

SS 2004

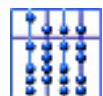
Diskrete Strukturen II

Ernst W. Mayr

Fakultät für Informatik

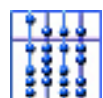
TU München

<http://www14.in.tum.de/lehre/2004SS/ds/index.html.de>



- **Annahmen über die Zufallsgrößen**

Was ist über die Verteilung der untersuchten Größe(n) bekannt? Bei entsprechenden Annahmen könnte es sich z.B. um die Art der Verteilung, den Erwartungswert oder die Varianz handeln.



Ein-Stichproben-Tests für Lageparameter

Beim approximativen Binomialtest wird ausgenutzt, dass die Binomialverteilung für große n nach dem Grenzwertsatz von DeMoivre (Korollar 46) gegen die Normalverteilung konvergiert. Aus diesem Grund kann man diesen Test auch als Spezialfall eines allgemeineren Testverfahrens ansehen, nämlich des **Gaußtest**, der nun dargestellt wird.

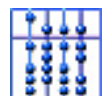


Tabelle 3.1: Gaußtest

Annahmen:

X_1, \dots, X_n seien unabhängig und identisch verteilt mit
 $X_i \sim \mathcal{N}(\mu, \sigma^2)$, wobei σ^2 bekannt ist.

Alternativ gelte $\mathbb{E}[X_i] = \mu$ und $\text{Var}[X_i] = \sigma^2$ und n sei groß genug.

Hypothesen:

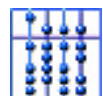
- a) $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$,
- b) $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$,
- c) $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$.

Testgröße:

$$Z := \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}.$$

Ablehnungskriterium für H_0 bei Signifikanzniveau α :

- a) $|Z| > z_{1-\alpha/2}$,
- b) $Z < z_\alpha$,
- c) $Z > z_{1-\alpha}$.



Der Gaußtest hat den Nachteil, dass man die Varianz σ^2 der beteiligten Zufallsgrößen kennen muss.

Wenn diese unbekannt ist, so liegt es nahe, die Varianz durch die Stichprobenvarianz S^2 (siehe [Definition](#)) anzunähern. Dies führt auf den so genannten t -Test, der in der folgenden Übersicht dargestellt ist.



Tabelle 3.2: t -Test

Annahmen:

X_1, \dots, X_n seien unabhängig und identisch verteilt mit
 $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

Alternativ gelte $\mathbb{E}[X_i] = \mu$ und $\text{Var}[X_i] = \sigma^2$ und n sei groß genug.

Hypothesen:

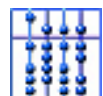
- a) $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$,
- b) $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$,
- c) $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$.

Testgröße:

$$T := \frac{\bar{X} - \mu_0}{S} \sqrt{n}.$$

Ablehnungskriterium für H_0 bei Signifikanzniveau α :

