Statistische Methoden der Datenanalyse

Wintersemester 2012/2013

Albert-Ludwigs-Universität Freiburg

Prof. Markus Schumacher, Dr. Stan Lai

Physikalisches Institut Westbau 2 OG

E-Mail: Markus.Schumacher@physik.uni-freiburg.de

stan.lai@cern.ch

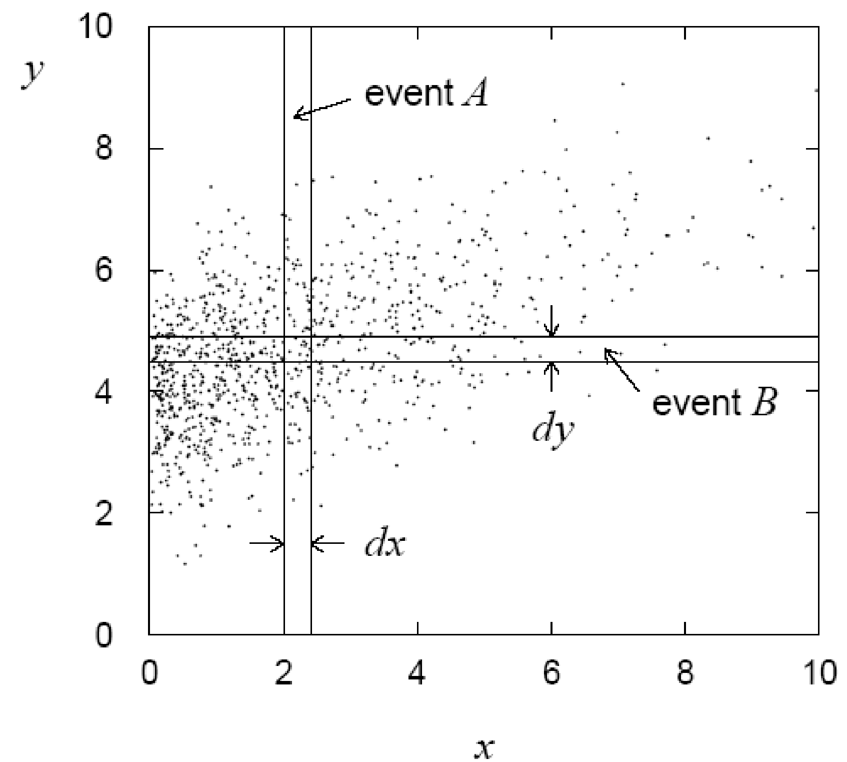
http://terascale.physik.uni-freiburg.de/lehre/ws_1213/statmethoden_ws1213

Kapitel 3

Wahrscheinlichkeitsdichtefunktionen (Fortsetzung)

Multivariate WDF: mehrere Zufallsvariablen

Ergebnis des Experiment wird durch mehrere Observablenwerte charakterisiert, z.B. einen N-komponentigen Vektor: $(x_1, \ldots x_n)$



$$P(A \cap B) = f(x, y)\, dx\, dy$$

Gemeinsame WDF f(x,y)

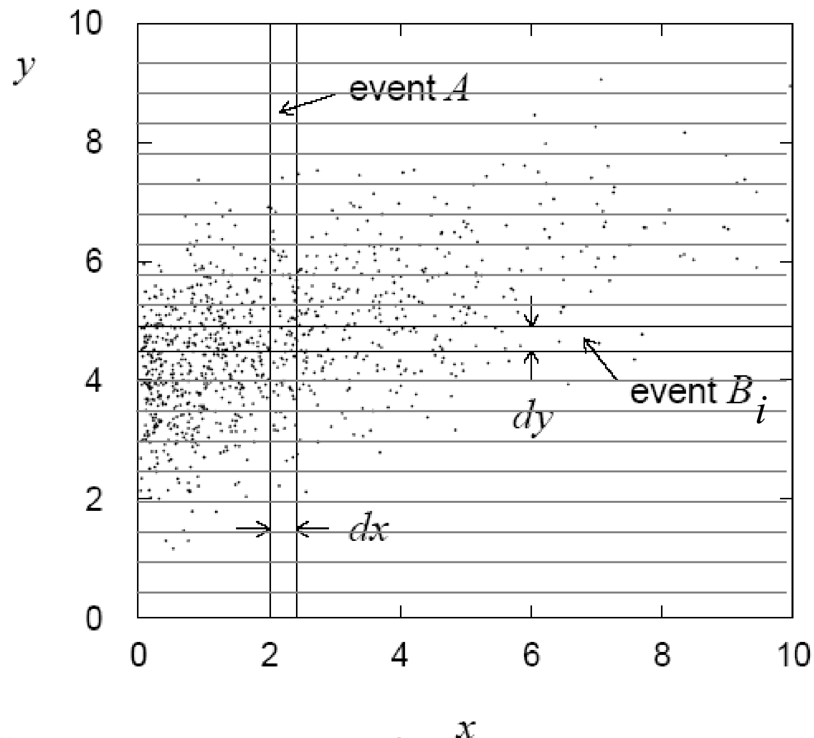
$$f(x, y) dx dy = \text{Wahrscheinlichkeit, dass } x \in [x, x + dx] \text{ und } y \in [y, y + dy]$$

Normierung: $$\int \cdots \int f(x_1, \ldots, x_n) dx_1 \cdots dx_n = 1$$

Marginal-WDF oder Randverteilung

Manschmal interessiert uns nur WDF für einige oder eine der Zuvallsvariablen:

$$
\begin{aligned}
P(A) &= \sum_i P(A \cap B_i) \\
&= \sum_i f(x, y_i)\, dy\, dx \\
&\to \int f(x, y)\, dy\, dx
\end{aligned}
$$



Integration über uninteressante ZV (“Marginalisieren”)

→ Marginal-WDF oder Randverteilung

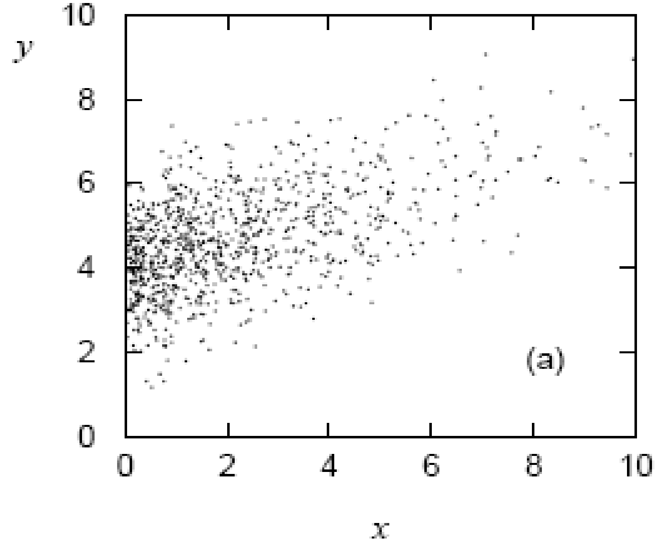
$$f_1(x_1) = \int \cdots \int f(x_1, \ldots, x_n)\, dx_2 \ldots dx_n$$

