Algorithms and Probability

Week 7

Georg Hasebe — 11/04/2024

Exercise Sheet 3

Theory Recap

Dichte-/Verteilungsfunktion

Sei X eine Zufallsvariable.

Dichte(funktion) von X:

$$f_X\colon \mathbb{R} \to [0,1], \quad x \mapsto \Pr[X = x]$$

Verteilungsfunktion von X:

$$F_X\colon \mathbb{R} \to [0,1], \quad x \mapsto \Pr[X \leq x]$$

Bernoulli-Verteilung

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

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

heisst **bernoulli-verteilt**. Wir schreiben: $X \sim \text{Ber}(p)$.

Nach Definition von f_X gilt also $\Pr[X = 0] = 1 - p$ und $\Pr[X = 1] = p$.

Es gilt: $\mathbb{E}[X] = p$.

Beispiele

- Werfen einer Münze, Indikator für Kopf
- Werfen eines Würfels, Indikator für “Augenzahl gerade”
- Werfen eines Würfels, Indikator für “1”

Indikatorvariablen

Beobachtung 2.35. 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 sind Bernoulli Variablen mit $p = \Pr[A]$.

Binomialverteilung

Bernoulli-verteilte Zufallsvariable erhalten wir zum Beispiel als Indikator für "Kopf", wenn wir eine Münze einmal werfen.

Werfen wir die Münze n-mal und fragen wie oft wir Kopf erhalten haben, so ist die entsprechende Zufallsvariable **binomialverteilt**. Wir schreiben: $X \sim \text{Bin}(n, p)$.

Für die **Dichtefunktion** einer **binomialverteilten** Zufallsvariable X gilt:

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

Es gilt: $\mathbb{E}[X] = np$

Beispiel (Blackboard)

Fünf faire Münzen werden geworfen. Sei X die Zufallsvariable, die zählt, wie oft Kopf erscheint. Bestimmen Sie den Wert $f_X(3)$ und erklären Sie, wie sie auf diesen Wert gekommen sind.

Geometrische Verteilung

Man stelle sich ein Experiment vor, bei dem eine Aktion so lange wiederholt wird, bis sie erfolgreich ist (z.B. bis das erste mal Kopf kommt).

Wenn ein einzelner Versuch mit (Erfolgs-)Wahrscheinlichkeit p gelingt, so ist die Anzahl der Versuche bis zum Erfolg **geometrisch verteilt** und man schreibt $X \sim \text{Geo}(p)$.

Für die Dichtefunktion einer **geometrisch verteilten** Zufallsvariable X gilt:

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

Es gilt: $\mathbb{E}[X] = 1/p$.

Beispiel (Blackboard)

Eine Urne enthält N weisse und M schwarze Bälle. Es werden zufällig Bälle aus der Urne gezogen, bis ein schwarzer Ball gezogen wird. Ein gezogener Ball wird vor dem nächsten Zug ersetzt (gleiche Farbe).

Was ist die Wahrscheinlichkeit, dass genau n Züge notwendig sind?

(*) Geometric Series

- $\displaystyle\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}$ for $|r| < 1$.
- $\displaystyle\sum_{k=1}^{n} ar^{k-1} = \begin{cases} a\left(\frac{1-r^n}{1-r}\right) & r \neq 1 \\ an & \text{otherwise} \end{cases}$

see Wikipedia.

please be careful with indices!

(*) not part of the lecture; might be worth knowing.

(*) Beispiel: Geometric Series

$X \sim \text{Geo}(p)$

$$F_X(u) = \Pr[X \leq n] = \sum_{k=1}^{n} \Pr[X = k]$$

$$= \sum_{k=1}^{n} p(1-p)^{k-1} = p \sum_{k=1}^{n} (1-p)^{k-1}$$

$$= p \frac{1-(1-p)^k}{1-(1-p)} = 1-(1-p)^k .$$

- $\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}$ for $|r| < 1.$
- $\sum_{k=1}^{n} ar^{k-1} = \begin{cases} a\left(\frac{1-r^n}{1-r}\right) & r \neq 1 \\ an & \text{otherwise} \end{cases}$

please be careful with indices!

see Wikipedia.

