

Statistik und Wahrscheinlichkeitsrechnung

6. Vorlesung

Dr. Jochen Köhler

1. Teilprüfung am Donnerstag 1. April

Wann?

- Donnerstag, 1. April, 8:00 Uhr – 9:30 Uhr
- Bitte um 7:45 Uhr am Hörsaal sein!

Wo?

- Studierende mit Nachnamen A – J : HIL E 4
- Studierende mit Nachnamen K – Z : HCI G 3

1. Teilprüfung am Donnerstag 1. April

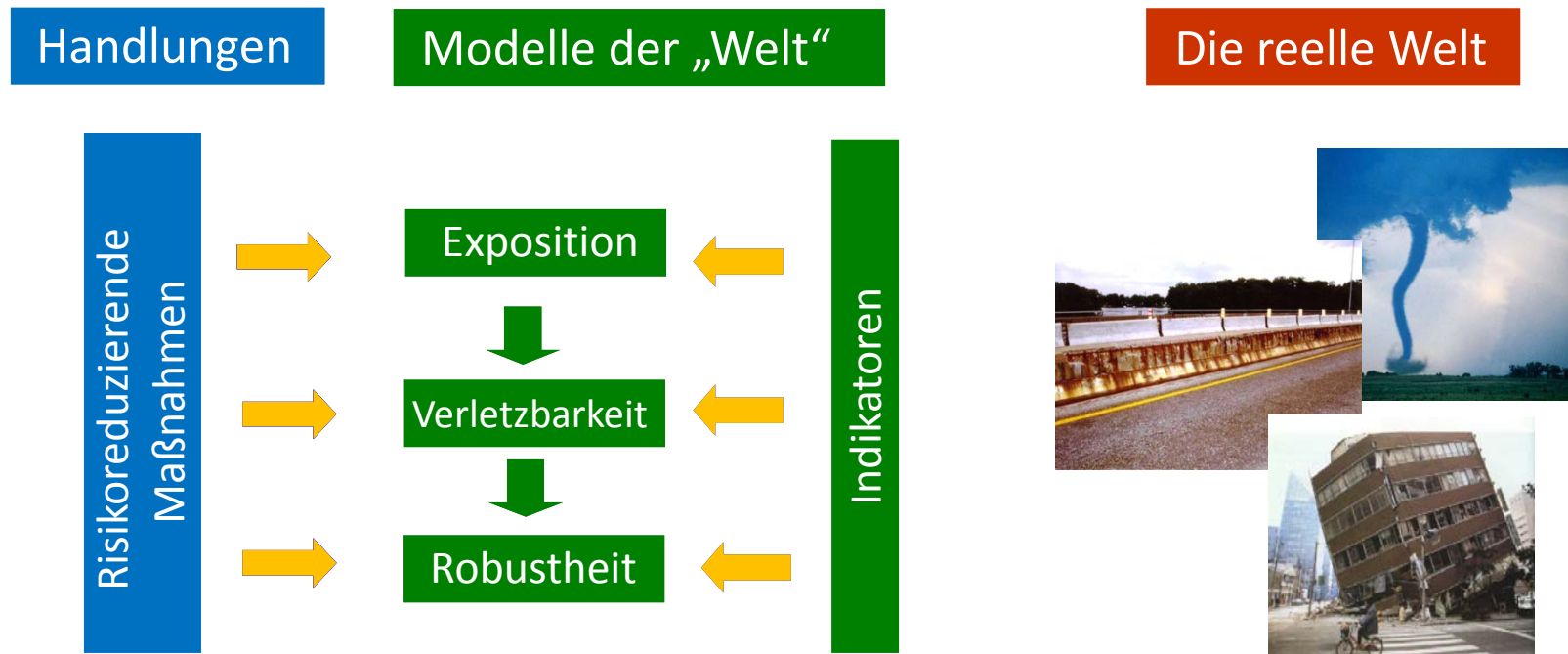
- **Inhalt**
- Multiple Choice + 1 Übung zum Rechnen
- Ganzer Stoff der Vorlesungen 1 bis 5, Übungen 1 bis 5.

Erlaubte Hilfsmittel

- Alle Unterlagen erlaubt.
- Taschenrechner (ohne Kommunikationsmittel) erlaubt, auch programmierbare.
- Keine Kommunikationsmittel (z.B. Telefon) erlaubt.

