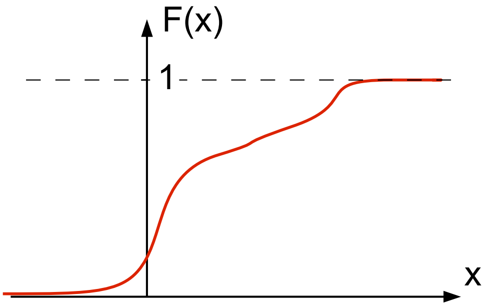
Statistik und Wahrscheinlichkeitsrechnung

Kapitel 05: Wichtige Zufallsverteilungen

Prof. Dr. Adrian Ulges

B.Sc. *Informatik*
Fachbereich DCSM
Hochschule RheinMain

Von Verteilungen zu Verteilungstypen Bild: [6]



- Verteilungsfunktionen F verraten uns wie Zufallsvariablen verteilt sind. *Aber: Wie ermitteln wir F?*
- Annahme: F gehört zu einem bestimmten Verteilungstyp.
- Im folgenden Kapitel werden wir uns mit einigen wichtigen Verteilungstypen auseinandersetzen, jeder mit anderen *Anwendungen, Parametern und Eigenschaften*.

Outline

Die Bernoulli-Verteilung Bild: [8]

Definition (Bernoulli-Variable)

Wir bezeichnen einen Zufallsvariable mit nur zwei möglichen Realisierungen (üblicher Weise 0 und 1) als **Bernoulli-Variable**.

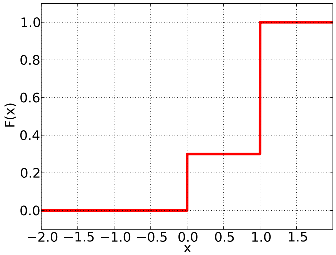
Anmerkungen

- Wir bezeichnen das Ereignis $X=1$ als "Erfolg" und $X=0$ als "Misserfolg".

Beispiele

- Münzwurf
- Defekt von Bauteilen
- …

Die Bernoulli-Verteilung: Eigenschaften



Parameter

(nur) die Erfolgswahrscheinlichkeit $p := P(X = 1)$

Erwartungswert und Varianz

$$E(X) = p$$

$$Var(X) = p \cdot (1 - p)$$

Verteilungsfunktion

$$F(x) = \begin{cases} 0 & \text{falls } x < 0 \\ 1 - p & \text{falls } 0 \leq x < 1 \\ 1 & \text{falls } 1 \leq x \end{cases}$$

Die Binomial-Verteilung Bilder: [2] [4]

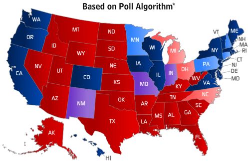
Aus der Bernoulli-Verteilung leiten wir die wichtige Binomial-Verteilung her:

- Wir betrachten n unabhängige Bernoulli-Variablen $X_1, \ldots, X_n$.
- Die Verteilung der Bernoulli-Variablen sei stationär, d.h. sie ändert sich zwischen den Wiederholungen nicht:

$$P(X_1{=}1) = P(X_2{=}1) = \ldots = P(X_n{=}1) = p$$

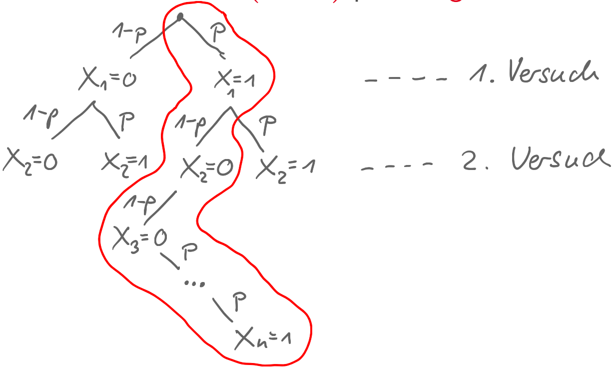
- Die neue Zufallsvariable X misst die Anzahl der erfolgreichen Versuche:

$$X := X_1 + X_2 + \ldots + X_n$$



Die Binomial-Verteilung: Herleitung

Um die Verteilung zu berechnen, leiten wir $P(X = k)$ per Ereignisbaum her:

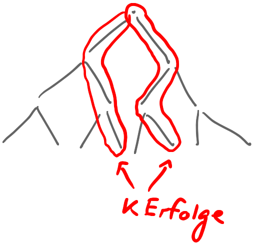


1. Wahrscheinlichkeit eines konkreten Pfades, z.B. (1,0,0,1,0):

$$= \underbrace{p \cdot p \cdot p \cdot \ldots \cdot p}_{k \text{ Stück}} \cdot \underbrace{(1-p) \cdot (1-p) \cdot \ldots \cdot (1-p)}_{n-k \text{ Stück}} = p^k \cdot (1-p)^{n-k}$$

Die Binomial-Verteilung: Herleitung

2. Wieviele Pfade mit k erfolgreichen Versuchen gibt es?



$$\underbrace{\circ\circ\circ}_{k \text{ Erfolge}}\;\underbrace{\bullet\bullet\bullet\bullet\bullet}_{n-k \text{ Misserfolge}}$$

Wieviele Permutationen?

→ Permutation mit Duplikate!

$$P(n;\, k,\, n-k) = \frac{n!}{k! \cdot (n-k)!} = \binom{n}{k}$$

Die Binomial-Verteilung

Definition (Binomial-Verteilung)

Eine diskrete Zufallsvariable X mit der Verteilung

$$P(X = k) = \begin{cases} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} & \text{falls } 0 \leq k \leq n \\ 0 & \text{sonst} \end{cases}$$

nennen wir binomial-verteilt.

