# 6. Vorlesung: Diskrete Zufallsvariablen

Nikolas Tapia

02. Mai 2024, Stochastik für Informatik(er)

# 6. Vorlesung: Diskrete Zufallsvariablen

Nikolas Tapia

02. Mai 2024, Stochastik für Informatik(er)

# Beispiel: Warten auf erste 6 bei *Mensch-ärgere-Dich-nicht!*

Was ist die Wahrscheinlichkeit, dass man genau *k* Versuche braucht bis man eine 6 würfelt?

Folge von ZV $X_1, X_2, \dots$

definiert durch $X_k := 1_{\{k\text{-ter Versuch zeigt } 6\}}$

$X_k \sim Ber\left(p = \frac{1}{6}\right)$ „unabhängig"

$A_k =$ „6 zum ersten Mal im k-ten Versuch"

$= \{X_1 = 0, X_2 = 0, \dots, X_{k-1} = 0, X_k = 1\}$

$$P(A_k) = P(X_1 = 0, X_2 = 0, \dots X_{k-1} = 0, X_k = 1) = \underbrace{P(X_1 = 0)}_{1-p} \underbrace{P(X_2 = 0)}_{1-p} \cdot \dots \cdot \underbrace{P(X_{k-1} = 0)}_{1-p} \underbrace{P(X_k = 1)}_{p}$$

$$= (1-p)^{k-1} p = \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6}$$

Allgemein: Sei $(X_k)_{k=1}^{\infty}$, $X_k \sim Ber(p)$

Wartezeit $T := \min\{k \geq 1 \mid X_k = 1\}$

$P(T = k) = (1-p)^{k-1} p \longrightarrow$ geometrische Verteilung

Beispiel (Wiederholter Dartwurf)

# Geometrische Verteilung

**Definition 6.1**

Sei $p \in [0,1]$. Eine Zufallsvariable $X$ heißt **geometrisch verteilt** mit Parameter $p$, falls $X(\Omega) = \mathbb{N}$ und

$$\mathbb{P}(X = k) = (1-p)^{k-1} p, \quad k \in \mathbb{N}.$$

**Anmerkung 1**

Die geometrische Verteilung beschreibt die Anzahl der Versuche bis zum ersten Erfolg in einem wiederholten Bernoulli-Experiment.

**Binomialverteilung**
X : **Anzahl der Erfolge** bei einer Zufallsstichprobe vom Umfang n

**Geometrische Verteilung**
X: **Anzahl der Versuche** bis zum ersten Erfolg
Oder

**Jetzt:** Modellierung von Zählvorgängen (Betrachtung von Ereignissen) innerhalb eines festen, vorgegebenen Intervalls.

# Poisson-Verteilung

## Definition 6.2

Sei $\lambda > 0$. Eine Zufallsvariable $X$ heißt **Poisson-verteilt** mit Parameter $\lambda$, falls $X(\Omega) = \mathbb{N}_0$ und

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N}_0.$$

## Theorem 1 (Poisson-Grenzwertsatz)

Sei $n \in \mathbb{N}$, und sei $(p_n)_{n\in\mathbb{N}}$ eine Folge von Zahle aus $[0, 1]$ mit $\lim_{n\to\infty} np_n = \lambda \in (0, \infty)$.
Sei $X_n \sim \text{Binom}(n, p_n)$ eine Folge von binomialverteilten Zufallsvariablen, und sei $X \sim \text{Poisson}(\lambda)$.
Dann gilt

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$$

für alle $k \in \mathbb{N}_0$.