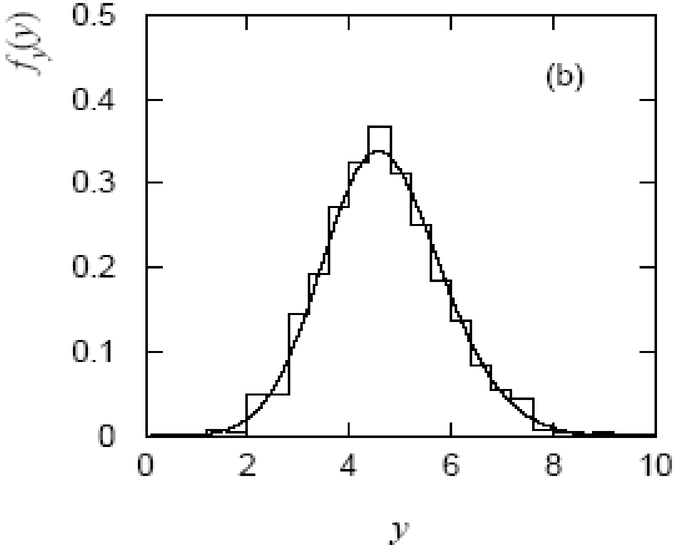
$$f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

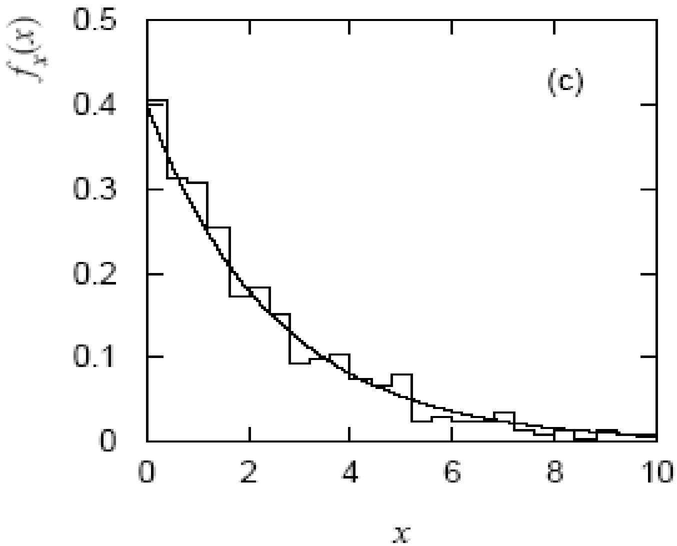
$$f_y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

x_1, x_2 sind unabhängig, genau dann wenn gilt $f(x_1, x_2) = f_1(x_1) f_2(x_2)$

Marginal-WDF oder Randverteilung







Marginal-WDF oder Randverteilung:
aus Projektion der gemeinsamen
WDF auf eine der ZV-Achsen

Bedingte WDF

Manchmal soll eine der Zufallsvariablen in der gemeinsamen WDF konstant gehalten werden.

Erinnerung an bedingte Wahrscheinlichkeit:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{f(x,y)\,dx\,dy}{f_x(x)\,dx}$$

$\rightarrow$ bedingte WDF: $$h(y|x) = \frac{f(x,y)}{f_x(x)}\,, \qquad g(x|y) = \frac{f(x,y)}{f_y(y)}$$

Bayes-Theorem wird zu: $$g(x|y) = \frac{h(y|x) f_x(x)}{f_y(y)}\,.$$

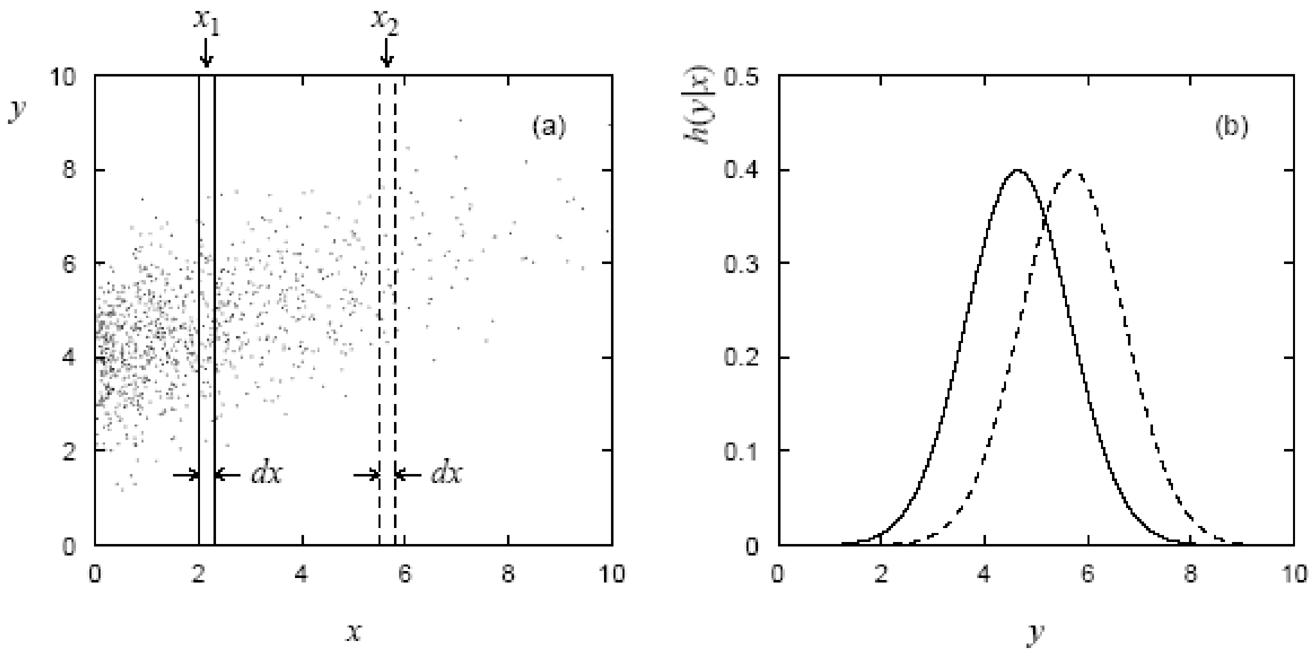
Erinnerung: *A*, *B* unabhängig, wenn $P(A \cap B) = P(A)P(B)\,.$

$\rightarrow$ *x*, *y* unabhängig wenn $f(x,y) = f_x(x) f_y(y)\,.$

Bedingte WDF (Fortsetzung)

