technische universität dortmund

Experimentelle Physik Vb
Astroteilchenphysik

Vorlesung
Statistische Methoden der Datenanalyse
Prof. Dr. Dr. Wolfgang Rhode

Wahrscheinlichkeitsverteilungen

Überblick

Definitionen von Wahrscheinlichkeiten

Kombination von Wahrscheinlichkeiten

Charakterisierung von Verteilungen

- 1-dim. Verteilungsfunktion und Wahrscheinlichkeitsdichte
- 1-dim. Momente
- 1-dim. Regeln über Mittelwerte und Varianzen
- mehr-dim. Verteilungsfunktion und Wahrscheinlichkeitsdichte
- mehr-dim. Bedingte Wahrscheinlichkeit und Randverteilungen
- mehr-dim. Erwartungswert, Varianz, Kovarianz
- Unabhängigkeit und Korrelation

Bestimmte ein- und mehrdimensionale Verteilungen

Tschebyscheff-Ungleichung, Gesetz der großen Zahl, Zentraler Grenzwertsatz

Definition (mathematisch)

- Wahrscheinlichkeit kann abhängig davon, ob a priori Wissen über den betrachteten Vorgang zur Definition benutzt werden kann, auf zwei Weisen eingeführt werden:

  1. Falls ein Ereignis auf *n* verschiedene und gleich wahrscheinliche Arten eintreten kann und *k* davon die Eigenschaft *A* haben (dies sei a priori bekannt), so ist die Wahrscheinlichkeit für das Auftreten von *A*

$$P(A) = \frac{k}{n} = \frac{\text{günstige}}{\text{mögliche}}\text{Fälle}$$

Definition (frequentistisch)

- Wahrscheinlichkeit kann abhängig davon, ob a priori Wissen über den betrachteten Vorgang zur Definition benutzt werden kann, auf zwei Weisen eingeführt werden:

  2. Ohne *a priori* Wissen: Die Eigenschaften A und nicht-A eines Experimentes werden n-fach unabhängig beobachtet. Dabei trete *k* mal die Eigenschaft *A* auf. Dann ist die Wahrscheinlichkeit *P(A)* gegeben durch

$$P(A) = \lim_{n \to \infty} \frac{k}{n}$$

- Problem: Identische Wiederholungen von Experimenten sind schwer zu gewährleisten

Rechenregeln für bedingte Wahrscheinlichkeiten

- Summenregel: $$p(A|B) + p(\bar{A}|B) = 1$$
  - $\bar{A}$ ist das Komplement von A
- Produktregel: $$\begin{aligned} p(A, B|C) &= p(A|C)p(B|A, C) \\ &= p(B|C)p(A|B, C) \end{aligned}$$
- A, B beschreibt dabei die Annahme, A **und** B seien wahr.

Kombination von Wahrscheinlichkeiten

- Gegeben seien die Ereignistypen *A* und *B* mit den Wahrscheinlichkeiten *P(A)* und *P(B)*, dann ist die Wahrscheinlichkeit für *A* **oder** *B*

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

- Wenn sich A und B ausschließen:

$$P(A \wedge B) = 0 \quad \text{und} \quad P(A \vee B) = P(A) + P(B)$$

- Als Spezialfall sei: B = Ā (nicht A), dann ist:

$$P(A \vee \overline{A}) = P(A) + P(\overline{A}) = 1$$

Kombination von Wahrscheinlichkeiten

- Gegeben seien die Ereignistypen *A* und *B* mit den Wahrscheinlichkeiten *P(A)* und *P(B)*, dann ist die Wahrscheinlichkeit für *A* <u>und</u> *B*

$$P(A \wedge B) = P(A) \cdot P(B|A)$$

- Sind *A* und *B* unabhängig

$$P(B|A) = P(B)$$

folgt

$$P(A \wedge B) = P(A) \cdot P(B)$$

Bayesischer Zugang

- Die Wahrscheinlichkeit *p(A|B)* ist ein quantitatives Maß der Plausibilität der Annahme *A* unter der Bedingung der bekannten Information gegeben durch die Annahme *B*

- *A* kann dabei eine beliebige logische Annahme sein

- *p(A|B)* lässt sich mit dem Satz von Bayes berechnen:

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

- Problem: Der vage Begriff „Plausibilität“ muss genau definiert werden
- Nutzung auch zur Parameterschätzung → Kapitel Schätzen

Verteilungsfunktion und Wahrscheinlichkeitsdichte

- Ziel ist die Klassifizierung von möglichen Endzuständen eines statistischen Vorganges
  - Beispiel zur Klassifizierung: Bei einem Münzwurf werden z.B. die Zuordnungen Kopf → 0 und Zahl → 1 vorgenommen.

- Allgemein:
  - Wird dem Ereignis A_i die ganze Zahl i zugewiesen, liegt eine *diskrete* Zufallsvariable vor.
  - *Kontinuierliche* Zufallsvariablen werden genutzt, wenn es nicht möglich ist, die Ereignisse ganzen Zahlen zuzuordnen.

Verteilungsfunktion und Wahrscheinlichkeitsdichte

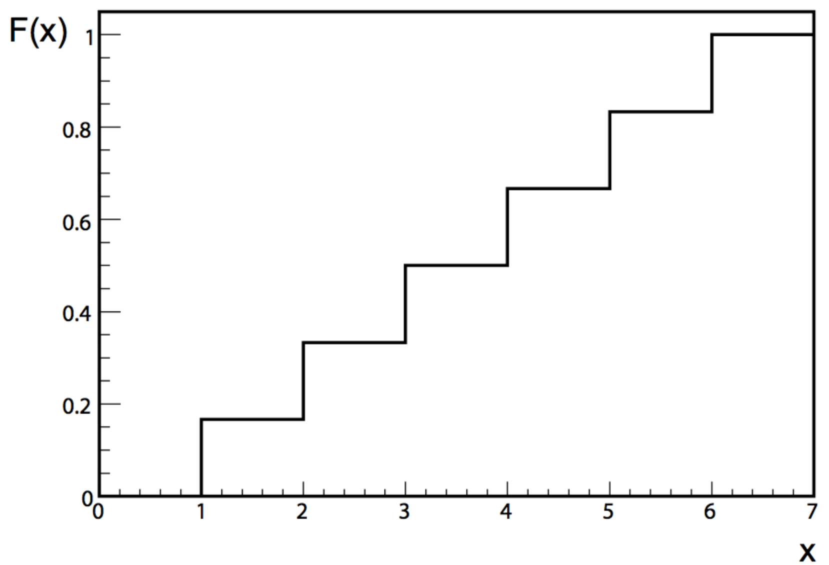
- Die Zufallsvariable *r* möge den möglichen Ausgang des Experimentes angeben. Sie wird mit der reellen Zahl *x* verglichen. Gesucht ist die Wahrscheinlichkeit dafür, dass ein Ereignis eintritt, bei dem die Zufallsvariable *r* kleiner ist als ein vorher gewähltes *x* ($r < x$). Dazu wird die Verteilungsfunktion gebildet:

$$F(x) = P(r \leq x)$$

- Die Verteilungsfunktion gibt die Summe aller Ereignisse unterhalb von *x* normiert auf die Gesamtzahl der Versuche an.

Verteilungsfunktion und Wahrscheinlichkeitsdichte

- Für einen Würfel mit sechs Seiten, ergibt sich für die Verteilungsfunktion F(x) eine sechsstufige Treppenfunktion, die monoton von 0 auf 1 ansteigt.



Verteilungsfunktion und Wahrscheinlichkeitsdichte

- Im Grenzfall einer kontinuierlichen Verteilung ist:

$$\lim_{x\to\infty} F(x) = \lim_{x\to\infty} P(r \leq x) = 1$$

- Da die Summe aus P(A) + P(Ā) = 1 ist, gilt:

$$P(r > x) = 1 - F(x) = 1 - P(r \leq x)$$

- Somit ist

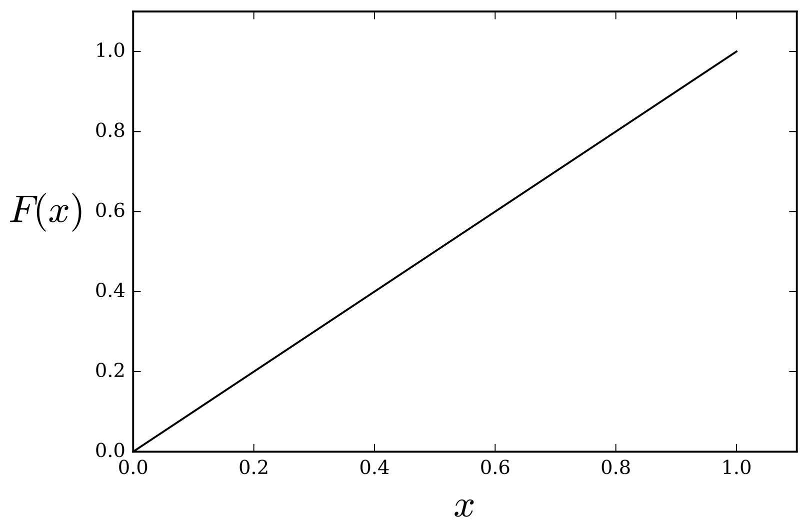
$$\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} P(r \leq x) = 1 - \lim_{x\to-\infty} P(r > x) = 0$$

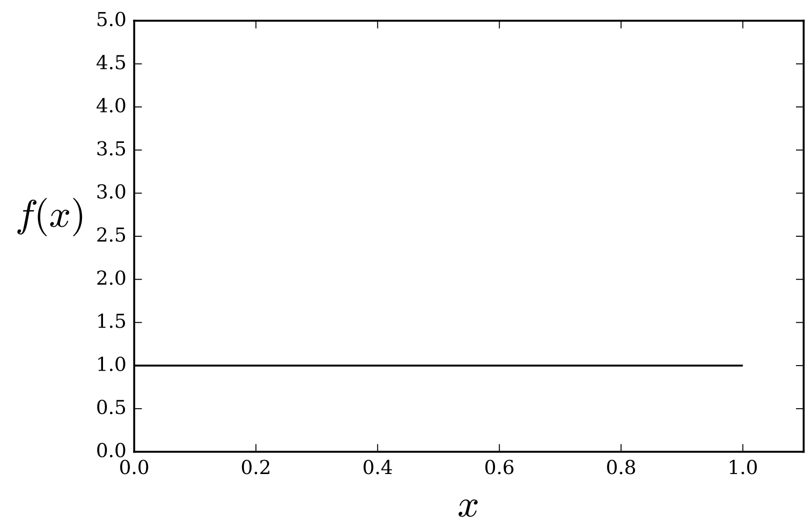
Verteilungsfunktion und Wahrscheinlichkeitsdichte

- Wenn die Verteilungsfunktion stetig differenzierbar ist, gilt:

$$\frac{dF(x)}{dx} = F'(x) = f(x)$$

- *f(x)* heißt dann Wahrscheinlichkeitsdichte von *r* und gibt ein Maß für die Wahrscheinlichkeit in dem Intervall *x* ≤ *r* ≤ *x* + d*x* an





