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

Anknüpfung an letzte Vorlesung

Axiomatische Defintion:

For all $A \subset S, P(A) \geq 0$

$$P(S) = 1$$

If $A \cap B = \emptyset, P(A \cup B) = P(A) + P(B)$

bedingte Wahrscheinlichkeit:
A gegeben *B* (mit *P(B) ≠ 0)*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Bayes-Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Interpretation der Wahrscheinlichkeit

2 Schulen:

a) Frequentisten / Klassische Wahrscheinlichkeit

b) Bayesianer / Subjektive Wahrscheinlichkeit

Frequentisten: Wahrscheinlichkeit = Grenzfall der relativen Häufigkeit

$$P(A) = \lim_{n\to\infty} \frac{\text{Anzahl der Ergebnisse } A \text{ in } n \text{ Messungen}}{n}$$

Nur anwendbar auf wiederholbare Ereignisse/Phänomene
Keine Aussagen über wahre Werte

Bayesianer: Wahrscheinlichkeit = Grad des Glaubens (Zuversicht), dass Hypothese A wahr ist

Kann allem Wahrscheinlichkeiten zuordnen.
Apriori Wahrscheinlichkeit im Prinzip beliebig.

Bayesianische Statistik: Generelle Philospohie

In Bayesianischer Statistik: weise "subjektive Wahrscheinlichkeit" = "Grad des Glaubens" den verschiedenen Hypothesen zu:

Wkt. solche Daten zu beobachten unter der Annahme der Hypothese *H* (the likelihood)

A-Priori-Wkt., i.e. vor der Datennahme

$$P(H|\vec{x}) = \frac{P(\vec{x}|H)\pi(H)}{\int P(\vec{x}|H)\pi(H)\,dH}$$

Posterior-Wkt., i.e. nach Auswertung der Daten

Normierung beinhaltet Summe über alle möglichen Hypothesen

Bayes-Theorem hat einen "wenn-dann"-Charakter:
Wenn Ihre A-Priori-Wkt $\pi(H)$ wären, dann sagt es Ihnen wie diese Wkt. sich im Licht der Beobachtung/Daten ändern.

Keine generelle Regel für Wahl der A-Priori-Wkt! →Subjektiv!

Bemerkung zu zwei Schulen

Subjektivität in der Annahme der A-Priori-Wahrscheinlichkeit in der Bayesianischen Schule erscheint vielen als nicht "wissenschaftlich". Das ist falsch! Aber Abhängigkeit von A-Priori-Wahrscheinlichkeiten enthält gewisse Beliebigkeit.

Häufige Annahme: flach=gleichverteilt
Aber in welcher Größe?
z.B. Erkrankung unabhängig vom Alter oder proportional zum Alter?

In Physik (Naturwissenschaften) frequentische Interpretation oft nützlicher, aber subjektive Wahrscheinlichkeit kann eine "natürlichere" Behandlung nicht wiederholbarer Messungen/ sich nicht wiederholender Phänomene liefern:
z.B. - systematische Fehler
- Wahrscheinlichkeit, dass ein neues Phänomen existiert, …

Kapitel 1

Grundlegende Konzepte der Wahrscheinlichkeitstheorie (Fortsetzung)

Zufallsvariablen und Wkt.dichteverteilungen

Eine Zufallsvariable (ZV) ist eine numerische Charakeristik,
die einem Element der Grundgesamtheit zugeordnet wird.
ZV können diskret oder kontinuierlich sein.

Annahme: das Ergebnis eines Experimentes ist kontinuierliche ZV x
Wkt. Dafür, genau den Wert x_i zu messen ist beliebig klein.

$$\text{Wahrscheinlichkeit } x \text{ im Intervall } [x, x+dx] \text{ zu erhalten } = f(x)dx.$$

→ $f(x)$ = Wahrscheinlichkeitsdichtefunktion (WDF)

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \quad x \text{ muss irgendwo sein}$$

Für diskrete Zufallsvariable mit Ergebnissen x_i mit z.B. $i = 1, 2, \ldots$

$$P(x_i) = p_i \quad \text{Wahrscheinlichkeits-Massenfunktion}$$

$$\sum_i P(x_i) = 1 \quad x \text{ muss einen der möglichen Werte annehmen}$$

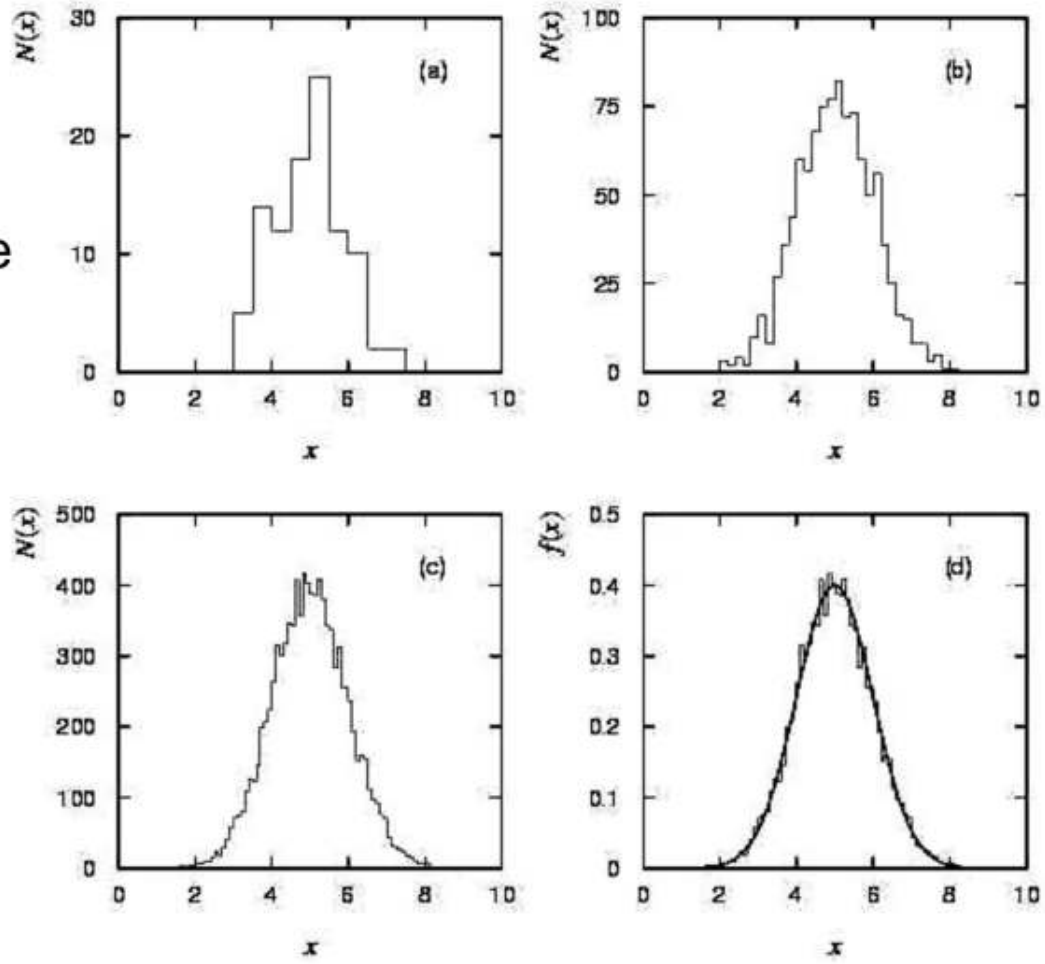
Approximative WDF aus Histogramm

WDF = Histogramm mit
unendlich vielen Einträgen
(unendlich großes Sample),
und verschwindender (0) Binbreite
normiert auf “eins”

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

$n =$ number of entries

$\Delta x =$ bin width

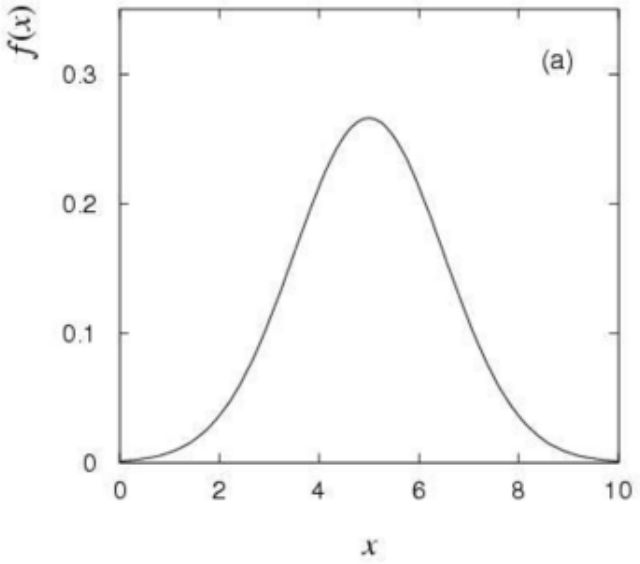


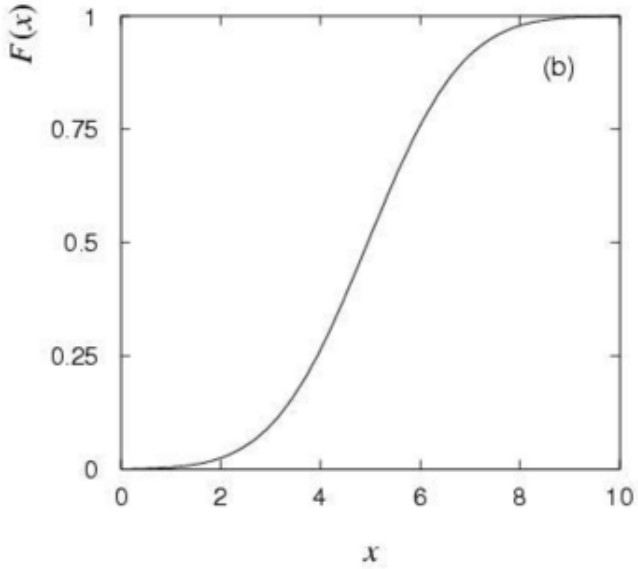
Kumulativverteilung

Wahrscheinlichkeit ein Ergebnis kleiner oder gleich *x* zu erhalten

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

Kumulativverteilung





Alternativ kann WDF über $f(x) = \dfrac{\partial F(x)}{\partial x}$ definiert werden

