Gleichmäßige Verteilung von Punkten in einem Würfel

Einstieg

Wie verteilt man vier Punkte möglichst gleichmäßig in einem Quadrat? Die Antwort erscheint einfach: man zerteilt das Quadrat in vier Quadrate mit halber Seitenlänge und wählt die vier Mittelpunkte der Teilquadrate.

Genauso kann man natürlich k^2 (mit $k \in \mathbb{N}$) Punkte *gleichmäßig* in einem Quadrat unterbringen. Hat man einen dreidimensionalen Würfel, so kann man ihn analog in k^3 Teilwürfel zerlegen und dann die $n = k^3$ Mittelpunkte dieser Teilwürfel betrachten. So eine Punktanordnung nennen wir ein *äquidistantes Gitter*. Diese Art der Punktanordnung funktioniert für jede Dimension d, falls man $n = k^d$ Punkte nimmt. Wo ist das Problem?

Zum Begriff der Gleichmäßigkeit

Das Problem besteht zunächst darin, dass wir gar nicht gesagt haben, wie wir die Gleichmäßigkeit von verschiedenen Punktmengen

$$M_n = \{x_1, x_2, \ldots, x_n\} \subset [0,1]^d$$

miteinander vergleichen wollen. Hier ist d die Dimension des „Würfels“, man kann sich $d = 2$ oder $d = 3$ oder auch $d = 1$ vorstellen. Hoffentlich stört sich niemand an der Bezeichnung Würfel. Im Alltag würde man natürlich von Quadraten ($d = 2$) und Strecken ($d = 1$) sprechen.

Wir wollen einen Begriff der Gleichmäßigkeit (oder auch Ungleichmäßigkeit) betrachten, den man *Diskrepanz* nennt. Schon bevor es mit der Definition losgeht, möchte ich betonen, dass es andere sinnvolle Arten gibt, wie man diesen Begriff präzisieren könnte. Der hier vorgestellte Diskrepanzbegriff hat sich aber als ziemlich wichtig für verschiedene Teilgebiete der Mathematik herausgestellt.

Die Diskrepanz einer Punktmenge

Zunächst brauchen wir ein paar Bezeichnungen. Seien $x, y, z \in [0,1]^d$, also Vektoren mit d Koordinaten, zum Beispiel

$$x = (x^1, x^2, \ldots, x^d)\,.$$

Die Hochzahlen bedeuten keine Exponenten, hier wird nichts potenziert! Wir schreiben nur die Indizes (für die Nummern der Koordinaten) oben hin, weil wir ja mit x_i bereits verschiedene Punkte bezeichnet haben. Es gilt also

$$x_i = (x_i^1, x_i^2, \ldots, x_i^d).$$

Es genügt, wenn man an den Fall $d = 3$ denkt. Höherdimensionale Räume sind in der Mathematik oft wichtig, und man muss sich dabei gar nichts Geometrisches vorstellen: Gemeint ist erstmal nur, dass ein „Objekt" x durch d Zahlen $x^1, x^2, \ldots, x^d$ vollständig beschrieben wird. Dieses Objekt x nennen wir jetzt einen Punkt, wie im Fall $d \in \{1, 2, 3\}$.

Hat man zwei Punkte $x, z \in [0,1]^d$, so kann es sein, dass $x^i \leq z^i$ für alle i gilt. Dann schreiben wir $x \leq z$. Wir betrachten Teilquader von $[0,1]^d$ der Form

$$[x, z] = \{y \mid x \leq y \leq z\},$$

wobei wir bei dieser Schreibweise immer $x \leq z$ unterstellen. (Siehe Abbildung 1.) Das Volumen von $[x, z]$ bezeichnen wir mit $V([x, z])$, es gilt

