Weierstraß-Institut für
Angewandte Analysis und Stochastik

Von der Binomialverteilung zur Normalverteilung

Wolfgang König
(WIAS und TU Berlin)

Ziele dieses Vortrags

speziell:

- Eigenschaften der Binomialverteilung
- ihre Konvergenz gegen die Normalverteilung (Satz von MOIVRE-LAPLACE)
- Illustration: GALTONbrett
- Anwendungsaufgaben

Ziele dieses Vortrags

speziell:

- Eigenschaften der Binomialverteilung
- ihre Konvergenz gegen die Normalverteilung (Satz von Moivre-Laplace)
- Illustration: Galtonbrett
- Anwendungsaufgaben

allgemeiner und universeller:

- Summen von unabhängigen Zufallsgrößen
- Abweichungen vom Erwartungswert (Gesetz der Großen Zahlen)
- Konvergenz gegen Normalverteilung (Zentraler Grenzwertsatz)
- Anwendungsaufgaben

Verteilung auf $\{0,\ldots,n\}$. Parameter: $n \in \mathbb{N}$, $p \in (0,1)$.

$$\mathrm{Bin}_{p,n}(k) = \mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k \in \{0,\ldots,n\}.$$

- Verteilung der Anzahl der Erfolge in einer Serie von n Glückspielen mit Erfolgswahrscheinlichkeit p
- Verteilung der Summe von n unabhängigen BERNOULLI-verteilten Zufallsgrößen
- eine der wenigen expliziten Verteilungen einer n-fachen unabhängigen Summe
- Erwartungswert $\mathbb{E}(X) = \sum_{k=0}^n k\mathrm{Bin}_{p,n}(k) = np$.
  Beweis mit Rechnung oder mit obiger Interpretation.
- Varianz $\mathbb{V}(X) = \sum_{k=0}^n k^2\mathrm{Bin}_{p,n}(k) - (np)^2 = np(1-p)$. Beweis dito.
- Falls $X_1 \sim \mathrm{Bin}_{p,n_1}$ und $X_2 \sim \mathrm{Bin}_{p,n_2}$ unabhängig sind,
  so ist $X_1 + X_2 \sim \mathrm{Bin}_{p,n_1+n_2}$.

(Es wird NICHT der POISSON'sche Grenzwertsatz behandelt, wo $np \approx \alpha$ vorausgesetzt wird.)

Wie verhält sich $X \sim \mathrm{Bin}_{p,n}$ für $n \to \infty$?

- Natürlich ist $X \approx \mathbb{E}(X) = np$, aber wie groß ist die Ordnung a_n der Abweichung der Größe X von $\mathbb{E}(X)$ (die Fluktuation)?
- Schwaches Gesetz der Großen Zahlen: Für $\varepsilon > 0$ gilt (TSCHEBYSCHEFF-Ungleichung!)

$$\mathbb{P}\big(|X - \mathbb{E}(X)| \geq \varepsilon n\big) \leq \frac{\mathbb{E}\big(|X - \mathbb{E}(X)|^2\big)}{(n\varepsilon)^2} = \frac{np(1-p)}{(n\varepsilon)^2} = \frac{C}{n} \to 0.$$

(Es wird NICHT der POISSON'sche Grenzwertsatz behandelt, wo $np \approx \alpha$ vorausgesetzt wird.)

Wie verhält sich $X \sim \mathrm{Bin}_{p,n}$ für $n \to \infty$?

- Natürlich ist $X \approx \mathbb{E}(X) = np$, aber wie groß ist die Ordnung a_n der Abweichung der Größe X von $\mathbb{E}(X)$ (die Fluktuation)?
- Schwaches Gesetz der Großen Zahlen: Für $\varepsilon > 0$ gilt (TSCHEBYSCHEFF-Ungleichung!)

$$\mathbb{P}\big(|X - \mathbb{E}(X)| \geq \varepsilon n\big) \leq \frac{\mathbb{E}\big(|X - \mathbb{E}(X)|^2\big)}{(n\varepsilon)^2} = \frac{np(1-p)}{(n\varepsilon)^2} = \frac{C}{n} \to 0.$$

  Also muss $a_n \ll n$ sein.
- Antwort: a_n ist so groß, dass die Varianz von $\frac{1}{a_n}(X - \mathbb{E}(X))$ von endlicher Größenordnung ist, sagen wir: $= 1$.
- Wegen $\mathbb{V}(\frac{1}{a_n}(X - \mathbb{E}(X))) = \frac{1}{a_n^2}\mathbb{V}(x) = \frac{1}{a_n^2} np(1-p)$ ist $a_n = \sqrt{np(1-p)}$.

