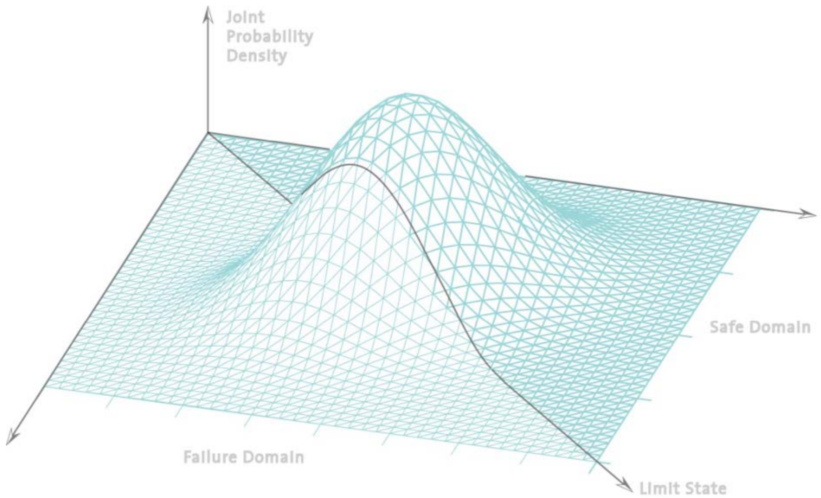


Statistik und Wahrscheinlichkeitsrechnung

Dr. Jochen Köhler

Inhalte der heutigen Vorlesung

- Fortsetzung
  - Kontinuierlicher Zufallsprozess
  - Extreme von Zufallsprozessen
- Überblick über Schätzung und Modellbildung
- Wahrscheinlichkeitsverteilungen in der Statistik
- Parameterschätzung
  - Statistische Charakteristiken von Stichproben: Mittelwert
  - Statistische Charakteristiken von Stichproben: Varianz
  - Konfidenzintervalle der Schätzer

Zufallsprozesse

Kontinuierlicher Zufallsprozess

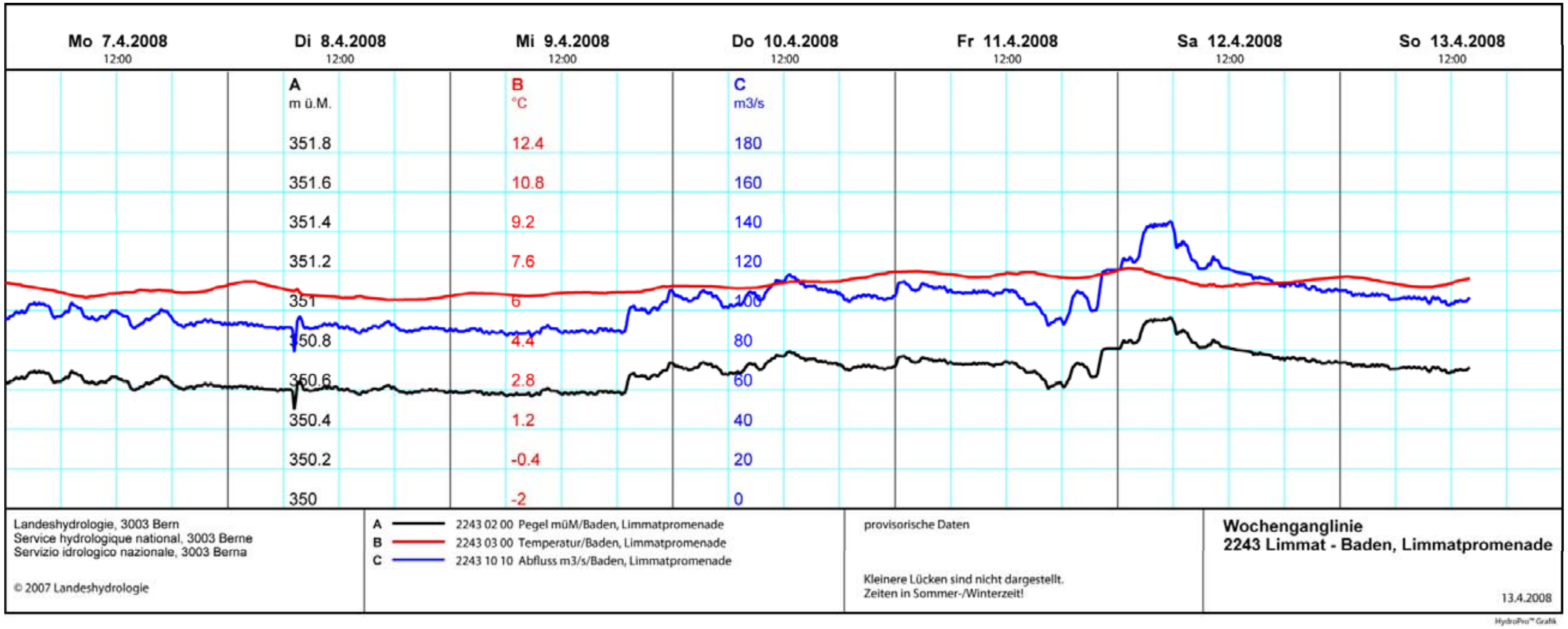
Mit einem kontinuierlichen Zufallsprozess werden Zufallsgrössen beschrieben, die sich über die Zeit kontinuierlich realisieren.

Variationen von:
Pegelständen
Windgeschwindigkeiten
Herzfrequenzen
Geschwindigkeiten von Objekten
.
.
.

Zufallsprozesse

Realisation eines kontinuierlichen Zufallsprozesses

– Pegel, Abfluss und Temperatur der Limmat bei Baden.



Zufallsprozesse

Kontinuierlicher Zufallsprozess

- Der Mittelwert der möglichen Realisation des Zufallsprozesses ist gegeben mit:

$$\mu_X(t) = E\left[X(t)\right] = \int_{-\infty}^{\infty} x\, f_X(x\,;t)dx$$

Funktion der Zeit!

- Die Autokorrelation zwischen den Realisationen an zwei Zeitpunkten ist gegeben mit:

$$R_{XX}(t_1,t_2) = E\left[X(t_1)X(t_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1\, x_2\, f_{XX}(x_1,x_2;t_1,t_2)dx_1 dx_2$$

Zufallsprozesse

Kontinuierlicher Zufallsprozess

– Die Auto-Kovarianz-Funktion ist definiert als:

$$C_{XX}(t_1,t_2) = E\left[(X(t_1)-\mu_X(t_1))(X(t_2)-\mu_X(t_2))\right]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1-\mu_X(t_1))\,(x_2-\mu_X(t_2))\,f_{XX}(x_1,x_2;t_1,t_2)dx_1dx_2$$

– für $t_1 = t_2 = t$ wird die Auto-Kovarianz-Funktion zur Kovarianz Funktion:

$$\sigma_X^2(t) = C_{XX}(t,t) = R_{XX}(t,t) - \mu_X^2(t)$$

$\sigma_X(t)$ **“Standardabweichungs Funktion”**

Zufallsprozesse

Kontinuierlicher Zufallsprozess

– Ein Vektor Zufallsprozess ist ein Zufallsprozess mit zwei oder mehr Komponenten:

$$\mathbf{X}(t) = (X_1(t), X_2(t), .., X_n(t))^T$$

mit Kovarianz Funktion:

$$\boxed{\begin{aligned}&C_{X_iX_j}(t_1,t_2) = \\ &E\left[(X_i(t_1)-\mu_{X_i}(t_1))(X_j(t_2)-\mu_{X_j}(t_2))\right]\end{aligned}}$$

$i = j$ **Auto-Kovarianz Funktion**

$i \neq j$ **Kreuz-Kovarianz Funktion**

Die Korrelationskoeffizienten Funktion ist definiert mit:

$$\rho\left[X_i(t_1), X_j(t_2)\right] = \frac{C_{X_iX_j}(t_1,t_2)}{\sigma_{X_i}(t_1)\cdot\sigma_{X_j}(t_2)}$$

Zufallsprozesse

Normal oder Gauss Prozess

– Ein Zufallsprozess $X(t)$ wird als Normal bezeichnet, wenn:

für beliebige; $X(t_1), X(t_2), X(t_3), \ldots, X(t_j)$

die gemeinsame Wahrscheinlichkeits-Verteilung von
$X(t_1), X(t_2), X(t_3), \ldots, X(t_j)$

die Normal Verteilung ist.

