Die Poisson und die Exponential-Verteilung

https://iuk.one/1012-1024

Clemens H. **Cap**
ORCID: 0000-0003-3958-6136

Department of Computer Science
University of **Rostock**
Rostock, Germany
clemens.cap@uni-rostock.de

Version 2.1

Ziele

Poisson- und Exponential-Verteilung sind
wichtige Verteilungen in den Ingenieurwissenschaften.

- Wir betrachten Ereignisse, die im Mittel λ mal pro Zeiteinheit eintreten.
- Die **Poisson-Verteilung** $P_{\lambda;t}(k)$ gibt die Wahrscheinlichkeit an, daß ein Ereignis in einem Zeitraum der Dauer t genau k-mal eintritt.
- Die **Exponential-Verteilung** $Q_\lambda(a,b)$ gibt die Wahrscheinlichkeit an, daß die Zeit zwischen zwei Ereignissen im Intervall $[a,b]$ liegt.

1. Ein Beispiel

Ziele: Wir betrachten ein Beispiel und leiten aus diesem die Formeln für Poisson- und Exponential-Verteilung ab.

Beispiel: Donald auf Twitter

Situation:
Donald ist eine bekannte Persönlichkeit auf Twitter.
Im Schnitt erhält er $\lambda = 0.5$ [Mentions/h] auf Twitter.
Wie groß ist die Wahrscheinlichkeit, daß Donald

- innerhalb von 4 Stunden 3 Mentions erhält?
- nach einem Mention auf das nächste mindestens 2 [h] aber höchstens 4 [h] wartet?

Zwischen-Überlegung

Achtung: Ohne Zusatzannahmen ist das Problem unterbestimmt!
Zusatzannahmen machen es wieder lösbar.
Vergleiche: Laplacesches Indifferenzprinzip.

Argument 1: Die Wahrscheinlichkeit kann von der **Uhrzeit** abhängen.

- Bsp: Mittags mehr Mentions als um Mitternacht.
- Aber: Wir nehmen an, Donald sei weltweit in allen Zeitzonen bekannt.
  Die Uhrzeit habe daher keinen Einfluß.

Argument 2: Die Wahrscheinlichkeit kann von der **Vorgeschichte** abhängen.

- Bsp: Wenn ein Tweet Mentions enthält, dann provoziert das Antworten.
  Diese Antworten enthalten wieder Mentions.
- Aber: Wir nehmen an, die Mentions stammen von Bots, die nur Tweets absetzen.
  Sie lesen keine Tweets und reagieren nicht auf diese.

Anschauliche Herleitung Poisson (1)

Sei $R(t_0, t_1)(k)$ die Wahrscheinlichkeit, daß Donald im Zeitraum t_0 bis t_1 genau k Mentions auf Twitter erhält.

Annahme 1: Die Wahrscheinlichkeit hängt nicht von der genauen Lage des Zeitintervalls ab, sondern **nur von seiner Länge**.

Daher dürfen wir die Wahrscheinlichkeit in der Form $R(t_0, t_1)(k) = P(t_1 - t_0)(k)$ schreiben.

Annahme 2: Die mittlere Anzahl von Ereignissen pro Zeiteinheit sei **unabhängig von Zeitdauer und Vorgeschichte** immer gleich λ.

Zerlege den Zeitraum t in n Abschnitte der Dauer $h = \frac{t}{n}$.

Näherung:

- In jedem kleinen Abschnitt sollen 0 oder 1 Ereignis stattfinden.
- Mehr als ein Ereignis pro Abschnitt werden nicht betrachtet.
- Die Wahrscheinlichkeit eines Ereignisses in einem Abschnitt sei $p = \lambda \cdot h$.

