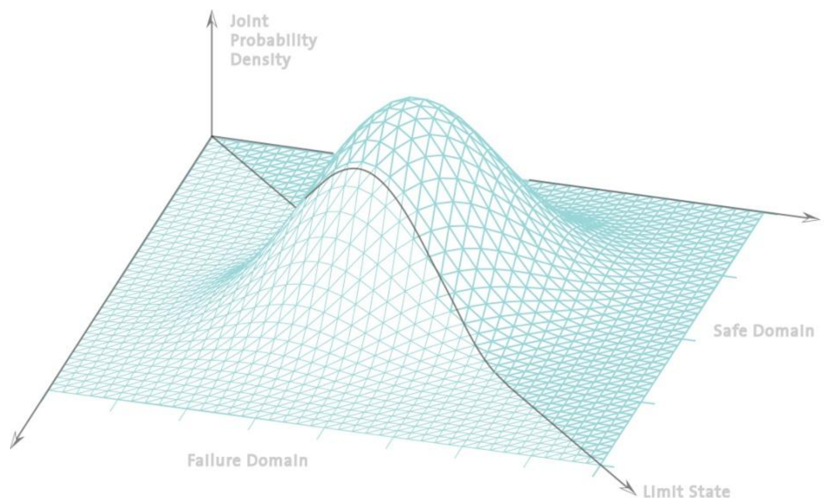


Statistik und Wahrscheinlichkeitsrechnung

11. Vorlesung

Jochen Köhler

Inhalt der heutigen Vorlesung

- Zusammenfassung Parameterschätzung
- 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
  - Modellvergleich
- Evaluation der Vorlesung

Zusammenfassung Parameterschätzung

- Die Parameter einer Verteilung können basierend auf Beobachtungen/Daten abgeschätzt werden.

  Was haben wir gelernt?

  Die Parameter einer Verteilung können z. B. anhand folgender Methoden geschätzt werden:

  - Methode der Momente (MoM)
  - Maximum-Likelihood-Methode (MLM)

Zusammenfassung der letzten 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.

Skript S. 88

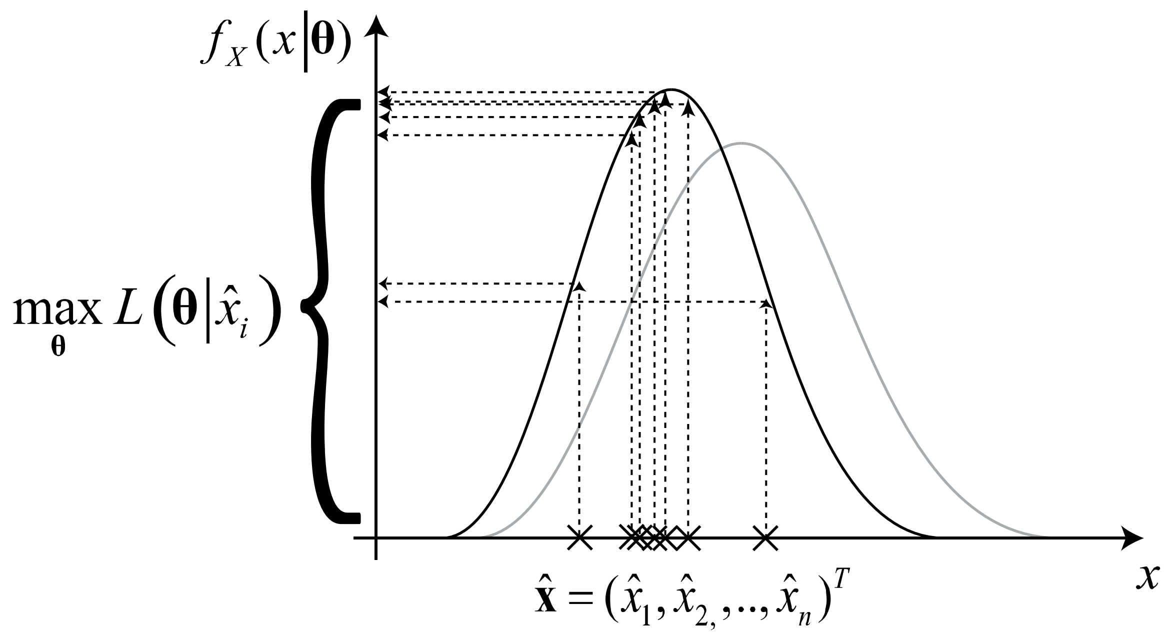
Verteilungstyp	Parameter	Erwartungswert und Streuung
Gleichverteilung $a \le x \le b$ $f_X(x) = \frac{1}{b-a}$ $F_X(x) = \frac{x-a}{b-a}$	a b	$\mu = \frac{a+b}{2}$ $\sigma = \frac{b-a}{\sqrt{12}}$
Normalverteilung $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$ $F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right) dt$	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, $x \ge \varepsilon$ $f_X(x) = \lambda \exp(-\lambda(x-\varepsilon))$ $F_X(x) = 1 - \exp(-\lambda(x-\varepsilon))$	ε $\lambda > 0$	$\mu = \varepsilon + \frac{1}{\lambda}$ $\sigma = \frac{1}{\lambda}$
Gamma, $x \ge 0$ $f_X(x) = \frac{b^p}{\Gamma(p)} \exp(-bx) x^{p-1}$ $F_X(x) = \frac{\Gamma(bx,p)}{\Gamma(p)}$ Mit: $\Gamma(bx,p) = \int\limits_0^{bx} t^{p-1} \exp(-t)\, dt$ $\Gamma(p) = \int\limits_0^{\infty} t^{p-1} \exp(-t)\, dt$	$p > 0$ $b > 0$	$\mu = \frac{p}{b}$ $\sigma = \frac{\sqrt{p}}{b}$