Zufallsprozesse

Stationarität und Ergodizität

- Ein Zufallsprozess wird als streng stationär bezeichnet, wenn alle seine Momente invariant sind über die Zeit.

- Ein Zufallsprozess wird als schwach stationär bezeichnet, wenn seine zwei ersten Momente, also der Mittelwert und die Autokorrelationsfunktion, invariant sind über die Zeit.

Zufallsprozesse

Stationarität und Ergodizität

- Ein Zufallsprozess wird als streng ergodisch bezeichnet, wenn er streng stationär ist und zusätzlich alle seine Momente aufgrund einer Realisation des Prozesses bestimmt werden können.

- Ein Zufallsprozess wird als schwach ergodisch bezeichnet, wenn er schwach stationär ist und zusätzlich seine zwei ersten Momente aufgrund einer Realisation des Prozesses bestimmt werden können.

Zufallsprozesse

Stationarität und Ergodizität

Wieso sind Annahmen bezüglich Stationarität und Ergodizität nützlich im Ingenieurwesen?

Wenn ein Zufallsprozess ergodisch ist, dann können wir probabilistische Modelle von Extremereignissen aus kurzen Referenzperioden für längere Referenzperioden extrapolieren .

Extremwertverteilungen

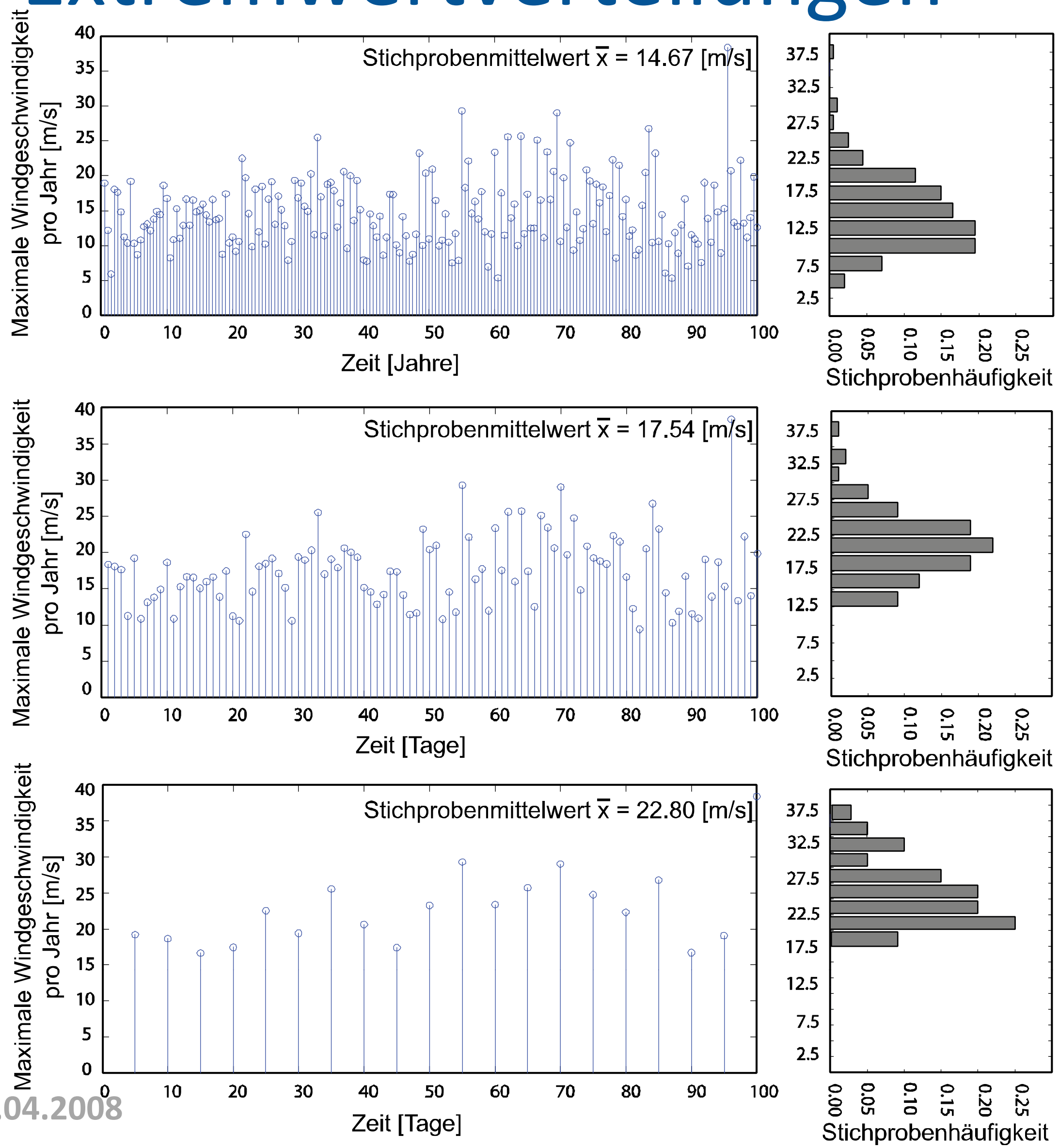
Bei Extremwerten handelt es sich jeweils um den grössten, oder um den kleinsten Wert einer bestimmten Grösse innerhalb eines gegebenen Zeitintervalls:

- Grösstes Erdbeben in einem Jahr
- Höchste Welle in einer Winterperiode
- Grösste Niederschlagsmenge in 100 Jahren

…oder um den grössten oder kleinsten Wert einer bestimmten Grösse innerhalb einer bestimmten Fläche oder Volumen:

- Die grösste Konzentration von Pestiziden in einem Kubikmeter Boden
- Das schwächste Glied in einer Kette
- Die kleinste Betonüberdeckung

