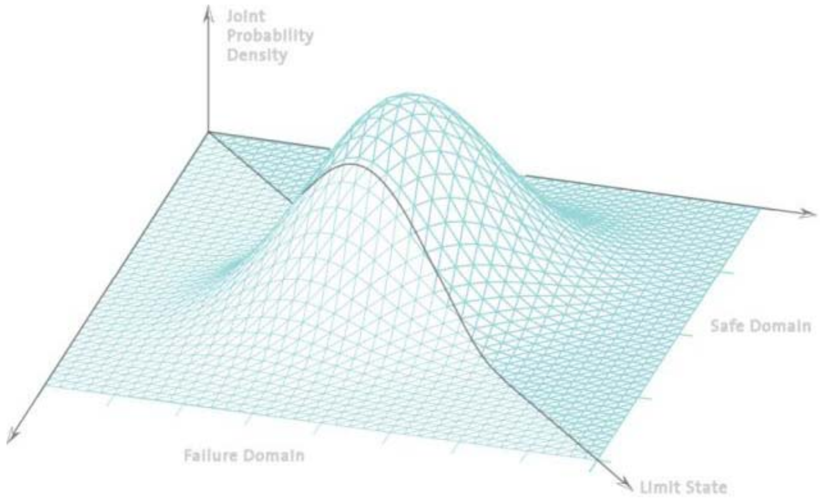
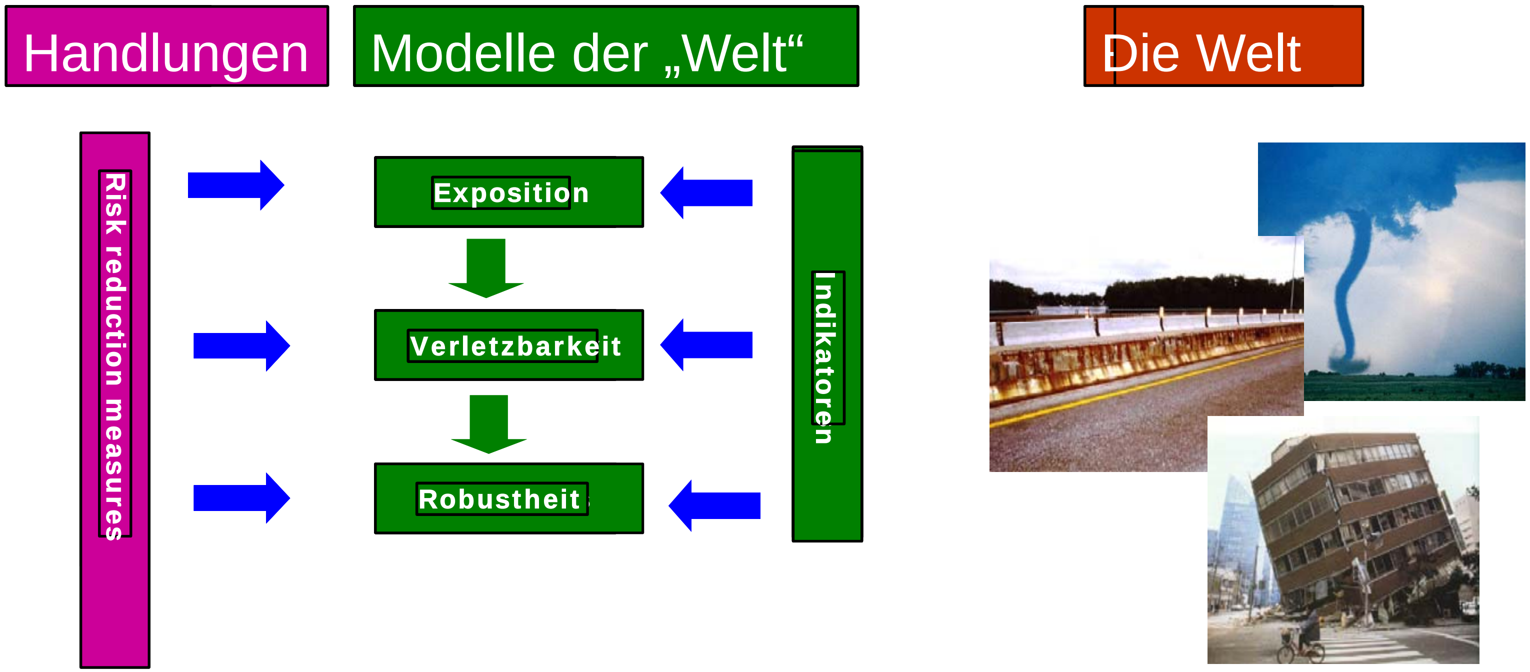


