

Angewandte Stochastik

Dr. C.J. Luchsinger

19 Miscellanea

Entsprechend dem Titel folgen ein paar lose Resultate aus diversen Gebieten. Die ersten drei Sektionen haben übrigens Anwendungen in der modernen statistischen Molekularbiologie.

19.1 Theorie der Komplexität

- * Endlich viele Teile, Mengen. Damit *theoretisch* lösbar
- * Probleme potentiell von Ordnung $N!$ (alles ausprobieren), braucht sehr lange, praktisch unmöglich

Welche Kategorien von Problemen gibt es (von der Ordnung her)?

- * **P** Polynomial lösbar. Beispiele: Multiplikation, Näherungsmethode für TSP
- * **E** Exponentiell oder überexponentiell lösbar: Man hat bewiesen, dass das Problem *nur* exponentiell (bzw. überexponentiell) lösbar ist (und nicht einfacher!). Beispiele: Spiel Roadblock oder Presburger-Arithmetik
- * **AU** Algorithmisch unlösbar: Man hat *bewiesen* (!), dass ein Problem nicht mit einem Algorithmus lösbar ist. Beispiele: Halteproblem für Turingmaschinen, Entscheidungsprobleme für die elementare Logik
- * **NP** Nichtdeterminiert und **P**olynomialeigenschaft, nachfolgend besprochen
- * **NP-vollständig** Nichtdeterminiert und **P**olynomeigenschaft, nachfolgend besprochen

Wir betrachten im Folgenden nur die Fälle E, NP und vor allem NP-vollständig. Worum geht es dabei?

Man kann sich fragen, ob TSP ein P oder ein E-Problem ist, also ob wir es in vernünftiger Zeit lösen können (P) oder nicht (E). Die Antwort ist nicht bekannt. Aber: mit vielen anderen Problemen gehört es zu einer Klasse von Problemen, die - bei aller Verschiedenheit im Detail - doch einige wichtige Merkmale gemeinsam haben:

- * Bei allen ist ein exponentielles (oder überexponentielles) Lösungsverfahren bekannt, in der Regel die (wahrhaft *erschöpfende*) Durchmusterung aller Möglichkeiten.
- * Bei allen ist jedoch die Suche nach einem *eleganten* Algorithmus bisher vergeblich gewesen.
- * Bei allen ist es denkbar, dass man die richtige Lösung in jedem Einzelfall durch Raten, also ein **Nichtdeterminiertes** Verfahren, schneller findet.
- * Bei allen gibt es dann für die Richtigkeit der vermuteten Lösung einen *eleganten* Beweis, der vielleicht schwer zu finden, aber leicht vorzuführen ist. Seine Schrittzahl wächst wieder höchstens wie eine Potenz der Eingabeparameter, ist also wieder ein **Polynom** in N .

Wegen der letzten Eigenschaften - **Nichtdeterminiertheit** und **Polynomeigenschaft** - nennt man diese Probleme NP-Probleme. TSP ist ein NP-Problem. In der Klasse der NP-Probleme gibt es Probleme, welche wechselseitig aufeinander reduzierbar sind: Hat man einen polynomialen Algorithmus welcher eines der Probleme löst, hat man auch automatisch einen polynomialen Algorithmus für das andere Problem. Es gibt aber sogar eine absolute Eliteklasse von wechselseitig aufeinander reduzierbarer Probleme derart, dass wenn eines dieser Probleme polynomial gelöst werden könnte, dann wären *alle* NP-Probleme polynomial lösbar. Dies sind die sogenannten NP-vollständigen Probleme. TSP ist sogar ein NP-vollständiges Problem. Wenn jemand also das TSP-Problem polynomial lösen kann, dann sind alle NP-Probleme (inklusive die NP-vollständigen) polynomial lösbar. Sollte sich das TSP Problem als ein Problem aus E erweisen, so wären alle NP-vollständigen Probleme aus E (hingegen könnten dann NP-Probleme, welche nicht gerade *NP-vollständig* sind, doch noch aus P sein). Es gibt übrigens mehrere Hundert NP-vollständige Probleme (auch optimale Stundenpläne für Schulklassen in Schulen). Heute gehen Forscher in diesem Gebiet davon aus, dass die NP-vollständigen Probleme aus E sind und versuchen dies auch zu beweisen.

