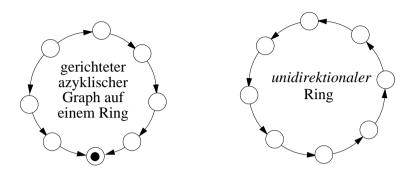
Sonderfall Ring?



- Auf unidirektionalen Ringen ist folgende Variante möglich:
 - Token wird "gejagt" (statt zurückgeholt)
 - Token kommt damit schliesslich beim Anforderer an
 - Jagdnachricht wird von einem Knoten nicht weitergeleitet, wenn dieser bereits eine solche (seit dem letzten Tokenbesuch) ausgesendet hat

- Frage: Sind FIFO-Kanäle notwendig?
- Variante: ständig kreisendes Token ("perpetuum mobile")
 - keine "Jagdnachrichten" nötig
 - --> Algorithmus hatten wir anfangs bereits genannt!

Welcher Baum beim O(log n)-Verfahren?

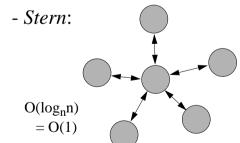
- Baum mit inneren Knoten vom *Grad k*:
 - --> längster Weg hat Länge O(log_k n); durchschnittliche Weglänge ebenfalls
 - --> bestimmt die Nachrichtenkomplexität bei schwacher Last
 - Knoten im "Zentrum" werden mehr belastet als weiter aussen liegende Knoten!

fast nie mehrere gleichzeitige Konkurrenten

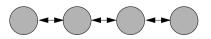
- Grad an Fairness?

Nachrichtenkomplexität bei starker Last:

- Idee: Jedes Weitersenden des Tokens ein "Treffer"
- Genauer: Token traversiert den Baum (wegen Fairness)
 - --> n Knotenbesuche bei 2(n-1) Token-Nachrichten
 - --> weiterer Faktor 2, da gleiche Anzahl von requests
 - --> ca. 4 Nachrichten pro Betreten des kritischen Abschnittes
- Algorithmus wird bei starker Last also "besser"!



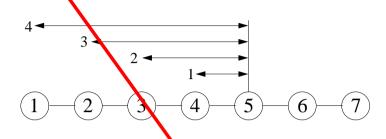
- Kürzeste Wege --> beste Topologie? (D.h. wenn schon Baum, dann diesen?)
- Unterschied zur Lösung mit zentralem Manager? (Und was war daran so schlecht?)
- Beachte: Einen Spannbaum gibt es immer; ein Stern erfordert aber u.U. zusätzliche (logische oder physische) Verbindungen!
- Lineare Kette als entarteter Baum



--> mittlere Weglänge ≈ n/3 (Beweis: einfaches kombinator. Nachrechnen)

Mittlere Weglänge im linearen Fall

(Für diejenigen, denen das Rechnen Spass macht)



Denn es gilt:
$$\sum_{k=1}^{n} k^{2} = \frac{n (n+1) (2n+1)}{6} = \frac{1}{2} \frac{(n-1) k (2(n-1)+1)}{6} + \frac{1}{2} \frac{n (n-1)}{2}$$

Gemittelt über alle n(n-1)/2 ungerichteten Paare $\{i,j\}$ mit $i\neq j$:

$$--> \frac{2(n-1)+1}{6} + \frac{1}{2} = \frac{2n+2}{6} = \frac{1}{3} (n+1) \approx \frac{n}{3}$$

