

STOCHASTIK FÜR INFORMATIK UND INGENIEURSWISSENSCHAFTEN

Noemi Kurt

INSTITUT FÜR MATHEMATIK
TU BERLIN

Vorlesungsskript¹

Vorläufige Version, unvollständig und fehlerhaft!

¹© Noemi Kurt
Date: 13. Juli 2016.

INHALTSVERZEICHNIS

11.	Konfidenzbereiche	3
11.1.	Grundprinzipien und Beispiele	3
11.2.	Einschub: χ^2 -Verteilung und t -Verteilung	9
11.3.	Konfidenzintervalle für Erwartungswert und Varianz	10
12.	Hypothesentests	11
12.1.	Grundprinzipien, p -Wert, Fehler erster und zweiter Art	11
12.2.	t -Test	14
12.3.	χ^2 -Test	17

11. KONFIDENZBEREICHE

Im Kapitel 10 hatten wir verschiedene Methoden kennen gelernt, einen unbekannten Parameter einer Verteilung aus Messwerten zu schätzen. Dabei haben wir auch die Frage angesprochen, wie “gut” ein Schätzer ist, denn natürlich möchte man immer erreichen, dass ein geschätzter Parameter möglichst nah am echten, unbekannten Wert ist.

Jedoch ist es auch bei Schätzern mit günstigen Eigenschaften prinzipiell immer möglich, dass die gemessenen Daten für die zugrundeliegende Verteilung “untypisch” sind, also eher extreme Ergebnisse darstellen. In solchen Fällen wird der aus den Daten berechnete Schätzer trotz guten Methoden vom echten Wert stark abweichen. Mit Hilfe von *Konfidenzbereichen* kann die Wahrscheinlichkeit von solch großen Abweichungen kontrolliert werden.

Die **Lernziele** dieses Kapitel sind:

- Das Grundprinzip von Konfidenzbereichen kennen
- Konfidenzintervalle in Beispielen berechnen können

11.1. Grundprinzipien und Beispiele. Wie bisher gehen wir von Messwerten x_1, \dots, x_n aus. Wir nehmen an, dass es sich dabei um Realisierungen von unabhängigen, identisch verteilten Zufallsvariablen X_1, \dots, X_n handelt, deren Verteilung von einem unbekannten Parameter θ abhängig ist.

Notation. In der obigen Situation bezeichnen wir mit Θ die Menge aller theoretisch möglicher Parameterwerte θ . Beispielsweise ist also $\Theta = \mathbb{R}$, falls $\theta = \mu$ der Erwartungswert einer normalverteilten Zufallsvariablen ist, oder $\Theta = [0, 1]$ falls $\theta = p$ die Erfolgswahrscheinlichkeit bei einem Bernoulli-Experiment ist.

Ein Schätzer $\bar{\theta}$ für den Parameter θ sollte also ein Element von Θ sein, bestimmt in Abhängigkeit von den Messwerten. Dabei sollte nach Möglichkeit $\bar{\theta} \approx \theta$ gelten, wobei jedoch $\bar{\theta} = \theta$ nicht zu erreichen sein wird. Die Idee hinter dem Konzept von Konfidenzbereichen ist es deshalb, eine (zufällige) Menge $J \subset \Theta$ zu finden, welche den richtigen Wert θ mit hinreichend großer Wahrscheinlichkeit enthält. Was “hinreichend groß” dabei genau bedeutet, muss vor dem Zufallsexperiment festgelegt werden, konkret wird dabei eine Zahl $\alpha \in]0, 1[$ gewählt, das sogenannte *Fehlerniveau*. Typische Werte für α sind dabei $\alpha = 0.05$ oder $\alpha = 0.01$.

Definition 11.1 (Konfidenzbereich). Seien X_1, \dots, X_n unabhängige, identisch verteilte Zufallsvariablen, deren Verteilung von einem Parameter $\theta \in \Theta$ abhängt. Sei ein Fehlerniveau $\alpha \in (0, 1)$ vorgegeben. Ein *Konfidenzbereich für den Parameter θ zum Fehlerniveau α* ist ein Intervall $J = J(X_1, \dots, X_n)$, so dass

$$\mathbb{P}_\theta(J \ni \theta) \geq 1 - \alpha \quad \text{für jedes } \theta \in \Theta$$

gilt. Dabei bezeichnet \mathbb{P}_θ die Wahrscheinlichkeit unter dem Parameterwert θ .

Hierbei ist es wichtig zu beachten, dass J (als Funktion der Zufallsvariablen X_1, \dots, X_n) zufällig ist, und nicht θ . Dies wird durch die Schreibweise $J \ni \theta$ (statt der üblicheren Variante $\theta \in J$) ausgedrückt. Insbesondere soll J nicht von θ abhängen.

Ein Konfidenzbereich ist also eine zufällige Teilmenge des Parameterbereichs, welcher den (wahren, aber unbekannten) Parameter θ mit einer Wahrscheinlichkeit von mindestens $1 - \alpha$ umfasst, egal welches $\theta \in \Theta$ tatsächlich auftritt.

Um möglichst viel Information zu enthalten, sollte J möglichst klein sein, also die Parameterwerte möglichst stark einschränken. Wie klein J sein kann, hängt jedoch von der Wahl von α ab. Je kleiner das Fehlerniveau α gewählt ist, desto größer muss J sein.

Beispiel 11.2. Seien X_1, \dots, X_n unabhängige, Bernoulli-verteilte Zufallsvariablen mit Parameter p . Wir geben uns das Fehlerniveau $\alpha = 0.01$ vor. Sei J ein Konfidenzbereich zum Fehlerniveau α . Seien x_1, \dots, x_n Realisierungen von X_1, \dots, X_n . Die Definition des Konfidenzbereichs besagt, dass mit Wahrscheinlichkeit $0.99 = 1 - \alpha$ der Bereich $J(X_1, \dots, X_n)$ den richtigen Wert p enthält.



Um einen Konfidenzbereich konkret zu bestimmen, verwendet man oft die sogenannte *Quantilmethode*.

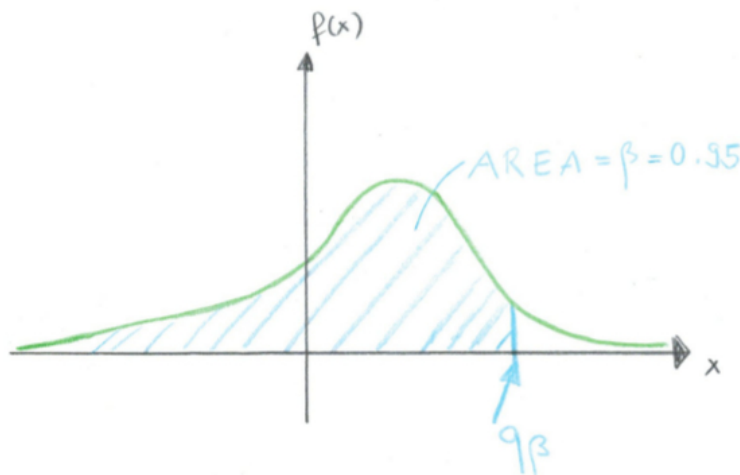
Definition 11.3 (Quantil). Sei $\beta \in [0, 1]$, und sei Y eine Zufallsvariable mit bekannter Verteilungsfunktion F . Das *Quantil* der Verteilung von Y zum Niveau β ist die kleinste Zahl q_β , für die gilt

$$F(q_\beta) = \mathbb{P}(Y \leq q_\beta) \geq \beta.$$

Die Skizze zeigt das Quantil zum Niveau $\beta = 0.95$ für eine Zufallsvariable mit einer Dichte f . Bekanntlich gilt dann die Beziehung $F(x) = \int_{-\infty}^x f(t)dt$. Das Quantil q_β ist also der kleinste Wert für x , für den $F(x) \leq \beta$ ist. Um q_β zu bestimmen, müssen wir also die Integralgleichung

$$F(q_\beta) = \int_{-\infty}^{q_\beta} f(t)dt = \beta$$

lösen. In anderen Worten muss der Flächeninhalt unter der Dichte links von q_β größer oder gleich $\beta = 0.95$ sein.



