② Verteilungen*

Zoltán Zomotor

Versionsstand: 23. März 2016, 11:36

Die nummerierten Felder bitte während der Vorlesung ausfüllen.

Bitte hier notieren, was beim Bearbeiten unklar geblieben ist:

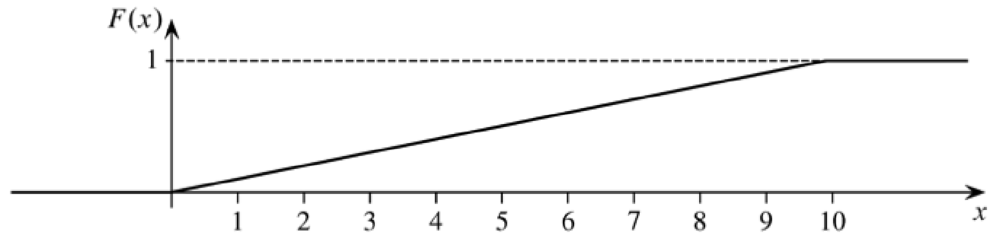
Inhaltsverzeichnis

*[Pap11, Teil II, Kapitel 6]

1 Verteilungsfunktion

- Zuweisung von Wahrscheinlichkeiten zu Realisationen.
  - Formal: 1 $F(x) = P(X \leqq x)$
- Eigenschaften:
  - $F(x) \in$ 2 $[0, 1]$
  - Definitionsbereich: $\mathbb{R}$ mit 3 $F(-\infty) = 0,\ F(+\infty) = 1$
  - monoton wachsend, d. h. 4 $x_1 < x_2 \Rightarrow F(x_1) \leqq F(x_2)$
  - Es gilt: 5 $P(a < X \leqq b) = F(b) - F(a)$
- Beispiel: Gleichverteilung zwischen 0 und 10:

6



2 Diskrete Verteilungen

2.1 Binomialverteilung

- Wiederholter Zufallsvorgang
- n Durchführungen
- Pro Durchführung: 7 A oder $\bar{A}$ ($\widehat{=}$ Ziehen mit Zurücklegen)
- 8 $X_i = \begin{cases} 1, & \text{falls } A \text{ bei } i\text{-ter Durchführung eintritt} \\ 0, & \text{falls } \bar{A} \text{ bei } i\text{-ter Durchführung eintritt} \end{cases}$

- Dann gibt

9
$$X = \sum_{i=1}^{n} X_i$$

an, wie oft A eintritt.

- Gesucht: Wahrscheinlichkeitsfunktion von X
- Herleitung:

10
1. $$P(X_i = 1) = P(A) = p$$
   $$P(X_i = 0) = P(\bar{A}) = 1 - p$$

11
2. $\sum_{i=1}^{n} x_i = x$ entspricht „x mal A und $n - x$ mal $\bar{A}$“

3. Wahrscheinlichkeit (bei Unabhängigkeit): 12 $p^x \cdot (1-p)^{n-x}$

4. Aber: Reihenfolge irrelevant! Anzahl Anordnungen: 13 $\binom{n}{x}$

⇒ Wahrscheinlichkeitsfunktion:

14
$$f(x) = \begin{cases} \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}, & \text{falls } x \in \{0, 1, \dots, n\} \\ 0, & \text{falls } x \notin \{0, 1, \dots, n\} \end{cases}$$

- Kurzschreibweise: 15 $X \sim B(n, p)$
- Wegen $F(x) = \sum_{x_i \leq x} f(x_i)$ lässt sich $f(x)$ auch berechnen als:

16
$$f(x) = F(x) - F(x-1)$$

- Beispiel: Aus einem 32-er Kartenblatt wird dreimal eine Karte mit Zurücklegen gezogen.

– Wie wahrscheinlich ist es, zweimal „Herz“ zu ziehen?