(ggf. abzüglich der Fälle, wo das Token bereits "vor Ort" ist

Klassifikation: Token \iff Request

- 1) Token-basierte Lösungen
 - Safety ist trivial
 - Fairness bei Tokenweitergabe beachten
 - Wie fordern Prozesse das Token an? --> unterschiedliche Lösungen
 - Topologie
 - --> "Reiseweg des Tokens"
 - --> Zeitaufwand durch sequentielle Nachrichtenketten
 - Fehlertoleranz:
 - wie Tokenverlust feststellen?
 - wer darf neues Token generieren? (Eindeutigkeit notwendig!)
 - Nur anwendbar, wenn für das exklusive Betriebsmittel von vornherein ein Token eingerichtet wird --> nicht immer möglich!
- Beispiel für a priori unbekannte exklusive Betriebsmittel: "Reservierungszeiten für einen Tennisplatz"
 - "Ich benötige ihn übermorgen von 10.28 12.17 exklusiv"
 - Dafür lässt sich nicht von vornherein ein Token generieren!
 - Vielleicht: ein Token "Tennis <Zeitintervall>" dynamisch generieren?
 - Aber: wer garantiert, dass ein anderer dies nicht gleichzeitig tut?
 - Zurück zu einer zentralen Lösung mit allen Nachteilen? ("Tennisplatz als Monitor")
- Anderes Bsp. für ein "abstraktes Betriebsmittel": Terminvereinbarung mit einer beliebigen Menge von Teilnehmern
- Exklusives Generieren eines Tokens unter symmetrischen Bedingungen --> *Election-Problem* (--> später)

Klassifikation (2)

2) Request-basierte Lösungen

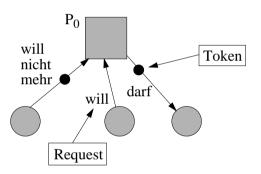
- wen sollen die Prozesse fragen? (request set)
- Safety sicherstellen!
- Deadlockfreiheit ist nicht trivial

auch hierfür verschiedene Lösungen

3) Zentraler Manager

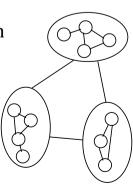
bzw. Server oder Monitor

- ist entarteter Fall *beider* Klassen!



4) Hierarchische / hybride Verfahren

- bei grossen Systemen mehrstufig (mittels "Stellvertreter")
- auf verschiedenen Stufen / in verschiedenen Clustern ggf. unterschiedliche Verfahren



Vergleich der Nachrichtenkomplexität

(pro Betreten des kritischen Abschnittes bzw. Anforderung des Betriebsmittels)

Token-Ring (LeLann,1977)	1 ∞
Lamport (1978) Vorlesung "vert. Systeme"	3 (n-1)
Ricart / Agrawala (1981)	2 (n-1)
Ricart / Agrawala (1983) · · · · · · · · · · · ·	n
Maekawa (1985)	$O(\sqrt{n})$
van de Snepscheut; Raymond; Naimi-Trehel (1987/89) · · · · · · · · · · · · · · · · · · ·	$O(\log_k n)$
zentraler Manager · · · · · · · · · · · · · · · · · · ·	2 ▼

- Offenbar soll es aber eben nicht alleine auf die Nachrichtenkomplexität ankommen!
- "Qualitative" Merkmale oft wichtiger!

Wechselseitiger Ausschluss: Kriterien

- Nachrichtenkomplexität

für die Qualität eines Lösungsalgorithmus

- Symmetrie
 - syntaktisch: gleicher Algorithmus für alle ◀
 - semantisch: gleiche Last für alle etc. ←

Inwiefern würde man den O(log n)-Algorithmus als symmetrisch bezeichnen?

- Fehlertoleranz
 - Verhalten des Algorithmus bei Fehlern
 - z.B. Nachrichtenverlust
 - oder: nach Abbruch von aussen wegen Deadlock der Anwendung
 - Zusatzaufwand, um (etwa bei erkanntem Fehler) wieder einen konsistenten Zustand herzustellen
- Grad an Fairness
 - inwieweit wird zeitlich globale Reihenfolge der Requests eingehalten?
- Zeitbedarf zwischen Freigabe und Benutzung durch einen anderen Prozess
 - (minimale) Länge sequentieller Nachrichtenketten
 - -Bsp.: bei O(log n)-Algorithmus ist Zeitbedarf auch O(log n) (statt O(1) wie bei einigen anderen Algorithmen)!
 - verschiedene Lastsituationen berücksichtigen:
 - schwache Last --> nur selten mehr als ein Konkurrent
 - hohe Last --> Betriebsmittel fast ständig in Benutzung
- Effizienz / Einfachheit der Implementierung
 - z.B.: wie wird broadcast / multicast ("request an alle") realisiert? (als effiziente Systemoperation; auf Ring; mit Echo-Algorithmus...)
 - wird eine spezielle Topologie vorausgesetzt (Ring, Baum,....) bzw. muss jeder Prozess jeden anderen kennen?

