Grundlagen der Mathematischen Statistik

Diskrete Verteilungen

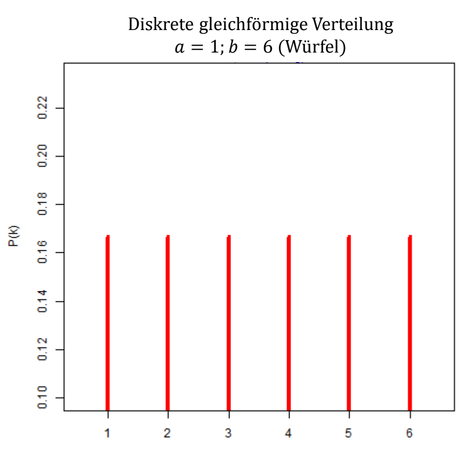
Uwe Menzel, 2018

uwe.menzel@matstat.org

www.matstat.org

1

Diskrete gleichförmige Verteilung



$X \sim U(a, b)$ Kurzbezeichnung

$\Omega = \{a, a+1, a+2, \ldots, b\}$

$N = b - a + 1$ Anzahl der Elementarereignisse

$p_X(k) = \dfrac{1}{N}$ Wahrscheinlichkeits-funktion

$E(X) = \dfrac{a+b}{2}$ Erwartungswert

$V(X) = \dfrac{(b-a+2)\cdot(b-a)}{12}$ Varianz

<u>Beispiel</u> Würfel: Zufallsvariable $X =$ Augenzahl;

$a = 1; b = 6; N = 6;\ p_X(k) = {}^{1}/_{6}\ ; E(X) = 3.5$

www.matstat.org

2

”bi”

Die Binomialverteilung

Zufallsversuch: **zwei** alternative Ereignisse: A, A^* (Bernoulli-Versuch)

$$\left.\begin{aligned} P(A) &= p \\ P(A^*) &= 1 - p \end{aligned}\right\}$$ p sei bekannt - wird man oft als ”Erfolgswahrscheinlichkeit” bezeichnet

Beispiel 1: Münze Ereignis A: Kopf, Ereignis A^*: Zahl $\Big\}$ $P(A) = P(A^*) = p = \frac{1}{2}$

Beispiel 2: Würfel Ereignis A: Augenzahl fünf, Ereignis A^*: andere Augenzahl $\Big\}$ $P(A) = p = \frac{1}{6}$, $P(A^*) = 1 - p = \frac{5}{6}$

Wir machen nun n unabhängige Wiederholungen dieses Versuches! Wir suchen die Wahrscheinlichkeit dass wir k mal Ereignis A erhalten! ($k \leq n$)
(und daher $n - k$ mal Ereignis A^*)

Beispiel 1, Münze: werfe 10 Mal und suche W., 6 Mal ”Kopf” zu erhalten ($n = 10; k = 6;\ p = {}^1/_2$)

Beispiel 2, Würfel: werfe 20 Mal und suche W. 3 Fünfen zu erhalten ($n = 20; k = 3; p = {}^1/_6$)

www.matstat.org

3

Die Binomialverteilung

$X \sim Bin(n, p)$

Wir machen n unabhängige Bernoulli-Versuche:

	A	A^*	Summe
Wahrscheinlichkeit	p	$1-p$	1
Anzahl	k	$n-k$	n

Zufallsvariable *X*: Die Anzahl der A-Ereignisse (”Erfolge”) bei insgesamt n Bernoulli- Versuchen

$\Omega_X = \{0, 1, 2, \ldots, n\}$ Ereignisraum

Welche **Wahrscheinlichkeitsfunktion** hat die Zufallsvariable X ?

$$P(X = k) = p_X(k) = ?$$

www.matstat.org

4

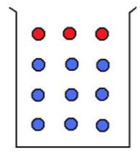
Die Binomialverteilung

$X \sim Bin(n,p) \qquad P(X=k) = p_X(k) = ?$

Wir versuchen, die Wahrscheinlichkeitsfunktionen mit Hilfe eines Beispiels zu finden:

Urne:

- 3 rote Kugeln
- 7 blaue Kugeln



Zufallsversuch:

- nimm eine Kugel, registriere dessen Farbe, lege zurück
- Ereignis A : rote Kugel ("Erfolg") $P(A) = p = 0.3$
- Ereignis A^*: blaue Kugel $P(A^*) = 1 - p = 0.7$
- wiederhole dies $n = 4$ mal → 4 Bernoulli-Versuche

Zufallsvariable *X*: Anzahl roter Kugeln (bei insgesamt 4 Ziehungen) ∩ bedeutet "und"

$$P(X=4) = P[(A_1 = rot) \cap (A_2 = rot) \cap (A_3 = rot) \cap (A_4 = rot)]$$

$$= P(A_1 = rot) \cdot (A_1 = rot) \cdot (A_1 = rot) \cdot (A_1 = rot) \quad \text{wegen Unabhängigkeit}$$

$$= p \cdot p \cdot p \cdot p = p^4 = 0.3^4$$

www.matstat.org

5

Die Binomialverteilung

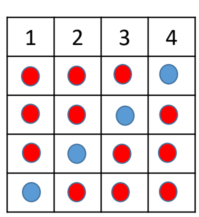
3 mal rot

$$P(X=3) = 4 \cdot p^3 \cdot (1-p)$$

1 mal blau

Warum der Faktor 4 ?

Siehe Beispiel "drei elektronische Komponenten" (Vorlesung **F3**): es gibt 4 Möglichkeiten, genau eine blaue und 3 rote Kugeln zu erhalten:



4 Möglichkeiten für Reihenfolge

Es kann die 1. Möglichkeit **oder** die 2. **oder** die 3. **oder** die 4. eintreffen → da die Ereignisse mit **oder** verknüpft und disjunkt sind, addieren sich deren W.

www.matstat.org

6

Die Binomialverteilung

$X \sim Bin(n,p) \qquad P(X=k) = p_X(k) = ?$

$P(X=2) = \binom{4}{2} \cdot p^2 \cdot (1-p)^2$ 2 rote, 2 blaue → $\binom{4}{2}$ Möglichkeiten der Anordnung (wähle 2 von 4)

$P(X=1) = \binom{4}{1} \cdot p \cdot (1-p)^3$ 1 rot, 3 blaue → $\binom{4}{1}$ Möglichkeiten der Anordnung (wähle 1 von 4)

$P(X=0) = \binom{4}{0} \cdot (1-p)^4$ 0 rote, 4 blaue → $\binom{4}{0} = 1$ Möglichkeit der Anordnung

allgemein: $\boxed{P(X=k) = p_X(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \qquad k = 0, 1, 2, \dots, n}$

Wenn eine Zufallsvariable X eine solche Wahrscheinlichkeitsfunktion hat, sagt man dass X **binomialverteilt** ist.