Anschauliche Herleitung Poisson (2)

Die Wahrscheinlichkeit, daß k Ereignisse in diesen n Abschnitten auftreten, ist durch die **Binomialverteilung** gegeben. Sie beträgt:

$$\binom{n}{k} p^k (1-p)^{n-k}$$

Was war nochmal die Binomialverteilung? (1)

Situation:

- Wir betrachten ein **dichotomes** Experiment.
- Das bedeutet: Es sind **genau zwei** Elementarereignisse möglich.
- Das E-Ereignis und sein Co-Ereignis.
- Die Wahrscheinlichkeit des E-Ereignisses sei p.
- Dann ist die Wahrscheinlichkeit des Co-Ereignisses $1 - p$.
- Bsp: Bestanden oder nicht-bestanden. Kopf oder Zahl.
- Wir wiederholen das Experiment n mal.
- Wir nehmen dabei an, daß die Wiederholungen unabhängig sind.

Binomialverteilung

Die **Binomialverteilung** $B_{p;n}(k)$ gibt die Wahrscheinlichkeit an, daß bei n-facher unabhängiger **Wiederholung eines dichotomen** E-Ereignisses der Wahrscheinlichkeit p das E-Ereignis k mal auftritt.

Was war nochmal die Binomialverteilung? (2)

Die einfachen Fälle sind leicht selber nachzuprüfen:

k	$B_{p;1}(k)$
1	p
0	$1-p$

k	$B_{p;2}(k)$
2	p^2
1	$2p(1-p)$
0	$(1-p)^2$

k	$B_{p;3}(k)$
3	p^3
2	$3p^2(1-p)$
1	$3p(1-p)^2$
0	$(1-p)^3$

k	$B_{p;4}(k)$
4	p^4
3	$4p^3(1-p)$
2	$6p^2(1-p)^2$
1	$4p(1-p)^3$
0	$(1-p)^4$

$$B_{p;n}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Anschauliche Herleitung Poisson (3)

Die Wahrscheinlichkeit, daß k Ereignisse in diesen n Abschnitten auftreten, ist durch die **Binomialverteilung** gegeben. Sie beträgt:

$$\binom{n}{k} p^k (1-p)^{n-k}$$

Wir bilden nun den Grenzwert $n \to \infty$ und erhalten:

$$\lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} = \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} =$$

$$\lim_{n\to\infty} \frac{n!}{k!(n-k)!n^k} (\lambda t)^k \left(1 - \frac{\lambda t}{n}\right)^n \left(1 - \frac{\lambda t}{n}\right)^{-k} =$$

$$\frac{(\lambda t)^k}{k!} \lim_{n\to\infty} \frac{n}{n} \frac{n-1}{n} \cdots \frac{n-k+1}{n} \lim_{n\to\infty} \left(1 - \frac{\lambda t}{n}\right)^n \left(1 - \frac{\lambda t}{n}\right)^{-k} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

Antwort Poisson

Wahrscheinlichkeit von genau k Ereignissen in einem Zeitraum der Dauer t:

$$P_{\lambda;t}(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

Es war: $\lambda = 0.5$ [Ereignisse/h], $t = 4$ [h] und $k = 3$ [Mentions].

Daraus erhalten wir:

$$P_{0.5;4}(3) = \frac{(0.5 \cdot 4)^3}{3!} e^{-0.5 \cdot 4} = \frac{8}{6} e^{-2} \sim \frac{4}{3} \frac{1}{(2.7)^2} \sim 0.18$$

Antwort: Die Wahrscheinlichkeit, daß Donald in 4 Stunden genau 3 Mentions erhält beträgt ca. 0.18.

Anschauliche Herleitung Exponential (1)

Annahmen:

- Donald hat seit letztem Mention y Zeiteinheiten gewartet und in diesem Zeitraum ist kein Mention eingetroffen.
- Dann trifft innerhalb einer sehr kurzen Zeitspanne dy ein Mention ein.

Daß y Zeiteinheiten lang *kein* Mention eintrifft hat eine Wahrscheinlichkeit, die wir aus der Poisson-Verteilung zu $P_{\lambda;y}(0) = e^{-\lambda y}$ bestimmen.

Daß nun innerhalb einer (sehr) kurzen Zeitspanne dy genau ein Mention eintrifft, hat die Wahrscheinlichkeit $\lambda \cdot dy$.