Die Binomial-Verteilung: Do-it-yourself

Berechnen Sie die Binomialverteilung für $n = 3$ und $p = \frac{3}{4}$.
Skizzieren Sie die zugehörige Verteilung als Säulendiagramm.

$$\binom{3}{0} = 1 = \binom{3}{3}$$

$$\binom{3}{1} = 3 = \binom{3}{2}$$

k	0	1	2	3
$\binom{n}{k}$	1	3	3	1
$P(X = k)$	$1/64$	$9/64$	$27/64$	$27/64$

$$P(X=0) = \binom{3}{0} \cdot \left(\frac{3}{4}\right)^0 \cdot \left(\frac{1}{4}\right)^3 = \frac{1}{64}$$

$$P(X=1) = \binom{3}{1} \cdot \left(\frac{3}{4}\right)^1 \cdot \left(\frac{1}{4}\right)^2 = 9/64$$

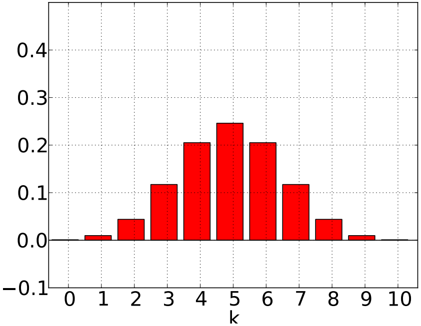
$$P(X=2) = \binom{3}{2} \cdot \left(\frac{3}{4}\right)^2 \cdot \left(\frac{1}{4}\right)^1 = 27/64$$

$$P(X=3) = \binom{3}{3} \cdot \left(\frac{3}{4}\right)^3 \cdot \left(1/4\right)^0 = 27/64$$

Die Binomial-Verteilung: Beispiel ($n = 10, p = 0.5$)

$$P(X = k) = \binom{10}{k} \cdot 0.5^k \cdot 0.5^{n-k}$$

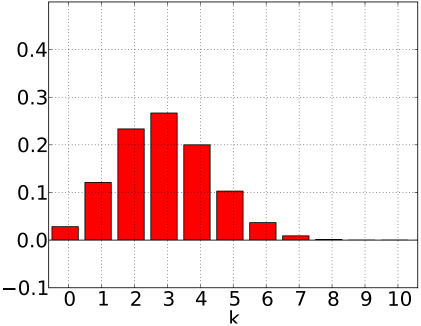
k	0	1	2	3	4	5	6	7	8	9	10
$\binom{n}{k}$	1	10	45	120	210	252	210	120	45	10	1
P(X=k)	0	0.01	0.04	0.12	0.21	0.25	0.21	0.12	0.04	0.01	0



Die Binomial-Verteilung: Beispiel ($n = 10, p = 0.3$)

$$P(X = k) = \binom{10}{k} \cdot 0.3^k \cdot 0.7^{n-k}$$

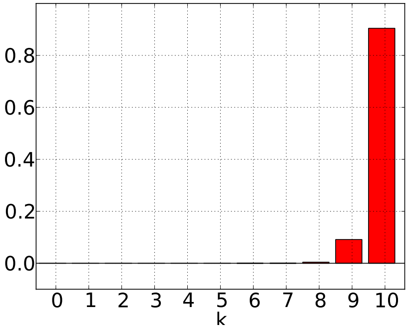
k	0	1	2	3	4	5	6	7	8	9	10
$\binom{n}{k}$	1	10	45	120	210	252	210	120	45	10	1
P(X=k)	0.03	0.12	0.23	0.27	0.20	0.10	0.04	0.01	0	0	0



Die Binomial-Verteilung: Beispiel ($n = 10, p = 0.99$)

$$P(X = k) = \binom{10}{k} \cdot 0.99^k \cdot 0.01^{n-k}$$

k	0	1	2	3	4	5	6	7	8	9	10
$\binom{n}{k}$	1	10	45	120	210	252	210	120	45	10	1
P(X=k)	0	0	0	0	0	0	0	0	0.004	0.091	0.904



Anwendungsbeispiel: Service Level Agreement

Die tägliche Ausfallwahrscheinlichkeit eines Servers betrage $p = 1\%$. Berechnen Sie die Wahrscheinlichkeit, dasss der Server in einem Monat an mehr als 2 Tagen ausfällt.

Wir definieren X als die Anzahl der Ausfälle in einem Monat:

$$
\begin{aligned}
P(X > 2) &= 1 - P(X \leq 2) \\
&= 1 - \big(P(X = 0) + P(X = 1) + P(X = 2)\big) \\
&= 1 - \underbrace{\binom{30}{0} \cdot 0.01^0 \cdot 0.99^{30}}_{=P(X=0)} - \underbrace{\binom{30}{1} \cdot 0.01^1 \cdot 0.99^{29}}_{P(X=1)} \\
&\qquad - \underbrace{\binom{30}{2} \cdot 0.01^2 \cdot 0.99^{28}}_{P(X=2)} \\
&\approx 0.33\%
\end{aligned}
$$

Die Binomial-Verteilung: Eigenschaften

Parameter

- n: Anzahl der Wiederholungen
- p: Erfolgswahrscheinlichkeit eines Versuchs.

Notation und Eigenschaften

- Ist eine Zufallsvariable X binomialverteilt, schreiben wir auch "$X \sim B(n, p)$".
- Nützliche Rekursionsformel (für $p < 1$):

$$P(X = k+1) = P(X=k) \cdot \left(\frac{p}{1-p}\right) \cdot \left(\frac{n-k}{k+1}\right) \quad \text{für } k = 0, \ldots, n-1$$

Die Binomial-Verteilung: Eigenschaften

Erwartungswert

$$E(X) = E(X_1 + X_2 + \ldots + X_n)$$

$$= \underbrace{E(X_1)} + E(X_2) + \ldots + E(X_n)$$

$$= p + p + \ldots + p$$

$$= n \cdot p$$

$(10 \quad 0{,}5 \quad = 5)$

Rechenregeln (Kapitel 04)

Def. Bernoulli-Variable

Varianz

$$Var(X) = Var(X_1 + X_2 + \ldots + X_n)$$

$$= \underbrace{Var(X_1)} + Var(X_2) + \ldots + Var(X_n)$$

$$= p \cdot (1-p) + p \cdot (1-p) + \ldots + p \cdot (1-p)$$

$$= n \cdot p \cdot (1-p)$$

Rechenregeln (weil unabhängig!)

