Universität Bielefeld

Fakultät für Mathematik
Prof. Dr. Barbara Gentz SS 2017

Vertiefung NWI: 9. Vorlesung zur Wahrscheinlichkeitstheorie

Mittwoch, 14.07.2017

9. Grenzwertsätze

9.1 Schwaches Gesetz der großen Zahlen

Satz: *Schwaches Gesetz der großen Zahlen*
Gegeben seien Zufallsgrößen $X_1, X_2, X_3, \dots$ auf einem Wahrscheinlichkeitsraum (Ω, p), deren Verteilung *nicht* identisch sein muß. Wir setzen voraus:

- Alle Erwartungswerte $\mathbb{E}[X_i]$ existieren, und es gilt $\mathbb{E}[X_i] = \mathbb{E}[X_1]$ für alle i
- $\operatorname{Var}(X_i) = \operatorname{Var}(X_1) < \infty$ für alle i
- Die Zufallsgrößen sind *paarweise unkorreliert:* $\mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] = 0$ für alle $i \neq j$

Sei $S_n = \sum_{i=1}^n X_i$. Dann gilt für jedes $\varepsilon > 0$:

$$\lim_{n\to\infty} \mathbb{P}\{|\tfrac{1}{n} S_n - \mathbb{E}[X_1]| \geq \varepsilon\} = 0$$

Beweis
Direkte Folgerung aus der Tschebyscheffschen Ungleichung: Für jedes feste $\varepsilon > 0$ gilt

$$\mathbb{P}\{|\tfrac{1}{n} S_n - \mathbb{E}[X_1]| \geq \varepsilon\} \leq \frac{\operatorname{Var}\left(\frac{1}{n} S_n\right)}{\varepsilon^2} = \frac{\operatorname{Var}(S_n)}{n^2\varepsilon^2} = \frac{n\mathbb{E}[X_1]}{n^2\varepsilon^2} = \frac{\mathbb{E}[X_1]}{n\varepsilon^2} \to 0 \qquad \text{für } n \to \infty$$

Bemerkung
Unabhängigkeit $\Rightarrow$ paarweise Unabhängigkeit $\Rightarrow$ paarweise Unkorreliertheit

Bemerkung
$\mathbb{E}[\frac{1}{n} S_n] = \mathbb{E}[X_1]$

Beispiel: *Ermitteln der Erfolgswahrscheinlichkeit aus Daten*
Ist S_n die Anzahl der Erfolge in einem n-mal unabhängig wiederholten Bernoulli-Experiment, so gilt $\mathbb{E}[\frac{1}{n}S_n] = \mathbb{E}[X_1] = p =$ Erfolgswahrscheinlichkeit (ggf. unbekannt). In diesem Fall zeigt das Schwache Gesetz der großen Zahlen, daß die relative Häufigkeit der Erfolge gegen die Erfolgswahrscheinlichkeit konvergiert (in dem Sinne, daß die Wahrscheinlichkeit, daß die relative Häufigkeit um mindestens ε von der Erfolgswahrscheinlichkeit abweicht, gegen Null konvergiert).

Werfen von Reißzwecken in der Vorlesung vom 3.6.2015 ergab, daß die Reißzwecken in 220 Würfen 97 Mal auf dem Kopf zu liegen kamen. Das führt zu einer Schätzung der Erfolgswahrscheinlichkeit: $p = \frac{97}{220} \sim 0,44.$

Wie gut ist diese Schätzung?

9.2 Lokaler zentraler Grenzwertsatz

$0 < p < 1$, $q := 1 - p$

Abkürzende Schreibweise: $x_k^{(n)} := (k - np)/\sqrt{npq}$ für $k = 0, \ldots, n$

Erinnerung
Für eine binomialverteilte Zufallsgröße S_n mit Parametern n und p ist $np = \mathbb{E}[S_n]$ und $npq = \mathrm{Var}(S_n)$.
Wir schreiben

$$b(k; n, p) = \binom{n}{k} p^k q^{n-k}$$

Satz: *Lokaler zentraler Grenzwertsatz*
Sei $A > 0$ eine beliebige Konstante. Dann gilt

$$(*) \qquad b(k; n, p) \sim \frac{1}{\sqrt{2\pi npq}} \, \mathrm{e}^{-(x_k^{(n)})^2/2} \quad \textit{gleichmäßig} \text{ für alle } k \text{ mit } |x_k^{(n)}| \le A$$