Extremwertverteilungen

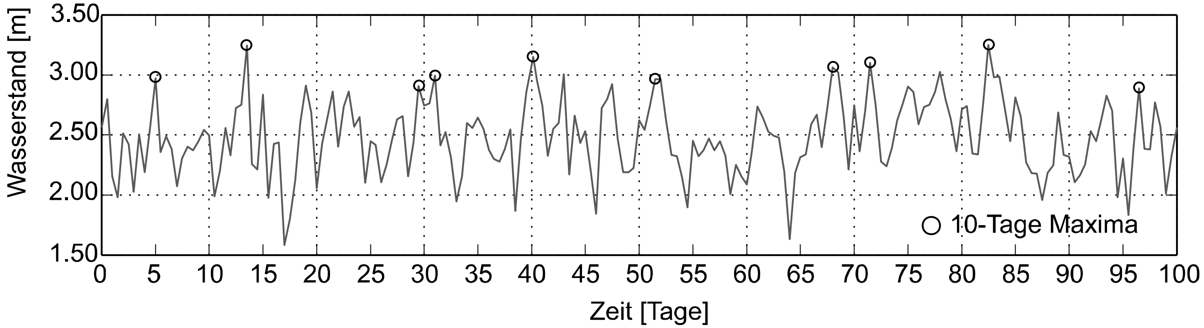


Beobachtete monatliche Extremwerte

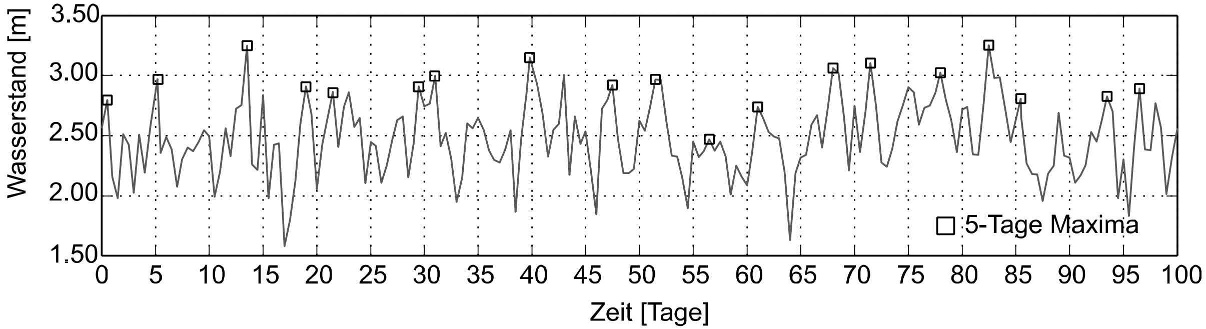
Beobachtete jährliche Extremwerte

Beobachtete 5-Jahres Extremwerte

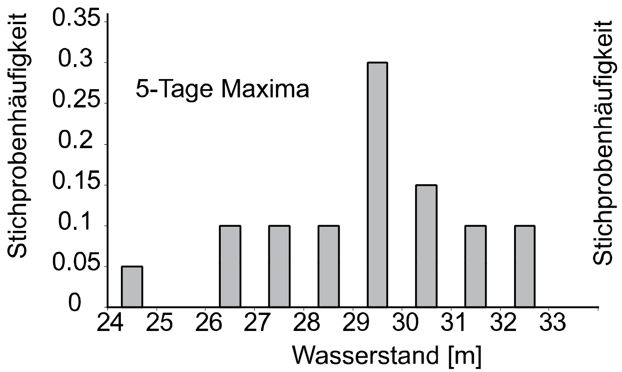
Extremwertverteilungen

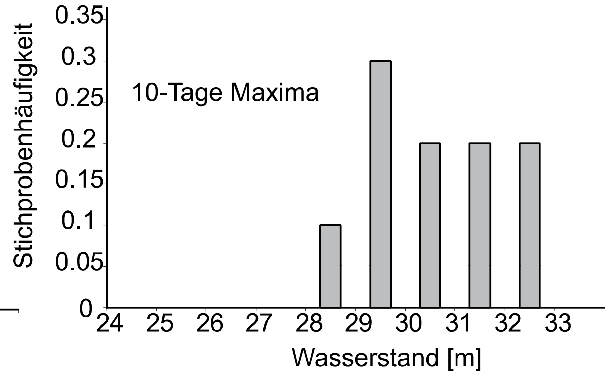


Beobachtete Extremwerte 10 Tage Intervalle



Beobachtete Extremwerte 5 Tage Intervalle





Extremwertverteilungen

Wenn die Extremwerte innerhalb einer Periode T eines ergodischen Zufallsprozesses $X(t)$ unabhängig sind und der Verteilung

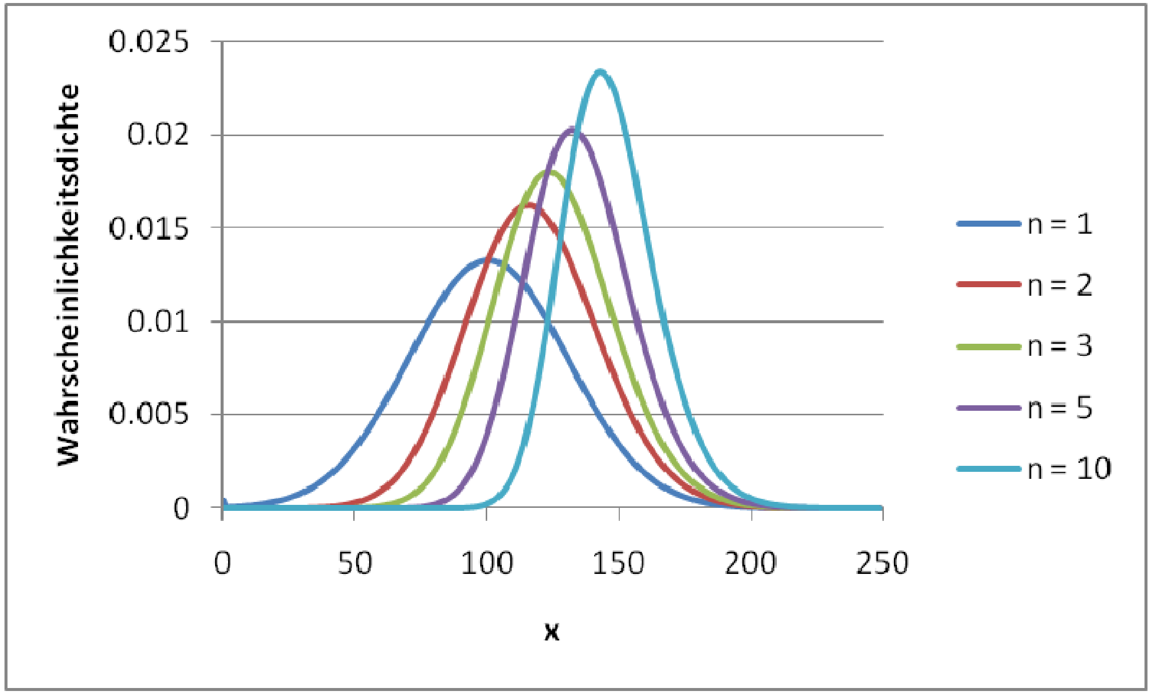
$$F_{X,T}^{\max}(x)$$

folgen, dann werden die Extremwerte des gleichen Prozesses innerhalb der Periode

$$n \cdot T$$

folgender Verteilung folgen:

$$F_{X,nT}^{\max}(x) = \left(F_{X,T}^{\max}(x)\right)^n$$



Extremwertverteilungen: Gumbel Max

Wenn der obere Schwanz der Wahrscheinlichkeitsdichtefunktion exponentiell abfällt (Exponential-, Normal- und Gammaverteilung), dann wird das Maximum im Zeitintervall T als Typ I extremverteilt bezeichnet.