(*) not part of the lecture; might be worth knowing.

Gedächtnislosigkeit

Satz 2.45. Ist $X \sim \mathrm{Geo}(p)$, so gilt für alle $s, t \in \mathbb{N}$:

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

Beispiel: Wahrscheinlichkeit im ersten Wurf Kopf zu bekommen ist identisch zur Wahrscheinlichkeit nach 1000 Fehlversuchen im 1001ten Wurf Kopf zu bekommen

Negative Binomialverteilung

Sei Z die Anzahl der Versuche bis zum n-ten Erfolg.

Für $n = 1$ haben wir $Z \sim \text{Geo}(p)$.

Für $n \geq 2$: Z ist **negativ binomialverteilt** (mit Ordnung n).

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

Es gilt: $\mathbb{E}[X] = n/p$.

Poisson-Verteilung

$$X \sim \mathrm{Po}(\lambda)$$

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

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

$\boxed{\text{Bin}(n,\lambda/n) \quad \text{konvergiert für } n \to \infty \quad \text{gegen } \mathrm{Po}(\lambda)}$

"Hier denke man zum Beispiel an die Möglichkeit, dass ein Bürger in der nächsten Stunde einen Herzinfarkt bekommt.
Für den Einzelnen ist dies unwahrscheinlich, schweizweit gibt es aber dennoch etwa drei bis vier pro Stunde."
see Skript p. 127

Coupon Collector

Szenario: es gibt n verschiedene Bilder
in jeder Runde erhalten wir (gleichw'lich) eines der Bilder

X := Anzahl Runden bis wir alle n Bilder besitzen

Frage: E[X] = ??

cont'd on the blackboard.

Betrachten Sie das Coupon-Collector-Problem mit nur zwei verschiedenen Coupons ($n = 2$). Die erwartete Anzahl Runden bis wir beide Coupons gesammelt haben ist 3.

Beispiel

Aufgabe 3 – *Fische im Teich*

Sie haben einen Teich mit n Lachsen und n Forellen. Sie haben einen Kescher, mit dem Sie Fische aus dem Teich fangen wollen. Jedes Mal wenn Sie einen Fisch fangen, fangen Sie einen Fisch zufällig gleichverteilt aus allen noch im Teich vorhandenen Fischen. Sie wollen nun für eine Grillparty alle Lachse fangen, um diese zu grillen. Da Sie die Forellenpopulation in Ihrem Teich erhalten wollen, werfen Sie jedes Mal, wenn Sie eine Forelle gefangen haben, diese zurück in den Teich (die Tiere leiden hierunter nicht!). Zeigen Sie, dass Sie in der Erwartung $\sum_{i=1}^{n} \frac{n+i}{i}$ Fische fangen müssen, um alle Lachse zu fangen. *(5 Punkte)*

siehe AlgoWahr FS21

$5/45 \approx 11\,\%$

Bedingte Zufallsvariable

Sei X eine Zufallsvariable und A ein Ereignis mit $\Pr[A] > 0$.

Die **bedingte Zufallsvariable** $X|A$ ist *dieselbe Funktion* wie X, aber der Definitionsbereich ist auf die Menge A eingeschränkt.

$$X|A \colon A \to \mathbb{R}.$$

Bedingte Zufallsvariable

Dichtefunktion

$$f_{X|A}: \mathbb{R} \to [0,1], \quad x \mapsto \Pr[X = x \,|\, A] = \frac{\Pr[\{\omega \in A : X(\omega) = x\}]}{\Pr[A]}$$

Verteilungsfunktion

$$F_{X|A}: \mathbb{R} \to [0,1], \quad x \mapsto \Pr[X \leq x \,|\, A] = \frac{\Pr[\{\omega \in A : X(\omega) \leq x\}]}{\Pr[A]}$$

