Algorithmische Mathematik II

Dozent
Professor Dr. Patrik Ferrari

Mitschrift
Maximilian Kessler

Version
2. Juni 2021 10:53

Zusammenfassung

Bei folgenden Vorlesungsnotizen handelt es sich um (inoffizielle) Mitschriften zur Vorlesung 'Algorithmische Mathematik II', die im Sommersemester 2021 an der Universität Bonn gehalten wird. Ich garantiere weder für Korrektheit noch Vollständigkeit dieser Notizen, und bin dankbar für jegliche Art von Korrektur, sowohl inhaltlich, als auch Tippfehler.

Bemerkungen, die nicht zum eigentlichen Vorlesungsinhalt gehören, wurden mit einem * gekennzeichnet. Sie werden nach eigenem Ermessen hinzugefügt, um weitere Details oder evtl. mündliche Anmerkungen beizufügen.

Manche Umgebungen sind mit einem † versehen. Das ist dann der Fall, wenn ihr Inhalt so, oder zumindest in sehr ähnlicher Form, in der Vorlesung vorkam (unter Umständen auch mündlich), ich aber die Umgebung der Aussage geändert habe. Insbesondere also, wenn ich aus Aussagen, die einfach erwähnt werden, ein **Lemma**† mache, um sie hervorzuheben.

Weitere Informationen finden sich bei GitHub oder auf der Vorlesungshomepage

Inhaltsverzeichnis

Übersicht der Vorlesungen

Vorlesung 1
Mo 31 Mai 2021 10:16

1 Simulationsverfahren und Monte Carlo Methode

1.1 Simulation von Zufallsvariablen

Definition 1.1. (a) U ist eine reellwertige Zufallsvariable, falls

$$\{\omega \in \Omega \mid U(\omega) \leqslant x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}.$$

(b) $U : \Omega \to [0,1]$ ist gleichverteilt auf $[0,1]$, falls

$$\mathbb{P}(U \leqslant x) = x \qquad \forall x \in [0,1].$$

Eine Familie von rellwertigen Zufallsvariablen $\{U_k\}_{k\in I}$ heißen unabhängig, falls $\forall x_1 x_2, \ldots, \in \mathbb{R} : \{U_k \leqslant x_k\}_{k\in I}$ unabhängig sind.

Notation. Wir schreiben hierfür $U \sim U([0,1])$ oder auch $U \sim \mathrm{Unif}([0,1])$

- Sei $\mathcal{S} = \{a_1, a_2, \ldots\}$ ein diskreter Zustandsraum (d.h. abzählbar) und μ eine Wahrscheinlichkeitsverteilung auf $\mathcal{S}$ mit $p_k = \mu(a_k)$, also $\sum_{k=1}^{\infty} p_k = 1$. Setze nun $s_0 := 0$, $s_n := \sum_{k=1}^{n} p_k$ für $n \geqslant 1$.

Lemma 1.2. Sei $U \sim \mathrm{Unif}([0,1])$ uniform verteilt und $X(\omega) := a_n$ falls $U(\omega) \in (s_{n-1}, s_n]$. Dann ist $X \sim \mu$.

Beweis. Für $n \geqslant 1$ ist

$$\begin{aligned}
\mathbb{P}(X = a_n) &= \mathbb{P}(s_{n-1} < U(\omega) \leqslant s_n) \\
&= \mathbb{P}(U(\omega \leqslant s_n)) - \mathbb{P}(U(\omega \leqslant s_{n-1})) \\
&= s_n - s_{n-1} \\
&= p_n
\end{aligned}$$

□

Mündliche Anmerkung. Wir können mit diesem Lemma theoretisch schon für beliebige diskrete Verteilungen einen enstprechenden Algorithmus schreiben, das ist aber unter Umständen äußerst unpraktikabel, wenn $\mathcal{S}$ sehr groß ist oder μ keine tolle Form hat, wie wir gleich sehen werden:

Algorithmus 1.3 : Simulation von μ

Eingabe : $p_1, p_2, \dots$
Ausgabe : Eine Zufallsvariable X mit $X \sim \mu$

```
n = 1
s = p_1
erzeuge u ~ Unif([0,1])
while n > s do
    n = n + 1
    s = s + p_n
return a_n
```

Frage. Wie viele Schritte brauchen wir?

$$\mathbb{E}(\#\text{'Schritte'}) = \sum_{n \geqslant 1} n \cdot p_n.$$

Mündliche Anmerkung. Abgesehen von potentiell sehr langer Rechendauer, kommen hier auch noch numerische Ungenauigkeiten / Probleme vor.

