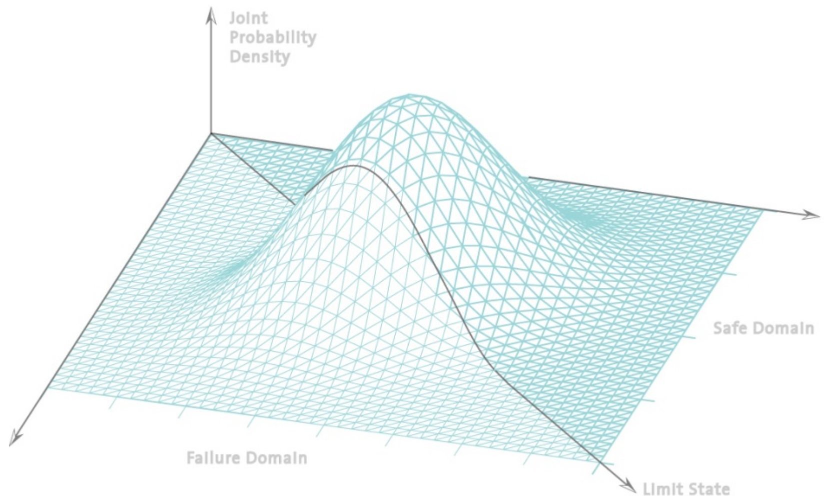


Statistik und Wahrscheinlichkeitsrechnung

Dr. Jochen Köhler

Inhalt der heutigen Vorlesung

- Zusammenfassung der vorherigen Vorlesung
- Übersicht über Schätzung und Modellbildung
- Modellevaluation durch Statistische Tests
  - Der χ^2 –Test für die Güte der Anpassung
  - Der Kolmogorov-Smirnov Test für die Güte der Anpassung
  - Modellvergleiche

Zusammenfassung der vorherigen Vorlesung

- Wir betrachteten die Möglichkeit, die Parameter einer Verteilung basierend auf Beobachtungen/Daten abschätzen zu können.

  Was haben wir gelernt?

  Dass die Parameter einer Verteilung geschätzt werden können mit Hilfe von z.B. der:

  – Methode der Momente MoM

  – Methode der maximalen Likelihood MLM

Zusammenfassung der vorangehenden Vorlesung

- Methode der Momente (MoM) – Punktschätzung

  Das Prinzip der MoM ist: Wir schätzen die Parameter, indem wir die analytisch berechneten Momente mit den Stichprobenmomenten gleichsetzen.

$$m_1 = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i \qquad \lambda_1 = \int_{-\infty}^{\infty} x \cdot f_X(x|\mu,\sigma)dx$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i^2 \qquad \lambda_2 = \int_{-\infty}^{\infty} x^2 \cdot f_X(x|\mu,\sigma)dx$$

Dies führt zu k Gleichungen, welche gelöst werden müssen um k Parameter abzuschätzen.

Zusammenfassung der vorangehenden Vorlesung

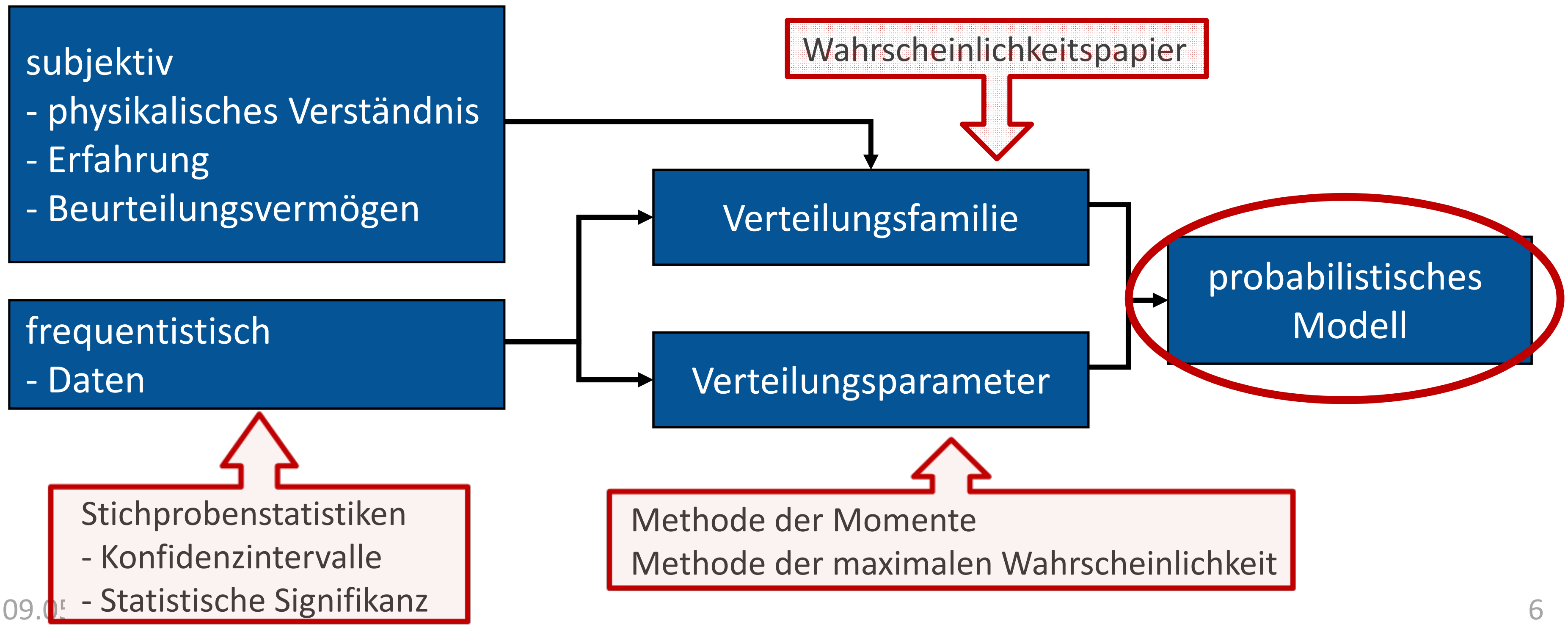
- Methode der maximalen Likelihood (MLM) – Schätzung der Parameter und ihrer Verteilung