Verteilungsfunktion und Wahrscheinlichkeitsdichte

- Die Wahrscheinlichkeit, dass *r* kleiner ist als ein vorgewählter Wert *a*, ist gegeben durch:

$$P(r < a) = \int_{-\infty}^{a} f(x)\,\mathrm{d}x = F(a),$$

- die Wahrscheinlichkeit, dass *r* in einem Intervall zwischen *a* und *b* liegt, ist:

$$P(a \leq r \leq b) = \int_{a}^{b} f(x)\,\mathrm{d}x = F(b) - F(a).$$

- Insbesondere gilt bei Integration über den gesamten Bereich in *x*:

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1.$$

Allgemeine Eigenschaften einer Zufallsverteilung: Momente

- Der Mittelwert oder Erwartungswert *E(x)* bei einer diskreten Verteilung:

$$\bar{x} = E(x) = \sum_{i=1}^{n} (x_i \cdot P(x = x_i))$$

- Der Erwartungswert einer Funktion von diskreten *r* ist:

$$E[H(x)] = \sum_{i=1}^{n} (H(x_i) \cdot P(x = x_i))$$

Momente

- Erwartungswert für kontinuierlich verteilte x

$$E(x) = \bar{x} = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x$$

- und für eine Funktion davon

$$E[H(x)] = \int_{-\infty}^{\infty} H(x) \cdot f(x)\,\mathrm{d}x$$

Momente

- Wichtige Charakteristika einer Verteilung sind ihre Breite und Symmetrie. Dazu betrachten wir als Spezialfall die Funktion:

$$H(x) = (x-c)^l$$

- Der Erwartungswert ergibt sich zu:

$$a_l = E\left[(x-c)^l\right]$$

Momente

- Berechnung von Momenten μ_l um den Mittelwert, heißen zentrale Momente:

$$\mu_l = E[(x - \bar{x})^l]$$

- Die Momente $\mu_0 = 1$ und $\mu_1 = 0$ sind trivial zu bestimmen.

Momente

- Für das zweite Moment gilt:

$$E[(r-\bar{x})^2] = \hat{\sigma}^2(x) = \mathrm{Var}(x) = \int_{-\infty}^{\infty} (x-\bar{x})^2 f(x)\,\mathrm{d}x$$

- Die so definierte (empirische) Varianz ist ein Maß für die Breite der Verteilung.

- Die Wurzel aus der Varianz heißt Streuung oder Standardabweichung

$$\sigma = \sqrt{\sigma^2(x)}$$

Momente

- Das dritte Momente um den Mittelwert normiert auf die Standardabweichung heißt Schiefe (*skewness*). Es beschreibt die Symmetrie der Verteilung.

$$\gamma = \frac{\mu_3}{\sigma^3}$$

- $\gamma < 0$ linksschief
- $\gamma > 0$ rechtsschief

Momente

- Der Quotient des vierten zentralen Moments und dem Quadrat der Varianz wird als Wölbung (*kurtosis*) bezeichnet:

$$C = \mu_4/\sigma^4$$

- *C* ist groß, wenn die Verteilung über größere Ausläufer verfügt als die Gauß-Verteilung.
- Die Gauß-Verteilung selbst liefert *C=3*

Regeln über Mittelwerte und Varianzen

- Multiplikation jeder Zahl einer Verteilung mit (derselben) Konstanten:

$$H(x) = cx,\ c = \text{const}$$

- Es folgt, dass

$$E(c \cdot r) = c \cdot E(r),\ \text{und}\ \sigma^2(c \cdot r) = c^2 \cdot \sigma^2(r)$$

- Daher ist

$$\sigma^2(r) = E[(r - \bar{x})^2] = E[r^2 - 2r\bar{x} + \bar{x}^2] = E(r^2) - \bar{x}^2$$

Regeln über Mittelwerte und Varianzen

- Die standardisierte Variable

$$u = \frac{r - \bar{x}}{\sigma(r)}$$

- hat den Erwartungswert

$$E(u) = \frac{1}{\sigma(r)} E(r - \bar{x}) = \frac{1}{\sigma(x)} (\bar{x} - \bar{x}) = 0$$

- und die Varianz

$$\sigma^2(u) = \frac{1}{\sigma^2(x)} E[(r - \bar{x})^2] = \frac{\sigma^2(x)}{\sigma^2(x)} = 1$$

Regeln über Mittelwerte und Varianzen

- Der <u>wahrscheinlichste Wert</u>

$$P(x = x_m) = \mathrm{maximal}$$

- Besitzt die Verteilung ein Maximum, heißt sie unimodal, sonst heißt sie multimodal

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) = 0 \qquad \frac{\mathrm{d}^2}{\mathrm{d}x^2} f(x) < 0$$

Regeln über Mittelwerte und Varianzen

- Der Median ist derjenige Wert einer Verteilung, für den die Verteilungsfunktion *F* = 0.5 ist

$$F(x_{0.5}) = P(r \leq x_{0.5}) = 0.5$$

- Ist *f(x)* stetig, gilt

$$\int_{-\infty}^{x_{0.5}} f(x)\,\mathrm{d}x = 0.5$$

- Ist die Verteilung unimodal, stetig und symmetrisch, dann ist Erwartungswert gleich dem wahrscheinlichsten Wert oder Median

Regeln über Mittelwerte und Varianzen

- Das Quartil einer Verteilung ist analog zu $x_{0.5}$ definiert als:

$$F(x_{1/4}) = 0.25, \quad F(x_{3/4}) = 0.75$$

unteres Quartil　　oberes Quartil

- Entsprechend sind Dezile (q=10%) und Quantile (q=beliebige Prozentsatze) definiert als

$$F(X_q) = \int_{-\infty}^{X_q} f(x)\,\mathrm{d}x = q$$

Regeln über Mittelwerte und Varianzen

- Der quadratische Mittelwert (*root mean square* = RMS) ist definiert als

$$x_{\mathrm{rms}} = \sqrt{E(x^2)} = \sqrt{\sigma^2(x) + \bar{x}^2}$$

- Ist der Erwartungswert gleich null, gilt:

$$x_{\mathrm{rms}} = \sigma(x)$$

Momente vs. Lagemaße & Streuungsmaße

- Gerade wurden Momente betrachtet. Momente beschreiben eine Wahrscheinlichkeitsverteilung, **nicht** die daraus gezogene Stichprobe.
- Lage- & Streuungsmaße werden zur Charakterisierung einer gezogenen Verteilung genutzt.
- Vorsicht: Als Mittelwert wird oft umgangssprachlich sowohl das 1. Moment (Erwartungswert) einer Wahrscheinlichkeitsverteilung, als auch der arithmetische Mittelwert einer Stichprobe bezeichnet. Letzterer ist jedoch ein Lagemaß.

Lagemaße

- Mittelwerte:

$$\underbrace{\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i}_{\text{arithmetisch}} \qquad \underbrace{\bar{x}_G = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}}_{\text{geometrisch}} \qquad \underbrace{\bar{x}_H = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\right)^{-1}}_{\text{harmonisch}}$$

- Median:
  auf geordneten, metrischen Zufallsvariablen

$$\tilde{x}_{0.5} = \begin{cases} x_{\frac{n+1}{2}} & \text{falls } n \text{ ungerade,} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & \text{falls } n \text{ gerade.} \end{cases}$$

Für p-Quantil 0.5 durch p ersetzen

Streuungsmaße

- Empirische Varianz:

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- Empirische Stichprobenvarianz:

$$\hat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- Entropie (mittlere Information in x_i):

$$I = \sum_{i=1}^{n} f(x_i)\ \underbrace{\ln\frac{1}{f(x_i)}}_{\text{Information}}$$

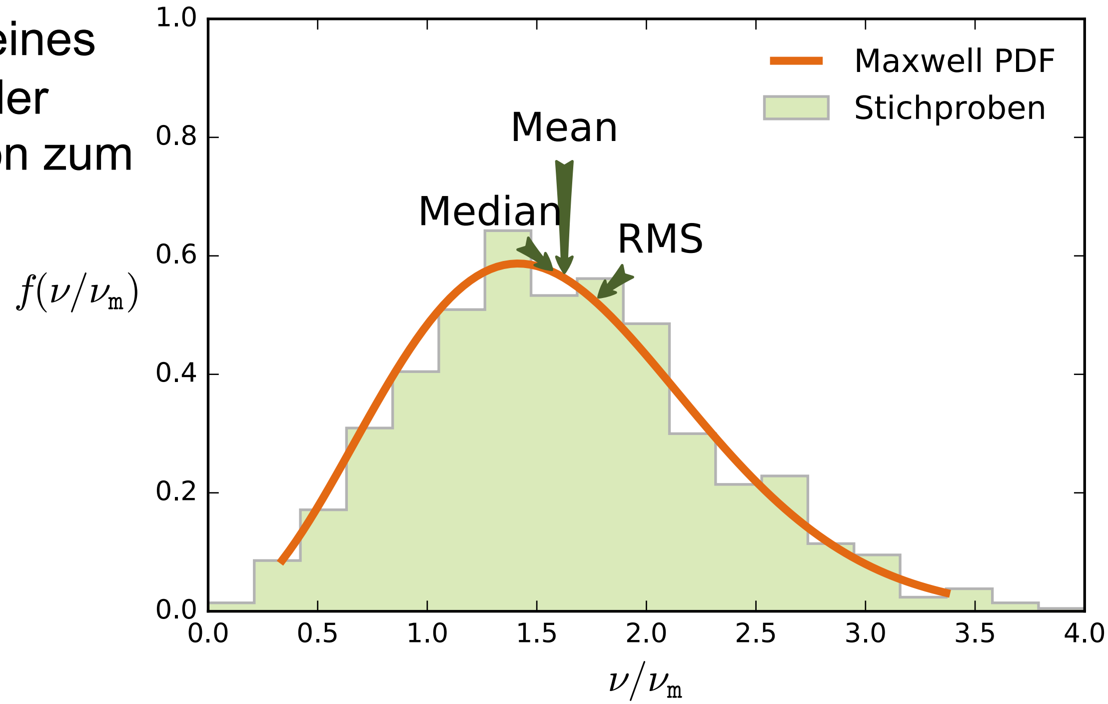
Keine Streuung: $f(x_i) = 1 \rightarrow I = 0$

Maximale Streuung (alle rel. Häufigkeiten sind gleich):

$$f(x_i) = \frac{1}{n} \rightarrow I = \ln n$$

Beispiel aus Blobel - Lohrmann

- Maxwellsche Geschwindigkeitsverteilung eines idealen Gases als Funktion der Geschwindigkeit ν in Relation zum wahrscheinlichsten Wert der Geschwindigkeit ν_{m}



Und wie sieht es mit mehrdimensionalen Verteilungen aus?

