

Das Achensee-Paradoxon

Joachim Engel, Ludwigsburg

Zusammenfassung:

Kann eine zufällige, dritte Zahl eine Schiedsrichterfunktion einnehmen, wenn zu entscheiden ist, ob eine Zahl A größer ist als eine zweite, unbekannte Zahl B ? Mit Hilfe eines elementaren Wahrscheinlichkeitsmodells und der Formel von der totalen Wahrscheinlichkeit wird eine zunächst überraschende Aussage bewiesen. Anhand von Spezialfällen wird die Aussage näher erläutert und plausibel gemacht.

Problem:

Sie greifen in Ihre Hosentasche und finden einen Zettel mit einer Zahl A darauf. Sie wissen, in der anderen Hosentasche befindet sich ein anderer Zettel mit einer zweiten Zahl B über deren Größe nichts bekannt ist. Ist A größer als B ? Wenn der Griff in die Hosentasche rein zufällig war (ob rechts oder links), dann ist (unter Ausschluß der Gleichheit) die Wahrscheinlichkeit für $P(A > B)$ offensichtlich $1/2$. Gibt es eine Möglichkeit, eine Entscheidung zwischen $A > B$ und $B > A$ zu treffen, die mit einer Wahrscheinlichkeit $p > 0,5$ die Richtige ist?

Die Geschichte der Wahrscheinlichkeitstheorie ist voll von Paradoxa - oft sind sie knapp und verständlich in ihrer Formulierung. Nicht selten haben sie zu lebhafter Diskussion und einer Verfeinerung der mathematischen Konzepte angeregt, siehe z.B. Székely (1990) für zahlreiche Beispiele. Paradoxa sind somit nicht nur ein Beitrag zur Geschichte der Mathematik, sondern auch eine heilsame Provokation für unser Denken. Darin liegt gerade auch ihr didaktischer Wert für einen Unterricht, der Schüler und Studenten zu eigenständigem Denken veranlassen will.

Die menschliche Intuition läßt sich durch den Begriff der bedingten Wahrscheinlichkeit leicht verwirren - eine Provokation, die didaktisch nutzbar zu machen ist. So führt Simpson's Paradox zu der Schlußfolgerung, daß für die Gesamtheit schlecht sein kann, was für jede einzelne Teilgruppe gut ist (Blyth 1972). Das Drei-Türen Paradox (Von Randow 1992) hat schon gestandene Mathematiker ins Schleudern gebracht. Über Erfahrungen, wie diese beiden Paradoxa dem schulischen Mathematikunterricht lebhafte Impulse versetzten, berichten Künzel (1991) und Klemisch (1993). Der renommierte Zauberkünstler und Harvard Statistik-Professor Persi Diaconis erläutert dazu in der Wochenzeitschrift *Der Spiegel* (1991), daß das menschliche Gehirn für ein Verständnis einfacher Grundbegriffe der elementaren Wahrscheinlichkeitsrechnung wohl schlecht verdrahtet sei.

Das obige Probleme über die Entscheidung zwischen zwei Zahlen sorgte am Rande des *10th International Workshop on Statistical Modelling* im Sommer 1995 am Achensee für heftige Diskussion.

Ähnlich wie bei den anderen erwähnten Paradoxa gründet seine Lösung auf dem bedingten Wahrscheinlichkeitsbegriff. Ein erster Schritt zur mathematischen Modellbildung charakterisiert die Zahlen A und B als identisch verteilte, unabhängige Zufallsvariablen mit der Dichte f und der Verteilungsfunktion F . Auch wenn vor der zu treffenden Entscheidung die Zahl A bekannt und B noch unbekannt ist, sind beide als vom Zufall abhängige Größen aufzufassen. Beide Zahlen haben irgendeinen zufälligen Wert, und dieser Wert ist unabhängig davon, daß ich nun eine Entscheidung treffen muß und dabei den Wert von Zahl A schon zur Kenntnis genommen habe. Von der Wahrscheinlichkeitsdichte f wird nichts als bekannt vorausgesetzt. Insbesondere wissen wir nicht, ob $f(x) > 0$ für alle $x \in \mathbb{R}$ oder ob f z.B. außerhalb eines Intervalles $[a, b]$ verschwindet. Es bezeichne \mathcal{A} das Ereignis "ich entscheide mich für $A > B$ " und \mathcal{B} entsprechend "ich entscheide mich für $B > A$ ". Gesucht ist eine Entscheidungsregel, d.h. eine Funktion d mit Wertebereich $\{\mathcal{A}, \mathcal{B}\}$, die von der Beobachtung A abhängt. Es sei $[A > B]$ das Ereignis, daß Zahl A größer ist als Zahl B , entsprechend $[B > A]$ " B ist größer als A ". Dann treffen wir eine korrekte Entscheidung genau dann, wenn das folgende Ereignis eingetroffen ist:

$$(\mathcal{A} \cap [A > B]) \cup (\mathcal{B} \cap [B > A]).$$

Da (unter Ausschaltung der Gleichheit) die Ereignisse $[A < B]$ und $[A > B]$ komplementär sind, können wir die Formel von der totalen Wahrscheinlichkeit anwenden. Damit ergibt sich für die Wahrscheinlichkeit einer korrekten Entscheidung:

$$\begin{aligned} P(\text{korrekte Entscheidung}) &= P((\mathcal{A} \cap [A > B]) \cup (\mathcal{B} \cap [B > A])) \\ &= P(\mathcal{A} | [A > B]) \cdot P([A > B]) + P(\mathcal{B} | [B > A]) \cdot P([B > A]). \end{aligned} \quad (1)$$

Abbildung 1 veranschaulicht Formel (1): eine getroffene Entscheidung ist genau dann richtig, wenn folgendes gilt: falls $[A > B]$ ist die Entscheidung \mathcal{A} getroffen, falls $[A < B]$ wird \mathcal{B} entschieden.

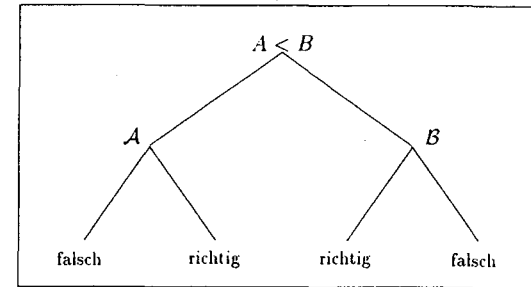


Abbildung 1: Baumdiagramm der Entscheidungsregel

Bisher ist noch nichts darüber gesagt, wie eine gute Entscheidung aussehen soll. \mathcal{A} und \mathcal{B} sind bisher lediglich abstrakte Bezeichnungen. Wenn der Griff in die erste Hosentasche rein zufällig ist (ob rechts oder links), so ergibt sich aus Symmetriegründen (und wird weiter unten auch formell bewiesen), daß $P(A > B) = 1/2$.

Entscheidungsregel 1:

Ziehe eine weitere, von A und B stochastisch unabhängige Zufallszahl Z von einer beliebigen Verteilung mit Dichte g . Nur eine einzige Einschränkung soll gelten: der Träger von Z ist die gesamte reelle Linie \mathbb{R} , d.h. $g(x) > 0$ für alle $x \in \mathbb{R}$. Falls $Z > A$ entscheide \mathcal{B} , falls $Z < A$ ist die Entscheidung \mathcal{A} .

Anstelle von "gesamte reelle Linie als Träger von Z " hat sich für Schüler auch die laxere Bezeichnung "die Verteilung ist für alle reellen Zahlen definiert" eingebürgert. Weiter unten werden wir uns auf die exakte Definition des Begriffs "Träger einer Verteilung" beziehen als die Menge der Punkte x , für die gilt: $F(x + \epsilon) - F(x - \epsilon) > 0$ für alle $\epsilon > 0$. Formell läßt sich die Entscheidungsregel als Abbildung darstellen.

$$\begin{aligned} d: \{(a, z) | a \neq z\} &\rightarrow \{\mathcal{A}, \mathcal{B}\} \\ d(a, z) &= \begin{cases} \mathcal{B} & \text{falls } a < z \\ \mathcal{A} & \text{falls } a > z. \end{cases} \end{aligned}$$

Bezeichnet man mit $[Z < A]$ das Ereignis, daß Zufallsvariable Z kleiner als A ist und entsprechend mit $[Z > B]$ das Ereignis " Z größer B ", so können wir Entscheidungsregel 1 auch in der Form darstellen

$$\mathcal{A} = [Z < A] \text{ und } \mathcal{B} = [Z > B].$$

Das erstaunliche an dem Ergebnis ist, daß die Lösung von einer unabhängigen und offensichtlich *beziehungslosen* dritten Zufallsvariable Z kommt. Information von Z gewonnen, hilft mit, die Entscheidung zwischen A und B zu verbessern.