  Das Prinzip der MLM ist: Die Parameter werden geschätzt, indem die Likelihood das die Parameter die Beobachtungen/Daten repräsentieren maximiert wird.

$$\left.\begin{array}{l} L(\boldsymbol{\theta}|\hat{\mathbf{x}}) = \displaystyle\prod_{i=1}^{n} f_X(\hat{x}_i|\boldsymbol{\theta}) \\[2ex] l(\boldsymbol{\theta}|\mathbf{x}) = \displaystyle\sum_{i=1}^{n} \log(f_X(\hat{x}_i|\boldsymbol{\theta})) \\[2ex] \displaystyle\min_{\boldsymbol{\theta}}(-l(\boldsymbol{\theta}|\hat{\mathbf{x}})) \end{array}\right\} \Rightarrow \left\{\begin{array}{l} \boldsymbol{\mu}_{\Theta} = (\theta_1^*, \theta_2^*, \ldots, \theta_n^*)^T \\[2ex] \mathbf{C}_{\Theta\Theta} = \mathbf{H}^{-1} \\[2ex] H_{ij} = \dfrac{\partial^2 - l(\boldsymbol{\theta}|\hat{\mathbf{x}})}{\partial\theta_i \partial\theta_j}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \end{array}\right.$$

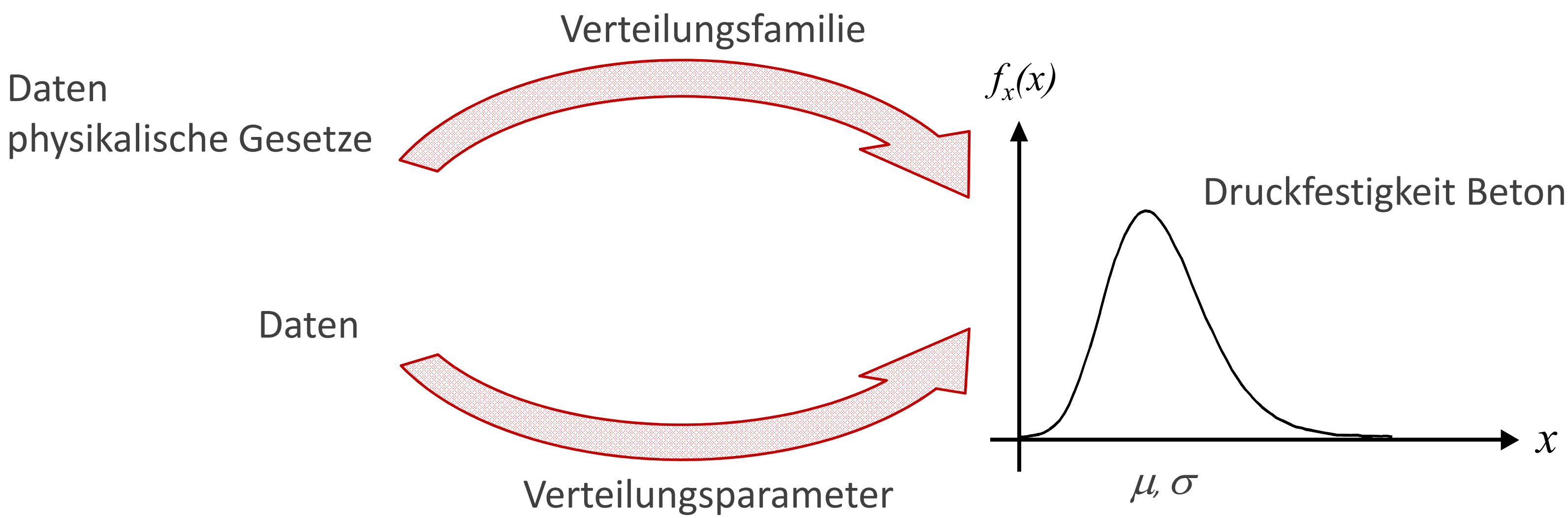
Schätzung und Modellentwicklung - Übersicht

- Unterschiedliche Typen an Information werden genutzt, wenn Ingenieurmodelle entwickelt werden.
  - Subjektive Information
  - Frequentistische Information



Modellevaluation durch Statistische Tests

Nehmen wir an, dass wir eine bestimmte Verteilungsfunktion gewählt haben, um die Unsicherheit eines unsicheren Ereignisses zu modellieren.



Nun wollen wir die Wahl unserer Verteilung prüfen –

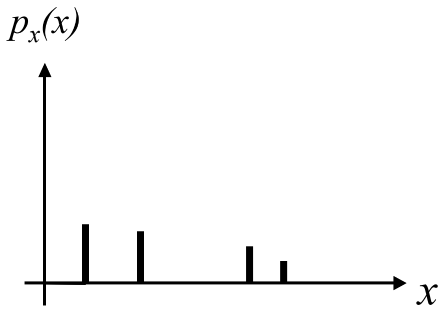
Durch statistische Tests.