Zusammenfassung der letzten Vorlesung

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

- Daten $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_{2,}, .., \hat{x}_n)^T$
- Parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, .., \theta_k)^T$

$$L_i = f_X(\hat{x}_i | \boldsymbol{\theta})$$



Zusammenfassung der letzten Vorlesung

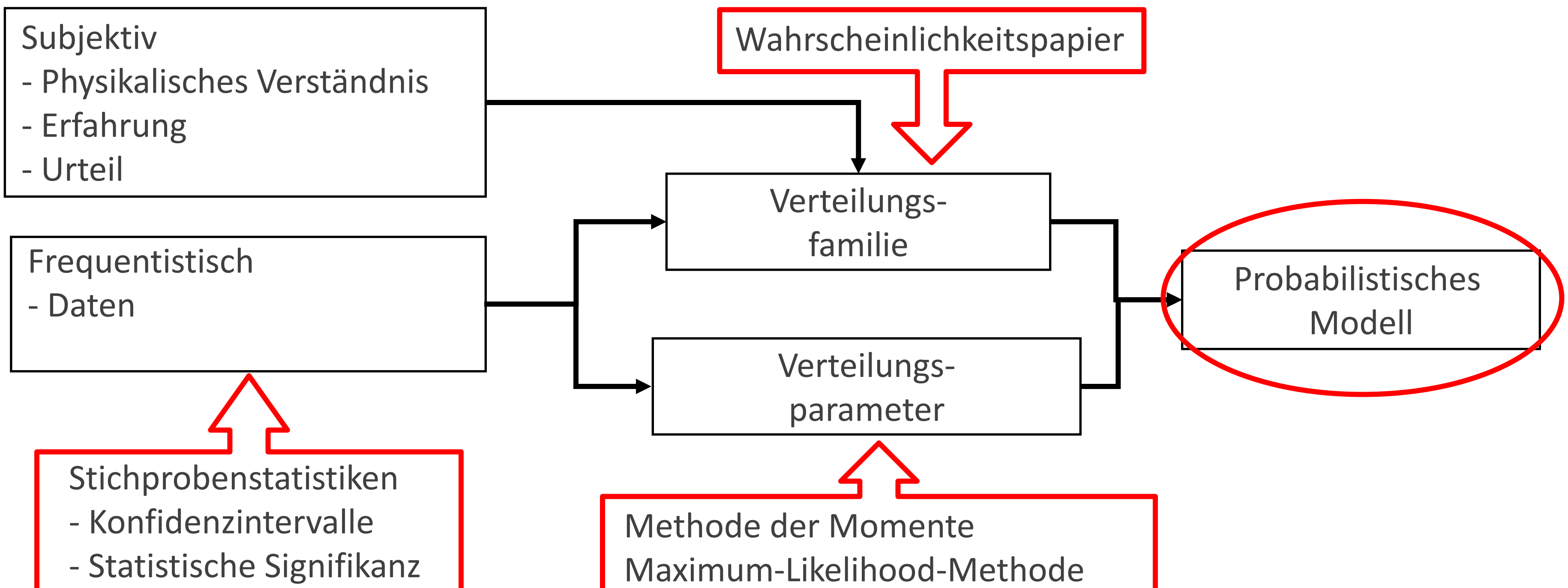
- Maximum-Likelihood-Methode (MLM) – Schätzung der Parameter und ihrer Verteilung

  Die Parameter werden geschätzt, indem die Likelihood, dass die Parameter die Beobachtungen/Daten repräsentieren, maximiert wird.

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

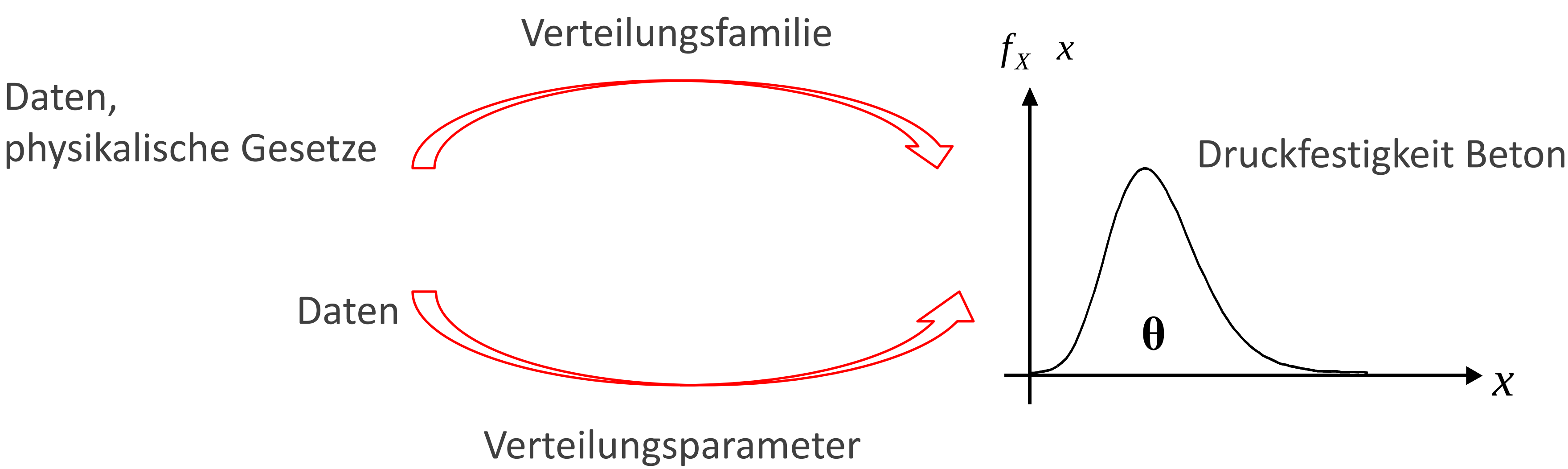
Übersicht Schätzung und Modellbildung

- Unterschiedliche Typen von Information werden genutzt, um Ingenieurmodelle zu entwickeln.
  - 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.