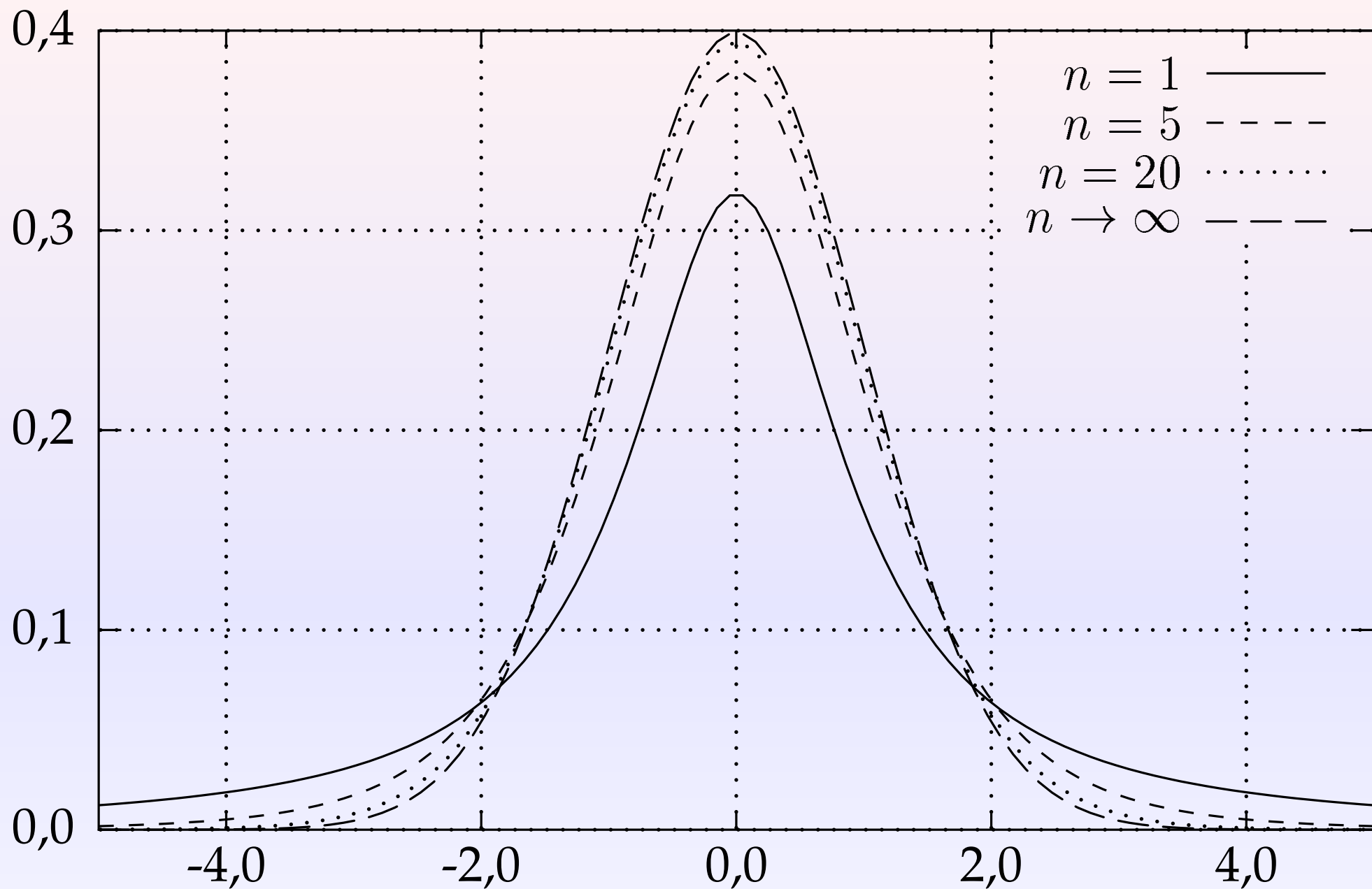
- a) $|T| > t_{n-1, 1-\alpha/2}$,
- b) $T < t_{n-1, \alpha}$,
- c) $T > t_{n-1, 1-\alpha}$.



Hierbei gibt $t_{n-1,1-\alpha}$ das $(1 - \alpha)$ -Quantil der t -Verteilung mit $n - 1$ Freiheitsgraden an. Die t -Verteilung taucht manchmal auch unter dem Namen **Student-Verteilung** auf, da sie ursprünglich unter dem Pseudonym „Student“ publiziert wurde.

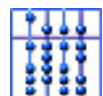
Wir gehen an dieser Stelle nicht darauf ein, wieso die Testgröße die t -Verteilung besitzt, sondern weisen nur darauf hin, dass die Dichte dieser Verteilung (eigentlich handelt es sich um eine ganze Familie von Verteilungen, da die Anzahl der Freiheitsgrade jeweils noch gewählt werden kann) der Dichte der Normalverteilung ähnelt. Für große n (Faustregel: $n \geq 30$) liegen die beiden Dichten so genau übereinander, dass man in der Praxis die t -Verteilung durch die Normalverteilung annähert.





Dichte der t -Verteilung mit n Freiheitsgraden

Als weitere Beispiele für gängige Ein-Stichproben-Tests zu Lageparametern seien der **Wilcoxon-Test** und der χ^2 -**Varianztest** genannt. Ersterer dient zum Testen von Hypothesen zum Median, während der zweite Test Hypothesen zur Varianz beinhaltet.



Zwei-Stichproben-Tests für Lageparameter

Bei Zwei-Stichproben-Tests wollen wir das Verhältnis von Lageparametern untersuchen. Besonders wichtig sind hierbei Tests zum Erwartungswert. Für zwei Zufallsgrößen X und Y könnten wir beispielsweise die Frage untersuchen, ob für die Erwartungswerte μ_X und μ_Y gilt, dass $\mu_X = \mu_Y$ ist.