Modellevaluation durch Statistische Tests

Zwei unterschiedliche Fälle werden betrachtet:

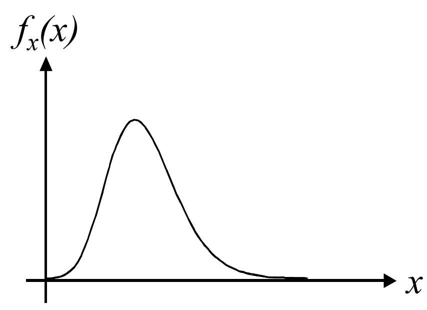
Verifizierung von

1: Diskreten Verteilungsfunktionen



CHI-Quadrat (χ^2) Test

2: Kontinuierlichen Verteilungsfunktionen

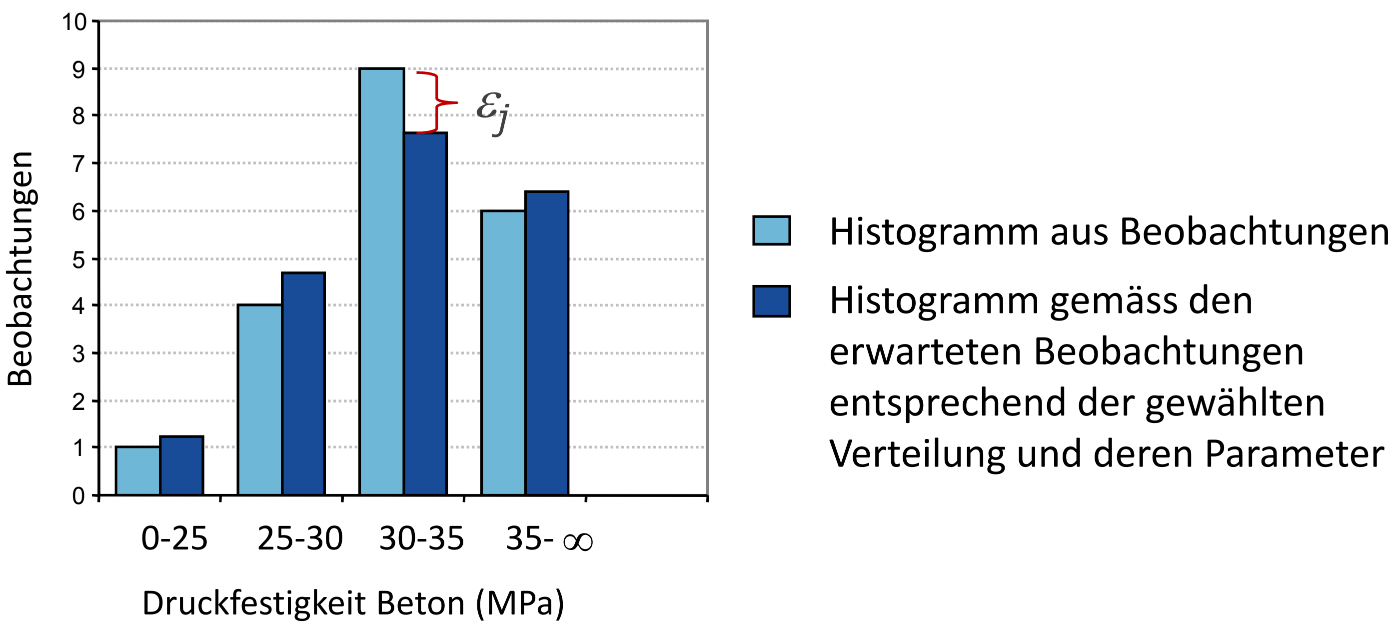


Kolmogorov Smirnov Test

Modellevaluation durch Statistische Tests

Der CHI-Quadrat Test der Güte der Anpassung

- Die Idee dahinter ist, dass die Differenzen ε_j zwischen der erwarteten und der beobachteten Datenverteilung klein sein sollte, wenn die gewählte Verteilungsfamilie die Stichprobe gut beschreiben kann.

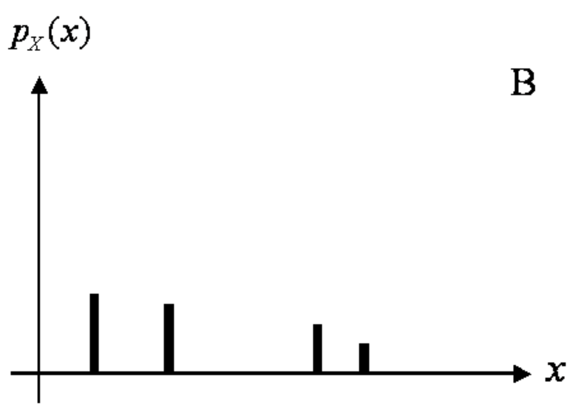


Modellevaluation durch Statistische Tests

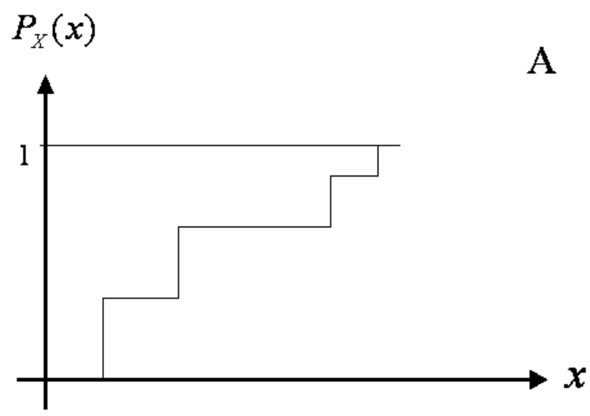
Der CHI-Quadrat Test der Güte der Anpassung

- Wie wir bereits wissen, ist eine diskrete kumulative Wahrscheinlichkeitsverteilungsfunktion wie folgt gegeben:

$$P(x_i) = \sum_{j=1}^{i-1} p(x_j)$$



Wahrscheinlichkeitsdichtefunktion



Kumulative Wahrscheinlichkeitsverteilungsfunktion

Modellevaluation durch Statistische Tests

Der CHI-Quadrat Test der Güte der Anpassung

- Sei n die Anzahl Beobachtungen einer diskreten Zufallsvariable X. Die Anzahl an Beobachtungen von $X = x_i$ d.h. N_i ist eine binomial verteilte Zufallsvariable mit folgendem Erwartungswert und Varianz:

$$E\left[N_i\right] = np(x_i) = N_{p,i}$$

Erwartete Anzahl Beobachtungen eines gewissen Wertes

$$Var\left[N_i\right] = np(x_i)(1 - p(x_i)) = N_{p,i}(1 - p(x_i))$$

Modellevaluation durch Statistische Tests

Der CHI-Quadrat Test der Güte der Anpassung

- Sei n die Anzahl Beobachtungen einer diskreten Zufallsvariable X. Die Anzahl an Beobachtungen von $X = x_i$ d.h. N_i ist eine binomial verteilte Zufallsvariable mit folgendem Erwartungswert und Varianz:

$$E[N_i] = np(x_i) = N_{p,i}$$

Erwartete Anzahl Beobachtungen eines gewissen Wertes

$$Var[N_i] = np(x_i)(1-p(x_i)) = N_{p,i}(1-p(x_i))$$

- Wenn das postulierte Modell korrekt und n gross genug ist, dann ist gemäss dem zentralen Grenzwertsatz die Differenz ε_i standard normalverteilt.

