Eine Abschätzung zur Binomialverteilung

Bernd Wollring, Münster

Zusammenfassung : Verwendet man für Tests oder Schätzungen zur Binomialverteilung die TSCHEBYSCHEW-Ungleichung, so erhält man ungünstige Abschätzungen mit schlechten Genauigkeiten oder großen Serienlängen. Approximiert man durch die Normalverteilung, erhält man günstigere Konditionen, muß aber Anpassungsbedingungen durch a priori Annahmen an p erfüllen, etwa $n \circ p \circ (1 - p) > 9$. Eine weitgehend unbekannte Abschätzung von HOEFFDING ist besser als die TSCHEBYSCHEW-Ungleichung, zwar schlechter als die bei Approximation durch Normalverteilung, erfordert dafür aber keine Bedingungen an p. Sie ermöglicht neues Nachdenken über den Umgang mit a priori Wahrscheinlichkeiten. Wir beweisen sie mit elementaren Mitteln, so daß eine Diskussion in der Schule im Zusammenhang mit dem Analysisunterricht zumindest möglich erscheint.

1 Abschätzungen zur Binomialverteilung

Ein von OKAMOTO und HOEFFDING seit 1958 bzw. 1963 publiziertes Ergebnis wollen wir neu erschließen, da es unseres Erachtens wenig bekannt, aber für den Unterricht möglicherweise interessant und mit hinreichend elementaren Mitteln zugänglich ist : Wir beweisen eine Abschätzung für Summen beschränkter unabhängiger Zufallsgrößen, die effizient auf die Binomialverteilung anzuwenden ist. Um Konvergenzprobleme auszuschließen, setzen wir alle Zufallsgrößen als reellwertig und endlich voraus.

1.1 Bekannt : Die Ungleichung von BIENAYMÉ und TSCHEBYSCHEW

Von so zentraler Bedeutung, daß man sie im Stochastikunterricht in jedem Fall diskutieren und wenn möglich auch beweisen sollte, ist die von BIENAYMÉ im Jahre 1853 und TSCHEBYSCHEW im Jahre 1866 bewiesene Ungleichung :

Satz 1 Ungleichung von BIENAYMÉ und TSCHEBYSCHEW

Ist X eine Zufallsgröße mit Erwartungswert μ und Varianz σ^2, so gilt für jedes $\varepsilon > 0$ die Ungleichung :

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Der Beweis basiert auf einer geeigneten Zerlegung der Varianz, die im Unterricht durchführbar ist und typisch für eine ganze Klasse analoger Abschätzungen, der "Momentenungleichungen" ist (siehe etwa das Schulbuch von BARTH und HALLER, [1], S. 184).

Eine binomial nach B(n,p) verteilte Zufallsgröße X läßt sich stets deuten als

relative Häufigkeit der "Treffer" bei n unabhängigen Einzelversuchen mit "Trefferchance" p im Einzelversuch. Wie bekannt, erhält man Erwartungswert und Varianz bequem, wenn man X als Mittelwert $X = (X_1 + .. + X_n)/n$ der n unabhängigen Zufallsgrößen

$$X_i = \text{Trefferzahl im i-ten Einzelversuch}$$

mit Werten in $\{0,1\}$ betrachtet, für die man elementar findet :

$$\mu_i = EX_i = p \quad , \quad \sigma^2{}_i = \sigma^2 X_i = p(1-p)$$

Die Gleichungen

$$\mu = EX = p \quad , \quad \sigma^2 = \sigma^2 X = \frac{p(1-p)}{n}$$

folgen mit den Rechenregeln für Erwartungswerte und Varianzen, denn Erwartungswerte sind linear, und Varianzen sind homogen vom Grad 2 und für unabhängige Zufallsgrößen additiv. Zusammen mit der Abschätzung $p - (1-p) \leq 1/4$ für alle $p \in [0,1]$ erhalten wir aus Satz 1 :

Satz 2 Ungleichung von BIENAYMÉ und TSCHEBYSCHEW für binomial verteilte Zufallsgrößen

