Effiziente Algorithmen

Sommersemester 2012

Welcome, Herzlich willkommen!

Worum gehts?

- Der Entwurf und die Analyse randomisierter Algorithmen:

  Was können würfelnde Algorithmen?
- Der Entwurf und die Analyse von Online Algorithmen:

  Wir müssen eine Folge von Entscheidungen treffen, deren Güte von der Zukunft abhängen. Wir kennen aber die Zukunft nicht!

Randomisierte Algorithmen

- Quicksort: Wähle das Pivotelement zufällig.
- Zufällige Primzahltests.
- Markoff-Ketten und Stichproben-Berechnung:
  - Der Page-Rank von Google.
  - Einige schnelle Algorithmen für $\mathcal{NP}$-vollständige Probleme.
  - Simulated Annealing: Eine approximative Lösung schwieriger Optimierungsprobleme.
- Fingerprinting:
  - Universelles Hashing.
  - Zählen in Datenströmen.
- Die einfache Konstruktion kombinatorischer Objekte wie etwa Graphen mit sehr vielen Kanten, aber kleinen Cliquen.
- Weitere Einsatzgebiete in parallelen Algorithmen (Routing in Würfeln, Symmetry Breaking) und Online Algorithmen (Randomisierung gegen die Zukunft).

Online Algorithmen: Typische Fragestellungen

Das Ski Problem

- Wir fahren in den Skiurlaub und sind vor die Entscheidung gestellt, möglicherweise jeden Tag Skier für 1 Euro pro Tag zu leihen, oder für K Euro Skier zu kaufen.
- Leider kennen wir die Dauer der Skisaison nicht!
  - Wenn wir Skier am ersten Tag kaufen, kann die Saison bereits am nächsten Tag beendet sein.
  - Wenn wir jeden Tag mieten, hätten wir bei einer Saison von mindestens $K + 1$ Tagen besser am ersten Tag Skier gekauft.

- Eine optimale Entscheidung ist unmöglich.
- Gibt es aber sehr gute Strategien?
- Und wenn wir würfeln dürfen?

Online Algorithmen: Typische Fragestellungen II

- Lastverteilung: Welcher Prozess sollte die nächste Aufgabe ausführen?
- Paging: Welche Seiten sollten wir aus dem Cache auslagern, wenn die Kapazität des Cache nicht mehr ausreicht?
- Expertenauswahl: Wie können wir fast so gut sein wie ein bester Experte, wenn eine Menge von Experten unterschiedlicher (und sich stetig ändernder) Qualität vorgegeben ist?
- Strategien für die Vermögensanlage.
- Selbstorganisierende Datenstrukturen:
  - Warum passen sich Splay-Bäume den Daten an?
  - Der Entwurf superschneller Priority Queues.
  - Wir müssen wiederholt in einer Liste suchen und dürfen dabei die Struktur der Liste verändern. Wie sollten wir die Liste verändern, um insgesamt möglichst schnell zu sein?

Warum randomisierte Algorithmen?

Häufig sind randomisierte Algorithmen effizienter und einfacher als deterministische Algorithmen. Um wieviel effizienter?

- Wenn die Eingabe vollständig bekannt ist:
  - Der Miller-Rabin Primzahltest läuft für eine Eingabe n in Zeit ungefähr $O(\log_2^2 n)$, der beste bekannte deterministische Algorithmus benötigt Zeit $\Omega(\log_2^{10} n)$ und ist in der Praxis nicht brauchbar.
  - Man vermutet aber, dass randomisierte Algorithmeb höchstens polynomielle Beschleunigungen erreichen. Warum?
- Wenn die Eingabe wie in Online Algorithmen nur teilweise bekannt ist:
  - Randomisierte Algorithmen sind beweisbar besser.
  - Was heißt „besser“?

Vorwissen

(1) Die Veranstaltung „Datenstrukturen“ führt Listen, Priority Queues und Splay-Bäume ein.

(2) Die Veranstaltung „Algorithmentheorie“, bzw. „Theoretische Informatik 1“ behandelt Entwurfsmethoden für effiziente deterministische Algorithmen.

(3) Eine Veranstaltung in elementarer Stochastik ist sehr hilfreich. Die notwendigen Grundlagen werden wir aber in Kürze behandeln.

