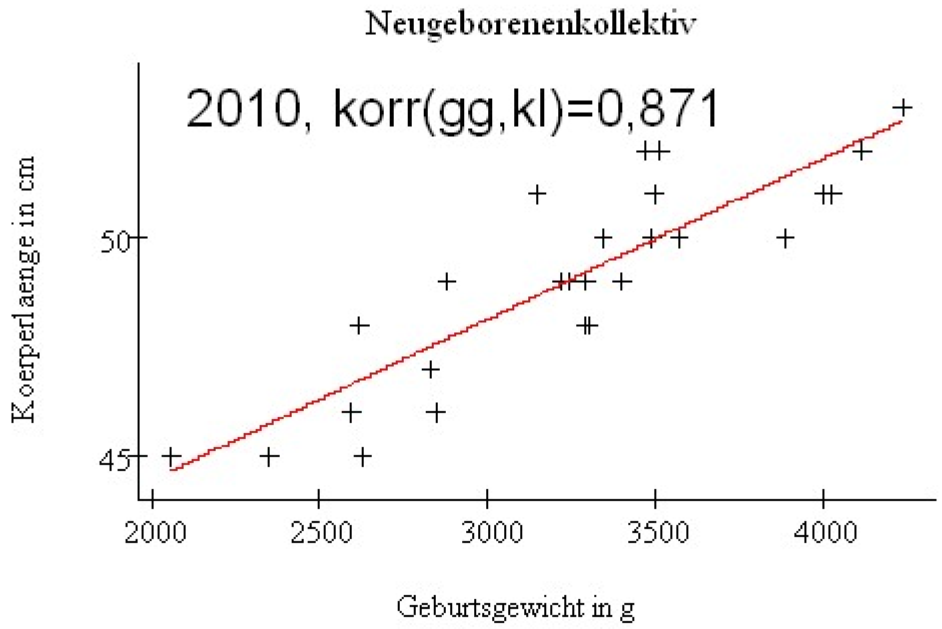

y_i[°C]
20
10
−5
−10
10
20
x_i[°C]
$\rho = 0{,}924$

Randbemerkung:

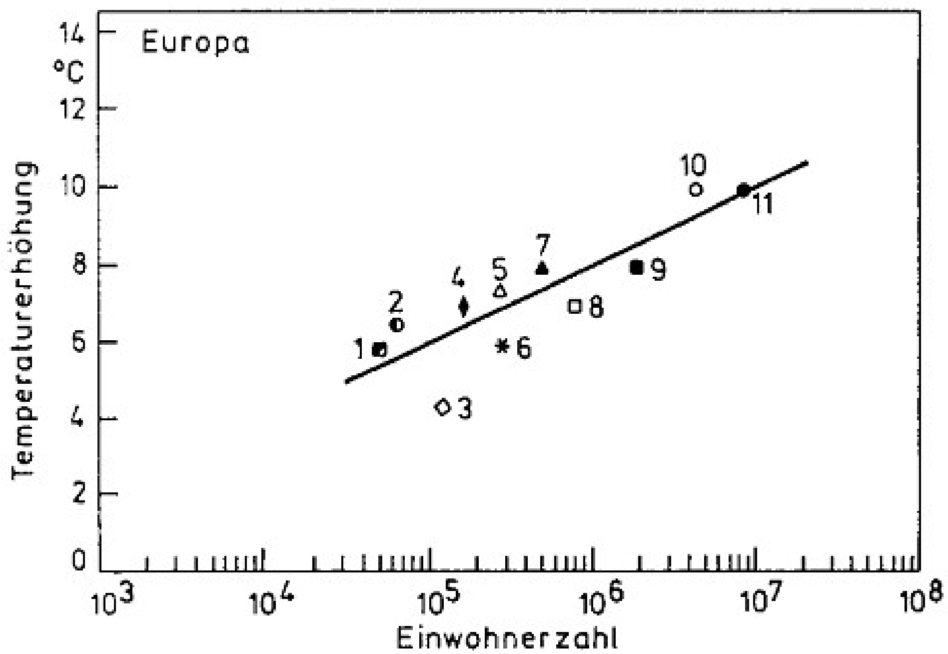
- Regressionsanalyse sucht Funktion, die Zusammenhang zwischen Wertepaaren annähert (quantifiziert)
- Korrelationskoeffizient bestimmt Grad des linearen Zusammenhangs

