Statistische Methoden der Datenanalyse

Prof. Markus Schumacher

ALU Freiburg, Wintersemsester 2009/2010

BOK-Veranstaltung im Rahmen des ZfS

- Kapitel 1: Grundlegende Konzepte der Wahrscheinlichkeitstheorie

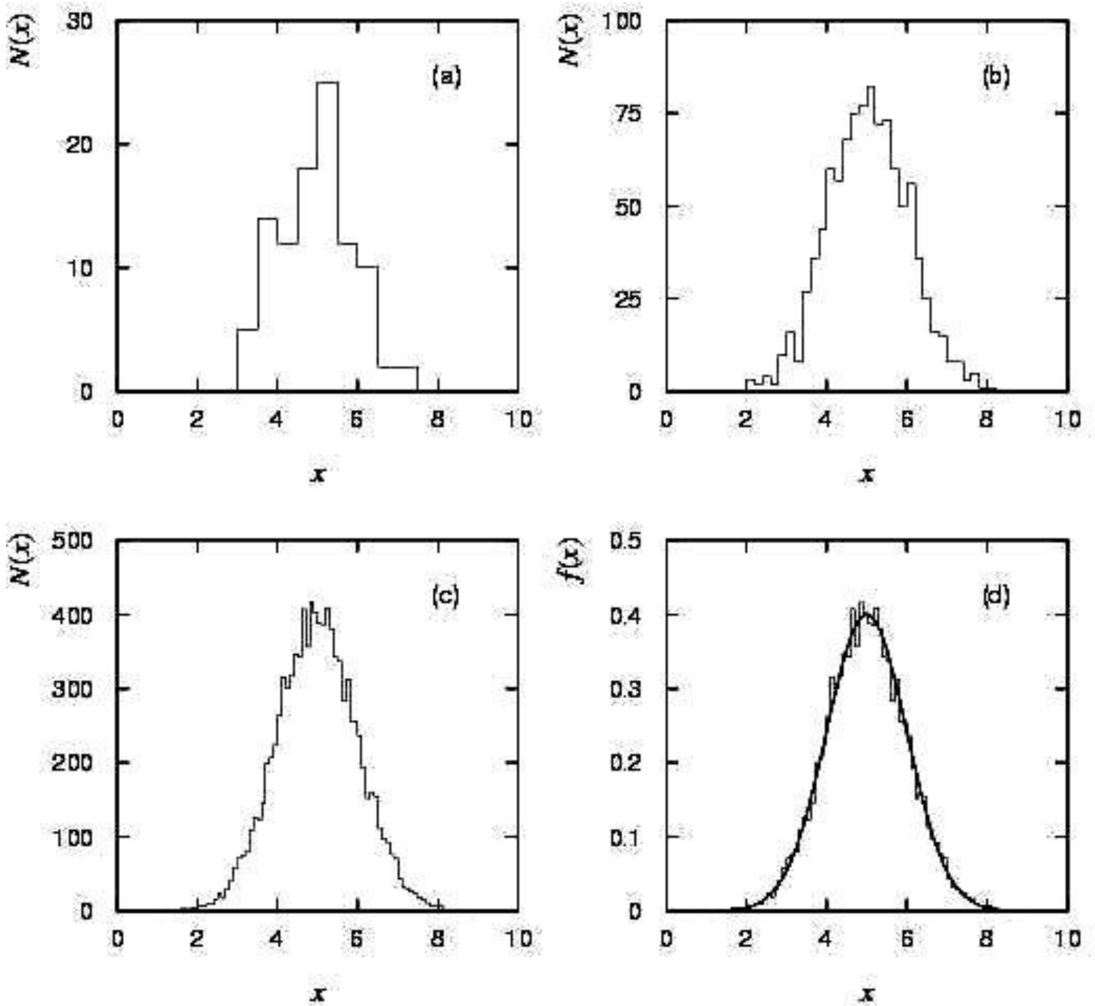
WDF = Histogram mit

unendlichen Datensatz,
verschindender Binbreite,
normier auf eins.

