

# Algorithmen und Datenstrukturen

Wintersemester 2022/23

7. Vorlesung

**Zufall!**

# Guten Morgen!

## Tipps für unseren ersten Zwischentest am Do, 17. November:

- Lesen Sie die Definitionen der Klassen  $O$ ,  $\Omega$  und  $\Theta$  gaaaaanz genau – bis Sie sie *restlos* verstehen! Besonders Beweise der Art  $f \notin O(g)$  machen erfahrungsgemäß Schwierigkeiten.
- Lesen Sie alle Vorlesungsfolien (Vorlesungen 1–6) und Kap. 1–4 & 6 im Buch [CLRS]!
- Machen Sie möglichst viele Übungsaufgaben in Kap. 3, 4, 6 [CLRS]!
- Programmieren Sie – z.B. Pseudocode aus der Vorlesung!
- Stellen Sie Fragen – Kommilitonen, Tutoren, Erklärhiwis, mir!
- Haben Sie schon ein Buch??
 

*Tip:* Das Buch „Algorithmen & Datenstrukturen: Die Grundwerkzeuge“ von Dietzfelbinger, Mehlhorn und Sanders (Springer, 2014) kann man im Uninetz kostenlos von der Unibib herunterladen.
- Erinnern Sie sich an die Linearzeitlösung für MaxSum?  
Beweisen Sie ihre Korrektheit mit einer Schleifeninvarianten!

*Lesen!!*

# Lösen der Übungsaufgaben

- Geben Sie auf Ihren Lösungen immer die Namen aller ( $\leq 3$ ) Autoren an – nur die bekommen Punkte!
- Lösen Sie Aufgaben möglichst nur mit Mitgliedern *Ihrer* Übungsgruppe. So sehen Sie sich automatisch jede Woche.
- Wenn Sie nicht immer *alle* Aufgaben lösen können – nicht verzweifeln! Wichtig ist, dass Sie's probiert haben!
- Benützen Sie zum Fragen/Diskutieren den Chat!  
Adresse: <https://chat.uni-wuerzburg.de/group/ads22>  
Eintritt: <https://chat.uni-wuerzburg.de/invite/TZFubc>

Bei Fragen im Diskussionforum in WueCampus bekommen 400 Leute eine Mail; im Chat sehen nur die Leute Ihre Frage, die sich freiwillig dort angemeldet haben.

# Inhaltsverzeichnis

- Ein Zufallsexperiment
- InsertionSort: erwartete bzw. Durchschnittslaufzeit
- Das Geburtstagsparadoxon

# Ein Experiment

Ein Franke und ein Münchner gehen (unabhängig voneinander)  $n$  mal in verschiedenen Restaurants essen und benoten nach jedem Besuch ihr Essen.

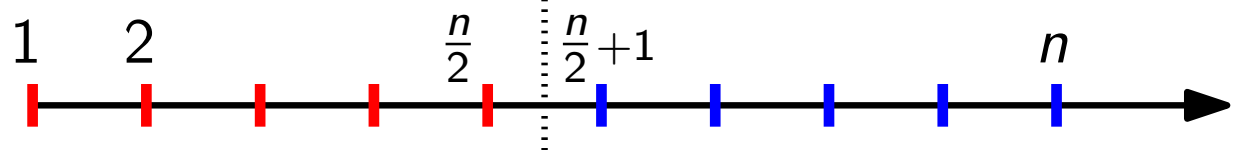
Der Franke ist zufrieden, wenn er überdurchschnittlich gut isst, d. h. wenn seine aktuelle Note über dem **Mittel** liegt.

Der Münchner ist zufrieden, wenn er besser isst als er jemals vorher gegessen hat.

*Wer is(s)t zufriedener?*

$$\frac{\sum_{i=1}^n i}{n} = \frac{n+1}{2}$$

Das Kleingedruckte:



Die Reihenfolge der Restaurants ist *zufällig*. Beide Gourmets müssen ihr Essen mit einer Zahl zwischen 1 (= sehr schlecht) und  $n$  (= sehr gut) bewerten und dürfen jede Zahl nur einmal vergeben. Die Zahl  $n$  sei gerade. Der Münchner ist beim ersten Essen zufrieden.