Def. Bernoulli-Variable

Binomialverteilung: Beispiel mit Karten Bild: [5]

- Wir ziehen 5 Karten aus 32 Karten mit Zurücklegen.
  Was ist die Wahrscheinlichkeit, 3 Kreuz-Karten zu ziehen?
- X := Anzahl der (gezogenen) Kreuzkarten.
- $X \sim B(n{=}5, p{=}\frac{1}{4})$.

$$\begin{aligned} P(X{=}3) &= \binom{5}{3} \cdot p^3 \cdot (1-p)^2 \\ &= \frac{5 \cdot 4 \cdot 3}{1 \cdot 2 \cdot 3} \cdot \left(\frac{1}{4}\right)^3 \cdot \left(\frac{3}{4}\right)^2 \\ &\approx 8.7\% \end{aligned}$$

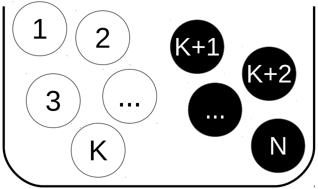
Gedankenspiel

- Können wir die Binomialverteilung verwenden, wenn wir ohne Zurücklegen ziehen?
- Nein! Die Wahrscheinlichkeit eine Kreuz-Karte zu ziehen ändert sich mit jedem Versuch – Wir können keine konstante Erfolgswahrscheinlichkeit p annehmen!
- Hier verwenden wir die hypergeometrische Verteilung *(nächstes Video)*.

Outline

Hypergeometrische Verteilung: Herleitung (Urnenmodell)

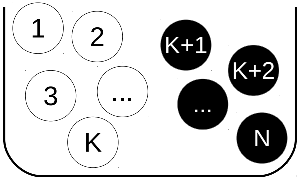
Wir leiten die gesuchte Verteilung mit Hilfe des Urnenmodells her:



- Analogie: Lotto-Spielen
  (wieviele der "Erfolgs"-Kugeln werden gezogen?)
- N Kugeln insgesamt,
  hiervon sind K hiervon weiß ("Erfolg"),
  die anderen $N - K$ schwarz ("Misserfolg")
- Wir ziehen n Kugeln ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge.
- Gesucht: $P(X = k)$ *("von den n gezogenen Kugeln sind k Stück weiß")*.

$$P(X = k) = \frac{\text{Anzahl der Kombinationen mit } k \text{ weißen und } n-k \text{ schwarzen Kugeln}}{\text{Anzahl aller möglichen Kombinationen}}$$

Die hypergeometrische Verteilung: Herleitung



- N = Anzahl aller Kugeln
- K = Anzahl aller weißen Kugeln
- $N - K$ = Anzahl aller schwarzen Kugeln

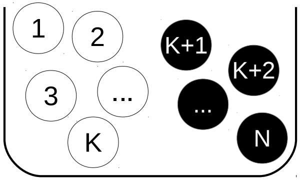
- n = Anzahl der gezogenen Kugeln
- k = Anzahl der gezogenen weißen Kugeln
- $n - k$ = Anzahl der gezogenen schwarzen Kugeln

Nenner

Wir betrachten zuerst den Nenner (alle möglichen Kombinationen): Standardformel für Ziehen ohne Zurücklegen, ungeordnet:

$$C(N; n) = \binom{N}{n}$$

Die hypergeometrische Verteilung: Herleitung



- N = Anzahl aller Kugeln
- K = Anzahl aller weißen Kugeln
- $N - K$ = Anzahl aller schwarzen Kugeln

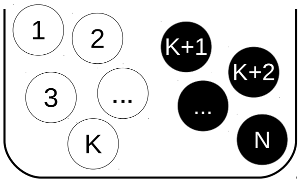
- n = Anzahl der gezogenen Kugeln
- k = Anzahl der gezogenen weißen Kugeln
- $n - k$ = Anzahl der gezogenen schwarzen Kugeln

Zähler

- Wieviele Kombinationen mit k (aus K) weißen Kugeln und $n - k$ (aus $N - K$) schwarzen Kugeln gibt es?
- Für die weißen Kugeln gibt es $C(K;k)$ Möglichkeiten:

$$C(K;k) = \binom{K}{k}$$

Die hypergeometrische Verteilung: Herleitung



- N = Anzahl aller Kugeln
- K = Anzahl aller weißen Kugeln
- $N - K$ = Anzahl aller schwarzen Kugeln

- n = Anzahl der gezogenen Kugeln
- k = Anzahl der gezogenen weißen Kugeln
- $n - k$ = Anzahl der gezogenen schwarzen Kugeln

Zähler

- Für die schwarzen Kugeln gibt es $C(N-K; n-k)$ Möglichkeiten, d.h.

