Klassische und bayesianische Schätzung der Wahrscheinlichkeit dessen, dass beim Würfelwerfen die Sechs herauskommt

Ödön Vancsó

Abstract: *This paper shows how the main ideas of estimation can be introduced and visually represented, simultaneously in classical and Bayes-statistics. We will see the difference of these theories and the question is embedded in the theorie of functions with two variables. The didactical relevance of this paper is clear to day. There are three different views about the teaching of estimation: i) only (first) classical; ii) only Bayes-statistics; iii) both of them simultaneously, see[1]. I stand with this paper by iii) showed for teachers a possible way to introduce the notion of statistical estimations in one concrete simple situation. The example of this paper can be used in teacher or in-service teacher training.*
Key words: teaching statistics, point- and interval-estimation, confidence-interval, Bayes-statistics, the Bayesian Highest Density Credible Region, β-distribution.

I. Einführung

In diesem Beitrag möchte ich durch ein Beispiel die Unterschiede zwischen dem klassischen und dem bayesschen Verfahren klarmachen. Die Ergebnisse haben didaktische Bedeutung, vor allem die Veranschaulichung beider Verfahren – im Fall der Apriori-Gleichverteilung – in einer gemeinsamen dreidimensionalen Abbildung. Anhand dieser Abbildung werden zwei grundsätzliche Schätzungen beschrieben und numerisch analysiert: die Punkt- und die Intervallschätzung. Die Hauptideen der im Folgenden *klassisch* genannten Schätzungen stammen aus den Arbeiten von R. A. Fisher, K. Pearson und J. Neyman (s. [2] pp. 91-123; [3]). Der Begriff des Konfidenzintervalls wurde erst durch Neyman in [4] eingeführt. Das Pendant in der Bayes-Statistik ist der Gamma-Bereich höchster Dichte (BHD), siehe z. B. [5] pp. 78-80. Nicht besonders kompliziert ist der BHD durch einen Computer im Fall einer anderen A-priori Verteilung veranschaulicht zu werden. Dieses Beispiel zeigt darüber hinaus die Wichtigkeit der neuen Technologien im Unterricht sowie bei der Veranschaulichung als auch bei dem Rechnen.

II. Das Beispiel

Betrachten wir die folgende Situation. Jemand wirft einen Würfel bis zur ersten Sechs. Viele Kinderspiele beginnen damit. Wir wissen aber nicht, was für einen Würfel gebraucht wird. Solche „falsche“ Würfel sind in den Geschäften zu kaufen; sie dienen verschiedenen Zwecken. Die Frage ist, wie die Wahrscheinlichkeit der Sechs geschätzt werden kann, wenn man die Anzahl der Würfe bis zur ersten Sechs kennt. Es ist wohlbekannt, dass diese Situation durch eine geometrische Verteilung modelliert werden kann. Mit der Bezeichnung p für die Wahrscheinlichkeit der Sechs ist die Chance dafür, dass die erste Sechs mit genau k

Würfen erreicht wird, ist: $P(X = k) = p \cdot (1 - p)^{k-1}$, wobei X die Anzahl der Würfe bedeutet. In einem dreidimensionalen Bild kann diese Verteilung mit einem Parameter (*p*) dargestellt werden. In diesem Artikel werden nur einige zweidimensionale Schnitte dargestellt. Sei der dargestellte Bereich anstatt [0, 1] x **R**, [0, 1] x {**N**∩[0, 15]}. Die zweite Variable ist diskret, das heiβt, dass der Bereich des Problems eigentlich [0,1] x **N** ist. Die Abbildungen 1.a, b und c zeigen das drei dimensionale Bild 1.a und die Schnitte k = 3 (diese spielt im Beispiel in II.2.b eine zentrale Rolle) bzw. k = 1, 2, 3, 4.

