Stochastik für die Informatik, Vorlesung 7

Inhalt

- Wichtige Verteilungen: Binomial, Geometrisch, Poisson, Zipf
- Vorkommen dieser Verteilungen
- Zufallsgraphen: Erdős-Rényi und Preferential Attachment

Lernziele

- Mit den Begriffen Zufallsvariable, Verteilung und Verteilungsfunktion umgehen können
- Wichtige Beispiele von Verteilungen, ihre Charakteristika und ihr Auftreten kennen
- Wahrscheinlichkeiten für diese Verteilungen berechnen können

Vorkenntnisse

Stoff der bisherigen Vorlesungen, Funktionen, Summen und Reihen

Beispiel: Binomial-Verteilung

(Def. 3.10) Seien $n \in \mathbb{N}, p \in [0,1]$. Eine Zufallsvariable X heißt binomialverteilt mit Parametern n und p falls gilt:
$X(\Omega) = \{0, ..., n\}$, und

$$\boxed{p_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad , k = 0, ..., n.}$$

(Satz 3.3)[Binomialverteilung im wiederholten Bernoulli-Experiment] Ein Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p wird n mal unter gleichbleibenden Bedingungen wiederholt. Sei X die Anzahl Erfolge. Dann ist X binomialverteilt mit Parameter n und p.

Binomialverteilung

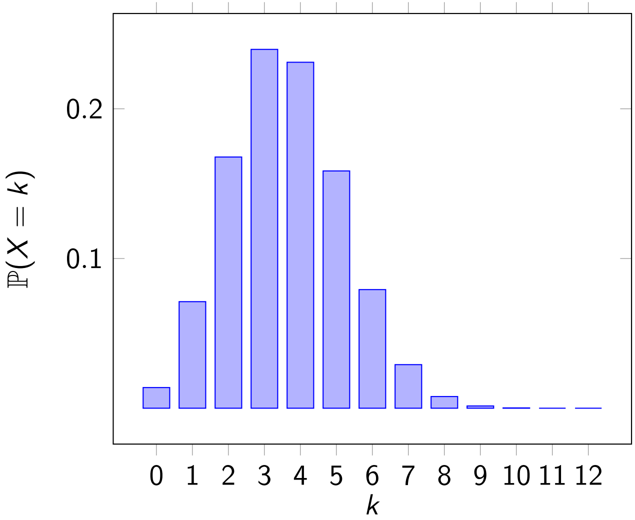


Abbildung: Binomialverteilung mit $n = 12, p = 0.3$

Binomialverteilung

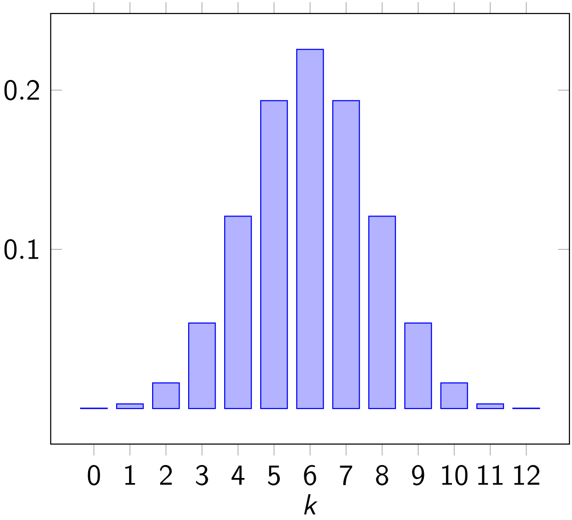


Abbildung: Binomialverteilung mit $n = 12, p = 0.5$

Beispiel: Geometrische Verteilung

(Def. 3.11) Sei $p \in [0,1]$ Eine Zufallsvariable X heißt geometrisch verteilt mit Parameter p falls gilt: $X(\Omega) = \mathbb{N}$, und

$$\boxed{p_X(k) = \mathbb{P}(X = k) = p(1-p)^{k-1}.}$$

- Notation: $X \sim \mathrm{Geo}(p)$
- Beispiel für einen abzählbaren Zustandsraum
- p_X ist tatsächlich eine Verteilung
- warum geometrisch?