Das Kapitel 1 im Skript stellt die wichtigsten Grundlagen vor allem aus der elementaren Stochastik zusammen.

Literatur

- Randomisierte Algorithmen:
  - M. Mitzenmacher and E. Upfal, Randomized algorithms and probabilistic analysis, Cambridge University Press, 2005.
  - R. Motwani und P. Raghavan, Randomized Algorithms, *Cambridge University Press*, 1995.
- Online Algorithmen:
  - A. Borodin und R. El-Yaniv, Online Computation and Competitive Analysis, *Cambridge University Press* 1998.
- Ein Skript wird zur Verfügung gestellt.

Weitere Informationen auf der Webseite der Veranstaltung

BITTE UNBEDINGT AN DEN ÜBUNGEN TEILNEHMEN!

Administratives

Die Veranstaltung wird für die Studiengänge

- Bachelor und Master Informatik,
- Master Bioinformatik,
- Master Computational Science
- Diplomhauptstudium Bioinformatik und Informatik

angeboten.

Prüfungs- und Studienleistungen:

(1) Für die Bachelor- und Master-Studiengänge: Mündliche Prüfung nach Absprache.
Die in den Übungen erzielten Punkte verbessern das Resultat einer bestandenen mündlichen Prüfung um bis zu 20%.

(2) Ein Leistungsnachweis für das Diplom Informatik/Bioinformatik: bei regelmäßiger, aktiver Teilnahme, wenn mindestens 60% aller Übungspunkte erzielt wurden.

Wahrscheinlichkeitsverteilungen

Sei U eine endliche Menge.

(a) Eine **Wahrscheinlichkeitsverteilung** über U ist eine Funktion $\pi : U \to [0,1]$ mit $\sum_{u \in U} \pi(u) = 1$.

(b) Die Elemente von U heißen **Elementarereignisse**, Teilmengen von U heißen **Ereignisse**.

(c) Die Wahrscheinlichkeit eines Ereignisses $V \subseteq U$ ist

$$\text{prob}[V] = \sum_{u \in V} \pi(u).$$

Zufallsvariablen

- Eine **Zufallsvariable** über U ist eine Funktion $X : U \to \mathbb{R}$.
- Der **Erwartungswert** von X wird durch

$$E[X] = \sum_{u \in U} \pi(u) \cdot X(u)$$

- und die **Varianz** durch

$$V[X] = E[\, X^2 \,] - E[X]^2$$

definiert.

Gleich- und Binomialverteilung

- Die **Gleichverteilung** oder **uniforme Verteilung** auf U weist jedem Elementarereignis $u \in U$ die Wahrscheinlichkeit $\frac{1}{|U|}$ zu:

  Alle Elementarereignisse sind also gleichwahrscheinlich.
- Wir werfen eine Münze n-mal mit Ausgang „Wappen“ oder „Zahl“.
  - Die Zufallsvariable X_n zählt wie oft wir „Wappen“ erhalten.
  - Wenn p die Wahrscheinlichkeit für Wappen ist, dann ist

$$p_k = \text{prob}[\, X = k \,] = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}.$$

  - $p = (p_0, \ldots, p_n)$ ist die **Binomialverteilung** mit den Parametern n und p.
    - ★ Es ist $E[\, X \,] = n \cdot p$ und
    - ★ $V[\, X \,] = n \cdot p \cdot (1-p)$.

Die geometrische Verteilung

Wir werfen solange eine Münze, bis wir den Ausgang Wappen erhalten. Die Zufallsvariable X halte diese Anzahl der Versuche fest.

- Sei p die Wahrscheinlichkeit für Wappen.
- Dann ist

$$p_k = \text{prob}[\ X = k\ ] = (1-p)^{k-1} \cdot p.$$

- $p = (p_0, \ldots)$ ist eine **geometrische Verteilung.**
- $E[X] = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \cdot p = \frac{1}{p}$.

Das Sekretärproblem

Wir möchten eine Stelle mit einem möglichst kompetenten Bewerber besetzen.

- Wann immer sich ein Bewerber vorstellt, müssen wir dem Bewerber absagen oder aber zusagen.
- Gibt es eine Strategie, die mit guter Wahrscheinlichkeit die beste Bewerbung auszuwählt?