Übungen (4)

(1) Fischer Alois fängt jede Minute einen Fisch.

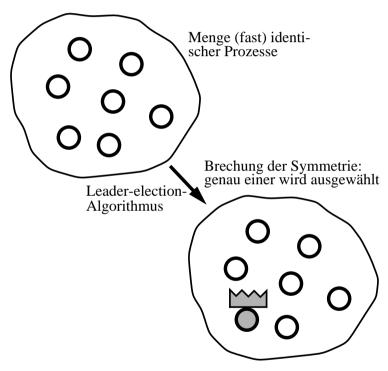
- (a) Wie lange fischt er im Mittel bei der "Stopregel" "grösserer Fisch als der erste gefangene oder kein Fisch mehr übrig" bei einem Teich/See/Meer von n verschieden grossen Fischen mit n=1000, $n=10^6$, $n=2^{31}$, $n=10^{13}$?
- (b) ... wenn schneller Fische geboren werden als Alois fangen kann?
- (c) ... wenn "grösserer" durch "kleinerer" in der Stopregel ersetzt wird? Diskutieren Sie dies im Vergleich zu (a)!

Sie mögen dies simulieren, wenn Sie an den theoretischen Ergebnissen zweifeln. Wenn Sie stochastisch simulieren, achten Sie auf einen guten Zufallszahlengenerator. (Lesen Sie dazu z.B.: PARK, S.K. and MILLER, K.W., Random Number Generators: Good Ones are Hard to Find, Comm. of the ACM 31:10, pp. 1192-1201, 1988.) Mitteln Sie jedes "Makroexperiment" über viele, typischerweise O(n) Einzelexperimente (also Kollegen von Alois). Vergleichen Sie die Mittelwerte mehrerer Makroexperiment (stabil, verlässlich?). Ermitteln Sie auch die relative Häufigkeit von m = 1,...,10 sowie von m = n - 1 und m = n ("leergefischt") Fängen.

(2) Wartezeit bis zum ersten echten Rekord:

- (a) Nach wievielen Jahren ist in einem Jahrhundert im Mittel der erste echte (d.h. verschieden vom ersten Jahr, das immer einen unechten "Rekord" darstellt) Rekord "kältester Januar" fällig?
- (b) Und in einem Jahrtausend? Darf hier etwas anderes herauskommen? Diskutieren! (Ungläubige mögen wieder simulieren.)
- (c) Wie wahrscheinlich ist es, dass in einem Jahrhundert bzw. Jahrtausend der erste echte Rekord gerade auf das 3. Jahr fällt? Und auf das 100.? Und, bei "Jahrtausend", auf ein Jahr nach dem 100.?
- (d) Gehen Sie spazieren. Wievielen Menschen begegnen Sie im Mittel, bis Sie einem grösseren begegnen? (Oder sind Sie selbst der grösste (:-)? Kann man diese Möglichkeit ganz vernachlässigen?)

Das Election-Problem

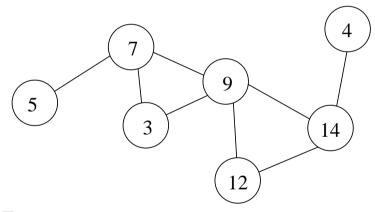


Anwendung z.B.:

- Monitorstation bei Token-Ring-LANs festlegen
- Generierung eines eindeutigen Tokens
- "Root bridge" für Spannbaum bei Ethernet-LANs
- "Symmetrisierung" anderer Algorithmen (Verwendung als vorgeschalteter Basisalgorithmus)

Election mit "message extinction"

- Alle Knoten haben unterschiedliche Nummern > 0
- Verteilter, symmetrischer Algorithmus zur Bestimmung der grössten Identität
- Jeder Knoten soll schliesslich grössten kennen
- Jeder Knoten darf unabhängig (gleichzeitig) den Algorithmus initiieren
- Schema der verteilten Approximation anwenden