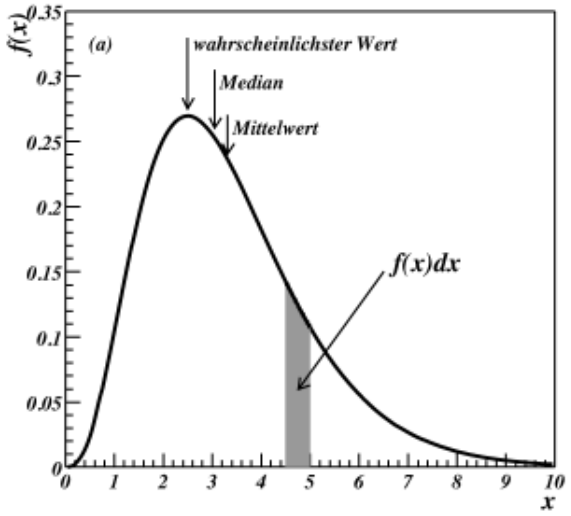
Für diskrete ZV gilt $F(x) = \sum_{x_i \le x} P(x_i)$

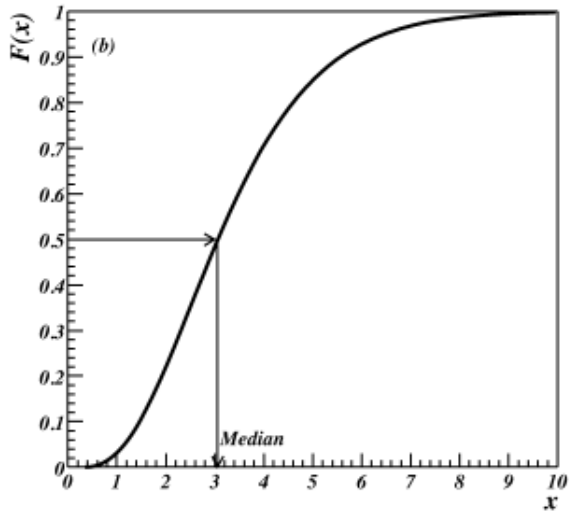
Quantile

Das Quantil x_α ist der Wert der ZV x für den die Wahrscheinlichkeit, dass die Messung einen Wert kleiner x_α ergibt gerade α ist.

$$F(x_\alpha) = \int_{-\infty}^{x_\alpha} f(x)dx = \alpha \Rightarrow x_\alpha = F^{-1}(\alpha)$$

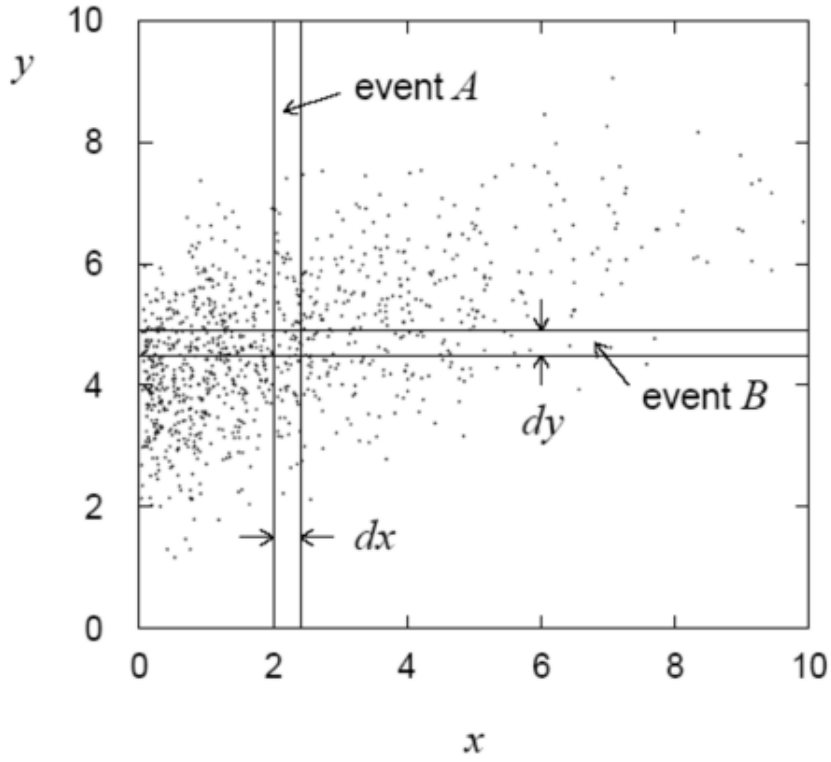
Das Quantil zu α=0.5 ist der Median $x_{1/2}$





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)

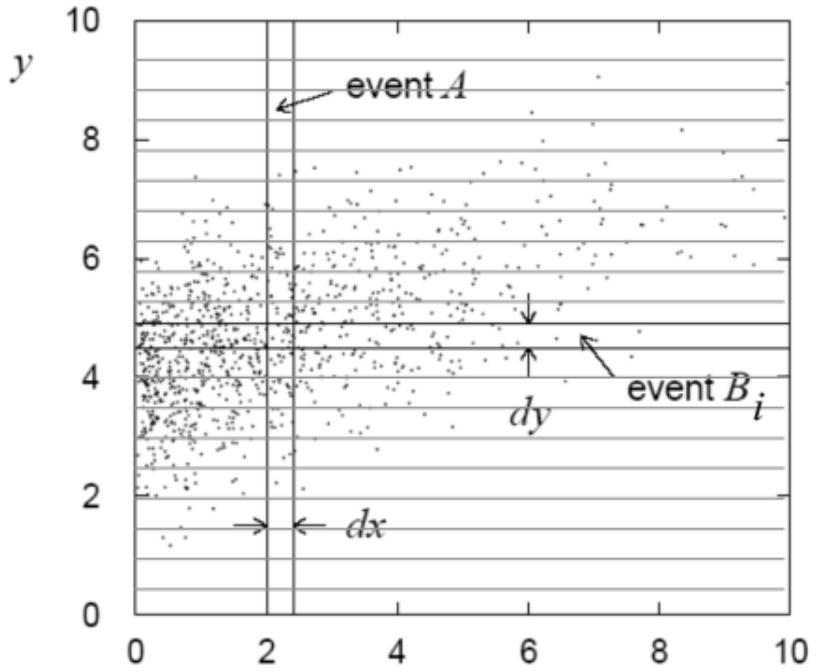
$$f(x,y)dxdy = \text{Wahrscheinlichkeit, dass } x \in [x, x+dx] \text{ und } y \in [y, y+dy]$$

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:

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



Integration über uninteressante ZV ("Marginalisieren")

→ Marginal-WDF oder Randverteilung

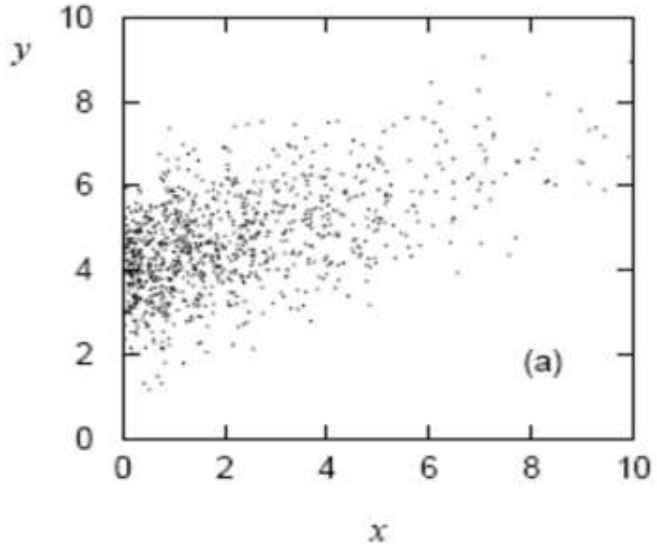
$$f_1(x_1) = \int \cdots \int f(x_1, \ldots, x_n)\, dx_2 \ldots dx_n$$

