

Stochastik

Markus Reichl, 21. Januar 2018

Inhaltsverzeichnis

| | | |
|----------|--|----------|
| 1 | Statistik | 2 |
| 1.1 | Begriffe | 2 |
| 1.1.1 | Boxplot | 2 |
| 2 | Wahrscheinlichkeit | 3 |
| 2.1 | Laplace Experiment | 3 |
| 2.2 | Zusammengesetzte Ereignisse | 3 |
| 2.3 | Bernoulli Versuch (Pfade) | 4 |
| 2.3.1 | Baumdiagramm | 4 |
| 2.4 | Abzähltechniken | 5 |
| 2.4.1 | Permutation | 5 |
| 2.4.2 | Kombination | 5 |
| 2.4.3 | Variation (Geordnete Auswahl) | 5 |
| 3 | Verteilung | 6 |
| 3.1 | Funktionen | 6 |
| 3.2 | Diskrete Verteilung | 6 |
| 3.2.1 | Hypergeometrische Verteilung (HgV) | 6 |
| 3.2.2 | Binomialverteilung (BV) | 7 |
| 3.3 | Stetige Verteilung | 7 |
| 3.3.1 | Normalverteilung (NV) | 7 |
| 3.3.2 | Standardisierte Normalverteilung (SNV) | 8 |
| 4 | Beurteilende Statistik | 9 |
| 4.1 | Zufallsstreuereich (ZSB) | 9 |
| 4.2 | Vertrauensbereich (Konfidenzintervall) | 10 |

1 Statistik

1.1 Begriffe

| | | |
|--------------------|--|--|
| Mittelwert | $\bar{x} = \frac{\text{Summe}}{\text{Anzahl}}$ | Durchschnittlicher Wert einer Liste. → maxima: <code>mean(<liste>)</code> |
| Standardabweichung | $\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$ | Mittlere Abweichung vom Mittelwert. → maxima: <code>std(<liste>)</code> |
| Varianz | σ^2 | Mittlere quadratische Abweichung vom Mittelwert. → maxima: <code>var(<liste>)</code> |
| Median | \hat{x} | Mittlerer Wert einer geordneten Liste. → maxima: <code>median(<liste>)</code> |
| Minimum | x_0 | Erster Wert einer geordneten Liste. → maxima: <code>min(<liste>)</code> |
| Maximum | x_n | Letzter Wert einer geordneten Liste. → maxima: <code>max(<liste>)</code> |
| Spannweite | $r = x_0 - x_n$ | Differenz zwischen Minimum und Maximum. → maxima: <code>range(<liste>)</code> |
| 1. Quartil | q_1 | Median der Werte vor dem Median einer geordneten Liste. → maxima: <code>quantile(<liste>, 0.25)</code> |
| 2. Quartil | $q_2 = \hat{x}$ | Median einer geordneten Liste. → maxima: <code>quantile(<liste>, 0.5)</code> |
| 3. Quartil | q_3 | Median der Werte nach dem Median einer geordneten Liste. → maxima: <code>quantile(<liste>, 0.75)</code> |

1.1.1 Boxplot

Aus dem Boxplot lassen sich die Quartile, der Median, sowie Minimum und Maximum auslesen. Dabei wird eine Skala vom Minimum zum Maximum gezogen und eine Box vom 1. bis zum 3. Quartil gespannt, die vom Median geteilt wird. Zusätzlich können auch Ausreißer vor oder nach dem Boxplot gezeichnet werden.

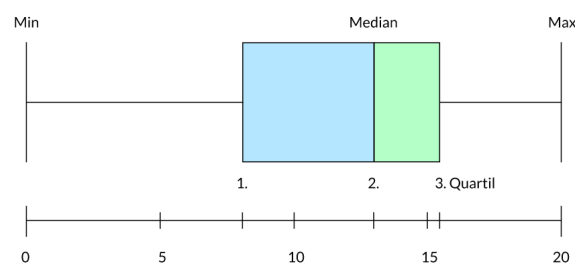


Abbildung 1: Boxplot

→ maxima: `boxplot(<<liste>>, box_orientation=horizontal)`

2 Wahrscheinlichkeit

Der Zufall unterliegt Regeln, die erst bei einer großen Anzahl von Versuchen sichtbar werden. Ein Zufallsexperiment muss n Mal durchgeführt werden, um zu erkennen, dass das Ereignis A k Mal auftritt.

