

# Baumdiagramme als roter Faden der Schulstochastik

Vorstellung eines baumorientierten Stochastiklehrgangs

STEFAN BARTZ, MECKEL

**Zusammenfassung:** Es wird einen Stochastiklehrgang vorgestellt, dem das Baumdiagramm als durchgängiges Lösungskonzept zu Grunde liegt. Über diesen gemeinsamen roten Faden können die einzelnen Themengebiete besser zueinander in Beziehung gesetzt und voneinander unterschieden werden. Die Gliederung der Schulstochastik wird klarer und überschaubarer. Die Lösungsverfahren können einheitlicher gestaltet werden. Die Schüler erhalten wesentlich mehr Sicherheit beim Lösen stochastischer Probleme. Anhand von 10 Beispielaufgaben wird der Lehrgang exemplarisch erläutert.

## Idee

Betrachtet man die Gliederung der Stochastik in Schulbüchern, so ergeben sich ungefähr folgende 10 Hauptthemenbereiche:

1	Bestimmen von Wahrscheinlichkeiten mit Pfadregeln
2	Bestimmen von Wahrscheinlichkeiten bei Bernoulli-Ketten; Kombinatorik
3	Bestimmen von Wahrscheinlichkeiten bei bedingten Ereignissen; Bayes'sche Regel
4	Bestimmen von Wahrscheinlichkeiten bei Markov-Ketten
	Bestimmen von Wahrscheinlichkeiten mit Hilfe verschiedener Verteilungsfunktionen.
5	a) Binomialverteilung
6	b) Hypergeometrische Verteilung
7	c) Normalverteilung
	Lösen von Problemen der beurteilenden Statistik
8	a) Stichprobenumfang bestimmen
9	b) Schätzen von Parametern
10	c) Testen von Hypothesen
	Exkurse und Projekte: Beschreibende Statistik, Simulationen, ...

Tabelle 1

Bei allen Themen geht es hauptsächlich darum, Wahrscheinlichkeiten zu bestimmen (1-7) oder umgekehrt, bei vorgegebenen Wahrscheinlichkeitsbereichen zugehörige Variablen- bzw. Parameterintervalle zu ermitteln (8-10).

Zu diesen gesuchten bzw. vorgegebenen Wahrscheinlichkeiten gehören immer interessierende Ereignisse. Die Grundidee ist nun, bei allen Themengebieten zunächst einmal diese interessierenden Ereignisse und ihre Zerlegungsstruktur mit Hilfe von Baumdiagrammen zu untersuchen. Die dabei gewonnen Erkenntnisse sind für den weiteren Lösungsvorgang von zentraler Bedeutung. Denn das Zerlegungsdiagramm des Ereignisses zeigt den Schülern, mit welchem Aufgabentyp sie es zu tun

haben und welche Zufallsvariable mit welchem Verteilungsmodell und mit welchen Parametern gegebenenfalls verwendet werden kann.

Mit anderen Worten: Hinter den 10 Themengebieten der Schulstochastik stehen jeweils spezifische Ereignisstrukturen, die in Form von spezifischen Baumdiagrammen visualisiert werden können. Nennt man die Themengebiete aus Tabelle 1 nun so um, dass die zugrundeliegenden Ereignisstrukturen besser zum Ausdruck kommen, so ergibt sich folgende neue Tabelle:

<b>I. Vorhersagen von Wkn mit Baumdiagrammen</b>	
1. Einfache Bäume	→ Pfadregeln
2. Mammutbäume	→ Kombinatorik
	(Exkurs: Unendliche Bäume → Markov-Ketten)
3. Bedingte Bäume	→ Bayes
<b>II. Vorhersagen von Wkn mit Funktionen<sup>1</sup></b>	
4. Binomiale Bäume	→ $\text{Bin}_{n,p}(x)$
5. Hypergeom. Bäume	→ $\text{Hyp}_{n,R,N}(x)$
6. Stetige Knotenwerte	→ $\text{Nor}_{\mu,\sigma}(x)$
<b>III. Umkehrprobleme bei binomialen Bäumen</b>	
7. n gesucht	→ Stichprobenumfang
8. p gesucht	→ Schätzen
9. x gesucht	→ Testen

Tabelle 2

Obwohl Tabelle 2 dieselben Themengebiete wie Tabelle 1 beschreibt, erscheint der Stochastiklehrgang jetzt überschaubarer und einprägsamer:

- Den Schülern kann bereits zu Beginn der Aufbau des gesamten Kurses vor Augen geführt werden. Die Beziehungen der einzelnen Kapitel zueinander können unmittelbar erfasst werden.
- Die neue Gliederung erweist sich in Anwendungssituationen als ausgesprochen hilfreich.
- Die Hauptkapitel lassen die Leitidee des Stochastiklehrgangs *Vorhersagen von Wahrscheinlichkeiten* erkennen. Gleichzeitig wird die Beziehung zur Analysis, bei der es um das *Vorhersagen von determinierten Vorgängen* geht, aufgezeigt. Mit den Begriffen *Funktionen* und *Umkehrprobleme* kann an bekanntes Vorwissen angeknüpft werden.

Im Folgenden wird der baumorientierte Stochastiklehrgang durch 10 Beispielaufgaben exemplarisch vorgestellt. Die Beispiele sollen zeigen, welche Baumtypen bei den einzelnen Themengebieten entstehen und inwiefern diese das weitere Lösungsverfahren bestimmen. Außerdem soll deutlich gemacht werden, dass bei allen 10 Themengebieten ein einheitlicher Lösungsansatz verwendet werden kann. Er besteht immer aus folgenden 3 Schritten:



man sich den Zustandsgraph als ein geschlossenes Wasserleitungssystem vor, durch das pro Zeiteinheit unterschiedliche Wassermengen von Knoten zu Knoten gepumpt werden und startet man mit einem Liter Wasser im Startknoten, so kann mithilfe von Excel schnell berechnet werden, nach welcher Zeiteinheit wie viel Wasser (Wahrscheinlichkeit) in welchen Knoten vorhanden ist. Eine Verallgemeinerung des Verfahrens führt unmittelbar zur Matrizenrechnung, die in diesem Zusammenhang anwendungsorientiert eingeführt werden kann. Im Anschluss an diese iterative Lösung kann dann der im Beispiel dargestellte explizite Lösungsweg entwickelt werden.

Den Gewinnzustand Z5 erlangt man hier mit 30%iger Wahrscheinlichkeit ab Z1. Befindet man sich bereits in Z2 oder Z3, so erhöht sich die Gewinnwahrscheinlichkeit auf 40% bzw. 60%. Die Wartezeit bis zum Spielende beträgt ab Z1 im Mittel 4,2 Würfe.

### 3 Bedingte Bäume

Die Alarmanlage eines Geschäftes gibt bei einem Einbruch mit einer Wk von 99% Alarm. Aber auch ohne Einbruch gibt sie mit einer Wk von 0,5% (falschen) Alarm. Die Einbruchswahrscheinlichkeit in der Nacht betrage 0,1%. Wie groß ist die Wahrscheinlichkeit, dass, wenn der Alarm ausgelöst wird, tatsächlich ein Einbruch stattfindet?

(1) E:  ${}_A B$ , d.h. es findet ein Einbruch B statt, unter der Bedingung A, dass die Alarmanlage angesprungen ist.

(2)

(3)  $P(A) = P(B \cap A) + P(\bar{B} \cap A) = 0,599\%$   
 $P(E) = P({}_A B) = \frac{P(A \cap B)}{P(A)} = \frac{0,001 \cdot 0,99}{0,00599} = 16,54\%$ .

Bei bedingten Bäumen hängt die Knotenwahrscheinlichkeit davon ab, was im vorherigen Knoten passiert; so ist im oberen Diagramm z. B.  $P({}_B A) \neq$

$P({}_B A)$ . Das bedeutet aber auch, dass die Knoten von Baum und Umkehrbaum unterschiedliche Astwahrscheinlichkeiten besitzen, so ist z.B.  $P(B) \neq P({}_A B)$ . Ist dieser Zusammenhang und die damit einhergehende Schreibweise geklärt, gelingt es relativ leicht, die Astwahrscheinlichkeiten des Umkehrbaums aus denen des ursprünglichen mit Hilfe der Pfadregeln zu ermitteln. So lässt sich im Beispiel die Astwahrscheinlichkeit  $P(A)$  des zweiten Baumes durch die Addition der Alarm-Pfade des ersten Baumes bestimmen. Die Wahrscheinlichkeit für einen berechtigten Alarm  $P({}_A B)$  ergibt sich aus dem Anteil von  $P(A \cap B)$  bezogen auf  $P(A)$ .