Quantile wichtiger Verteilungen werden in Tabellen angegeben, bzw. können von Statistik-Programmen wie R ausgegeben werden.

Hängt die Verteilung der Zufallsvariablen Y von einem Parameter θ ab, so schreiben wir $q_{\theta,\beta}$ für das β -Quantil.

Die Grundidee der Quantilmethode zur Bestimmung eines Konfidenzbereichs basiert auf folgender einfacher Überlegung: Für große β (z.B. $\beta = 0.95$) ist die Wahrscheinlichkeit, dass eine Realisierung einer Zufallsvariablen Y größer als q_β ist, sehr klein (nämlich $1 - \beta$.) Umgekehrt ist für kleine β die Wahrscheinlichkeit, dass eine Realisierung kleiner ist als q_β sehr klein. Falls nun die Verteilung von Y von einem Parameter $\theta \in \Theta$ abhängt, so suchen wir, in Abhängigkeit von einer Realisierung x , einen Bereich $J(x) \subset \Theta$, für den die Wahrscheinlichkeit x zu beobachten groß genug ist, nämlich mindestens $1 - \alpha$, wenn als Fehlerniveau α gewählt wurde. Somit bietet sich, je nach Problemstellung, eine der drei folgenden Möglichkeiten an. Dabei sei X eine Zufallsvariable deren Verteilung von $\theta \in \Theta$ abhängt, und sei $\alpha \in]0, 1[$. Sei x eine Realisierung von X .

- Der **obere Konfidenzbereich** zum Niveau α ist gegeben durch

$$J(x) = \{\theta \in \Theta : x \leq q_{\theta,1-\alpha}\},$$

d.h. man arbeitet mit dem *oberen* Quantil.

- Der **untere Konfidenzbereich** zum Niveau α ist gegeben durch

$$J(x) = \{\theta \in \Theta : x \geq q_{\theta,\alpha}\}$$

- Der **beidseitige Konfidenzbereich** zum Niveau α ist gegeben durch

$$J(x) = \{\theta \in \Theta : q_{\theta,\alpha/2} \leq x \leq q_{\theta,1-\alpha/2}\}$$

Aus der Definitionen der Begriffe Quantil und Konfidenzbereich folgt unmittelbar, dass jeder der drei Fälle einen Konfidenzbereich zum Fehlerniveau α definiert. Allerdings haben wir hier nur eine einzelne Zufallsvariable betrachtet, und nicht, wie das meistens geschieht, mehrere unabhängige Zufallsvariablen X_1, \dots, X_n .

Diese Idee ist nützlich um ein Konfidenzintervall für die Erfolgswahrscheinlichkeit p in einem Bernoulli-Experiment zu finden, da wir ausnutzen können, dass die Anzahl Erfolge in einem n -fach ausgeführten Bernoulli-Experiment eine binomialverteilte Zufallsvariable ist. Somit stecken hier in einer einzigen binomialverteilten Zufallsvariable, zumindest bei hinreichend großem n , implizit n unabhängige Bernoulli-Variablen mit Parameter p . Wir wenden nun die Methode in zwei Beispielen an.

Beispiel 11.4 (Binomialverteilung, einseitiges Konfidenzintervall). Bei einer Produktion eines Massenartikels wird eine gewisse Menge an Ausschuss produziert. Wir suchen ein Konfidenzintervall für die unbekannte Wahrscheinlichkeit p , mit der ein zufällig ausgewähltes Stück Ausschuss ist. Dafür testen wir n zufällig ausgewählte Stücke, wobei wir entweder mit Zurücklegen ziehen, oder eine kleine Stichprobe aus einer sehr großen Gesamtheit ziehen. Die Anzahl X an Ausschuss ist somit (zumindest annähernd) binomialverteilt mit Parametern n und p . Wenn in unserem Probelauf k Ausschuss-Stücke enthalten sind, so erhalten wir einen Schätzer für p , nämlich $\bar{p} = \frac{k}{n}$. Nun suchen wir jedoch einen Konfidenzbereich. Es gilt $\Theta = [0, 1]$, da dies die Werte sind, die p annehmen kann.

Fixiere das Fehlerniveau α . Nun müssen wir uns zuerst überlegen, welche der drei Varianten für den Konfidenzbereich wir verwenden wollen. Dazu müssen wir uns klar machen, welche Information wir aus dem Konfidenzbereich ziehen wollen. Da p die Wahrscheinlichkeit für Ausschuss ist, möchte man ein möglichst kleines p haben. Somit suchen wir eine obere Schranke für den Konfidenzbereich, denn man will sich dagegen absichern, versehentlich mit einem zu großen p zu rechnen. Wir benutzen also den zweiten Fall aus obiger Liste, denn unteren Konfidenzbereich, denn je größer p ist, desto größer ist $q_{p,\alpha}$ bzw. $q_{p,1-\alpha}$.

Wir erhalten also für $k \in \{0, \dots, n\}$

$$\begin{aligned} J(k) &= \{p \in [0, 1] : k \geq q_{p,\alpha}\} \\ &= \{p \in [0, 1] : \mathbb{P}_p(X \leq k) \geq \mathbb{P}_p(X \leq q_{p,\alpha}) = \alpha\} \\ &= \left\{p \in [0, 1] : \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i} \geq \alpha\right\}. \end{aligned}$$

Falls wir also $k = 0$ Stück Ausschuss in unserem Test gezählt haben, so erhalten wir

$$J(0) = \{p \in [0, 1] : (1-p)^n \geq \alpha\} = [0, 1 - \alpha^{1/n}].$$

Für andere Werte von k ist das Auflösen von $\sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i} \geq \alpha$ nicht mehr so leicht möglich. Zur expliziten Bestimmung des Konfidenzbereichs kommen deshalb entweder numerische Methoden zur Anwendung, oder man verwendet die Approximation der Binomialverteilung durch die Normalverteilung (Satz 7.6), der besagt dass

$$\mathbb{P}_p(X \leq k) \approx \Phi_{0,1}\left(\frac{k + 1/2 - np}{\sqrt{np(1-p)}}\right)$$

ist. Sei z_α das α -Quantil der Standardnormalverteilung. Dieses kann aus der Tabelle der Normalverteilungsquantile abgelesen werden. Nach unseren Überlegungen muss also gelten

$$\Phi_{0,1}\left(\frac{k + 1/2 - np}{\sqrt{np(1-p)}}\right) \geq \alpha,$$

d.h.

$$\frac{k + 1/2 - np}{\sqrt{np(1-p)}} \geq z_\alpha.$$

Auflösen nach p ergibt die obere Schranke.

Beispiel 11.5 (Binomialverteilung, zweiseitiges Konfidenzintervall). **siehe Vorlesung**

Die Grundidee der Quantilmethode ist auch in vielen anderen Fällen nützlich. Wir betrachten ein weiteres Beispiel.