- Zufallsvariable X: Anzahl "Erfolge" (ohne moralische Wertung!)
- $P(X=k)$: W., dass die Zufallsvariable X den Wert k annimmt, also Wahrscheinlichkeit für k "erfolgreiche" Versuche
- n: totale Anzahl der Bernoulli-Versuche
- $p = P(A)$: Wahrscheinlichkeit für "Erfolg" in jedem einzelnen Bernoulli-Versuch

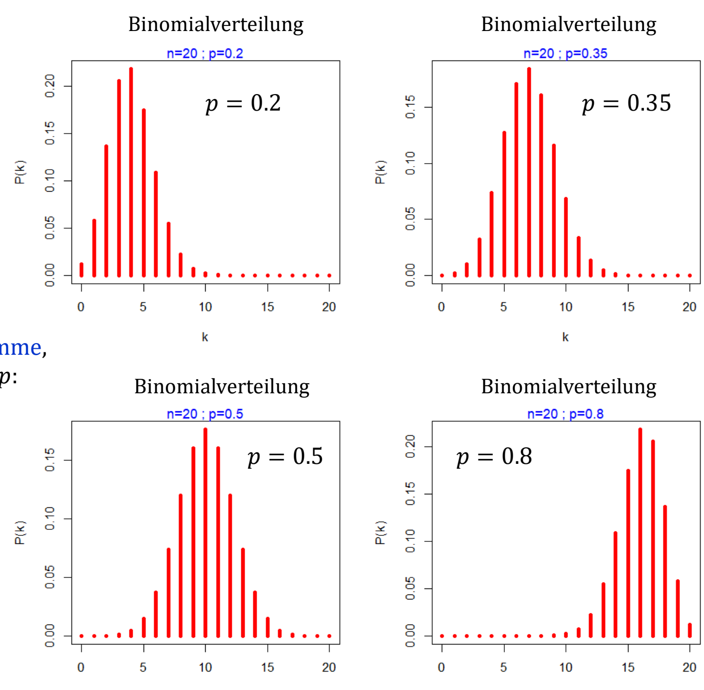
Kurzbezeichnung: $X \sim Bin(n,p)$ Ereignisraum: $\Omega_X = \{0, 1, 2, \dots, n\}$

www.matstat.org

7

$X \sim Bin(n,p)$

Säulendiagramme, verschiedene p:



8

Die Binomialverteilung

Beispiel: Münzwurf, 20 mal, mit P(Kopf) = 0.5.
Wahrscheinlichkeit 8 Mal "Kopf" zu werfen?:

Zufallsvariable X: Anzahl "Kopf" bei insgesamt 20 Würfen

$X \sim Bin(20, 0.5)$

$$P(X = 8) = p_X(8) = \binom{20}{8} \cdot 0.5^8 \cdot 0.5^{12} = 0.12$$

R $P = dbinom(8, size = 20, prob = 0.5)$

www.matstat.org

9

Die Binomialverteilung

Normierung:

$$\sum_{k=0}^{n} p_X(k) = \sum_{k=0}^{n} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = 1$$

Beweis mit Binomialtheorem
siehe Anhang

Erwartungswertt: $E(X) = \sum_{k=0}^{n} k \cdot p_X(k) = \ ... = n \cdot p$

Beispiel Münzwürf: $p = {}^1/_2\,; n = 10 \ \rightarrow \ E(X) = 5$

Erwartungswert, Varianz, Standardabweichung:

$$\boxed{\begin{array}{lcl} E(X) & = & n \cdot p \\ V(X) & = & n \cdot p \cdot (1-p) \\ D(X) & = & \sqrt{V(X)} \end{array}}$$

www.matstat.org

10

Die Binomialverteilung

Verteilungsfunktion:

$$\boxed{F_X(a) = P(X \leq a) = \sum_{k=0}^{a} p_X(k)}$$

a: reelle Zahl
keine analytische Formel möglich

Die Verteilungsfunktionen für $\mathrm{Bin}(n, p)$ ist **tabelliert:**

- nur für eine gewisse Auswahl von p
- meist nur für $n = 2, 3, \ldots, 20$

Zur Erinnerung: Die Formel $p_X(k) = F_X(k) - F_X(k-1)$ kann angewendet werden, um mit Hilfe der Verteilungsfunktion F_X die Wahrscheinlichkeits-funktion p_X auszurechnen.

www.matstat.org

11

Tabell 6. Binomialfördelningen

$P(X \leq x)$ där $X \in \mathrm{Bin}(n, p)$.
För $p > .5$ utnyttja att $P(X \leq x) = P(Y \geq n - x)$ där $Y \in \mathrm{Bin}(n, 1 - p)$

Tabelle:

$X \sim Bin(n, p)$

n	x \ p	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50
2	0	.90250	.81000	.72250	.64000	.56250	.49000	.36000	.25000
	1	.99750	.99000	.97750	.96000	.93750	.91000	.84000	.75000
3	0	.85737	.72900	.61412	.51200	.42188	.34300	.21600	.12500
	1	.99275	.97200	.93925	.89600	.84375	.78400	.64800	.50000
	2	.99987	.99900	.99662	.99200	.98438	.97300	.93600	.87500
4	0	.81451	.65610	.52201	.40960	.31641	.24010	.12960	.06250
	1	.98598	.94770	.89048	.81920	.73828	.65170	.47520	.31250
	2	.99952	.99630	.98802	.97280	.94922	.91630	.82080	.68750
	3	.99999	.99990	.99949	.99840	.99609	.99190	.97440	.93750
5	0	.77378	.59049	.44371	.32768	.23730	.16807	.07776	.03125
	1	.97741	.91854	.83521	.73728	.63281	.52822	.33696	.18750
	2	.99884	.99144	.97339	.94208	.89648	.83692	.68256	.50000
	3	.99997	.99954	.99777	.99328	.98438	.96922	.91296	.81250
	4	1.00000	.99999	.99992	.99968	.99902	.99757	.98976	.96875
6	0	.73509	.53144	.37715	.26214	.17798	.11765	.04666	.01562
	1	.96723	.88574	.77648	.65536	.53394	.42017	.23328	.10938
	2	.99777	.98415	.95266	.90112	.83057	.74431	.54432	.34375
	3	.99991	.99873	.99411	.98304	.96240	.92953	.82080	.65625
	4	1.00000	.99995	.99960	.99840	.99536	.98906	.95904	.89063
	5	1.00000	1.00000	.99999	.99994	.99976	.99927	.99590	.98438
7	0	.69834	.47830	.32058	.20972	.13348	.08235	.02799	.00781
	1	.95562	.85031	.71658	.57672	.44495	.32942	.15863	.06250
	2	.99624	.97431	.92623	.85197	.75641	.64707	.41990	.22656