Beispiel: : aus gemeinsamer WDF $f(x,y)$ werden

bedingte WDFs $h(y|x_1)$, $h(y|x_2)$ bestimmt

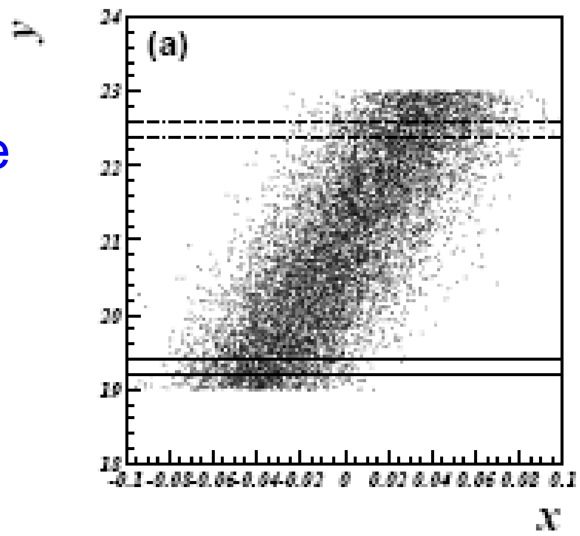


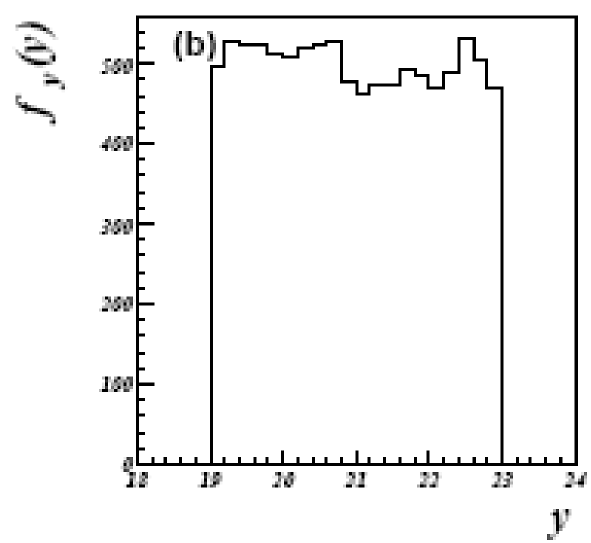
$$\int h(y|x)\,dy = 1\ .$$

- behandele eine der ZV als konstant in gemeinsamer WDF
- teile die gemeinsame WDF durch die Margnial-WDF der Variablen, die konstant gehalten werden
- was ist übrig bleibt ist bedingte WDF mit korrekter Normierung

Marginal- und Bedingte-WDF

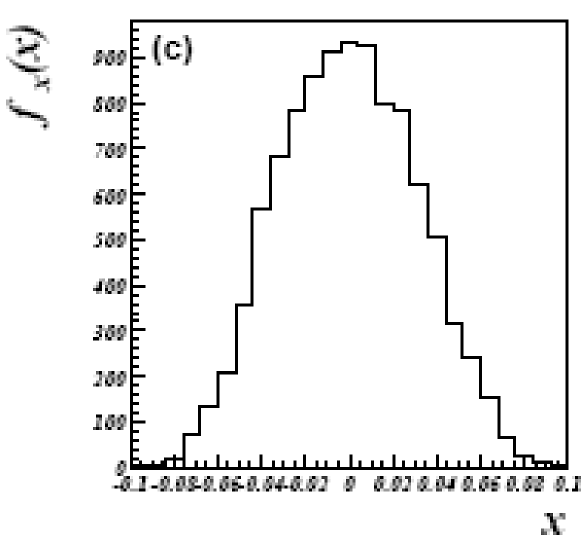
2-dim.
gemeinsame
WDF

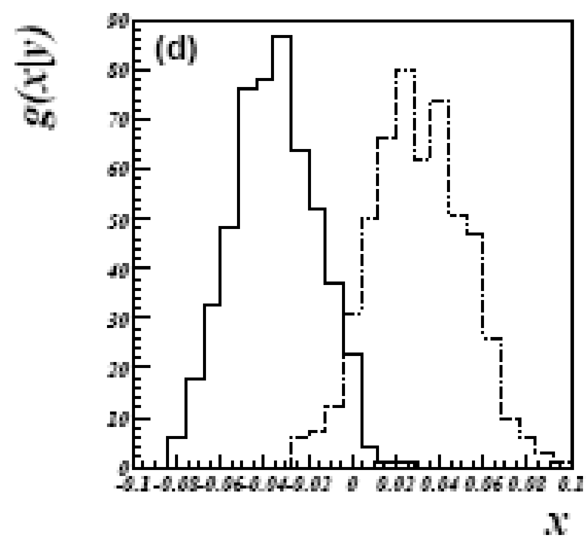




Rand-
verteilung
für y

Rand-
verteilung
für x





Bedingte
WDFs
für 2
Y-Werte

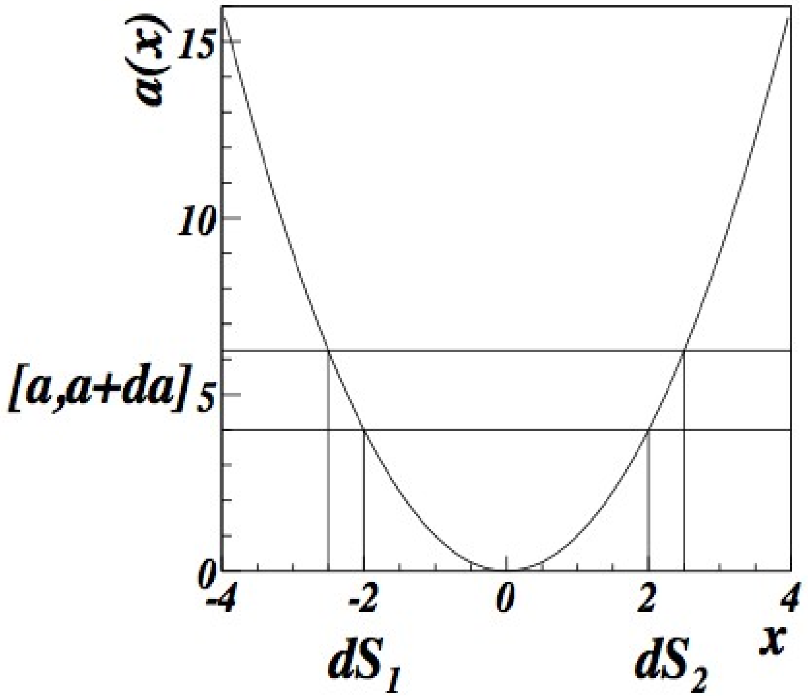
Funktionen / Transformation von Zufallsvariablen

Ein Funktion *a(x)* von einer ZV *x* ist selbst ebenfalls eine Zufallsvariable.

Annahme: ZV *x* folgt einer WDF *f*(*x*)

Betrachte Abhängigkeit/Funktion: *x*→ *a*(*x*).

Was ist die WDF für die ZV *g*(*a*)?



