Statistik

für Betriebswirtschaft, Internationales Management, Wirtschaftsinformatik und Informatik

Raum: B3.05 oder B3.06

Sommersemester 2017

HSA Statistik SS 2017 Sessionlist

Datum	Statistik für BW/IM/I/Winf	Nr.
15.03.2017	Einführung Statistik	1
22.03.2017	Differentialrechnung, 2-dim Diff.Rechnung	2
29.03.2017	univ. deskr. Stat., Quantile, Plots	3
05.04.2017	Streuung, Konzentrationsmaße	4
12.04.2017	Kontingenztabellen, Mosaikplots, Korrelation	5
19.04.2017	Preisindizes, lineare Regression	6
26.04.2017	Wahrscheinlichkeitsbegriff	7
03.05.2017	Bedingte Wahrscheinlichkeit, Bayes	8
10.05.2017	diskrete Zufallsvariablen	9
17.05.2017	Stetige ZV, Gleichverteilung	10
24.05.2017	Pyramid	
31.05.2017	Normalverteilung, Verteilungsparameter	11
07.06.2017	Schätzfunktionen und Punktschätzer	12
14.06.2017	Konfidenzintervalle	13
21.06.2017	Wiederholung, Besprechung Probeklausur	14
28.06.2017	Prüfungswoche	15


AW-Berufungsvorlesung

Umfrage von Tamara Kartheininger | 10 | 0 | vor einer Stunde

Bitte tragt euch für den Zeitplan der AW-Berufung Politikwissenschaften, Ethik und Philoso

1. Ethik/Philosophie "Der Gerechtigkeitsbegriff nach John Rawls am Beispiel der Digitalisie
2. Politikwissenschaften - freie Themenwahl

Tabellen-Ansicht | Kalender-Ansicht

	Mai 2017 Di 23		Mi 24			
	14:30 – 15:30	16:30 – 17:30	09:00 – 10:00	11:00 – 12:00	14:00 – 15:00	16:00 – 17:00
10 Teilnehmende						

Stefan Etschberger

Hypergeometrische Verteilung

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

$$X = \text{Anzahl gezogener Objekte mit Markierung}$$

heißt hypergeometrisch verteilt mit den Parametern N, M, n.

- Kurzschreibweise: $X \sim \text{Hyp}(N; M; n)$
- Wahrscheinlichkeitsfunktion:

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

- Ist $n \leqq \frac{N}{20}$, so gilt: $\text{Hyp}(N; M; n) \approx B(n; \frac{M}{N})$

Beispiel: Hypergeometrische Verteilung

- Aus einem 32-Kartenblatt wird 3-mal eine Karte ohne Zurücklegen gezogen.
- Wie wahrscheinlich ist es, 2-mal "'Herz"' zu ziehen?
- D.h.: $N = 32,\ M = 8,\ n = 3,\ x = 2.$

$$P(X = 2) = f(2) = \frac{\binom{8}{2}\binom{32-8}{3-2}}{\binom{32}{3}} = \frac{\binom{8}{2}\binom{24}{1}}{\binom{32}{3}} = \frac{\frac{8!}{2!\cdot 6!}\cdot 24}{\frac{32!}{3!\cdot 29!}}$$

$$= \frac{29!\cdot 8!\cdot 3!\cdot 24}{32!\cdot 6!\cdot 2!} = \frac{8\cdot 7\cdot 3\cdot 24}{32\cdot 31\cdot 30} = \frac{4032}{29760} = \frac{21}{155}$$

$$= 0{,}1355$$

Dabei wurde verwendet:

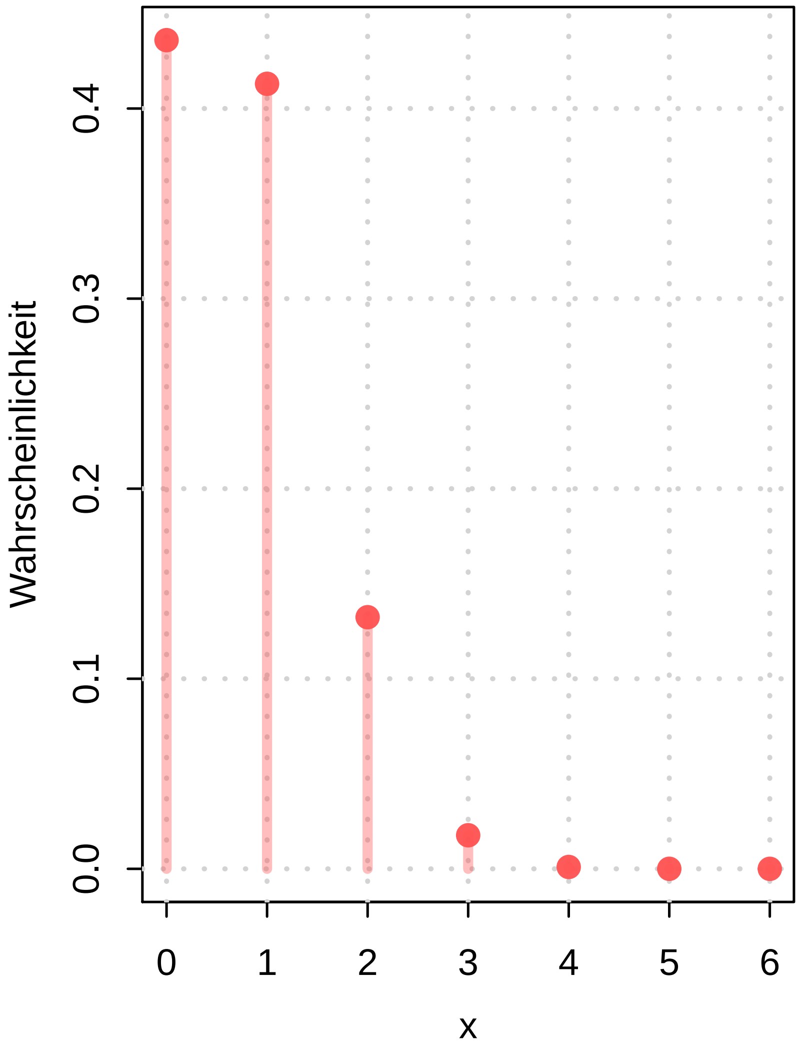
$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad \text{und} \quad \binom{n}{1} = n.$$

Hypergeometrische Verteilung

Beispiel: x Treffer im Lotto 6 aus 49

- $X \sim \text{Hyp}(49, 6, 6)$

x	$P(X = x)$ (in %)
0	43.596498
1	41.301945
2	13.237803
3	1.765040
4	0.096862
5	0.001845
6	0.000007



Poisson-Verteilung

- Approximation für $B(n;p)$ und Hyp$(N;M;n)$
- Geeignet, wenn
  p klein ($\leqq 0{,}1$), n groß ($\geqq 50$) und $np \leqq 10$.

➡ „Verteilung der seltenen Ereignisse“
(z.B. Anzahl 6-er pro Lottoausspielung)

- X ist poissonverteilt mit Parameter λ: $X \sim P(\lambda)$
- Wahrscheinlichkeitsfunktion:

$$f(x) = \begin{cases} \dfrac{\lambda^x}{x!} \cdot e^{-\lambda}, & \text{falls } x = 0,1,2,\ldots \\ 0, & \text{sonst} \end{cases}$$

- $F(x)$ in Tabelle
- Überblick: Approximation

$$\text{Hyp}(N;M;n) \xrightarrow{\;p = \frac{M}{N}\;} B(n;p) \xrightarrow{\;\lambda = np = n\frac{M}{N}\;} P(\lambda)$$

Poissonverteilung: $X \sim P(\lambda)$, Tabelle der Verteilungsfunktionen

$x \setminus \lambda$	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
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
1	0.5249	0.4933	0.4628	0.4338	0.4060	0.3796	0.3546	0.3309	0.3085	0.2873	0.2674	0.2487	0.2311	0.2146	0.1992
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
3	0.9212	0.9068	0.8913	0.8747	0.8571	0.8387	0.8194	0.7994	0.7787	0.7576	0.7360	0.7141	0.6919	0.6696	0.6472
4	0.9763	0.9704	0.9636	0.9559	0.9474	0.9379	0.9275	0.9163	0.9041	0.8912	0.8774	0.8629	0.8477	0.8318	0.8153
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
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
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
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	0.9970	0.9962
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	0.9992	0.9989
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
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	1.0000	0.9999
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