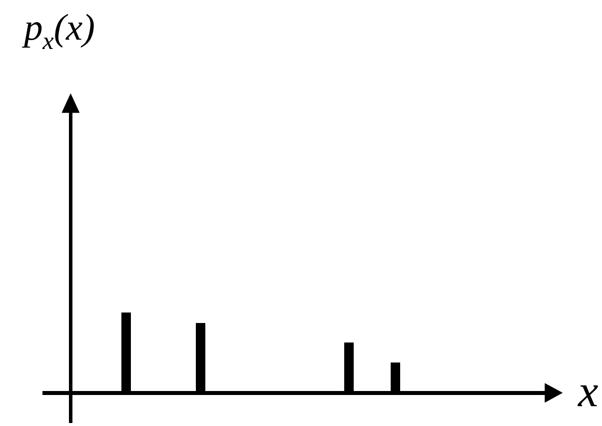
Nun wird die Wahl der Verteilung und der Parameter geprüft – durch statistische Tests.

Modellevaluation durch Statistische Tests

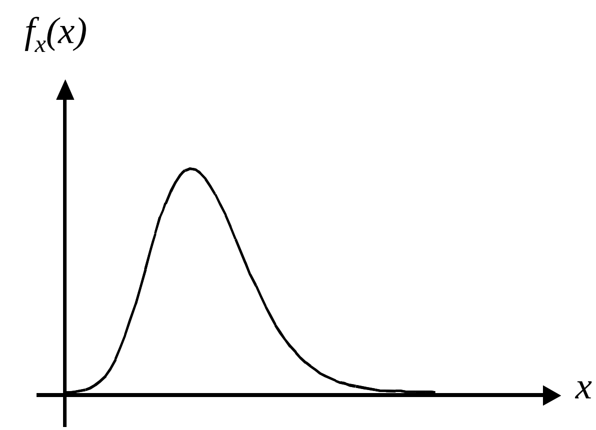
Zwei unterschiedliche Fälle werden betrachtet:

Verifizierung von

1 Diskreten Verteilungsfunktionen
χ^2-Test



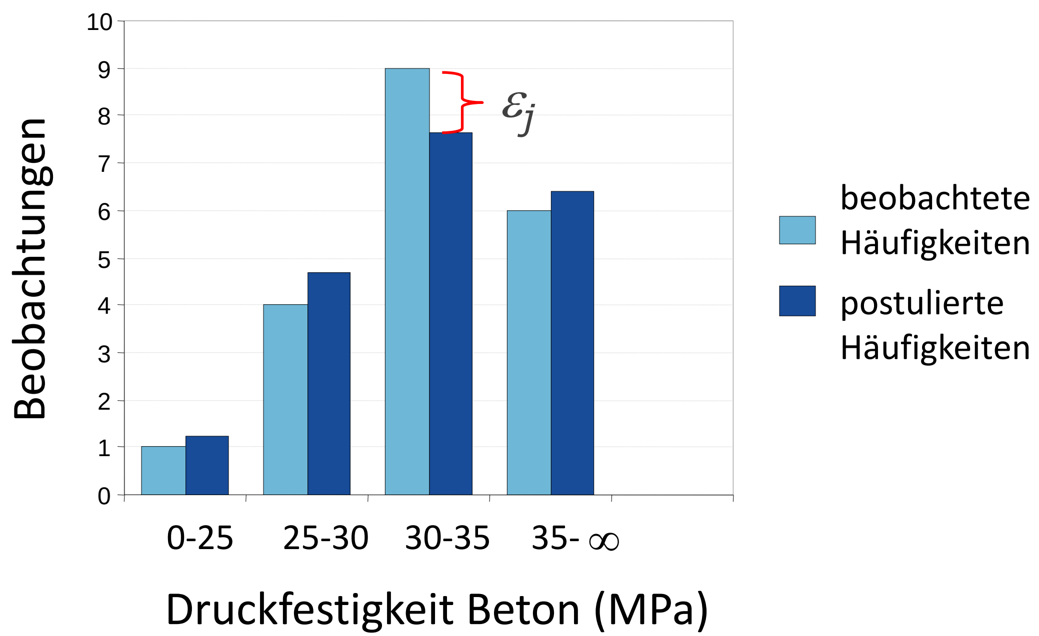
2 Kontinuierlichen Verteilungsfunktionen
Kolmogorov-Smirnov-Test