1.2 Acceptance-Rejection-Verfahren

- Sei μ eine Wahrscheinlichkeitsverteilung mit Massenfunktion p und ν eine Wahrscheinlichkeitsverteilung mit Massenfunktion q, wobei wir ν simulieren wollen.
- Nehmen wir an, dass $\exists c \in [1, \infty)$, sodass

$$\begin{aligned} p(x) &\leqslant c \cdot q(x) \qquad \forall x \in \mathcal{S} \\ \Leftrightarrow \quad 0 &\leqslant \frac{p(x)}{c \cdot q(x)} \qquad \forall x \in \mathcal{S} \end{aligned}$$

Algorithmus 1.4 :

Eingabe : $p(x), q(x)$ für $x \in \mathcal{S}$, Konstante c
Ausgabe : Zufallsvariable X mit $X \sim \mu$

repeat:
erzeuge $x \sim \nu$
erzeuge $u \sim \mathrm{Unif}([0,1])$
until ($\frac{p(x)}{c \cdot q(x)} \geqslant u$)
return x.

Wir nehmen also den Vorschlag $X = x$ mit der Wahrscheinlichkeit $\frac{p(x)}{c \cdot q(x)}$ an.

Mündliche Anmerkung. Die Erzeugung von $x \sim \nu$ ist im Algorithmus nicht genauer beschrieben. Der Algorithmus ist also nur dann (sinnvoll) anwendbar, wenn wir $x \sim \nu$ mit einem anderen Algorithmus sinnvoll schnell erzeugen können.
Wir gehen ebenfalls davon aus, dass wir $u \sim \mathrm{Unif}([0,1])$ mit einem Pseudo-Zufallsgenerator mit dem Computer erzeugen können.

Seien $X_1, X_2, \ldots \sim \nu$ die Vorschläge, die wir erhalten, und seien $U_1, U_2, \ldots \sim \mathrm{Unif}([0,1])$ die uniformen Zufallsvariablen. Sei

$$T := \min\left\{n \geqslant 1 \mid \frac{p(X_n)}{c \cdot q(X_n)} \geqslant U_n\right\}.$$

der erste Zeitpunkt, bei dem die entsprechende Ungleichung gilt, d.h. wenn wir unsere Schleife abbrechen. Also ist $X_T(\omega) := X_{T(\omega)}(\omega)$ der Output unseres Algorithmus.

Satz 1.5. a $X_T \sim \mu$, dh. der Algorithmus ist korrekt.
b $T - 1 \sim \mathrm{Geo}(\frac{1}{c})$, also $\mathbb{E}(T) = c$.

Beweis. Die Ereignisse

$$A_n = \left\{\frac{p(X_n)}{c \cdot q(X_n)} \geqslant U_n\right\} \qquad n = 1, 2, \ldots.$$

sind unabhängig, also ist

$$\begin{aligned}\mathbb{P}(T = n) &= \mathbb{P}(A_1^c \cap \ldots \cap A_{n-1}^c \cap A_n) \\ &\overset{\text{unabhängig}}{=} \mathbb{P}(A_1) \cdot (1 - \mathbb{P}(A_1))^{n-1}\end{aligned}$$

Nun ist

$$\begin{aligned}\mathbb{P}(A_1) &= \sum_{a \in \mathcal{S}} \mathbb{P}(\left\{U_1 \leqslant \frac{p(a)}{c \cdot q(a)}\right\} \cap \{X_1 = a\}) \\ &= \sum_{a \in \mathcal{S}} \frac{p(a)}{c \cdot q(a)} \cdot q(a) \\ &= \frac{1}{c}\end{aligned}$$

Also folgt bereits Teil b.

Es ist

$$\begin{aligned}\mathbb{P}(X_T) = a) &= \sum_{n \geqslant 1} \mathbb{P}(X_T = a, T = n) \\ &= \sum_{n \geqslant 1} \mathbb{P}(A_1^c \cap \ldots \cap A_n \cap \{X_n = a\}) \\ &= \sum_{n \geqslant 1} \left(1 - \frac{1}{c}\right)^{n-1} \frac{p(a)}{c} \\ &= p(a)\end{aligned}$$

□

Beispiel. Sei $Y \sim \nu = \mathrm{Geo}(q)$ geometrische verteilt.

- In Kapitel 1.7.4 haben wir gesehen, dass wir $u \sim \mathrm{Unif}([0,1])$ erzeugen, und dann

  $$Y = \left\lfloor \frac{\ln(1-u)}{\ln(q)} \right\rfloor.$$

  setzen können, um Y zu simulieren, und dann eben $\mathbb{P}(Y = k) = (1-q)q^k$ für $k = 0, 1, \ldots$.