Verhalten für große n

(Es wird NICHT der POISSON'sche Grenzwertsatz behandelt, wo $np \approx \alpha$ vorausgesetzt wird.)

Wie verhält sich $X \sim \mathrm{Bin}_{p,n}$ für $n \to \infty$?

- Natürlich ist $X \approx \mathbb{E}(X) = np$, aber wie groß ist die Ordnung a_n der Abweichung der Größe X von $\mathbb{E}(X)$ (die Fluktuation)?
- Schwaches Gesetz der Großen Zahlen: Für $\varepsilon > 0$ gilt (TSCHEBYSCHEFF-Ungleichung!)

$$\mathbb{P}\big(|X - \mathbb{E}(X)| \geq \varepsilon n\big) \leq \frac{\mathbb{E}\big(|X - \mathbb{E}(X)|^2\big)}{(n\varepsilon)^2} = \frac{np(1-p)}{(n\varepsilon)^2} = \frac{C}{n} \to 0.$$

  Also muss $a_n \ll n$ sein.
- Antwort: a_n ist so groß, dass die Varianz von $\frac{1}{a_n}\big(X - \mathbb{E}(X)\big)$ von endlicher Größenordnung ist, sagen wir: $= 1$.
- Wegen $\mathbb{V}\big(\frac{1}{a_n}(X - \mathbb{E}(X))\big) = \frac{1}{a_n^2}\mathbb{V}(x) = \frac{1}{a_n^2} np(1-p)$ ist $a_n = \sqrt{np(1-p)}$.

$$\Longrightarrow \qquad \text{Vermutung:} \qquad \frac{X - np}{\sqrt{np(1-p)}} \qquad \text{konvergiert.}$$

Die Limes-Zufallsgröße sollte Erwartungswert Null und Varianz Eins haben.

Aber was sollte sie sein? Und was soll "konvergiert" heißen?

Relevante Übungsaufgaben

(mit den üblichen Voraussetzungen: Unabhängigkeit, exakte Verteilung etc.)

Würfeln

Wenn man einen fairen Würfel 1200 Mal wirft, wie groß ist dann die Wahrscheinlichkeit, dass dabei zwischen 190 und 200 Sechsen fallen?

Relevante Übungsaufgaben

(mit den üblichen Voraussetzungen: Unabhängigkeit, exakte Verteilung etc.)

Würfeln

Wenn man einen fairen Würfel 1200 Mal wirft, wie groß ist dann die Wahrscheinlichkeit, dass dabei zwischen 190 und 200 Sechsen fallen?

Wahlprognose

Bei einer Wahl erhält Kandidat A einen unbekannten Anteil p der Stimmen. Um den Wert von p zu ermitteln, werten wir die ersten n Wahlzettel aus (bzw. befragen wir n zufällig gewählte Wähler). Wie groß sollte n sein, damit die Wahrscheinlichkeit eines Irrtums von mehr als einem Prozent nicht größer als 0,05 ist?

Relevante Übungsaufgaben

(mit den üblichen Voraussetzungen: Unabhängigkeit, exakte Verteilung etc.)

Würfeln

Wenn man einen fairen Würfel 1200 Mal wirft, wie groß ist dann die Wahrscheinlichkeit, dass dabei zwischen 190 und 200 Sechsen fallen?

Wahlprognose

Bei einer Wahl erhält Kandidat A einen unbekannten Anteil p der Stimmen. Um den Wert von p zu ermitteln, werten wir die ersten n Wahlzettel aus (bzw. befragen wir n zufällig gewählte Wähler). Wie groß sollte n sein, damit die Wahrscheinlichkeit eines Irrtums von mehr als einem Prozent nicht größer als 0,05 ist?