Nach einer ausreichenden Zahl von Versuchen kann man als Schätzwert für die Wahrscheinlichkeit $P(A)$, die Anzahl der Ereignisse k relativ zur Anzahl der Versuche n nehmen.

$$P(A) = \frac{k}{n}$$

$$0 \leq P(A) \leq 1$$

| | | |
|--------|-----|-----------------------|
| $P(A)$ | ... | Wahrscheinlichkeit |
| A | ... | Ereignis |
| k | ... | Anzahl der Ereignisse |
| n | ... | Anzahl der Versuche |

2.1 Laplace Experiment

Gilt für ein Zufallsexperiment, dass

- endlich viele Ereignisse existieren und
- jedes Ereignis gleich wahrscheinlich ist,

kann man ein Laplace Experiment durchführen. Man unterscheidet dabei die Zahl der günstigen g und der möglichen m Fälle für ein Ereignis A . Die Wahrscheinlichkeit $P(A)$ für ein Ereignis A ist als Quotient der günstigen g und der möglichen m Fälle definiert.

$$P(A) = \frac{g}{m}$$

2.2 Zusammengesetzte Ereignisse

2 Ereignisse A, B sollen zu einem Ereignis C zusammengesetzt werden.

| | | |
|-------------------------|------|----------------------------|
| $A \text{ UND } B$ | bzw. | $A \text{ ODER } B$ |
| (A und B zugleich) | | (mindestens A oder B) |

- **Gegenwahrscheinlichkeit**

$$P(\bar{A}) = 1 - P(A) \quad \bar{A} \quad \dots \quad \text{„Non-A“}$$

- **Unvereinbarkeit**

$$P(A_{\text{UND}} B) = 0$$

- **Additionssatz (ODER-Regel)**

$$P(A_{\text{ODER}} B) = P(A) + P(B) - P(A_{\text{UND}} B) \quad \text{wenn A und B beliebig}$$

$$P(A_{\text{ODER}} B) = P(A) + P(B) \quad \text{wenn A und B unvereinbar}$$

- **Multiplikationssatz (UND-Regel)**

$$P(A_{\text{UND}} B) = P(A) * P(B_{\text{ODER}} A) \quad \text{wenn A und B beliebig}$$

$$P(A_{\text{UND}} B) = P(A) * P(B) \quad \text{wenn A und B unvereinbar}$$

2.3 Bernoulli Versuch (Pfade)

1. Pfadregel (UND-Regel) Die Wahrscheinlichkeit eines Ereignisses ist gleich dem Produkt der Wahrscheinlichkeiten entlang seines Pfades.

2. Pfadregel (ODER-Regel) Die Wahrscheinlichkeit eines Ereignisses ist gleich der Summe der Wahrscheinlichkeiten seiner Pfade.

2.3.1 Baumdiagramm

Der Bernoulli Versuch kann über sogenannte Baumdiagramme dargestellt werden. In diesem Beispiel werden zwei Münzen geworfen und es soll die Wahrscheinlichkeit berechnet werden, zu der mindestens einmal Zahl fällt.