$$\varepsilon_i = \frac{N_{o,i} - N_{p,i}}{\sqrt{N_{p,i}(1-p(x_i))}}$$

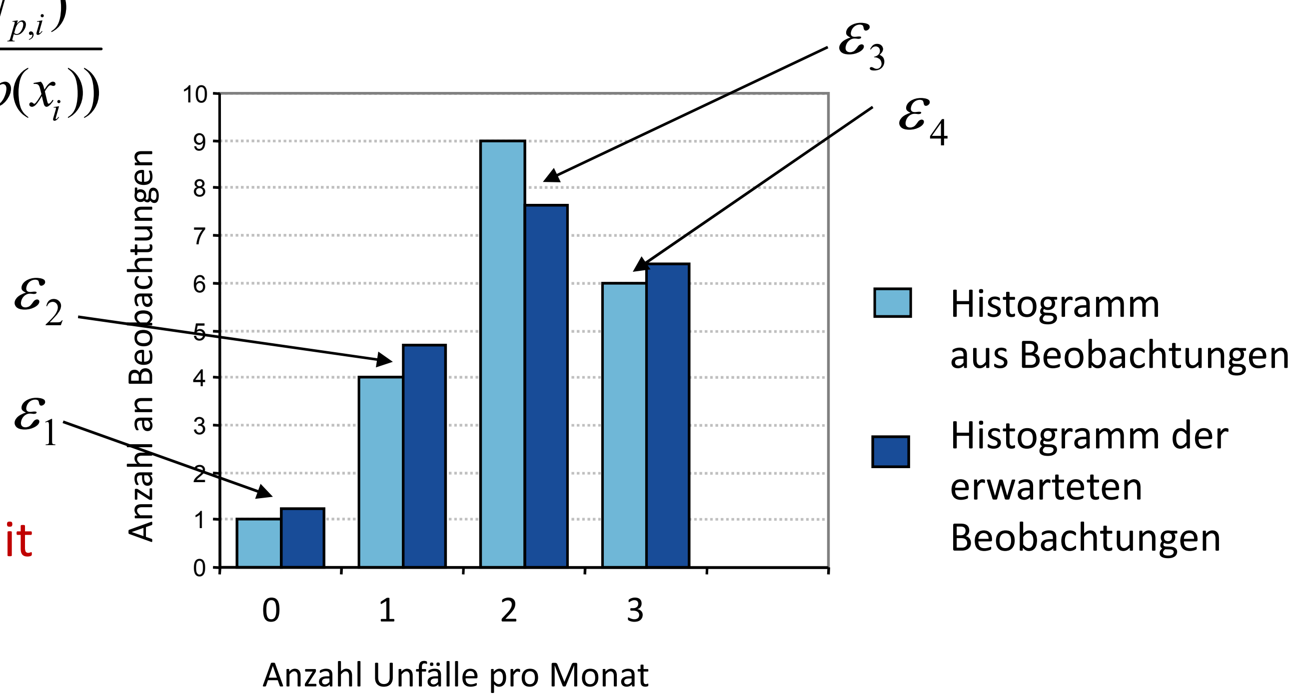
Beobachtete Anzahl Beobachtungen eines gewissen Wertes

Modellevaluation durch Statistische Tests

Der CHI-Quadrat Test der Güte der Anpassung

- Werden die quadrierten Differenzen von beobachteten und erwarteten Anzahl Beobachtungen summiert, dann erhalten wir:

$$\varepsilon^2 = \sum_{i=1}^{k} \varepsilon_i^2 = \sum_{i=1}^{k} \frac{(N_{o,i} - N_{p,i})^2}{N_{p,i}(1 - p(x_i))}$$



CHI-Quadrat verteilt mit k-1 Freiheitsgraden

Modellevaluation durch Statistische Tests

Der CHI-Quadrat Test der Güte der Anpassung

- Es wird nun auf einem Signifikanzniveau α getestet, ob die Summe aller beobachteten quadrierten Differenzen plausibel ist, d.h.

  Es wird die Nullhypothese H_0 aufgestellt, dass die gewählte Verteilungsfunktion die beobachtete Stichprobe repräsentiert. Die Vorgehensregel lautet dann

$$P(\varepsilon_m^2 \geq \Delta) = \alpha$$

Die Alternativhypothese H_1 ist weit weniger informativ, weil mit ihr ausser der gewählten Verteilung alle anderen Verteilungen akzeptiert werden.

Δ ist der α Fraktilwert der χ^2 Verteilung mit $v = k - 1 - j$ Freiheitsgraden.

Modellevaluation durch Statistische Tests

Der CHI-Quadrat Test der Güte der Anpassung

- Wir betrachten folgendes Beispiel:

  Als Verteilungsfunktion für 20 Beobachtungen der Betondruckfestigkeit nehmen wir die Normalverteilung an.

