Westfälische Wilhelms-Universität Münster

Fachbereich Informatik/Mathematik

Seminararbeit

im Studiengang Mathematik
Fachrichtung: Wahrscheinlichkeitstheorie

Thema: aus dem Buch "Finite Markov Chains and Algorithmic Applications" von Olle Häggström
Kapitel 7: Markov-Ketten-Monte-Carlo
Kapitel 8: Schnelle Konvergenz von MCMC-Algorithmen am Beispiel der Random q-Colourings

von: Stefanie Schmutte

Datum des Vortrags: 17. April 2012

7 Markov-Ketten-Monte-Carlo (MCMC)

Nachdem in den letzten Kapiteln bestimmte Eigenschaften von Markov-Ketten erläutert wurden, befassen wir uns in den folgenden Kapiteln mit ihrer Anwendung. Dazu ist es hilfreich, eine Markov-Kette mit Zustandsraum $S = \{s_1, ..., s_n\}$, Anfangsverteilung μ und Übergangsmatrix $P \in R^{n \times n}$ simulieren zu können. Das funktioniert mit einer Folge von Zufallszahlen, die auf dem Intervall [0,1] gleichverteilt sind. Wir nehmen an, dass ein Computer uns eine solche Folge U_0, U_1, U_2, ... zur Verfügung stellt. Im Weiteren seien Zufallszahlen immer gleichmäßig aus [0,1] gezogen. Zur Simulation von X_0 benötigen wir eine Anfangsfunktion $\psi : [0,1] \to S$, die folgende Bedingungen erfüllt:

1. ψ ist stückweise konstant.

2. Die Länge der Intervalle mit $\psi(x) = s_i$ entspricht μ_i $\forall i = 1, ..., n$.

Wir setzen $X_0 = \psi(U_0)$.
Für alle weiteren Schritte konstruieren wir eine Update-Funktion $\phi : S \times [0,1] \to S$ mit den zwei Eigenschaften:

1. $\phi(s_i, .)$ ist stückweise konstant für jedes feste $i \in \{1, ..., n\}$.

2. Die Länge der Intervalle mit $\phi(s_i, x) = s_j$ gleicht $P_{i,j}$ für alle festen s_i, s_j mit $i, j \in \{1, ..., n\}$.

Wir definieren $X_1 = \phi(X_0, U_1),\ X_2 = \phi(X_1, U_2), ...$.
Dass diese Simulation unserer Markov-Kette entspricht, ist leicht nachzurechnen.

Die MCMC-Methoden stellen Lösungen für folgendes Problem dar: Gegeben sei eine Wahrscheinlichkeitsverteilung μ auf der Menge $S = \{s_1, ..., s_n\}$. Wie kann man gemäß der Verteilung μ ein $s \in S$ ziehen?
Es liegt nahe, eine passende Anfangsfunktion (wie oben beschrieben) zu konstruieren und diese mit einer Zufallszahl auszuwerten. Warum dieser Plan in bestimmten Fällen misslingt, wird das folgende Beispiel verdeutlichen. An ihm werden wir später die Idee der MCMC-Methoden erarbeiten.

Beispiel 7.1 Random q-Colourings

Sei G = (V,E) ein Graph, $q \geq 2$ eine natürliche Zahl und $V = \{v_1, ..., v_k\}$. Ordnet man den Ecken Zahlen aus $T = \{1, ..., q\}$ zu, wobei die Zahlen für Farben stehen sollen, und achtet man darauf, dass keine Nachbarn die gleiche Zahl haben, so erhält man ein q-Colouring. Wir definieren das Wahrscheinlichkeitsmaß $p_{G,q}$ auf T^V:

$$p_{G,q}(\xi) = \begin{cases} \frac{1}{Z_{G,q}} & wenn \quad \xi \; ein \; q - Colouring \; ist \\ 0 & sonst \end{cases}$$

wobei $Z_{G,q} = \#\{\xi \in \{1, ..., q\}^V : \xi \text{ ist ein q-Colouring}\} = \#S$.
Es liegt also eine Laplace-Verteilung auf den q-Colourings vor.

