# Statistische Methoden der Datenanalyse

# Wintersemester 2011/2012

# Albert-Ludwigs-Universität Freiburg

Prof. Markus Schumacher

Physikalisches Institut Westbau 2 OG Raum 008

Telefonnummer 07621 203 7612

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

Vorlesung basiert in weiten Teilen auf der von Glen Cowan.

http://terascale.physik.uni-freiburg.de/lehre/ws_1112/statmethoden_ws1112

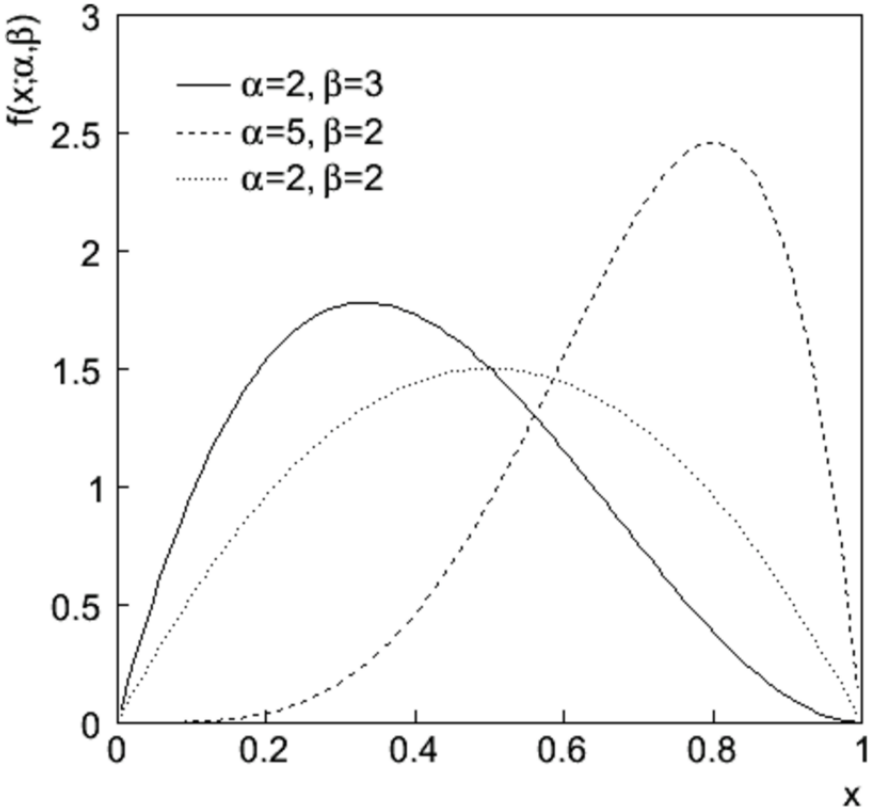
# Beta-Verteilung

Die Beta-WDF für eine kontinuierliche ZV *x* im Intervall [0,]| ist definiert als

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

$$E[x] = \frac{\alpha}{\alpha+\beta}$$

$$V[x] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$




Wird oft verwendet um WDF zu approximieren, die eine kontinuierliche ZV in eine gewissen Bereich beschrieben [a,b] beschreiben.

# Beta-Verteilung: Eigenschaften

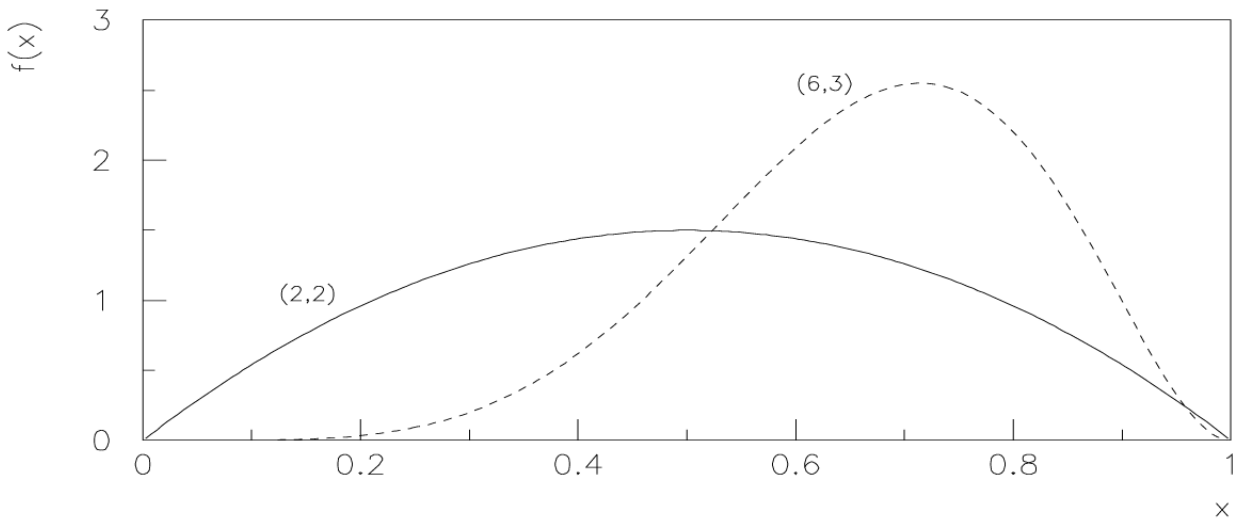
Für a=b=1 ergibt sich die Gleichverteilung in [0,1]

Für a=1, b=2 (oder umgekehrt) → Dreiecksverteilung f(x)=2-2x und f(x)=2x

Für a=b=2 → Parabelverteilung f(x)=6x(1-x)

Für a,b>1: ein Modalwert bei x=(a-1)/(a+b-2)

Für a und/oder b<1: f(0) und/oder f(1) → unendlich (J-Form)



# Beta-Verteilung: Herleitung

Seien $y_m$ und $y_n$ zwei ZV die nach Chi2-WDF mit m bzw. n Freiheitsgraden verteilt sind

Dann folgt ZV $x = y_m/(y_m+y_n)$ einer Beta-WDF mit $\alpha$=m/2 und $\beta$=n/2

$$x = y_m/(y_m + y_n) \text{ and } y = y_m + y_n \qquad y_m = xy \text{ and } y_n = y(1-x)$$

$$\begin{aligned} f(x,y) &= \left\| \begin{matrix} \frac{\partial y_m}{\partial x} & \frac{\partial y_m}{\partial y} \\ \frac{\partial y_n}{\partial x} & \frac{\partial y_n}{\partial y} \end{matrix} \right\| f(y_m, y_n) = \\ &= \left| \begin{matrix} y & x \\ -y & 1-x \end{matrix} \right| \left\{ \frac{\left(\frac{y_m}{2}\right)^{\frac{m}{2}-1} e^{-\frac{y_m}{2}}}{2\Gamma\left(\frac{m}{2}\right)} \right\} \left\{ \frac{\left(\frac{y_n}{2}\right)^{\frac{n}{2}-1} e^{-\frac{y_n}{2}}}{2\Gamma\left(\frac{n}{2}\right)} \right\} = \\ &= \left\{ \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} x^{\frac{m}{2}-1}(1-x)^{\frac{n}{2}-1} \right\} \left\{ \frac{\left(\frac{y}{2}\right)^{\frac{m}{2}+\frac{n}{2}-1} e^{-\frac{y}{2}}}{2\Gamma\left(\frac{m+n}{2}\right)} \right\} \end{aligned}$$

Wir erkennen Produkt von Beta-WDF für x mal Chi-Quadrat-WDF mit n+m Freiheitsgraden für y

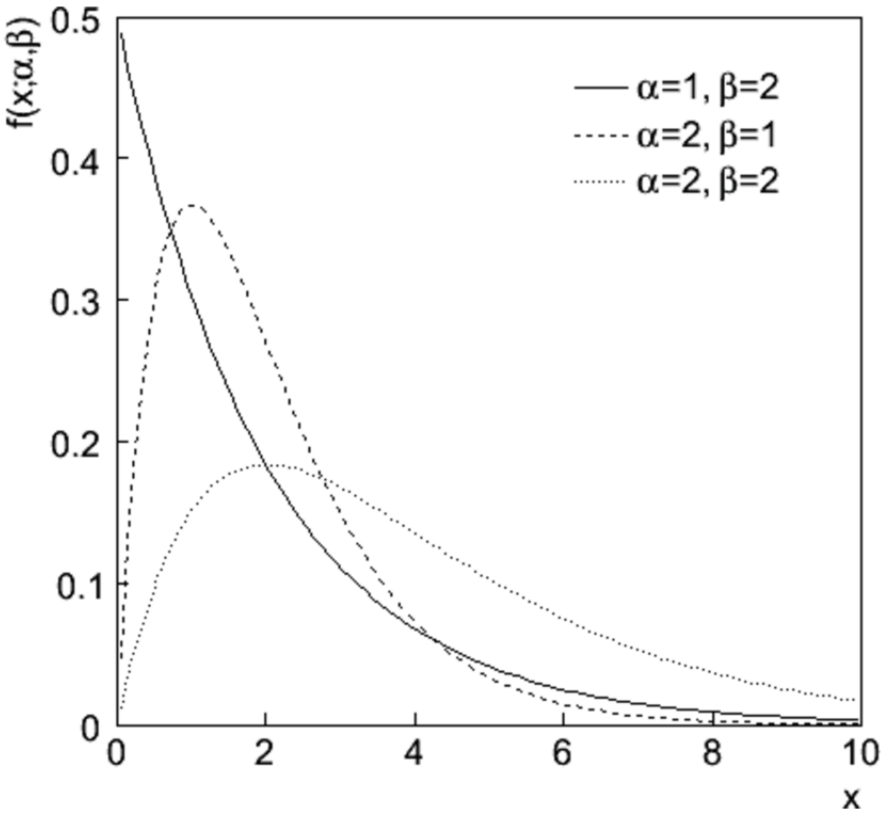
# Gamma-Verteilung

Die Gamma-WDF für eine kontinuierliche ZV *x* im Intervall [0,∞] ist definiert als