19.2. Simuliertes Annealing

Annealing = Abkühlung, Methode aus der Physik, Metropolis et al. (1953)

- * Problem: $f : V \rightarrow \mathbb{R}$ zu minimieren, wo $|V| < \infty$. Minimum in v^* .
- * Motivation: aus statistische Physik; Menge V ist dort die Menge alle möglichen Konfigurationen eines physikalischen Systems; $f(v)$ ist die Energie, welche das System hat, wenn es in Konfiguration v ist. $T, T > 0$ sei die Temperatur.
- * Könnte mit Gradientenmethode Minimum suchen (Problem: lokale Minima).
- * Monte Carlo Methode: wähle Punkte in V mit gleicher Wahrscheinlichkeit $1/|V|$ aus. Registriere jeweils den kleinsten Wert von $f(v)$. Vorteile: kein Startbias, keine Probleme mit lokalen Minima; dagegen findet man kaum ein Minima wenn $|V|$ gross ist.
- * Nun: Simulated Annealing = geschickte Kombination von Gradientenmethode und Monte Carlo:

Der im Folgenden beschriebene Simulated Annealing-Ansatz wird auch Metropolis-Algorithmus genannt. Dazu führen wir die Gibbs-Verteilung auf V ein:

$$\pi_T(v) := \frac{1}{K} \exp\{-f(v)/T\},$$

wobei $K = K_{f,T}$ eine normalisierende Konstante sei. Folgende interessante Eigenschaften:

- * Wenn T gross wird, strebt diese Verteilung gegen eine Gleichverteilung (**entspricht Monte Carlo**).
- * Wenn T klein ist, wird dieses Mass gleichverteilt auf allen Elementen v , wo $f(v)$ minimiert wird.

Wie soll man aber die Verteilung π_T kennen, ohne alle $f(v)$ berechnen zu müssen? Man möchte ja das Minimum von f finden, *ohne* alle $f(v)$ überprüfen zu müssen. Dazu bedient man sich der Theorie der Markov-Ketten: Wir werden Markov-Ketten in den Metropolis-Algorithmus einbauen: Die Idee ist, eine Markov Kette auf V so zu definieren, dass die stationäre Verteilung dieser Markov Kette dann genau π_T ist (siehe das kommende Theorem 19.1).

- * Ganz V ist eine Kommunikationsklasse.
- * Definieren für jedes $v \in V$ Menge $N_v \subset V$ derart, dass alle Elemente aus N_v von v in einem Schritt erreichbar sind. N steht dabei für Nachbarschaft.
- * Wir fordern noch $|N_v| = |N_w|$ und $v \in N_w$ genau dann wenn $w \in N_v$.

Jetzt definieren wir die Übergangswahrscheinlichkeiten: Für $v \notin N_w$ haben wir $p_T(w, v) = 0$ und für $v \in N_w$, $v \neq w$ haben wir

$$p_T(w, v) := \frac{\alpha \exp\{-(f(v) - f(w))^+/T\}}{|N_w|}$$

und $p_T(w, w)$ wird derart gewählt, dass dann

$$\sum_{v \in N_w} p_T(w, v) = 1.$$

α brauchen wir allenfalls, damit die Wahrscheinlichkeiten nicht auf mehr als 1 summieren. Dann gilt folgendes Theorem:

Theorem 19.1 [Metropolis] *Die oben definierte Markov-Kette $p_T(v, w)$ mit Zustandsraum V hat Gleichgewichtsverteilung π_T ,*

$$\pi_T(v) = \frac{1}{K} \exp\{-f(v)/T\}.$$

Beweis: Wir zeigen zuerst, dass lokale Gleichgewichtsbedingungen erfüllt sind:

$$\pi_T(w)p_T(w, v) = \pi_T(v)p_T(v, w). \quad (\text{GGW loc})$$

Dies ist trivialerweise richtig für $w = v$. Sonst haben wir

$$\begin{aligned} \pi_T(w)p_T(w, v) &= \frac{1}{K} \exp\{-f(w)/T\} \frac{\alpha \exp\{-(f(v) - f(w))^+/T\}}{|N_w|} \\ &= \frac{\alpha \exp\{-[(f(v) - f(w))^+ + f(w)]/T\}}{K|N_w|} \\ &= \frac{\alpha \exp\{-\max(f(v), f(w))/T\}}{K|N_w|}. \end{aligned}$$