Geometrische Verteilung

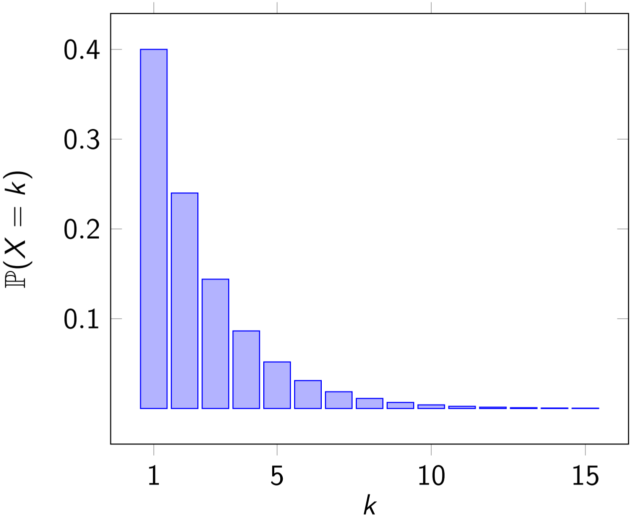


Abbildung: Geometrische Verteilung mit $p = 0.4$

Geometrische Verteilung

(Satz 3.4) Wir betrachten ein Bernoulli-Experiment mit Parameter p. Sei X die Anzahl Versuche die gemacht werden müssen, bis zum ersten Mal "Erfolg" eintritt. Dann ist X geometrisch verteilt mit Parameter p.

- (Beispiel: Würfeln)
- (Beispiel 3.18 Stochastischer Algorithmus)

Beispiel: Poisson-Verteilung

(Def. 3.12) Sei $\lambda > 0$. Eine Zufallsvariable X heißt Poisson-verteilt mit Parameter λ, falls $X(\Omega) = \mathbb{N}_0$, und

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

- Auftreten der Poisson-Verteilung: Anzahl (seltener) Ereignisse in einem bestimmten Zeitraum. Beispiele:
  - Anzahl Anrufe, die in einer Telefonzentrale pro Minute eingehen
  - Anzahl Bestellungen pro Minute in einem Online-Shop
  - Anzahl Mutationen auf einem Stück DNA
  - Anzahl Schäden pro Tag in einem Versicherungsunternehmen
- Der Parameter λ bezeichnet die Rate mit der die Ereignisse eintreten

Poisson-Verteilung

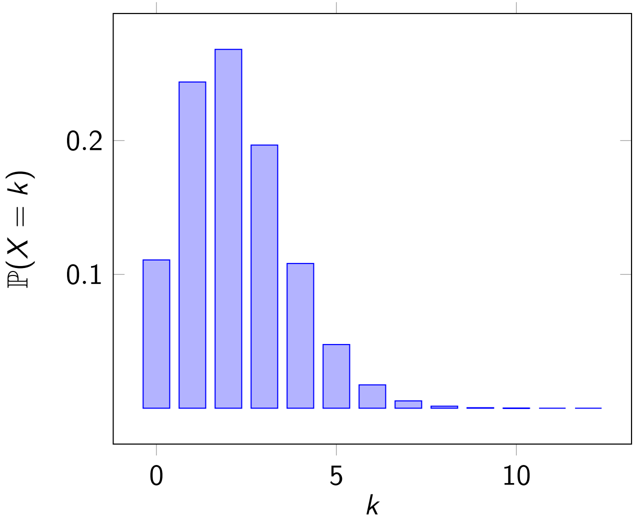


Abbildung: Poisson-Verteilung mit $\lambda = 2.2$

Poisson-Grenzwertsatz

(Satz 3.5) Sei $n \in \mathbb{N}$, und sei $(p_n)_{n\in\mathbb{N}}$ eine Folge von Zahlen aus $[0,1]$ mit $\lim_{n\to\infty} n \cdot p_n = \lambda \in]0,\infty[$. Sei X_n binomial verteilt mit Parametern n und p_n, und sei X Poisson-verteilt mit Parameter λ. Dann gilt

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

- Anwendung (Satz 3.6): Bernoulli-Experiment mit einer *großen* Anzahl Versuche n und *kleiner* Erfolgswahrscheinlichkeit p_n, so dass $n \cdot p_n = \lambda$ gilt. Dann ist die Anzahl Erfolge X_n im wiederholten Bernoulli-Experiment *ungefähr* Poisson-verteilt mit Parameter λ,
$$\mathbb{P}(X_n = k) \approx \frac{\lambda^k e^{-\lambda}}{k!}.$$
- Interpretation der Poisson-Verteilung als Anzahl seltener Ereignisse
- (Beispie 3.20: Erdős-Rényi-Zufallsgraph)

Realisierungen

(Def.) Der Wert der Zufallsvariablen für eine Ausführung eines Zufallsexperiments nennt man Realisierung.