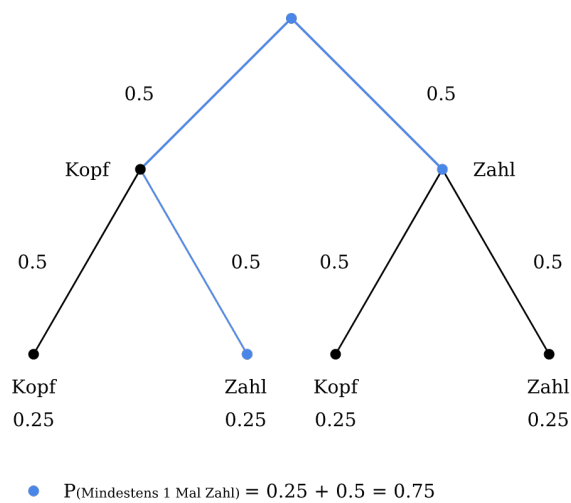


Abbildung 2: Baumdiagramm

Die Wahrscheinlichkeit bei einem Wurf Zahl zu werfen, liegt bei 50%. Am linken Pfad wird erst ein Kopf und dann eine Zahl, am rechten Pfad gleich eine Zahl geworfen. Zur Berechnung der Wahrscheinlichkeit werden nun die Produkte der Pfade aufsummiert.

$$P_{(\text{Mindestens 1 Mal Zahl})} = \text{Kopf} * \text{Zahl} + \text{Zahl} = 0.5 * 0.5 + 0.5 = 0.75$$

→ Zu 75% wird mindestens einmal Zahl geworfen.

2.4 Abzähltechniken

2.4.1 Permutation

Es seien 4 Personen gegeben und es wird die Anzahl verschiedener Reihenfolgen gesucht. Jede Person kann nur an einer Position stehen, es handelt sich also um ein *ziehen ohne zurücklegen*.

Nach jeder Auswahl einer Position stehen eine Person und ein Platz weniger zur Verfügung.

| | | | | |
|----|----|----|----|---------------|
| 1. | 2. | 3. | 4. | Position |
| 4 | 3 | 2 | 1 | Möglichkeiten |

Es bestehen also $4 * 3 * 2 * 1$ Möglichkeiten (Permutationen) oder 4 Fakultät ($4!$).

Die Anzahl der Permutationen von n Elementen entspricht also $n! = \sum_{i=0}^{n-1} n - i$.

In Maxima kann die Fakultät mittels `factorial(n)` berechnet werden.

2.4.2 Kombination

Aus $n = 7$ Personen sollen $k = 3$ Personen ausgewählt werden. Die Reihenfolge ist dabei belanglos, es handelt sich also um ein *ziehen mit zurücklegen*.

$7 * 6 * 5$ sind wählbar
 $3 * 2 * 1$ sind gleich für jede Auswahl

Es bestehen also $\frac{7*6*5}{3*2*1} = \frac{210}{6} = 35$ Möglichkeiten. Allgemein ist die Kombination als $\frac{n*(n-1)*...*(n-k+1)}{1*...*k}$ definiert, was dem Binomialkoeffizienten entspricht.

Binomialkoeffizient

$$\binom{n}{k} = \frac{n!}{(n-k)! * k!}$$

In Maxima kann der Binomialkoeffizient mittels `binomial(n, k)` berechnet werden.

2.4.3 Variation (Geordnete Auswahl)

Aus $n = 7$ Personen sollen $k = 3$ ausgewählt werden. Die Reihenfolge der Kombination ist dabei zu beachten, es handelt sich also um ein *ziehen ohne zurücklegen*.

1. 2. 3. Position
 7 6 5 $\rightarrow 7 * 6 * 5$ Möglichkeiten

Die Variation kann auch anhand der Kombination berechnet werden.

$$\binom{7}{3} * 3! = 7 * 6 * 5 = 210 \text{ Variationen}$$

3 Verteilung

3.1 Funktionen

Verteilungsfunktion $G(x)$ Die Zufallsvariable nimmt einen Wert kleiner gleich x an.

$$P(\leq x) = G(x)$$

$$P(\geq x) = 1 - G(x - 1)$$

Wahrscheinlichkeitsfunktion $g(x)$ Die Zufallsvariable nimmt den Wert x an.

$$P(x) = g(x) = G(x) - G(x - 1)$$

3.2 Diskrete Verteilung

Eine Wahrscheinlichkeitsverteilung ist diskret, wenn die Menge entweder

- endlich
- unendlich aber abzählbar oder
- beliebig definiert ist und eine abzählbare Zahl von positiven Elementen aufweist ($P(M) = 1$).