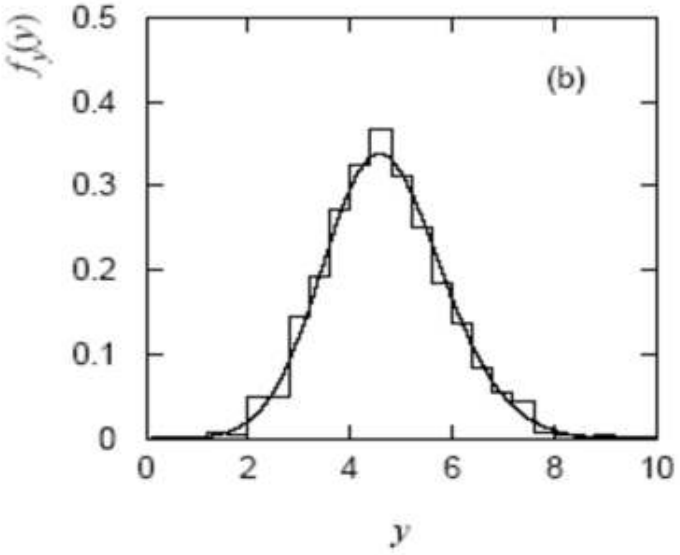
$$f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

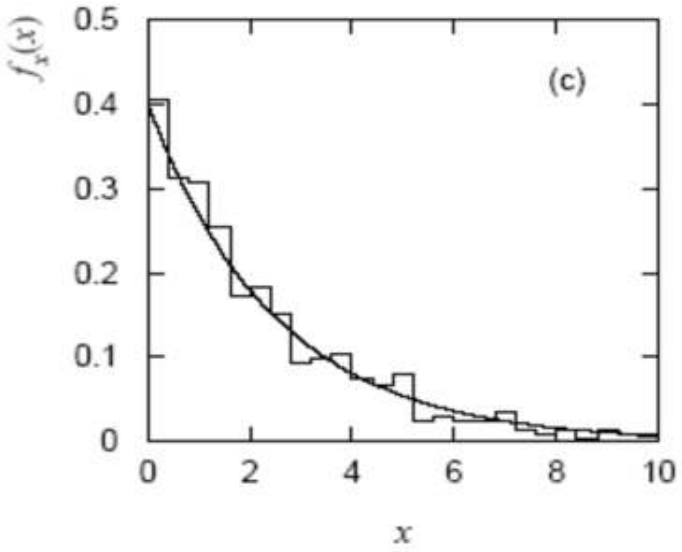
$$f_y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

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:

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

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

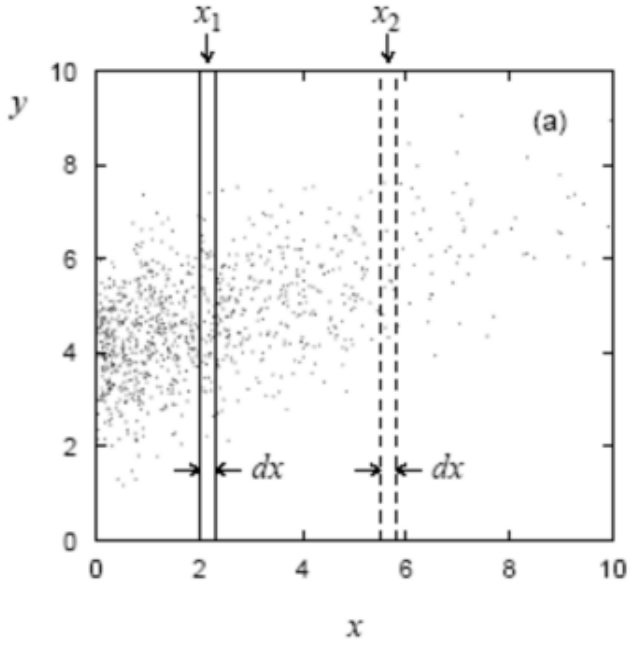
Bayes-Theorem wird zu: $$g(x|y) = \frac{h(y|x) f_x(x)}{f_y(y)}\,.$$

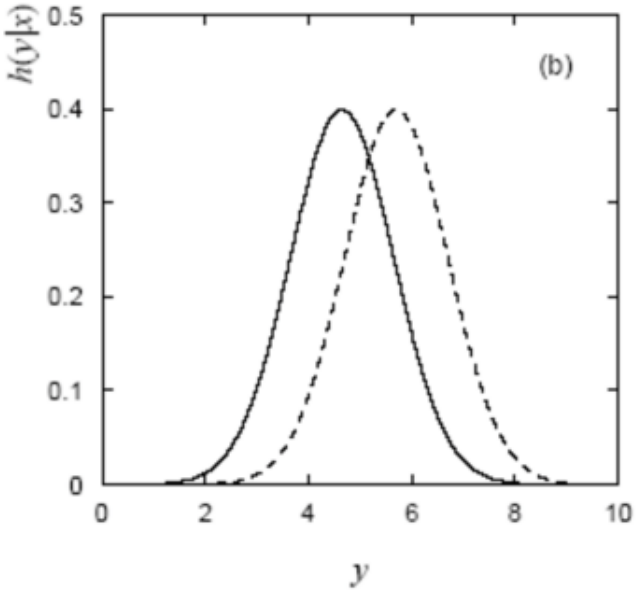
Erinnerung: *A*, *B* unabhängig, wenn $P(A \cap B) = P(A)P(B)\,.$