Sinnvolle Bedingung:

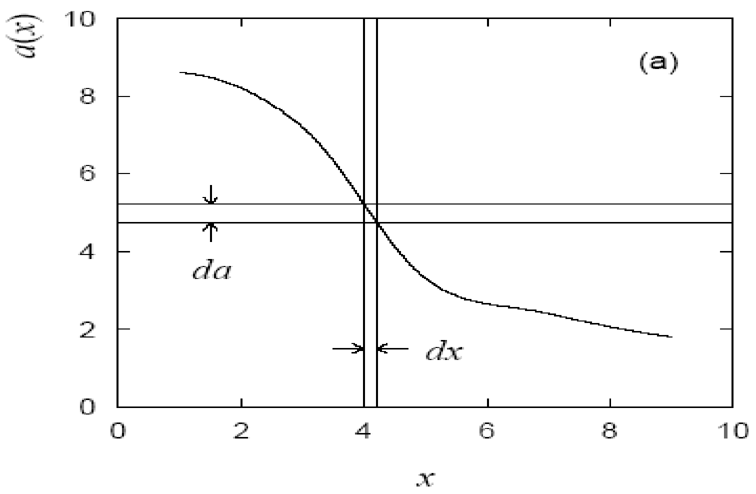
$$g(a)\,da = \int_{dS} f(x)\,dx$$

dS = Region im Raum der *x*-Werte für die *a* in [*a*, *a*+*da*] liegt.

Transformation von Zufallsvariablen (II)

i) a(x) eindeutig umkehrbar

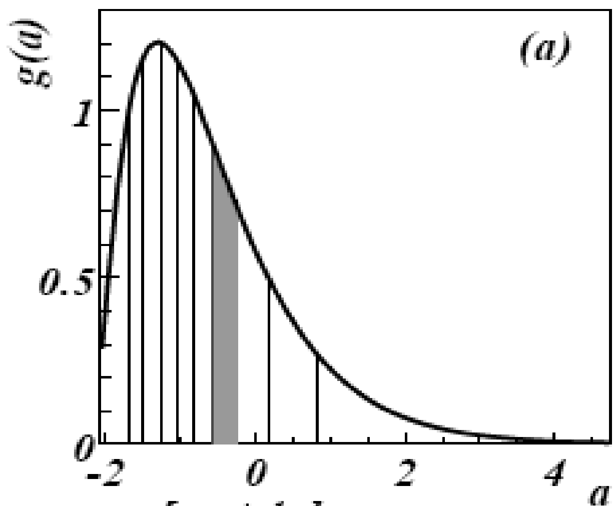
Forderung:
Wahrscheinlichkeit, dass
a in $[a, a + da]$
= Wahrscheinlichkeit,
dass x in $[x(a), x(a + da)]$

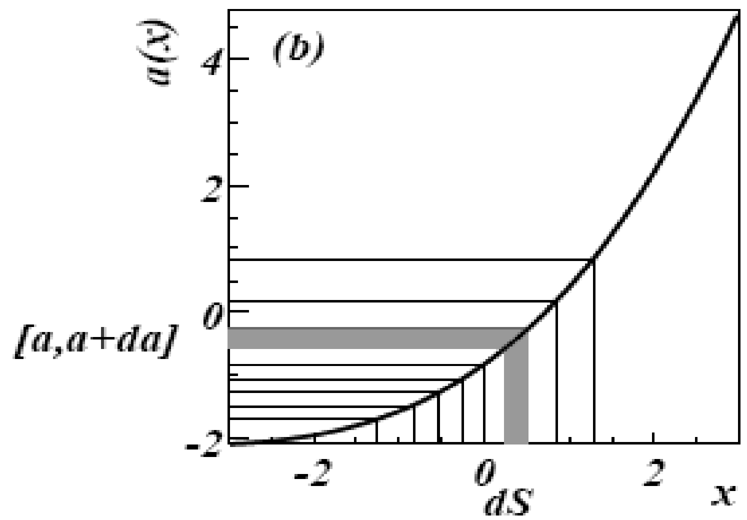


Falls gilt $x(a) > x(a + da)$ wäre Integral < 0 → Betragsstriche

$$g(a)da = \left| \int_{x(a)}^{x(a+da)} f(x')dx' \right| = \int_{x(a)}^{x(a)+\left|\frac{dx}{da}\right|da} f(x')dx' = f(x(a)) \left|\frac{dx}{da}\right| da$$

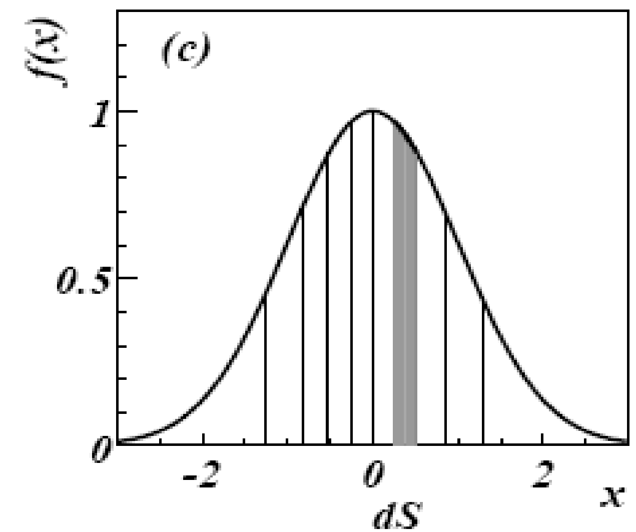
Also gilt zusammengefasst: $$g(a) = f(x(a)) \left|\frac{dx}{da}\right|$$