Ist X eine binomial nach $B(n,p)$ verteilte Zufallsgröße, so gilt für jedes $\varepsilon > 0$ die Ungleichung :

$$P(|X-p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

Auf dieser Abschätzung lassen sich Hypothesentests und Parameterschätzungen zu binomial verteilten Zufallsgrößen aufbauen, ferner führt sie direkt auf das von JACOB BERNOULLI wohl bereits um 1685 gefundene schwache Gesetz der großen Zahlen :

$$\lim_{n \to \infty} P(|X-p| < \varepsilon) = 1 \quad \text{für jedes } \varepsilon > 0$$

Das Rekapitulieren des bekannten Beweises soll darauf verweisen, daß wesentliche seiner Argumente im Beweis der folgenden stärkeren Sätze dieselben sind, einige analytische Argumente werden hinzukommen.

Der Erkenntniswert der Sätze im Unterricht ist hoch, da sie ein Gesetz der großen Zahlen erschließen. Die Varianzzerlegung und die Abschätzung sind aber derart "grob", daß sich bei praktischen Anwendungen als Schwäche zeigt, daß 22das TSCHEBYSCHEW-Risiko $1/4n\varepsilon^2$ ungünstig groß ist. Fordert man wie

üblich, daß das Risiko 5% nicht überschreitet, so folgt mit

$$P(|X-p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2} \leq 0.05$$

aus Satz 2 die n-ε-Bedingung :

$$\textbf{"5\%-Regel T"}: \quad 5 < n\varepsilon^2$$

Die Konstante 5 ist allerdings so hoch, daß man diese Abschätzung bei praktischen Anwendungen oft gar nicht erst in Betracht zieht. Könnte man eine entsprechende Abschätzung mit einer kleineren positiven Konstanten auf der linken Seite beweisen, so wäre dies bei Anwendungen ein großer Vorteil.

Stattdessen wählt man in der Praxis häufig den Weg, die Binomialverteilung durch die Normalverteilung zu "approximieren". Wird akzeptiert, daß man "brauchbare Werte" erhält, wenn die Approximationsbedingung $n \cdot p \cdot (1-p) > 9$ erfüllt ist (siehe etwa [1], S. 291), so findet man die als "2σ-Regel" bekannte Abschätzung :

$$P(|X-p| \geq \varepsilon) \leq 0.05 \quad \text{für} \quad \varepsilon \geq 1.96\sqrt{\frac{p(1-p)}{n}}$$

Dies ist erfüllt für

$$\varepsilon \geq 1.96\sqrt{\frac{1}{4n}} \geq 1.96\sqrt{\frac{p(1-p)}{n}}$$

wegen $p - (1-p) \leq 1/4$ für alle $p \in [0,1]$ und führt auf die n-ε-Bedingung:

$$\textbf{"5\%-Regel N"}: \quad 1 \leq n\varepsilon^2$$

Damit erhält man bei Anwendungen wesentlich günstigere Konditionen.

Problematisch bleibt allerdings, daß man Art und Güte der Approximation letztlich nicht elementar erläutern kann. *"Eine theoretische Untersuchung des Fehlers übersteigt unsere Möglichkeiten"* heißt es denn auch entwaffnend ehrlich in einem anspruchsvollen Schulbuch ([1], S. 293) Insbesondere hat man keine Aussage, die gleichmäßig für alle $p \in [0,1]$ gilt. Man steht damit vor einem zentralen Problem, das im Unterricht in jedem Fall zu diskutieren ist, wenn man Approximationen durch die Normalverteilung benutzt : **Die Approximationsbedingung erfordert a priori Annahmen an die zu schätzende Größe.** Bei einer Schätzung des Parameters p sind zur Bestimmung hinreichender Serienlängen n somit a priori Annahmen an p zu stellen, die zu begründen sind. Bisweilen ist dies bei Anwendungen kein Problem, man kann eventuell mit plausiblen

Annahmen $0 < p_u < p$ und $p < p_o < 1$ starten. Aber es gibt auch Fragestellungen, bei denen derartige a priori Annahmen an p problematisch erscheinen. Man stößt auf die grundsätzliche Frage : Soll man den BAYESschen Standpunkt einnehmen, demzufolge jede Schätzung im Prinzip die Verbesserung von a priori Annahmen bedeutet, oder sind a priori Annahmen bei manchen oder allen Fragestellungen grundsätzlich abzulehnen ?