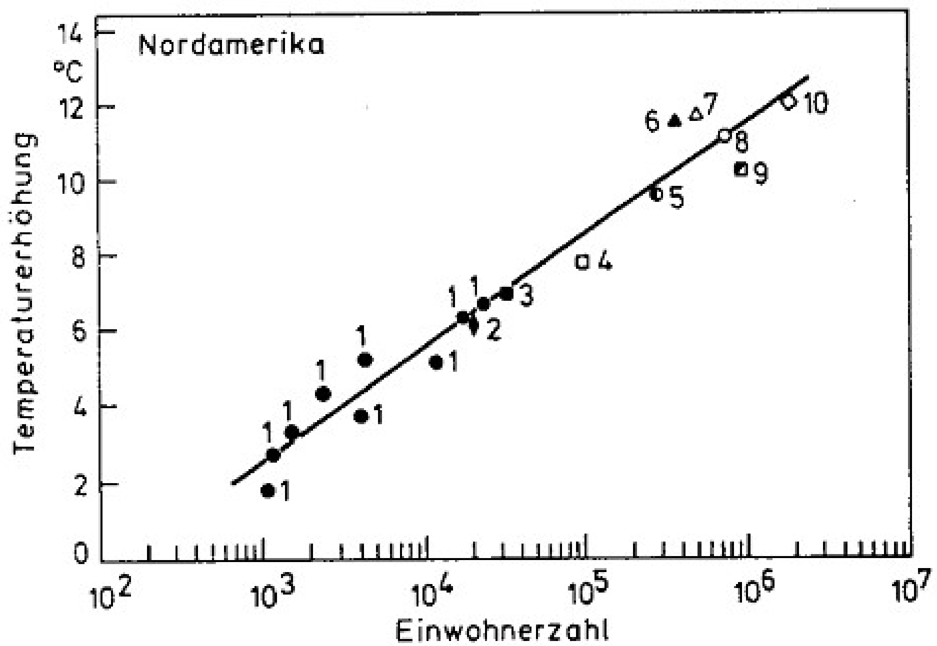




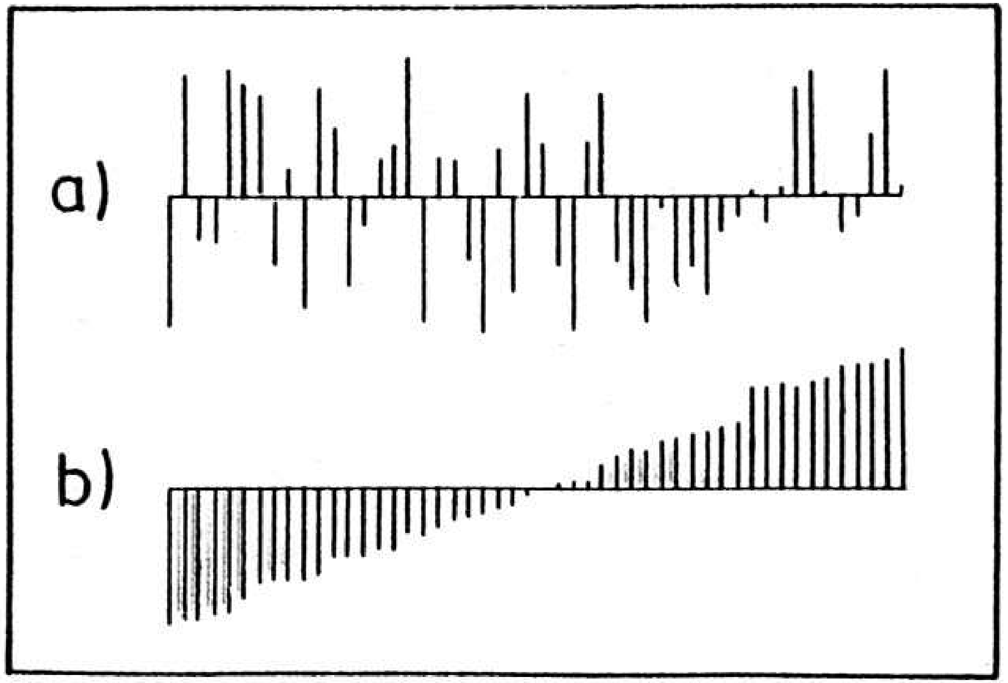


Jahr	gg	kl
1982	3509	51,5
2005	3367	50,1
2010	3264	49,1


Europa
Temperaturerhöhung
°C
Einwohnerzahl


Nordamerika
Temperaturerhöhung
°C
Einwohnerzahl

2 Signalstatistik



Zwei Signale mit unterschiedlichem Informationsgehalt. Bei a) sind aufeinanderfolgende Signalwerte in keiner Weise vorhersehbar, das Signal enthält daher ein Maximum an Information, bei b) ist eine gute Vorhersage möglich und das Signal ist somit arm an Information. Beide Signale bestehen aus identischen, im Bereich −1 ... 1 gleichverteilten Werten, die bei a) in zufälliger Weise aufeinander folgen, bei b) nach ihrer Größe geordnet sind. Der durch Anzahl der Werte und Wertebereich bestimmte Aufwand zur Signaldarstellung ist in beiden Fällen gleich, der Informationsgehalt jedoch max mal unterschiedlich [Martini: 175, SVO-2]