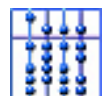


Tabelle 3.3: Zwei-Stichproben- t -Test

Annahmen:

X_1, \dots, X_m und Y_1, \dots, Y_n seien unabhängig und jeweils identisch verteilt, wobei $X_i \sim \mathcal{N}(\mu_X, \sigma_X^2)$ und $Y_i \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ gelte. Die Varianzen seien identisch, also $\sigma_X^2 = \sigma_Y^2$.

Hypothesen:

- a) $H_0 : \mu_X = \mu_Y$ gegen $H_1 : \mu_X \neq \mu_Y$,
- b) $H_0 : \mu_X \geq \mu_Y$ gegen $H_1 : \mu_X < \mu_Y$,
- c) $H_0 : \mu_X \leq \mu_Y$ gegen $H_1 : \mu_X > \mu_Y$.

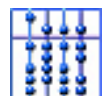
Testgröße:

$$T := \sqrt{\frac{n+m-2}{\frac{1}{m} + \frac{1}{n}}} \cdot \frac{\bar{X} - \bar{Y}}{\sqrt{(m-1) \cdot S_X^2 + (n-1) \cdot S_Y^2}}.$$

Ablehnungskriterium für H_0 bei Signifikanzniveau α :

- a) $|T| > t_{m+n-2, 1-\alpha/2}$,
- b) $T < t_{m+n-2, \alpha}$,
- c) $T > t_{m+n-2, 1-\alpha}$.

Vom Zwei-Stichproben- t -Test findet man in der Literatur noch zusätzliche Varianten, die auch dann einsetzbar sind, wenn die beteiligten Zufallsgrößen nicht dieselbe Varianz besitzen. Der beim Ein-Stichproben-Fall erwähnte Wilcoxon-Test kann ebenfalls auf den Zwei-Stichproben-Fall übertragen werden.



Nicht an Lageparametern orientierte Tests

Wir betrachten in diesem Abschnitt exemplarisch den χ^2 -Anpassungstest. Bei einem Anpassungstest wird nicht nur der Lageparameter einer Verteilung getestet, sondern es wird die Verteilung als Ganzes untersucht.

Beim approximativen Binomialtest (siehe Tabelle 3.1) haben wir streng genommen bereits einen Anpassungstest durchgeführt. Bei der Nullhypothese $H_0 : p = p_0$ wird untersucht, ob es sich bei der betrachteten Zufallsgröße um eine Bernoulli-verteilte Zufallsvariable mit Parameter p_0 handelt. Beim χ^2 -Test gehen wir nun einen Schritt weiter: Wir nehmen an, dass die Zufallsgröße X genau k verschiedene Werte annimmt. Ohne Beschränkung der Allgemeinheit sei $W_X = \{1, \dots, k\}$. Die Nullhypothese lautet nun

$$H_0 : \Pr[X = i] = p_i \quad \text{für } i = 1, \dots, k.$$

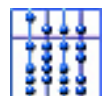


Tabelle 3.4: χ^2 -Anpassungstest

Annahmen:

X_1, \dots, X_n seien unabhängig und identisch verteilt mit
 $W_{X_i} = \{1, \dots, k\}$.

Hypothesen:

H_0 : $\Pr[X = i] = p_i$ für $i = 1, \dots, k$,

H_1 : $\Pr[X = i] \neq p_i$ für mindestens ein $i \in \{1, \dots, k\}$,

Testgröße:

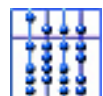
$$T = \sum_{i=1}^k \frac{(h_i - np_i)^2}{np_i},$$

wobei h_i die Häufigkeit angibt, mit der X_1, \dots, X_n den Wert i angenommen haben.