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}\, x^{\alpha-1}\, e^{-x/\beta}$$

$$E[x] = \alpha\beta$$

$$V[x] = \alpha\beta^2$$



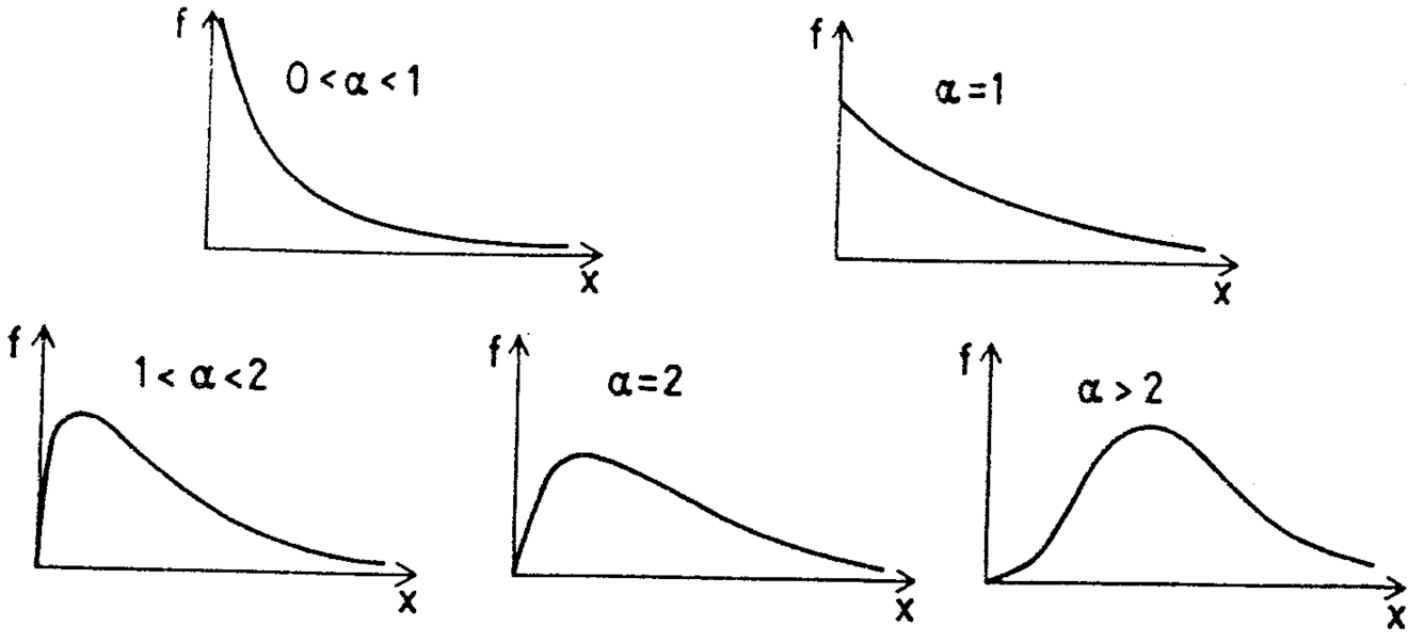

Oft benutzt um eine WDF im Bereich [0,∞] zu parametrisieren.

Beschreibt auch z.B.
- Summe von *n* ZV aus Exponentialverteilung
- Zeit bis zum *n*ten Ereignis in einem Poisson-Prozess

# Gamma-Verteilung

Für $\alpha$<=1 ist Gamma-WDF J-förmig für $\alpha$>1 ist Modalwert bei x=(a-1)

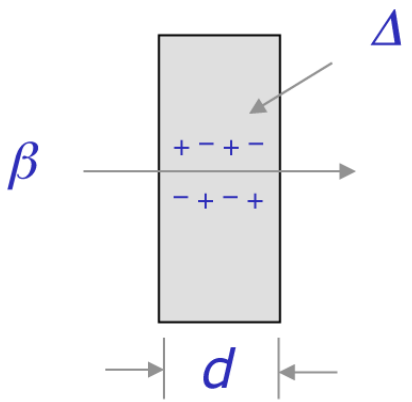
Schiefe: = 2/sqrt($\alpha$) Wölbung=6/$\alpha$




für $\alpha$= 1 → Exponential-WDF
für $\beta$=2, $\alpha$ = n (ganzzahlig) /2 → Chi-Quadrat-WDF mit n Freiheitsgraden
für $\alpha$ ganzzahlig auch Erlangen-Verteilung genannt.

# Landau-Verteilung

Der Energieverlust $\Delta$ eines geladenen Teilchens mit Geschwindigkeit $\beta = v/c$ , welches eine Materialschicht der Dicke d durchquert, folgt der Landau-Verteilung.




$$f(\Delta;\beta) = \frac{1}{\xi}\phi(\lambda) \ ,$$

$$\phi(\lambda) = \frac{1}{\pi}\int_0^\infty \exp(-u\ln u - \lambda u)\sin \pi u \, du \ ,$$

$$\lambda = \frac{1}{\xi}\left[\Delta - \xi\left(\ln\frac{\xi}{\epsilon'} + 1 - \gamma_{\mathrm{E}}\right)\right] \ ,$$

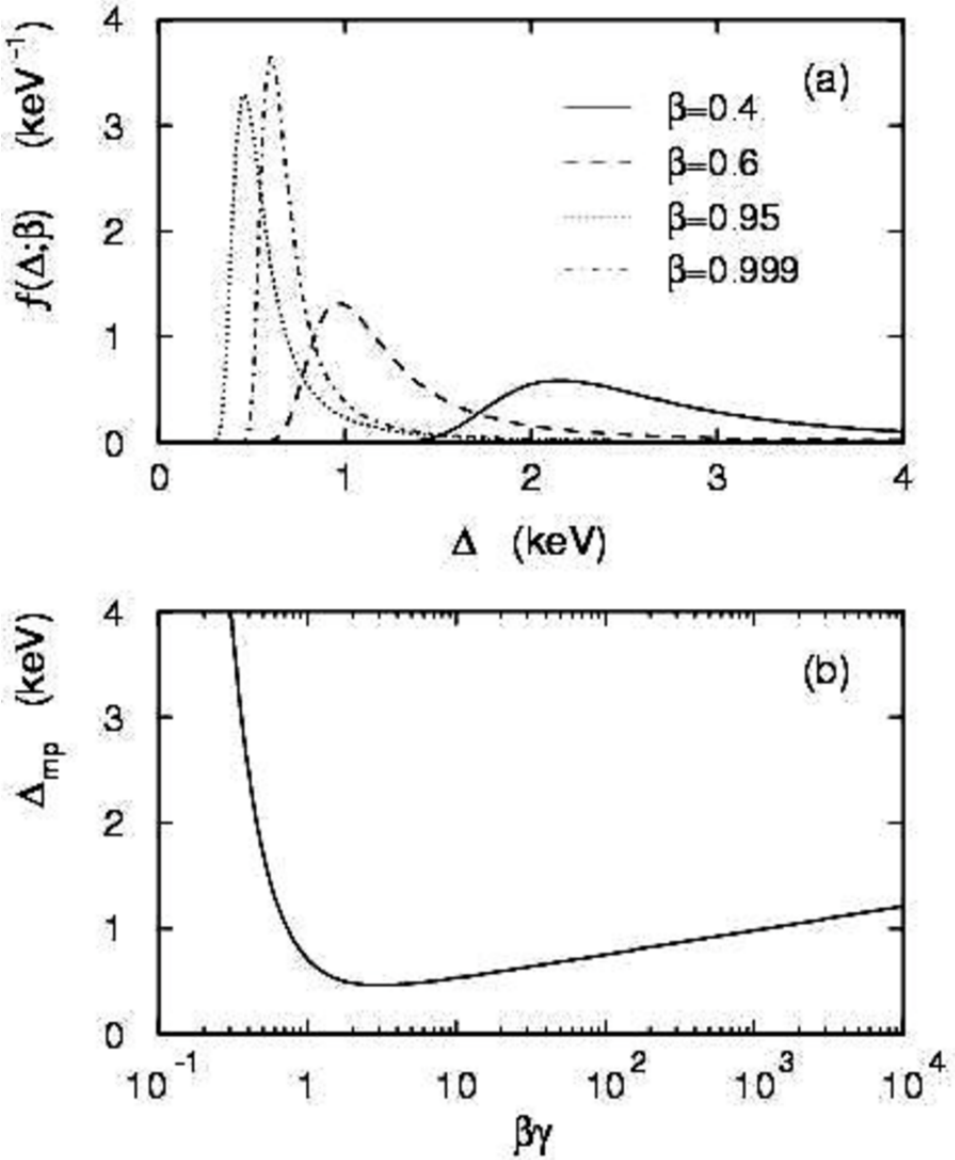
$$\xi = \frac{2\pi N_{\mathrm{A}} e^4 z^2 \rho \sum Z}{m_{\mathrm{e}} c^2 \sum A}\frac{d}{\beta^2} \ , \qquad \epsilon' = \frac{I^2 \exp\beta^2}{2 m_{\mathrm{e}} c^2 \beta^2 \gamma^2} \ .$$

L. Landau, J. Phys. USSR **8** (1944) 201; see also
W. Allison and J. Cobb, Ann. Rev. Nucl. Part. Sci. **30** (1980) 253.

