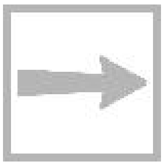
# 1 Ein Modell für den Qualitätstest - Welche Fehler sind möglich?



Das einführende Beispiel von den Knallkörpern schildert einen statistischen Qualitätstest. Anhand dieses praktischen Beispiels erfahren Sie im ersten Kapitel, wie man Tests in der Statistik modelliert und dass Fehler bei solchen Tests unvermeidlich sind. Im Einzelnen werden Sie die Begriffe *Grundgesamtheit*, *Stichprobe und Fehler 1. und 2. Art* kennen lernen.

Der Abschnitt informiert abschliessend ganz kurz über das Wesen der Statistik und die Rolle, welche die Wahrscheinlichkeitsrechnung dabei spielt.

### Urnenmodell

Statistiker und Statistikerinnen ziehen es vor, anstelle von Knallkörpern neutrale (gleich grosse) Kugeln zu betrachten. Die Kugeln liegen in ihrer Vorstellung in einer Kiste, der sog. *Urne* (Fig. 1.1). In unserem Fall enthält die Urne N=100‘000 Kugeln. Den *defekten* Knallkörpern entsprechen *schwarze* Kugeln; ihre (unbekannte) Anzahl nennen wir S.

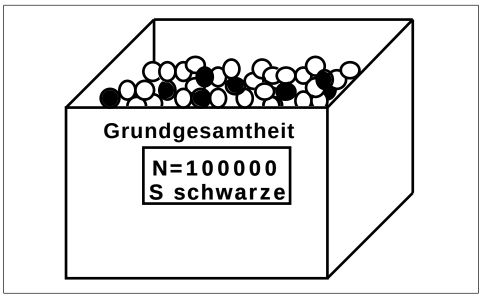



Fig. 1.1: Die Grundgesamtheit im Urnenmodell

Alle anderen Kugeln, ihre Anzahl beträgt N-S, sind *weiss*; sie repräsentieren die *funktionierenden* Knallkörper. Die weissen und scharzen Kugeln insgesamt bilden die sog. *Grundgesamtheit*.

Dieses Beispiel zeigt Ihnen, wie die Statistik die Realität in ein universales, abstraktes und übersichtliches Gedankenmodell, eben das *Urnenmodell*, übersetzt.

### Stichprobe

Dem zufälligen Auswählen von 200 Knallkörpern in der Realität entspricht im Urnenmodell etwa der folgende Vorgang: Wir mischen die Kugeln kräftig, ziehen mit verbundenen Augen eine Kugel und legen sie beiseite. Dieses Prozedere wiederholen wir 199-mal.

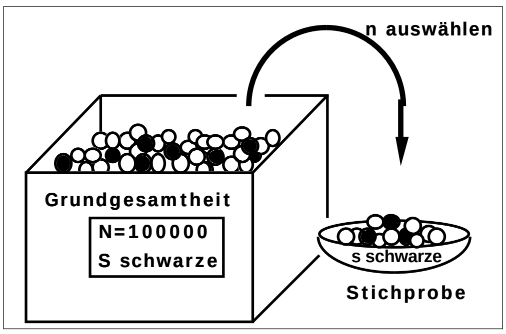



Fig. 1.2: Eine Stichprobe aus der Grundgesamtheit im Urnenmodell.

Wir haben so eine sog. *Stichprobe* erhalten (Fig. 1.2). Die Zahl n=200 nennen wir den *Umfang* der Stichprobe.

Die nachstehende Figur 1.3 zeigt drei mögliche Stichproben vom Umfang n=200 und s=1, s=5 und s=10 schwarzen Kugeln.

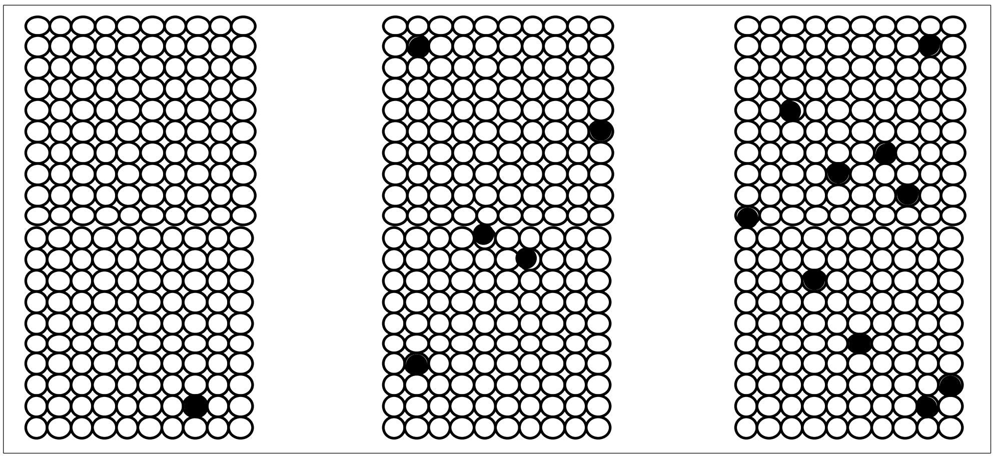
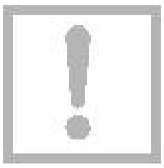


Fig. 1.3: Drei mögliche Stichproben mit dem Umfang 200.

***Aufgabe 1.1***

*Mit seinem Qualitätstest möchte unser Einkäufer – in der Sprache des Urnenmodells formuliert – herausfinden, ob die Urne (die Grundgesamtheit) nicht mehr als 4000 schwarze Kugeln enthält. Weshalb zählt er nicht einfach in der Urne nach?*

### *Die Grundidee des Tests*



Der Einkäufer hofft, dass er sich mit seiner Stichprobe ein kleines Spiegelbild der Grundgesamtheit verschafft hat. Wenn die Grundgesamtheit gerade noch den Qualitätsanforderungen entspricht, so enthält sie, wie gesagt, genau S=4000 schwarze Kugeln. Der Einkäufer müsste dann mit s=8 schwarzen Kugeln in der Stichprobe rechnen, aber selbstverständlich können es mehr oder weniger sein, in den Extremfällen sogar 0 oder 200.

Gefühlsmässig ist uns jedoch klar: *Grosse* Abweichungen von 8 sind *unwahrscheinlich*. Und unser Einkäufer will in *diesem* Sinne entscheiden! Wenn die Stichprobe mehr als 13 schwarze Kugeln enthält, glaubt er nicht mehr daran, dass die Grundgesamtheit nur 4000 schwarze Kugeln enthält. Deshalb lehnt er die Sendung ab.

Hand aufs Herz! Sehen Sie eine grundsätzlich andere, realistische Möglichkeit, die Qualität von Knallkörpern zu überprüfen? Aber natürlich ist dieses bestechend einfache Verfahren mit gewissen Risiken behaftet. Diese müssen wir untersuchen und irgendwie messen.

### *Der Fehler 1. