Algorithmen und Wahrscheinlichkeit

Verteilungen

Bernoulli-Verteilung

$$X \sim \mathrm{Bernoulli}(p)$$

Dichtefunktion:

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

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

Beispiele: Werfen einer Münze, Indikator für Kopf
Werfen eines Würfels, Indikator für "Augenzahl gerade"
Werfen eines Würfels, Indikator für "1"

Binomialverteilung

$$X \sim \mathrm{Bin}(n, p).$$

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

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

Beispiel: Werfen einer Münze n mal, X = Anzahl Kopf

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

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

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

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

Beispiel: Modellierung seltener Ereignisse, zum Beispiel
X:= Anzahl Herzinfarkte in der Schweiz in der nächsten Stunde

Geometrische Verteilung

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

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

$$F_X(n) = 1 - (1-p)^n \qquad \text{für alle } n=1,2,\ldots$$

$$\mathbb{E}[X] = 1/p$$

Beispiel: Wiederholtes Werfen einer Münze,
X = Anzahl Würfe bis zum ersten Mal Kopf

Negative Binomialverteilung

$$X \sim \text{NegativeBinomial}(n,p)$$

$$f_X(k) = \begin{cases} \binom{k-1}{n-1}(1-p)^{k-n}p^n, & \text{für } k = 1, 2, \ldots \\ 0, & \text{sonst} \end{cases}$$

$$\mathbb{E}[X] = n/p$$

Beispiel: Warten auf den n-ten Erfolg

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] = ??

X_i := Anzahl Runden in Phase i, $X_i \sim \mathrm{Geo}(\frac{n-(i-1)}{n})$

$X = \sum_{i=1}^{n} X_i$

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

$$H_n = \ln n + O(1)$$

Geometrische Verteilung

Wichtige Eigenschaft: Gedächtnislosigkeit

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

$$\Pr[X \overset{=}{\cancel{\geq}} s + t \mid X > s] = \Pr[X \overset{=}{\cancel{\geq}} t].$$

Beweis: nachrechnen

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

Kapitel 2.6

Mehrere Zufallsvariablen

Dichte- und Verteilungsfunktion, Erwartungswert

Ω Wahrscheinlichkeitsraum

Zufallsvariable: $X : \Omega \to \mathbb{R}$

Dichtefunktion

$$f_X : \mathbb{R} \to [0,1], \quad x \mapsto \Pr[X = x] \qquad (= \Pr[\{\omega : X(\omega) = x\}])$$

Verteilungsfunktion

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

Erwartungswert

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

Bedingte Zufallsvariable

Ω Wahrscheinlichkeitsraum

Zufallsvariable: $X : \Omega \to \mathbb{R}$

$A \subseteq \Omega$ Ereignis mit $\Pr[A] > 0$

bedingte W'keit: $\Pr[\,\cdot\,|A]$

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

Zufallsvariable: $X|A : A \to \mathbb{R}$

Bedingte Zufallsvariable

Ω Wahrscheinlichkeitsraum, $A \subseteq \Omega$ Ereignis mit $\Pr[A] > 0$

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

Zufallsvariable: $X|A : A \to \mathbb{R}$

Dichtefunktion

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

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

Verteilungsfunktion

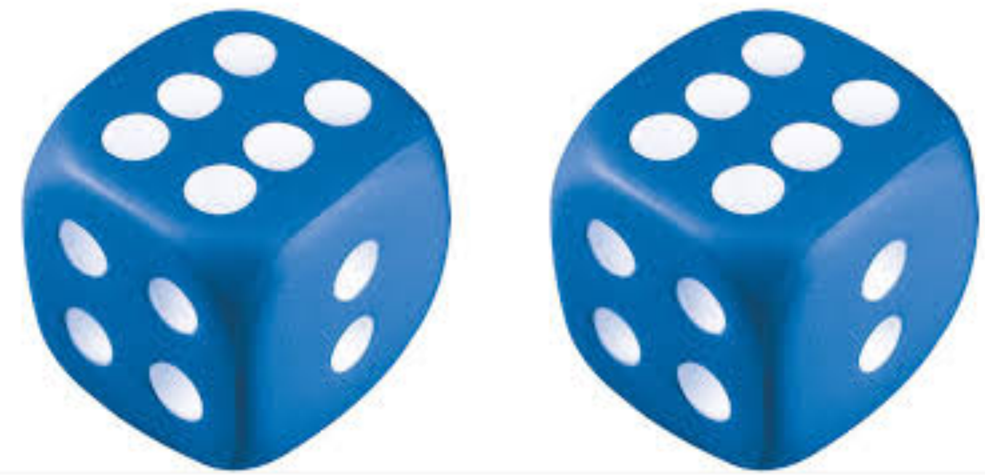
$$F_{X|A} : \mathbb{R} \to [0,1], \quad x \mapsto \Pr[X \leq x \mid A]$$

Erwartungswert

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

Beispiel

X_1 := Augenzahl des ersten Würfels
X_2 := Augenzahl des zweiten Würfels
X := Summe der Augenzahlen = $X_1 + X_2$



$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = 3.5$$
$$\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 3.5 + 3.5 = 7$$

A := "X_2 ist gerade"

$$\mathbb{E}[X \mid A] = \mathbb{E}[X_1 \mid A] + \mathbb{E}[X_2 \mid A]$$
$$= \mathbb{E}[X_1] + \sum_{i \in \{2,4,6\}} i \cdot \Pr[X_2 = i \mid A] = 3.5 + 4 = 7.5$$

Aber: $X = X_1 + X_2$ ist **nicht unabhängig** von B.