$$V([x, z]) = \prod_{i=1}^{d} (z^i - x^i).$$

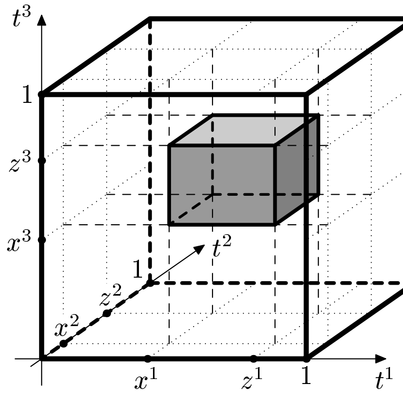


Abbildung 1. Teilquader $[x, z]$ von $[0,1]^3$

Jeder kennt die Formel (für $d = 2$ und $d = 3$) aus der Schule (Fläche gleich Höhe mal Breite bzw. Volumen gleich Höhe mal Breite mal Tiefe). Jetzt fragt sich der Leser vielleicht: Was hat das Volumen (wir sprechen vom Volumen bei jeder Dimension und meinen also sowas wie Länge bzw. Fläche bzw. Volumen, je nach d) mit der Gleichmäßigkeit einer Punktmenge M_n zu tun?

Gegeben seien also

$$M_n = \{x_1, x_2, \ldots, x_n\} \subset [0,1]^d$$

und ein Quader $[x, z] \subset [0,1]^d$. Wir haben n Punkte und fragen nach der Anzahl von Punkten, die im Quader $[x, z]$ liegen. Diese Anzahl bezeichnen wir mit $\#(M_n, [x, z])$, wir interessieren uns für das Verhältnis

$$\frac{\#(M_n, [x, z])}{n}$$

der Punkte, die in $[x, z]$ liegen, zur Gesamtzahl n der Punkte. Idealerweise sollte dieses Verhältnis ähnlich dem Verhältnis der Volumina sein, d. h. $V([x, z])$ zu 1. Die Größe

$$\operatorname{dis}(M_n, [x, z]) = \left| V([x, z]) - \frac{\#(M_n, [x, z])}{n} \right|$$

misst also die Abweichung oder Diskrepanz der Punkte in Bezug auf den Quader $[x, z]$.

Wir wollen wissen, wie gut die Menge M_n überhaupt zur Volumenmessung für Quader taugt, nicht nur für einen einzigen Quader. Daher definieren wir die Diskrepanz $\operatorname{dis}(M_n)$ als das Maximum über alle $\operatorname{dis}(M_n, [x, z])$,

$$\operatorname{dis}(M_n) = \max_{x,z} \operatorname{dis}(M_n, [x, z]) = \max_{x,z} \left| V([x, z]) - \frac{\#(M_n, [x, z])}{n} \right| .$$

Das Problem

Jetzt sind wir in der Lage, unsere Eingangsfrage präzise zu stellen: Gegeben seien natürliche Zahlen n und d. Wie muss man die Punkte

$$M_n = \{x_1, x_2, \ldots, x_n\} \subset [0, 1]^d$$

wählen, damit $\operatorname{dis}(M_n)$ möglichst klein wird? Sind (im Fall $n = k^d$) äquidistante Gitter optimal oder gibt es bessere Punktanordnungen?

Äquidistante Gitter

Wir wollen äquidistante Gitter untersuchen und definieren

$$G_k^d = \left\{ x = (x^1, x^2, \ldots, x^d) \;\middle|\; x^i \in \left\{ \frac{1}{2k}, \frac{3}{2k}, \ldots, \frac{2k-1}{2k} \right\},\ i = 1, \ldots, d \right\}.$$

Die Anzahl der Punkte ist dann $n = k^d$.

Satz 1. *Es gilt*

$$\operatorname{dis}(G_k^d) = 1 - \left(\frac{k-1}{k} \right)^d .$$

Beweis. Wir werden diese Aussage nicht beweisen, sondern nur die Ungleichung

$$\operatorname{dis}(G_k^d) \geq 1 - \left(\frac{k-1}{k} \right)^d .$$

Dazu genügt es, den Würfel $[x,z]$ zu betrachten mit

$$x = \left(\frac{1}{2k}, \frac{1}{2k}, \ldots, \frac{1}{2k}\right) \quad \text{und} \quad z = \left(\frac{2k-1}{2k}, \frac{2k-1}{2k}, \ldots, \frac{2k-1}{2k}\right).$$