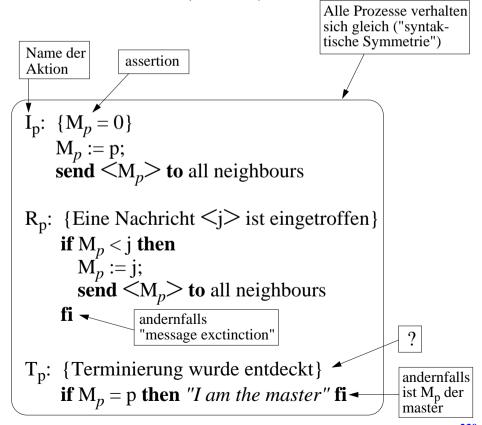
Fragen:

- Anzahl der Nachrichten?
- Terminierung?
- Bessere Algorithmen für gleiches Problem?
- Spezielle Topologien (Ring, Baum)...?

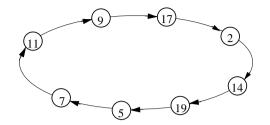
Der Election-Algorithmus

- Nachrichtengesteuerte Spezifikation des Algorithmus ("message driven")
- Atomare Aktionen vereinfacht u.a. auch die Verifikation
- Jeder Prozess mit Identität p hat lokale Variable M_p

- M_p ist initial 0; am Ende enthält M_p die Identität des Masters ("leader")



Leader-Election auf einem Ring



- Voraussetzung hier: unidirektionaler Ring, wobei alle Identitäten verschieden sind

1. Idee: - Jeder Prozess wacht irgendwann auf (spätestens, wenn er eine Nachricht von einem anderen erhält)

"Bully-Algorithmus"

- Startet vollständigen Ringumlauf

- Meldet, ob unterwegs einen grösseren getroffen

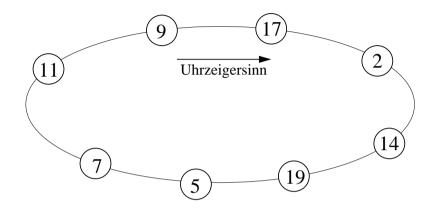
==> n² Einzelnachrichten (n = Anzahl der Prozesse)

2. Idee: Message-extinction anwenden!

 I_p : $\{M_p = 0\}$ $M_p := p;$ **send** $\leq M_p \geq$ **to** next R_p : {Eine Nachricht $\leq j \geq$ ist eingetroffen} if $M_p < j$ then $M_p := j$; send $\leq M_p \geq$ to next **if** j = p **then** "I am the master" **fi**

Election: Nachrichtenkomplexität

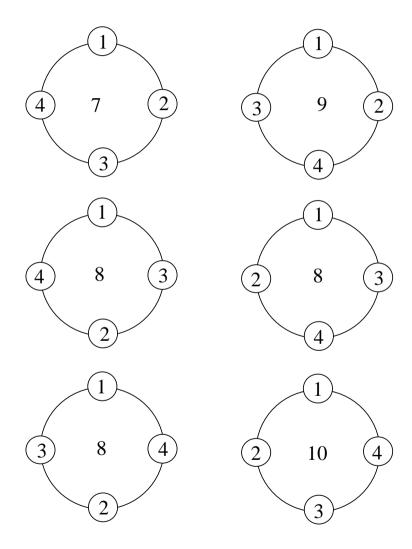
- Message-extinction-Prinzip von Chang und Roberts 1979
 - war einer der ersten verteilten Algorithmen



- Worst-case Nachrichtenkomplexität bei k Startern: n + (n-1) + (n-2) + ... + (n-k+1) = nk - k(k-1)/2--> O(n²) bei Ringgrösse n und k=n
- Wie hoch ist die *mittlere* Nachrichtenkomplexität?
 - bei "zufälliger" Permutation der Identitäten
- Wie hoch ist die (Einheits)zeitkomplexität?
 - wenn alle gleichzeitig starten?
 - beim zeitversetzten Starten?