$$f_{X,T}^{\max}(x) = \alpha \exp(-\alpha(x-u) - \exp(-\alpha(x-u)))$$

$$F_{X,T}^{\max}(x) = \exp(-\exp(-\alpha(x-u)))$$

$$\mu_{X_T^{\max}} = u + \frac{\gamma}{\alpha} = u + \frac{0.577216}{\alpha}$$

$$\sigma_{X_T^{\max}} = \frac{\pi}{\alpha\sqrt{6}}$$

Für zunehmende Zeitintervalle ist die Varianz konstant, aber der Mittelwert wird grösser:

$$\mu_{X_{nT}^{\max}} = \mu_{X_T^{\max}} + \frac{\sqrt{6}}{\pi}\sigma_{X_T^{\max}} \ln(n)$$

Extremwertverteilungen: Frechet Max

Wenn eine Wahrscheinlichkeitsdichtefunktion unten bei Null begrenzt ist, und in ihrem oberen Ende abfällt in der Form

$$F_X(x) = 1 - \beta(\frac{1}{x})^k$$

dann wird das Maximum im Zeitintervall T als Typ II extremverteilt bezeichnet.

$$F_{X,T}^{\max}(x) = \exp(-\left(\frac{u}{x}\right)^k)$$

$$f_{X,T}^{\max}(x) = \frac{k}{u}\left(\frac{u}{x}\right)^{k+1} \exp(-\left(\frac{u}{x}\right)^k)$$

Mittelwert und Standardabweichung

$$\mu_{X_T^{\max}} = u\Gamma(1 - \frac{1}{k})$$

$$\sigma^2_{X_T^{\max}} = u^2\left[\Gamma(1 - \frac{2}{k}) - \Gamma^2(1 - \frac{1}{k})\right]$$

Extremwertverteilungen: Weibull Min

Wenn eine Wahrscheinlichkeitsdichtefunktion unten bei ε begrenzt ist, und das untere Ende zu ε hin abfällt in der Form

$$F(x) = c(x-\varepsilon)^k$$

dann wird das Maximum im Zeitintervall T als Typ III extremverteilt bezeichnet.

$$F_{X,T}^{\min}(x) = 1 - \exp\left(-\left(\frac{x-\varepsilon}{u-\varepsilon}\right)^k\right)$$

$$f_{X,T}^{\min}(x) = \frac{k}{u-\varepsilon}\left(\frac{x-\varepsilon}{u-\varepsilon}\right)^{k-1}\exp\left(-\left(\frac{x-\varepsilon}{u-\varepsilon}\right)^k\right)$$

Mittelwert und Standardabweichung

$$\mu_{X_T^{\min}} = \varepsilon + (u-\varepsilon)\Gamma(1+\frac{1}{k})$$

$$\sigma_{X_T^{\min}}^2 = (u-\varepsilon)^2\left[\Gamma(1+\frac{2}{k}) - \Gamma^2(1+\frac{1}{k})\right]$$

Wiederkehrperiode

Die Wiederkehrperiode für Extremereignisse T_R wird definiert als:

$$T_R = n \cdot T = \frac{1}{(1 - F_{X,T}^{\max}(x))} T$$

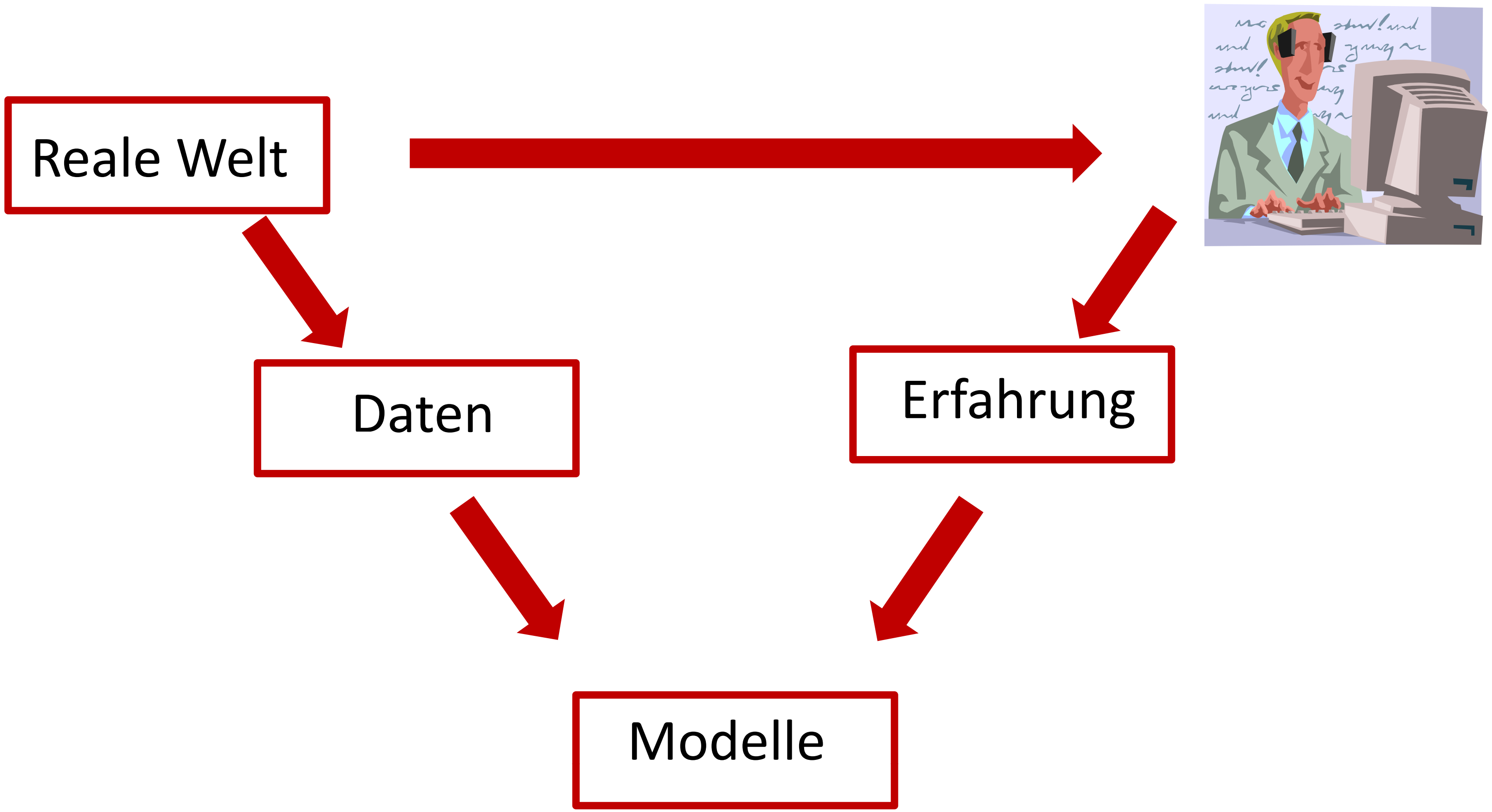
Beispiel:
Nehmen wir an, dass gemäss der kumulativen Wahrscheinlichkeits-verteilungsfunktion des jährlichen Maximums der Verkehrslasten, die jährliche Wahrscheinlichkeit einer Verkehrslast grösser als 100 Tonnen gleich 0.02 ist - dann ist die Wiederkehrperiode einer solchen Last:

$$T_R = n \cdot T = \frac{1}{0.02} 1 = \frac{1}{0.02} 1 = 50 \text{ Jahre}$$

T=1 (jährliche Wahrscheinlichkeit)

