Stochastik für die Informatik, Vorlesung 8

Inhalt

- Zipf-Verteilung und skalenfreie Netzwerke
- Erwartungswert

Lernziele

- Die Zipf-Verteilung und ihr Auftreten kennen
- Den Erwartungswert kennen
- Erwartungswerte berechnen können, insbesondere bei wichtigen Verteilungen

Vorkenntnisse

Stoff der bisherigen Vorlesungen

Erinnerung: 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	$p \in [0, 1]$	$\mathbb{N}$	$(1 - p)^{k-1} p$
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

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

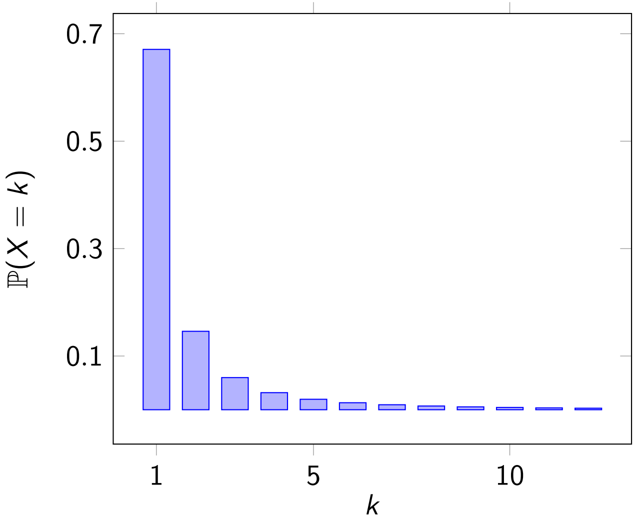


Abbildung: Zipf-Verteilung mit $a = 2.2$

Zipf-Verteilung: Auftreten

(Beispiel 3.41: $a = 1$)
(Häufigkeit von Buchstaben)

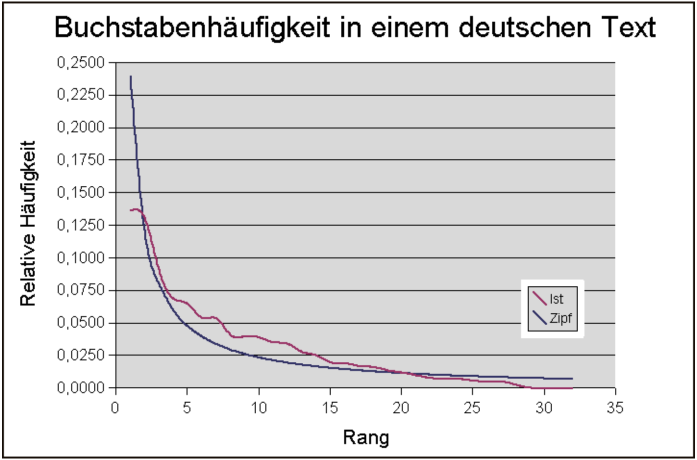


Bild: Zipf-Verteilung-Buchstaben von Anton aus der deutschsprachigen Wikipedia. Lizenziert unter CC BY-SA 3.0 ber Wikimedia Commons

Zipf-Verteilung: Auftreten

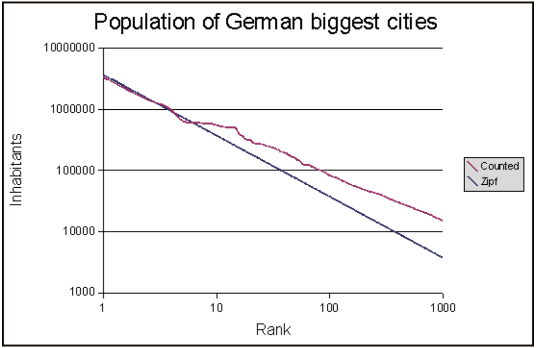


Bild: “Powercitieszipl nrp”. Lizenziert unter CC BY-SA 2.5 ber Wikimedia Commons. Achtung: Logarithmische Skala!

Anwendung: Skalenfreie Netzwerke/Preferential attachment Modell

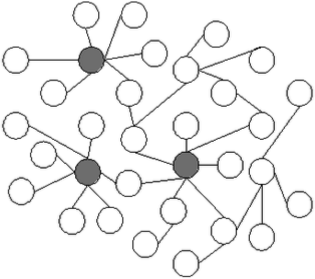


Bild: Bearbeitung von "Scale-free network deutsch" von von CollectiveStupidity - Bearbeitung des englischen Originals. Lizenziert unter CC BY-SA 3.0 ber Wikipedia

(Vereinfachtes) Bildungsgesetz: Sukzessives Hinzufügen von Knoten, welche sich mit einem bereits vorhandenen Knoten verbinden. Dieser wird mit einer Wahrscheinlichkeit proportional zur schon bestehenden Anzahl Nachbarknoten ausgewählt.