Statistik und Wahrscheinlichkeitsrechnung

Dr. Jochen Köhler
Prof. Michael H. Faber

Modellierung von Unsicherheiten - Praxisbezug

- Zufallsvariablen und deren Charakteristiken



Modellierung von Unsicherheiten - Praxisbezug

- Z.B. Steinschlag

Modellierung von Unsicherheiten - Praxisbezug

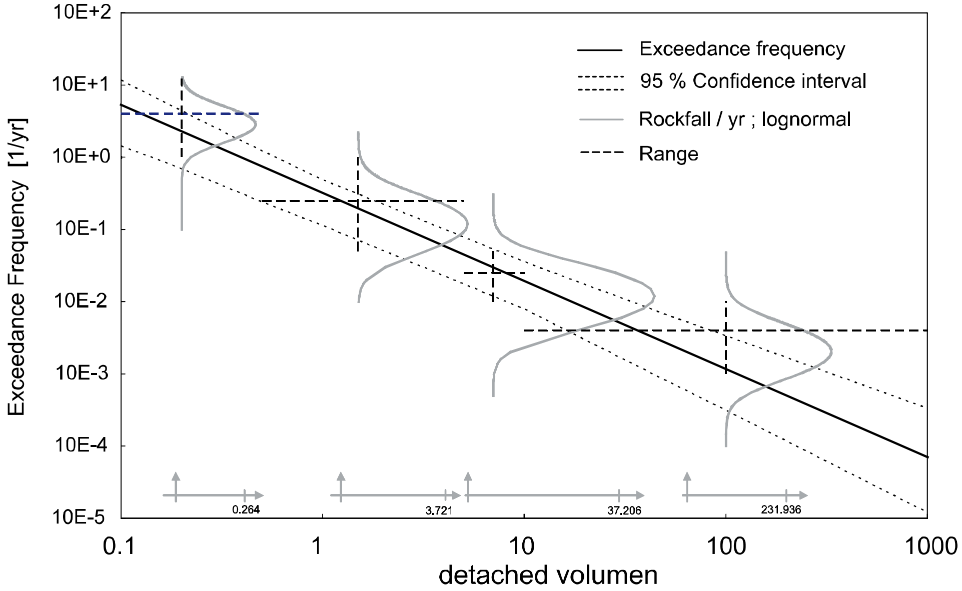
- Z.B. Steinschlag

- Entscheidungen unter Unsicherheiten
  - Sollten wir in eine Steinschlaggalerie investieren?
  - Wenn ja, von wo bis wo und wie dick?
  - Oder doch besser Steinschlagnetze?
  - Oder sollten wir die Strecke umlegen?
  - …

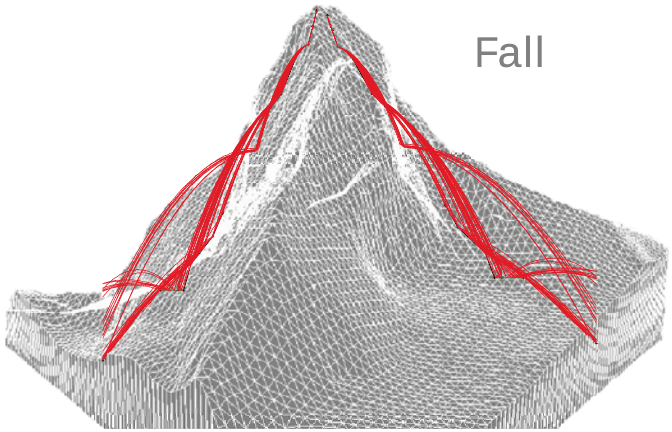
Modellierung von Unsicherheiten - Praxisbezug

- Z.B. Steinschlag

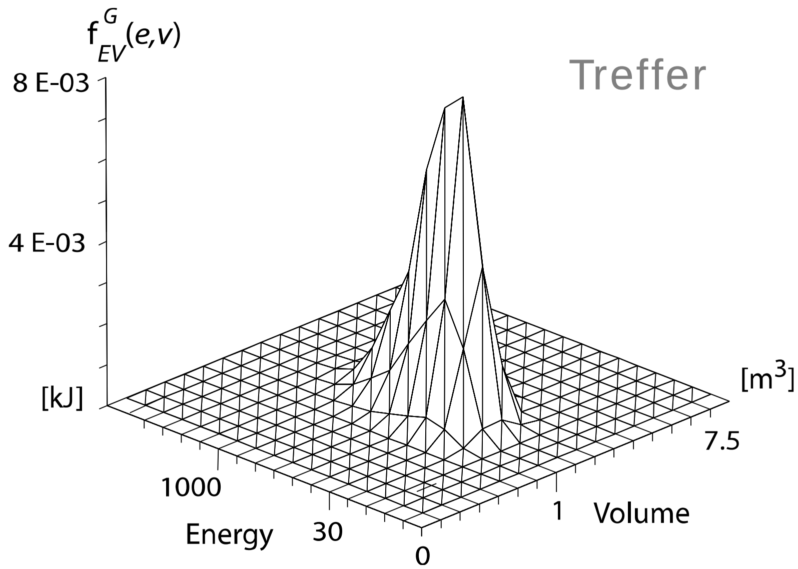
Ablösung



Fall



Treffer



Unsicherheiten bei Problemen im Ingenieurwesen

Ingenieurprobleme sind oftmals sehr spezifisch – bzw. einmalig.

Solche Probleme lösen zu können erfordert:

- Basis ‚Werkzeuge' (Physik, Mathematik, Naturwissenschaften, Ingenieurwissenschaften,…)
- Innovation (in der Lage sein, Lösungswege zu finden)
- Übung

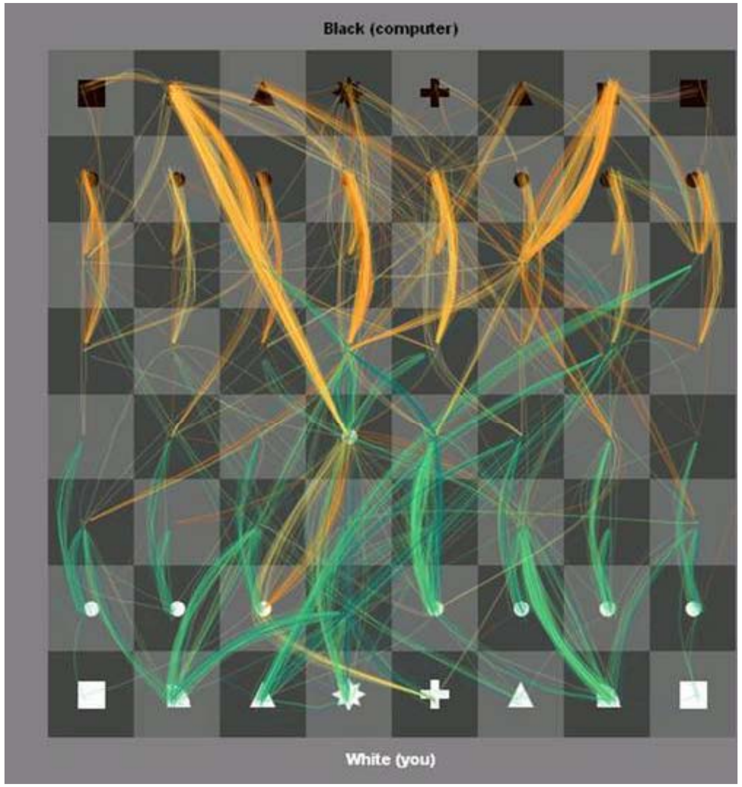
  Übung ist wichtig! – Bedingung für Entwicklung der Mustererkennung

Unsicherheiten bei Problemen im Ingenieurwesen

- Mustererkennung fördert die Identifikation und Einschätzung:

  - der Anwendbarkeit und Angemessenheit der Lösungsstrategien von früheren Problemlösungen