1.2 Ein Kompromiß : Ungleichung von OKAMOTO und HOEFFDING

Angesichts der skizzierten Alternative zwischen der TSCHEBYSCHEW-Ungleichung mit "5%-Regel T" und der Approximation durch die Normalverteilung mit Anpassungsbedingung und "5%-Regel N" erscheint es für manche Anwendungen hilfreich, über eine Abschätzung zu verfügen, die einerseits zu besseren Konditionen als "5%-Regel T" führt, andererseits nicht die in der Anpassungsbedingung zur "5%-Regel N" implizierten a priori Annahmen erfordert. Eine solche bietet bei akzeptablen Voraussetzungen folgender Satz :

Satz 3 Ungleichung von OKAMOTO und HOEFFDING für binomial verteilte Zufallsgrößen

Ist X eine binomial nach B(n,p) verteilte Zufallsgröße, so gilt für jedes $\varepsilon > 0$ die Ungleichung :

$$P(|X - p| \geq \varepsilon) \leq \frac{2}{e^{2n\varepsilon^2}}$$

Dieses Ergebnis wurde 1958 von OKAMOTO veröffentlicht (siehe [3]). Es ist ein Spezialfall des folgenden Satzes :

Satz 4 Ungleichung von HOEFFDING für Summen beschränkter unabhängiger Zufallsgrößen

Sind $X_1, ..., X_n$ unabhängige Zufallsgrößen mit dem Erwartungswert μ , die zudem beschränkt sind gemäß $0 \leq X_i \leq 1$ für alle i, ist ferner $X = (X_1 + .. + X_n)/n$ ihr arithmetisches Mittel (mit dem Erwartungswert μ), so gilt für alle positiven ε :

$$P(|X - \mu| \geq \varepsilon) \leq \frac{2}{e^{2n\varepsilon^2}}$$

Diesen Satz publizierte HOEFFDING 1963 (siehe [2]). Er läßt sich verallgemeinern für unabhängige Zufallsgrößen mit verschiedenen Erwartungswerten und

anderen Schranken. Die angegebene obere Schranke kann verbessert werden, eine optimale Schranke nennt HOEFFDING (vgl. Theorem 1 in [2], S. 15). Entscheidende Voraussetzung in Satz 4 ist die Annahme der Beschränktheit der unabhängigen Zufallsgrößen. Die Voraussetzung der einheitlichen Beschränktheit auf [0,1] wurde nur aus technischen Gründen getroffen, sie erlaubt einen Beweis mit einigermaßen elementaren Hilfsmitteln aus der Analysis, der im Unterricht zugänglich ist. Wir führen diesen gegenüber dem allgemeineren Beweis bei HOEFFDING reduzierten Beweis in Abschnitt 2 vollständig aus.

Für Anwendungen ist das HOEFFDING-Risiko $2/\exp(-2n\varepsilon^2)$ günstiger als das TSCHEBYSCHEW-Risiko $1/4n\varepsilon^2$. Verlangt man, daß das Risiko 5% nicht überschreitet, so folgt mit

$$P(|X - p| \geq \varepsilon) \leq \frac{2}{e^{2n\varepsilon^2}} \leq 0.05$$

via $(\ln(2/0.05))/2 \leq 1.845$ aus Satz 3 die n-ε-Bedingung :

"5%-Regel H" : $1.845 < n\,\varepsilon^2$

Dieses Ergebnis findet man bei WARMUTH (1991, S.166). Es ist eine deutliche Verbesserung gegenüber der "5%-Regel T" $5 < n\,\varepsilon^2$, und es gilt gleichmäßig für alle $p \in [0,1]$.

1.3 Beispiele

Alle drei 5%-Regeln haben dieselbe Struktur $C < n\,\varepsilon^2$. Die folgenden kurzen Tabellen zeigen, wie sich die unterschiedlichen Konstanten C in der Praxis auswirken.

Die erste Tabelle gibt an, wie genau man jeweils bei gegebener Länge n der Versuchsreihe schätzen kann.