Die Poisson-Verteilung tritt oft in Situationen auf, in denen die Anzahl (seltener) Ereigniss
in einem bestimmten Zeitraum untersucht werden. Beispiele dafür sind

- Anzahl Anrufe, die in einer Telefonzentrale pro Minute eingehen,
- Anzahl Fahrzeuge, die pro Stunde eine bestimmte Brücke befahren,

**Beispiel:**
Eine Rückversicherung will die Prämien für Versicherungen gegen Großunfälle kalkulieren. Aufgrund von Erfahrungswerten geht sie davon aus, dass die Zufallsvariable

**X = „Anzahl der Großunfälle im Winterhalbjahr** (Oktober bis März)" Poisson-verteilt ist mit der Rate λ = 3.

**Y = „Anzahl der Großunfälle im Sommerhalbjahr** (April bis September)" als Poisson-verteilt mit der Rate λ = 6

## Schadensfälle bei einer Versicherung

$P(X=x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$

**Beispiel:**
X ist P(λ = 3) verteilt.
Y ist P(λ = 6) verteilt.

Berechnen Sie die Wkt.,
1) - Für genau 2 Großunfälle in einem Winter
2) - Mehr als ein Großunfall im Winter
3) - Für genau 2 Großunfälle im Sommerhalbjahr
4) - Mehr als ein Großunfall im Sommerhalbjahr
- Sowohl im Winter als auch im Sommer mehr als 1 Unfall

1) $P(X=2) = e^{-3} \frac{3^2}{2!} = \underline{0{,}224}$

$= F(2) - F(1) = 0.224$

2) $P(X \geq 2) = 1 - P(X \leq 1) = 1 - e^{-3}\left(\frac{3^0}{0!} + \frac{3^1}{1!}\right) = \underline{0.8009}$

3) $P(Y=2) = e^{-6} \frac{6^2}{2!} = 0{,}0446$

4) $P(Y \geq 2) = 0{,}9826$

Beispiel (Kundenbesuche)

Im Durchschnitt kommen in ein Fachgeschäft unabhängig von der Tageszeit 5 Kunden pro Stunde. Wie hoch ist die Wahrscheinlichkeit, dass kein Kunde innerhalb eines Ein-Stunden-Zeitraums den Laden betritt?

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

其中：

- $k$ 是我们要找的事件次数（在这个例子中，$k = 0$）。
- $\lambda$ 是期望的事件平均次数（这里 $\lambda = 5$）。
- $e$ 是自然常数（大约为 2.71828）。

代入参数计算概率：

$$P(X = 0) = \frac{e^{-5} \cdot 5^0}{0!} = e^{-5} \approx 0.0067$$

Beispiel (Kundenbesuche)

Wie hoch ist die Wahrscheinlichkeit, dass weniger als 2 Kunden (d.h. maximal 1 Kunde) innerhalb eines Ein-Stunden-Zeitraums den Laden betreten?

- $k$ 是要计算的顾客数量，
- $\lambda$ 是期望的平均顾客数量（这里 $\lambda = 5$），
- $e$ 是自然常数（大约为 2.71828）。

计算在一小时内少于2个顾客进入商店的概率（即 $P(X < 2)$），我们需要计算以下两个概率并相加：

1. $P(X = 0)$：

$$P(X = 0) = \frac{e^{-5} \cdot 5^0}{0!} = e^{-5} \approx 0.0067$$

2. $P(X = 1)$：

$$P(X = 1) = \frac{e^{-5} \cdot 5^1}{1!} = e^{-5} \cdot 5 \approx 0.0337$$

两者相加：

$$P(X < 2) = P(X = 0) + P(X = 1) \approx 0.0067 + 0.0337 = 0.0404$$

# Zipf-Verteilung

## Definition 6.3

Sei $a > 1$. Eine Zufallsvariable $X$ heißt **Zipf-verteilt** mit Parameter $a$, falls $X(\Omega) = \mathbb{N}$ und

$$\mathbb{P}(X = k) = \frac{k^{-a}}{\zeta(a)}, \quad k \in \mathbb{N},$$

wobei $\zeta(a) := \sum_{k \geq 1} k^{-a}$ die Riemannsche Zeta-Funktion ist.