Überblick Schätzung und Modellbildung

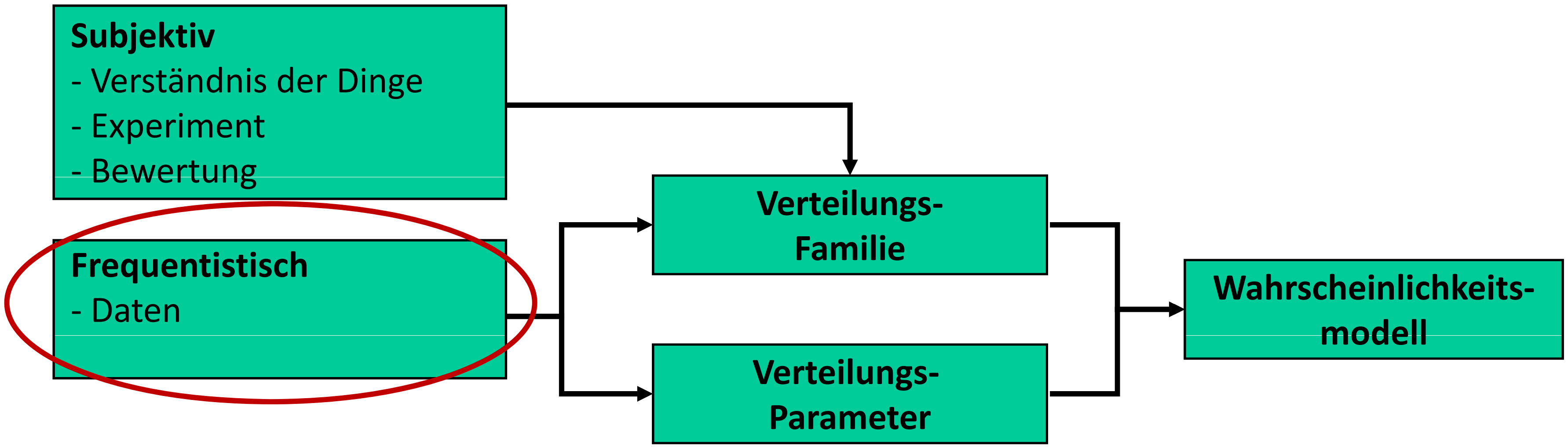
Wie kommen Ingenieure zu Wissen?



Überblick Schätzung und Modellbildung

Unterschiedliche Typen an Informationen werden zur Bildung von Ingenieurmodellen verwendet

- Subjektive Information
- Frequentistische Information



Überblick Schätzung und Modellbildung

Die Modellbildung kann in fünf Schritten erfolgen:

1) Bewertung und statistische Quantifizierung vorhandener Daten
2) Wahl der Verteilungsfunktion
3) Schätzung der Verteilungsparameter
4) Modellvalidierung
5) Modellaktualisierung

Wahrscheinlichkeitsverteilungsfunktionen

In der klassischen Statistik werden häufig bestimmte Wahrscheinlichkeitsverteilungsfunktionen, welche alle von der Normalverteilung abgeleitet werden können, verwendet und zur Bewertung und zum Testen verwendet.

Diese Wahrscheinlichkeitsverteilungsfunktionen sind:

- Chi-Quadrat Verteilung
- Chi-Verteilung
- t-Verteilung
- F-Verteilung

Wahrscheinlichkeitsverteilungsfunktionen

Chi-Quadrat Verteilung (χ^2)

Wenn $X_i, i = 1,2,\ldots n$
standardnormal verteilt und unabhängige Zufallsvariablen sind , dann ist die Summe der Quadrate der Zufallsvariablen Chi-Quadrat verteilt.

$$Y_n = \sum_{i=1}^{n} X_i^2$$

Die Chi-Quadrat Verteilung ist regenerativ, d.h. die Summe der Chi-Quadrat verteilten Zufallsvariablen ist auch wieder Chi-Quadrat verteilt.

Wahrscheinlichkeitsverteilungsfunktionen

Chi-Quadrat Verteilung (χ^2)

Betrachte den einfachsten Fall mit n=*1*, d.h. $Y_1 = X^2$
dann können wir schreiben

$$F_{Y_1}(y) = P(Y_1 \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le +\sqrt{y})$$

$$= F_X(\sqrt{y}) - F_X(-\sqrt{y}) = F_X(\sqrt{y}) - (1 - F_X(\sqrt{y})) =$$

$$= 2F_X(\sqrt{y}) - 1$$

und bekommen

$$f_{Y_1}(y) = \frac{dF_{Y_1}(y)}{dy} = \frac{d(2F_X(\sqrt{y})-1)}{dy} = y^{-\frac{1}{2}} f_X(\sqrt{y}) = \frac{1}{\sqrt{2\pi y}} \exp(-\frac{1}{2} y)$$

Wahrscheinlichkeitsverteilungsfunktionen

Chi-Quadrat Verteilung (χ^2)

Chi-Quadrat Wahrscheinlichkeitsverteilung ist gegeben durch

$$f_{Y_n}(y_n) = \frac{y_n^{(n/2-1)}}{2^{n/2}\Gamma(n/2)}\exp(-y_n/2), \qquad y_n \geq 0$$

Der Mittelwert ist $\mu_{Y_n} = n$ ← **Freiheitsgrade**

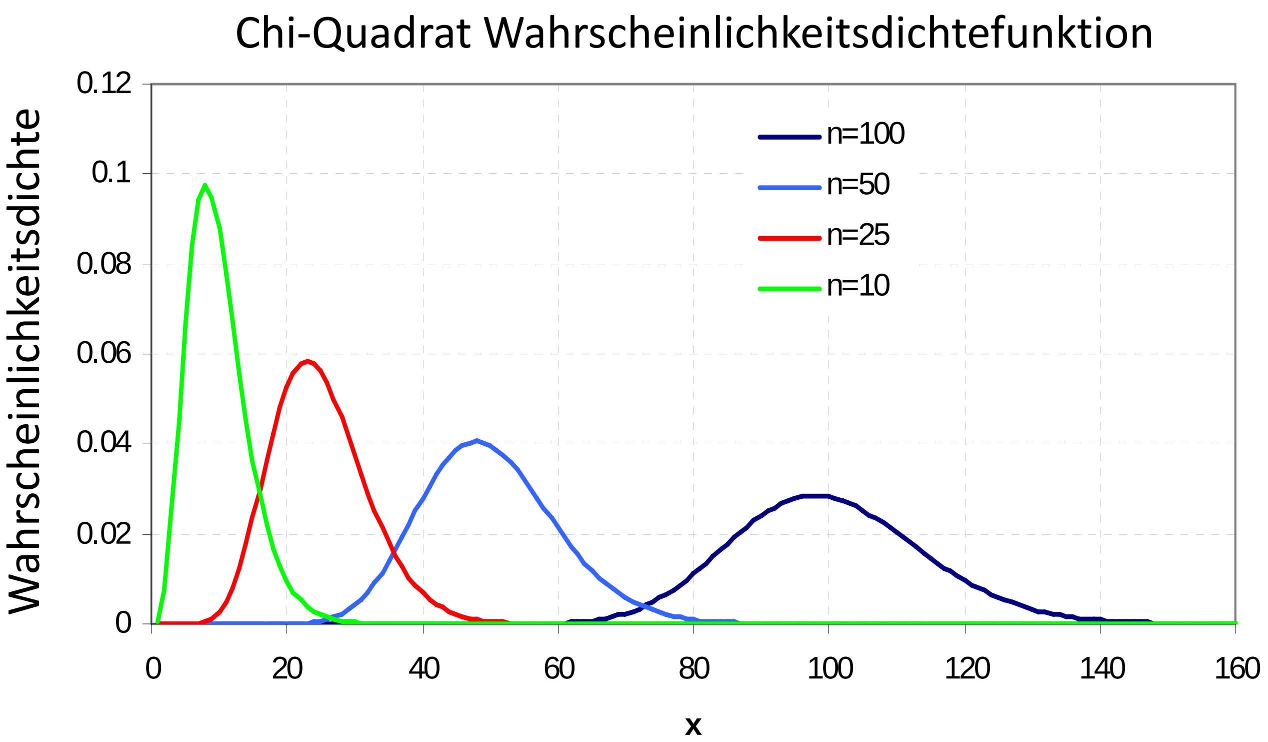
Die Varianz ist $\sigma_{Y_n}^2 = 2n$

$$\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}dt$$ ist die komplette Gamma Funktion

Für grosse n konvergiert die Chi-Quadrat Verteilung zu einer Normalverteilung.