Skalenfreie Netzwerke und die Zipf-Verteilung

(Beispiel 3.22: Preferential attachment)

Sei X die Anzahl Nachbarknoten eines zufällig ausgewählten Knoten (= der Grad des Knoten).

Falls die Gesamtzahl der Knoten groß ist, so ist X (annähernd) Zipf-verteilt. Der Wert des Parameters a hängt von der genauen Konstruktion des Netzwerkes ab. Im beschriebenen Bildungsgesetz ist $a \approx 3$.

Simulationsprogramm: `http://ccl.northwestern.edu/netlogo/models/PreferentialAttachment`
Beispielvideo (youtube):
`https://www.youtube.com/watch?v=4GDqJVtPEGg`
R-code für verschiedene Zufallsgraphen: `https://www.dam.brown.edu/MSF/misc/MSF_SimulationsClass3.html`

Skalenfreie Netzwerke

Vorkommen von skalenfreien Netzwerken (vermutet):

- WorldWideWeb, Internet
- Soziale Netzwerke, Netzwerke wissenschaftlicher Zusammenarbeit
- Stromnetze
- usw.

(Mathematisch) verwandte Modelle

- Lässt sich in ein bestimmtes Urnenmodell übersetzen
- Bildung biologischer Spezies

Kapitel 5: Kenngrößen von Zufallsvariablen

- Die Verteilung enthält im Prinzip alle Informationen über eine (diskrete) Zufallsvariable, ist aber möglicherweise aufwändig zu bestimmen oder vollständig anzugeben
- Gesucht: Einfach zu bestimmende Größen, welche bereits wichtige Informationen über die Zufallsvariable enthalten
- Beispiel: Welchen Wert nimmt die Zufallsvariable im Mittel an, wie stark ist die Streuung um diesen Wert?
- Zentrale Größen: Erwartungswert, Varianz
- Kenngrößen für den Zusammenhang zwischen mehreren Zufallsvariablen: Kovarianz, Korrelation

Beispiel

Wir werfen einen fairen Würfel 100 Mal, und notieren die Ergebnisse. Sei y_i das Ergebnis des $i-$ten Wurfs. Wie groß ist dann der Mittelwert über die 100 Versuche, also

$$\frac{1}{100}\sum_{i=1}^{100} y_i?$$

Da die Ergebnisse $\{1, ..., 6\}$ gleich wahrscheinlich sind, erwarten wir im Durchschnitt

$$\frac{1}{100}\sum_{i=1}^{100} y_i \approx \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5.$$

Beispiel

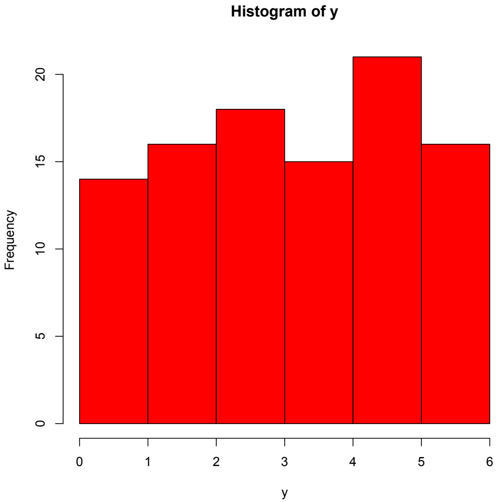
Wir werfen einen fairen Würfel 100 Mal, und notieren die Ergebnisse. Sei y_i das Ergebnis des $i-$ten Wurfs. Wie groß ist dann der Mittelwert über die 100 Versuche, also

$$\frac{1}{100}\sum_{i=1}^{100} y_i?$$

Da die Ergebnisse $\{1, ..., 6\}$ gleich wahrscheinlich sind, erwarten wir im Durchschnitt

$$\frac{1}{100}\sum_{i=1}^{100} y_i \approx \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5.$$