Allerdings kommt es bei bedingten Bäumen – also bei abhängigen Teilereignissen – häufig zu Fehlurteilen. Im ersten Moment unterscheiden die Schüler nicht zwischen  $P({}_B A)=99\%$  und  $P({}_A B)=16,54\%$ . Sie können nicht glauben, dass eine so zuverlässige Alarmanlage nur in 16,54% der Fälle berechtigterweise Alarm gibt und trauen diesem errechneten Wert nicht. Sie sehen nicht, dass sich die 99% nur auf ganz wenige Fälle beziehen. Aus diesem Grund wird bei der Einführung des Themas oft mit sogenannten Häufigkeitsbäumen gearbeitet, bei denen die Knoteneinträge um absolute Häufigkeitsangaben ergänzt werden. Anhand der absoluten Zahlen können die Schüler die tatsächlichen Verhältnisse deutlicher erkennen (vgl. Gigerenzer [www.max-wissen.de/Tools/drucken/3859.html](http://www.max-wissen.de/Tools/drucken/3859.html))

### 4 Binomiale Bäume<sup>1</sup>

Der Anteil der Linkshänder betrage 14%. Wie groß ist die Wahrscheinlichkeit, dass von 15 Personen mindestens 5 Linkshänder sind?

(1) E: mind. 5 von 15 Personen sind Linkshänder

(2)

(3)  $P(E) = P(X \geq 5) = \text{Bin}_{15;0,14}(X \geq 5) = 1 - \text{Bin}_{15;0,14}(X \leq 4) = 4,78\%$

Wenden die Schüler den 3-schrittigen Lösungsansatz bei diesem Aufgabentyp an, stoßen sie auf spezielle Mammutbäume, bei denen jeder Knoten genau 2 Nachfolger mit jeweils gleichbleibender Astwahrscheinlichkeit hat (Bernoulli-Kette). Es müssen nur wenige Stufen des Baumes skizziert werden, um zu erkennen, dass hier das Modell der Binomialverteilung verwendet werden kann. Auch die Zufallsvariable X und die benötigten Parameter ergeben sich anhand der Baumskizze. Damit ist geklärt, dass  $P(E)$  mit  $\text{Bin}_{15;0,14}(X \geq 5)$  bestimmt

werden kann. Dieser Wert, und auch die Wahrscheinlichkeits- und Parameterwerte der nachfolgenden Musteraufgaben, werden am besten mithilfe der Exceldatei *stochastik.xls* ermittelt. Sie kann die Verteilungstabellen im Schulunterricht ersetzen, erleichtert das Verständnis und steht als Freedownload unter [www.stefanbartz.de](http://www.stefanbartz.de) zur Verfügung (vgl. Bartz 2007).

## 5 Hypergeometrische Bäume

Wie groß ist die Wk, beim Lotto höchstens 2 Richtige zu haben?

(1) E: *Es werden höchstens 2 Richtige gezogen.*  
 (2)

hypergeometrischer Baum mit  
 $n = 6$   
 $N = 49$   
 $R = 6$ ;  
 $E: X \leq 2$

(3)  $P(E) = P(X \leq 2) = \text{Hyp}_{6; 6; 49}(X \leq 2) = 98,14\%$

Auch hier genügen wenige Stufen um das Vorliegen eines hypergeometrischen Baumes zu erkennen: jeder Knoten hat 2 Nachfolger, die Treffer- und Nichttreffer-Wahrscheinlichkeiten vermindern sich kontinuierlich (Ziehen ohne Zurücklegen). Die Baumskizze liefert den Schülern zuverlässig die zugrunde liegende Verteilung, die Zufallsvariable  $X$  sowie die zu wählenden Parameter.

## 6 Stetige Knotenwerte

Auf einer Hühnerfarm mit sehr vielen Hühnern stellt sich heraus, dass ein Ei im Durchschnitt 50g wiegt (Standardabweichung 5g). Wie groß ist die Wahrscheinlichkeit, dass ein zufällig herausgegriffenes Ei zwischen 48g und 54g wiegt?