Modellierung von Unsicherheiten - Praxisbezug

- Zufallsvariablen und deren Charakteristiken



Modellierung von Unsicherheiten - Praxisbezug

- Z.B. Steinschlag

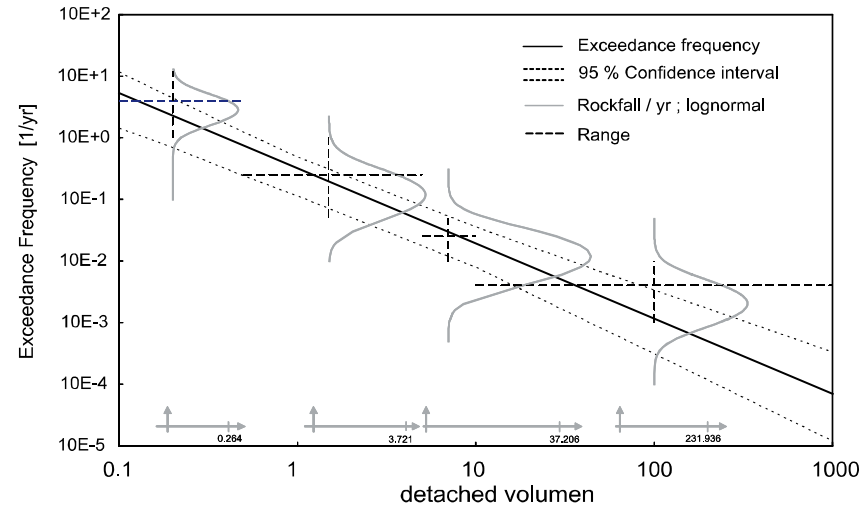


Modellierung von Unsicherheiten - Praxisbezug

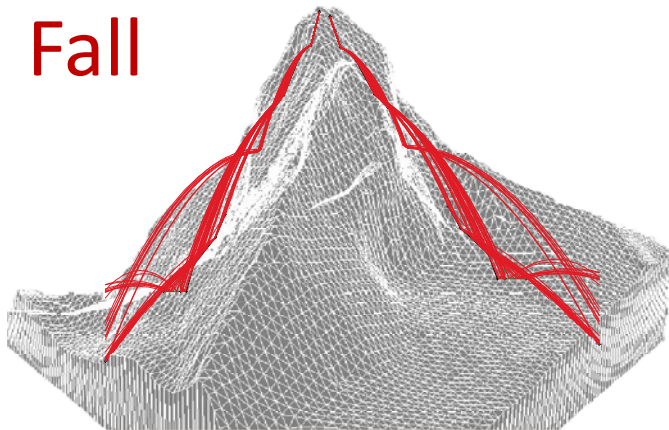
- Z.B. Steinschlag
- Entscheidungen unter Unsicherheiten
 - Sollten wir in eine Steinschlaggalerie investieren?
 - Wenn ja, von wo bis wo und wie dick?
 - Oder doch besser Steinschlagnetze?
 - Oder sollten wir die Strecke umlegen?
 - ...

Modellierung von Unsicherheiten - Praxisbezug

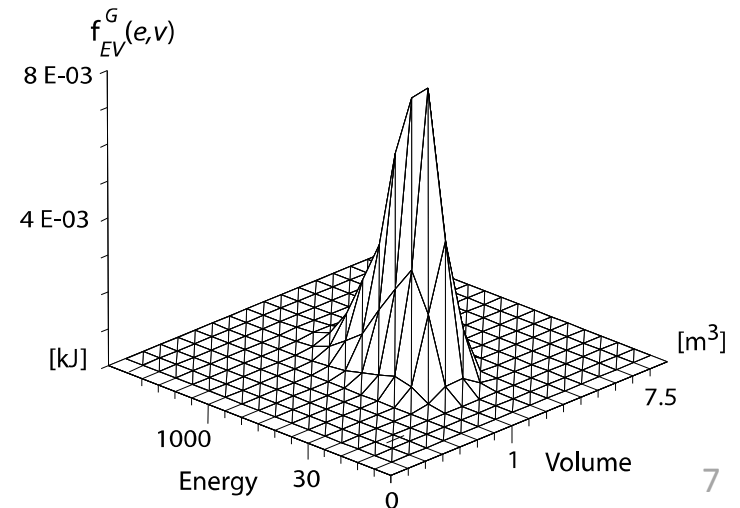
Ablösung



Fall



Treffer



Unsicherheiten bei Problemen im Ingenieurwesen

Problemstellungen im Ingenieurwesen sind oftmals sehr spezifisch bzw. einmalig.

Das Lösen dieser Probleme erfordert:

- grundlegende ‚Werkzeuge‘ (Physik, Mathematik, Naturwissenschaften, Ingenieurwissenschaften,...)
- Innovation (in der Lage sein, Lösungswege zu finden)
- Übung

Übung ist wichtig! – Bedingung für Entwicklung der Mustererkennung

Werkzeuge zur Modellierung von Unsicherheiten

- Erwartungswertoperator:

$$E[c] = c$$

$$E[cX] = cE[X]$$

$$E[a + bX] = a + bE[X]$$

$$E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$$

- Varianzoperator:

$$\text{Var}[X] = E[(X - \mu_X)^2]$$

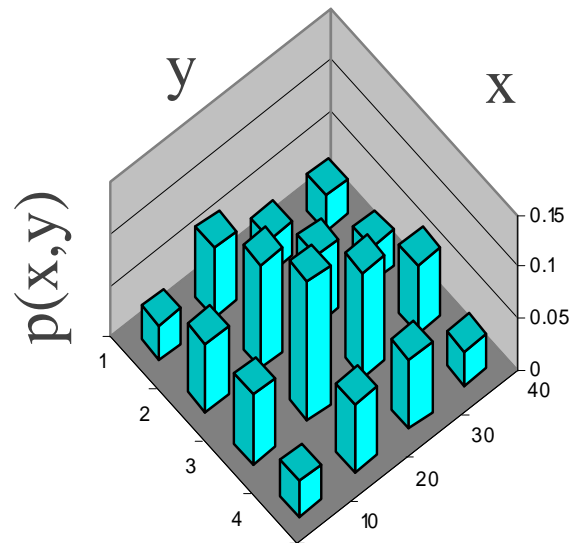
$$= E[X^2 + \mu_X^2 - 2\mu_X X]$$

$$= \mu_X^2 + E[X^2] - 2\mu_X E[X]$$

$$= \mu_X^2 + E[X^2] - 2\mu_X^2 = E[X^2] - \mu_X^2$$

Werkzeuge zur Modellierung von Unsicherheiten

- Multivariate Verteilungen:



$$F_{\mathbf{X}}(\mathbf{x}) = P(X_1 \leq x_1 \cap X_2 \leq x_2 \cap \dots \cap X_n \leq x_n)$$