# Landau-Verteilung (2)

Lange “Landau-Ausläufer”
→ alle Momente ∞

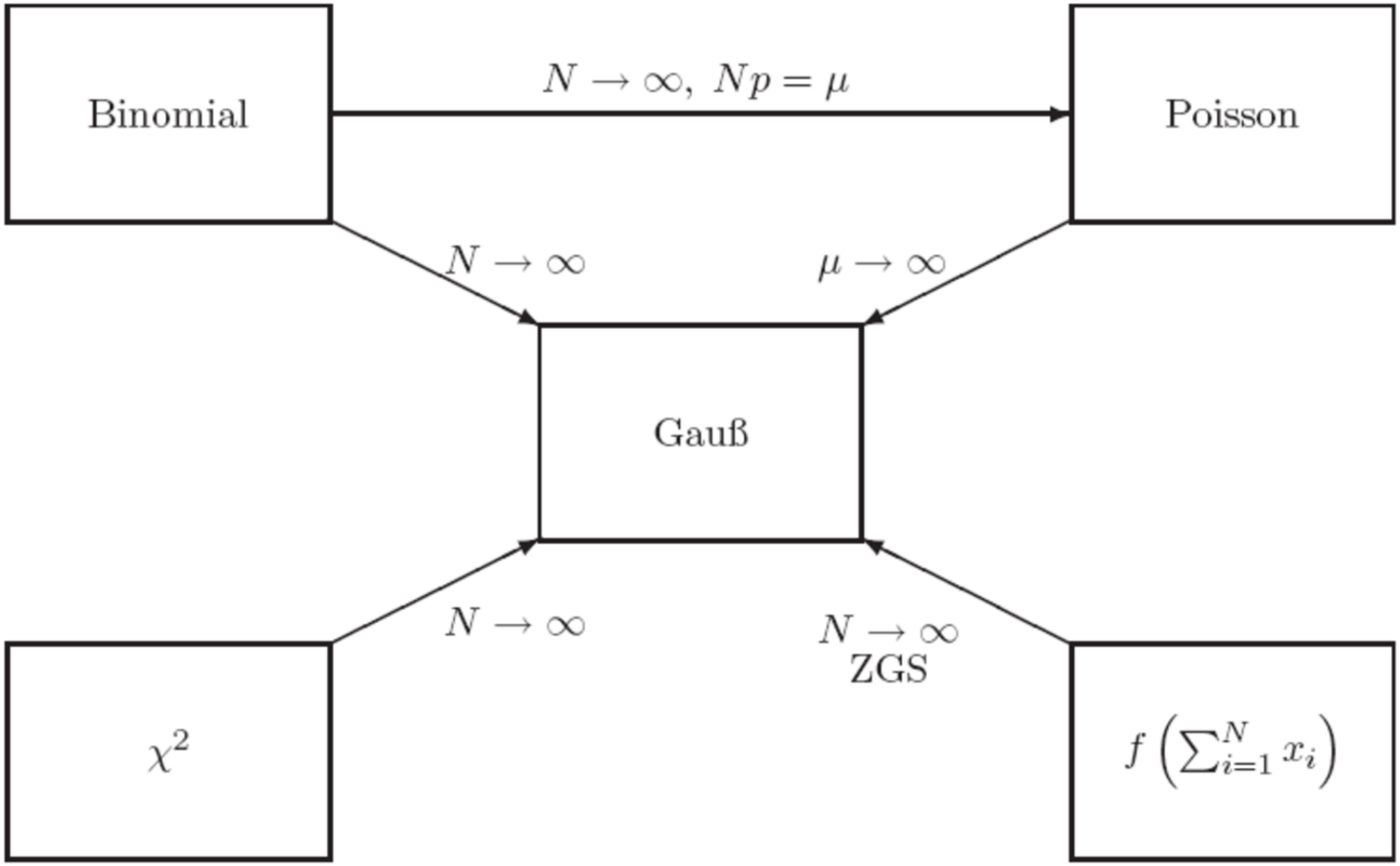
Modalwert sensitiv
auf Geschwindigkeit $\beta$ ,
→ Teilchenidentifikation



# Tabellarische Übersicht

| Name | Definitionsbereich | Wahrscheinlichkeits(dichte) | $\langle x \rangle = E[x]$ | $\sigma_x = \sqrt{V[x]}$ | charakter. Funktion |
|---|---|---|---|---|---|
| Binomial | $\{0, 1, 2, 3, ..., N\}$ | $f(x; p, N) = \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x}$ | $Np$ | $\sqrt{Np(1-p)}$ | $(p(\exp(ik) - 1) + 1)^N$ |
| Poisson | $\{0, 1, 2, 3, ...\}$ | $f(x; \lambda) = \frac{\lambda^x}{x!} \exp(-\lambda)$ | $\lambda$ | $\sqrt{\lambda}$ | $\exp(\lambda(\exp(ik) - 1))$ |
| Gleichvert. | $[a, b]$ | $f(x; a, b) = \frac{1}{(b-a)}$ | $\frac{a+b}{2}$ | $\frac{b-a}{\sqrt{12}}$ | $\frac{\exp(ibk) - \exp(iak)}{(b-a)ik}$ |
| Gauß | $[-\infty, \infty]$ | $f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ | $\mu$ | $\sigma$ | $\exp\left(i\mu k - \frac{1}{2}\sigma^2 k^2\right)$ |
| Chi-Quadrat | $[0, \infty]$ | $f(x; n) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{\frac{n}{2}-1} \exp\left(-\frac{x}{2}\right)$ | $n$ | $2n$ | $(1 - 2ik)^{-n/2}$ |
| Exponentiell | $[0, \infty]$ | $f(x; \xi) = \frac{1}{\xi} \exp(-x/\xi)$ | $\xi$ | $\xi$ | $\frac{1}{1-ik\xi}$ |
| Cauchy | $[-\infty, \infty]$ | $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ | $(0)$ | $\infty$ | $\exp(-\lvert k \rvert)$ |
| Landau | $[0, \infty]$ | $f(x; \beta) = ...$ | $\infty$ | $\infty$ | |

# Zusammenhang zwischen WDFs



# Zusammenhang zwischen WDFs

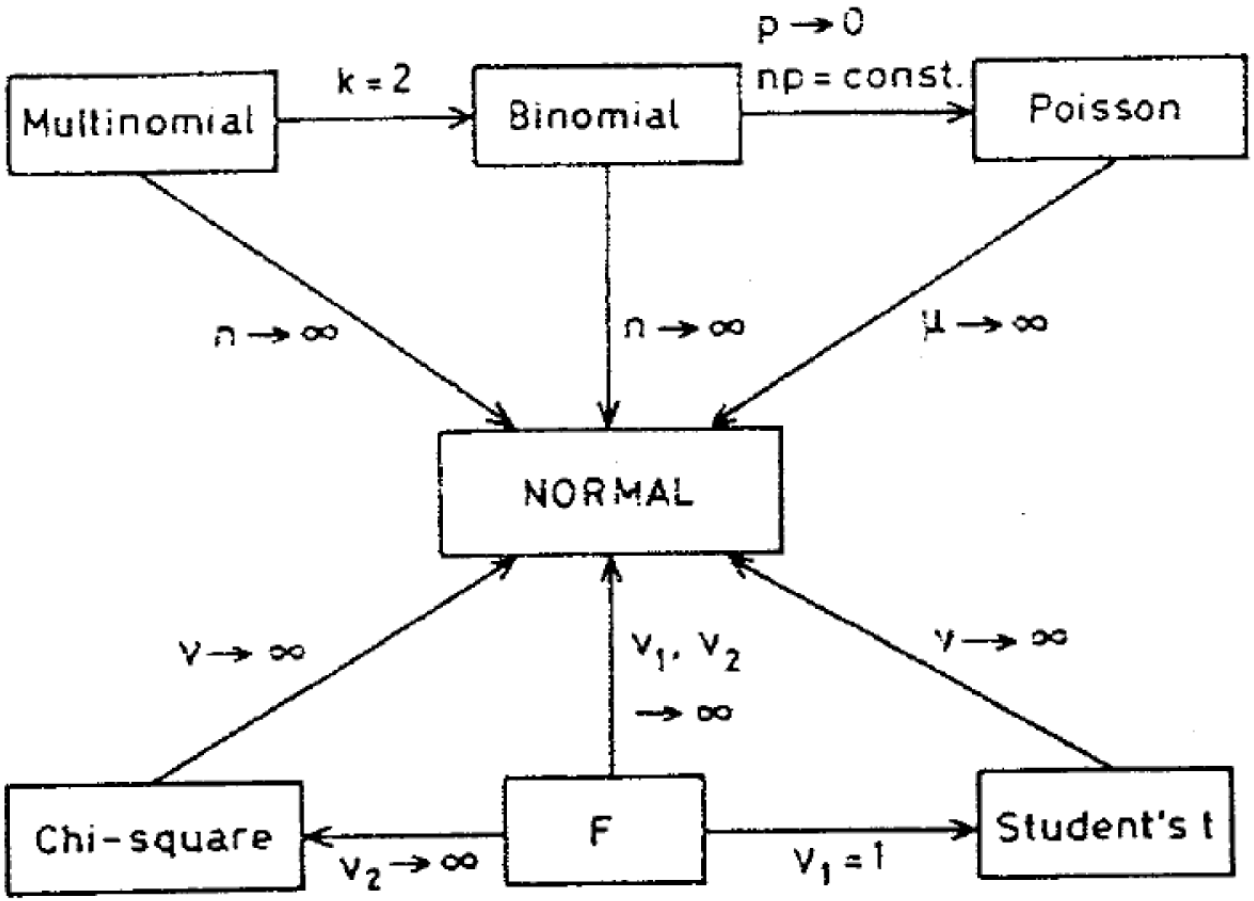



Fig. 5.5. Relations between probability distributions.

# Kapitel 3

# Die Monte-Carlo-Methode für Erstellung von Wahrscheinlichkeitsdichtefunktionen

# Die Transformationsmethode (Zsfg.)

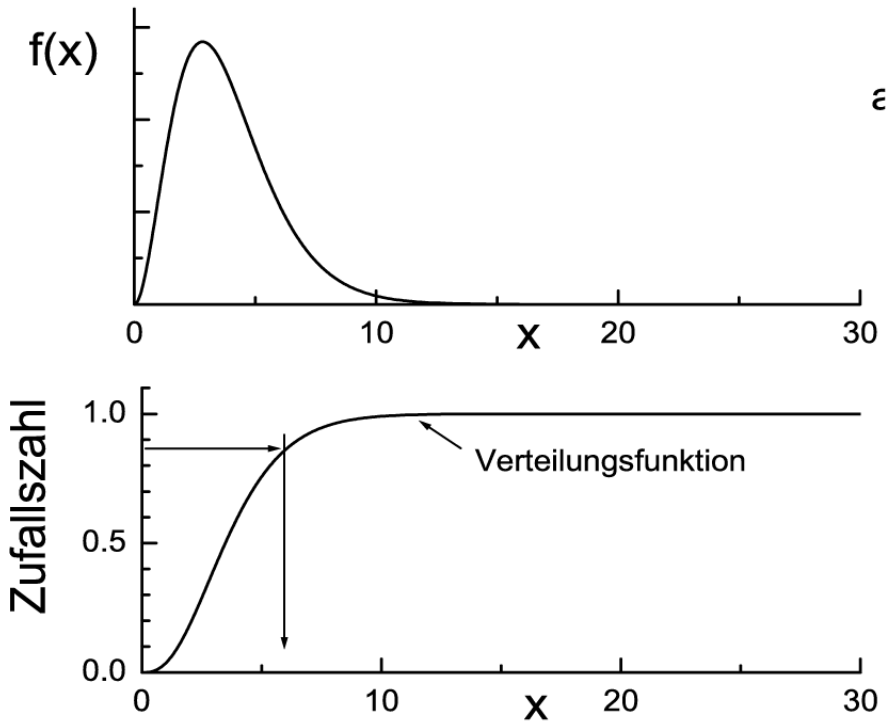
Aus r gleichförmig in [0,1] erzeuge x, die WDF f(x) folgt, gemäß:

$$f(x)\mathrm{d}x = u(r)\mathrm{d}r,$$

$$\int_{-\infty}^{x} f(x')\mathrm{d}x' = \int_{0}^{r(x)} u(r')\mathrm{d}r' = r(x)$$

$$F(x) = r,$$

$$x(r) = F^{-1}(r)\,.$$




Voraussetzung: Kumulativverteilung analytisch integrierbar u. invertierbar

Effizienz des Verfahrens 100% (jedes r erzeugt ein x)

# Zufallszahlengenerator

Ziel: Erzeugung gleichmäßig verteilter Werte im Intervall [0, 1].