## Anmerkung 1

Die Zipf-Verteilung, als funktion von $k$, fällt *polynomial* mit Exponent $a$, d.h.

$$\mathbb{P}(X = k) = \frac{1}{k^a}\mathbb{P}(X = 1).$$

# Gemeinsame Verteilung

## Definition 6.4

Seien $X$, $Y$ zwei diskrete Zufallsvariablen, die auf demselben Wahrscheinlichkeitsraum definiert sind. Die **gemeinsame Verteilung** von $X$ und $Y$ ist gegeben durch

$$\mathbb{P}(\{X = x\} \cap \{Y = y\}) := \mathbb{P}(X = x, Y = y), \quad x, y \in X(\Omega) \times Y(\Omega).$$

Beispiel (Gemeinsame Verteilung)

# Randverteilung

**Definition 6.5**

Seien $X$, $Y$ zwei diskrete Zufallsvariablen mit gemeinsamer Verteilung $\mathbb{P}(X = x, Y = y)$. Die **Randverteilungen** von $X$ bzw. von $Y$ sind die Verteilungen der einzelnen Zufallsvariablen $X$ bzw. $Y$. Sie sind gegeben durch

$$\mathbb{P}(X = x) := \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y), \quad x \in X(\Omega),$$

$$\mathbb{P}(Y = y) := \sum_{x \in X(\Omega)} \mathbb{P}(X = x, Y = y), \quad y \in Y(\Omega).$$

Beispiel (Randverteilung)

| $Y \backslash X$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | $\mathbb{P}(Y=y)$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1/16 | 1/8 | 1/8 | 1/8 | 0 | 0 | 0 | 7/16 |
| 2 | 0 | 0 | 1/16 | 1/8 | 1/8 | 0 | 0 | 5/16 |
| 3 | 0 | 0 | 0 | 0 | 1/16 | 1/8 | 0 | 3/16 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 1/16 | 1/16 |
| $\mathbb{P}(X=x)$ | 1/16 | 1/8 | 3/16 | 1/4 | 1/4 | 1/8 | 1/16 | 1 |

# Bedingte Verteilung

**Definition 6.6**

Seien $X$, $Y$ zwei diskrete Zufallsvariablen mit gemeinsamer Verteilung $\mathbb{P}(X = x, Y = y)$. Die **bedingte Verteilung** von $X$ gegeben $Y = y$ ist definiert als

$$\mathbb{P}(X = x | Y = y) := \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}, \quad x \in X(\Omega), y \in Y(\Omega).$$

Beispiel (Bedingte Verteilung)

**Beispiel 4.3 (Urnenmodell, Fortsetzung von Beispiel 4.1)**
Betrachte eine Urne mit 4 Kugeln, nummeriert von 1 bis 4. Wir ziehen zwei Kugeln mit Zurücklegen. Sei wieder $X$ die Summe der beiden Zahlen, und $Y$ das Minimum der beiden gezogenen Zahlen. Gesucht ist die bedingte Verteilung von $Y$ gegeben $X$.

Betrachte erst die bedingte Verteilung von $Y$ gegeben $X = 2$. Aus Tab. 4.1 sehen wir: $\mathbb{P}(X = 2) = \frac{1}{16}$, und $\mathbb{P}(X = 2, Y = 1) = \frac{1}{16}$, $\mathbb{P}(X = 2, Y = 2) = \mathbb{P}(X = 2, Y = 3) = \mathbb{P}(X = 2, Y = 4) = 0$. Somit erhalten wir

$$\mathbb{P}(Y = 1 \mid X = 2) = \frac{1/16}{1/16} = 1,$$

Beispiel (Bedingte Verteilung)

| $Y\backslash X$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 2/3 | 1/2 | 0 | 0 | 0 |
| 2 | 0 | 0 | 1/3 | 1/2 | 2/3 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 1/3 | 1 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

$$\mathbb{P}(Y = 2 \,|\, X = 2) = \mathbb{P}(Y = 3 \,|\, X = 2) = \mathbb{P}(Y = 4 \,|\, X = 2) = \frac{0}{1/16} = 0.$$

# Unabhängigkeit

**Definition 6.7**

Zwei diskrete Zufallsvariablen $X$, $Y$ heißen **unabhängig**, falls für alle $x \in X(\Omega)$ und $y \in Y(\Omega)$ die Ereignisse $\{X = x\}$ und $\{Y = y\}$ unabängig sind, d.h.