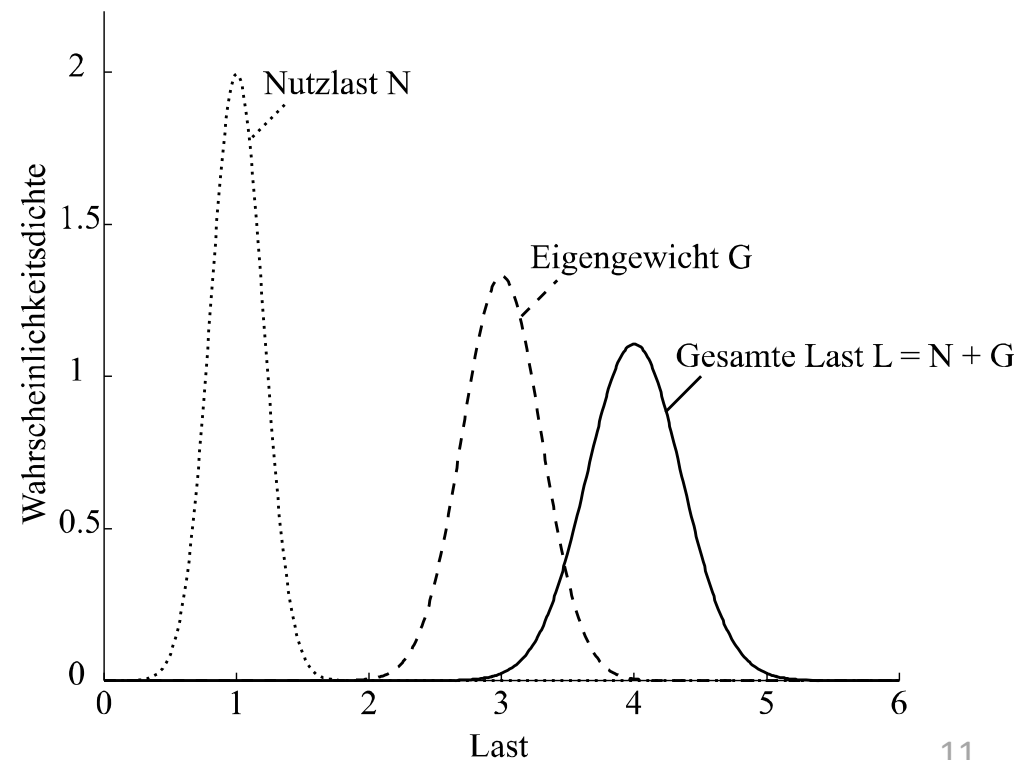
Werkzeuge zur Modellierung von Unsicherheiten

- Funktionen aus Zufallsvariablen
 - Summe aus Zufallsvariablen:

$$Y = X_1 + X_2$$

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

Faltungsintegral



Werkzeuge zur Modellierung von Unsicherheiten

- Funktionen aus Zufallsvariablen
 - Summe aus Zufallsvariablen:

$$Y = X_1 + X_2$$

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

- Nichtlineare monotone Funktionen von Zufallsvariablen:

$$Y = g(X) \quad f_Y(y) = \left| \frac{dx}{dy} \right| f_X(x)$$

Werkzeuge zur Modellierung von Unsicherheiten

- Funktionen aus Zufallsvariablen
 - Nichtlineare monotone Funktionen von Zufallsvariablen:

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

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

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

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

Inhalte der heutigen Vorlesung

- Zufallsvariablen
 - Der zentrale Grenzwertsatz
 - Die Normalverteilung
 - Die Lognormalverteilung
 - Weitere Verteilungsfamilien

- Zufallsprozesse
 - Zufällige Folge (Bernoulli trials)
 - Binomialverteilung
 - Geometrische Verteilung