$x \setminus \lambda$	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4	4.1	4.2	4.3	4.4	4.5
0	0.0451	0.0408	0.0369	0.0334	0.0302	0.0273	0.0247	0.0224	0.0203	0.0183	0.0166	0.0150	0.0136	0.0123	0.0111
1	0.1847	0.1712	0.1586	0.1469	0.1359	0.1257	0.1162	0.1074	0.0992	0.0916	0.0845	0.0780	0.0719	0.0663	0.0611
2	0.4012	0.3799	0.3594	0.3397	0.3209	0.3028	0.2854	0.2689	0.2531	0.2381	0.2238	0.2102	0.1974	0.1852	0.1736
3	0.6248	0.6025	0.5803	0.5584	0.5366	0.5152	0.4942	0.4735	0.4533	0.4335	0.4142	0.3954	0.3772	0.3595	0.3423
4	0.7982	0.7806	0.7626	0.7442	0.7255	0.7064	0.6872	0.6679	0.6484	0.6288	0.6093	0.5898	0.5704	0.5512	0.5321
5	0.9057	0.8946	0.8829	0.8706	0.8576	0.8441	0.8301	0.8156	0.8006	0.7851	0.7693	0.7532	0.7367	0.7199	0.7029
6	0.9612	0.9554	0.9490	0.9422	0.9347	0.9267	0.9182	0.9091	0.8995	0.8893	0.8787	0.8675	0.8558	0.8437	0.8311
7	0.9858	0.9832	0.9802	0.9769	0.9733	0.9692	0.9648	0.9599	0.9546	0.9489	0.9427	0.9361	0.9290	0.9214	0.9134
8	0.9953	0.9943	0.9931	0.9917	0.9901	0.9883	0.9863	0.9840	0.9815	0.9786	0.9755	0.9721	0.9683	0.9642	0.9598
9	0.9986	0.9982	0.9978	0.9973	0.9967	0.9960	0.9952	0.9942	0.9931	0.9919	0.9905	0.9889	0.9871	0.9851	0.9829
10	0.9996	0.9995	0.9994	0.9992	0.9990	0.9987	0.9984	0.9981	0.9977	0.9972	0.9966	0.9959	0.9952	0.9943	0.9933
11	0.9999	0.9999	0.9998	0.9998	0.9997	0.9996	0.9995	0.9994	0.9993	0.9991	0.9989	0.9986	0.9983	0.9980	0.9976
12	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9998	0.9998	0.9997	0.9997	0.9996	0.9995	0.9994	0.9992
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998
14	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	0.9999
15	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

Poisson-Verteilung: Beispiel

Beispiel

- $X \sim B(10\,000; 0{,}0003)$; In Tabelle der Binomialverteilung nicht vertafelt! Approximation:

$$\left.\begin{aligned} p &= 0{,}0003 < 0{,}1 \\ n &= 10\,000 \;> 50 \\ np &= 3 \qquad\;\; < 10 \end{aligned}\right\} \Rightarrow B(10\,000; 0{,}0003) \approx P(3)$$

- Mithilfe der Wahrscheinlichkeitsfunktion:

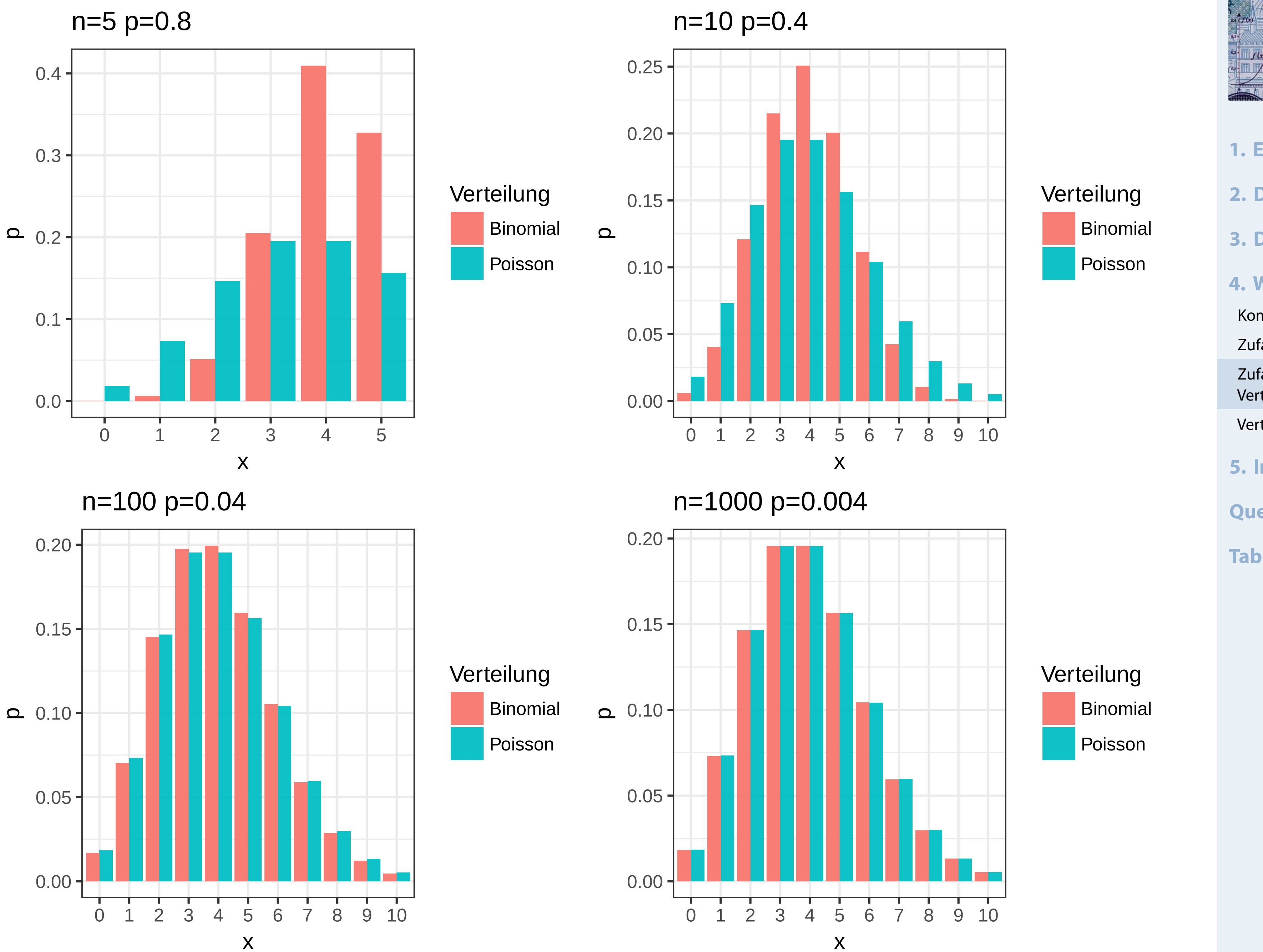
$$P(X=5) = \frac{3^5}{5!} \cdot e^{-3} = 0{,}1008188$$

- Mithilfe der Tabelle der Poissonverteilung:

$$P(X=5) = F(5) - F(4) = 0{,}9161 - 0{,}8153 = 0{,}1008$$

- Exakter Wert: $P(X=5) = 0{,}1008239$

Poisson- versus Binomialverteilung: Vergleich

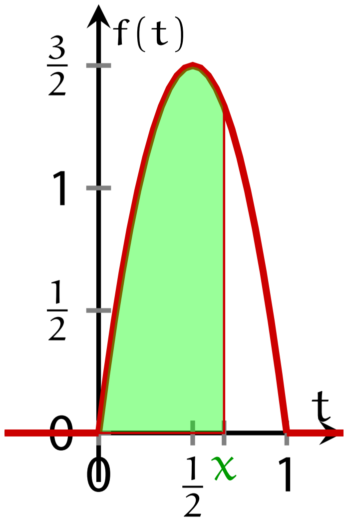


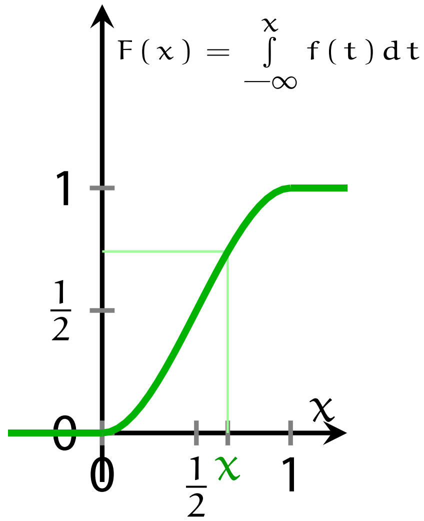
Stetige Zufallsvariablen

- X heißt stetig, wenn $F(x)$ stetig ist.
- Dann existiert ein $f(t)$ mit:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

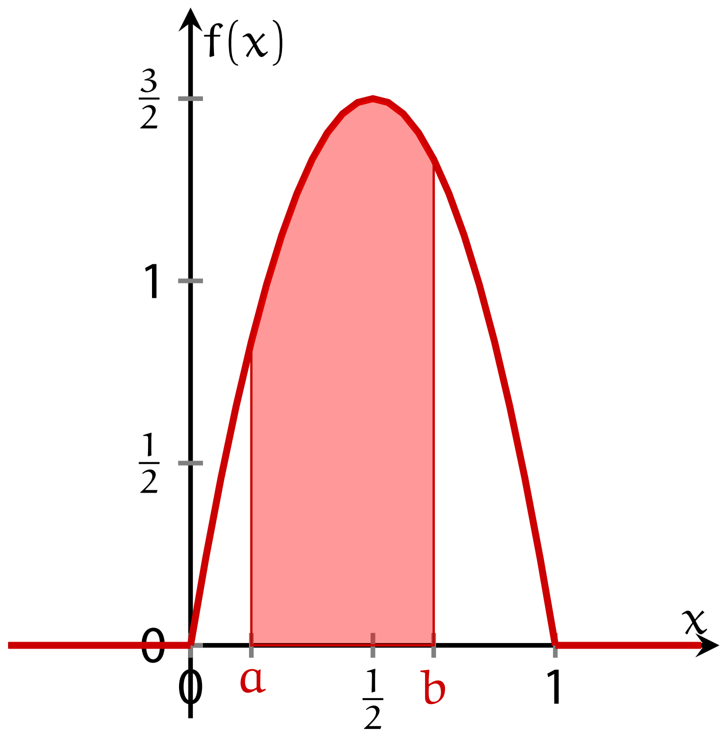
$f(x)$ heißt Dichtefunktion von X.





- Dann:

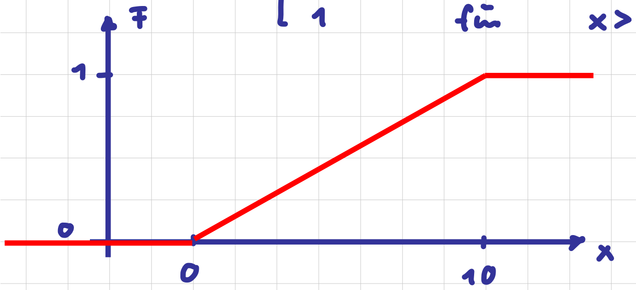
$$
\begin{aligned}
P(a < X < b) &= P(a \leqq X < b) \\
&= P(a < X \leqq b) \\
&= P(a \leqq X \leqq b) \\
&= \int_a^b f(x)\,dx \\
&= F(b) - F(a)
\end{aligned}
$$



Stetige Zufallsvariablen

Beispiel: Wartezeit bis Maschine frei: Zufallsvariable X

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0.1x & \text{für } 0 \le x \le 10 \\ 1 & \text{für } x > 10 \end{cases}$$



d.h. zum Beispiel

Wahrsch., dass Wartezeit höchstens 8 Min.

$$P(X \le 8) = F(8) = 0.1 \cdot 8 = 0.8 = 80\%$$

Zufallsvariable X heißt stetig, wenn es eine Funktion f gibt, mit

$$\boxed{F(x) = \int_{-\infty}^{x} f(t)\,dt}$$

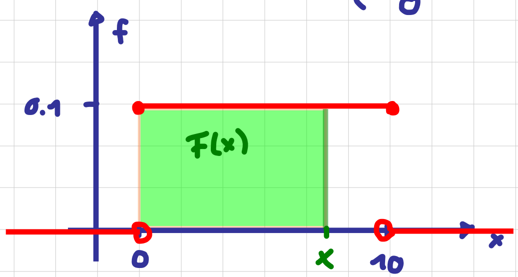
f heißt **Dichtefunktion** der ZV. X

(F ist Verteilungsfunkt.)