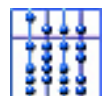
Ablehnungskriterium für H_0 bei Signifikanzniveau α :

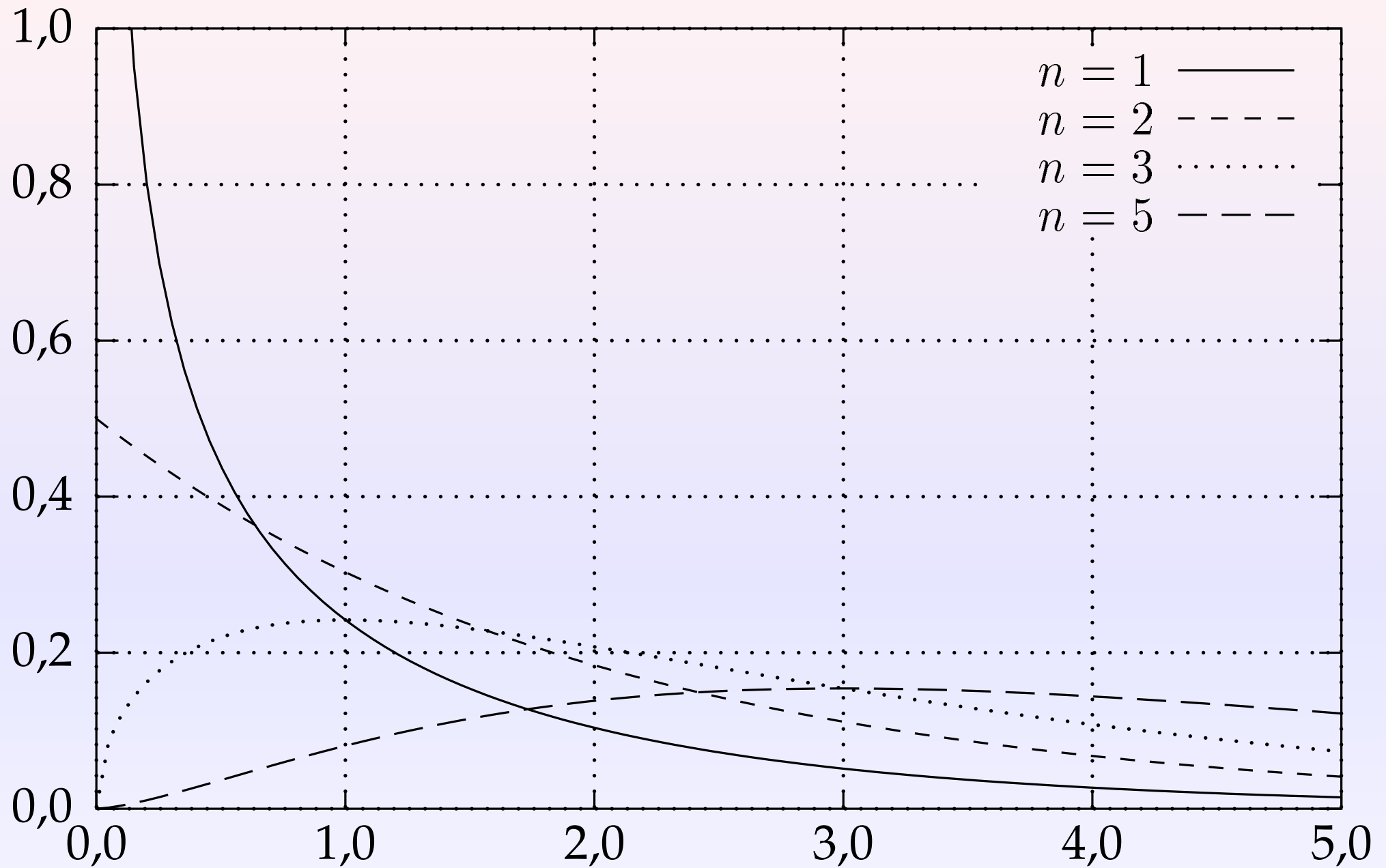
$$T > \chi_{k-1, 1-\alpha}^2;$$

dabei sollte gelten, dass $np_i \geq 1$ für alle i und $np_i \geq 5$ für mindestens 80% der Werte $i = 1, \dots, k$.

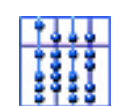


Für die Testgröße T wird näherungsweise eine χ^2 -Verteilung mit $k - 1$ Freiheitsgraden angenommen. Die Werte dieser Verteilung finden sich in entsprechenden Tabellen in der Literatur. Damit diese Approximation gerechtfertigt ist, sollte gelten, dass $np_i \geq 1$ für alle i und $np_i \geq 5$ für mindestens 80% der Werte $i = 1, \dots, k$. Das γ -Quantil einer χ^2 -Verteilung mit k Freiheitsgraden bezeichnen wir mit $\chi_{k,\gamma}^2$.