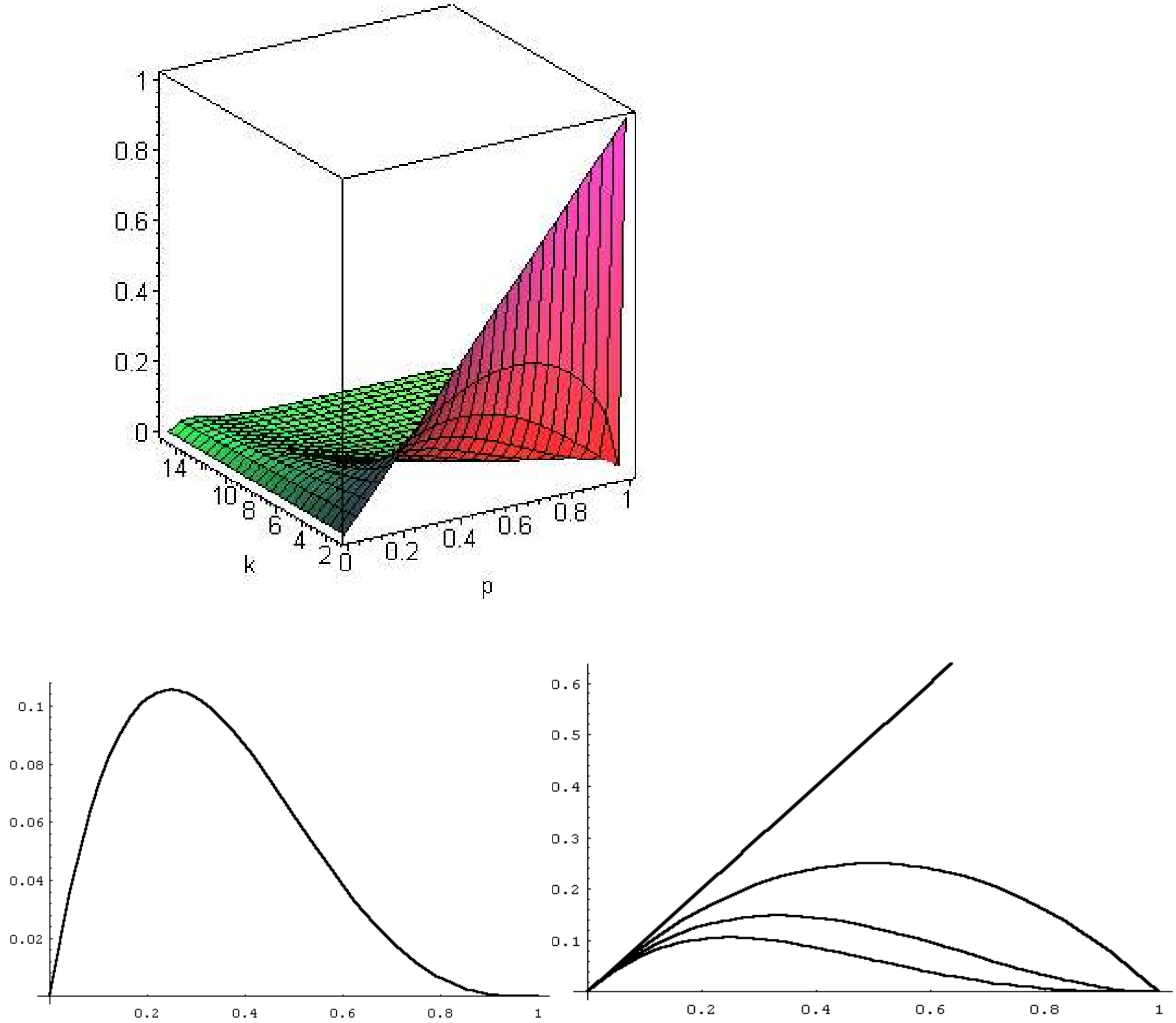

Abb. 1. a, b, c

Sei das Ergebnis bekannt: X = k. Welche Schätzung kann für *p* gegeben werden? Zuerst nehmen wir die Punktschätzung, danach die Intervallschätzung durch.

II. 1. Punktschätzung

a) klassisch

Mit der wohlbekannten Maximum Likelihood Method (MLM) [3] wählen wir ein solches *p* aus, wofür dieses Ergebnis das wahrscheinlichste ist. Die Frage lautet also folgenderweise: Mit welchem Parameterwert *p* ist P(X = k) am gröβten, das heiβt, wo liegt der Maximumwert bei einem horizontalen Schnitt (X=k)?