Man beachte, daß bei den Ungleichungen

$$X - \varepsilon < p < X + \varepsilon \quad \Leftrightarrow \quad |X - p| < \varepsilon \quad \Leftrightarrow \quad p - \varepsilon < X < p + \varepsilon$$

ε nur die halbe Länge des jeweiligen Intervalls ist.

Die zweite kurze Tabelle zeigt, welche Serienlängen bei geforderter Genauigkeit ε erforderlich sind.

Serienlänge n	5%-Regel T $5 < n\,\varepsilon^2$	5%-Regel H $1.845 < n\,\varepsilon^2$	5%-Regel N $1 < n\,\varepsilon^2$
10 Versuche	$\varepsilon \geq 0.707$	$\varepsilon \geq 0.430$	$\varepsilon \geq 0.317$
50 Versuche	$\varepsilon \geq 0.317$	$\varepsilon \geq 0.193$	$\varepsilon \geq 0.142$
60 Versuche	$\varepsilon \geq 0.289$	$\varepsilon \geq 0.176$	$\varepsilon \geq 0.130$
300 Versuche	$\varepsilon \geq 0.129$	$\varepsilon \geq 0.079$	$\varepsilon \geq 0.058$

Genauigkeit ε	5%-Regel T $5 < n\,\varepsilon^2$	5%-Regel H $1.845 < n\,\varepsilon^2$	5%-Regel N $1 < n\,\varepsilon^2$
$\varepsilon = 0.2$	$n \geq 125$	$n \geq 47$	$n \geq 25$
$\varepsilon = 0.1$	$n \geq 500$	$n \geq 185$	$n \geq 100$
$\varepsilon = 0.05$	$n \geq 2000$	$n \geq 738$	$n \geq 400$
$\varepsilon = 0.01$	$n \geq 50000$	$n \geq 18445$	$n \geq 10000$

Bei der "5%-Regel N" ist n aufgrund der a priori Annahme eventuell höher zu wählen.

Beispiele findet man, indem man die üblichen Lehrbuchbeispiele mit der "5%-Regel H" neu diskutiert und die Ergebnisse mit den von der "5%-Regel N" oder den von der "5%-Regel T" stammenden vergleicht.

Wir analysieren hier als Beispiel einige Simulationsergebnisse zum "Drei-Türen-Problem" :

Ein Spieler steht vor drei Türen. Hinter einer steckt ein Gewinn, hinter den anderen Nieten. Gespielt wird so :

S0 Der Spielleiter hat Gewinn und Nieten hinter die Türen gesetzt.
S1 Der Spieler wählt eine der drei Türen, alle Türen bleiben noch zu.
S2 Der Spielleiter öffnet eine Tür vor einer Niete, die gewählte Tür bleibt noch zu.

S3 Der Spieler bleibt bei der gewählten Tür (Nichtwechsel) oder wählt die andere geschlossene Tür (Wechsel).

Ist seine Gewinnchance bei Nichtwechsel oder bei Wechsel höher ?

Da die "5%-Regel H" die beste uns bekannte ist, bei der keine a priori Annahmen erforderlich sind, schätzen wir in [5] die Gewinnchance p_W bei Wechsel und die Gewinnchance p_{NW} bei Nichtwechsel aus Versuchsserien (siehe WOLLRING 1992). Genau die Spiele, die man bei Wechseln gewinnt, verliert man bei Nichtwechsel, also $p_W + p_{NW} = 1$.

- SANDRA und SYLVIA erhielten in 60 Versuchen bei Wechsel 38 Gewinne und 22 Verluste und folgerten, die Gewinnchance bei Wechsel sei größer. Nach der "5%-Regel H" sind aber p_W und p_{NW} daraus nur mit der Genauigkeit $\varepsilon = 0.175$ zu schätzen :

  $$0.192 < p_{NW} < 0.542 \quad \text{und} \quad 0.458 < p_W < 0.808$$

  Und $p_{NW} = 1/2$ und $p_W = 1/2$ können aufgrund dieser Versuchsergebnisse nicht abgelehnt werden. Dies wurde erst akzeptiert, als die Studierenden die Ergebnisse mehrerer Serien der Länge 60 mit verschiedenen Ergebnissen nebeneinander sahen und auffaßten, daß ihr Serienergebnis eines von vielen möglichen war.