Mittlere Nachrichtenkomplexität (1)

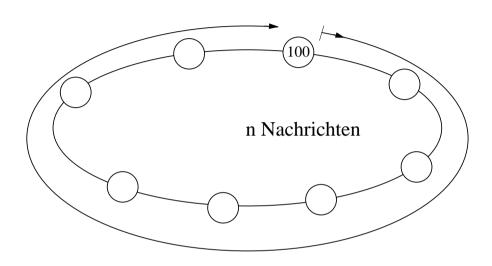
- Beispiel: Sei k = n = 4
- Über alle Permutationen mitteln (wieviele?)

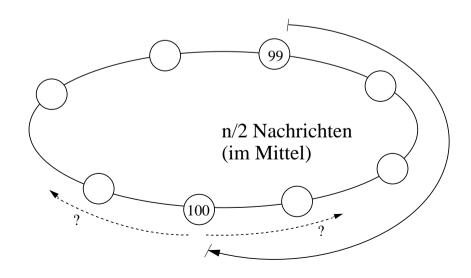


50/6 = 8.333... Nachrichten im Mittel

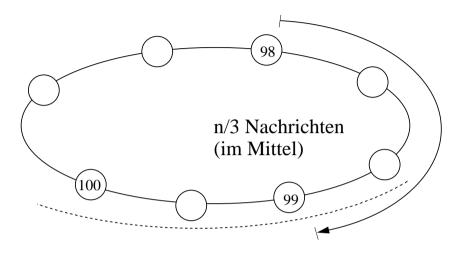
Mittlere Nachrichtenkomplexität (2)

Sei n = 100; id = 1, 2, ..., 100





Mittlere Nachrichtenkomplexität (3)



Vermutung: Drittgrösster sollte im Mittel nach n/3 Schritten auf einen grösseren Prozess (99 oder 100) treffen

- Lässt sich durch explizites Nachrechnen aller ca. n² Fälle auch beweisen...
- Allgemeiner Beweis für die Vermutung "n/i Schritte beim i-t grössten"?

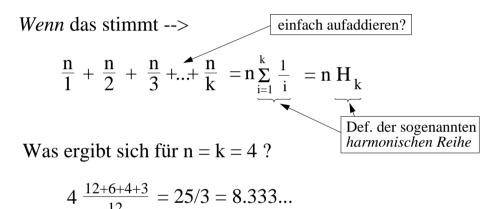
Mittlere Nachrichtenkomplexität (4)

- Grösste Identität: n Nachrichten (immer)

- Zweitgrösste: im Mittel n/2

- Drittgrösste: im Mittel n/3 ??

stimmt das? - i-t grösster: im Mittel n/i



Stimmt mit früherem Wert überein! ==> Vermutung: n H_k ist korrekt

- Man kann tatsächlich exakt zeigen:

Die mittlere Nachrichtenkomplexität beträgt

$$n \sum_{i=1}^{k} \frac{1}{i} = n H_k \approx \mathbf{n} \ln \mathbf{k}$$

Wir wollen dies gleich genauer herleiten...

Unabhängige Fälle?

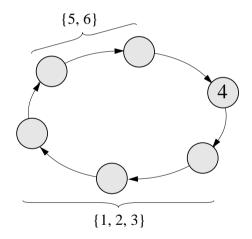
Im allgemeinen sind die "Zufallsvariablen", die die Länge der von Prozess i initiierten Nachrichtenkette repräsentieren, nicht unabhängig voneinander!

Beispiel: Gegeben 6 Prozesse 1, 2, ... 6 im Ring. Sei A das Zufallsereignis "Nachrichtenkette von Prozess 4 hat die Länge 4" und B das Zufallsereignis "Nachrichtenkette von Prozess 5 hat die Länge 2"

Aus A folgt nebenstehende Situation: dann kann die von 5 initiierte Nachrichtenkette jedoch nur die Länge 1 oder 5 haben, nicht jedoch 2, wie von B gefordert

 $==> \operatorname{prob}(A \cap B) = 0$, obwohl prob(A) \times prob(B) \neq 0

==> A und B sind *nicht* unabhängig voneinander!



