Algorithms and Probability

Week 7

G08 - mkilic

3.IV.2025

Overview

1. Indikatorvariablen
2. Randomisiertes Quicksort
3. Varianz
4. Wichtige Diskrete Verteilungen
5. Bedingte Zufallsvariablen
6. Mehrere Zufallsvariablen
7. Exercises

Roadmap

1. Graphentheorie
   - Zusammenhang
   - Kreise
   - Matchings
   - Färbungen
2. W'keitstheorie
   - Bedingte W'keit
   - Unabhängigkeit
   - (mehrere) Zufallsvariablen
   - Diskrete Verteilungen
   - Abschätzen von W'keiten
   - Randomisierte Algorithmen
3. Algorithmen
   - Lange-Bunte Pfade
   - MaxFlow
   - MinCut
   - Kleinster umschliessender Kreis
   - Konvexe Hülle

Indikatorvariablen

Indikatorvariablen

Für ein Ereignis $A \subseteq \Omega$ ist die zugehörige Indikatorvariable X_A definiert durch:

$$X_A(\omega) := \begin{cases} 1, & \text{falls } \omega \in A \\ 0, & \text{sonst} \end{cases}$$

Für den Erwartungswert von X_A gilt: $\mathbb{E}[X_A] = Pr[A]$

Indikatorvariablen - II

Indikatorvariable $I_A(\omega) := \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$

- Schnitt von zwei Indikatorvariablen I_A, I_B:

$$I_{A \cap B} = I_A \cdot I_B$$

Das Produkt von Indikatorvariablen ist ebenfalls eine Indikatorvariable, nämlich der Indikator für den Schnitt der beteiligten Ereignisse.

Indikatorvariablen - II

Indikatorvariable $I_A(\omega) := \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$

- Schnitt von zwei Indikatorvariablen I_A, I_B:

$$I_{A \cap B} = I_A \cdot I_B$$

Das Produkt von Indikatorvariablen ist ebenfalls eine Indikatorvariable, nämlich der Indikator für den Schnitt der beteiligten Ereignisse.

- Komplement von einer Indikatorvariable I_A:

$$I_{\bar{A}} = 1 - I_A$$

Indikatorvariablen - II

Indikatorvariable $I_A(\omega) := \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$

- Schnitt von zwei Indikatorvariablen I_A, I_B:

$$I_{A \cap B} = I_A \cdot I_B$$

Das Produkt von Indikatorvariablen ist ebenfalls eine Indikatorvariable, nämlich der Indikator für den Schnitt der beteiligten Ereignisse.

- Komplement von einer Indikatorvariable I_A:

$$I_{\bar{A}} = 1 - I_A$$

- Vereinigung von zwei Indikatorvariablen $I_A und I_B$:

$$I_{A \cup B} \rightarrow Siebformel$$

Indikatorvariablen - Beweis der Siebformel

Siebformel: Beweis

Wir wollen die W'keit von Vereinigung allgemeiner Ereignissen berechnen. Sei

$$B := A_1 \cup \cdots \cup A_n$$

Weiter seien $I_i = I_{A_i}$ und I_B die zugehöhrigen Indikatorvariablen. Beobachte

$$\begin{aligned}
\bar{B} &= \overline{A_1 \cup \cdots \cup A_n} = \bar{A}_1 \cap \cdots \cap \bar{A}_n \\
I_{\bar{B}} &= \prod_{i=1}^{n} I_{\bar{A}_i} = \prod_{i=1}^{n} (1 - I_{A_i}) \\
&= 1 - \sum_i I_{A_i} + \sum_{i<j} I_{A_i} I_{A_j} - \sum_{i<j<k} I_{A_i} I_{A_j} I_{A_k} + \cdots + (-1)^n I_{A_1} I_{A_2} \cdots I_{A_n}
\end{aligned}$$