Kommen wir zurück zur Simulation von Zufallsobjekten. Zu Beginn haben wir gesagt, dass zur Simulation lediglich eine Anfangsfunktion und eine Zufallszahl benötigt wird. Im Falle der Random q-Colourings müssen wir dazu $Z_{G,q}$ kennen. Denn sei $S = \{s_1, ..., s_{Z_{G,q}}\}$, so muss die Länge der Inter-

valle, auf denen $\psi(x) = s_i$ gilt, $\frac{1}{Z_{G,q}}$ betragen ($i \in \{1, ..., Z_{G,q}\}$).
Dann ist

$$\psi(x) = \begin{cases} s_1 & wenn\ x \in [0, \frac{1}{Z_{G,q}}) \\ s_2 & wenn\ x \in [\frac{1}{Z_{G,q}}, \frac{2}{Z_{G,q}}) \\ ... & \\ s_{Z_{G,q}} & wenn\ x \in [\frac{Z_{G,q}-1}{Z_{G,q}}, \frac{Z_{G,q}}{Z_{G,q}}] \end{cases}$$

eine passende Funktion für unsere Simulation.
Doch betrachten wir folgenden Graphen G' mit $q = 9$:

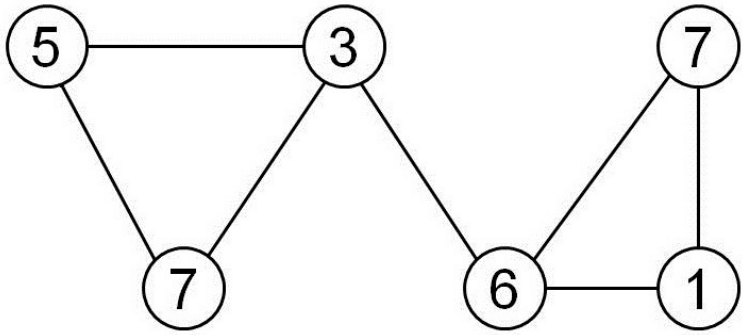

Abbildung 1: q-Colouring auf G'

Selbst bei diesem Graphen und $q = 9$ ist die Aufgabe $Z_{G',q}$ zu ermitteln nicht trivial. Hinzu kommt, dass die Random q-Colourings häufig in Bereichen der Physik Anwendung findet. Dort sind die betrachteten Graphen wesentlich größer. Meist sind sie so groß, dass selbst die Hilfe eines Computers nicht ausreicht, um $Z_{G,q}$ korrekt zu ermitteln.
Genau an dieser Stelle greift ein MCMC-Algorithmus ein. Die Idee dahinter: Wir versuchen eine irreduzible, aperiodische Markov-Kette auf S zu konstruieren, deren stationäre Verteilung π der Wahrscheinlichkeitsverteilung μ (hier $p_{G,q}$) entspricht. Gelingt dies, so sagt uns der Konvergenzsatz für Markov-Ketten aus Kapitel 5, dass, egal mit welchem $\mu^{(0)}$ wir die Kette starten, $\mu^{(n)}$ gegen $\pi = \mu$ konvergiert. Also können wir mit diesem Algorithmus ein näherungsweise μ-verteiltes Zufallsobjekt X_n ziehen, wenn wir n groß genug wählen.
Das folgende Beispiel liefert uns einen passenden MCMC-Algorithmus für die Random q-Colourings:

Beispiel 7.2

Sei G = (V,E) ein Graph und seien die Ecken durchnummeriert, sodass $V = \{v_1, ..., v_k\}$ (wie in Beispiel 7.1). Als Zustandsraum wählen wir $S = \{\xi \in \{1, ..., q\}^V : \xi$ ist ein q-Colouring$\}$, das heißt die Elemente aus $\{1, ..., q\}^V$, die kein q-Colouring sind, werden rausgelassen. Diese haben gemäß $p_{G,q}$ sowieso Wahrscheinlichkeit Null. Für unseren MCMC-Algorithmus auf S beginnen wir mit einem beliebigen q-Colouring $X_0 = \xi$ auf G. Dann benutzen wir $\forall n \in \mathbb{N}$ folgenden Übergangsmechanismus:

① Wähle $v \in V$ gemäß Laplace-Verteilung.
② Wähle $X_n(v)$ gemäß Laplace-Verteilung auf der Menge aller möglichen Zahlen, die kein Nachbar von v in X_{n-1} trägt.
③ $\forall w \in V \backslash \{v\}$ setze $X_n(w) = X_{n-1}(w)$.

