Extremwertuntersuchungen bei der Binomialverteilung

von *Wolfgang Göbels*, Bergisch Gladbach

Zusammenfassung: Im folgenden werden Bedingungen für die Verwendung der Poissonverteilung als Näherung für die Binomialverteilung angegeben. Der Beitrag soll motivieren, mit Schülern aus Leistungskursen die Zusammenhänge zwischen diesen Verteilungen zu erörtern und auf die Bedingungen einzugehen, unter denen diese Näherung brauchbar ist.

Die Aufgabe

Die nachfolgende Aufgabe soll die Schüler in die Problematik einführen: In einer Urne, die weiße und schwarze Kugeln enthält, sei der Anteil der weißen Kugeln mit p bezeichnet. Ein Spieler zieht n Kugeln mit Zurücklegen. Laut Spielregel soll er einen Preis gewinnen, wenn sich unter seinen n gezogenen Kugeln genau k weiße befinden.

(a) Für welchen - von k und n abhängigen - Wert wird die Gewinnwahrscheinlichkeit des Spielers maximal und welchen Wert hat diese maximale Gewinnwahrscheinlichkeit?

(b) Gibt es einen von k abhängigen Grenzwert der Folge seiner maximalen Gewinnwahrscheinlichkeiten?

Diese Aufgabe kann mit Hilfe des nachfolgenden Satzes gelöst werden.

Satz

Nimmt p den Wert $\frac{k}{n}$ an, so ist die Wahrscheinlichkeit, in einer Bernoulli-Kette der Länge n mit der Trefferwahrscheinlichkeit p genau k Treffer zu erzielen, maximal. Diese maximalen Wahrscheinlichkeiten $\mathbf{B}_{n,k}$ hängen von n und k ab.

Es gilt:

$$\mathrm{w}(k) \;:=\; \lim_{n\to\infty} \mathbf{B}_{n,k}\left(\frac{k}{n}\right) \;=\; \frac{k^k}{k!}\cdot \mathrm{e}^{-k} \quad . \qquad (1)$$

Stochastik in der Schule **11** (1991), Nr. 3, 42-45.

Diese Konvergenz zeigt, daß man die Wahrscheinlichkeiten $\mathbf{B}_{n,k}(\frac{k}{n})$ mittels der Näherungsformel von Poisson approximieren kann. Der rechte Ausdruck in (1) ist die Poisson-Näherung mit dem Parameter $\mu = k$, dieser entspricht dem Mittelwert

$$\mu = n\cdot p = n\cdot\frac{k}{n} = k\,.$$

Beweis

Es beginnt mit der Untersuchung der Funktion B; ihre Gleichung lautet:

$$\mathbf{B}_{n,k}(p) = \binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}$$

Die erste Ableitung liefert $p = \frac{k}{n}$ als mögliche Extremstelle:

$$\mathbf{B}'_{n,k}(p) = \binom{n}{k}\cdot[\,k\cdot p^{k-1}\cdot(1-p)^{n-k} - p^k\cdot(n-k)\cdot(1-p)^{n-k-1}\,]$$

$$= \binom{n}{k}\cdot[\,p^{k-1}\cdot(1-p)^{n-k-1}\cdot(k(1-p) - (n-k)p)\,]$$

$$= \binom{n}{k}\cdot[\,p^{k-1}\cdot(1-p)^{n-k-1}\cdot(k-np)\,]$$

Sie ist nämlich genau dann Null, wenn $p = \frac{k}{n}$. Die zweite Ableitung bzw. eine Untersuchung auf Wechsel des Vorzeichens bestätigt, daß die Funktion $\mathbf{B}_{n,k}$ an der Stelle $\frac{k}{n}$ ein strenges relatives Maximum hat. Im vorliegenden Fall ist der Nachweis am schnellsten über das Kriterium mit dem Vorzeichenwechsel zu erhalten. Für hinreichend kleine $\varepsilon > 0$ verifiziert man:

$$\mathbf{B}'_{n,k}(\tfrac{k}{n} - \varepsilon) > 0 \quad\text{und}\quad \mathbf{B}'_{n,k}(\tfrac{k}{n} + \varepsilon) < 0\,.$$