- DIRK und RUTH erhielten in 60 Versuchen bei Wechsel 32 Gewinne und 28 Verluste und hielten daher die Gewinnchance bei Wechsel und bei Nichtwechsel für gleich groß, die kleine Abweichung sei "Zufall". Die Daten sind authentisch, sie überraschten nicht nur den Autor. Wie oben sind nach der "5%-Regel H" aber p_W und p_{NW} daraus nur mit der Genauigkeit $\varepsilon = 0.175$ zu schätzen :

  $$0.292 < p_{NW} < 0.642 \quad \text{und} \quad 0.358 < p_W < 0.708$$

  So können $p_{NW} = 1/3$ und $p_W = 2/3$ aufgrund dieser Versuchsergebnisse nicht abgelehnt werden. Auch dies wurde erst nach Vergleich mehrerer Serien der Länge 60 mit verschiedenen Ergebnissen akzeptiert. RUTH hatte zu Beginn der Simulationen den Standpunkt der "Nichtwechsler" vertreten, sie hielt es für kaum möglich, daß ihr Ergebnis bei Vorliegen von $p_{NW} = 1/3$ und $p_W = 2/3$ auftritt.

Bei Versuchsserien mit der Serienlänge 300 erzielt man bei einer Irrtumswahrscheinlichkeit von 5% eine Genauigkeit der Schätzungen von $\varepsilon = 0.079$. Beispiele :

Team I : $X = 96/300$ $P(p_{NW} \in [0.242, 0.398]) \geq 95\%$

Team II : $X = 91/300$ $P(p_{NW} \in [0.225, 0.381]) \geq 95\%$

Team III : $X = 91/300$ $P(p_{NW} \in [0.225,0.381]) \geq 95\%$

Team IV : $X = 82/300$ $P(p_{NW} \in [0.195,0.351]) \geq 95\%$

Investiert man die a priori Annahmen $1/3 \leq p_W \leq 2/3$ und $1/3 \leq p_{NW} \leq 2/3$, wie vorgeschlagen wurde, so gilt die "5%-Regel N" für Versuchsserien der Länge $n \geq 41$, und wir hätten die Genauigkeit von $\varepsilon = 0.079$ schon mit Serienlängen ab 161 erreichen können. Wir haben bei den Simulationen keine a priori Bedingung akzeptiert.

2 Beweis der Ungleichung von HOEFFDING

2.1 Differenzabschätzung statt Abstandsabschätzung

Anders als im Beweis der TSCHEBYSCHEW-Ungleichung, der mit Abständen argumentiert, werden im folgenden Wahrscheinlichkeiten für Differenzen abgeschätzt. Es gilt :

$$P(|X - \mu| \geq \varepsilon) = P(X - \mu \geq \varepsilon) + P(-X + \mu \geq \varepsilon)$$

Bei geeigneten Voraussetzungen sind beide Summanden durch dieselbe Schranke $C(n,\varepsilon)$ nach oben abzuschätzen (siehe 2.3). Somit folgt :

$$P(|X - \mu| \geq \varepsilon) \leq 2\, C(n,\varepsilon)$$

2.2 Abschätzung durch Produktzerlegung

Zunächst notieren wir die gefragte Wahrscheinlichkeit als eine für die Nichtnegativität einer anderen Zufallsgröße :

$$\begin{aligned} P(X - \mu \geq \varepsilon) &= P(X - \mu - \varepsilon \geq 0) \\ &= P(\Sigma X_i - n\mu - n\varepsilon \geq 0) \end{aligned}$$

Die rechte Wahrscheinlichkeit stellen wir nach einer auf S.N.BERNSTEIN zurückgehenden Methode als Erwartungswert einer "Indikator-Zufallsgröße" dar, und den schätzen wir dann unter Ausnutzen der Eigenschaften von Erwartungswerten geeignet nach oben ab. Wir betrachten die Zufallsgröße I mit :

$$\begin{aligned} I(\bullet,\varepsilon) &= 1 \quad , \text{falls } \Sigma X_i - n\mu - n\varepsilon \geq 0 \\ I(\bullet,\varepsilon) &= 0 \quad , \text{falls } \Sigma X_i - n\mu - n\varepsilon < 0 \end{aligned}$$