Dichte der χ^2 -Verteilung mit n Freiheitsgraden



Beispiel: Als Anwendung für den χ^2 -Test wollen wir überprüfen, ob der Zufallszahlengenerator von Maple eine gute Approximation der Gleichverteilung liefert. Dazu lassen wir Maple $n = 100000$ Zufallszahlen aus der Menge $\{1, \dots, 10\}$ generieren. Wir erwarten, dass jede dieser Zahlen mit gleicher Wahrscheinlichkeit $p_1 = \dots = p_{10} = 1/10$ auftritt. Dies sei unsere Nullhypothese, die wir mit einem Signifikanzniveau von $\alpha = 0,05$ testen wollen.

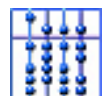
Beispiel:

i	1	2	3	4	5	6	7	8	9	10
h_i	10102	10070	9972	9803	10002	10065	10133	9943	10009	9901

Für den Wert der Testgröße gilt $T = 8,9946$. Ferner erhalten wir $\chi_{9,0,95}^2 \approx 16,919$. Der Test liefert also keinen Grund, die Nullhypothese abzulehnen.



Das Prinzip des χ^2 -Anpassungstests kann in leicht abgewandelter Form auch noch zum Testen einiger anderer Hypothesen verwendet werden: Beim χ^2 -Homogenitätstest wird überprüft, ob zwei oder mehrere Verteilungen identisch sind, während beim χ^2 -Unabhängigkeitstest zwei Zufallsgrößen auf Unabhängigkeit untersucht werden. Beschreibungen dieser Tests findet man in der Literatur.

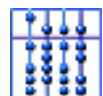


4 Stochastische Prozesse

4.1 Einführung

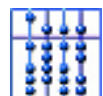
Wir betrachten zeitliche Folgen von Zufallsexperimenten.

Mathematisch beschreibt man diese durch einen so genannten **stochastischen Prozess**. Darunter versteht man eine Folge von Zufallsvariablen $(X_t)_{t \in T}$, die das Verhalten des Systems zu verschiedenen Zeitpunkten t angeben.



Wenn wir $T = \mathbb{N}_0$ annehmen, sprechen wir von einem stochastischen Prozess mit diskreter Zeit. Lässt man andererseits $T = \mathbb{R}_0^+$ zu, so spricht man von stochastischen Prozessen mit kontinuierlicher Zeit.

Eine besonders einfache Art von stochastischen Prozessen sind so genannte **Markov-Ketten**. Diese haben die Eigenschaft, dass der nächste Zustand des Prozesses zwar vom aktuellen Zustand abhängen darf, nicht aber von der Historie, d.h. davon, wie der aktuelle Zustand erreicht wurde.



4.2 Prozesse mit diskreter Zeit

4.2.1 Einführung

Definition: Eine (endliche) Markov-Kette (mit diskreter Zeit) über der Zustandsmenge $S = \{0, \dots, n - 1\}$ besteht aus einer unendlichen Folge von Zufallsvariablen $(X_t)_{t \in \mathbb{N}_0}$ mit Wertemenge S sowie einer Startverteilung q_0 mit $q_0^T \in \mathbb{R}^n$. Die Komponenten von q_0 sind hierbei positiv und addieren sich zu 1. Für jede Indexmenge $I \subseteq \{0, \dots, t - 1\}$ und beliebige Zustände i, j, s_k ($k \in I$) gilt

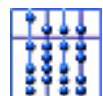
$$\Pr[X_{t+1} = j \mid X_t = i, \forall k \in I : X_k = s_k] = \Pr[X_{t+1} = j \mid X_t = i]. \quad (4.2.1)$$



Sind die Werte

$$p_{ij} := \Pr[X_{t+1} = j \mid X_t = i]$$

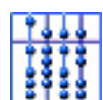
von t unabhängig, so nennt man die Markov-Kette (zeit)homogen.
In diesem Fall definiert man die Übergangsmatrix durch
 $P = (p_{ij})_{0 \leq i, j < n}$. Wenn man $S = \mathbb{N}_0$ zulässt, so spricht man von
einer unendlichen Markov-Kette.