Werfe Münze für z.B. 32 Bitnummer... (too tiring).

→ 'Zufallszahlengenerator'

= Computeralgorithmus zur Erzeugung von $r_1, r_2, ..., r_n$.

Beispiel: Multiplikativer linear kongruenter Generator (MLCG)

$n_{i+1} = (a\, n_i) \bmod m$ , wobei

$n_i$ = ganze Zahl

$a$ = Multiplikator (ganze Zahl)

$m$ = Modulus (ganze Zahl)

$n_0$ = "Seed" ="Saatzahl" (Startwert)

Bemerkung: mod = Modulus z.B. 27 mod 5 = 2.

Diese Regel erzeugt eine Sequenz von Zahlen $n_0, n_1, ...$

# Zufallszahlengenerator (2)

Die Sequenz ist (unglücklicherweise) periodisch!

Beispiel: (siehe Brandt Kapitel 4): $a = 3$, $m = 7$, $n_0 = 1$

$$
\begin{aligned}
n_1 &= (3 \cdot 1) \mathrm{mod} 7 = 3 \\
n_2 &= (3 \cdot 3) \mathrm{mod} 7 = 2 \\
n_3 &= (3 \cdot 2) \mathrm{mod} 7 = 6 \\
n_4 &= (3 \cdot 6) \mathrm{mod} 7 = 4 \\
n_5 &= (3 \cdot 4) \mathrm{mod} 7 = 5 \\
n_6 &= (3 \cdot 5) \mathrm{mod} 7 = 1 \quad \leftarrow \text{Folge wiederholt sich}
\end{aligned}
$$

Wähle $a$, $m$ so, dass eine lang Periode (Maximum = $m$ - 1) erreicht wird; $m$ normalerweise nahe an der größten “Integer”-Zahl, die auf dem Computer repräsentiert werden kann.

# Zufallszahlengenerator (3)

$r_i = n_i/m$ sind in [0, 1] aber sind sie “zufällig”?

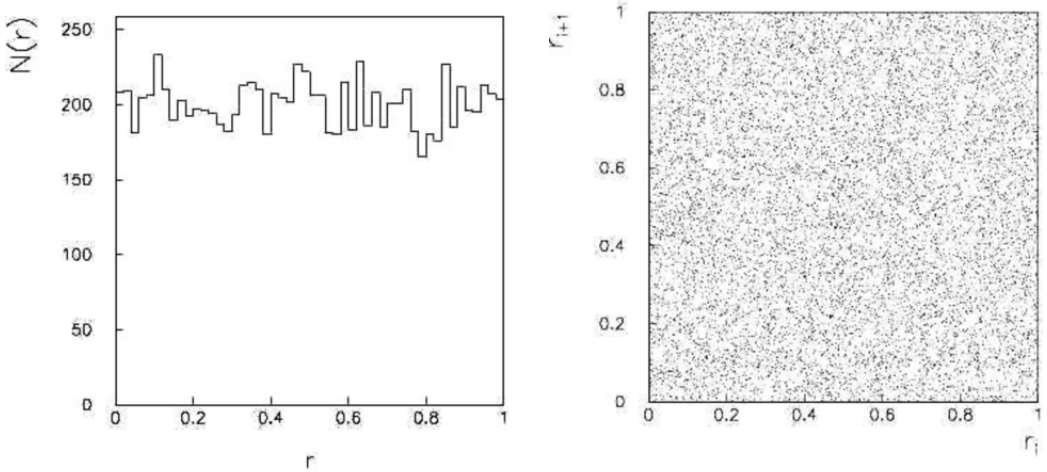
Wähle *a*, *m* so, dass die $r_i$ eine Reihe von Tests für Zufälligkeit erfüllen:

gleichförmig verteilt in [0, 1],

alle Werte unabhängig (keine Korrelationen zwischen Paaren),

z.B. L’Ecuyer, Commun. ACM **31** (1988) 742 schlägt

*a* = 40692
*m* = 2147483399




Es gibt weit bessere Generatoren, z.B. **TRandom3**, basierend auf “Mersenne twister”-Algorithmus, Periode = $2^{19937} - 1$.

Siehe F. James, Comp. Phys. Comm. 60 (1990) 111; Brandt Ch. 4

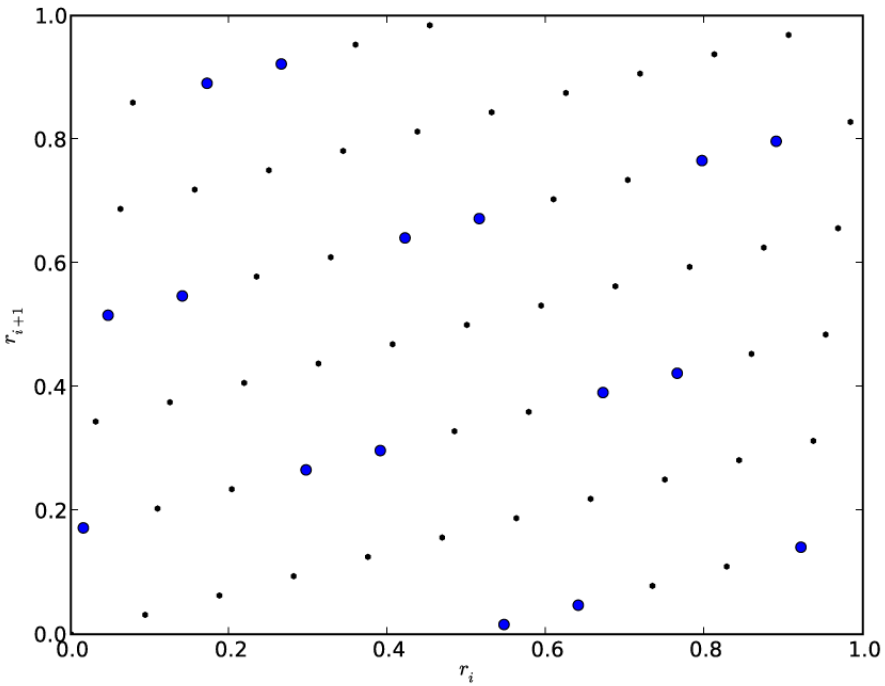
# Zufallszahlengenerator (3)

$r_i = n_i/m$ sind [0, 1] aber sind sie “zufällig”?

$$x_{i+1} = (ax_i + b) \mod m$$

Schlechtes Beispiel

*a* = 11
b = 0
*m* = 64

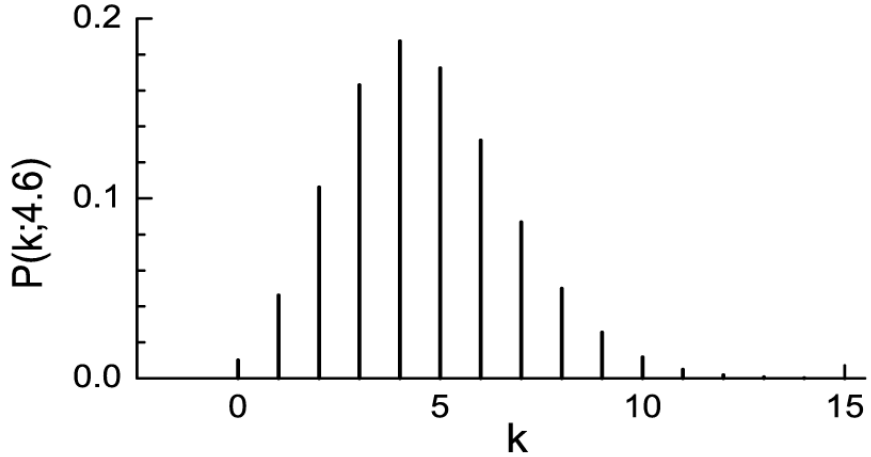



Liefert 2 Sequenzen der Periodenlänge 16

Für gerade Startwerte sind Perioden noch kürzer.

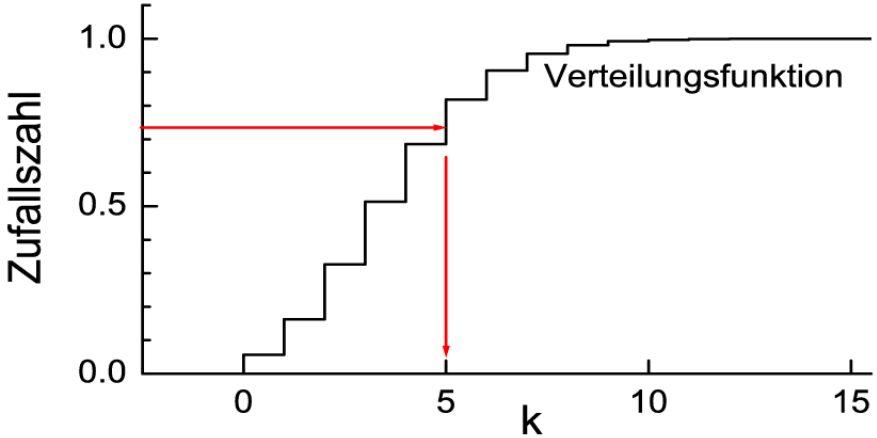
# Trafomethode für diskrete ZV und Histogramme

Ordne r die Zahl n zu,
die dem kleinsten S(k) entspricht,
welches S>r erfüllt.
S(k) ist Kumulativverteilung
für diskrete Zufallsvariable.



