12.4 Die probabilistische Methode

Bisher haben wir die Stochastik als eine Theorie betrachtet, die uns »reelle« Zufallsexperimente analysieren lässt. Es gibt aber auch eine andere Seite der Stochastik: Man kann mit ihrer Hilfe Aussagen auch in einigen Situationen treffen, wo der Zufall überhaupt keine Rolle spielt!

Die Hauptidee der sogenannten *probabilistischen Methode* ist die folgende: Will man die *Existenz* eines Objekts mit bestimmten Eigenschaften zeigen, so definiert man einen entsprechenden Wahrscheinlichkeitsraum und zeigt, dass ein zufällig gewähltes Element mit einer *positiven* Wahrscheinlichkeit die gewünschte Eigenschaft hat.

Im Allgemeinen ist eine Menge M der Objekte sowie eine Funktion $f : M \to \mathbb{R}$ gegeben. Für einen Schwellenwert t will man wissen, ob es ein Objekt $x \in M$ mit $f(x) \geq t$ gibt. Dazu wählt man eine entsprechende Wahrscheinlichkeitsverteilung $\Pr : M \to [0,1]$ und betrachtet den resultierenden Wahrscheinlichkeitsraum $(M, \Pr)$. In diesem Raum ist f eine Zufallsvariable. Man berechnet dann den Erwartungswert $\mathrm{E}(f)$ dieser Zufallsvariable und testet, ob $\mathrm{E}(f) \geq t$ oder $\Pr(f(x) \geq t) > 0$ gilt. Ist mindestens eines davon der Fall, so muss es mindestens ein Element $x_0 \in M$ mit $f(x_0) \geq t$ geben: Würde es nämlich $f(x) < t$ für *alle* $x \in M$ gelten, so hätten wir

$$\Pr(f(x) \geq t) = \Pr(\emptyset) = 0$$

und

$$\mathrm{E}(f) = \sum_{x \in M} f(x) \cdot \Pr(f = x) < \sum_{x \in M} t \cdot \Pr(f = x) = t \cdot \sum_{x \in M} \Pr(f = x) = t\,.$$

Die Eigenschaft

aus $\mathrm{E}(f) \geq t$ folgt $f(x) \geq t$ für mindestens ein $x \in M$

nennt man auch das *Taubenschlagprinzip des Erwartungswertes*. Ein Prototyp dieser (überraschend mächtigen) Methode ist das folgende »Mittel-Argument«: Ist der arithmetische Mittel

$$\frac{x_1 + \cdots + x_n}{n}$$

der Zahlen $x_1, \ldots, x_n \in \mathbb{R}$ größer als a, so muss es mindestens ein j mit $x_j > a$ geben. Die Nützlichkeit dieses Argument liegt in der Tatsache, dass es oft viel leichter ist, eine Abschätzung für das Mittel zu finden als ein j mit $x_j > a$ zu bestimmen. Wir demonstrieren die probabilistische Methode an ein paar typischen Beispielen.

Sei $G = (V, E)$ ein ungerichteter Graph. Eine Knotenmenge $S \subseteq V$ heißt *Clique*, falls zwischen je zwei Knoten in S eine Kante liegt. Liegt zwischen keinen zwei Knoten eine Kante, so heißt S *unabhängige Menge*. Sei $r(G)$ die kleinste Zahl r, so dass der Graph G weder eine Clique noch eine unabhängige Menge mit r Knoten besitzt.

Frank Plumpton Ramsey hat im Jahre 1930 bewiesen, dass jeder Graph G mit n Knoten entweder eine Clique oder eine unabhängige Menge mit $\frac{1}{2}\log_2 n$ Knoten enthalten muss, also $r(G) > \frac{1}{2}\log_2 n$ gilt. Eine natürliche Frage ist daher, ob es Graphen mit $r(G) \leq c\log_2 n$ für eine Konstante $c > 0$ überhaupt gibt. Solche Graphen nennt man *Ramsey-*

Graphen. Mit Hilfe der probabilistischen Methode hat Paul Erdős in 1947 bewiesen, dass solche Graphen doch existieren!

