

2/1 (2004), 133–150

tmcs@math.klte.hu  
<http://tmcs.math.klte.hu>

**Teaching**  
Mathematics and  
Computer Science

## Die Stichprobe als ein Beispiel dafür, wie im Unterricht die klassische und die bayesianische Auffassung gleichzeitig dargestellt werden kann

ÖDÖN VANCÓS

*Abstract.* Teaching statistics and probability in the school is a new challenge of the Hungarian didactics. It means new tasks also for the teacher- and in service-teacher training. This paper contains an example to show how can be introduced the basic notion of the inference statistics, the point- and interval-estimation by an elementary problem of the public pole. There are two concurrent theories of the inference statistics the so called classical and the Bayesian Statistics. I would like to argue the importance of the simultaneously introduction of both methods making a comparison of the methods. The mathematical tool of our elementary model is combinatorial we use some important equations to reach our goal. The most important equation is proved by two different methods in the appendix of this paper.

*Key words and phrases:* teaching statistics, inference statistics, point- and interval-estimation, confidence-interval, the Bayesian Highest Density Credible Region, binomial- and hyper geometrical distribution, Binomial coefficient, different notions of probability.

*ZDM Subject Classification:* K24, K55, K74, K79, M19.

### 1. Einführung

Es ist eine sehr aktuelle Frage, welcher Weg (klassisch oder/und bayesianisch) auf verschiedenen Stufen des Unterrichtes (Mittelschule oder Hochschulen und Universitäten) verfolgt werden kann. Dazu wäre es wichtig beide Theorien und ihre Anwendungsbereiche begrifflich klar zu kennen. Es gab eine interessante Diskussion über diese Fragen im Journal American Statistician von 1996. Auf

dem klassischen bzw. Bayesianischen Standpunkt standen u. a. D. Moore [1] bzw. D. Berry [2], J. Albert [3] und Iversen [4], beide Standpunkte wurden mit besonderer Rücksicht auf den Unterricht von D. K. Stangl [5] vertreten. Der Letztere leitete ein Seminar für Universitätsstudenten, dessen Inhalt und Struktur in [5] beschrieben worden ist. Ich finde solche Diskussionen interessant, denn sie helfen diese Begriffe und Ideen ausführlicher darzulegen, gegenüberzustellen, und beide Standpunkte besser zu erklären.

Einen möglichen Vergleich der beiden Auffassungen – nicht auf der Universitätsstufe (wie etwa in [5]), sondern früher – möchte ich in diesem Beitrag an einem schulgemäßen Problem vorstellen. Ähnliche Beispiele aus der Schulmathematik wie die Meinungsumfrage sind z. B. in [6] oder [7] zu finden; diese haben für den vorliegenden Artikel auch als eine Motivation gedient. Mein Beispiel kann aber die Rolle eines Ausgangspunktes für den ganzen Erkenntnisprozess der beurteilenden Statistik in der Schule spielen. Dabei kommt die moderne Computertechnologie zum Einsatz sowohl in der Veranschaulichung als auch im Rechnen. Bei der Umsetzung der Ergebnisse des Computereinsatzes hat Csaba Sárvári mitgearbeitet.

Es ist wichtig bei einer konkreten Situation ein solches Modell zu entwickeln, an dem beide Auffassungen gleichzeitig vorgestellt und interpretiert werden können. Die wohlbekannteste Meinungsumfrage bietet didaktisch ein ausgezeichnetes Beispiel dafür. Dadurch kann die Leitidee beider Theorien vorgestellt werden. Daneben ist die Meinungsumfrage heutzutage ein oft diskutiertes, alltägliches Gebiet. Vom Gesichtspunkt der Motivation her ist es also ein geeignetes Beispiel.

In diesem Beitrag werden wir das Problem der Meinungsumfrage analysieren. Es werden zwei verschiedene Schätzungen (die Punkt- und die Intervallschätzung) aus beiden Aspekten (klassisch und bayesianisch) vorgestellt. Das Novum dieses Beitrags ist, dass eine Hauptfrage der beurteilenden Statistik gleichzeitig aus den zwei verschiedenen Blickwinkeln analysiert, anschaulich verglichen und interpretiert wird.

Die von uns klassisch genannte Methode ist im Fall der Intervallschätzung (Konfidenz-Intervall) von Neyman genau deshalb eingeführt worden, damit die von ihm als „subjektiv“ betrachtete Apriori-Verteilung (in der Bayes-Statistik) vermieden werden kann. Er hat das nur so erreichen können, dass er nicht direkt die uns interessierende Frage beantwortet, sondern eine umgekehrte Frage. Wir möchten ein solches Intervall finden, worin der gesuchte Parameterwert mit einer vorgegebenen (großen) Wahrscheinlichkeit liegt; eine solche Wahrscheinlichkeit existiert aber klassisch gar nicht! (S. z. B. in [8].) Stattdessen wird ein Intervall

mit der folgenden Eigenschaft bestimmt: Es existiert zu jedem Parameterwert aus diesem Intervall ein  $\gamma$ -Schätzbereich, so dass das Ergebnis des Experimentes in diesem  $\gamma$ -Schätzbereich liegt. Das ist aber eine andere Auffassung als die sehr oft vorkommende falsche Interpretation, dass nämlich der gesuchte Parameterwert mit der Wahrscheinlichkeit  $1 - \gamma$  in dem Konfidenzintervall liegt (s. wieder [8]).

In dem bayesianischen Ansatz gibt der  $\gamma$ -BHD (siehe z. B. [9]) eine erwünschte Antwort auf unsere Frage, aber der „Preis“ dieser Antwort ist, dass wir unsere subjektive Einschätzung (vor dem Experiment) einbeziehen müssen. Es ist noch wichtig zu bemerken, dass im Hintergrund des bayesianischen Ansatzes eine *andere Wahrscheinlichkeitsauffassung* steht, die meistens gar nicht als klassische – auf Häufigkeiten basierte – Wahrscheinlichkeit interpretiert werden kann.