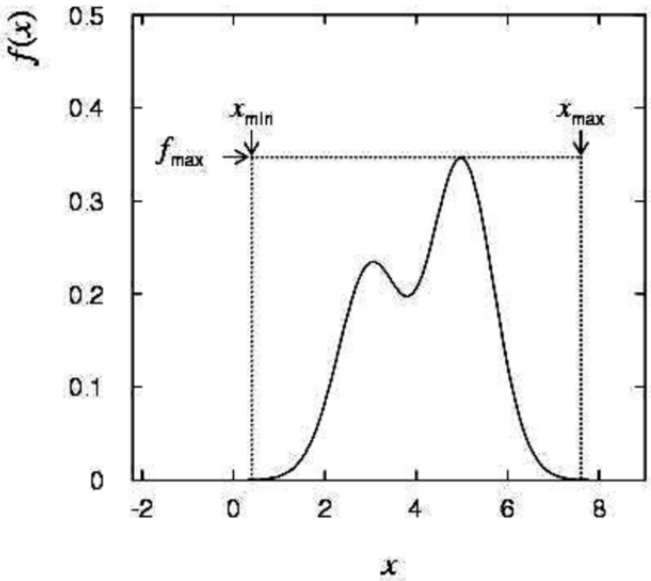

Kann für empirische Verteilung
in Form eines Histogrammes
angewendet werden.

Rest r-S(j-1) wird für lineare
Interpolation verwendet.



# Die von-Neumannsche-Zurückweisungsmethode

Schliesse WDF in Box ein




(1) Generiere Zufallszahl $x$, gleichförmig in $[x_{min}, x_{max}]$, i.e.

$$x = x_{\mathrm{min}} + r_1(x_{\mathrm{max}} - x_{\mathrm{min}}) \ ,$$ $r_1$ ist gleichverteilt in [0,1].

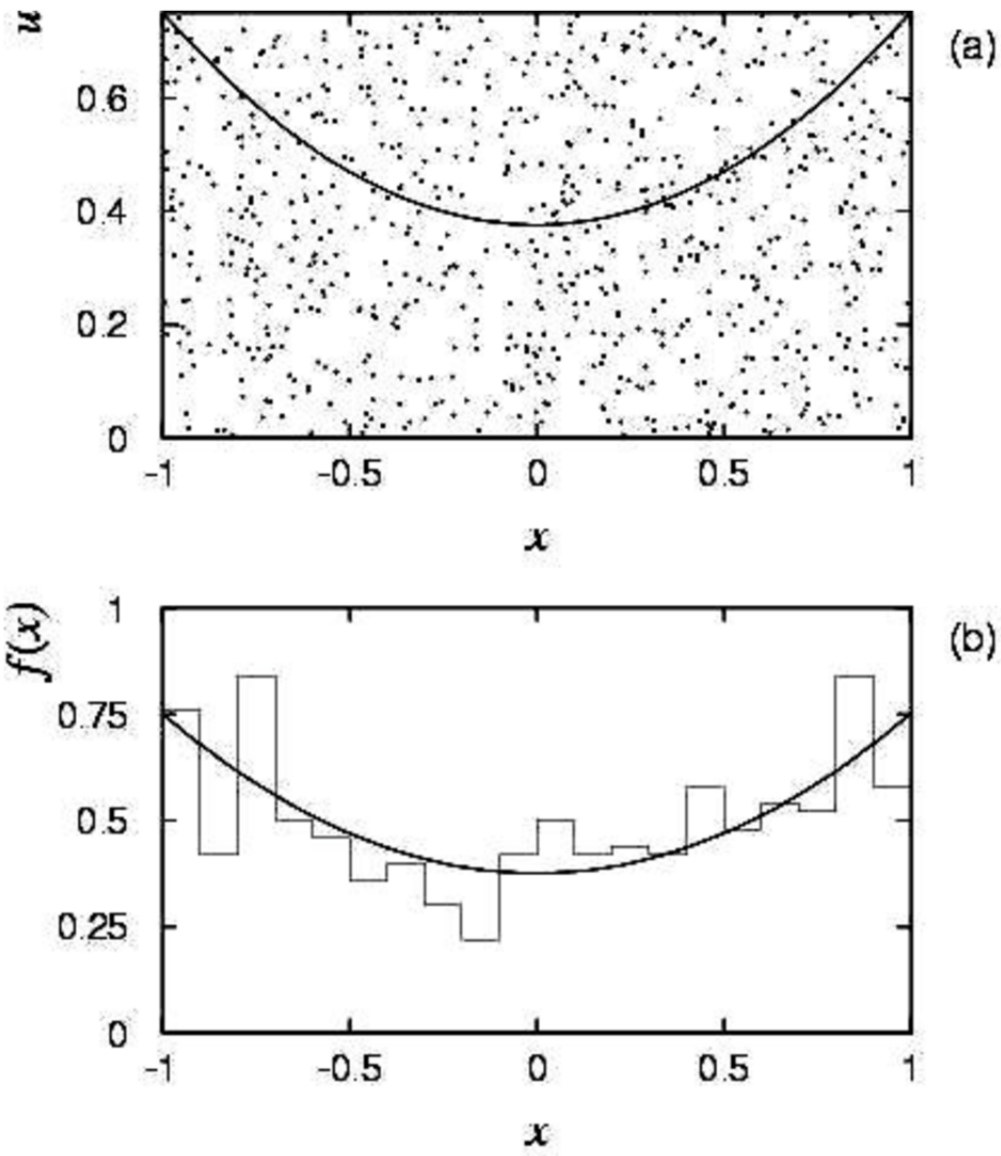
(2) Generiere eine 2te unabhängige Zufallszahl $u$ gleichverteilt zwischen 0 und $f_{max}$, i.e. $u = r_2 f_{\mathrm{max}}$ .

(3) Wenn $u < f(x)$, dann akzeptiere $x$. Wenn nicht, verwerfe $x$ and versuche es erneut.

# Die von-Neumannsche-Zurückweisungsmethode

$$f(x) = \frac{3}{8}(1 + x^2)$$

$$(-1 \leq x \leq 1)$$

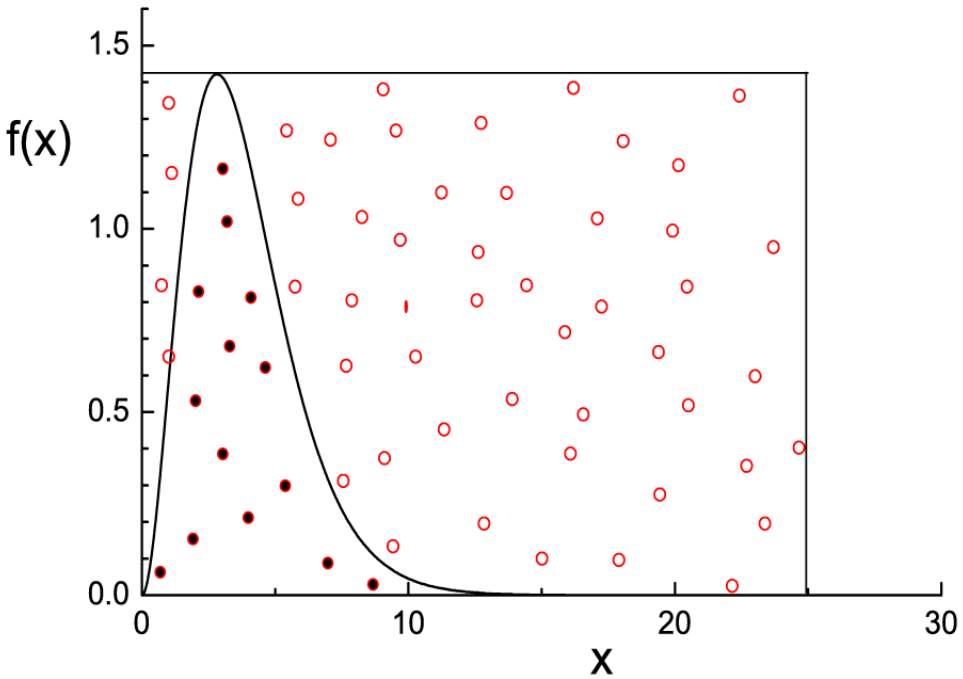



Wenn Punkt unterhalb der Kurve, dann behalte ihn und Fülle *x*-Wert in Histogramm.

# Effiizienzerhöhung für Zurückweisungmethode

Einfaches Beispiel:
Plankspektrum

$$f(x) = c\frac{x^3}{\mathrm{e}^x - 1}$$




Effizienz: hier ca. 10%
Weiteres Problem: Wertebereich x geht bis unendlich
Box wird nur bis $x_{max}$ gewürfelt.

# Majorantenmethode (importance sampling)

Suche geignete Funktion (Majorante) $m \geq f$ für alle $x$

Mit invertierbarer Stammfuktion M(x) $M(x) = \int_{-\infty}^{x} m(x')\mathrm{d}x'$

Zufallszahlen gemäß m(x) werden über erzeugt $x = M^{-1}(r)$

Erzeuge zweite Zufallszahl in Abhängigkeit von x im Bereich: null und $m(x)$

Behalte (verwerfe) diese wenn sie ≤ (>) f(x) ist

Bruchteil $[m(x) - f(x)]/f(x)$

der zweiten Zufallszahlen wird lokal verworfen

Besonders gut, wenn Majorante m(x) nahe an f(x) → große Effizienz
Vorteil: Generierung von Verteilungen die bis “unendlich gehen”

# Majorantenmethode: Beispiel

Zielfunktion: $$f(x) = c(\mathrm{e}^{-0.2x}\sin^2 x) \quad \text{für} \quad 0 < x < \infty$$

Geeignete Majorante: $$m(x) = c\,\mathrm{e}^{-0.2x}\ .$$

Bedingung für Kumulativfkt.: $$r = \int_0^x \frac{1}{0.2}\mathrm{e}^{-0.2x'}\,\mathrm{d}x' = 1 - \mathrm{e}^{-0.2x}\ .$$

Transformation von gleichverteilten r auf x gemäß Majorante m(x) $$x = -\frac{1}{0.2}\ln(1 - r_1)$$