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

Beispiel (Unabhängigkeit)

# Funktionen von unabhängigen Zufallsvariablen

## Aussage 6.1

Seien $X$, $Y$ zwei unabhängige diskrete Zufallsvariablen, und seien $f, g$ zwei Funktionen. Dann sind $f(X)$ und $g(Y)$ ebenfalls unabhängige Zufallsvariablen.

## Definition 6.8

Seien $X_1, \dots, X_n$ diskrete Zufallsvariablen. Die heißen **unabhängig**, falls für alle $x_1, \dots, x_n$ die Ereignisse $\{X_1 = x_1\}, \dots, \{X_n = x_n\}$ unabhängig sind, d.h.

$$\mathbb{P}(X_{i_1} = x_{i_1}, \dots, X_{i_k} = x_{i_k}) = \mathbb{P}(X_{i_1} = x_{i_1}) \cdots \mathbb{P}(X_{i_k} = x_{i_k})$$

für alle $k \leq n$, für alle paarweise verschiedenen $i_1, \dots, i_k \in \{1, \dots, n\}$, und $x_{i_1} \in X_{i_1}(\Omega), \dots, x_{i_k} \in X_{i_k}(\Omega)$.

# Faltungsformel

## Aussage 6.2

Seien $X$, $Y$ zwei unabhängige diskrete Zufallsvariablen. Dann hat die Zufallsvariable $X + Y$ die Verteilung

$$\mathbb{P}(X + Y = k) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x)\mathbb{P}(Y = k - x)$$

für alle $k \in (X + Y)(\Omega) = \{m + n : m \in X(\Omega), n \in Y(\Omega)\}$.

Beweis (Faltungsformel)

# Summe von unabhängigen Poisson-verteilten Zufallsvariablen

## Aussage 6.3

Seien $X, Y$ unabhängige Poisson-verteilte Zufallsvariablen mit Parametern $\lambda, \mu > 0$. Dann ist die Zufallsvariable $X + Y$ Poisson-verteilt mit Parameter $\lambda + \mu$.

$$
\begin{aligned}
\mathbb{P}(X+Y=k) &= \sum_{m=0}^{\infty} \mathbb{P}(X=m)\mathbb{P}(Y=m-k) \\
&= \sum_{m=0}^{k} \frac{\lambda^m \mathrm{e}^{-\lambda}}{m!} \cdot \frac{\mu^{k-m}\mathrm{e}^{-\mu}}{(k-m)!} \\
&= \mathrm{e}^{-(\lambda+\mu)} \sum_{m=0}^{k} \lambda^m \mu^{k-m} \frac{\binom{k}{m}}{k!} \\
&= \mathrm{e}^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!},
\end{aligned}
$$

wobei wir im zweitletzten Schritt $\binom{k}{m} = \frac{k!}{m!(k-m)!}$ und im letzten Schritt den binomischen Lehrsatz $(\lambda+\mu)^k = \sum_{m=0}^{k} \binom{k}{m} \lambda^m \mu^{k-m}$ benutzt haben.

Die rechte Seite in obiger Rechnung ist nun wieder die Formel für die Poisson-Verteilung, mit Parameter $\lambda + \mu$. Wir haben somit eine wichtige Eigenschaft der Poisson-Verteilung hergeleitet: Die Summe zweier unabhängiger Poisson-verteilter Zufallsvariablen ist wiederum Poisson-verteilt, und der Parameter ist durch die Summe der Parameter gegeben. ■

Beweis