  - des Potentials von bekannten Werkzeugen in einem neuen Kontext



Werkzeuge zur Modellierung von Unsicherheiten

- Erwartungswert

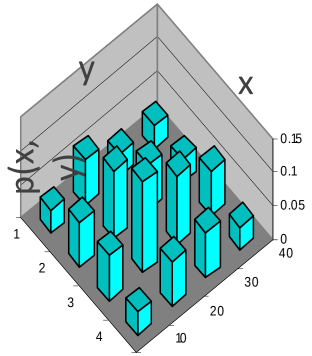
$$E[c] = c$$
$$E[cX] = cE[X]$$
$$E[a + bX] = a + bE[X]$$
$$E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$$

- Varianz

$$\begin{aligned} Var[X] &= E\left[(X-\mu_X)^2\right] \\ &= E\left[X^2 + \mu_X^2 - 2\mu_X X\right] \\ &= \mu_X^2 + E\left[X^2\right] - 2\mu_X E[X] \\ &= \mu_X^2 + E\left[X^2\right] - 2\mu_X^2 = E\left[X^2\right] - \mu_X^2 \end{aligned}$$

- Multivariate Verteilungen

$$F_{\mathbf{X}}(\mathbf{x}) = P\left(X_1 \le x_1 \cap X_2 \le x_2 \cap \ldots \cap X_n \le x_n\right)$$



Werkzeuge zur Modellierung von Unsicherheiten

- Funktionen aus Zufallsvariablen
  - Summe aus Zufallsvariablen:

$$Y = X_1 + X_2$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_2}(y - x_1) f_{X_1}(x_1) dx_1$$

  - Nichtlineare monotone Funktionen von Zufallsvariablen:

$$Y = g(X)$$

$$f_Y(y) = \left| \frac{\partial x}{\partial y} \right| f_X(x)$$

Werkzeuge zur Modellierung von Unsicherheiten

- Funktionen aus Zufallsvariablen
  - Nichtlineare monotone Funktionen von Zufallsvariablen:

$$\mathbf{Y} = (Y_1, Y_2, ..Y_n)^T$$

$$Y_i = g_i(\mathbf{X}), \qquad X_i = f_i(\mathbf{Y})$$

$$f_{\mathbf{Y}}(\mathbf{y}) = |\mathbf{J}|\, f_{\mathbf{X}}(\mathbf{x})$$

$$\mathbf{J} = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \cdots\cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots\cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix}$$

Inhalte der heutigen Vorlesung

- Zufallsvariablen

  - Der zentrale Grenzwertsatz

  - Die Normalverteilung

  - Die Lognormalverteilung

  - Weitere Verteilungen

- Zufallsprozesse

  - Zufällige Folge (Bernoulli trials)

  - Binomial Verteilung

  - Geometrische Verteilung

Zufallsvariablen

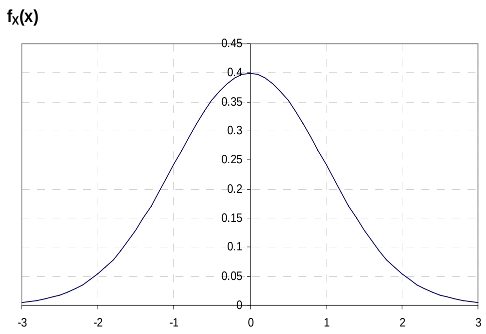
- Der zentrale Grenzwertsatz lautet:

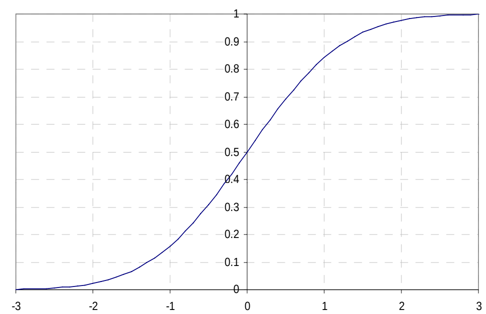
Die Wahrscheinlichkeitsdichteverteilung der Summe von Zufallsvariablen konvergiert zu einer Normalverteilung wenn deren Anzahl gross wird.

$$Y = X_1 + X_2 + \ldots + X_n$$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right)$$

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right) dx$$





Zufallsvariablen

- Bedingungen für den zentralen Grenzwertsatz:

$$Y = X_1 + X_2 + \ldots + X_n$$

  - Die Summe sollte nicht von einer oder wenigen Komponenten dominiert werden.
  - Die statistische Abhängigkeit zwischen den Variablen sollte klein sein.
  - Keine Anforderungen an die Verteilungsfunktion der Komponenten.

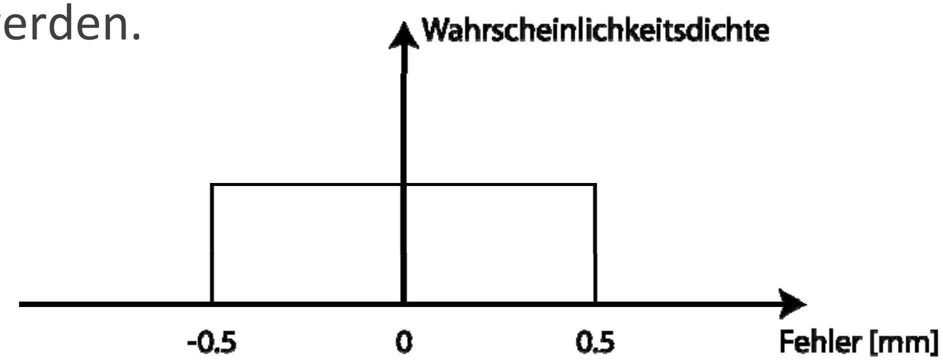
Zufallsvariablen

- Illustration:
- Eine Strecke wird mit einem Zollstock vermessen.

  - Der Zollstock hat eine Länge von 2 m.

  - Die kleinste Einheit auf dem Zollstock ist 1mm.

- Alle Messungen werden gerundet, d.h. die Messungenauigkeit jeder einzelnen Messung kann mit einer gleichverteilten Zufallsvariablen modelliert werden.