$$f(x) = \frac{N(x)}{n\Delta x}$$

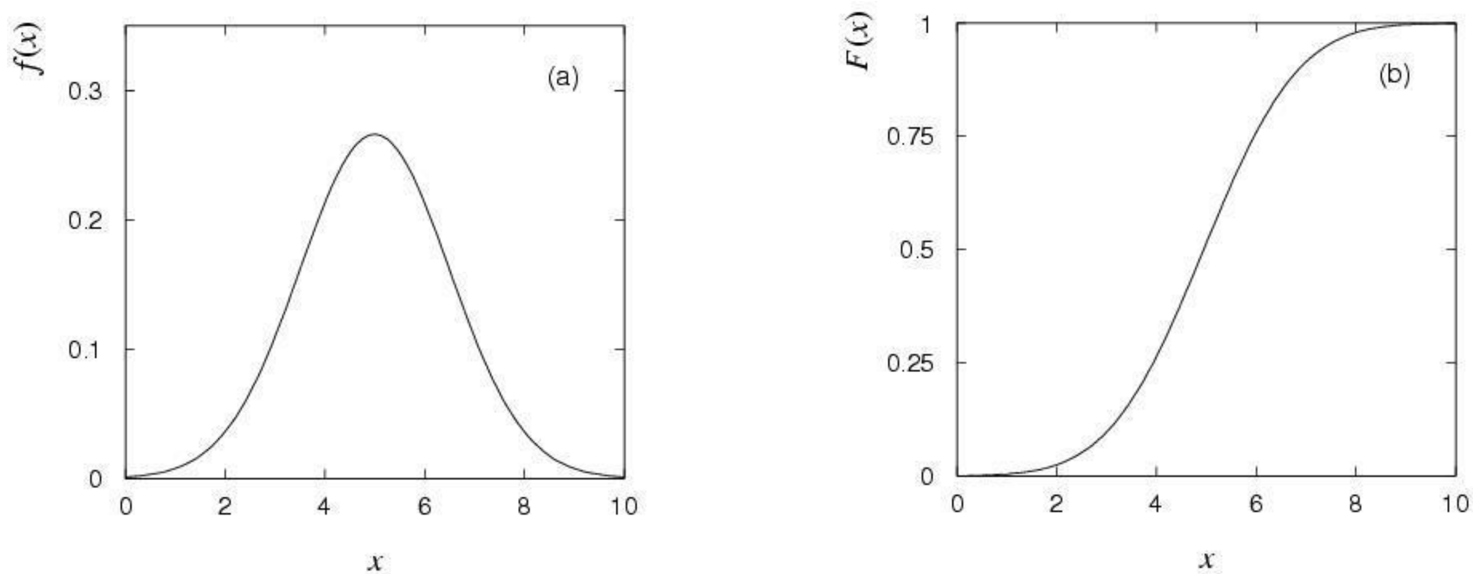
$n =$ number of entries

$\Delta x =$ bin width



Wahrscheinlichkeit ein Ergebnis keiner oder gleich x zu haben

$$\int_{-\infty}^{x} f(x')\,dx' \equiv F(x)$$ Kumulativfunktion



Alternativ: Definition der WDF via

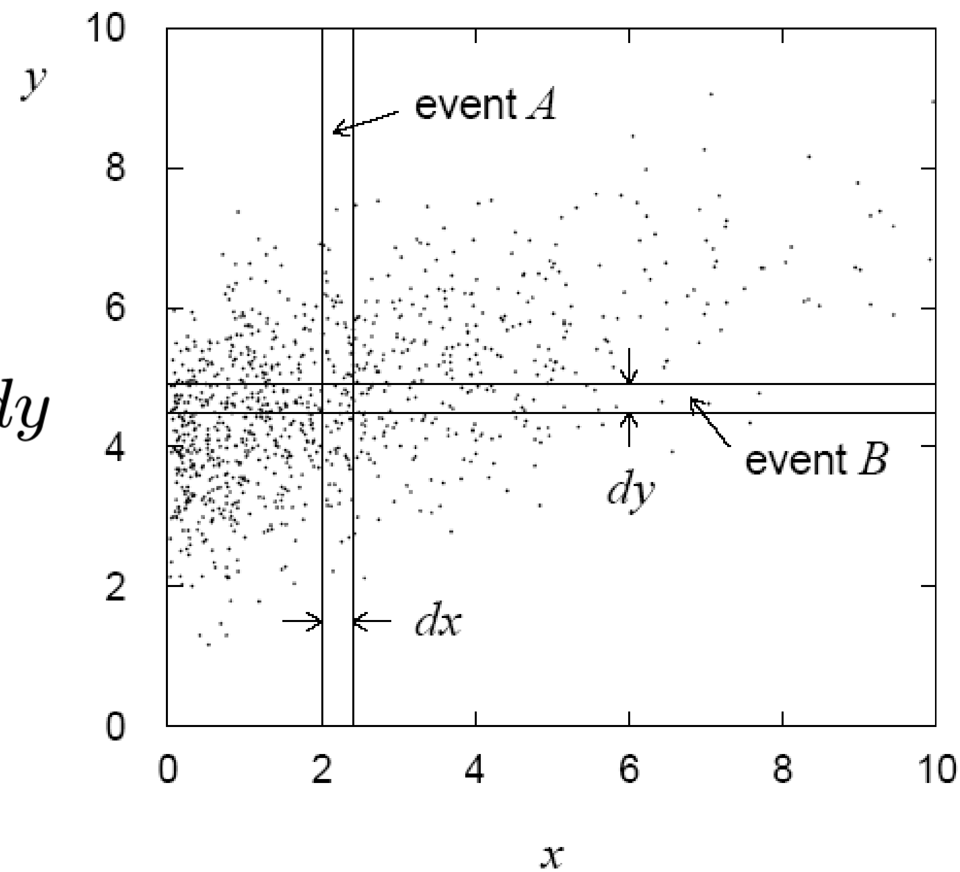
$$f(x) = \frac{\partial F(x)}{\partial x}$$

Gemeinsame WDF für mehrere Zufallsvariablen

Ergebnis eines Experimentes charakterisiert durch mehrer Werte, z.B. ein n-Komponenten Vektor (x_1, ... x_n)