Jetzt bleibt zu überprüfen, ob diese Markov-Kette $p_{G,q}$ als stationäre Verteilung hat sowie aperiodisch und irreduzibel ist.

1. $p_{G,q}$ ist stationäre Verteilung: Aus Kapitel 6 wissen wir, dass aus reversibel stationär folgt. Wir zeigen deshalb
   $p_{G,q}(\xi) \cdot P_{\xi,\xi'} = p_{G,q}(\xi') \cdot P_{\xi',\xi}\ \forall \xi, \xi' \in S$.
   Den Beweis führen wir mit Hilfe von Fallunterscheidungen:

   a) $\xi = \xi'$
   Hier ist die Gleichung trivialerweise erfüllt.

   b) Die beiden Konfigurationen ξ, ξ' unterscheiden sich in mehr als einer Ecke.
   Durch den in Beispiel 7.2 definierten Übergangsmechanismus wird bei jedem Schritt nur eine Ecke upgedatet. Daher wissen wir $P_{\xi,\xi'} = P_{\xi',\xi} = 0$ und somit ist auch hier die Gleichung erfüllt.

   c) Die beiden Konfigurationen unterscheiden sich in genau einer Ecke v.
   ② bezieht sich nur auf die Nachbarn von v und diese sind in ξ und ξ' gleich. Daraus leitet man $p_{G,q}(\xi) \cdot P_{\xi,\xi'} = \frac{1}{Z_{G,q}} \cdot \frac{1}{k} \cdot \frac{1}{l} = p_{G,q}(\xi') \cdot P_{\xi',\xi}$ ab. Dabei entspricht l der Anzahl an Zahlen die kein Nachbar von v in X_{n-1} trägt. □

2. Aperiodizität: Betrachte $\xi \in S$, dann ist die Wahrscheinlichkeit, dass beim Übergang zum nächsten Zustand die Ecke v_1 zum Updaten ausgesucht wird gleich $\frac{1}{k}$. Weiter ist die Wahrscheinlichkeit, dass dann die Zahl an dieser Ecke bestehen bleibt (also $X_{n-1}(v_1) = X_n(v_1)$) größer/ gleich $\frac{1}{q}$.
   $\Rightarrow P_{\xi,\xi} \geq$ P(in① wird v_1 gewählt)·P($X_n(v_1) = X_{n-1}(v_1)$ | in① wird v_1 gewählt)$\geq \frac{1}{k} \cdot \frac{1}{q} > 0$
   $\Rightarrow$ Jeder Zustand hat Periode 1. Das impliziert wiederum, dass die Markov-Kette aperiodisch ist. □

3. Irreduzibilität: Ob die Markov-Kette irreduzibel ist, hängt von G und q ab und muss für den speziellen Fall geprüft werden.

Damit haben wir bewiesen, dass die vorgestellte Methode ein geeigneter MCMC-Algorithmus für die Random q-Colourings ist, sofern Irreduzibilität vorliegt.

8 Schnelle Konvergenz von MCMC-Algorithmen am Beispiel der Random q-Colourings

Im vorigen Kapitel haben wir einen MCMC-Algorithmus für Random q-Colourings kennengelernt um ein $p_{G,q}$-verteiltes Zufallsobjekt zu ziehen. In Wahrheit war es aber $\mu^{(n)}$-verteilt, was der Verteilung von X_n entsprach. Für großes n liegt $\mu^{(n)}$ nahe bei $p_{G,q}$. Um diese vage Formulierung zu konkretisieren, besteht unsere Aufgabe nun darin, ein n zu ermitteln, sodass

$$d_{TV}(\mu^{(n)}, p_{G,q}) = max_{A \subseteq \{1,...,q\}^V} \mid \mu^{(n)}(A) - p_{G,q}(A) \mid \leq \epsilon.$$

Das ist für jedes $\epsilon > 0$ möglich, da wir wissen, dass $\mu^{(n)}$ gegen $p_{G,q}$ konvergiert, sofern die Markov-Kette irreduzibel ist. Für die folgende Abschätzung wandeln wir den Übergangsmechanismus an Punkt ① (siehe S.4) ein wenig ab.
Wir ziehen nicht irgendeine Ecke, sondern gehen systematisch vor: Beim ersten Schritt wählen wir die erste Ecke, beim zweiten die zweite,...,beim $k+1$-ten Schritt wieder die erste Ecke usw.. Allgemein ausgedrückt aktualisieren wir beim n-ten Schritt die (n *modulo* k)-te Ecke. Dieser neue Algorithmus liefert eine inhomogene Markov-Kette, die aperiodisch ist und $p_{G,q}$ als stationäre Verteilung hat. Die resultierende Markov-Kette ist genau dann irreduzibel, wenn auch der erste Algorithmus eine Markov-Kette mit dieser Eigenschaft liefert. In diesem Fall gilt für die inhomogene Variante ebenfalls der Konvergenzsatz.