Aus Symmetriegründen und da $|N_v| = |N_w|$ folgen die (GGW loc). Jetzt können wir einfach zeigen, dass π_T eine Gleichgewichtsverteilung ist:

$$\sum_{w \in V} \pi_T(w) p_T(w, v) = \sum_{w \in V} \pi_T(v) p_T(v, w) = \pi_T(v) \sum_{w \in V} p_T(v, w) = \pi_T(v).$$

□

Wie wird davon Gebrauch gemacht? Wir wollen eigentlich Realisationen haben aus der Verteilung π_T wo $T > 0$ eine tiefe Temperatur hat. Aber: Dazu müssten wir ja auch f überall kennen. Dies vermeiden wir mit Hilfe obiger Markov-Kette folgendermassen: Wenn wir in w sind, wählen wir aus N_w zufällig einen Kandidaten v und berechnen dann $f(v)$. Der Übergang von w nach v wird dann gemacht mit Wahrscheinlichkeit

$$p := \exp\{-(f(v) - f(w))^+/T\}.$$

Ansonsten bleibt die Markov-Kette in w . Diese Übergangswahrscheinlichkeiten sind konsistent mit denen unserer alten Markov-Kette $p_T(v, w)$. Diese Übergänge führen dazu, dass der Prozess eher in diejenigen Zustände v geht, an denen $f(v)$ klein ist (**entspricht Gradientenmethode**).

Bis jetzt war die Temperatur T konstant. Wenn T kleiner wird, dann wird die Markovkette sich noch eher in denjenigen Zuständen v aufhalten, an denen $f(v)$ klein ist (siehe früher). Man könnte also das System laufen lassen und gleichzeitig abkühlen (Annealing!). Aber was wenn man das System *zu schnell abkühlt* und es sich in einem lokalen Minimum “verfängt”? Man muss sich also Gedanken machen, wie schnell man das System abkühlen “darf”. Das folgende Theorem gibt eine Antwort darauf.

Theorem 19.2 [Extended Metropolis Algorithm; Geman und Geman] *Gegeben sei der Metropolis-Algorithmus, wobei im Schritt n die Temperatur T_n gebraucht wird. Es gelte $\lim_{n \rightarrow \infty} T_n = 0$ und $T_n \geq c/\log(n)$, wobei $c = c_f$ eine Konstante (von f abhängig). Dann gilt:*

$$\left(\lim_{n \rightarrow \infty} \mathbb{P}[X_n = v] \right)_{v \in V} = \pi_0.$$

Beweisskizze Der vollständige Beweis ist zu langwierig. Es gibt jedoch eine einfache Heuristik, mit welcher man die Rate

$$\frac{c}{\log(n)}$$

erklären kann: Mit dem Metropolis Algorithmus kann man mit positiver Wahrscheinlichkeit ein allfälliges lokales Minimum wieder verlassen. Sei $\Delta := f(v) - f(w) > 0$ der notwendige "Sprung", um dieses lokale Minima zu verlassen. Die entsprechende Wahrscheinlichkeit hierfür ist

$$e^{-\Delta/T_n} = e^{-(f(v)-f(w))^+/T_n}.$$

Die Wahrscheinlichkeit dafür, dass das lokale Minimum nie verlassen wird ist von der Ordnung

$$\prod_{n \geq 1} (1 - e^{-\Delta/T_n}) =: \prod_{n \geq 1} (1 - p_n)$$

Es gilt von der Analysis I:

$$1 - p_i \leq 1 - p_i + \frac{p_i^2}{2!} - \frac{p_i^3}{3!} + \dots = e^{-p_i}$$

und damit

$$\prod_{n=1}^N (1 - p_n) \leq \exp\left(-\sum_{n=1}^N p_i\right).$$

Von Übungsaufgabe 28 wissen wir, dass

$$\sum_{n \geq 1} p_i = \infty \Leftrightarrow \prod_{n \geq 1} (1 - p_n) = 0.$$

Wir wollen die Wahrscheinlichkeit, das lokale Minimum nie zu verlassen gleich 0 haben.

Dazu brauchen wir also

$$\sum_{n \geq 1} p_i = \sum_{n \geq 1} e^{-\Delta/T_n} = \infty.$$

Mit der Wahl

$$T_n := \frac{c}{\log(n)}$$

haben wir

$$\sum_{n \geq 1} e^{-\Delta/T_n} = \sum_{n \geq 1} \frac{1}{n^{\Delta/c}}.$$

Mit $\Delta/c \leq 1$ haben wir die gewünschte Divergenz.