Zunächst der **zweidimensionale** Fall:

Gegeben: Zufallsvariablen X und Y

Gesucht: $P((X < x) \wedge (Y < y))$

Analog zum eindimensionalen Fall, ist Verteilungsfunktion definiert durch:

$$F(x, y) = P(X < x, Y < y)$$

Zweidim. Wahrscheinlichkeitsdichte und Verteilungsfunktion

F sei stetig differenzierbar, dann

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

$$P(a \leq x < b, c \leq y < d) = \int_a^b \int_c^d f(x,y)\,dx\,dy$$

Randverteilungen

Wahrscheinlichkeitsdichten der Randverteilungen:

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

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

Daraus folgt beispielsweise:

$$P(a \leq x < b, -\infty \leq y < \infty) = \int_{a}^{b} \int_{-\infty}^{\infty} f(x,y)dydx = \int_{a}^{b} g(x)dx$$

Bedingte Wahrscheinlichkeit

Bedingte Wahrscheinlichkeit für *Y*
= Wahrscheinlichkeit für *Y* bei bekanntem *X*

$$P(y \leq Y \leq y + dy | x \leq X \leq x + dx)$$

Entsprechende Wahrscheinlichkeitsdichte:

$$f(y|x) = \frac{f(x,y)}{g(x)}$$

Daraus folgt für die Randverteilung:

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

Stochastische Unabhängigkeit

Zufallsvariablen *X* und *Y* sind **unabhängig**, wenn gilt:

$$f(x,y) = g(x) \cdot h(y)$$

Bei Unabhängigkeit gilt also:

$$f(y|x) = \frac{f(x,y)}{g(x)} = \frac{g(x) \cdot h(y)}{g(x)} = h(y)$$

Achtung:

$$\text{Unabhängigkeit} \neq \text{Unkorreliertheit}$$
$$\text{Unabhängigkeit} \not\leftarrow \text{Unkorreliertheit}$$
$$\text{Unabhängigkeit} \rightarrow \text{Unkorreliertheit}$$

Erwartungswert

Analog zu 1-dim. Fall:

$$E[H(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(x,y) f(x,y) dx dy$$

Beispiel: *H(x,y)=ux+by*

$$E(ux + by) = uE(x) + bE(y)$$

Varianz

$$\sigma^2[H(x,y)] = E\{[H(x,y) - E\{H(x,y)\}]^2\}$$

Beispiel: *H(x,y)=ax+by*

$$\begin{aligned}
\sigma^2(ax+by) &= E\left[\left((ax+by) - E\left[ax+by\right]\right)^2\right] \\
&= E\left[\left(a(x-\bar{x}) + b(y-\bar{y})\right)^2\right] \\
&= E\left[a^2(x-\bar{x})^2 + b^2(y-\bar{y})^2 + 2ab(x-\bar{x})(y-\bar{y})\right] \\
&= a^2\sigma^2(x) + b^2\sigma^2(y) + 2ab\cdot\mathrm{cov}(x,y)
\end{aligned}$$

Beispiel: *H(x,y)=xy* mit *x,y* unabh.

$$E[xy] = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} xy\, g(x)h(y)\, dxdy = \int\limits_{-\infty}^{\infty} x\, g(x)\, dx \cdot \int\limits_{-\infty}^{\infty} y\, h(y)\, dy = E[x]\cdot E[y]$$

Mehrdimensionale Verteilungen

Noch einen Schritt weiter...

Die *n*-dim Verteilungen!

Verteilungsfunktion:

$$F(x_1, x_2, ..., x_n) = P(X_1 < x_1, X_2 < x_2, ..., X_n < x_n)$$

Wahrscheinlichkeitsdichte:

$$f(x_1, x_2, ..., x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 ... \partial x_n} F(x_1, x_2, ..., x_n)$$

Mehrdimensionale Verteilungen

Randverteilung einer Variablen

$$g(x_r) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_{r-1} dx_{r+1} ... dx_n$$

Mehrdimensionale Verteilungen

Erwartungswert:

$$E[H(x_1,\ldots,x_n)] = \int\limits_{-\infty}^{\infty} \ldots \int\limits_{-\infty}^{\infty} H(x_1,\ldots,x_n) f(x_1,\ldots,x_n) dx_1 \ldots dx_n$$

Mehrdimensionale Verteilungen

Varianz:

$$\sigma^2[H(x_1,\ldots,x_n)] = E\{[H(x_1,\ldots,x_n) - E\{H(x_1,\ldots,x_n)\}]^2\}$$

Kovarianz

Es gilt:

$$cov(X_i, X_j) = E[(X_i - E[X_i]) \cdot (X_j - E[X_j])]$$

Alternativ (über den Verschiebungssatz):

$$cov(X_i, X_j) = E[X_i \cdot X_j] - E[X_i]E[X_j]$$

Kovarianz

Die Kovarianz ist:

- positiv, wenn $X_i > (<)\ E[X_i]$ mit $X_j > (<)\ E[X_j]$;
- negativ, wenn $X_i > (<)\ E[X_i]$ mit $X_j < (>)\ E[y]$
- =0, wenn $X_i,\ X_j$ unabhängig sind.

Korrelationskoeffizient

Grobes Maß für die Abhängigkeit zweier Zufallsvariablen

$$\rho(X_i, X_j) = \frac{\mathrm{cov}(X_i, X_j)}{\sigma(X_i) \cdot \sigma(X_j)}$$

Unkorreliertheit zweier Variablen, wenn der Korrelationskoeffizient Null ist

Kovarianzmatrix

Allgemein:

$$\begin{pmatrix} cov(X_1, X_1) & \dots & cov(X_1, X_n) \\ \vdots & \ddots & \vdots \\ cov(X_n, X_1) & \dots & cov(X_n, X_n) \end{pmatrix}$$

Bei Unkorreliertheit:

$$\begin{pmatrix} \sigma^2(X_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma^2(X_n) \end{pmatrix}$$

Spezielle eindimensionale Verteilungen:

(diskrete) Gleichverteilung

- Einzelwahrscheinlichkeiten

$$P(X = a_m) = \frac{1}{n}$$

- Parameterbereich

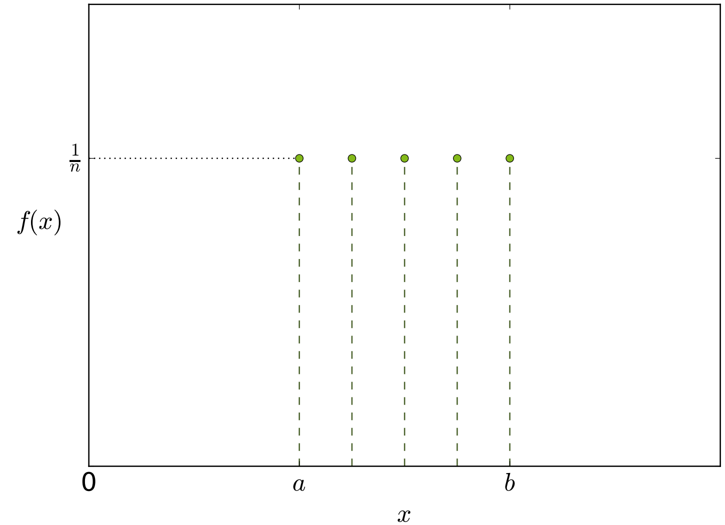
$$a, b \in \mathbb{R}, a < b$$

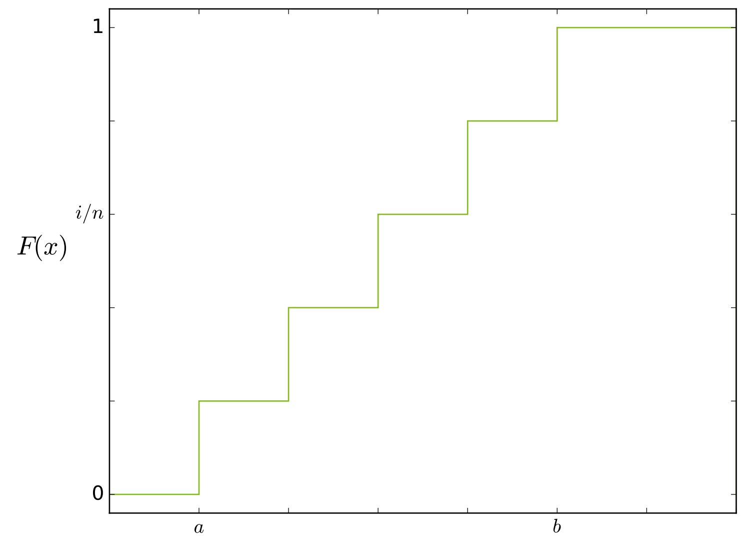
- Momente

$$m_k = \sum_{m=1}^{n} a_m^k \cdot P(X = a_m) = \sum_{m=1}^{n} a_m^k \cdot \frac{1}{n}$$

- Erwartungswert

$$\frac{1}{n} \sum_{m=1}^{n} a_m$$





(diskrete) Gleichverteilung

- Anwedungen
  - Zufällige Versuche mit *n* gleichwahrscheinlichen Ausgängen (Würfeln)

(stetige) Gleichverteilung

- Dichtefunktion

$$\frac{1}{b-a}\,,\; a \leqq x \leqq b$$

- Parameterbereich

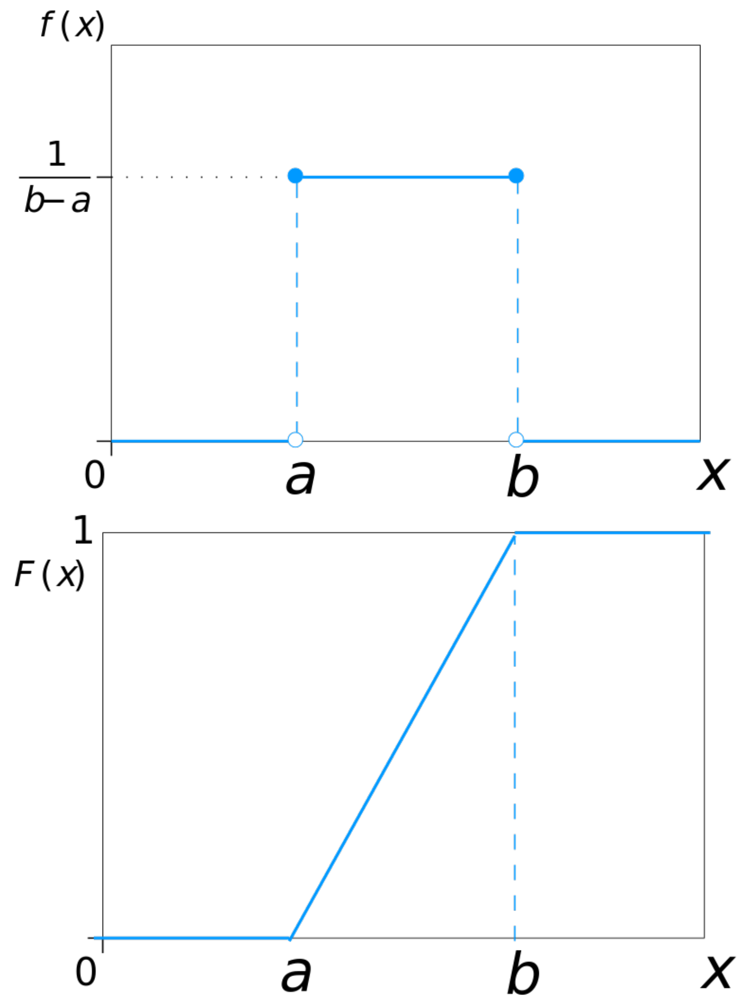
$$a, b \in \mathbb{R}, a < b$$

- Momente

$$\mu_{2k} = \frac{1}{2k+1}\left(\frac{b-a}{2}\right)^{2k}$$
$$\mu_{2k-1} = 0$$

- Erwartungswert

$$\frac{a+b}{2}$$



(stetige) Gleichverteilung

- Anwendungen
  - Physikalische Prozesse ohne Vorzugseigenschaft:
    - Richtung im Zweikörperzerfall
  - Geometrische Wahrscheinlichkeit
  - Erzeugung von Zufallszahlen