Beispiel 11.6 (Konfidenzintervall für den Erwartungswert der Normalverteilung bei bekannter Varianz). Seien x_1, \dots, x_n Realisierungen von unabhängigen, identisch verteilten Zufallsvariablen X_1, \dots, X_n , welche einer Normalverteilung mit bekannter Varianz σ^2 und unbekanntem Erwartungswert μ folgen. Wir wählen das Fehlniveau $\alpha = 0.05$. Dazu suchen wir ein Konfidenzintervall für μ , d.h. $J \subset \mathbb{R}$ so dass

$$\mathbb{P}_\mu(\mu \in J) \geq 1 - \alpha = 0.95 \quad \text{für alle } \mu \in \mathbb{R}.$$

Wir wählen dazu einen Ansatz: J sei um das empirische Mittel zentriert, d.h. es gibt $h > 0$ so dass

$$J = [\bar{\mu} - h, \bar{\mu} + h]$$

gilt, mit $\bar{\mu} = \bar{\mu}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$. Da $\bar{\mu}$ einfach berechnet werden kann, müssen wir also $h = h(\alpha, n, \sigma^2)$ so bestimmen, dass die Bedingung für einen Konfidenzbereich zum Fehlniveau α erfüllt ist. Wir wollen also

$$\begin{aligned} 1 - \alpha &\leq \mathbb{P}_\mu\left(\mu \in \left[\frac{1}{n} \sum_{i=1}^n X_i - h, \frac{1}{n} \sum_{i=1}^n X_i + h\right]\right) \\ &= \mathbb{P}_\mu\left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu) \in [-h, h]\right) \\ &= \mathbb{P}_\mu\left(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \in \left[\frac{-h}{\sigma\sqrt{n}}, \frac{h}{\sigma\sqrt{n}}\right]\right). \end{aligned}$$

Nun ist nach Satz muss in Kapitel zur Normalverteilung noch rein!!! die Zufallsvariable

$$Y := \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (X_i - \mu)$$

standardnormalverteilt, da die X_i unabhängig und $\mathcal{N}(\mu, \sigma^2)$ -verteilt sind. Somit müssen wir h so wählen, dass für $Y \sim \mathcal{N}(0, 1)$ gilt

$$\mathbb{P}\left(Y \in \left[\frac{-h}{\sigma\sqrt{n}}, \frac{h}{\sigma\sqrt{n}}\right]\right) \geq 1 - \alpha,$$

beziehungsweise, was wegen der Symmetrie der Standardnormalverteilung äquivalent ist,

$$\mathbb{P}\left(Y \geq \frac{h}{\sigma\sqrt{n}}\right) \leq \frac{\alpha}{2}.$$

Letzteres besagt aber genau, dass $\frac{h}{\sigma\sqrt{n}}$ gleich dem Quantil der Standardnormalverteilung zum Wert $1 - \frac{\alpha}{2}$ sein muss. In anderen Worten haben wir den folgenden Satz bewiesen:

Satz 11.7 (Konfidenzintervall für den Erwartungswert einer Normalverteilung). *Das Konfidenzintervall zum Fehlerniveau α für den Erwartungswert μ einer Normalverteilung bei bekannter Varianz σ^2 und n unabhängigen Messungen ist gegeben durch*

$$J = J(n, \alpha, \sigma) = \left[\bar{\mu} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{\mu} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$$

wobei $z_{1-\alpha/2}$ das Quantil der **Standardnormalverteilung** ist, und $\bar{\mu}$ das empirische Mittel.

Wenn man sich die Herleitung im letzten Beispiel genauer anschaut, so stellt man fest, dass die Normalverteilung nur an einer Stelle benutzt wurde, um zu argumentieren, dass Y standardnormalverteilt ist. Nun hat aber Y (bis auf den konstanten Vorfaktor $\frac{1}{\sigma}$) genau die Form der Zufallsvariable, die im zentralen Grenzwertsatz vorkommt. Da der zentrale Grenzwertsatz für Folgen von unabhängigen, identisch verteilten Zufallsvariablen $X_i, i \in \mathbb{N}$ mit beliebiger Verteilung gilt (sofern die Varianz existiert), erhält man als Konfidenzintervall für den Erwartungswert bei bekannter Varianz, zumindest im Grenzwert für $n \rightarrow \infty$, das gleiche Ergebnis wie für die Normalverteilung.

Satz 11.8 (Konfidenzintervall für den Erwartungswert bei bekannter Varianz). *Seien X_1, \dots, X_n unabhängige, identisch verteilte Zufallsvariablen mit existierender (und bekannter) Varianz $\sigma^2 > 0$. Falls n hinreichend groß ist, ist das Konfidenzintervall zum Fehlerniveau α für den Erwartungswert μ gegeben durch*

$$J = J(n, \alpha, \sigma) = \left[\bar{\mu} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{\mu} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$$

wobei $z_{1-\alpha/2}$ das Quantil der **Standardnormalverteilung** ist, und $\bar{\mu}$ das empirische Mittel.

Wir wollen in dieser Vorlesung noch zwei weitere wichtige Konfidenzintervalle berechnen, nämlich

- Konfidenzintervall für die Varianz bei bekanntem Erwartungswert, und
- Konfidenzintervall für den Erwartungswert bei unbekannter Varianz.

Dafür benötigen wir jedoch noch zwei Verteilungen die wir noch nicht kennen, die jedoch auch in Kapitel 12 für das Testen von Hypothesen relevant werden.

11.2. Einschub: χ^2 -Verteilung und t -Verteilung.

Definition 11.9. Eine Zufallsvariable X ist χ^2 -verteilt, bzw. folgt einer χ^2 -Verteilung mit Parameter $n \in \mathbb{N}$, wenn sie die Dichte

$$f_n(x) = \begin{cases} \frac{x^{n/2-1} e^{-x/2}}{2^{n/2} \Gamma(n/2)} & \text{falls } x \geq 0 \\ 0 & \text{falls } x < 0 \end{cases}$$

hat. Dabei ist Γ die sogenannte Gamma-Funktion, welche für Werte der Form $n/2, n \in \mathbb{N}$ rekursiv aus $\Gamma(\frac{1}{2}) = \sqrt{\pi}, \Gamma(1) = 1$ und $\Gamma(r+1) = r\Gamma(r)$ für alle $r \in [0, \infty[$ berechnet werden kann.

Die genaue Form dieser Dichte ist für unsere Zwecke nicht so wichtig, Sie brauchen sie sich entsprechend nicht einzuprägen. Für uns werden die Quantile der χ^2 -Verteilung relevant sein, welche wiederum in Tabellen aufgelistet werden. Die Bedeutung der χ^2 -Verteilung für die Statistik liegt in folgendem Satz:

Satz 11.10. Seien Y_1, \dots, Y_n unabhängige standardnormalverteilte Zufallsvariablen. Dann folgt die Zufallsvariable

$$Y_1^2 + \dots + Y_n^2$$

einer χ^2 -Verteilung mit Parameter n .

(ohne Beweis)

Definition 11.11. Eine Zufallsvariable X heißt t -verteilt zum Parameter $n \in \mathbb{N}$, falls sie die Dichte

$$f_n(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi} \Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

Hat. Dabei ist Γ wieder die Gamma-Funktion.

Auch hier ist weniger die genaue Form der Dichte relevant, sondern die tabellierten Quantile. Die t -Verteilung tritt in folgender Form auf:

Satz 11.12. Seien X_1, \dots, X_n unabhängige, $\mathcal{N}(\mu, \sigma^2)$ -verteilte Zufallsvariablen. Seien $\bar{\mu}_n(X_1, \dots, X_n)$ und $\bar{\sigma}_n(X_1, \dots, X_n)$ das empirische Mittel bzw. die empirische Standardabweichung dieser Zufallsvariablen. Dann folgt die Zufallsvariable

$$\frac{\sqrt{n}(\bar{\mu}_n(X_1, \dots, X_n) - \mu)}{\bar{\sigma}_n(X_1, \dots, X_n)}$$

einer t -Verteilung mit Parameter $n-1$.

