Brilliant & Simple | DSci

Field of Data Science

Paul Julitz

Notizen

03.01.2020

Inhaltsverzeichnis

Teil I

Mathematics

Kapitel 1

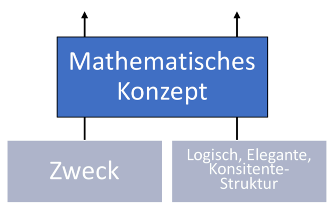
Lose Gedanken

1 LEK-Struktur

Ein mathematisches Konzept kann sich durch zwei *Treiber* bestimmen lassen. Den **Zweck** und durch eine **Logisches, elegantes und konsitenze Strutkur**.

Die *Logisches, elegantes und konsistente Struktur* gibt den klaren Rahmen vor, wie ein mathematisches Konzept aufgebaut und begründet wird. Diese muss jedoch nicht den Zweck direkt erkennen lassen. Es ist auch möglich, dass mit der *Logisches, elegantes und konsistente Struktur* nur gezorgt wird, dass dieses in den Korpus der Mathematik eingefügt werden kann.

Der Zweck ist kann einer eine Fragestellung heraus gefunden werden oder selbst wieder auf ein anders mathematisches Konzept verweisen (*B-to-B Beziehung*)



Beispiel *Wahrscheinlichkeitsraum*. Dieser dient dem Zweck, später mit Zufallsvariablen arbeiten zu können und das Chaos realer Phänomen zu bändigen. Dies bildet den Zweck ab. Der Aufbau ist jedoch nach dem Treiber: *Logisches, elegantes und konsistente Struktur* konzipiert.

Eine Herausforderung ist, dass der Zweck eines mathematischen Konzepts sich manchmal erst am Ende ergibt, weshalb der Aufbau sich manchmal erst ganz am Ende erschließt.

Kapitel 2

Stochastik

)

1 Überblick - Zum Beginn stand der Zufall

1.1 Einführung

1.1.1 Theorie und Reale Werte

Das Gebiet der Stochastik beschreibt zwei Gebiete

- Wahrscheinlichkeitstheorie: Diese betrachtet die theoretische Modellierung von Zufallsereignissen.
- Statistik: Dieses Gebiet versucht anhand von realen Daten Parametern und Verteilungen theoretisch modellierte Zufallsereignisse zu schätzen.

In der Wahrscheinlichkeitstheorie wird angenommen, dass alle Parameter eines Zufallsexperimentes vorhanden sind. Gefragt wird, welche Ausgänge mit welcher Wahrscheinlichkeit beobachtet werden können.

7 Eine Münze, die mit der Wahrscheinlichkeit $p = 0,50$ *Kopf* zeigt, wird $n = 100$ geworfen. Mit welcher Wahrscheinlichkeit wird $k = 60$ mal *Kopf* getroffen?

In der Statistik wird angenommen, dass das Zufallsexperiment schon durchgeführt wurde, jedoch dies nicht vollkommen beschrieben ist, und sein Ausgang (am Beispiel k) schon bekannt ist. Gefragt wird, nach den Parametern die dieses Zufallsexperiments beschreiben, um zukünftige Ausgänge mit ihren jeweiligen Wahrscheinlichkeiten zu beschreiben.

Eine Münze wurde $n = 100$ geworfen und hat dabei $k = 60$ *Kopf* angezeigt. Bestimmen (*schätzen*) die Wahrscheinlichkeit p, mit der die Münze bei einem Wurf *Kopf* zeigt.

Die Statistik versucht nicht Zufälligkeit aus der Welt zu eliminieren, sondern nur die Unsicherheit zu reduzieren, in welcher Welt man sich befindet.

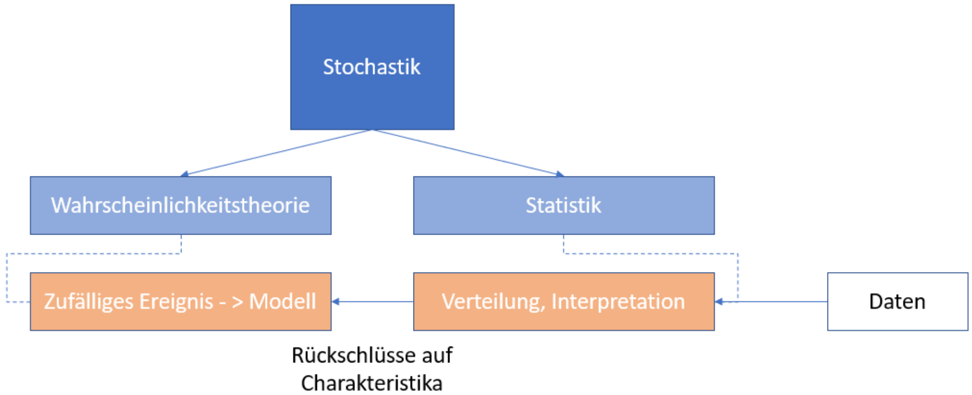


Abbildung 2.1: Feld der Stochastik

Um wichtige mathematische Objekte wieder aufzufrischen, wird im Folgenden kurz das Konzept einer Abbildung erklärt.

- In der Mathematik versteht man unter einer Abbildung oder unter gewissen Bedingungen Funktion, eine Beziehung zwischen zwei Mengen, wobei Elementen der einen Menge (Definitionsmenge, *unabhängige Variablen, x-Werte, Funktionsargument*) Elementen der anderen Menge (Ergebnismenge, *abhängige Variable, y-Werte, Funktionswert*) zugeordnet werden. Man spricht von einer Funktion, wenn die Abbildung in einen reelle- oder komplexwertigen Körper abbildet.*Notation.Abbildung* o. D.

Definition I.1: Abbildung

Eine Abbildung ordnet jedem Element der Definitionsmenge D genau ein Element der Zielmenge Z zu: $f : D \rightarrow Z, x \mapsto y := f(x)$

- Die Begrifflichkeiten von Unabhängiger und Zielvarialben hat Ihren Ursprung in der Physik und Physik nahen Bereichen.

1.1.2 Gebiete der Statistik

Das Gebiet von Stochastik gliedert sich grob in

Exploritory Data Analysis (EDA) Die Explorative Datenanalyse überlappt sich von der Zielsetzung mit der Deskriptiven Statistik. Es werden *realen* Daten dargestellt. Ebenso werden dardurch Informationen gewonnen und Hypothesen gewonnen, welche auf das Statistisches Model (SM) schließen lassen, welche die Daten generiert. Mit dem Wachstum neuer technischer Möglichkeiten liegt der Fokus darauf, computergestützte Algorithmen zu nutzen, um Erkenntnisse über die Daten zu gewinnen. Teilaufgaben

- wie Ausreißer finden,
- Objekte mit ähnlichen Eigenschaften finden **Clusteranalyse**,
- Kategorie schaffen **Klassifikation**,
- stochastisch-funktionale Zusammenhängen zwischen verschiedenen Teilen der Daten zu finden **Regression**,

etc. gehen einen Schritt weiter, als die klassische *Deskriptiven Statistik*.

Wahrscheinlichkeitsrechnung Dieses Gebiet der Stochastik baut das theoretische Fundament auf, um später mit den realen Daten arbeiten zu können. Hierbei werden Werkzeug geschaffen, welche mit anderem Zweck und Bedeutung in der *Deskriptiven* und *Induktiven Statistik* Die *Schließende Statistik* bewegt sich von den Daten und den ersten Annahmen aus, hin zu einem sicheren Verständnis, ob die getroffene Modellierung der Welt richtig ist. Dabei wird getestet, ob Hypothesen über die Annahmen der Grundgesamtheit verworfen werden müssen, oder sie den Testkriterien genügen.

Deskripten Statistik Wie in EDA beschrieben, ist der Zweck die Darstellung der Daten und Zusammenhänge dieser. Auch hier überlappen sich die Werkzeuge aus der *Schätztheorie*, *Stichprobentheorie*. Dies werden jedoch mit dem Zweck eingesetzt, erst "Vermutung" sichtbar zu machen.

Induktiven Statistik Die *Schließende Statistik* beschäftig sich mit dem Prozess: Wenn man die Daten, Einblicke und Hypothesen über den Sachverhalt formuliert hat, wie kann man sauber eine Rückschluss auf die Grundgesamtheit (Das Theoretische Fundament) bekommen. Dabei werden unterschieden

- Stichprobentheorie,
- Schätztheorie,
- Testtheorie

Die Auseinanderdividierung der Gebiet und deren Werkzeuge lässt sich somit nicht so sauber trennen, wie man es vielleicht anfänglich vermutet. Die Gebiete greifen mehr oder minder Hand in Hand, um aus

Richtig Daten zu erheben $\rightarrow$ Vermutungen in *Statistische Modelle* umzusetzten $\rightarrow$ Erste Abschätzung über die Reale "modellierte" Welt zu erhalten.

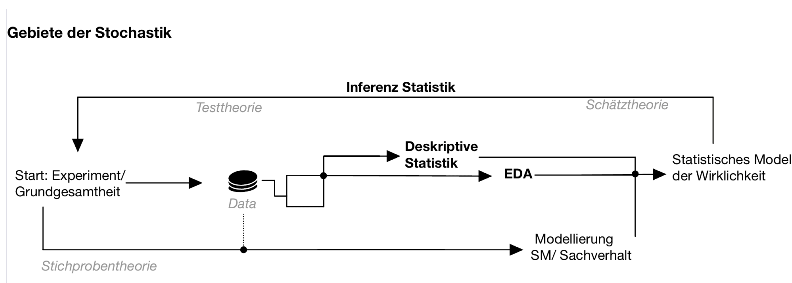


Abbildung 2.2: Prozess der Abbildung der Realität in eine Statisches Modell

Dabei wird mit der Stichproben aus der Grundgesamtheit der Einblick gewonnen, welcher nötig ist, um eine Rückschluss auf diese zu gewinnen.

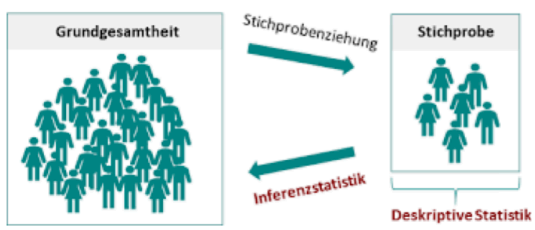


Abbildung 2.3: Rückschluss auf die Grundgesamtheit

1.1.3 Parameter der Grundgesamtheit

Im späteren Verlauf wird sich zeigen, dass einige Werkzeuge zwischen den Teilgebieten der Deskriptiv und Induktiven Statistik in Modifikation wieder vorkommen.
Das Gebiet von SM beschäftigt sich ausschließlich damit, *Parameter der Verteilung* einer oder mehrer Zufallsvariablen zu schätzen. Das Gebiet der Regression beschäftigt sich ebenfalls mit den *Parametern der Grundgesamtheit* hierbei wird jedoch nicht die Parameter der Verteilung herangezogen, sondern die Parameter, welche das Modell von Beziehung zwischen den Zufallsvariablen beschreiben. [1]

[1] Wichtig: Persönlich habe ich eine sauber Trennung der Methoden und neu Definition nicht direkt gefunden.

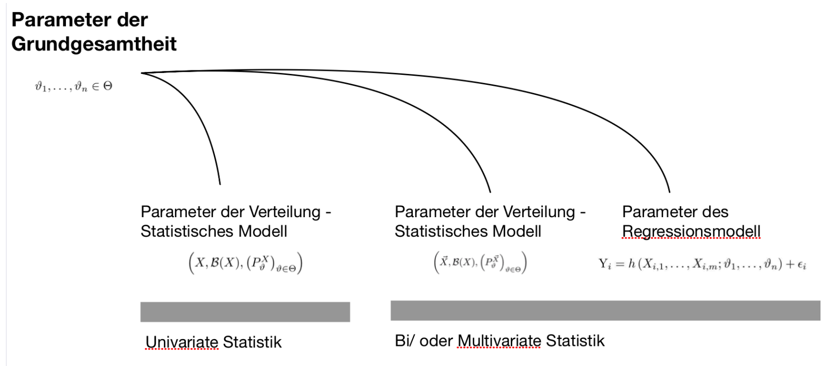


Abbildung 2.4: Parameter der Grundgesamtheit

1.2 Wahrscheinlichkeitsraum

Zu beginn steht der *Zufall*.

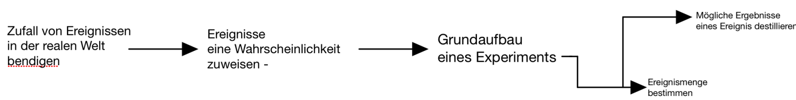


Abbildung 2.5: Gedanklicherprozess, um Zufall in geordnete Bahnen zu lenken.

Mit Hilfe des Konstrukt eines Wahrscheinlichkeit(s)(W.) Raums. Werden Ereignisse mit Wahrscheinlichkeiten versehen. Damit es mathematisch Kohärent ist, werden den einzelnen Elemente des Raums Eigenschaften abverlangt.

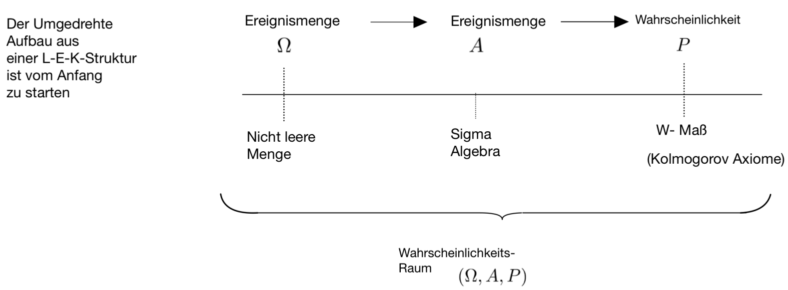


Abbildung 2.6: Der Wahrscheinlichkeitsraum

Auch hier gilt, das Problem mit dem gestartet wurde, wird ans Ende gestellt. Das L.E.K-Struktur lässt auch hier nicht direkt erkennen, was das Ziel von all dem Ausbau ist.

1.3 Vererbung von Wahrscheinlichkeiten mit einer ZV

1.3.1 Mehrer Ausprägungen

Angenommen es soll ein Würfelwurf modelliert werden. Dann wird benötigt

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- $\mathcal{A} = \mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{1, 2\}, \ldots\}$
- $P(\{\omega_i\}) = \frac{1}{\#\Omega} = \frac{1}{6}$
  Die anderen Elemente von $\mathcal{A}$ werden im Regelfall nicht aufgelistet, weil sie entweder implizit verstanden werden oder aus anderen Rechenregeln für Wahrscheinlichkeiten gelten. Als Beispiel für $P(\{\emptyset\})$ wird

Null implizit verstanden. Für Ergebnisse, welche auch mehreren Ergebnissen ω bestehen, werden meist Rechenregeln angefügt oder implizit angenommen. Für diesen Fall wird jedem Element nichtleeren Element $\omega \in \Omega$ in $A \in \mathcal{A}$ die Wahrscheinlichkeit $\frac{1}{6}$ zugewiesen und diese addiert. Sodass gilt $P = \frac{3}{6}$, für $P(\{1, 2, 3\})$.

Die Zufallsvariable bildet die Ereignisse aus Ω auf $\mathbb{R}$ ab.

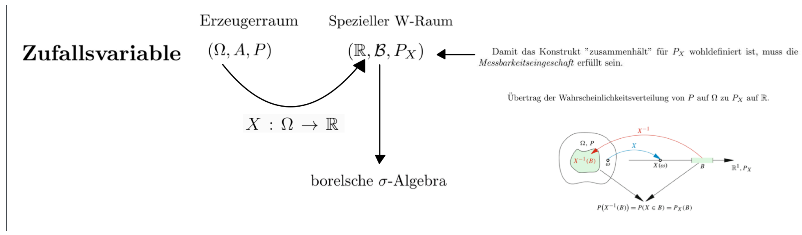


Abbildung 2.7: Neu Schaffung eines speziellen $(\Omega, \mathcal{A}, P)$

Einfacher Wurf eines Würfels Für diese Beispiel soll die Zufallsvariable eine obsolete Frage beantworten:

$$X : \Omega \to \mathbb{R}, \omega \mapsto \omega$$
$$- \text{Augenzahl nach einem Wurf}$$

Die Wahrscheinlichkeitfunktion P_X definiert sich unter I.10 - Verteilung einer Zufallsvariable/ Wahrscheinlichkeitsfunktion (S.: 17) , und findet für $B = \{1\} \in \mathcal{B}$

$$P_X(B) = P(X^{-1}(B)) = P(\{1\}) = \frac{1}{6}$$

Zweifacher Wurf eines Würfels Für zwei Würfel, welche geworfen werden gilt

- $\Omega = \{1, 2, 3, 4, 5, 6\}^2 = \{(1, 1), (1, 2), \ldots\}$
- $\mathcal{A} = \mathcal{P}(\Omega) = \{\emptyset, \{(1, 1)\}, \{(1, 1), (1, 2)\}, \ldots\}$
- $P(\{\omega_i, \omega_j\}) = \frac{1}{\#\Omega} = \frac{1}{36}$ für $\omega_i, \omega_j \in \{1, 2, 3, 4, 5, 6\}$ Auch hier gilt, dass für die restlichen Elemente aus $\mathcal{A}$ Regeln oder Herleitungen impliziert angenommen oder durch die $\mathcal{A}$ definiert werden. Für das Beispiel $P(\{(1, 1), (2, 1)\})$ gilt $P(\{(1, 2)\}) + P(\{(2, 1)\}$. Dies leitet sich aus der $\mathcal{A} -$ *additivit* her.

Auf diese Ω angewandt, werden hier drei verschiedenen Zufallsvariablen gebildet.

$$X_1 : \Omega \to \mathbb{R}, (\omega_i, \omega_j) \mapsto min(\omega_i, \omega_j)$$
$$- \text{Geringste Augenzahl zwei Würfel}$$
$$X_2 : \Omega \to \mathbb{R}, (\omega_i, \omega_j) \mapsto \omega_i + \omega_j$$
$$- \text{Summe der Augenzahl zweier Würfel}$$
$$X_3 : \Omega \to \mathbb{R}, (\omega_i, \omega_j) \mapsto \omega_j$$
$$- \text{Zweite Augenzahl zweier Würfel}$$

Durch die Messbarkeitseigenschaft I.10 - Verteilung einer Zufallsvariable/ Wahrscheinlichkeitsfunktion (S.: 17) wird auch hier die Wahrscheinlichkeit von P übertragen. Für X_2 lässt sich P_X grafisch aufzeigen.

1.3.2 Einfache Ausprägung

Münzwurf Für den $(\Omega, \mathcal{A}, P)$eines Münzwurfes, wird dieser wie folgt aufgespannt

- $\Omega = \{\textit{Kopf}, \textit{Zahl}\}$
- $\mathcal{A} = \mathcal{P}(\Omega) = \{\emptyset, \{\textit{Kopf}\}, \ldots, \{\textit{Kopf, Zahl}\}\}$
- $P(\{\omega_i\}) = \frac{1}{2}$

n-facher Münzwurf Für einen mehrfachen Münzwurf wird der $(\Omega, \mathcal{A}, P)$wie folgt aufgespannt

- Für $i = 1, 2, \dots, n$ wird $\Omega_i = \{1, 0\}$ gesetzte.
- Für die $\mathcal{A}$ gilt $\mathcal{A} = \mathcal{P}(\Omega)$ mit $\Omega = \Omega_1 \times \cdots \times \Omega_n = \{\omega = (\omega_1, \dots, \omega_n), \omega_i \in \{0, 1\}\}$
- $P(\{\omega\}) = \prod P_i(\{\omega_i\})$

Hierbei wird die Situation so abgebildet, das die Versucht

Binomial Verteilung aus Wahrscheinlichkeitsmaß Die gängigen Wahrscheinlichkeitsfunktionen für Zufallsvariablen wie Bernoulli, Poison oder ähnliche sind Funktionen, leitet sich aus dem Grundraum $(\Omega, \mathcal{A}, P)$ und der zu modellierenden Situation her. Die Wahrscheinlichkeitfunktion muss dabei ebenfalls die Messbarkeitseigenschaft einhalten.
Am Beispiel eines $n-$fachen Münzwurfs folgt. Sei $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ ein beliebiges Elementarereignis gilt

$$P(\{\omega\}) = \prod_{i:w_i=1} a_i \prod_{j:w_j=0} (1 - a_j)$$

mit $P_i(\{1\}) = a$ und $P_i(\{0\}) = 1 - a$, wobei mit einem *identischen Versuchsaufbau* gilt $p_1 = \dots p_n = p \in [0, 1]$. Für eine bestimmte Anzahl $k = |\{i : w_i = 1\}|$ von Ausprägungen von mit 1 eines Elements, gilt für das Element $\omega \in \Omega$:

$$P(\{\omega\}) = p^k (1 - p)^{n-k}, \quad \forall k = 0, 1, \dots, n$$

Um zu überprüfen, ob die Annahmen über die zu modellierenden Situation akkurat und zutreffen sind, kommt das Feld der Inferenzstatistik zum tragen.

1.4 Dualität (Drei Faltigkeit) Input Zufallsvariable/Stichprobenwerte

Im Umgang mit Schätzern, Statistiken und der Konstruktion von Zufallsvariablen ist die Dualität zwischen der theoretischen und realen Konstruktion dieser Abbildungen meist implizit angenommen.
Als Beispiel, es werden selten beide Varianten einer Abbildung für eine Schätzfunktion notiert. Zum einen werden als Inputvariablen $X_1, \dots, X_n$ verwendet, zum anderen $x_1, \dots, x_n$ als Realisation der Stichprobe.

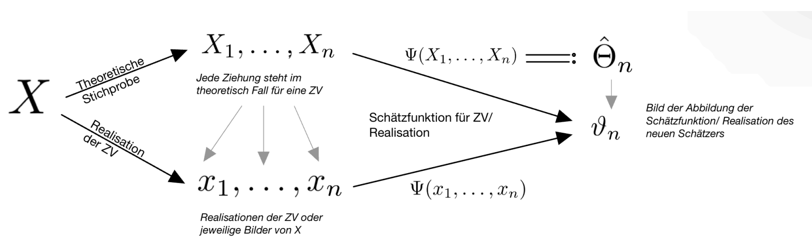


Abbildung 2.8: Dualität - Theoretischer und Realisationsverlauf

Es besteht ein weiter Zusammenhang zwischen der

- tatsächlichen Abbildung, der Schätzfunktion,
- der neuen Konstruktion einer Zufallsvariable (ZV) und
- der tatsächlichen Realisation, als Schätzwert für den Parameter oder die Parameterfunktion

Merkmal X	Zufallsvariable X
Ω ist die Grundgesamtheit	Ω ist die Ergebnismenge und gehört zu einem W-Raum
M ist der Merkmalsraum	$\mathbb{R}$ ist Bildraum
$X(\omega)$ ist die Ausprägung	$X(\omega)$ ist die Realisation
X besitzt eine Häufigkeitsverteilung	X besitzt eine W- Maß (W-verteilung)
X besitzt eine empirische Verteilungsfunktion	X besitzt eine theoretische Verteilungsfunktion

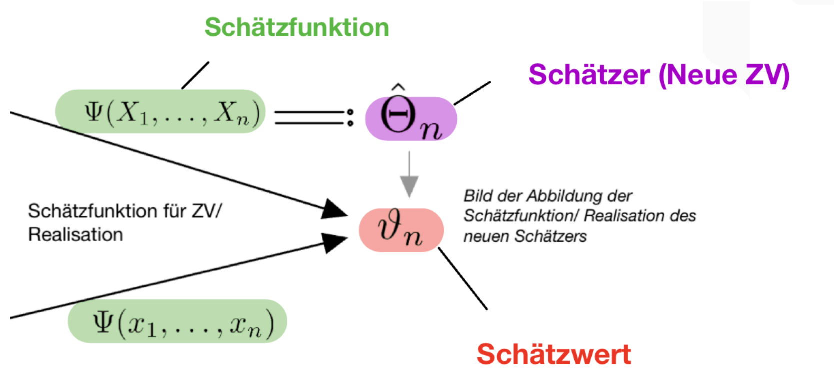


Abbildung 2.9: Drei Faltigkeit - Schätzer, Schätzfunktion und Schätzwert

1.5 Abriss Deskriptive Statistik

In der Deskriptiven Statistik wird der Gesichtpunkt geändert. Die Realisationen der Zufallsvariablen liegen vor. Es wird jetzt kein Rückschluss auf die wahren W-Raum gezogen, sondern eher eine Bewertung und Darstellung der Daten vollzogen.

Achtung: Die Begrifflichkeiten spiegeln sich und es gibt viele Gleichheiten. Es handelt sich aber hier nicht um das Feld der Induktiven Statistik.

Wie schon erwähnt, gibt es ähnliche Konzept, welche sich schwach zuordnen.

Merkmal vs Zufallsvariable Anders als im vorherigen Kapitel ist ein Merkmal zwar das Kernelement der Despriptiven Statistik, jedoch nicht so strikt definiert wie die Zufallsvariable. Der Begriff der ZV wird jedoch oft als Synonym verwandt.

Definition I.2: Merkmal

Ein *Merkmal* ist eine Abbildung von Ω in eine Menge M:

$$X : \Omega \to M$$

Wenn die Menge M ohne weitere Struktur ist, so ist X ein **nominales Merkmal**. Ist $X(\Omega)$ eine geordnete Menge, so ist X ein **ordinales Merkmal**. Ist $X(\Omega) \subset \mathbb{R}$ und übernimmt die $X(\Omega)$ die euklidische Metrik, so ist X ein **quantitatives** oder **kardinales** Merkmal.

Die Skalierung von Merkmalen erlaubt Vergleichsmöglichkeiten:

Nominal $X(\omega_1) = X(\omega_2)$ oder $X(\omega_1) \neq X(\omega_2)$

Ordinal $X(\omega_1) \preceq X(\omega_2)$ oder $X(\omega_1) \simeq X(\omega_2)$

Kardinal $X(\omega_1) - X(\omega_2)$

2 Wahrscheinlichkeitstheorie

2.1 Wahrscheinlichkeitsraum

Der Wahrscheinlichkeitsraum[2] ist das Kernstück der Wahrscheinlichkeitstheorie. Diese gibt die Werkzeuge, Zufallsexperimente aus der realen Welt zu modellieren.

Im Folgenden wird das Konzept von *Ergebnis* uns *Ereignis* konzeptioniert. Mit der gewonnen Ergebnismenge wird das W-Maß bestimmt.[3]

Definition I.3: Wahrscheinlichkeitsraum

Sein Ω eine nichtleere Menge, sei $\mathcal{A}$ eine σ-Algebra auf Ω und P das W-Maß auf $\mathcal{A}$, heißt das Tupel $(\Omega, \mathcal{A}, P)$ **Wahrscheinlichkeitsraum**.

[4] Mit den drei Komponenten Das Konzept der ZV wird helfen, einen leichteren Umgang mit dem W-Raum zu erhalten. Arens, 2018, Henze, 2018

2.1.1 Ergebnismenge

Die Menge der möglichen Ergebnisse eines stochastischen Vorgangs wird üblicherweise mit Ω abgekürzt. Am Anfang einer stochastischen Modellierung ist diese der entscheidenste Schritt, zu definieren, wie diese Menge aussieht.

Definition I.4: Ergebnismenge/raum

Die Menge Ω ist die *Ergebnismenge* oder der *Grundraum* eines Zufallsexperiments und enthält alle möglichen Ergebnisse ω.

Dies Definition greift einen naiven Mengenbegriff auf, welcher nur erklärt, ob $\omega \in \Omega$ oder $\omega \notin \Omega$ ist. Diese Eigenschaft scheint in der Theorie trival zu sein, in der Praxis ist es nicht immer klar, ob ein Ergebnis zu einer Grundgesamtheit oder nicht gehört. Ein Beispiel hierfür ist, das *Zufallsvorgang (Experiment): Schwere Unfälle in Berlin*. Ist das Zufallsexperiment ein *Münzwurf*, dann ist der Grundraum leichter zu bestimmen:

$$\begin{aligned} \Omega &= \{K, Z\} \text{ oder} \\ &= \{Kopf, Zahl\}. \end{aligned}$$

In diesem Fall ist die Mächtigkeit gleich, die genaue Ausprägung von ω kann jedoch für das gleichgemeinte Zufallsexpermient ungleich sein. Im Falle eines Kartenspiels ist gut zu zeigen, wie unterschiedlich die Grundgesamtheit sogar in ihrer Machtigkeit ausfallen kann:

$$\begin{aligned} \Omega &= \{Rot, Schwarz\} \text{ oder} \\ &= \{Rot, Schwarz\} \text{ oder} \\ &= \{RotBild, RotZahl, SchwarzBild, SchwarzZahl\} \text{ oder} \\ &= \{R2, R3, \dots, SAss\} \text{ oder} \\ &\text{etc.} \end{aligned}$$

[2] engl. Probability

[3] Wenn im Folgenden W- als Prefix verwendet wird, ist der Begriff Wahrscheinlichkeit gemeint.

[4] Begrifflichkeit aus der Maß Theorie:

Definition I.1: Maßraum

Das Triplet $(\Omega, \mathcal{A}, \mathbb{P}))$ heißt *Maßraum*,

- wenn Ω eine beliebige, nicht leere Menge ist. Diese wird *Grundraum* genannt.
- $\mathcal{A}$ eine $\sigma-$Algebra über den Grundraum Ω ist.
- $\mathbb{P}$ ein *Maß*, dass auf $\mathcal{A}$ definiert ist.

[5]

Werden n hintereinander durchgeführte Einzelexperimente durchgeführt, können dies als $n-$Tupel eines Gesamtexperiments gesehen werden. Angenommen es liegen zwei Zufallsexperimente vor: 1. Experiment ist ein Münzwurf, das 2. Experiment ist ein Würfelwurf, welche hintereinander durchgeführt werden. Ein adäquater Grundraum für das Gesamtexperiment ist

$$\Omega_1 \times \Omega_2 = \{(a_1, a_2) : a_1 \in \{K, Z\}, a_2 \in \{1, 2, \ldots, 6\}\}$$

2.1.2 Ereignismenge

Bei stochastischen Vorgängen interessiert oft nur, ob ein Ergebnis ω zu einer *gewissen Menge von Ergebnissen* gehört. Es wird also eine Menge gesucht, welche alle Fragestellungen an das Zufallsexperiment beantworten kann. Jede Teilmenge von Ω heißt jetzt ein **Ereignis**. Diese Teilmengen werden mit Großbuchstaben abgekürzt.

Am Beispiel es *Würfel* wird jedes Ergebnis als Ereignis dargestellt: $\{1\}, \{2\}, \ldots, \{6\}$. Die Grundmenge Ω ist ein **sicheres Ereignis**. Das Ereignis A, dass eine gerade Zahl fällt, ist $\{2, 4, 6\}$. Das keine Zahl getroffen wird, wird **Unmögliches Ereignis** und entspricht $\emptyset$.

Um hinreichend komplexe Modelle zu entwickeln, müssen Anforderungen aus der Mengenlehre erfüllt werden. In der Sprache der Mengenlehre, muss die Gesamtheit der Ereignisse eine $\mathcal{A}$ bilden, sie I.2 - Sigma-Algebra (S.: 13) .

Definition I.5: Ereignismenge

Sei Ω eine nichtleere Ergebnismenge, dann heißt das Mengensystem $\mathcal{A}$ von Teilmengen auf Ω eine **Ergebnismenge**, wenn $\mathcal{A}$ eine $\sigma-$Algebra ist (siehe I.2 - Sigma-Algebra (S.: 13)).

[6]

Ist die Ω eine endliche oder abzählbar unendliche Menge, so die größte $\sigma-$Algebra die aufgespannte Potenzmenge. Die kleinste ist $\{\emptyset, \Omega\}$.

Beim *Wurf* einer Münze ist $\Omega = \{K, Z\}$. Die erzeugte von Ω erzeugenden $\sigma-$Algebra die Potenzmenge $\{\emptyset, \{K\}, \{Z\}, \{K, Z\}\}$.

2.1.3 Wahrscheinlichkeitsmaß

Der Begriff der Wahrscheinlichkeit ist inhaltlich leer. Dieser beruht auf das axiomatische Fundamente von *Kolmogorov*. Dies lässt ein großen Spielraum für die Interpretationen und Berechnungen offen. Solang die Axiomen eingehalten werden, kann eine Wahrscheinlichkeit als relative Häufigkeit, Grenzwert von Häufigkeiten als Expertenmeinung oder etc. verstanden werden.

---

[5] Rechtschreibung: Am Ende eines Ganzsatzes setzt man nach Abkürzungen nur einen Punkt. Ausnahmen sind Fragesätze oder Anforderungen.(§ 103 des amtlichen Regelwerks zur deutschen Rechtschreibung)

[6] Begrifflichkeit aus der Maß Theorie:

Definition I.2: Sigma-Algebra

Sei Ω eine nichtleere Obermenge und sei $\mathcal{A}$ eine Menge von Teilmengen der Obermenge, dann heißt $\mathcal{A}$ eine $\sigma-$Algebra, wenn gilt:

- $\mathcal{A}$ enthält die Grundmenge: $\Omega \in \mathcal{A}$.
- $\mathcal{A}$ ist stabil bezüglich der Komplimentbildung: Ist $A \in \mathcal{A}$, dann ist auch $A^C \in \mathcal{A}$. Dabei ist A^C das Komplement von A bezüglich Ω.
- $\mathcal{A}$ ist stabil bezüglich abzählbarer Vereinigungen: Sind $A_i \in \mathcal{A}, i \in \mathbb{N}$, so ist auch $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Definition I.6: Drei Axiome von Kolmogorov/ Wahrscheinlichkeitsmaß

Sei Ω eine Obermenge, sei $\mathcal{A}$ eine $\sigma-$Algebra von Ω, heißt die Abbildung P von $\mathcal{A}$ nach $\mathbb{R}$ **Wahrscheinlichkeitsverteilung** oder **Wahrscheinlichkeitsmaß** von Ω (genauer: auf den Teilmengen von Ω), wenn für P gilt:

- Axiom Nichtnegativität und Normiertheit: $\forall A \in \mathcal{A} : 0 \leq P(A) \leq 1$
- Axiom Normiertheit: $P(\Omega) = 1$
- Axiom σ-Additivität: Für jede abzählbare Folge von disjunkten Mengen $A_i \in \mathcal{A}$ gilt

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Der Wert von $P(A)$ heißt Wahrscheinlichkeit.

[7] Das W-Maß kann somit als Abbildung in $[0, 1]$ verstanden werden.

Für den Wurf eines Münze würde für $\Omega = \{K, Z\}$ das W-Maß wie folgt aussehen:

$$P : \mathcal{A} \to [0, 1], A \mapsto P(A) = \begin{cases} P(\{K\}) & = 0,5 \\ P(\{Z\}) & = 0,5 \\ P(\{Z, K\}) & = 1 \\ P(\emptyset) & = 0 \end{cases}$$

Nicht immer wird eine Vorschrift für jede Menge $A \in \mathcal{A}$ definiert oder auch der das Ergebnis richtig definiert.

- Sei $\Omega = 1, 2, \ldots, 6$ die Ergebnismenge für einen Würfelwurf. Es kann beobachtet werden, dass $P(i) = \frac{1}{6} \forall i \in \Omega$ angegeben wird. Das W-Maß nimmt jedoch nur Mengen auf, weshalb es richt wäre $P(\{i\})$ zu notieren.
- Selbst wenn $P(\{i\})$ notiert wird, wird durch die $\sigma-$Additivität $P(\{2, 4, 6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6}$ hergeleitet.
- Wenn bei wiederholten Experimenten *Totale Unabhängigkeit* vorliegt, kann bei einem dreimaligen Münzwurf $P(\{KZK\}) = P(\{K\})P(\{Z\})P(\{K\})$ mit $\left(\frac{1}{2}\right)^3$.

2.2 Zufallsvariable und Verteilungen

Das Konzept der ZV dient des leichteren Umgangs mit dem W-Raum. Rechnungen und weitere aufgesetzten Konzepte werden nicht mehr in einem abstrakten Raum Ω bestimmt, sondern im $\mathbb{R}^1$ oder auch bei mehrdimensionale ZV der $\mathbb{R}^k$ mit $k \in \mathbb{N}$.

2.2.1 Herleitung

Schlussendlich erzeugt die Zufallsvariable einen neuen W-Raum, welcher auf einem Erzeugerraum $(\Omega, \mathcal{A}, P)$ definiert wird. Dabei werden unter gewissen Bedingungen Eigenschaften dieses Erzeugerraums auf die Zufallsvariable übertragen, so auch die Bildung des neuen W-Maß P_X.

[7] Alternative Beschreibung aus der Maß Theorie:

Definition I.3: Maß

Es sein $\mathcal{A}$ eine $\sigma-$Algebra einer nichtleeren Grundmenge Ω. Eine Abbildung $f : \mathcal{A} \to [0, \infty]$ heißt *Maß* auf $\mathcal{A}$, wenn die beiden Bedingungen erfüllt sind:

- $f(\emptyset) = 0$
- $\sigma-$Algebra Additivität.

Bemerkung

- ZV werden mit Großbuchstaben aus dem hinteren Bereich des Alphabetes notiert: $X, Z, W,$ etc.
- Es gibt auch anderen Begrifflichkeiten für ZV. Vom Prinzip handelt sich sich bei einer ZV nicht um eine Variable sondern um eine Abbildung. Weil ZV jedoch im Verbindung mit dem W-Raum betrachtet und weiterverarbeitet werden, kommt diese Bezeichnung zustanden. Realisationen einer ZV werden nie einzeln betrachtet, sie werden immer in der ein oder anderen Weise in Abhängigkeit ihres zugehörigen W-Maß betrachtet. Aus diesem Zusammenhang leitet sich auch der Begriff des Zufalls ab, für die Interpretation einer Realisierung kommt immer in Frage, wie wahrscheinlich das eintreten dieses Event ($\mathcal{A}$) ist. Der Zufall entscheidet über die Verteilung und gibt so eine variable Größe wieder. Daher kommt auch der andere Name *Zufallsgröße* zustande. Betrachtet man höherdimensionale ZV spricht man von *Zufallsvektoren*.

Definition In der einfachsten Form ist eine ZV eine Abbildung, welche aus Ω eine W-Raums c auf $\mathbb{R}$ abbildet. Diese Abbildung muss nicht bijektiv oder injektiv sein. Somit könnte gelten

Definition I.7: Zufallsvariable

Sei $(\Omega, \mathcal{A}, P)$ ein W-Raum. Die Abbildung

$$X : \Omega \to \mathbb{R}$$

heißt *Zufallsvariable*.

Allgemeiner Vermerk: Eine mathematische Definition kann auf verschiedenen Weise beschreiben werden. Welche Konzepte wann angebracht werden, kann aus den verschiedensten Gründen erfolgen. Schlussendlich besteht das Ziel, ein Problem lösen zu können. Bei dem Konzept der Zufallsvariable gilt das gleiche. Es können Element der Maßtheorie in die Definition gezogen werden. Ebenso können Bestandteile des W-Maß P_X vorgezogen werden, und an die Definition des Zufallsvariablen angefügt werden. Bsp.:

- **Definition I.4: Alternative Definition I der Zufallsvariable**

  Sei $(\Omega, \mathcal{A}, P)$ ein W-Raum. Die Abbildung

  $$X : \Omega \to \mathbb{R}$$

  heißt **Zufallsvariable**, wenn bei der das vollständige Urbild $X^{-1}(B)$ jeder Borel-Menge $B \in \mathcal{B}$ ein Element von $\mathcal{A}$ ist.

- **Definition I.5: Alternative Definition II der Zufallsvariable**

  Sei $(\Omega, \mathcal{A}, P)$ ein W-Raum und $(\Omega_2, \mathcal{A}_2)$ ein Messraum, so heißt die Abbildung

  $$X : \Omega \to \Omega_2$$

  Zufallsvariable. Ist $\Omega_2 = \mathbb{R}^k$, so wird X als $k-$*dimensionale reelle Zufallsvariable* und im Fall $k = 1$ als *reelle Zufallsvariable* bezeichnet.

Die ZV ist eine Vorschrift, welche jedem Ergebnis ω eines stochachstischen Vorgang eine Wert $X(\omega)$ zu weißt. Dies Werte werden als *Realisationen von* X bezeichnet. Ebenso können sie als Messung einer stochastischen Experiment interpretiert werden.

Borel Sigma Algebra Um das W-Maß P_X zu definieren, wird die σ-Algebraauf $\mathbb{R}$ benötigt. Für abzählbare Mengen Ω war die größte σ-Algebradie Potenzmenge $\mathcal{P}(\Omega)$. Für $\mathbb{R}$ ist es die Borel-σ-Algebra.[8]

[8] Referenz: Wikipedia/MessbareFunktion

Definition I.8: Borelsche Sigma-Algebra

Sei Ω eine topologischer Raum und O das System der offenen Teilmengen von Ω. Dann heißt

$$\otimes := \sigma(O) \tag{2.1}$$

die *Borelsche* σ-Algebraüber Ω. Ihre Elemente $B \in \mathcal{B}$ heißen *Borelsche-Mengen* oder *Borel-Mengen*. Im Fall $\Omega = \mathbb{R}^k$ setzen wir $\mathcal{B}^k := \mathbb{R}^{\|}$, im Fall $k = 1$ ist $\mathcal{B} := \mathbb{R}^{\infty}$.

Bemerkung

- Borel-σ-Algebrawird über einen Erzeuger E definiert und wird nicht direkt von Ω. Beispiel für $\Omega = \mathbb{R}$ sind

$$E_1 = \{(-\infty, a] : a \in \mathbb{R}\}, E_2 \qquad = \{(-\infty, a) : a \in \mathbb{R}\}, E_3 = \{(a, b) \subset \mathbb{R} : a \leq b\}, \text{etc.}$$

- Die *Borel*-σ-Algebraist die kleinste σ-Algebradie alle Teilmengen von Ω besitzt.
- Für den weiteren Verlauf wird nicht weiter darauf eingegangen, welcher Erzeuger die *Borel*-σ-Algebraerzeugt.

Verteilung einer Zufallsvariable Es besteht folgendes Problem. Wie kann sicherstellt werden, dass die Eigenschaften der Wahrscheinlichkeitsverteilung des Erzeugerraums $(\Omega, \mathcal{A}, P)$ erhalten bleiben oder sich in gewisserweise übertragen. Hierfür wird für X die *Messbarkeitseingeschaft* verlangt.

Diese besagt im Kontext der Maß Theorie

Definition I.6: Messbarkeitseigenschaft

Eine Abbildung f heißt *messbar*, wenn das Urbild jeder Menge $B \in \mathcal{B}$ unter f ein Element aus $\mathcal{A}$ist.

Formaler bedeutet diese

$$f^{-1}(B) \in \mathcal{A} \forall B \in \mathcal{B} \tag{2.2}$$

Bemerkung Messbarkeitseigenschaft

- Der Ausdruck $f^{-1}(B)$ bezeichnet dabei keine Abbildung, sondern die Urbildmenge.
- Alternative Schreibweisen sind $\{\omega \in \Omega : f(\omega) = x\} \in \mathcal{A}$ für $\forall x \in \mathbb{R}$.
- Diese Anforderung stellt für die Definition oder Konzeption einer Zufallsvariable sicher, dass das Urbild eines Ereignis in $\mathcal{B}$ auch ein Ereignis in $\mathcal{A}$unter Ω ist.

Hintergrund ist, diese Mengengebilde $\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{A}$ wird für das W-Maß P_X benötigt. Denn im Bezug auf die ZV X beleibt die Frage bestehen, wie kann die Wahrscheinlichkeitsverteilung von Ω über X auf P_X übertragen. Die Antwort: I.6 - Messbarkeitseigenschaft (S.: 16) . Damit diese Definition der *Verteilung von* X greift, muss zur Definition einer Zufallsvariable die Messbarkeitseigenschaft (messbar) ergänzt werden.

Definition I.9: Zufallsvariable und Messbarkeitseigenschaft

Sei $(\Omega, \mathcal{A}, P)$ ein W-Raum und $(\mathbb{R}, \mathcal{B})$ ein Messraum, so heißt die die $\Omega - \mathbb{R}$-messbare Abbildung

$$X : \Omega \to \mathbb{R}$$

Zufallsvariable.

Definition I.10: Verteilung einer Zufallsvariable/ Wahrscheinlichkeitsfunktion

Sei $(\Omega, \mathcal{A}, P)$und $(\mathbb{R}, \mathcal{B})$ ein Messraum und $X : \Omega \to \mathcal{B}$ eine Zufallsvariable, so heißt das Bildmaß

$$P_X(B) := P(X^{-1}(B))$$

Verteilung von X oder auch *Wahrscheinlichkeitsverteilung* im diskreten oder *Dichtefunktion* im stetigen Fall.

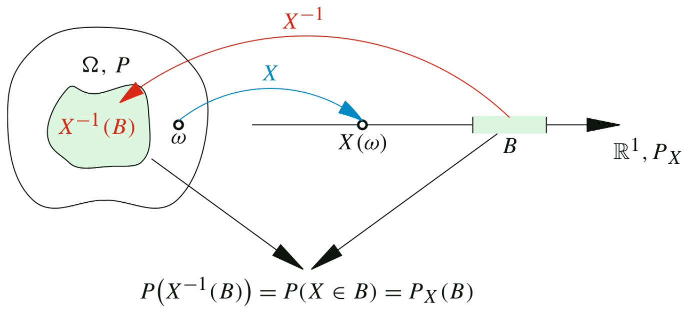


Abbildung 2.10: Die Zufallsvariable X bildet Ω in $\mathbb{R}$ ab und überträgt die Wahrscheinlichkeiten der Ereignisse

- Wegen der *Messbarkeitseigenschaft* ist das Bildmaß P_X auf Ω gilt $P_X(\{\mathbb{R}\}) = P(\Omega) = 1$
- Eigenschaften der Verteilung P_X werden als Eigenschaften von X deklariert. Z. B.: Wenn P_X diskret ist, ist auch X diskret. Wenn $P_X = \mathcal{N}\left(\mu, \sigma^2\right)$, so wird

  $$X \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

  notiert. Das Symbole $\sim$ wird nicht einheitlich ausgedrückt. Für den weiterem Umgang, wird hier angenommen, dass es für *verteilt nach* steht.
- Andere Schreibweisen für das W-Maß P_X sind

$$\begin{aligned} P(X \in B) &:= P_X(B) := P(\{\omega \in \Omega : X(\omega) \in B\}), \\ P(X = x) &:= P_X(x) := P(\{\omega \in \Omega : X(\omega) = x\}), \\ P(X \leq x) &:= P_X((-\infty, x]) := P(\{\omega \in \Omega : X(\omega) <= x\}), \text{etc.} \end{aligned}$$

[9]

Auf der Verteilung von X aufbauend wird die *Verteilungfunktion von X* definiert.

Definition I.11: Verteilungsfunktion von X

Ist X eine Zufallsvarialbe auf einen WS-Raum $(\Omega, \mathcal{F}, \mathbb{P})$, so heißt die durch

$$F_X(x) := \mathbb{P}(X \leq x), x \in \mathbb{R}$$

definierte Abbildung $F_X : \mathbb{R} \to [0, 1]$ die *Verteilungsfunktion von* X. Diese besitzt die Eigenschaft *monoton wachsend*, *rechtsseitig stetig* und F_X kommt von 0 und geht nach 1.

[9] Kommt die Frage auf, ist der Ausdruck $P(X \leq x)$ überhaupt möglich, wenn nicht sichergestellt werden kann, ob die Zufallsvariable mindestens eine ordinale Skalierung besitzt? Die Antwort: Ja. Das W-Maß gibt nur wieder, wie die Verteilung von X aussieht und die Realisationen in $\mathbb{R}$ zu verwerten sind. Dabei gibt es eine gegeben Ordnung in $\mathbb{R}$. Ob diese Sinn ergibt, hängt von der Skala ab. Dies ist Thema im Kapitel *Deskriptive Statistik*.

2.2.2 Eigenschaften und Besonderheiten

Diskret ZV Handelt es sich bei X um eine stetige oder diskrete ZV ist damit die *Wahrscheinlichkeitsfunktion* oder *Dichtefunktion* gemeint. Dies Eigenschaft wird sprachlich auf die Abbildung X gemünzt.

Definition I.12: Diskrete Wahrscheinlichkeitsverteilung einer diskreten zufälligen Variablen

Eine *diskrete* Zufallsvariable X besitzt endlich oder abzählbar unendlich viele Realisationen x_i, die mit Wahrscheinlichkeiten

$$p_i = P(X = x_i) > 0$$

angenommen werden. Für x_i gilt

$$\sum_{i=1}^{\infty} P(X = x_i) = 1.$$

Die Angaben aller p_i heißen *Wahrscheinlichkeitsverteilung von* X.

Verteilung im Erzeugerraum bekannt Im Fall, dass Ω sich nicht bestimmen lässt oder P sich nicht abbilden lässt, ist das Wissen gut, dass für den W-Raum von X es ausreicht, wenn die F_X bekannt ist. Dies erlaubt, dass direkt im aufgespannten Raum zu bleiben.

Kanonische Konstruktion Lässt sich Ω nicht definieren oder ist zu kompliziert darzustellen. Kann für Ω gleich mit $\mathbb{R}$ gesetzt werden. Für $\mathcal{A}$ wird $\mathcal{B}$ gesetzt. Für die Abbildung von Ω nach $\mathbb{R}$ wird die Indentität gewählt.

Die Art des Aufbaues des neuen W-Raums nennt sich **Kanonische Konstruktion** bei $X(x) := x$ ist. Dies ist nicht untypisch in der Statistik, dass der W-Raum $(\Omega, \mathcal{A}, P)$in den Hintergrund gerät.

Verteilungen Verteilungen wie Bernoulli, Poison, Normalverteilt, etc. bestimmen sich aus den gegeben Parametern des zu Grunde liegenden Sachverhalt der W-Raums.

Nicht messbare Abbildung Sei $X = Y = \{1, 2\}$ und $\mathcal{A} = \{\emptyset, X\}$ (Kleinste $\sigma-$Algebra) und $\mathcal{B} = \mathcal{P}(X)$.

Die Frage: Ist $\mathcal{A} - \mathcal{B} - f$ messbar, wenn $f(1) := 2$ und $f(2) := 1$ definiert ist?

Antwort: Nein, die Abbildung ist nicht messbar. Für $f^{-1}(\{2\})$ müsste das Ereignis $\{1\}$ in $\mathcal{A}$ liegen.[10]

Hinweis: Für *messbare* ZV wird im diskreten Fall die Potenzmenge für Ω herangezogen.

2.2.3 Beispiele

Beispiel zweifacher Würfelwurf Der Erzeugerraum wird für den stochastischen zweifachen Würfelwurf

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2 = \big\{\omega = (\omega_i)_{i \in \{1,2\}} : \omega_i \in \{1, 2, 3, 4, 5, 6\}\big\}$$

Ein Ereignis $A = \{2, 4, 6\}$ könnte die Menge der geraden Zahlen sein. Bei der Zufallsvariable X interessiert *Die Summe der Augenzahlen*. Die Abbildungsvorschrift sieht dabei wie folgt aus

$$X = \begin{cases} 1 & \text{, wenn } (1, 1) \\ 2 & \text{, wenn } (1, 2) \\ \vdots & \vdots \\ 36 & \text{wenn}(6, 6) \end{cases}$$

[10] Referenz: Matheraum/nicht messbare Funktion

Das W-Maß P auf Ω definiert sich gleichverteilt: $P(\{\omega\}) = \frac{1}{36}$. Für P_X leitet sich aus der I.6 - Messbarkeitseigenschaft (S.: 16) ab, die Kombinationen von ω welche die Summenzahl x_i ergeben, werden aufaddiert. Als Beispiel:

$$
\begin{aligned}
&X = 4 - \text{Summe 4 der Augenzahlen zweier Würfel}\\
&\to P_X(\{X = 4\}) = P(\{(1,3),(2,2),(3,1)\})\\
&= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36}\\
&= \frac{4}{36}
\end{aligned}
$$

Beispiel Plurale Informationen Gegeben sei eine Population von Menschen mit $\Omega = \{\text{Silvia}, \text{Manfred}, \ldots\}$ von diesen können Informationen (Plural) gezogen werden.

- Körpergewicht $\mathbb{R}_>$. Diese genau *Wahrscheinlichkeitsverteilung* P_X ergibt sich aus den konkreten Urbilder zu jeden Wert von x.
- Die Information Parteizugehörigkeit kann ebenfalls als Zufallsvariable abgebildet werden. Die Parteien werden einer belieben Zahl zugeordnet. Je nach dem wie in der Population die Personen den Parteien zugeordnet sind, bestimmen die Urbilder über der Verteilung der ZV.

2.2.4 Momente einer Zufallsvariable

Momente von ZV sind Parameter/Kenngrößen der Deskriptiven Statistik. Das Konzept des Moment wird auch genutzt in der Inferenz Statistik, um Schätzer zu konstruieren.
Die bekanntesten Momente sind Erwartungswert, Varianz, Schiefe und Wölbung. Es gilt, eine Verteilung einer ZV ist eindeutig bestimmt, durch die Angabe aller Momente, falls diese existieren und die momentenerzeugenden Funktionen konvergieren.

k-tes theoretische Moment Sei X eine Zufallsvariable und k eine natürliche Zahl. Man bezeichne das $k-te$ Moment von X als Erwartungswert der $k-ten$ Potenz von X:

$$m_k := E\left[X^k\right] \tag{2.3}$$

Das $1-$ Moment von X ist der **Erwartungswert** μ.

k-tes absolutes Moment Das absoluten Moment ist erhält die Ergänzung

$$m_k^a := E\left[|X|^k\right] \tag{2.4}$$

k-tes zentrales Momente Beim $k-ten$ zentralen Moment wird der Erwartungswert von X von X abgezogen:

$$\mu_k := E\left[(X - \mu_1)^k\right]. \tag{2.5}$$

Das zweite zentrale Moment ist die Varianz $\sigma^2 := \mu_2$. Bei dritten spricht man von der Schiefe und vom vierten von der Wölbung. In beiden Fällen wird das Moment noch mit der Standardabweichung σ normiert:

$$\mu_{3/4}^s := E\left[\left(\frac{X-\mu}{\sigma}\right)^{3/4}\right] \tag{2.6}$$

Beispiel stetiges und diskrete Zufallsvariable Am Beispiel des Erwartungswertes wird das Moment für eine stetige und diskrete ZV aufgezeigt.

Definition I.13: Erwartungswert (stetig)

Ist $X : \Omega \to \mathbb{R}$ eine quasiintegrierbare reelle Zufallsvariable, so heißt

$$E[X] := \int X dP = \int x dP_X(x)$$

der Erwartungswert von X.

Ist X diskret, so definiert sich der Erwartungswert wie folgt:

Definition I.14: Erwartungswert (Diskret)

Für eine reele Zufallsvariable X auf eine endlichen W-Raum $(\Omega, \mathcal{A}, P)$heißt

$$E[X] := E[X]_P := \sum_{\omega \in \Omega} X(\omega) P_X(\{\omega\})$$

der Erwartungswert von X.

Das $P_X(\{\omega\})$ angegeben wurde, bezieht sich darauf, dass zu jedem ω es unterschiedliche oder auch gleiche x abgebildet werden können.[11]

2.3 Statistisches Modell

Wie in der Einleitung für Stochastik erwähnt, die Kategorisierung in *Deskripter/ EDA* und Inferenzstatistik muss sich nicht immer in den Methoden unterscheiden. Oftmals werden die gleichen Sachverhalten etwas modifiziert, um die Fragestellung beantworten zu können. Zum Beispiel auch hier, beim SM. Das SM in der theoretischen Form beschreibt die Variation des $(\Omega, \mathcal{A}, P)$durch die Para- und Nichtparametisierung der Verteilungsfunktion. Wenn im Folgenden vom Stichprobenraum und angepasste Parameterräumen gesprochen wird dann wird das Werkzeug des SMs in Richtung Inferenz Statistik getrieben.

Motivation Das SM vereinfacht oder passt das vorherige Verständnis eines W-Raum an. Für die Modellierung eines W-Raum wird eine konkrete W-Verteilung/ W-Maß festgelegt. Dies wird im SM durch eine Tupel von W-Maße ersetzt. Dieses Tupel ist

$$(P^X_\vartheta)_{\vartheta \in \Theta} \; - \; \text{die parametrische Familie möglicher Verteilungen von } X.$$

Es wird angenommen, dass die *wahre* Verteilung zu diesem Tupel gehört. Im Allgemeinen wird von einem *Parameter* gesprochen. Dieser kann auch mehrdimensional sein, wie auch bei der Normalverteilung mit (μ, σ^2). Ebenso wird im Allgemeinen von einer ZV gesprochen, wenn auch hier eine mehrdimensional ZV vorliegt.

Definition I.7: Parameterraum

Die Menge aller möglichen Werte des Parameter ϑ wird *Parameterraum* Θ genannt.

Ein Parameter ϑ kann auch mehrdimensional sein.

Der Grundgesamtheit liegt ein unbekannter aber fester Parameter $\vartheta \in \Theta$ zugrunde - Beim SM ist der Parameter auf die Verteilungsfunktion bezogen. Im Regressionsmodell beziehen sich die Parameter auf die Beziehungen zwischen den ZV. Dieser Parameter wird aus der Stichprobe geschätzt. **Bsp.:** Sei $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ wird mit Hilfe der Stichproben $\vartheta = (\mu, \sigma^2) \in \Theta \subset \mathbb{R} \times \mathbb{R}^+$ versucht zuschätzen.

[11] Quelle: Moment(Wiki)) .

Erzeugerraum Um alle Formalitäten abzudecken, der Erzeugerraum mit dem W-Maß auf $(\Omega, \mathcal{A})$ ist auch eine Familien von W-verteilungen:

$$P_\vartheta \text{ für alle } \vartheta \in \Theta$$

Die Eigenschaft überträgt sich auf X, somit gilt für das W-Maß von X

$$P_\vartheta^X \text{ für alle } \vartheta \in \Theta$$

Beispiel 0-1-wertige Zufallsvariable Es sei n 0-1-wertige Zufallsvariablen $X_1, \dots, X_n$ auf $(\Omega, \mathcal{A}, P_p)$ gegeben. Diese sind unter P_p stochastisch unabhängig. Wie oben hergeleitet, gilt für das W-Maß bezüglich des Messbarkeits-Eigenschaft, für die Zufallsvariablen

$$P_p(X_i = 1) = p, \quad \forall i = 1, \dots, n. \tag{2.7}$$

] Es ist zu vermerken, dass es einen darauf zu achten ist, ob bei der Schreibweise mehrere Zufallsvariablen zu einer zusammengefasst werden, dann wird die $\bigotimes$-Schreibweise verwendet. Im Allgemeinen Fall kann für die Zufallsvariable als *Ergebnisraum* das Symbole M verwendet werden.

Normalverteilung Gegegeben $X = (X_1, \dots, X_n)$ stochastisch unabhängige Zufallsvariablen, ein SM, wenn angenommen wird, dass X_i normalverteilt ist, sieht wie folgt aus:

$$\left(\overset{n}{\mathbb{R}}, \mathcal{A}B^n, \bigotimes_{i=1}^{n} \mathcal{N}\left(\mu, \sigma^2\right)_{\mu \in \mathbb{R}, \sigma^2 \in (0,\infty]}\right)$$

Wenn von $X \sim \mathcal{N}\left(()\, \mu, \sigma^2\right)$ mit $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$ gesprochen wird, dann handelt es sich um ein *(parametrisches) statistisches Modell*.

Binomialverteilung Sei $M = \{0, 1, \dots, n\} \subset \mathbb{R}$, $\mathcal{A} = \mathcal{P}(M)$ und $P_p = Bi(n, p)$ mit $p \in (0, 1)$.

Binomialverteilung Sei $M = \{0, 1\}^n \subset \mathbb{R}^n$, $\mathcal{A} = \mathcal{P}(M)$ und $P_p = \bigotimes_{i=1}^{n} Bi(1, p)$ mit $p \in (0, 1)$.

Definition Statistisches Modell Die Definition für das SM ist:

Definition I.15: Statistisches Modell

(a) Sei $(\mathbb{R}, \mathcal{B})$ ein Messraum und $(P_\vartheta^X)_{\vartheta \in \Theta}$ eine Familien von W-Verteilungen auf diesen Messraum, dann nennt man das Triplet

$$\left(\mathbb{R}, \mathcal{B}, (P_\vartheta^X)_{\vartheta \in \Theta}\right)$$

ein *parametrisches statistisches Modell*.
(b) Sei $(\mathbb{R}, \mathcal{B})$ ein Messraum und $\mathbb{P}$ eine Menge von W-Verteilungen auf diesen Messraum, dann nennt man das Triplet

$$(\mathbb{R}, \mathcal{B}, \mathbb{P})$$

ein *nicht-parametrisches statistisches Modell*. [a]

[a] In der Regel ist $\vartheta \in \Theta \subset \mathbb{R}^k$ eine ein reeler Vektor.

Ein *nicht-parametrisches statistisches Modell* ist ein Triplet gleich eines *parametrisches statistisches Modell* mit der Ausnahme von Familie von W-Verteilungen. In diesem Fall besteht dieses auch aus einer Familie von W-Verteilungen, diese sind aber nicht von $\vartheta \in \Theta$ abhängig, sondern von einer Menge $\mathbb{Q}$ von W-Verteilungen auf $(\mathbb{R}, \mathbb{R}$:

$$\mathbb{Q} = \left\{ \bigotimes_{i=1}^{n} Q_i : Q_i \text{ ist eine W-Verteilung auf } (\mathbb{R}, \mathcal{A}) \right\}^{12}$$

3 Inferenz Statistik

3.1 Motivation

In der Schließenden Statistik werden die Ergebnisse (Parameter der Grundgesamtheit) überprüft und zu verifizieren, ob die usprüngelichen Annahmen über das Statisches Experiment zu treffen und ob die Stichprobe ausreichend Informationen enthält, eine solche Aussage zu treffen.
Die Essenz ist daher, anhand einer bekannten Stichproben x einen unbekannten Parameter Θ zu schätzen. Dafür konstruiert man einen *Schätzer*.

Im Folgenden wird der Schwerpunkt der Beispiele darauf liegen, dass es sich bei den *Parametern der Grundgesamtheit* um Parameter des SM handelt - der Verteilungsfunktion. Bei den Regressionparameter gelten in den meisten Fällen die gleichen Definitionen. Die Definitionen beschreiben im Kern, wie mit den Daten der Stichproben umgegangen werden soll. Eine Einschränkung auf Parameter der Verteilungsfunktion ist in der Regel nicht gegeben.

3.2 Schätztheorie

3.2.1 Statistik

Definition Statistik und Abgrenzung zum Schätzer Eine *Statistik* unterscheidet sich strukturell nicht von

- von einer Schätzfunktion,
- einer Zufallsvariable (Abbildung),
- einem Punktschätzer oder
- einer messbaren Funktion (I.6 - Messbarkeitseigenschaft (S.: 16)).

Definition I.16: Statistik

Sei $(\Omega, \mathcal{A})$ ein *messbarer Raum*[a] und (β, β) ein weitere *messbarer Raum*, so heißt eine *messbare Abbildung*[b]

$$S : (\Omega, \mathcal{A}) \to (\beta, \beta) \tag{3.1}$$

eine **Statistik**.

[a] Hinweis zur Regression: Es wird hier nicht auf eine Wahrscheinlichkeitsraum verwiesen. Für die Abbildung werden nur die Werte aus Ω auf β abgebildet. Für Ω kann auch gelten, dass es sich hierbei um einen Vektor handelt. So könnten verschiedenen Variablen X und Y auf einen andere Menge abgebildet werden.

[b] Siehe Definition Messbare Abbildung I.6 - Messbarkeitseigenschaft (S.: 16)

Im Kern steht, dass mit dieser Konstruktion Verteilungen und Bildmaße ermöglichen.
Der Unterschied, zwischen *Statistik* und *Schätzfunktion*, liegt in der Verwendung und Interpretation[13]

Statistik Ziel ist es Ordnung und Struktur aus den vorhanden Informationen (Daten) zu gewinnen. Dabei soll die Beobachtungstiefe reduziert werden.

Schätzfunktion Ziel ist hier mit den vorhanden Daten ($(\Omega, \mathcal{A})$) eine Auswertung zu gewinnen, um unbekannte Parameter/ Werte bestmöglich zu *schätzen*. Die Schätzfunktion unterliegt hier Gütekriterien.

Beispiele für Statistiken und Schätzfunktionen Wie schon oben mehrmals erwähnt, eine Statistik teilt sich den Aufbau mit der Schätzfunktion und weiteren Objekten. Das Ziel ist jedoch eine Struktur oder eine Beschreibung (Deskriptive Statistik) zu bewinnen.

- Die **Ordnungsstatistik** gibt den i-ten kleinsten Wert einer Stichprobe an.

[13] Bei der Schätzfunktion werden nicht binden, weitere Eigenschaften gefordert. Somit bleibt die Definition die gleiche.

- Eine **Teststatistik**, auch ***Prüfgröße*** oder ***Testgröße*** genannt, wird als Hilfsfunktion in der mathematischen Statistik im Bereich der Testtheorie verwendet. Bei einem Hypothesentest wird die *Nullhypothese* abgelehnt, wenn die *Teststatistik* einen gewissen Wert k über oder unterschreitet.

Am Beispiel der Tesstatistik: Gegeben sei

$$T : \mathcal{X} \to \mathbb{R}, \tag{3.2}$$

die *Teststatistik*, sowie ein *statistischer Test*

$$\phi : \mathcal{X} \to [0, 1], \tag{3.3}$$

mit

$$\phi(X) = \begin{cases} 1 & \text{falls} T(X) > k, \\ 0 & \text{falls} T(X) <= k \end{cases} \tag{3.4}$$

hierbei ist k eine feste Zahl, die auch kriterischer Wert genannt. Ist T ein $z-$Statistik,

$$\overline{X} = \frac{1}{n}\left(X_1 + X_2 + \ldots, X_n\right) \tag{3.5}$$

das Stichprobenmittel, $\mathcal{X} = \mathbb{R}^n$, ist eine typische Teststatistik

$$T(X) = \sqrt{n}\frac{\overline{X} - \mu}{\sigma} \tag{3.6}$$

Die **Teststatistik** ist standardnormalverteilt mit $T \sim \mathcal{N}\left(0, 1\right)$. Für den $t-$Test ist die $t-$Statistik $t-$verteilt mit $T \sim t_{n-1}$.

Sei $X = \{0, 1\}^n, \mathcal{A} = \mathcal{P}(X)$ und $(P_\vartheta)_{\vartheta\in\Theta} = \{Ber(n, \vartheta)|\vartheta \in \Theta\}$ das SM. Das *Stichprobenmittel*, als Schätzfunktion für den Parameter ϑ, ist gegeben durch das

$$\hat{\Theta} : X \to [0, 1] \text{ mit}, \tag{3.7}$$

$$\hat{\Theta}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.8}$$

Stichprobenmittel. Eine Statistik könnte sein

$$M : X \to \mathbb{R} \tag{3.9}$$

mit $M(x) = \sum x_i$ als Statistik. Mit der Statistik M wird nicht der Parameter $\vartheta = p$ geschätzt, sondern nur eine Informationsreduktion vorgenommen.[14]

3.2.2 Schätzer

Wie erwähnt, ein *Schätzer* oder auch *Schätzfunktion* oder *Schätzstatistik* genannt, ist eine Funktion aus den Daten der Stichproben eine Aussage über unbekannte Parameter einer Grundgesamtheit zu tätigen.

Definition I.8: Stichprobenraum

Ein *Stichprobenraum* $\mathcal{X}_n$ besteht aus der Menge aller möglichen Stichprobenwerte $(x_1, \ldots, x_n) \in X_n \subseteq \mathbb{R}^n$.

[14] Quelle: Statistik (Funktion), Ordnungsstatistik

Definition I.17: Schätzer/Schätzfunktion

Sei Ψ eine Statistik auf $(\Omega, \mathcal{A})$ nach (β, β) (I.16 - Statistik (S.: 22)) und $X_1, \ldots, X_n$ eine Stichprobe mit $(X_1, \ldots, X_n) \in \Omega$.
Eine **Schätzer** für den Parameter ϑ oder den Funktionswert γ ist eine Abbildung

$$\Psi : \Omega \to \beta. \tag{3.10}$$

die messbar von $X_1, \ldots, X_n$ abhängt.

Bei dem Schätzer[15] sind drei Punkte besonders wichtig

- Zu schätzen kann ein Wert ϑ oder eine Parameterfunktion/Kenngröße γ sein.
- Ein Schätzer ist auch eine Zufallsvariable $\hat{\Theta}_n$.
- Im Theoretischen wird mit den Zufallsvariablen $X_1, \ldots, X_n$ im praktischen mit den Bildern der Zufallsvariablen/Stichprobenwerte $x_1, \ldots, x_n$ gearbeitet.

Parameter oder Parameterfunktion In Verknüpfung mit der Entscheidungstheorie wird der Parameter nicht als Skalar abgebildet, sondern per Funktion, welche aus dem Parameterraum Θ in den Entscheidungsraum β abbildet.

$$\gamma : \Theta \to \beta \tag{3.11}$$

Der Raum (β, β) wird auch *Entscheidungsraum* genannt. [16].
Um genau zu sein, bildet die Entscheidungsfunktion in einen *Entscheidungsraum* $(\beta, \beta$.

Definition I.9: Entscheidungsraum

Ein *Entscheidungsraum* ist ein Messraum, bei dem die $\sigma-$Algebra alle einelementigen Mengen enthält, sodass $\forall \omega \in \Omega$ die Menge $\{\omega\} \in \beta$ liegt. Beispielsweise ist $(\mathbb{R}, \mathbb{R}$ ein Entscheidungsraum.

Beispiel Parameterfunktion Für den Verteilungsparameter eine Bernoulli Verteilung ϑ, wird dieser in der Regel mit dem Stichprobenmittel auf

$$\gamma(\vartheta) = \vartheta \tag{3.12}$$

abgebildet.
Für die Normalverteilung wird für den Parameter σ das Stichprobenmittel auf

$$\gamma(\sigma) = \sigma^2 \tag{3.13}$$

abgebildet.

Symmetrie Schätzer als Zufallsvariable Wie in der Übersicht geschildert, wird aus der Schätzfunktion auch gleichzeitig eine eigene Zufallsvariable, siehe 1.4.
Die Abbildung Ψ als Schätzer des Parameters ϑ ist gleichzeitig eine Zufallsvariable:

$$\hat{\Theta_n} := \Psi(X_1, \ldots, X_n) \tag{3.14}$$

mit eigener Verteilung.

[15] Schätzfunktion
[16] Parameterfunktion

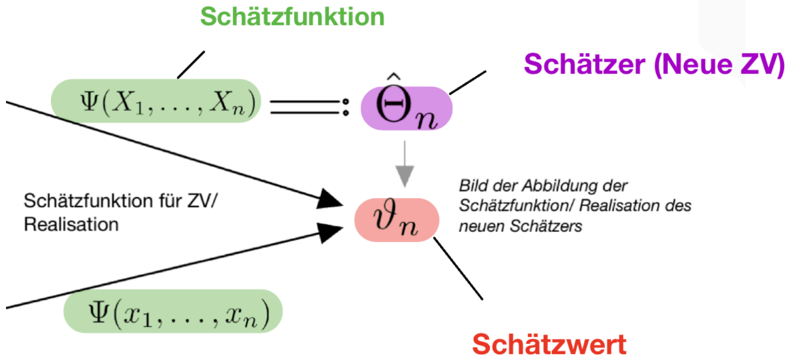


Abbildung 2.11: Drei Faltigkeit - Schätzer, Schätzfunktion und Schätzwert

Ebenso ist nochmals daraufhinzuweisen, dass es eine Unterschied zwischen der Funktion des Schätzers und des tatsächlichen Schätzwertes gibt.

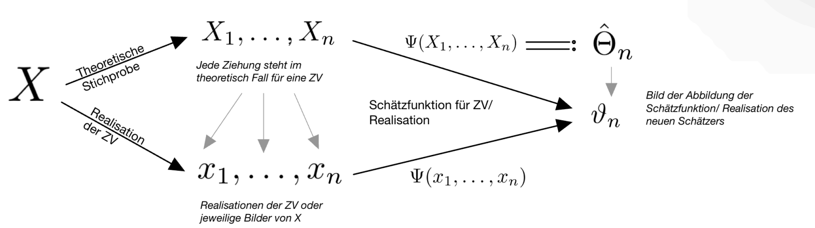


Abbildung 2.12: Theoretische und reale Betrachtung einer Stichprobe

Die Realisation von $\hat{\Theta}_n$ wird ebenfalls mit einem Dach versehen: $\hat{\vartheta}_n$

Verteilung der Schätzfunktion Wie im vorherigen Abschnitt erwähnt, ist die Schätzfunktion ebenso eine eigenen Zufallsvariaben $\hat{\Theta}$. Je nach Schätzfunktion besitzt diese eine eigene Verteilung.

Sei $X_1, \dots, X_n$ unabhängig und identisch verteilt ((engl.) independent and identically distributed) (i.i.d) Normalverteilt gilt für die Schätzfunktion, Stichprobenmittel

$$\hat{\Theta} \sim \mathcal{N}\left(\mu_X, \sigma_X / n\right) \tag{3.15}$$

Die Varianzschätzung S_n^2 um konstante Terme angepasst, verteilt sich wie folgt:

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \mathbb{X}(n-1) \tag{3.16}$$

[17]

3.2.3 Eigenschaften von Schätzern

In der klassischen Inferenz Statistik sind

- Erwartungstreue/ Mediantreue
- Effizienz
- Konsitenz,

[17] Quelle: Schätzfunktion, Wiki

- Suffizienz
- Normalität
- Linearität
- Robustheit

die wichtigsten Eigenschaften eines Schätzer $\hat{\Theta}_n$. Einige der Eigenschaften sind asymptotisch: Verhalten von $\hat{\Theta}_n$ für $n \to \infty$.

Erwartungs- und Mediantreue Für den Erwartungswert gilt wie unter 2.2.4:

$$E\left[\hat{\Theta}_n\right] := \int \hat{\vartheta} dP_{\hat{\Theta}_n}(\hat{\vartheta}_n) \tag{3.17}$$

Wiederholter Hinweis: Diese Eigenschaft ist <u>nur</u> wichtig, dass die Eigenschaft erfüllt ist!

Definition I.18: Erwartungstreue

Ein Schätzer $\hat{\Theta}_n$ heißt *erwartungstreu* oder *unverzerrt*, wenn

$$E\left[\hat{\Theta}_n\right] = \gamma(\vartheta)\, \forall \vartheta \in \Theta. \tag{3.18}$$

oder simpler $E\left[\hat{\Theta}_n\right] = \vartheta\, \forall$.

Plausible Eigenschaft Sei $\hat{\Theta}_n = \overline{X}_n = \frac{1}{n}\sum_i^n X_i$ das *Stichprobenmittel* und $\vartheta = E[X] = \mu$ mit $X_i \sim i.i.d.$ gilt

$$\begin{aligned} E\left[\hat{\Theta}_n\right] &= E\left[\overline{X}_n\right] && (3.19)\\ &= E\left[\frac{1}{n}\sum_i^n X_i\right] && (3.20)\\ &= \frac{1}{n}\sum_i^n E[X_i] && (3.21)\\ &= \frac{1}{n} n\mu \text{ mit } E[X_i] = \mu_i = \mu && (3.22)\\ &= \mu && (3.23) \end{aligned}$$

Der Schätzer $\overline{X}_n$ ist erwartungstreu. Wäre $\gamma(\mu) = \mu$, wäre der Schätzer ebenso erwartungstreu.
Bezüglich der Verteilung von $\overline{X}_n$ ist gegeben durch $X_i \sim \mathcal{N}\left(\mu, \sigma^2\right)$[18] $\overline{X}_n \sim \mathcal{N}\left(\mu, \sigma^2/n\right)$.

Erwartungstreu, aber riskant Sei der Schätzer

$$\Psi(X_1, \ldots, X_n) \mapsto X_1, \tag{3.24}$$

so ist der Schätzer erwartungstreu für μ. Dies ist jedoch ein unplausibler Schätzer, weil nur die erste Beobachtung berücksichtigt wird.

Abschluss: Es gibt viele *erwartungstreue* Schätzer die unplausible sind. Ein *idealer* Schätzer ist ein solcher welcher für $\forall \vartheta \in \Theta$ und $(x_1, \ldots, x_n)$ den wahren Wert für $\gamma(\vartheta)$ schätzt.

[18] Jede Zufallsvariable steht in dem Fall für das gleiche Merkmal. In Stichproben können jedoch auch mehrere Merkmale betrachtet werden.

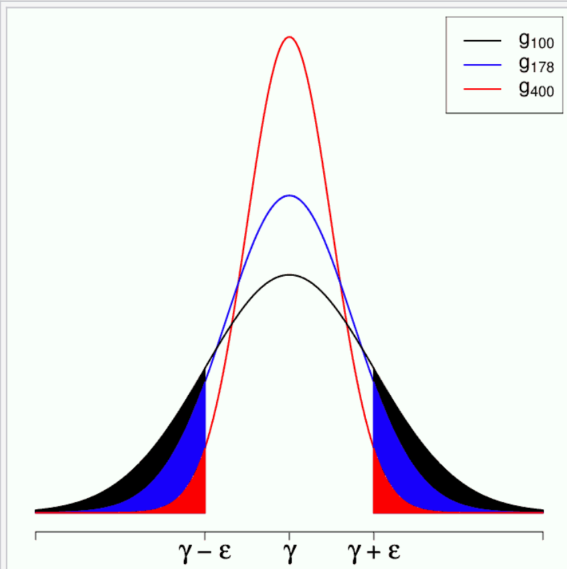
Konsitenz Bei der *Konsitenz* geht es darum, dass mit einer höhere Stichprobe, die Abweichung, zwischen den Schätzerwert und den Parameter gegen null geht.

Definition I.19: Erwartungstreue

Sei $\Psi(X_1, ..., X_n)$ eine *Schätzer* für die Funktion $\gamma(\vartheta)$ für alle $\vartheta \in \Theta$. Für den unbekannten Parameter ϑ der Grundgesamtheit heißt Ψ für ein beliebiges $\epsilon > 0$ ***konsitenz***, wenn

$$\lim_{n\to\infty} P_\vartheta \left(|\Psi(X_1, ..., X_n) - \gamma(\vartheta)| >= \epsilon\right) = 0 \tag{3.25}$$

Mit zunehmenden n wird die Wahrscheinlichkeitsdichte des Parameters $\gamma(\vartheta)$ schmaler. Somit wird für jedes höhere n mehr Fläche zwischen $\Psi \pm \epsilon$ liegen.



Wahrscheinlichkeitsdichten für die konsistenten Schätzfunktionen g_n ($n = 100, 178, 400$). Mit steigendem Stichprobenumfang wird der unbekannte Parameter γ immer genauer geschätzt.

Abbildung 2.13: Wahrscheinlichkeitsdichte wird konzentierter um γ

Mittlere Quadratische Fehler (MSE) Die Mittlere quadratischer Fehler (MQE) (Engl.: Root Mean Squard Error (MSE)) ist die Eigenschaft, welche untersucht, ob eine systematische Verzehrung vorliegt - ein Bias.

Definition I.20: Mittler Quadratischer Fehler (MSE)

Sei $\Psi(X_1, ..., X_n)$ eine *Schätzer* für die Funktion $\gamma(\vartheta)$ für alle $\vartheta \in \Theta$. Der MSE definiert sich über

$$MSE(\Psi, \gamma(\vartheta)) := E\left[|\Psi(X_1, ..., X_n) - \gamma(\vartheta)|^2\right] = Var\left[\Psi\right] - Bias(\Psi)^2, \vartheta \in \Theta \tag{3.26}$$

Der Bias einer Zufallsvariablen definiert sich wie folgt:

Definition I.10: Bias

Sei Ψ ein Schätzer für $\vartheta \in \Theta$:

$$Bias(\Psi) := Bias(\Psi, \vartheta) := E\left[\Psi\right] - \vartheta \tag{3.27}$$

Herleitung des Mittleren Quadratischen Fehlers

Durch gleichzeitige Addition und Subtration von $E[\Psi]$ bleibt das Ergebnis unverändert:

$$MSE(\Psi) = E\left[(\Psi(X_1,\dots,X_n) - \gamma(\vartheta))^2\right] \tag{3.28}$$
$$= E\left[(\Psi(X_1,\dots,X_n) - E[\Psi] + E[\Psi] - \gamma(\vartheta))^2\right] \tag{3.29}$$
$$\text{mit } \Psi = \Psi(X_1,\dots,X_n) \text{ und } \gamma = \gamma(\vartheta) \tag{3.30}$$
$$= E\left[(\Psi - E[\Psi])^2 + 2(\Psi - E[\Psi])(E[\Psi] - \vartheta) - (E[\Psi] - \gamma)\right] \tag{3.31}$$
$$= E\left[(\Psi - E[\Psi])^2 + 2(\Psi - E[\Psi])(E[\Psi] - \vartheta) - (E[\Psi] - \gamma)^2\right] \tag{3.32}$$
$$\tag{3.33}$$

Die Erwartungswert-Zerlegung erlaubt, dass der Erwartungswert auf die einzelnen additiven Teile angewandt werden kann.

$$E\left[(\Psi - E[\Psi])^2\right] + E[2(\Psi - E[\Psi])(E[\Psi] - \vartheta)] - E\left[(E[\Psi] - \gamma)^2\right] \tag{3.34}$$

Der Mittlere Termin mit $E[\Psi] - E[\Psi]$ wird zu Null. Der letzte Teil wir zu

$$E\left[(E[\Psi] - \gamma)^2\right] = E[(E[\Psi] - \gamma)(E[\Psi] - \gamma)] \tag{3.35}$$
$$= (E[\Psi] - \gamma)^2 \tag{3.36}$$
$$= Bias(\Psi)^2 \tag{3.37}$$

Der erste Termin ist die Varianz, was zum Schluss ergibt:

$$MSE(\Psi) = Var[\Psi] - Bias(\Psi)^2 \tag{3.38}$$

[19]

3.3 Methoden zur Schätzfunktionen Konstruktion

In Folgenden werden Methode,

- Momentummethode,
- Maximum-Likelihood Methode,
- Kleinstes Quadrate Methode,
- Bayer Methode

aufgezeigt, mit welche Schätzfunktionen konstruiert werden. Im Abschnitt zuerst wurde aufgezeigt, welche Eigenschaften eine Schätzer bewerten.[20]
Die verschiedenen Verfahren liefern je nach Beispiel gleiche Ergebnisse, habe jedoch unterschiedliche Eigenschaften und Güte für ihre Ergebnisse. Somit konkurrieren die Methoden zum einen, zum anderen ergänzen sie sich.

3.3.1 Momenten Methode

Sei x_i die Beobachtung von n stochastisch unabhängig Zufallsvariablen X_i mit bekanntem Verteilungstyp.

[19] Quellen hierfür sind: Uni Berlin, Schätzfunktion
[20] Schaetzmethoden, Uni UlmStaMa

Herleitung Am Beispiel der Normalverteilung werde die Parameter μ und σ theretisch durch

$$m_1 = E\left[X^1\right] = \mu \,\text{mit}\, m_2 = E\left[X^2\right] = \sigma^2 + \mu^2 \tag{3.39}$$

bestimmt.[21]
Es gibt noch das *empirische Moment* $\hat{m_k}$, welches sich aus den Realisation x_i mit

$$\frac{1}{n}\sum_{i=1} n x_i^k \tag{3.40}$$

bestimmt. Als Beispiel ist $\hat{m_k}$ der *empirische Mittelwert*.
Die **simple Idee** ist, die *theoretischen* mit dem *empririschen Momenten* gleichzusetzen:

$$\hat{m_k} = m_k \, k = 1, \ldots, n \tag{3.41}$$

Dabei benötigt es p Gleichungen, um p Parametern zu finden.

Beispiel Um die Parameter der Normalverteilung μ und σ^2 zu schätzen, wird

$$m_1(\mu, \sigma^2) = E\left[X_i\right] = \frac{x_1 + \ldots, x_n}{n} \,\text{und} \tag{3.42}$$

$$m_2(\mu, \sigma^2) = E\left[X_i^2\right] = \frac{x_1^2 + \ldots, x_n^2}{n} \tag{3.43}$$

gleichgesetzt. Sei $m_1(\mu, \sigma^2) = \overline{x}$. Es wird erhalten

$$\hat{\sigma^2} = \frac{1}{n}\sum (x_-\overline{x})^2. \tag{3.44}$$

Konzept

Vorteil Die Vorteil dieser Methode liegen in der einfachen Berechnung, selbst wenn das Gleichungssystem durch nicht-lineare Gleichungen bestimmt wird und die Lösung iterativ gefunden werden muss. Die Momentenmethoden kann auch bei abhängigen Zufallsvariablen eingesetzt werden. Während die Maximum Likelikhood Estimation (MLE) eine komplizierte Lösung benötigt.

Nachteil Die Einfachkeit der Berechnung ist auch gleichzeitig der Nachteil. Nicht alle Informationen aus einer Stichprobe werden genutzt. So kann bei kleinen Stichproben der Schätzwert außerhalb des Parameterraums liegen. Ebenso sind Schätzer meist nicht effizient. So zum Beispiel ist bei einer Gleichverteilung Momentenschätzer weniger effizient als MLE.

3.3.2 Maximum Likelihood Schätzer

Sei x_i die Beobachtung von n stochastisch unabhängig Zufallsvariablen X_i mit bekanntem Verteilungstyp.

Die MLE wurde von Carl Friedrich Gauß entwickelt. Bei der MLE wird der Schätzwert für den unbekannten Parameter ϑ bestimmt. Dieser Schätzer bestimmt sich, dass gefragt, wird, welcher Parameter weißt die höchste Wahrscheinlichkeit auf, dass die Stichprobe $x = (x_1, \ldots, x_n)$ vorkommt.[22]

Vorteile Der Vorteil der MLE liegt in den Eigenschaft von Schätzern der MLE:

- effizient und
- asymthotische Effizienz, ebenso können vergleichsweise einfach
- verschiedenen Signifikanztests durchführen.

Nachteile Ein großer Nachteil ist, dass der Verteilungstype für X_i vorher bekannt sein muss. Ist diese Wahl falsch, so kommt es zu Verzerrungen in den Ergebnissen. Wenn die Parameter durch numerische Maximierung gefunden werden müssen, besteht das Risiko, dass es sich um ein lokales nicht ein globales Maximum handelt.

Vergleich Ist X_i normalverteilt, liefert MLE und die Momentenmethode das gleiche Ergebnis für die Parameter. Der Unterschied liegt im systematischen Fehler. Bei der Momentenmethode kommt es zu einem kleineren Fehler in der Standardabweichung.

[21] Conjektor: Der 2-te theoretische Erwartungswert ist gleich des 1-ten und 2-ten zentralen Moments.
[22] Im Regelfall werden die Parameter der Verteilung gesucht.

Herleitung Für MLE wird entweder angenommen, dass die Verteilung diskret oder absolute stetig ist.

Gegeben eine Familie diskreter Familie von Verteilungen $P^X = \left\{P_\vartheta^X, \vartheta \in \Theta\right\}$, soll auf Grundlager der Beobachtungen von X der Parameter ϑ geschätzt werden.

Definition I.21: Maximum-Likelihood Funktion

Die Funktion

$$\mathbb{L}(\vartheta|x) \overset{def}{=} P_\vartheta^X(x) = P_\vartheta(\{X = x\}) \tag{3.45}$$

heißt *Maximum-Likelihood Funktion*. Die $\mathbb{L}(\vartheta|x)$wird als Funktion abhängig nur vom Parameter ϑ verstanden, weshalb es ausreicht, wenn $\mathbb{L}(\vartheta)$ geschrieben wird.

- Bei einem fixen ϑ ist die Maximum-Likelihood Funktion (MLF) gleich der Zähldichte von X.
- Die Funktion
$$\mathbb{L}(\vartheta|x) = \mathbb{L}\left(x_1, \ldots, x_n; \vartheta\right) = P_\vartheta^X\left(X_1 = x_1, \ldots, X_n = x_n\right) \tag{3.46}$$
gibt an, mit welcher Wahrscheinlichkeit die Kombination (Stichprobe) gegeben des unbekannten Parameters ϑ ist.
- Wenn die Ziehungen X_i unabhängig und identisch ist, gilt
$$\mathbb{L}(\vartheta|x) = P_\vartheta^X(X_1 = x_1) \cdot P_\vartheta^X(X_n = x_n) = \prod P_\vartheta^X(X_i = x_i) \tag{3.47}$$

Das Grundprinzip ist

$$\mathbb{L}(\vartheta|x) \to \max. \tag{3.48}$$

Der Maximum-Likelihood Schätzer ist definiert durch

$$\Psi_{ML} = \arg\max_{\vartheta\in\Theta} \mathbb{L}(\vartheta|x). \tag{3.49}$$

Es kann sein, dass es mehrer Lösungen gibt.

3.3.3 Methode der Kleinsten Quadrate

Allgemeine Anwendung Die MKQ kann in allgemeiner Form und angepasst für die Regression definiert werden. Der erste Fall ist der wenig verwendete Fall. Der Schwerpunkt dieser Methode liegt darin, Schätzer für die Parameter des General Linear Model (GML) zu bestimmen.

Im Allgemeinen Fall[23] wird ein Parameter $\hat{\mu}$ bestimmt, welcher einen Abstand zwischen den Werten der gesuchten Stichproben y und und einer festgelegten Grenze m minimiert. Für m können verschiedenen Werte vorliegen. Meist handelt es sich um den zu bestimmenden Erwartungswert. Im Fall einer Regression handelt es sich um den Vektor, der vom GML produziert wird und gespeißt durch mit erhaltenen Werte x. Dieser soll nach den erzeugten Stichprobenwerten y kommen. Dabei ist die Gleichung abhängig von den ***variablen*** Parametern β. Im Regressionsfall ist es jedoch auch der Erwartungswert $E\left[Y\right]$ des GML.

Definition I.22: Kleinste-Quadrat Methode

Die MKQ schätzt den Wert für den unbekannten Wert μ durch $\hat{\mu} \in \Theta$, den *Vektor der minimalsten Abstände zu* y:

$$\hat{\mu} = \arg\min_{m\in\Theta} ||y - m||^2. \tag{3.50}$$

Es gilt $||y - \hat{\mu}||^2 \leq ||y - m||^2$ für alle $m \in \Theta$.

Dies bedeutet, gegeben der Stichproben y der Zufallsvariable Y wird ein Vektor m gesucht, welcher den Abstand zwischen $\hat{\mu}$ und y minimal ist. [24]

[23] Mathematik, Arens, Seite 1533

[24] Die Operator arg min gibt nicht den Funktionswert wie min zurück, sondern die Menge der gesuchten Parameter. In dem Fall m.

Anwendung bei der Regression In Fall eines GML ist m durch den Erwartungswert des Regressionsmodell in Abhängigkeit des Parameter β gegeben. Gleichfalls wird in das Modell die Ausprägungen der abhängigen Variable X eingesetzt. Zu vergleichen sind ist die Stichprobe y:

$$\hat{\beta} = \arg\min_{\beta\in\Theta} ||y - E[y]||^2 \tag{3.51}$$

$$= \arg\min_{\beta\in\Theta} ||y - E[\beta x]||^2 \tag{3.52}$$

$$\tag{3.53}$$

Hier hängt der Erwartungswert $E[Y]$ direkt oder durch eine bekannte Funktion von den Parametern (den zu schätzenden Parametern) sowie einer Störgröße ab. Daher bestimmt man die gesuchten Parameter so, dass die Summer der Störgrößen möglichst klein sind.

Das einfache Beispiel ist die lineare Regression $Y = \beta_0 + \beta_1 X$, mit den Parametern β_0 und β_1.[25] Diese wird von einer Störgröße, einer Zufallsvariablen ϵ_i überlagert, somit beobachtet man nur (y_i, x_i) mit $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$. Für die Zufallsvariable Y_i gilt

$$E[Y_i] = \beta_0 + \beta_1 X_i, \tag{3.54}$$

wobei $E[\epsilon_i] = 0$ ist. Die Funktion der Summe der quadrierten Störgrößen

$$\sum (y_i - (\beta_0 + \beta_1 x_i))^2 \rightarrow \min \tag{3.55}$$

um die Schätzwerte für β_0 und β_1 zu finden.
Für ein lineares Modell gilt, sind die Schätzer für β_0, β_1 gegeben durch die Minimierung der Summer der Quadrate

$$(\Psi_{KQ,\beta_0}, \Psi_{KQ,\beta_1}) := \arg\min \sum (Y_i - \beta_0 - \beta_1 X_i)^2. \tag{3.56}$$

Versteht man die Schätzfunktion Ψ_{KQ} als Funktion von β_0 und β_1, und wie im Beispiel unten, wird diese minimiert, erhält man für β_0 die Lösung

$$\beta_1 = \frac{Cov_{X,Y}}{Var_X} \tag{3.57}$$

und β_0 wird durch einsetzen gefunden.

Vorteil Es benötigt eine Annahmen über die Verteilung von Y_i. Es wird nur der Zusammenhang zwischen dem Erwartungswert $E[Y_i]$ und den Parametern benötigt. Die Methode ist daher für ein breites Spektrum anzuwenden. Die große Anwendung findet die Methoden, wenn nicht eine einzelne Variablen im Fokus steht, sondern der Zusammenhang verschiedener Variablen.

Nachteil Der Vorteil ist auch gleichzeitig der Nachteil. Weil Informationen über die Verteilung von Y_i nicht genutzt werden, weißen die Schätzer schlechtere Eigenschaften als Funktionen der MLE auf. Weißt die Regressionsfunktion einen nicht linearen Zusammenhang auf, müssen die Parameter numerisch geschätzt werden.

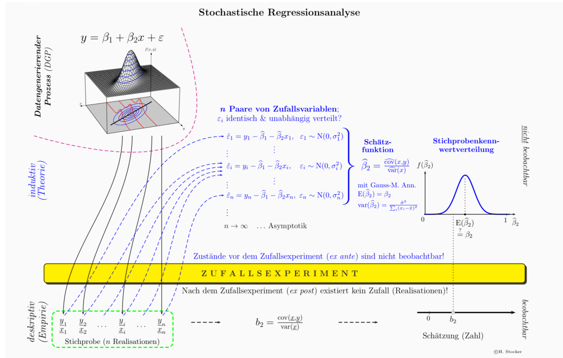
Schätzeigenschaften Bei der Schätzfunktion und den Eigenschaften muss unterschieden werden, zwischen den Lösungen von Ψ_{KQ,β_1} und Ψ_{KQ,β_1} selbst.
Die *erwartungstreue* für β_1 ergibt sich, wenn

$$E\left[\hat{\beta}_1\right] = E\left[\frac{Cov(X,Y}{Var[X]}\right] \tag{3.58}$$

$$= \beta_1 + E\left[\frac{Cov(X,\epsilon}{Var[X]}\right], \tag{3.59}$$

[25] Je nach Modell, wird hier schon von einer Zufallsvariablen Y oder nur einer Variablen Y gesprochen. In dem Beispiel wird der zweite Fall betrachtet.

und somit der Erwartungswert $Cov(X, \epsilon) = 0$ ist, und somit der Schätzer für β_1 *erwartungstreu*.



3.3.4 Anwendung an einem Beispiel

In einem Spiel kann man 1 Euro mit der Wahrscheinlichkeit p verlieren, $1-2p$ gewinnen und p weder gewinnen noch verlieren.
Das Spiel wird sechs mal gespiel mit der Ziehung

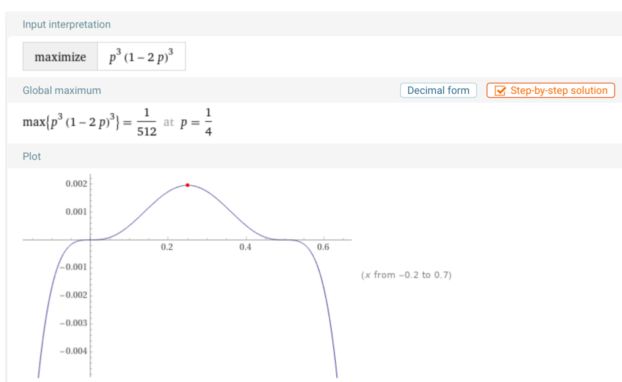
$$x = (-1, 1, -1, 0, 1, 1). \tag{3.60}$$

Wie groß ist der Wert für p?

Maximum-Likelihood Methode (MLE) Mit der MLE wir der Erwartungswert für p wie folgt berechnet.

$$\arg\max_p P(X_1 = -1, X_2 = 1, X_3 = -1, X_4 = 0, X_5 = 1, X_6 = 1) \tag{3.61}$$
$$= P(X_1 = -1) \cdot P(X_2 = 1) \cdot P(X_3 = -1) \cdot P(X_4 = 0) \cdot P(X_5 = 1) \cdot P(X_6 = 1) \tag{3.62}$$
$$= (1-2p)^3 p^3 \tag{3.63}$$
$$\to p = 0,25 \tag{3.64}$$
$$\tag{3.65}$$

Lösung:



Methode der Kleinsten Quadrate Der interessante Teil ist, dass für den KQS der Erwartunswert $E[X_i]$ benötigt wird, siehe Methode oben. Dieser wird als *nicht geschätzer* Erwartungswert betrachtet. Jede Ausprägung wird betrachtet und mit der zugeordneten Wahrscheinlichkeit betrachtet:[26]

$$E[X_i] = (1) \cdot (1-2p) + (0) \cdot (p) + (-1) \cdot p \tag{3.66}$$
$$= 1 - 2p - p \tag{3.67}$$
$$= 1 - 3p \tag{3.68}$$

[26] Achtung: Es für für $E[X_i]$ nicht ein separater Schätzer eingesetzt, $E[X_i] = -p + p - p + 0 + p + p$ ist falsch.

Für jede Ziehung wird jetzt die quadratische Differenz zwischen der tatsächlichen Ausprägung und der erwarteten Ausprägung bestimmt.

$$\Psi_{KQ,p} = \text{arg}\min_{p\in[0,1]}(-1-\cdot E[X])^2+(1-\cdot E[X])^2 \tag{3.69}$$

$$+(-1-\cdot E[X])^2+(0-\cdot E[X])^2+(1-\cdot E[X])^2+(1-\cdot E[X])^2 \tag{3.70}$$

$$=\text{arg}\min_{p\in[0,1]}(-1-(1-3p))^2+(1-(1-3p))^2+(-1-(1-3p))^2 \tag{3.71}$$

$$+(0-(1-3p))^2+(1-(1-3p))^2+(1-(1-3p))^2 \tag{3.72}$$

$$=\text{arg}\min_{p\in[0,1]} 9-30p+54p^2 \tag{3.73}$$

In im Konzept der Methode besprochen, wird nicht der Funktionswert gesucht, sondern die p, welche die Funktion $9-30p+54p^2$ minimiert. Mit der pq-Formel erhält man $\Psi_{KQ,p} = p_{KQ} = 5/18$.
Wie man sieht, unterscheidet sich bei diesem Beispiel, der Wert für den Schätzer von p abhängig von der Methode.