Allerdings (hier ohne Rechtfertigung):

- die Zufallsvariablen in obigem Beispiel sind paarweise unkorreliert
- Erwartungswerte (d.h. die mittleren Längen der Nachrichtenketten) dürfen daher aufaddiert werden, um den Mittelwert bzgl. der Gesamtnachrichtenzahl zu erhalten (in "Formeln": $E(X+Y) = \tilde{E}(X) + E(Y)$)

Die harmonische Reihe Hn ...

Beh.: Nn divergiert

Bew.: Fasse 4 Reihenglieder ab 1/4 zusammen, (jew. grösser als 1/8), 8 Reihenglieder ab 1/8...

Beh. (o. Bew.)
$$\sum_{i} \frac{1}{i^r}$$
 konvergiert gdw. $r > 1$
 ($r = 2 --> \pi^2$)6 ... Riemann'sche Zeta-Funktion)

==> Harmonische Reihe "divergiert gerade noch"

Beh. (o.Bew.)
$$\gamma = \lim_{n \to \infty} (H_n - \ln n) = 0.3772156649...$$

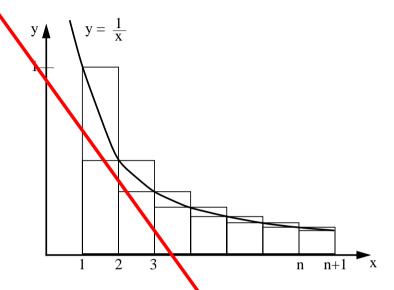
Euler'sche Konstante

- ==> Hn lässt sich zumindest für grosse n durch ln n abschätzen
 - diese Abschätzung lässt sich sogar noch präzisieren...

Abschätzung der harmonischen Reihe Hn

Beh.: $\ln n < H_n < 1 + \ln n$

Bew.:



- Fläche der Obertreppe von 1 bis n+1: H_n
- Fläche unter Kurve von 1 bis n+1: $\int_{1}^{1+1} dx = \ln(n+1)$ $==> H_n > \ln(n+1) ==> \underline{H_n > \ln n}$ (wg. $\ln(1) = 0$)
- Fläche der Untertreppe von 1 bis n: H_n-1
- Fläche unter Kurve von 1 bis n: $\int_{1}^{1} \frac{1}{x} dx = \ln(n)$

$$==> H_n -1 < ln(n) ==> H_n < 1 + ln n$$

Wartezeit bis zum ersten Rekord

- Wie weit kommt die von einem "x-beliebigen" Knoten initiierte Nachrichtenkette im Mittel?



In einem Teich schwimmen n Fische verschiedener Grösse - wieviele Fische muss man im Mittel noch fangen, bis man einen fängt, der grösser als der erste gefangene Fisch ist? (Oder bis der Teich leer ist)

Lösungsansatz (aber darf man wirklich so vorgehen?):

Fängt man den grössten zuerst --> n Fänge ("Pech", Teich wird leer) Fängt man den zweitgrössten zuerst --> n/2 Fänge im Mittel Fängt man den drittgrössten zuerst --> ...

Zurückführung auf ein bek. math. Modell: Urnenmodell ohne Zurücklegen

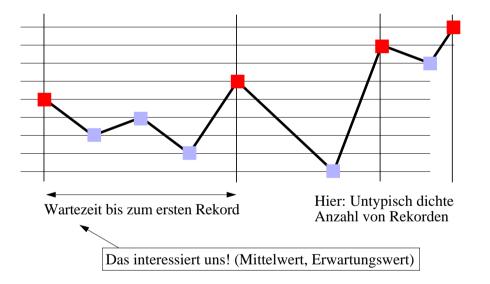
Z.B. n=100, 7.grösster Fisch initial:

- --> 6 schwarze und 93 weisse Kugeln
- --> Sei p die Wahrscheinlichkeit, eine schwarze Kugel zu ziehen ("Treffer")
- --> Beispiel: Wahrscheinlichkeit, eine weisse und dann eine schwarze zu ziehen (d.h. genau die Entfernung 2 zu schaffen): (1-p) mal Wahrscheinlichkeit für "Treffer bei 6 schwarzen und 92 weissen"
- --> ... alles gewichtet aufsummieren... (aufwendig!)