Modellevaluation durch Statistische Tests

Der χ^2-Test für die Güte der Anpassung

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

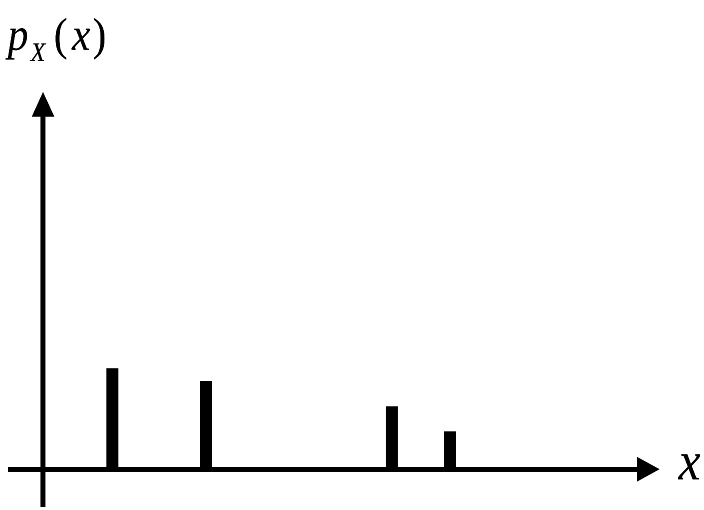


Modellevaluation durch Statistische Tests

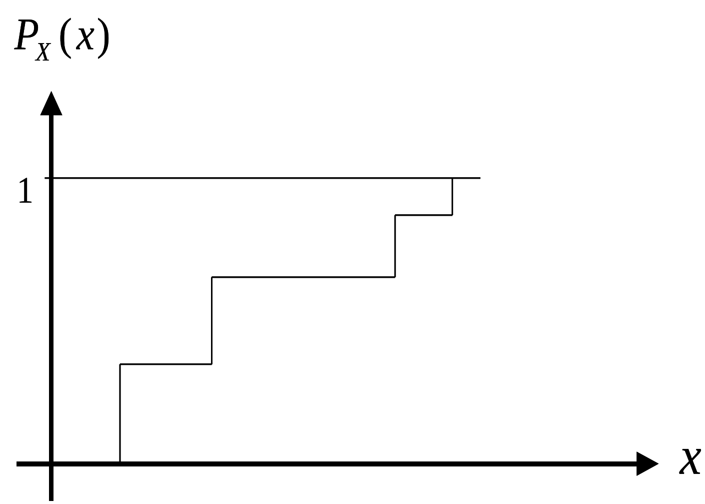
Der χ^2-Test für die Güte der Anpassung

- 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 χ^2-Test für die Güte der Anpassung

- Es sei n die Anzahl Beobachtungen einer diskreten Zufallsvariable X_j. Die Anzahl an Beobachtungen von $X_j = x_j$ ist N_j, eine poisson verteilte Zufallsvariable:

$$E\left[X_j\right] = np(x_j) = N_{p,j} \longrightarrow \text{Postulierte Häufigkeiten}$$

$$Var\left[X_j\right] = np(x_j) = N_{p,j}$$

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

Beobachtete Häufigkeiten

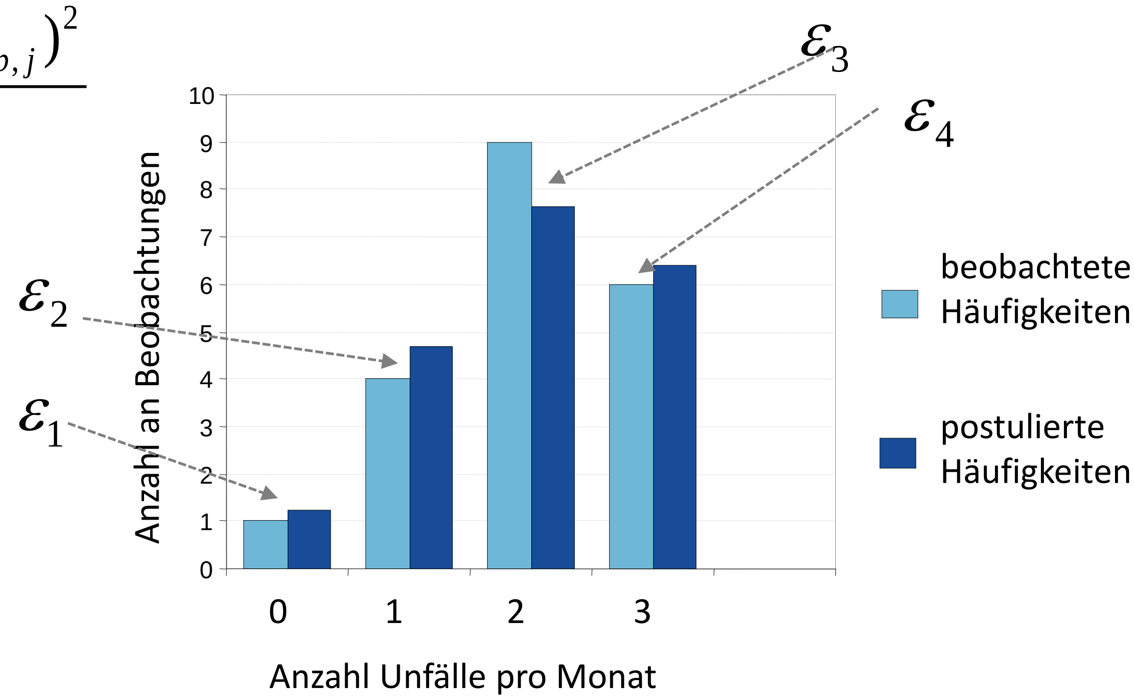
$$\varepsilon_j = \frac{N_{o,j} - N_{p,j}}{\sqrt{N_{p,j}}}$$

Modellevaluation durch Statistische Tests

Der χ^2-Test für die Güte der Anpassung

- Werden die quadrierten Differenzen der beobachteten und erwarteten Häufigkeiten summiert, dann erhalten wir:

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



χ^2 verteilt mit k-1 Freiheitsgraden

Wahrscheinlichkeitsverteilungsfunktionen

Chi-Quadrat Verteilung (χ^2- Verteilung)