Dreiecksverteilung

- Dichtefunktion

$$\frac{2}{b-a}\left(1-\frac{2}{b-a}\left|x-\frac{a+b}{2}\right|\right)$$

- Parameterbereich

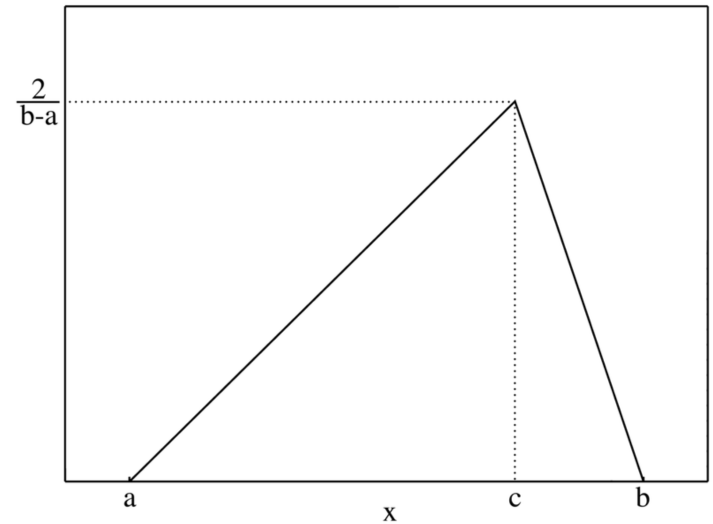
$$a, b \in \mathbb{R}, a < b$$

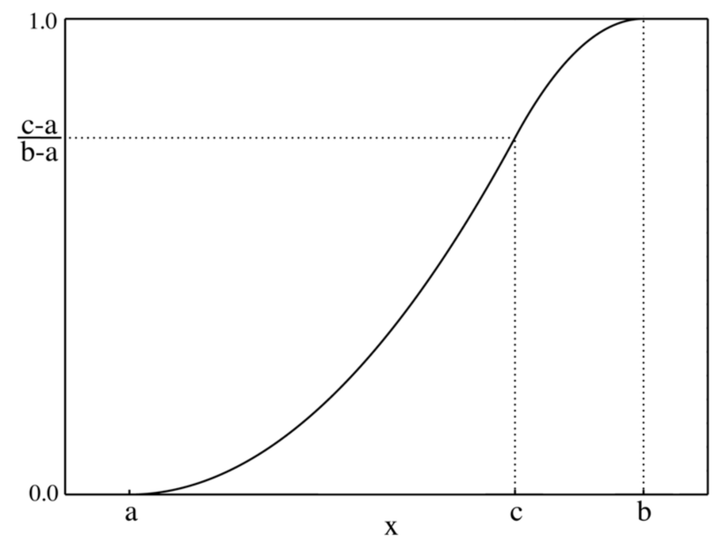
- Momente

$$\mu_{2k} = \frac{(a-b)^{(2k)}}{2^{(2k-1)}(2k+1)(2k+2)}\ ,\ \mu_{2k-1} = 0$$

- Erwartungswert

$$\frac{a+b}{2}$$

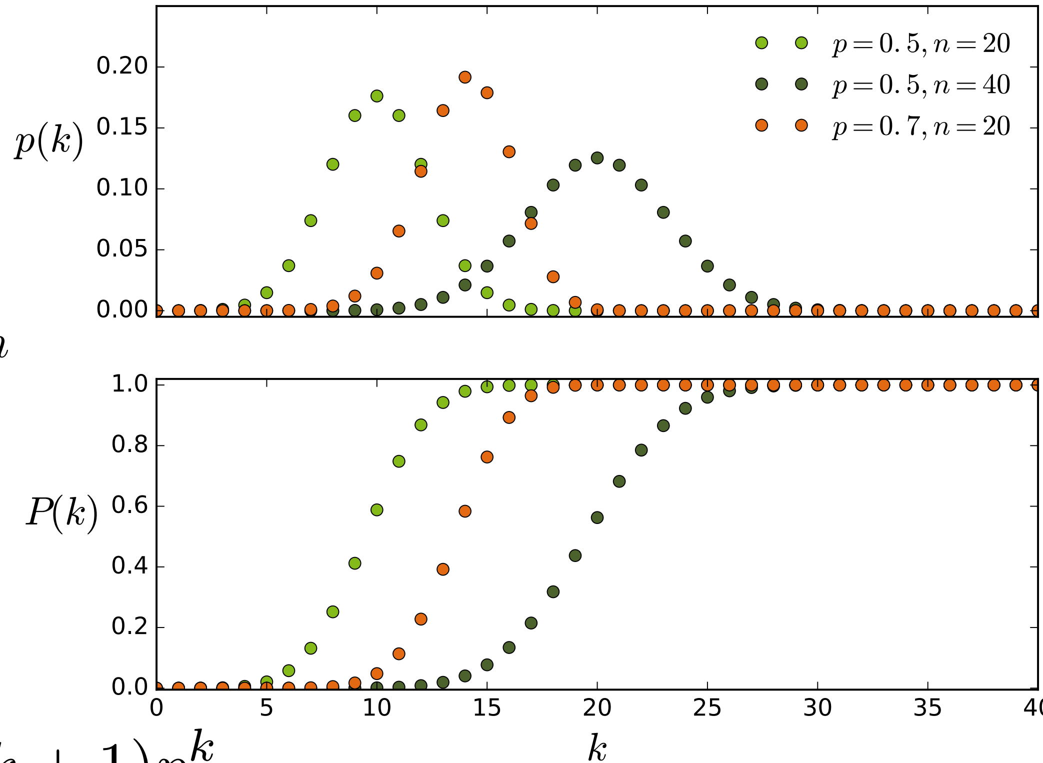




Dreiecksverteilung

- Anwendungen
  - Verteilung der Summe zweier unabh. identisch (stetig) gleichmäßig verteilter Zufallszahlen

Binominalverteilung



- Dichtefunktion

$$\binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, \dots, n$$

- Parameterbereich

$$0 < p < 1, n \in \mathbb{N}$$

- Momente

$$m_{(k)} = n(n-1)...(n-k+1)p^k$$

- Erwartungswert

$$np$$

Binominalverteilung

- Anwendungen
  - Bernoulli-Schema
  - Statistische Qualitätskontrolle
  - Fehlerrechnung
  - “Signalauslösung durch ein Myon in hintereinander liegenden Kammern”
  - Moderate Wahrscheinlichkeit bei einer kleinen Zahl von Versuchen

Poissonverteilung

- Dichtefunktion

$$e^{-\mu}\frac{\mu^k}{k!}, k = 0, 1, \ldots$$

- Parameterbereich

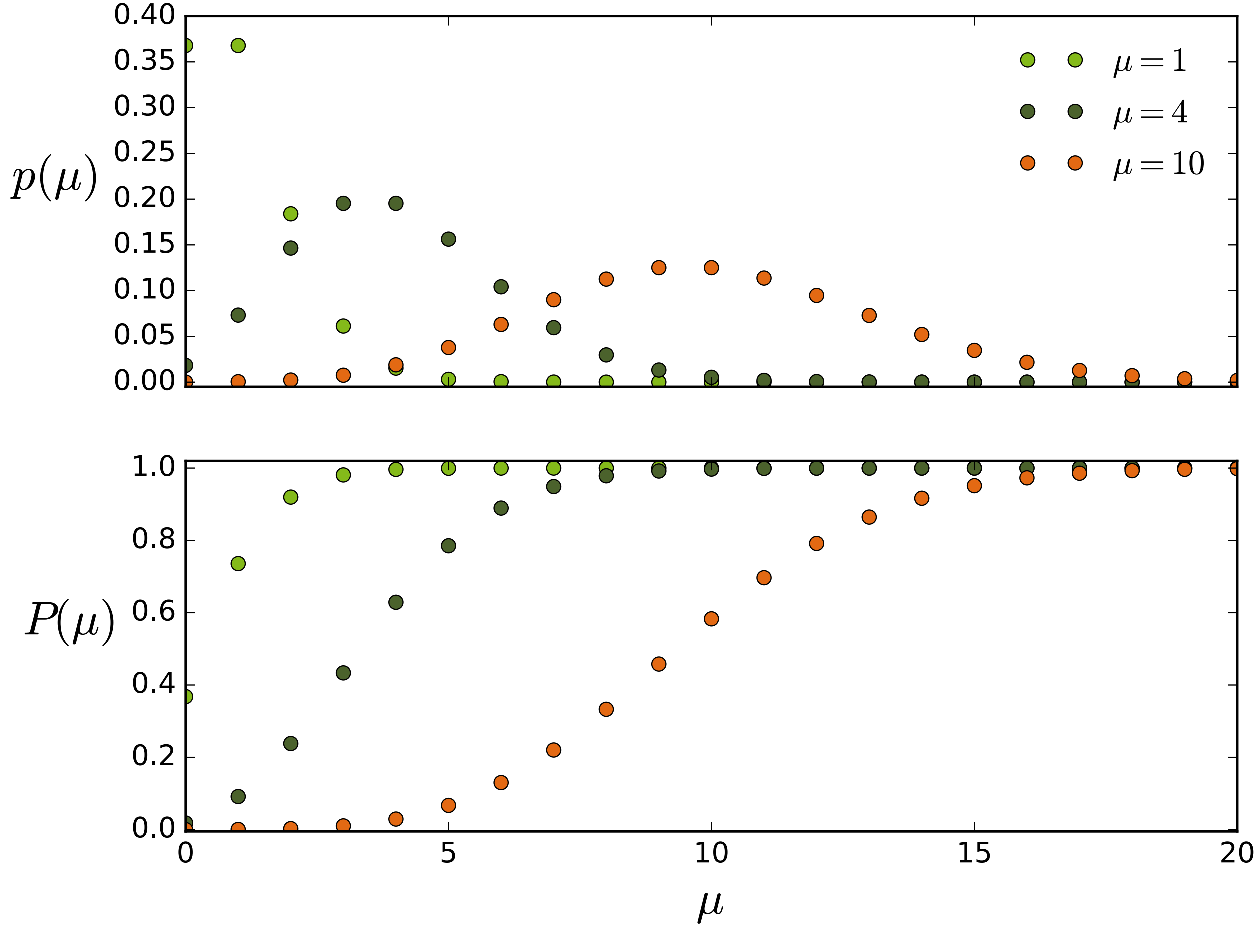
$$\mu > 0$$

- Momente

$$m_{(k)} = \mu^k$$

- Erwartungswert

$$\mu$$



Poissonverteilung

- Anwendungen
  - Reifenpannen, Zerfälle, Detektion weniger Elementarteilchen eines Typs in Experimenten, die große Datenmengen aufzeichnen.
  - Kleine Wahrscheinlichkeit bei einer großen Zahl von Versuchen