Die Extremwerte einer Polinomfunktion können wir bestimmen.

$f'(p) = (1-p)^{k-1} + p(k-1)(1-p)^{k-2}\cdot(-1) = (1-p)^{k-2}([1-p]-p[k-1]) = (1-p)^{k-2}(1-pk)$.

Da p = 1 unmöglich ist, wenn k > 1, haben wir f'(p) = 0 nur dann, wenn 1-pk = 0, also $p = \frac{1}{k}$.

Es ist nicht kompliziert um zu zeigen, dass in diesem Fall ein richtiges Maximum gefunden worden ist.

Die klassische Punktschätzung ist also $p = \frac{1}{k}$. Diese Schätzung ist erwartungswerttreu.

b) bayesianisch

Wie sieht die Antwort in der Bayes-Statistik aus? Nehmen wir an, dass die Apriori-Verteilung eine Gleichverteilung ist: $\pi(p) = 1$ für alle $p \in [0;1]$.

Die Aposteriori-Verteilung, bei X = k ist:

$$\pi(p|X{=}k) = \frac{p(1-p)^{k-1}\cdot\pi(p)}{\int_0^1 p(1-p)^{k-1}\cdot\pi(p)dp} = \frac{p(1-p)^{k-1}}{\int_0^1 p(1-p)^{k-1}dp} = \frac{p(1-p)^{k-1}}{\frac{1}{k(k+1)}} = k(k+1)p(1-p)^{k-1}.$$

Zum Ausrechnen des Nenners brauchten wir den Satz der partiellen Integration.

Die Frage ist: mit welchem Parameterwert p ist diese Aposteriori-Verteilung maximal? Diese Verteilung hat sich aber nur in einer Konstante von $p(1-p)^{k-1}$ unterschieden, also ergibt sich das Maximum wie beim klassischen Fall bei dem Wert $p = \frac{1}{k}$. Das zeigt, dass in diesem Fall das Ergebnis nicht davon abhängt, welche Theorie benutzt wird. Es gibt hier also keinen numerischen Unterschied zwischen den Theorien im Fall der Apriori-Gleichverteilung. Bei der Interwallschätzung ist es aber anders.

Noch eine Bemerkung ohne Beweis (der allerdings nicht zu kompliziert ist): Im Bayesschen Fall ist die Schätzung nur asymptotisch erwartungswerttreu.

Wir können aber sogar bei der Punktschätzung ein ganz anderes Ergebnis erhalten, wenn uns am Anfang irgendeine Information über den Würfel zur Verfügung steht. Wir wissen zum Beispiel, dass der Würfel gefälscht (gezinkt) ist, dass daher viel häufiger die Sechs, als die Eins herauskommt. Unser Vorwissen soll in Form einer Verteilung ausgedrückt werden, z.B. in einer β-Verteilung (damit können wir einfacher rechnen). – Es ist bemerkenswert, dass von D. Wickmann ein Programm für graphische Taschenrechner entwickelt wurde, dies es ermöglicht mit der Maus eine Verteilungsskizze zu zeichnen und darüber hinaus eine Apos-teriori-Verteilung zu berechnen. - Sei eine β-Verteilung gegeben, deren zwei Parameter *a* und *b* nach Belieben gewählt werden können:

$$f(p) = \frac{(a+b+1)!}{a!\cdot b!}p^a(1-p)^b.$$

Was kommt in diesem Fall als Punktschätzung heraus?