□

- * Warum die Gibbs-Verteilung: gut zu rechnen, es ging; gingen auch andere, aus Physik bekannt.
- * Probleme mit Metropolis-Algorithmus: gute Nachbarschaftsstruktur von V .

19.3 Anwendung von "Simulated Annealing" auf TSP

- * Wählen V als Menge der Permutationen S_n auf $\{1, \dots, n\}$.
- * Für jede Permutation $\sigma \in S_n$ steht die Reihenfolge der Zahlen dann für die Tour, welche der Verkäufer wählen sollte. Am Schluss muss der Verkäufer wieder an den Anfang zurück. Man mache sich klar, dass es damit - für unser Problem - egal ist, wo wir beginnen!
- * Wir fügen in σ die erste Zahl am Schluss nochmals hinzu. Damit besteht also σ aus $(n + 1)$ Zahlen.
- * Energie, f ist dann die Länge dieser Tour (eine monotone Transformation davon geht genauso).

Wir brauchen jetzt für den Metropolis-Algorithmus eine *Nachbarschaftsstruktur*. Bekannt dafür sind die $k - opt$ -Strukturen für $k \geq 2$. Man bricht eine Permutation an k Stellen in somit $k + 1$ Teile. Jetzt darf man alle Teile entweder gegenseitig auswechseln oder reversieren (alle ausser den ersten und den letzten Teil). Zur Nachbarschaft gehören also alle Permutationen, welche man mit solchen Transformationen erreichen kann. Wir nennen eine Tour σ $k - opt$, wo $1 \leq k \leq n$, wenn σ die kürzeste innerhalb ihrer Nachbarschaft ist. Damit sind insbesondere alle Touren $1 - opt$ und nur die besten Touren sind $n - opt$ (es könnte mehrere geben). Man kann sehen, dass man mit Transformationen innerhalb von $2 - opt$ -Nachbarschaften sukzessive von jeder Permutation zu jeder anderen Permutation gelangen kann (damit hat die zu konstruierende Markovkette eine einzige Kommunikationsklasse (und ist aperiodisch)). Wir wählen also als Nachbarschaftsregion sinnvollerweise

$$N_\sigma := \{\tau \in S_n \mid \tau \text{ kann von } \sigma \text{ durch 2 Schnitte \& nachfolgende Reversion erreicht werden}\}.$$

Man hat dann auch trivialerweise alle anderen Eigenschaften erfüllt: $|N_\sigma| = |N_\tau|$ und $\sigma \in N_\tau$ genau dann wenn $\tau \in N_\sigma$.

Zusammenfassung: Wir haben jetzt also eine Umformulierung gefunden, welche TSP mit dem Metropolis-Algorithmus lösbar macht.

19.4 Erwartete Anzahl Rekorde

Das folgende Problem aus dem Sport hat viele Verwandte in anderen Gebieten bis in die theoretische Informatik hinein. Die Probleme sind gegenseitig aufeinander reduzierbar. Worum geht es?

Stellen Sie sich vor, N Spieler nehmen an einem Wettkampf teil. Es handelt sich um eine Disziplin, bei der bis auf die letzte Kommastelle gemessen werden kann. Keine zwei Spieler werden exakt mit dem gleichen Wert abschliessen. Im 100 Meter Lauf Männer würde man also so genau messen, dass aus drei 9.9, vielleicht ein 9.88, 9.87 und ein 9.92 werden. Setzen wir einfach stetige Zufallsgrößen voraus. Die Spieler gehen unabhängig, zufällig ausgewählt an den Start. Was kann jetzt mit der Anzahl Rekorde passieren? Der erste Teilnehmer wird sicher einen Rekord aufstellen. Der zweite nur dann, wenn er besser ist als der erste. Falls gleich zu Beginn der Beste an den Start geht, dann haben wir im ganzen Spiel nur einen Rekord. Falls die Wettkampfteilnehmer sich so organisieren würden, dass zuerst der schlechteste kommt, dann der zweitschlechteste und so weiter, so hätten wir N Rekorde - aber dies wäre ja nicht mehr zufällig ausgewählt. Wir fragen uns jetzt nach der erwarteten Anzahl Rekorde.