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

Gemeinsame WDF



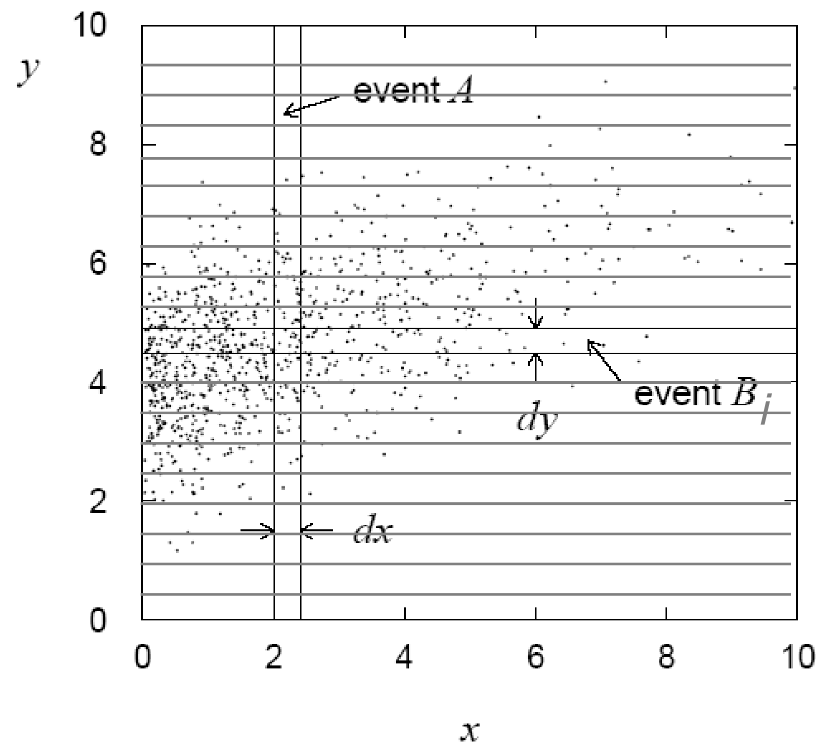
Normierung: $$\int \cdots \int f(x_1, \ldots, x_n) dx_1 \cdots dx_n = 1$$

Randverteilung / Marginal-WDF

Manschmal nur Interesse an WDF an einigen (oder einer) der Komponenten:

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

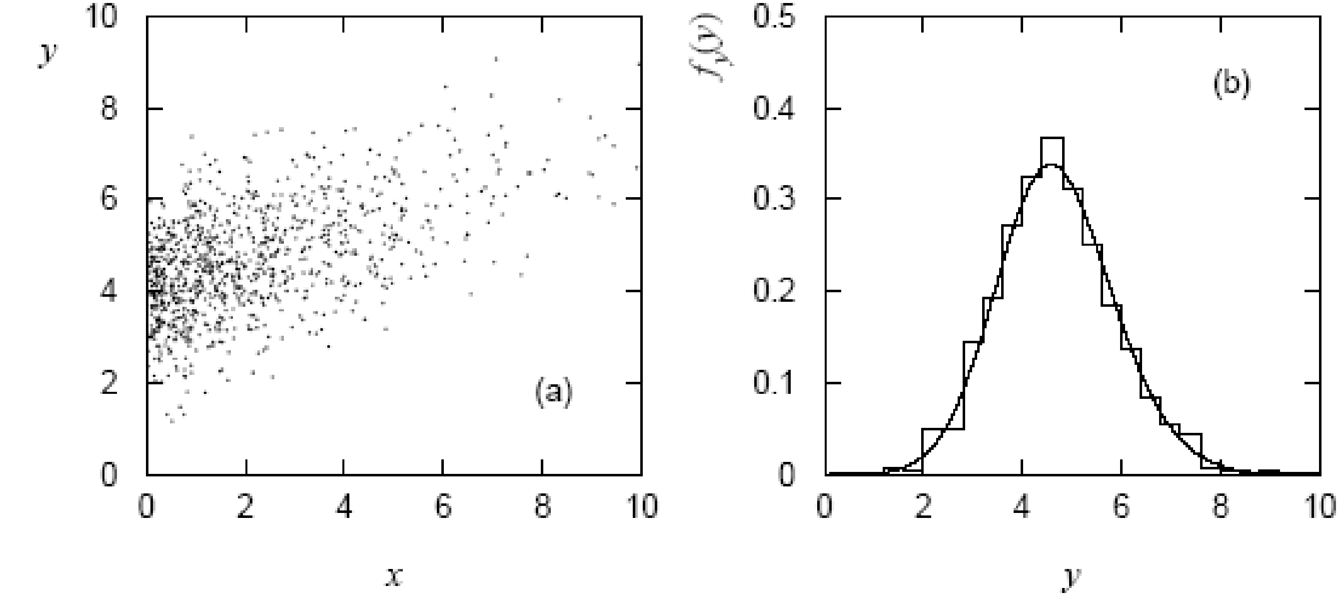
$$f_x(x) = \int f(x, y)\, dy$$

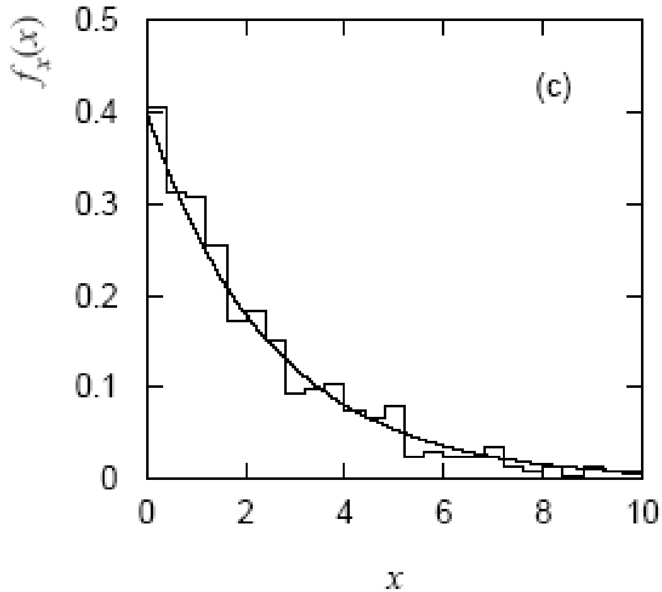


$\rightarrow$ Marginal-WDF $\quad f_1(x_1) = \int \cdots \int f(x_1, \ldots, x_n)\, dx_2 \ldots dx_n$