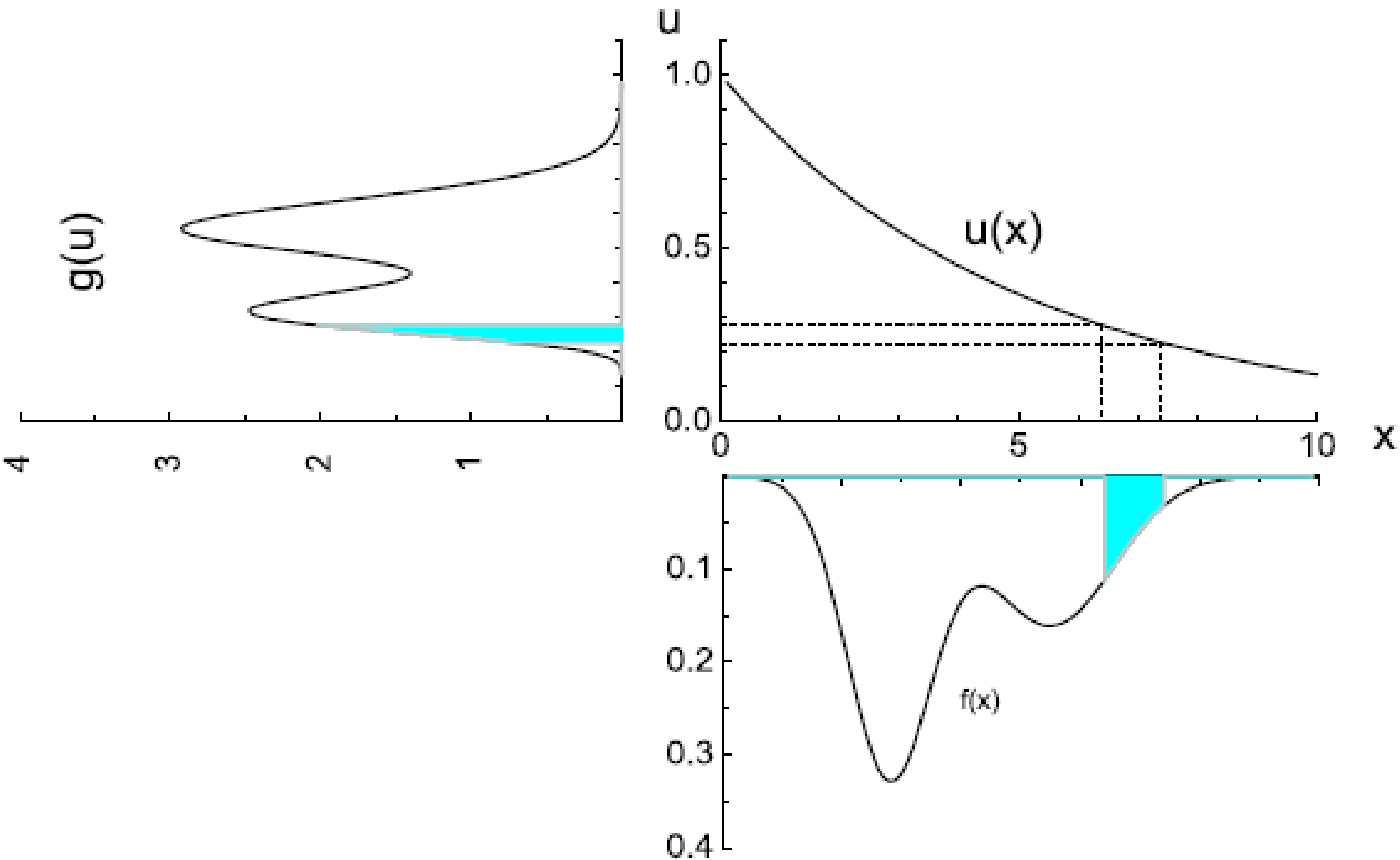


Wahrscheinlichkeit im a-Raum g(a)da

Entspricht der im x-Raum f(x)dS



Transformation von Zufallsvariablen (IV)



Analytisches Beispiel

$$g(a) = f(x(a)) \left| \frac{dx}{da} \right|$$

Transformation / Funktion: $a(x) = \exp(x)$

Wktdichtefuntion für ZV x: $f(x) = \frac{1}{2}x^2, x \text{ aus } [0, 2]$

Eindeutige Umkehrfuntion: $x(a) = \ln(a)$

Jakobifaktor / Differential: $\frac{dx}{da} = \frac{1}{a} > 0$

Und alles zusammengesetzt ….

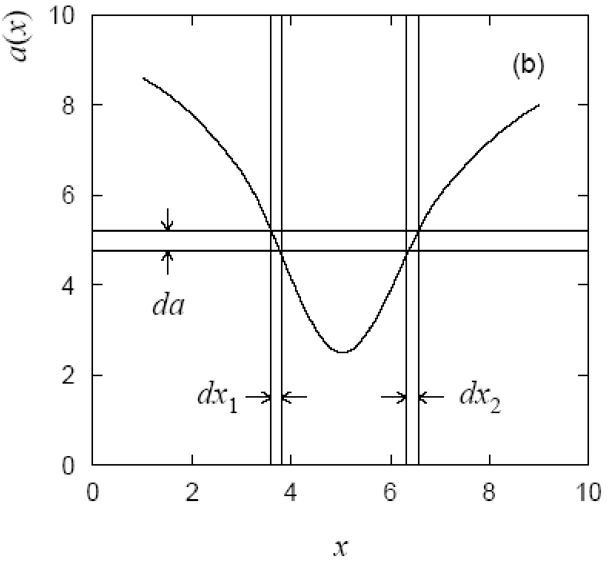
$$g(a) = \frac{1}{2}(\ln(a))^2 \frac{1}{a}$$

oder kompakt

$$g(a) = \frac{\ln^2(a)}{2a}$$

Funktionen ohne eindeutige Umkehrung

Wenn es keine eindeutige Inverse von *a*(*x*) gibt, dann summiere über alle *dx*-Intervalle in *dS,* die nach *da* abgebildet werden



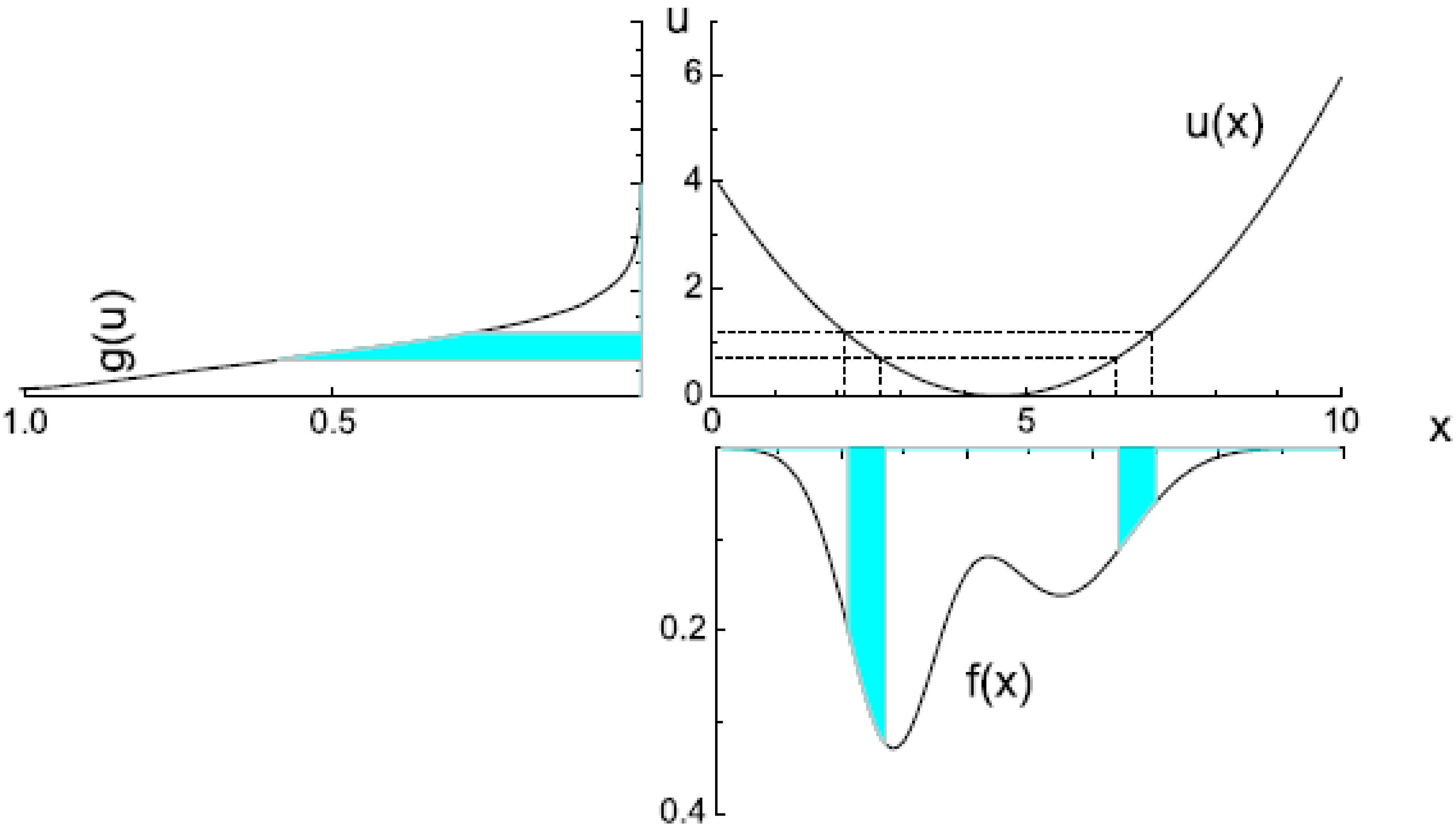
Beispiel:

$$a = x^2,\ \ x = \pm\sqrt{a},\ \ dx = \pm\frac{da}{2\sqrt{a}}\ .$$

$$dS = \left[\sqrt{a}, \sqrt{a} + \frac{da}{2\sqrt{a}}\right] \cup \left[-\sqrt{a} - \frac{da}{2\sqrt{a}}, -\sqrt{a}\right]$$

$$g(a) = \frac{f(\sqrt{a})}{2\sqrt{a}} + \frac{f(-\sqrt{a})}{2\sqrt{a}}$$

Transformation von ZV: Mehrdeutige Inverse



Funktionen von mehreren Zufallsvariablen

Betrachte Zufallsvariablen $\vec{x} = (x_1, \ldots, x_n)$

Und eine Funktion $a(\vec{x})$.

Die Bedingung der Gleichheit von Wahrscheinlichkeiten lautet:

$$g(a')da' = \int \ldots \int_{dS} f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

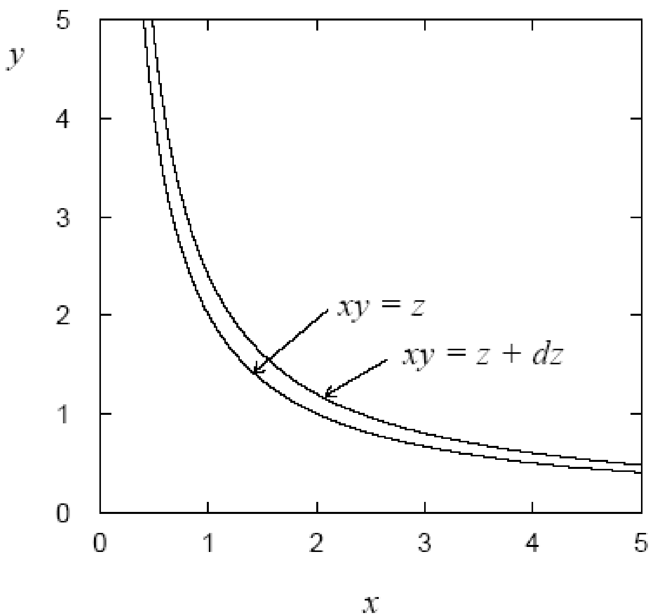
Wobei *dS* = Region im *x*-Raum zwischen den Hyperflächen,
Die definiert werden durch

$$a(\vec{x}) = a', \; a(\vec{x}) = a' + da'$$

Funktionen von mehreren ZV: Mellin-Faltung

Beispiel: ZV x, $y > 0$ mit gemeinsamer WDF $f(x,y)$,

Betrachte Funktion $z = xy$. Was ist $g(z)$?



$$g(z)\,dz = \int \ldots \int_{dS} f(x,y)\,dx\,dy = \int_0^{\infty} dx \int_{z/x}^{(z+dz)/x} f(x,y)\,dy$$

$$\rightarrow \quad g(z) = \int_0^{\infty} f\left(x, \frac{z}{x}\right) \frac{dx}{x} = \int_0^{\infty} f\left(\frac{z}{y}, y\right) \frac{dy}{y}$$

(Mellin-Faltung)

Funktionen von mehreren ZV: Fourier-Faltung

Beispiel: 2 unabhängige ZV x, y mit gemeinsamer WDF $f_x(x)\ f_y(y)$

Betrachte Funktion $a = x + y$. Was ist $g(a)$?

Anwendung: x Meßgröße , y Auflösungsvermögen des Messinstrumentes

$$g(a)da = \int_{-\infty}^{\infty} \int_{(a-y)}^{(a-y+da)} f_X(x) f_Y(y) dx dy$$

$$= \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) da dy$$

$$\Rightarrow \quad g(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy$$

(Fourier-Faltung)

Mehr zu Variablentransformationen

Betrachte eine Vektor von ZV $\vec{x} = (x_1, \ldots, x_n)$ mit gemeinsamer WDF $f(\vec{x})$.

Bilde *n* linear unabhängige Funktionen $\vec{y}(\vec{x}) = (y_1(\vec{x}), \ldots, y_n(\vec{x}))$

Für die, die inversen Funktionen existieren $x_1(\vec{y}), \ldots, x_n(\vec{y})$

Dann ist die gemeinsame WDF für den Vektor der y-Funktionen gegeben durch

$$g(\vec{y}) = |J| f(\vec{x})$$

wobei *J* die Jacobi-Determinante ist

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \ldots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \ldots & \frac{\partial x_2}{\partial y_n} \\ \vdots & & & \vdots \\ & & \ldots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

Um z.B.. $g_1(y_1)$ zu erhalten, integriere (marginalisiere) über andere y in $g(\vec{y})$