Über die einschlägigen, tiefen, begrifflichen und philosophischen Fragen hat D. Wickmann einen lesenswerten und sehr gedankenreichen Artikel [10] geschrieben. Ich möchte diesmal darauf nicht tiefer eingehen, da ich beabsichtige dies später in einem anderen Beitrag durchzuführen.

Nach meiner Meinung wäre es sehr wichtig, dass die Schüler so früh wie möglich beide Auffassungen kennenlernen, unter anderem deshalb, damit sie verstehen: Es gibt kein absolutes Modell; jedes Modell hat Vor- und Nachteile, und von der jeweiligen Situation abhängig versuchen wir für unseren Zweck eines von den möglichen zu wählen. Das hilft dem Schüler den Dogmatismus zu vermeiden. Die Mathematik ist also keine absolute Wahrheit, wie sich viele unter dem Einfluss der Schulmathematik falsch vorstellen.

Endlich möchte ich mich für die vielen fruchtbaren Gespräche, die sowie inhaltlichen als auch sprachlichen Korrekturen und die guten Ratschläge von Ervin Deák recht herzlich bedanken. In seinen mathematikdidaktischen Arbeiten pflegt er solchen begrifflichen und erkenntnistheoretischen strukturellen Problemen auf den Grund zu gehen, allerdings nicht in der Wahrscheinlichkeitsrechnung und Statistik, sondern in vielen anderen Gebieten der Schul- und Hochschulmathematik; diese Probleme liegen aber eben so tief, dass sie nicht an ein konkretes Wissensgebiet gebunden sind. (Eine diesbezügliche, umfassende Studie von E. Deák ist in Arbeit.)

## 2. Das Beispiel

Betrachten wir die wohlbekannte Situation der Meinungsumfrage. Es gibt eine Population mit  $N$  Elementen, unter denen  $K$  Gefragte auf eine bestimmte Frage mit „ja“ antworten. Wir kennen aber diese Zahl  $K$  nicht. Um diese Zahl zu

schätzen nehmen wir eine Stichprobe mit  $n$  Elementen. Wir wollen den Wert von  $K$  schätzen, unter der Annahme, dass genau  $k$  von den  $n$  Gefragten eine positive Antwort geben.

Da wir ein Maß der Zuverlässigkeit unserer Schätzung bestimmen möchten, muss die Auswahl bei der Stichprobe wichtige Bedingungen erfüllen. Jedes Element der Grundgesamtheit (der Population) soll gleichwahrscheinlich durch einen zufälligen Prozess ausgewählt werden.

Es gibt zwei Typen der Schätzung, die sogenannte Punktschätzung und die Intervallschätzung. Beide haben eine klassische und eine bayesianische Version. Im Folgenden werden diese untersucht, und im Fall der Intervallschätzung stellen wir eine neue Veranschaulichung vor. Diese Struktur folgt der Aufbau des Kapitels 13 der Dissertation von S. Götz [11].

Das Modell dieser Situation ist kombinatorisch; es wird eine hypergeometrische Verteilung für alle möglichen  $K$  angegeben, als das Ergebnis einer Ziehungsreihe ohne Zurücklegen mit  $n$  Ziehungen. Wir können die Anzahl der Möglichkeiten in einer Matrix der Form  $(N + 1) \times (n + 1)$  einordnen.

$$\begin{pmatrix} \binom{0}{0} \binom{N}{n} & \binom{1}{0} \binom{N-1}{n} & \cdots & \binom{K}{0} \binom{N-K}{n} & \cdots & \binom{N}{0} \binom{0}{n} \\ \binom{0}{1} \binom{N}{n-1} & \binom{1}{1} \binom{N-1}{n-1} & \cdots & \binom{K}{1} \binom{N-K}{n-1} & \cdots & \binom{N}{1} \binom{0}{n-1} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \binom{0}{k-1} \binom{N}{n-k+1} & \binom{1}{k-1} \binom{N-1}{n-k+1} & \cdots & \binom{K}{k-1} \binom{N-K}{n-k+1} & \cdots & \binom{N}{k-1} \binom{0}{n-k+1} \\ \binom{0}{k} \binom{N}{n-k} & \binom{1}{k} \binom{N-1}{n-k} & \cdots & \binom{K}{k} \binom{N-K}{n-k} & \cdots & \binom{N}{k} \binom{0}{n-k} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \binom{0}{n} \binom{N}{0} & \binom{1}{n} \binom{N-1}{0} & \cdots & \binom{K}{n} \binom{N-K}{0} & \cdots & \binom{N}{n} \binom{0}{0} \end{pmatrix}$$

Diese Matrix werden wir für die verschiedenen Schätzungen benutzen.

### 3. Die Schätzungen

#### 3.1. Punktschätzung

##### 3.1.1. Klassische Methode

Die Idee dieser Methode wurde von R. A. Fisher eingeführt und wird die Maximum Likelihood Methode (MLM) genannt. Das Wesen dieses Verfahrens ist einen solchen Schätzwert für  $K$  zu geben, für den die Wahrscheinlichkeit, die

tatsächlich vorliegende Stichprobe zu ziehen, maximal ist. Genauer formuliert, wird ein solcher Wert von  $K$  als Schätzwert angenommen, bei dem es am wahrscheinlichsten ist, dass unter  $n$  Gefragten  $k$  positiv antworten. Angenommen, dass  $K$  unter den  $N$  Personen mit „ja“ antwortet, wird die Wahrscheinlichkeit für das gegebene Stichprobenergebnis  $k$  unter  $n$ :

$$\frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}.$$

Dieses Ergebnis kann mit der Division der Elemente der oben gegebenen Matrix durch „ $N$  über  $n$ “ gewonnen werden. In diesem Fall gibt jede Spalte eine (i. Allg.) unvollständige hypergeometrische Verteilung an.