- Chi-Quadrat Wahrscheinlichkeitsverteilung ist gegeben durch

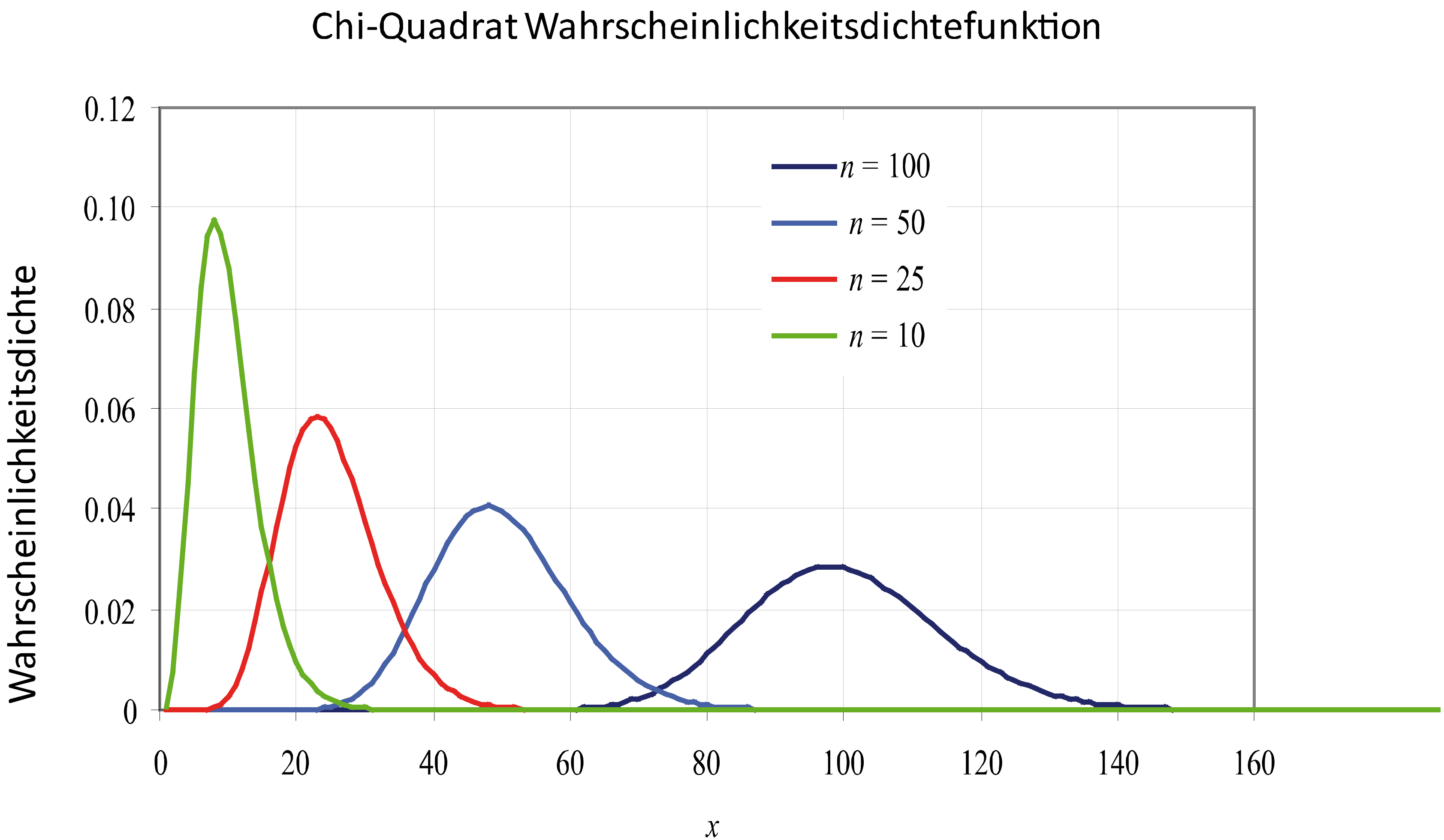
$$f_{Y_n}(y_n) = \frac{y_n^{\left(\frac{n}{2}-1\right)}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} exp\left(\frac{-y_n}{2}\right), \qquad y_n \geq 0$$

- Der Mittelwert ist $\mu_{Y_n} = n$
- Die Varianz ist $\sigma_{Y_n}^2 = 2n$

Freiheitsgrade

- $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$ ist die komplette Gamma Funktion.
- Für grosse n konvergiert die Chi-Quadrat Verteilung zu einer Normalverteilung.

Wahrscheinlichkeitsverteilungsfunktionen



Modellevaluation durch Statistische Tests

Der χ^2-Test für die Güte der Anpassung

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

  Dafür wird die Nullhypothese H_0 aufgestellt, die besagt, 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 alle anderen Verteilungen ausser der postulierten Verteilung akzeptiert werden.

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

Modellevaluation durch Statistische Tests

Der χ^2-Test für die Güte der Anpassung

- Wir betrachten folgendes Beispiel:

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

Der Mittelwert beträgt	33 MPa.
Die Standardabweichung	5 MPa.

Die Parameter werden nicht aus den vorhandenen Beobachtungen geschätzt.