Beispiel: $F(x)$ wie oben

(für alle x mit $f(x)$ ist stetig gilt: $f(x) = F'(x)$)

$$\Rightarrow f(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0.1 & \text{für } 0 \le x \le 10 \\ 0 & \text{für } x > 10 \end{cases}$$



Beispiel: Zufallsvar. X: Anteil Öltankfüllung, der bis zum Zeitpunkt 1 Jahr verbraucht ist

Dichte: $f(x) = \begin{cases} ax^2 + bx + c & \text{für } 0 \le x \le 1 \\ 0 & \text{sonst} \end{cases}$

Gesucht: Parameter a, b, c

Bekannt (gilt allgemein): ① $f(x) \ge 0$ ② $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Zusatzangabe (hier) ③ $f(0) = f(1) = 0$

Setze ③ ein: $f(0) = a \cdot 0^2 + b \cdot 0 + c = 0 \iff c = 0$

$f(1) = a \cdot 1^2 + b \cdot 1 + 0 = 0 \iff a = -b$

② : $\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 ax^2 + bx + c\,dx = \int_0^1 -bx^2 + bx\,dx$

$= b \cdot \left[-\frac{1}{3}x^3 + \frac{1}{2}x^2\right]_0^1 = b \cdot \left(-\frac{1}{3} + \frac{1}{2}\right) = b \cdot \frac{1}{6} = 1$

$\iff b = 6 \quad (a = -6,\ c = 0)$

$$\Rightarrow f(x) = \begin{cases} -6(x^2 - x) & \text{für } 0 \le x \le 1 \\ 0 & \text{sonst} \end{cases}$$

[Maximum von $f(0.5) = -6 \cdot (0.25 - 0.5) = 1.5$]

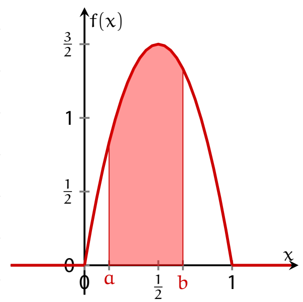
Beobachtung: $f(x)$ kann größer als 1 sein

damit möglich

$P(0.2 \le X \le 0.6) = F(0.6) - F(0.2)$

$\int_{0.2}^{0.6} (-6x^2 + 6x)\,dx = \left[-2x^3 + 3x^2\right]_{0.2}^{0.6}$

$= (-2 \cdot 0.6^3 + 3 \cdot 0.6^2) - (-2 \cdot 0.2^3 + 3 \cdot 0.2^2) \approx 0.544$



Dichtefunktion

Eigenschaften der Dichtefunktion

- $f(x) \geqq 0$ für alle $x \in \mathbb{R}$
- Wegen $F(\infty) = 1$ muss stets gelten:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

- $P(X = x) = 0$ für alle $x \in \mathbb{R}$
- $f(x) > 1$ ist möglich
- für $x \in \mathbb{R}$ ist $F(x)$ differenzierbar $\Rightarrow F'(x) = f(x)$.
- Intervallgrenzen spielen keine Rolle:

$$\begin{aligned} P(X \in [a;b]) &= P(X \in (a;b]) \\ &= P(X \in [a;b)) \\ &= P(X \in (a;b)) \\ &= F(b) - F(a) \end{aligned}$$

Dichtefunktion: Beispiel

Beispiel

$$f(x) = \begin{cases} 0, & \text{falls} \quad x < 0 \\ \frac{1}{10}, & \text{falls} \quad 0 \leqq x \leqq 10 \\ 0, & \text{falls} \quad x > 10 \end{cases}$$

Verteilungsfunktion:

$$\int_0^x f(t)\,dt = \int_0^x \frac{1}{10}\,dt = \left[\frac{t}{10}\right]_0^x = \frac{x}{10} \Rightarrow$$