Dabei bedeutet $\sim$:

$$f(n) \sim g(n) \iff \lim_{n\to\infty} \frac{f(n)}{g(n)} = 1 \, .$$

Die Aussage (*) kann also geschrieben als

$$\lim_{n\to\infty} \sup_{k: |x_k^{(n)}| \le A} \left| \frac{b(k; n, p)}{\frac{1}{\sqrt{2\pi npq}} \, \mathrm{e}^{-(x_k^{(n)})^2/2}} - 1 \right| = 0 \, .$$

Beweis

- Ausschreiben des Binomialkoeffizienten
- Approximieren der Fakultäten mit Hilfe der Stirlingschen Formel (siehe nächste Seite)
- Sortieren der Terme
- Sorgfältiges Berechnen des Limes $n \to \infty$

Beispiel: *1200-mal Würfeln mit einem fairen Würfel*
Wie groß ist die Wahrscheinlichkeit, *genau* k Sechsen zu werfen? ($k = 200$, $k = 250$)

k	b(k;1200,1/6)	Approximation gemäß lokalem zentralem Grenzwertsatz
200	0,030888…	0,030901…
250	0,0000244…	0,0000170…

Sind das wirklich die Wahrscheinlichkeiten, die uns typischerweise interessieren?

Neue Fragen

- Wie groß ist die Ws., zwischen 150 und 250 Sechsen zu werfen?
- Wie groß ist die Ws., mindestens 250 Sechsen zu werfen?

Ergänzung: *Stirlingsche Formel*

Für große n gilt

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n$$

n	n!	Approximation
1	1	0.922
2	2	1.919
3	6	5.836
4	24	23.506
5	120	118.019
6	720	710.078
7	5040	4980.4
8	40320	39902.4
9	362880	359537
10	3628800	$3.5987 \cdot 10^6$
11	39916800	$3.96156 \cdot 10^7$
12	479001600	$4.75687 \cdot 10^8$

9.3 Satz von de Moivre-Laplace (Spezialfall des zentralen Grenzwertsatzes)

Satz: *Satz von de Moivre-Laplace*
Es sei S_n die Anzahl der Erfolge in einem n-fach unabhängig wiederholten Bernoulli-Experiment mit Erfolgsws. $p \in (0,1)$. Dann gilt für jede Wahl von Konstanten a, b mit $-\infty < a < b < \infty$

$$\lim_{n\to\infty} \mathbb{P}\Big\{a < \frac{S_n - np}{\sqrt{npq}} \leq b\Big\} = \frac{1}{\sqrt{2\pi}} \int_a^b \mathrm{e}^{-x^2/2}\ \mathrm{d}x$$

Dabei verstehen wir die rechte Seite als Riemann-Integral $\int_a^b \varphi(x)\ \mathrm{d}x$ mit

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}\ .$$

Wir nennen φ die Dichte der Standardnormalverteilung. Die Stammfunktion

$$\Phi(y) = \int_{-\infty}^{y} \varphi(x)\ \mathrm{d}x$$

heißt auch Verteilungsfunktion der Standardnormalverteilung. (Dazu mehr in der 11. Vorlesung.)

Der Graph von φ ist die berühmte Gaußsche Glockenkurve, Φ ist die Fläche unter der Kurve φ.

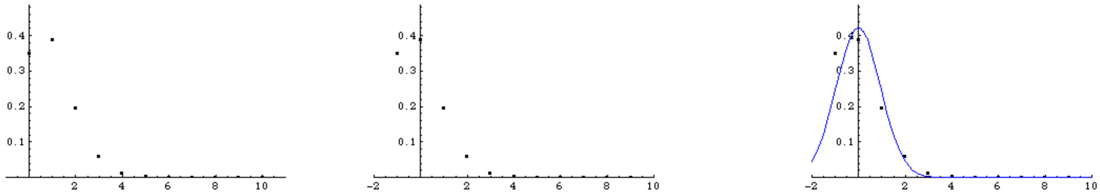
Beweis
Zunächst Approximieren der Summanden $b(k; n, p)$ durch den lokalen zentralen Grenzwertsatz. Dann approximieren der Summe mit Hilfe eines Riemann-Integrals.