Die Frage ist, für welchen Wert  $K$  wird diese Zahl maximal, wenn  $k$ ,  $n$ ,  $N$  gegeben sind. Von  $K$  hängt nur der Zähler ab, also müssen wir den folgenden Ausdruck maximieren:

$$K(K-1)(K-2)\dots(K-k+1)(N-K)(N-K-1)\dots(N-K-n+k+1).$$

Mit den Mitteln der Differentialrechnung ist es nicht einfach den Maximalwert zu bestimmen. Es gibt aber eine bessere Methode für solche diskrete Fälle, wo die Verteilung eine monoton zunehmende und abnehmende (Teil-)Folge ist. Das Maximum liegt bei dem „Wechselpunkt“. Dazu müssen wir die folgende Ungleichung untersuchen:

$$\begin{aligned} & K(K-1)\dots(K-k+1)(N-K)(N-K-1)\dots(N-K-n+k+1) < \\ & < (K+1)K\dots(K-k+2)(N-K-1)(N-K-2)\dots(N-K-n+k). \end{aligned}$$

Gekürzt ergibt sich die Ungleichung:

$$(K-k+1)(N-K) < (K+1)(N-K-n+k),$$

umgeordnet

$$K < \frac{k}{n}(N+1) - 1.$$

Wenn  $K = \lceil \frac{k}{n}(N+1) - 1 \rceil$  (wobei  $[x]$  der ganze Anteil der Zahl  $x > 0$  bedeutet) so gilt noch die Ungleichung, bei dem nächsten Wert nicht mehr, also ist der Maximalwert die nächste Zahl:  $K = \lceil \frac{k}{n}(N+1) \rceil$ . Wenn  $\frac{k}{n}(N+1) - 1$  ganz ist, dann ist das der Maximalwert.

Wir können diesen Wert der hypergeometrischen Verteilung ziemlich genau auch durch Approximation mit der Binomialverteilung erhalten. Diese Approximation kann genutzt werden wenn  $N$  wenigstens um zwei Größenordnungen größer ist als  $n$ , und  $K$  sich ähnlich zu  $k$  verhält. Wir untersuchen also die Binomialverteilung:  $\binom{n}{k} \left(\frac{K}{N}\right)^k \left(\frac{N-K}{N}\right)^{n-k}$ . Der Maximalwert dieser Wahrscheinlichkeit liegt dort (in Abhängigkeit von  $K$ ), wo der Wert der Funktion  $K^k(N-K)^{n-k}$  maximal ist; diese Stelle zu bestimmen differenzieren wir diese Funktion nach  $K$ :

$$\begin{aligned} kK^{k-1}(N-K)^{n-k} + K^k(n-k)(N-K)^{n-k-1}(-1) &= \\ &= K^{k-1}(N-K)^{n-k-1}[k(N-K) - (n-k)K] \end{aligned}$$

Die Nullstelle dieses Ausdrucks ist, angenommen  $K \neq 0$  und  $K \neq N$ : die Lösung der Gleichung  $k(N-K) - (n-k)K = 0$ , also  $K = \frac{k}{n}N$ .

Dies Ergebnis ist nicht wirklich überraschend, das heißt, dass der relative Anteil in der Grundgesamtheit (in der Population) mit dem relativen Anteil in der Stichprobe zusammenfällt. In dem Fall, dass  $\frac{k}{n}N$  keine ganze Zahl ist, ist das Maximum entweder der ganze Anteil oder um 1 größer. Dieser Schätzwert ist entweder gleich dem Ergebnis im hypergeometrischen Fall, oder der Unterschied zwischen den zwei Werten ist gleich 1. Dies zeigt, dass die Approximation bei dieser Schätzung nur einen sehr kleinen Fehler verursachen kann.

### 3.1.2. Bayesianische Methode

Das Wesen dieser Methode ist Folgendes. *Vor* dem Experiment (Ziehung  $n$ -mal ohne Zurücklegen aus einer Urne mit  $N$  Elementen) nehmen wir eine Verteilung für die möglichen Werte von  $K$  an, es soll also für alle  $K$  eine Wahrscheinlichkeitswert angegeben werden. Diese Verteilung drückt *unser Vorwissen* aus. Das hängt also von der Person und ihrem Wissensgrad ab und ist daher eine *subjektive* Wahrscheinlichkeitsschätzung. Es ist vielleicht günstiger die eine Zustandswahrscheinlichkeitsverteilung zu nennen. Genau diesen Schritt wollten Neyman und Pearson (und allgemeinen die heute „klassisch“ genannten Statistiker) vermeiden. Zu der bayesianischen Theorie soll ein anderer, etwas breiterer Wahrscheinlichkeitsbegriff eingeführt werden, als zu der klassischen, wobei der Anwendungsbereich nicht auf die zufälligen Prozesse eingeschränkt wird. Dies ist tatsächlich möglich, siehe z. B. [10], [12] oder [13]. Durch den Bayes'schen Satz wird eine Aposteriori Verteilung ausgerechnet. Das drückt unser Wissen *nach* der Erfahrung des Experimentes aus.

In diesem Fall wird auch die MLM gebraucht, aber zuerst soll eine Aposteriori Verteilung berechnet werden. Im Folgenden wird diese im Fall der Apriori

Gleichverteilung berechnet. In dem einfachsten Fall (keine Vorinformation) wird angenommen, dass  $\pi(K) = \frac{1}{N+1}$ , wobei  $K \in \{0, 1, \dots, N\}$  ist, und die Aposteriori Verteilung wird durch den Satz von Bayes ausgerechnet:

$$\begin{aligned} \pi(K | X = k) &= \frac{P(X = k | K) P(K)}{\sum_{K=k}^{N-n+k} P(X = k | K) P(K)} = \\ &= \frac{\binom{K}{k} \binom{N-K}{n-k} \cdot \frac{1}{N+1}}{\sum_{K=k}^{N-n+k} \binom{K}{k} \binom{N-K}{n-k} \cdot \frac{1}{N+1}} = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\sum_{K=k}^{N-n+k} \binom{K}{k} \binom{N-K}{n-k}}. \end{aligned}$$

Dabei zählt die Zufallsvariable  $X$  die Ja-Stimmen in der Stichprobe.