(1) E: *Das gegriffene Ei wiegt 48g – 54g.*  
 (2)

stetige Knotenwerte  
 $E: 48 \leq X \leq 54$

(3)  $P(E) = \text{Nor}_{50; 5}(48 \leq X \leq 54)$  | Vermutung  
 $= \text{Nor}_{50; 5}(X \leq 54) - \text{Nor}_{50; 5}(X \leq 48)$   
 $= 44,35\%$

Beim Versuch, das interessierende Ereignis in Teilereignisse zu zerlegen, erkennt der Lernende, dass als Knotenwerte alle Zwischenwerte angenommen werden können. Das angedeutete Baumdiagramm kann den Schülern zwar zeigen, dass die Zufallsvariable  $X$  hier *Gewicht eines Eies* lautet und dass diese stetig ist d.h. unendlich viele Ausprägungen

annehmen kann. Die Verteilung der Zufallsvariable lässt sich jedoch nicht mehr anhand des Baums herausfinden. Da sich im obigen Fall die Zufallsvariable  $X$  aus sehr vielen Einflüssen wie z.B. Gewicht des Huhns, Alter, Gesundheit, Standort, Vererbung, usw. zusammensetzt, kann nach dem zentralen Grenzwertsatz eine Normalverteilung für  $X$  zwar vermutet werden, es besteht jedoch die Notwendigkeit dies nachzuweisen.

Die Grenzen des baumorientierten Ansatzes werden hier deutlich: Bei stetigen Knotenwerten benötigt man zusätzliche Methoden, um die zugrunde liegende Verteilung aufklären zu können.

## 7 Umkehrprobleme bei binomialen Bäumen ( $n$ gesucht)

Wie oft muss man einen gerechten Würfel mind. werfen, um mit einer Wk von mind. 95% mind. 2-mal eine „6“ zu erhalten?

(1) E: *Man erhält mindestens 2-mal eine „6“*  
 (2)

binomialer Baum mit  
 $n = ?$   
 $p = 1/6$   
 $E: X \geq 2$

(3)  $P(X \geq 2) \geq 0,95$   
 $\Rightarrow \text{Bin}_{n; 0,167}(X \geq 2) \geq 0,95$   
 $\Leftrightarrow 1 - \text{Bin}_{n; 0,167}(X \leq 1) \geq 0,95$   
 $\Leftrightarrow \text{Bin}_{n; 0,167}(X \leq 1) \leq 0,05$   
 $\Rightarrow n \in [27; \infty)$

Auch bei Umkehraufgaben kann an dem baumorientierten 3-Schritt festgehalten werden. Wenige Stufen des Baumes erbringen den Nachweis, dass dieses Problem im Modell der Binomialverteilung gelöst werden kann und liefern Zufallsvariable und Parameter. Das bereits oben erwähnte Excelblatt liefert dann direkt die Werte von  $n$ .

## 8 $p$ gesucht (Schätzen)

Man hat keinen Anhaltspunkt, wie groß die Trefferwahrscheinlichkeit  $p$  für eine „6“ bei einem verbeulten Würfel ist. Deshalb wirft man ihn 50-mal und erhält 32-mal eine „6“. Schätze  $p$  mit einem 95% Vertrauensintervall.

(1) E: *Man erhält 32-mal eine „6“*  
 (2)

binomialer Baum mit  
 $n = 50$   
 $p = ?$   
 $E: X = 32$

(3) Annahme: E liegt im 95%igen Hauptstreu-bereich von  $\text{Bin}_{50; p}(X = x)$   
 $\Rightarrow 0,025 \leq \text{Bin}_{50; p}(X \leq 32) \leq 0,975$

$$\Rightarrow p \in [0.5123; 0.7708]$$

Wenn man davon ausgeht, dass das erzielte Stichprobenergebnis innerhalb des Hauptstrebereichs ( $\geq 95\%$ ) liegt, muss sich  $p$  im angegebenen Intervall befinden.

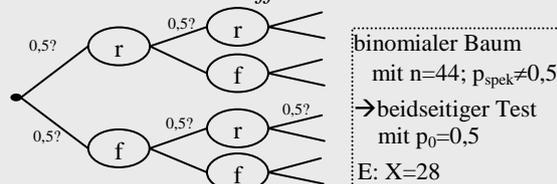
Auf den ersten Blick erscheint diese Aufgabe ähnlich der vorherigen: dort wurde der Parameter  $n$  gesucht, hier ist es  $p$ . Probleme bereitet hier jedoch die Variable: Wieso betrachtet man  $\text{Bin}_{50;p}(X \leq 32)$  obwohl das interessierende Ereignis doch  $X=32$  lautet? Den Schülern muss deutlich gemacht werden, dass der Wert  $\text{Bin}_{50;p}(X=32)$  nur die Wahrscheinlichkeit der 32-er Histogrammsäule liefert und nichts darüber aussagt, ob diese 32-er Säule innerhalb des Hauptstrebereichs liegt. Dazu muss der Wert der kumulierten Funktion  $\text{Bin}_{50;p}(X \leq 32)$  zu Rate gezogen werden. Ist dieser Zusammenhang geklärt, lässt sich  $p$  mit dem angesprochenen Excelblatt direkt bestimmen (Andere Lösungsverfahren benötigen dafür einen deutlich größeren Aufwand; vgl. Bartz 2007.).