x_1, x_2 unabhängig, wenn $\quad f(x_1, x_2) = f_1(x_1) f_2(x_2)$

Marginal-WDF durch Projektion auf Achsen

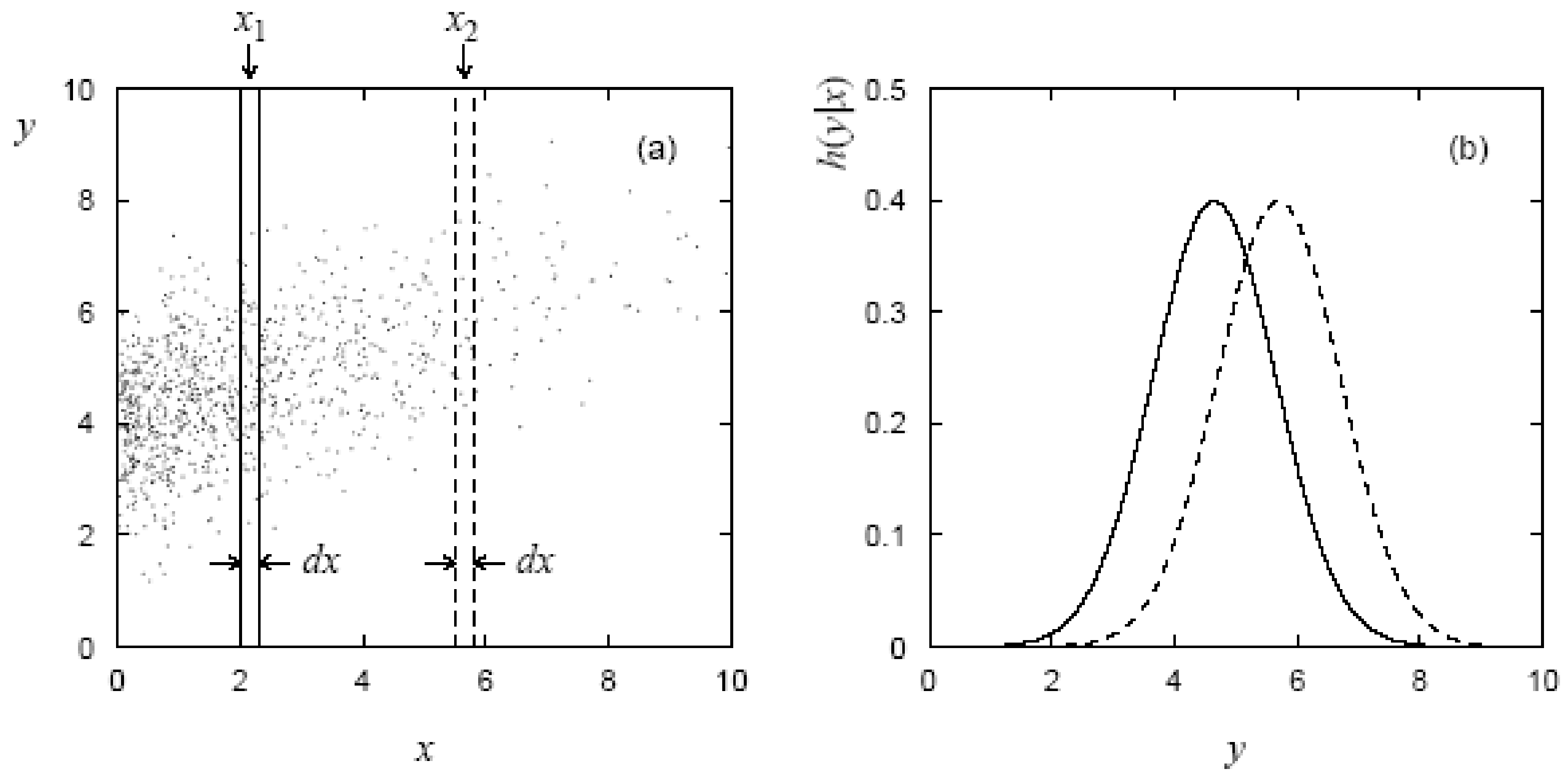




Marginal-WDF ~ Projektion der gemeinsamen WDF

auf individuelle Achsen

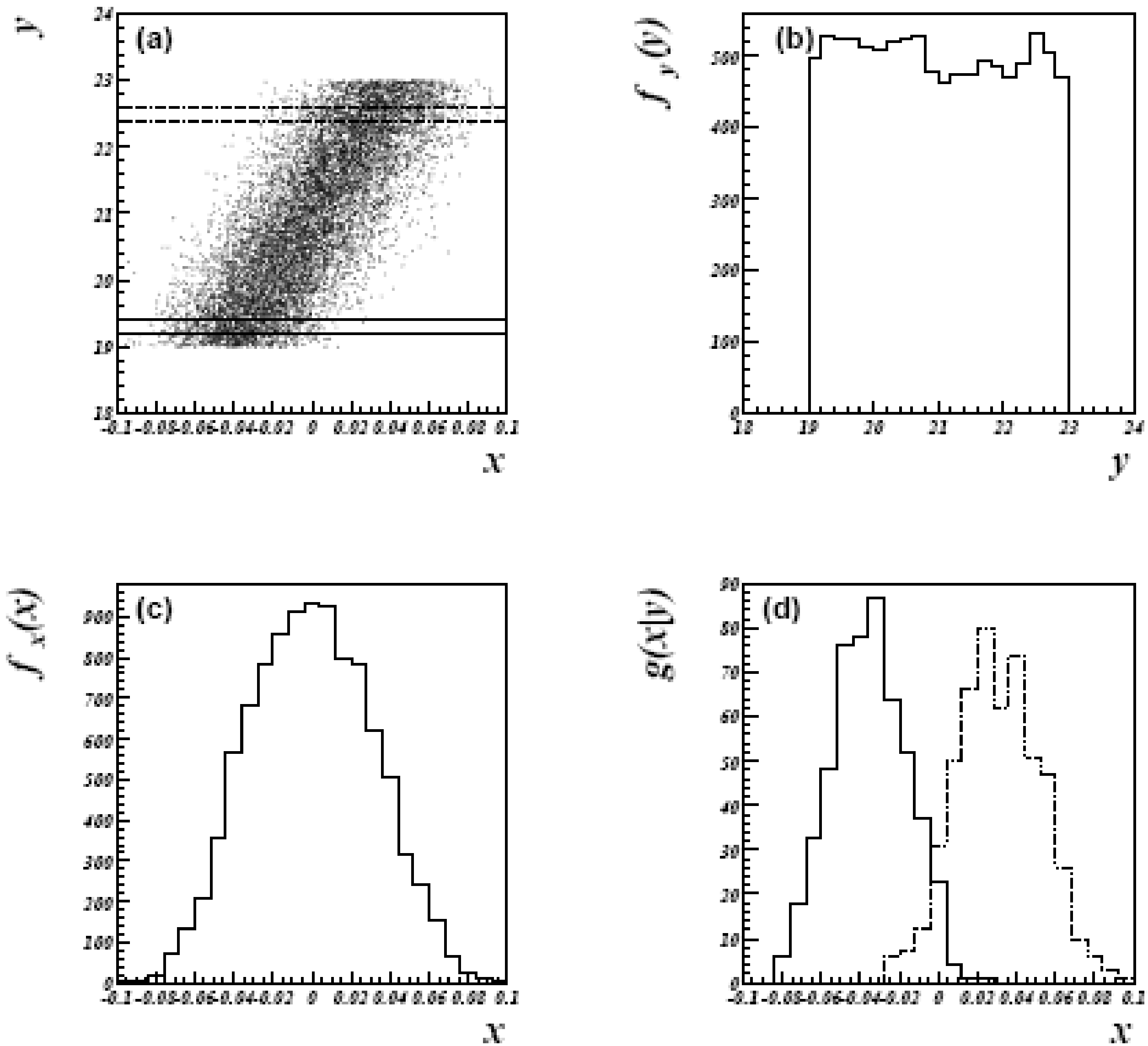