Der Nenner ist konstant (unabhängig von  $K$ ), suchen wir also das Maximum des Zählers. Dieses ist aber gleich mit jenem im klassischen Fall, siehe oben 3.1.1.

### 3.1.3. Fazit

In dem Fall der Apriori Gleichverteilung ist *der numerische Wert* beider Schätzungen gleich. Das ist eine bekannte Behauptung. Es ist wichtig zu bemerken, dass die Interpretation der beiden Ergebnisse völlig anders ist, siehe z. B. in [10]. Die Situation kann aber ganz anders sein, wenn die Apriori Verteilung keine Gleichverteilung ist. Dafür ist *ein konkretes Beispiel* das Folgende:

Sei eine Binomialverteilung a priori mit den Parametern  $N$  und  $p$  für die Verteilung von  $K$  gegeben. In diesem Fall schätzen wir vor dem Experiment so, dass der wahrscheinlichste Wert für  $K$   $Np$  ist, und die Chance der anderen Werte durch eine sehr schmale „glockenförmige“ Verteilung angegeben wird. In diesem Fall wird die Aposteriori Verteilung:

$$\begin{aligned} \pi(K | X = k) &= \frac{P(X = k | K) P(K)}{\sum_{K=k}^{N-n+k} P(X = k | K) P(K)} \\ &= \frac{\binom{K}{k} \binom{N-K}{n-k} \cdot \binom{N}{K} p^k (1-p)^{N-K}}{\sum_{K=k}^{N-n+k} \binom{K}{k} \binom{N-K}{n-k} \cdot \binom{N}{K} p^K (1-p)^{N-K}}, \end{aligned}$$

wobei die möglichen Werte von  $K$   $k, k+1, \dots, N-n+k$  sind.

In diesem Fall ist es nur komplizierter, aber im Wesentlichen nicht schwieriger

den Maximalwert zu bestimmen. Wenn wir der oben gesehenen Methode folgen, untersuchen wir die Ungleichung:

$$\binom{K}{k} \binom{N-K}{n-k} \binom{N}{K} p^k (1-p)^{N-K} < \binom{K+1}{k} \binom{N-K-1}{n-k} \binom{N}{K+1} p^{K+1} (1-p)^{N-K-1}.$$

Nach dem Ausrechnen der Binomialkoeffizienten und nach Kürzung erhalten wir:

$$(K-k+1)(N-K)(K+1)(1-p) < (K+1)(N-K-1-n+k+1)(N-K)p.$$

Angenommen, dass  $K \geq 0$  und  $K < N$ , kann die Ungleichung weiter gekürzt werden:

$$(K-k+1)(1-p) < (N-K-1-n+k+1)p$$

woraus

$$K < (N-n)p + k - 1 + p$$

folgt, also ist der Maximalwert  $K = [(N-n)p + k - 1 + p] + 1$  oder  $(N-n)p + k - 1 + p$ , wenn dieser Wert ganz ist, wobei  $[x]$  den ganzen Anteil der reellen Zahl  $x > 0$  bezeichnet. Dieses Ergebnis unterscheidet sich wesentlich von dem Fall der Apriori-Gleichverteilung. In dem Fall dass  $N$  viel größer als  $n$  (und deshalb noch größer als  $k$ ) ist wird sich der wahrscheinlichste Wert von  $K/N$  nur wenig von dem a priori  $p$ -Wert unterscheiden. Unser Vorurteil wird also stark unsere Schätzung beeinflussen. Ein Beispiel:

Seien  $N = 1\,000\,000$ ,  $n = 800$ ,  $k = 160$  und  $p = 0,5$ .

In diesem Fall beträgt der oben bestimmten Wert von  $K$  499 760. Der a posteriori relative Anteil ist also: 0,49976. Das klassische (und das a priori gleichverteilte) Ergebnis ist aber ungefähr  $160/800 = 0,2$ !

Diese Apriori Binomialverteilung (angenommen, dass wir uns geirrt haben) kann man nur mit einer viel größeren Stichprobe beeinflussen. Z. B. ist bei  $N = 1\,000\,000$ ,  $n = 10\,000$  und  $k = 2000$  immer noch  $K/N = 0,497$ ; wenn  $n = 100\,000$  und  $k = 20\,000$  ist immer noch  $K/N = 0,47$  anstatt 0,2.

Der Grund ist, dass diese a priori Verteilung eine sehr starke Überzeugung ausdrückt. Mit der 99%-Sicherheit schätzen wir vor dem Experiment den  $K$ -Wert zwischen 498 500 und 501 500, also liegt der relative Anteil zwischen 0,4985 und 0,5015. Eine kleine Stichprobe kann den relativen Anteil ( $p$ ) nicht bedeutsam verändern.