→ *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



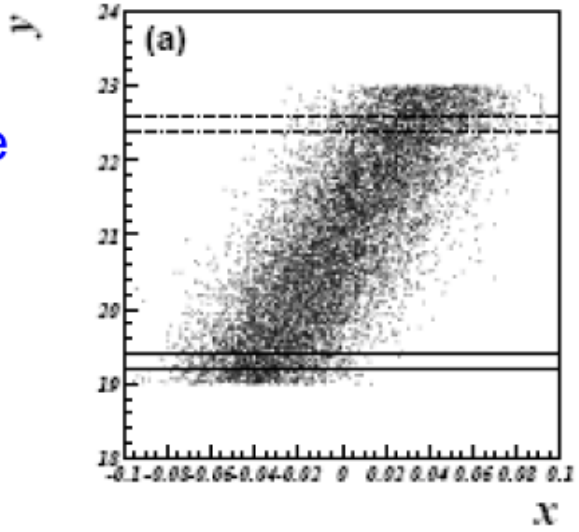


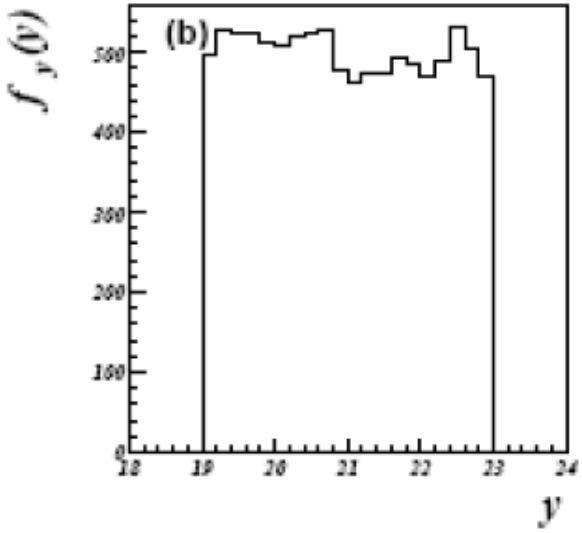
$$\int h(y|x)\,dy = 1\ .$$

- 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

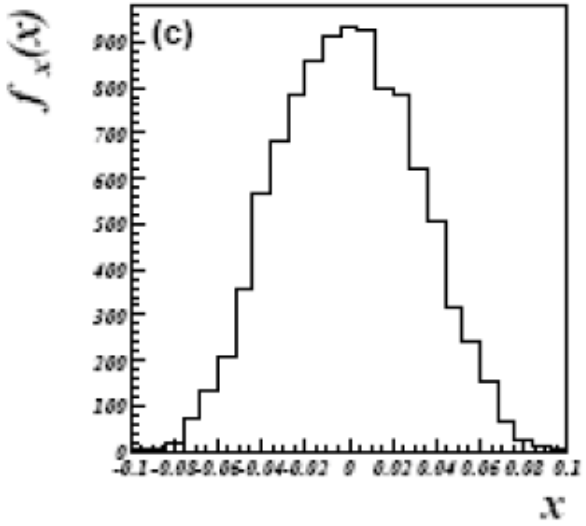
2-dim. gemeinsame WDF

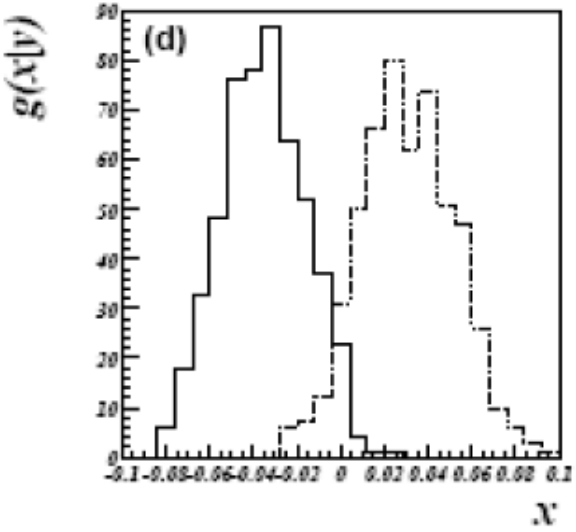




Rand-verteilung für y

Rand-verteilung für x





Bedigte WDFs für 2 Y-Werte

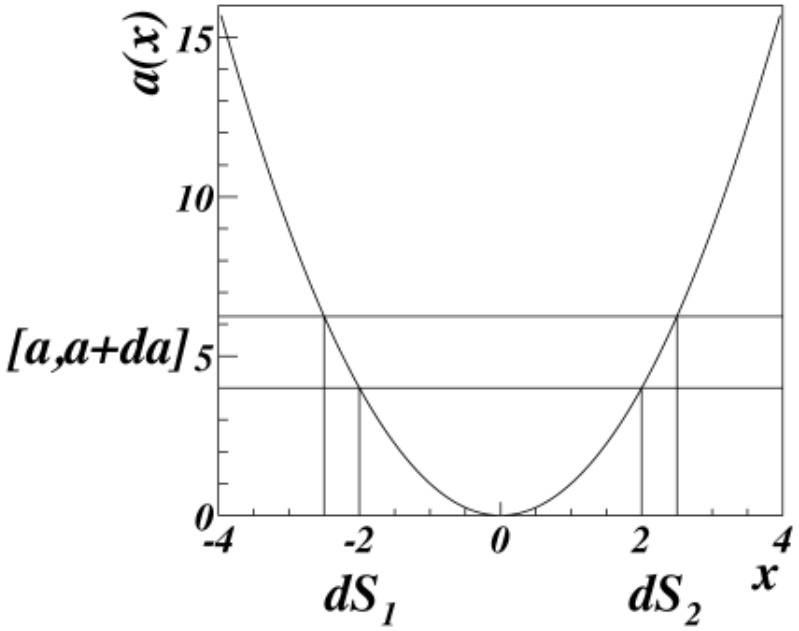
Funktionen / Transformation von Zufallsvariablen

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:

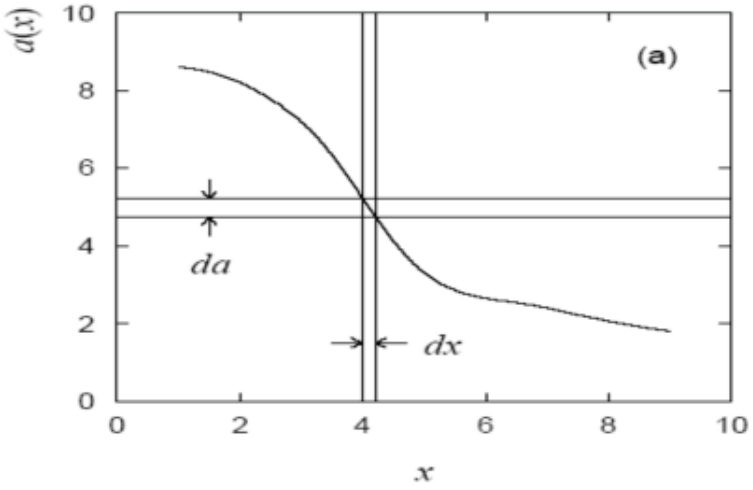
$$g(a)\,da = \int_{dS} f(x)\,dx$$

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 $[x(a), x(a + da)]$



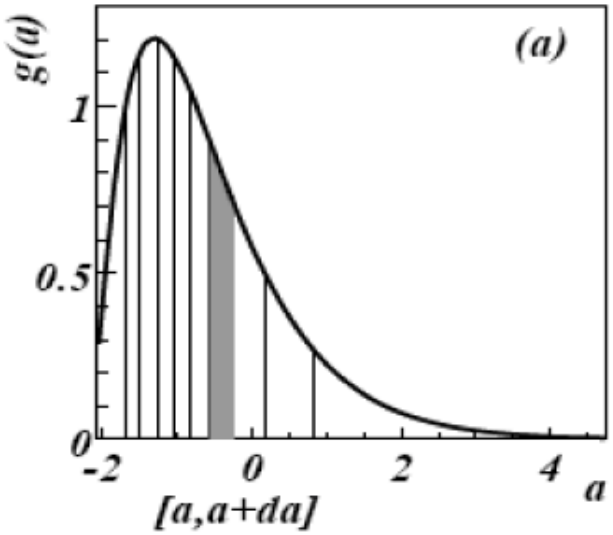
Falls gilt $x(a) > x(a + da)$ wäre Integral < 0 → Betragsstriche

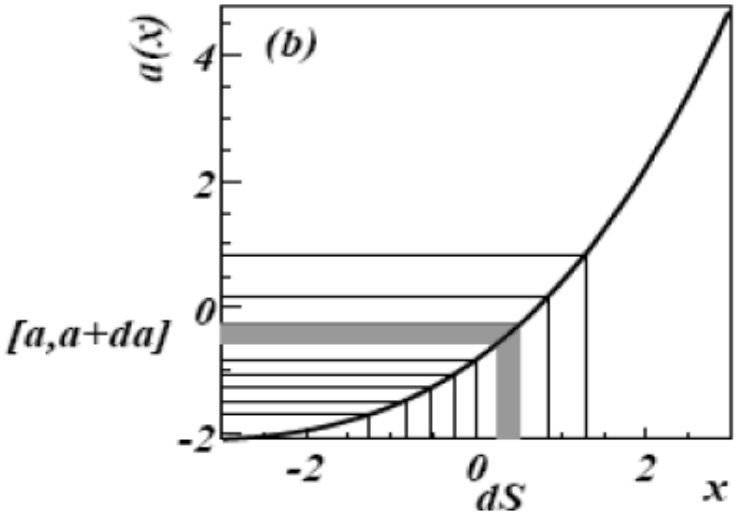
$$g(a)da = \left| \int_{x(a)}^{x(a+da)} f(x')dx' \right| = \int_{x(a)}^{x(a)+\left|\frac{dx}{da}\right|da} f(x')dx' = f(x(a)) \left| \frac{dx}{da} \right| da$$

Also gilt zusammengefasst:

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

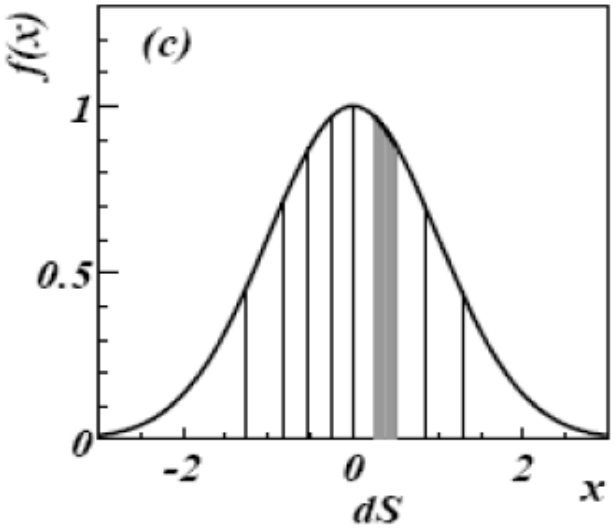
Transformation von Zufallsvariablen (III)