Annahmen:

(1) Am Ende eines jeden Bewerbungsgesprächs wird eine Note vergeben:
   - Alle Bewerber erhalten verschiedene Noten.
   - Wir suchen die Bewerbung mit der besten Note.

(2) Wir kennen die Anzahl n aller Berwerbungen.

(3) Die Bewerbungen erscheinen in zufälliger Reihenfolge:

   Alle Reihenfolgen sind gleichwahrscheinlich.

Die r-Strategie

(1) Wir sagen den ersten r Bewerbern ab, bestimmen aber die Höchstnote N, die von einem dieser Bewerber erreicht wird.

(2) Wir sagen dem ersten verbleibenden Bewerber zu, der besser als mit N benotet wird.

  Wird die Höchstnote nicht mehr erreicht, dann müssen wir dem letzten Bewerber zusagen.

> Wie gut ist diese Strategie und insbesondere, wie sollte r gewählt werden?

Die zweitbeste Bewerbung

Wir nehmen an, dass der beste Bewerber an Position opt erscheint.

- Wenn $\text{opt} \leq r$, dann haben wir verloren, denn wir haben dem besten Bewerber abgesagt.
- Nehmen wir also an, dass $\text{opt} > r$ gilt.
  - Wir sind genau dann erfolgreich, wenn der **zweitbeste unter den ersten opt Bewerbern** zu den ersten r Bewerbern gehört.

  Warum?
  - Wenn der zweitbeste zu den ersten r Bewerbern gehört, der beste aber nicht, dann wählen wir den besten Bewerber aus.
  - Wenn der zweitbeste nicht zu den ersten r Bewerbern gehört, dann brechen wir den Bewerbungsprozess vor dem besten Bewerber ab: Der zweitbeste erscheint ja vor Position opt.

Wo ist die zweitbeste Bewerbung?

Es gelte $\text{opt} > r$.

- Die zweitbeste Bewerbung erscheint mit Wahrscheinlichkeit $r/(\text{opt} - 1)$ unter den ersten r Bewerbungen. Warum?
  - ★ Die zweitbeste Berbung erscheint an jeder der opt-1 Positionen mit gleicher Wahrscheinlichkeit.
- Unsere Strategie ist, bei fixiertem opt, mit Wahrscheinlichkeit $p_r = r/(\text{opt} - 1)$ erfolgreich, wenn $\text{opt} > r$.

- Die beste Bewerbung erscheint mit Wahrscheinlichkeit $1/n$ auf irgendeiner fixierten Position.
- Es gilt

$$p_r = \sum_{\text{opt},\text{opt}=r+1}^{n} \frac{1}{n} \cdot \frac{r}{\text{opt} - 1}.$$

Analyse

p_r ist die Erfolgswahrscheinlichkeit. Dann

$$p_r = \sum_{\text{opt},\text{opt}=r+1}^{n} \frac{1}{n} \cdot \frac{r}{\text{opt}-1} \approx \frac{r}{n} \cdot \int_{x=r}^{n} \frac{1}{x}\, dx = \frac{r}{n} \cdot (\ln(n) - \ln(r)) = -\frac{r}{n} \cdot \ln(\frac{r}{n}).$$

- Wie sollten wir r wählen, so dass p_r größtmöglich ist?
- Ersetze r/n durch $x \in [0, 1]$. Wir müssen $g(x) = -x \cdot \ln(x)$ maximieren.
  - Wann ist $g'(x) = 0$?
  - $g'(x) = -\ln(x) - x \cdot \frac{1}{x} = -\ln(x) - 1.$
  - Setze $\frac{r}{n} = x = \frac{1}{e}$.

Für $r = \frac{n}{e}$ sind wir mit Wahrscheinlichkeit ungefähr $1/e$ erfolgreich!

Bedingte Wahrscheinlichkeiten

Es seien V_1 und V_2 zwei Ereignisse. Dann ist

$$\text{prob}[\ V_1 \mid V_2\ ] = \frac{\text{prob}[V_1 \cap V_2]}{\text{prob}[V_2]}$$

die bedingte Wahrscheinlichkeit von V_1 bezüglich V_2.

Das Monty Hall Problem

In einer Game Show ist hinter einer von drei Türen ein Preis verborgen.