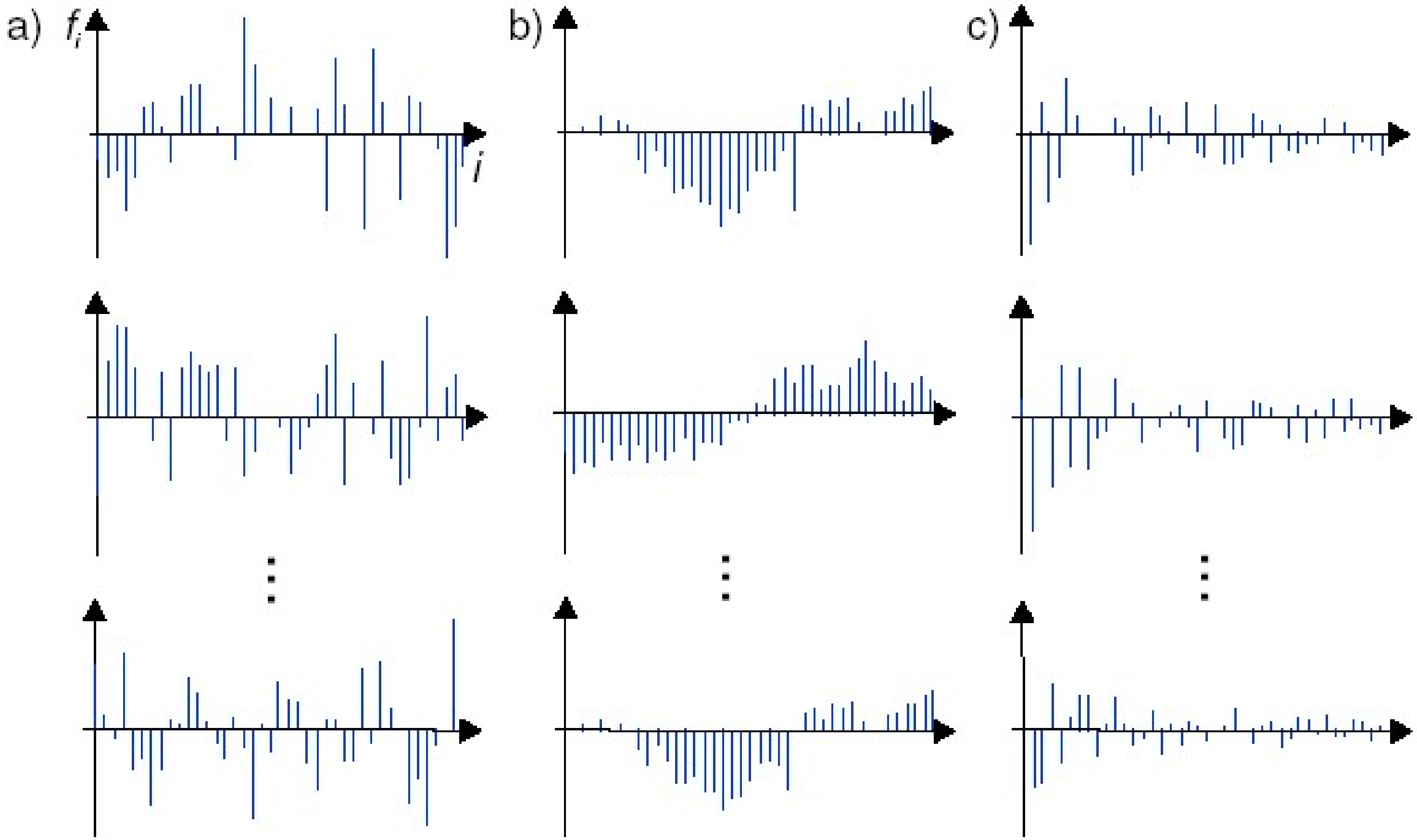


Abbildung 3.11: Drei Episoden typischer diskreter Signalverläufe
a) zufällige, b) langsam veränderliche und c) zufällige und abklingende Signale