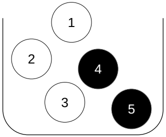
$$C(N-K; n-k) = \binom{N-K}{n-k}$$

- Insgesamt lautet die Anzahl der Möglichkeiten im Zähler:

$$C(K; k) \cdot C(N-K; n-k) = \binom{K}{k} \cdot \binom{N-K}{n-k}$$

Die hypergeometrische Verteilung: Beispiel

Wir ziehen aus der abgebildeten Urne 3 Kugeln und berechnen P(X=2):



Weiße Kombina-tionen (2 aus 3) $\binom{3}{2}$ x Schwarze Kombina-tionen (1 aus 2) $\binom{2}{1}$ = Kombinationen Mit 2 weißen *und* 1 schwarzen $\binom{3}{2} \cdot \binom{2}{1}$

1 2, 1 3, 2 3 x 4, 5 = 1 2 4, 1 2 5, 1 3 4, 1 3 5, 2 3 4, 2 3 5

P(X=2) = (1 2 4, 1 2 5, 1 3 4, 1 3 5, 2 3 4, 2 3 5) / (1 2 3, 1 2 4, 1 2 5, 1 3 4, 1 3 5, 1 4 5, 2 3 4, 2 3 5, 2 4 5, 3 4 5) = $\frac{6}{10}$

Alle Kombinationen (3 aus 5) $\binom{5}{3}$

Die hypergeometrische Verteilung: Herleitung

Definition (Hypergeometrische Verteilung)

Eine diskrete Zufallsvariable X mit der Verteilung

$$P(X = k) = \begin{cases} \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}} & \text{falls } 0 \leq k \leq n \text{ und } k \leq K \\ 0 & \text{sonst} \end{cases}$$

nennen wir hypergeometrisch verteilt.

Parameter

- $N > 0$: Gesamtzahl der Kugeln
- $K \geq 0$: Gesamtzahl der weißen Kugeln
- $n > 0$: Anzahl der gezogenen Kugeln

Hypergeometrische vs. Binomialverteilung

Wann sind sich hypergeometrische und Binomialverteilung "ähnlich"?

- Wir betrachten die Erfolgswahrscheinlichkeit p der Binomialverteilung (mit Zurücklegen) und der hypergeometrischen Verteilung (ohne Zurücklegen):

	Binomialverteilung	hypergeometrische Verteilung
1. Zug	p = K/N	p = K/N
2. Zug	p = K/N	p = K/(N-1) oder (K-1)/(N-1)
3. Zug	p = K/N	p = K/(N-2) oder (K-1)/(N-2) oder (K-2)/(N-2)
...	...	...

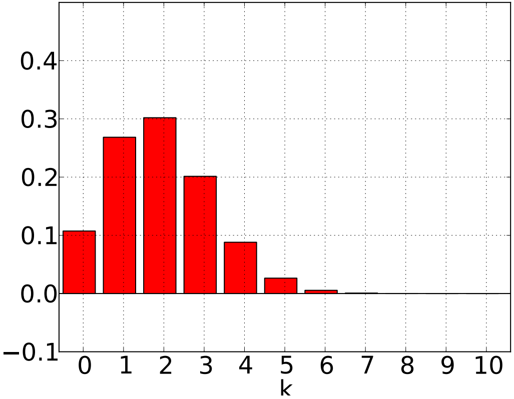
- Für $N >> n$ sind die Unterschiede gering, d.h.

$$p_1 \approx p_2 \approx \ldots \approx p_n \approx p.$$

- In diesem Fall können wir die hypergeometrische Verteilung durch die Binomialverteilung annähern!
- Faustregel: Die Approximation ist sinnvoll, wenn $n < 0.05 \cdot N$.

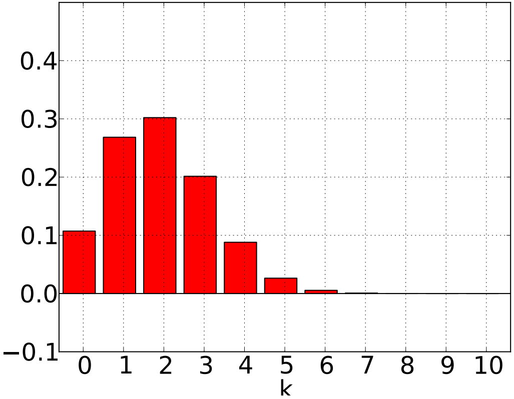
Hypergeometrische vs. Binomialverteilung

- Die Binomialverteilung mit $n = 10$, $p = 0.2$



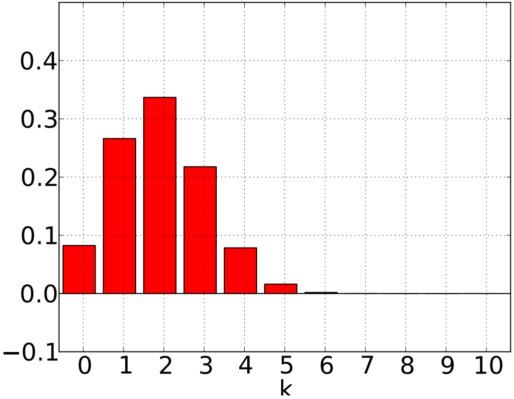
Hypergeometrische vs. Binomialverteilung

- Wir reduzieren schrittweise die Kugeln in der Urne...



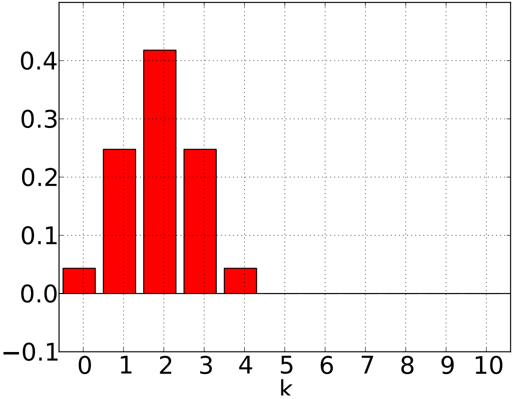
Hypergeometrische vs. Binomialverteilung

- $N = 50, K = 10$

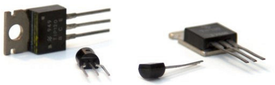


Hypergeometrische vs. Binomialverteilung

- $N = 20, K = 4$

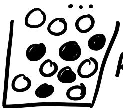


Hypergeometrisch: Do-it-yourself Bild: [3]



Eine Lieferung aus 100 Transistoren enthält 5 defekte.
In einer Kontrolle werden 4 Transistoren geprüft.