Dieser Würfel enthält alle $n = k^d$ Punkte und hat das Volumen $\left(\frac{k-1}{k}\right)^d$. Daraus folgt die angegebene Ungleichung. □

Speziell betrachten wir den Fall $k = d$, also $n = d^d$, und erhalten

$$\operatorname{dis}(G_2^2) = 0{,}75, \qquad \operatorname{dis}(G_3^3) \approx 0{,}70370, \qquad \operatorname{dis}(G_4^4) \approx 0{,}68359.$$

Man kann auch $k = d$ sehr groß wählen und erhält zum Beispiel

$$\operatorname{dis}(G_{100}^{100}) \approx 0{,}63397.$$

Mit Hilfe der Zahl e kann man den Grenzwert für $d \to \infty$ betrachten und bekommt

$$\lim_{d\to\infty} \operatorname{dis}(G_d^d) = 1 - \frac{1}{e} \approx 0{,}63212.$$

Um es mal in Worten klar auszusprechen: Wenn man äquidistante Gitter benutzt, so braucht man in der Dimension d ungefähr $n = d^d$ Punkte, um die (sehr bescheidene) Genauigkeit (Diskrepanz) $0{,}63$ zu erreichen.

Es gilt $101^{101}/100^{100} \approx 273$. Beim Übergang von $d = 100$ zu $d = 101$ erhöht sich also der Aufwand ungefähr um den Faktor 273, wenn man wieder die Diskrepanz $0{,}63$ erreichen will – zumindest dann, wenn man äquidistante Gitter benutzt.

Sind Gitter optimal?

Äquidistante Gitter sind natürlich leicht zu definieren. Sie sind auch relativ leicht zu analysieren und man könnte vermuten, dass sie optimal sind oder wenigstens „ungefähr optimal". Was immer das auch genau heißen mag. Seltsamerweise sind diese Gitter sehr schlecht, es gibt sehr viel bessere Punktmengen. Um diese Behauptung zu belegen, geben wir zwei Ergebnisse an:

- Das erste Ergebnis geht von einer festen Dimension d aus, und es werden gewisse Punktmengen für $n \to \infty$ mit kleiner Diskrepanz definiert.

- Im zweiten Ergebnis betrachten wir eine feste Diskrepanz $\varepsilon < 1$, die wir erzielen wollen. Wir fragen nach der Anzahl der Punkte, die man (abhängig von der Dimension) braucht.

Die Punktmengen von Hammersley (1960)

Wir betrachten nochmal kurz die äquidistanten Gitter G_k^d. Für festes d und große k gilt

$$\left(\frac{k-1}{k}\right)^d \approx 1 - \frac{d}{k}$$

und damit $\operatorname{dis}(G_k^d) \approx d/k$. Wegen $n = k^d$ kann man auch schreiben

$$\operatorname{dis}(G_k^d) \approx d \cdot \left(\frac{1}{n}\right)^{1/d}.$$

Den Exponenten (hier $1/d$) bei $1/n$ bezeichnet man mit *Konvergenzordnung*. Diese Konvergenzordnung wird also bei Gittern immer schlechter (kleiner), wenn d größer wird.

Jetzt betrachten wir Punktmengen, die HAMMERSLEY im Jahr 1960 untersucht hat. Sei $b \geq 2$ eine natürliche Zahl und

$$i = \sum_{j=0}^{\infty} d_j(i)\, b^j, \quad d_j(i) \in \{0, 1, \ldots, b-1\}$$

die Entwicklung von $i \in \mathbb{N}_0$ bezüglich der Basis b. Dann ist die VAN-DER-CORPUT-*Folge* $\big(\gamma_b(i)\big)_{i \in \mathbb{N}_0}$ definiert durch