# Modellierung

Ein Ergebnis unseres Experiments entspricht einer Permutation der Zahlen  $1, 2, \dots, n$ .

Sei  $S_n$  die Menge all dieser Permutationen.  $\Rightarrow |S_n| = n!$

*Ergebnismenge*  $\Omega$   $\rightarrow$  *Beobachtungsmenge*  $\Omega'$

Sei  $M: S_n \rightarrow \{1, \dots, n\}$  eine *Zufallsvariable*, die angibt, wie oft der Münchner zufrieden ist.

Uns interessiert der erwartete Wert von  $M$ ,  
kurz: der *Erwartungswert*  $\mathbf{E}[M]$  von  $M$ .

**Definition:**  $\mathbf{E}[M] = \sum_{i \in \Omega'} i \cdot \mathbf{Pr}[M = i]$  (Def. für *diskrete ZV*)

„**gewichtetes Mittel**“ der Werte in  $\Omega'$

Es gilt:  $\sum_{i \in \Omega'} \mathbf{Pr}[M = i] = \mathbf{1}$

**Problem:** Was ist  $\mathbf{Pr}[M = 7]??$

# Ein Trick

**Definition:**  $\mathbf{E}[M] = \sum_{i \in \Omega'} i \cdot \mathbf{Pr}[M = i]$

Führe **Indikator-Zufallsvariable** ein (für  $i = 1, \dots, n$ ):

Sei  $M_i = \begin{cases} 1, & \text{falls Münchner nach dem } i. \text{ Essen zufrieden,} \\ 0 & \text{sonst.} \end{cases}$

$\Rightarrow \mathbf{E}[M_i] \stackrel{\text{laut Def.}}{=} 0 \cdot \mathbf{Pr}[M_i = 0] + 1 \cdot \mathbf{Pr}[M_i = 1] = \mathbf{Pr}[M_i = 1]$

**Beispiel:** Zahlenfolge =  $(7, 2, 8, 5, 9, 1, 4, 6, 3)$

$\mathbf{Pr}[M_i = 1] =$  WK, dass  $i$ . Zahl die bisher größte ist

$$\underline{\underline{=}} \frac{\text{Anz. der „guten“ Ergebnisse}}{\text{Anz. aller Ergebnisse}}$$

$$\underline{\underline{=}} \frac{\text{Anz. Perm., bei denen } i. \text{ Zahl am größten}}{\text{Anz. aller Permutationen von } i \text{ Zahlen}} = \frac{(i-1)!}{i!} = \frac{1}{i}$$

**Voraussetzung:**  
Alle Ergebnisse sind  
gleich wahrscheinlich!

# Ein Trick

**Definition:**  $\mathbf{E}[M] = \sum_{i \in \Omega'} i \cdot \mathbf{Pr}[M = i]$

Führe **Indikator-Zufallsvariable** ein (für  $i = 1, \dots, n$ ):

Sei  $M_i = \begin{cases} 1, & \text{falls Münchner nach dem } i. \text{ Essen zufrieden,} \\ 0 & \text{sonst.} \end{cases}$

$\Rightarrow \mathbf{E}[M_i] \stackrel{\text{laut Def.}}{=} 0 \cdot \mathbf{Pr}[M_i = 0] + 1 \cdot \mathbf{Pr}[M_i = 1] = \mathbf{Pr}[M_i = 1]$

**Beispiel:** Zahlenfolge =  $(7, 2, 8, 5, 9, 1, 4, 6, 3)$

$\mathbf{Pr}[F_i = 1] = \text{WK, dass } i. \text{ Zahl größer als } \frac{n+1}{2} \text{ ist}$

$$= \frac{\text{Anz. der „guten“ Ergebnisse}}{\text{Anz. aller Ergebnisse}}$$

$$= \frac{\text{Anz. Perm., bei denen } i. \text{ Zahl} > n/2}{\text{Anz. aller Permutationen von } n \text{ Zahlen}} = \frac{(n-1)! \cdot \frac{n}{2}}{n!} = \frac{1}{2}$$