Warten auf einen neuen Rekord...

- Rekord = grösserer Wert als alle vorangehenden
- "left to right maxima" (einer Zahlenfolge)

- Z.B. "heissester August seit Anfang des Jahrhunderts"



- Anwendung: z.B. Verlobungshäufigkeit bei deletion-sort

Festhalten und weitersuchen

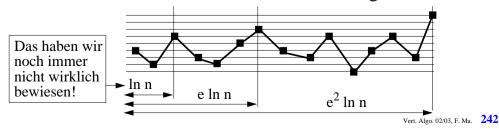
Rekorde ...werden immer seltener

- Wieviele Rekorde gibt es eigentlich?
 - z.B. heissester August in einem Jahrhundert...
 - Annahme: keine Korrelation (d.h. zufällige Permutation vorausgesetzt)
- - Wieviele 1en bei gegebener Länge n?

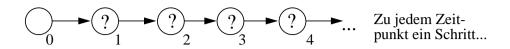
Antwort: - damit i-tes Bit auf 1 steht, muss der i-te Zahlenwert der Folge ≥ alle ersten i Zahlenwerte sein

- Wahrscheinlichkeit dafür ist 1 / i (stimmt's? wieso?)

- --> Mittlere Anzahl von 1en = 1 + 1/2 + 1/3 + ...
 - darf man wirklich einfach so aufaddieren?
- --> Mittlere Anzahl von Rekorden ist $H_n \approx \ln n$
 - Bsp: es gibt ca. 5 heisseste Auguste im Jahrhundert ... und wieviele in einem Jahrtausend?
- ==> Für einen Rekord mehr (im Mittel) muss das Zeitintervall ca. e = 2.71828 mal länger sein



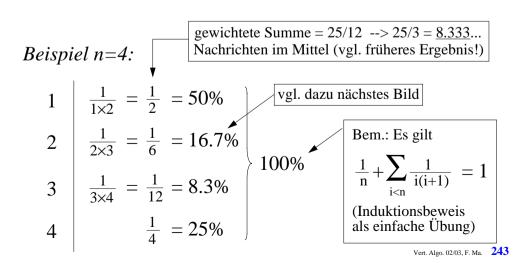
Verteilung der Lebensdauer



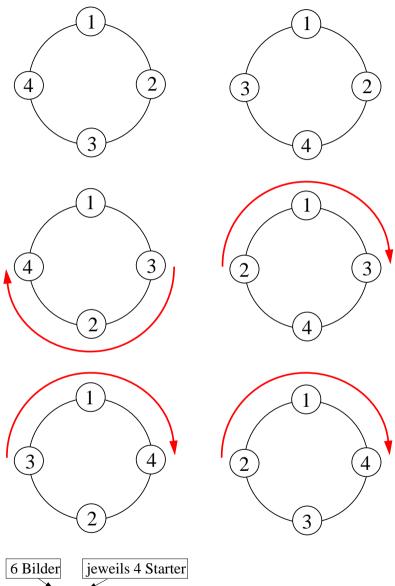
- Wahrscheinlichkeit, *genau* bis zur Position i zu gelangen (und dort besiegt zu werden)?
- $p(Lebenszeit = i) = p(Lebenszeit \ge i)$ p(Lebenszeit > i)
 - Vgl.: 18% Studis ≥ 16. Semester; 13% Studis ≥ 17. Semester ==> 5% *im* 16. Semester > 16. Semester

Wahrscheinlichkeit für "grösster unter den ersten i"

- Also: $p(Lebenszeit = i) = \frac{1}{i} \frac{1}{i+1} = \frac{1}{i(i+1)}$ (für i<n)
- Aber: $p(Lebenszeit = n) = \frac{1}{n}$ ("erfolgreicher Durchmarsch")