z.B. gemeinsame WDF $f(x,y)$ wird verwendet um bedingte WDFs $h(y|x_1)$, $h(y|x_2)$ zu finden:



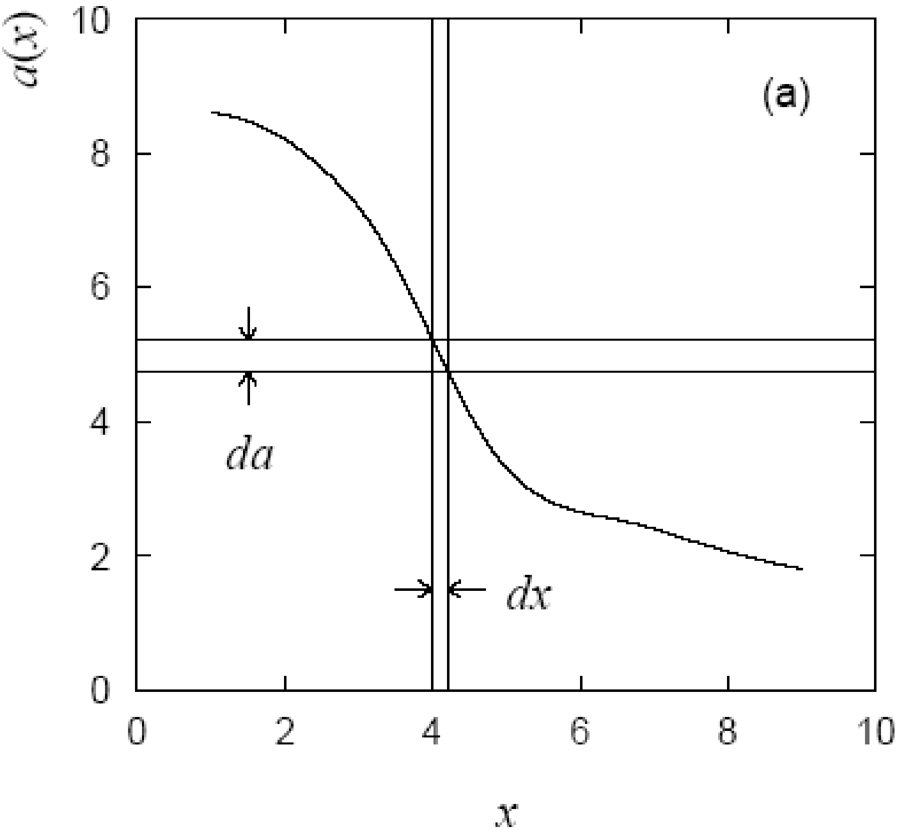
Behandele eine der ZV als konstant. Dividiere die gemeinsame WDF durch die Marginal-WDF für die Variablen, die konstant gehalten werden. Normierung per constructionem korrekt.