Satz 8.1

Sei $G = (V, E)$ ein Graph, k die Anzahl an Ecken in G und jede Ecke $v \in V$ habe maximal d Nachbarn. Es gelte $q > 2d^2$, was uns die Irreduzibilität der Markov-Kette sichert. Sei außerdem $\epsilon > 0$. Dann benötigt man maximal

$$k \cdot \left(\frac{log(k) + log(\epsilon^{-1}) - log(d)}{log\left(\frac{q}{2d^2}\right)} + 1 \right) = O(k \cdot log(k))$$

Iterationen des neuen Algorithmus (der bei einem festen q-Colouring startet) um
$d_{TV}(\mu^{(n)}, p_{G,q}) \leq \epsilon$ sicher zu stellen.

Beweis von Satz 8.1

Den Beweis führen wir wie schon in Kapitel 5 mit Hilfe des Kopplungsarguments. Dazu lassen wir zwei Markov-Ketten $X = (X_0, X_1, ...)$ und $X' = (X'_0, X'_1, ...)$ mit Zustandsraum $Q = \{\xi \in \{1, ..., q\}^V : \xi \; ist \; ein \; q - Colouring\}$ parallel laufen. Der zugehörige Übergangsmechanismus entspricht dem modifizierten Algorithmus. Wir setzen $X_0 = \xi$ für beliebiges $\xi \in Q$, wohingegen X'_0 gemäß $p_{G,q}$ ermittelt wird. Also wissen wir, dass X'_n die Verteilung $p_{G,q}$ $\forall n$ hat, da $p_{G,q}$ die (einzige) stationäre Verteilung ist. $\mu^{(n)}$ sei die jeweilige Verteilung von X_n.
Für das Updaten der Ecken, sprich Punkt ② unseres Mechanismus, benötigen wir eine Folge R_1, $R_2, ...$ von Zufallspermutationen der Menge $\{1, ..., q\}$. Jede Permutation werde gleichmäßig aus den q! Permutationen gezogen. Wird die Ecke v im n-ten Schritt aktualisiert, so erhält sie in X_n und X'_n die

erste Zahl der Permutation R_n, die keiner seiner Nachbarn in X_{n-1} bzw. X'_{n-1} trägt.
Formal bedeutet das:
Sei $R_n = (R_n^1, ..., R_n^q)$ und v werde im n-ten Schritt upgedatet, dann setze $X_n(v) = R_n^i$ mit
$i = min\{j : X_{n-1}(w) \neq R_n^j \forall$ Nachbarn w von $v\}$.
(Für X' ist die Definition identisch.)
Beide Markovketten werden folglich mit denselben Permutationen upgedatet. Die Ketten entwickeln sich nicht unabhängig, was die Kernidee unserer Kopplung darstellt. Es ist leicht zu verifizieren, dass wir durch dieses Vorgehen den Übergangsmechanismus des neuen MCMC-Algorithmus beibehalten.
Im Weiteren werden wir sehen, dass $d_{TV}(\mu^{(n)}, p_{G,q})$ nahe bei 0 ist (bzw. $< \epsilon$), wenn $P(X_n \neq X'_n)$ auch genügend klein ist.
Zuerst müssen wir dazu eine obere Schranke für $P(X_n \neq X'_n)$ in Abhängigkeit von n herleiten. Dafür schätzen wir $P(X_n(v) \neq X'_n(v))$ für beliebiges $v \in V$ ab.
Sei zunächst $n \leq k$. Dann befinden wir uns im ersten Durchgang des zyklischen Algorithmus. Wir sagen, dass ein Update der Ecke v im n-ten Schritt fehlschlägt, wenn der Fall $X_n(v) \neq X'_n(v)$ eintritt. Wir definieren B_2, B_1 und B_0 wie folgt:

$B_2 = \{s \in \{1, ..., q\} : s$ ist eine Farbe, die in der Nachbarschaft von v sowohl in X_{n-1} als auch in X'_{n-1} vorkommt$\}$
$b_2 = \#B_2$

$B_1 = \{s \in \{1, ..., q\} : s$ ist eine Farbe, die in der Nachbarschaft von v entweder in X_{n-1} oder in X'_{n-1} vorkommt$\}$
$b_1 = \#B_1$

$B_0 = \{s \in \{1, ..., q\} :$ s ist eine Farbe, die in der Nachbarschaft von v weder in X_{n-1} noch in X'_{n-1} vorkommt$\}$
$b_0 = \#B_0$

$$\Rightarrow b_2 + b_1 + b_0 = q \tag{1}$$

Steht bei der Permutation R_n ein Element aus B_2 an erster Stelle, so gehen wir über zur nächsten Stelle, denn diese Farbe ist für v weder in X_n noch in X'_n zulässig. Kommen wir auf diese Weise zuerst zu einem Element aus B_1, bevor ein Element aus B_0 an die Reihe kommt, so ist das Update fehlgeschlagen.

$$\Rightarrow P(fehlgeschlagenes\ Update) = \frac{b_1}{b_1 + b_0} \tag{2}$$

Es gilt die Abschätzung

$$b_1 \leq 2d - 2b_2, \tag{3}$$

denn v hat in X und X' zusammen maximal $2d$ Nachbarn, von denen mindestens $2b_2$ Farben aus B_2 tragen.
Eingesetzt in (2) folgt daraus:

$$P(fehlgeschlagenes\ Update) = \frac{b_1}{b_1+b_0} \overset{(1)}{=} \frac{b_1}{q-b_2} \overset{(3)}{\leq} \frac{2d-2b_2}{q-b_2} \leq \frac{2d-b_2}{q-b_2} = \frac{2d\left(1-\frac{b_2}{2d}\right)}{q\left(1-\frac{b_2}{q}\right)} \leq \frac{2d}{q},$$

wobei die letzte Ungleichung aus der Annahme $q > 2d^2 > 2d$ in Satz 8.1 resultiert. Nach k Schritten gilt dann für jede Ecke $v \in V$

$$P(X_k(v) \neq X'_k(v)) \leq \frac{2d}{q}. \tag{4}$$

Nun betrachten wir den zweiten Durchgang, also $k < n \leq 2k$. Damit ein Update von v fehlschlagen kann, müssen die Konfigurationen der Nachbarn in X und X' an mindestens einer Stelle voneinander abweichen. Für jeden der maximal d Nachbarn von v gilt (4)

$$P(X_{n-1}(w) \neq X'_{n-1}(w)) \leq \tfrac{2d}{q}$$

und somit

$$P(\exists\ Nachbar\ w\ von\ v\ mit X_{n-1}(w) \neq X'_{n-1}(w)) = P(Abweichung) \leq \tfrac{2d^2}{q}.$$

Gegeben, dass es eine Abweichung gibt, ist die Wahrscheinlichkeit für ein fehlschlagendes Update $\leq \frac{2d}{q}$, gemäß den Überlegungen zum ersten Durchgang. Somit erhalten wir:

$$P(fehlgeschlagenes\,Update) = P(Abweichung) \cdot P(fehlgeschlagenes\,Update \mid Abweichung) \leq \frac{2d^2}{q} \cdot \frac{2d}{q}.$$

Am Ende des zweiten Durchgangs kann man $\forall v \in V$

$$P(X_{2k}(v) \neq X'_{2k}(v)) \leq \tfrac{2d^2}{q} \cdot \tfrac{2d}{q}$$

feststellen.
Übertragen wir dieses Vorgehen auf den dritten Durchgang, so gilt:

$$P(X_{3k}(v) \neq X'_{3k}(v)) \leq \left(\tfrac{2d^2}{q} \cdot \tfrac{2d}{q}\right) \cdot d \cdot \tfrac{2d}{q} = \left(\tfrac{2d^2}{q}\right)^2 \cdot \tfrac{2d}{q},$$

sodass wir induktiv zu folgender Abschätzung kommen:

$$P(X_{mk}(v) \neq X'_{mk}(v)) \leq \left(\frac{2d^2}{q}\right)^{m-1} \cdot \frac{2d}{q}. \tag{5}$$