Wir beweisen zunächst, daß Entscheidungsregel 1 zu einer korrekten Entscheidung führt mit einer Wahrscheinlichkeit strikt größer als 0,5. Dann betrachten wir einige Sonderfälle, aus denen sich alternative Entscheidungsregeln ableiten lassen und bei denen das obige Problem in einem Licht erscheint, das für unsere Intuition leichter zu fassen ist.

Behauptung 1:

Bei der Entscheidungsregel 1 gilt :

$$\begin{aligned}
 P(\text{korrekte Entscheidung}) &= P(Z > A | A < B) \cdot P(A < B) + P(Z < A | A > B) \cdot P(A > B) \\
 &> 0,5.
 \end{aligned}$$

Beweis:

Nach obiger Formel (1) ergibt sich unmittelbar

$$\begin{aligned}
 P(\text{korrekte Entscheidung}) &= P(A > Z | A > B) \cdot P(A > B) + P(A < Z | A < B) \cdot P(A < B).
 \end{aligned}$$

Nun ist $P(A > B) = 0,5$ wie man entweder intuitiv aus Symmetriegründen sofort, oder durch Anwendung der partiellen Integration sehen kann:

$$\begin{aligned}
 P(A > B) &= \int_{-\infty}^{\infty} \int_{-\infty}^a f(b)f(a)dbda \\
 &= \int_{-\infty}^{\infty} F(a)f(a)da = F(a)^2 \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(a)F(a)da \\
 &= 1 - P(A > B).
 \end{aligned}$$

Somit ergibt sich für die Wahrscheinlichkeit einer korrekten Entscheidung

$$\begin{aligned}
 P(\text{ korrekte Entscheidung }) &= 0,5 \cdot P(A > Z | A > B) + 0,5P \cdot (A < Z | A < B) \\
 &= 0,5(s + t) + 0,5(t + u) = 0,5 + 0,5t. \quad (2)
 \end{aligned}$$

wobei mit der Bezeichnung $M_1 = \min(A, B)$, $M_2 = \max(A, B)$ und g als Dichtefunktion der Zufallsvariablen Z die Größen s, t und u wie folgt definiert sind (siehe Abbildung 2):

$$S = \int_{-\infty}^{M_1} g(x)dx, T = \int_{M_1}^{M_2} g(x)dx, U = \int_{M_2}^{\infty} g(x)dx, s = ES, t = ET, u = EU.$$

Die Größen S, T und U sind also selbst Zufallsgrößen, da sie als Integrale mit zufälligen Grenzen $\min(A, B)$ und $\max(A, B)$ definiert sind. Die Werte von s, t und u als Erwartungswerte berechnen sich als Dreifachintegrale, z.B.

$$t = ET = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\int_{\min(a,b)}^{\max(a,b)} g(x)dx \right) f(a)f(b)dadb.$$

Die Abschätzung $P(\text{korrekte Entscheidung}) > 0,5$ ergibt sich dann wegen $s + t + u = 1$ aus der folgenden Überlegung: das Integral t ist sicherlich dann positiv, wenn es ein Intervall I in \mathbb{R} gibt, über dem $f(\cdot)$ und $g(\cdot)$ positiv sind. \square

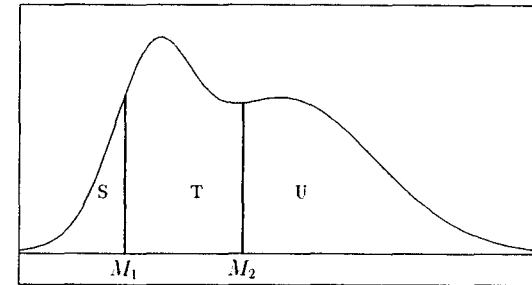


Abbildung 2: Dichte der Zufallsvariable Z

Für den vollständigen Beweis muß die Gültigkeit der Gleichung (2) noch erklärt werden. Abbildung 2 macht die Aussage plausibel und ist für den Schulunterricht vielleicht ausreichend. Ein exakter mathematischer Beweis kommt wohl ohne Berechnung eines Dreifachintegrals nicht aus (oder, alternativ, unter Bezugnahme auf den Begriff der bedingten Wahrscheinlichkeitsdichte).