(ohne Beweis)

Die empirische Standardabweichung ist dabei die Wurzel aus der empirischen Varianz, vgl. Kapitel 10. Zu beachten ist, dass der Parameter der t -Verteilung tatsächlich $n-1$ ist, und nicht etwa n , obwohl wir n Zufallsvariablen X_1, \dots, X_n betrachten.

11.3. Konfidenzintervalle für Erwartungswert und Varianz. In den Sätzen 11.7 und 11.8 haben wir gesehen, wie das Konfidenzintervall für den Erwartungswert bei bekannter Varianz bestimmt wird. In diesem Abschnitt betrachten wir wie angekündigt zwei weitere Situationen.

Theorem 11.13 (Konfidenzintervall für die Varianz der Normalverteilung bei bekanntem Erwartungswert). Seien x_1, \dots, x_n Messwerte von Realisierungen von unabhängigen normalverteilten Zufallsvariablen X_1, \dots, X_n mit bekanntem Erwartungswert μ und unbekannter Varianz σ^2 . Das Konfidenzintervall für σ^2 erhält man als

$$J = \left[\bar{\sigma}^2 \frac{n-1}{\chi_{n-1, 1-\alpha/2}^2}, \bar{\sigma}^2 \frac{n-1}{\chi_{n-1, \alpha/2}^2} \right]$$

wobei $\bar{\sigma}^2$ die empirische Varianz ist, und $\chi_{n-1, \beta}^2$ das β -Quantil der χ^2 -Verteilung mit Parameter $n-1$.

Beweis. Der Beweis verläuft ähnlich wie der für Satz 11.7, unter Verwendung von Satz 11.10. Details siehe Vorlesung. \square

Für den Fall dass die X_1, \dots, X_n eine beliebige Verteilung haben, also nicht zwingend normalverteilt sind, können wir uns im Fall von unbekannter Varianz nicht einfach auf den zentralen Grenzwertsatz zur Normalapproximation berufen. Hier sind andere Methoden notwendig, die den Rahmen dieser Vorlesung sprengen würden.

Kennt man weder Erwartungswert noch Varianz, so kann man im Fall von normalverteilten Zufallsvariablen dennoch ein Konfidenzintervall für den Erwartungswert angeben.

Theorem 11.14 (Konfidenzintervall für den Erwartungswert der Normalverteilung bei unbekannter Varianz). Seien x_1, \dots, x_n Messwerte von Realisierungen von unabhängigen normalverteilten Zufallsvariablen X_1, \dots, X_n mit unbekanntem Erwartungswert μ und unbekannter Varianz σ^2 . Das Konfidenzintervall für μ erhält man als

$$J = \left[\bar{\mu} - \frac{\bar{\sigma}}{\sqrt{n}} t_{n-1, 1-\alpha/2}, \bar{\mu} + \frac{\bar{\sigma}}{\sqrt{n}} t_{n-1, 1-\alpha/2} \right]$$

wobei $\bar{\mu}$ das empirische Mittel, $\bar{\sigma}$ die empirische Standardabweichung, und $t_{n-1, 1-\alpha/2}$ das Quantil der t -Verteilung mit Parameter $n-1$ ist.

Proof. Wiederum folgt der Beweis dem Vorgehen von Satz 11.7, unter Verwendung von Satz 11.12. \square

12. HYPOTHESENTESTS

Hypothesentests gehören zu den Grundproblemen der Statistik. Sie beruhen auf folgender Idee: Vor der Messung der Daten wird eine Hypothese betreffend der zugrundeliegenden Zufallsvariablen aufgestellt. Mögliche Hypothesen können sein:

- Die Daten folgen einer Normalverteilung
- Der Erwartungswert ist 2
- Die Realisierungen sind unabhängig

Nach der Messung der Daten wird durch Rechnung überprüft, ob die gemessenen Daten der Hypothese widersprechen, d.h. ob unter der Annahme, dass die Hypothese gilt, die Daten sehr unwahrscheinlich sind. Sind die Messwerte sehr unwahrscheinlich, wird die Hypothese verworfen, andernfalls wird sie angenommen.

Die **Lernziele** dieses Kapitel sind:

- Die Grundprinzipien von Hypothesentests kennen
- t -Test und χ^2 -Test ausführen und interpretieren können

12.1. Grundprinzipien, p -Wert, Fehler erster und zweiter Art. Beim Testen von Hypothesen wird entsprechend der in der Einleitung beschriebenen Grundidee wie folgt vorgegangen:

Vorgehen bei Hypothesentests

1. Formuliere die *Nullhypothese* H_0 .
2. Konstruiere den *Annahmebereich* (bzw. den Ablehnungsbereich).
3. Führe das Experiment durch, erhalte Daten x_1, \dots, x_n .
4. Überprüfe, ob die Daten im Annahmebereich liegen.

Je nach Ergebnis bei 4. nimmt man die Hypothese an oder verwirft sie:

- Falls in 4. das Ergebnis *nein* lautet: Nullhypothese verwerfen
- Andernfalls: Nullhypothese annehmen

Definition 12.1 (Nullhypothese, Annahmebereich, Fehlerniveau). Seien X_1, \dots, X_n Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) . Seien $x_1, \dots, x_n \in \mathbb{R}$ Messwerte. Eine *Nullhypothese* über X_1, \dots, X_n ist eine Aussage, deren Wahrheitsgehalt nur von der gemeinsamen Verteilung von X_1, \dots, X_n abhängt, und für die $\{H_0 \text{ ist wahr}\}$ ein Ereignis ist. Die *Alternativhypothese* H_A ist die Negation von H_0 . Ein *Annahmebereich* zum *Fehlerniveau* $\alpha \in [0, 1]$ für die Nullhypothese H_0 ist eine zufällige Teilmenge A von \mathbb{R}^n , so dass

$$\mathbb{P}((x_1, \dots, x_n) \in A \mid H_0 \text{ ist wahr}) \geq 1 - \alpha$$

ist.

Grundsätzlich kann jede Aussage über die Daten x_1, \dots, x_n bzw. ihre zugrundeliegenden Zufallsvariablen eine *Nullhypothese* sein. Wir werden etliche der gängigsten Nullhypothesen in Beispielen kennenlernen. Ein Annahmebereich ist a priori einfach eine Teilmenge von

\mathbb{R}^n , wenn n Messwerte x_1, \dots, x_n gemessen wurden. Er wird je nach Situation auf Grund von theoretischen Überlegungen konstruiert. Zum vorgegebenen Fehlerniveau α soll er also die Eigenschaft haben, dass *falls* die Nullhypothese erfüllt ist, die Messwerte mit einer Wahrscheinlichkeit von mindestens $1 - \alpha$ im Annahmebereich liegen

Beispiel 12.2 (Materialbelastung). Es soll ein Stabilitäts-Belastungstest durchgeführt werden, um die Frage zu klären, wie viel Gewicht ein Bauteil (in Erwartung) tragen kann, ohne dass es zu stark verformt wird, es wird angenommen dass das im Mittel 12 kg sind. Die mathematische Modellierung nimmt an, dass das zu tragende Gewicht (in kg) durch eine Zufallsvariable X mit Werten in $[0, \infty)$ gegeben ist, bzw. für n baugleiche Teile die jeweiligen Gewichte durch unabhängige, identisch verteilte Zufallsvariablen X_1, \dots, X_n mit Verteilungsfunktion F_X gegeben sind. In einer Messreihe wird die Belastung von 20 baugleichen Teilen gemessen, d.h. Daten x_1, \dots, x_{20} als unabhängige Realisierungen der Zufallsvariablen X ermittelt. Weiter nehmen wir an, dass $\sigma^2 = \mathbb{V}(X) = 2$ ist. Wir führen nach obigem Vorgehen einen Hypothesentest durch.