Satz 12.35:

Ramsey-Graphen mit beliebig vielen Knoten existieren: Für alle $n \geq 2$ gibt es Graphen G auf n Knoten mit $r(G) \leq 2\log_2 n$.

Beweis:

Um die Existenz von Ramsey-Graphen zu beweisen, betrachten wir Zufallsgraphen über der Knotenmenge $V = \{1, \ldots, n\}$: Wir werfen für jede potenzielle Kante uv eine faire Münze und setzen die Kante ein, wenn das Ergebnis »Wappen« ist.

Wir fixieren eine Knotenmenge $S \subseteq V$ der Größe k und sei A_S das Ereignis »S ist eine Clique oder eine unabhängige Menge«. Es ist $\Pr(A_S) = 2 \cdot 2^{-\binom{k}{2}}$, denn entweder ist S eine Clique und alle $\binom{k}{2}$ Kanten sind vorhanden oder S ist eine unabhängige Menge und keine der $\binom{k}{2}$ Kanten ist vorhanden. Wir sind vor Allem an der Wahrscheinlichkeit p_k interessiert, dass ein Zufallsgraph G eine Clique der Größe k oder eine unabhängige Menge der Größe k besitzt. Da wir nur $\binom{n}{k}$ k-elementigen Mengen $S \subseteq V$ haben, gilt nach der Summen-Schranke für Wahrscheinlichkeiten (Behauptung 11.4):

$$p_k \leq \binom{n}{k} \cdot 2 \cdot 2^{-\binom{k}{2}} < \frac{n^k}{k!} \cdot \frac{2 \cdot 2^{k/2}}{2^{k^2/2}}.$$

Wir setzen $k = 2\log_2 n$ und erhalten $n^k = (2^{k/2})^k = 2^{k^2/2}$. Da andererseits $2 \cdot 2^{k/2} < k!$ für $k \geq 4$ gilt, folgt somit $p_k < 1$ für $k \geq 4$. Es gibt somit Graphen, die keine Cliquen oder unabhängige Mengen der Größe $2\log_2 n$ besitzen. $\square$

Sei $K_n = (V, E)$ ein vollständiger ungerichteter Graph mit der Knotenmenge $V = \{1, \ldots, n\}$. Der Graph besitzt also alle $|E| = \binom{n}{2}$ mögliche Kanten. Eine *bipartite Clique* ist ein bipartiter Graph von der Form $H = L \times R$ mit $L, R \subseteq V$ und $L \cap R = \emptyset$. (Hier steht »L« bzw. »R« für die »linke« bzw. für die »rechte« Seite der Clique.) Das *Gewicht* einer solchen Clique ist die Anzahl $v(H) = |L| + |R|$ ihrer Knoten.

Unser Ziel ist alle Kanten von K_n mit bipartiten Cliquen $H_i = L_i \times R_i$, $i = 1, \ldots, t$ so zu überdecken, dass das Gesamtgewicht $v(H_1) + \cdots + v(H_t)$ der dabei beteiligten Cliquen möglichst klein wird.

Eine ähliche Frage haben wir bereits in Abschnitt 6.3.1 behandelt. Da wollten wir die Kanten von K_n in möglichst wenigen disjunkten bipartiten Cliquen zerlegen. Mit Hilfe der linearen Algebra haben wir da gezeigt, dass man dafür mindestens $n - 1$ Cliquen benötigt. Nun interessiert uns nicht die *Anzahl* der Cliquen, sondern ihr Gesamtgewicht. Dabei verlangen wir nicht mehr, dass die Cliquen disjunkt sein müssen – sie können auch gemeinsame Kanten haben.