Zufallsvariablen

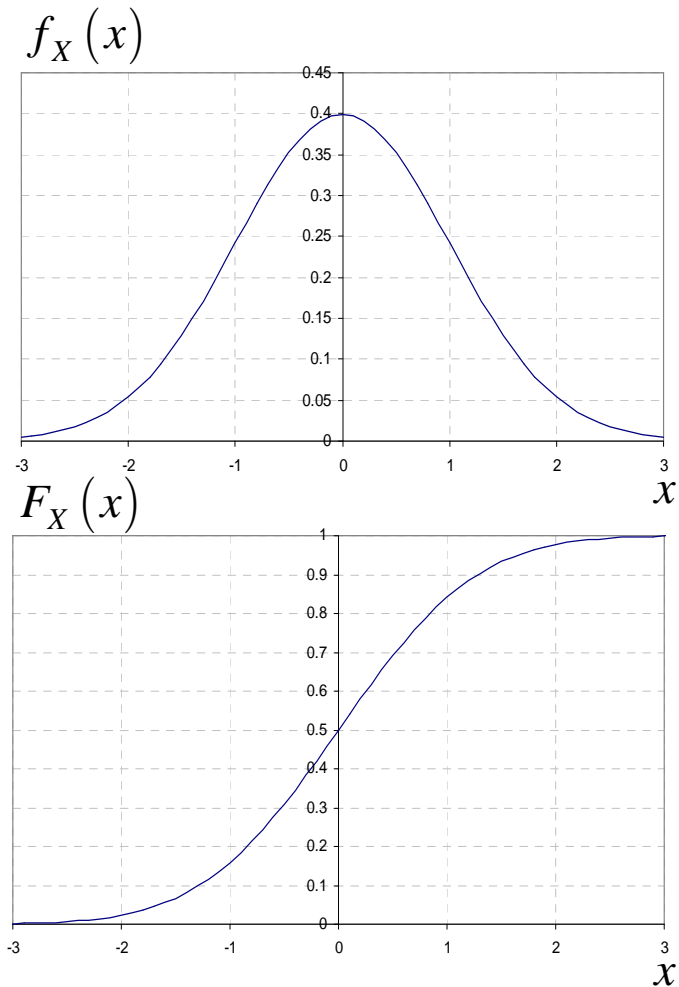
- Der zentrale Grenzwertsatz lautet:

Die Verteilung einer Summe von Zufallsvariablen nähert sich der **NORMALVERTEILUNG**, wenn die Anzahl der Summanden gross wird.

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

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

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



Zufallsvariablen

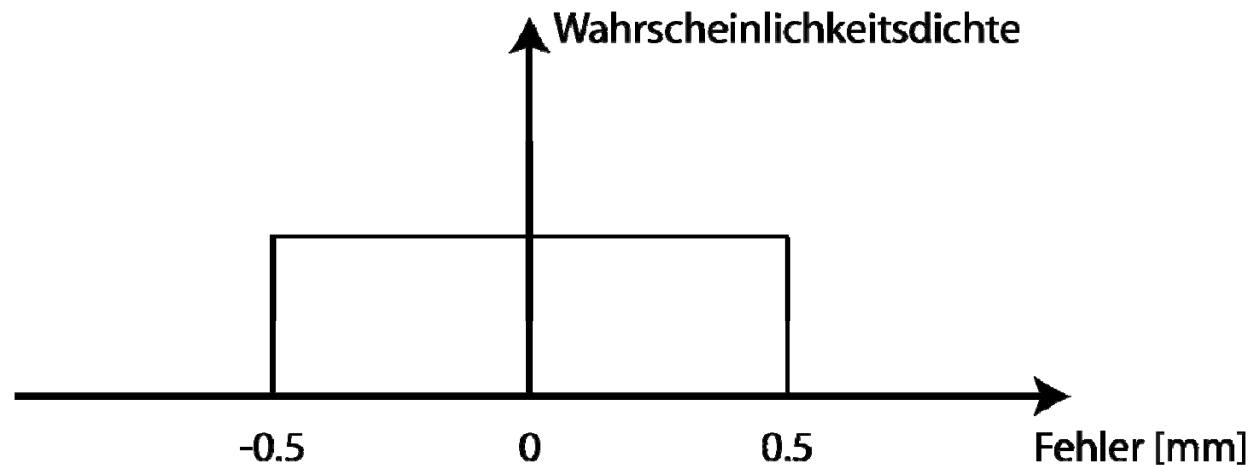
- Bedingungen für den zentralen Grenzwertsatz:

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

1. Die Summe sollte nicht von einer oder wenigen Komponenten dominiert werden.
2. Die Summanden dürfen nicht oder nur schwach voneinander abhängig sein.
3. Es gibt keine Bedingungen betreffend der Verteilungstypen der Summanden.

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.