Indikatorvariablen - Beweis der Siebformel - II

Siebformel: Beweis

Linearität des Erwartungswertes und $\mathbb{E}[I_A] = Pr[A]$ besagt:

$$\mathbb{E}[I_{\bar{B}}] = 1 - \sum_i \mathbb{E}[I_{A_i}] + \sum_{i<j} \mathbb{E}[I_{A_i} I_{A_j}] - \sum_{i<j<k} \mathbb{E}[I_{A_i} I_{A_j} I_{A_k}] + \cdots + (-1)^n \mathbb{E}[I_{A_1} I_{A_2} \cdots I_{A_n}]$$

also gilt

$1 - \Pr[B] =$

$$1 - \sum_i \Pr[A_i] + \sum_{i<j} \Pr[A_i \cap A_j] - \sum_{i<j<k} \Pr[A_i \cap A_j \cap A_k] + \cdots + (-1)^n \Pr\left[\bigcap_{i=1}^{n} A_i\right]$$

Das ist genau die Siebformel!

Randomisiertes Quicksort

```
1         if < r then
2             p ← Uniform({ℓ, ℓ + 1, ..., r})
3             t ← PARTITION(A, ℓ, r, p)
4             QUICKSORT(A, ℓ, t − 1)          \\Block I
5             QUICKSORT(A, t + 1, r)          \\Block II
6
```

Quicksort

Wir sind interessiert für Laufzeit. Sie ist von der Anzahl der Vergleiche bestimmt.

Randomisiertes Quicksort

Definiere: $X_{i,j} :=$ *Indikatorvariable für "a_i wird mit a_j verglichen"*

- Pivotelement wird mit allen anderen Elementen im selben Block verglichen. Sonst gibt es keine Vergleiche und das Pivotelement ist danach in keinem Block mehr.
- Damit das Paar (a_i, a_j) verglichen werden können, muss einer der beiden als Pivotelement gewählt werden. Jedes Paar wird maximal einmal verglichen.

Randomisiertes Quicksort

Definiere: $X_{i,j} :=$ *Indikatorvariable für "a_i wird mit a_j verglichen"*

- Pivotelement wird mit allen anderen Elementen im selben Block verglichen. Sonst gibt es keine Vergleiche und das Pivotelement ist danach in keinem Block mehr.
- Damit das Paar (a_i, a_j) verglichen werden können, muss einer der beiden als Pivotelement gewählt werden. Jedes Paar wird maximal einmal verglichen.

$$X = \sum_{1 \leq i < j \leq n} X_{i,j}$$

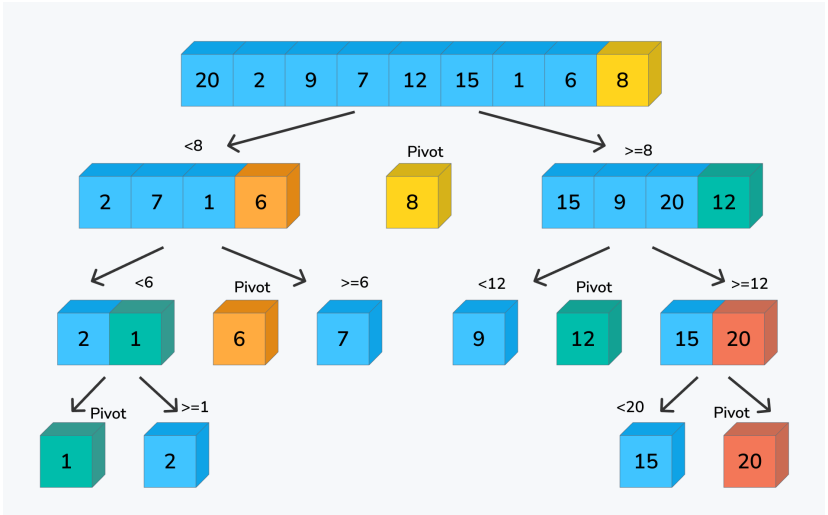
Für ein Paar (a_i, a_j) gibt es zwei Möglichkeiten:

1. a_i oder a_j wird als Pivot gewählt: $X_{i,j} = 1$ mit W'keit $\frac{2}{j-i+1}$
2. Einer von $a_{i+1}, \ldots a_{j-1}$ wird gewählt: $X_{i,j} = 0$

Also haben wir: $\mathbb{E}[X_{i,j}] = Pr[X_{i,j} = 1] = \frac{2}{j-i+1}$

$$\mathbb{E}[X] \overset{lin.}{=} \sum_{1 \leq i < j \leq n} \mathbb{E}[X_{i,j}] = \sum_{1 \leq i < j \leq n} \frac{2}{j-i+1} \cdots \leq 2n \ln(n)$$

Randomisiertes Quicksort



Varianz

Varianz

Für eine Zufallsvariable X mit $\mu = \mathbb{E}[X]$ definieren wir die *Varianz* $Var[X]$ durch

$$Var[X] := \mathbb{E}[(X - \mu)^2] = \sum_{x \in W_x} (x - \mu)^2 \cdot Pr[X = x]$$

Die Grösse $\sigma := \sqrt{Var[X]}$ heisst *Standardabweichung* von X

Wir sind eigentlich interessiert für die erwartete Abweichung vom Erwartungswert.

Varianz

Varianz und Erwartungswert

Für eine beliebige Zufallsvariable X gilt

$$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Beweis:

$$\begin{aligned}
\mathbb{E}[(X-\mu)^2] &= \mathbb{E}[X^2 - 2\mu X + \mu^2] && \text{(Linearität des } \mathbb{E}\text{)} \\
&= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 && (\mu = \mathbb{E}[X]) \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2
\end{aligned}$$

□

Varianz II

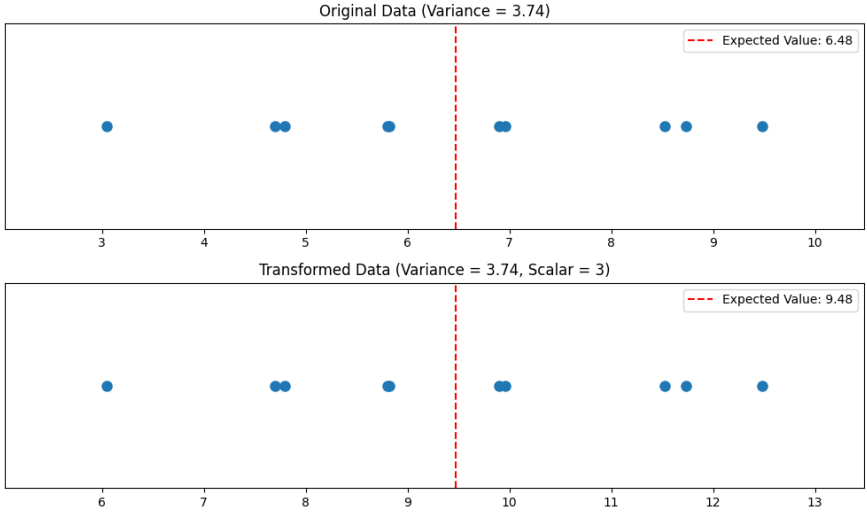
Satz 2.41.