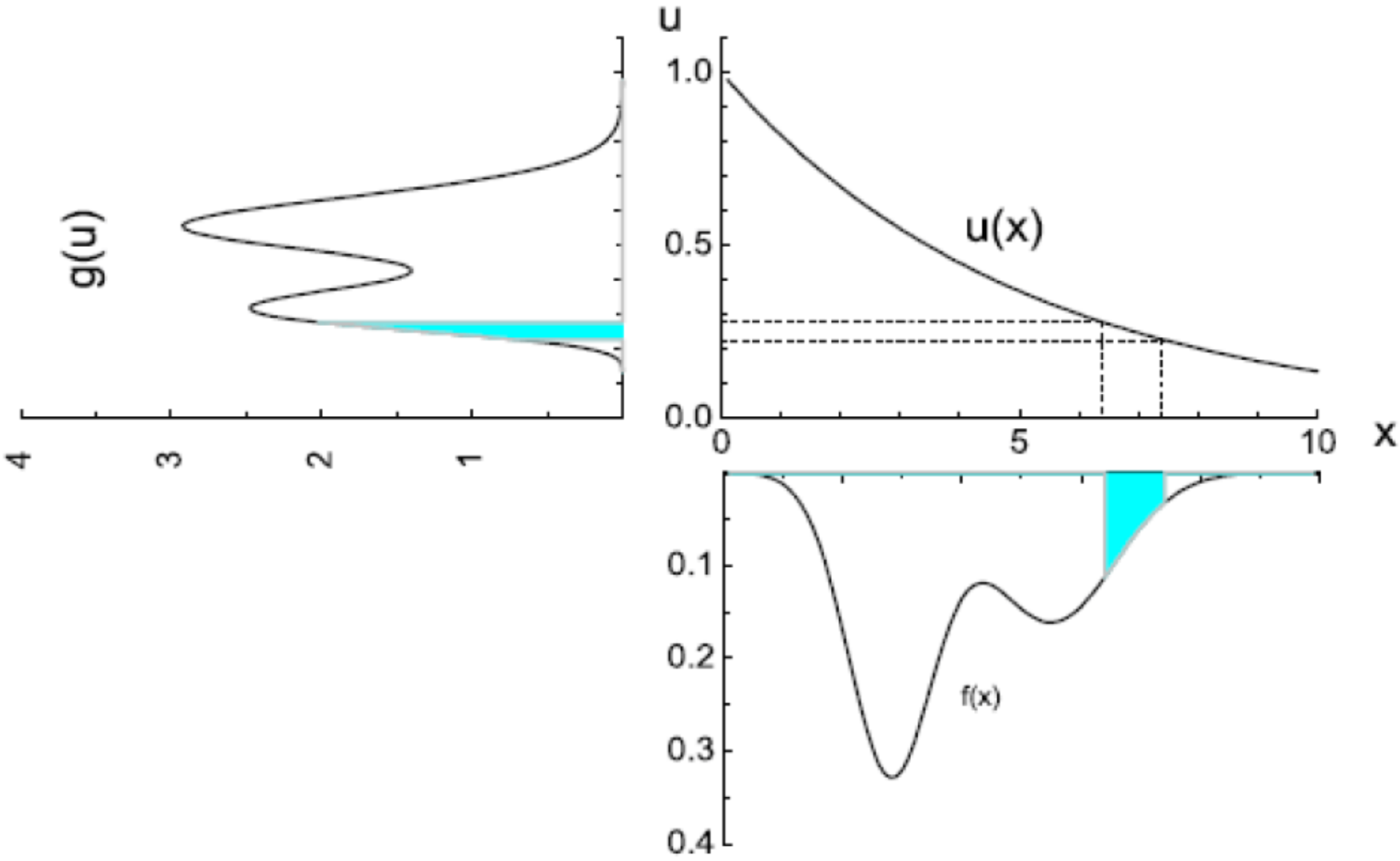


Wahrscheinlichkeit im a-Raum g(a)da

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

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

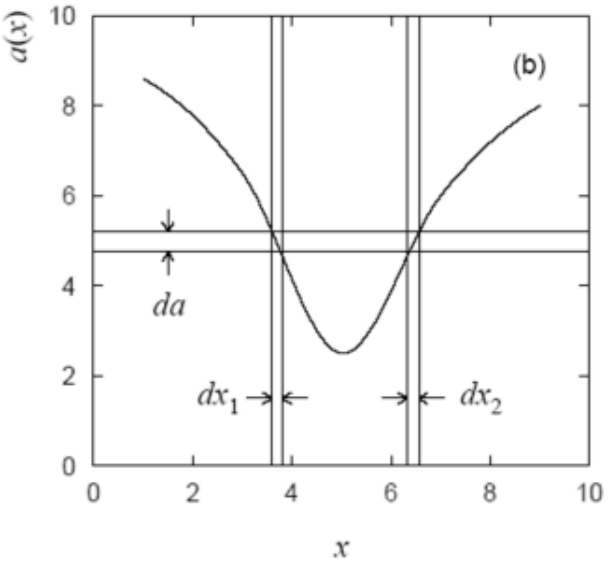
Jakobifaktor / Differential: $\frac{dx}{da} = \frac{1}{a} > 0$

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

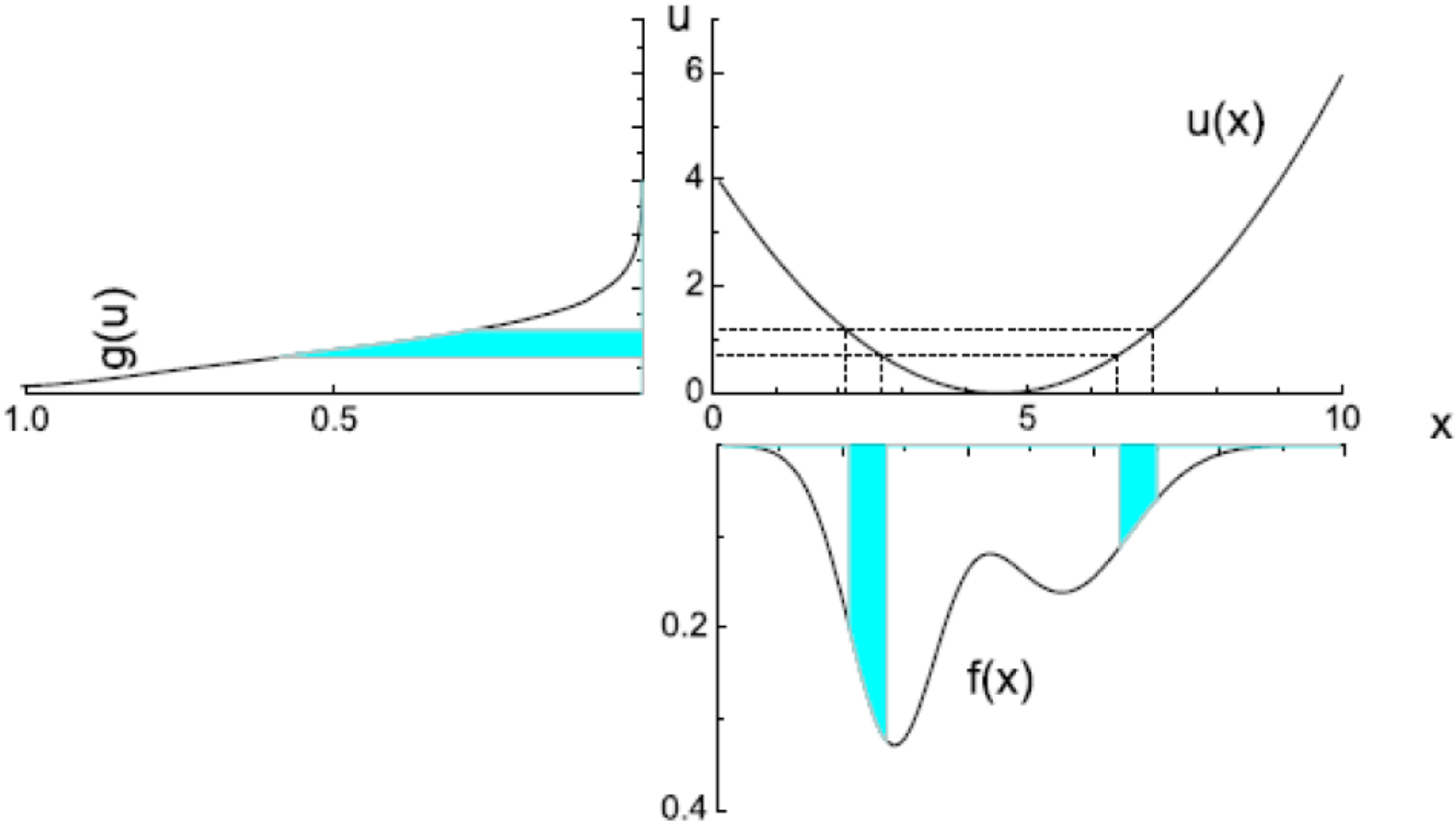


Beispiel: $a = x^2, \;\; x = \pm\sqrt{a}, \;\; dx = \pm\frac{da}{2\sqrt{a}} \; .$

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

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

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:

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

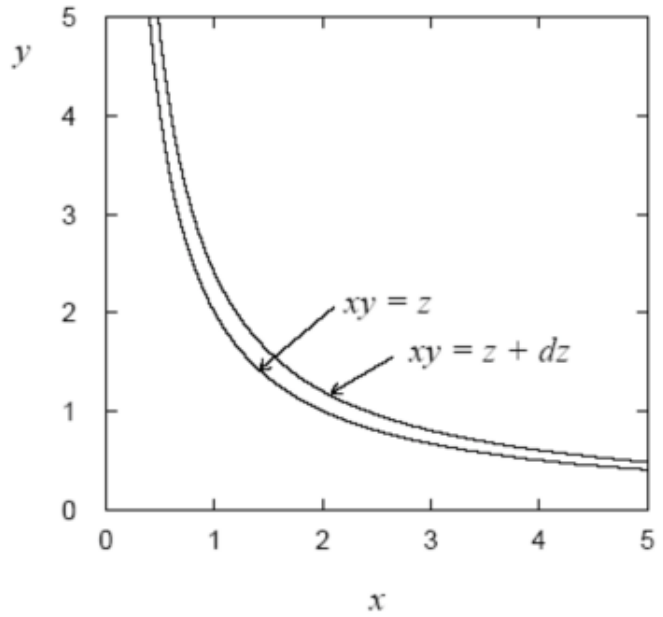
Wobei *dS* = Region im *x*-Raum zwischen den Hyperflächen,
Die definiert werden durch

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

Funktionen von mehreren ZV: Mellin-Faltung

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

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



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

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

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

$$= \int_0^\infty f(\frac{z}{y}, y)\,\frac{dy}{y}$$

(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

$$
\begin{aligned}
g(a)da &= \int_{-\infty}^{\infty} \int_{(a-y)}^{(a-y+da)} f_x(x) f_y(y) dx dy \\
&= \int_{-\infty}^{\infty} f_x(a-y) f_y(y) da dy \\
\Rightarrow g(a) &= \int_{-\infty}^{\infty} f_x(a-y) f_y(y) dy
\end{aligned}
$$

(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

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

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

Erwartungswerte

Betrachte kontinuierliche Zufallsvariable *x* with WDF *f* (*x*).