Bedingung (4.2.1) heißt **Markov-Bedingung** und besagt:

Wenn wir den Zustand i zum Zeitpunkt t kennen, so hängt die Übergangswahrscheinlichkeit zum Folgezustand j nur von i und j ab. Die Vergangenheit (Zustände zu Zeitpunkten $< t$) der Markov-Kette spielt keine Rolle. Das „Gedächtnis“ der Markov-Kette besteht also nur aus ihrem aktuellen Zustand und sie „weiß“ nicht, wie sie dorthin gekommen ist.

Bei einer zeithomogenen Markov-Kette hat die (absolute) Zeit t keinen Einfluss auf die Übergangswahrscheinlichkeiten p_{ij} , d.h. das Systemverhalten wird nur durch den aktuellen Zustand bestimmt und nicht durch eine absolute Uhr.



Wahrscheinlichkeitsraum einer Markov-Kette

Nehmen wir an, dass wir die Kette von der Zeit 0 bis zur Zeit t_0 beobachten wollen. Wir bezeichnen die Folge von Zuständen, die von der Kette in dieser Zeit durchlaufen wurde, mit $\vec{x} = (x_0, x_1, \dots, x_{t_0})$. $\Omega = S^{t_0+1}$ sei die Menge möglicher Zustandsfolgen. Einer beliebigen Folge $\omega := (x_0, x_1, \dots, x_{t_0}) \in \Omega$ ordnen wir die Wahrscheinlichkeit

$$\Pr[\omega] = (q_0)_{x_0} \cdot \prod_{i=1}^{t_0} \Pr[X_i = x_i \mid X_{i-1} = x_{i-1}]$$

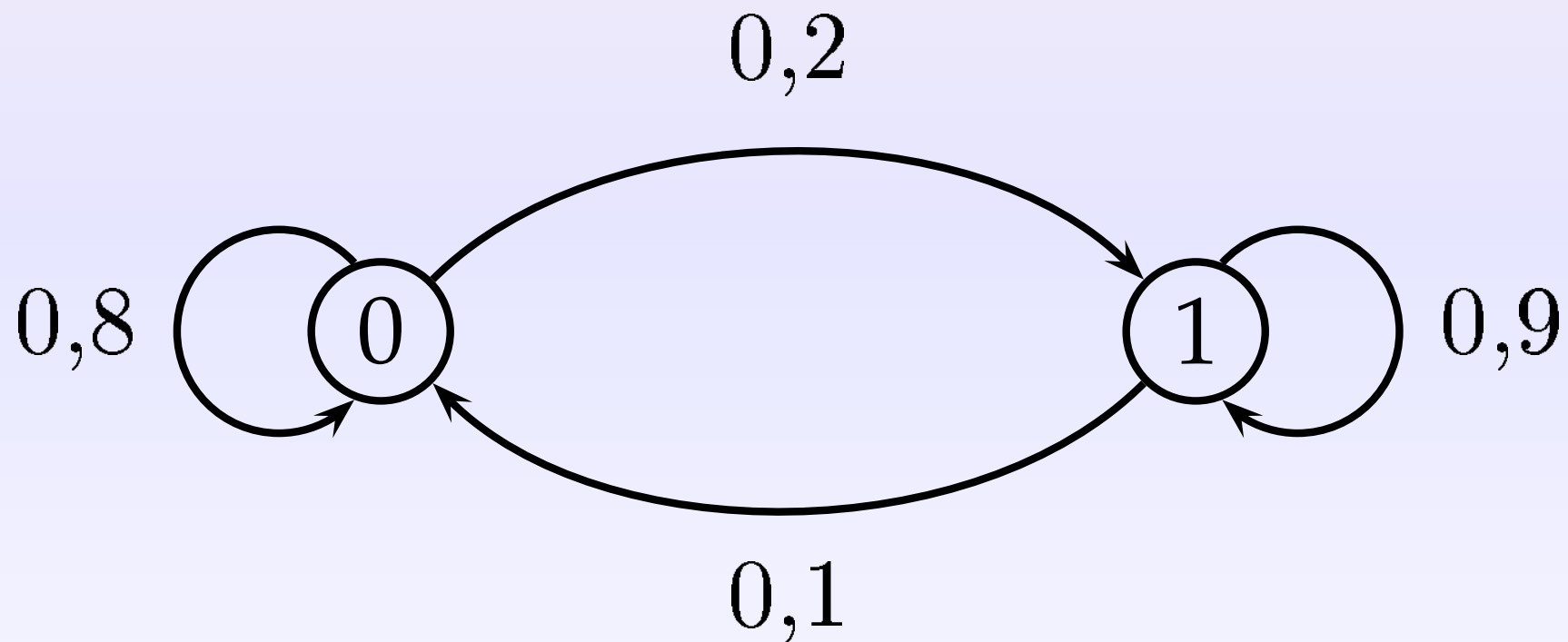
zu. Dadurch erhalten wir einen diskreten Wahrscheinlichkeitsraum im Sinne der Definition.