- Sei $X \sim \mu = \mathrm{Poi}(\lambda)$, d.h.

$$|(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

- Sei

$$\mathcal{R}(k) := \frac{e^{-\lambda}\lambda^k}{k!} \cdot \frac{1}{(1-q)q^k}.$$

- Wähle $q = \frac{\lambda}{\lambda+1}$, denn dann ist $\lambda = \frac{q}{1-q}$, und die beiden Verteilungen erhalten den gleichen Mittelwert, also ist

$$\mathcal{R}(k) = \frac{e^{-\lambda}(\lambda+1)^{k+1}}{k!}.$$

Wegen

$$\mathcal{R}(k+1) - \mathcal{R}(k) = \mathcal{R}(k)\left(\frac{1+\lambda}{1+k} - 1\right) \begin{cases} > 0 & k < \lambda \\ < 0 & k > \lambda \end{cases}.$$

Veranschaulichung des maximalen k

erhalten wir

$$\max_{k \geqslant 0} \mathcal{R}(k) \leqslant \frac{e^{-\lambda}(1+\lambda)^{\lfloor\lambda\rfloor+1}}{\lfloor\lambda\rfloor!} =: c.$$

Wir können nun also $\mathrm{Poi}(\lambda)$ mit Algorithmus 2 simulieren, weil wir ein entsprechendes c gefunden haben.

1.3 Monte-Carlo-Verfahren

Ziel. es ist

$$\mathbb{P}_\mu(f) := \sum_{x \in \mathcal{S}} f(x)\mu(x). \tag{1}$$

für $|S| \gg 1$ zu berechnen.
Beweisstrategie. Man simuliert eine Folge von Zufallsvariablen oder eine Markovkette mit $\mu = \mu P$.

Beispiel. Sei $\mathcal{S} = \{-1, +1\}^{100}$, also $|\mathcal{S}| = 2^{100} \approx 10^{30}$. Das könnte z.B. dann der Fall sein, wenn wir bei einem 10×10-Gitter in jedem Punkt eine Markierung haben oder nicht.

Mündliche Anmerkung. Wenn wir nun den Erwartungswert einer Funktion über diesem Zustandsraum berechnen wollen, so ergeben sich Probleme, da wir einerseits absurd viele Summanden haben, aber auch numerische, denn μ ist für jeden einzelnen Zustand verschwindend klein. Unsere Formel (1) stimmt also, wir haben aber keine Chance, diese auszurechnen.

- Sei $\mathcal{S}$ abzählbar und μ eine Wahrscheinlichkeitsverteilung auf $\mathcal{S}$.
- Sei f eine Funktion $\mathcal{S} \to \mathbb{R}$, sodass

$$\mathbb{E}_\mu(f^2) = \sum_{x \in \mathcal{S}} (f(x))^2 \mu(x) < \infty.$$

Wir wollen z.B. $\theta := \mathbb{E}_\mu(f)$ bestimmen. Als Schätzung definieren wir

$$\theta_n := \frac{1}{n} \sum_{k=1}^{n} f(X_k).$$

wobei X_1, X_2,... Zufallsvariablen mit Verteilung μ sind. Falls $X_1, X_2, \ldots$ unabhängig:

Satz 1.6. $\forall \varepsilon > 0$ ist $\mathbb{P}(|\theta_n - \theta| \geqslant \varepsilon) \leqslant$

$$\leqslant \frac{\operatorname{Var}_\mu(f^2)}{n\varepsilon^2} \leqslant \frac{\mathbb{E}_\mu(f^2)}{n\varepsilon^2} \to 0.$$

Beweis. Tchebishev Ungleichung, siehe Gesetz der großen Zahlen. ☐

Beispiel. (1) Schätzung von $\theta = \int_0^1 f(x)\,\mathrm{d}x$.

- Seien $u_1, u_2, \ldots \sim \operatorname{Unif}([0,1])$, dann setze

$$\theta_n := \frac{1}{n} \sum_{k=1}^{n} f(u_k).$$

und diese ist eine Schätzung des entsprechenden Integrals mit

$$\sqrt{\mathbb{E}(|\theta_n - \theta|^2} = O\left(\frac{1}{n}\right).$$

(2) Schätzung einer Wahrscheinlichkeitsverteilung.

- Sei $\mathcal{S}$ ein diskreter Zustandsraum und $B \subseteq \mathcal{S}$. Gesucht ist $p := \mu(B)$ (unbekannt).
- Es ist $\mu(B) = \mathbb{E}_\mu(\mathbb{1}_B)$, also setze

$$p_n := \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_B(X_k).$$

wobei $X_1, X_2, \ldots \sim \mu$ (unabhängige) Zufallsvariablen sind und erhalte damit eine Näherung.

Frage. Wie gut ist diese Annäherung?

$$\mathbb{P}(|p_n - p| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^2} \operatorname{Var}(p_n) = \frac{1}{n\varepsilon^2} \operatorname{Var}(\mathbb{1}_B) = \frac{p(1-p)}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2}.$$

Gegeben ε wählen wir n also groß genug, sodass $\frac{1}{4n\varepsilon^2} \leqslant$ angegebene Fehlerschranke.

