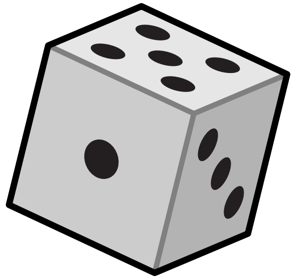
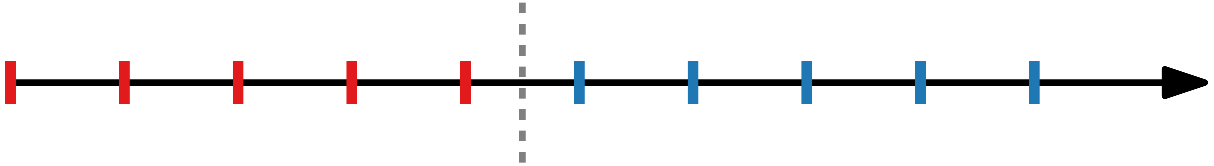
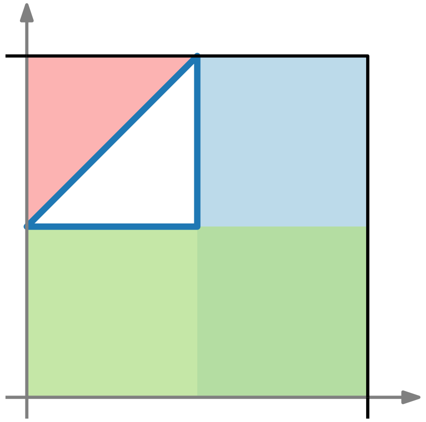
Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Algorithmen und Datenstrukturen



Vorlesung 7:
Zufallsexperimente





Alexander Wolff

Wintersemester 2025

Ein Experiment

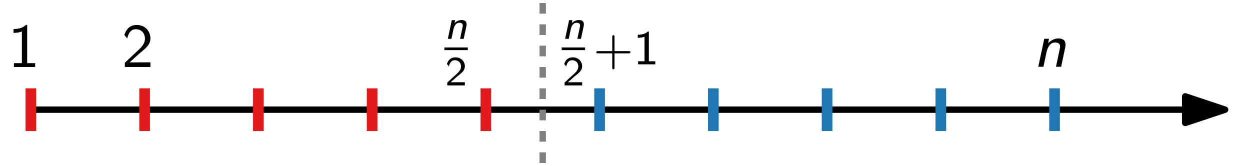
Ein Franke und ein Münchner gehen (unabhängig voneinander) n mal in verschiedenen Restaurants essen und benoten nach jedem Besuch ihr Essen.

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?

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



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 $1, 2, \ldots, n$.

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

Sei $M\colon \underbrace{S_n}_{\text{Ergebnismenge } \Omega} \to \underbrace{\{1, \ldots, n\}}_{\text{Beobachtungsmenge } \Omega'}$ 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 Ω'

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

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

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

Beispiel. Zahlenfolge $= (7, 2, 8, 5, \overset{i}{9}, 1, 4, 6, 3)$

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

$$= \frac{\text{Anzahl der „guten" Ergebnisse}}{\text{Anzahl aller Ergebnisse}}$$

Voraussetzung:
Alle Ergebnisse sind gleich wahrscheinlich!

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

Ein Trick

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

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

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

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

Beispiel. Zahlenfolge $= (7, 2, 8, 5, \overset{i}{9}, 1, 4, 6, 3)$

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

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

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

Zurück zum Erwartungswert

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

$\Rightarrow M = \sum_{i=1}^{n} M_i$ (Anzahl Essen, nach denen der Münchner 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

$= \sum_{i=1}^{n} \frac{1}{i} \leq \boxed{1 + \ln n}$

3) $\sum_{i=1}^{n} \frac{1}{i} \leq 1 + \ln n$ **harmonische Reihe**

Entsprechend...

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

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

Average-Case-Laufzeit von InsertionSort

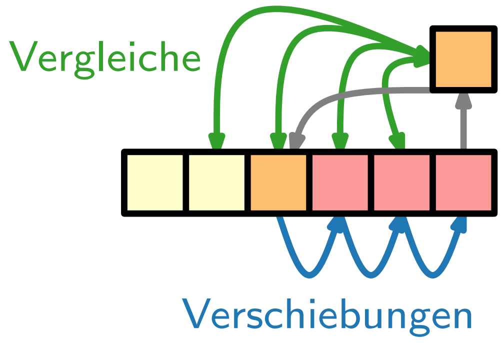
Beobachtung. Der „durchschnittliche Fall“ ist i. A. schwer fassbar.