$$g(p) = \frac{p(1-p)^{k-1}p^a(1-p)^b}{\int_0^1 p(1-p)^{k-1}p^a(1-p)^b\,dp} = \frac{p^{a+1}(1-p)^{k+b-1}}{\int_0^1 p^{a+1}(1-p)^{k+b-1}dp} = \frac{p^{a+1}(1-p)^{k+b-1}}{\frac{(a+1)!\cdot(k+b-1)!}{(k+a+b+1)!}} =$$

$$= \frac{(k+a+b+1)!}{(a+1)!\cdot(k+b-1)!}p^{a+1}(1-p)^{k+b-1}$$

Wir suchen das Maximum dieser Verteilung. Bis auf eine Konstante haben wir:

$$g'(p) = (a+1)p^a(1-p)^{k+b-1} + p^{a+1}(k+b-1)(1-p)^{k+b-2}(-1) = 0$$

$$p^a(1-p)^{k+b-2}\{(a+1)(1-p) - p(k+b-1)\} = 0$$

Angenommen $0 < p < 1$, bekommen wir aus der zweiten Gleichung $p = \frac{a+1}{k+a+b}$. Der Wert dieses Ausdrucks ist nur in sehr speziellen Fällen gleich dem im klassischen Fall erhaltenen $p = \frac{1}{k}$. Wenn z. B. $b = (k\text{-}1)a$, so gibt es keinen Unterschied zwischen den verschiedenen ML-Schätzungen. Seien $a = 3$, $b = 9$ und $k = 4$; dann kommt durch die ML-Schätzung anstatt $p = 1/4$ der Wert $p = 1/2$ heraus. Natürlich hängt der Wert der Parameter a und b von unseren Vorkenntnissen ab.

II. 2. Intervallschätzung

a) klassisch

Es soll in unserem Fall das Konfidenzintervall im klassischen Sinn bestimmt werden. Dazu soll zuerst der Schätzbereich gegeben werden. Da die geometrische Verteilung abnehmend ist, ist der Schätzbereich immer vom Typ [1, K], wobei K von p abhängt. Diese Abhängigkeit kann durch die Formel

$$\sum_{k=1}^{K} p(1-p)^{k-1} \geq 1-\alpha$$

ausgedrückt werden, und K ist der kleinste unter den Werten K, die diese Ungleichung erfüllt. Mit der Summenformel der endlichen geometrischen Reihen bekommen wir:

$$p\frac{1-(1-p)^K}{1-(1-p)} = 1-(1-p)^K,$$

also gilt für K:

(*) $\qquad (1-p)^K \leq \alpha$

Daraus folgt $K = \left\lceil \frac{\log\alpha}{\log(1-p)} \right\rceil$, wobei wir mit dieser Klammer den oberen Ganzteil bezeichnen. Hier ist α gegeben, K hängt von p ab. K(p) ist eine monoton abnehmende Funktion, wie leicht zu sehen ist. Die folgende Abbildung 2 zeigt den Schätzbereich in der Funktion von p. Zu einem gegebenen K gehört ein Konfidenzintervall [0; p(K)]. Wir möchten vom bekannten Wert von K auf jenen von p schließen. Es soll also p aus der Ungleichung (*) $(1\text{-}p)^K < \alpha$ berechnet werden. So ergibt sich die Ungleichung

$p < 1-\sqrt[K]{\alpha}$, das heißt $p \in [0; 1-\sqrt[K]{\alpha}]$. Das ist eine Intervallschätzung für p, wenn α und K gegeben sind.

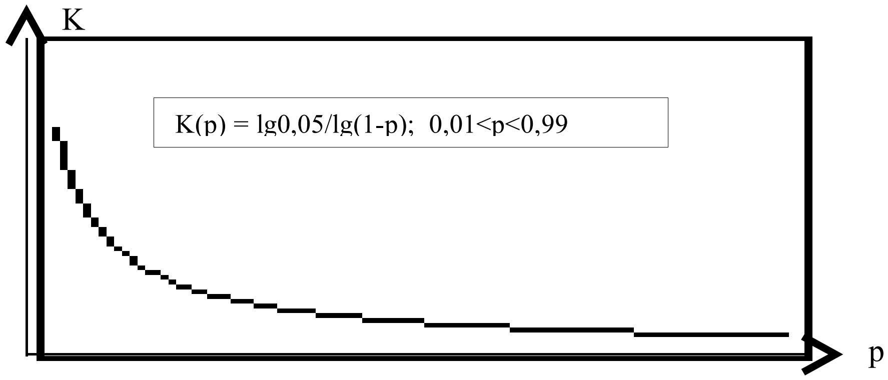

Abb. 2.