12

Die Binomialverteilung

<u>Beispiel</u>: $X \sim Bin(5, 0.2)$, d.h. $n = 5$ und $p = 0.2$

Suche: Wahrscheinlichkeit für 3 "Erfolge" (d.h. 3 mal das Ereignis mit $p = 0.2$)

a) $p_X(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$ wird gebraucht für $n = 5$; $k = 3$; $p = 0.2$

$$p_X(3) = \binom{5}{3} \cdot 0.2^3 \cdot 0.8^2 = \frac{5!}{3! \cdot 2!} \cdot 0.2^3 \cdot 0.8^2 = \frac{4 \cdot 5}{2} \cdot 0.2^3 \cdot 0.8^2 = \underline{\underline{0.0512}}$$

b) **oder** mit Hilfe der Tabelle für die Verteilungsfunktion:

$p_X(3) = F_X(3) - F_X(2) = 0.99328 - 0.94208 = \underline{\underline{0.0512}}$ (siehe Tabelle oben)

große n, k : Tabelle meist einfacher und schneller (wenn die Tabelle p enthält)

c) noch einfacher: R

dbinom : Wahrscheinlichkeitsfunktion ("probability density")
pbinom : Verteilungsfunktion ("probability distribution")

dbinom(x, size, prob)

3 5 0.2

dbinom(3, 5, 0.2)=$\underline{\underline{0.0512}}$

uwe.menzel@matstat.org

13

Die Binomialverteilung

Verteilung für komplementäre Ereignisse:

$P(A) = p$ $Z.v.$ X: $Anzahl\ A$ $X \sim Bin(n, p)$

$P(A^*) = 1 - p$ $Z.v.$ Y: $Anzahl\ A^*$ $Y \sim Bin(n, 1-p)$

$$X + Y = n$$

Anzahl A   Anzahl A*   Anzahl der Versuche

Zusammenhang zwischen den Wahrscheinlichkeitsfunktionen für X und Y:

$$\boxed{P(X = k) = P(Y = n - k)}$$

Trat k mal Ereignis A ein, dann muss $(n - k)$ mal Ereignis A^* eingetreten sein.
<u>Beispiel</u> Münze: 10 kast, 3 mal Kopf → 7 mal Zahl

www.matstat.org

14

Die Binomialverteilung

Verteilung für komplementäre Ereignisse:

$P(A) = p$ $\quad$ Z.v. X: Anzahl A $\quad$ $X \sim Bin(n, p)$

$P(A^*) = 1 - p$ $\quad$ Z.v. Y: Anzahl A^* $\quad$ $Y \sim Bin(n, 1 - p)$

Zusammenhang zwischen den Verteilungsfunktionen für X und Y:

$$F_X(k) = P(X \leq k) = P(n - Y \leq k) = P(-Y \leq k - n) = P(Y \geq n - k)$$
$$= 1 - P(Y \leq n - k - 1) = 1 - F_Y(n - k - 1)$$

$$\Rightarrow \boxed{F_X(k) = 1 - F_y(n - k - 1)}$$

$$\Rightarrow \boxed{P(X \leq k) = P(Y \geq n - k)}$$

uwe.menzel@matstat.org

15

Die Binomialverteilung

Zusammenhang zwischen den Verteilungsfunktionen für X und Y:

Beispiel: $n = 5$; $\quad k = 2$

X	Y
0	5
1	4
2	3
3	2
4	1
5	0

$$P(X \leq k) = P(Y \geq n - k)$$
$$\Downarrow$$
$$P(X \leq 2) = P(Y \geq 3)$$

www.matstat.org

16

Die Binomialverteilung

Zusammenhang zwischen den Verteilungsfunktionen für X und Y:

Beispiel: $X \sim Bin(n, p)$ mit $P(A) = p = 0.7\,; n = 5 \quad \rightarrow \quad X \sim Bin(5, 0.7)$

Suche: $F_X(1)$

a) nutze $F_X(k) = 1 - F_y(n - k - 1)$ und Tabelle

!! Keine Werte für $p = 0.7$ in der Tabelle, daher wenden wir $F_Y(k)$ an:

$$F_X(1) = 1 - F_Y(\overset{n}{5} - \overset{k}{1} - 1) = 1 - F_Y(3) \quad \text{där } Y \sim Bin(5, 0.3)$$

$$= 1 - 0.969 = \underline{\underline{0.031}}$$

b) nutze $F_X(a) = \sum_{k=0}^{a} p_X(k)$

$$F_X(1) = p_X(0) + p_X(1) = \binom{5}{0} \cdot 0.7^0 \cdot 0.3^5 + \binom{5}{1} \cdot 0.7^1 \cdot 0.3^4 = \underline{\underline{0.031}}$$

www.matstat.org

17

Die Binomialverteilung

Additionssatz für die Binomialverteilung:

$$\left.\begin{array}{l} X \sim Bin(n_1, p) \\ Y \sim Bin(n_2, p) \end{array}\right\} \quad X + Y \sim Bin(n_1 + n_2, p)$$

!! Die Zufallsvariablen X und Y müssen denselben Parameter p haben!