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

Marginal- und bedingte WDF



Transformation der Zufallsvariablen

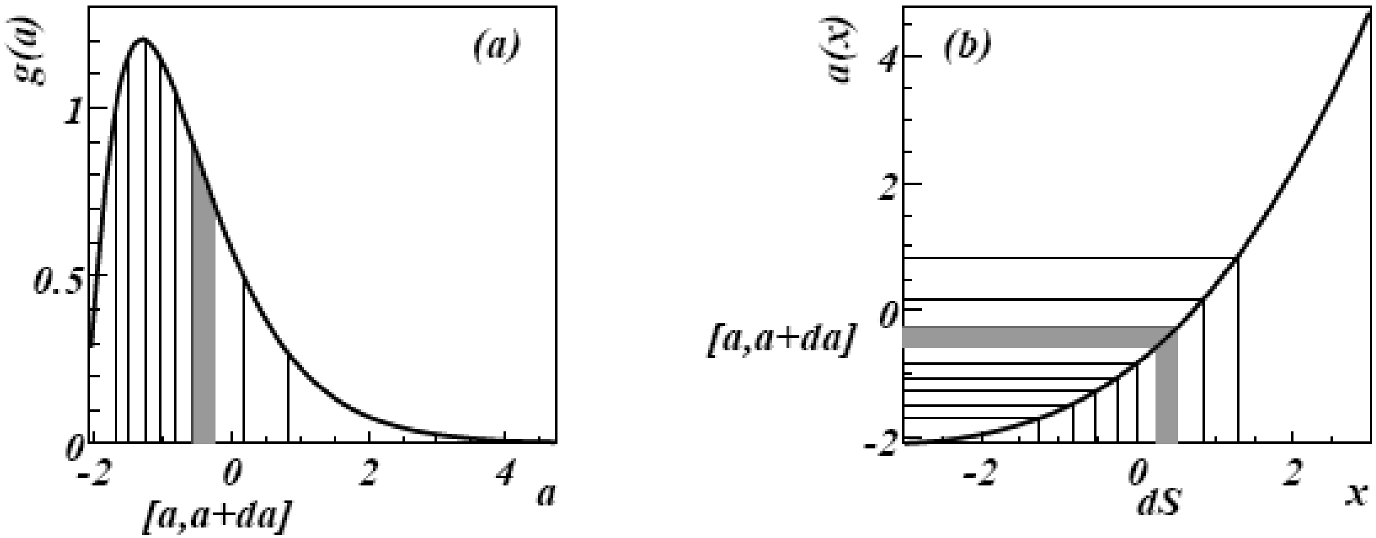


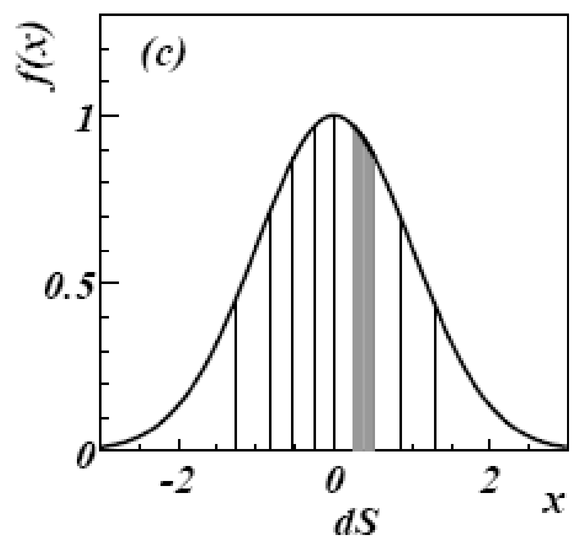
$$g(a)\,da = \int_{dS} f(x)\,dx$$

dS = Region im *x*-Raum für die *a(x)* in [*a*, *a*+*da*] ist.
Für eine ZV mit eindeutig invertierbarer Funktion a(x)

$$g(a)\,da = f(x)\,dx$$

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



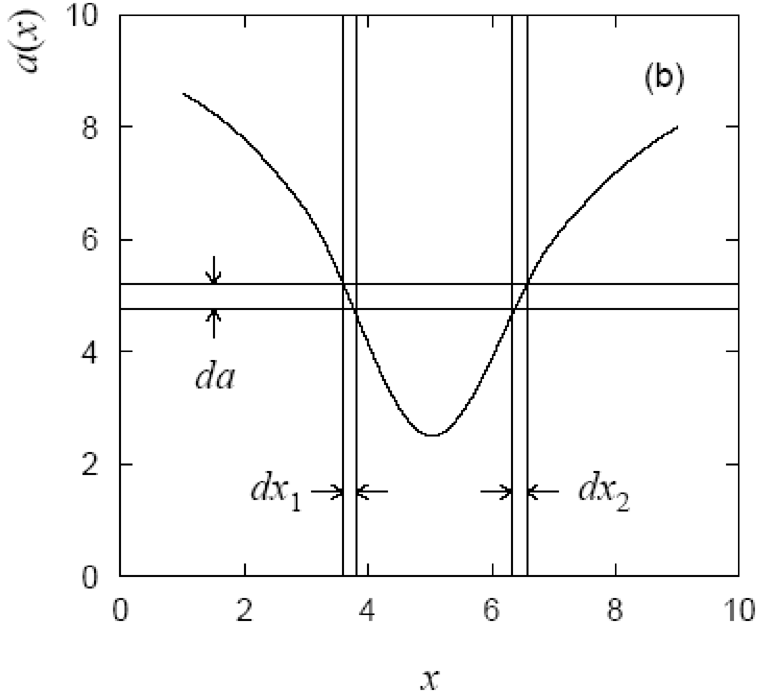


Wahrscheinlichkeit im a-Raum g(a)da

Entspricht der im x-Raum f(x)dS

Trafo von ZV: mehrdeutige Inverse

Wenn Inverse von *a*(*x*) nicht eindeutig,
Dann alle Intervalle *dx* in *dS*
Berücksictigen, die zu *da* korrespondieren