b) bayesianisch

Wie sieht eine Intervallschätzung für p bayesianisch, im Fall der Apriori-Verteilung aus? Die Aposteriori-Verteilung ist: $k(k+1)p(1\text{-}p)^{k\text{-}1}$, wobei k die Anzahl der Würfe bis zur ersten Sechs bezeichnet. Wir sollen ein solches $[p_1; p_2]$ Intervall suchen, welches die Bedingungen

$$\int_{p_1}^{p_2} k(k+1)p(1-p)^{k-1}\,dp \ge 1-\alpha \quad \text{und } p_2\text{-}p_1 \text{ ist minimal}$$

erfüllt. Das bedeutet z.B. bei $\alpha = 0{,}05$, und $k = 3$, dass

$$12\int_{p_1}^{p_2} p(1-p)^2\,dp \ge 0{,}95\,.$$

In diesem Fall haben wir

$$12\left[\frac{p^4}{4}-\frac{2p^3}{3}+\frac{p^2}{2}\right]_{p_1}^{p_2} = 3p_2^4-8p_2^3+6p_2^2-(3p_1^4-8p_1^3+6p_1^2) \ge 0{,}95\,.$$

Die Analyse der Verteilung – angenommen, die Differenz p_2-p_1 ist die kleinste – bekommen wir entweder mit dem Bedingten-Extremalwert-Satz (Lagrange Multiplikator-Methode) oder einfach nach der Anschauung, dass der Funktionswert der Aposteriori-Verteilung in p_1 gleich dem Wert in p_2 sein soll. Mit einem Computer (siehe z. B. BINOBETA in [5]) kann ein solches Paar (p_1, p_2) ausgerechnet werden, wofür das Integral gleich 0,95 ist.

In diesem konkreten Fall kann die Frage sogar mit einem Taschenrechner gelöst werden. Darüber hinaus ergibt sich – wegen $p_1(1-p_1)^2 = p_2(1-p_2)^2$ – ein Zusammenhang zwischen p_1 und p_2:

$$0 = p_2^3 - p_1^3 - 2(p_2^2 - p_1^2) + (p_2 - p_1) = (p_2 - p_1)(p_2^2 + p_1p_2 + p_1^2 - 2p_1 - 2p_2 + 1)\,.$$

Wir nehmen $p_1 \neq p_2$ an, also $0 = p_2^2 + p_1p_2 + p_1^2 - 2p_1 - 2p_2 + 1$. Ausgedrückt z. B. den Wert p_2 bekommen wir $p_2 = 1 - \frac{p_1}{2} - \sqrt{p_1(1-0{,}75p_1)}$, die andere Wurzel ist größer als 1, kommt also nicht in Frage. Wir können mit dem Taschenrechner für verschiedene gegebene Werte p_1 die entsprechenden Werte p_2 ausrechnen und kontrollieren, ob dass Integral den Wert 0,95 erreicht oder nicht. Nach einigem Probieren gelingt es derart den Wert $p_1 = 0{,}043$ und den entsprechenden Wert $p_2 = 0{,}7745$ zu erhalten (mit der Genauigkeit $5\cdot 10^{-4}$). Also ist [0,043; 0,7745] ist der gesuchte Bereich, welcher schon bemerkbar vom klassischen Ergebnis [0; 0,63] abweicht. Im diesen Fall fallen die Ergebnisse also schon auch numerisch nicht zusammen.

Diese Methode kann mit anderem Wert k durchgeführt werden, es ist bemerkbar aber: je gröβer k gewählt ist desto komplizierter wird das Rechnen. Im allgemeinen ist es nicht einfach zu beweisen, dass es immer ein solches Paar (p_1, p_2) in [0, 1] gibt, das die kürzeste Intervall angibt, also passt zur Idee der BHD. Bei der Rechnung hilft der Computer.

Es ist möglich mit verschiedenen anderen A-priori Verteilungen zu untersuchen und numerisch bewerten. Jetzt hätte ich nur die Hauptidee dieser Visualisation und das Verfahren des numerischen Rechnens durch Computer vorstellen mögen.