Für eine beliebige Zufallsvariable X und $a, b \in R$ gilt

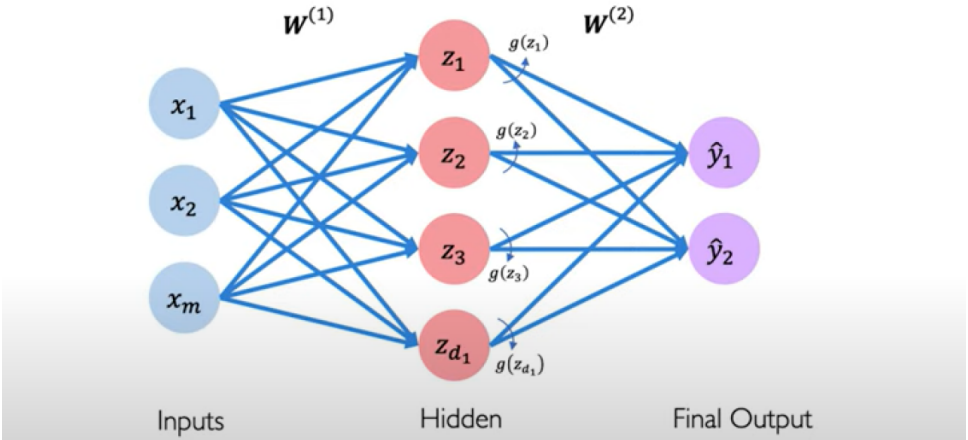
$$Var[a \cdot X + b] = a^2 \cdot Var[X]$$

Folgt aus beiden vorherigen Folien

Invariance of the variance under scalar addition



Bonus: Vanishing Gradients & Propagation of Variance in Deep Neural Networks



Variance Basics

- Variance of a random variable X:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

- For independent random variables X and Y:
  - $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$
  - $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
  - $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$
  - $\text{Var}[X \cdot Y] = \mathbb{E}[X^2]\mathbb{E}[Y^2] - (\mathbb{E}[X])^2(\mathbb{E}[Y])^2$
- If $\mathbb{E}[Y] = 0$, then:

$$\text{Var}[X \cdot Y] = \mathbb{E}[X^2] \cdot \text{Var}[Y]$$

ReLU Activation

- ReLU function: $v = \varphi(z) = \max(z, 0)$
- $z = \sum_{i=1}^{d} w_i x_i$
- Assume x_i are independent random variables with $\mathbb{E}[x_i^2] = 1$ (e.g., ensured by standardization).
- Let $w_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d.
- Then

$$\mathbb{E}[z] = \mathbb{E}[\sum w_i x_i] = \sum \mathbb{E}[w_i x_i] = \sum \mathbb{E}[w_i]\mathbb{E}[x_i] = 0$$

$$\mathrm{Var}(z) = \mathrm{Var}[\sum w_i x_i] = \sum \mathrm{Var}[w_i x_i] = \sum \mathrm{Var}[w_i]\mathbb{E}[x_i^2] = d \cdot \sigma^2$$

- Observation: z is symmetric, $p(z > 0) = p(z < 0) = \frac{1}{2}$

Variance Propagation

$$\begin{aligned}\mathbb{E}[v^2] &= \mathbb{E}[\max(z,0)^2] \\ &= p(z>0)\mathbb{E}[\max(z,0)^2|z>0] + p(z\leq 0)\mathbb{E}[\max(z,0)^2|z\leq 0]\end{aligned}$$

- $\mathbb{E}[v^2] = \frac{1}{2}d \cdot \sigma^2$
- $\sigma = \sqrt{\frac{2}{d}}$

Bernoulli-Verteilung

Bernouilli-Verteilte Zufallsvariablen

Eine Zufallsvariable X mit $W_X = \{0, 1\}$ und der Dichte

$$f_X(x) = \begin{cases} p & \text{für } x = 1, \\ 1 - p & \text{für } x = 0, \\ 0 & \text{sonst} \end{cases}$$

heisst bernouilli-verteilt wobei $p =$ Erfolgswahrscheinlichkeit.
Man notiert: $X \sim \mathit{Bernouilli}(p)$