R-Simulation: 100 Würfe



Mittelwert in der Beispielsimulation: $\frac{1}{100}\sum_{i=1}^{100} y_i = 3.61$

Erwartungswert

(Def. 5.1) Sei X eine diskrete Zufallsvariable. Der Erwartungswert von X ist definiert als

$$\boxed{\mathbb{E}[X] := \sum_{k \in X(\Omega)} k \cdot \mathbb{P}(X = k).}$$

- (Beispiel 5.1: Würfeln)
- (Beispiel 5.2: Tabelle)
- (Beispiel 5.4: Zipf, Bemerkung: Existenz des Erwartungswertes)
- (Satz 5.2) Falls Ω höchstens abzählbar ist, so gilt

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega).$$

Eigenschaften des Erwartungswerts

(Satz 5.1) Seien X und Y diskrete Zufallsvariablen auf $(\Omega, \mathbb{P})$. Dann gelten

(a) $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

(b) $\mathbb{E}[a \cdot X] = a \cdot \mathbb{E}[X]$ für jedes $a \in \mathbb{R}$

(c) Falls $X(\omega) \geq 0 \forall \omega$, so ist $\mathbb{E}[X] \geq 0$.

(d) Falls $X = c$ konstant ist, so ist $\mathbb{E}[X] = c$.

- (a) und (b) zusammengenommen besagen, dass der Erwartungswert linear ist.
- (Beispiel 5.6: Bernoulli- und Binomialverteilung, Beispiel 5.7 Zufallsgraph)

Binomialverteilung

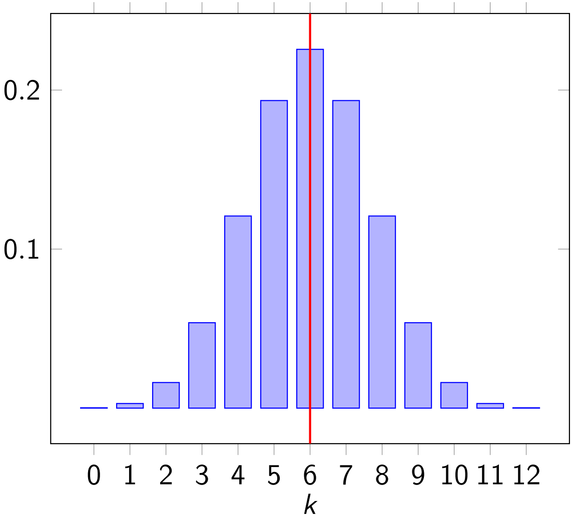


Abbildung: $n = 12, p = 0.5$. Erwartungswert $\mathbb{E}[X] = 12 \cdot 0.5 = 6$

Beispiel: Binomialverteilung

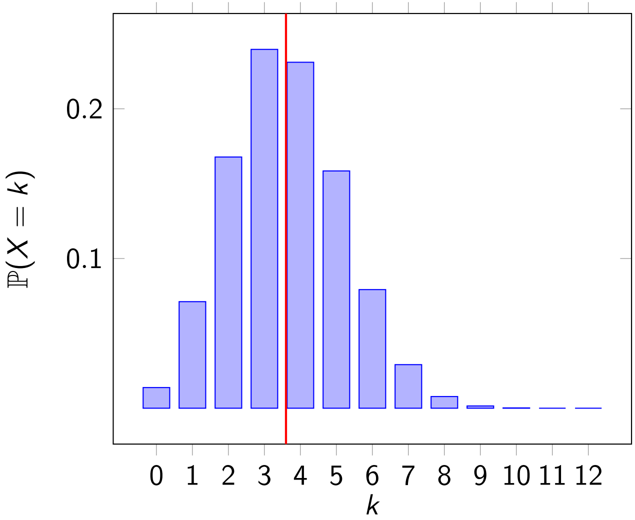


Abbildung: $n = 12, p = 0.3$ Erwartungswert $\mathbb{E}[X] = 12 \cdot 0.3 = 3.6$

Eigenschaften des Erwartungswerts

(Satz 5.1, Fortsetzung) Seien X und Y diskrete Zufallsvariablen auf $(\Omega, \mathbb{P})$. Dann gelten

(e) Sei $f : \mathbb{R} \to \mathbb{R}$ eine Funktion. Dann gilt

$$\mathbb{E}[f(X)] = \sum_{k \in X(\Omega)} f(k)\mathbb{P}(X = k).$$

- (Beispiel)
- (Beispiel 5.3/5.8: Geometrische Verteilung und coupon collector problem)
- (Randomisierter Quicksort: Siehe Kapitel 12)