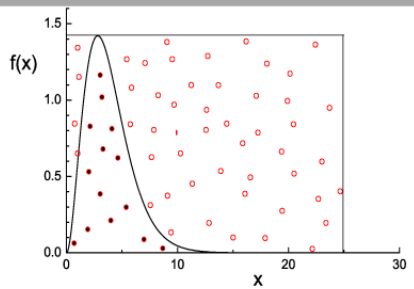
Würfele zweite Zufallszahl r2 und bilde Produkt: $r_2\, m(x)$

Akzeptanz/Zurückweisung gemäß:

$$\text{für } r_2 < \sin^2 x \quad \rightarrow \quad \text{behalte } x,$$
$$\text{für } r_2 > \sin^2 x \quad \rightarrow \quad \text{verwerfe } x$$

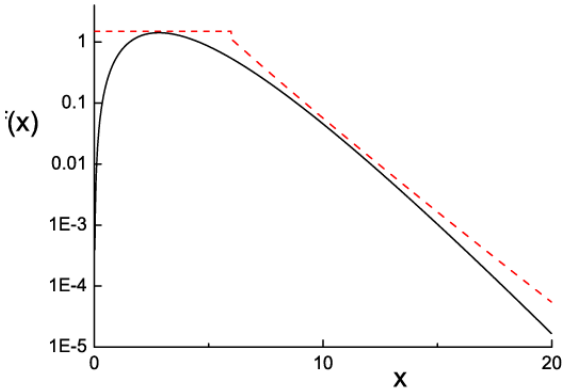
# Majorantenmethode: Beispiel 2

Zielfunktion: $$f(x) = c\frac{x^3}{\mathrm{e}^x - 1}$$



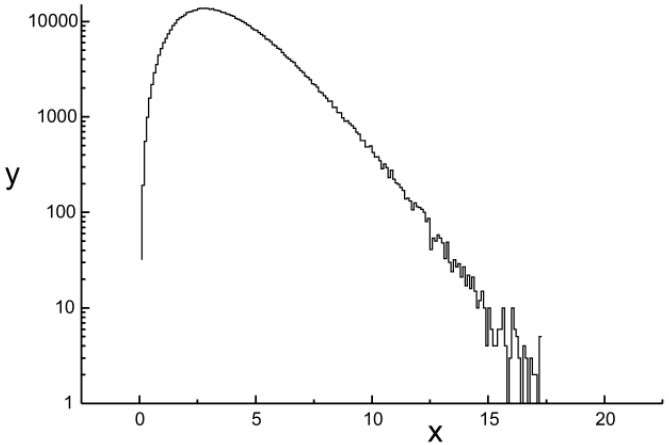

Stückweise Majorante: $x < x_1 \qquad m_1(x) = 6\,c$

$X^{-0.1}$ ermöglicht analytische Integration $\qquad x > x_1 \qquad m_2(x) = 200\,c\,x^{-0.1}\mathrm{e}^{-x^{0.9}}$.




Für $x > x_1$:

$$M_2(x) = \int_{x_1}^{x} m_2(x')\mathrm{d}x',$$
$$= \frac{200c}{0.9}\left[\mathrm{e}^{-x_1^{0.9}} - \mathrm{e}^{-x^{0.9}}\right]$$



# Behandlung additiver WDFs

Oft ist WDF Summe von mehreren Termen: $f(x) = f_1(x) + f_2(x)$

mit $S_1 = \int_{-\infty}^{\infty} f_1(x)\mathrm{d}x, \quad S_2 = \int_{-\infty}^{\infty} f_2(x)\mathrm{d}x \quad S_1 + S_2 = 1$

Wähle mit Wkt $S_1$ bzw $S_2$ ein Zufallszahl die nach $f_1(x)$ bzw. $f_2(x)$ verteilt ist.

Falls Stammfunktionen $F_1(x) = \int_{-\infty}^{x} f_1(x')\mathrm{d}x' \quad F_2(x) = \int_{-\infty}^{x} f_2(x')\mathrm{d}x'$ invertierbar

Dann generiere x aus gleichverteilten r gemäß

$$x = F_1^{-1}(r) \text{ für } r < S_1 \qquad x = F_2^{-1}(r - S_1) \text{ für } r > S_1$$

# Behandlung additiver WDFs: Beispiel

Ziel-WDF: $$f(x) = \varepsilon \frac{\lambda \mathrm{e}^{-\lambda x}}{1 - \mathrm{e}^{-\lambda a}} + (1-\varepsilon)\frac{1}{a} \quad \text{für} \quad 0 < x < a$$

Bestimme Stammfunktion und deren Inverse für beide Summanden.

Transformiere gleichverteilte Zufallszahl r gemäß:

$$r < \varepsilon \qquad x = \frac{-1}{\lambda} \ln\left(1 - \frac{1 - \mathrm{e}^{-\lambda a}}{\varepsilon} r\right)$$

$$r > \varepsilon \qquad x = a \frac{r - \varepsilon}{1 - \varepsilon}$$

Methode liefert 100% Effizienz.
Ohne Separation wäre Inverse hier nicht analytisch berechenbar gewesen.

Seperation auch meist sinnvoll, wenn Terme nicht analytisch invertierbar.

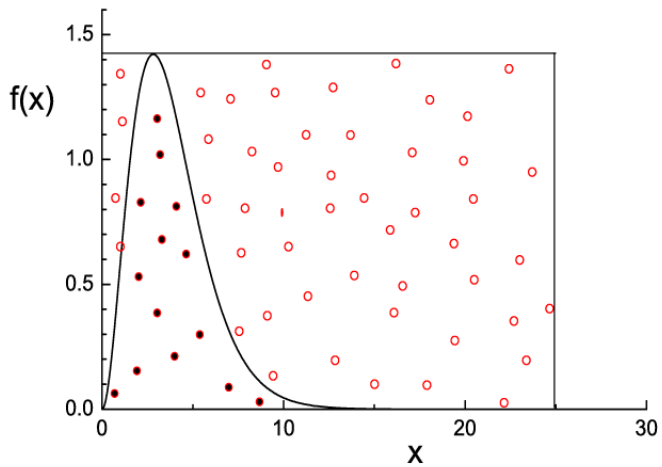
# MC-Integration

Inegrale bei der Integrand das Vorzeichen wechselt werden in Bereiche zwischen Nullstellen zerlegt. Dann gilt o.B.d.A:

$$I = \int_{x_a}^{x_b} y(x)\,\mathrm{d}x \quad \text{mit} \quad y > 0$$

Methode 1: Primitive Zurückweisungsmethode

$$\widehat{I} = I_0 \frac{N}{N_0}$$




N: Anzahl der Erfolge (akzeptierten Zufallszahlen)

$N_0$: Anzahl aller Zufallszahlen

$I_0$: Integral der constanten = Fläche der Box

Unsicherheit aus Binomialverteilung

$$\varepsilon = N/N_0 \qquad \delta N = \sqrt{N_0 \varepsilon (1-\varepsilon)}\,,$$

$$\frac{\delta I}{I} = \frac{\delta N}{N} = \sqrt{\frac{1-\varepsilon}{N}}$$

# Primitive Zurückweisungsmethode: Beispiel

Szintillationsausbeute beim Durchqueren ionisierender Strahlung durch Fieber

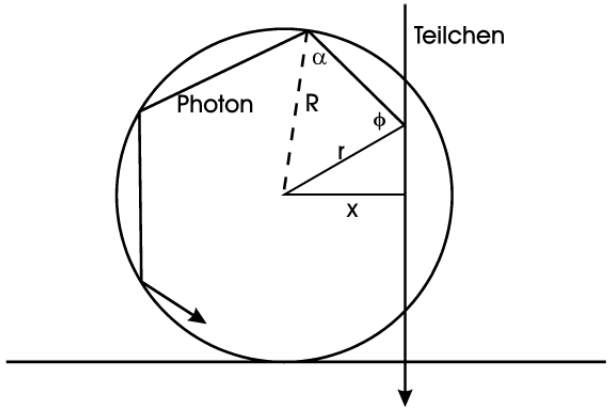
Photonen räumlich isotrop verteilt

Photonen werden totalreflektiert wenn

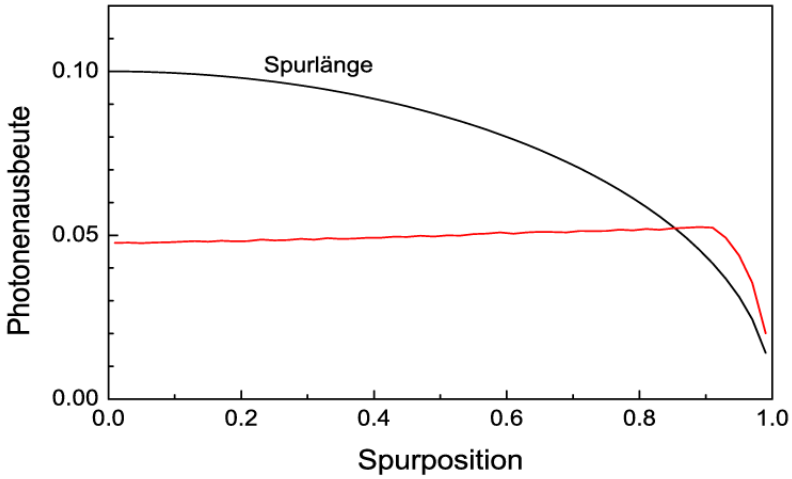
Winiel b zur Flächennormalen gr. 60° ist.