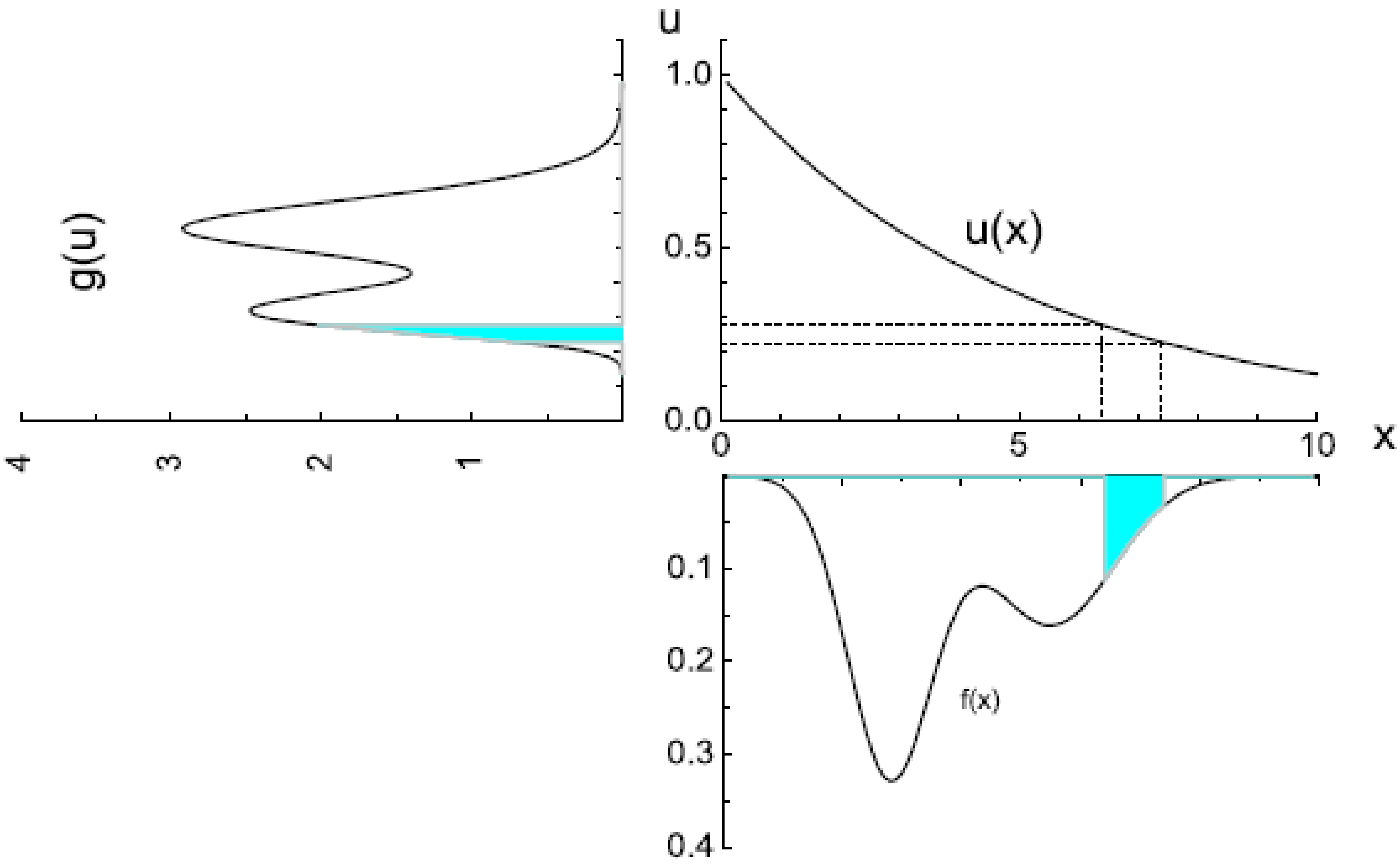


Example: $$a = x^2, \;\; x = \pm\sqrt{a}, \;\; dx = \pm\frac{da}{2\sqrt{a}} \; .$$