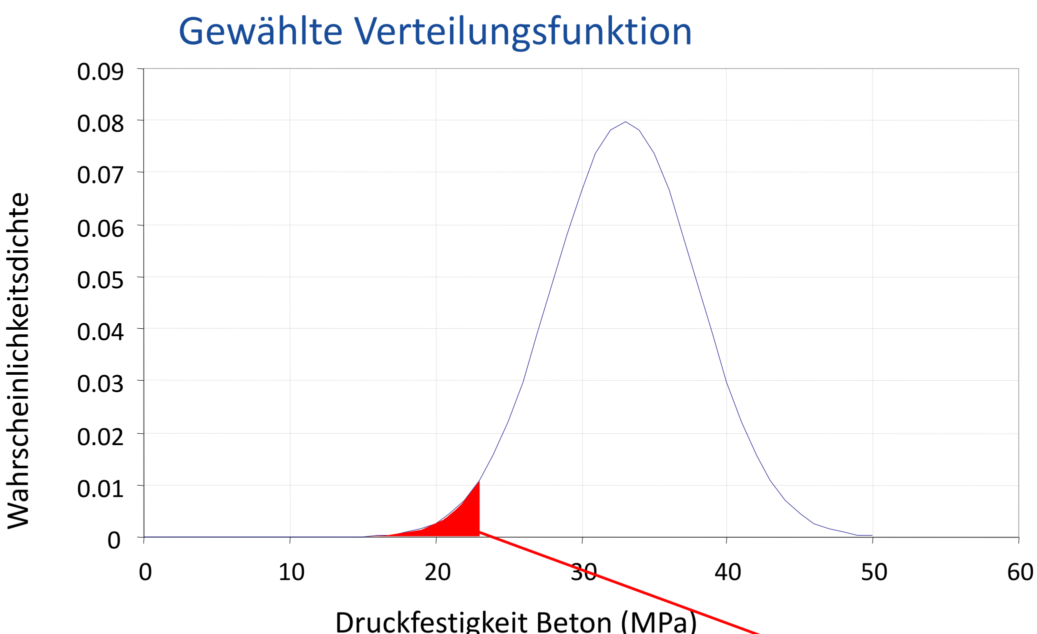
Die Normalverteilung ist eine kontinuierliche Verteilung.

Sie kann jedoch ganz einfach diskretisiert werden.

Modellevaluation durch Statistische Tests

Der χ^2-Test für die Güte der Anpassung

Die Dichtefunktion der gewählten Verteilungsfunktion wird diskretisiert:



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

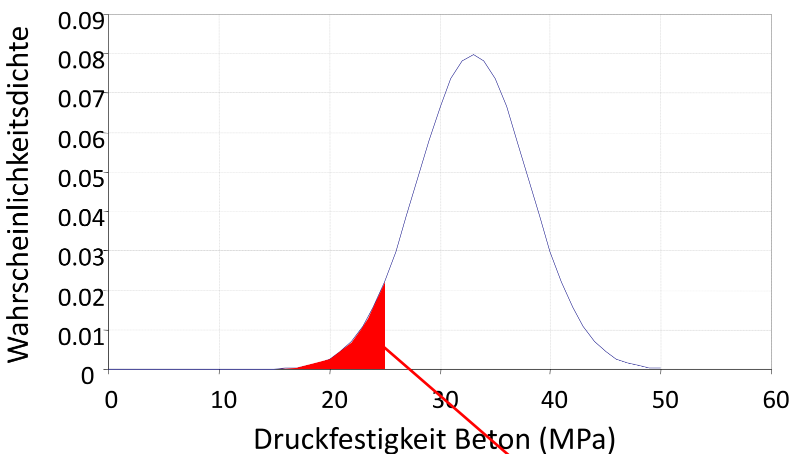
Totale Anzahl an Versuchen

Modellevaluation durch Statistische Tests

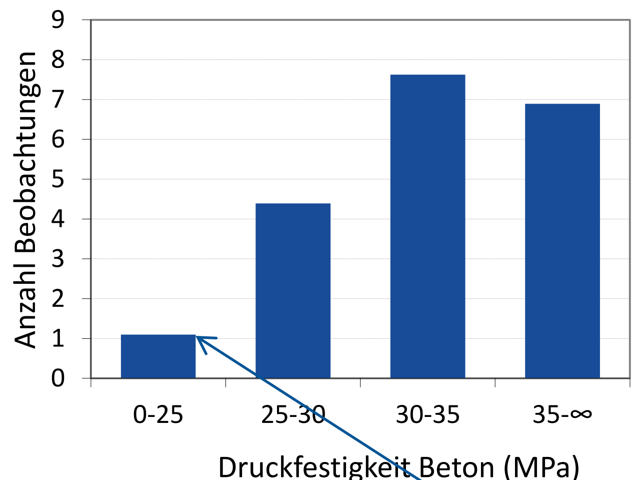
Der χ^2-Test für die Güte der Anpassung

Die Dichtefunktion der gewählten Verteilungsfunktion wird diskretisiert:

Gewählte Verteilungsfunktion



Erwartetes Histogramm



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

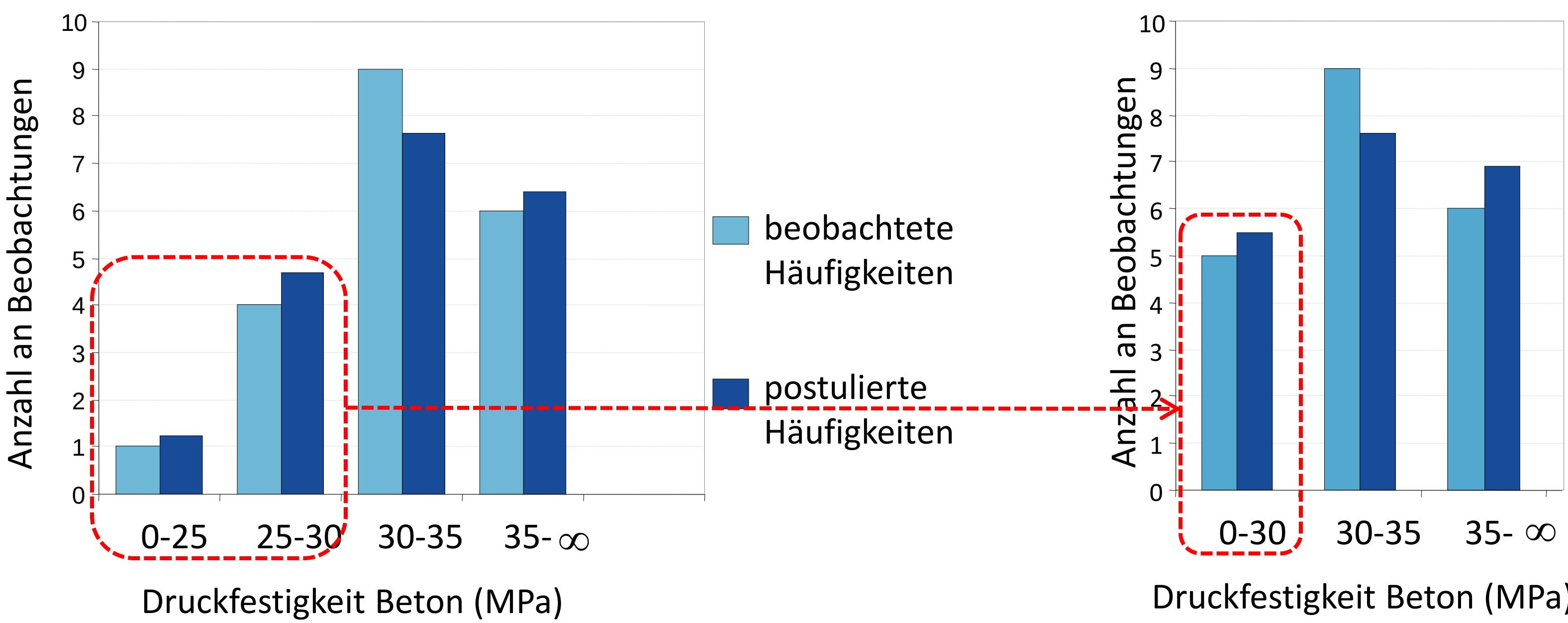
Totale Anzahl an Versuchen

Modellevaluation durch Statistische Tests

Der χ^2-Test für die Güte der Anpassung

- Die beobachteten und erwarteten Histogramme können nun verglichen werden.

Aufgrund der kleinen Anzahl an Stichproben werden die zwei unteren Intervalle zusammengeführt.



Modellevaluation durch Statistische Tests

Der χ^2-Test für die Güte der Anpassung

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

Berechnungen zum genannten Beispiel $\mu = 33, \sigma = 5$

Intervall x_j [MPa]	beobachtete Häufigkeiten $N_{o,j}$	vorausgesagte Wahrscheinlichkeiten $p(x_j)$	postulierte Häufigkeiten $N_{p,j}$	Stichproben-Statistik
0-30	5	0.2743	5.4860	0.0431
30-35	9	0.3812	7.6240	0.2483
35- ∞	6	0.3446	6.8920	0.1155
			Summe:	**0.4069**

Auf einem Signifikanzniveau von 5% erhalten wir für die χ^2-Verteilung
Mit N=3-1=2 Freiheitsgraden aus der Tabelle: $N_{o,j}$ = 5.99.
Da **0.4069** kleiner ist als 5.99, kann die Nullhypothese H_0 nicht verworfen werden.

Modellevaluation durch Statistische Tests

Der χ^2-Test für die Güte der Anpassung

- Wird einer oder mehrere Parameter der gewählten Verteilung aus dem gleichen Datensatz bestimmt, welcher auch für den Test verwendet wurde, 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 Freiheitsgrade.

Modellevaluation durch Statistische Tests

Der χ^2-Test für die Güte der Anpassung

- Wenn wir eine Normalverteilung annehmen mit den Parametern μ = 33.00 und σ = 4.05, erhalten wir folgendes Ergebnis:

Intervall x_j [MPa]	beobachtete Häufigkeiten $N_{o,j}$	vorausgesagte Wahrscheinlichkeiten $p(x_j)$	postulierte Häufigkeiten $N_{p,j}$	Stichproben-Statistik
0-30	5	0.2294	4.588507	0.036902
30-35	9	0.4599	9.197211	0.004229
35-∞	6	0.3107	6.214282	0.007389
			Summe:	**0.04852**

Auf einem Signifikanzniveau von 5% erhalten wir für die χ^2-Verteilung mit N=3-1-1=1 Freiheitsgraden aus der Tabelle: Δ= 3.84.
Da 0.04852 kleiner ist als 3.84, kann die Nullhypothese H_0 nicht verworfen werden.

Modellevaluation durch Statistische Tests

Der Kolmogorov-Smirnov-Test für die Güte der Anpassung

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

Wenn die kumulative Wahrscheinlichkeitsverteilungsfunktion der gewählten Verteilung für die Beobachtungen betrachtet wird, dann sollte die maximale Differenz zwischen der beobachteten und der postulierten kumulativen Wahrscheinlichkeitsverteilungsfunktion $\varepsilon_{\max}$ klein sein.