- erst n_1 unabhängige Bernoulli-Versuche mit W. p
- dann n_2 unabhängige Bernoulli-Versuche mit W. p

} dasselbe wie $n_1 + n_2$ Bernoulli-Versuche mit einem Mal!

www.matstat.org

18

Die Binomialverteilung

<u>Beispiel</u>:
Ein (Bernoulli-) Versuch "glückt" mit der Wahrscheinlichkeit $p = 0.8$.
Man führt 12 unabhängige Versuche aus.

a) Welche Verteilung hat die Zufallsvariable $X =$ Anzahl geglückter Versuche ?

$$X \sim Bin(12, 0.8)$$

b) Welche Verteilung hat die Zufallsvariable $Y =$ Anzahl missglückter Versuche ?

$$Y \sim Bin(12, 0.2)$$

Vorsicht Falle!

c) Berechne $P(2 \leq Y \leq 4)$

Tabelle

$$P(2 \leq Y \leq 4) = P(1 < Y \leq 4) = F_X(4) - F_X(1) = 0.927 - 0.275 = \underline{\underline{0.652}}$$

19

Die Binomialverteilung

<u>Beispiel</u>:

Ein (Bernoulli-) Versuch "glückt" mit der Wahrscheinlichkeit $p = 0.8$.
Man führt 12 unabhängige Versuche durch.

d) Berechne die Wahrscheinlichkeit dass die Anzahl geglückter Versuche größer als 7, jedoch nicht größer als 10 ist.

$$P(7 < X \leq 10) = F_X(10) - F_X(7)$$

$p = 0.8$ nicht in Tabelle, wende Y an!

$$= 1 - F_Y(12 - 10 - 1) - 1 + F_Y(12 - 7 - 1) = F_Y(4) - F_Y(1)$$

$$= 0.927 - 0.275 = \underline{\underline{0.652}}$$

!! Dies ist dasselbe wie in Frage c), denn

$X \in \{8, 9, 10\}$ bedeutet gleichzeitig $Y \in \{2, 3, 4\}$

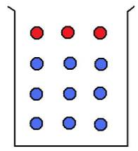
X	Y	Σ
8	4	12
9	3	12
10	2	12

20

Die Hypergeometrische Verteilung

Wir versuchen wieder, die Verteilung mit Hilfe eines Beispiels zu verstehen:

<u>Urne</u>:
- 3 rote Kugeln
- 7 blaue Kugeln



<u>Zufallsversuch</u>:
- ziehe Kugel, ohne Zurücklegen (Unterschied zu $Bin(n, p)$!!)
- ziehe insgesamt n Kugeln

Wenn die Kugeln **nicht** zurückgelegt werden:

$$P(rot\ in\ 2.\ Ziehung \mid rot\ in\ 1.\ Ziehung) = {}^{2}/_{9}$$

$$P(rot\ in\ 2.\ Ziehung \mid blau\ in\ 1.\ Ziehung) = {}^{3}/_{9}$$

Die einzelnen Ziehungen sind **nicht** mehr unabhängig → **keine** Binomialverteilung !

Zufallsvariable X = Anzahl roter Kugeln unter den n gezogenen Kugeln. Wir suchen die **Wahrscheinlichkeitsfunktion**:

$$P(X = k) = p_X(k) = ?$$

www.matstat.org

21

Die Hypergeometrische Verteilung

Anmerkung: Wir hatten

$$P(rot\ in\ 2. \mid rot\ in\ 1.) = \frac{2}{9}$$

$$P(rot\ in\ 2. \mid blai\ in\ 1{:}\, a) = \frac{3}{9}$$

Dies bedeutet jedoch **nicht,** dass die totale W., eine rote Kugel in der 2. Ziehung zu erhalten, verschieden ist von der W. eine rote Kugel in der 1. Ziehung zu erhalten (was man vielleicht vermuten könnte):

Zur Erinnerung, **totale Wahrscheinlichkeit**:

$$P(rot\ in\ 2.) = P(rot\ in\ 2. \mid rot\ in\ 1.) \cdot P(rot\ in\ 1.) + P(rot\ in\ 2. \mid blau\ in\ 1.) \cdot P(blau\ in\ 1.)$$

$$= \frac{2}{9} \cdot \frac{3}{10} + \frac{3}{9} \cdot \frac{7}{10} = \frac{6 + 21}{90} = \frac{27}{90} = \frac{3}{10} = \underline{\underline{P(rot\ i\ 1{:}\, a)}}$$

Die Zufallsvariable X ist deshalb **nicht binomialverteil,** weil die einzelnen Ziehungen **nicht unabhängig sind** (was eine Voraussetzung für das Vorliegen der Binomialverteilung ist)!

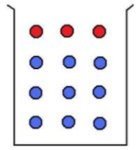
www.matstat.org

22

Die Hypergeometrische Verteilung

Urne:

- 3 rote Kugeln
- 7 blaue Kugeln



Zufallsversuch:

- ziehe Kugel, ohne Zurücklegen (Unterschied zu $Bin(n,p)$)
- ziehe insgesamt n Kugeln (hier $n = 5$)

Zufallsvariable X = Anzahl der roten Kugeln unter den n gezogenen $\Omega_X = \{0, 1, 2, 3\}$

Die **Wahrscheinlichkeitsfunktion** kann mittels der klassischen Wahrscheinlichkeits-definition gefunden werden: $P = {}^g/_m$
Angenommen, wir haben 3 rote und 7 blaue Kugeln; und ziehen insgesamt 5 Kugeln. Wie groß ist die Wahrscheinlichkeit, dass unter diesen 5 Kugeln genau 2 rote sind?:

$$P(X=2) = \frac{g}{m} = \frac{\binom{3}{2}\cdot\binom{7}{3}}{\binom{10}{5}}$$

$\binom{3}{2}$ = Anzahl der Möglichkeiten, 2 von den 3 roten zu ziehen

$\binom{7}{3}$ = Anzahl der Möglichkeiten, 3 von den 7 blauen zu ziehen (wenn 2 rote von insgesamt 5 Kugeln gezogen wurden, dann müssen 3 blaue Kugeln gezogen worden sein!)

$\binom{10}{5}$ = Anzahl der Möglichkeiten, 5 von insgesamt 10 Kugeln zu ziehen

www.matstat.org

23

Die Hypergeometrische Verteilung

Wahrscheinlichkeitsfunktion $P(X=k)$:

Zufallsvariable X = Anzahl der "roten" unter den n gezogenen

$$\boxed{P(X=k) = \frac{\binom{m}{k}\cdot\binom{N-m}{n-k}}{\binom{N}{n}}}$$

$\Omega_X = \{0, 1, 2, \dots, \min(n, m)\}$
N : totale Anzahl der "Kugeln" in der Urne
n: Anzahl der gezogenen Kugeln
m: Anzahl der "roten Kugeln" in der Urne
k: Anzahl der gezogenen "roten Kugeln"

$k \le m$

Anzahl der gezogenen roten Kugeln — Anzahl der roten K. in der Urne

$n - k \le N - m$

Anzahl der blauen gezogenen Kugeln — Anzahl der blauen Kugeln in der Urne

R : dhyper

Wenn diese Restriktionen nicht erfüllt sind, wird $P(X=k)$ nach obenstehender Formel automatisch Null, denn es gilt $\binom{m}{k} = 0$ wenn $k > m$.

www.matstat.org

24

Die Hypergeometrische Verteilung

Kurzbezeichnung: $X \sim Hyp(N, n, m)$

N: Gesamtanzahl der Objekte
n: Anzahl der gezogenen Objekte
m: Anzahl der Objekte der 1. Sorte

Natürlich gilt all dies nicht nur für Kugeln, die aus einer Urne gezogen werden. Die Verteilung gilt allgemein für das Auswählen ohne Zurücklegen aus einem Reservoir mit zwei Sorten von Objekten.