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

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

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

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



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

$\#$ Verschiebungen $\leq$ $\#$ Vergleiche $\leq$ $\#$ Verschiebungen $+1$

d.h. insg. gilt: $T_{\text{IS}} \;\leq\; V_{\text{IS}} \;\leq\; T_{\text{IS}} + (n-1)$.

Average-Case-Laufzeit vs. erwartete Laufzeit

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

Warum?

Betrachte Definition:

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

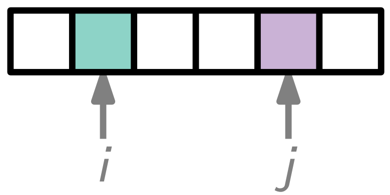
hier: # Verschiebungen (d.h. Laufzeit) — **Anteil** der Permutationen, die i Verschiebungen verursachen.

Erwartete Laufzeit von InsertionSort

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

Was wäre eine gute Indikatorvariable, um T auszudrücken?

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



$\Rightarrow \; T_j = \sum_{i=1}^{j-1} T_{ij} =$ Anzahl Positionen, um die $A[j]$ nach links verschoben wird.

$\Rightarrow \; T = \sum_{j=2}^{n} T_j = \sum_{j=2}^{n} \sum_{i=1}^{j-1} T_{ij}$ und $\mathbf{E}[T] = \sum_{j=2}^{n} \sum_{i=1}^{j-1} \boxed{\mathbf{E}[T_{ij}]}$

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

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

1) $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$ **arithmetische Reihe**

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

$$\text{Average Case} =_{\text{asymptotisch}} \text{Worst Case!}$$

Geburtstagswahrscheinlichkeiten

Frage: Wie groß ist die **Wahrscheinlichkeit**, dass mindestens zwei Leute hier in der Vorlesung 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 in der Vorlesung, die am gleichen Tag Geburtstag haben?

Was wäre eine gute Indikatorvariable, um X auszudrücken?

$$X_{ij} := \begin{cases} 1 & \text{falls } i \text{ und } j \text{ gleichen Geburtstag haben} \\ 0 & \text{sonst.} \end{cases} \Bigg\} \text{ für } 1 \le i < j \le k;\ k = \text{Anz. Leute}$$

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

Geburtstagserwartungen

Geburtstag von Person j

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

Annahme: Alle n Tage sind gleich wahrscheinlich Geburtstage.

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

Ereignisse schließen sich gegenseitig aus!

(Geht auch einfacher!)

Linearität des Erwartungswerts!

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

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

$n = 365$
$k(k-1) \ge 730$
$k = 28 \Rightarrow$
$28 \cdot 27 = 756 \ge 730$

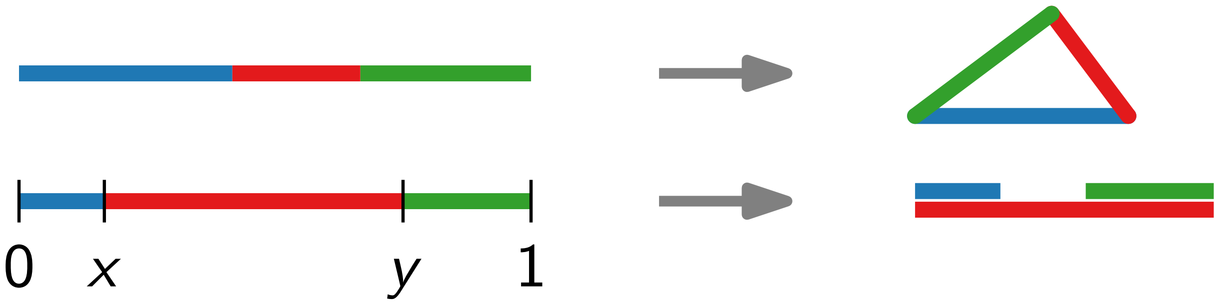
Für ein Jahr mit $n = 365$ Tagen braucht man also nur $k \ge 28$ Personen um ein Pärchen mit gleichem Geburtstag erwarten zu können. □

Bonustrack

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 $[0,1] \times [0,1]$.