Zufallsvariablen

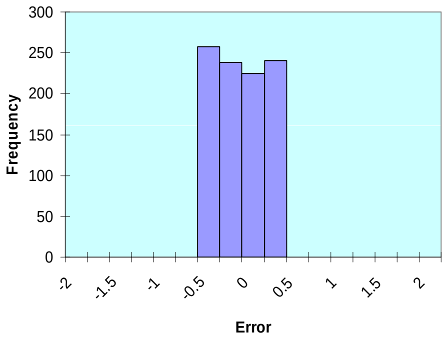
- Illustration:
  - Wir betrachten nun den akkumulierten Fehler über die gemessene Länge:

    - bis 2 m (eine Messung)
    - zwischen 2 und 4 m (zwei Messungen)
    - zwischen 6 und 8 m (vier Messungen)
    - zwischen 14 und 16 m (acht Messungen)

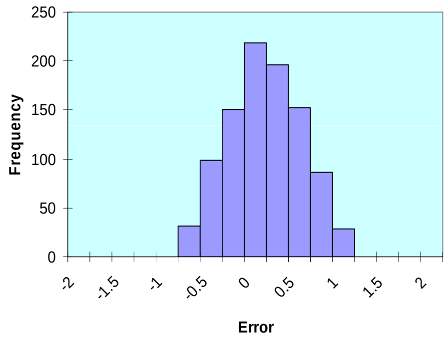
Zufallsvariablen

- Illustration:

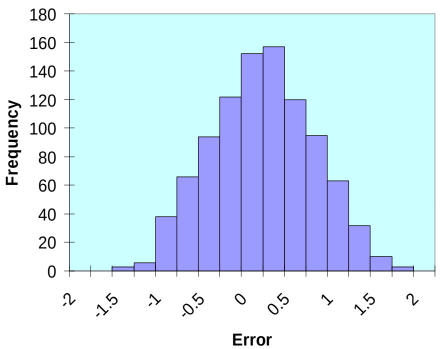
N = 1



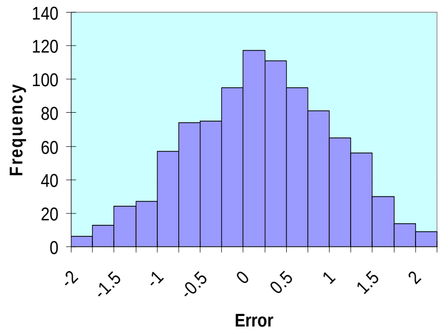
N = 2



N = 4



N = 8



Zufallsvariablen

- Normalverteilung

Analytisch kann die Normalverteilung durch wiederholte Anwendung der Rechenregel für die Dichte der Summe von zwei Zufallsvariablen hergeleitet werden.

Die Normalverteilung wird im Ingenieurwesen oft verwendet wenn die Zufallsgrösse als Summe von mehreren zufälligen Einflüssen angesehen werden kann: $X_i, i = 1, 2, .., n$

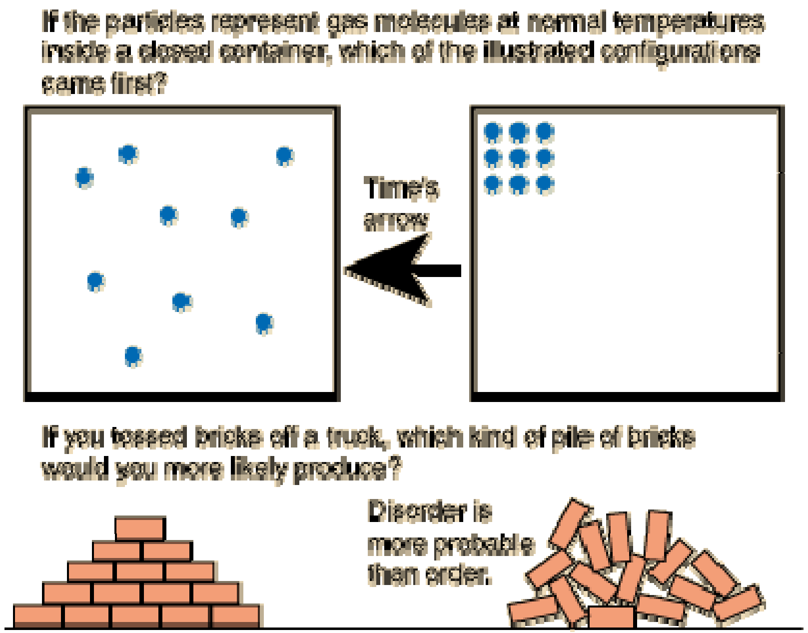
Eine Linearkombination von Normalverteilten Zufallsvariablen ist deswegen auch Normalverteilt: $S = a_0 + \sum_{i=1}^{n} a_i X_i$

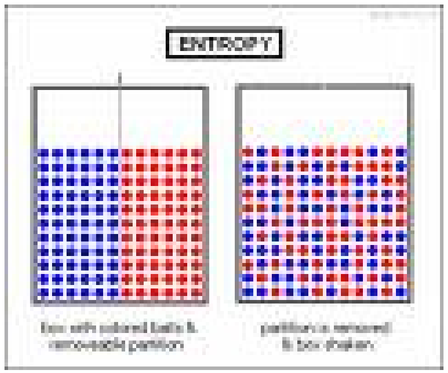
Zufallsvariablen

- Normalverteilung

  Die Normalverteilung resultiert auch aus anderen Überlegungen.

