Algorithmen, Wahrscheinlichkeit und Information

Geburtstags-Effekt

- Menge n Personen
- m mögliche Geburtstage (m = 365)
- WSK, dass zwei Personen am selben Tag Geburtstag haben?
- Universum von m Elementen
- n uniforme und gleich verteilte Auswahlen
- WSK dafür, dass zwei Auswahlen gleich
  …dass zwei Auswahlen kollidieren

$$\rightarrow n = \Theta(\sqrt{m})$$

Herleitung 1 - Untere Schranke
m Elemente, n Auswahlen
G_i sei Auswahl i (Geburtstag von i)
∀ d: P_{G_i}(d) = $\frac{1}{m}$

- Für zwei Personen l und k, und für bestimmten Tag d:
  $P[G_l = d \cap G_k = d] = \frac{1}{m^2}$
- WSK für den gleichen Geburtstag von zwei:
  $P[G_l = G_k) = \frac{1}{365} = \frac{1}{m}$

P[mind. zwei gleiche] = 1-P[keine zwei gleiche]
Person i:
$X_i \hat{=}$ Ereignis G_i verschieden von $G_1, \ldots, G_{i-1}$
$Y_i \hat{=} \bigcap_{j=1}^{i-1} X_i$
→ gesucht ist P[Y_n]
$Y_i = Y_{i-1} \cap X_i$
P[Y_i] = P[X_i | Y_{i-1}] · P[Y_{i-1}]
P[Y_1] = 1
P[X_i | Y_{i-1}] = $\frac{m-i+1}{m}$

P[Y_n] = P[X_n | Y_{n-1}] · P[Y_{n-1}]
= P[X_n | Y_{n-1}] · … · P[X_i | Y_{i-1}] · … · P[Y_1]
= $\prod_{i=1}^{n}$ P[X_i | Y_{i-1}]
= $\prod_{i=1}^{n} \frac{m-i+1}{m} = \prod_{i=1}^{n} \left(1 - \frac{i-1}{m}\right)$
P[Y_n] ≤ $\prod_{i=1}^{n} e^{\frac{-(i-1)}{m}}$
.....
P[Y_n] ≤ p: ln(p) ≥ $-\frac{n\cdot(n-1)}{2m}$
$\frac{n\cdot(n-1)}{2m} \geq ln(\frac{1}{p})$
Für p = $\frac{1}{2}$: $n \cdot (n-1) \geq ln(2) \cdot 2m$
dann P[Y_n] ≤ p = $\frac{1}{2}$
m = 365 Tage:

- n ≥ 23, dann P[Y_n] ≤ $\frac{1}{2}$
- n ≥ 68, dann P[Y_n] ≤ 0.002

Herleitung 2 - Indikator-ZV

ZV $G_{ij} = \{^{1,\ falls\ G_i\ =\ G_j}_{0,\ sonst}$
E[X_{ij}] E[$G_i = G_j$] = $\frac{1}{m}$
$X = \sum_{i,j} X_{ij} = \sum_i = 1^n \sum_j = 1^n X_{ij}$
E[X] = E[$\sum_i \sum_j X_{ij}$] = $\sum_i \sum_j$ E[X_{ij}] = $\sum_i \sum_j \frac{1}{m} = \frac{1}{m} \sum_i \sum_j 1 = \frac{1}{m} \cdot \frac{n \cdot (n-1)}{2}$
Falls n = $\Theta(\sqrt{m})$, dann im Erwartungswert mind. eine Kollision

Hashfunktionen

- Berechnen kurzer, eindeutiger Werte für einen beliebig langen Input
- H:$\{0,1\}^* \rightarrow \{0,1\}^k$ (fixes k)
- SHA-256, SHA-512
- Sicherheit: Es ist praktisch nicht möglich zwei x und x' zu finden → H(x) = H(x')
  → "keine Kollision"
- Angenommen: Output von H ist zufälliger k-bit String
  $m = 2^k$ → mit n = $O(\sqrt{m}) = O(2^{\frac{k}{2}})$ Operationen → Kollision