$N=100$



$K=5$

$n=4$

- Wie lauten N, n, und K ?
- Mit welcher Wahrscheinlichkeit ist genau ein geprüfter Transistor defekt?
- Wie lautet die Wahrscheinlichkeit, wenn wir sie mit der Binomialverteilung approximieren?

$$P(X=1)=\frac{\binom{5}{1}\cdot\binom{95}{3}}{\binom{100}{4}}=\frac{5\cdot\frac{95\cdot 94\cdot 93}{\cancel{1}\cdot\cancel{2}\cdot 3}}{\frac{100\cdot 99\cdot 98\cdot 97}{\cancel{1}\cdot\cancel{2}\cdot\cancel{3}\cdot 4}}\approx 17{,}65\,\%$$

Binomial: $n=4;\ p=5\,\%$

$$P(X=1)=\binom{4}{1}\cdot 0{,}05^{1}\cdot 0{,}95^{3}\approx 17{,}15$$

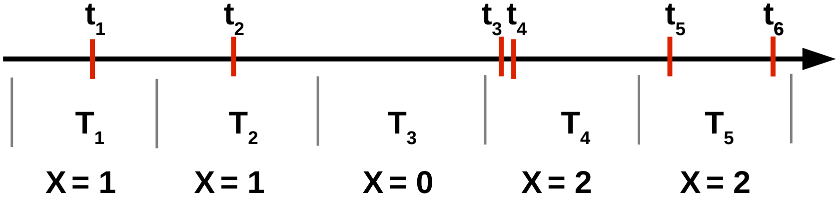
Hypergeometrisch: Do-it-yourself

Outline

Die Poisson-Verteilung: Anwendungsszenarien Bild: [1]

- Wir möchten nun Ereignisse zählen, die unabhängig voneinander zu unregelmäßigen Zeitpunkten auftreten:
  - Kunden die ein Kaufhaus betreten
  - Jobs die in eine Warteschlange eingefügt werden
  - Autounfälle
  - Server-Ausfälle
  - ...
- Wir beobachten einen festen Zeitraum und möchten wissen, wieviele Ereignisse zu erwarten sind.

Die Poisson-Verteilung: Formalisierung



Es sei X die Anzahl der Ereignisse über den gegebenen Zeitraum.

Definition (Die Poisson-Verteilung)

Gegeben $\lambda \in \mathbb{R}$, nennen wir eine diskrete Zufallsvariable X mit

$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \quad \textit{für } k = 0, 1, 2, \ldots$$

Poisson-verteilt.

Die Poisson-Verteilung: Eigenschaften

Anmerkungen

- Die Formel für die Poisson-Verteilung lässt sich mit Hilfe der Binomialverteilung herleiten *(hier nicht)*.

Parameter

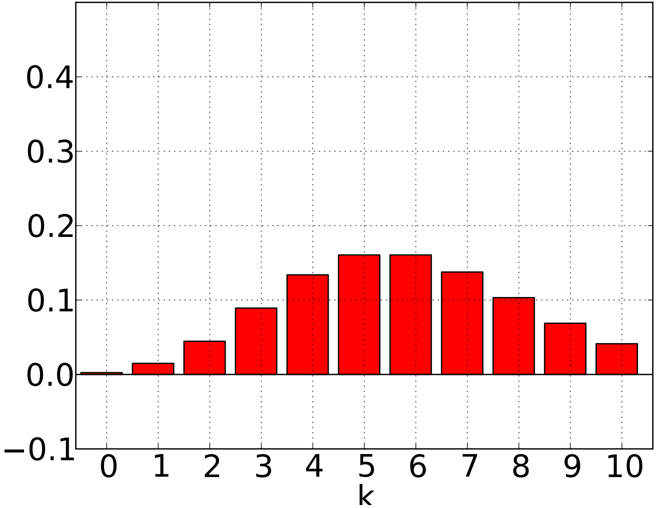
- $\lambda > 0$ (= die erwartete Anzahl an Ereignissen über den Beobachtungszeitraum)

Erwartungswert und Varianz

Ist X Poisson-verteilt (mit Parameter $\lambda > 0$), gilt:

$$E(X) = Var(X) = \lambda.$$

Die Poisson-Verteilung: Beispiel ($\lambda = 6$)



Do-Poisson-Yourself

Ein Fußballfeld hat eine Fläche von 0.75 ha. Pro Quadratkilometer schlägt im Schnitt alle zwei Jahre * ein Blitz ein. Was ist die Wahrscheinlichkeit, dass binnen eines Monats der Blitz häufiger als einmal auf dem Fußballfeld einschlägt?

$X :=$ Anzahl der Blitzeinschläge pro Fußballfeld + Monat

Gesucht: $P(X > 1)$ $\quad$ X ist Poisson-verteilt! (Zeit + Fläche)

λ? ⇝ Wieviele Blitze pro Fußballfeld + Monat im Mittel?

* $0{,}5$ Blitze pro Jahr pro Quadr. Kilometer

$0{,}5/12$ " " Monat " " "

$\underbrace{\frac{0{,}5}{12} \cdot \frac{0{,}75}{100}}$ " " " " Fußballfeld

→ $\lambda = 0{,}0003125$

Do-Poisson-Yourself

$$P(X>1) = 1 - P(X \leq 1)$$

$$= 1 - \big(\underbrace{P(X=0)} + P(X=1)\big)$$

$$= 1 - \left(\cancel{\frac{\lambda^0}{0!}} \cdot e^{-\lambda} + \frac{\lambda^1}{\cancel{1!}} \cdot e^{-\lambda}\right)$$

$$= 4{,}88 \cdot 10^{-8}$$

Outline

Die Uniformverteilung

Wir kommen nun zu den stetigen Verteilungen. Die einfachste ist die Uniformverteilung: Alle Werte innerhalb eines Intervalls $[a,b]$ sind gleich wahrscheinlich.