Erwartungswert, Varianz, Standardabweichung:

Sei die Zufallsvariable X hypergeometrisch verteilt, also $X \sim Hyp(N, m, n)$
Es gilt :

$$\boxed{\begin{aligned} E(X) &= n \cdot p \\ V(X) &= n \cdot p \cdot (1-p) \cdot \frac{N-n}{N-1} \\ D(X) &= \sqrt{V(X)} \end{aligned}}$$

$p = m/N$ Anteil der "roten", also der Objekte der 1. Sorte im Reservoir (z.B. Urne)

Die Hypergeometrische Verteilung

Approximation mittels Binomialverteilung:

Annahme: $n/N < 0.1$ "Daumenregel": es werden nicht mehr als ca. 10% aller Objekte im Reservoir gezogen. Dann kann man nähern

$$Hyp(N, m, n) \rightarrow Bin\left(n, \frac{m}{N}\right) \qquad p = \frac{m}{N} \text{ (Anteil "roter Kugeln/Objekte Sorte 1")}$$

z.B. $N = 10000 \,; n = 10 \rightarrow \frac{N-n}{N-1} = \frac{9990}{9999} \approx 1$

$\Rightarrow$

$V(X) = n \cdot p \cdot (1-p) \cdot \frac{N-n}{N-1} \approx n \cdot p \cdot (1-p)$ ... wie für $Bin(n, p)$

Die Poisson-Verteilung

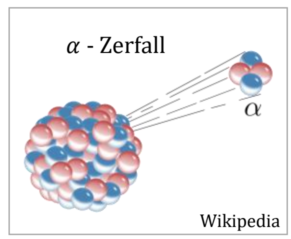
☆ = Ereignis ("event")

Vorkommen und Voraussetzungen:

- Die Poisson-Verteilung beschreibt die Wahrscheinlichkeit, dass eine gewisse Anzahl von zufälligen Ereignissen in einem festgelegten Zeit- oder Längen-intervall eintritt.
- Es wird angenommen, dass (in Zeit oder Raum) aufeinander folgende Ereignisse ("Events") unabhängig eintreffen, d.h. wenn ein oder mehrere Ereignisse gerade eingetroffen sind, so hat dies keinen Einfluss auf die Anzahl der Ereignisse in einem späteren Intervall (es ist kein "Gedächtnis" vorhanden)

Beispiele für Ereignisse:

- Anfragen an einen Server (unanhängig in der Zeit?)
- Unfälle auf der A9 (unanhängig in Zeit/Raum ?)
- Eintreffen von e-Mails (unanhängig in der Zeit ?)
- Eintreffen von Kunden an einer Kasse
- radioaktiver Zerfallsprozess (unanhängig in der Zeit!)



www.matstat.org

27

Die Poisson-Verteilung $X \sim Po(\mu)$

- **Zufallsvariable** X : Anzahl der Ereignisse in einem bestimmten Zeit- oder Längenintervall
- **Wahrscheinlichkeitsfunktion** : $P(X=k) = p_X(k)$ = Wahrscheinlichkeit, dass in dem Intervall genau k Ereignisse eintreffen

$$p_X(k) = \frac{\mu^k}{k!} e^{-\mu} \qquad k = 0, 1, 2, 3, \ldots$$ (μ muss dimensionslos sein)

$$p_X(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad k = 0, 1, 2, 3, \ldots$$ (λ hat Dimenion 1/Zeit)

$$p_X(k) = \frac{(\lambda l)^k}{k!} e^{-\lambda l} \qquad k = 0, 1, 2, 3, \ldots$$ (λ hat Dimenion 1/Länge)

Erwartungswert, Varianz, Standardabweichung:

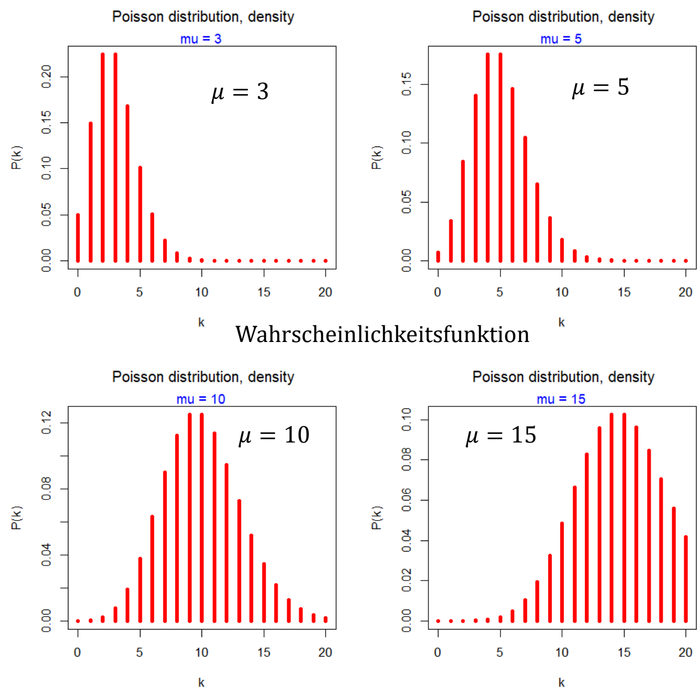
$$\begin{aligned} E(X) &= \mu \\ V(X) &= \mu \\ D(X) &= \sqrt{\mu} \end{aligned}$$

$E(X)$: "Durchschnitt" für ein Zeit- oder Längenintervall

www.matstat.org

28

$X \sim Po(\mu)$



29

Die Poisson-Verteilung $X \sim Po(\mu)$

Zufallsvariable X : Anzahl der Ereignisse in einem bestimmten Zeit- oder Längenintervall → **Normierung**:

$$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} \frac{\mu^k}{k!} \cdot e^{-\mu} = e^{-\mu} \cdot \underbrace{\sum_{k=0}^{\infty} \frac{\mu^k}{k!}}_{= e^{\mu} \text{ (Reihenentwicklung)}} = e^{-\mu} \cdot e^{\mu} = \underline{\underline{1}} \quad ✓$$

