

1 Grundbegriffe der Stichprobentheorie

Egal, ob es sich um eine physikalische Messung, die Einkommenssituation eines deutschen Haushaltes oder den durch einen Waldbestand gegebenen Holzvorrat handelt – jede Beobachtung und jede Sachlage resultiert aus verschiedenen und sich überlagernden Einflüssen. Dabei werden die Einflussfaktoren, welche außerhalb unserer Wahrnehmung oder Vorstellungskraft liegen, unter dem Begriff „Zufall“ zusammengefasst. Die Gesetzmäßigkeit einer Beobachtung oder eines Merkmals ist oft zufallsabhängig. Etwas umfassender, nämlich die Gesamtheit aller Ausprägungen, organisiert durch sämtliche Einflussfaktoren, bezeichnet schließlich die sogenannte *Grundgesamtheit* oder *Population*. In der Statistik verfolgt man nun das Ziel, an Hand von Beobachtungsdaten die Population einer interessierenden Erscheinung bestmöglich zu beschreiben.

Selbst eine Naturkonstante lässt sich nach dieser Auffassung von den uns umgebenen Zuständen nicht direkt, also fehlerfrei, ermitteln. Um sie dennoch in Erfahrung zu bringen, ist eine Charakterisierung der Fehlerverteilung naheliegend. Hierfür sollen die mit einer Methode wiederholt durchgeführten Messungen $y_1, y_2, \dots, y_n$ dienen und die Bezeichnung *Stichprobe* erhalten. In dieser Situation lassen sich die Daten dahingehend interpretieren, dass jeder Wert der Stichprobe der gleichen Verteilung folgt und zugleich kein Wert die sonstigen Messungen beeinflusst. Es handelt sich also um einen bequemen Sonderfall, den einer sogenannten unabhängigen, identisch verteilten (u. i. v.) Zufallsstichprobe.

Oftmals werden Stichproben jedoch durch kompliziertere *Auswahlschemata* erzeugt, da zusätzliche Informationen vor oder während der Erhebung der Daten zur Verfügung stehen und mit ihrer Verwendung eine bessere Beschreibung der interessierenden Gegebenheit möglich ist. Aber auch ein

gewisser Informationsmangel oder Einschränkungen bei der praktischen Umsetzung des Ziehungsmechanismus verbieten eine Auffassung der Daten als u. i. v. Stichprobe.

Letztlich hängt eine Aussage über die Grundgesamtheit der Einflussfaktoren eines interessierenden Sachverhaltes natürlich genau von dem Ausgang der Stichprobenziehung ab, zumal man diese im Kontext aller denkbaren Stichprobenverteilungen betrachten muss. Die Vorstellung von der Gesamtheit aller plausiblen Stichprobenverteilungen wird durch den Begriff (*statistisches Stichproben-*)*Experiment* oder *Modell* formalisiert.

In diesem Kapitel sollen die drei grundlegenden Begriffe, nämlich der Begriff einer Population, eines Auswahlschemas und eines statistischen Modells präzisiert und veranschaulicht werden. Nicht zuletzt gelangen wir zu dem sehr wichtigen Objekt namens *Stichprobendesign*, welches die drei Begriffe gewissermaßen verbindet.

Anders als in der gegenwärtig geläufigen Literatur zur Mathematischen Statistik und zur Stichprobentheorie werden diese Begriffsbildungen dabei derart herausgebildet, dass sich schließlich alle der eingangs genannten Beispiele in diesem allgemeinen Rahmen behandeln lassen. Das Messen einer physikalischen Größe, die Einkommensverteilungsschätzung und die Waldinventur werden uns somit als Anwendungsfälle durchweg begleiten.

1.1 Populationen

Ein Wahrscheinlichkeitsmaß P ordnet Ereignissen, also gewisse Mengen von möglichen Ausprägungen einer zufälligen Erscheinung, eine (relative) reelle Zahl aus $[0, 1]$ zu. Es ist das zentrale Objekt zur Modellierung und Analyse des Zufalls in der Wahrscheinlichkeitstheorie. Für die Zwecke der Statistik betrachten wir nun etwas darüber hinaus die Menge aller Ausprägungen, welche oft in einem natürlich gegebenen Register angeordnet sind. Zum Beispiel sind die zu untersuchenden Elemente gelegentlich an eine räumliche oder zeitliche Komponente gekoppelt. In anderen Fällen lassen sich diese durchnummerieren.