Die gefragte Wahrscheinlichkeit ist ihr Erwartungswert :

$$P(X - \mu \geq \varepsilon) = E\, I(\bullet,\varepsilon)$$

Entscheidende Idee dieses Beweisteils ist die Einführung der Abschätzung :

$$I(\bullet,\varepsilon) \leq e^{\Sigma X_i - n\mu - n\varepsilon}$$

Es gilt sogar für jedes positive h :

$$I(\bullet,\varepsilon) \leq e^{h(\Sigma X_i - n\mu - n\varepsilon)}$$

Denn h steuert "Verbiegungen" des Graphen der e-Funktion mit Fixpunkt (0,1) , die die Abschätzung nicht beeinflussen **(siehe Abb. 1).**

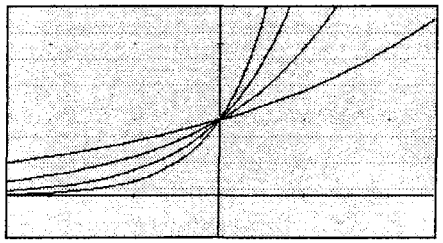

Abb. 1: Graphen zu $x \to \exp(x)$ für $h \in \{0.3, 0.6, 1, 1.6\}$

Aus den Eigenschaften der e-Funktion und der Monotonie und Linearität des Erwartungswertes erhalten wir :

$$\begin{aligned} P(X - \mu \geq \varepsilon) = E\, I(\bullet,\varepsilon) &\leq E\, e^{h(\Sigma X_i - n\mu - n\varepsilon)} \\ &= e^{-hn\varepsilon} \cdot E \prod_i e^{hX_i - h\mu} \end{aligned}$$

Sind nun die X_i stochastisch unabhängig, so kann man das Bilden des Produktes mit dem des Erwartungswertes vertauschen, und wir finden folgendes Zwi-

schenergebnis :

Satz 5 Abschätzung durch Produktzerlegung

Sind $X_1 ,.., X_n$ stochastisch unabhängige Zufallsgrößen mit dem Erwartungswert μ, ist ferner $X = (X_1 +..+ X_n)/n$ ihr arithmetisches Mittel, so gilt für alle positiven ε und alle positiven h :

$$P(X - \mu \geq \varepsilon) \leq e^{-hn\varepsilon} \cdot \prod_i E\, e^{hX_i - h\mu}$$

2.3 Abschätzung bei beschränkten Zufallsgrößen

Wenden wir die Abschätzung durch Produktzerlegung nun auf beschränkte Zufallsgrößen X_i an, so können wir mit gar nicht mal abseitigen Mitteln der Analysis Schranken für die Erwartungswerte in dem rechten Produkt gewinnen und diese bezüglich h minimieren. Die Anwendung auf die Binomialverteilung im Blick setzen wir $0 \leq X_i \leq 1$ für alle i voraus. Zunächst gilt :

$$E\, e^{hX_i - h\mu} = e^{-h\mu} \cdot E\, e^{hX_i}$$

Entscheidendes Hilfsmittel bei der Abschätzung des rechten Faktors ist nun die Konvexität der e-Funktion : Jede Sekante ihres Graphen verläuft zwischen den Schnittpunkten oberhalb des Graphen. Dies genau drückt die JENSENsche Ungleichung aus, etwa für eine Sekante mit den Schnittstellen 0 und h (Abb. 2):