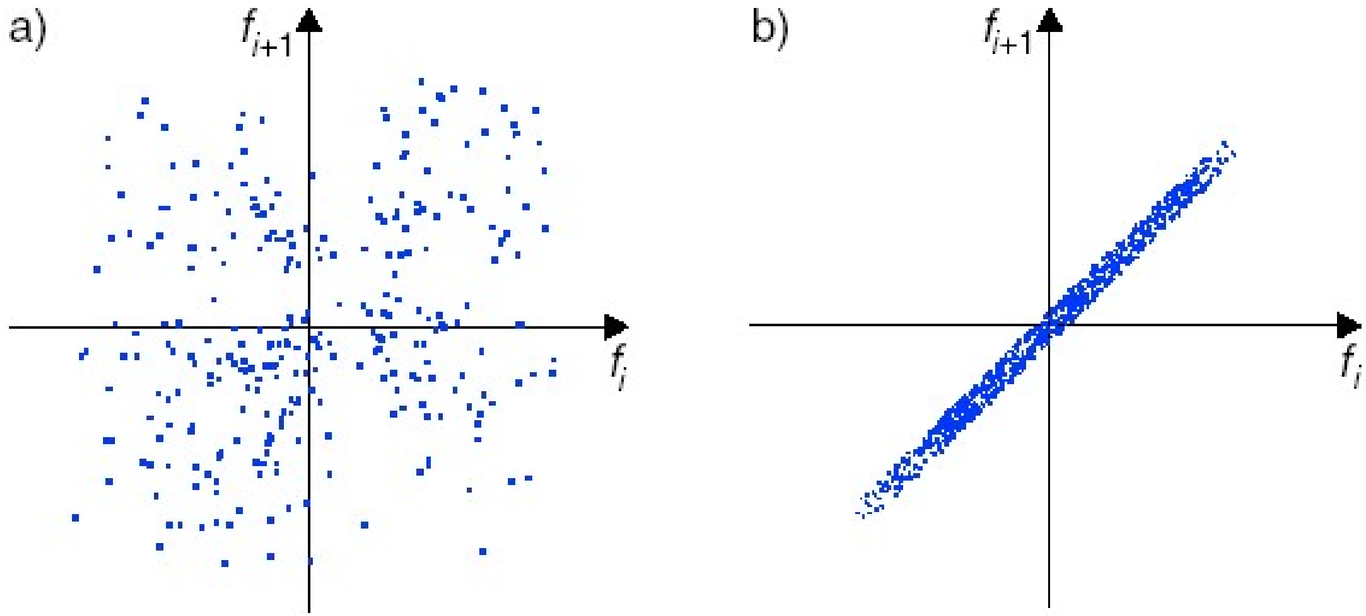
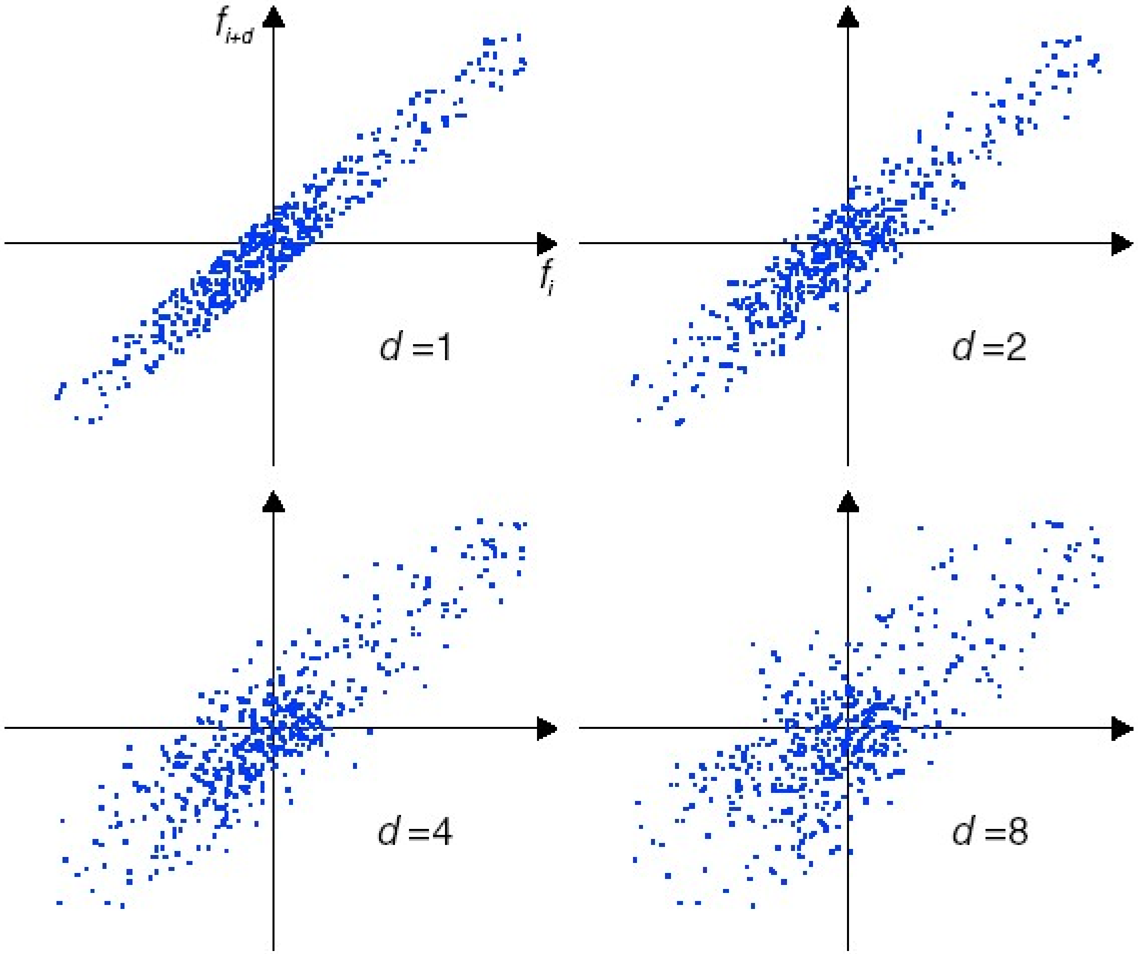