Definition (Uniformverteilung)

Es seien $a, b \in \mathbb{R}$ mit $a < b$. Dann nennen wir eine stetige Zufallsvariable X mit der Dichtefunktion

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

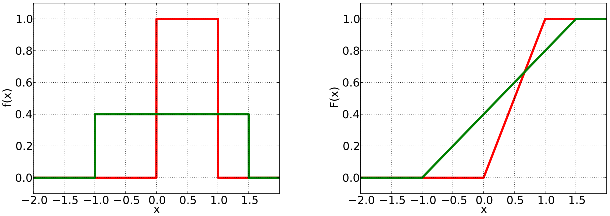
und der Verteilungsfunktion

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{falls } x < a \\ \frac{x-a}{b-a} & \text{falls } x \in [a,b] \\ 1 & \text{falls } x > b \end{cases}$$

uniformverteilt.

Die Uniformverteilung

Illustration



Parameter

- Intervallgrenzen a, b (mit $a < b$)

Anwendungsbeispiele

- Zeitpunkt von HTTP Requests an einen Server *(alle Sekunden in einer Minute sind gleich wahrscheinlich)*.
- Die Wartezeit auf den Bus.

Die Uniformverteilung: Eigenschaften

Erwartungswert und Varianz

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

$$Var(X) = \frac{1}{12}(b-a)^2$$

Outline

Die Exponentialverteilung

Definition (Die Exponentialverteilung)

Eine stetige Zufallsvariable X mit der Dichtefunktion

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

(mit $\lambda > 0$) und der zugehörigen Verteilungsfunktion

$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x} & \textit{falls } x \geq 0 \\ 0 & \textit{sonst} \end{cases}$$

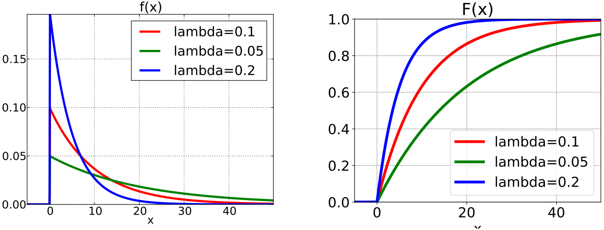
nennen wir exponentialverteilt.

Anmerkungen

- Häufigster Anwendungsfall ist die Funktionsdauer von Bauteilen/Systemen *(siehe unser Beispiel "Festplatte")*.

Die Exponentialverteilung: Eigenschaften

Dichte- und Verteilungsfunktion



Parameter

- Einziger Parameter: $\lambda > 0$

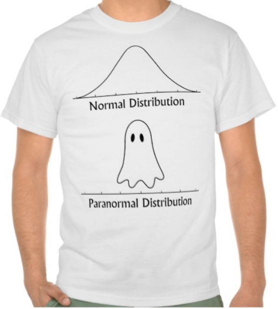
Erwartungswert und Varianz

- Erwartungswert: $E(X) = \frac{1}{\lambda}$ // siehe Kapitel 4
- Varianz: $Var(X) = \frac{1}{\lambda^2}$

Outline

1. Die Binomialverteilung
2. Die Hypergeometrische Verteilung
3. Die Poisson-Verteilung
4. Die Uniformverteilung
5. Die Exponentialverteilung
6. Die Gauß'sche Normalverteilung

Die Normalverteilung: Bedeutung



Die Normalverteilung ist die Verteilung

- Körpergröße, Gewicht (je Geschlecht)
- Reisezeit (auf fester Strecke)
- Größe von Schrauben
- Gewicht von Kartoffeln, Schokotafeln, ...
- Messfehler, Pixelwerte in Bildern
- IQ
- ...

Warum ist die Normalverteilung so weit verbreitet?

- Mathematische Begründung: Der Zentrale Grenzwertsatz.
- Umgangssprachlich: "Mitteln wir viele Zufallsvariablen, so konvergiert die Verteilung des Mittelwerts gegen eine Normalverteilung."

Die Gauß'sche Normalverteilung

Definition (Gauß'sche Normalverteilung)

Eine stetige Zufallsvariable X mit der Dichtefunktion

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

(wobei $\mu \in \mathbb{R}$ und $\sigma^2 > 0$) und der Verteilungsfunktion

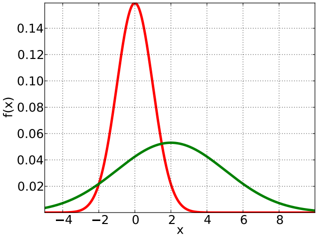
$$\mathcal{N}(x;\mu,\sigma^2) := F(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt$$

nennen wir normalverteilt.

Anmerkungen

- μ entspricht dem Erwartungswert ($E(X)=\mu$), σ^2 der Varianz ($Var(X)=\sigma^2$).
- Für $\mu = 0$ und $\sigma^2 = 1$ nennen wir X standardnormalverteilt.
  Statt $\varphi(x;0,1)$ schreiben wir einfach nur $\varphi(x)$,
  statt $\mathcal{N}(x;0,1)$ nur $\mathcal{N}(x)$.

Die Gauß'sche Normalverteilung: Eigenschaften



Die Gestalt der Dichtefunktion φ erinnert an eine Glocke. Deshalb sprechen wir auch von der *Gauß'schen Glockenkurve* oder Gauß-Glocke (engl. *"bell curve"*).

- φ besitzt das einzige lokale Maximum bei $x = \mu$.
- φ ist symmetrisch zu $x = \mu$.
- Die einzigen beiden Wendepunkte von φ liegen bei $\mu \pm \sigma$.
- Das Beispiel zeigt: $\varphi(x; 2, 3^2)$ und $\varphi(x; 0, 1^2)$
  (die Standardnormalverteilung).

Die Gauß'sche Normalverteilung: Berechnung

Wie berechnen wir die Wahrscheinlichkeit $\mathbf{P(a \leq X \leq b)}$
(z.B. dass die Körpergröße zwischen 170 und 190 cm liegt)?