- Ein Zuschauer rät eine der drei Türen.
- Der Showmaster Monty Hall wird daraufhin eine weitere Tür öffnen, hinter der sich aber kein Preis verbirgt.
- Der Zuschauer erhält jetzt die Möglichkeit, seine Wahl zu ändern.

Sollte er dies tun?

Bedingte Wahrscheinlichkeiten

Die Ereignisse

- P_i: Der Preis befindet sich hinter Tür i,
- Z_i: Der Zuschauer wählt zuerst Tür i und
- M_i: Monty Hall öffnet Tür i nach der ersten Wahl des Zuschauers.

- Wir nehmen o.B.d.A. an, dass der Zuschauer zuerst Tür 1 wählt und dass Monty Hall daraufhin Tür 2 öffnet.
- Entscheidend sind die bedingten Wahrsacheinlichkeiten

  prob[$P_1 \mid Z_1, M_2$], die Wahrscheinlichkeit, dass sich der Preis hinter Tür 1 befindet

  und

  prob[$P_3 \mid Z_1, M_2$], die Wahrscheinlichkeit, dass sich der Preis hinter Tür 3 befindet.

Bedingte Wahrscheinlichkeiten II

Der Preis befindet sich nicht hinter Tür 2. Also gilt

$$\text{prob}[\ P_1 \mid Z_1, M_2\ ] + \text{prob}[\ P_3 \mid Z_1, M_2\ ] = 1.$$

- Es ist $\text{prob}[\ E_1 \mid E_2\ ] = \text{prob}[\ E_1 \cap E_2\ ]/\text{prob}[\ E_2\ ]$. Also folgt

$$\text{prob}[\ E_1 \cap E_2\ ] = \text{prob}[\ E_1 \mid E_2\ ] \cdot \text{prob}[\ E_2\ ].$$

- Wir erhalten

$$\begin{aligned}
& \text{prob}[\ P_1 \mid Z_1, M_2\ ] \cdot \text{prob}[\ Z_1, M_2] \\
= \quad & \text{prob}[\ Z_1, M_2 \mid P_1\ ] \cdot \text{prob}[\ P_1\ ] = (\frac{1}{3} \cdot \frac{1}{2}) \cdot \frac{1}{3} = \frac{1}{18} \\
\text{und} \quad & \text{prob}[\ P_3 \mid Z_1, M_2\ ] \cdot \text{prob}[\ Z_1, M_2] \\
= \quad & \text{prob}[\ Z_1, M_2 \mid P_3\ ] \cdot \text{prob}[\ P_3\ ] = (\frac{1}{3} \cdot 1) \cdot \frac{1}{3} = \frac{1}{9}.
\end{aligned}$$

Bedingte Wahrscheinlichkeiten III

Was wissen wir?

- Wir müssen entscheiden, ob
$$\mathrm{prob}[\ P_1 \mid Z_1, M_2\ ] \overset{?}{>} \mathrm{prob}[\ P_3 \mid Z_1, M_2\ ] \text{ gilt.}$$
- Es ist $\mathrm{prob}[\ P_1 \mid Z_1, M_2\ ] + \mathrm{prob}[\ P_3 \mid Z_1, M_2\ ] = 1$ und
$$\begin{aligned} \mathrm{prob}[\ P_1 \mid Z_1, M_2\ ] \cdot \mathrm{prob}[\ Z_1, M_2] &= \frac{1}{18}, \\ \mathrm{prob}[\ P_3 \mid Z_1, M_2\ ] \cdot \mathrm{prob}[\ Z_1, M_2] &= \frac{1}{9}. \end{aligned}$$

Der Zuschauer sollte seine Wahl ändern, denn
$$\mathrm{prob}[\ P_3 \mid Z_1, M_2\ ] = \frac{2}{3} \text{ und } \mathrm{prob}[\ P_1 \mid Z_1, M_2\ ] = \frac{1}{3}.$$

Eine kompaktere Argumentation

Sei W die *Zufallsvariable*, die die erste Wahl des Zuschauers beschreibt und sei T die Zufallsvariable, die die richtige Tür beschreibt.