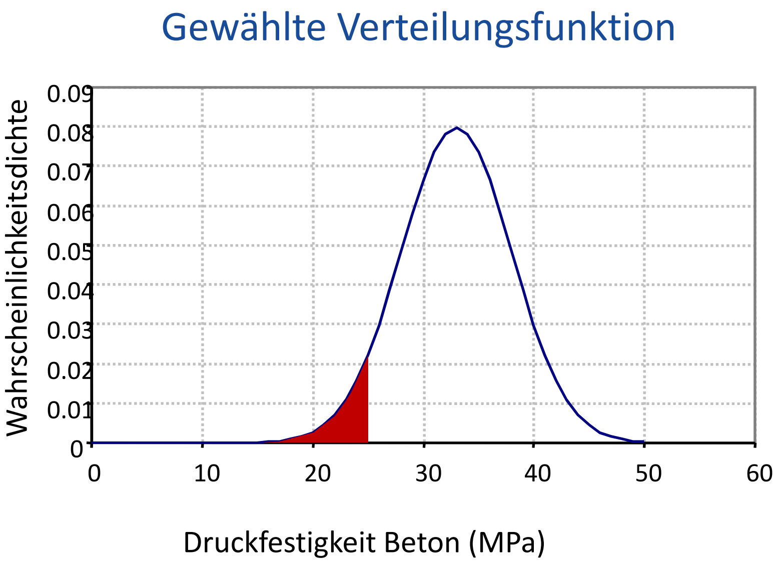
Der Mittelwert beträgt	33 Mpa
Und die Standardabweichung	5 Mpa

  Wobei die Parameter nicht aus den vorhandenen Beobachtungen geschätzt werden.
  Die Normalverteilung ist eine kontinuierliche Verteilung.
  Aber sie kann ganz einfach diskretisiert werden!

Modellevaluation durch Statistische Tests

Der CHI-Quadrat Test der Güte der Anpassung

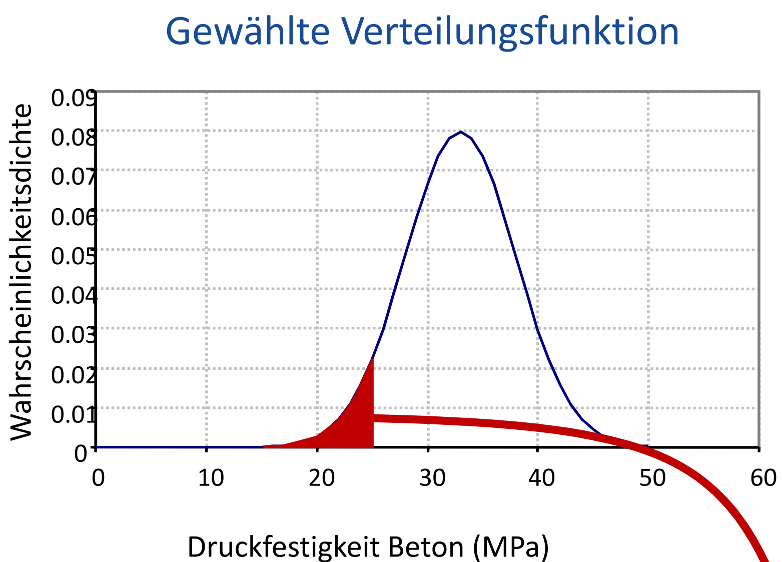
- Die Dichtefunktion der gewählten Verteilungsfunktion wird diskretisiert:



Modellevaluation durch Statistische Tests

Der CHI-Quadrat Test der Güte der Anpassung

- Die Dichtefunktion der gewählten Verteilungsfunktion wird diskretisiert:



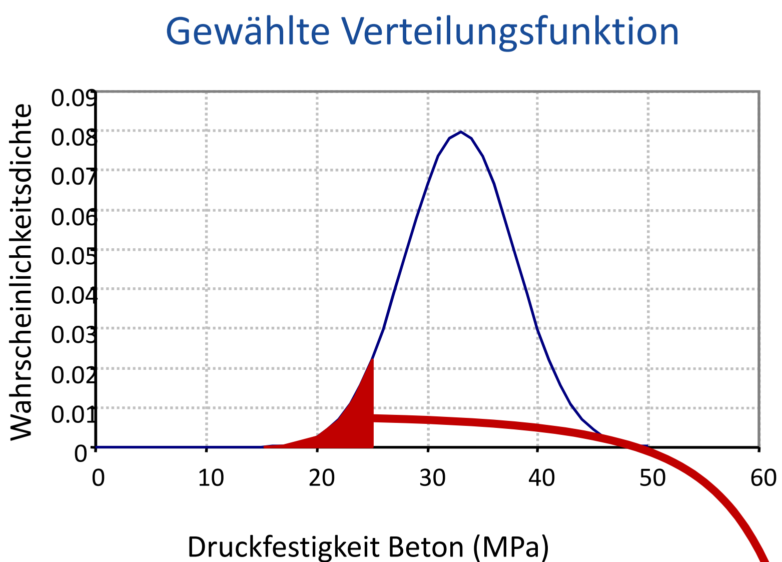
Intervall 0-25: $$20\left[\Phi\left(\frac{25-33}{5}\right)-\Phi\left(\frac{-\infty-33}{5}\right)\right]=20\cdot 0.055=1.10$$

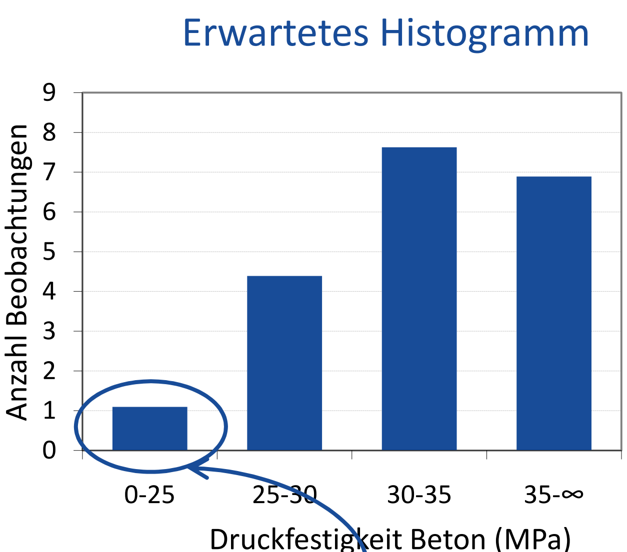
Totale Anzahl an Versuchen

Modellevaluation durch Statistische Tests

Der CHI-Quadrat Test der Güte der Anpassung

- Die Dichtefunktion der gewählten Verteilungsfunktion wird diskretisiert:





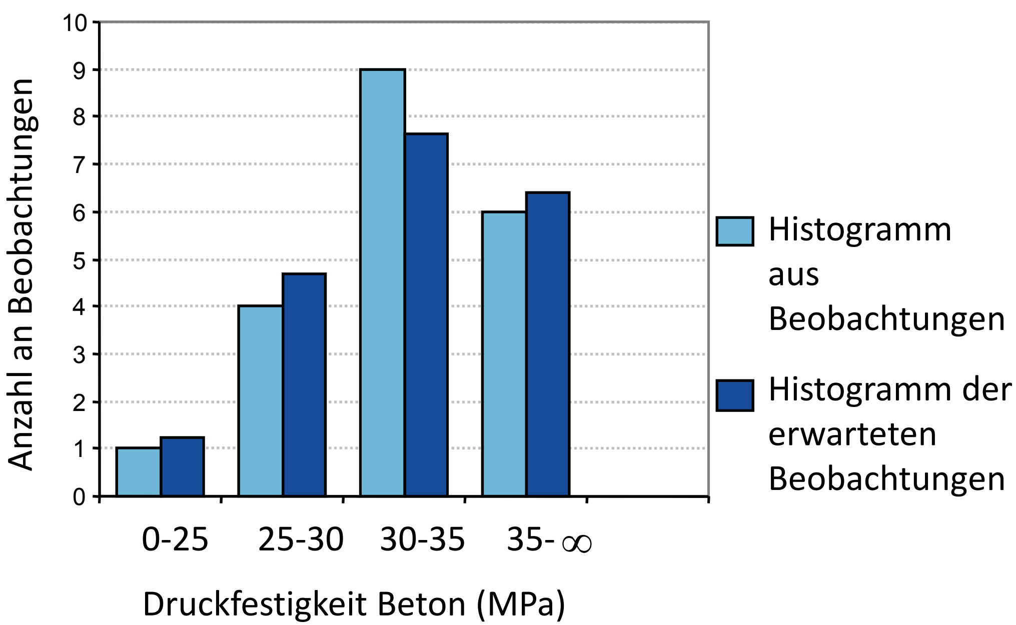
Intervall 0-25: $$20\left[\Phi\left(\frac{25-33}{5}\right)-\Phi\left(\frac{-\infty-33}{5}\right)\right]=20\cdot 0.055=1.10$$

Totale Anzahl an Versuchen

Modellevaluation durch Statistische Tests

Der CHI-Quadrat Test der Güte der Anpassung

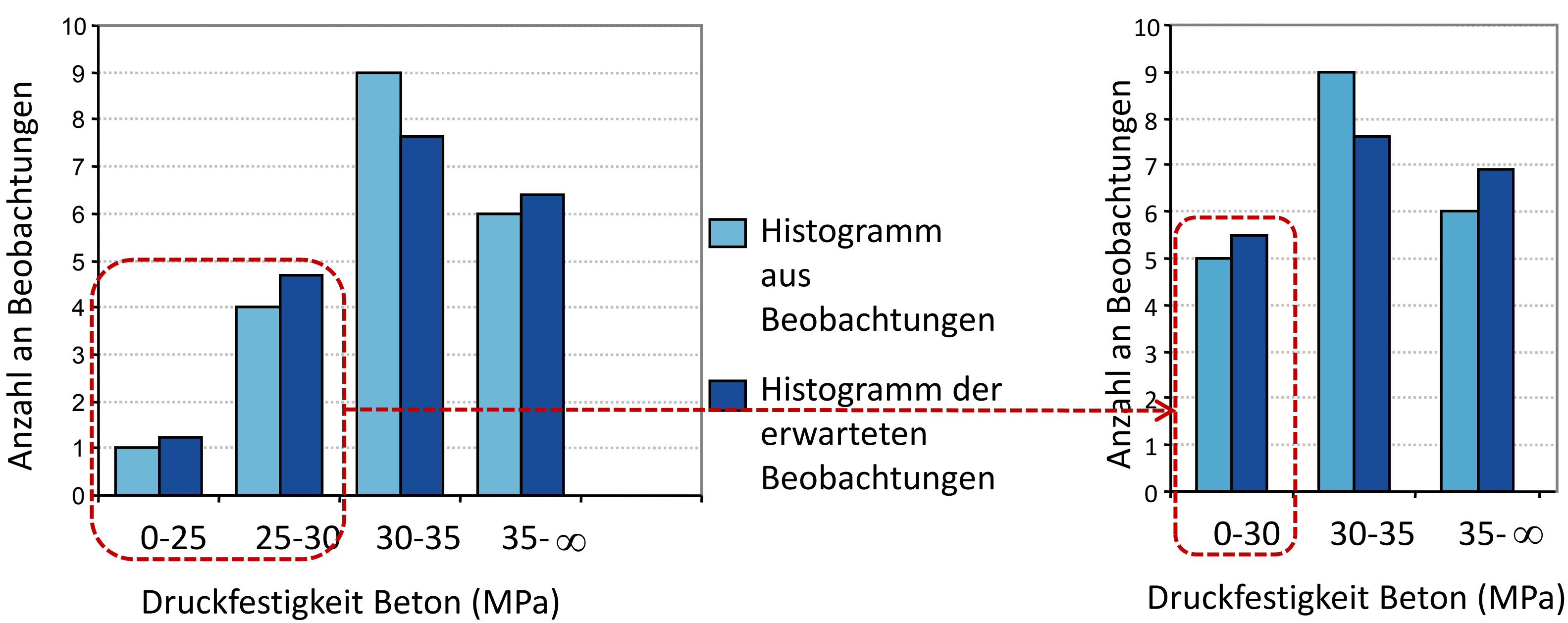
- Die beobachteten und erwarteten Histogramme können nun verglichen werden.