Definiere den Erwartungswert (arithmetischen.Mittelwert) als

$$E[x] = \int x\, f(x)\, dx$$

Notation (meist): $E[x] = \mu$

Achtung:

E[ ] ist keine Zufallsvariable, keine Funktion der Stichprobe

ist ein linearer Operator auf der Funktion z.B. a(x)=x

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

ist eine Charakteristik der Wahrscheinlichkeitsdichtefunktion

ist abhängig von den Parametern der WDF

Erwartungswerte

Für ein Funktion $y(x)$ with WDF $g(y)$, gilt

$$E[y] = \int y\, g(y)\, dy = \int y(x) f(x)\, dx \qquad \text{(äquivalent)}$$

n-tes algebraisches Moment (n=1 ergibt Mittelwert):

$$E[x^n] = \int_{-\infty}^{\infty} x^n f(x) dx = \mu'_n$$

n-tes zentrales Moment (n>1):

$$E\left[(x-\mu)^n\right] = \int_{-\infty}^{\infty} (x-\mu)^n f(x) dx = \mu_n$$

Erwartungswerte (II)

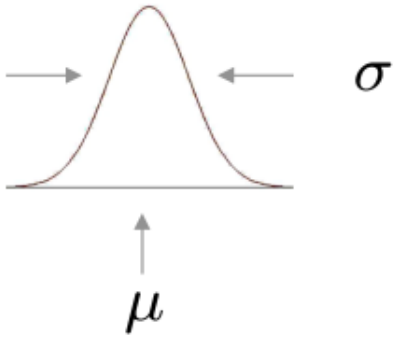
Varianz = 2tes zentrales Moment:

$$V[x] = E[(x-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx \quad V[x] = \sigma^2$$

$$V[x] = E[(x-\mu)^2] = E[x^2] - 2E[x]\mu + \mu^2 = E[x^2] - \mu^2.$$

Standardabweichung: $\sigma = \sqrt{\sigma^2}$

$$\sigma_x = \sqrt{V[x]}$$



Erwartungswerte (III)

Weitere zentrale Momente (selten gebraucht):

Schiefe $\gamma = \mu_3/\sigma^3$

misst Asymmetrie der Verteilung:
positiv (negativ) für Ausläufer nch rechts (links)

Wölbung oder **Kurtosis** $\kappa = \mu_4/\sigma^4 - 3$

misst Anteil im Zentrum zum Anteil in Ausläufern
relativ zur Gauss-/Normalverteilung

0: für Gaussverteilung
positiv: für mehr Ausläufer als Gaussverteilung bei gleicher Varianz
negativ: für weniger Ausläufer als Gaussverteilung bei gleicher Varianz

Erwartungswerte (IV)

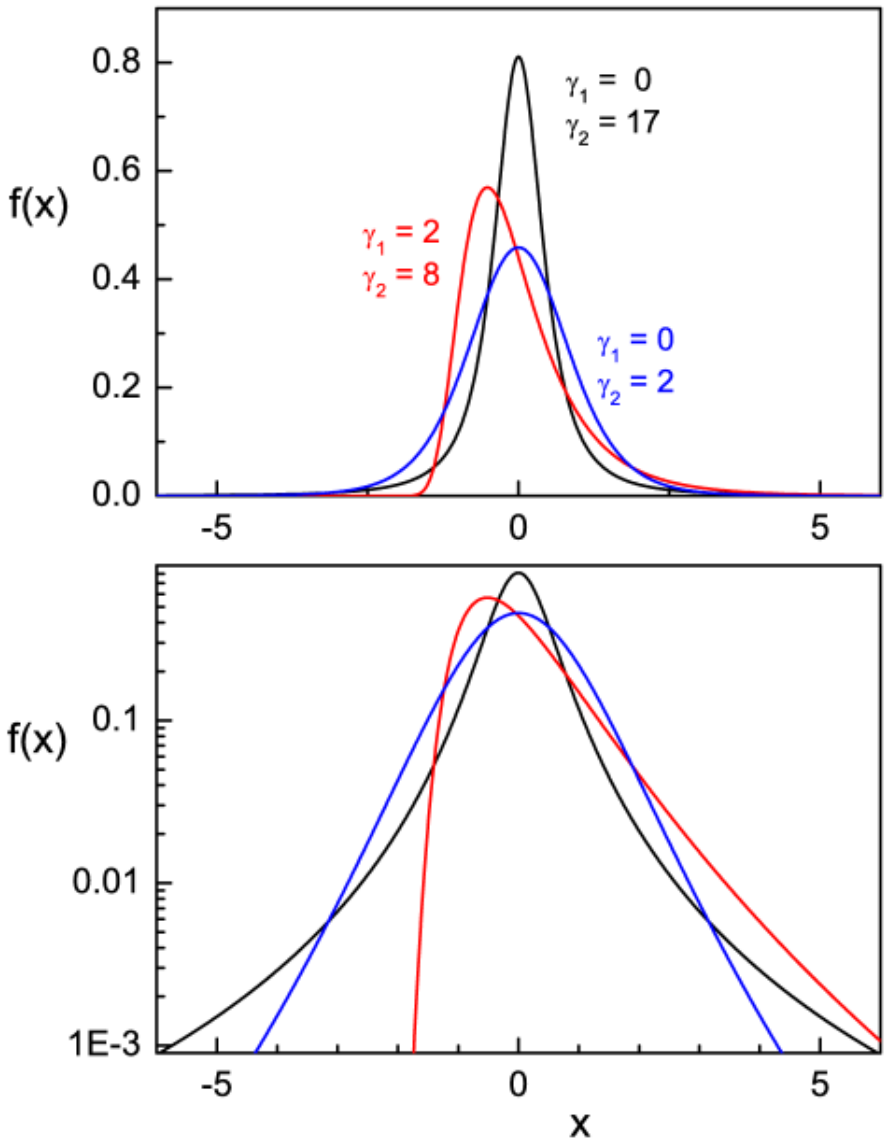
Schiefe $\gamma = \mu_3/\sigma^3$

$$\begin{aligned}\gamma_1 &= E\left[(x-\mu)^3\right]/\sigma^3 \\ &= E\left[x^3 - 3\mu x^2 + 3\mu^2 x - \mu^3\right]/\sigma^3 \\ &= \left\{E(x^3) - 3\mu\left[E(x^2) - \mu E(x)\right] - \mu^3\right\}/\sigma^3 \\ &= \frac{E(x^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3}\,.\end{aligned}$$

Wölbung oder **Kurtosis** $\kappa =$

$$\mu_4/\sigma^4 - 3$$

Alle Verteilungen haben
Mittelwert = 0 und Varianz =1



Kovarianz und Korrelation

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

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

Korrelationskoeffizient (dimensionlos) definiert als

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

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

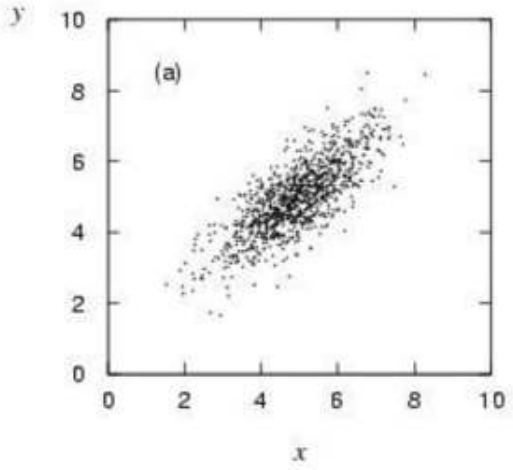
$$E[xy] = \int\int xy\, f(x,y)\, dxdy = \mu_x \mu_y$$

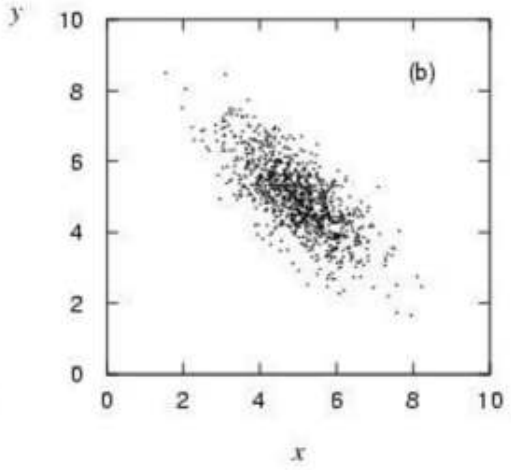
→ $\mathrm{cov}[x,y] = 0$ *x* und *y* sind 'unkorreliert'

Bemerkung: Umkehrschluss i. a. nicht wahr.