1. Nullhypothese H_0 : Der Erwartungswert μ von X ist gleich $\mu_0 := 12$ (kg).
2. Annahmebereich: Hier gibt es verschiedene Möglichkeiten. Zuerst wählen eine Fehlerschranke, z.B. 5%, also $\alpha = 0.05$. Wir suchen nun einen Annahmebereich $A \subset \mathbb{R}^{20}$, für den

$$\mathbb{P}_{\mu_0}((x_1, \dots, x_{20}) \in A) \geq 1 - \alpha$$

gilt. Das erinnert an das Vorgehen bei Konfidenzintervallen: Falls die Nullhypothese gilt, falls also μ_0 der richtige Erwartungswert ist, soll das gemessene empirische Mittel $\bar{\mu}$ mit einer Wahrscheinlichkeit von mindestens $1 - \alpha$ im Annahmebereich A liegen, d.h. die Wahrscheinlichkeit, dass wir die Nullhypothese verwerfen, obwohl sie wahr ist, soll höchstens α sein. Wir können uns weiter an Konfidenzintervallen orientieren, und als Annahmebereich die folgende Teilmenge von \mathbb{R}^{20} wählen:

$$A := \{x = (x_1, \dots, x_{20}) : \bar{\mu}(x) \in [\mu_0 - h, \mu_0 + h]\},$$

wobei $\bar{\mu}$ wie bisher das empirische Mittel der Daten bezeichnet, und h geeignet gewählt werden muss. Mit dieser Form von A gilt

$$\mathbb{P}_{\mu_0}(x \in A) = \mathbb{P}(\mu_0 \in J(x))$$

wobei $J(x)$ das Konfidenzintervall für den Erwartungswert bei bekannter Varianz für das Fehlerniveau α ist. Aus Satz 11.8 wissen wir, dass

$$J = [\bar{\mu} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{\mu} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}]$$

ist, dass wir also

$$h = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} = z_{0.975} \frac{\sqrt{2}}{\sqrt{20}} \approx 0.62$$

wählen sollten. Somit erhalten wir als Annahmebereich

$$A = \{x = (x_1, \dots, x_{20}) : \bar{\mu}(x) \in [11.38, 12.62]\}.$$

3. Experiment durchführen ergibt die Daten $x = (x_1, \dots, x_{20}) = (9.50, 11.01, 12.97, 12.29, 12.34, 10.80, 13.51, 13.65, 13.01, 11.16, 14.77, 8.87, 11.15, 7.56, 12.77, 12.71, 12.10, 14.24, 9.35, 15.10)$
Das empirische Mittel ist dann $\bar{\mu}(x) = 11.94$.
4. Da $11.94 \in [11.38, 12.62]$ liegt, folgt somit $x \in A$, also wird die Nullhypothese angenommen, bzw. sie wird nicht verworfen – die Daten sprechen also nicht gegen die Wahrheit der Nullhypothese.

Die Konstruktion des Annahmebereiches im obigen Beispiel erfüllt das Grundprinzip, welches für Annahmebereiche bei Hypothesentests immer gelten soll: Falls *unter Annahme der Gültigkeit Nullhypothese* die beobachteten Daten *sehr unwahrscheinlich* sind, so verwirft man die Nullhypothese. Im obigen Beispiel wurde deshalb mit der Wahrscheinlichkeit \mathbb{P}_0 gerechnet, d.h. es wurde angenommen, dass die Nullhypothese “der Erwartungswert ist gleich μ_0 ” gilt. Das Kriterium “sehr unwahrscheinlich” wurde mit Hilfe des Konfidenzintervalls überprüft. Jedoch benötigt man im Allgemeinen eine etwas konkretere Vorstellung davon, was unter “sehr unwahrscheinlich” genau zu verstehen ist. Deshalb führen wir nun einige weitere Begriffe ein.

Definition 12.3 (Fehler 1. und 2. Art). Sei H_0 eine Nullhypothese, und A ein Annahmebereich. Dann sind zwei Typen von “Fehlern” möglich:

- Der *Fehler 1. Art*: Die Nullhypothese wird verworfen, obwohl sie eigentlich richtig wäre
- Der *Fehler 2. Art*: Die Nullhypothese wird angenommen, obwohl sie eigentlich falsch ist

Satz 12.4. Sei H_0 eine Nullhypothese, und A ein Annahmebereich zum Fehlerniveau $\alpha \in [0, 1]$. Dann ist die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen, kleiner oder gleich α .

Beweis. Die Wahrscheinlichkeit für einen Fehler 1. Art ist nach Definition

$$\mathbb{P}(H_0 \text{ wird abgelehnt} \mid H_0 \text{ ist wahr}).$$

Andererseits gilt nach Definition des Annahmebereichs

$$\mathbb{P}((x_1, \dots, x_n) \in A \mid H_0 \text{ ist wahr}) \geq 1 - \alpha.$$

Da gemäß unserem Vorgehen die Nullhypothese genau dann angenommen wird, wenn $(x_1, \dots, x_n) \in A$ ist, folgt

$$\mathbb{P}(H_0 \text{ wird abgelehnt} \mid H_0 \text{ ist wahr}) = 1 - \mathbb{P}((x_1, \dots, x_n) \in A \mid H_0 \text{ ist wahr}) \leq \alpha.$$

□

Man kann beweisen (aber das würde den Rahmen dieser Vorlesung sprengen), dass *je kleiner die Wahrscheinlichkeit für einen Fehler 1. Art ist, desto größer die Wahrscheinlichkeit für einen Fehler 2. Art ist*. Im weiteren kann im der Fehler 2. Art (normalerweise) nicht explizit quantifiziert werden, auch wenn er durch das Testdesign beeinflusst werden

kann, beispielsweise durch Wahl einer größeren Stichprobe. Deshalb ist es für die Konstruktion eines statistischen Tests eminent wichtig, wie die Nullhypothese H_0 gewählt wird, da die Rollen von Nullhypothese und Alternativhypothese nicht vertauscht werden können, und nur einer der Fehler, nämlich der Fehler 1. Art, kontrolliert werden kann, nämlich durch Wahl des Fehlerniveaus α . Wenn man α sehr klein wählt, also den Fehler 1. Art sehr unwahrscheinlich macht, hat man im Gegenzug *zwingend* eine große Unsicherheit hinsichtlich der Wahrscheinlichkeit für einen Fehler 2. Art.

Beispiel 12.5.

Ein weiterer Begriff dient der Quantifizierung des Begriffs “sehr unwahrscheinlich”.

Definition 12.6 (*p*-Wert). *p*-Wert: Seien X_1, \dots, X_n Zufallsvariablen, und $x_1, \dots, x_n \in \mathbb{R}$ Daten. Sei H_0 eine Nullhypothese diese Daten betreffend. Der *rechtsseitige p-Wert* ist definiert als

$$p = p(x_1, \dots, x_n) = \mathbb{P}(X_i \geq x_i, i = 1, \dots, n \mid H_0 \text{ ist wahr}),$$

der *linksseitige p-Wert* als

$$p = p(x_1, \dots, x_n) = \mathbb{P}(X_i \leq x_i, i = 1, \dots, n \mid H_0 \text{ ist wahr}),$$

und der *beidseitige p-Wert* als

$$p = 2 \min\{\mathbb{P}(X_i \geq x_i, i = 1, \dots, n \mid H_0 \text{ ist wahr}), \mathbb{P}(X_i \leq x_i, i = 1, \dots, n \mid H_0 \text{ ist wahr})\}.$$

Der *p*-Wert ist also die Wahrscheinlichkeit, dass das Tupel von Zufallsvariablen (X_1, \dots, X_n) Werte “gleich oder extremer als (x_1, \dots, x_n) ” annimmt, unter der Annahme dass die Nullhypothese gilt. Dabei bedeutet “extremer” je nach Situation “größer”, “kleiner” oder “größer oder kleiner”. Der *p*-Wert hängt also mit dem Quantil zusammen, welches durch die Daten x_1, \dots, x_n begrenzt wird.

Skizze

Satz 12.7. Sei $p(x_1, \dots, x_n)$ ein *p*-Wert zu den Daten $x = (x_1, \dots, x_n)$. Dann ist

$$A_\alpha := \{x : p(x) \geq \alpha\}$$

ein Annahmebereich zum Fehlerniveau α .

Proof. Nach Definition gilt $\mathbb{P}(p(X) < \alpha \mid H_0) \leq \alpha$. □

Wird mit dieser Form von Annahmebereichen gearbeitet, so wird H_0 angenommen wenn $p \geq \alpha$ ist, und abgelehnt wenn $p < \alpha$ ist.

12.2. *t*-Test. Der *t*-Test ist ein Hypothesentest, der auf der *t*-Verteilung (Kapitel 11) beruht. Er wird verwendet, wenn folgende Annahmen erfüllt sind:

Annahme 12.8 (Annahme beim *t*-Test). Gegeben Daten x_1, \dots, x_n , von denen angenommen werden kann, dass sie Realisierungen von unabhängigen, identisch verteilten Zufallsvariablen X_1, \dots, X_n sind, welche außerdem (annähernd) normalverteilt sind (z.B. weil n groß ist und der zentrale Grenzwertsatz gilt, insbesondere muss die Varianz existieren).

Dann können mit Hilfe des t -Tests die folgenden Hypothesen getestet werden:

- **Zweiseitiger t -Test**: Test ob der Erwartungswert μ der X_i gleich einer gegebenen Zahl μ_0 ist, d.h. Nullhypothese

$$H_0 : \mu = \mu_0$$

- **Einseitiger t -Test**: Test ob der Erwartungswert μ der X_i größer (kleiner) als eine gegebenen Zahl μ_0 ist, d.h.

$$H_0 : \mu \geq \mu_0$$

bzw.

$$H_0 : \mu \leq \mu_0.$$

Außerdem kann mit einer Variante des t -Test getestet werden, ob zwei unabhängige Zufallsvariablen denselben Erwartungswert haben.

Die theoretische Grundlage für den t -Test bildet Satz 11.12, welcher besagt dass unter der Annahme 12.8 die Zufallsvariable

$$T := \sqrt{n} \cdot \frac{\bar{\mu}(X_1, \dots, X_n) - \mathbb{E}[X_1]}{\bar{\sigma}_n(X_1, \dots, X_n)}$$

(annähernd) t -verteilt mit Parameter $n-1$ ist. Um den t -Test durchzuführen, wird deshalb die **Testgröße**

$$t := \sqrt{n} \cdot \frac{\bar{\mu}(x_1, \dots, x_n) - \mu_0}{\bar{\sigma}_n(x_1, \dots, x_n)}$$

in Abhängigkeit von den gemessenen Daten bestimmt. Ist dieser Wert unter der t -Verteilung *sehr unwahrscheinlich* (was durch das Fehlniveau α quantifiziert wird), so wird die Nullhypothese verworfen.

Vorgehen 12.9 (Zweiseitiger t -Test). Die Annahme 12.8 sei erfüllt. Der zweiseitige t -Test läuft folgendermaßen ab:

1. Stelle die Nullhypothese auf: $H_0: \mu = \mu_0$
2. Wähle Fehlniveau $\alpha \in (0, 1)$
3. Erhebe die Daten x_1, \dots, x_n .
4. Bestimme den sogenannten *Freiheitsgrad* f : Im t -Test ist $f = n - 1$
5. Bestimme $t_{1-\alpha/2, f}$, das $1 - \alpha/2$ -Quantil der t -Verteilung zum Parameter f (aus der Tabelle der Quantile der t -Verteilung, oder mit einem geeigneten Programm)
6. Berechne $\bar{\mu}$ und $\bar{\sigma}_n$ aus den Daten
7. Berechne den Testwert

$$t = \sqrt{n} \cdot \frac{\bar{\mu} - \mu_0}{\bar{\sigma}_n}$$

8. Vergleiche den Testwert t mit dem Quantil $t_{1-\alpha/2, f}$. Entscheidungsregel:
 - Ist $|t| \geq t_{1-\alpha/2, f}$, so wird H_0 verworfen
 - Andernfalls wird H_0 angenommen

Beispiel 12.10 (Lineare Regression). siehe Vorlesung

Vorgehen 12.11 (Einseitiger t -Test). Die Annahme 12.8 sei erfüllt. Der einseitige t -Test läuft folgendermaßen ab:

1. Stelle die Nullhypothese auf: $H_0: \mu \leq \mu_0$
2. Wähle Fehlerniveau $\alpha \in (0, 1)$
3. Erhebe die Daten x_1, \dots, x_n .
4. Bestimme den sogenannten *Freiheitsgrad* f : Im t -Test ist $f = n - 1$
5. Bestimme $t_{1-\alpha, f}$, das $1 - \alpha$ -Quantil der t -Verteilung zum Parameter f (aus der Tabelle der Quantile der t -Verteilung, oder mit einem geeigneten Programm)
6. Berechne $\bar{\mu}$ und $\bar{\sigma}_n$ aus den Daten
7. Berechne den Testwert

$$t = \sqrt{n} \cdot \frac{\bar{\mu} - \mu_0}{\bar{\sigma}_n}$$

8. Vergleiche den Testwert t mit dem Quantil $t_{1-\alpha, f}$. Entscheidungsregel:
 - Ist $t \geq t_{1-\alpha, f}$, so wird H_0 verworfen
 - Andernfalls wird H_0 angenommen

Wird als Nullhypothese $H_0: \mu \geq \mu_0$ gewählt, so geht man genauso vor, nur die Entscheidungsregel wird umgedreht: H_0 wird verworfen falls $t \leq t_{\alpha, f} = -t_{1-\alpha, f}$ ist.

Beispiel 12.12 (Wirkung eines Medikaments). siehe Vorlesung

Der t -Test kann auch verwendet werden, um zu testen, ob zwei verschiedene Zufallsvariablen denselben Erwartungswert haben (bzw. ob der Erwartungswert der ersten Zufallsvariablen größer oder kleiner als der der zweiten ist). Dabei ist das Vorgehen praktisch gleich, lediglich die Berechnung des Freiheitsgrades und des Testwertes weicht etwas ab.

Annahme 12.13 (Annahme beim t -Test mit zwei zu messenden Größen). Gegeben Daten x_1, \dots, x_n und y_1, \dots, y_n , von welchen angenommen werden kann, dass sie Realisierungen von unabhängigen Zufallsvariablen X_1, \dots, X_n und Y_1, \dots, Y_n sind, wobei die X_i und die Y_i jeweils identisch verteilt sind, und $\mathbb{V}(X_i) = \mathbb{V}(Y_i) < \infty$ gilt. Außerdem seien die X_i und Y_i zumindest annähernd normalverteilt. Es sei $\mu_x = \mathbb{E}[X_1]$ und $\mu_y = \mathbb{E}[Y_1]$.

Vorgehen 12.14 (Zweiseitiger t -Test, zwei zu messende Größen). Die Annahme 12.13 sei erfüllt. Der zweiseitige t -Test läuft folgendermaßen ab:

1. Stelle die Nullhypothese auf: $H_0: \mu_x = \mu_y$
2. Wähle Fehlerniveau $\alpha \in (0, 1)$
3. Erhebe die Daten $x_1, \dots, x_n, y_1, \dots, y_n$.
4. Bestimme den sogenannten *Freiheitsgrad* f : Im t -Test mit zwei Größen ist $f = n + m - 2$
5. Bestimme $t_{1-\alpha/2, f}$, das $1 - \alpha/2$ -Quantil der t -Verteilung zum Parameter f (aus der Tabelle der Quantile der t -Verteilung, oder mit einem geeigneten Programm)
6. Berechne die empirischen Mittel und empirischen Varianzen $\bar{\mu}_x, \bar{\mu}_y$ und $\bar{\sigma}_x^2, \bar{\sigma}_y^2$ aus den Daten, sowie

$$S^2 := \frac{(n-1)\bar{\sigma}_x^2 + (m-1)\bar{\sigma}_y^2}{n+m-2}$$

7. Berechne den Testwert

$$t = \sqrt{\frac{nm}{n+m}} \frac{\bar{\mu}_x - \bar{\mu}_y}{S}$$

8. Vergleiche den Testwert t mit dem Quantil $t_{1-\alpha/2,f}$. Entscheidungsregel:

- Ist $|t| \geq t_{1-\alpha/2,f}$, so wird H_0 verworfen
- Andernfalls wird H_0 angenommen

Im Falle dass man einen einseitigen Test für zwei Größen durchführen will, wird das Vorgehen 12.14 folgendermaßen abgewandelt: Die Nullhypothese lautet

$$\mu_x \leq \mu_y,$$

dann wird sie verworfen wenn $t \geq t_{1-\alpha,f}$ ist. Umgekehrt wird die Nullhypothese

$$\mu_x \geq \mu_y$$

verworfen, wenn $t \leq t_{\alpha,f}$ ist.

Beispiel 12.15 (Marihuanakonsum und Kurzzeitgedächtnis). siehe Vorlesung

12.3. χ^2 -Test. Der χ^2 -Test beruht auf der χ^2 -Verteilung aus Kapitel 11. Er findet in zwei verschiedenen Situationen Anwendung:

- **χ^2 -Test auf Verteilung:** Test ob die den gemessenen Daten zugrundeliegenden Zufallsvariablen einer bestimmten Verteilung folgen, z.B. einer Normalverteilung, oder einer Poisson-Verteilung, uniformen Verteilung...
- **χ^2 -Test auf Unabhängigkeit:** Test ob die gemessenen Daten von Paaren von unabhängigen Zufallsvariablen stammen.

Annahme 12.16 (χ^2 -Test auf Verteilung). Geben Daten x_1, \dots, x_n , von denen angenommen werden kann, dass sie Realisierungen von unabhängigen, identisch verteilten Zufallsvariablen X_1, \dots, X_n sind.

Der χ^2 -Test auf Verteilung beruht darauf, dass die Daten in *Klassen* gruppiert werden, also ähnliche Werte zusammengefasst werden. Für jede Klasse wird Häufigkeit bestimmt, d.h. gezählt, wie viele Messwerte in dieser Klasse enthalten sind. Sind die Klassen hinreichend groß, so sind diese Zufallsvariablen annähernd normalverteilt. Unter der Nullhypothese, dass die Daten einer bestimmten Verteilung folgen, können die theoretischen Häufigkeiten unter dieser Annahme bestimmt werden. Falls die Nullhypothese wahr ist, folgt nach Satz 11.10 die Summe der mittleren quadratischen Abweichungen der gemessenen Häufigkeiten vom theoretischen einer χ^2 -Verteilung. Mit Wahrscheinlichkeit $1 - \alpha$ weichen unter der Nullhypothese die gemessenen Häufigkeiten um höchstens das $1 - \alpha$ -Quantil der χ^2 -Verteilung vom echten Wert ab. Eine größere Abweichung ist also sehr unwahrscheinlich und deutet darauf hin, dass die Nullhypothese falsch ist, weshalb sie in diesem Fall abgelehnt wird. Konkret geht man so vor:

Vorgehen 12.17 (χ^2 -Test auf Verteilung). Die Annahme 12.16 sei erfüllt. Der χ^2 -Test auf Verteilung läuft folgendermaßen ab:

1. Gruppieren der Daten in k Klassen A_1, \dots, A_k , wobei die Klassen A_i Teilmengen von \mathbb{R} sind, so dass $A_i \cap A_j = \emptyset$ gilt, und so dass jeder Messwert x_j in einer Klasse A_i enthalten ist. Bestimmung der Häufigkeiten

$$N_i := |\{j : x_j \in A_i\}|, i = 1, \dots, k.$$

2. Aufstellen der Nullhypothese: H_0 : Die Daten folgen einer bestimmten Verteilung (z.B. Normalverteilung, Poisson-Verteilung, uniforme Verteilung...). Falls nötig Schätzung der Parameter der Verteilung aus den Messwerten.
3. Wahl des Fehlerniveaus $\alpha \in (0, 1)$
4. Berechnung des *Freiheitsgrades*: $f = k - m - 1$, wobei m die Anzahl geschätzter Parameter in der angenommenen Verteilung ist.
5. Bestimmung des $1-\alpha$ -Quantils $\chi^2_{1-\alpha, f}$ der χ^2 -Verteilung mit Parameter $k-m-1$.
6. Berechnung der *theoretischen Häufigkeiten* unter der Nullhypothese:

$$F_i = n \cdot \mathbb{P}(X \in A_i | H_0), i = 1, \dots, k$$

7. Berechnung des *Testwerts*

$$\chi^2 := \sum_{i=1}^k \frac{(N_i - F_i)^2}{F_i}.$$

8. Vergleich des Testwerts χ^2 mit dem Quantil $\chi^2_{1-\alpha, f}$. Entscheidungsregel:
 - Ist $\chi^2 > \chi^2_{1-\alpha, f}$ so wird H_0 verworfen.
 - Andernfalls wird H_0 angenommen.

Beispiel 12.18. Daten: Gewinnspiel einer Getränkemarkte zur Fußball-WM 2014. Sammelbilder zu den teilnehmenden Ländern, eingeteilt in die Gruppen A bis H. Mit allen Bildern der Gruppen A und B gewinnt man den Hauptpreis, mit allen Bildern einer anderen Gruppe gewinnt man Trostpreise. Fragestellung: Sind alle Gruppen gleich häufig vertreten?

1. Daten (gesammelte Bilder): Gesamtzahl $n = 30$. Einteilung in die 8 Gruppen, $k = 8$, ist bereits in natürlicher Weise vorgegeben.

Gruppe A_i	A	B	C	D	E	F	G	H
Häufigkeit N_i	2	4	2	3	3	7	5	4

2. Nullhypothese: H_0 : Die Daten sind diskret gleichverteilt. Da bei einer diskreten Gleichverteilung der Parameter durch die Anzahl Gruppen gegeben ist, hier also $k = 8$, muss kein Parameter geschätzt werden.
3. Wähle das Fehlerniveau $\alpha = 0.05$
4. Da $k = 8$ ist und kein Parameter geschätzt wurde, ist der Freiheitsgrad $f = 8 - 1 = 7$.
5. Aus der Tabelle bestimmen wir das 0.95-Quantil der χ^2 -Verteilung mit Parameter $n = 7$. Dieses ist $\chi^2_{0.95, 7} = 14.07$.