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

Modellevaluation durch Statistische Tests

Der Kolmogorov-Smirnov-Test für die Güte der Anpassung

- Die kumulative Wahrscheinlichkeitsverteilungsfunktion der Beobachtungen kann berechnet werden als:

$$F_o(\hat{x}_i^o) = \frac{i}{n}$$

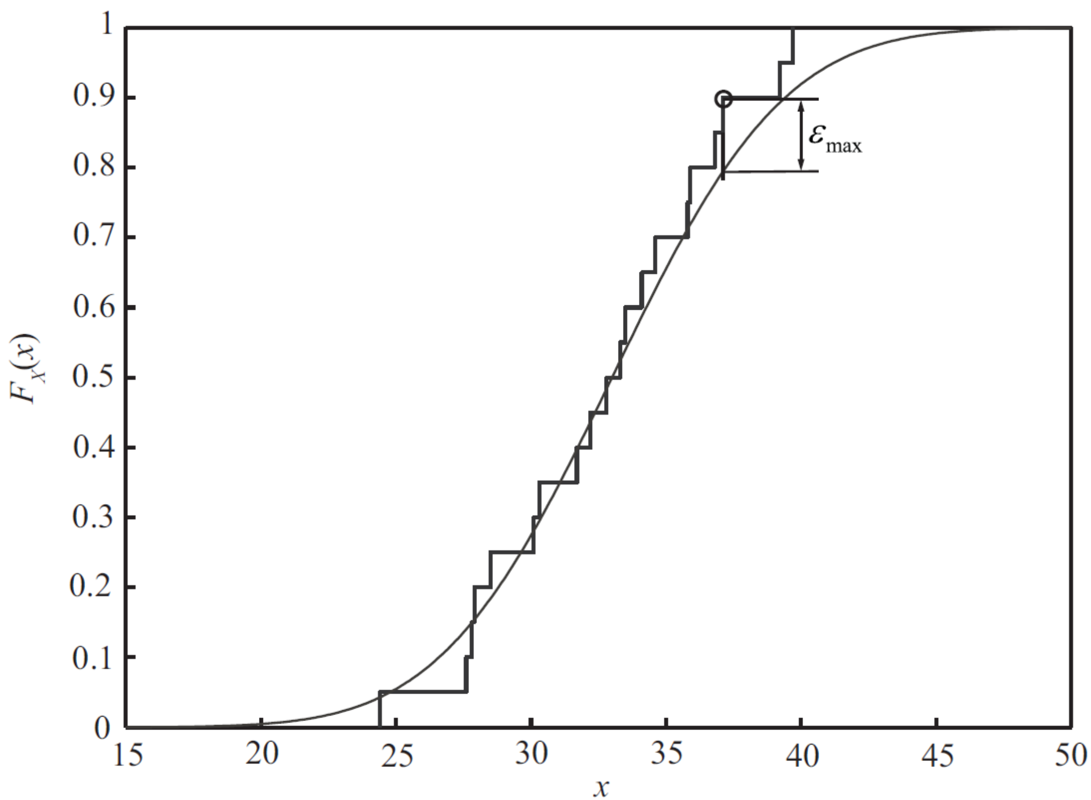
Folgende Stichprobenstatistik wird benutzt:

$$\varepsilon_{\max} = \max_{i=1}^{n}\left[\left|F_o\left(\hat{x}_i^o\right) - F_p\left(\hat{x}_i^o\right)\right|\right] = \max_{i=1}^{n}\left[\left|\frac{i}{n} - F_p\left(\hat{x}_i^o\right)\right|\right]$$

Modellevaluation durch Statistische Tests

Der Kolmogorov-Smirnov-Test für die Güte der Anpassung

- Die Kolmogorov-Smirnov-Stichprobenstatistik wird folgendermassen ermittelt:



i	x_i	$F_{xo}(x_i)$	$F_{xp}(x_i)$	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 für die 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	$\boxed{0.2941}$	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 einen bestimmten Datensatz 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

- Betrachten wir ein Beispiel mit zwei unterschiedlichen Modellen:

Modell 1: N(33;5)

Parameter nicht aus den gleichen Daten geschätzt
n=3-1=2
χ^2-Stichprobenstatistik = 0.40987
Stichproben-Likelihood = 0.8151

Modell 2: N(33;4.05)

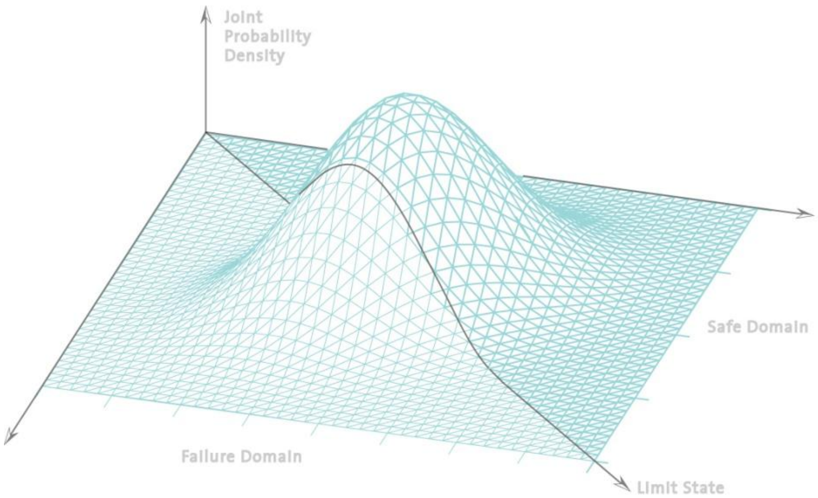
Parameter (Standardabweichung) aus den gleichen Daten geschätzt
n=3-1-1=1
χ^2-Stichprobenstatistik = 0.40683
Stichproben-Likelihood = 0.5236

Zusammenfassung

- Die Wahl eines geeigneten probabilistischen Modells kann durch Signifikanztests unterstützt werden.
- Der χ^2-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 Stichproben-Likelihoods geprüft werden.

Evaluation der Vorlesung

LV Nummer: 101-0012-00



Vielen Dank für die Aufmerksamkeit.