Rechenregeln

- $\displaystyle\int_{-\infty}^{\infty} \varphi(x)\ \mathrm{d}x = 1$
- $\Phi(y) \to \begin{cases} 0 & \text{für } y \to -\infty \\ 1 & \text{für } y \to +\infty \end{cases}$
- $\Phi(0) = \dfrac{1}{2}$
- $\displaystyle\Phi(z) = \int_{-\infty}^{z} \varphi(x)\ \mathrm{d}x = 1 - \int_{z}^{\infty} \varphi(x)\ \mathrm{d}x = 1 - \int_{-\infty}^{-z} \varphi(x)\ \mathrm{d}x = 1 - \Phi(-z)$

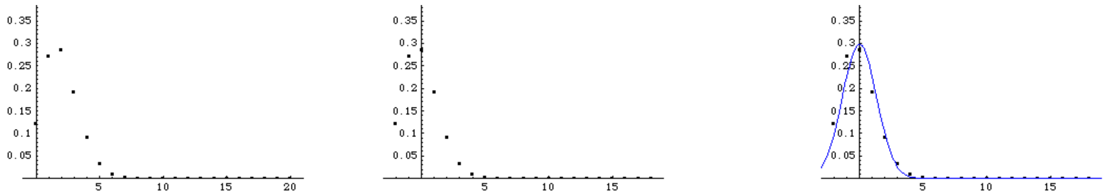
Ergänzung: *Lokaler zentraler Grenzwertsatz und Satz von de Moivre-Laplace*

$$n = 10,\ p = 1/10$$



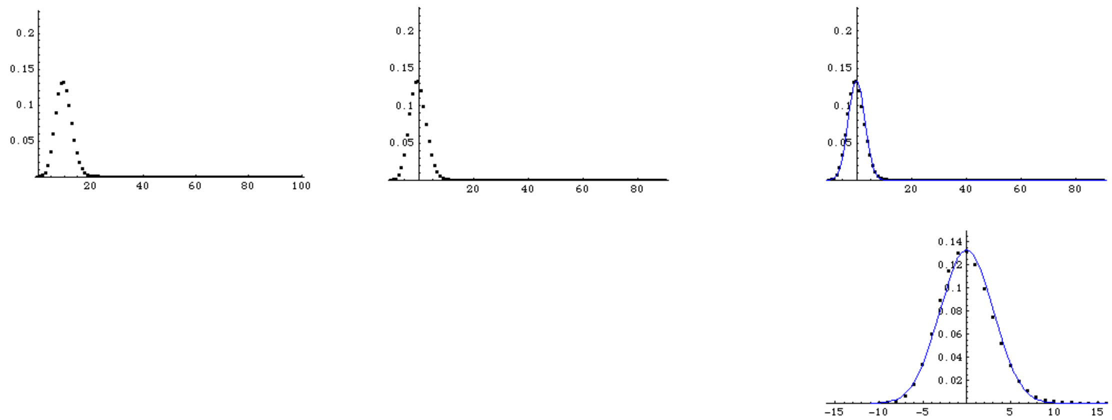
Binomialverteilung zentrierte Binomialverteilung approximierende Normalverteilung