Sei $(Y_i)_{i=1}^N$ eine iid Folge von stetigen Zufallsgrößen (die identische Verteilung ist keine relevante Einschränkung, siehe unten). Aus Symmetriegründen haben wir für $1 \leq j \leq N$

$$P[Y_N = \max\{Y_1, \dots, Y_N\}] = P[Y_j = \max\{Y_1, \dots, Y_N\}] = \frac{1}{N}.$$

Für die weiteren Überlegungen kehren wir die Zeitachse um: Es gibt eine Zeit X_1 , wo der absolut grösste Wert stattfindet. X_1 ist die Zeit, wo der absolute Weltrekord stattfindet. Danach kommen keine Rekorde mehr. Es gibt aber Zeiten vor X_1 , bei denen auch Rekorde stattfanden (ausser $X_1 = 1$). Damit haben wir also (Zeitachse umgedreht) folgende **Zeiten der Weltrekorde**:

$$X_1 > X_2 > X_3 > X_4 > \dots > X_k = 1.$$

Des weiteren führen wir den Indikator für Weltrekord ein: $I_j = 1$ wenn Zeit j einen Weltrekord hervorbrachte und $I_j = 0$, wenn Zeit j keinen Weltrekord hervorbrachte. Die

gesamte Anzahl Rekorde ist demnach

$$R := \sum_{j=1}^N I_j.$$

Es gilt

$$P[I_j = 1] = \frac{1}{j}.$$

Damit haben wir (letztes Gleichheitszeichen für N gross)

$$E[R] = E\left[\sum_{j=1}^N I_j\right] = \sum_{j=1}^N E[I_j] = \sum_{j=1}^N P[I_j = 1] = \sum_{j=1}^N \frac{1}{j} \doteq \ln(N) + 0.57721566.$$

Zuletzt haben wir das folgende Resultat aus der Analysis I gebraucht:

$$\lim_{N \rightarrow \infty} \left[\sum_{j=1}^N \frac{1}{j} - \ln(N) \right] = 0.57721566\dots$$

Man nennt diese Zahl (0.57721566..) auch "Euler-Mascheronische Konstante". Bei 10 Spielern ist mit etwa 2-3 Rekorden zu rechnen, bei 100 Spielern mit 5-6 Rekorden, bei 1000 Spielern mit etwa 7-8 Rekorden. Konstante Faktoren mehr Teilnehmer führen also nur linear zu mehr Rekorden:

$$\log(a^k N) = k \log(a) + \log(N)$$

(z.B. $a = 10$ und $k = 1, 2, 3, 4$, etc.).

Kleine Nebenbemerkung (freiwilliger Beweis): die $(I_j)_{j=1}^N$ sind unabhängig voneinander.

19.5 Das Sankt Petersburger Paradox und die Nutzentheorie

Wir haben schon verschiedentlich darauf hingewiesen, dass Versicherungen mindestens den **erwarteten Schaden** als Prämie verlangen müssen. Man kann solche Überlegungen auch von der anderen Seite her betrachten. Das folgende Paradox ist als Sankt Petersburger Paradox bekannt:

Eine faire Münze wird so häufig geworfen, bis zum ersten Mal "Zahl" erscheint. Die Münze wird also mindestens einmal geworfen; die Verteilung ist demnach geometrisch auf den natürlichen Zahlen grösser oder gleich 1. Ein Spieler erhält den Betrag

$$X = 2^k,$$

wenn erst beim k -ten Wurf "Zahl" zum ersten Mal kommt. Wie viel sind Sie bereit zu bezahlen, um an diesem Spiel teilnehmen zu können?

Von der geometrischen Verteilung wissen wir, dass der Gewinn dieses Spiels endlich ist:

$$P[X < \infty] = \sum_{k=1}^{\infty} P[X = 2^k] = \sum_{k=1}^{\infty} 2^{-k} = \sum_{k=0}^{\infty} 2^{-k-1} = \frac{1}{2} \sum_{k=0}^{\infty} 2^{-k} = \frac{1/2}{1 - 1/2} = 1.$$

Der erwartete Gewinn hingegen ist:

$$E[X] = \sum_{k=1}^{\infty} 2^k P[X = 2^k] = \sum_{k=1}^{\infty} 1 = \infty.$$

Der faire Preis ist also eigentlich $+\infty$. Was meinen Sie dazu?