### 3.2. Intervallschätzung

#### 3.2.1. Klassisch

Das Konfidenz-Intervall kann an einem zweidimensionalen Bild visualisiert werden; auf der  $x$ -Achse werden die verschiedenen relativen Anteile  $(0, 1/N, 2/N, \dots, K/N, \dots, 1)$ , auf der  $y$ -Achse die Werte der relativen Häufigkeit der Stichprobe  $(0, 1/n, 2/n, \dots, k/n, \dots, 1)$  dargestellt. Zu jedem Wert  $K$  gehört ein Intervall  $[k_1/n; k_2/n]$  als ungefährer 0,95-Schätzbereich – also ist 0,95 die Wahrscheinlichkeit, dass die relative Häufigkeit in dieses Intervall fällt –, angenommen, dass der relative Anteil in der Grundgesamtheit  $K/N$  ist. Es muss hier bemerkt werden, dass dieses Intervall nicht eindeutig ist, dazu muss noch eine Bedingung erfüllt werden. Im Allgemeinen ist diese die Folgende: das Intervall soll das kürzeste sein. Eine ähnliche Bedingung soll bei auch der bayesianischen Methode angenommen werden. Wir erhalten ein „wurstförmiges“ Bild (aus diskreten Punkten, nicht stetig, aber wenn  $N$  und  $n$  genug groß sind, scheint das Bild auf dem Bildschirm dennoch stetig), die Konfidenzellipse (die Punkte der oberen bzw. unteren Kurve entstehen aus den Werten  $k_2/n$  bzw.  $k_1/n$ , angenommen bei verschiedenen  $K/N$  Werten). Das Konfidenz-Intervall kann als ein Schnittintervall der Inversfunktion gesehen werden, wobei also eine zur  $y$ -Achse senkrechte Gerade die Ellipse schneidet. Siehe Abbildung 1.

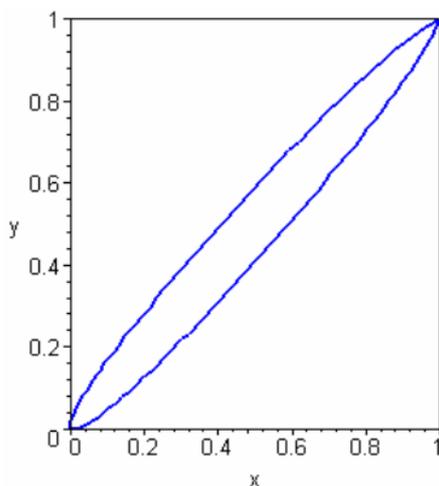


Abbildung 1. Die 0,95-Konfidenz-Ellipse

Zur Bestimmung eines Konfidenz-Intervalls muss der  $\gamma$ -Schätzbereich zuerst berechnet werden. Im Fall  $K/N = p$  ist dieser

$$\left[ p - \Phi^{-1}\left(\frac{1-\gamma}{2}\right) \cdot \sqrt{\frac{p(1-p)}{n}}; p + \Phi^{-1}\left(\frac{1-\gamma}{2}\right) \cdot \sqrt{\frac{p(1-p)}{n}} \right],$$

unter der Voraussetzung, dass wir die Approximation der hypergeometrischen Verteilung durch die Binomialverteilung und diese durch die Normalverteilung anwenden dürfen.

Auf einem anderen Weg kann das Konfidenz-Intervall direkt mit einem Computer berechnet werden. In 3.2.2 wird das Ergebnis gegeben.

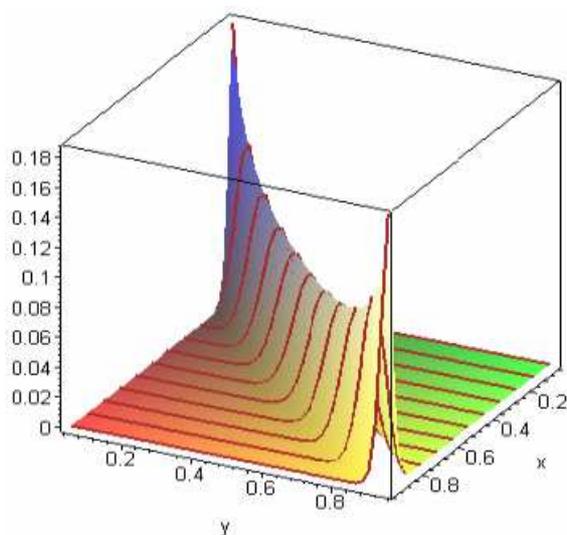


Abbildung 2. Das dreidimensionale Bild

Auch ein dreidimensionales Bild kann in dieser Situation gezeichnet werden, wobei über das Einheitsquadrat eine Fläche gezeichnet wird, so dass jedem Punkt des Quadrats  $[K/N; k/n]$  der Wert  $\frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$  zugeordnet wird, siehe Abbildung 2. In diesem Fall wird durch jeden zur  $x$ -Achse senkrechten Schnitt (eine Ebene) eine hypergeometrische Verteilung angegeben. Auf der  $[x; y]$  Ebene kann der oben erwähnte Schätzbereich und das darüber hinaus bekommene Konfidenz-Intervall gezeichnet werden. Das zeigt die klassische Lösung. Wir werden noch auf demselben Bild auch die bayesianische Lösung (siehe nächster Abschnitt) wenigstens in dem Fall der Apriori Gleichverteilung zeigen. Siehe Abbildung 3.

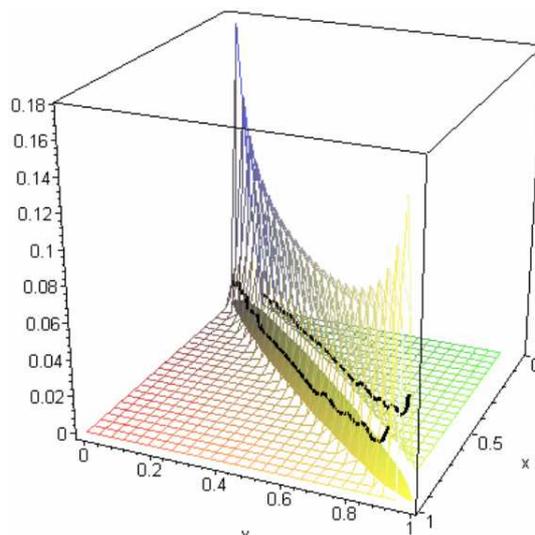


Abbildung 3. Gleichzeitiges Bild des Konfidenz-Intervalls (gelb) und des BHD (schwarz)