Unser bekannter Ansatz

$$P(a \leq X \leq b) = \int_a^b \underbrace{\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}}}_{\varphi(t;\mu,\sigma^2)} dt$$

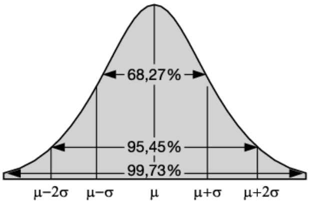
Schlüsselproblem: Nicht-Integrierbarkeit der Normalverteilung

- Wir können für die Stammfunktion von φ keine geschlossene Form bestimmen!
- **Übung**: *Versuchen Sie die Stammfunktion von e^{-t^2} zu finden.*

Die Gauß'sche Normalverteilung: Berechnung Bild: [9]

Ansatz 1: "Faustregeln"

- Es gibt Faustregeln für die Bereiche $[\mu \pm 1/2/3 \cdot \sigma]$
- *Beispiel:* $P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 68.27\%$



Ansatz 2: Standardisierung und Wertetabelle

- Allgemeine Wahrscheinlichkeiten $P(X \leq x)$ berechnen wir mittels numerischer Integration *(siehe Analysis: Rechteckregel, Trapezregel, Simpson-Regel...)*.
- Die Ergebnisse speichern wir in Tabellen.
- Hierzu standardisieren wir die Zufallsvariable zuvor *(wird im Folgenden erklärt)*.

u	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5639	0,5675	0,5714	0,5754
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7258	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7518	0,7549
0,7	0,7580	0,7612	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7996	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8398
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0.9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936

Standardisierung normalverteilter Variablen

Satz (Standardisierung normalverteilter Variablen)

Es sei $X \sim \mathcal{N}(x; \mu, \sigma^2)$ eine normalverteilte Zufallsvariable mit Erwartungswert μ und Varianz σ^2. Dann ist

$$Y := \frac{X - \mu}{\sigma}$$

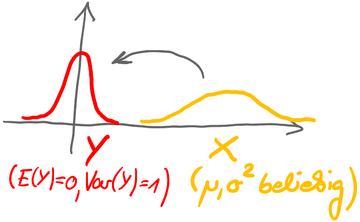
eine standardnormalverteilte Zufallsvariable.

Anmerkung

- Mit Hilfe dieses Satzes "verwandeln" wir ein beliebiges X in eine standardnormalverteilte Variable Y. Wir bezeichnen diesen Prozess als **Standardisierung**.
- Durch die Standardisierung brauchen wir keine separate Tabelle für jedes μ und σ: Eine einzige Tabelle für die Standardnormalverteilung reicht!

Standardisierung normalverteilter Variablen

Beweis



zu zeigen: $E(Y) = 0; \ Var(Y) = 1$

- $E(Y) = E\Big(\underbrace{\frac{X-\mu}{\sigma}}_{Y}\Big)$

  Rechenregeln (Kapitel 04)

  $= \frac{1}{\sigma} \cdot E(X-\mu) = \frac{1}{\sigma} \cdot \Big(\cancel{\underbrace{E(X)}_{\mu}} - \cancel{\mu}\Big) = 0$

- $Var(Y) = Var\Big(\frac{X-\mu}{\sigma}\Big) = \frac{1}{\sigma^2} \cdot Var(X - \cancel{\mu}) = \frac{1}{\sigma^2} \cdot \underbrace{Var(X)}_{\sigma^2} = 1 \quad \checkmark$

  Rechenregeln

Standardisierung normalverteilter Variablen

Wie wir $P(a \leq X \leq b)$ berechnen

1. Standardisiere X, d.h. definiere $Y \sim \mathcal{N}(0,1)$ als

$$Y := (X - \mu)/\sigma$$

2. Berechne die Grenzen a', b':

$$a' := (a - \mu)/\sigma$$
$$b' := (b - \mu)/\sigma$$

3. Es gilt:

$$\begin{aligned} P(a \leq X \leq b) &= P\Big(\underbrace{\frac{a-\mu}{\sigma}}_{a'} \leq \underbrace{\frac{X-\mu}{\sigma}}_{Y} \leq \underbrace{\frac{b-\mu}{\sigma}}_{b'}\Big) \\ &= P(a' \leq Y \leq b') \\ &= \mathcal{N}(b') - \mathcal{N}(a') \end{aligned}$$

4. Lese $\mathcal{N}(b')$ und $\mathcal{N}(a')$ aus Tabelle der Standardnormalverteilung ab.

Beispiel: IQ

- Der IQ lässt sich als normalverteilte Zufallsvariable X mit $\mu = 100$ und $\sigma = 15$ beschreiben.
- Bob hat einen IQ von 120, Charly einen IQ von 100.
- Wieviele Prozent der Bevölkerung besitzen einen IQ zwischen Charlys und Bobs?
- Wieviele Prozent besitzen einen höheren IQ als Bob?

$$P(100 \le X \le 120) = P\left(\underbrace{\frac{100-\mu}{\sigma}} \le \underbrace{\frac{X-\mu}{\sigma}} \le \underbrace{\frac{120-\mu}{\sigma}}\right)$$

$$= P\left(0 \le Y \le \frac{4}{3}\right)$$

$$= N(4/3) - N(0) = 0{,}9082 - 0{,}5 = 0{,}4082$$

$\mu = 100$
$\sigma = 15$

Beispiel: IQ

Bild: [7]

// 2. Nachkomma-stelle *

u	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5639	0,5675	0,5714	0,5754
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7258	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7518	0,7549
0,7	0,7580	0,7612	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7996	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8398
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545

Beispiel: IQ

Tabelle!

$$P(X > 120) = 1 - P(X \leq 120)$$

$$= 1 - N\left(\frac{120-100}{15}\right)$$

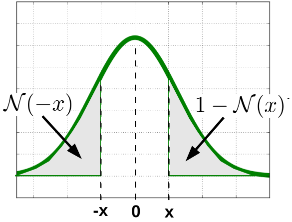
$$= 1 - N\left(\tfrac{4}{3}\right)$$

$$= 1 - 0{,}9082 = 0{,}0918 \quad \checkmark$$

Standardisierung: Anmerkungen

Symmetrie

- Warum fehlen in der Tabelle negative Werte ?
- Aufgrund der Symmetrie von φ gilt: $\mathcal{N}(-x) = 1 - \mathcal{N}(x)$.
- Wir können negative Werte also auch aus der Tabelle gewinnen.