$$\gamma_b(i) = \sum_{j=0}^{\infty} d_j(i)\, b^{-j-1}.$$

Es gilt also z. B. $\gamma_2(0) = 0$, $\gamma_2(1) = 1/2$, $\gamma_2(2) = 1/4$, $\gamma_2(3) = 3/4$, $\gamma_2(4) = 1/8$ und so weiter. Die Zahl 35 (im Dezimalsystem) schreibt man im Dualsystem als 100011 (wegen $35 = 32 + 2 + 1$), daher gilt

$$\gamma_2(35) = 1/2 + 1/4 + 1/64.$$

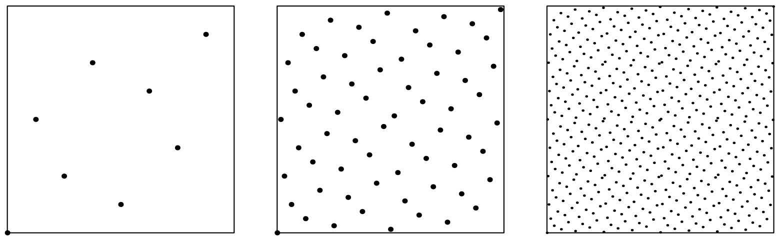
Noch einfacher ist der Fall $b = 10$. Dann gilt beispielsweise

$$\gamma_{10}(2579) = 0{,}9752\,,$$

und man spricht von der *Spiegelung am Dezimalkomma*. Seien $b_1, \ldots, b_{d-1}$ die ersten $d-1$ Primzahlen, und sei $n \in \mathbb{N}$ gegeben. Dann ist die d-dimensionale HAMMERSLEY-*Folge* der Länge n gegeben durch

$$x_i = \big(i/n, \gamma_{b_1}(i), \ldots, \gamma_{b_{d-1}}(i)\big), \qquad i = 0, \ldots, n-1.$$

Wir bezeichnen die Menge $\{x_0, \ldots, x_{n-1}\}$ mit H_n^d. In Abbildung 2 sind 2-dimensionale HAMMERSLEY-Folgen mit verschiedenen Längen zu sehen. HAMMERSLEY hat Folgendes bewiesen.



$n = 8$ $n = 64$ $n = 512$

Abbildung 2. HAMMERSLEY-Folgen für $d = 2$

Satz 2. *Es existiert eine Konstante $c_d > 0$, so dass*

$$\operatorname{dis}(H_n^d) \le c_d \cdot n^{-1} \cdot (\log n)^{d-1}$$

für alle $n > 1$ gilt.

Man erreicht also mit den HAMMERSLEY-Punkten bis auf logarithmische (d. h. „kleine“) Faktoren die Konvergenzordnung 1, unabhängig von der Dimension.

Bei festem d und großen n sind also die HAMMERSLEY-Punkte viel besser als die Punkte eines äquidistanten Gitters. Leider kennt man nur ziemlich grobe Abschätzungen für die Diskrepanz der Mengen H_n^d. So ist zum Beispiel nicht bekannt (auch nicht näherungsweise), wie groß man n wählen muss, damit man in Abhängigkeit von d eine Diskrepanz ε (etwa $\varepsilon = 1/3$) erreicht.

Wie muss n mit d wachsen?

Gibt es $n = 3^d$ Punkte $x_1, x_2, \ldots, x_n \in [0,1]^d$ mit

$$\operatorname{dis}(\{x_1, x_2, \ldots, x_n\}) \le 1/3\,?$$

Dies war bis zum Jahr 2000 noch unbekannt. Wir wissen schon, dass man viel mehr Punkte benötigt, sofern man äquidistante Gitter wählt. Unklar ist auch, wie schon gesagt, wie groß man n im Fall der HAMMERSLEY-Punkte H_n^d wählen muss. Im Jahr 2001 wurde folgendes Ergebnis veröffentlicht, siehe [1].