Beide Aufgaben können praktisch approximativ gelöst werden mit Hilfe des folgenden Satzes und einer Tabelle für die Werte der Verteilungsfunktion der Normalverteilung,

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-\frac{1}{2}x^2} \, \mathrm{d}x.$$

Satz von MOIVRE-LAPLACE

Wenn $X \sim \mathrm{Bin}_{p,n}$ ist, so konvergiert $\frac{X-np}{\sqrt{np(1-p)}}$ gegen die Standard-Normalverteilung $\mathcal{N}$.

Das heißt, für alle $a < b$ gilt

$$\lim_{n\to\infty} \mathbb{P}\left(a \leq \frac{X-np}{\sqrt{np(1-p)}} \leq b\right) = \mathbb{P}(\mathcal{N} \in [a,b]) = \Phi(b) - \Phi(a) = \int_a^b \frac{\mathrm{e}^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}\,\mathrm{d}x.$$

Beweis: Stirling-Formel $n! \sim \binom{n}{\mathrm{e}}\sqrt{2\pi n}$, Taylor-Approximation $\log(1+\varepsilon) \approx \varepsilon - \frac{1}{2}\varepsilon^2$ und Exponentialsatz $(1+\frac{c}{n})^n \to \mathrm{e}^c$ und etwas Rechnung.

Satz von Moivre-Laplace

Wenn $X \sim \mathrm{Bin}_{p,n}$ ist, so konvergiert $\frac{X-np}{\sqrt{np(1-p)}}$ gegen die Standard-Normalverteilung $\mathcal{N}$.

Das heißt, für alle $a < b$ gilt

$$\lim_{n\to\infty} \mathbb{P}\Big(a \leq \frac{X-np}{\sqrt{np(1-p)}} \leq b\Big) = \mathbb{P}(\mathcal{N} \in [a,b]) = \Phi(b) - \Phi(a) = \int_a^b \frac{\mathrm{e}^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}\,\mathrm{d}x.$$

Beweis: Stirling-Formel $n! \sim \left(\frac{n}{\mathrm{e}}\right)\sqrt{2\pi n}$, Taylor-Approximation $\log(1+\varepsilon) \approx \varepsilon - \frac{1}{2}\varepsilon^2$ und Exponentialsatz $(1+\frac{c}{n})^n \to \mathrm{e}^c$ und etwas Rechnung.

Aber warum die Normalverteilung?

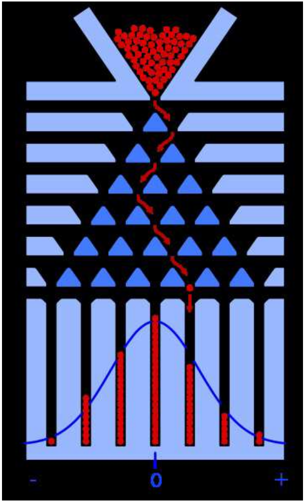
Im Satz von Moivre-Laplace sieht man es nicht, aber der Grenzwert ist universell:

Er hängt nicht von der Binomialverteilung ab!

Wir sind einem der wichtigsten Prinzipien der Wahrscheinlichkeitstheorie auf der Spur!

Visualisierung mit dem Galtonbrett

- N Kugeln fallen durch ein regelmäßiges Muster von Hindernissen. Jedes Mal werden “links” und “rechts” mit gleicher Wahrscheinlichkeit unabhängig gewählt.
- Die Schritte sind also gleichverteilt auf $\{-1, 1\}$ (standardisiert).
- Die Summe Y der Abweichungen vom Startort zeigt sich unten im Ort der Kugel nach ihrem Lauf. Es ist $Y = 2X - n$ mit $X \sim \mathrm{Bin}_{n,\frac{1}{2}}$ und $n = 6$.
- Nach dem Satz von MOIVRE-LAPLACE ist $Y = 2(X - \frac{n}{2}) \approx 2(\sqrt{n}\frac{1}{2}\mathcal{N}) = \sqrt{6}\mathcal{N}$.
- Die Glockenkurve zeigt sich im Histogramm von N unabhängigen Wiederholungen.



Aber wie wurde MOIVRE-LAPLACE wirklich angewendet? Auf N oder auf $n = 6$?