17

$$X_i = \begin{cases} 1, & \text{falls } i\text{-te Karte Herz} \\ 0, & \text{sonst} \end{cases} \qquad X_i \sim B(1, \frac{8}{32})$$

$$X = \sum_{i=1}^{n} X_i = X_1 + X_2 + X_3 \qquad X \sim B(3, \frac{1}{4})$$

– Bestimmung mithilfe der Wahrscheinlichkeitsfunktion $f(x)$:

18

$$P(X = 2) = f(2) = \binom{3}{2} \cdot 0.25^2 \cdot 0.75^1 = 0.1406$$

– Bestimmung mithilfe der Wahrscheinlichkeitsverteilung $F(x)$ (Tabelle 1):

19

$$P(X = 2) = F(2) - F(1) = 0.9844 - 0.8438 = 0.1406$$

Tabelle 1: Binomialverteilung $p = 0.25$

x \ n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	n / x
0	0,7500	0,5625	0,4219	0,3164	0,2373	0,1780	0,1335	0,1001	0,0751	0,0563	0,0422	0,0317	0,0238	0,0178	0,0134	0
1	1,0000	0,9375	0,8438	0,7383	0,6328	0,5339	0,4449	0,3671	0,3003	0,2440	0,1971	0,1584	0,1267	0,1010	0,0802	1
2		1,0000	0,9844	0,9492	0,8965	0,8306	0,7564	0,6785	0,6007	0,5256	0,4552	0,3907	0,3326	0,2811	0,2361	2
3			1,0000	0,9961	0,9844	0,9624	0,9294	0,8862	0,8343	0,7759	0,7133	0,6488	0,5843	0,5213	0,4613	3
4				1,0000	0,9990	0,9954	0,9871	0,9727	0,9511	0,9219	0,8854	0,8424	0,7940	0,7415	0,6865	4
5					1,0000	0,9998	0,9987	0,9958	0,9900	0,9803	0,9657	0,9456	0,9198	0,8883	0,8516	5
6						1,0000	0,9999	0,9996	0,9987	0,9965	0,9924	0,9857	0,9757	0,9617	0,9434	6
7							1,0000	1,0000	0,9999	0,9996	0,9988	0,9972	0,9944	0,9897	0,9827	7
8								1,0000	1,0000	1,0000	0,9999	0,9996	0,9990	0,9978	0,9958	8
9									1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9992	9
10										1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	10
11											1,0000	1,0000	1,0000	1,0000	1,0000	11
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	

2.2 Anmerkungen zur Binomialverteilung

- Es werden nur Werte für $p \leqq 0.5$ tabelliert. Falls $p > 0.5$:

20

$$P(X \leqq x) = 1 - F(n - x - 1) \text{ mit } F\colon B(n, 1 - p)$$

- Beispiel: $X \sim B(20, \frac{3}{4})$

21

$$P(X \leqq 10) = 1 - F(20 - 10 - 1) = 1 - F(9)$$
$$= 1 - 0.9861 = 0.0139 \text{ mit } F\colon B(20, \frac{1}{4})$$

- Sind $X_i \sim B(1, p)$ unabhängig, so ist 22 $\sum_i^n X_i \sim B(n, p)$
- Urnenmodell:
  - N Objekte, davon sind M Stück markiert.
  - Ziehe n Objekte *mit* Zurücklegen.
  - Anzahl gezogener Objekte mit Markierung 23 $\sim B(n, \frac{M}{N})$

2.3 Hypergeometrische Verteilung

- n-faches Ziehen *ohne* Zurücklegen aus N Objekten, davon M markiert.

  $X =$ Anzahl gezogener Objekte mit Markierung

  ist *hypergeometrisch verteilt* mit den Parametern N, M, n.
- Kurzschreibweise: 24 $X \sim Hyp(N, M, n)$
- Wahrscheinlichkeitsfunktion: 25

$$f(x) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} & \text{, falls } x \text{ möglich} \\ 0 & \text{, sonst} \end{cases}$$

- Ist $n < \frac{N}{20}$, so ist folgende Approximation möglich: 26 $Hyp(N, M, n) \approx B(n, \frac{M}{N})$
- Beispiel: Aus einem 32-er Kartenblatt wird dreimal eine Karte ohne Zurücklegen gezogen.
  - Wie wahrscheinlich ist es, zweimal „Herz“ zu ziehen? Für N, M, n und x gilt:

27 $N = 32,\ M = 8,\ n = 3,\ x = 2$

- Für die Wahrscheinlichkeit ergibt sich:

28

$$P(X=2) = f(2) = \frac{\binom{8}{2}\binom{32-8}{3-2}}{\binom{32}{3}} = \frac{\frac{8!}{2!\cdot 6!}\cdot 24}{\frac{32!}{3!\cdot 29!}}$$
$$= \frac{29!\cdot 8!\cdot 3!\cdot 24}{32!\cdot 6!\cdot 2!} = \frac{8\cdot 7\cdot 3\cdot 24}{32\cdot 31\cdot 30} = \frac{4032}{29760}$$
$$= \frac{21}{155} = 0.1355$$

$$\binom{n}{k} = \frac{n!}{k!\cdot(n-k)!},\quad \binom{n}{1} = n$$

2.4 Poisson-Verteilung

- Approximnation für $B(n,p)$ und $Hyp(N,M,n)$ für 29 $n \to \infty$ und $p \to 0$
- Faustregel: geeignet, falls 30 $n \cdot p < 10$ und $n > 1500p$
- „Verteilung der seltenen Ereignisse“ (zum Beispiel Anzahl der 6-er pro Lottoausspielung)
- Kurzschreibweise: 31 $X \sim P(\mu)$
- Wahrscheinlichkeitsfunktion: 32

$$f(x) = \begin{cases} \frac{\mu^x}{x!}\cdot e^{-\mu} & \text{, falls } x \in \{0,1,2,\dots\} \\ 0 & \text{, sonst} \end{cases}$$

- Approximationsmöglichkeiten:

33

$$Hyp(N,M,n) \xrightarrow{p=\frac{M}{N}} B(n,p) \xrightarrow{\mu=np=n\cdot\frac{M}{N}} P(\mu)$$

- Beispiel ([BBK12, Beispiel 8.18]): In einem Telefonteilnetz mit 10 000 Telefonanschlüssen sei die Wahrscheinlichkeit dafür, dass pro Tag und pro Anschluss eine Störung auftritt, gleich 0.0003. Wie groß ist die Wahrscheinlichkeit, dass die Anzahl X der Störungen pro Tag genau gleich 5 oder größer als 9 ist? Annahmen: Störungen treten unabhängig voneinander auf, höchstens eine Störung pro Tag und Anschluss.
  - Ziehen 34 mit Zurücklegen: $\Rightarrow X \sim B(10\,000, 0.0003)$

Tabelle 2: Poissonverteilung

x \ μ	1,6	1,7	1,8	1,9	2	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	3	n / μ
0	0,2019	0,1827	0,1653	0,1496	0,1353	0,1225	0,1108	0,1003	0,0907	0,0821	0,0743	0,0672	0,0608	0,0550	0,0498	0
1	0,5249	0,4932	0,4628	0,4337	0,4060	0,3796	0,3546	0,3309	0,3084	0,2873	0,2674	0,2487	0,2311	0,2146	0,1991	1
2	0,7834	0,7572	0,7306	0,7037	0,6767	0,6496	0,6227	0,5960	0,5697	0,5438	0,5184	0,4936	0,4695	0,4460	0,4232	2
3	0,9212	0,9068	0,8913	0,8747	0,8571	0,8386	0,8194	0,7993	0,7787	0,7576	0,7360	0,7141	0,6919	0,6696	0,6472	3
4	0,9763	0,9704	0,9636	0,9559	0,9473	0,9379	0,9275	0,9162	0,9041	0,8912	0,8774	0,8629	0,8477	0,8318	0,8153	4
5	0,9940	0,9920	0,9896	0,9868	0,9834	0,9796	0,9751	0,9700	0,9643	0,9580	0,9510	0,9433	0,9349	0,9258	0,9161	5
6	0,9987	0,9981	0,9974	0,9966	0,9955	0,9941	0,9925	0,9906	0,9884	0,9858	0,9828	0,9794	0,9756	0,9713	0,9665	6
7	0,9997	0,9996	0,9994	0,9992	0,9989	0,9985	0,9980	0,9974	0,9967	0,9958	0,9947	0,9934	0,9919	0,9901	0,9881	7
8	1,0000	0,9999	0,9999	0,9998	0,9998	0,9997	0,9995	0,9994	0,9991	0,9989	0,9985	0,9981	0,9976	0,9969	0,9962	8
9	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9998	0,9997	0,9996	0,9995	0,9993	0,9991	0,9989	9
10	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9998	0,9998	0,9997	10
11	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	11
12	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	12
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	