Satz 3. *Es gibt eine Konstante $c > 0$, so dass für beliebige $d, n \in \mathbb{N}$ eine Menge*

$$M_n = \{x_1, x_2, \ldots, x_n\} \subset [0,1]^d$$

existiert, für die

$$\operatorname{dis}(M_n) \leq c \cdot \sqrt{\frac{d}{n}} \,.$$

Zu diesem Satz geben wir ein paar Kommentare:

- Wir haben gesehen: Wenn man Gitter nimmt, dann erhöht sich (für die erstrebte Diskrepanz 0,63) der Aufwand n beim Übergang von $d = 100$ zu $d = 101$ um den Faktor 273. Wenn man dagegen optimale Punktmengen nimmt, so erhöht sich der Aufwand beim Übergang von $d = 100$ zu $d = 200$ nur etwa um den Faktor 2. Beim Übergang von $d = 100$ zu $d = 200$ mit äquidistanten Gittern müsste man den Aufwand um den Faktor

  $$\frac{200^{200}}{100^{100}} \approx 1,6 \cdot 10^{260}$$

  vergrößern. Dies zeigt, dass es viel bessere Punktmengen gibt als äquidistante Gitter.

- Satz 3 sagt aus, dass die Diskrepanz von geeigneten M_n ziemlich klein ist. Es gibt auch Abschätzungen in der anderen Richtung: In [2] wird gezeigt, dass ein $c > 0$ existiert, so dass

  $$\operatorname{dis}(M_n) \geq c \cdot \min\left(\frac{d}{n}, 1\right)$$

  für beliebige $d, n \in \mathbb{N}$ und $M_n = \{x_1, \ldots, x_n\} \subset [0,1]^d$ gilt.

- Leider ist Satz 3 nur ein Existenzsatz. Wir haben keine Ahnung, wie entsprechende Punktmengen $M_n \subset [0,1]^d$ ausschauen. Der Beweis benutzt Argumente aus der Wahrscheinlichkeitstheorie. Grob gesprochen gilt: Wählt man zufällig irgendeine Menge

  $$M_n = \{x_1, \ldots, x_n\} \subset [0,1]^d$$

  aus, so ist die Wahrscheinlichkeit groß, dass man eine Menge mit $\operatorname{dis}(M_n) \leq c \cdot \sqrt{d/n}$ erwischt.

Offene Probleme

Aufgrund der letzten Bemerkung stellen sich Fragen, die bisher niemand beantworten kann: Gibt es einen konstruktiven Beweis von Satz 3 in dem Sinn, dass man geeignete Punktmengen M_n konkret beschreiben kann?

Genügt es vielleicht sogar, die HAMMERSLEY-Punktmengen H_n^d zu nehmen? Gilt also

$$\mathrm{dis}\,(H_n^d) \leq c \cdot \sqrt{\frac{d}{n}}$$

oder eine ähnlich gute Abschätzung?

Schlussbemerkung

Ist das beschriebene Diskrepanz-Problem interessant oder nur eine x-beliebige Rätselaufgabe? Diese Frage ist schwierig und verschiedene Mathematiker werden sie wohl verschieden beantworten. Probleme gelten z. B. dann als interessant, wenn sie in verschiedenen mathematischen Teilgebieten immer wieder auftauchen.

Das Problem der Diskrepanz von Punktmengen hat sicher eine elementargeometrische Bedeutung. Es taucht aber auch auf in der Numerischen Mathematik (wie gut sind Quadraturformeln?), in der Zahlentheorie (Theorie der Gleichverteilung), in der Kombinatorik (kombinatorische Probleme erhält man, wenn man gute Punktmengen konstruieren möchte; bei Beweisen wird die sog. VC-Dimension benutzt), in der Statistik und in der Wahrscheinlichkeitstheorie (BROWNsche Bewegung). Daher würde ich das Problem der Diskrepanz mit Sicherheit zu den interessanten Problemen zählen.

Literatur

[1] S. HEINRICH, E. NOVAK, G. W. WASILKOWSKI, H. WOŹNIAKOWSKI: *The inverse of the star-discrepancy depends linearly on the dimension.* Acta Arithmetica **96**, (2001), 279–302.

[2] A. HINRICHS: *Covering numbers, Vapnik-Červonenkis classes and bounds for the star-discrepancy.* (2003), erscheint im Journal of Complexity.

Prof. Erich Novak, Jena