7. Vorlesung: Gemeinsame Verteilung

Nikolas Tapia

06. Mai 2024, Stochastik für Informatik(er)

Wichtige diskrete Verteilungen – Resümee

Name	Parameter	$X(\Omega)$	$\mathbb{P}(X = k)$	Notation
Gleichverteilung	$n \in \mathbb{N}$	$\{1, \ldots, n\}$	$1/n$	$X \sim \mathrm{Unif}(n)$
Bernoulli	$p \in [0, 1]$	$\{0, 1\}$	$\mathbb{P}(X = 1) = p$	$X \sim \mathrm{Ber}(p)$
Binomial	$p \in [0, 1], n \in \mathbb{N}$	$\{0, \ldots, n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$	$X \sim \mathrm{Bin}(n, p)$
Geometrisch	$p \in [0, 1]$	$\mathbb{N}$	$(1-p)^{k-1} p$	$X \sim \mathrm{Geo}(p)$
Poisson	$\lambda > 0$	$\mathbb{N}_0$	$\frac{e^{-\lambda}\lambda^k}{k!}$	$X \sim \mathrm{Poi}(\lambda)$
Zipf	$a > 1$	$\mathbb{N}$	$\frac{k^{-a}}{\sum_{k=1}^{\infty} k^{-a}}$	$X \sim \mathrm{Zipf}(a)$

- n Realisierungen von $X \sim \mathrm{Ber}(p) \to$ # Erfolge: $Y = \sum_{i=1}^{n} X_i \sim \mathrm{Bin}(n,p)$
- Wiederhole $X \sim \mathrm{Ber}(p)$ bis zum 1. Erfolg: # Versuche $\sim \mathrm{Geo}(p)$
- $X \sim \mathrm{Bin}(n,p) \xrightarrow[\substack{n \to \infty \\ p \to 0 \\ n \cdot p = \lambda \ll n}]{} Y \sim \mathrm{Pois}(\lambda)$

Leibniz Gemeinschaft

Gemeinsame Verteilung

Definition 7.1

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

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

- Beispiele: Größe und Gewicht eines Menschen; Anzahl Bestellungen und Bearbeitungszeit in einem Logistikzentrum; Anzahl und Höhe von Schäden bei einer Versicherung...
- Analog wird die gemeinsame Verteilung von mehr als zwei Zufallsvariablen definiert $P(X=x, Y=y, Z=z)$

Randverteilung

Definition 7.2

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

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

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

Bedingte Verteilung

Definition 7.3

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)

Tab. 4.1 Gemeinsame Verteilung mit Randverteilung

$Y \backslash X$	2	3	4	5	6	7	8	$\mathbb{P}(Y = \cdot)$
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
$\mathbb{P}(X = \cdot)$	1/16	1/8	3/16	1/4	3/16	1/8	1/16	1

Bedingte Verteilung $\mathbb{P}(Y|X)$ von Y gegeben X

$Y \backslash X$	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

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

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

Unabhängigkeit

Definition 7.4

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.

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

Definition 7.5

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

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

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

Beispiel (Unabhängigkeit)

Beispiel 4.4 (Urnenmodell, Fortsetzung von Beispiel 4.2)

Sind X und Y aus dem Urnenbeispiel 4.1 unabhängig? Aus Tab. 4.1 sehen wir beispielsweise gleich in der ersten Zeile, dass $\mathbb{P}(X = 2, Y = 1) \neq \mathbb{P}(X = 2)\mathbb{P}(Y = 1)$. Also können die Zufallsvariablen *nicht* unabhängig sein. Bei unabhängigen Zufallsvariablen müsste *jeder Eintrag im Inneren der Tabelle* gleich dem Produkt der beiden zugehörigen Randwahrscheinlichkeiten sein.

Man kann sich auch anschaulich klar machen, dass X und Y nicht unabhängig sein können, da beispielsweise die Information $X = 8$ zwingend festlegt, dass $Y = 4$ sein muss. Aus der Kenntnis von X lassen sich also Rückschlüsse auf Y ziehen, und umgekehrt. ■

Beispiel 4.5 (Unabhängigkeit)

Wir betrachten Zufallsvariablen X und Y, deren gemeinsame Verteilung in Tab. 4.3 gegeben ist. Wie man direkt nachrechnen kann, gilt für *jede* Wahrscheinlichkeit im Inneren der Tabelle, dass sie das Produkt der zugehörigen Randwahrscheinlichkeiten ist, also

$$\mathbb{P}(X = 0, Y = 0) = 0{,}12 = 0{,}2 \cdot 0{,}6 = \mathbb{P}(X = 0)\mathbb{P}(Y = 0),$$

$$\mathbb{P}(X = 1, Y = 0) = 0{,}18 = 0{,}3 \cdot 0{,}6 = \mathbb{P}(X = 1)\mathbb{P}(Y = 0)$$

usw. Somit sind die Zufallsvariablen X und Y in diesem Beispiel unabhängig. ■

Ordnungsstatistik

Definition 7.6

Seien $X_1, \dots, X_n$ n diskrete Zufallsvariablen. Die **Ordnungsstatistik** ist die Folge $(X_{(1)}, \dots, X_{(n)})$ der Zufallsvariablen, die durch Sortieren der Zufallsvariablen $X_1, \dots, X_n$ in aufsteigender Reihenfolge entsteht.