Wahrscheinlichkeitsverteilungsfunktionen

Chi-Quadrat Wahrscheinlichkeitsdichtefunktion



Wahrscheinlichkeitsverteilungsfunktionen

Chi-Verteilung (χ)

Wenn die Zufallsvariable Z durch die Wurzel von der Chi-Quadrat verteilten Zufallsvariable gegeben ist, d.h.

$$Z = \sqrt{Y_n}$$

dann ist sie Chi-verteilt mit n Freiheitsgraden.

Wahrscheinlichkeitsverteilungsfunktionen

Chi-Verteilung (χ)

Angenommen, dass Y_n Chi-Quadrat verteilt ist mit n Freiheitsgraden.

Wenn $Z=\sqrt{Y_n}$ dann können wir schreiben

$$F_Z(z)=P(Z\leq z)=P(\sqrt{Y_n}\leq z)=P(Y_n\leq z^2)=F_{Y_n}(z^2)$$

Und wir bekommen

$$f_Z(z)=\frac{dF_Z(z)}{dz}=\frac{dF_{Y_n}(z^2)}{dz}=2zf_{Y_n}(z^2)=\frac{z^{n-1}}{2^{n/2-1}\Gamma(n/2)}\exp(-\frac{1}{2}z^2)$$

Wahrscheinlichkeitsverteilungsfunktionen

Chi-Wahrscheinlichkeitsverteilung

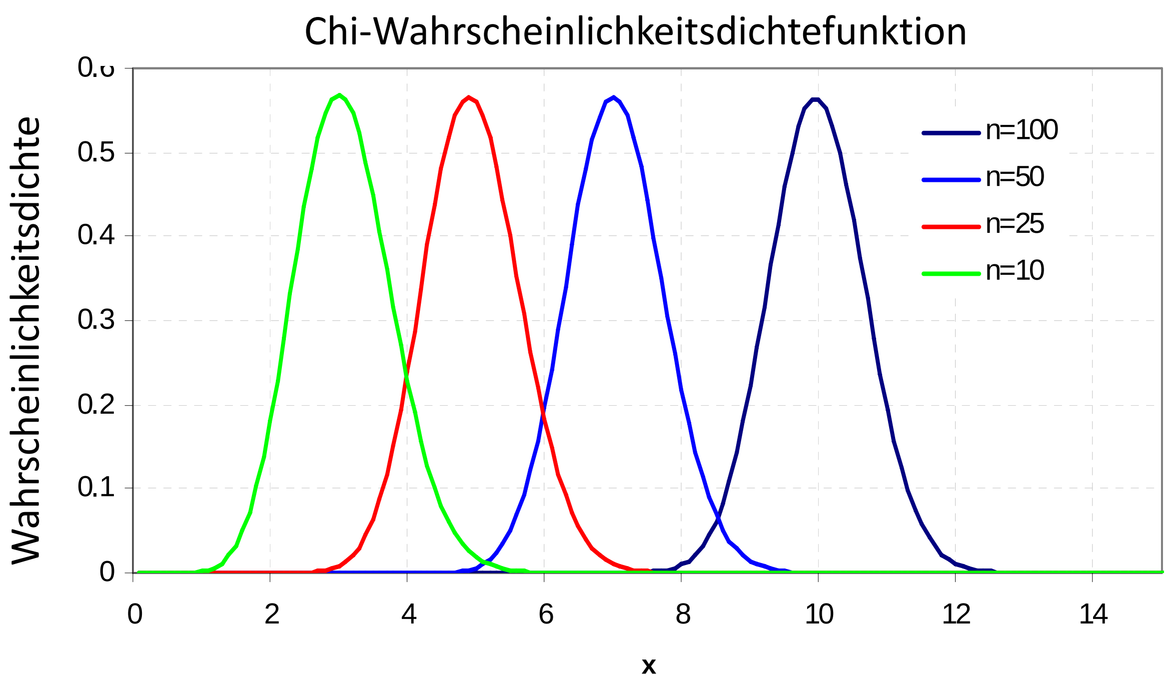
Ist gegeben durch $f_Z(z) = \frac{z^{(n-1)}}{2^{n/2-1}\Gamma(n/2)} \exp(-z^2/2), \qquad z \geq 0$

Der Mittelwert ist $\mu_z = \sqrt{2}\frac{\Gamma((n+1)/2)}{\Gamma(n/2)}$

Die Varianz ist $\sigma_z^2 = n - 2\frac{\Gamma^2((n+1)/2)}{\Gamma^2(n/2)}$

Wahrscheinlichkeitsverteilungsfunktionen

Chi-Wahrscheinlichkeitsdichtefunktion



Wahrscheinlichkeitsverteilungsfunktionen

t-Verteilung (Student-Verteilung)

Wenn eine Standardnormalverteilte Zufallsvariable S durch eine Chi-verteilte Zufallsvariable geteilt wird, d.h.

$$S = \frac{X}{\frac{\sqrt{\sum_{i=1}^{n} X_i^2}}{n}} = \frac{X}{\frac{\sqrt{Y_n}}{n}} = \frac{X}{\frac{Z}{n}} = \frac{nX}{Z}$$

Dann heisst die Verteilung von S *t-Verteilung* bzw. *Student-Verteilung* mit n Freiheitsgraden.

Für grosse n konvergiert die *t-Verteilung* zu einer Normalverteilung

Wahrscheinlichkeitsverteilungsfunktionen

t-Wahrscheinlichkeitsverteilung (Student`s)