Erwartungswert :

Subst.: $m = k - 1$

$$E(X) = \sum_{k=0}^{\infty} k \cdot p_X(k) = \sum_{k=0}^{\infty} k \cdot \frac{\mu^k}{k!} \cdot e^{-\mu} = e^{-\mu} \cdot \sum_{k=1}^{\infty} \frac{\mu \cdot \mu^{k-1}}{(k-1)!}$$

$$= e^{-\mu} \cdot \mu \cdot \underbrace{\sum_{m=0}^{\infty} \frac{\mu^m}{m!}}_{= e^{\mu} \text{ (Reihenentwicklung)}} = e^{-\mu} \cdot \mu \cdot e^{\mu} = \underline{\underline{\mu}}$$

www.matstat.org

30

Die Poisson-Verteilung $X \sim Po(\mu)$

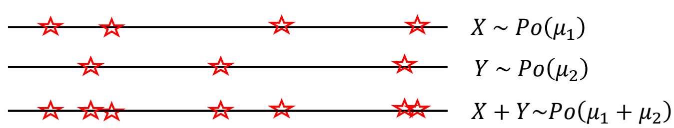
Zufallsvariable X : Anzahl der Ereignisse in einem bestimmten Zeit- oder Längenintervall → **Verteilungsfunktion:**

$$F_X(a) = P(X \leq a) = \sum_{k=0}^{a} p_X(k) = \sum_{k=0}^{a} \frac{\mu^k}{k!} \cdot e^{-\mu}$$ Tabelle!

Additionseigenschaft:

$$\left.\begin{array}{l} X \sim Po(\mu_1) \\ Y \sim Po(\mu_2) \end{array}\right\} X + Y \sim Po(\mu_1 + \mu_2)$$

Vorsicht!: Die Differenz $X - Y$ ist **nicht** Poisson-verteilt!



www.matstat.org

Verteilungsfunktion:

Tabelle:

$X \sim Po(\mu)$

Tabell 5. Poissonfördelningen
$P(X \leq x)$ där $X \in \mathrm{Po}(\mu)$.

x \ μ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	.90484	.81873	.74082	.67032	.60653	.54881	.49659	.44933	.40657
1	.99532	.98248	.96306	.93845	.90980	.87810	.84420	.80879	.77248
2	.99985	.99885	.99640	.99207	.98561	.97688	.96586	.95258	.93714
3	1.00000	.99994	.99973	.99922	.99825	.99664	.99425	.99092	.98654
4		1.00000	.99998	.99994	.99983	.99961	.99921	.99859	.99766
5			1.00000	1.00000	.99999	.99996	.99991	.99982	.99966
6					1.00000	1.00000	.99999	.99998	.99996
7							1.00000	1.00000	1.00000

x \ μ	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6
0	.36788	.30119	.24660	.20190	.16530	.13534	.11080	.09072	.07427
1	.73576	.66263	.59183	.52493	.46284	.40601	.35457	.30844	.26738
2	.91970	.87949	.83350	.78336	.73062	.67668	.62271	.56971	.51843
3	.98101	.96623	.94627	.92119	.89129	.85712	.81935	.77872	.73600
4	.99634	.99225	.98575	.97632	.96359	.94735	.92750	.90413	.87742
5	.99941	.99850	.99680	.99396	.98962	.98344	.97509	.96433	.95096
6	.99992	.99975	.99938	.99866	.99743	.99547	.99254	.98841	.98283
7	.99999	.99996	.99989	.99974	.99944	.99890	.99802	.99666	.99467
8	1.00000	1.00000	.99998	.99995	.99989	.99976	.99953	.99914	.99851
9			1.00000	.99999	.99998	.99995	.99990	.99980	.99962
10				1.00000	1.00000	.99999	.99998	.99996	.99991
11						1.00000	1.00000	.99999	.99998
12								1.00000	1.00000

x \ μ	2.8	3.0	3.2	3.4	3.6	3.8	4.0	4.2	4.4
0	.06081	.04979	.04076	.03337	.02732	.02237	.01832	.01500	.01228

Die Poisson-Verteilung

Skalierung:

$$p_X(k) = \frac{\mu^k}{k!} \cdot e^{-\mu}$$

μ muss dimensionslos sein !

Manchmal ist:

$\mu = \lambda \cdot t$ (für einen Prozess in der Zeit) oder
$\mu = \lambda \cdot l$ (für eine Prozess im Raum)

mit den Einheiten: $[\lambda] = \frac{1}{s}$ bzw. $[\lambda] = \frac{1}{m}$

$$\Rightarrow \quad p_X(k) = \frac{(\lambda \cdot t)^k}{k!} \cdot e^{-\lambda \cdot t}$$

www.matstat.org

33

Die Poisson-Verteilung

Beispiel: Eintreffende e-Mails

Ein Student bekommt durchschnittlich1.5 e-Mails pro Stunde. Wir nehmen an, dass die Anzahl der e-Mails pro Zeiteinheit Poisson-verteilt ist.
Zufallsvariable X: Anzahl e-Mails im Zeitintervall.

$$X \sim Po(\mu) \qquad \mu = \lambda \cdot t \qquad \lambda = \frac{1.5}{h}$$

a) Wahrscheinlichkeit, dass keine e-Mail innerhalb einer Stunde kommt?

$$t = 1h \qquad \mu = \lambda \cdot t = \frac{1.5}{h} \cdot 1\text{h} = 1.5 \qquad k = 0 \ \text{(keine Mail)}$$

$$p_X(0) = \frac{\mu^0}{0!} \cdot e^{-1.5} = e^{-1.5} \cong \underline{\underline{0.2231}}$$

R: dpois(0, 1.5)

b) Wahrscheinlichkeit, dass mehr als 3 Mails innerhalb von zwei Stunden kommen?:

$$t = 2h \qquad \mu = \lambda \cdot t = \frac{1.5}{h} \cdot 2\text{h} = 3$$

(Erwartungswert für 2 Stunden!)

$$P(X > 3) = 1 - P(X \leq 3) = 1 - F_X(3) \ \text{wobei} \ X \sim Po(3)$$

Tabelle

$$= 1 - 0.6472 = \underline{\underline{0.3528}}$$

R: ppois(3, 3) berechnet $F_X(3)$

www.matstat.org

34

Die Poisson-Verteilung

Beispiel: Eintreffende e-Mails (Fortsetzung)