In diesem Abschnitt wollen wir nun einen Begriff für diese organisierte Gesamtheit der Ausprägungen präzisieren. Bevor wir aber zur Definition dieses zentralen Objektes der Statistik kommen, soll das folgende Beispiel

die bereits angedachten Aspekte veranschaulichen und einen Ausblick auf die Verwendung des Begriffs geben.

1.1 Beispiel (Umfragen) Politik und Wirtschaft interessieren sich häufig für die Einkommensverteilung einer Bevölkerung oder auch nur für die einer gewissen Gesellschaftsschicht. Eine zeitnahe Ermittlung dieses Sachverhaltes soll per Umfrage realisiert werden. Vollständig bekannt ist eine Liste der Haushalte, für welche die Menge $\mathbb{R} := \{1, \dots, N\}$ steht. Bei der Untersuchung der Einkommensverteilung der gesamten Bevölkerung ist jedes Element von $\mathbb{R}$ Träger des Merkmals „Einkommen“. Die Menge

$$G := \{(1, Y(1)), \dots, (N, Y(N))\} \subset \mathbb{R} \times \mathbb{R}_+$$

stellt das zu untersuchende Objekt „Population“ dar, wobei $Y(k)$ für das Einkommen des Haushaltes k steht. Ferner lässt sich die Merkmalszuordnung $Y : \mathbb{R} \rightarrow \mathbb{R}_+$ mit G identifizieren, so dass^{1 2} $G \sqsubset \frac{1}{N}\zeta(\cdot)$ für die Merkmalsverteilung, d. h. Einkommensverteilung, steht.

Eine Untersuchung der Einkommensverhältnisse einer speziellen Teilgruppe von $M \leq N$ Haushalten ist dagegen etwas schwieriger. Mit der Teilgruppe kann zum Beispiel eine gewisse Einkommensschicht gemeint sein, d. h.

$$G_0 := \{(k, Y(k)) : Y(k) \in [a, b], k = 1, \dots, N\} \subset \mathbb{R} \times \mathbb{R}_+.$$

Die Registerpositionen dieser Mitglieder der interessierenden Gesellschaftsschicht sind dabei nicht bekannt, in diesem Sinne also in der Liste $\mathbb{R}$ (zufällig) verstreut und können nicht direkt erhoben werden. Statt an G sind wir nun an der Beschreibung von G_0 interessiert. Mit³ $M := |G_0| := \zeta(G)$ ist hier $G_0 \sqsubset \frac{1}{M}\zeta(\cdot)$ die Merkmalsverteilung. ♦

Dieses Beispiel betont in seiner Darbietung drei wesentliche Aspekte: (1) Für einen Ziehungsmechanismus steht uns hier der Rahmen $\mathbb{R} \times \mathbb{R}_+$ zur Verfügung, jedoch im Allgemeinen nicht ausschließlich die eigentlich interessierenden Individuen, wie z. B. beim zweiten Fall G_0 . (2) Das uns interessierende Objekt wird als spezielle Teilmenge von $\mathbb{R} \times \mathbb{R}_+$ aufgefasst,

¹Für ein Maß μ auf $(\mathfrak{X}, \mathcal{A})$ und einer messbaren Abbildung $T : (\mathfrak{X}, \mathcal{A}) \rightarrow (\mathfrak{Y}, \mathcal{B})$ bezeichnen wir mit $T \sqsubset \mu := \mu(T^{-1}(\cdot))$ das Bildmaß von μ unter T .

²Hierbei ist ζ das Zählmaß.

³Statt $\zeta(M)$ schreiben wir häufig auch $|M|$.

nämlich als Funktion⁴. (3) Auf der hier endlichen Menge $\mathbf{R}$, welche wir im Folgenden als *Registermenge* bezeichnen, verwenden wir das Zählmaß ζ .