Experimentelle Erfahrung am Histogramm

Das Histogramm von vielen unabhängigen Wiederholungen einer beliebigen Zufallsgröße approximiert die Dichte der Standardnormalverteilung:

Satz von MOIVRE-LAPLACE für Histogramme

Wenn $Y_1, Y_2, \dots, Y_N$ eine Folge von N unabhängigen, identisch verteilten Zahlen ist, so ist für jedes $a < b$ die Anzahl der k mit $Y_k \in [a, b]$ ungefähr Np plus $\sqrt{Np(1-p)}$ Mal eine standardnormalverteilte Zufallsgröße, wobei $p = \mathbb{P}(Y_1 \in [a, b])$.

(Zum Beweis wendet man den Satz von MOIVRE-LAPLACE an auf die Indikatorgrößen $1\!\!1_{\{Y_k \in [a,b]\}}$, denn ihre Summe ist die obige Anzahl.)

Das Histogramm von vielen unabhängigen Wiederholungen einer beliebigen Zufallsgröße approximiert die Dichte der Standardnormalverteilung:

Satz von Moivre-Laplace für Histogramme

Wenn $Y_1, Y_2, \dots, Y_N$ eine Folge von N unabhängigen, identisch verteilten Zahlen ist, so ist für jedes $a < b$ die Anzahl der k mit $Y_k \in [a, b]$ ungefähr Np plus $\sqrt{Np(1-p)}$ Mal eine standardnormalverteilte Zufallsgröße, wobei $p = \mathbb{P}(Y_1 \in [a, b])$.

(Zum Beweis wendet man den Satz von Moivre-Laplace an auf die Indikatorgrößen $\mathbb{1}_{\{Y_k \in [a,b]\}}$, denn ihre Summe ist die obige Anzahl.)

Ist das das Prinzip hinter dem Galtonbrett?

Nein, nicht ganz, wir verwenden nur das Gesetz der Großen Zahlen:

Wende den obigen Satz für $i = -3, -2, \dots, 3$ an auf $p_i = \mathbb{P}(Y = i) = \mathrm{Bin}_{6,\frac{1}{2}}(2i) - \frac{n}{2}$, dann sammeln sich in der i-ten Box etwa Np_i Kugeln. Die p_i approximieren die Dichte von $\sqrt{6}\mathcal{N}$ nach dem Satz von Moivre-Laplace für $n = 6$.

Man muss N mit $n = 6$ geschickt koppeln!

Universalität: Warum die Normalverteilung?

Welche Verteilung mit Erwartungswert Null und Varianz Eins hat die Eigenschaft, dass für unabhängige Realisationen $X_1, \dots, X_n$ gilt:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \qquad \text{hat die selbe Verteilung wie} \quad X_1?$$

Antwort (unter technischen Voraussetzungen): nur die Standardnormalverteilung $\mathcal{N}$! Denn nur sie hat die Eigenschaft: Falls $X_1, X_2 \sim \mathcal{N}$ unabhängig sind, so ist $X_1 + X_2 \sim \sqrt{2}\mathcal{N}$.

Welche Verteilung mit Erwartungswert Null und Varianz Eins hat die Eigenschaft, dass für unabhängige Realisationen $X_1, \dots, X_n$ gilt:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \qquad \text{hat die selbe Verteilung wie} \quad X_1?$$

Antwort (unter technischen Voraussetzungen): nur die Standardnormalverteilung $\mathcal{N}$! Denn nur sie hat die Eigenschaft: Falls $X_1, X_2 \sim \mathcal{N}$ unabhängig sind, so ist $X_1 + X_2 \sim \sqrt{2}\mathcal{N}$.

Zentraler Grenzwertsatz

Seien $X_1, \dots, X_n$ unabhängige, identisch verteilte Zufallsgrößen mit Erwartungswert Null und Varianz Eins, dann konvergiert $n^{-1/2}(X_1 + \dots + X_n)$ in Verteilung gegen $\mathcal{N}$, das heißt, für jedes $a < b$ gilt

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \in [a,b]\Big) = \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_a^b \mathrm{e}^{-\frac{1}{2}x^2}\, \mathrm{d}x.$$

Universelle Aussage: Die unabhängige Überlagerung von n standardisierten Fehlern ist ungefähr normalverteilt mit Varianz n.

Der Zentrale Grenzwertsatz ist eine der fundamentalsten Aussagen der W'Theorie!