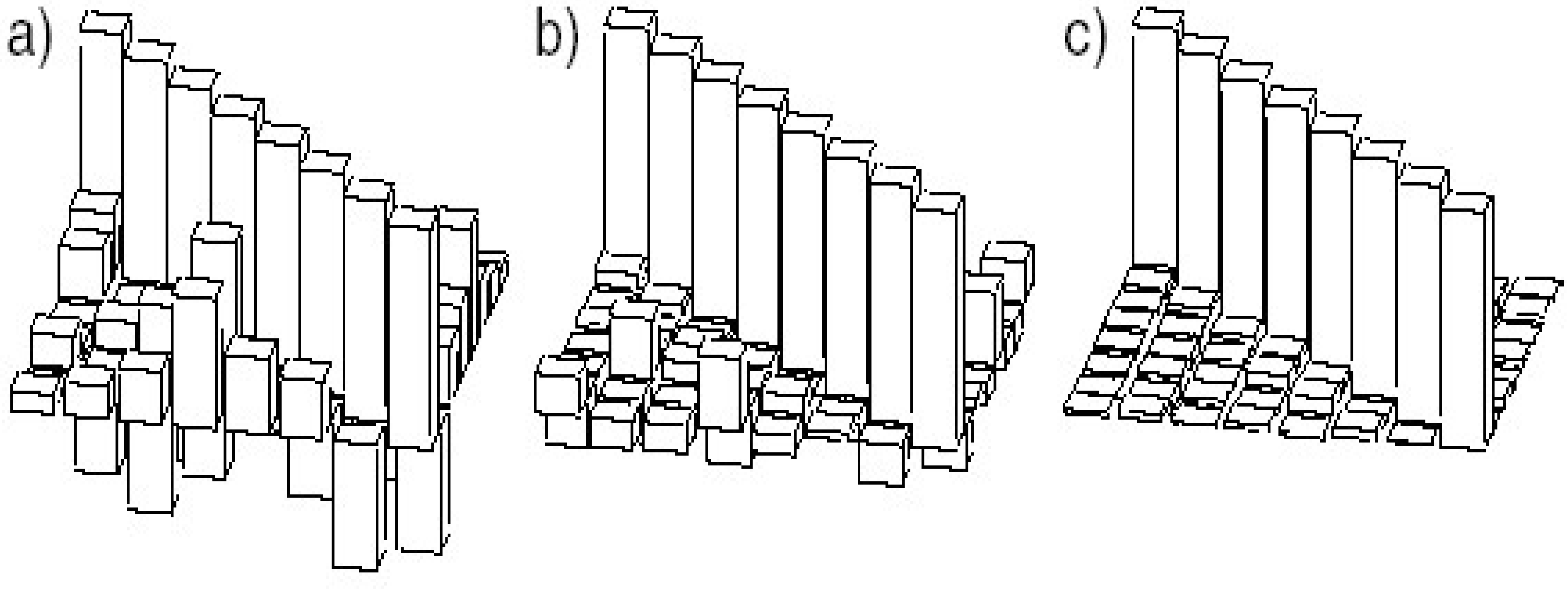