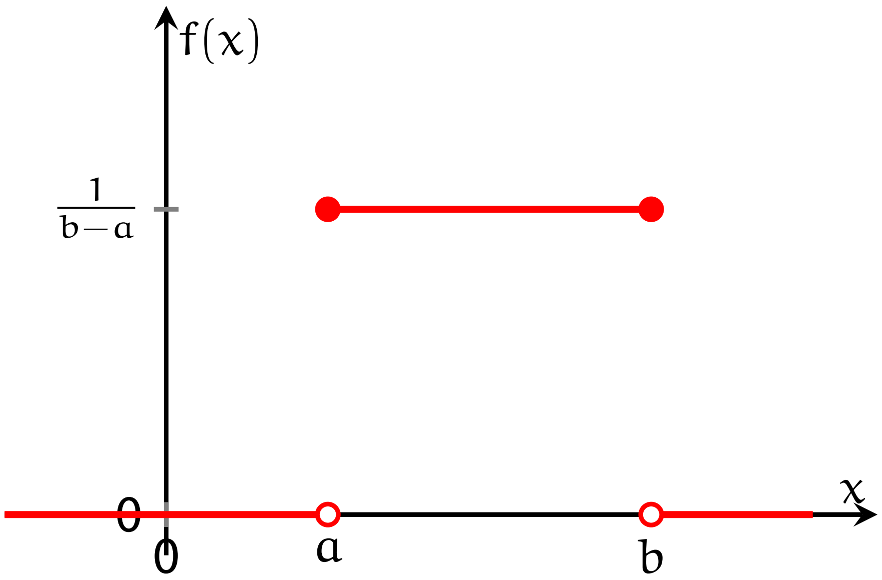
$$F(x) = \begin{cases} 0, & \text{falls} \quad x < 0 \\ \frac{x}{10}, & \text{falls} \quad 0 \leqq x \leqq 10 \\ 1, & \text{falls} \quad x > 10 \end{cases}$$

Gleichverteilung

Eine Zufallsvariable X mit

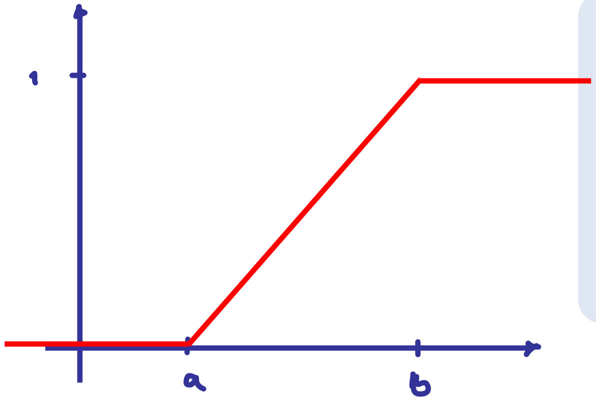
$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{falls } a \leqq x \leqq b \\ 0\;, & \text{sonst} \end{cases}$$

heißt gleichverteilt im Intervall $[a; b]$.



Gleichverteilung

- **Verteilungsfunktion der Gleichverteilung:**



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

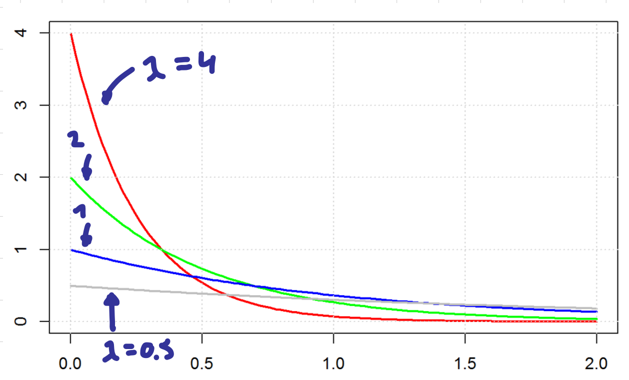
- **Beispiel:** X gleichverteilt in $[1;20]$

$$P(2 \leqq X \leqq 12) = F(12) - F(2) = \frac{12-1}{20-1} - \frac{2-1}{20-1}$$
$$= \frac{12-2}{20-1} = \frac{10}{19}$$
$$= 0{,}5263$$