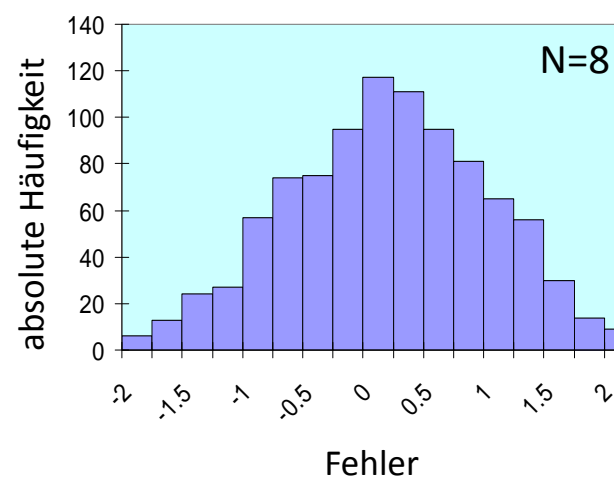
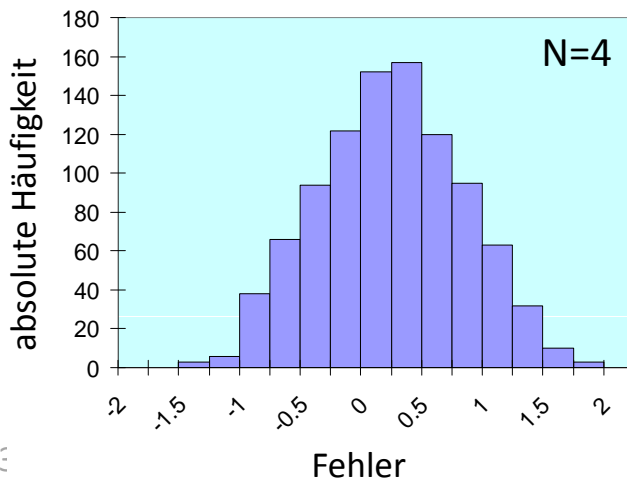
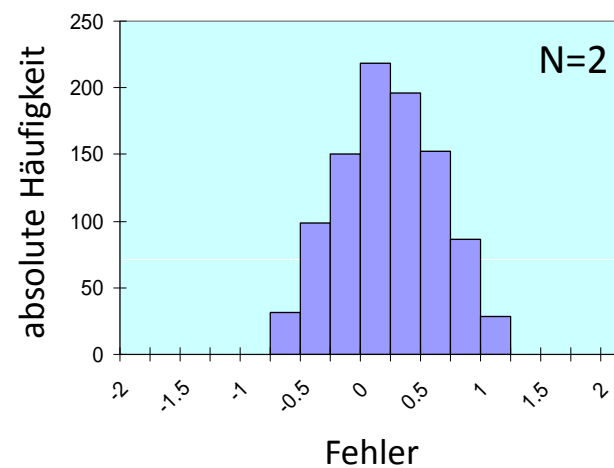
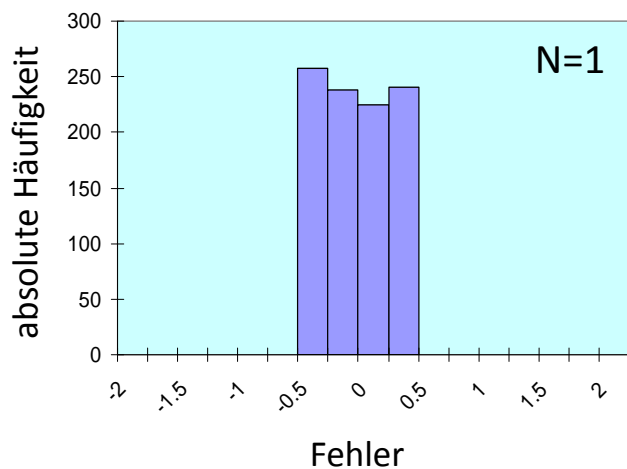


Zufallsvariablen

- 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:



Zufallsvariablen

- Normalverteilung

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

Die Normalverteilung wird im Ingenieurwesen oft verwendet, wenn die Zufallsgrösse als Summe von mehreren zufälligen Einflüssen angesehen werden kann:

$$X_i, i = 1, 2, \dots, n$$

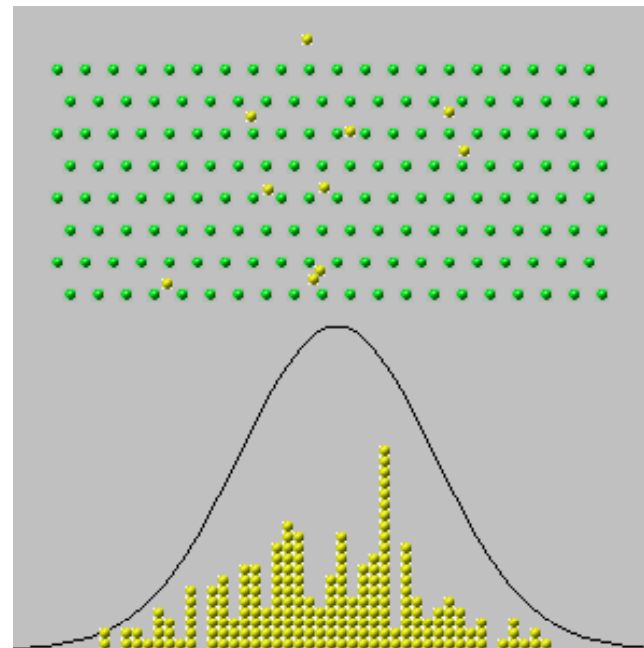
Eine Linearkombination von Normalverteilten Zufallsvariablen ist deswegen auch Normalverteilt:

$$S = a_0 + \sum_{i=1}^n a_i X_i$$

Zufallsvariablen

- Normalverteilung

Die Akkumulation von zufälligen Bewegungen



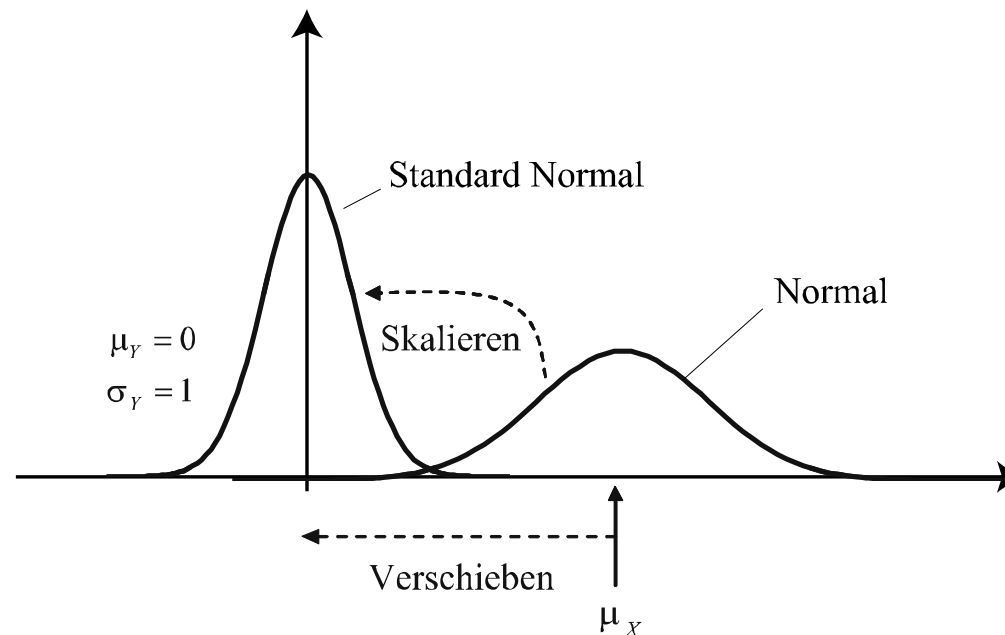
Zufallsvariablen

- Normalverteilung

Wenn der Mittelwert der Normalverteilung gleich Null ist und die Standardabweichung gleich Eins, erhält man die *Standardnormalverteilung*.

Standardisierte Dichtefunktion der Normalverteilung

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



Zufallsvariablen

- Normalverteilung

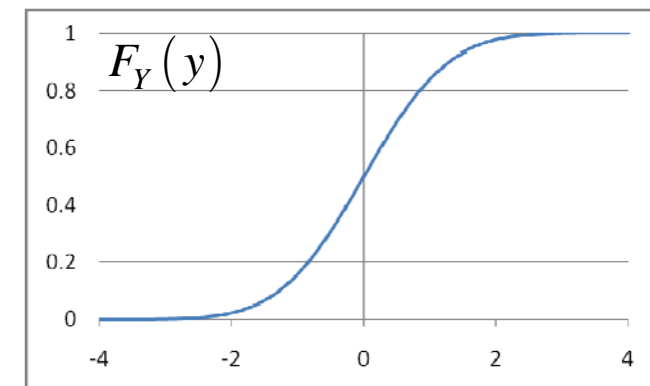
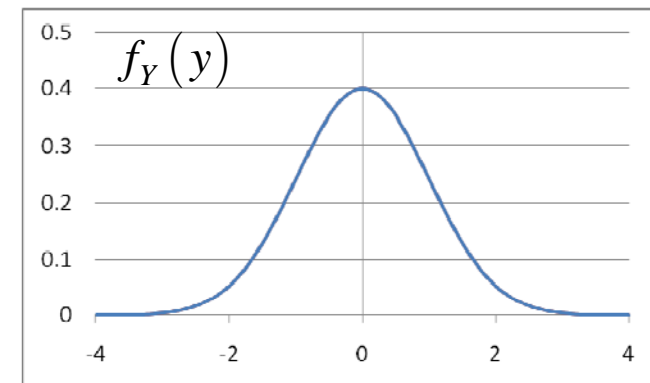
Wenn der Mittelwert der Normalverteilung gleich Null ist und die Standardabweichung gleich Eins, erhält man die *Standardnormalverteilung*.

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

standardisiert

$$f_Y(y) = \varphi(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right)$$

$$F_Y(y) = \Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y \exp\left(-\frac{1}{2}x^2\right) dx$$



Zufallsvariablen

- Lognormalverteilung

Wenn der Logarithmus der Zufallsvariablen X , d.h.

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

normalverteilt ist, ist die Variable X lognormalverteilt $X : LN(\lambda_X, \xi_X)$

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

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

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

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

Zufallsvariablen

Die Normalverteilung ergibt sich aus der **Summe** von mehreren Zufallsvariablen (**zentraler Grenzwertsatz**).

Die Lognormalverteilung ergibt sich aus dem **Produkt** von mehreren 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 Lognormalverteilung hat die nützliche Eigenschaft, dass:

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

wobei Y_i eine unabhängige lognormalverteilte Zufallsvariable mit Parametern ζ_i , λ_i und $\varepsilon_i = 0$ ist.

Somit ist auch P lognormalverteilt mit Parametern:

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

Zufallsvariablen

Die Lognormalverteilung wird häufig angewendet für die Modellierung von

- unsicheren Parametern, welche nicht-negative Realisationen annehmen können.
- Ermüdungslebensdauern.
- Festigkeit von Holz, Stahl und Beton.
- täglichen Niederschlagsmengen.
- unsicheren Einflussparametern aus dem Produkt von Zufallsvariablen.