## 9 x gesucht (Testen)

Annika behauptet, dass sie nur am Geschmack erkennt, ob der Tee mit entkalktem oder nicht entkalktem Wasser hergestellt wurde. Bei 44 Versuchen stimmt ihre Angabe in 28 Fällen. Ist damit bestätigt, dass Annika einen sehr sensiblen Geschmackssinn hat (Signifikanzniveau 5%)?

(1) E: Man erhält 28 Treffer bei 44 Versuchen

(2)



(3) Liegt E noch im HSB von  $\text{Bin}_{44;p_0}$ ?

$$0,025 \leq P(X \leq x) \leq 0,975$$

$$\Rightarrow 0,025 \leq \text{Bin}_{44; 0,5}(X \leq x) \leq 0,975$$

$$\Rightarrow x \in [16; 27]$$

Da das Stichprobenergebnis E außerhalb des Hauptstrebereichs ( $\geq 95\%$ ) liegt, kann  $p_0$  abgelehnt und somit  $p_{\text{spek}}$  bestätigt werden; d.h. man kann davon ausgehen, dass Annika nicht nur rät.

Analog zu oben liefert die Baumskizze die Zufallsvariable  $X$ , den Nachweis ihrer Binomialverteilung und den Parameter  $n=44$ .  $p_{\text{spek}}$  beschreibt die spektakuläre neue Hypothese, die man verifizieren will,  $p_0$  die ihr entgegenstehende Nullhypothese. Diese kann hier abgelehnt werden, da das Stichprobenergebnis E außerhalb des 95%-igen Hauptstrebereichs (HSB) der  $\text{Bin}_{44; p_0}(X=x)$ -Verteilung liegt.

## Fazit

Durch das vorgestellte Konzept der Baumorientierung kann die Stochastik den Schülern kompakter, zusammenhängender und überschaubarer präsentiert werden. Die Lösungsverfahren können auf einen gemeinsamen roten Faden zurückgeführt werden. Die Schüler erlangen dadurch mehr Sicherheit beim Lösen von Problemen. Sie können sich die vermittelten Inhalte besser einprägen und so nachhaltiger Kompetenzen aufbauen.

Weitere Verteilungen und Umkehrprobleme können in das Konzept integriert werden; z.B. *Multi-nomialen Bäumen* und *Umkehrprobleme bei normalverteilten Zufallsvariablen*.

Der Lehrgang ist vereinbar mit bisherigen Gliederungen. Gängige Schulbücher und die darin gestellten Aufgaben können weiterhin verwendet werden. Die Kapitel müssen im Unterricht lediglich umbenannt und die Aufgaben – wie oben gesehen – anhand von Baumdiagrammen gelöst werden.

Der baumorientierte Ansatz stößt bei stetigen Knotenwerten an seine Grenzen. Dort werden zusätzliche Verfahren benötigt, um die zugrunde liegende Verteilung erkennen zu können.

## Anmerkungen

<sup>1</sup> Aus didaktischen Gründen werden die Begriffe *Binomialfunktion* für  $\text{Bin}_{n,p}(X=x)$  und *kumulierte Binomialfunktion* für  $\text{Bin}_{n,p}(X \leq x)$ , statt der für Schüler verwirrenden Begriffe „Binomialverteilung“ und „Verteilungsfunktion der Binomialverteilung“ verwendet.

Die Schreibweise knüpfen an die der Analysis an: Funktionsname mit 3 Buchstaben,  $x$  als Variablenname und in Klammern, Scharparameter als Index, also  $\text{Bin}_{n,p}(x)$  analog zu  $\sin(x)$  bzw.  $f_i(x)$ . Entsprechend wird bei den Verteilungen  $\text{Hyp}_{n,R,N}(x)$  und  $\text{Nor}_{\mu,\sigma}(x)$  verfahren.

## Literatur

Bartz, S. (2007): Excelblatt vereinfacht Stochastik. *Stochastik in der Schule* 27(2), S. 25-29.  
[www.stefanbartz.de/materialien.htm](http://www.stefanbartz.de/materialien.htm)