- $\mathbb{E}[X] = p$
- $\mathit{Var}[X] = p(1 - p)$

Binomialverteilung

Binomialverteilung

Eine Zufallsvariable X mit $W_X = \{0, 1, \ldots, n\}$ und der Dichte

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x \in \{0, 1, \ldots, n\} \\ 0, & \text{sonst.} \end{cases}$$

heisst binomialverteilt. Man schreibt $X \sim Bin(n, p)$

- $\mathbb{E}[X] = np$
- $Var[X] = np(1-p)$

mehrere Bernouilli-verteilte Variablen

Geometrische Verteilung

so lange wiederholen bis man erfolgreich ist

Geometrische Verteilung

Wenn ein einzelner Versuch mit Wahrscheinlichkeit p gelingt, so ist die Anzahl der Versuche bis zum Erfolg geometrisch verteilt: $X \sim Geo(p)$ Für die Dichte und Verteilung einer geometrsich verteilten Zufallsvariable gilt:

$$f_X(i) = \begin{cases} p(1-p)^{i-1} & \text{für } i \in \mathbb{N}, \\ 0 & \text{sonst.} \end{cases}$$

(Vert) $F_X(n) = \Pr[X \leq n] = \sum_{i=1}^{n} \Pr[X = i] = \sum_{i=1}^{n} p(1-p)^{i-1} = 1 - (1-p)^n$.

- $\mathbb{E}[X] = \frac{1}{p}$
- $Var[X] = \frac{1-p}{p^2}$

Gedächtnislosigkeit

"Die Wahrscheinlichkeit, dass wir gleich im ersten Versuch 'Kopf' sehen, ist identisch zu der Wahrscheinlichkeit, dass wir nach 1000 Fehlversuchen beim 1001ten Wurf 'Kopf' erhalten. Die Tatsache, dass wir die Münze schon lange (aber erfolglos) geworfen haben, hat also keinerlei Einfluss auf den Erfolg im nächsten Wurf."

Gedächtnislosigkeit - Formal

Gedächtnislosigkeit

Ist $X \sim Geo(p)$, so gilt für alle $s, t \in \mathbb{N}$:

$$Pr[X \geq s + t \mid X > s] = Pr[X \geq t]$$

Beweis: $F_X(n) = 1 - (1-p)^n, n \in \mathbb{N}$. Dann gilt $Pr[X \geq n] = (1-p)^n$. Dann mithilfe der Def. von bedingten W'keit folgt:

$$\Pr[X \geq s + t \mid X > s] = \frac{\Pr[X \geq s + t]}{\Pr[X > s]} = \frac{(1-p)^{s+t-1}}{(1-p)^s} = (1-p)^{t-1} = \Pr[X \geq t],$$

[0] *Engl. Memorylessness*

Warten auf den n-ten Erfolg

Z bezeichnet die Anzahl der Versuche bis zum n-ten erfolgreichen Experiment. Wenn $Z = z$ ist, so wurden also genau n erfolgreiche und zn nicht erfolgreiche Experimente durchgeführt. Da nach Definition von Z das letzte Experiment erfolgreich sein muss, steht der Zeitpunkt des n-ten erfolgreichen Experimentes fest. Die übrigen $n - 1$ erfolgreichen Experimente können beliebig auf die restlichen $z - 1$ Experimente verteilt werden.

Warten auf den n-ten Erfolg

Warten auf den n-ten Erfolg

Falls Z **negativ binomialverteilt** für die Dichte gilt:

$$f_Z(z) = \binom{z-1}{n-1} \cdot p^n (1-p)^{z-n}.$$

- $\mathbb{E}[Z] = \frac{n}{p}$ (*Linearität des Erwartungswertes*)

Coupon-Collector-Problem: Bio-Twist

Problem: Coupon Collector - Bio Twist

There is a new virus that gets a random mutation every time a new person is infected. We know that there exists only n many possible mutations. How many patients (including patient 0) must get infected on average such that we can see all the mutations and develop medication for them?

Coupon-Collector-Porblem: Bio-Twist - II

Coupon Collector - Bio Twist

Teile den ganzen Prozess in Phasen. In der Phase i, hatte das Virus schon $i-1$ Mutationen.