# Zurück zum Erwartungswert

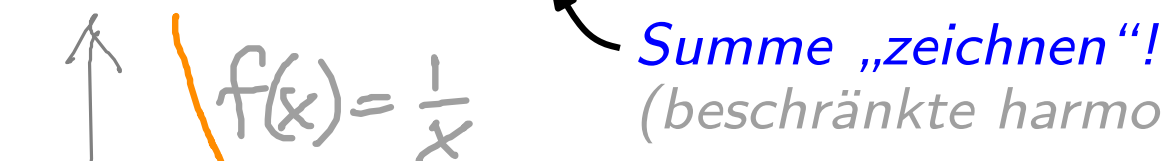
$$M_i = \begin{cases} 1, & \text{falls Münchner nach dem } i. \text{ Essen zufrieden,} \\ 0 & \text{sonst.} \end{cases}$$

$$\Rightarrow M = \sum_{i=1}^n M_i \quad (\text{Anz. Male, die Münchener zufrieden ist.})$$

$$\Rightarrow \mathbf{E}[M] = \mathbf{E}[\sum_i M_i] = \sum_i \mathbf{E}[M_i] = \sum_i \mathbf{Pr}[M_i = 1]$$

*Linearität des Erwartungswerts*      *Indikatorvariable*

$$= \sum_{i=1}^n \frac{1}{i} \leq 1 + \int_1^n \frac{1}{x} dx = 1 + (\ln x \Big|_1^n) = 1 + \ln n$$



*Summe „zeichnen“!*  
(beschränkte harmonische Reihe)

$$\frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \leq \int_1^n \frac{1}{x} dx$$

# Zurück zum Erwartungswert

$$M_i = \begin{cases} 1, & \text{falls Münchner nach dem } i. \text{ Essen zufrieden,} \\ 0 & \text{sonst.} \end{cases}$$

$$\Rightarrow M = \sum_{i=1}^n M_i \quad (\text{Anz. Male, die Münchener zufrieden ist.})$$

$$\Rightarrow \mathbf{E}[M] = \mathbf{E}[\sum_i M_i] = \sum_i \mathbf{E}[M_i] = \sum_i \mathbf{Pr}[M_i = 1]$$

*Linearität des Erwartungswerts*      *Indikatorvariable*

$$= \sum_{i=1}^n \frac{1}{i} \leq 1 + \int_1^n \frac{1}{x} dx = 1 + (\ln x \Big|_1^n) = 1 + \ln n$$

*Summe „zeichnen“!*

*(beschränkte harmonische Reihe)*

Entsprechend...

$$\mathbf{E}[F] = \sum_{i=1}^n \mathbf{Pr}[F_i = 1] = \sum_{i=1}^n \frac{1}{2} = \frac{n}{2}$$

M.a.W.: man kann erwarten,  
dass der Franke *exponentiell zufriedener* ist als der Münchner!

; -)

# Average-Case-Laufzeit von InsertionSort

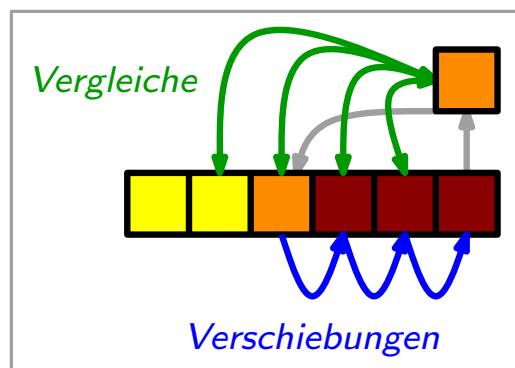
**Beob.** Der „durchschnittliche Fall“ ist i. A. schwer fassbar.