- Approximation: 35

$$\begin{array}{rcccl} n \cdot p & = & 3 & < & 10 \\ n & > & 1500 \cdot p & = & 0.45 \end{array}$$

- Bestimmung mithilfe der Wahrscheinlichkeitsfunktion $f(x)$:

36

$$P(X=5) = f(5) = \frac{3^5}{5!} e^{-3} = 0.1008188$$

- Bestimmung mithilfe der Wahrscheinlichkeitsverteilung $F(x)$ (Tabelle 2)

37

$$P(X=5) = F(5) - F(4) = 0.9161 - 0.8153 = 0.1008$$

- Exakter Wert: 38

$$P(X=5) = 0.1008239$$

3 Stetige Verteilungen

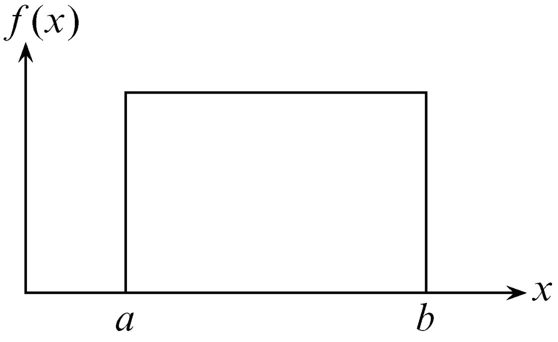
3.1 Gleichverteilung

Eine Zufallsvariable X mit 39

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{falls } a \leqq x \leqq b \\ 0 & \text{sonst} \end{cases}$$

heißt *gleichverteilt* im Intervall $[a, b]$

40



41

- Verteilungsfunktion:

$$F(x) = \begin{cases} 0 & \text{falls } x < a \\ \dfrac{x-a}{b-a} & \text{falls } a \leqq x \leqq b \\ 1 & \text{falls } x > b \end{cases}$$

- Beispiel: X ist gleichverteilt in $[1, 20]$. Wie groß ist die Wahrscheinlichkeit, dass x zwischen 2 und 12 liegt?

42

$$P(2 \leqq X \leqq 12) = \begin{aligned} F(12) - F(2) &= \frac{12-1}{20-1} - \frac{2-1}{20-1} \\ &= \frac{12-2}{20-1} = \frac{10}{19} \\ &= 0.5263 \end{aligned}$$

3.2 Exponentialverteilung

Eine Zufallsvariable X mit

43

$$f(x) = \begin{cases} \lambda \cdot \mathrm{e}^{-\lambda x} & \text{falls } x \geqq 0 \\ 0 & \text{sonst} \end{cases}$$

und $\lambda > 0$ heißt *exponentialverteilt.*

- Kurzschreibweise: 44 $X \sim Exp(\lambda)$

- Verteilungsfunktion: 45

$$F(x) = \begin{cases} 0 & \text{falls } x < 0 \\ 1 - \mathrm{e}^{-\lambda x} & \text{falls } x \geqq 0 \end{cases}$$

- Gedächtnislosigkeit: 46 $P(X \leqq t + s | X \geqq t) = P(X \leqq s)$

47

$$\begin{aligned} P(X \leqq t + s | X \geqq t) &= \frac{P(t \leqq X \leqq t + s)}{P(X \geqq t)} = \frac{F(t+s) - F(t)}{1 - F(t)} \\ &= \frac{(1 - \mathrm{e}^{-\lambda(t+s)}) - (1 - \mathrm{e}^{-\lambda t})}{1 - (1 - \mathrm{e}^{-\lambda t})} = \frac{\mathrm{e}^{-\lambda t} - \mathrm{e}^{-\lambda(t+s)}}{\mathrm{e}^{-\lambda t}} \\ &= 1 - \frac{\mathrm{e}^{-\lambda(t+s)}}{\mathrm{e}^{-\lambda t}} = 1 - \mathrm{e}^{-\lambda(t+s-t)} = 1 - \mathrm{e}^{-\lambda s} \\ &= F(s) = P(X \leqq s) \end{aligned}$$

3.3 Normalverteilung

Eine Zufallsvariable X mit

48

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

und $\sigma > 0$ heitßt *normalverteilt.*

Kurzschreibweise: 49 $X \sim N(\mu, \sigma)$

3.4 Eigenschaften der Normalverteilung

- Dichte ist achsensysymmetrisch zu μ: 50 $f(\mu - x) = f(\mu + x)$

- Maximum bei 51 $x = \mu$, Wendepunkte bei 52 $x = \mu - \sigma$ und $x = \mu + \sigma$

$\Rightarrow$ 53 μ ist Lageparameter und 54 σ ist Streuungsparameter.