Beispiel:

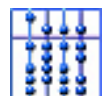
$$\Pr[X_{t+1} = 1 \mid X_t = 1] = 0,9 \text{ und } \Pr[X_{t+1} = 1 \mid X_t = 0] = 0,2$$

$$\Pr[X_{t+1} = 0 \mid X_t = 1] = 0,1 \text{ und } \Pr[X_{t+1} = 0 \mid X_t = 0] = 0,8$$



Einen bestimmten Ablauf des Systems kann man sich als so genannten **Random Walk** vorstellen.

Wenn wir beispielsweise uns zum Zeitpunkt $t = 0$ im Knoten **1** (also $X_0 = 1$), dann führen von dort zwei Kanten weiter, nämlich zu den Knoten **0** und **1**. Diese Kanten sind mit Wahrscheinlichkeiten beschriftet, die sich zu Eins addieren. Gemäß dieser Wahrscheinlichkeiten entscheiden wir zufällig, wohin wir uns im nächsten Schritt begeben.



Wir können auch die Frage beantworten, mit welcher Wahrscheinlichkeit wir uns zum Zeitpunkt $t = 2$ im Knoten 1 befinden. Da wir vereinbarungsgemäß beim Knoten 1 starten, gibt es zwei mögliche Wege der Länge zwei durch den Graphen mit Endknoten 1, nämlich „111“ und „101“. Die Wahrscheinlichkeiten für diese Wege lauten $0,9 \cdot 0,9 = 0,9^2$ bzw. $0,1 \cdot 0,2$. Insgesamt erhalten wir also eine Wahrscheinlichkeit von $0,81 + 0,02 = 0,83$.

Auch eine Aussage über die erwartete Anzahl Schritte, die wir im Knoten 1 bis zum ersten Übergang zu Knoten 0 verbleiben, ist schnell getroffen. Die Wahrscheinlichkeit, dass man genau k Schritte verbleibt, ist $(0,9)^k \cdot 0,1$. Die Anzahl Schritte ist also geometrisch verteilt mit Erfolgswahrscheinlichkeit $0,1$. Der Erwartungswert ist daher $1/0,1 = 10$.



4.2.2 Berechnung von Übergangswahrscheinlichkeiten

Wir beschreiben die Situation zum Zeitpunkt t durch einen Zustandsvektor q_t (den wir als Zeilenvektor schreiben). Die i -te Komponente $(q_t)_i$ bezeichnet dabei die Wahrscheinlichkeit, mit der sich die Kette nach t Schritten im Zustand i aufhält.

Es gilt

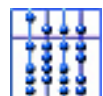
$$\Pr[X_{t+1} = k] = \sum_{i=0}^{n-1} \Pr[X_{t+1} = k \mid X_t = i] \cdot \Pr[X_t = i],$$

also

$$(q_{t+1})_k = \sum_{i=0}^{n-1} p_{ik} \cdot (q_t)_i,$$

bzw. in Matrixschreibweise

$$q_{t+1} = q_t \cdot P.$$



Mit der Matrixschreibweise können wir q_t einfach durch die Startverteilung q_0 ausdrücken:

$$q_t = q_0 \cdot P^t .$$

Ebenso gilt wegen der Zeithomogenität allgemein für alle $t, k \in \mathbb{N}$:

$$q_{t+k} = q_t \cdot P^k .$$

Die Einträge von P^k geben an, mit welcher Wahrscheinlichkeit ein Übergang vom Zustand i zum Zustand j in genau k Schritten erfolgt.

$$p_{ij}^{(k)} := \Pr[X_{t+k} = j \mid X_t = i] = (P^k)_{ij} .$$