Die bei einer Befragung erhobenen Werte sind jedoch auch oft fehlerbehaftet. Zum Beispiel sind sensible personenbezogene Daten irrtümlichen oder bewussten Falschangaben ausgesetzt. Auch Fehler in der Übertragung oder beim Messen selbst sind zu beachten. Deshalb wird gelegentlich die Betrachtung der zu untersuchenden Population als Realisation einer Verteilung notwendig. Das ist noch ein vierter Aspekt, der in die Definition des Begriffs „Population“ Einzug erhalten soll.

Fassen wir nun diese Vorbemerkungen zusammen, so steht für die Populationsbegriffsbildung allgemein die Forderung im Raum, für beliebige Mengen $\mathbf{R}$ und $\mathbf{Y}$ auf

$$\mathcal{G} := \mathcal{G}(\mathbf{R}, \mathbf{Y}) := \{G \subset \mathbf{R} \times \mathbf{Y} : G \text{ Funktion}\} \quad (1.1)$$

Wahrscheinlichkeitsverteilungen zu betrachten. Sei dazu $\mathcal{F}$ eine σ -Algebra auf $\mathcal{G}$, so definieren wir nun:

1.2 Definition Eine Verteilung γ auf $(\mathcal{G}, \mathcal{F})$ heißt (*randomisierte*) *Population im Register* $\mathbf{R}$ mit *Zustandsraum* $\mathbf{Y}$. Speziell⁵ $\gamma = \delta_G$ für ein $G \in \mathcal{G}$ heißt *deterministisch*, andernfalls nennen wir γ *echt randomisiert*. ♦

Statt Population verwendet man auch die Bezeichnung *Grundgesamtheit*, während eine randomisierte Population in der Literatur häufig auch *Superpopulation* genannt wird. Eine deterministische Population δ_G identifizieren wir mit dem zugehörigen Massepunkt G selbst und nennen dann⁶ $\mathbf{R}_0 := \text{dom}(G)$ *Menge der Merkmalsträger* und G , interpretiert als Abbildung $G : \mathbf{R}_0 \rightarrow \mathbf{Y}$, auch *Merkmal*. Gelegentlich betrachtet man einen Produktraum $\mathbf{X} \times \mathbf{Y}$ als Zustandsraum und nennt dann z. B. $\text{pr}_{\mathbf{X}} \circ G$ *Hilfs-* und $\text{pr}_{\mathbf{Y}} \circ G$ *Untersuchungsvariable*. Man beachte auch, dass der Sonderfall der *leeren Population* entsprechend der Mengendefinition von $\mathcal{G}$ in (1.1) nicht ausgeschlossen ist.

Die Registermenge mit ihrer aus dem Anwendungskontext natürlich gegebenen Struktur stellt also den Rahmen der Merkmalsträger dar. Während in dem Eingangsbeispiel 1.1 lediglich eine endliche Registermenge

⁴Zum Funktionsbegriff vergleiche man Definition A.1 (Seite 85).

⁵Mit δ_a bezeichnen wir das *Dirac-Maß* im Punkt a .

⁶Für eine Funktion $F \subset \mathbf{X} \times \mathbf{Y}$ bezeichnet $\text{dom}(F)$ den Definitionsbereich von F (vgl. Definition A.1, Seite 85).

$\{1, \dots, N\}$ auftrat, deren übliche Wohlordnung ggf. verwendet wird,⁷ ist es häufig auch eine räumliche oder zeitliche Struktur der Merkmalsträger, die mit der Registermenge codiert wird. Zusammen mit den Ausprägungen formt sie letztendlich die zu untersuchende Population.

Gelegentlich interessiert eine weitere sehr wichtige kanonische „Ausstattung“ der Registermenge bzw. der Menge der Merkmalsträger. Sie ist durch ein Maß gegeben. Wir betrachten dazu nur die beiden wesentlichen Fälle $0 < |\mathbb{R}| < \infty$ sowie $\mathbb{R} \in \mathcal{B}(\mathbb{R}^n)$ mit $0 < \lambda^n(\mathbb{R}) < \infty$. Damit definieren wir nun

$$\eta := \eta_{\mathbb{R}} := \begin{cases} \zeta, & \text{falls } 0 < |\mathbb{R}| < \infty \\ \lambda^n|_{\mathbb{R}}, & \text{falls } \mathbb{R} \in \mathcal{B}(\mathbb{R}^n) \text{ und } 0 < \lambda^n(\mathbb{R}) < \infty \end{cases}$$

und nennen η *Registermaß*. Ist η zur Menge der Merkmalsträger $\mathbb{R}_0$ definiert, so verwenden wir dann außerdem die *Gleichverteilung* $\mathbf{U}_{\mathbb{R}_0} := \frac{1}{\eta(\mathbb{R}_0)} \cdot \eta$ und nennen das Bildmaß davon unter einer deterministischen Population G *Merkmalsverteilung*.

Bei dem folgenden Beispiel handelt es sich nun um einen Fall, bei dem wir die Registermenge lediglich im Zusammenhang mit dem zugehörigen Registermaß verwenden.

1.3 Beispiel (Fehlertheorie) Messungen physikalischer Größen sind aufgrund von *unzählbaren* Einflüssen außerhalb unserer Wahrnehmung stets fehlerbehaftet. Jede Messung lässt sich also als ein Element einer Population auffassen, wie sie zum Beispiel durch $G :]a, b[\rightarrow \mathbb{R}$ gegeben ist. Eine solche Population wird dann insofern untersucht, als dass die Verteilung um den wahren Wert der physikalischen Größe $G \sqsubset \eta_{]a, b[}$ interessiert.

Wir können uns dabei auf die Betrachtung des Intervalls $]a, b[:=]0, 1[$ als Registermenge beschränken. Ferner geht man häufig davon aus, dass die Messung durch Elementarfehler verfälscht ist, die sich additiv überlagern, voneinander unabhängig und für den Totalfehler jeweils nahezu gleichbedeutend sind. Somit ist nach dem Zentralen Grenzwertsatz eine solche Messung annähernd normalverteilt. Wir bezeichnen im Folgenden mit $\mathbf{N}(\mu, \sigma^2)$ die *Normalverteilung* mit Erwartungswert $\mu \in \mathbb{R}$ und Varianz $\sigma^2 \in]0, \infty[$.

Beschreibt F_p^{-1} die *Quantilfunktion* einer Verteilung $P \in \text{Prob}(\mathbb{R})$, so ist schließlich die Messung der physikalischen Größe μ als eine „zufällige“

⁷Wir konkretisieren diesen Gedanken in einem einführenden Absatz in Abschnitt 1.2.

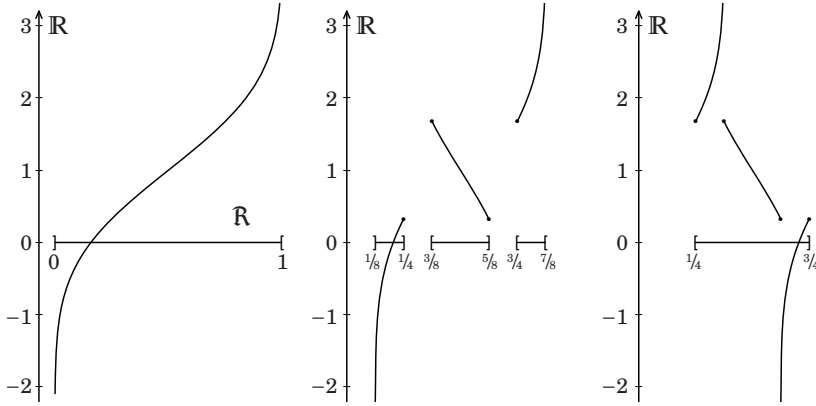


Abbildung 1.1: Drei Populationen, deren Merkmale identisch normalverteilt sind mit $\mu = \sigma = 1$. Links: Population G aus (1.2). Mitte: Population G' aus (1.3). Rechts: Population G'' aus (1.4).

Auswahl eines Elementes aus der (deterministischen) Population

$$G := \left\{ (u, F_{\mathcal{N}(\mu, \sigma^2)}^{-1}(u)) : u \in]0, 1[\right\} \quad (1.2)$$

in der Registermenge $\mathbb{R} =]0, 1[$ mit dem Zustandsraum $\mathfrak{Y} = \mathbb{R}$ formalisierbar. Hier ist jedes Element $r \in]0, 1[$ Merkmalsträger und es gilt $G \sqsubset \lambda_{|]0, 1[} = \mathcal{N}(\mu, \sigma^2)$.

Genauso sind mit $H(u) := F_{\mathcal{N}(\mu, \sigma^2)}^{-1}(2(u - 1/4))$ für $u \in]1/4, 3/4[$ und $\mathbb{R}_0 :=]1/8, 1/4[\cup]3/8, 5/8[\cup]3/4, 7/8[$ auch

$$G'(u) := \begin{cases} H(u + 1/8), & \text{falls } u \in]1/8, 1/4[\\ H(3/8) + H(5/8) - H(u), & \text{falls } u \in]3/8, 5/8[\\ H(u - 1/8), & \text{falls } u \in]3/4, 7/8[\end{cases} \quad (1.3)$$

sowie

$$G''(u) := \begin{cases} G'(u + 1/2), & \text{falls } u \in [1/4, 3/8[\\ G'(u), & \text{falls } u \in [3/8, 5/8] \\ G'(u - 1/2), & \text{falls } u \in]5/8, 3/4[\end{cases} \quad (1.4)$$

Populationen in $\mathcal{G}(]0, 1[, \mathbb{R})$, deren Merkmale identisch normalverteilt sind, d. h.

$$G \sqsubset \mathbf{U}_{|]0, 1[} = G' \sqsubset \mathbf{U}_{\mathbb{R}_0} = G'' \sqsubset \mathbf{U}_{[1/4, 3/4]} = \mathcal{N}(\mu, \sigma^2).$$

Abbildung 1.1 skizziert diese drei Populationen für den Fall $\mu = \sigma = 1$. ♦

Beispiele, bei denen die Registermengenstruktur als eine räumliche interpretiert wird, sind z. B. die bei Waldinventuren formulierten Populationen. Je nach Untersuchungsziel sind diese dann unterschiedlich zu wählen.

1.4 Beispiel (Waldinventur) Als eine interessierende Landfläche, die ein Waldgebiet darstellt, denken wir uns eine Menge der Ebene wie z. B. $\mathbb{R} := [a, b] \times [c, d] \subset \mathbb{R}^2$. Ein gängiges Merkmal bei der forstwirtschaftlichen Bestandsaufnahme ist der *Brusthöhendurchmesser*⁸ (BHD) eines Baumes. Idealisieren wir jeden Baum zu einem Punkt, so sei $\mathbb{R}_0 \subset \mathbb{R}$ genau die Menge aller dieser Punkte, also die Koordinaten der in $\mathbb{R}$ stehenden Bäume. Dabei ist offensichtlich von $|\mathbb{R}_0| < \infty$ auszugehen. Als Waldpopulation betrachten wir dann die Grundgesamtheit

$$G := \{(r, Y(r)) : r \in \mathbb{R}_0\}$$

in $\mathbb{R}$ mit dem Zustandsraum $\mathfrak{Y} =]0, \infty[$, wobei $Y(r)$ für den BHD des Baumes mit dem Koordinatenpaar r stehe. $\blacklozenge$

Gelegentlich interessiert man sich für die räumliche Verteilungsstruktur eines Waldbestandes und interpretiert den Waldbestand selbst als Realisation einer Verteilung. Wir erweitern jetzt das vorangegangene Beispiel, indem wir mit G allgemein ein zufälliges endliches Punktfeld in $\mathbb{R} \times \mathfrak{Y}$ betrachten. Wir haben damit ein Beispiel für eine randomisierte Population.

1.5 Beispiel (Waldinventur) Wieder repräsentiere $\mathbb{R} := [a, b] \times [c, d] \subset \mathbb{R}^2$ das interessierende Waldgebiet, wobei die Bäume nach dem Merkmal BHD unterschieden werden. Bezeichnen wir mit

$$\mathbb{G}_e(\mathbb{R}, \mathfrak{Y}) := \{G \subset \mathbb{R} \times \mathfrak{Y} : |G| < \infty, G \text{ Funktion}\}$$

die Menge aller endlichen Funktionen in $\mathbb{R} \times \mathfrak{Y}$, so betrachten wir nun darauf die σ -Algebra

$$\mathcal{G}_e := \sigma(\{G \in \mathbb{G}_e(\mathbb{R}, \mathfrak{Y}) : G \cap U \neq \emptyset\} : U \text{ offen}).$$

Sei ferner $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, so interpretieren wir eine Zufallsgröße

$$\Psi : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{G}_e(\mathbb{R}, \mathfrak{Y}), \mathcal{G}_e)$$

⁸Der Brusthöhendurchmesser ist der Durchmesser eines Baumstammes in 1,3 Meter Höhe.

als eine randomisierte (Wald-)Population im Register R mit Zustandsraum $\mathfrak{Y} =]0, \infty[$. Solche Zufallsgrößen heißen auch *Punktprozesse*.⁹ ♦

Mit dem Begriff Population haben wir in diesem Abschnitt eines der grundlegendsten Objekte der Stichprobentheorie präzisiert. Mit ihm lässt sich einerseits die Merkmalsverteilung und somit die Zufallsstruktur der interessierenden Erscheinung beschreiben. Andererseits enthält er durch die Registerstruktur Informationen zur Anordnung der Merkmalsträger. Letzteres ist von entscheidender Bedeutung für mögliche und geplante Ziehungsmechanismen bzw. für das Entstehungsmuster der uns bereits vorliegenden, unkontrollierten Auswahl von Populationselementen.

1.2 Stichproben und deren Design

Wir lassen uns nun von der noch etwas unklaren Vorstellung leiten, dass eine Stichprobe stets einen Teil der Population darstellt. Anscheinend unzählige Möglichkeiten liefern Stichproben auf unterschiedlichste Weise. Im Beispiel 1.1 bietet sich mit der n -fachen zufälligen und unabhängigen Auswahl eines Elementes der deterministischen Population G eine sehr elementare Methode an, eine Stichprobe zu erzeugen. Hierbei kann jedoch auch ein und dasselbe Element mehrfach gezogen werden. Eine andere einfache und bekannte Konstruktionsvariante erhält man durch das Ausnutzen der Wohlordnung der Registermenge $\{1, \dots, N\}$. So steht die systematische Befragung jedes k -ten Haushaltes mit zufälligem Startwert $k_0 \leq k$ für eine besonders bequeme Erhebungsmethode. Für die Praxis sind dabei noch gewisse Modifikationen zu berücksichtigen, wie z. B. eine aus Kostengründen gewählte Beschränkung auf eine maximale Anzahl an Stichprobenelementen.

Ein Teil dieser kurzen Einführung erklärt, dass wir nun unter einer *realisierten Stichprobe* aus einer deterministischen Population G eine Funktion in dem Produktraum $G \times N$ verstehen wollen. Die Hinzunahme des Faktors N erhält seine Bedeutung in den Fällen, wo eine wiederholte Auswahl desselben Elementes möglich ist. Sei $\mathfrak{S}_G \subset \mathfrak{G}(G, N)$ eine Menge von (zuzulassenden) realisierten Stichproben, auf der wir eine σ -Algebra $\mathcal{S}_G$ betrachten, so sind doch die $\mathfrak{S}_G$ -wertigen Zufallsgrößen und ihre Verteilungen

⁹Vgl. Abschnitt B.2, Seite 106 ff.

das, was wir uns unter einer (*zufälligen*) *Stichprobenwahl* aus G vorstellen und mit denen wir arbeiten müssen.

Im Folgenden geht es um allgemeinere Begriffsbildungen. Wir definieren hierfür zunächst einen ganz zentralen Begriff der Stichprobentheorie.¹⁰

1.6 Definition Seien $\Delta \subset \mathcal{G}(\mathbb{R}, \mathfrak{Y})$ nichtleer, $\mathcal{F}$ eine σ -Algebra darauf sowie $\mathfrak{S} \subset 2^{\mathbb{R} \times \mathfrak{Y} \times \mathbb{N}}$ mit einer σ -Algebra $\mathcal{S}$ ausgestattet. Ein Markov-Kern q von $(\Delta, \mathcal{F})$ nach $(\mathfrak{S}, \mathcal{S})$ heißt $\mathfrak{S}$ -wertiges *Stichprobendesign* für $(\Delta, \mathcal{F})$, falls für jedes $G \in \Delta$ die Menge¹¹ $\mathfrak{G}(G, \mathbb{N}) \cap \mathfrak{S}$ eine $q(G, \cdot)$ -sichere Menge ist. Die Menge aller solcher Stichprobendesigns bezeichnen wir mit $\text{Samp}(\Delta, \mathcal{F})$. ♦