### 3.2.2. Bayesianisch

In diesem Fall muss ein 0,95-BHD gerechnet werden, angenommen zuerst die Apriori-Gleichverteilung, die schon unter 3.1.2 bestimmt worden ist. Die Aposteriori Verteilung ist

$$\pi \left( K \mid X = \frac{k}{n} \right) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\sum_{K=k}^{N-n+k} \binom{K}{k} \binom{N-K}{n-k}},$$

mit  $K = k, k+1, \dots, N-n+k$ . Diese Verteilung kann als zur  $y$ -Achse senkrechter Schnitt dargestellt werden. Dieser Schnitt kann durch ein zwei-dimensionales Bild repräsentiert werden, zur Normierung dient der Faktor  $\sum_{K=k}^{N-n+k} \binom{K}{k} \binom{N-K}{n-k}$ . Es ist eine schöne kombinatorische Aufgabe zu beweisen, dass dieser Faktor für alle  $k$  gleich  $\binom{N+1}{n+1}$ . Zwei Beweise dieser Tatsache sind in Anhang dieses Artikels zu lesen.

Mit dem Computer kann einen 0,95-BHD gerechnet werden.

Das Ergebnis ist in Abbildung 3. zu sehen. Wir müssen in diesem Fall die zur  $y$ -Achse senkrechten Schnitte betrachten, und es soll ein solches Intervall als 0,95-BHD bestimmt werden, wo die Summe der Aposteriori Wahrscheinlichkeiten größer gleich 0,95 und das Intervall des kürzeste ist. Siehe [9], S. 78–79.

Das Rechnen des  $\gamma$ -BHD's kann direkt auf einem zweidimensionalen Bild animiert werden. Daran wird noch mit Csaba Sárvári gearbeitet.

Wir geben eine Tabelle um zu zeigen, wie dieser 0,95-BHD konkret aussieht, wenn z. B.  $N = 1\,000\,000$ ,  $n = 800$  und  $k = 237, 238, \dots, 245$ . Anstatt eine zu lange Computerzeit in Anspruch zu nehmen um den 0,95-BHD genau zu berechnen, wurden zwei Approximationen (die in diesem Fall erlaubt sind) benutzt:

Zuerst die Annäherung des Ziehens ohne Zurücklegung durch die Ziehung mit Zurücklegung (also die Annäherung der Hypergeometrischen Verteilung durch die Binomialverteilung) und danach die Annäherung der  $\beta$ -Verteilung (welche dann als Aposteriori Verteilung auftritt) durch die Normalverteilung (s. [9] S. 119–120).

So erhalten wir die folgenden BHDs:

$k = 237$	[265 161; 328 356]
$k = 238$	[266 365; 329 636]
$k = 239$	[267 578; 330 808]
$k = 240$	[268 787; 332 211]
$k = 241$	[269 996; 333 494]
$k = 242$	[210 207; 334 778]
$k = 243$	[272 417; 336 062]
$k = 244$	[273 627; 337 345]
$k = 245$	[274 837; 338 629]

Zum Vergleich sei das Konfidenz-Intervall in diesen Fällen (durch einen graphischen Taschenrechner berechnet) gegeben:

$k = 237$	[264 610; 327 890]
$k = 238$	[265 821; 329 180]
$k = 239$	[267 032; 330 468]
$k = 240$	[268 244; 331 756]
$k = 241$	[269 457; 333 043]

$k = 242$	[270 670; 334 331]
$k = 243$	[271 882; 335 618]
$k = 244$	[273 095; 336 904]
$k = 245$	[274 309; 338 191]

### 3.2.3. Fazit

Bei der Intervallschätzung gibt es numerische Differenzen zwischen den beiden Methoden auch noch im Fall der Apriori-Gleichverteilung. Diese Differenzen wären auch als Folge der Approximation vorstellbar, sie scheinen aber tendenziell: ein Konfidenz-Intervall etwa kann durch eine kleine Verschiebung eines dazu gehörenden BHDs erhalten werden. Diese Bemerkung ist einfach zu beweisen. (S. [9] S. 119–120.) Die numerischen Differenzen werden aber viel größer, wenn wir verschiedene Apriori-Verteilungen gebrauchen.

Es soll aber nochmal betont werden, dass sowohl das Berechnen als auch die Interpretation der Ergebnisse beider Methoden unterschiedlich sind, unabhängig von der Möglichkeit, dass die Ergebnisse numerisch zusammenfallen.

## Anhang

Der Satz lautet:

$$\sum_{K=k}^{N-n+k} \binom{K}{k} \binom{N-K}{n-k} = \binom{N+1}{n+1}.$$

Wir geben zwei verschiedene Beweise dieses Satzes.

### I.

Im ersten Beweis wird eine Methode benutzt, die eigentlich auf Fermat’sche „Methode des Herabsteigens“ (oder des „Zurückführens“) zurückgeht.

Die Methode besteht darin zu zeigen, dass zwei benachbarte Zeilen der Matrix in 2. jeweils dieselbe Summe haben. (Da demzufolge sämtliche Zeilensummen gleich sind und die Summe aller Elemente der Matrix leicht berechnet werden kann, ist somit die Aufgabe gelöst, jede einzelne Zeilensumme ermittelt.)

Da die totale Summe der Matrix  $(N+1)\binom{N}{n}$  (jede Spaltensumme beträgt  $\binom{N}{n}$ ) wegen  $\sum_{i=0}^n \binom{K}{i} \binom{N-K}{n-i} = \binom{N}{n}$  ist, wird die Summe einer Zeile

$$\frac{N+1}{n+1} \binom{N}{n} = \binom{N+1}{n+1} \quad (\text{I.1})$$

betragen, wenn alle Zeilen die gleiche Summe haben.