6. Falls die Daten uniform verteilt sind, hat jede Gruppe die gleiche Häufigkeit. Somit sind die theoretische Häufigkeiten gegeben durch

$$F_i = 30/8 = 3.75, i = 1, \dots, 8$$

7. Daraus berechnet sich nun der Testwert als

$$\chi^2 = \sum_{i=1}^k \frac{(N_i - F_i)^2}{F_i} = \frac{1}{3.75} \left((2 - 3.75)^2 + (4 - 3.75)^2 + \dots + (4 - 3.75)^2 \right) = 5.2$$

8. Somit ist der Testwert $\chi^2 = 5.2 < 14.07 = \chi_{0.95,7}^2$, die Nullhypothese wird also angenommen.

Beispiel 12.19 (Umsätze von 197 börsennotierten Unternehmen). *siehe Vorlesung*

Der zweite Anwendungsbereich des χ^2 -Tests liegt darin, zu testen, ob Paare von Daten von unabhängigen Zufallsvariablen herrühren.

Annahme 12.20 (χ^2 -Test auf Unabhängigkeit). Gegeben Paare von Messwerten $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, für welche angenommen werden kann, dass sie Realisierungen von Zufallsvariablen $X_1, Y_1, X_2, Y_2, \dots, X_n, Y_n$ sind, wobei die X_1, \dots, X_n jeweils untereinander unabhängig und identisch verteilt sind, und die Y_1, \dots, Y_n ebenfalls untereinander unabhängig und identisch verteilt sind. Die Verteilung der X_i und der Y_i muss dabei nicht bekannt und auch nicht gleich sein. Zu beachten ist, dass nicht gefordert wird, dass die X_i und Y_i jeweils voneinander unabhängig sind.

Die Fragestellung im Test auf Unabhängigkeit lautet jeweils, zu überprüfen, ob die X_i unabhängig von den Y_i sind.

Auch bei dieser Fragestellung gruppiert man die Daten, und bestimmt die entsprechenden Häufigkeiten, welche man mit den theoretischen Häufigkeiten unter der Nullhypothese der Unabhängigkeit berechnet. Wiederum folgen nach Satz 11.10 die quadratischen Abweichungen der gemessenen und der theoretischen Häufigkeiten einer χ^2 -Verteilung.

Vorgehen 12.21 (χ^2 -Test auf Unabhängigkeit). Unter der Annahme 12.20 wird der χ^2 -Test auf Unabhängigkeit folgendermaßen durchgeführt:

1. Gruppieren der x -Werte in k verschiedene Gruppen, und die y -Werte in m verschiedene Gruppen. Bestimme die Häufigkeiten

$$N_{i,j} := |\{l : x_l \text{ gehört zur Gruppe } i, y_l \text{ gehört zur Gruppe } j\}|$$

Aufstellen der *Kontingenztafel* mit diesen Häufigkeiten (siehe Tabelle). Berechnung der Randhäufigkeiten $N_{i,*} := \sum_{j=1}^m N_{i,j}$ und $N_{*,j} := \sum_{i=1}^k N_{i,j}$.

2. Aufstellen der Nullhypothese: H_0 : Die X_i sind unabhängig von den Y_i .
3. Wahl des Fehlerniveaus $\alpha \in (0, 1)$.
4. Berechnung des Freiheitsgrades $f = (k - 1)(m - 1)$
5. Bestimmung des $1 - \alpha$ -Quantils $\chi_{1-\alpha,f}^2$ der χ^2 -Verteilung mit Parameter f

6. Berechnung der theoretischen Häufigkeiten unter der Nullhypothese:

$$F_{i,j} = \frac{N_{i,*} \cdot N_{*,j}}{n}$$

7. Berechnung des Testwerts $\chi^2 = \sum_{i,j} \frac{(F_{i,j} - N_{i,j})^2}{F_{i,j}}$

8. Vergleich des Testwerts χ^2 mit dem Quantil $\chi^2_{1-\alpha, f}$. Entscheidungsregel:

- Ist $\chi^2 > \chi^2_{1-\alpha, f}$ so wird H_0 verworfen.
- Andernfalls wird H_0 angenommen.

Kontingenztafel: Unter Annahme 12.20 stellt man folgende Kontingenztafel auf:

X \ Y	1	2	...	m	Total
1	$N_{1,1}$	$N_{1,2}$...	$N_{1,m}$	$N_{1,*}$
2	$N_{2,1}$				
\vdots					
k	$N_{k,1}$			$N_{k,m}$	$N_{k,*}$
Total	$N_{*,1}$			$N_{*,m}$	n

Dabei bezeichnet $N_{i,j}$ die Anzahl der Paare mit x -Wert i und y -Wert j , $N_{i,*}$, $N_{*,j}$ die Randhäufigkeiten, d.h. $N_{i,*}$ die Anzahl Paare mit x -Wert i , und $N_{*,j}$ die Anzahl Paare mit y -Wert j . Diese Randhäufigkeiten werden zur Berechnung der theoretischen Häufigkeiten verwendet. Sind die Zufallsvariablen unabhängig, so sind nach Definition die theoretischen Häufigkeiten gleich dem Produkt der Randhäufigkeiten (vgl. Kapitel 4).

Beispiel 12.22 (Unabhängigkeit von Genveränderungen). 100 Personen werden auf zwei verschiedene Genveränderungen (A und B) getestet. Dabei findet man folgende Häufigkeiten:

$N_{i,j}$	A vorhanden	A nicht vorhanden	Total
B vorhanden	16	14	30
B nicht vorhanden	24	46	70
Total	40	60	100

Mittels eines χ^2 -Tests soll nun ermittelt werden, ob das Auftreten der beiden Genveränderungen unabhängig voneinander ist. Wir folgen dem Vorgehen 12.21.

1. Die obige Tabelle ist bereits die Kontingenztafel.
2. Nullhypothese: Die Merkmale sind unabhängig.
3. Wir wollen zwei α untersuchen: $\alpha = 0.05$ und $\alpha = 0.1$.
4. Wir haben $k = m = 2$, also ist $f = 1$.
5. Die Quantile sind $\chi^2_{0.95,1} = 3.84$ und $\chi^2_{0.9,1} = 2.71$
6. Wir berechnen die theoretischen Häufigkeiten $F_{i,j}$, unter der Nullhypothese dass A und B unabhängig voneinander auftreten (bei gleicher Gesamtzahl der einzelnen Merkmale, d.h. die Randhäufigkeiten sind dieselben wie in der Kontingenztafel der gemessenen Häufigkeiten):

$F_{i,j}$	A vorhanden	A nicht vorhanden	Total
B vorhanden	12	18	30
B nicht vorhanden	28	42	70
Total	40	60	100

7. Der Testwert wird somit

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(N_{i,j} - F_{i,j})^2}{F_{i,j}} = \frac{(16 - 12)^2}{12} + \frac{(14 - 18)^2}{18} + \frac{(24 - 28)^2}{28} + \frac{(46 - 42)^2}{42} = 3.17$$

8. Wir sehen, dass $\chi^2 < \chi_{0.95,1}^2$ ist, aber $\chi^2 > \chi_{0.9,1}^2$. Somit wird die Nullhypothese für $\alpha = 0.05$ angenommen, jedoch für $\alpha = 0.1$ verworfen.