Aussage 7.1

Seien $X_1, \dots, X_n$ n unabhängige und **identisch verteilte** Zufallsvariablen. Dann gilt

$$\mathbb{P}(X_{(1)} \leq x) = 1 - (1 - F_X(x))^n, \quad \mathbb{P}(X_{(n)} \leq x) = F_X(x)^n.$$

Beweis

Poisson-Approximation

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 $X_n \sim \mathrm{Binom}(n, p_n)$ eine Folge von binomialverteilten Zufallsvariablen, und sei $X \sim \mathrm{Poisson}(\lambda)$.
Dann gilt

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

für alle $k \in \mathbb{N}_0$.

Poisson-Approximation

Aussage 7.2

Sei $\lambda > 0$. Sei X binomialverteilt mit Parametern $n \in \mathbb{N}$ und $p = \frac{\lambda}{n}$. Dann gilt

$$\mathbb{P}(X = k) \approx \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \{0, \dots, n\},$$

d.h. X ist approximativ Poisson-verteilt mit Parameter $\lambda = np$. Die Approximation wird besser, je größer n bzw. kleiner p ist.

Ein Geschäft hergestellt eine große Anzahl von gleichartigen Produkten, aus denen n zufällig getestet werden. Mit Wahrscheinlichkeit $p \in (0, 1)$ ist ein Produkt defekt.
Wie groß ist die Wahrscheinlichkeit, dass aus einer Stichprobe von 10 Produkten höchstens 1 defekt ist?

使用泊松分布来近似这个问题是一个合理的选择，尤其是当产品数量较大，而每个产品出现缺陷的概率 p 较小时。在这种情况下，泊松分布可以作为二项分布的一个近似。

在这个特定的例子中，如果我们考虑每个产品出现缺陷的概率 p 较小，而样本数量 $n = 10$ 并不非常大，泊松近似可能不是非常精确，但仍可提供一个合理的估计。

泊松分布的参数 λ 等于 $n \times p$，即 $\lambda = 10p$。

泊松分布的概率质量函数为：

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

我们想计算最多有一个产品缺陷的概率，即 $P(X \leq 1)$：

$$P(X \leq 1) = P(X = 0) + P(X = 1)$$

$$P(X = 0) = e^{-\lambda}\frac{\lambda^0}{0!} = e^{-10p}$$

$$P(X = 1) = e^{-\lambda}\frac{\lambda^1}{1!} = 10pe^{-10p}$$

所以：

$$P(X \leq 1) = e^{-10p} + 10pe^{-10p}$$

$$P(X \leq 1) = (1 + 10p)e^{-10p}$$

Funktionen von unabhängigen Zufallsvariablen

Aussage 7.3

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

Faltungsformel

Aussage 7.4

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: $P(Z=k)=\sum_{m\in X(\Omega)} P(X=m, Z=k)$ Summenregel

$$= \sum_{m\in X(\Omega)} P(\underbrace{\{X=m\}\cap\{X+Y=k\}}_{\{X=m\}\cap\{Y=k-m\}}) = \sum_m P(X=m, Y=k-m)$$

$$= \sum_{m\in X(\Omega)} P(X=m)\cdot P(Y=k-m)$$

Beweis (Faltungsformel)

Summe von unabhängigen Poisson-verteilten Zufallsvariablen

Aussage 7.5

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

Beispiel 4.7: Poisson-Verteilung

$X \sim \mathrm{Pois}(\lambda) \qquad Y \sim \mathrm{Pois}(\mu) \qquad$ unabhängig

$Z := X + Y$

$$P(Z=k) = \sum_{m=0}^{\infty} P(X=m)\, P(Y=k-m)$$

$$= \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} \frac{\mu^{k-m}}{(k-m)!} e^{-\mu} = e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{m=0}^{\infty} \underbrace{\underbrace{\frac{1 \cdot k!}{m!(k-m)!}}_{\binom{k}{m}} \lambda^m \mu^{k-m}}_{(\lambda+\mu)^k}$$

$$= \frac{(\lambda+\mu)^k}{k} e^{-(\lambda+\mu)}$$

$$\Rightarrow \quad Z \sim \mathrm{Pois}(\lambda+\mu)$$

Beweis

Beispiel

Seien $X \sim \text{Poisson}(\lambda)$,
$Y \sim \text{Poisson}(\mu)$.
Welche bedingte Verteilung hat X,
gegeben $X + Y = n$?