Dazu müssten X_1, X_2, B unabhängig sein. Sie sind aber nur *paarweise* unabhängig!

B := "X ist gerade"

$$\mathbb{E}[X \mid B] = \sum_{i \in \{2,4,6,8,10,12\}} i \cdot \Pr[X = i \mid B] = \ldots$$

X_1 ist unabhängig von B, denn für alle i ist $\Pr[B \mid X_1 = i] = \Pr[B] = 1/2$.

$\Rightarrow$ Es ist auch $\Pr[X_1 = i \mid B] = \Pr[X_1 = i]$. Ebenso für X_2.

$$\Rightarrow \mathbb{E}[X \mid B] = \mathbb{E}[X_1 \mid B] + \mathbb{E}[X_2 \mid B] = 3.5 + 3.5 = 7$$

Mehrere Zufallsvariablen

   ... 

X_1 Indikator für Kopf im 1ten Wurf

X_2 Indikator für Kopf im 2ten Wurf

X_n Indikator für Kopf im nten Wurf

$X = X_1 + X_2 + \ldots + X_n$ = Anzahl Kopf

Beispiel

- werfe zunächst einen Würfel

  X := Augenzahl des Würfels

- werfe danach X-mal eine Münze

  Y := Anzahl Kopf

Beispiel

- werfe zunächst einen Würfel

X := Augenzahl des Würfels

- werfe danach X-mal eine Münze

Y := Anzahl Kopf

$$\Pr[X = x, Y = y] = \begin{cases} \frac{1}{6} \cdot \frac{\binom{x}{y}}{2^x} & x \in \{1,2,3,4,5,6\} \text{ and } y \in \{0,1,\dots,x\} \\ 0, & x \notin \{1,2,3,4,5,6\} \text{ or } y \notin \{0,1,\dots,x\} \end{cases}$$

$$\Pr[Y = 3] = \sum_{x=1}^{6} \Pr[X = x, Y = 3] = \frac{1}{6} \cdot \left(0 + 0 + \frac{1}{8} + \frac{4}{16} + \frac{10}{32} + \frac{20}{64}\right) = \frac{1}{6}$$

X, Y Zufallsvariablen

gemeinsame Dichte

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

“Randdichte von X” (= Dichte von X)

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

(nach dem Additionssatz)

Unabhängigkeit von Zufallsvariablen

Definition (Version 1):

$x_1 \in W_{X_1}, \ldots, x_n \in W_{X_n}$

$X_1,\ldots,X_n$ heissen unabhängig, wenn **für alle** $\cancel{x_1, \ldots, x_n \in \mathbb{R}}$ die Ereignisse "$X_1 = x_1$", "$X_2 = x_2$", …, "$X_n = x_n$" unabhängig sind, d.h.

$$\underbrace{\Pr[X_1 = x_1, \ldots, X_n = x_n]}_{=f_{X_1,\ldots,X_n}(x_1,x_2,\ldots,x_n)} = \underbrace{\Pr[X_1 = x_1]}_{=f_{X_1}(x_1)} \cdot \ldots \cdot \underbrace{\Pr[X_n = x_n]}_{=f_{X_n}(x_n)},$$

~~und genauso für jede Teilmenge von Indizes!~~

Aber: Viele der Gleichungen sind *redundant*.

Beispiel für n = 3: Falls X_3 Wertebereich {0,1} hat, dann folgt aus

$$\Pr[X_1 = x_1, X_2 = x_2, X_3 = 0] = \Pr[X_1 = x_1] \cdot \Pr[X_2 = x_2] \cdot \Pr[X_3 = 0], \text{ und}$$
$$\Pr[X_1 = x_1, X_2 = x_2, X_3 = 1] = \Pr[X_1 = x_1] \cdot \Pr[X_2 = x_2] \cdot \Pr[X_3 = 1]$$

$$\begin{aligned}
\Pr[X_1 = x_1, X_2 = x_2] &= \Pr[X_1 = x_1, X_2 = x_2, X_3 = 0] + \Pr[X_1 = x_1, X_2 = x_2, X_3 = 1] \\
&= \Pr[X_1 = x_1] \cdot \Pr[X_2 = x_2] \cdot \Pr[X_3 = 0] + \Pr[X_1 = x_1] \cdot \Pr[X_2 = x_2] \cdot \Pr[X_3 = 1] \\
&= \Pr[X_1 = x_1] \cdot \Pr[X_2 = x_2] \cdot (\Pr[X_3 = 0] + \Pr[X_3 = 1]) \\
&= \Pr[X_1 = x_1] \cdot \Pr[X_2 = x_2].
\end{aligned}$$