  Die Verteilung von Energie in einem isolierten System

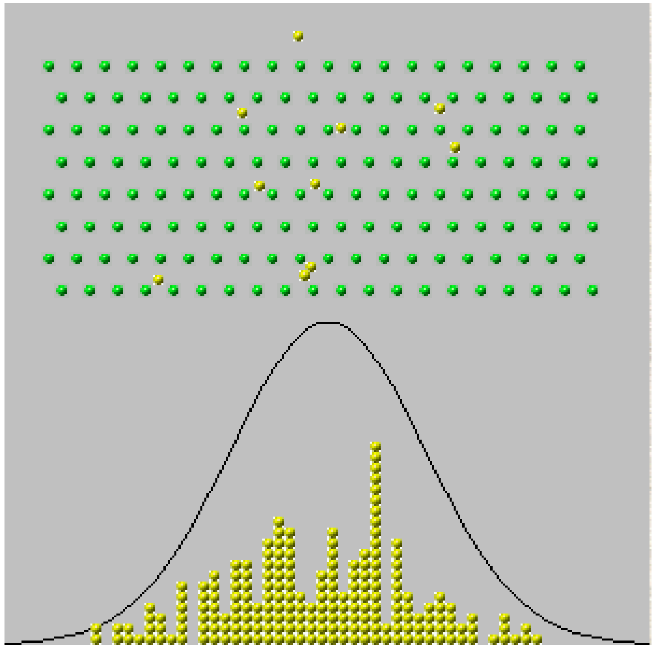




Zufallsvariablen

- Normalverteilung

  Die Akkumulation von zufälligen Bewegungen

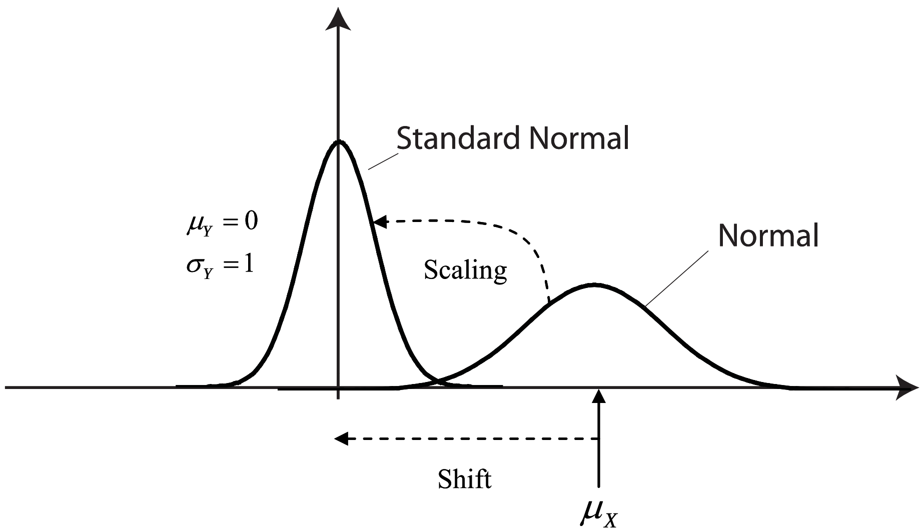


Zufallsvariablen

- Normalverteilung

Wenn der Mittelwert der Normalverteilung gleich null ist und die Standardabweichung gleich 1 nennt man das *standardisierte Normalverteilung*.

$$Y = \frac{X - \mu_X}{\sigma_X}$$ Standardisierte Normalverteilung

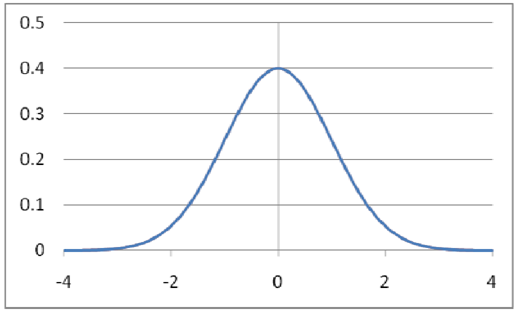


Zufallsvariablen

- Normalverteilung

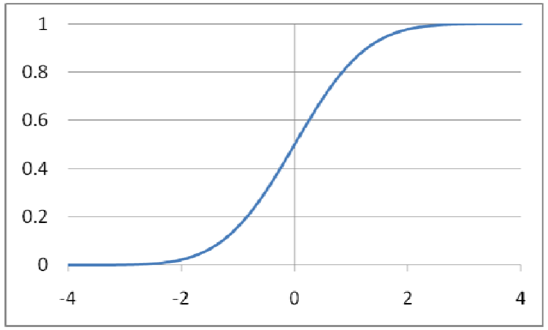
Wenn der Mittelwert der Normalverteilung gleich null ist und die Standardabweichung gleich 1 nennt man das *standardisierte Normalverteilung*.

$$Y = \frac{X - \mu_X}{\sigma_X}$$ Standardisierte Normalverteilung



Standard normal

$$\begin{cases} f_Y(y) = \varphi(y) = \dfrac{1}{\sqrt{2\pi}} \exp\left(-\dfrac{1}{2} y^2\right) \\ F_Y(y) = \Phi(y) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{y} \exp\left(-\dfrac{1}{2} x^2\right) dx \end{cases}$$



Zufallsvariablen

- Log-Normalverteilung

  Wenn der Logarhitmus der Zufallsvariablen X, d.h.

  $$Y = \ln(X), \qquad Y : N(\mu_Y, \sigma_Y)$$

  Normalverteilt ist, ist die Variable X Log-Normalverteilt $\quad X : \ln(\lambda, \zeta)$

$$f_X(x) = \frac{1}{x\zeta\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln(x)-\lambda}{\zeta}\right)^2\right)$$

$$F_X(x) = \Phi\left(\frac{\ln(x)-\lambda}{\zeta}\right)$$

$$\mu_X = \exp\left(\lambda + \frac{\zeta^2}{2}\right)$$

$$\sigma_X = \exp\left(\lambda + \frac{\zeta^2}{2}\right)\sqrt{\exp(\zeta^2)-1}$$

Zufallsvariablen

Während die Normalverteilung aus der Summe von vielen Zufallsvariablen folgt – zentraler Grenzwertsatz

ergibt sich die Log-Normalverteilung aus dem Produkt von vielen Zufallsvariablen

$$\ln(X_1 \cdot X_2 \cdots X_n) = \ln\left(\prod_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \ln(X_i)$$

Zufallsvariablen

Die Log-Normalverteilung hat die nützlichen Eigenschaft dass wenn:

$$P = \prod_{i=1}^{n} Y_i^{a_i}$$

wo Y_i eine unabhängige Log-Normalverteilte Zufallsvariable mit Parametern ζ_i , λ_i und $\varepsilon_i = 0$ ist,
dann ist auch P Log-Normalverteilt mit Parametern:

$$\lambda_P = \sum_{i=1}^{n} a_i \lambda_i \qquad \zeta_P^2 = \sum_{i=1}^{n} a_i^2 \zeta_i^2 \qquad f_P(p) = \frac{1}{p\zeta_P\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln(p)-\lambda_P}{\zeta_P}\right)^2\right)$$