- Die Änderungsstrategie findet genau dann die richtige Tür, wenn W und T unterschiedliche Werte annehmen.
- Das Ereignis $W \neq T$ hat die Wahrscheinlichkeit $\frac{2}{3}$, denn der Zuschauer wird im ersten Versuch die richtige Tür mit Wahrscheinlichkeit $1/3$ auswählen.
- Fazit: Mit höherer Wahrscheinlichkeit war die erste Wahl schlecht und wird durch die veränderte Wahl richtig.

Unabängige Zufallsvariablen

Zufallsvariablen $X_1, \ldots, X_n$ heißen **unabhängig** genau dann, wenn für alle $x_1, \ldots, x_n \in \mathbb{R}$

$$\text{prob}[\ X_1 = x_1 \wedge \cdots \wedge X_n = x_n\ ] = \Pi_{i=1}^{n} \text{prob}[\ X_i = x_i\ ].$$

- $X_1, \ldots, X_n$ sind genau dann unabhängig, wenn $X_1, \ldots, X_{n-1}$ unabhängig sind und

  $$\text{prob}[\ X_n = x_n \mid X_1 = x_1 \wedge \cdots \wedge X_{n-1} = x_{n-1}\ ] = \text{prob}[\ X_n = x_n\ ]$$

  für alle $x_1, \ldots, x_n$ gilt.
- Insbesondere bleibt der Ausgang von X_n völlig offen, selbst wenn der Ausgang von $X_1, \ldots, X_{n-1}$ bekannt ist.

Wiederholte Experimente

Angenommen, wir wiederholen ein Experiment n-mal. Den Ausgang der $i$ten Wiederholung bezeichnen wir mit X_i.

- Der Ausgang X_n des letzten Experiments ist völlig offen, selbst wenn der Ausgang aller früheren Experimente bekannt ist.
- Mit induktivem Argument:

Die Wiederholung eines Zufallsexperiments führt auf unabhängige Zufallsvariablen.

- Eine häufige Anwendung:
  - Ein Zufallsexperiment sei mit Wahrscheinlichkeit p erfolgreich und mit der Gegenwahrscheinlichkeit $1 - p$ erfolglos.
  - Wenn X_i der Ausgang der $i$ten Wiederholung des Experiments ist, dann sind die Zufallsvariablen unabhängig.

Abhängige Zufallsvariablen

- Zufallsvariablen $X_1, \ldots, X_n$ mit

$$\sum_{i=1}^{n} X_i = 0$$

  sind stets abhängig: Der Ausgang von X_n ist determiniert, wenn der Ausgang von $X_1, \ldots, X_{n-1}$ bekannt ist.
- Im Monty Hall Problem sei
  - T die Zufallsvariable, die die richtige Tür beschreibt und
  - A die von der Änderungsstrategie gewählte Tür.

  Dann ist $\mathrm{prob}[A = T] = 2/3$:

  Die Zufallsvariablen A und T sind korreliert (und damit abhängig), denn bei Unabhängigkeit müsste $\mathrm{prob}[A = T] = \frac{1}{3}$ gelten.

Was muss man wissen?

X und Y seien Zufallsvariablen.

(a) Es ist stets
$$E[\, X + Y \,] = E[X] + E[Y].$$

Wenn X und Y unabhängig sind, dann gilt
$$E[\, X \cdot Y \,] = E[X] \cdot E[Y] \;\text{ und }\; V[X + Y] = V[X] + V[Y].$$

(b) Für Ereignisse $V_1, V_2, \ldots, V_n$ gilt die „Union Schranke"
$$\text{prob}[\, \bigcup_{i=1}^{n} V_i \,] \leq \sum_{i=1}^{n} \text{prob}[V_i].$$

Die Markoff Ungleichung

Für eine Zufallsvariable $X \geq 0$ and jedes $a > 0$ gilt

$$\text{prob}[\, X > a \,] \leq \frac{E[X]}{a}.$$

Z.B. Mit Wahrscheinlichkeit höchstens $1/2$ ist X größer als das Doppelte seines Erwartungswerts.

- $\pi(u)$ sei die Wahrscheinlichkeit des Elementarereignisses $u \in U$.
- Nach Definition des Erwartungswerts ist

$$\begin{aligned} E[X] &= \sum_{u \in U} \pi(u) \cdot X(u) \\ &\geq \sum_{u \in U, X(u) > a} \pi(u) \cdot X(u) \\ &\geq \text{prob}[\, X > a \,] \cdot a. \end{aligned}$$

Welche Rolle spielt die Varianz?