Die Menge besteht dabei meist aus natürlichen Zahlen \mathbb{N} .

3.2.1 Hypergeometrische Verteilung (HgV)

Eine Grundgesamtheit N enthält K Merkmalsträger^I, die zu einer Wahrscheinlichkeit P , k Mal in einer Stichprobe n zu finden sind. Durch das Ziehen der Stichprobe wird die Grundgesamtheit verändert, es handelt sich also um ein **Ziehen ohne Zurücklegen**.

$$P(k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

$$x \leq K \quad n - k \leq N - K^{\text{II}}$$

| | | |
|--------|-----|---|
| $P(k)$ | ... | Wahrscheinlichkeit, dass k Ereignisse auftreten |
| N | ... | Grundmenge |
| K | ... | Merkmalsträger |
| k | ... | Anzahl der Ereignisse |
| n | ... | Umfang der Stichprobe |

Für die HgV kann ein Erwartungswert μ gefunden werden.

$$\mu = n * \frac{K}{N}$$

^I Im Qualitätsmanagement werden Merkmalsträger auch als Defekte D und die Grundmenge als Prüflös bezeichnet. Im deutschen Sprachraum ist für die Merkmalsträger die Variable M gängig.

^{II} Gilt $k \leq K \quad n - k \leq N - K$ nicht, ist $P(k) = 0$

3.2.2 Binomialverteilung (BV)

In einer Stichprobe n sind zu einer Wahrscheinlichkeit P , k Merkmalsträger zu finden. Diese treten immer zur Wahrscheinlichkeit p^1 auf, es handelt sich also um ein **Ziehen mit Zurücklegen**.

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

| | | |
|--------|-----|---|
| $P(k)$ | ... | Wahrscheinlichkeit, dass k Ereignisse auftreten |
| k | ... | Anzahl der Ereignisse |
| n | ... | Umfang der Stichprobe |
| p | ... | Wahrscheinlichkeit eines Ereignisses |

Für die BV kann ein Erwartungswert μ gefunden werden.

$$\mu = n * p$$

3.3 Stetige Verteilung

Im Gegensatz zur diskreten Verteilung kann die Menge beliebige Werte, innerhalb eines definierten Intervalls annehmen. Die Verteilungsfunktion von P ist dabei stetig, mit $P \in \mathbb{R}$.

3.3.1 Normalverteilung (NV)

Eine stetige Zufallsvariable x mit der Wahrscheinlichkeitsdichte $g : \mathbb{R} \rightarrow \mathbb{R}_0$, definiert als

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

| | | |
|----------|-----|---------------------------|
| $g(x)$ | ... | Wahrscheinlichkeitsdichte |
| x | ... | Zufallsvariable |
| μ | ... | Erwartungswert |
| σ | ... | Standardabweichung |

heißt normalverteilt. Die resultierende Kurve wird als „Gaußsche Glockenkurve“ bezeichnet und nimmt im Integral $\int_{-\infty}^{+\infty}$ immer den Wert 1 an.

¹ p wird meist einzeln angegeben, kann jedoch auch als Quotient der Merkmalsträger K und der Grundmenge N berechnet werden.

$$p = \frac{K}{N}$$

Eigenschaften

- Symmetrisch zu μ
- Keine Nullstellen
- 2 Wendepunkte $\mu \pm \sigma$
- Eine Veränderung von μ verschiebt die Kurve an der x-Achse
- Eine Veränderung von σ verbreitert die Kurve
- Die Fläche unter der Kurve ist immer 1
- Die Wahrscheinlichkeit, dass ein Merkmal $\leq x_0$ ist, entspricht der Fläche unter der Kurve zwischen $-\infty$ bis x_0

$$\int_{-\infty}^{x_0} g(x; \mu, \sigma) dx$$

- Die Wahrscheinlichkeit, dass ein Merkmal zwischen a und b ist, entspricht der Fläche unter der Kurve zwischen a bis b

$$\int_a^b g(x; \mu, \sigma) dx$$

- Die Dichtefunktion ist nicht analytisch, sondern nur numerisch zu berechnen

3.3.2 Standardisierte Normalverteilung (SNV)

Eine NV ist standardisiert wenn der Erwartungswert $\mu = 0$ und die Standardabweichung $\sigma = 1$ ist. Durch eine z-Transformation kann aus jeder beliebigen NV eine SNV gebildet werden.