Mit charakteristischen Funktionen:

Wenn für jedes $t \in \mathbb{R}$ gilt: $\lim_{n\to\infty} \mathbb{E}(\mathrm{e}^{\mathrm{i}tZ_n}) = \mathbb{E}(\mathrm{e}^{\mathrm{i}t\mathcal{N}})$, dann konvergiert Z_n gegen $\mathcal{N}$.

Man rechnet mit Hilfe der Unabhängigkeit und einer Taylor-Approximation:

$$\begin{aligned}
\mathbb{E}(\mathrm{e}^{\mathrm{i}t(X_1+\dots+X_n)/\sqrt{n}}) &= \mathbb{E}\Big(\prod_{i=1}^{n} \mathrm{e}^{\mathrm{i}tX_i/\sqrt{n}}\Big) \\
&= \prod_{i=1}^{n} \mathbb{E}\big(\mathrm{e}^{\mathrm{i}tX_i/\sqrt{n}}\big) \\
&= \Big(\mathbb{E}\big(\mathrm{e}^{\mathrm{i}tX_1/\sqrt{n}}\big)\Big)^n \\
&\approx \Big(1 + \frac{\mathrm{i}t}{\sqrt{n}}\mathbb{E}(X_1) + \frac{1}{2}\Big(\frac{\mathrm{i}t}{\sqrt{n}}\Big)^2 \mathbb{E}(X_1^2)\Big)^n \\
&= \Big(1 - \frac{t^2}{2n}\Big)^n,
\end{aligned}$$

und dies konvergiert nach dem Exponentialsatz gegen $\mathrm{e}^{-t^2/2} = \mathbb{E}(\mathrm{e}^{\mathrm{i}t\mathcal{N}})$.

Eine Aufgabe zum ZGWS, die nicht mit dem Satz von MOIVRE-LAPLACE gelöst werden kann:

Rundungsfehler

Eine Bank tauscht hundert Millionen Geldbeträge von DM in Euro gemäß der Regel $1\,\mathrm{Euro} = 1.95583\ \mathrm{DM}$ um. Das Euro-Ergebnis wird auf ganze Cent gerundet.

(Beispiel: $50\,\mathrm{DM} = 25.56459406\,\mathrm{Euro} \approx 25.56\,\mathrm{Euro}$).

Wie ist der Gesamt-Rundungsfehler verteilt?

Innerhalb welcher Schranken bleiben die Rundungsfehler mit einer Wahrscheinlichkeit von 99 Prozent?

Beispiel

Eine Aufgabe zum ZGWS, die nicht mit dem Satz von MOIVRE-LAPLACE gelöst werden kann:

Rundungsfehler

Eine Bank tauscht hundert Millionen Geldbeträge von DM in Euro gemäß der Regel $1\,\mathrm{Euro} = 1.95583\ \mathrm{DM}$ um. Das Euro-Ergebnis wird auf ganze Cent gerundet.

(Beispiel: $50\,\mathrm{DM} = 25.56459406\,\mathrm{Euro} \approx 25.56\,\mathrm{Euro}$).

Wie ist der Gesamt-Rundungsfehler verteilt?

Innerhalb welcher Schranken bleiben die Rundungsfehler mit einer Wahrscheinlichkeit von 99 Prozent?

Hinweise zur Lösung: Annahmen: (1) Jeder einzelne Rundungsfehler (in Euro) ist eine auf $[-0.005, 0.005]$ gleichverteilte Zufallsgröße, und (2) die 10^8 Fehler sind unabhängig.

Dann ist der Gesamtrundungsfehler (in Euro) gleich $R = X_1 + \cdots + X_{10^8}$, wobei $X_1, \ldots, X_{10^8}$ unabhängige, auf $[-0.005, 0.005]$ gleichverteilte Zufallsgrößen sind. Sie haben den Erwartungswert Null und die Varianz $\frac{1}{3}0.005^2 = \frac{1}{12}10^{-4}$. Nach dem ZGWS gilt $R \approx \sqrt{10^8}\frac{1}{\sqrt{12}}10^{-2}\mathcal{N} = \frac{100}{\sqrt{12}}\mathcal{N}$.

Die Zusatzfrage löst man approximativ mit Tabellen der Werte der Verteilungsfunktion Φ; eine Schranke ist etwa 74 Euro.