Unabhängigkeit von Zufallsvariablen

Definition (Version 2) 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:

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

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

Unabhängigkeit: Beispiel

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

$$X(\omega) = \begin{cases} 1 & \text{wenn } \omega \text{ durch } 2 \text{ teilbar} \\ 0 & \text{sonst} \end{cases} \qquad f_X(i) = \begin{cases} 1/2 & \text{für } i=0,1 \\ 0 & \text{sonst} \end{cases}$$

$$Y(\omega) = \begin{cases} 1 & \text{wenn } \omega \text{ durch } 3 \text{ teilbar} \\ 0 & \text{sonst} \end{cases} \qquad f_Y(i) = \begin{cases} 1/2 & \text{für } i=0,1 \\ 0 & \text{sonst} \end{cases}$$

$f_{X,Y}(1,1) = 0 \neq f_X(1) \cdot f_Y(1)$

$\Rightarrow$ X und Y sind **nicht** unabhängig

Unabhängigkeit: Beispiel

$\Omega = \{1,2,3,6\}$ mit $\Pr[\omega] = 1/4$ für alle $\omega \in \Omega$

$$X(\omega) = \begin{cases} 1 & \text{wenn } \omega \text{ durch } 2 \text{ teilbar} \\ 0 & \text{sonst} \end{cases} \qquad f_X(i) = \begin{cases} 1/2 & \text{für } i=0,1 \\ 0 & \text{sonst} \end{cases}$$

$$Y(\omega) = \begin{cases} 1 & \text{wenn } \omega \text{ durch } 3 \text{ teilbar} \\ 0 & \text{sonst} \end{cases} \qquad f_Y(i) = \begin{cases} 1/2 & \text{für } i=0,1 \\ 0 & \text{sonst} \end{cases}$$

$$f_{X,Y}(1,1) = 1/4 = f_X(1) \cdot f_Y(1)$$

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: 1. Möglichkeit:
-> nachrechnen, dass aus $f_{X,Y}(1,1) = f_X(1)\, f_Y(1)$ auch die übrigen drei Glg folgen
2. Möglichkeit:
-> verwende, dass aus A,B unabhängig auch A,$\bar{B}$ unabhängig folgt, und analog für $\bar{A}$,B und $\bar{A}$,$\bar{B}$

Achtung: Für drei und mehr Variablen gilt analoges **nicht**

Summe von Zufallsvariablen

Satz 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: Additionssatz und Definition der Unabhängigkeit, siehe Skript

Satz 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($\lambda_1+\lambda_2$)

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

Waldsche Identität

- werfe zunächst einen Würfel

  X := Augenzahl des Würfels

- werfe danach X-mal eine Münze

  Y := Anzahl Kopf

E[Y] = ??

Satz 2.64 (Waldsche Identität). N und X seien zwei unabhängige Zufallsvariable, wobei für den Wertebereich von N gelte: $W_N \subseteq \mathbb{N}$. Weiter sei

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

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

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

Beweis:

$$\mathbb{E}[Z] = \sum_{n \in W_N} \mathbb{E}[Z \mid N = n] \cdot \Pr[N = n]$$

$$\mathbb{E}[Z \mid N = n] = \mathbb{E}[X_1 + \ldots + X_n] = n \cdot \mathbb{E}[X],$$

$$\mathbb{E}[Z] = \sum_{n \in W_N} n \cdot \mathbb{E}[X] \cdot \Pr[N = n] = \mathbb{E}[X] \sum_{n \in W_N} n \cdot \Pr[N = n] = \mathbb{E}[X] \cdot \mathbb{E}[N].$$

- werfe zunächst einen Würfel

X := Augenzahl des Würfels

- werfe danach eine Münze und zwar so oft wie die Augenzahl des Würfels angibt

Y := Anzahl Kopf

- aus Waldscher Identität folgt: $E[Y] = E[X]*1/2 = 7/4$

2. Beispiel:

- wir werfen ein faire Münze bis zum ersten Mal Kopf kommt,
  N := Anzahl Würfe

  1.Phase

- wir werfen die Münze nochmals N mal
  Z := Anzahl Kopf insgesamt

  2.Phase

1.Phase: 1 mal Kopf
2.Phase: Waldsche Identität

N ~ Geo(1/2)

… gibt an wie oft wir das Experiment wiederholen

X ~ Bernoulli(1/2)

… in jedem Experiment mit W'keit 1/2 Kopf

$$\mathbb{E}[\text{Anzahl Kopf in 2. Phase}] = \mathbb{E}[N] \cdot \mathbb{E}[X] = 2 \cdot \frac{1}{2} = 1$$