Mit einem Stichprobendesign wird also die zufällige Konstruktion einer Stichprobe in messbarer Abhängigkeit von der deterministischen Population G beschrieben. Die Messbarkeit bezieht sich dabei auf den Messraum $(\Delta, \mathcal{F})$, den wir im Folgenden *Populationsklasse* nennen.¹² Auf den Messraum $(\mathfrak{S}, \mathcal{S})$ gehen wir später ein.

Für den Moment seien $(\mathfrak{X}, \mathcal{A})$ und $(\mathfrak{Y}, \mathcal{B})$ zwei Messräume, P eine Verteilung auf $\mathfrak{X}$ und K ein Kern von $\mathfrak{X}$ nach $\mathfrak{Y}$, so interpretieren wir im Folgenden die zweite Randverteilung

$$K \sqcap P := \int_{\mathfrak{X}} K(x, \cdot) P(dx)$$

als *Bild* von P unter K . Diese Auffassung, Bezeichnung und Notation rechtfertigt sich dadurch, dass im Fall eines deterministischen Kerns δ_T mit einer messbaren Abbildung T von $\mathfrak{X}$ nach $\mathfrak{Y}$ schließlich $\delta_T \sqcap P = T \sqcap P$ gilt. Wir erhalten somit eine Fortsetzung der üblichen Bildmaßbildung $\cdot \sqcap P$ von der Menge aller messbaren Abbildungen auf die Menge aller Kerne jeweils von $\mathfrak{X}$ nach $\mathfrak{Y}$.

1.7 Definition Seien $(\Delta, \mathcal{F})$ eine Populationsklasse, $\gamma \in \text{Prob}(\Delta, \mathcal{F})$ eine Population und $q \in \text{Samp}(\Delta, \mathcal{F})$, so heißt $q \sqcap \gamma$ *Stichprobenwahl* aus γ unter q . ♦

Statt mit der Verteilung $q \sqcap \gamma$ arbeitet man auch mit Zufallsgrößen $S \sim q \sqcap \gamma$, die wir dann *nichtrealisierte Stichprobe* nennen, während $s \in \mathfrak{S}$

¹⁰Wir verwenden im Folgenden die Bezeichnungen: Ist μ ein Maß auf einem Messraum $(\mathfrak{X}, \mathcal{B})$, so heißt ein $M \subset \mathfrak{X}$ μ -Nullmenge, falls ein $N \in \mathcal{B}$ existiert mit $M \subset N$ und $\mu(N) = 0$. Für eine Verteilung P auf $(\mathfrak{X}, \mathcal{B})$ heißt dementsprechend $M \subset \mathfrak{X}$ P -sichere Menge, falls M^c eine P -Nullmenge ist.

¹¹Wir identifizieren ein Paar $((x, y), z)$ bzw. $(x, (y, z))$ mit dem Tripel (x, y, z) .

¹²Wie auch bei Messräumen allgemein üblich schreiben wir kurz Δ , sofern klar ist, welche σ -Algebra gemeint ist, oder kein Bedarf am Konzept der Messbarkeit besteht.

realisierte Stichprobe heißt.¹³ Den Messraum $(\mathfrak{S}, \mathcal{S})$, auf dem die Stichprobenwahl $q \sqcap \gamma$ lebt, bezeichnet man ferner als *Stichprobenraum*.