Korrelation (Fortsetzung)

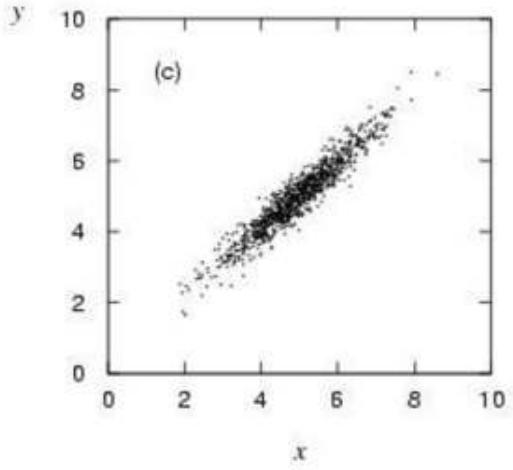
$\rho = 0.75$

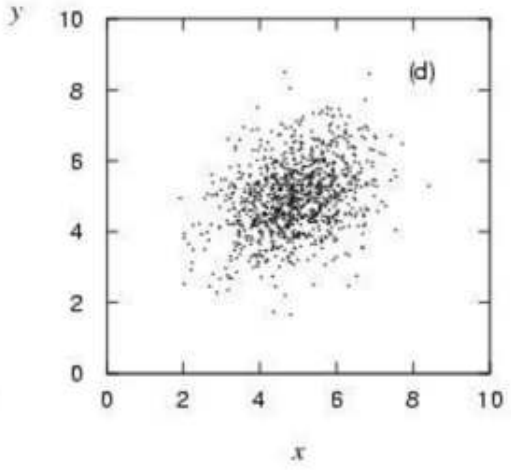




$\rho = -0.75$

$\rho = 0.95$





$\rho = 0.25$

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 $y(\vec{x})$

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

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

Fehlerfortpflanzung (IV)

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

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

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

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

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 Sticprobe der ZV x_i vom Umfang n.
Die unkorellierten Messungen haben Messfehler: $\sigma^2 = \sigma_i^2$

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

Was ist der Fehler auf den Stichprobenmittelwert?

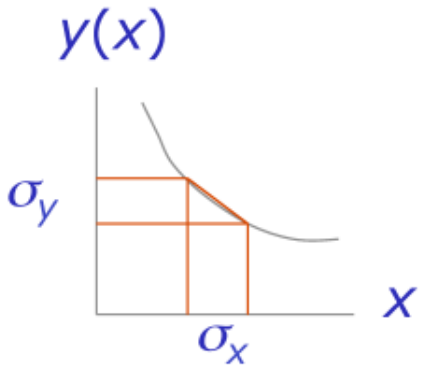
Partielle Ableitungen: $\frac{\partial y}{\partial x_i} = \frac{1}{n}$

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

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

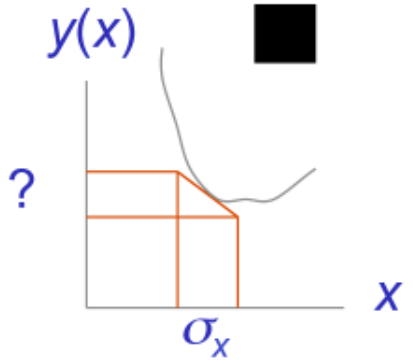
Fehlerfortpflanzung (VI)

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 Annhme über die WDF von x_i, i.e., sie muss nicht die Gaußverteilung sein

Fehlerfortpflanzung (VII) – Speziallfall

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

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

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 unkorelliert und nes gelte $\sigma_r = 0, \sigma_\phi, \sigma_z$

Die Kovarianzmatrix lautet dann:

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

Wie sehen die Fehler in karthesischen Koordinaten aus?

$$x = r \cos \phi$$

$$y = r \sin \phi$$

$$z = z$$

Fehlerfortpflanzung (X): Beispiel 2

Die Jakobi-Matrix der Ableitungen lautet:

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

Die Kovarianzmatrix in x,y, z lautet:

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

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

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

Charakteristische Funktion

Nützlich für Berechungen wie z.B. Momente, Faltungsintegrale, Beweise etc.

Definiert als: $\phi_x(k) = E[\exp(ikx)]$

d.h.: charakteristische Funktion ist Fouriertransformierte der WDF:

$$\phi_x(k) = \int_{-\infty}^{\infty} \exp(ikx) f(x) dx$$

und Rücktransformation: $f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-ikx) \phi_x(k) dk.$

Wegen Eindeutigkeit der Fouriertrafo gilt:
wenn charakteristische Funktion gleich, dann auch die WDF

Für diskrete ZV: $\phi_x(k) = \sum_i \exp(ikx_i) P(x_i)$

$$F(b) - F(a) = \frac{i}{2\pi} \int_{-\infty}^{\infty} \frac{\exp(ikb) - \exp(ika)}{k} \phi_x(k) dk$$

Charakteristische Funktion (II)

Seien $x_1, ..., x_n$ unabhängige zufallsvariablen mit WDFs $f_1(x_1), ..., f_n(x_n)$

Betrachte ZV: $z = \sum_{i=1}^{n} x_i$

D.h.: charakteristische Funktionen zu den f_i seinen: $\phi_1(k_1), ..., \phi_n(k_n)$

Dann ist die charakterischte Funktion $\phi_z(k)$ zu z gegeben durch:

$$\begin{aligned}
\phi_z(k) &= \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \exp\left(ik \sum_{i=1}^{n} x_i \right) f_1(x_1)...f_n(x_n)dx_1...dx_n \\
&= \int_{-\infty}^{\infty} \exp(ikx_1) f_1(x_1)dx_1 ... \int_{-\infty}^{\infty} \exp(ikx_n) f_n(x_n)dx_n \\
&= \phi_1(k)...\phi_n(k).
\end{aligned}$$

i.e. durch das Produkt der charakteristischen Funktionen

Charakteristische Funktion (III)

Das *m*-te algebraische Moment der WDF ist berechenbar über

$$\begin{aligned}
\frac{d^m}{dk^m}\phi_x(k)\,|_{k=0} &= \frac{d^m}{dk^m}\int_{-\infty}^{\infty}\exp(ikx)f(x)dx\Big|_{k=0} \\
&= i^m\int_{-\infty}^{\infty}x^m f(x)dx \\
&= i^m\mu'_m.
\end{aligned}$$

d.h. als *m*-te Ableitung der char. Funktion bei Null.

Kapitel 2

Katalog von ausgewählten Wahrscheinlichkeitsdichtefunktionen

Binomial-Verteilung

Betrachte *N* unabhängige Messungen/Experimente (Bernoulli-Versuche):

Ausgang jedes Experimentes ist entweder ”Erfolg” oder “Misserfolg”
Die Wahrscheinlichkeit für Erfolg (Misserfolg) sind p (1-p)

Definiere diskrete ZV als Anzahl der Erfolge n ($0 \leq n \leq N$).

Wkt. für eine spezielle Reihenfolge von n Erfolgen und (N-n) Misserfolgen ist:

$$pp(1-p)p(1-p) = p^n(1-p)^{N-n}$$

Aber Anordnung ist unwichtg →Kombinatorik es gibt $\frac{N!}{n!(N-n)!}$

Möglichkeiten (Permutationen) *n* Erfolge in *N* Versuchen einzuordnen.
Geamtwkt. für *n* Erfolge ist die Summe der Wkt. für jede Permutation.

Binomial-Verteilung (2)

Die Binomialverteilung lautet:

$$f(n; N, p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

Zufalls-variable

Parameter

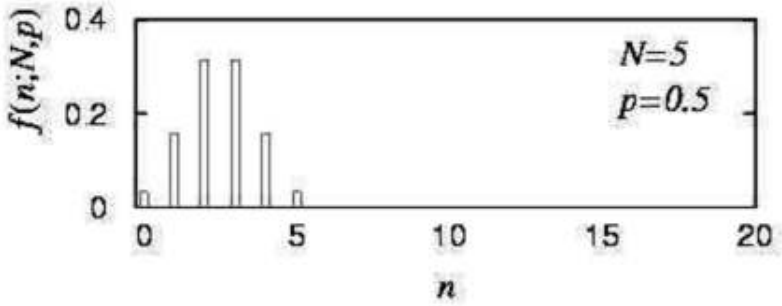
Für den Erwartungswert und die Varianz findet man:

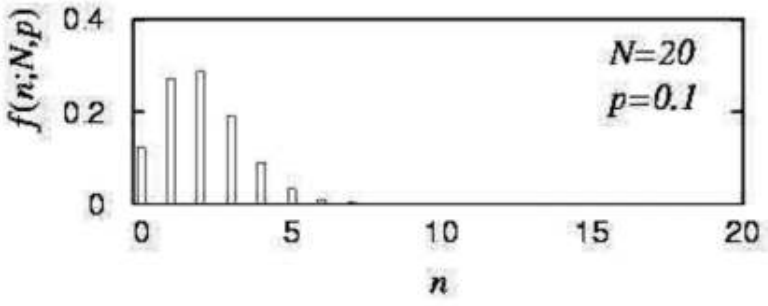
$$E[n] = \sum_{n=0}^{N} n f(n; N, p) = Np \qquad V[n] = E[n^2] - (E[n])^2 = Np(1-p)$$

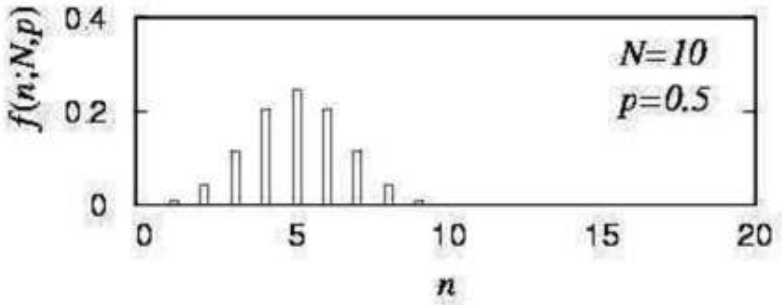
Die charakteristische Funktion lautet: $\phi_n(k; p, N) = (p(\exp(ik) - 1) + 1)^N$