Mündliche Anmerkung. Wir können nun, wenn wir $p := \mu(B)$ bestimmen wollen, also für ein gewünschtes ε ein geeignetes n wählen, sodass wir p approximieren und mit höchstens der gewünschten Wahrscheinlichkeit um mindestens ε von der echten Wahrscheinlichkeit abweichen.

Vorlesung 2
Mi 02 Jun 2021 10:15

1.4 Gleichgewicht von Markovketten

Oft ist μ nicht explizit bekannt, z.B. wenn

$$\mu = \lim_{n \to \infty} \mu_0 P^n.$$

wobei P die Übergangsmatrix einer Markovkette ist.

Mündliche Anmerkung. Selbst wenn der obige Limes existiert und eindeutig ist (d.h. nicht von μ_0 abhängt), heißt das nicht, dass wir eine explizite Formel für μ kennen. Allerdings können wir natürlich approximieren, indem wir mit einem μ_0 starten und $\mu_0 P^n$ für ein hinreichend großes n bestimmen.

Betrachten wir eine homogene Markovkette mit Übergangsmatrix P und Anfangsverteilung μ_0.

Definition 1.7 (Stationäre Verteilung)**.** (a) Eine Wahrscheinlichkeitsverteilung μ auf $\mathcal{S}$ ist eine **stationäre Verteilung** einer Markovkette mit Übergangsmatrix P, falls $\mu = \mu P$.

(b) μ erfüllt die **Detailed-Balance-Bedingung** bezüglich P, falls

$$\mu(x)P(x,y) = \mu(y)P(y,x) \qquad \forall x, y \in \mathcal{S}.$$

Gleichgewicht zwischen x, y durch $\mu(x)P(x,y) \equiv$ Massenflus von x nach y

Satz 1.8. Falls μ die Detailed-Balance-Bedingung erfüllt, so ist μ stationär.

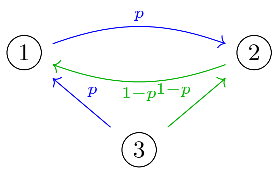
Beweis. Es ist

$$\begin{aligned}(\mu P)(x) &= \sum_{y\in\mathcal{S}} \mu(y)P(y,x)\\ &\overset{\text{Detailed Balance}}{=} \sum_{y\in\mathcal{S}} \mu(x)P(x,y)\\ &= \mu(x) \underbrace{\sum_{y\in\mathcal{S}} P(x,y)}_{=1(P\text{ stochastisch})}\\ &= \mu(x)\end{aligned}$$

□

Warnung. μ stationär $\nRightarrow$ μ erfüllt die Detailed Balance Bedingung

Beispiel. Sei $\mathcal{S} = \{1,2,3\}$ und $p \in \left(\frac{1}{2}, 1\right)$.



.

also

$$P = \begin{pmatrix} 0 & p & 1-p \\ 1-p & 0 & p \\ p & 1-p & 0 \end{pmatrix}.$$

Dann ist $\mu = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ eine stationäre Verteilung, wie man leicht prüft (Syemmtriegründe oder einfaches Nachrechnen). Allerdings ist

$$\mu(1)P(1,2) = \frac{1}{3}p \neq \frac{1}{3}(1-p) = \mu(2)P(2,1).$$

also erfüllt μ nicht die Detailed-Balance-Bedingung.

Ordentliches Diagramm

1.5 Konvergenz ins Gleichgewicht

Um Konvergenz messen zu können, brauchen wir einen Abstandsbegriff für Wahrscheinlichkeitsverteilungen. Sei hierzu

$$\mathcal{M}(\mathcal{S}) := \left\{ \mu = (\mu(x))_{x\in\mathcal{S}} \mid \mu(x) \geqslant 0 \ \forall x, \sum_{x\in\mathcal{S}} \mu(x) = 1 \right\}.$$

der Raum aller Wahrscheinlichkeitsverteilungen.

Definition 1.9. Die **totale Variatonsdistanz** zweier Wahrscheinlichkeitsverteilungen μ, ν auf $\mathcal{S}$ ist defniert durch:

$$\begin{aligned} d_{TV}(\mu,\nu) &:= \frac{1}{2}\|\mu-\nu\|_1 \\ &= \frac{1}{2}\sum_{x\in\mathcal{S}} |\mu(x)-\nu(x)| \end{aligned}$$

Bemerkung. (a) d_{TV} ist eine Metrik.
(b) $\forall \mu, \nu \in \mathcal{M}(\mathcal{S})$ ist

$$d_{TV}(\mu, \nu) \leqslant \frac{1}{2} \sum_x (\mu(x) + \nu(x)) = 1.$$

C Stichwortverzeichnis