$$n = 20,\ p = 1/10$$



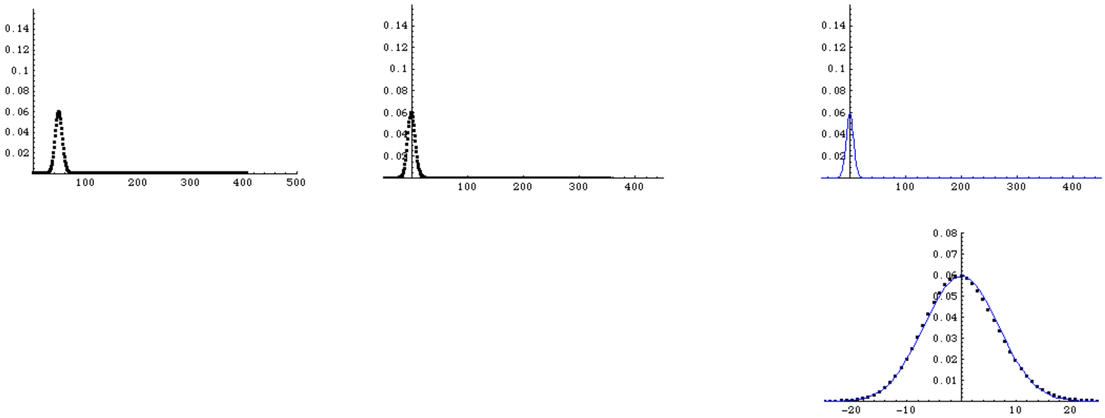
Binomialverteilung zentrierte Binomialverteilung approximierende Normalverteilung

$$n = 100,\ p = 1/10$$



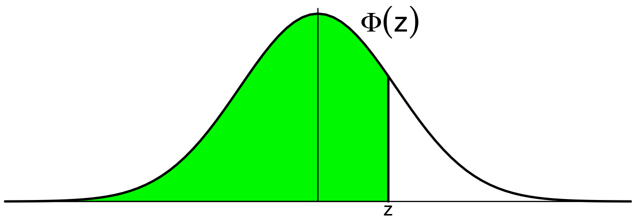
Binomialverteilung zentrierte Binomialverteilung approximierende Normalverteilung

$n = 500,\ p = 1/10$



Binomialverteilung zentrierte Binomialverteilung approximierende Normalverteilung

Tabelle der Kumulativen Normalverteilung $\Phi(z) = \mathrm{P}\left[Z \leq z\right], \;\; Z \sim \mathcal{N}(0, 1$



Bsp.: $\mathrm{P}\left[Z \leq 1.96\right] = 0.975$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

9.4 Beispiele zum Satz von de Moivre-Laplace

Beispiel 1
Eine Fabrik stellt Chips her mit einer Ausschußrate von 10%. Mit welcher Wahrscheinlichkeit sind unter 400 produzierten Chips mehr als 50 defekt?

Wir betrachten jeden defekten Chip unter den 400 Chips als "Erfolg" und modellieren das Experiment als 400-mal unabhängig wiederholtes Bernoulli-Experiment mit Erfolgswahrscheinlichkeit $1/10$. Dabei soll die Anzahl der Erfolge S_n gerade die Anzahl defekter Chips sein.

- $\Omega = \{0,1\}^n$ mit $n = 400$ und $p(\omega) = p^{\sum_{i=1}^n \omega_i}(1-p)^{n-\sum_{i=1}^n \omega_i}$
- $X_i(\omega) = \omega_i$, $X_i = 1$ bedeute einen "Erfolg" im i-ten Experiment, d.h., der i-te Chip ist defekt
- Annahme: Die Zufallsgrößen X_i sind u.i.v. mit $\mathbb{P}\{X_i = 1\} = p = 1/10$
  (Wie realistisch ist diese Annahme?)
- Unter dieser Annahme ist S_n binomialverteilt mit Parametern n und p
- $\mathbb{E}S_n = np = 40$ und $\mathrm{Var}(S_n) = npq = 6^2$ mit $q = 1-p = 9/10$

Gesucht ist $\mathbb{P}\{S_n > 50\}$.

$$\begin{aligned}
\mathbb{P}\{S_n > 50\} &= \mathbb{P}\{50 < S_n \leq 400\} \\
&= \mathbb{P}\Big\{\frac{50-40}{6} < \frac{S_n - np}{\sqrt{npq}} \leq \frac{400-40}{6}\Big\} && \text{(Standardisieren)} \\
&\approx \Phi(60) - \Phi(5/3) && \text{(Satz von de Moivre-Laplace)} \\
&\approx 1 - \Phi(1,67) \\
&\approx 1 - 0,9525 && \text{(Tabelle)} \\
&\approx 0,05
\end{aligned}$$