Quantile

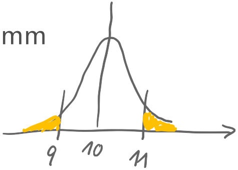
- Oft ist eine Wahrscheinlichkeit α gegeben, und wir suchen das α-Quantil, d.h. jenes x_α für das gilt $P(X \leq x_\alpha) = \alpha$
- Diese Werte sind ebenfalls in vorgefertigten Tabellen verfügbar.

α	x_α
0.001	-3.090
0.005	-2.576
0.01	-2.326
0.025	-1.960
0.05	-1.645
0.1	-1.282
0.9	1.282
0.95	1.645
0.975	1.960
0.99	2.326
0.995	2.576
0.999	3.090

Standardnormalverteilung: Do-it-yourself

Bei der Massenproduktion von Schrauben ist der Durchmesser des Gewindes (in mm) eine Zufallsvariable $X \sim \mathcal{N}(x; \mu{=}10, \sigma^2{=}4)$.

1. Bei wievielen Schrauben weicht der Durchmesser um mehr als 1mm vom Sollwert (= 10 mm) ab?
2. Wie müsste die Standardabweichung lauten, damit dies nur für 2% der Schrauben gilt?



1.) $1 - P(9 \le X \le 11) = 1 - \left[\mathcal{N}\left(\frac{11-10}{2}\right) - \mathcal{N}\left(\frac{9-10}{2}\right) \right]$

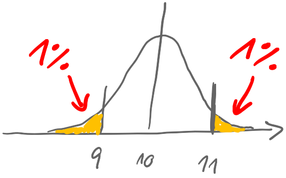
$$= 1 - \left[\mathcal{N}\left(\tfrac{1}{2}\right) - \mathcal{N}\left(-\tfrac{1}{2}\right) \right]$$

$$= 1 - \left[\quad '' \quad - \left(1 - \mathcal{N}\left(\tfrac{1}{2}\right)\right) \right]$$

$$= 2 - 2 \cdot \underbrace{\mathcal{N}\left(\tfrac{1}{2}\right)}_{0{,}6915} = 0{,}617$$

Standardnormalverteilung: Do-it-yourself

Gesucht: σ ?



Wir fordern: $\underline{P(X \le 11) = 99\%}$

$P(X \le 11) = 99\%$

$P\Big(\underbrace{\frac{X-10}{\sigma}}_{Y} \le \underbrace{\frac{11-10}{\sigma}}_{y_{99\%}}\Big) = 99\%$

99%-Quantil der Std. norm. Vtlg.

$\frac{11-10}{\sigma} = y_{99\%}$

$\frac{1}{\sigma} = 2{,}326$

$\sigma = 1/2{,}326 \approx 0{,}43$

Standardnormalverteilung: Do-it-yourself

Praxisbeispiel: “Six Sigma”

- Produktionsprozesse werden immer weiter perfektioniert – hierbei spielen statistische Verfahren eine große Rolle.
- Ziel ist in der Regel die Produktion “möglichst gleicher” Teile, d.h. die Standardabweichung der Artikel einer Produktionslinie soll möglichst gering sein.
- Unter dem Namen “Six Sigma” wurde eine Methode des Qualitätsmanagements bekannt, um die Standardabweichung σ soweit zu senken, dass nur noch Werte außerhalb eines Toleranzbereiches von 6 Standardabweichungen (6σ, daher der Name) als Ausschuss betrachtet werden müssen.
- Da in der Praxis die Lage des Mittelwerts driften kann, wird das Qualitätslimit in der Praxis von 6σ auf 4.5σ gesenkt.
- Wir suchen also: $P(|X-\mu| > 4.5\sigma)$

$$P(|X-\mu| > 4.5\sigma) = 1 - \Big(\mathcal{N}(4.5) - \mathcal{N}(-4.5)\Big) = 0.0000068$$

- Das bedeutet: Ziel sind höchstens 6.8 Defekte unter 1 Million produzierter Artikel.

References I

[1] Dylan Meconis: Parent Line.
https://flic.kr/p/5R2khv (changed to black and white, CC license, retrieved: Dec 2016).

[2] Fairphone: Fairphone quality control check.
https://flic.kr/p/iQLoHE (retrieved: Dec 2016).

[3] Martin Bircher: Transistors.
https://flic.kr/p/94gcGL (retrieved: Dec 2016).

[4] Nationwide Poll Results for 2008 Presidential Election.
https://commons.wikimedia.org/wiki/File:Nationwide_Poll_Results_for_2008_Presidential_Election.svg (retrieved: Oct 2016).

[5] Olli Henze: 4 Asse.
http://www.ohenze.de/ (retrieved: Nov 2016).

[6] Sean Fallon: Statistical Distribution Plushies: Hug The Math.
http://nerdapproved.com/approved-products/statistical-distribution-plushies-hug-the-math/ (retrieved: Dec 2016).

[7] L. Papula.
Mathematik für Ingenieure und Naturwissenschaftler, volume 3.
Friedrich Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig/Wiesbaden, 4 edition, 2001.

[8] unknown.
Jakob Bernoulli.
https://commons.wikimedia.org/wiki/File:Jakob_Bernoulli.jpeg (retrieved: Dec 2016).

[9] Eggert Winter, editor.
Gabler Wirtschaftslexikon.
Gabler, 18. edition, 2013.