Modellevaluation durch Statistische Tests

Der CHI-Quadrat Test der Güte der Anpassung

- Die beobachteten und erwarteten Histogramme können nun verglichen werden.
  - Aufgrund der kleinen Anzahl an Stichproben im unteren Bereich werden die zwei unteren Intervalle zusammengeführt.



Modellevaluation durch Statistische Tests

Der CHI-Quadrat Test der Güte der Anpassung

– Berechnungen zum genannten Beispiel

Intervall x_j (MPa)	Anzahl Beobach-tungen $N_{o,j}$	Erwartete Wahrschein-lichkeiten	Erwartete Anzahl Beo-bachtungen $N_{p,j}$	Stich-proben-statistik
0- 30	5	0.296671	5.933415	0.14684
30 -35	9	0.381169	7.65443	0.236537
35- ∞	6	0.344578	6.412155	0.026492
			Summe	**0.40987**

$$\varepsilon_m^2 = \sum_{j=1}^{k} \frac{(N_{o,j} - N_{p,j})^2}{N_{p,j}}$$

Auf einem Signifikanzniveau von 5% erhalten wir für die CHI-Quadrat-Verteilung
Mit N=3-1=2 Freiheitsgraden aus der Tabelle: Δ = 5.99.
Da **0.40987** kleiner ist als 5.99, kann die Nullhypothese H_0 nicht verworfen werden.

Modellevaluation durch Statistische Tests

Der CHI-Quadrat Test der Güte der Anpassung

- Wird eines, oder mehrere (m) Parameter der gewählten Verteilung aus den gleichen Daten bestimmt wurde, welche auch für den Test verwendet wurden, dann muss die Anzahl der Freiheitsgrade entsprechend reduziert werden:

$$v = k - 1 - j$$

  - Unter der Annahme, dass die Varianz aus den Daten bestimmt wurde, aber nicht der Mittelwert, erhalten wir $n = 3-1-1=1$ Freiheitsgrad.

Modellevaluation durch Statistische Tests

Der CHI-Quadrat Test der Güte der Anpassung

- Wenn wir eine Normalverteilung annehmen mit folgenden Parametern:

  $\mu =$ 33.00

  $\sigma =$ 4.05

  erhalten wir folgendes Ergebnis:

Intervall x_j (MPa)	Anzahl Beobach-Tungen $N_{o,j}$	Erwartete Wahrschein-lichkeiten $p(x_j)$	Erwartete Anzahl Beo-bachtungen $N_{p,j}=20p(x_j)$	Stich-proben-statistik
0 - 30	5	0.274253	5.485061	0.042896
30 - 35	9	0.381169	7.623373	0.248591
35 - ∞	6	0.344578	6.891566	0.115342
			Summe	0.406829

Auf einem Signifikanzniveau von 5% erhalten wir für die CHI-Quadrat-Verteilung mit N=3-1-1=1 Freiheitsgraden aus der Tabelle: Δ = 3.84.
Da 0.406829 kleiner ist als 3.84, kann die Nullhypothese H_0 nicht verworfen werden.

Modellevaluation durch Statistische Tests

Der Kolmogorov-Smirnov Test der Güte der Anpassung

- Die Idee hinter dem Kolmogorov-Smirnov Test ist folgende:

Wenn die kumulative Wahrscheinlichkeitsverteilungsfunktion der gewählten Verteilung für die Beobachtungen in Betracht kommt, dann sollte die maximale Differenz zwischen der beobachteten und der erwarteten kumulativen Wahrscheinlichkeitsverteilungsfunktion $\varepsilon_{\max}$ klein sein.

$$\varepsilon_{\max} < \Delta_{n,\alpha}$$

Modellevaluation durch Statistische Tests

Der Kolmogorov-Smirnov Test der Güte der Anpassung

- Die kumulative Wahrscheinlichkeitsverteilungsfunktion der Beobachtungen kann berechnet werden als:

$$F_o(x_i) = \frac{i}{n}$$

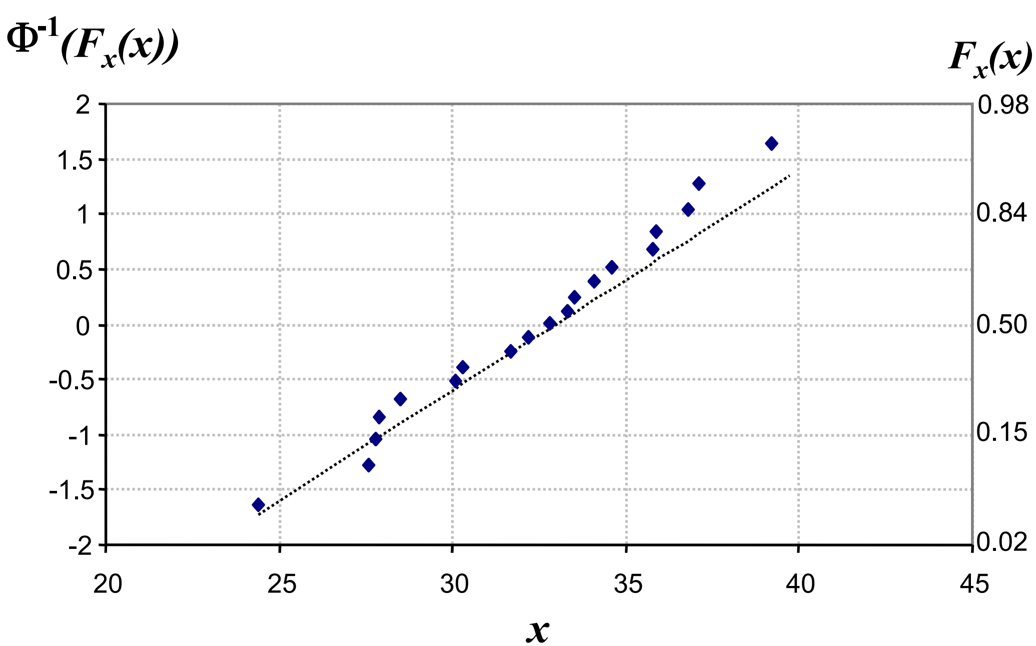
Folgende Stichprobenstatistik wird benutzt:

$$\varepsilon_{\max} = \max\left[\left|F_o\left(x_i\right) - F_p\left(x_i\right)\right|\right] = \max\left[\left|\frac{i}{n} - F_p\left(x_i\right)\right|\right]$$