Mit welcher Wahrscheinlichkeit sind zwischen 35 und 45 Chips defekt?

$$\mathbb{P}\{34 < S_n \leq 45\} = \mathbb{P}\Big\{-1 < \frac{S_n - np}{\sqrt{npq}} \leq \frac{5}{6}\Big\} \approx \Phi(5/6) - \Phi(-1) \ = \Phi(5/6) - (1 - \Phi(+1)) \approx 0,64$$

Beispiel 2
Zwei Kinos A und B konkurrieren um Kunden. Von den 1000 Kunden wollen wir der Einfachheit halber annehmen, daß sie sich unabhängig voneinander und rein zufällig für eins der beiden Kinos entscheiden.

Wie viele Plätze sollte jedes der Kinos haben, damit die Ws., einen Kunden abweisen zu müssen, höchstens 1% beträgt?

- Wahl des Wahrscheinlichkeitsraums (Ω, p):

  $\Omega = \{0,1\}^n$ mit $n = 1000$ und Gleichverteilung

  Dabei bedeute $\omega_i = 1$, daß der i-te Kunde sich für Kino A entscheidet.

- S_n sei die Anzahl Kunden, die sich für Kino A entscheiden, wobei wir insgesamt $n = 1000$ Kunden haben. Wir setzen

  $X_i(\omega) = \omega_i$ für $i = 1, \ldots, n$ und $S_n = \sum_{i=1}^n X_i$

- Somit ist anhand der Unabhängigkeitsannahme S_n binomialverteilt mit $n = 1000$ und $p = 1/2$. Setze $q = 1 - p$. Wir wissen: $\mathbb{E}[S_n] = np = 500$ und $\mathrm{Var}(S_n) = npq = 250$

- Aus Symmetriegründen können wir annehmen, daß beide Kinos gleich viele Plätze haben. Sei N die Anzahl der Plätze, die jedes Kino hat. Wir suchen N. Dabei wissen wir bereits, daß nur $N \geq 500$ sinnvoll ist.

- Alle Kunden sind zufrieden, wenn die folgenden beiden Bedingungen erfüllt sind:

$$S_n \leq N \quad \text{und} \quad n - S_n \leq N \ .$$

Gesucht ist das minimale N derart, daß

$$1 - \mathbb{P}\{S_n \leq N \text{ und } n - S_n \leq N\} \leq 0,01 \ .$$

Wir betrachten die Komplementärwahrscheinlichkeit

$$\begin{aligned}
\mathbb{P}\{n - N \leq S_n \leq N\} &= \mathbb{P}\{n - N - 1 < S_n \leq N\} \\
&= \mathbb{P}\Big\{\frac{n - N - 1 - np}{\sqrt{npq}} < \frac{S_n - np}{\sqrt{npq}} \leq \frac{N - np}{\sqrt{npq}}\Big\} \\
&\approx \Phi\Big(\frac{N - 500}{\sqrt{250}}\Big) - \Phi\Big(\frac{499 - N}{\sqrt{250}}\Big) \\
&= \Phi\Big(\frac{N - 500}{\sqrt{250}}\Big) - 1 + \Phi\Big(\frac{N - 499}{\sqrt{250}}\Big) \ .
\end{aligned}$$

Wären beide Argumente von Φ gleich $z = (N - 500)/\sqrt{250}$, so müßten wir die Lösung von

$$0,99 \leq 2\Phi(z) - 1$$

finden. Es würde $z \geq 2,58$ folgen (mit Hilfe der Tabelle), also

$$N \geq 2,58 \cdot \sqrt{250} + 500 = 540,793 \ldots$$

Probieren zeigt, daß

$$\Phi\Big(\frac{N-500}{\sqrt{250}}\Big)+\Phi\Big(\frac{N-499}{\sqrt{250}}\Big)-1=0,9893\cdots<0,99 \qquad \text{für } N=540$$

und

$$\Phi\Big(\frac{N-500}{\sqrt{250}}\Big)+\Phi\Big(\frac{N-499}{\sqrt{250}}\Big)-1=0,9912\cdots>0,99 \qquad \text{für } N=541$$

Folglich sollte jedes der Kinos 541 Plätze haben.