$$g(x; 0, 1) = \varphi(z)$$

$$z = \frac{x - \mu}{\sigma}$$

4 Beurteilende Statistik

Mithilfe der beurteilenden Statistik wird aus empirisch gewonnenen Daten auf unbekannte Größen, wie Wahrscheinlichkeiten und Erwartungswerte geschlossen.

Speziell geht es dabei um den Zusammenhang zwischen Grundgesamtheit und Stichprobe.

4.1 Zufallsstrebereich (ZSB)

Der ZSB gibt an, zu welcher Wahrscheinlichkeit eine Stichprobe in einem bestimmten Bereich liegt. Diese Wahrscheinlichkeit ist abhängig

- von den Eigenschaften der Grundgesamtheit
 - z.B.: μ, σ
- vom Zufall

ZSB von μ, σ

- Zweiseitig begrenzter ZSB

$$x_{o,u} = \mu \pm z_{1-\frac{\alpha}{2}} * \sigma$$

- Einseitig begrenzter ZSB

$$x_o = \mu + z_{1-\alpha} * \sigma$$

$$x_u = \mu - z_{1-\alpha} * \sigma$$

ZSB für Stichprobenmittelwerte

- Zweiseitig begrenzter ZSB

$$x_{o,u} = \mu \pm z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$$

- Einseitig begrenzter ZSB

$$\bar{x}_o = \mu + z_{1-\alpha} * \frac{\sigma}{\sqrt{n}}$$

$$\bar{x}_u = \mu - z_{1-\alpha} * \frac{\sigma}{\sqrt{n}}$$

| | | |
|-----------|-----|---|
| x | ... | Zufallsstrebereich (ZSB) |
| \bar{x} | ... | Mittelwert |
| o | ... | nach oben |
| u | ... | nach unten |
| z | ... | Zufallsvariable der Standardnormalverteilung |
| α | ... | Irrtumswahrscheinlichkeit (Signifikanzniveau) |
| μ | ... | Erwartungswert aus Grundgesamtheit |
| σ | ... | Standardabweichung aus Grundgesamtheit |
| n | ... | Umfang der Stichprobe |

4.2 Vertrauensbereich (Konfidenzintervall)

Der Vertrauensbereich gibt die Wahrscheinlichkeit an, zu der ein Parameter der Grundgesamtheit in einem bestimmten Bereich liegt.

Vertrauensbereich für μ

- σ bekannt

$$\mu_{o,u} = \bar{x} \pm z_{1-\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$$

- σ unbekannt

$$\mu_{o,u} = \bar{x} \pm t_{f;1-\frac{\alpha}{2}} * \frac{s}{\sqrt{n}}$$

Vertrauensbereich für σ

$$\sigma_o = s * \sqrt{\frac{f}{x_{f;\frac{\alpha}{2}}^2}} \quad ; \quad \sigma_u = s * \sqrt{\frac{f}{x_{f;1-\frac{\alpha}{2}}^2}}$$

| | | |
|-----------|-----|---|
| \bar{x} | ... | Mittelwert |
| o | ... | nach oben |
| u | ... | nach unten |
| z | ... | Zufallsvariable der Standardnormalverteilung |
| t | ... | Zufallsvariable der Student-Verteilung |
| f | ... | Freiheitsgrad $f = n - 1$ |
| α | ... | Irrtumswahrscheinlichkeit (Signifikanzniveau) |
| μ | ... | Erwartungswert aus Grundgesamtheit |
| σ | ... | Standardabweichung aus Grundgesamtheit |
| n | ... | Umfang der Stichprobe |