III. Interpretation und Vergleich

Die Logik des klassischen Verfahrens ist folgende: solche p Werte, dabei das Ergebnis k nicht in einem α-Schätzbereich fallen, werden nicht akzeptiert. Das Konfidenzintervall bedeutet al-so nicht, dass der richtige Wert p mit der Wahrscheinlichkeit (1-α) im Konfidenzintervall liegt, sondern dass bei der Wahl eines Wertes von p das statistische Ergebnis in den dazu ge-hörenden Schätzbereich fällt. Durch Formeln ausgedrückt: anstatt $P(p \in [p_1, p_2] | X = k) \ge 1-\alpha$ wissen wir nur, dass für alle $p \in [p_1, p_2]$, $P(X = k | p) \ge 1-\alpha$ gilt.

Das bayessche Verfahren antwortet auf die Frage wie wahrscheinlich es ist, dass p im BHD liegt. Dazu soll aber eine (subjektive) Apriori-Verteilung eingeführt werden. Das ist aber ge-

nau der Punkt, der von klassischen Statistikern angegriffen wird. J. Neyman hat den Begriff des Konfidenzintervalls genau deshalb eingeführt, weil er die Apriori-Verteilung vermeidet wollte (siehe in [4] die Seiten 194-229).
Ich möchte hier nicht auf die philosophische Frage eingehen, was überhaupt im unseren Fall die Wahrscheinlichkeit bedeutet; ich wollte nur die Verfahren vorstellen.
Wir haben die zwei Verfahren veranschaulicht und die numerischen Ergebnisse verglichen.

IV. Meine Experimente

Seit mehreren Jahren schon biete ich für Lehramtstudenten ein Bayes-Statistik-Seminar an der Eötvös Loránd Universität in Budapest an. Ich lege jedes Mal ein großes Gewicht auf den Vergleich des klassichen und bayesschen Verfahrens und auf ihre Interpretationen. Die gleichzeitige Veranschaulichung des Konfidenzintervalls und des BHD's im Fall der Gleichverteilung scheint mir ein sehr hilfreiches Mittel, wodurch der Unterschied klar ablesbar ist, und wodurch wir verstehen, warum es solche Fälle gibt, in denen die numerischen Ergebnisse die gleichen sind (im Fall der Normalverteilung).

Ein anderes Beispiel, wo die Kombinatorik ebenfalls eine wichtige Rolle spielt, ist die Analyse einer Meinungsumfrage. In diesem Beitrag habe ich keine Möglichkeit, auch dieses Beispiel in obigem Stil vorzustellen. Zum Schluss möchte ich noch ein interessantes Buch von Mignon und Gamerman [6] erwähnen, welches ein Versuch ist, beide Theorien parallel vorzustellen.

Literatur:

[1] D. K. Stangl: Classical and Bayesian Paradigms: Can we teach both?
The American Statistician 1998, 52(3), 253-258.

[2] G. Gigerenzer e.a.: Das Reich des Zufalls. Wissen zwischen Wahrscheinlichkeiten, Häufigkeiten und Unschärfen. Spektrum Akademischer Verlag 1999.

[3] R.A. Fisher: On the mathematical foundations of theoretical statistics
in: Phil. Trans. Royal Soc., London, Ser. A, Vol. 222 (1922), pp. 309-368

[4] J. Neyman: Lectures and Conferences on Mathematical Statistics and Probability
Second Edition, Graduate School U.S. Department of Agriculture
Washington 1952

[5] D. Wickmann: Bayes-Statistik Einsicht gewinnen und Entscheiden bei Unsicherheit
BI Wissenschaftsverlag Mannhein/Wien/Zürich 1990

[6] H.S. Mignon-D. Gamerman: Statistical Inference - An integrated Approach
Oxford University Press, New York 1999

Adresse des Autors: dr. Ödön Vancsó, Eötvös Lóránd Universität Budapest
H-1117 Budapest, Pázmány Péter sétány 1/C.
Ungarn
e-mail: vancso@ludens.elte.hu