Für die **sehr kurze Zeitspanne** dy ist die Wahrscheinlichkeit, daß zunächst y Zeiteinheiten lang kein Mention eintrifft und dann innerhalb der sehr kurzen Zeitspanne dy genau ein Mention eintrifft, das Produkt dieser Wahrscheinlichkeiten (weil diese Ereignisse als unabhängig angenommen wurden): $e^{-\lambda y} \cdot \lambda \cdot dy$

Anschauliche Herleitung Exponential (2)

Für **längere Zeitspannen** ergibt sich die Wahrscheinlichkeit als Summe über viele (sehr) kurze Zeitspannen, in die wir die längere Zeitspanne aufteilen können.

Die Summe ist zu benutzen, da die Ereignisse disjunkt sind.

Als Wahrscheinlichkeit, daß die Wartezeit im Intervall $[a, b]$ liegt, ergibt sich das Integral:

$$Q_\lambda(a, b) = \int_a^b e^{-\lambda y} \lambda dy = e^{-\lambda a} - e^{-\lambda b}$$

Antwort Exponential

Wahrscheinlichkeit, daß die Wartezeit von Mention zu Mention im Intervall $[a, b]$ liegt:

$$Q_\lambda(a, b) = e^{-\lambda a} - e^{-\lambda b}$$

Es war: $\lambda = 0.5$ [Ereignisse/h], $a = 2$ [h] und $b = 4$ [h].

Daraus erhalten wir:

$$Q_{0.5}(2, 4) = e^{-0.5 \cdot 2} - e^{-0.5 \cdot 4} = e^{-1} - e^{-2} = \frac{1}{e} - \frac{1}{e^2} = \frac{e-1}{e^2} \sim \frac{1.7}{(2.7)^2} \sim 0.23$$

Antwort: Die Wahrscheinlichkeit, daß Donald zwischen 2 und 4 Stunden auf das nächste Mention wartet, beträgt ca. 0.23.

2. Poisson-Verteilung

Ziele: Wir definieren die Poisson-Verteilung formal und betrachten einige ihrer Eigenschaften.

Definition und Eigenschaften

Die **Poisson-Verteilung** zum *Parameter* $\lambda \in \mathbb{R}^+$ über die *Zeitdauer* $t \in \mathbb{R}^+$ ist die Funktion $P_{\lambda;t} \colon \mathbb{N}_0 \to [0,1]$ mit

$$P_{\lambda;t}(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

Eigenschaften:

- $P_{\lambda;t}$ ist tatsächlich eine **Wahrscheinlichkeit** über $\mathbb{N}_0$.
  Insbesondere ist $\sum_{k=0}^{\infty} P_{\lambda;t}(k) = 1$.
- Die **mittlere Anzahl von Ereignissen** in einem Zeitraum der Dauer t ist $\lambda \cdot t$.

Nachweis der Eigenschaften

$$\sum_{k=0}^{\infty} P_{\lambda;t}(k) = \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} = e^{-\lambda t} \underbrace{\sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!}}_{=e^{\lambda t}} = 1$$

Die mittlere Anzahl von Ereignissen ist definiert als die Summe über die jeweilige Wahrscheinlichkeit mal dem entsprechenden Wert. Das ist:

$$\sum_{k=0}^{\infty} P_{\lambda;t}(k) \cdot k$$

Das ist aber:

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

Gedächtnisfreiheit der Poisson-Verteilung

Die Wahrscheinlichkeit des Eintretens von Ereignissen ist **unabhängig** davon, ob und wann davor bereits Ereignisse eingetreten sind.

Anschaulich: Die Verteilung merkt sich nicht, was in früheren Zeiträumen geschehen ist. Sie hat kein Gedächtnis.

Unabhängigkeit ist gegeben, wenn die Wahrscheinlichkeit **gemeinsam** auftretender Ereignisse gleich dem **Produkt** der Wahrscheinlichkeiten der Ereignisse ist.