Zufallsvariablen

Beton Druckfestigkeit

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

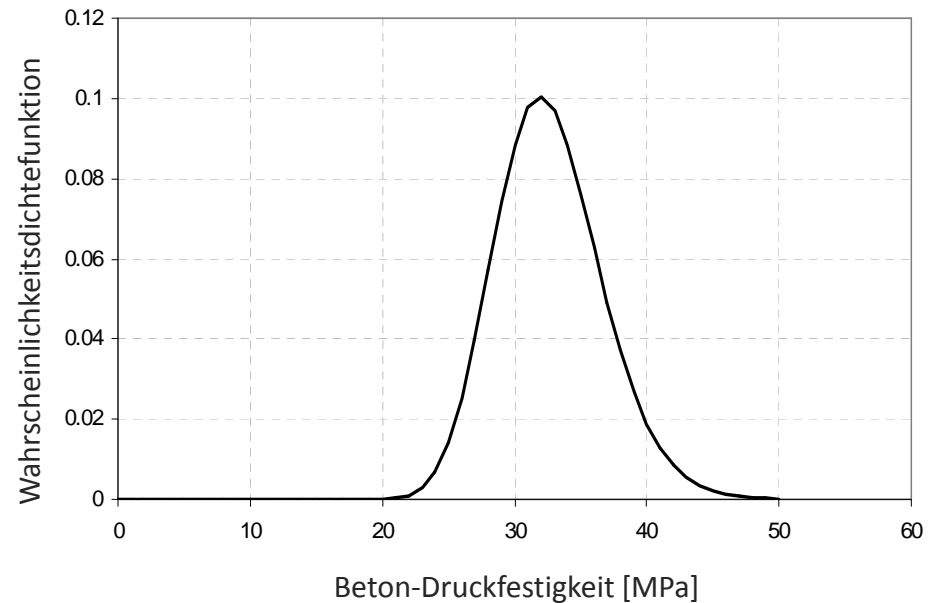
Wahrscheinlichkeit
eines Ereignisses
kleiner als 25 MPa

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

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

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

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



Zufallsvariablen

Es besteht eine grosse Anzahl von verschiedenen Wahrscheinlichkeitsverteilungen:

Gleichverteilung

Normalverteilung

Lognormalverteilung

Exponentialverteilung

Betaverteilung

Gammaverteilung

...

Verteilungstyp	Parameter	Erwartungswert und Streuung
Gleichverteilung $a \leq x \leq b$ $f_X(x) = \frac{1}{b-a}$ $F_X(x) = \frac{x-a}{b-a}$	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
Verschobene Lognormalverteilung, $x > \varepsilon$ $f_X(x) = \frac{1}{(x-\varepsilon)\zeta\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln(x-\varepsilon)-\lambda}{\zeta}\right)^2\right)$ $F_X(x) = \Phi\left(\frac{\ln(x-\varepsilon)-\lambda}{\zeta}\right)$	λ $\zeta > 0$ ε	$\mu = \varepsilon + \exp\left(\lambda + \frac{\zeta^2}{2}\right)$ $\sigma = \exp\left(\lambda + \frac{\zeta^2}{2}\right) \sqrt{\exp(\zeta^2) - 1}$
Verschobene Exponentialverteilung, $x \geq \varepsilon$ $f_X(x) = \lambda \exp(-\lambda(x-\varepsilon))$ $F_X(x) = 1 - \exp(-\lambda(x-\varepsilon))$	ε $\lambda > 0$	$\mu = \varepsilon + \frac{1}{\lambda}$ $\sigma = \frac{1}{\lambda}$

Stochastische Prozesse und Extreme

Zufallsgrößen können zeitabhängig sein, so dass sie im Laufe der Zeit neue Realisationen annehmen.

- 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, Wellenhöhe,...

Stochastische Prozesse und Extreme

▪ Zufallsreihen

- Eine Reihe von Experimenten mit nur zwei möglichen sich gegenseitig ausschliessenden Ergebnissen nennt man einen **Bernoulli-Versuch**.
- Typischerweise nennt man die Ereignisse eines Bernoulli-Versuchs **Erfolg und Versagen**.

Wenn die Wahrscheinlichkeit p eines Erfolgs in einem Versuch konstant ist, dann kann die Wahrscheinlichkeitsdichte von y Erfolgen in n Versuchen wie folgt bestimmt werden:

$$p_Y(y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, 1, 2, \dots, n$$

↑
Wahrscheinlichkeitsdichtefunktion

↑
Binomialkoeffizient

Stochastische Prozesse und Extreme

- Zufallsreihen
 - Eine Reihe von Experimenten mit nur zwei möglichen sich gegenseitig ausschliessenden Ereignissen nennt man einen **Bernoulli Versuch**.