**Hier:** Was ist die durchschnittliche Laufzeit über alle Permutationen der Eingabe  $A[1..n]$  (für festes  $n$ )?

**Einfacher:** Was ist die durchschnittliche Laufzeit über alle Permutationen der Eingabe  $\langle 1, 2, \dots, n \rangle$ ?

**Wissen:**  $n - 1 \leq V_{IS}(n) \leq n(n - 1)/2$

**Beob.** Statt Vergleiche können wir auch die Anzahl der Verschiebungen  $T_{IS}$  zählen, denn



wenn wir **ein** Element einfügen (innere Schleife von InsertionSort), gilt:

$\# \text{ Verschiebungen} \leq \# \text{ Vergleiche} \leq \# \text{ Verschiebungen} + 1$

Also **insgesamt:**  $T_{IS} \leq V_{IS} \leq T_{IS} + (n - 1)$ .

# Average-Case-Laufzeit vs. erwartete Laufzeit

**Beob.** Statt der durchschnittlichen Laufzeit über alle Permutationen können wir auch die *erwartete* Laufzeit einer zufälligen Permutation betrachten.

## Warum?

Betrachte Definition:

$$\mathbf{E}[T] = \sum_{i=0}^{n^2} i \cdot \mathbf{Pr}[T = i]$$

*hier:* # Verschiebungen (d.h. Laufzeit) Anteil der Permutationen, die  $i$  Verschiebungen verursachen.

# Erwartete Laufzeit von InsertionSort

$T$  := Zufallsvariable für die Anzahl von Verschiebungen, die IS benötigt, um eine zufällige Permutation  $A[1..n]$  von  $\langle 1, 2, \dots, n \rangle$  zu sortieren

Was wäre eine gute Indikatorvariable um  $T$  auszudrücken?

$$T_{ij} := \begin{cases} 1, & \text{falls } A[i] > A[j] \\ 0 & \text{sonst.} \end{cases} \quad \text{für } 1 \leq i < j \leq n.$$

$$\Rightarrow T_j = \sum_{i=1}^{j-1} T_{ij} = \text{Anz. Pos., um die } A[j] \text{ n. li. verschoben wird.}$$

$$\Rightarrow T = \sum_{j=2}^n T_j = \sum \sum T_{ij} \quad \text{und} \quad \mathbf{E}[T] = \sum \sum \mathbf{E}[T_{ij}]$$

Aber was ist  $\mathbf{E}[T_{ij}]$ ? Laut Def.  $\mathbf{E}[T_{ij}] = \mathbf{Pr}[T_{ij} = 1] = \frac{1}{2}$

$$\Rightarrow \mathbf{E}[T] = \sum_{j=2}^n \sum_{i=1}^{j-1} \mathbf{E}[T_{ij}] = \sum_{j=2}^n \frac{j-1}{2} = \frac{1}{2} \sum_{j=1}^{n-1} j = \frac{n(n-1)}{4} \quad \square$$

# Zusammenfassung InsertionSort

## Satz. [alt]

Im *besten Fall* benötigt InsertionSort  $n - 1 \in \Theta(n)$  Vergleiche und 0 Verschiebungen.

Im *schlechtesten Fall* benötigt InsertionSort  $n(n - 1)/2 \in \Theta(n^2)$  Vergleiche/Verschiebungen.

## Satz. [neu]

Im *Durchschnitt* benötigt InsertionSort  $n(n - 1)/4 \in \Theta(n^2)$  Verschiebungen und zwischen  $n(n - 1)/4$  und  $n(n - 1)/4 + (n - 1)$ , d.h.  $\Theta(n^2)$ , Vergleiche.

*Kurz:* Bei InsertionSort gilt

**Average Case =<sub>asymptotisch</sub> Worst Case!**

# Geburtstagswahrscheinlichkeiten

**Frage:** Wie groß ist die *Wahrscheinlichkeit*, dass mindestens zwei Leute hier im Hörsaal am gleichen Tag Geburtstag haben? [siehe Abschnitt 5.4, CLRS]

**Frage':** Wie groß ist der *Erwartungswert* für die Anzahl  $X$  von Paaren von Leuten hier im Hörsaal, die am gleichen Tag Geburtstag haben?