Lemma 1 Unter den obigen Bezeichnungen gilt:

$$P(A > Z | A > B) = s + t = E \left(\int_{-\infty}^{M_2} g(x)dx \right)$$

Beweis:

Es gilt

$$\begin{aligned}
 P(A > Z | A > B) &= P(A > Z, A > B) / P(A > B) = 2P(A > Z, A > B) \\
 &= 2 \int_{-\infty}^{\infty} \int_{-\infty}^a \int_{-\infty}^a g(z)f(b)f(a)dzdbda.
 \end{aligned}$$

Wir betrachten folgende Mengen in \mathbb{R}^3 : $A_1 = \{(a, b, z) | z < a, b < a\}$, $A_2 = \{(a, b, z) | z < b, a < b\}$ und $A = \{(a, b, z) | z < \max(a, b)\}$. Es gilt: $A = A_1 \cup A_2$. Offensichtlich sind A_1 und A_2 symmetrisch in der ersten und zweiten Variable und es gilt $A_1 \cap A_2 = \emptyset$. Unsere gesuchte bedingte Wahrscheinlichkeit ist dargestellt als Integral über A_1 und der Integrand ist symmetrisch in a und b . Daher ist $\int_{A_1} = 1/2 \int_A$ und somit

$$\begin{aligned} P(A > Z | A > B) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\max(a,b)} g(z) f(b) f(a) dz db da. \\ &= E \left(\int_{-\infty}^{\max(a,b)} g(x) dx \right) \\ &= s + t. \square \end{aligned}$$

Der Beweis von $P(A < Z | A > B) = E(\int_{M_1}^{\infty} g(x) dx)$ erfolgt ganz analog. Aus Abbildung 2 wird außerdem deutlich, warum der Träger von Z die gesamte Menge der reellen Zahlen sein muß: wie auch immer die Lage von A und B , $t = E(\int_{M_1}^{\max(a,b)} g(x) dx)$ wird genau dann immer größer als 0 sein, falls ganz \mathbb{R} zum Träger von g gehört. Andernfalls kann der Schnitt der Trägermengen von A und B leer sein und dann ist $t = 0$.

Die Zahlen in den beiden Hosentaschen A und B sind selbst Zufallsvariablen, gegeben durch eine unbekannt Dichte f . Die bisherige Argumentation ging von der Annahme aus, daß nichts über die Dichte f bekannt ist. Wenn doch etwas von der Dichte f bekannt ist, erhellt sich das Paradoxon und auch andere Entscheidungsregeln sind denkbar. Zunächst genügt es, daß der Träger von f im Träger von g enthalten ist. Denn dadurch ist gewährleistet daß $A, B \in \text{Träger}(g)$ und somit $t > 0$. Falls der Träger von f bekannt ist, dann sind bei leichter Modifikation des Ausgangsproblems auch andere Entscheidungsregeln denkbar.

Entscheidungsregel 2:

Wähle zuerst (noch vor dem Nachschauen in Hosentasche 1) eine beliebige Zahl z_0 aus dem Inneren des Träger von f . Dann schau Dir die Zahl A in der ersten Hosentasche an. Ist $z_0 > A$, so entscheide Dich für $B > A$. Andernfalls soll die Entscheidung ausfallen: $A > B$.

Falls zum Beispiel f das Intervall $[-1,1]$ zum Träger hat, könnte man als Entscheidungskriterium nehmen: Ist $A > 0$? Aber jede andere Zahl $z_0 \in (-1, 1)$ erfüllt die Bedingung genauso, daß die darauf basierende Entscheidung mit einer Wahrscheinlichkeit $p > 0,5$ korrekt ist.