Ohne hier eine präzisere Form der Unabhängigkeit zu formulieren (eine solche ist möglich!) beobachten wir hier einfache Gleichungen, die eine solche Unabhängigkeit in einzelnen Beispielen ausdrücken:

$$P_{\lambda;a}(0) \cdot P_{\lambda;b}(0) = P_{\lambda;a+b}(0)$$

$$P_{\lambda;a}(0) \cdot P_{\lambda;b}(1) + P_{\lambda;a}(1) \cdot P_{\lambda;b}(0) = P_{\lambda;a+b}(1)$$

Graph Poisson-Verteilung

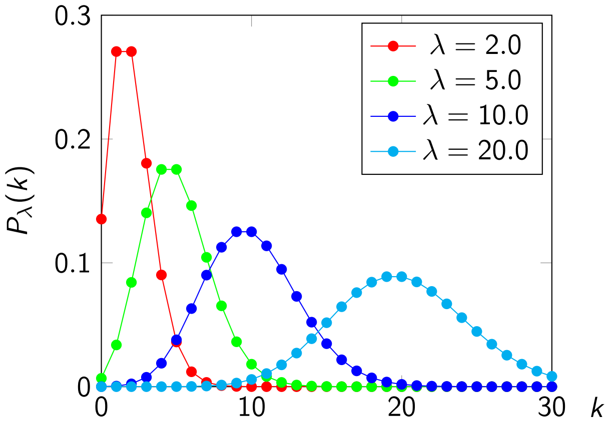

Abb. 1: Verteilungsfunktion der Poisson-Verteilung. Beachte: Obwohl die Poisson-Verteilung eine diskrete Verteilung ist, zeichnen wir sie hier interpoliert durchgezogen, um optisch besser herauszuarbeiten, daß sie unter bestimmten Bedingungen auch als Approximation der Normalverteilung angesehen werden kann.

Anwendung

Charakteristika:

- Ereignisse treffen mit einer gewissen zeitlichen Rate ein.
- Die einzelnen Ereignisse sind voneinander statistisch unabhängig.
- Die Rate ist unabhängig von der Zeit.

Dann gilt:

- Die Anzahl der Ereignisse in einem gegebenen Zeitraum ist **Poisson-verteilt**.
- Die Wartezeit von einem Ereignis auf das nächste ist **Exponential-verteilt**.

Beispiele für die Poisson-Verteilung

Typische Beispiele:

- Anzahl Datenpakete, die zum Versand bei einem Router eintreffen.
- Anzahl Zugriffe auf einen Webserver.
- Anzahl Käufe auf einem Webshop.
- Anzahl Tippfehler einer Schreibkraft pro Seite.
- Anzahl Kunden, die pro Stunde ein Kaufhaus betreten.
- Anzahl Blitze, die pro Jahr eine Fläche bestimmter Größe treffen.

Beachte: Das gilt nicht immer ganz streng!
Bsp: Zu Beginn und gegen Ende des Arbeitstags wird eine Schreibkraft vermutlich eher etwas mehr Fehler machen.

Die Anwendung ist daher meist als (oft gutes) **Modell** zu verstehen.

Kritik am Modell der Poisson-Verteilung

Am Beispiel der Datenpakete:

- **Zeitabhängigkeit** der Wahrscheinlichkeit
  Bsp: Beginn oder Ende einer Kauf-Session bei Amazon.
- **Statistische Abhängigkeit** der Datenpakete.
  Bsp: Retransmission, Protokoll-Mechanismen,
  Übertragung mehrerer Pakete pro Request, usw.
- **Erzeugender Prozeß** viel komplizierter.
  Bsp: Nicht ein Prozeß, sondern mehrere (Surfer, Mailer, Server, usw.) die Traffic nach unterschiedlichsten, voneinander abhängigen Vorgängen erzeugen.
- **Systemeffekte** werde vernachlässigt.
  Bsp: Ein Paket erzeugt Folgepakete (ARP, DNS, HTTP-request, HTTP-reply).

Daher: Muß andere, viel komplexere Verteilungen betrachten.
Aber: Dann oft nur mehr Simulation statt geschlossener Lösung.