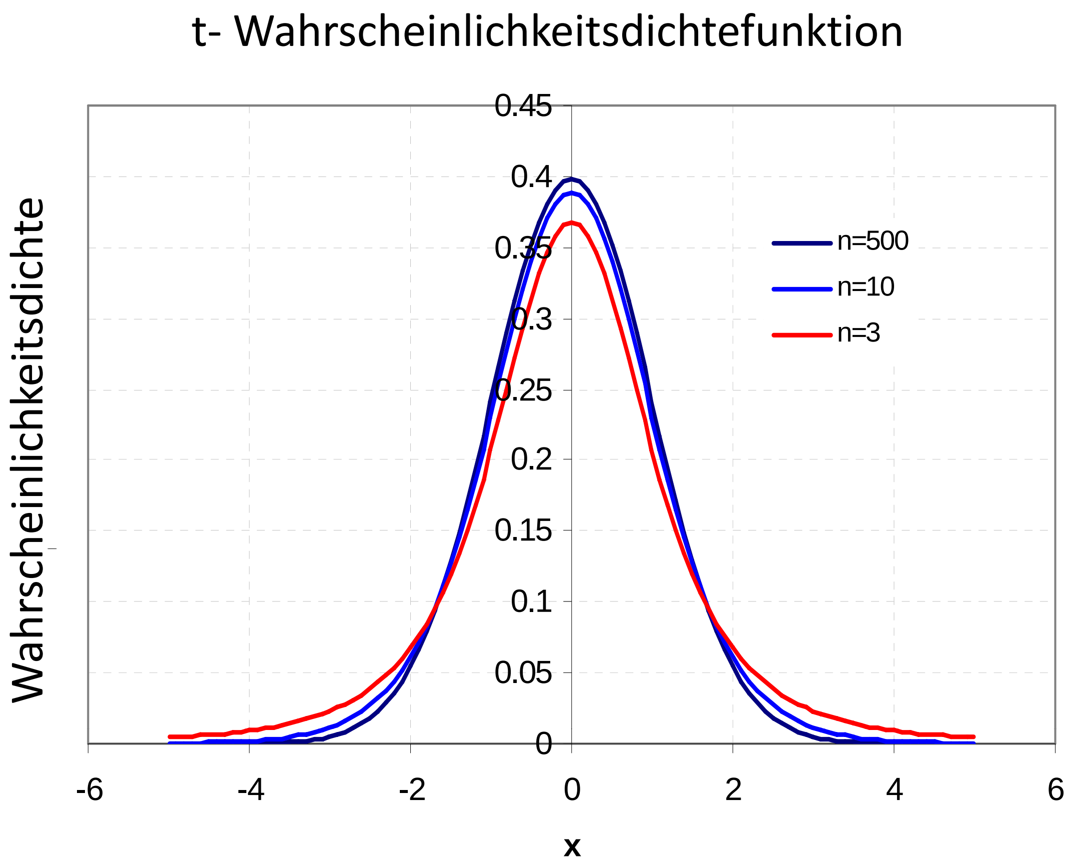
Ist gegeben durch $f_S(s) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\ \Gamma(n/2)} \left(1 + \frac{s^2}{n}\right)^{-(n+1)/2}$, $-\infty \leq s \leq \infty$

Der Mittelwert ist Null

Die Varianz ist $\sigma_S^2 = \frac{n}{n-2}$

Wahrscheinlichkeitsverteilungsfunktionen

t-Wahrscheinlichkeitsdichtefunktion



Wahrscheinlichkeitsverteilungsfunktionen

F-Verteilung

Wenn eine Zufallvariable Q ist gegeben als das Verhältnis zwischen zwei Chi-Quadrat verteilten Zufallsvariablen, d.h.

$$Q = \frac{Y_{n_1}}{Y_{n_2}}$$

Die Verteilung heisst F-Verteilung mit den Freiheitsgraden n_1, n_2

Wahrscheinlichkeitsverteilungsfunktionen

F-Wahrscheinlichkeitsverteilung

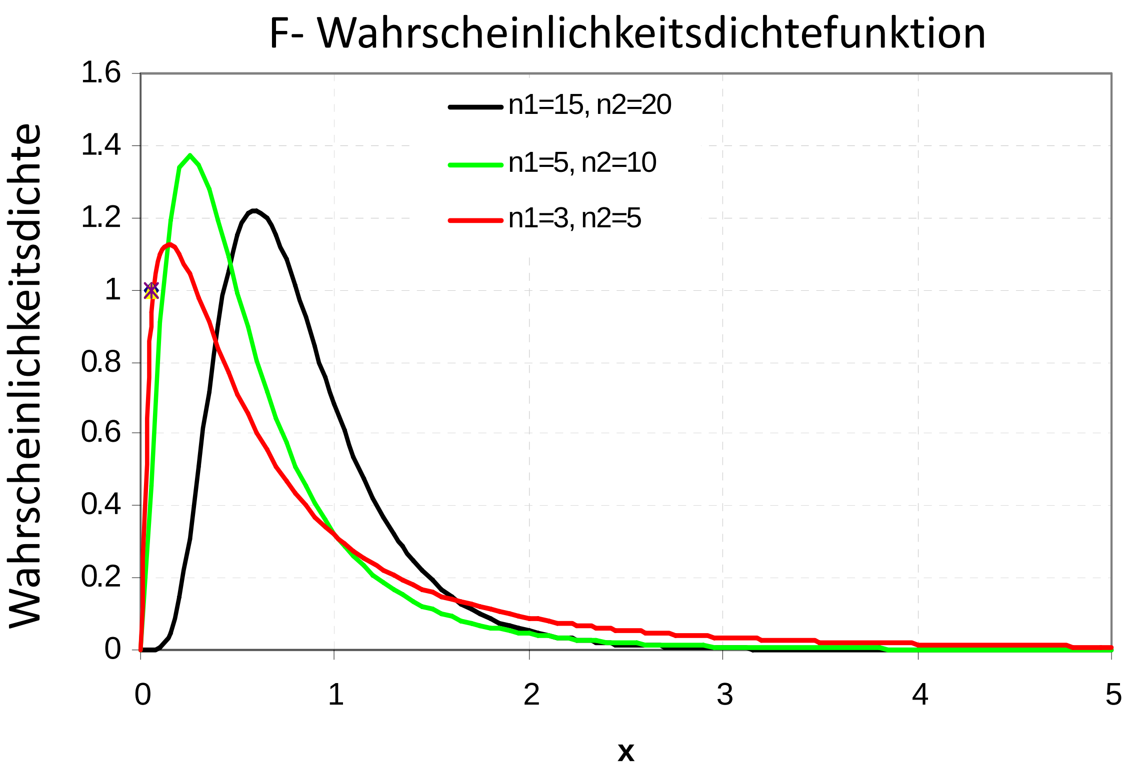
Ist gegeben durch $f_Q(q) = \dfrac{\Gamma((n_1+n_2)/2)q^{(n_1-2)/2}(1+q)^{-(n_1+n_2)/2}}{\Gamma(n_1/2)\Gamma(n_2/2)}, \qquad q \geq 0$

Der Mittelwert ist $\mu_Q = \dfrac{n_2}{n_2-2}, \qquad n_2 > 2$

Die Varianz ist $\sigma_Q^2 = \dfrac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}, \qquad n_2 > 4$

Wahrscheinlichkeitsverteilungsfunktionen

F-Wahrscheinlichkeitsdichtefunktion



Wahrscheinlichkeitsverteilungsfunktionen

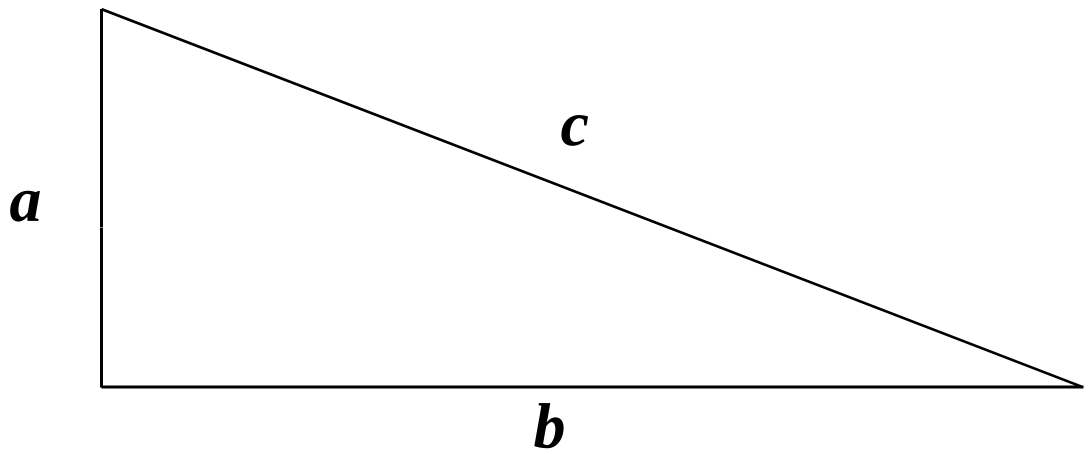
Zusammenfassung von abgeleiteten Wahrscheinlichkeits-funktionen