Die Tschebyscheff Ungleichung

Für jede Zufallsvariable X gilt

$$\text{prob}[\ |X - E[X]| > t\ ] \leq \frac{V[X]}{t^2}.$$

Je kleiner die Varianz, umso unwahrscheinlicher sind große Abweichungen vom Erwartungswert!

- Mit der Markoff-Ungleichung

$$\text{prob}[\ (X - E[X])^2 > t^2\ ] \leq \frac{E[(X - E[X])^2]}{t^2}.$$

- Jetzt beachte

$$\begin{aligned} E[\ (X - E[X])^2\ ] &= E[\ X^2 - 2X \cdot E[X] + E[X]^2\ ] \\ &= E[X^2] - 2E[X]^2 + E[X]^2 = V[X]. \end{aligned}$$

Abweichungen für unabhängige Experimente

Die Chernoff Ungleichungen

$X_1, \ldots, X_n$ seien **unabhängige**, **binäre** Zufallsvariablen, wobei

$$p_i = \text{prob}[\ X_i = 1\ ]$$

die Erfolgswahrscheinlichkeit von X_i ist. Dann ist $E = \sum_{i=1}^{n} p_i$ die erwartete Anzahl von Erfolgen und

$$\begin{aligned} \text{prob}[\ \sum_{i=1}^{n} X_i > (1+\beta) \cdot E\ ] &\leq e^{-E \cdot \beta^2/3} \\ \text{prob}[\ \sum_{i=1}^{n} X_i < (1-\beta) \cdot E\ ] &\leq e^{-E \cdot \beta^2/2} \end{aligned}$$

für jedes $\beta > 0$ (bzw. $0 < \beta \leq 1$ im zweiten Fall).

Eine typische Anwendung

Es gelte $\mathrm{prob}[X_i = 1] = p$.
Wenn $X_1, \ldots, X_n$ unabhängig sind und $E = n \cdot p$ gilt, dann folgt

$$\mathrm{prob}[\sum_{i=1}^{n} X_i > (1+\beta) \cdot E\,] \quad \leq \quad e^{-pn \cdot \beta^2/3}$$
$$\mathrm{prob}[\sum_{i=1}^{n} X_i < (1-\beta) \cdot E\,] \quad \leq \quad e^{-pn \cdot \beta^2/2}.$$

- **Markoff**: Eine Zufallsvariable übertrifft das Doppelte ihres Erwartungswerts mit Wahrscheinlichkeit höchstens 1/2.
- **Chernoff**: Eine Summe unabhängiger, binärer Zufallsvariablen übertrifft das Doppelte ihres Erwartungswerts mit Wahrscheinlichkeit höchstens

$$e^{-pn/3}.$$

Wie zuverlässig sind wiederholte Experimente?

- Angenommen, wir wissen dass ein binäres Experiment mit Wahrscheinlichkeit mindestens $2/3$ eine “richtige” Antwort liefert.
- Wir wiederholen das Experient n Mal und übernehmen das Mehrheitsergebnis.
- Wie groß muss n, sein, damit das Mehrheitsergebnis „hochwahrscheinlich“ richtig ist?

- Wenn wir einen Fehler machen, dann ist

$$\sum_{i=1}^{n} X_i < n/2 = (3/4) \cdot (2n/3) = (1 - 1/4) \cdot (2n/3).$$

- Wir machen einen Fehler mit Wahrscheinlichkeit höchstens

$$e^{-2n/3 \cdot (1/4^2)/2} = e^{-n/48}.$$

Und wann ist der Fehler klein genug?

Wie zuverlässig sind wiederholte Experimente? II

Man schätzt, dass es ungefähr

$$10^{78}$$

Atome im Universum gibt.

- Es ist $10^{-78} \geq e^{-3 \cdot 78}$.
- Wir erhalten eine Fehlerwahrscheinlichkeit von höchstens 10^{-78}, wenn $n/48 > 3 \cdot 78$.

Die inverse Fehlerwahrscheinlichkeit ist größer als die Atome im Universum, wenn $n = 11232$.
Ein Fehler ist damit wesentlich unwahrscheinlicher als ein Sechser im Lotto.