Normalverteilung

- Dichtefunktion

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Parameterbereich

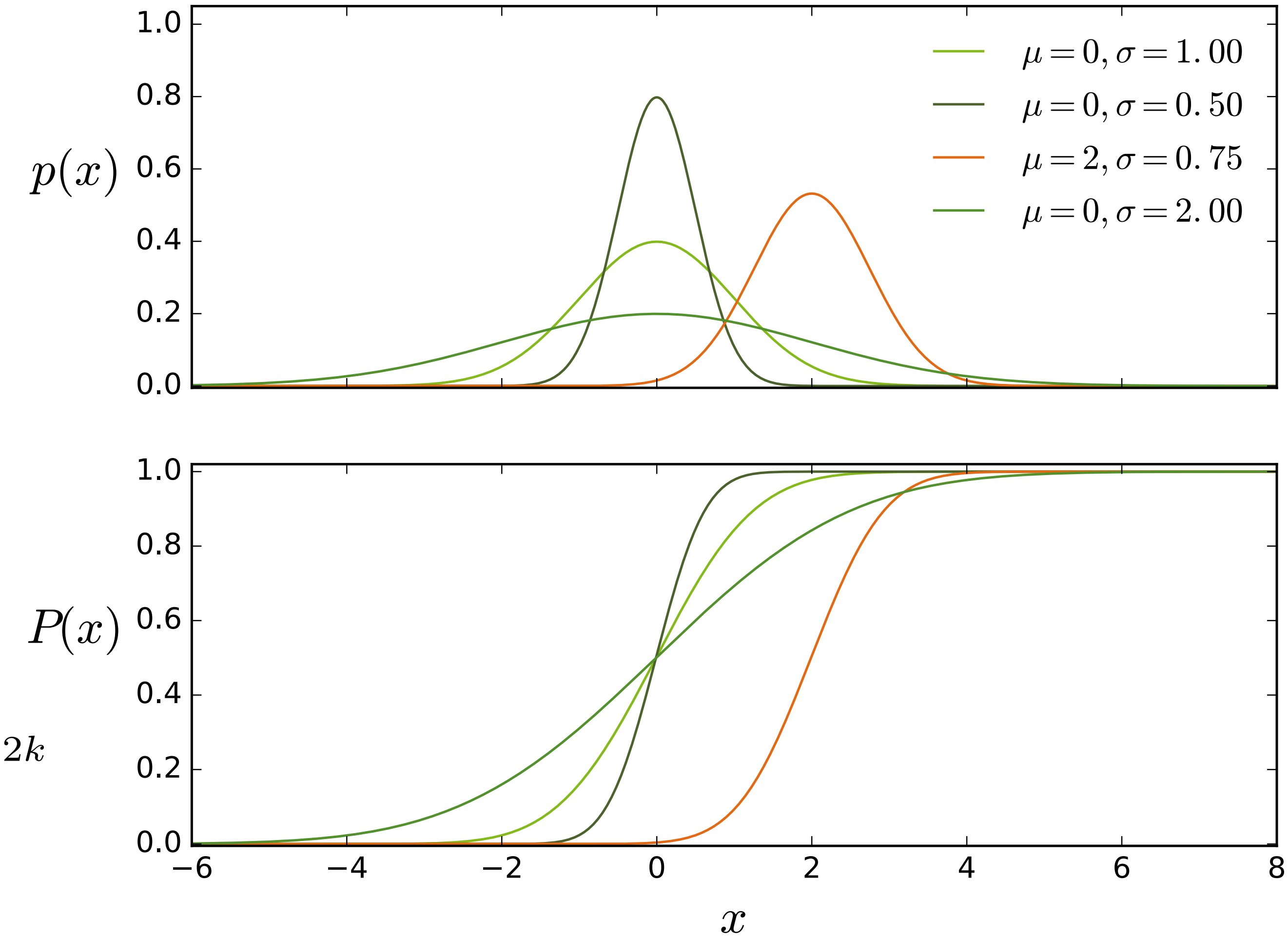
$$\mu \text{ reell}, \sigma > 0$$

- Momente

$$\mu_{2k-1} = 0$$
$$\mu_{2k} = (2k-1)!!\sigma^{2k}$$

- Erwartungswert

$$\mu$$



$$(2k)!! = 2^k \cdot k!$$

Normalverteilung

- Anwendungen
  - Grundlegende Verteilung in der Wahrscheinlichkeitsrechnung und zur statistischen Auswertung von Versuchs-, Beobachtungs – und Messergebnissen.
  - Moderate Wahrscheinlichkeiten bei einer großen Zahl von Versuchen

Logarithmische Normalverteilung

- Dichtefunktion

$$\frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(ln(x)-\mu)^2}{2\sigma^2}}\;, x > 0$$

- Parameterbereich

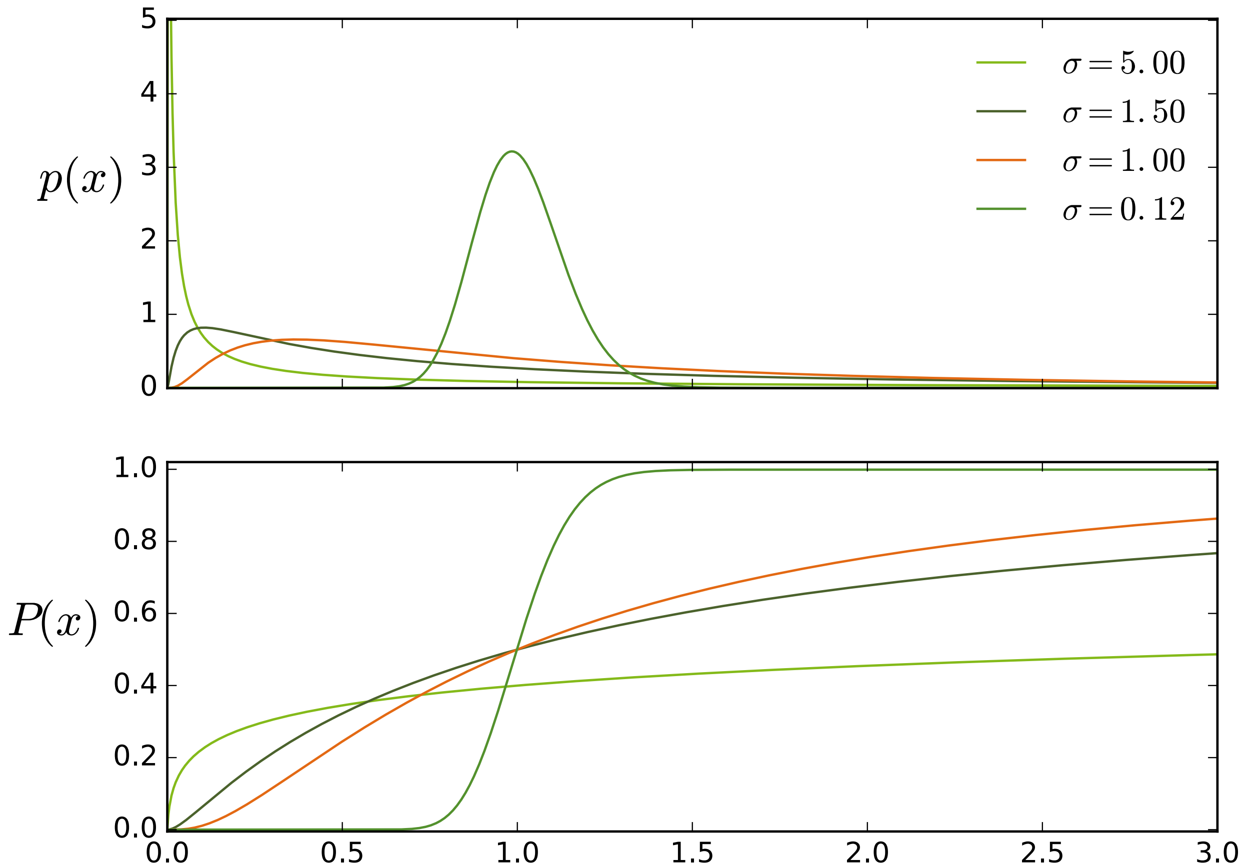
$$\mu \text{ reell}, \sigma > 0$$

- Momente

$$m_k = e^{\left(k_m u + \frac{k^2\sigma^2}{2}\right)}$$

- Erwartungswert

$$e^{\left(\mu + \frac{\sigma^2}{2}\right)}$$



Logarithmische Normalverteilung

- Anwendungen
  - Lebensdauer- sowie Festigkeitsprobleme
  - Konzentrationsuntersuchungen