Die binomiale Wahrscheinlichkeitsverteilungsfunktion ist wie folgt definiert:

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

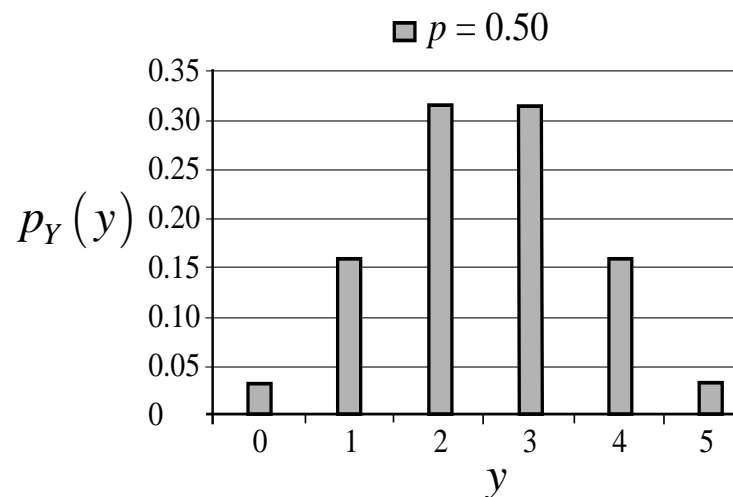
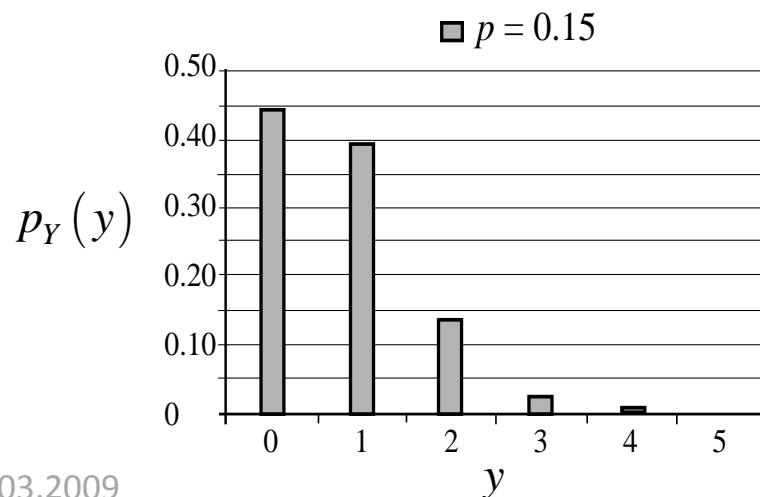
Stochastische Prozesse und Extreme

- Zufallsreihen

- Eine Reihe von Experimenten mit nur zwei möglichen sich gegenseitig ausschliessenden Ereignissen nennt man einen **Bernoulli Versuch**.

- Illustration:

Binomiale Wahrscheinlichkeitsdichtefunktion mit $n = 5$ und $p = 0.15$ bzw. $p = 0.50$



Kleines Beispiel 1

Wir erinnern uns, dass wir die Wahrscheinlichkeitsdichtefunktion einer Funktion einer Zufallsvariable $Y = g(X)$ bestimmen können durch:

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

$$Y = X^2$$

⇓

$$X = \sqrt{Y}$$

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

⇓

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

Stochastische Prozesse und Extreme

- Zufallsreihen

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

$$E[Y] = np$$

$$\text{Var}[Y] = np(1 - p)$$

Stochastische Prozesse und Extreme

- Zufallsreihen

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

$$p_N(n) = p(1-p)^{n-1} \quad \leftarrow \text{Geometrische Wahrscheinlichkeitsdichtefunktion}$$

Die Wahrscheinlichkeitsverteilungsfunktion ist somit:

$$P_N(n) = \sum_{i=1}^n p(1-p)^{i-1} = 1 - (1-p)^n \quad \leftarrow \text{Geometrische Wahrscheinlichkeitsverteilungsfunktion}$$

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:

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

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

⇓

Wir logarithmieren auf beiden Seiten...

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

und wenden folgende Umschreibung an:

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

⇓

$$\ln(1 - p) \approx -p \quad \text{für kleine } p$$

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

Stochastische Prozesse und Extreme

Anwendungen des Resultats

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

$$n = 0.7 \cdot 6 = 4 \text{ Würfe}$$

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

$$n = 0.7 \cdot 36 = 25 \text{ Würfe}$$

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

$$n = 0.7 \cdot 1000 = 700 \text{ Jahre}$$

Stochastische Prozesse und Extreme

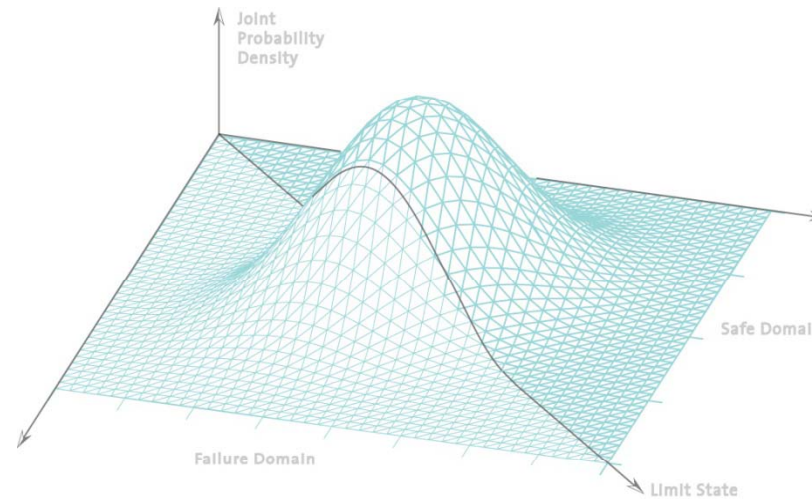
- Zufallsreihen

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

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

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

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



Statistik und Wahrscheinlichkeitsrechnung

Dr. Jochen Köhler