:

$$e^{hx} \leq (1 - x)e^0 + xe^h \quad \text{für} \quad 0 \leq x \leq 1$$

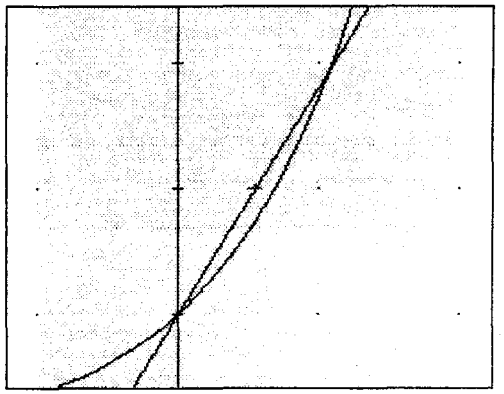

Abb. 2: Graph zu exp(x) mit Sekante

Man kann dies im Unterricht dem Bild glauben oder mit der zweiten Ableitung beweisen (siehe 3.1). Genau hier geht nun die angenommene Beschränktheit $0 \leq X_i \leq 1$ ein, für die Zufallsgrößen hX_i gilt die Abschätzung :

$$e^{hX_i} \leq (1 - X_i) + X_i e^h$$

Und überträgt sich auf ihre Erwartungswerte :

$$E\, e^{hX_i} \leq (1 - \mu) + \mu e^h$$

Wir erhalten so :

$$\begin{aligned} E\, e^{hX_i - h\mu} &= e^{-h\mu} \cdot E\, e^{hX_i} \\ &\leq e^{-h\mu} \cdot ((1 - \mu) + \mu e^h) \\ &= e^{-h\mu + \ln((1 - \mu) + \mu e^h)} \end{aligned}$$

Jetzt ist "nur noch" der rechte Term nach oben abzuschätzen. Die e-Funktion wächst streng monoton, wir schätzen daher statt der Funktion den Exponenten in Abhängigkeit von h

$$g(h) = -h\mu + \ln(1 - \mu + \mu e^h)$$

nach oben ab und benutzen dazu die "kurze" TAYLOR-Entwicklung mit Rest nach LAGRANGE (siehe 3.2) :

$$g(h) = g(0) + g'(0) \cdot h + \frac{1}{2} \cdot g''(\xi) \cdot h^2 \quad \text{mit einem } \xi \in]0,h[$$

Wir bilden die Ableitungen

$$g'(h) = -\mu + \frac{\mu}{(1 - \mu)e^{-h} + \mu} \quad ,$$

$$g''(h) = \frac{\mu(1 - \mu)e^{-h}}{((1 - \mu)e^{-h} + \mu)^2} \quad ,$$

finden zunächst $g(0) = 0$ und $g'(0) = 0$, und erhalten mit eben jenem schönen Trick, mit dem man die Varianz der Binomialverteilung nach oben abschätzt :

$$0 \le g''(\xi) = \frac{\mu}{(1-\mu)e^{-\xi}+\mu} \cdot \left(1 - \frac{\mu}{(1-\mu)e^{-\xi}+\mu}\right) \le \frac{1}{4}$$

Denn beide Faktoren sind aus [0,1] und haben die Summe 1. Diese **Abschätzung ist unabhängig von $\mu \in [0,1]$**, was sich auf alle folgenden Abschätzungen vererbt ! Der Exponent ist somit quadratisch abschätzbar :

$$g(h) = \frac{1}{2} g''(\xi) \cdot h^2 \le \frac{1}{8} h^2$$

Und wir finden für alle X_i :

$$E\, e^{hX_i - h\mu} \le e^{\frac{1}{8}h^2}$$

Damit erhalten wir aus Satz 5, der Abschätzung durch Produktzerlegung :

$$P(X - \mu \ge \varepsilon) \le e^{-hn\varepsilon} \cdot \left(e^{\frac{1}{8}h^2}\right)^n = e^{n\left(\frac{1}{8}h^2 - \varepsilon h\right)}$$

Mit elementaren Mitteln findet man schließlich, daß der Exponent im rechten Term bei $h = 4\varepsilon$ sein Minimum $-2\varepsilon^2$ annimmt, es folgt :

$$P(X - \mu \ge \varepsilon) = C(n,\varepsilon) \le e^{-2n\varepsilon^2}$$

Eine analoge Rechnung mit $-1 \le -X_i \le 0$ führt auf :

$$P(-X + \mu \ge \varepsilon) = C(n,\varepsilon) \le e^{-2n\varepsilon^2}$$

Damit ist Satz 4 vollständig bewiesen. Die benutzten Argumente beschränken sich auf Linearität und Monotonie des Erwartungswertes, Homogenität der Varianz und ihre Additivität bei unabhängigen Zufallsgrößen und die folgend kurz skizzierten Hilfsmittel aus der Analysis.