Erwartungswert

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

Bedingte Zufallsvariable

X ist unabhängig von A, falls:

$$f_{X|A} = f_X.$$

Es gilt also:

$$\Pr[X = x \,|\, A] = \frac{\Pr[\{\omega \in A : X(\omega) = x\}]}{\Pr[A]} = \Pr[X = x]$$

Beispiel

$X_1 :=$ "Augenzahl des ersten Würfels"

$X_2 :=$ "Augenzahl des zweiten Würfels"

$X := X_1 + X_2 =$ "Summe der beiden Augenzahlen"

$A :=$ "X_2 ist gerade"

Berechnen Sie $\mathbb{E}[X_1], \mathbb{E}[X_2], \mathbb{E}[X \mid A]$.

Zur Erinnerung:

Dichtefunktion

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

Verteilungsfunktion

$$F_{X|A}: \mathbb{R} \to [0,1], \quad x \mapsto \Pr[X \leq x \mid A] = \frac{\Pr[\{\omega \in A : X(\omega) \leq x\}]}{\Pr[A]}$$

Erwartungswert

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

cont'd on the blackboard.

Mehrere Zufallsvariablen

Mehrere Zufallsvariablen zugleich?

$$\Pr[X = x, Y = y] = \Pr[\{\omega \in \Omega \mid X(\omega) = x, Y(\omega) = y\}]$$

Weiters haben wir:

- $\Pr[X = x, Y = y]$ ist als Abkürzung von $\Pr[X = x \cap Y = y]$ zu verstehen
- gemeinsame Dichte:

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

Mehrere Zufallsvariablen

Randdichte von X (= Dichte von X)

$$f_X(x) = \sum_{y \in W_Y} f_{X,Y}(x, y)$$

Good to know!

Beweis. Die Ereignisse "$Y = y$" bilden eine **disjunkte Zerlegung** des Wahrscheinlichkeitsraumes.

$$\Rightarrow \quad f_X(x) \overset{\text{Def.}}{=} \Pr[X = x] \overset{2.13}{=} \sum_{y \in W_Y} \Pr[X = x, Y = y]\,.$$

Beispiel

$X :=$ "Augenzahl des Würfels"

$Y :=$ "Anzahl Kopf bei X-mal Münze geworfen"

Bestimmen Sie $f_{X,Y}$.

Bestimmen Sie $f_Y(6)$. Benutzen Sie $f_{X,Y}$.

Zur Erinnerung:

Mehrere Zufallsvariablen zugleich?

$$\Pr[X = x, Y = y] = \Pr[\{\omega \in \Omega \mid X(\omega) = x, Y(\omega) = y\}]$$

Weiters haben wir:

- $\Pr[X = x, Y = y]$ ist als Abkürzung von $\Pr[X = x \cap Y = y]$ zu verstehen
- gemeinsame Dichte:

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

cont'd on the blackboard.

Unabhängigkeit von mehreren Zufallsvariablen

Definition 2.52. Zufallsvariablen $X_1, \ldots, X_n$ heissen unabhängig genau dann, wenn für alle $(x_1, \ldots, x_n) \in W_{X_1} \times \ldots \times W_{X_n}$ gilt

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

Alternativ kann man Definition 2.52 auch über die Dichten formulieren: $X_1, \ldots, X_n$ sind genau dann unabhängig, wenn für alle $(x_1, \ldots, x_n) \in W_{X_1} \times \ldots \times W_{X_n}$ gilt

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

siehe Skript S. 132.

Beispiel

Sei $\Omega = \{2,3\}$ mit $\Pr[\omega] = 1/2$ für alle $\omega \in \Omega$.

Bestimmen Sie die Indikatorvariablen X, Y wobei X bzw. Y genau dann 1 ist, wenn ω durch 2 bzw. 3 teilbar ist.

Bestimmen Sie f_X, f_Y.

Sind X und Y voneinander unabhängig?