Zufallsvariablen

Die Log-Normalverteilung wird oft angewendet für die Modellierung von:

- Unsicheren Parametern, welche nicht negative Realisationen annehmen können

- Ermüdungslebensdauer

- Festigkeit von Stahl und Beton

- Täglichen Niederschlagsmengen

- Falls ein unsicherer Einflussparameter aus einem Produkt von Zufallsvariablen folgt

Zufallsvariablen

Beton Druckfestigkeit

Wahrscheinlichkeit eines Ereignisses kleiner als 25 MPa

$$\mu_X = \exp\left(\lambda + \frac{\zeta^2}{2}\right)$$

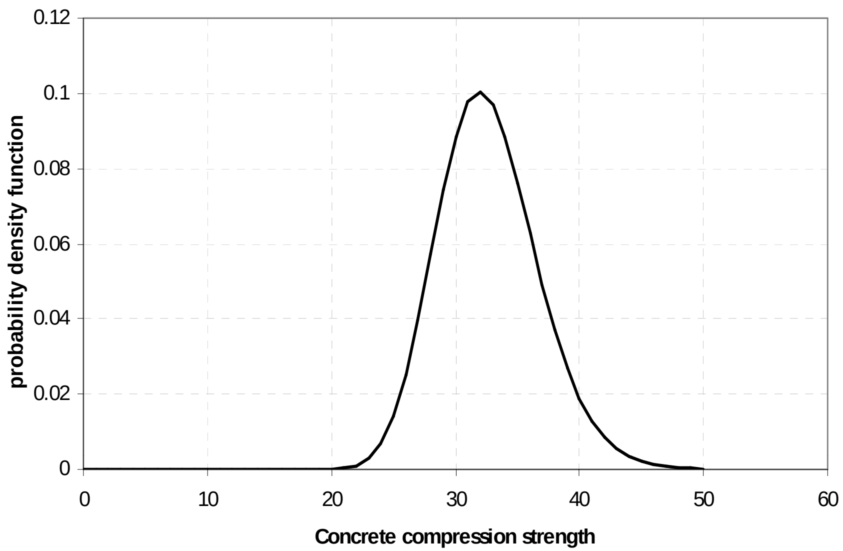
$$\sigma_X = \exp\left(\lambda + \frac{\zeta^2}{2}\right)\sqrt{\exp(\zeta^2)-1}$$

$$\Downarrow$$

$$V_X = \frac{\sigma_X}{\mu_X} = \sqrt{\exp(\zeta^2)-1} = \frac{4.05}{32.67} = 0.12 \Rightarrow \zeta = 0.12, \lambda = 3.48$$

$$F_X(25) = \Phi\left(\frac{\ln(25)-3.48}{0.12}\right) = 0.018$$

i	x_i
1	24.4
2	27.6
3	27.8
4	27.9
5	28.5
6	30.1
7	30.3
8	31.7
9	32.2
10	32.8
11	33.3
12	33.5
13	34.1
14	34.6
15	35.8
16	35.9
17	36.8
18	37.1
19	39.2
20	39.7



Zufallsvariablen

Es besteht eine grosse Anzahl von verschiedenen Wahrscheinlichkeitsverteilungen:

Gleichverteilung
Normalverteilung
Log-Normalverteilung
Exponentialverteilung
Betaverteilung
Gammaverteilung
…

…

Distribution type	Parameters	Moments
Uniform, $a \le x \le b$ $f_X(x) = \frac{1}{b-a}$ $F_X(x) = \frac{x-a}{b-a}$	a b	$\mu = \frac{a+b}{2}$ $\sigma = \frac{b-a}{\sqrt{12}}$
Normal $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$ $F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right) dt$	μ $\sigma > 0$	μ σ
Shifted Lognormal, $x > \varepsilon$ $f_X(x) = \frac{1}{(x-\varepsilon)\zeta\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln(x-\varepsilon)-\lambda}{\zeta}\right)^2\right)$ $F_X(x) = \Phi\left(\frac{\ln(x-\varepsilon)-\lambda}{\zeta}\right)$	λ $\zeta > 0$ ε	$\mu = \varepsilon + \exp\left(\lambda + \frac{\zeta^2}{2}\right)$ $\sigma = \exp\left(\lambda + \frac{\zeta^2}{2}\right)\sqrt{\exp(\zeta^2)-1}$
Shifted Exponential, $x \ge \varepsilon$ $f_X(x) = \lambda \exp(-\lambda(x-\varepsilon))$ $F_X(x) = 1 - \exp\left(-\lambda(x-\varepsilon)\right)$	ε $\lambda > 0$	$\mu = \varepsilon + \frac{1}{\lambda}$ $\sigma = \frac{1}{\lambda}$
Gamma, $x \ge 0$ $f_X(x) = \frac{b^p}{\Gamma(p)} \exp(-bx) x^{p-1}$ $F_X(x) = \frac{\Gamma(bx, p)}{\Gamma(p)}$	$p > 0$ $b > 0$	$\mu = \frac{p}{b}$ $\sigma = \frac{\sqrt{p}}{b}$

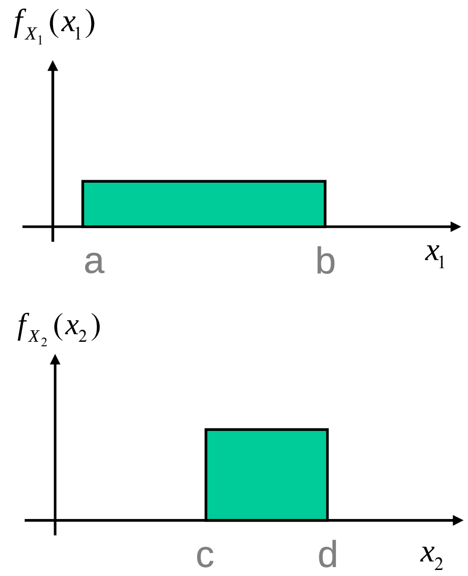
Kleines Beispiel 1

Wir erinnern uns an das Faltungsintegral das wir angewendet haben um die Wahrscheinlichdichtefunktion der Summe von zwei Zufallsvariablen zu bestimmen:

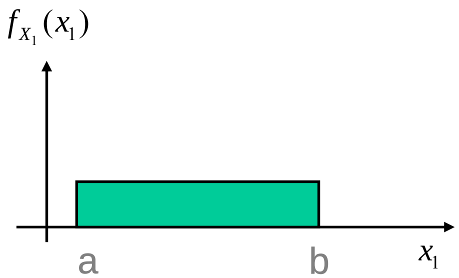
$$Y = X_1 + X_2$$

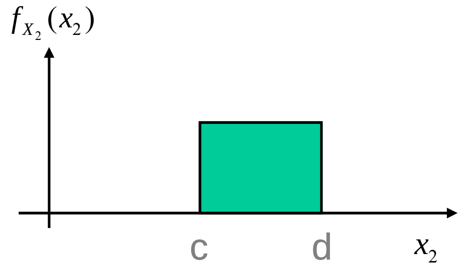
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_2}(y - x_1) f_{X_1}(x_1) dx_1$$

Lass uns sehen wie dies auszuführen ist für zwei gleichverteilte Zufallsvariablen:



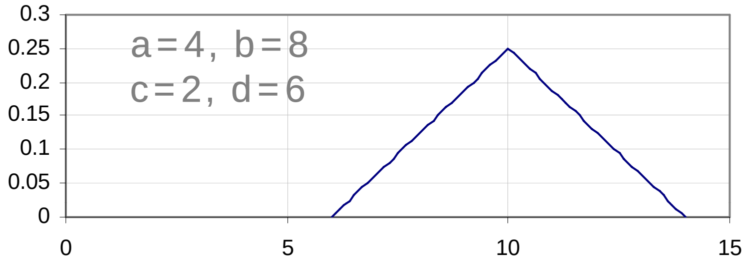
Kleines Beispiel 1

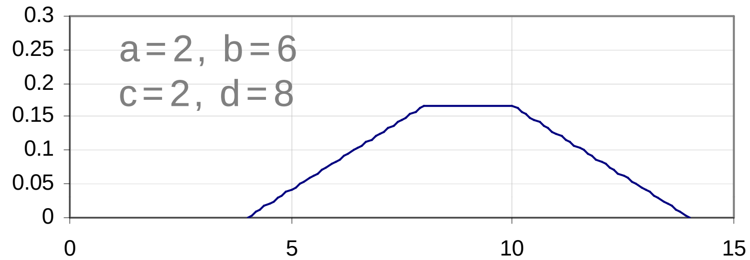




Wenn die zwei Zufallsvariablen voneinander unabhängig sind, dann können wir das Faltungsintegral wie folgt schreiben:

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{X_2}(y-x_1)\, f_{X_1}(x_1)\, dx_1 \\
&= \frac{1}{(b-a)(d-c)} \int_a^b \Uparrow (y - x_1 \in [c;d])\, dx_1 \\
&= \frac{1}{(b-a)(d-c)} \left[x_1\right]_{\max(c,y-a)}^{\min(d,y-b)}, \quad a+c \le y \le b+d
\end{aligned}
$$





Stochastische Prozesse und Extreme

Zufallsgrössen können zeitabhängig sein im Sinne dass sie im Laufe der Zeit neue Realisationen annehmen können.

- Falls neue Realisationen zu diskreten Zeiten vorkommen und diskrete Werte annehmen nennt man das Phänomen eine Zufallsreihe.

  Versagensereignissen, Verkehrsstörungen, …

- Falls neue Realisationen kontinuierlich über die Zeit vorkommen und kontinuierliche Werte annehmen nennt man das Phänomen einen Zufallsprozess oder einen Stochastischen Prozess.

  Windgeschwindigkeit, höhe von Wellen,…

Stochastische Prozesse und Extreme

- Zufallsreihen

  - Eine Reihe von Experimenten mit nur zwei möglichen und exklusiven Ereignissen nennt man einen Bernoulli Versuch.
  - Typischerweise nennt man die Ereignisse eines Bernoulliversuches Erfolg und Versagen.

  Wenn die Wahrscheinlichkeit *p* eines Erfolgs in einem Versuch konstant ist, dann kann die Wahrscheinlichkeitsdichte von *y* Erfolgen in *n* Versuchen d.h. $p_Y(y)$ bestimmt werden durch:

$$p_Y(y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0,1,2 \ldots n \qquad \binom{n}{y} = \frac{n!}{y!(n-y)!}$$

Binomiale Wahrscheinlichkeitsdichte

Binomiale Funktion

Stochastische Prozesse und Extreme

- Zufallsreihen

  - Eine Reihe von Experimenten mit nur zwei möglichen und exklusiven Ereignissen nennt man einen Bernoulli Versuch.

Die Binomiale Wahrscheinlichkeitsverteilungsfunktion folgt als:

$$P_Y(y) = \sum_{i=0}^{y} \binom{y}{i} p^i (1-p)^{n-i}, \qquad y = 0,1,2,\ldots n$$

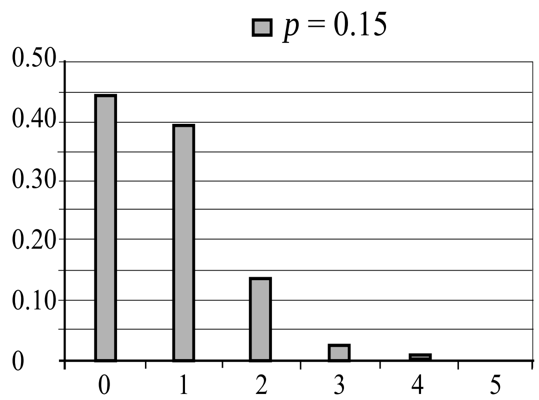
Stochastische Prozesse und Extreme

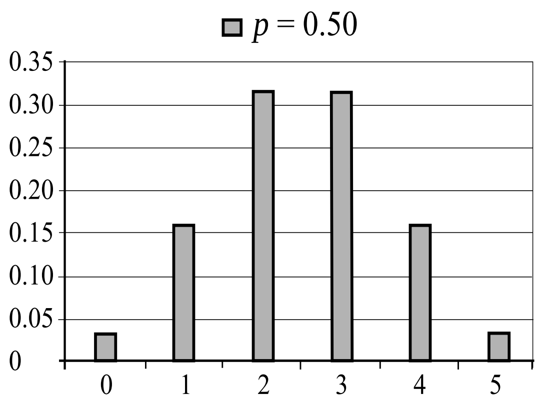
- Zufallsreihen

  - Eine Reihe von Experimenten mit nur zwei möglichen und exklusiven Ereignissen nennt man einen Bernoulli Versuch.