Behauptung 2:

Bei der Entscheidungsregel 2 gilt:

$$P(\text{korrekte Entscheidung}) > 0,5.$$

Beweis:

$$\begin{aligned} P(\text{korrekte Entscheidung}) &= P(A > z_0, A > B) + P(A < z_0, A < B) \\ &= \int_{z_0}^{\infty} \int_{-\infty}^u f(b) f(a) db da + \int_{-\infty}^{z_0} \int_a^{\infty} f(b) f(a) db da \\ &= \int_{z_0}^{\infty} F(a) f(a) da + \int_{-\infty}^{z_0} (1 - F(a)) f(a) da \\ &= \int_{z_0}^{\infty} F(a) f(a) da + F(z_0) - \int_{-\infty}^{z_0} F(a) f(a) da \\ &= \int_{-\infty}^{\infty} F(a) f(a) da + F(z_0) - 2 \int_{-\infty}^{z_0} F(a) f(a) da \\ &= \frac{1}{2} + F(z_0) - 2 \int_{-\infty}^{z_0} f(a) F(a) da. \end{aligned}$$

Da aber nach partieller Integration $\int_{-\infty}^{z_0} F(a) f(a) da = F(z_0)^2 - \int_{-\infty}^{z_0} f(a) F(a) da$ ist, folgt $\int_{-\infty}^{z_0} F(a) f(a) da = 1/2 F(z_0)^2$ und somit

$$\begin{aligned} P(\text{korrekte Entscheidung}) &= \frac{1}{2} + F(z_0) - F(z_0)^2 \\ &= \frac{1}{2} + F(z_0)(1 - F(z_0)) \\ &> \frac{1}{2}, \text{ falls } z_0 \in \text{Träger}(f) \square \end{aligned}$$

Entscheidungsregel 2 führt also nur dann zum gewünschten Erfolg (korrekte Entscheidung mit Wahrscheinlichkeit $> 1/2$), falls für die gewählte Zahl z_0 gilt: $0 < F(z_0) < 1$. In diesem Fall löst sich auch das Paradoxe schnell auf: dann nämlich hat die Beobachtung, daß $A > z_0$ wichtigen Informationsgehalt. A hat somit einen "Größentest" schon bestanden, im Gegensatz zu B , das der Testzahl z_0 unterlegen sein könnte (d.h. $B < z_0$). A hat sich schon einmal bewährt, während von B einfach Nichts bekannt ist. Falls aber $F(z_0) = 0$ oder $F(z_0) = 1$, dann ist ein Vergleich von z_0 mit A trivial, da dann immer gilt: $z_0 > A$ bzw. $z_0 < A$. Der Vergleich der beiden Zahlen enthält keinerlei Informationsgehalt. Welche Zahl $z_0 \in \text{Träger}(f)$ ist die beste Wahl, d.h. führt mit höchst möglicher Wahrscheinlichkeit zu einer korrekten Entscheidung? Äquivalent dazu ist die Frage, für welches z_0 der Ausdruck $\phi(z_0) := F(z_0)(1 - F(z_0))$ maximal wird. Dies führt auf die Bedingung $F(z_0) = 1/2$, d.h. der Median von f ist das optimale Entscheidungskriterium. Ist der Median von f bekannt, so kann eine korrekte Entscheidung mit einer Wahrscheinlichkeit von $p = 0,75$ getroffen werden. Ist aber gar nichts über die Verteilung von A und B bekannt, dann führt lediglich die Entscheidungsregel 1 zum gewünschten Resultat.

Danksagung:

Der Autor dankt seinen Kollegen Yonathan Anson, Antoine de Falguerolles, Brian Marx und Peter Van der Heijden für die lebhafte und amüsante Diskussion am Rande des 10th International Workshop on Statistical Modelling, 10.– 14. Juli 1995 in Innsbruck.

Literatur:

Blyth, C. (1972): On Simpson's Paradox and the Sure-things Principle, *Journal of the American Statistical Association* **67** 364–273

Klemisch, I. (1993): Ein Einstieg in das Drei-Türen Problem. *Stochastik in der Schule* **13**, Nr. 1, 8–14

Künzel, E. (1991): Über Simpson's Paradoxon. *Stochastik in der Schule*. **11**. Nr. 1, 54 – 62

Von Randow, G. (1992): *Das Ziegenproblem*. Reinbek bei Hamburg

Der Spiegel, 1991, Nr. 34, 212 – 213

Székely, G. (1990): *Paradoxa. Klassische und neue Überraschungen aus Wahrscheinlichkeitsrechnung und mathematischer Statistik*. Verlag Harri Deutsch, Frankfurt a.M.

Dr. Joachim Engel, PH Ludwigsburg, Postfach 220, 71602 Ludwigsburg
email: engel@ph-ludwigsburg.de