Ein Student bekommt durchschnittlich1.5 e-Mails pro Stunde. Wir nehmen an, dass die Anzahl der e-Mails pro Zeiteinheit Poisson-verteilt ist.
Zufallsvariable X: Anzahl e-Mails im Zeitintervall.

$$X \sim Po(\mu) \qquad \mu = \lambda \cdot t \qquad \lambda = \frac{1.5}{h}$$

c) Wahrscheinlichkeit, dass der Student höchstens 10 Mails innerhalb eines Arbeitstages von 8 Stunden bekommt?

$$t = 8h \qquad \mu = \lambda \cdot t = \frac{1.5}{h} \cdot 8\text{h} = 12$$ (Erwartungswert für 8 Stunden!)

$$P(X \leq 10) = F_X(10) \text{ där } X \sim Po(12)$$ Tabelle

$$= \underline{\underline{0.3472}}$$

R: ppois(10, 12) berechnet $F_X(10)$

www.matstat.org

35

Die Poisson-Verteilung

Beispiel: Kernphysik, Versuch I

Eine radioaktive Substanz emittiert durchschnittlich 8 Partikel pro Sekunde. Die Anzahl der pro Sekunde emittierten Partikel ist Poisson-verteilt.
Zufallsvariable X: Anzahl der Emissionen pro Zeiteinheit.

$$X \sim Po(\mu) \qquad \mu = \lambda \cdot t \qquad \lambda = \frac{8}{s}$$

a) Wahrscheinlichkeit, dass 7 Partikel innerhalb einer bestimmten Sekunde emittiert werden?

$$t = 1s \qquad \mu = \lambda \cdot t = \frac{8}{s} \cdot 1\text{s} = 8 \qquad k = 7 \text{ (sieben Partikel)}$$

$$P(X = 7) = p_X(7) = \frac{8^7}{7!} \cdot e^{-8} = \frac{2097152}{5040} \cdot 3.355 \cdot 10^{-4} \cong \underline{\underline{0.1395}}$$

oder: $P(X = 7) = p_X(7) = F_X(7) - F_X(6) = 0.453 - 0.3134 = \underline{\underline{0.1396}}$ Tabelle

oder: R dpois(7,8) $\rightarrow$ 0.1395865

www.matstat.org

36

Die Poisson-Verteilung

<u>Beispiel</u>: Kernphysik, Versuch I (Fortsetzung)
Eine radioaktive Substanz emittiert durchschnittlich 8 Partikel pro Sekunde.
Zufallsvariable X: Anzahl der Emissionen pro Zeiteinheit.

$$X \sim Po(\mu) \qquad \mu = \lambda \cdot t \qquad \lambda = \frac{8}{s}$$

b) Wahrscheinlichkeit, dass höchstens 6 Partikel in einer Sekunde emittiert werden?:

$$t = 1s \qquad \mu = \lambda \cdot t = \frac{8}{s} \cdot 1s = 8$$

$$P(X \le 6) = F_X(6) = \underline{\underline{0.3134}}$$

c) Wahrscheinlichkeit, dass mehr als 8 Partikel pro Sekunde emittiert werden?:

$$P(X > 8) = 1 - P(X \le 8) = 1 - F_X(8) \text{ där } X \sim Po(8) \qquad \text{Tabelle}$$
$$= 1 - 0.5925 = \underline{\underline{0.4075}}$$

d) W. , dass mindestens 6 und höchstens 10 Partikel pro Sekunde emittiert werden?:

$$P(6 \le X \le 10) = P(5 < X \le 10) = F_X(10) - F_X(5) \text{ där } X \sim Po(8) \qquad \text{Tabelle}$$
$$= 0.8159 - 0.1912 = \cdots$$

www.matstat.org $\quad P = ppois(10{,}8) - ppois(5{,}8)$

37

Die Poisson-Verteilung

<u>Beispiel</u>: Kernphysik, Versuch II

Es ist bekannt, dass $X \sim Po$. Man hat viele Sekunden-Intervalle vermessen und dabei festgestellt, dass die Wahrscheinlichkeit, **kein** Signal pro Sekunde zu erhalten gleich 0.07 ist. Welche Anzahl von Signalen pro Sekunde kommt am häufigsten vor ?

$$X \sim Po(\mu) \qquad \mu = ?? \text{ (wir versuchen zunächst, } \mu \text{ zu berechnen)}$$

$$p_X(0) = 0.07 = \frac{\mu^0}{0!} \cdot e^{-\mu} = e^{-\mu} \rightarrow -\mu = \ln(0.07) \rightarrow \mu = 2.66$$

$$p_X(1) = \frac{\mu^1}{1!} \cdot e^{-\mu} = 2.66 \cdot e^{-2.66} = 0.186$$

$$p_X(2) = \frac{\mu^2}{2!} \cdot e^{-\mu} = \frac{2.66^2}{2} \cdot e^{-2.66} = \underline{\underline{0.247}} \quad \Leftarrow \quad \text{am größten!}$$

$$p_X(3) = \frac{\mu^3}{3!} \cdot e^{-\mu} = \frac{2.66^3}{6} \cdot e^{-2.66} = 0.219$$

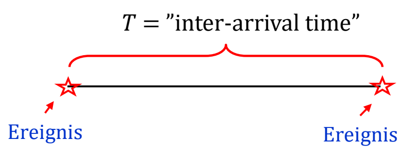
<u>Antwort</u>: Die am häufigsten vorkommende Anzahl ist 2 Emissionen pro Sekunde.

www.matstat.org

38

Die Poisson-Verteilung

Wartezeit ("Inter-arrival time") für Poisson-Prozess: $X \sim Po(\mu)$ $X \sim Po(\lambda \cdot t)$



Zur Erinnerung: die Ereignisse geschehen unabhängig (zufällig) in Zeit oder Raum (bei T kann es sich auch um eine Länge handeln!) Daher ist T eine Zufallsvariable.

Wie ist T verteilt?

Die Wahrscheinlichkeit dass T größer als eine Zeit/Länge t ist, also $P(T > t)$, ist gleich der Wahrscheinlichkeit dass kein Ereignis im Intervall t eintritt, also gleich $p_X(0)$ bezogen auf das Intervall t.