- Standardnormalverteilung: 55 $N(0, 1)$ mit Verteilungsfunktion $\Phi(x)$ ([Pap11, Tabelle 1, Seite 740])

- Kenntnis von $\Phi(x)$, μ und σ genügt, denn:

56
$$X \sim N(\mu, \sigma) \Leftrightarrow \frac{X - \mu}{\sigma} \sim N(0, 1) \Rightarrow F(x) = \Phi(\frac{X - \mu}{\sigma})$$

- Nur positive x tabelliert wegen: 57 $\Phi(-x) = 1 - \Phi(x) =$

- Beispiel: Die Klausurnoten sind normalverteilt mit $\mu = 2.5$ und $\sigma = 1.0$. Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Student eine Note zwischen 1 und 2 hat?

58

$$
\begin{aligned}
P(1 \leqq X \leqq 2) &= F(2) - F(1) \\
&= \Phi(\frac{2-2.5}{1}) - \Phi(\frac{1-2.5}{1}) \\
&= \Phi(-0.5) - \Phi(-1.5) \\
&= 1 - \Phi(0.5) - 1 + \Phi(1.5) \\
&= 0.9332 - 0.6915 \\
&= 0.2417
\end{aligned}
$$

- Wahrscheinlichkeit, um höchstens c von μ abzuweichen:

59

$$
\begin{aligned}
P(\mu - c \leqq X \leqq \mu + c) &= F(\mu + c) - F(\mu - c) \\
&= \Phi(\frac{\mu + c - \mu}{\sigma}) - \Phi(\frac{\mu - c - \mu}{\sigma}) \\
&= \Phi(\frac{c}{\sigma}) - \Phi(-\frac{c}{\sigma}) \\
&= \Phi(\frac{c}{\sigma}) - (1 - \Phi(\frac{c}{\sigma})) \\
&= 2 \cdot \Phi(\frac{c}{\sigma}) - 1
\end{aligned}
$$

$\Rightarrow$ $k\sigma$-Bereiche: $[\mu - k\sigma, \mu + k\sigma]$. Setze $c = k\sigma$ (mit $k = 1, 2, \ldots$):

60

$$
P(\mu - k\sigma \leqq X \leqq \mu + k\sigma) = 2\Phi(k) - 1 = \begin{cases} 0.683 & \text{für } k = 1 \\ 0.954 & \text{für } k = 2 \\ 0.997 & \text{für } k = 3 \end{cases}
$$

- Reproduktionseigenschaft:
  - Ist X normalverteilt, so ist

    61 $Y = a + bX$ mit $b \neq 0$

    ebenfalls normalverteilt.
  - Sind $X1, \ldots, X_n$ normalverteilt, so ist

    62

    $$Z = \sum_{i=1}^{n} X_i$$

ebenfalls normalverteilt.

4 Unterschiede diskrete vs. stetige Verteilungen

	Verteilung	
	diskret	stetig
	63	
$f(x)$	Wahrscheinlichkeitsfunktion ($\sum = 1$)	Dichtefunktion ($\int = 1$)
$f \to F$	$F(x) = \sum\limits_{x_i \leqq x} f(x_i)$	$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$
$F \to f$	$f(x) = F(x) - F(x-1)$	$f(x) = F'(x)$

5 Übungsaufgaben

[Pap11, Teil II, Zu Abschnitt 6, Seite 462ff]

Literatur

[BBK12] Günter Bamberg, Franz Baur und Michael Krapp. *Statistik*. 17. Auflage. Oldenbourg Verlag, 2012.

[Pap11] Lothar Papula. *Mathematik für Ingenieure und Naturwissenschaftler*. 6. Auflage. Bd. 3. Vieweg + Teubner, 2011.