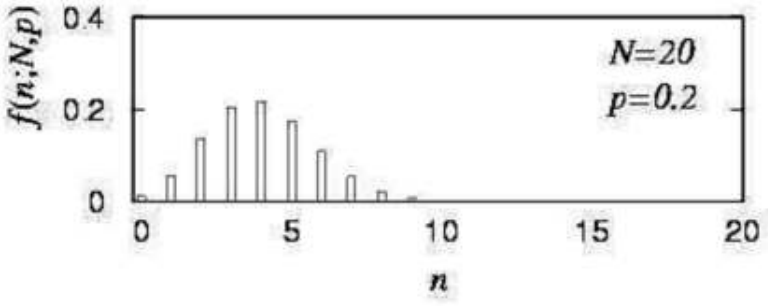
Binomial-Verteilung (3)

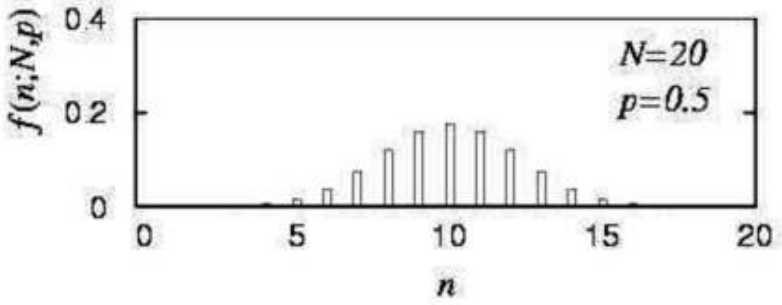
Verteilungen für verschiedene Parameter p und Anzahl der Versuche N:

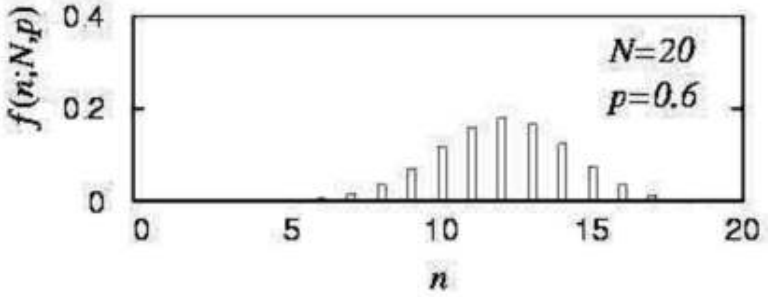












Binomial-Verteilung: Beispiele

Nachweis eines Objektes mit Detektor
Ansprechwahrscheinlichkeit eines Detektors p=0.95

Bedingung: mindestes Treffer in drei Detektorlagen für Rekonstruktion der Spur
Frage: wieviel Kammern werden benötigt um >99% effizient zu sein?

3 Kammern: $P(k=3;0.95,3) = 0.95^3 = 0.857$
4 Kammern: $P(k>=3;0.95,4)=0.171+0.815=0.986$
5 Kammern: $P(k>=3;0.95,5)=0.021+0.204+0.774=0.999$

Wahrscheinlichkeit für Erfolg p
Wiviele Versuche damit mindestens 1 Erfolg mit Wkt >= α ?
$P(k>=1;p,n) >= \alpha$ oder $1-P(K=0;p,n)>=\alpha$
$P(k=0)<=1-\alpha$
$(1-p)^N<=1-\alpha$
$N>=\ln(1-\alpha)/\ln(1-p)$

Binomial-Verteilung: Beweise

Normierung:

$$\begin{aligned}\sum_{n=0}^{N} f(n;N,p) &= \sum_{n=0}^{N} \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \\ &= (p+(1-p))^N \\ &= 1.\end{aligned}$$

Erwartungswert: $E[n]$

$$\begin{aligned}E[n] &= \sum_{n=0}^{N} n \, \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \\ &= \sum_{n=1}^{N} n \, \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \\ &= Np \sum_{n=1}^{N} \frac{(N-1)!}{(n-1)!((N-1)-(n-1))!} p^{(n-1)} (1-p)^{(N-1)-(n-1)} \\ &= Np \sum_{n=0}^{N-1} \frac{(N-1)!}{(n)!((N-1)-(n))!} p^{n} (1-p)^{(N-1)-n} \\ &= Np\end{aligned}$$

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 Bustaben des Alphabets.
(nA, …,nZ) folgt Multinomialverteilung.

Poissonverteilung (i)

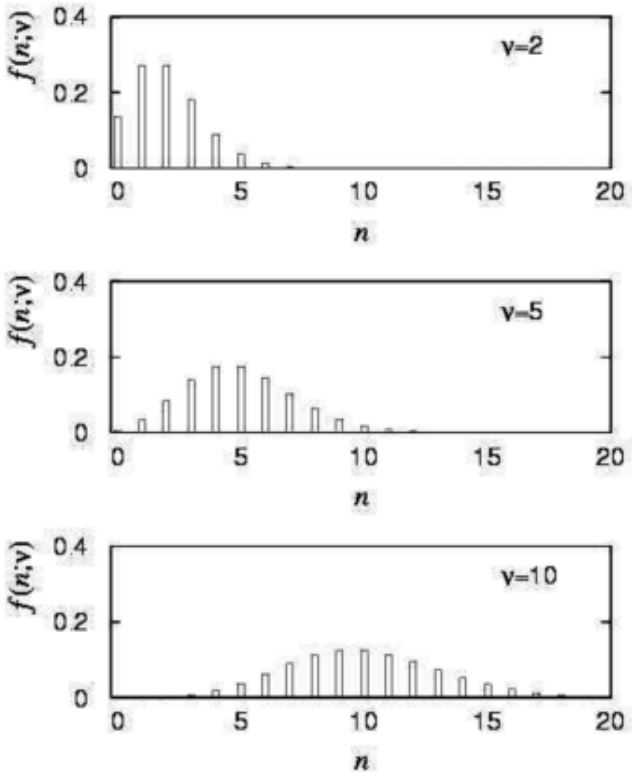
Betrachte Grenzfall der Binomialverteilung mit

$$N \to \infty, \qquad p \to 0, \qquad E[n] = Np \to \nu \; . \quad \text{konstant}$$

→ *n* folgt der Poissonverteilung:

$$f(n;\nu) = \frac{\nu^n}{n!} e^{-\nu} \quad (n \geq 0)$$

$$E[n] = \nu \, , \quad V[n] = \nu \; .$$



Example: number of scattering events *n* with cross section σ found for a fixed integrated luminosity, with

$$\nu = \sigma \int L \, dt \; .$$

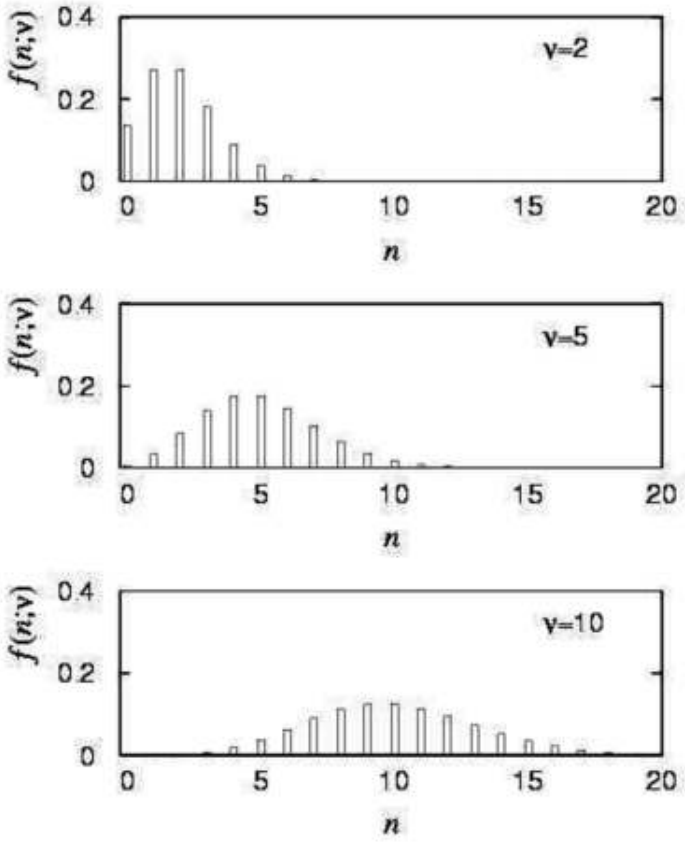
Poissonverteilung: Herleitung 1

Betrachte Grenzfall der Binomialverteilung mit

$$N \to \infty, \qquad p \to 0, \qquad E[n] = Np \to \nu \text{ . konstant}$$

→ *n* folgt der Poissonverteilung:

$$f(n;\nu) = \frac{\nu^n}{n!} e^{-\nu} \quad (n \geq 0)$$

$$E[n] = \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} \exp(-\lambda) = \lambda,$$

$$V[n] = \sum_{n=0}^{\infty} (n-\lambda)^2 \frac{\lambda^n}{n!} \exp(-\lambda) = \lambda.$$