Man kann eine Überdeckung von K_n mit dem Gesamtgewicht höchstens $2n\log_2 n$ folgendermaßen konstruieren. Einfachheitshalber sei $n = 2^k$ eine Zweierpotenz. Zunächst weisen wir jedem Knoten $v \in V$ einen eindeutigen Vektor $\boldsymbol{v} = (v_1, \ldots, v_k) \in \{0,1\}^k$ zu und betrachten die folgenden $2n$ Cliquen

$$H_i^a = \{(u, v) \colon u_i = a,\ \boldsymbol{v}_i = 1 - a\} \qquad (a = 0,1;\ i = 1, \ldots, k).$$

Da sich je zwei *verschiedene* Vektoren in mindestens einer der k Koordinaten unterscheiden, liegt jede Kante von K_n in mindestens einer dieser Cliquen. Da für jedes $i = 1, \ldots, k$ genau die Hälfte der Vektoren eine Eins bzw. eine Null in der i-ten Koordinate haben, enthält jede Clique H_i genau $v(H_i) = (n/2) + (n/2) = n$ Knoten. Das Gesamtgewicht dieser Überdeckung ist also $2nk = 2n \log_2 n$.

Mit der probabilistischen Methode zeigen wir nun, dass es viel besser auch nicht geht.

Satz 12.36:

Jede Überdeckung von K_n mit bipartiten Cliquen muss das Gesamtgewicht mindestens $n \log_2 n$ haben.

Beweis:

Sei $H_i = L_i \times R_i$, $i = 1, \ldots, t$ eine Überdeckung der Kanten von $K_n = (V, E)$ mit $V = \{1, \ldots, n\}$. Sei $g = \sum_{i=1}^{n}(|L_i| + |R_i|)$ das Gesamtgewicht dieser Überdeckung. Für jeden Knoten $v \in V$ sei $m_v = |\{i \colon v \in L_i \cup R_i\}|$ die Anzahl der Cliquen, die diesen Knoten enthalten. Das Prinzip der doppelten Abzählung ergibt (siehe Aufgabe 3.14)

$$g = \sum_{i=1}^{t}(|L_i| + |R_i|) = \sum_{v=1}^{n} m_v \,.$$

Es reicht also die letzte Summe nach unten abzuschätzen. Dazu werfen wir für jede Clique $H_i = L_i \times R_i$ eine faire 0-1 Münze. Kommt 0, so entfernen wir alle Knoten L_i aus V; sonst entfernen wir alle Knoten R_i. Sei $X = X_1 + \cdots + X_n$, wobei X_v die Indikatorvariable für das Ereignis »Knoten v überlebt« ist.

Da je zwei Knoten in K_n durch eine Kante verbunden sind und diese Kante durch mindestens eine der Cliquen H_i überdeckt wird, kann am Ende höchstens ein Knoten überleben. Somit gilt $\mathrm{E}(X) \leq 1$. Andererseits wird jeder einzelne Knoten v mit Wahrscheinlichkeit 2^{-m_v} überleben: Es gibt nur m_v für den Knoten v »gefährliche« Schritte und in jedem dieser Schritten wird der Knoten mit Wahrscheinlichkeit $1/2$ überleben. Nach der Linearität des Erwartungswertes erhalten wir

$$\sum_{v=1}^{n} 2^{-m_v} = \sum_{v=1}^{n} \Pr(v \text{ überlebt}) = \sum_{v=1}^{n} \mathrm{E}(X_v) = \mathrm{E}(X) \leq 1 \,.$$

Wir wissen, dass das arithmetische Mittel der Zahlen $a_1, \ldots, a_n$ mindestens so gross wie ihr geometrisches Mittel ist (Aufgabe 3.12):

$$\frac{1}{n} \sum_{v=1}^{n} a_v \geq \Big(\prod_{v=1}^{n} a_v \Big)^{1/n} .$$

Angewand mit $a_v = 2^{-m_v}$ ergibt dies

$$\frac{1}{n} \geq \frac{1}{n} \sum_{v=1}^{n} 2^{-m_v} \geq \Big(\prod_{v=1}^{n} 2^{-m_v} \Big)^{1/n} = 2^{-\frac{1}{n} \sum_{v=1}^{n} m_v} ,$$

woraus $2^{\frac{1}{n} \sum_{v=1}^{n} m_v} \geq n$ und somit auch $\frac{1}{n} \sum_{v=1}^{n} m_v \geq \log_2 n$ folgt. $\square$