Zur Erinnerung:

Definition 2.52. Zufallsvariablen $X_1, \ldots, X_n$ heissen unabhängig genau dann, wenn für alle $(x_1, \ldots, x_n) \in W_{X_1} \times \ldots \times W_{X_n}$ gilt

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

Alternativ kann man Definition 2.52 auch über die Dichten formulieren: $X_1, \ldots, X_n$ sind genau dann unabhängig, wenn für alle $(x_1, \ldots, x_n) \in W_{X_1} \times \ldots \times W_{X_n}$ gilt

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

cont'd on the blackboard.

Unabhängigkeit: Indikatorvariablen

Für zwei **Indikator**variablen X und Y gilt

X und Y sind **unabhängig** $\iff$ $f_{X,Y}(1,1) = f_X(1) \cdot f_Y(1)$

Beweis. Von **Übungsblatt 3** wissen wir, dass wenn A und B unabhängig sind, auch A und $\bar{B}$, $\bar{A}$ und B und $\bar{A}$ und $\bar{B}$ unabhängig sind.

Summe von Zufallsvariablen

Satz 2.58. Für zwei unabhängige Zufallsvariablen X und Y sei $Z := X + Y$. Es gilt

$$f_Z(z) = \sum_{x \in W_X} f_X(x) \cdot f_Y(z - x).$$

Beweis. Mit Hilfe des Satzes von der totalen Wahrscheinlichkeit folgt, dass

$$\begin{aligned} f_Z(z) &= \Pr[Z = z] = \sum_{x \in W_X} \Pr[X + Y = z \mid X = x] \cdot \Pr[X = x] \\ &= \sum_{x \in W_X} \Pr[Y = z - x] \cdot \Pr[X = x] = \sum_{x \in W_X} f_X(x) \cdot f_Y(z - x). \end{aligned}$$

$\square$

Bemerkung: Die Ereignisse “$X = x$” bilden eine **disjunkte Zerlegung** des Wahrscheinlichkeitsraumes.

Summe von Zufallsvariablen

Satz 2.58. Für zwei unabhängige Zufallsvariablen X und Y sei $Z := X + Y$. Es gilt

$$f_Z(z) = \sum_{x \in W_X} f_X(x) \cdot f_Y(z - x).$$

Folgerungen:

Poisson(λ_1) + Poisson(λ_2) = Poisson(λ_1+λ_2)

Bin(n, p) + Bin(m, p) = Bin(n+m, p)

Waldsche Identität

Satz 2.65 (Waldsche Identität). N und X seien zwei unabhängige Zufallsvariable, wobei für den Wertebereich von N gilt: $W_N \subseteq \mathbb{N}$. Weiter sei

$$Z := \sum_{i=1}^{N} X_i,$$

wobei $X_1, X_2, \ldots$ unabhängige Kopien von X seien. Dann gilt:

$$\mathbb{E}[Z] = \mathbb{E}[N] \cdot \mathbb{E}[X].$$

Proof on the blackboard.

Beispiel

$X :=$ "Augenzahl des Würfels"

$Y :=$ "Anzahl Kopf bei X-mal Münze geworfen"

Was ist $\mathbb{E}[Y]$?

Zur Erinnerung:

Satz 2.65 (Waldsche Identität). N und X seien zwei unabhängige Zufallsvariable, wobei für den Wertebereich von N gilt: $W_N \subseteq \mathbb{N}$. Weiter sei

$$Z := \sum_{i=1}^{N} X_i,$$

wobei $X_1, X_2, \ldots$ unabhängige Kopien von X seien. Dann gilt:

$$\mathbb{E}[Z] = \mathbb{E}[N] \cdot \mathbb{E}[X].$$

cont'd on the blackboard.

Varianz

Es gibt Zufallsvariablen, die niemals Werte in der Grössenordnung des Erwartungswerts annehmen.

Beispiel: $\Pr[X = -10^6] = \Pr[X = 10^6] = 1/2$ mit $\mathbb{E}[X] = 0$.

Mass für die Abweichung vom Erwartungswert? Die **Varianz** $\mathrm{Var}[X]$.