Exponentialverteilung Zufallsvariable mit Dichte

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

heißt **exponentialverteilt**



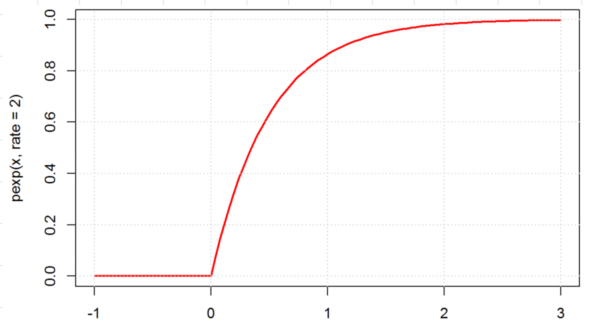
Beispiel: geg.: X ist exponentialverteilt mit $\lambda = 2$

gesucht: Verteilungsfunktion $F(x)$

$$F(x) \underset{x \geq 0}{=} \int_{-\infty}^{x} f(t)\,dt = \int_0^x 2e^{-2t}\,dt = \left[(-1)\cdot e^{-2t}\right]_0^x$$

$$= -e^{-2x} - (-e^{-2\cdot 0}) = 1 - e^{-2x}$$

$$\Rightarrow \quad F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 - e^{-2x} & \text{für } x \geq 0 \end{cases}$$



Anwendungsbeispiele:

- Zeit zwischen zwei Anrufen in Callcenter
- Zeit bis zum nächste Serverausfall
- Service zeit (Abfertigungszeit für Kunde)

Beispielaufgabe aus Bamberg, Günter, Franz Baur und Michael Krapp (2017). Statistik - Arbeitsbuch. 10. Aufl. De Gruyter Oldenbourg.
E-Book innerhalb des Hochschulnetzes frei erhältlich:
https://goo.gl/JtnOxL

Aufgabe 2.30

Das monatliche Einkommen betrage mindestens x_0 Geldeinheiten, wobei x_0 durch Tarifverträge, das soziale Netz und dergleichen bestimmt wird. Zur approximativen Beschreibung der Einkommensverteilung wird häufig eine Dichtefunktion der Form

$$f(x) = \begin{cases} \frac{c}{x^{\alpha+1}} & \text{für } x \geqq x_0 \\ 0 & \text{sonst} \end{cases}$$

benutzt, wobei α ein positiver vorgegebener Parameter und c eine noch zu bestimmende Normierungskonstante ist.

a) Bestimmen Sie c.
b) Bestimmen Sie die Verteilungsfunktion F des Einkommens X.
c) Setzen Sie speziell $\alpha = 1$ sowie $x_0 = 1\,000$ Euro und berechnen Sie die bedingte Wahrscheinlichkeit

$$P(X > 10\,000 | X \geqq 5\,000) ,$$

d. h. den Anteil derjenigen unter den mindestens 5 000 Euro Verdienenden, die sogar über 10 000 Euro verdienen.

a) $\int_{-\infty}^{\infty} f(x)\,dx = \int_{x_0}^{\infty} \frac{c}{x^{\alpha+1}}\,dx = c \cdot \int_{x_0}^{\infty} x^{-(\alpha+1)}\,dx$

$= c \cdot \left[\frac{1}{-\alpha-1+1} x^{-\alpha-1+1} \right]_{x_0}^{\infty}$

$= c \cdot \left[\frac{1}{-\alpha} x^{-\alpha} \right]_{x_0}^{\infty} = c \cdot \left(\frac{1}{-\alpha} \cdot 0 - \left(\frac{1}{-\alpha} \cdot x_0^{-\alpha} \right) \right)$

$= \frac{c}{\alpha} x_0^{-\alpha} = 1$

$\Leftrightarrow \quad c = \alpha \cdot x_0^{\alpha}$

$\Rightarrow \quad f(x) = \begin{cases} \frac{\alpha x_0^{\alpha}}{x^{\alpha+1}} & \text{für } x \geq x_0 \\ 0 & \text{sonst} \end{cases}$

b), c) $\rightarrow$ Hausaufgabe