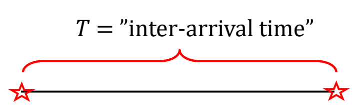
$\Rightarrow$ $P(T > t) = p_X(0 \mid t)$ $p_X(0 \mid t)$ = Wahrscheinlichkeit, dass 0 Ereignisse im Intervall t eintreten

www.matstat.org

39

Die Poisson-Verteilung

Väntetid (Inter-arrival time) för Poisson process: $X \sim Po(\lambda \cdot t)$



$\Rightarrow$ $P(T > t) = p_X(0 \mid t)$ $p_X(0 \mid t)$ = Wahrscheinlichkeit, dass 0 Ereignisse im Intervall t eintreten

$$p_X(k) = \frac{(\lambda \cdot t)^k}{k!} \cdot e^{-\lambda \cdot t}$$ W., dass k Ereignisse im Zeitraum t eintreffen

$$P(T > t) = p_X(0) = e^{-\lambda \cdot t} \Rightarrow P(T \leq t) = 1 - e^{-\lambda \cdot t}$$

$\Rightarrow$ $\boxed{F_T(t) = 1 - e^{-\lambda \cdot t}}$ Dies ist die Verteilungsfunktion für die Exponentialverteilung.

<u>Antwort</u>: Die "Wartezeit" T ist exponentialverteilt (siehe Vorlesung **F5**).

www.matstat.org

40

Die Geometrische Verteilung

Man wiederholt einen Versuch welcher mit der Wahrscheinlichkeit p "glückt". Man fährt solange fort bis der Versuch zum 1. Mal glückt.
Zufallsvariable X: Anzahl der Versuche bis zum ersten geglückten Versuch.
Beispiel: Man wirft eine Münze so lange bis man "Kopf" erhält ("Glück" = "Kopf"). Wie groß ist die Wahrscheinlichkeit dass man 1 Mal, 2 Mal, 3 Mal, ... usw. werfen muss?

Wahrscheinlichkeitsfunktion:

$$P(X = k) = p_X(k) = \underbrace{(1-p)^{k-1}}_{(k-1)\text{ missglückte Versuche}} \cdot p$$

geglückt im k-ten Versuch

Ereignisraum:

$$\Omega_X = \{1, 2, 3, 4, \ldots . .\}$$

Kurzbezeichnung: $X \sim Ge(p)$

Anm: X wird manchmal als Anzahl missglückter Versuche bis zum 1. geglückten definiert!

$\Rightarrow$ $p_X(k) = (1-p)^k \cdot p$ $\quad k = 0, 1, 2, \ldots$ siehe Wikipedia "Geometric distribution"

Die Geometrische Verteilung

$X \sim Ge(p)$

Väntevärde:

$$E(X) = \sum_{k=1}^{\infty} k \cdot p_X(k) = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \cdot p = p \cdot \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1}$$

Substitution: $q = 1 - p$

$$E(X) = (1-q)\sum_{k=1}^{\infty} k \cdot q^{k-1} = (1-q)\sum_{k=0}^{\infty} k \cdot q^{k-1} = (1-q)\sum_{k=0}^{\infty} \frac{d}{dq} q^k$$

$$= (1-q)\frac{d}{dq}\sum_{k=0}^{\infty} q^k = (1-q)\frac{d}{dq}\left(\frac{1}{1-q}\right) = (1-q) \cdot \frac{-1}{(1-q)^2} \cdot (-1)$$

$$= \frac{1}{1-q} = \underline{\underline{\frac{1}{p}}}$$

$$\boxed{E(X) = \frac{1}{p}}$$

Ist p klein, dann ist im Schnitt eine größere Anzahl von Versuchen erforderlich um den 1. geglückten Versuch zu erhalten

Die Geometrische Verteilung

$X \sim Ge(p)$

Varianz :

$$\boxed{V(X) = \frac{1-p}{p^2}}$$

p kleiner $\to$ größere Varianz

Standardabweichung :

$$D(X) = \sqrt{V(X)}$$

Verteilungsfunktion :

siehe auch Anhang

$P(X > k) =$ Wahrsch. für Misserfolg in den ersten k Versuchen $= (1-p)^k$

$$F_X(k) = P(X \leq k) = 1 - P(X > k) = 1 - \underbrace{(1-p)}_{q = 1 - p \text{ (Definition)}}{}^k = 1 - q^k$$

$$\boxed{F_X(k) = 1 - q^k}$$

www.matstat.org

43

Die Geometrische Verteilung

Beispiel: Würfel

Wie groß ist die Wahrsch. dass die erste Fünf im 4. Wurf kommt?:

Zufallsvariable X: Anzahl der Würfe bis zur 1. Fünf $\qquad X \sim Ge(p)$

$p = \dfrac{1}{6}$ Wahrsch. dass Wurf glückt = Wahrsch. dass Fünf kommt

$$P(X = 4) = p_X(4) = \underbrace{(1-p)^3}_{\text{3 missglückte Versuche}} \cdot p = \left(\frac{5}{6}\right)^3 \cdot \frac{1}{6} = \underline{\underline{0.096}}$$

www.matstat.org

44

Anhang

Diskrete Verteilungen

Uwe Menzel, 2018
uwe.menzel@matstat.org
www.matstat.org

Die Geometrische Verteilung

$P(X > k) =$ W. dass die ersten k Versuche missglücken $= (1-p)^k$

$$\left.\begin{aligned} P(X > k) = &(1-p)^k \cdot p \\ &+ (1-p)^k \cdot (1-p) \cdot p \\ &+ (1-p)^k \cdot (1-p)^2 \cdot p \\ &+ (1-p)^k \cdot (1-p)^3 \cdot p \\ &+ \ldots \end{aligned}\right\}$$

... alle möglichen Ereignisse nach k missglückten Versuchen

$q = 1 - p$ (definition)

$= (1-p)^k \cdot [p + (1-p) \cdot p + (1-p)^2 \cdot p + (1-p)^3 \cdot p + \ldots]$

$= (1-p)^k \cdot p \cdot [1 + q + q^2 + q^3 + q^4 + \ldots]$ geometrische Reihe

$= (1-p)^k \cdot p \cdot \frac{1}{1-q} = (1-p)^k \cdot p \cdot \frac{1}{p} = (1-p)^k$ wie zuvor

www.matstat.org

Binomialverteilung; Normierung

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \cdot y^{n-k}$$ Binomialtheorem

$\Downarrow$ setze: $x = p$, $y = 1 - p$

$$1 = \sum_{k=0}^{n} \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

www.matstat.org

47