Kovarianz und Korrelation

Definiere Kovarianz cov[*x*,*y*] (oft auch Matrixnotation V_{xy}) als

$$\text{cov}[x, y] = E[xy] - \mu_x \mu_y = E[(x - \mu_x)(y - \mu_y)]$$

Korrelationskoeffizient (dimensionlos) definiert als

$$\rho_{xy} = \frac{\text{cov}[x, y]}{\sigma_x \sigma_y}$$

Wenn *x*, *y*, unabhängig, $f(x, y) = f_x(x) f_y(y)$, dann gilt

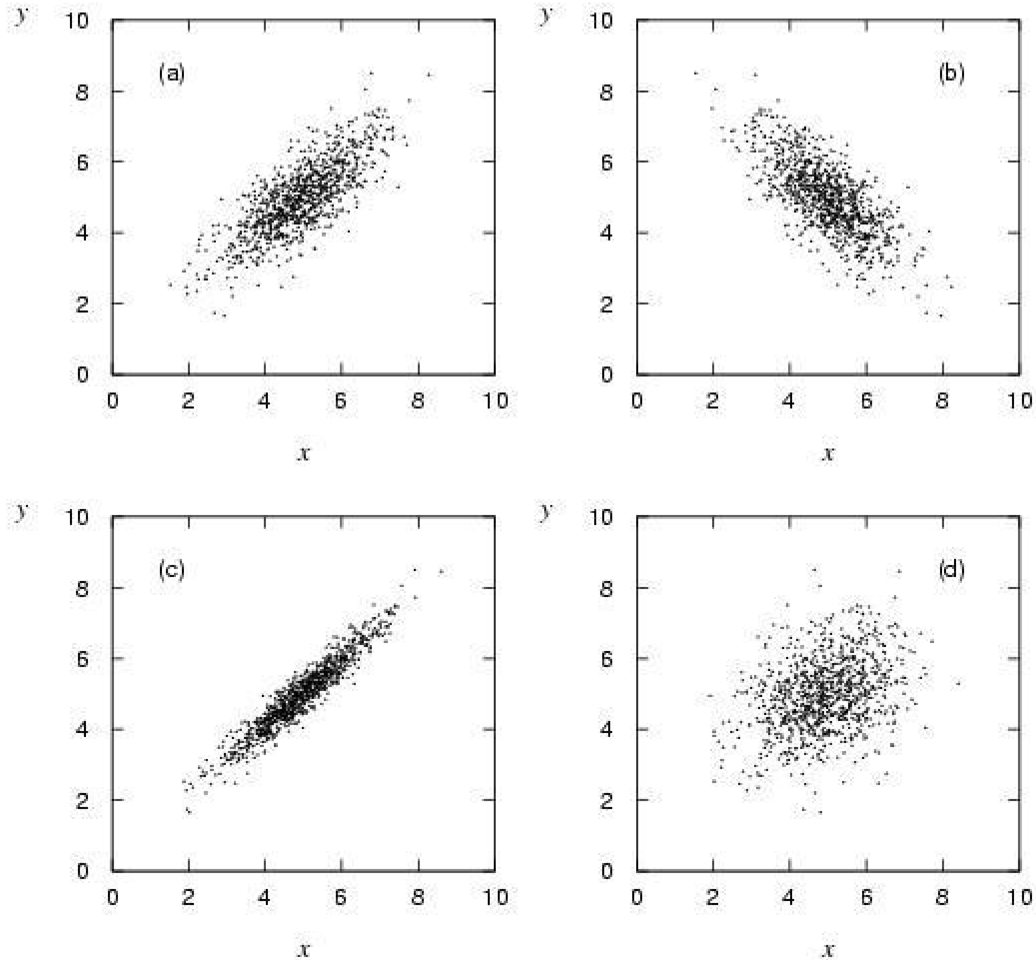
$$E[xy] = \int\int xy\, f(x, y)\, dx dy = \mu_x \mu_y$$

→ $\text{cov}[x, y] = 0$ *x* und *y* sind 'unkorreliert'

Bemerkung: Umkehrschluss i. a. nicht wahr.

Korrelation (Fortsetzung)

$\rho = 0.75$



$\rho = -0.75$

$\rho = 0.95$

$\rho = 0.25$

Fehlerfortpflanzung

Annahme: wir haben eine Stichprobe der ZV *x* $\vec{x} = (x_1, \ldots, x_n)$

und wir kennen/schätzen die Kovarianzen

die die Messfehler in the x_i quantifizieren $V_{ij} = \mathrm{cov}[x_i, x_j]$

Nun betrachten wir ein Funktion $y(\vec{x})$.

Frage: was ist die Varianz von $y(\vec{x})$?

a) WDF $f(\vec{x})$ unbekannt
b) WDF $f(\vec{x})$ bekannt → "orthodoxe" Weg:
nutze gemeinsame WDF $f(\vec{x})$ um die WDF $g(y)$, zu bestimmen.
Dann bestimmte aus $g(y)$ die Varianz $V[y] = E[y^2] - (E[y])^2$.

Oft nicht praktikabel, $f(\vec{x})$ nicht vollständig bekannt.

Fehlerfortpflanzung (II)

Annahme: wir kennen: $\vec{\mu} = E[\vec{x}]$

In der Praxis nur Schätzwert aus der Stichprobe $\vec{x}$

Entwickle $y(\vec{x})$ in einer Taylor-Serie bis zum Glied erster Ordnung um $\vec{\mu}$

$$y(\vec{x}) \approx y(\vec{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i)$$

Um die Varianz $V[y]$ zu finden, brauchen wir $E[y^2]$ and $E[y]$.

$E[y(\vec{x})] \approx y(\vec{\mu})$ weil gilt $E[x_i - \mu_i] = 0$

Fehlerfortpflanzung (III)

$$E[y^2(\vec{x})] \approx y^2(\vec{\mu}) \; + \; 2y(\vec{\mu}) \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}} E[x_i - \mu_i]$$

$$+E\left[\left(\sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i)\right) \left(\sum_{j=1}^{n} \left[\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} (x_j - \mu_j)\right)\right]$$

$$= y^2(\vec{\mu}) + \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} V_{ij}$$

Zusammenfügen der Teilergebnisse liefert Varianz für $y(\vec{x})$

$$\boxed{\sigma_y^2 \approx \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right]_{\vec{x}=\vec{\mu}} V_{ij}}$$

$$V_y = \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \frac{\partial y}{\partial x_3} \\ \ldots \\ \frac{\delta y}{\delta x_n} \end{pmatrix}^T V_x \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \frac{\partial y}{\partial x_3} \\ \ldots \\ \frac{\partial y}{\partial x_n} \end{pmatrix} = (\vec{\nabla} y)^t V_x \vec{\nabla} y$$

Fehlerfortpflanzung (IV)

Wenn die x_i unkorelliert sind, i.e. $V_{ij} = \sigma_i^2 \delta_{ij}$, dann wird daraus

$$\sigma_y^2 \approx \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i} \right]^2_{\vec{x}=\vec{\mu}} \sigma_i^2$$

Ähnlich gilt für einen Satz von Funktionen $\vec{y}(\vec{x}) = (y_1(\vec{x}), \ldots, y_m(\vec{x}))$

$$\boxed{U_{kl} = \mathrm{cov}[y_k, y_l] \approx \sum_{i,j=1}^{n} \left[\frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij}}$$

Oder in Matrixnotation $U = AVA^T$, wobei

$$A_{ij} = \left[\frac{\partial y_i}{\partial x_j} \right]_{\vec{x}=\vec{\mu}}$$

Fehlerfortpflanzung (V): Beispiel 1

Betrachte Stichprobe der ZV x_i vom Umfang n.

Die unkorellierten Messungen haben Messfehler: $\sigma^2 = \sigma_i^2$

Die Funktion $y(x_i)$ ist der Stichprobenmittelwert: $y(\vec{x}) = \bar{x} = \frac{1}{n}\sum_i^n x_i$

Was ist der Fehler auf den Stichprobenmittelwert?