Lebensdauer 2 bei n=4



In 4 von $6 \times 4 = 24$ Situationen, also 1/6 aller Fälle (= 16.7%), wird für einen Initiator die Lebensdauer 2 exakt erreicht

Induktionsbeweis der Summenformel

Behauptung:

$$\sum_{i=1}^{n-1} \frac{1}{i(i+1)} = 1 - \frac{1}{n}$$

Beweis durch vollständige Induktion:

Induktionsanfang:

$$n = 1 --> 0 = 0$$

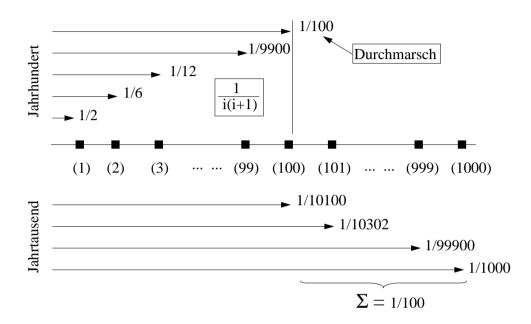
$$n = 2 --> \frac{1}{1 \cdot 2} = 1 - \frac{1}{2}$$

Induktionsschritt:

Ind. annahme
$$\sum_{i=1}^{n} \frac{1}{i(i+1)} = 1 - \frac{1}{n} + \frac{1}{n(n+1)}$$
$$= 1 + \frac{-(n+1)+1}{n(n+1)}$$
$$= 1 + \frac{-n}{n(n+1)}$$
$$= 1 - \frac{1}{n+1}$$

Rekord im Jahr i eines Jahrhunderts?

- Oder: Wahrscheinlichkeit, bei Position (i) "unterzugehen":

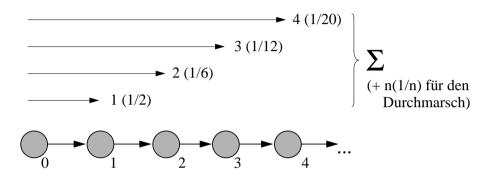


- "Durchmarsch" ist wahrscheinlicher, als an der Position davor "unterzugehen"
- Nur 50% Chance, über den ersten hinwegzukommen
- p(kein echter Rekord in einem Jahrhundert) = 1/100,
- p(... in einem Jahrtausend) = 1/1000

Rekord würde jenseits liegen: Geht bei Jahrhundert "nur" mit einem Gewicht von 100 ein, statt mehrere hundert bei Jahrtausend

Nun also: Wartezeit bis zum ersten Rekord

- Gesucht: Die *mittlere* Lebensdauer
- Dazu: Wahrscheinlichkeit mit Weglänge gewichten und alle Einzelfälle aufsummieren



- Daraus folgt allgemein für den Erwartungswert:

$$n \frac{1}{n} + 1 \frac{1}{1 \times 2} + 2 \frac{1}{2 \times 3} + 3 \frac{1}{3 \times 4} + \dots + (n-1) \frac{1}{(n-1)n} = \underline{\underline{Hn}}$$

Also: Eine von einem "x-beliebigen" Prozess initiierte Nachricht kommt im Mittel Hn weit

- darf man das nun mit n multiplizieren, da dies jeder Prozess tut?
- ja; aber generell Vorsicht bei solchen Überlegungen!

Damit haben wir also nHn als mittl. Nachrichtenkomplexität!

Qualität des Chang/Roberts-Algorithmus

- *Satz* (o. Bew.):

Für unidirektionale Ringe der Grösse n und n Initiatoren ist n Hn die optimale mittlere Nachrichtenkomplexität beim Election-Problem

- mit anderen Worten: es gibt keinen anderen Algorithmus, der "besser" ist!
- --> Chang/Roberts-Algorithmus ist "average case optimal"!

- Satz (Rotem et al., o. Bew.):

Für die Nachrichtenzahl m des Chang/Roberts-Algorithmus gilt für alle ε : $\lim_{n\to\infty}$ prob $(m<(1+\varepsilon) n H_n)=1$

Interpretation: Bei grösseren Ringen fast nie mehr als n Hn Nachrichten...