Exponentialverteilung

- Dichtefunktion

$$\lambda e^{-\lambda x}$$

- Parameterbereich

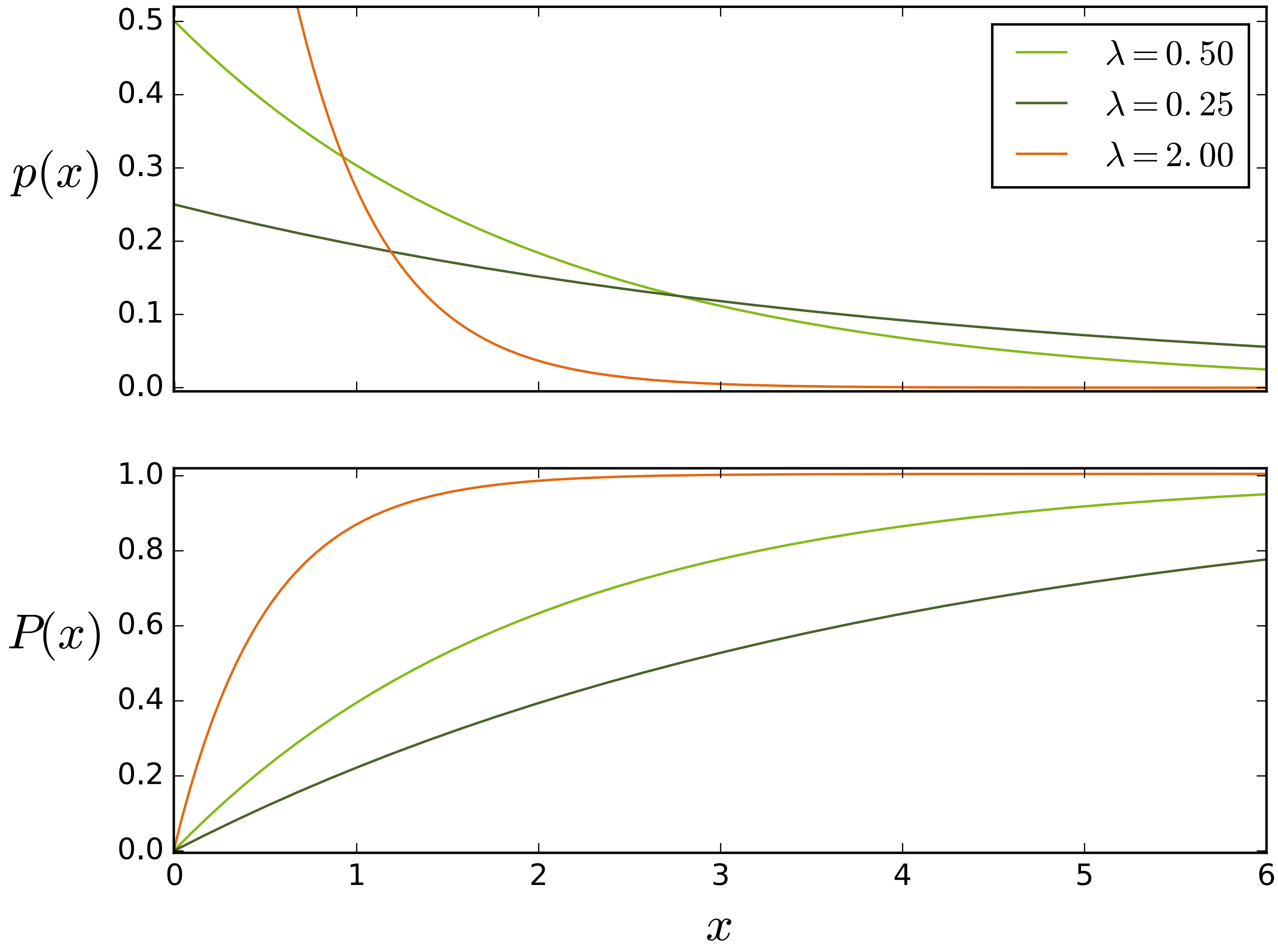
$$\lambda > 0$$

- Momente

$$m_k = \frac{\lambda}{\lambda - k}$$

- Erwartungswert

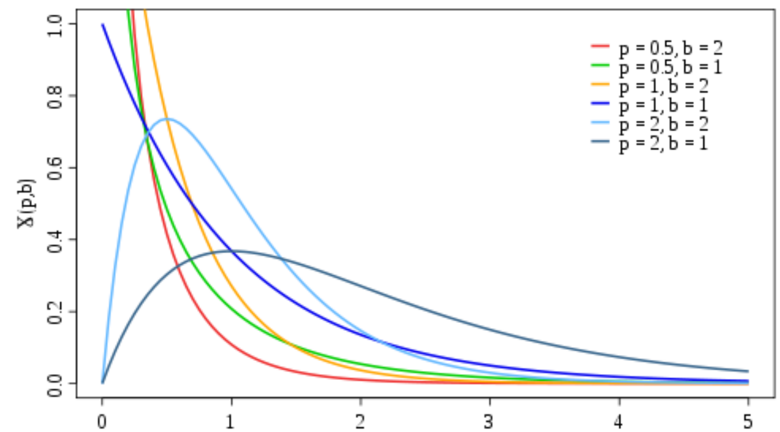
$$\frac{1}{\lambda}$$



Exponentialverteilung

- Anwendungen
  - Zeit zwischen zwei Ereignissen / Anrufen
  - Lebensdauern bei radioaktiven Zerfall

Gamma-Verteilung

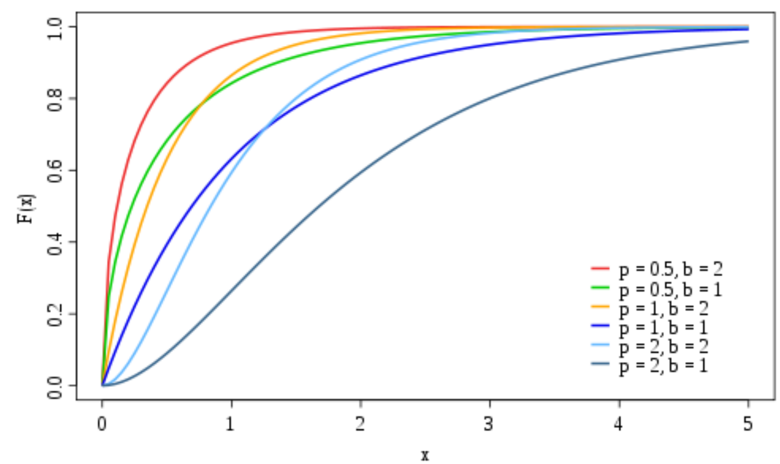


- Einzelwahrscheinlichkeiten

$$\frac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$$

- Parameterbereich

$$\lambda > 0$$



- Momente

$$m_k = \left(\frac{\lambda}{\lambda - k)}\right)^{\alpha}$$

- Erwartungswert

$$\frac{\alpha}{\lambda}$$

Gamma-Verteilung

- Anwendungen
  - Verallgemeinerung der Exponentialverteilung
  - Wahrscheinlichkeitstheorie (Reparaturzeiten)
  - Versicherungsmathematik: (Schadenshäufigkeit)

t-Test - Einstichproben Test (Vorgriff)

Fragestellung:

Teste, ob die gegebenen Daten zu einer Normalverteilung mit vorgegeben Erwartungswert $\mu=\mu_0$ passen

Test-Statistik t, mit arithmetischem Mittel x ...

$$t = \sqrt{n}\,\frac{x - \mu_0}{\sqrt{S^2}}$$

Die Test-Statistik ist unter der Nullhypothese t-verteilt mit n-1 Freiheitsgraden.

t-Verteilung

- Dichtefunktion

$$\frac{\Gamma(\frac{\nu+1}{2})(1+\frac{x^2}{\nu})^{-\frac{n\nu+1}{2}}}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}}$$

- Parameterbereich

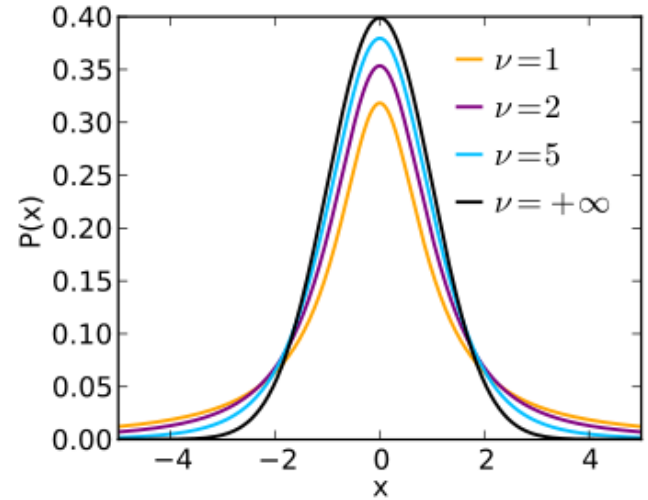
$$\nu \in \mathbb{N}$$

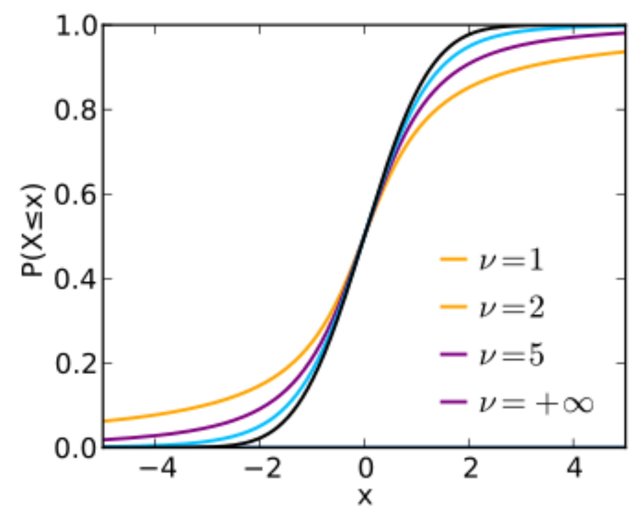
- Momente

$$m_{2k-1} = 0, 2k \leq \nu$$
$$m_{2k} = \nu^k \frac{(2k-1)!!(\nu-2k-2)!!)}{(\nu-2)!!},$$
$$2k \leq \nu - 1$$

- Erwartungswert

$$0, \nu \geq 2$$





t-Verteilung

- Anwendungen
  - Prüfen von Erwartungswerten
  - Regressions- und Korrelationsanalysen
  - T-Test

F-Test (Vorgriff)

Fragestellung:

Test, ob die Varianzen zweier Stichproben X, Y mit Umfang n_X, n_Y aus unterschiedlichen, normalverteilten Grundgesamtheiten gleich sind

H_0: $\sigma_x^2 = \sigma_y^2$

H_1: $\sigma_x^2 < \sigma_y^2$

Test-Statistik ist der Quotient der geschätzten Varianzen

$$F = \frac{S_Y^2}{S_X^2} = \frac{\frac{1}{n_Y - 1} \sum_{i=1}^{n_Y} (y_i - y)^2}{\frac{1}{n_X - 1} \sum_{i=1}^{n_X} (x_i - x)^2}$$

Unter H_0 ist die Test-Statistik F-verteilt mit n_Y-1 Freiheitsgraden im Zähler und n_X-1 Freiheitsgraden im Nenner.