Wie groß ist Zahl der Eingefangen Photonen

Vom Durchgangsort x (R sei 1)



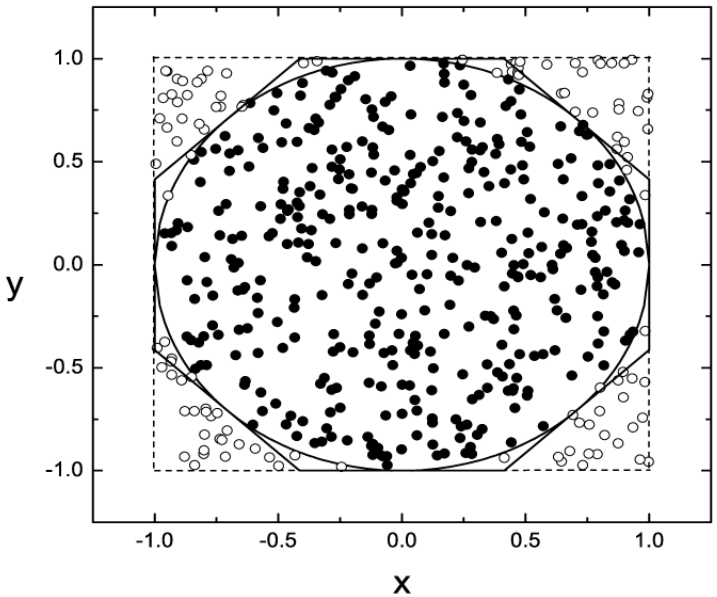

1) Gleichverteilt in 0≤x,y≤1
2) Für x2+x2>r verwerfe Ereignis
3) 3) Würfele Photonrichtung
4) Gleichförmig in 0,2pi
5) berechne Winkel a zur flächenonrmalen
6) Sin a = r sin phi



# Verbesserte Zurückweisungsmethode

Verkleinerung des Referenzbereichs durch einführen einer Majorante

Beispiel: Bestimmung von “pi”




Referenzbereich = Quadrat: $\langle N \rangle = \frac{\pi}{4} N_0 \quad \widehat{\pi} = \frac{4N}{N_0} \quad \frac{\delta\widehat{\pi}}{\pi} = \frac{\sqrt{1-\pi/4}}{\sqrt{N_0\pi/4}} \approx 0.52 \frac{1}{\sqrt{N_0}}$

Referenzbereich = Achteck → Reduzierung des Fehlers um Faktor ~2 bei gleicher Anzahl Versuche

# Verbesserte Zurückweisungsmethode

Zu bestimmen: Integral über y(x) $I = \int_{x_a}^{x_b} y(x)\mathrm{d}x$

Majorante m(x) für y(x) mit invertierbarer Stammfunktion $M(x) = \int_{x_a}^{x} m(x')\mathrm{d}x'$

Generiere Zufallszahlen $x_i$ gemäß m(x) durch Transformationsmethode

Erzeuge weitere gleichverteilte Zufallszahl $y_i$ im Bereich $0 < y < m(x_i)$

Zähle Anzahl der Erfolge N, definiert über $y_i \leq y(x_i)$

Integral ist dann gegeben durch: $I = M(x_b)\frac{N}{N_0}$

$$\delta N = \sqrt{N_0 \varepsilon (1-\varepsilon)}\,,$$

$$\frac{\delta I}{I} = \frac{\delta N}{N} = \sqrt{\frac{1-\varepsilon}{N}}$$

Fehler reduziert sich mit “Anschmiegen” der Majorante m(x) an y(x).

# Wichtungsmethode

Zu bestimmen: Integral über y(x) $I = \int_{x_a}^{x_b} y(x)\mathrm{d}x$

Würfele gleichverteilte Zufallszahlen im Intervall $x_a < x < x_b$

Bestimmte Mittelwert der Stichprobe für y(x): $\overline{y} = \sum_{i=1}^{N} y(x_i)/N$

Schätzwert für Integral gegeben durch: $\widehat{I} = (x_b - x_a)\overline{y}$

Entspricht numerischer Integration, aber Stützstellen zufällig.

Unsicherheit aus Varianz von y(x): $(\delta\overline{y})^2 = \frac{1}{N}\int_{x_a}^{x_b} (y(x) - \langle y \rangle)^2 y(x)\mathrm{d}x / \int_{x_a}^{x_b} y(x)\mathrm{d}x$

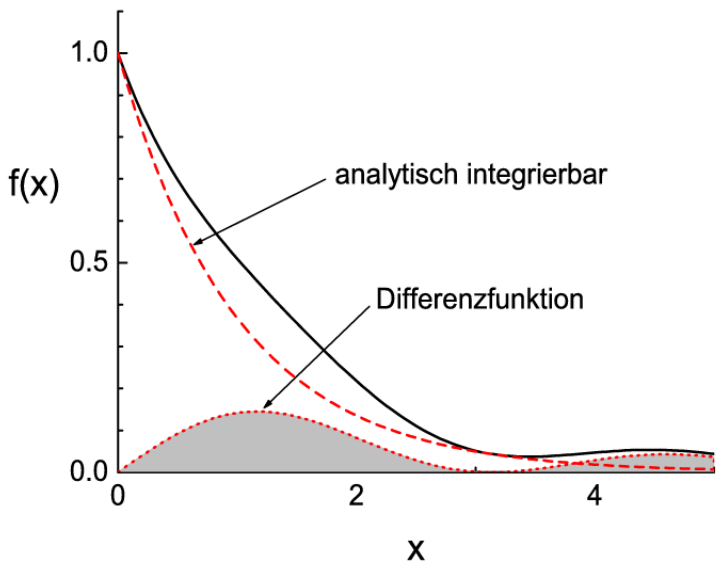
Bei MC-Integration abgeschätzt über:

$$(\delta\overline{y})^2 \approx \frac{1}{N(N-1)} \sum_i (y(x_i) - \overline{y})^2 \qquad \frac{\delta\hat{I}}{\hat{I}} = \frac{\delta\overline{y}}{\overline{y}}$$

# Subtraktionsmethode

Wenn wir Funktion $\widetilde{y}(x)$ kenne, die analytisch integrierbar ist und nur wenig von $y(x)$ abweicht, lässt sich Integral umschreiben auf:

$$\int_{x_a}^{x_b} y(x)\mathrm{d}x = \int_{x_a}^{x_b} \widetilde{y}(x)\mathrm{d}x + \int_{x_a}^{x_b} \left(y(x) - \widetilde{y}(x)\right)\mathrm{d}x$$




Es muss nur noch die Differenzfunktion (schattierte Fläche) mit der Gewichtungsmethode integriert werden.

# Rückführung auf Erwartungswerte

Oft Faktorisierung des Integranden sinnvoll: $y(x) = f(x)y_1(x)$

$f(x)$ WDF, nach der einfach Zufallszahlen erzeugt werden können

Integral ergibt sich als Erwartungswert in der Form:

$$\int_{x_a}^{x_b} y(x)\mathrm{d}x = \int_{x_a}^{x_b} f(x)y_1(x)\mathrm{d}x$$
$$= \langle y_1 \rangle \, .$$

Geschätzt wird der Wert durch den Stichprobenmittelwert:

$$\widehat{I} = \frac{\sum_i y_1(x_i)}{N}$$

mit $x_i$ gemäß $f(x)$ erzeugt.

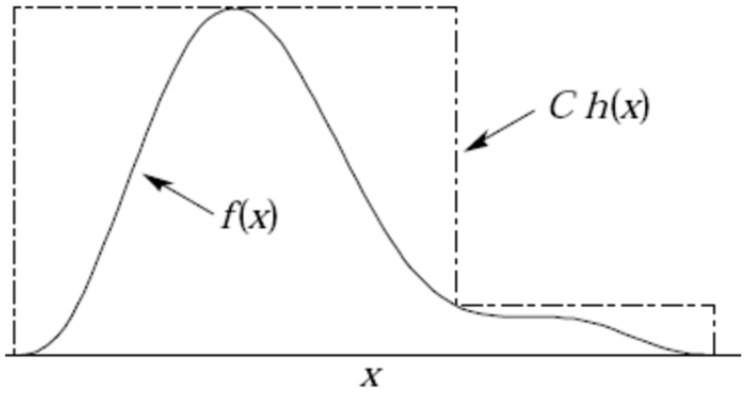
Fehler ergibt sich wiederum über: $(\delta\bar{y})^2 \approx \frac{1}{N(N-1)} \sum_i (y(x_i) - \bar{y})^2$ $\quad \frac{\delta\hat{I}}{\hat{I}} = \frac{\delta\bar{y}}{\bar{y}}$

# Effiizienzerhöhung für Zurückweisungmethode

Der Anteil akzeptierter Punkte ist gleich dem Verhältnis des Intergrals über f(X) zur Fläche der Box.

Für Verteilungen mit grüßen “Spitzen” mag die Effizienz des Algorithmus sehr gering werden,

Verbesserung: Fine Improve by enclosing the pdf $f(x)$ in a curve $C\,h(x)$ that conforms to $f(x)$ more closely, where $h(x)$ is a pdf from which we can generate random values and $C$ is a constant.




Generate points uniformly over $C\,h(x)$.

If point is below $f(x)$, accept $x$.

# Kapitel 4

# Grundlagen der Parameterschätzung

# Grundgesamtheit, Stichprobe, Statistik, Schätzer

Betrachte ZV x mit WDF $f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}$ oft abhängig von Parametern

Grundgesamtheit = Menge möglicher Ergebnisse beschrieben durch f(x)
Stichprobe vom Umfang n = Satz von n unabhängigen Messungen der Zv x

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

Ziel: Ableitung von Eigenschaften von f(x) aus der Stichprobe

Statistik: beliebige Funktion der Stichprobenwerte der ZV

Schätzer: Statistik um Eigenschaften der WDF zu bestimmen

a) Form der WDF unbekannt → Schätzwerte für Erwartungswert, Varianz, …
b) Form der WDF bekannt → Schätzwerte für Parameter $\hat{\theta}(\vec{x})$

Schätzer = Funktioneller Zusammenhang, Schätzwert = numerischer Wert
beide werden mit “Hut ^” beschrieben

# Geminsame WDF der Stichprobe

Betrachte $n$ Beobachtungen der ZV $x$: $x_1, \ldots, x_n$,
wobei $x$ der WDF $f(x; \theta)$ folgt.

Unter den Annahmen:

1) Messungen sind unabhängig
2) WDF/Grundgesamtheit ändert sich zwischen Beobachtung nicht

folgt die gemeinsame WDF der Stichprobe:

$$f(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Achtung! Ist nicht immer erfüllt in der Praxis!