- $X :=$ Anzahl Infektionen insgesamt
- Sei $X_i :=$ Anzahl Infektionen in der Phase i

Also gilt $X = \sum_{i=1}^{n} X_i$ Wir wissen auch, dass $X_i \sim Geo(p_i)$ Wir wollen einer der $n-i+1$ Mutationen sehen, die noch nicht aufgetacuht sind. Somit erhalten wir

$$p_i = \frac{n-i+1}{n} \text{ und } \mathbb{E}[X_i] = \frac{n}{n-i+1}$$

$$\mathbb{E}[X] \overset{Lin.des\mathbb{E}}{=} \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \cdot \sum_{i=1}^{n} \frac{1}{i} = n \cdot H_n,$$

Coupon-Collector-Porblem: Bio-Twist - III

Wir wissen $H_n := \sum_{i=1}^{n} \frac{1}{i} = \ln n + O(1)$ also folgt

$$\mathbb{E}[X] = n \ln n + O(n)$$

Fazit: $n \ln n + O(n)$ Menschen im Durschschnitt müssen infiziert werden damit wir alle Mutationen beobachten können.

Poisson-Verteilung

Poisson-Verteilung

Poisson-Verteilung mit Parameter λ ist definiert durch die Dichtefunktion

$$f_X(i) = \begin{cases} \frac{e^{-\lambda}\lambda^i}{i!} & \text{für } i \in \mathbb{N}_0 \\ 0 & \text{sonst.} \end{cases}$$

Es gilt:

$$\mathbb{E}[X] = Var[X] = \lambda$$

Man kann die Poisson-Verteilung als Grenzwert einer Bionomialverteilung erhaletn.

Bedingte Zufallsvariable

Bedingte Zufallsvariable

Sei X Zufallsvariable auf W'raum Ω, $A \subseteq \Omega$ Ereignis mit $\Pr[A] > 0$. Die bedingte Zufallsvariable $X|A$ ist dieselbe Funktion wie X, aber der Definitionsbereich ist auf die Menge A eingeschränkt.

$$f_{X|A} : \mathbb{R} \to [0,1], \quad x \mapsto \Pr[X = x \mid A]$$

X ist unabhängig von A, falls $f_{X|A} = f_X$.

$$F_{X|A} : \mathbb{R} \to [0,1], \quad x \mapsto \Pr[X \leq x \mid A]$$

$$\mathbb{E}[X \mid A] := \sum_{x \in W_X} x \cdot \Pr[X = x \mid A] = \frac{1}{\Pr[A]} \sum_{\omega \in A} X(\omega) \cdot \Pr[\omega]$$

Erinnerung: $X = x$ ist auch ein Ereignis, also $''X = x \subseteq \Omega''$

Gemeinsame Dichten und Randdichten

Gemeinsame Dichte

$$f_{X,Y}(x,y) = \Pr[X = x, Y = y]$$

"Randdichte von X" (= Dichte von X)

$$f_X(x) = \sum_{y \in W_Y} f_{X,Y}(x,y) \quad \text{(nach dem Additionssatz)}$$

Unabhängigkeit von Zufallsvariablen

Unabhängigkeit von Zufallsvariablen

Die Ereignisse $X_1 = x_1, \ldots, X_n = x_n$ heißen unabhängig, wenn für alle $x_1, \ldots, x_n$ die Zufallsvariablen $X_1, \ldots, X_n$ unabhängig sind, dh.

$$\Pr[X_1 = x_1, \ldots, X_n = x_n] = \Pr[X_1 = x_1] \cdots \Pr[X_n = x_n].$$

Aber Viele der Gleichungen sind redundant. Siehe Beispiel in Vorlesungsslides.
Alternativ:

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

für alle $(x_1, \ldots, x_n) \in W_{X_1} \times \cdots \times W_{X_n}$

Exercises

Quiz auf der Webseite!

The End

♠