Hierzu genügt die Untersuchung des Terms $k - np$ aus der ersten Ableitung an der Stelle $p = \frac{k}{n}$, da alle übrigen Terme stets positiv sind. Es gilt

$$k - n\cdot(\frac{k}{n} - \varepsilon) = n\cdot\varepsilon > 0\ , \quad k - n\cdot(\frac{k}{n} + \varepsilon) = -n\cdot\varepsilon < 0\ ,$$

womit die Existenz des strengen relativen Maximums gezeigt ist. Als geeignete Schülerübung empfiehlt sich auch der Existenznachweis mit Hilfe der zweiten Ableitung, wenn dieser auch aufwendiger ist, sei es auf direktem Wege oder über die Beziehung

$$B'_{n,k}(p) = \frac{n^2}{k}\cdot B_{n-1,k-1}(p) - \frac{n(n-1)}{k}\cdot B_{n-2,k-1}(p)\ .$$

Das Maximum wird nun berechnet.

$$\begin{aligned} B_{n,k}(\frac{k}{n}) &= \binom{n}{k}\cdot(\frac{k}{n})^k\cdot(1-\frac{k}{n})^{n-k} \\ &= \frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{1\cdot 2\cdot\ldots\cdot k}\cdot\frac{k^k}{n^k}\cdot(1-\frac{k}{n})^n\cdot(1-\frac{k}{n})^{-k} \\ &= (1-\frac{1}{n})\cdot(1-\frac{2}{n})\cdot\ldots\cdot(1-\frac{k-1}{n})\cdot\frac{k}{1}\cdot\frac{k}{2}\cdot\ldots\cdot\frac{k}{k}\cdot(1-\frac{k}{n})^n\cdot(1-\frac{k}{n})^{-k} \end{aligned}$$

Aufgrund der Grenzwertsätze für Zahlenfolgen und der Grenzwerteigenschaften der Eulerschen Zahl e folgt nun:

$$w(k) \ := \ \lim_{n\to\infty} B_{n,k}(\tfrac{k}{n}) = \frac{k^k}{k!}\cdot e^{-k} \quad \diamond$$

Weiterführung

Es bietet sich an, die Thematik eventuell weiterzuführen und mögliche Eigenschaften der Folge $w(k)$ hinsichtlich der Monotonie, Beschränktheit und Konvergenz zu diskutieren. Falls diese Folge nämlich beschränkt ist, so gibt ihr dann existierender Grenzwert eine von n und k unabhängige absolute Obergrenze für die Gewinnwahrscheinlichkeit des Spielers an. Da die Suche nach

einem solchen Grenzwert auch für einen Leistungskursschüler auf zu große Schwierigkeiten stoßen dürfte, kann ersatzweise auch eine experimentelle Untersuchung mit geeigneten Rechnern durchgeführt werden, welche bestätigen wird, daß der fragliche Grenzwert Null beträgt und somit eine positive Obergrenze für die Gewinnwahrscheinlichkeit des Spielers nicht existiert.

Zum Schluß sei noch darauf hingewiesen, daß ähnliche Überlegungen wie die in diesem Beitrag geschilderten beim 'Ziehen ohne Zurücklegen' im Zusammenhang zwischen der Hypergeometrischen Verteilung und der Binomialverteilung bei der Ermittlung des besten Schätzers gemäß der Maximum-Likelihood-Methode angestellt werden können.

Brosamen für den rauhen Alltag

Stimmt die Vermutung, daß die Aufgabe des Statistikers darin besteht, Signifikanzen zu machen?

* * *

Was "Stochastik" heißt, ist mir bekannt. Können Sie mir aber sagen, wie man dieses Wort in Silben trennt?

* * *

Die treue Begleiterin "Wahrscheinlichkeit" wird dich nie im Stich lassen. In weiser Voraussicht haben es die Begründer unserer Lehre so eingerichtet, daß die Muse der statistischen Kunst stets im Bereiche zwischen Null und Eins zu finden ist. Andrejs Dunkels hat in einem Cartoon dies so umschrieben: "Ach, ich mag die Wahrscheinlichkeitsmenschen so sehr" - "Wieso?" - "Sie sind nie negativ!"