Abbildung 3.12: Statistische Abhängigkeit unmittelbar aufeinanderfolgender Abtastwerte
a) zufälliges Signal aus Abbildung 3.11 a)
b) langsam veränderliches Signal aus Abbildung 3.11 b)

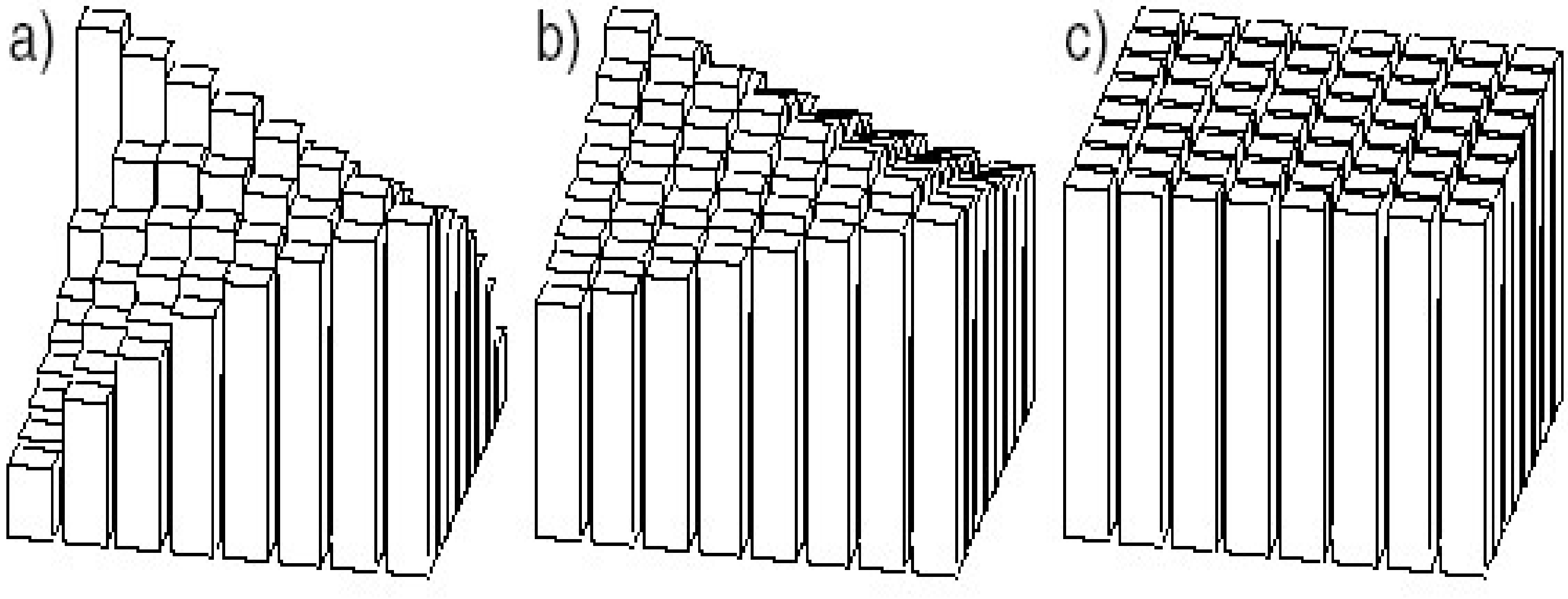


Abhängigkeit zwischen zwei Abtastwerten in einem langsam veränderlichen Signal für verschiedene Abstände *d*



Korrelationsmatrizen eines (richtigen) Zufallssignals

a) 10, b) 50 und c) 1000 Episoden mit je 8 Abtastwerten