3. Exponential-Verteilung

Ziel: Die Exponential-Verteilung beschreibt einen anderen Aspekt des Poisson-Verhaltens.

Verteilungsfunktionen

Wir betrachten einen **Poisson-Prozeß** zum Parameter λ.

Exponential-Verteilung: Kontinuierliche Verteilung.
Wahrscheinlichkeit, daß die Wartezeit im Intervall $[0, x]$ liegt:

$$R_\lambda(x) = 1 - e^{-\lambda x} = \int_0^x \rho(x)dx$$

Wahrscheinlichkeit, daß die Wartezeit im Intervall $[a, b]$ liegt:

$$Q_\lambda(a, b) = e^{-\lambda a} - e^{-\lambda b} = R_\lambda(b) - R_\lambda(a) = \int_a^b \rho(x)dx$$

Dichte der Exponential-Verteilung:

$$\rho_\lambda(x) = \lambda e^{-\lambda x} \qquad \text{für} x \geq 0$$

Exponential-Verteilung

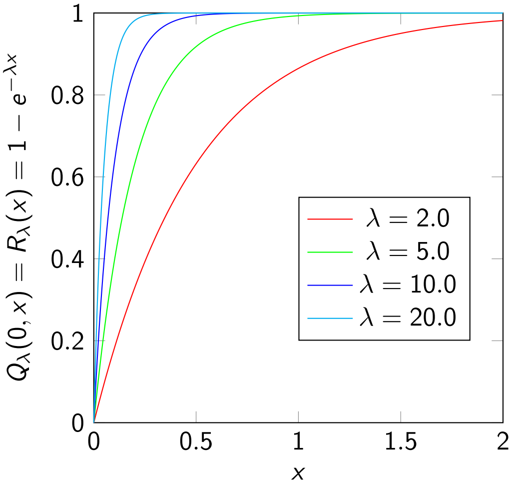

Abb. 2: Exponential-Verteilung.

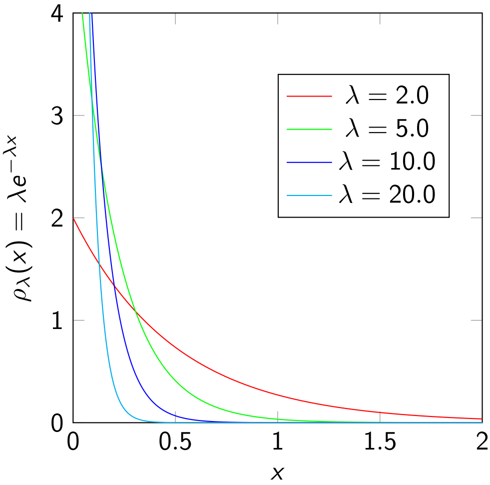

Abb. 3: Dichte-Funktion der Exponential-Verteilung.

Anhang

Übersicht

Verzeichnis aller Abbildungen

Rechtliche Hinweise

Zitierweise dieses Dokuments

Wenn Sie Inhalte aus diesem Werk nutzen oder darauf verweisen wollen, zitieren Sie es bitte wie folgt:

Clemens H. Cap: Die Poisson und die Exponential-Verteilung. Electronic document. https://iuk.one/1012-1024 25. 1. 2021.

Bibtex Information: https://iuk.one/1012-1024.bib

```
@misc{doc:1012-1024,
   author       = {Clemens H. Cap},
   title        = {Die Poisson und die Exponential-Verteilung},
   year         = {2021},
   month        = {1},
   howpublished = {Electronic document},
   url          = {https://iuk.one/1012-1024}
}
```

Typographic Information:
Typeset on January 25, 2021
This is pdfTeX, Version 3.14159265-2.6-1.40.21 (TeX Live 2020) kpathsea version 6.3.2
This is pgf in version 3.1.5b
This is preamble-slides.tex myFormat©C.H.Cap

Verzeichnis aller Folien

Legende:

- Fortsetzungsseite
- Seite ohne Überschrift
- Bildseite