F-Verteilung

- Dichtefunktion

$$\frac{\frac{m}{n}^{m/2} x^{m/2-1}\left(1-\frac{mx}{n}\right)^{-\frac{m+n}{2}}}{B(m/2,n/2)}$$

- Parameterbereich

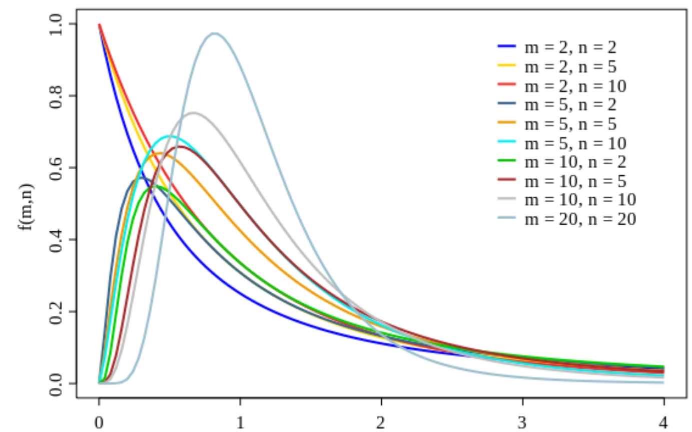
$$m, n \in \mathbb{N}$$

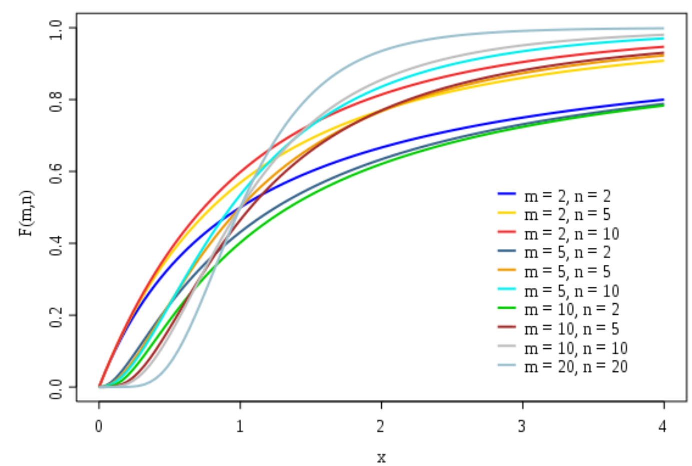
- Momente

$$m_k = \frac{\Gamma(\frac{m}{2}+k)\Gamma(\frac{n}{2}-k)n^k}{\Gamma(m/2)\Gamma(n/2)m^k}$$

- Erwartungswert

$$n/(n-2), n \geq 3$$





F-Verteilung

- Anwendungen
  - Stimmen die Streuungen zweier Verteilungen überein?
  - Varianz- und Kovarianzanalyse
  - F-Test

χ^2-Verteilung

- **Fragestellung:**
  Teste, ob n gemessene Daten y_i einem angenommenen Modell $f(x_i)$ folgen

- Dichtefunktion

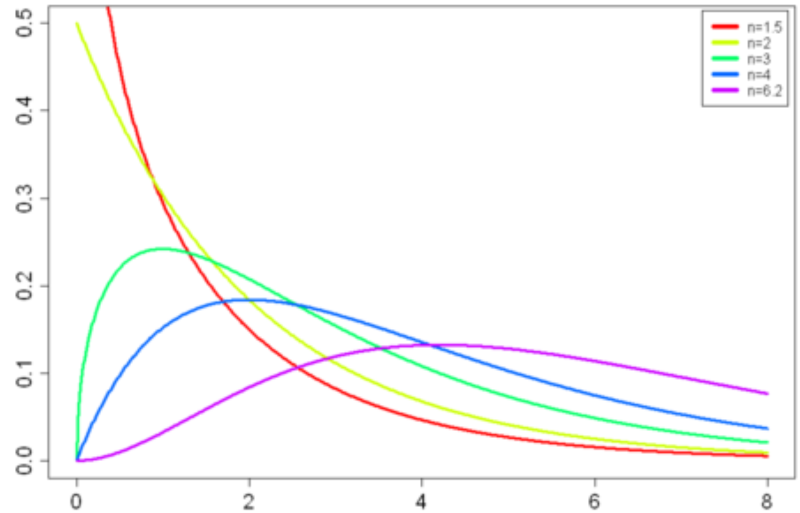
$$\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\left(\frac{n}{2}-1\right)}e^{-\frac{x}{2}}$$

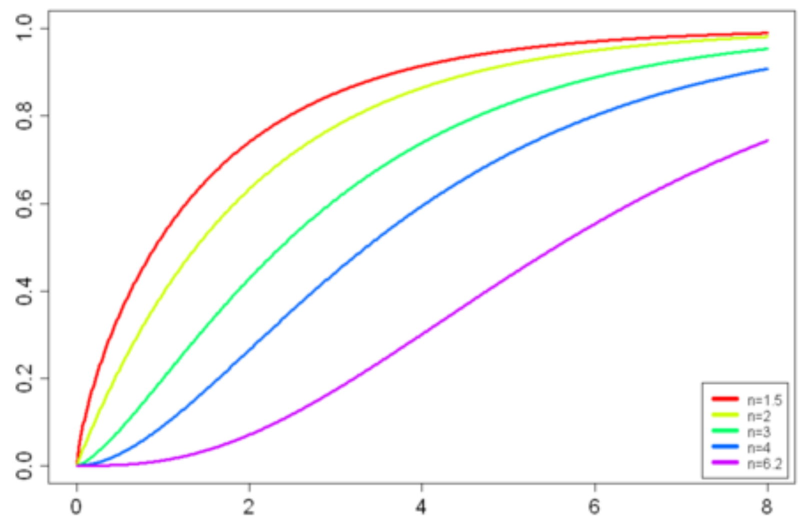
- Parameterbereich: $n \in \mathbb{N}$

- Momente

$$m_k = 2^k \frac{\Gamma\left(k+\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$$

- Erwartungswert: n





χ^2-Verteilung

- Anwendungen
  - Prüfen von Streuungen, z.B. ist eine Verteilung Gauß-verteilt?
  - χ^2 -Test

Beta-Verteilung

- Dichtefunktion

$$\frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}\,,\ 0 < x < 1$$

- Parameterbereich

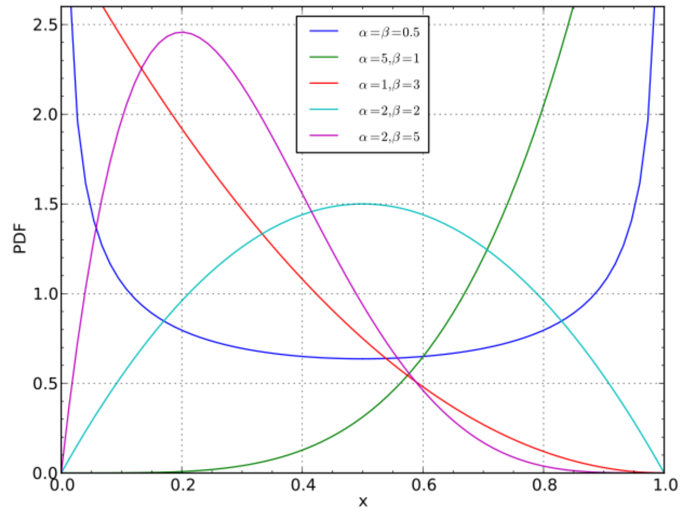
$$p > 0\,,\ q > 0$$

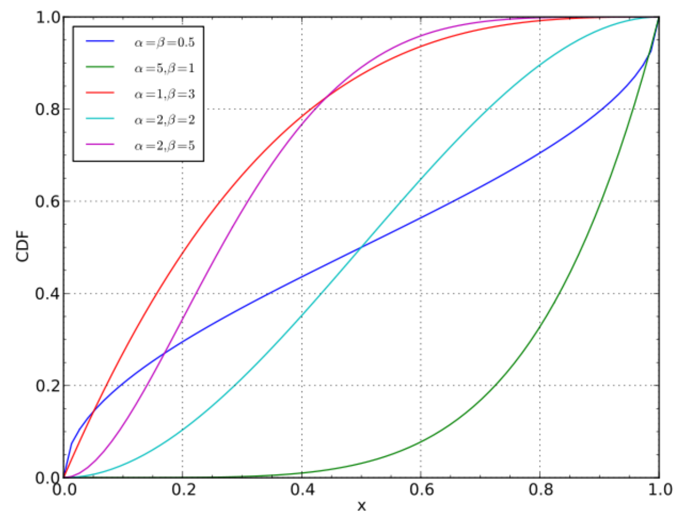
- Momente

$$m_k = \frac{B(k+p,q)}{B(p,q)}$$

- Erwartungswert

$$\frac{p}{p+q}$$





Beta-Verteilung

- Anwendungen
  - Korrelationsanalyse

Wahrscheinlichkeitsverteilungen

Vorlesung
Statistische Methoden der Datenanalyse
Prof. Dr. Dr. Wolfgang Rhode

Spezielle mehrdimensionale Verteilungen:

Mehrdimensionale Gaußverteilung

- Vektor mit *n* Variablen $\overrightarrow{X} = (x_1, x_2, ..., x_n)$
- Wahrscheinlichkeitsdichte:

$$f\left(\vec{X}\right) = k \cdot e^{-\frac{1}{2}\left(\vec{X}-\vec{a}\right)^{\top} B\left(\vec{X}-\vec{a}\right)} = k \cdot e^{-\frac{1}{2} g\left(\vec{X}\right)}$$

$$k_n = \left(\frac{\det B}{(2\pi)^n}\right)^{1/2}$$

$$\overrightarrow{a} : \quad n - \text{Komponenten Vektor},$$

$$B : \quad n \times n - \text{Matrix, symmetrisch und positiv definit}$$

Mehrdimensionale Gaußverteilung

Eigenschaften

Wahrscheinlichkeitsdichte symmetrisch um $\vec{X} = \vec{a}$

Erwartungswerte:

$$E(\vec{X} - \vec{a}) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(\vec{X}) d\vec{x} = 0$$

$$E(\vec{X}) = \vec{a}$$

Kovarianzmatrix *C*:

$$C = E[(\vec{X} - \vec{a})(\vec{X} - \vec{a})^\top] = B^{-1}$$

Mehrdimensionale Gaußverteilung

Wenn Zufallsvariablen abhängig voneinander:
Übergang zu standardisierten Variablen sinnvoll

$$u_i = \frac{x_i - a_i}{\sigma_i}, \quad i = 1, 2...$$

$$\phi(u_1, u_2) = k \cdot e^{-\frac{1}{2}\vec{u}^\top B \vec{u}} = k \cdot e^{-\frac{1}{2} g(\vec{u})}$$

Mehrdimensionale Gaußverteilung

Bei Verwendung standardisierter Variablen:

$$\rho = \frac{\mathrm{cov}(x_1, x_2)}{\sigma_1 \sigma_2} = \mathrm{cov}(u_1, u_2),$$

$$B = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$$

Zweidimensionale Gaußverteilung

$$C = B^{-1} = \begin{pmatrix} \sigma_1^2 & \mathrm{cov}(x_1, x_2) \\ \mathrm{cov}(x_1, x_2) & \sigma_2^2 \end{pmatrix}$$

$$B = \frac{1}{\sigma_1^2 \sigma_2^2 - \mathrm{cov}(x_1, x_2)^2} \begin{pmatrix} \sigma_2^2 & -\mathrm{cov}(x_1, x_2) \\ -\mathrm{cov}(x_1, x_2) & \sigma_1^2 \end{pmatrix}$$

Wenn Kovarianzen verschwinden: diagonale Matrizen

$$B = B_0 = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{pmatrix}$$

Zweidimensionale Gaußverteilung

- Daraus folgt die Wahrscheinlichkeitsdichte:

$$\phi(x) = k \cdot e^{-\frac{1}{2}(\vec{X}-\vec{a})^\top B_0 (\vec{X}-\vec{a})} = k \cdot e^{-\frac{1}{2}\frac{(x_1-a_1)^2}{\sigma_1^2}} e^{-\frac{1}{2}\frac{(x_2-a_2)^2}{\sigma_2^2}},$$

- Mit der Normierung:

$$k = k_0 = \frac{1}{2\pi\sigma_1\sigma_2}$$

Zweidimensionale Gaußverteilung

Linien gleicher Wahrscheinlichkeit als Höhenlinien

$$\phi(u_1, u_2) = \text{const} \;\Rightarrow\; -\frac{1}{2} g(\vec{u}) = \text{const}$$

$$\Rightarrow \quad -\frac{1}{2} \frac{1}{1-\rho^2} (u_1^2 + u_2^2 - 2u_1 u_2 \rho) = \text{const}$$

Zweidimensionale Gaußverteilung

Betrachte $g(\vec{u}) = 1$, so ergibt sich:

$$\frac{(x_1 - a_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - a_1)}{\sigma_1} \cdot \frac{(x_2 - a_2)}{\sigma_2} + \frac{(x_2 - a_2)^2}{\sigma_2^2} = 1 - \rho^2$$

Dies ist eine Ellipsengleichung!