Korrelationsmatrizen eines langsam veränderlichen Signals mit von links nach rechts zunehmender Korrelation

$$\begin{bmatrix} \overline{f_i^2} & \mathrm{cor}\left(f_i, f_{i+1}\right) & \cdots & \mathrm{cor}\left(f_i, f_{i+n}\right) \\ \mathrm{cor}\left(f_{i+1}, f_i\right) & \overline{f_{i+1}^2} & \cdots & \mathrm{cor}\left(f_{i+1}, f_{i+n}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cor}\left(f_{i+n}, f_i\right) & \mathrm{cor}\left(f_{i+n}, f_{i+1}\right) & \cdots & \overline{f_{i+n}^2} \end{bmatrix}$$

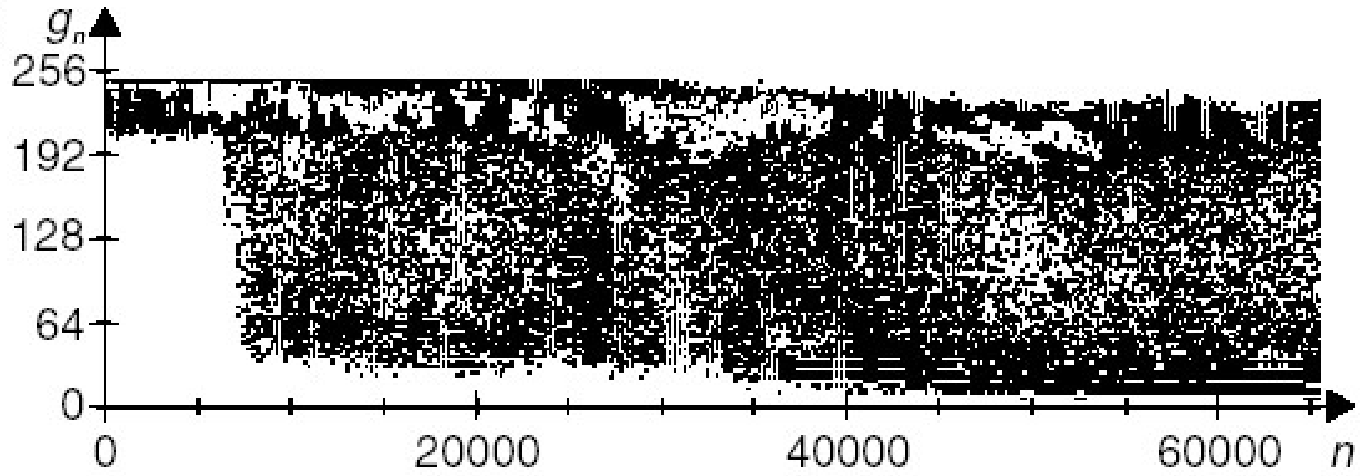
a)