Gegenbeispiele:
- Lotterie (Ziehen von Kugeln ohne zurücklegen)
- Längenmessung von Stab bei Temperaturschwankungen

# Die Likelihoodfunktion

Das Ergebnis eines Experimente (Satz von Messungen) $\vec{x} = (x_1, \ldots, x_n)$ sei eine Menge von Zahlen ***x***, und die gemeinsame WDF der Stichprobe ist eine Funktion/Statistik, welche von den Parametern der WDF ***θ***: abhängt:

$$f(\vec{x}; \vec{\theta})$$

Nun werte diese Funktion mit den Werten der Stichprobe aus und betrachte sie als Funktion der Parameter:

$$L(\vec{\theta}) = f(\vec{x}; \vec{\theta})$$ Likelihoodfunktion

(***x*** konstant)

Im Falle von unabhängigen, identischen Messungen gilt:

$$L(\vec{\theta}) = \prod_{i=1}^{n} f(x_i; \vec{\theta})$$ ($x_i$ konstant)

# Schätzer für den Erwartungswert μ von f(x)

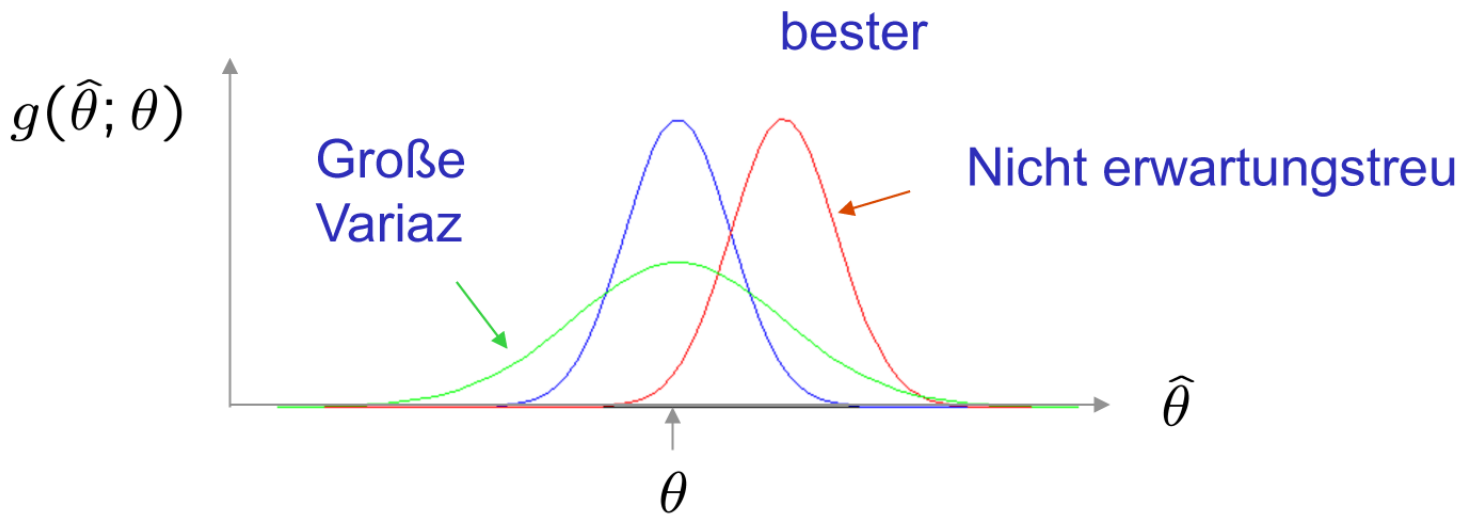
## Mögliche Schätzer für den Erwartungswert der WDF der Grundgesamtheit

| | |
|---|---|
| $\widehat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ | den Mittelwert der Stichprobe |
| $\widehat{\mu} = \frac{1}{10}\sum_{i=1}^{10} x_i$ | den Mittelwert der ersten 10 Punkte der Stichprobe |
| $\widehat{\mu} = \frac{1}{n-1}\sum_{i=1}^{n} x_i$ | $n/(n-1)$ mal den Mittelwert der Stichprobe |
| $\widehat{\mu} = 42$ | |
| $\widehat{\mu} = (\min(x_i) + \max(x_i))/2$ | Mittelwert des größten und kleinsten Wertes |
| $\widehat{\mu} =$ Median der Stichprobe | |

Fragen: - welcher Schätzer ist gut, der “Beste”?
- welche Kriterien sollte ein guter Schätzer erfüllen?
- wie findet man den optimalen Schätzer für ein Problem?

# Eigenschaften von Schätzern

Wenn mir das Experiment (aus m Messungen bestehend) oft wiederholen.
Würden die Schätzwerte einer WDF folgen:




Wir wollen kleinen (oder Null) Bias (systematischen. Fehler): $b = E[\hat{\theta}] - \theta$

$\rightarrow$ Mittelwert der wiederholten Messungen sollte = wahren Wert sein

Und wir wollen kleine Varianz (statistischer Fehler): $V[\hat{\theta}]$

$\rightarrow$ kleiner Bias und keine Varianz sind i.a. gegenläufige Kriterien

# Eigenschaften von Schätzern

Aus der Informationstheorie lässt sich zeigen, dass es eine untere Schranke für die Varianz eines Schätzers für einen Parameter gibt

$$V\left[\hat{\theta}\right] \geq \frac{\left(1+\frac{\partial b}{\partial \theta}\right)^2}{E\left[-\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right]}. \qquad V\left[\hat{\theta}\right] \geq \frac{\left(1+\frac{\partial b}{\partial \theta}\right)^2}{E\left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta}\right)^2\right]}.$$

**Schranke Minimaler Varianz (SMV)**

**Rao-Cramér-Frechet-Ungleichung**

Information nach R.A Fisher: $I(\theta) \equiv E\left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta}\right)^2\right] = E\left[-\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right]$

$\rightarrow$ je größer die Information, desto kleiner der statistische Fehler.

# Kriterien für gute Schätzer

**Konsistenz** $\widehat{\theta}^{(n)} \xrightarrow{n\to\infty} \theta$.

für jedes $\varepsilon > 0$ gilt $\lim_{n\to\infty} P(|\widehat{\theta}^{(n)} - \theta| > \varepsilon) = 0$.

**Verzerrung (Bias)** $b^{(n)} = E[\widehat{\theta}^{(n)}] - \theta$

möglichst keinen Bias d.h. Schätzer ist erwartungstreu.
Konsistente Schätzer mit endlicher Varianz sind asymptotisch (n→unendlich) erwartungstreu

**Effizienz**

$$\text{Effizienz}\left[\widehat{\theta}^{(n)}\right] = \frac{SMV}{V\left[\widehat{\theta}^{(n)}\right]}$$

Effizienz sollte nahe an “1” sein.

**Robustheit** Varianz des Schätzers ist unabhängig von der WDF der Grundgesamtheit

# Ein Schätzer für den Erwartungswert

Parameter: $\mu = E[x]$

Schätzer: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \equiv \overline{x}$ ('Stichprobenmittelwert')

Man kann zeigen: ist konsistent

$b = E[\hat{\mu}] - \mu = 0$ ist erwartungstreu

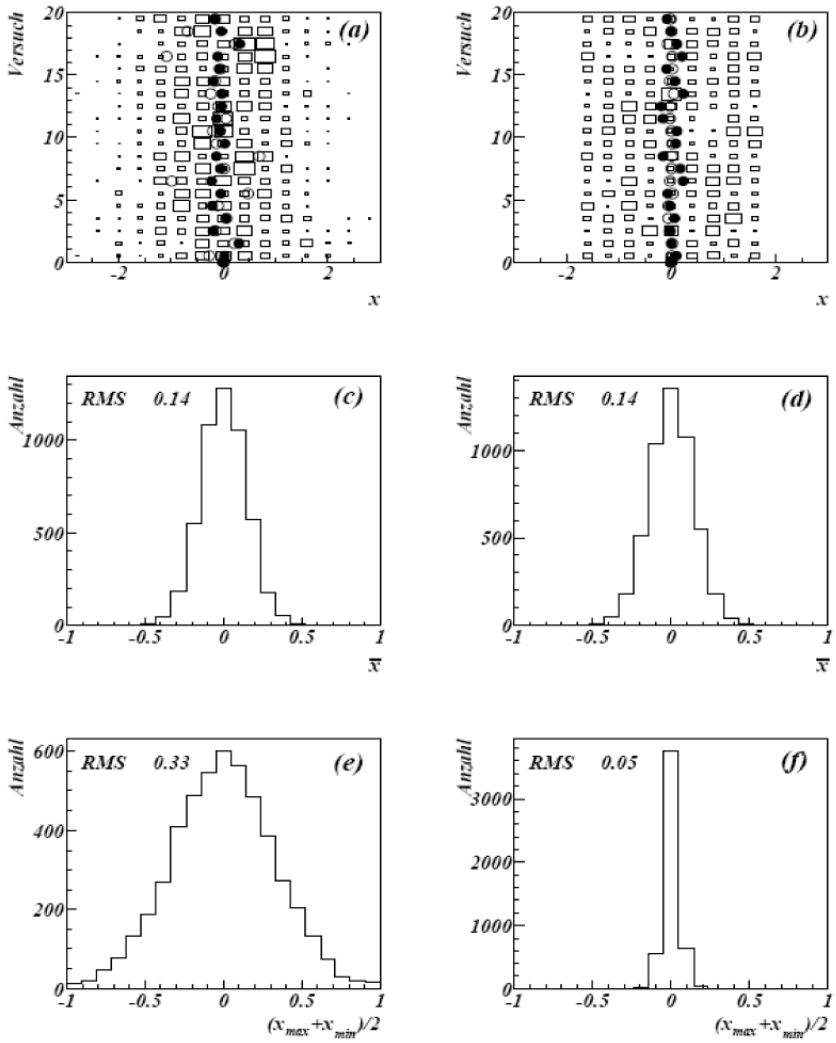
ist effizient für Gauss-WDF (aber nicht für alle WDFs)

Der Fehler auf den Schätzer für den Erwartungswert ist gegeben durch

$$V[\hat{\mu}] = \frac{\sigma^2}{n} \qquad \left( \sigma_{\hat{\mu}} = \frac{\sigma}{\sqrt{n}} \right)$$

# Vergleich von 2 Schätzern für den Erwartungswert

Gauss-WDF | Gleichverteilung




$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \equiv \overline{x}$$

robust

$$\hat{\mu} = (\min(x_i) + \max(x_i))/2$$

effizienter für Gleich-verteilung

# Ein Schätzer für die Varianz

Parameter: $\sigma^2 = V[x]$

Schätzer: $$\widehat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \equiv s^2$$ ('Stichprobenvarianz')

Man kann zeigen: ist konsistent (mit und ohne Besselkorrektur (*n/n*-1)

$$b = E[\widehat{\sigma^2}] - \sigma^2 = 0$$

erwartungstreu (ohne *n/n*-1 Korrektur nur asympotisch erwartungstreu)

Der Fehler auf den Schätzer für die Varianz ist gegben durch

$$V[\widehat{\sigma^2}] = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\mu_2\right) , \qquad \mu_k = \int (x-\mu)^k f(x)\, dx$$