Die Fragestellung

Wir möchten den Erwartungswert $E[X]$ eines Zufallsexperiments X möglichst exakt messen.
Leider ist das Experiment instabil, d.h. die Varianz von X ist groß.

Wir führen $n = m \cdot k$ unabhängige Experimente aus und werten die jeweiligen Experimente in zwei Phasen aus.

- In Phase 1 fassen wir jeweils k Experimente $X_{i,1}, \ldots, X_{i,k}$ zu einem Großexperiment zusammen und berechnen die Schätzungen

$$Y_i = \frac{1}{k} \cdot \sum_{j=1}^{k} X_{i,j}.$$

- In der zweiten Phase müssen wir unsere endgültige Schätzung aus den Schätzungen $Y_1, \ldots, Y_m$ bestimmen.

Was sollte unsere endgültige Schätzung sein?

Analyse von Phase 1

- In Phase 1 senken wir die Varianz, denn

$$V[\ Y_i\ ] = \frac{1}{k^2} \cdot V[\ \sum_{j=1}^{k} X_{i,j}\ ] = \frac{1}{k^2} \cdot \sum_{j=1}^{k} V[\ X_{i,j}\ ] = \frac{1}{k} \cdot V[\ X\ ].$$

  Der Erwartungswert bleibt natürlich erhalten, denn
  $E[\ Y\ ] = E[\ \frac{1}{k} \cdot \sum_{i=1}^{k} X_i\ ] = \frac{1}{k} \cdot \sum_{i=1}^{k} E[\ X\ ] = E[\ X\ ].$
- Die Tschebyscheff Ungleichung liefert

$$\text{prob}[\ |Y_i - E[X]| > t\ ] = \text{prob}[\ |Y_i - E[Y_i]| > t\ ] \leq \frac{V[Y_i]}{t^2} = \frac{V[X]}{k \cdot t^2}.$$

  Vergleiche mit der ursprünglichen Abschätzung

$$\text{prob}[\ |X - E[X]| > t\ ] \leq \frac{V[X]}{t^2}.$$

Analyse von Phase 2

Angenommen, wir haben erreicht, dass jedes Y_i mit Wahrscheinlichkeit mindestens $1-\varepsilon$ in das „**Toleranzintervall**“ $T = [E[\,X\,] - \delta, E[\,X\,] + \delta\,]$ fällt.

- Jedes Y_i liegt mit Wahrscheinlichkeit mindestens $1-\epsilon$ in T.
- Wenn der Median nicht in T liegt, dann
  - liegen mindestens $\frac{m}{2}$ Einzelschätzungen außerhalb,
  - während nur $\varepsilon \cdot m$ außerhalb liegende Einzelschätzungen zu erwarten sind.
- Wir wenden die Chernoff Ungleichung an:
  - Es ist $(1+\frac{1-2\cdot\varepsilon}{2\cdot\varepsilon})\cdot\varepsilon\cdot m = \frac{m}{2}$.
  - Und Chernoff liefert mit $\beta = \frac{1-2\cdot\varepsilon}{2\cdot\varepsilon}$,

$$\text{prob}[\,M \notin T\,] \leq e^{-\varepsilon\cdot m\cdot\beta^2/3} = e^{-(1-2\cdot\varepsilon)^2\cdot m/(12\cdot\varepsilon)}.$$

Die Fehlerwahrscheinlichkeit fällt negativ exponentiell, falls $\varepsilon < \frac{1}{2}$.

Das Handwerkzeug

- Wir müssen wissen, wie man mit Erwartungswerten rechnet
- und das Konzept unabhängiger Zufallsvariablen verstehen.
- Die „Union Schranke“

$$\mathrm{prob}[\bigcup_{i=1}^{n} E_i] \leq \sum_{i=1}^{n} \mathrm{prob}[E_i]$$

  werden wir häufig anwenden.
- Später werden wir Algorithmen bauen, die gut sind, wenn sie ihr erwartetes Verhalten erreichen.

  Die Ungleichungen von Markoff, Tschebyscheff und **Chernoff** zeigen, dass große Abweichungen unwahrscheinlich sind.

Das war's, wenn da nicht noch die Mutter aller Ungleichungen

$$e^{x/(1+x)} \leq 1 + x \leq e^x$$

(für alle $x > -1$) wäre.