3 Hilfsmittel aus der Analysis

Damit sie bei Bedarf schnell zur Hand sind. sind die verwendeten Hilfsmittel aus der Analysis hier kurz genannt :

3.1 Konvexität der e-Funktion

Zum Beweis der JENSENschen Ungleichung für die e-Funktion analysieren wir

die Ordinatendifferenz von Funktionsgraph und Sekante als Funktion, eine ganz typische Etüde zur Analysis, die auch als Schülerübung vorstellbar ist. Für $f(x) = e^x$ starten wir mit $0 < f(x) = f'(x) = f''(x)$, fixieren $h > 0$ und betrachten für $x \in [0,1]$ die Differenz

$$d(x) = (1-x)e^0 + xe^h - e^{hx} = e^0 + x(e^h - e^0) - e^{hx} \quad \text{mit}$$

$$d'(x) = e^h - e^0 - he^{hx} \quad \text{und} \quad d''(x) = -h^2 e^{hx} \ .$$

Da wegen $f''(x) > 0$ global $d''(x) < 0$ gilt, fällt $d'(x)$ auf [0,1] streng monoton, hat also dort höchstens eine Nullstelle. Daher nimmt $d(x)$ in $]0,1[$ höchstens ein Extremum an, ein Maximum. Minimal wird $d(x)$ somit nur am Rand. Wegen $d(0) = d(1) = 0$ folgt dann $d(x) \ge 0$ für alle $x \in [0,1]$, die Behauptung. Die Argumentation gilt analog für jede Funktion mit global positiver zweiter Ableitung.

3.2 "kurze" TAYLOR-Entwicklung

Die zur Abschätzung des Exponenten g(x) erforderliche Beweisidee ist zugleich die allgemeine zur TAYLOR-Entwicklung mit Rest nach LAGRANGE, sie läßt sich auch - vielleicht aus dem hier gegebenen Anlaß - ad hoc mit folgendem Ansatz entwickeln : Analysiere die Abweichung von der Schmiegeparabel als Funktion des Entwicklungspunktes mit dem Mittelwertsatz. Wir fixieren $h > 0$ und betrachten für Entwicklungspunkte $x \in [0,h]$ die Differenz :

$$\Delta(x) = g(h) - g(x) - g'(x)(h-x) - C(h-x)^2$$

$$\Delta'(x) = -g''(x)(h-x) + 2C(h-x)$$

Wählt man $$C = \frac{g(h) - g(0) - g'(0)\cdot h}{h^2} \ ,$$

so ist $\Delta(0) = 0$ und $\Delta(h) = 0$, und der Mittelwertsatz liefert ein $\xi \in]0,h[$ mit :

$$0 = \Delta'(\xi) = (-g''(\xi) + 2C)(h - \xi)$$

Das führt auf $C = g''(\xi)/2$ und so zur Behauptung.

4 Literatur

[1] BARTH, F.; HALLER, R.: Stochastik Leistungskurs .- Ehrenwirth Verlag, München 1983

[2] HOEFFDING, W.: Probability inequalities for sums of bounded random variables .- Journal of the American Statistical Assossiation 58(1963), S. 13 - 30

[3] OKAMOTO, M.: Some inequalities relating to the partial sum of binomial probabilities .- Annals of the Institute of Statistical Mathematics 10(1958), S. 29 - 35

[4] WARMUTH, E.: Was ist die Wahrscheinlichkeit ? Gedanken zur Erklärung und Entwicklung des Begriffs in der Schule .- DdM 3, 1991, S. 165 - 170

[5] WOLLRING, B.: Ein Beispiel zur Konzeption von Simulationen bei der Einführung des Wahrscheinlichkeitsbegriffs, Aus der Vorbereitung einer Unterrichtsreihe für die Jahrgangsstufe 6 .- Stochastik in der Schule 12(1992), Heft 3, S. 2 - 25

Dr. Bernd Wollring ,
Institut für Didaktik der Mathematik,
Universität Münster

Einsteinstr.62
4400 Münster