- Illustration:

  Binomiale Wahrscheinlichkeitsdichtefunktion mit n=5 und p=0.15, p=0.5





Kleines Beispiel 2

Wir erinnern uns, dass wir die Wahrscheinlichkeitsdichtefunktion einer Funktion einer Zufallsvariable

$$Y = g(X)$$

bestimmen können durch:

$$f_Y(y) = \left| \frac{\partial x}{\partial y} \right| f_X(x)$$

Kleines Beispiel 2

Lasst uns sehen wie dies auszuführen ist:

$$Y = X^2$$

$$\Downarrow$$

$$X = \sqrt{Y}$$

$$f_Y(y) = \left|\frac{\partial x}{\partial y}\right| f_X(x)$$

$$\boxed{\frac{\partial x}{\partial y} = \frac{\partial \sqrt{y}}{\partial y} = \frac{1}{2} y^{-\frac{1}{2}}}$$

$$\boxed{f_Y(y) = \left|\frac{1}{2} y^{-\frac{1}{2}}\right| f_X(\sqrt{y})}$$

Stochastische Prozesse und Extreme

- Zufallsreihen

Der Erwartungswert und die Varianz einer Binomialverteilten Zufallsvariable Y ist gegeben durch:

$$E[Y] = np$$

$$Var[Y] = np(1-p)$$

Stochastische Prozesse und Extreme

- Zufallsreihen

Die Wahrscheinlichkeitsdichtefunktion für die Anzahl von (unabhängigen) Versuchen bis zum ersten Erfolg ist gegeben durch:

$$p_N(n) = p(1-p)^{n-1}$$

Geometrische Wahrscheinlichkeitsdichtefunktion

und die Wahrscheinlichkeitsverteilungsfunktion ist dann:

$$P_N(n) = \sum_{i=1}^{n} p(1-p)^{i-1} = 1-(1-p)^n$$

Geometrische Wahrscheinlichkeitsverteilungsfunktion

Kleines Beispiel 3

Wir erinnern uns, dass die Wahrscheinlichkeitsdichtefunktion von Funktionen ***Y*** von Zufallsvariablen **X** durch die Jakobi Determinante gegeben ist:

$$\mathbf{Y} = (Y_1, Y_2, ..Y_n)^T \qquad Y_i = g_i(\mathbf{X})$$

$$\mathbf{X} = (X_1, X_2, ..X_n)^T \qquad X_i = f_i(\mathbf{Y})$$

$$f_{\mathbf{Y}}(\mathbf{y}) = |\mathbf{J}|\, f_{\mathbf{X}}(\mathbf{x}) \qquad \mathbf{J} = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & ..... & \frac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & ..... & \frac{\partial x_n}{\partial y_n} \end{bmatrix}$$

Kleines Beispiel 3

Lasst uns sehen wie dies auf folgendes Problem anzuwenden ist:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \cdots\cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots\cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix}$$

$$Y_1 = X_1 + X_2 \qquad X_1 = Y_1 - Y_2$$

$$Y_2 = X_2 \qquad X_2 = Y_2$$

$$\mathbf{J} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \qquad \det(\mathbf{J}) = 1 \times 1 - 0 \times 1 = 1 \Rightarrow |\mathbf{J}| = 1$$

$$\boxed{f_{\mathbf{Y}}(y_1, y_2) = f_{\mathbf{X}}(y_1 - y_2, y_2)}$$

Stochastische Prozesse und Extreme

Der Median der Geometrischen Verteilungsfunktion beinhaltet Informationen darüber, wie lange (wie viele Wiederholungen) ein Spiel mit Erfolgswahrscheinlichkeit *p* (pro Zeiteinheit) gespielt werden muss um eine faire Gewinnchance zu haben:

Zeiteinheiten können folgende sein:

- Würfe (Würfeln)
- Jahre (Erdbeben)

Der Median ist definiert durch $$P_N(n) = 0.5 = 1 - (1 - p)^n$$

Wir brauchen nur *n* zu bestimmen als Funktion von *p*.

Stochastische Prozesse und Extreme

Der Median der Geometrischen Verteilungsfunktion beinhaltet Informationen darüber, wie lange (wie viele Wiederholungen) ein Spiel mit Erfolgswahrscheinlichkeit *p* (pro Zeiteinheit) gespielt werden muss um eine faire Gewinnchance zu haben:

$$P_N(n) = 0.5 = 1-(1-p)^n$$

Wir logarithmieren auf beiden Seiten

$$\ln(0.5) = n\ln(1-p)$$

$$\Downarrow$$

$$0.7 \approx -n\ln(1-p)$$

Und wenden folgende Umschreibung an:

$$\ln(1-p) = -p + \frac{1}{2}p^2 - \frac{1}{3}p^3 + \ldots = \sum_{k=1}^{\infty}(-1)^k \frac{p^k}{k}$$

$$\Downarrow$$

$$\ln(1-p) \approx -p \quad \text{for small } p$$

$$\boxed{0.7 \approx np \Rightarrow n = \frac{0.7}{p}}$$

Stochastische Prozesse und Extreme

Anwendungen des Resultats

Um eine 50% Chance zu haben, einen 6‘er zu würfeln, werden *n* Würfe benötigt:

n = 0.7 x 6 = 4 Würfe

50% Chance für zwei 6‘er (mit zwei Würfeln) benötigt:

n = 0.7 x 36 = 25 Würfe

50% Chance, ein Erdbeben mit jährlicher Auftretenswahrscheinlichkeit von 0.001 zu erleben, benötigt *n* Jahre:

n = 0.7 x 1000 = 700

Stochastische Prozesse und Extreme

- Zufallsreihen

Der Erwartungswert und die Varianz einer Geometrisch verteilten Zufallsvariable sind gegeben durch:

$$E[N] = \frac{1}{p}$$

$$Var[N] = \frac{1-p}{p^2}$$

z.B. falls *p* die jährliche Wahrscheinlichkeit eines extremen Erdbebens ist, ist die Wiederkehrperiode gegeben durch den Erwartungswert *E[N]*