Darstellung der Korrelation benachbarter Grauwerte:

alle Bildzeilen nacheinander in Vektor schreiben und seine Elemente über ihrem Index abtragen

ρ = -0,605

c)



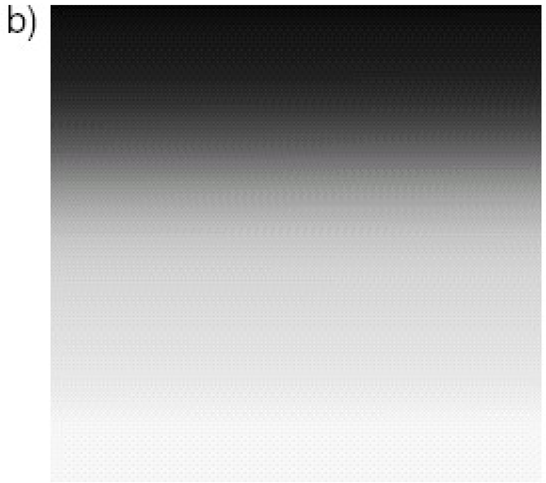
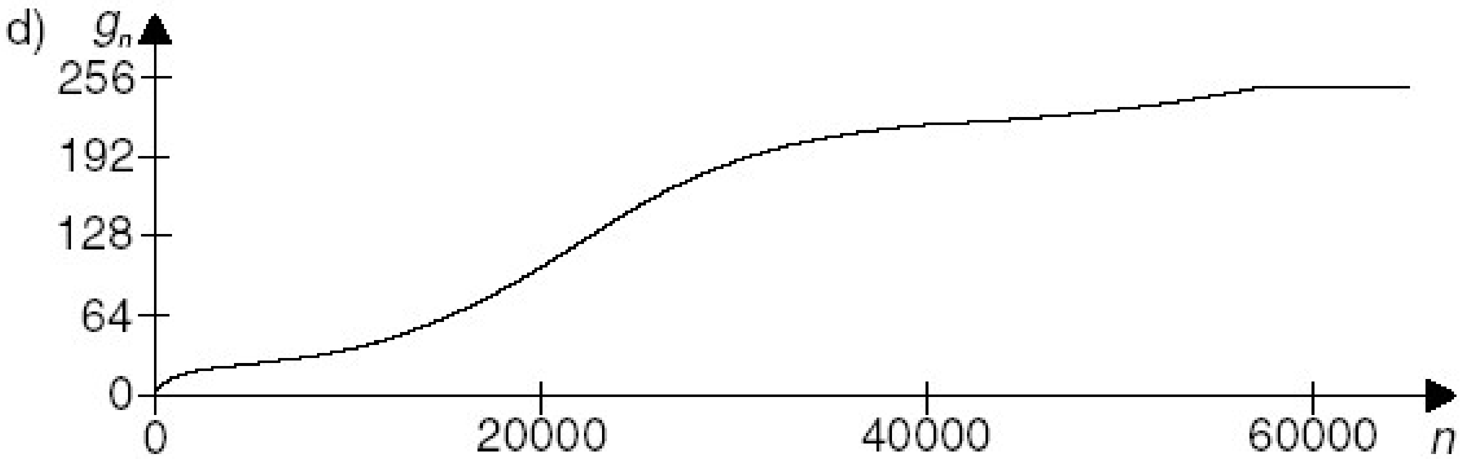


Abbildung enthält dieselben Grauwerte, allerdings von links oben nach rechts unten der Größe nach zu einem Graukeil geordnet

ρ = 0,954



- aufeinanderfolgende Bildpunkte im Graukeil sind viel stärker korreliert
- Informationsgehalt ist entsprechend geringer
- Informationsgehalt ist also hoch, wenn Korrelation gering
- sinnvolles Ziel von signalverarbeitenden Operationen ist demnach Verringerung der Korrelation