Varianz

Definition 2.39. Für eine Zufallsvariable X mit $\mu = \mathbb{E}[X]$ definieren wir die *Varianz* $\mathrm{Var}[X]$ durch

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

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

Satz 2.40. Für eine beliebige Zufallsvariable X gilt

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

Satz 2.41. Für eine beliebige Zufallsvariable X und $a, b \in \mathbb{R}$ gilt

$$\mathrm{Var}[a \cdot X + b] = a^2 \cdot \mathrm{Var}[X].$$

both Satz 2.40 and Satz 2.41 can be proven using just Definition 2.39.

Varianz

Definition 2.39. Für eine Zufallsvariable X mit $\mu = \mathbb{E}[X]$ definieren wir die *Varianz* $\mathrm{Var}[X]$ durch

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

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

Satz 2.40. Für eine beliebige Zufallsvariable X

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

Satz 2.41. Für eine beliebige Zufallsvariable X und $a, b \in \mathbb{R}$ gilt

$$\mathrm{Var}[a \cdot X + b] = a^2 \cdot \mathrm{Var}[X].$$

ETH Zürich
Institute of Theoretical Computer Science
Dr. Johannes Lengler, Prof. Emo Welzl
Dr. Anders Martinsson, Charlotte Knierim

FS 2021

Keine Garantie! Selbst nachsehen!

Algorithmen und Wahrscheinlichkeit
Formelsammlung

Varianz

- **Definition:** $\mathrm{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
- **Translation:** Für $a, b \in \mathbb{R}$ gilt $\mathrm{Var}[a \cdot X + b] = a^2 \cdot \mathrm{Var}[X]$.
- **Standardabweichung:** $\sigma[X] := \sqrt{\mathrm{Var}[X]}$.
- **Additivität:** Für unabhängige $X_1, \ldots, X_n$ gilt $\mathrm{Var}[X_1 + \ldots + X_n] = \mathrm{Var}[X_1] + \ldots + \mathrm{Var}[X_n]$.

Wichtige Rechenregeln

$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$	$\forall X, Y$	siehe Satz 2.60 im Skript.
$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$	$\forall X, Y$ unabhängig	siehe Satz 2.61 im Skript.
$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$	$\forall X, Y$ unabhängig	siehe Satz 2.62 im Skript.
$\mathrm{Var}[X \cdot Y] \neq \mathrm{Var}[\mathrm{X}] \cdot \mathrm{Var}[Y]$	i.A. (auch für unabhängige ZV)	

Ungeleichung von Markov

Satz 2.67. *(Ungleichung von Markov)* Sei X eine Zufallsvariable, die nur nicht-negative Werte annimmt. Dann gilt für alle $t \in \mathbb{R}$ mit $t > 0$, dass
$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

Beweis. Wir rechnen direkt nach, dass

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x \in W_X} x \cdot \Pr[X = x] \geq \sum_{x \in W_X,\, x \geq t} x \cdot \Pr[X = x] \\ &\geq t \cdot \sum_{x \in W_X,\, x \geq t} \Pr[X = x] = t \cdot \Pr[X \geq t].\end{aligned}$$

Ungleichung von Chebyshev

Satz 2.68. *(Ungleichung von Chebyshev)* Sei X eine Zufallsvariable und $t \in \mathbb{R}$ mit $t > 0$. Dann gilt

$$\Pr[|X - \mathbb{E}[X]| \geq t] \;\leq\; \frac{\mathrm{Var}[X]}{t^2}$$

Beweis. Es gilt

$$\Pr[|X - \mathbb{E}[X]| \geq t] \;=\; \Pr[(X - \mathbb{E}[X])^2 \geq t^2].$$

Die Zufallsvariable $Y := (X - \mathbb{E}[X])^2$ ist nicht-negativ und hat nach Definition der Varianz den Erwartungswert $\mathbb{E}[Y] = \mathrm{Var}[X]$. Damit folgt die Behauptung durch Anwendung der Markov-Ungleichung:

$$\Pr[|X - \mathbb{E}[X]| \geq t] \;=\; \Pr[Y \geq t^2] \;\leq\; \frac{\mathbb{E}[Y]}{t^2} \;=\; \frac{\mathrm{Var}[X]}{t^2}.$$

□