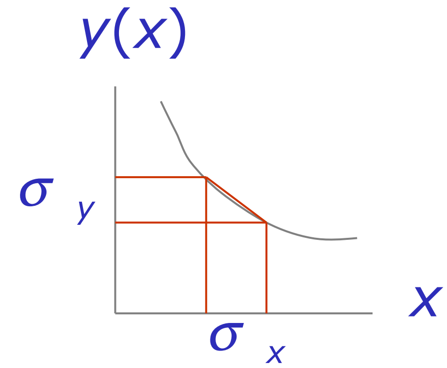
Partielle Ableitungen: $\frac{\partial y}{\partial x_i} = \frac{1}{n}$

$$V(y) = V(\bar{x}) = \sum_i^n \left(\frac{\partial y}{\partial x_i}\right)^2 \sigma^2 = \frac{n}{n}\sigma^2 = \frac{\sigma^2}{n}$$

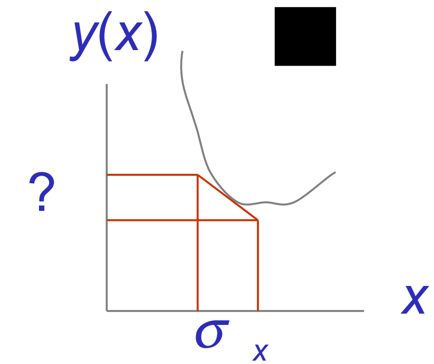
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Fehlerfortpflanzung (VI)

Die ‘Fehlerfortpflanzungsformel’ sagt uns, wie Wir die Kovarianzen eines Satzes von Funktionen $\vec{y}(\vec{x}) = (y_1(\vec{x}), \ldots, y_m(\vec{x}))$ durch die Kovarianzen der ursprünglichen Variablen approximativ auswerten können.



Limitierung: exakt nur wenn $\vec{y}(\vec{x})$ linear.



Näherung bricht zusammen wenn die Funktion nicht linear ist in einem Bereich der Vergleichbar ist mit der Größe von σ_x.

Bemerkung: keine Annahme über die WDF von x_i,

i.e., sie muss nicht die Gaußverteilung sein

Fehlerfortpflanzung (VII) − Spezialfall

$$y = x_1 + x_2 \quad \rightarrow \quad \sigma_y^2 = \sigma_1^2 + \sigma_2^2 + 2\mathrm{cov}[x_1, x_2]$$

$$y = x_1 x_2 \quad \rightarrow \quad \frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} + 2\frac{\mathrm{cov}[x_1, x_2]}{x_1 x_2}$$

wenn die x_i unkorreliert sind, gilt

addiere Fehler quadratisch für die Summe (oder Differenz),

addiere relative Fehler quadratisch für Produkt (oder Quotient).

Aber Korrelationen können dies komplett ändern…

Fehlerfortpflanzung (VIII) – Spezialfall (2)

Betrachte $y = x_1 - x_2$ mit

$$\mu_1 = \mu_2 = 10, \quad \sigma_1 = \sigma_2 = 1, \quad \rho = \frac{\mathrm{cov}[x_1, x_2]}{\sigma_1 \sigma_2} = 0\,.$$

$$V[y] = 1^2 + 1^2 = 2, \rightarrow \sigma_y = 1.4$$

Nur nehmen wir an, dass $\rho = 1$. Dann gilt

$$V[y] = 1^2 + 1^2 - 2 = 0, \rightarrow \sigma_y = 0$$

i.e. für 100% Korrelation verschwindet der Fehler in der Differenz.

Analog: für die Summe werden die Fehler linear addiert

Fehlerfortpflanzung (IX): Beispiel 2

Betrachte Messung einer Spur/Trajektorie in Zylinderkkoordinaten r, φ, z

Die Messfehler seinen unkorrelliert und es gelte $\sigma_r = 0, \sigma_\varphi, \sigma_z$

Die Kovarianzmatrix lautet dann:

$$V_{r,\varphi,z} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sigma_\varphi^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}$$

Wie sehen die Fehler in karthesischen Koordinaten aus?

$$x = r\cos\varphi$$

$$y = r\sin\varphi$$

$$z = z$$

Fehlerfortpflanzung (X): Beispiel 2

Die Jakobi-Matrix der Ableitungen lautet:

$$A = \begin{pmatrix} \cos\varphi & -r\sin\varphi & 0 \\ \sin\varphi & r\cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Die Kovarianzmatrix in x,y, z lautet:

$$V_{x,y,z} = A V_{r,\varphi,z} A^T = \begin{pmatrix} \sigma_\varphi^2 y^2 & -\sigma_\varphi^2 xy & 0 \\ -\sigma_\varphi^2 xy & \sigma_\varphi^2 x^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}$$

Bem.: - Fehler auf z unverändert
- Fehler in x,y σ_x, σ_y abhängig von x,y korreliert

Die Korrelation ergibt sich zu: $$\rho_{xy} = -\frac{\sigma_\varphi^2\, x\, y}{\sigma_\varphi\, x\, \sigma_\varphi\, y} = -1$$

Multinomialverteilung

Wie Binomial aber nun *m* verschiedene Ausgänge des anstelle von zwei,
Mit Wahrscheinlickieten für die einzelnen Ausgänge:

$$\vec{p} = (p_1, \ldots, p_m)\,, \quad \text{with} \quad \sum_{i=1}^{m} p_i = 1\,.$$

Für *N* Versuche suchen wir die Wahrscheinlichkeit das folgende Ergebnis zu erhalten:

n_1 von Möglichkeit 1,
n_2 von Möglichkeit 2,
...
n_m von Möglichkeit *m*.

Die Wktdichtefunktion ist die Multinomialverteilung für $\vec{n} = (n_1, \ldots, n_m)$

$$f(\vec{n}; N, \vec{p}) = \frac{N!}{n_1! n_2! \cdots n_m!} p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}$$

Multinomialverteilung II

Nun betrachte Möglichkeit *i* as ‘Erfolg’, alle anderen als “Misserfolg”.

→ alle individuellen n_i binomialverteilt mit Parametern N, p_i

$$E[n_i] = Np_i, \quad V[n_i] = Np_i(1-p_i) \quad \text{für alle } i$$

Für die Kovarianzen (i ungleich j) ergibt sich:

$$\begin{aligned}
\mathrm{cov}[n_i, n_j] &= E[(n_i - E[n_i])(n_j - E[n_j])] \\
&= E[n_i n_j] - E[n_i]E[n_j] \\
&= N(N-1)p_i p_j - (Np_i)(Np_j) \\
&= -Np_i p_j .
\end{aligned}$$

Negative Korrelation:
wenn in einer Klasse mehr
Ereignisse müssen
irgendwo welche fehlen

$$\rho_{n_i n_j} = \frac{\mathrm{cov}[n_i, n_j]}{\sigma_{n_i}\sigma_{n_j}} = -\sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}}.$$

Multinomialverteilung (III): Beispiele

$\vec{n} = (n_1, \ldots, n_m)$ representiert ein Histogram

mit *m* bins, *N* (bekannt) Gesamteinträgen, alle Einträge unabhängig.

Zerfall eines instabilen Teilchens A: (1) A→BC, (2) A→DE, (3) A→FG
Mit bekannten Zerfallswahrscheinlichkeiten für alle drei Modi.
(nBC, nDE, nfg) folgte Multinomialverteilung.

Textanalyse: Auftreten der einzelnen Buchstaben des Alphabets.
(nA, …,nZ) folgt Multinomialverteilung.