Zweidimensionale Gaußverteilung

Eigenschaften der Ellipse:

Mittelwert $u_1 u_2$

Winkel *α* zwischen Ellipsen-Hauptachsen und Koordinatenachsen

Halbmesser $p_1,\ p_2$ der Ellipsen-Hauptachsen

Zweidimensionale Gaußverteilung

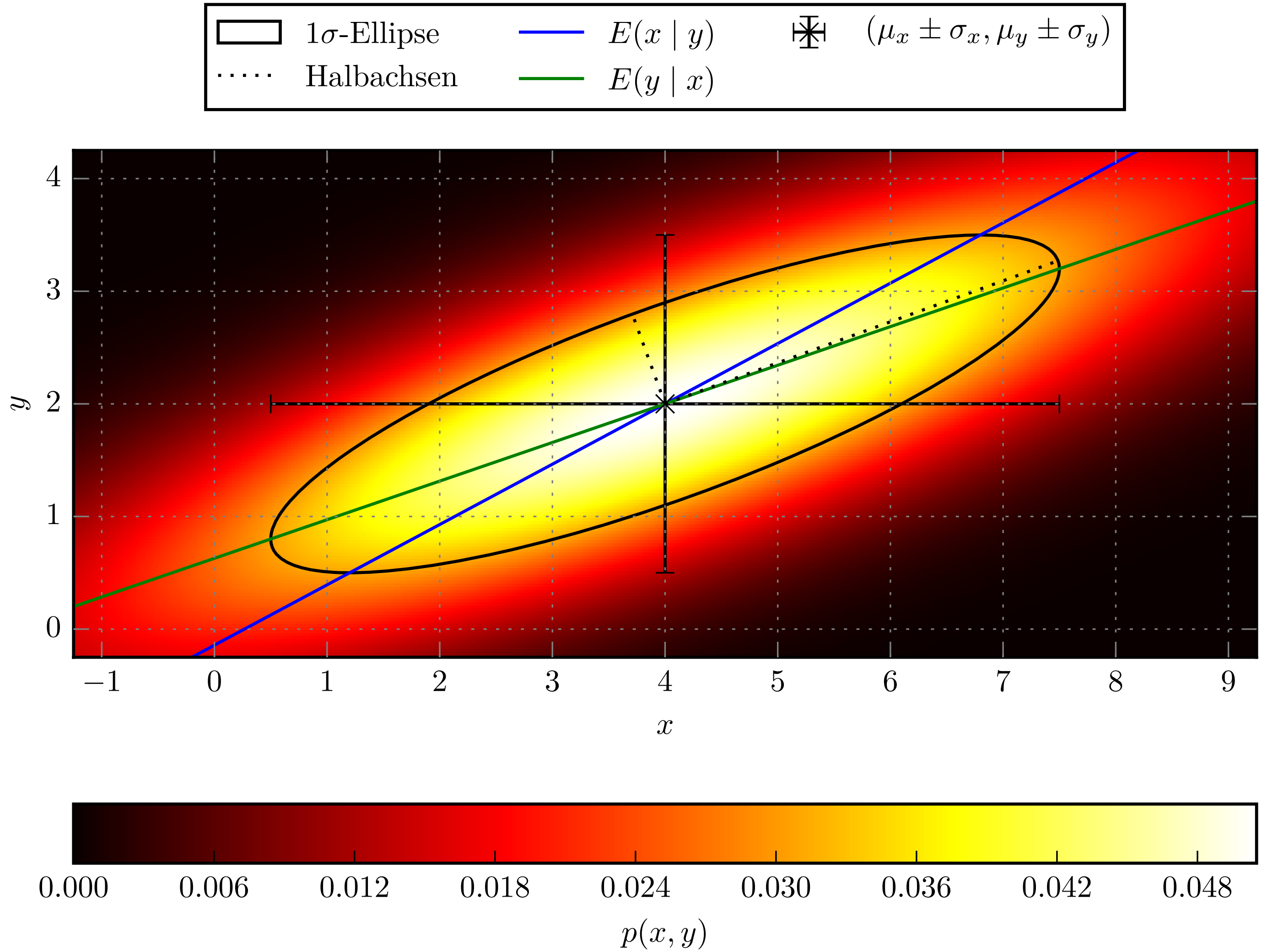
Spezielle Ellipse: Kovarianzellipse

$$\alpha = \frac{1}{2}\arctan\left(\frac{2\rho\sigma_1\sigma_2}{\sigma_1^2-\sigma_2^2}\right)$$

$$p_1^2 = (1-\rho^2)\left(\frac{\cos^2\alpha}{\sigma_1^2} - \frac{2\rho\sin\alpha\cos\alpha}{\sigma_1\sigma_2} + \frac{\sin^2\alpha}{\sigma_2^2}\right)^{-1}$$

$$p_2^2 = (1-\rho^2)\left(\frac{\sin^2\alpha}{\sigma_1^2} + \frac{2\rho\sin\alpha\cos\alpha}{\sigma_1\sigma_2} + \frac{\cos^2\alpha}{\sigma_2^2}\right)^{-1}$$

Innerhalb des 1σ-Bereichs liegen analog zum eindim. Fall 68,39%


1σ-Ellipse
Halbachsen
$E(x \mid y)$
$E(y \mid x)$
$(\mu_x \pm \sigma_x, \mu_y \pm \sigma_y)$
x
y
-1 0 1 2 3 4 5 6 7 8 9
0 1 2 3 4
0.000 0.006 0.012 0.018 0.024 0.030 0.036 0.042 0.048
$p(x, y)$

Theoreme und Sätze

- Tschebyscheff-Ungleichung
- Gesetz der großen Zahlen
- Zentraler Grenzwertsatz

Tschebyscheff-Ungleichung

- Obere Schranke für die Wahrscheinlichkeit, dass eine Zufallsvariable mehr als *k* Standardabweichungen vom Mittelwert abweicht

- Für die Wahrscheinlichkeit, dass eine Zufallsvariable einen Wert aus dem Bereich $|x - \langle x \rangle| \geq k\sigma$ stammt, ist gegeben durch:

$$\int_{-\infty}^{\langle x \rangle - k\sigma} f(x)\mathrm{d}x + \int_{\langle x \rangle + k\sigma}^{\infty} f(x)\mathrm{d}x \leq \frac{1}{k^2}$$

- Gilt unter sehr allgemeinen Bedingungen (für alle Wahrscheinlichkeitsdichten)
- Ist jedoch im Gegenzug eine sehr schwache Bedingung

Tschebyscheff-Ungleichung – Herleitung

- Die Herleitung basiert auf der Definition der Varianz:

$$\sigma^2 = \int\limits_{-\infty}^{\infty} (x - \langle x \rangle)^2 \cdot f(x) \mathrm{d}x$$

$$= \left(\int\limits_{-\infty}^{\langle x \rangle - k\sigma} + \int\limits_{\langle x \rangle - k\sigma}^{\langle x \rangle + k\sigma} + \int\limits_{\langle x \rangle + k\sigma}^{\infty} \right) (x - \langle x \rangle)^2 \cdot f(x) \mathrm{d}x$$

- Das Weglassen des mittleren Terms führt dann auf eine Ungleichung:

$$\sigma^2 \geq \left(\int\limits_{-\infty}^{\langle x \rangle - k\sigma} + \int\limits_{\langle x \rangle + k\sigma}^{\infty} \right) (x - \langle x \rangle)^2 \cdot f(x) \mathrm{d}x$$

Tschebyscheff-Ungleichung – Herleitung

- Für die Integrale gilt nun aufgrund der Grenzen:

$$\begin{aligned} x &< \langle x \rangle - k\sigma \\ x - \langle x \rangle &< -k\sigma \\ (x - \langle x \rangle)^2 &> k^2\sigma^2 \end{aligned} \qquad \text{und} \qquad \begin{aligned} x &> \langle x \rangle + k\sigma \\ x - \langle x \rangle &> k\sigma \\ (x - \langle x \rangle)^2 &> k^2\sigma^2 \end{aligned}$$

- Einsetzen liefert dann die Ungleichung:

$$\sigma^2 \geq k^2\sigma^2 \int\limits_{-\infty}^{\langle x \rangle - k\sigma} f(x)\mathrm{d}x + \int\limits_{\langle x \rangle + k\sigma}^{\infty} f(x)\mathrm{d}x$$

Gesetz der großen Zahlen

- Gegeben seien *n* unabhängige Experimente, in denen das Ereignis *j* n_j mal aufgetreten ist
- Die n_j seien binomialverteilt und $h_j = n_j / n$ sei die entsprechende Zufallsvariable
- Dann gilt für den Erwartungswert von h_j: $E(h_j) = \frac{1}{n} E(n_j) = p_j$
- Wie genau wird die unbekannte Wahrscheinlichkeit p_j damit geschätzt?
- Berechne die Varianz von h_j:

$$V(h_j) = \sigma^2(h_j) = \sigma^2(n_j/n) = \frac{1}{n^2} \cdot \sigma^2(n_j) = \frac{1}{n^2} n p_j (1 - p_j)$$

- Da $p_j(1-p_j) \leq \frac{1}{4}$ ist, gilt für die Varianz: $\sigma^2(h_j) \leq \frac{1}{4n}$
- Somit kann für große Zahlen ($n \to \infty$) der Fehler der Schätzung h_j so klein gemacht werden wie gewünscht. Der Fehler ist durch $1/2\sqrt{n}$ beschränkt

Der zentrale Grenzwertsatz

- Die Wahrscheinlichkeitsdichte der Summe $\omega = \sum_{i=1}^{n} x_i$ einer Stichprobe aus n unabhängigen Zufallsvariablen x_i mit einer beliebigen Wahrscheinlichkeitsdichte mit dem Mittelwert $\langle x \rangle$ und der Varianz σ^2 geht im Grenzfall $n \to \infty$ gegen eine Gaußverteilung mit dem Mittelwert $\langle \omega \rangle = n \cdot \langle x \rangle$ und einer Varianz $V(\omega) = n\sigma^2$

- Größen, die auf Summen von zufallsverteilten Ereignissen basieren sind gaußverteilt