Eine Stichprobe $S \sim q \sqcap \gamma$ bzw. deren Stichprobenwahl $q \sqcap \gamma$ heißt *einfach*, falls $S \in \{A \times \{1\} : A \subset \mathbb{R} \times \mathfrak{Y}\}$ fast sicher bzw. $\{A \times \{1\} : A \subset \mathbb{R} \times \mathfrak{Y}\}$ eine $q \sqcap \gamma$ -sichere Menge ist.¹⁴ Außerdem nennen wir ein Stichprobendesign q *einfach*, falls $q(G, \cdot) = q \sqcap \delta_G$ für jedes $G \in \Delta$ einfach ist. Wir sind dann häufig in der Lage, uns auf einen entsprechend einfacheren Stichprobenraum $\mathfrak{S}' \subset 2^{\mathbb{R} \times \mathfrak{Y}}$ mit einer σ -Algebra $\mathcal{S}'$ darauf zu beschränken. Das Stichprobendesign betrachten wir schließlich als ein $\mathfrak{S}'$ -wertiges.

Das primitivste Beispiel für ein Stichprobendesign ist die *Inventur*. Sie ist jedoch eher eine entartete Form der Stichprobenziehung und verdient den Namen eigentlich nicht.

1.8 Beispiel (Vollerhebung) Für jede Populationsklasse Δ ist mit $\mathbf{v}_\Delta(G, \cdot) := \delta_{G \times \{1\}}$ für $G \in \Delta$ ein $\{G \times \{1\} : G \in \Delta\}$ -wertiges Stichprobendesign definiert. Es beschreibt eine vollständige Bestandsaufnahme der Population und heißt daher *Vollerhebung* beziehungsweise *Inventur*. Nach Definition ist $\mathbf{v}_\Delta$ einfach, so dass dieses im Folgenden als Δ -wertiges Stichprobendesign betrachtet wird. ♦

Für Dirac-Verteilungen ist trivialerweise jede Menge messbar, so dass man in dem eben notierten Beispiel genauso auch $(2^{\mathbb{R} \times \mathfrak{Y} \times \mathbb{N}}, 2^{\mathbb{R} \times \mathfrak{Y} \times \mathbb{N}})$ als Stichprobenraum betrachten kann. Um jedoch auch atomfreie Stichprobenwahlen auf einem Stichprobenraum $(\mathfrak{S}, \mathcal{S})$ wählen zu können, wollen wir uns nun auf ein Mengensystem $\mathfrak{S} \subset 2^{\mathbb{R} \times \mathfrak{Y} \times \mathbb{N}}$ zurückziehen, auf dem eine hinreichend starke (topologische) Struktur eingeführt werden kann.

Wir betrachten deshalb, sofern nichts anderes gesagt wird, für den gesamten weiteren Verlauf der Arbeit mit $(\mathbb{R}, \mathcal{T}_\mathbb{R})$ und $(\mathfrak{Y}, \mathcal{T}_\mathfrak{Y})$ lokalkompakte Hausdorff-Räume mit abzählbarer Basis.¹⁵ Dann ist $(\mathbb{R} \times \mathfrak{Y} \times \mathbb{N}, \mathcal{T}_\mathbb{R} \otimes \mathcal{T}_\mathfrak{Y} \times 2^\mathbb{N})$ nach Korollar A.8 (Seite 90) wieder ein LKHA-Raum, welcher als polnischer Raum insbesondere metrisierbar ist (vgl. Lemma A.5, Seite 90). Als eine sowohl wichtige als auch praktische Wahl für einen Stichprobenraum wird sich das System aller kompakten Teilmengen von $\mathbb{R} \times \mathfrak{Y} \times \mathbb{N}$ erweisen, das

¹³Den sich auf eine Stichprobe beziehenden Zusatz „realisiert“ bzw. „nichtrealisiert“ vernachlässigen wir, wenn sich aus dem Kontext ergibt, was gemeint ist.

¹⁴Ist $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathfrak{X}, \mathcal{B})$ eine Zufallsgröße und $M \subset \mathfrak{X}$, so schreiben wir $X \in M$ *P-fast sicher*, falls $\{X \in M\}$ eine P -sichere Menge ist (vgl. Fußnote 10, Seite 15).

¹⁵Wir schreiben für topologische Räume dieser Art das Kürzel: LKHA-Raum.

Beiträge zur Mathematischen Stichprobentheorie
Statistische Modellbildung mit Stichprobendesigns und
anderen Morphismen

Höllwarth, H.

2015, XIX, 126 S. 7 Abb., Softcover

ISBN: 978-3-658-10380-4