Dazu soll die folgende Gleichung für alle Werte von  $k$  ( $0 \leq k \leq n$ ) bewiesen werden:

$$(\text{I.2}). \quad \sum_{K=k}^{N-n+k} \binom{K}{k} \binom{N-K}{n-k} = \sum_{K=k+1}^{N-n+k+1} \binom{K}{k+1} \binom{N-K}{n-k-1}$$

Das stimmt im Fall  $k=0$ , da

$$\begin{aligned} & \binom{0}{0} \binom{N}{n} + \binom{1}{0} \binom{N-1}{n} + \dots + \binom{N-n}{0} \binom{n}{n} \\ &= \binom{1}{1} \binom{N-1}{n-1} + \binom{2}{1} \binom{N-2}{n-1} + \dots + \binom{N-n+1}{1} \binom{n-1}{n-1}. \end{aligned} \quad (\text{I.3})$$

Auf der linken Seite kann der zweite Faktor zerlegt werden:

$$\binom{N-K}{n} = \binom{N-K-1}{n} + \binom{N-K-2}{n-1}.$$

Auf der rechten Seite kann der erste Faktor werden:

$$\binom{K}{1} = \binom{K-1}{1} + \binom{K-1}{0}.$$

So erhalten wir:

$$\begin{aligned} & \binom{0}{0} \left[ \binom{N-1}{n} + \binom{N-1}{n-1} \right] + \binom{1}{0} \left[ \binom{N-2}{n} + \binom{N-2}{n-1} \right] + \dots \\ & \quad + \binom{N-n}{0} \left[ \binom{n-1}{n-1} + \binom{n-1}{n} \right] \end{aligned} \quad (\text{I.4})$$

und

$$\begin{aligned} & \left[ \binom{0}{1} + \binom{0}{0} \right] \binom{N-1}{n-1} + \left[ \binom{1}{1} + \binom{1}{0} \right] \binom{N-2}{n-1} + \dots \\ & \quad + \left[ \binom{N-n}{1} + \binom{N-n}{0} \right] \binom{n-1}{n-1}. \end{aligned} \quad (\text{I.5})$$

Nach dem Auflösen der eckigen Klammern kann jedes zweite Glied der Summen beidseitig abgezogen werden, da sie gleich sind, und es bleibt nur die Gleichung:

$$\begin{aligned} \binom{0}{0} \binom{N-1}{n} + \binom{1}{0} \binom{N-2}{n} + \dots + \binom{N-n-1}{0} \binom{n}{n} &= \\ &= \binom{1}{1} \binom{N-2}{n-1} + \binom{2}{1} \binom{N-3}{n-1} + \dots + \binom{N-n}{1} \binom{n-1}{n-1}. \end{aligned} \quad (\text{I.3a})$$

Offensichtlich ist dies die Gleichung (I.3), bloß mit  $N-1$  an Stelle von  $N$ . Durch Wiederholung dieser Umformung können wir die Gleichung (I.3a) bis  $n+1$  zurückführen. In diesem Fall bleibt die Gleichung  $\binom{0}{0} \binom{n+1-1}{n} = \binom{1}{1} \binom{n+1-2}{n-1}$  die richtig ist, da beide Seiten gleich 1 sind.

Ähnlich geht dieses Verfahren im Allgemeinen, wenn  $k$  nicht gleich 0 ist:

$$\sum_{K=k}^{N-n+k} \binom{K}{k} \binom{N-K}{n-k} = \sum_{K=k+1}^{N-n+k+1} \binom{K}{k+1} \binom{N-K}{n-k-1}, \quad (\text{I.2})$$

dass heißt:

$$\begin{aligned} \binom{k}{k} \binom{N-k}{n-k} + \binom{k+1}{k} \binom{N-k-1}{n-k} + \binom{k+2}{k} \binom{N-k-2}{n-k} + \dots \\ + \binom{N-n+k}{k} \binom{n-k}{n-k} = \binom{k+1}{k+1} \binom{N-k-1}{n-k-1} + \binom{k+2}{k} \binom{N-k-2}{n-k-1} + \dots \\ + \binom{N-n+k+1}{k+1} \binom{n-k-1}{n-k-1}. \end{aligned}$$

Wieder werden zwei Umformungen gebraucht: auf der linken Seite für alle zweiten Faktoren, auf der rechten Seite für alle ersten Faktoren in den Produkten:

$$\binom{N-k-i}{n-k} = \binom{N-k-i-1}{n-k} + \binom{N-k-i-1}{n-k-1} \quad (i = 0, 1, \dots, N-n)$$

beziehungsweise

$$\binom{k+j+1}{k+1} = \binom{k+j}{k+1} + \binom{k+j}{k} \quad (j = 0, 1, \dots, N-n).$$

Durch diese Zerlegungen wird die Anzahl der Glieder auf beiden Seiten verdoppelt, aber jedes zweite gekürzt und wir erhalten die folgende Gleichung, die um ein

Glied kürzer ist (weil ein Glied auf beiden Seiten null ist):

$$\begin{aligned}
 & \binom{k}{k} \left\{ \binom{N-k-1}{n-k-1} + \binom{N-k-1}{n-k} \right\} + \binom{k+1}{k} \left\{ \binom{N-k-2}{n-k} + \binom{N-k-2}{n-k-1} \right\} \\
 & + \binom{k+2}{k} \left\{ \binom{N-k-3}{n-k} + \binom{N-k-3}{n-k-1} \right\} + \dots \\
 & + \binom{N-n+k}{k} \left\{ \binom{n-k-1}{n-k} + \binom{n-k-1}{n-k-1} \right\} = \\
 & = \left\{ \binom{k}{k+1} + \binom{k}{k} \right\} \binom{N-k-1}{n-k-1} + \left\{ \binom{k+1}{k} + \binom{k+1}{k+1} \right\} \binom{N-k-2}{n-k-2} \\
 & + \left\{ \binom{k+2}{k} + \binom{k+2}{k+1} \right\} \binom{N-k-3}{n-k-1} + \dots \\
 & + \left\{ \binom{N-n+k}{k} + \binom{N-n+k}{k+1} \right\} \binom{n-k-1}{n-k-1},
 \end{aligned}$$

nach dem Kürzen bekommen wir:

$$\begin{aligned}
 & \binom{k}{k} \binom{N-k-1}{n-k} + \binom{k+1}{k} \binom{N-k-2}{n-k} + \binom{k+2}{k} \binom{N-k-3}{n-k} + \dots \\
 & + \binom{N-n+k}{k} \binom{n-k-1}{n-k} = \\
 & = \binom{k+1}{k+1} \binom{N-k-2}{n-k-1} + \binom{k+2}{k+1} \binom{N-k-3}{n-k-1} + \dots + \binom{N-n+k}{k+1} \binom{n-k-1}{n-k-1}.
 \end{aligned}$$