Verteilungstyp	Wann
Chi-Quadrat Verteilung	Summe der Quadrate N(0;1)
Chi-Verteilung	Wurzel von Chi-Quadrat
t-Verteilung	Verhältnis von N(0;1) zu Chi/n
F-Verteilung	Verhältnis von zwei Chi-Quadrat

Wahrscheinlichkeitsverteilungsfunktionen

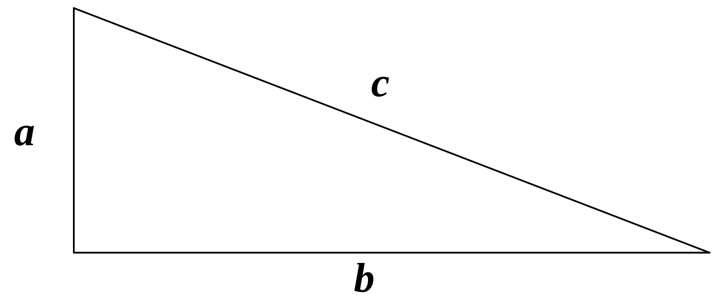
Beispiel Chi-Verteilung

Es wurden Messungen von den Seiten a und b durchgeführt, mit der Absicht, die Seite c bestimmen zu können.



Verteilungsfamilien in der Statistik

Beispiel einer Chi Verteilung

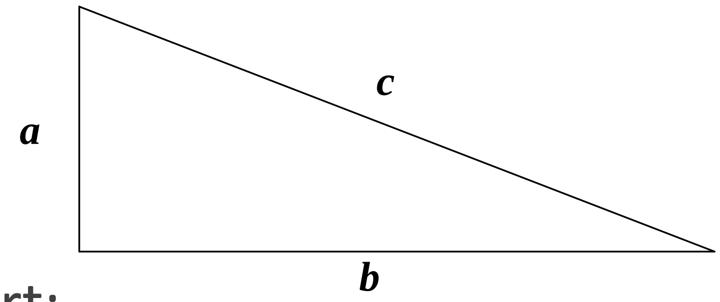


Es wird angenommen, dass die Messungen von a und b mit dem selben absoluten Fehler ε durchgeführt werden, welcher als $N(0;\, \sigma_\varepsilon)$ angenommen wird (Normalverteilt, nicht verzerrt und mit der Standardabweichung σ_ε.).

Bestimme die statistischen Charakteristiken des Fehlers in c, welcher durch a und b bestimmt wurde.

Verteilungsfamilien in der Statistik

Beispiel einer Chi Verteilung



Der Fehler setzt sich folgendermassen fort:

$$\varepsilon_c = \sqrt{\varepsilon_a^2 + \varepsilon_b^2}$$

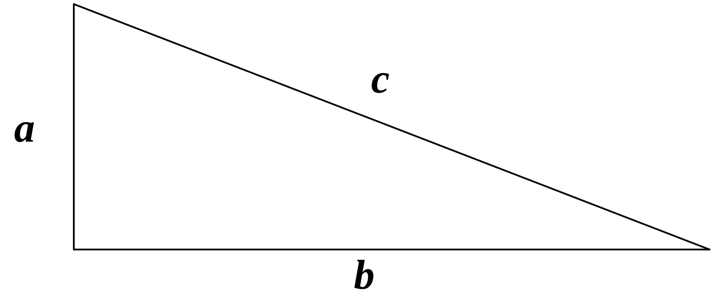
Daraus lässt sich folgen, dass

$$\frac{\varepsilon_c}{\sigma_\varepsilon} = \sqrt{\left(\frac{\varepsilon_a}{\sigma_\varepsilon}\right)^2 + \left(\frac{\varepsilon_b}{\sigma_\varepsilon}\right)^2}$$

Chi-verteilt ist, mit zwei Freiheitsgraden.

Verteilungsfamilien in der Statistik

Beispiel einer Chi Verteilung



Die Wahrscheinlichkeitsdichtefunktion von $Z = \dfrac{\varepsilon_c}{\sigma_\varepsilon}$ kann bestimmt werden durch

$$f_Z(z) = z\exp(-0.5z^2), \qquad z \geq 0$$

Unter Einhaltung von

$$f_{\varepsilon_c}(\varepsilon_c) = \frac{\varepsilon_c}{\sigma_\varepsilon}\exp\left(-0.5\varepsilon_c^2 / \sigma_\varepsilon^2\right), \qquad \varepsilon_c \geq 0$$

Parameterschätzung für Stichproben

Wenn neue Daten verfügbar sind, besteht der erste Schritt darin, diese zu beurteilen.

Daten/Beobachtungen

n	x_n	$F_X(x_n)$
1	24.4	0.047619048
2	27.6	0.095238095
3	27.8	0.142857143
4	27.9	0.19047619
5	28.5	0.238095238
6	30.1	0.285714286
7	30.3	0.333333333
8	31.7	0.380952381
9	32.2	0.428571429
10	32.8	0.476190476
11	33.3	0.523809524
12	33.5	0.571428571
13	34.1	0.619047619
14	34.6	0.666666667
15	35.8	0.714285714
16	35.9	0.761904762
17	36.8	0.80952381
18	37.1	0.857142857
19	39.2	0.904761905
20	39.7	0.952380952

→ **Mittelwert**

→ **Varianz**

→ **Median**

→

→ **Usw..**

Funktion von Stichproben

Stichproben Charakteristik

oder

Stichproben Statistik

Parameterschätzung für Stichproben

Die statistischen Charakteristiken von Stichproben Statistiken werden im folgenden genauer betrachtet, um die darin enthaltenen Informationen, besser zu verstehen.

Angenommen wir haben noch unbekannte Stichproben aus einem Experimentergebnis $X_i, \ i = 1,2,..n$

generiert durch die kumulative Verteilungsfunktion

$$F_{X_i}(x_i, \mathbf{p}) = F_X(x, \mathbf{p}), i = 1,2,..n$$

können wir die Stichproben Statistiken beschreiben für den Stichprobenmittelwert und die Stichprobenvarianz.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Parameterschätzung für Stichproben

Die Stichproben Statistiken sind Zufallsvariablen, solange die Ergebnisse das Experiments noch nicht realisiert sind.

Daher kann der Erwartungswert und die Varianz der Stichproben Statistik folgendermassen bestimmt werden:

$$E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n} n \cdot \mu_X = \mu_X$$

$$Var\left[\bar{X}\right] = Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} Var\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var\left[X_i\right] = \frac{1}{n}\sigma_X^2$$

Parameterschätzung für Stichproben

Die Wahrscheinlichkeitsdichtefunktion für den Stichprobendurchschnitt kann als eine Normalverteilung angenommen werden – Zentraler Grenzwertsatz