Momente und Abweichungen
Markov-Ungleichung
Theorem: Sei X eine ZV in R^+

$$\forall a > 0, \quad P[X \geq a] \leq \frac{E[x]}{a} \qquad \leftrightarrow \qquad P[X \geq c \cdot E[X]] \leq \frac{1}{c}$$

Beweis:
• I: Indikatorfunktion := $\{^{1, \geq a}_{0, X < a}$

$$E[X] = \sum_x x \cdot P_X(x) \geq \sum_{x \geq a} x \cdot P_X(x) \geq \sum_{x \geq a} a \cdot P_X(x) = a \cdot \sum_{x \geq a} P_X(x) = a \cdot P[X \geq a]$$

Beispiel: Fairer Münzwurf mit n Wiederholungen
X Anzahl Münze = Kopf
E[X] = $\frac{n}{2}$
$P[X \geq \frac{7}{8} \cdot n] \leq \frac{n}{2} \cdot \frac{8}{7n} = \frac{4}{7}$

Momente einer ZV

- Momente charakterisieren ZV
- Erwartungswert ist das erste Moment

Definition: Das letzte Moment einer ZV X

$$E[X^k]$$

Das k-te zentrale Moment von X:

$$E[(X - \mu)^k] \ wobei\ \mu\ = E[X]$$

Defintion: Varianz einer ZV X ist

$$Var[X] = E[(X - E[X])^2]$$

Definition: Standardabweichung

$$\sigma = \sqrt{Var[X]}$$

Theorem: $Var[X] = E[X^2] - E[X]^2$

Beweis:

$$\mu = E[X]$$

$$Var[X] = E[(X-\mu)^2] = E[X^2-2X\mu+\mu^2] = E[X^2]-2\mu \underbrace{E[X]}_{\mu} + \mu^2 = E[X^2]-\mu^2 = E[X^2]-(E[X])^2$$

Beispiel:

- X $\in_R$ [1,6]
  $Var[X] = \frac{1}{6} \cdot (1^2 + 2^2 + \ldots + 6^2) - (\frac{7}{2})^2 = \frac{35}{12} = 2,916666...$
- X $\in_R$ [a, b]
  $Var[X] = \frac{(b-a+1)^2-1}{12}$

Theorem: ZV X und Y unabhängig

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

Beweis:

$$E[X \cdot Y] = \sum_{x,y} x \cdot y \cdot P_{XY}(x,y) = \sum_x \sum_y x \cdot y \cdot P_X(x) \cdot P_Y(y) = \sum_x x \cdot P_X(x) \cdot \sum_y y \cdot P_Y(y) = E[X] \cdot E[Y]$$

Theorem: ZV X und Y unabhängig

$$Var[X+Y] = Var[X] + Var[Y]$$

Beweis:

$$Var[X+Y] = E[(X+Y-E[X+Y])^2] = E[(X+Y-E[X]-E[Y])^2]$$
$$= E[(X-E[X])^2 + (Y-E[Y])^2 + 2(X-E[X]) \cdot (Y-E[Y])]$$
$$= E[(X-E[X])^2] + E[(Y-E[Y])^2] + \underbrace{E[2 \cdot \underbrace{(X-\mu_X)}_{0} \cdot \underbrace{(Y-\mu_Y)}_{0}]}_{0} = Var[X] + Var[Y]$$

Varianz einer ZV mit Binomialverteilung

$$X \sim B(n,p)$$
$$Var = ?$$
$$X = \sum_{i=1}^{n} Z_i \qquad Z_i := \{ {}^{1,\ mit\ WSK\ p}_{0,\ sonst}$$
$$Z_i\ unabhängig \qquad E[Z_i] = p$$
$$Var[Z_i] = E[(Z_i-p)2] = p \cdot (1-p)^2 + (1-p) \cdot p^2 = p \cdot (1-p) \cdot (1-p+p) = p \cdot (1-p)$$
$$Var[X] = \sum_{i=1}^{n} Var[Z_i] = n \cdot p \cdot (1-p)$$

Chebyshev-Ungleichung

Theorem:

$$ZV\ X\ und\ a > 0$$
$$P[|\ X - E[X]\ | \geq a] \leq \frac{Var[X]}{a^2}$$