Das ist die ursprüngliche Gleichung (I.2) nur um ein Glied kürzer, da statt  $N-k$   $N-k-1$  steht. Setzen wir dieses Verfahren fort, so erhalten wir am Ende auf beiden Seiten 0, also ist auch die ursprüngliche Gleichung gültig, womit der Beweis vollständig ist.

## II.

Der andere Beweis ist – im Gegensatz zum ersten, wo nur der Formelapparat der Binomialkoeffizienten ausgenutzt worden ist – auf die kombinatorische Bedeutung der Binomialkoeffizienten gegründet. Die kombinatorische Grundidee des Beweises verdanke ich der Mitteilung von A. Bereczky und R. Freud; sie kann auch als ein Beispiel für den Satz von der totalen Wahrscheinlichkeit aufgefasst werden.

Wir versuchen zuerst ein totales Ereignissystem konstruieren. Nehmen wir eine Urne mit den ersten  $N + 1$  natürlichen Zahlen. Davon werden ohne Zurücklegen  $n + 1$  Zahlen nacheinander herausgezogen. Seien die herausgezogenen Zahlen wie beim Lottospiel in eine aufsteigende Reihenfolge gestellt. Die Frage ist, welche Wahrscheinlichkeit dem Ereignis „die  $(k + 1)$ -ste Zahl bei einer solchen Reihenfolge ist genau  $K + 1$ “ zukommt. Es ist klar, dass diese Ereignisse von  $K = k$  bis  $K = N - n + k$  ein totales Ereignissystem bilden.

Die Anzahl der Elementarereignisse ist in diesem Fall natürlich  $\binom{N+1}{n+1}$ , das ist nämlich die Anzahl aller möglichen Ziehungen. Die Anzahl der günstigen Elementarereignisse ist  $\binom{K}{k} \binom{N-K}{n-k}$ .

Die Wahrscheinlichkeit, dass die  $(k + 1)$ -ste Zahl genau  $K + 1$  ist, beträgt also

$$\frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N+1}{n+1}}.$$

Wenn wir diese Wahrscheinlichkeiten addieren, muss Eins herauskommen, also haben wir tatsächlich

$$\sum_{K=k}^{N-n+k} \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N+1}{n+1}} = 1, \quad \sum_{K=k}^{N-n+k} \binom{K}{k} \binom{N-K}{n-k} = \binom{N+1}{n+1}.$$

### Literatur

- [1] David S. Moore, Bayes for Beginners? Some reasons to hesitate, *The American Statistician* **51**, no. 3 (1997), 254–261.
- [2] D. A. Berry, Teaching elementary Bayesian statistics with real applications in science, *The Am. Stat.* **51**, no. 3, 241–246.
- [3] J. Albert, Teaching Bayes’rule: a data-oriented approach, *The Am. Stat.* **51**, 247–253.
- [4] G. Iversen, Student perceptions of Bayesian Statistics, *The Am. Stat.* **51**, 231–237.
- [5] D. K. Stangl, Classical and Bayesian paradigms: Can we teach both?, *The Am. Stat.* **51**, 253–258.
- [6] S. Götz, *Klassische und bayesianische Behandlung von Stochastik-Aufgaben in österreichischen Schulbüchern*, in: Borovcnik, J. Engel, D. Wickmann (Hrsg): Anregungen zum Stochastikunterricht, Hildesheim: Franzbecker, 2001, S. 147–162.
- [7] D. Wickmann, *Die Übungstunde*, in: Borovcnik, J. Engel, D. Wickmann (Hrsg): Anregungen zum Stochastikunterricht, Hildesheim: Franzbecker, 2001, S. 117–122.

- [8] S. Krauss und G. Giegerenzer, *Statistischen Denken oder statistische Rituale: Was sollte man unterrichten?*, in: Borovcnik, J. Engel, D. Wickmann (Hrsg): Anregungen zum Stochastikunterricht, Hildesheim: Franzbecker, 2001, S. 53–62.
- [9] D. Wickmann, *Bayes-Statistik – Einsicht gewinnen Entscheidung treffen*, BI Verlag, 1991.
- [10] D. Wickmann, Zur Begriffsbildung im Stochastikunterricht, *JMD* **19**, Heft 1 (1998), 46–80.
- [11] S. Götz, *Bayes-Statistik – ein alternativer Zugang zur beurteilenden Statistik in der siebenten und achten Klasse AHS*, Wien im, Dezember 1997.
- [12] B. de Finetti, *Wahrscheinlichkeitsrechnung*, Birkhauser Verlag, 1953.
- [13] D. W. Lindley, *Introduction to Probability and Statistics from a Bayesian viewpoint*, Cambridge University Press, 1965.

ÖDÖN VANCÓSÓ  
EÖTVÖS LORÁND UNIVERSITÄT  
NATURWISSENSCHAFTLICHE FAKULÄT  
MATHEMATISCH-FACHDAKTISCHE GRUPPE  
H-1117 BUDAPEST, PÁZMÁNY PÉTER SÉTÁNY 1/C HUNGARY  
*E-mail:* vancso@ludens.elte.hu

(Received February, 2004)