Anstelle der realen Ausführung eines Zufallsexperiments simulieren wir die entsprechenden Zufallsvariablen am Rechner.

- Simulation von diskret gleichverteilten Zufallsvariablen mit R: `sample(1:n, k, replace=T)` ergibt k Realisierungen einer Gleichverteilung auf $\{1, ..., n\}$.
- Simulation von gleichverteilten Zufallsvariablen auf $[0, 1]$ mit R: `runif(k,0,1)` ergibt k Realisierungen einer Gleichverteilung auf $[0, 1]$. Allgemeiner ergibt `runif(k,a,b)` entsprechend k Realisierungen einer Gleichverteilung auf $[a, b]$.
- Simulation von endlichen Verteilungen: Siehe Vorlesung und Hausaufgaben

Simulation mit R

Zufallsvariablen der wichtigsten Verteilunge können mit R einfach simuliert werden, da diese bereits implementiert sind.

- `rbinom(k,n,p)` gibt k Realisierungen einer binomial verteilten Zufallsvariablen mit Parametern n und p aus
- `rgeom(k,p)` gibt k Realisierungen einer geometrisch verteilten Zufallsvariablen mit Parameter p aus
- `rpois(k, λ)` gibt k Realisierungen einer Poisson-verteilten Zufallsvariablen mit Parameter λ aus

Ausgabe der entsprechenden Verteilungen: Befehl `dbinom(n,p)`, `dgeom(p)`, `dpois(λ)`

Beispiel 3.8: Gleichverteilung auf $[0,1]$

Analog zur diskreten Gleichverteilung heißt eine Zufallsvariable X gleichverteilt, wenn jeder Wert von $X(\Omega)$ gleich wahrscheinlich ist.

Problem: Ist $X(\Omega)$ nicht endlich, so hat ein einzelner Wert damit Wahrscheinlichkeit 0.

(Def.) Eine Zufallsvariable X mit $X(\Omega) = [0,1]$ heißt gleichverteilt (auf $[0,1]$), wenn

$$\boxed{\mathbb{P}(X \leq x) = F_X(x) = x \quad \text{für jedes } x \in [0,1]}$$

gilt.

- (Interpretation)
- (Beispiel 3.9: Simulation von Zufallsvariablen mit endlichem Wertebereich)

Häufigkeiten und Histogramme

(Beispiel 3.10) Mit R erzeugen wir 100 Realisierungen der Summe zweier fairer Würfel:

```
sum<-sample(1:6, 100, replace = T) + sample(1:6, 100,
replace = T)
```

Die Ausgabe ist eine Liste von 100 Werten.

- Befehl `sort` um die Liste zu sortieren
- Befehl `table` um eine Tabelle der Häufigkeiten auszugeben
- Befehl `hist` um ein Histogramm zu erzeugen

Histogramm

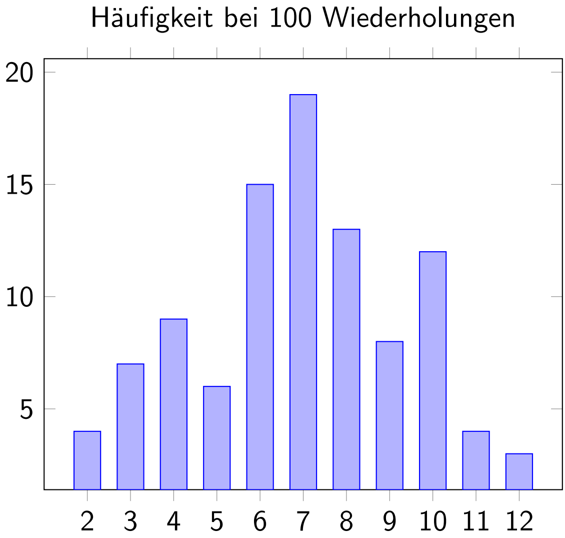


Abbildung: Summe der Augenzahlen beim zweifachen Würfeln

Beispiel: Zipf-Verteilung

(Def. 3.13) Sei $a > 1$. Eine Zufallsvariable X heißt Zipf-verteilt (oder Zeta-verteilt) mit Parameter a falls gilt: $X(\Omega) = \mathbb{N}$, und

$$\boxed{p_X(k) = \mathbb{P}(X = k) = \frac{k^{-a}}{Z(a)},}$$

wobei $Z(a) := \sum_{k=1}^{\infty} k^{-a}$ ist.