Modellevaluation durch Statistische Tests

Der Kolmogorov-Smirnov Test der Güte der Anpassung

- Die Kolmogorov-Smirnov Stichprobenstatistik wird folgendermassen ermittelt:



i	x_i	$F_{x0}(x_i)$	$F_{xp}(x_i)$	ε_i
1	24.4	0.05	0.042716	0.007284
2	27.6	0.1	0.140071	0.040071
3	27.8	0.15	0.14917	0.00083
4	27.9	0.2	0.153864	0.046136
5	28.5	0.25	0.18406	0.06594
6	30.1	0.3	0.280957	0.019043
7	30.3	0.35	0.294598	0.055402
8	31.7	0.4	0.397432	0.002568
9	32.2	0.45	0.436441	0.013559
10	32.8	0.5	0.484047	0.015953
11	33.3	0.55	0.523922	0.026078
12	33.5	0.6	0.539828	0.060172
13	34.1	0.65	0.587064	0.062936
14	34.6	0.7	0.625516	0.074484
15	35.8	0.75	0.71226	0.03774
16	35.9	0.8	0.719043	0.080957
17	36.8	0.85	0.776373	0.073627
18	37.1	0.9	0.793892	0.106108
19	39.2	0.95	0.892512	0.057488
20	39.7	1	0.909877	0.090123

Modellevaluation durch Statistische Tests

Der Kolmogorov-Smirnov Test der Güte der Anpassung

- Die Kolmogorov-Smirnov Statistik ist tabelliert:

	n											
α	**1**	**5**	**10**	**15**	**20**	**25**	**30**	**40**	**50**	**60**	**70**	**80**
0.01	0.9950	0.6686	0.4889	0.4042	0.3524	0.3166	0.2899	0.2521	0.2260	0.2067	0.1917	0.1795
0.05	0.9750	0.5633	0.4093	0.3376	0.2941	0.2640	0.2417	0.2101	0.1884	0.1723	0.1598	0.1496
0.1	0.9500	0.5095	0.3687	0.3040	0.2647	0.2377	0.2176	0.1891	0.1696	0.1551	0.1438	0.1347
0.2	0.9000	0.4470	0.3226	0.2659	0.2315	0.2079	0.1903	0.1654	0.1484	0.1357	0.1258	0.1179

- Für n = 20 und α = 5% erhalten wir 0.2941, im Vergleich zur beobachteten Statistik von 0.1061 – die Nullhypothese H_0 kann nicht verworfen werden auf einem Signifikanzniveau von 5%.

Modellevaluation durch Statistische Tests

Modellvergleich

- Modellverifizierung durch statistische Tests kann genutzt werden, um die Plausibilität eines bestimmten Modells in Bezug auf ein bestimmtes Datenset zu quantifizieren.

Zwei Fälle müssen in Betracht gezogen werden:

1. Es kann gezeigt werden, dass die Hypothese akzeptiert werden kann.
2. Es kann gezeigt werden, dass die Hypothese verworfen werden muss.

Welche Information ist in diesen beiden Fällen enthalten?

Modellevaluation durch Statistische Tests

Modellvergleich

- Wenn ein Signifikanztest zeigt, dass eine Hypothese akzeptiert werden kann:

Wir müssen uns daran erinnern, dass auch andere Modelle (Verteilungen) in Frage kommen… tatsächlich ist es oft der Fall, dass mehrere Modelle den Signifikanztest „bestehen“!

- Wenn ein Signifikanztest zeigt, dass eine Hypothese verworfen werden muss:

Dies heisst nicht unbedingt, dass das gewählte Modell schlecht ist – es könnte bedeuten, dass der Beweis einfach nicht stark genug ist, um die entsprechende Signifikanz zu zeigen – zu wenig Daten!

Modellevaluation durch Statistische Tests

Modellvergleich

- Wenn zwei Modellhypothesen akzeptiert werden können, d.h. beide Modelle plausibel sind, dann können wir die Güte der Anpassung der zwei Modelle testen entweder durch:

- Die Stichprobenstatistik direkt vergleichen – kann nicht beweiskräftig sein, u.a. aufgrund unterschiedlicher Freiheitsgrade

- Die Likelihood vergleichen !!!

Modellevaluation durch Statistische Tests

Modellvergleich

- Betrachten wir ein Beispiel mit zwei unterschiedlichen Modellen:

Modell 1: N(33;5)

Parameter nicht aus den gleichen Daten geschätzt
n=3-1=2
CHI-Quadrat Stichprobenstatistik = 0.40987
Stichprobenwahrscheinlichkeit = 0.8151

Modell 2: N(33;4.05)

Parameter aus den gleichen Daten geschätzt
n=3-1-1=2
CHI-Quadrat Stichprobenstatistik = 0.40683
Stichprobenwahrscheinlichkeit = 0.5236

Modellevaluation durch Statistische Tests

Zusammenfassung

- Die Wahl eines geeigneten probabilistischen Modells kann durch Signifikanztests an der Modellhypothese unterstützt werden.
- Der CHI-Quadrat Test wurde für diskrete Verteilungen entwickelt.
- Der Kolmogorov-Smirnov Test wurde für kontinuierliche Verteilungen entwickelt.
- Die Güte der Anpassung verschiedener Modellalternativen kann durch den Vergleich verschiedenen Likelihoods geprüft werden.