Was wäre eine gute Indikatorvariable, um  $X$  auszudrücken?

$$X_{ij} := \begin{cases} 1 & \text{falls } i \text{ und } j \text{ gleichen Geburtstag haben} \\ 0 & \text{sonst.} \end{cases} \quad \left. \vphantom{X_{ij}} \right\} \begin{array}{l} \text{für} \\ 1 \leq i < j \leq k; \\ k = \text{Anz. Leute} \end{array}$$

$$\text{Dann gilt } X = \sum_{1 \leq i < j \leq k} X_{ij} = \sum_{i=1}^{k-1} \sum_{j=i+1}^k X_{ij}.$$

# Geburtstagserwartungen

Es gilt:

$$X_{ij} := \begin{cases} 1 & \text{falls } G_i = G_j \\ 0 & \text{sonst.} \end{cases} \quad \text{und } X = \sum_{1 \leq i < j \leq k} X_{ij}.$$

Geburtstag von Person  $j$

**Annahme:** Alle  $n$  Tage sind gleich wahrscheinlich Geburtstage.

$$\Rightarrow \mathbf{E}[X_{ij}] = \mathbf{Pr}[X_{ij} = 1] = \sum_{t=1}^n \mathbf{Pr}[G_i = G_j = t] = \frac{n}{n^2} = 1/n.$$

Ereignisse schließen sich gegenseitig aus!

(Geht auch einfacher!)

**Gesucht:**

$$\mathbf{E}[X] = \mathbf{E} \left[ \sum_{1 \leq i < j \leq k} X_{ij} \right] = \sum_{1 \leq i < j \leq k} \mathbf{E}[X_{ij}] =$$

$$= \binom{k}{2} \cdot \frac{1}{n} = \frac{k(k-1)}{2n} \geq 1 \Leftrightarrow k(k-1) \geq 2n.$$

Linearität des Erwartungswerts!

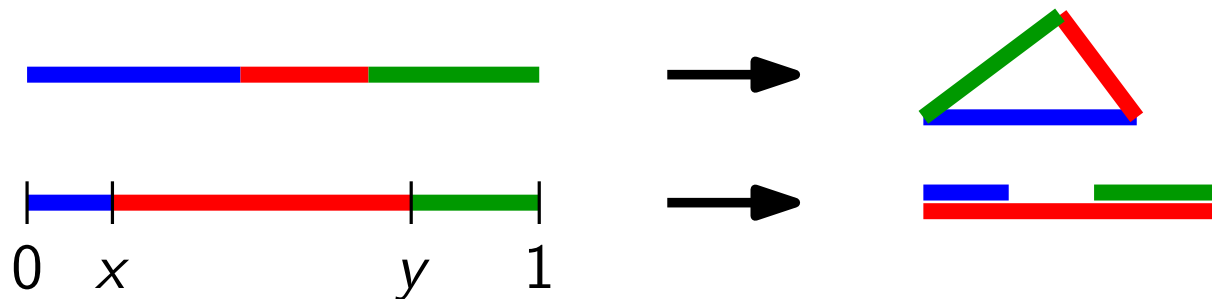
Für ein Jahr mit  $n = 365$  Tagen braucht man also nur  $k \geq 28$  Personen um ein Pärchen mit gleichem Geburtstag erwarten zu können.  $\square$



# Bonustrack (Nur eine Frage, keine Antwort!)

Angenommen, Sie haben einen Stock der Länge 1 und zerbrechen ihn an zwei zufällig gewählten Stellen  $x \in [0, 1]$  und  $y \in [0, 1]$ . (Die ZV  $x$  &  $y$  sind also *nicht* diskret!)

Was ist die Wahrscheinlichkeit, dass Sie aus den drei Teilen ein Dreieck legen können?



Mit anderen Worten – was ist die Wahrscheinlichkeit, dass  $\max\{x, y - x, 1 - y\} < 1/2$ ?

**Tipp:** Betrachten Sie das Einheitsquadrat  $[0, 1] \times [0, 1]$ .