- p_X ist tatsächlich eine Verteilung, Rolle von $Z(a)$
- Parameter a ist keine Wahrscheinlichkeit, er spielt eine andere Rolle als (z.B.) das p bei der geometrischen Verteilung
- (Fall $a \leq 1$)
- (Vergleich mit geometrischer Verteilung)

Zipf-Verteilung

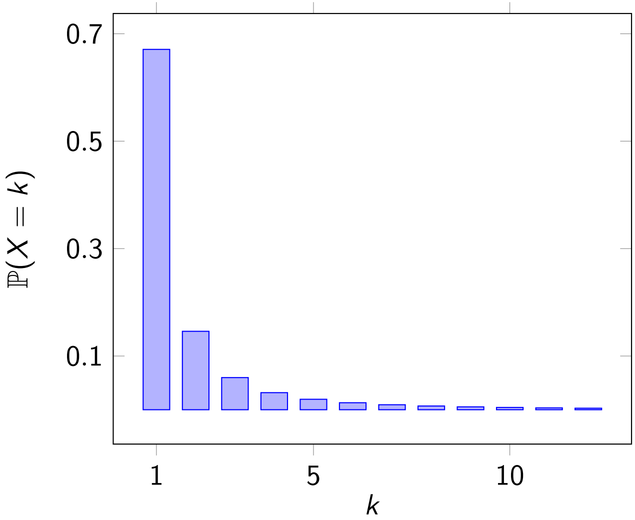
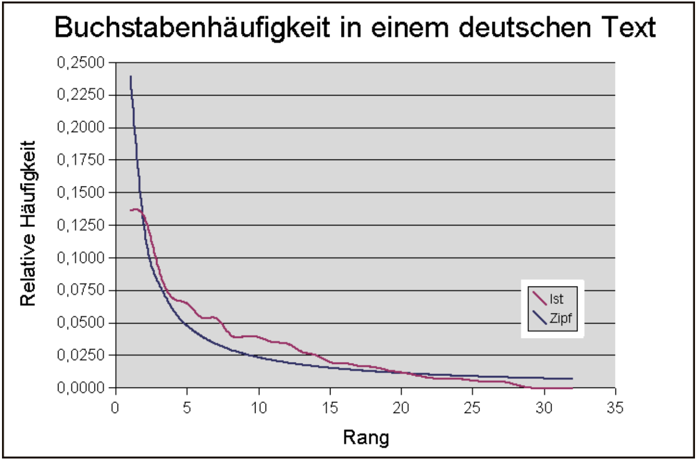


Abbildung: Zipf-Verteilung mit $a = 2.2$

Zipf-Verteilung: Auftreten

(Beispiel 3.21: $a = 1$)
(Häufigkeit von Buchstaben)



Zusammenfassung: Wichtige diskrete Verteilungen

Name	Parameter	$X(\Omega)$	$\mathbb{P}(X=k)$
Gleichverteilung	$n \in \mathbb{N}$	$\{1, ..., n\}$	$1/n$
Bernoulli	$p \in [0,1]$	$\{0,1\}$	p falls $k=1$, $1-p$ falls $k=0$
Binomial	$p \in [0,1], n \in \mathbb{N}$	$\{0, ..., n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$
Geometrisch	$p \in [0,1]$	$\mathbb{N}$	$(1-p)^{k-1} p$
Zipf	$a > 1$	$\mathbb{N}$	$\frac{k^{-a}}{\sum_{k=1}^{\infty} k^{-a}}$
Poisson	$\lambda > 0$	$\mathbb{N}_0$	$\frac{e^{-\lambda}\lambda^k}{k!}$

Notation: $X \sim$Ber(p), $X \sim$Bin(n,p), $X \sim$Geo(p), $X \sim$Poi(λ)...

Auftreten:

Bernoulli	Bernoulli-Experiment
Binomial	Anzahl Erfolge im wiederholten Bernoulli-Experiment
Geometrisch	Anzahl Versuche bis zum ersten Erfolg im Bernoulli-Experiment
Poisson	Anzahl (seltener) Ereignisse in einem gegebenen Zeitraum