Damit ist die Analyse eingeschränkt auf eine beliebige Ecke beendet und wir sind in der Lage $P(X_{mk} \neq X'_{mk})$ abzuschätzen:

$$\begin{aligned} P(X_{mk} \neq X'_{mk}) &\leq \sum_{v \in V} P(X_{mk}(v) \neq X'_{mk}(v)) \\ &\overset{(5)}{\leq} k \cdot \left(\frac{2d^2}{q}\right)^{m-1} \cdot \frac{2d}{q} \\ &= \frac{k}{d} \cdot \left(\frac{2d^2}{q}\right)^{m}. \end{aligned}$$

Kommen wir zurück zu unserem eigentlichen Ziel: $d_{TV}(\mu^{(n)}, p_{G,q}) \leq \epsilon$. In Kapitel 5 haben wir gelernt:

$$\begin{aligned} d_{TV}(\mu^{(mk)}, p_{G,q}) &= max_{A \subseteq \{1,\ldots,q\}^V} \mid \mu^{(mk)}(A) - p_{G,q}(A) \mid \\ &= max_{A \subseteq \{1,\ldots,q\}^V} \mid P(X_{mk} \in A) - P(X'_{mk} \in A) \mid \end{aligned}$$

Für beliebiges $A \subseteq \{1, ..., q\}^V$ gilt dann:

$$\begin{aligned} P(X_{mk} \in A) - P(X'_{mk} \in A) &= P(X_{mk} \in A, X'_{mk} \in A) + P(X_{mk} \in A, X'_{mk} \notin A) \\ &\quad - (P(X'_{mk} \in A, X_{mk} \in A) + P(X'_{mk} \in A, X'_{mk} \notin A)) \\ &= P(X_{mk} \in A, X'_{mk} \notin A) - P(X'_{mk} \in A, X_{mk} \notin A) \\ &\leq P(X_{mk} \in A, X'_{mk} \notin A) \\ &\leq P(X_{mk} \neq X'_{mk}) \\ &\leq \frac{k}{d} \cdot \left(\frac{2d^2}{q}\right)^m. \end{aligned}$$

Genauso zeigt man:

$$P(X'_{mk} \in A) - P(X_{mk} \in A) \leq \tfrac{k}{d} \cdot \left(\tfrac{2d^2}{q}\right)^m.$$

Das impliziert

$$| P(X_{mk} \in A) - P(X'_{mk} \in A) | \leq \frac{k}{d} \cdot \left(\frac{2d^2}{q}\right)^m \tag{6}$$

sowie

$$d_{TV}(\mu^{(mk)}, p_{G,q}) \leq \tfrac{k}{d} \cdot \left(\tfrac{2d^2}{q}\right)^m,$$

da (6) für beliebige A gilt und somit insbesondere für das Maximum über alle A. Außerdem folgt aus der Konstruktion von X und X', dass X_{mk} gerade $\mu^{(mk)}$-verteilt ist und X'_{mk} die Verteilung $p_{G,q}$ hat. Aufgrund der Annahme $2d^2 < q$ aus Satz 8.1 haben wir eine gegen Null konvergierende obere Schranke von $d_{TV}(\mu^{(mk)}, p_{G,q})$ für m$\to \infty$ gefunden. Um unsere ϵ-Schranke einzuhalten, setzen wir:

$$\tfrac{k}{d} \cdot \left(\tfrac{2d^2}{q}\right)^m = \epsilon \Rightarrow m = \tfrac{log(k)+log(\epsilon^{-1})-log(d)}{log\left(\frac{q}{2d^2}\right)}.$$

Mit $n = mk$ folgert man:

$$n = \tfrac{k \cdot (log(k)+log(\epsilon^{-1})-log(d))}{log\left(\frac{q}{2d^2}\right)}.$$

Wählen wir unser n so, so muss $m = \frac{n}{k}$ (die Zahl der benötigten Durchgänge) nicht zwangsläufig eine ganze Zahl sein. Wir wollen aber, dass der letzte Durchgang auf jeden Fall beendet wird. Deshalb zählen wir noch k Iterationen hinzu, sodass wir

$$n = k \cdot \left(\tfrac{log(k)+log(\epsilon^{-1})-log(d)}{log\left(\frac{q}{2d^2}\right)} + 1\right),$$

wie im Satz behauptet, bekommen. □