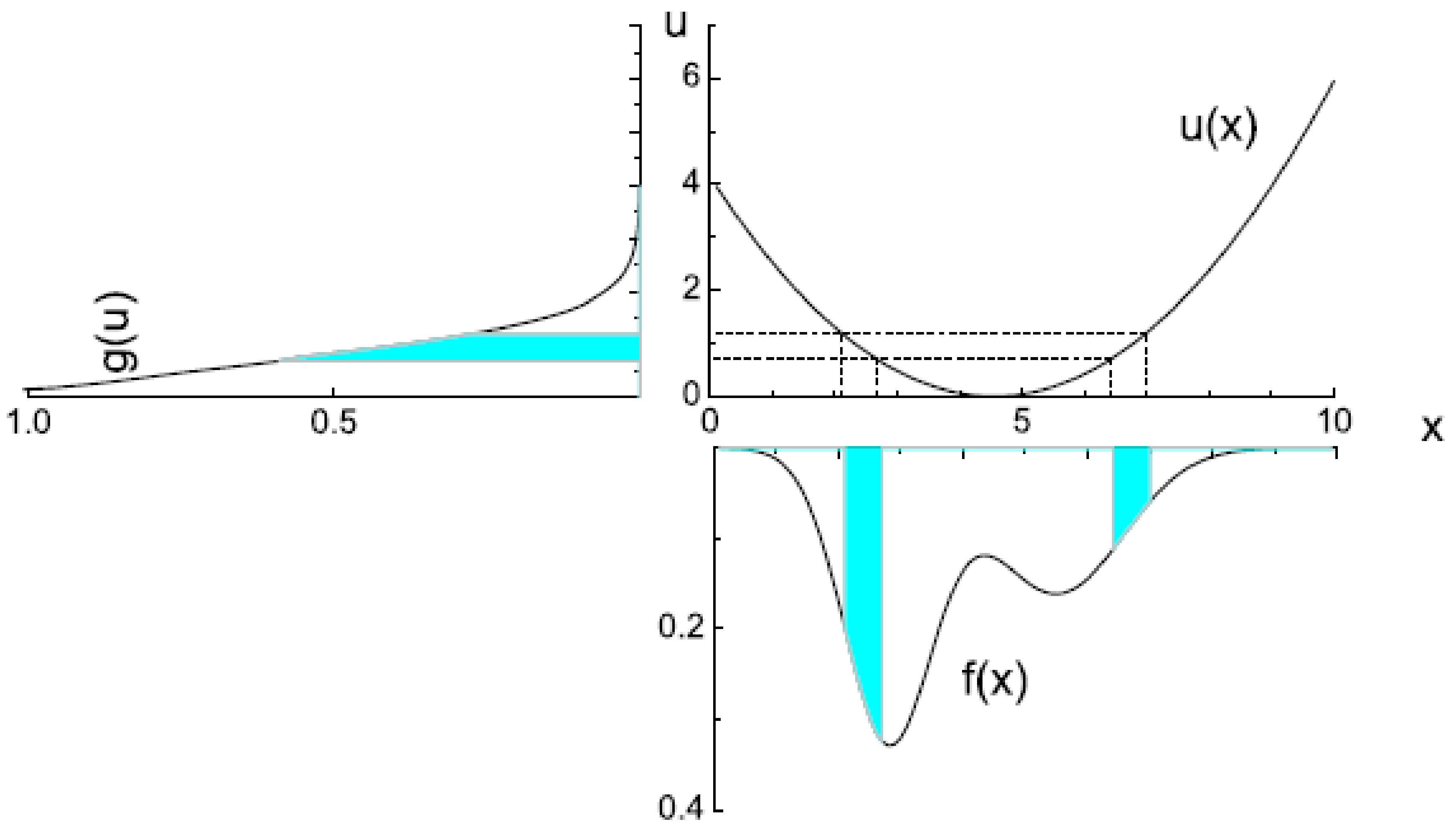
$$dS = \left[\sqrt{a}, \sqrt{a} + \frac{da}{2\sqrt{a}}\right] \cup \left[-\sqrt{a} - \frac{da}{2\sqrt{a}}, -\sqrt{a}\right]$$

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

Transformation von Zufallsvariablen

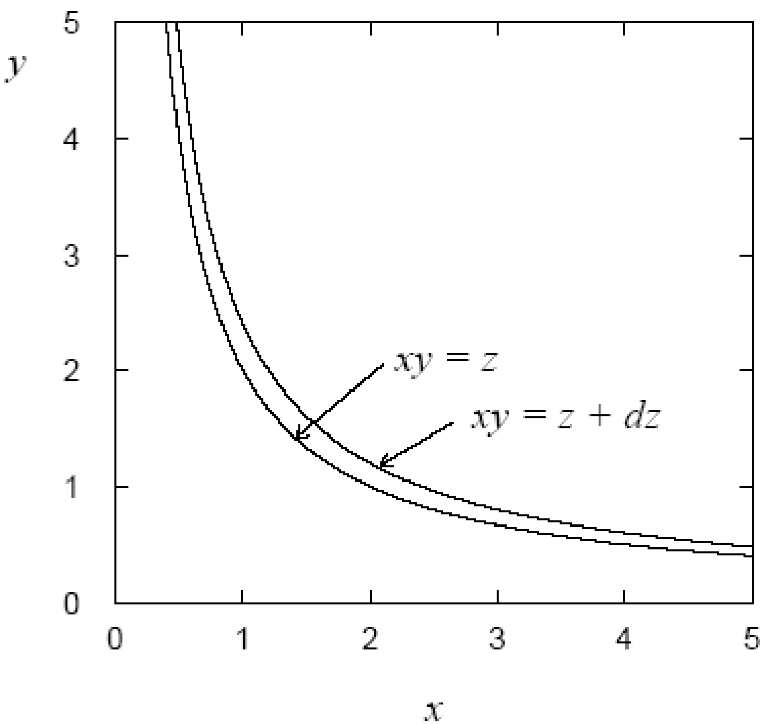


Trafo von ZV: mehrdeutige Inverse



ZV z=a(x,y)=xy : Mellin-Faltung

x, *y* > 0 mit gemeinsamer WDF *f*(*x*,*y*). Betrachte Funktion z = *xy*.



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

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

(Mellin-Faltung)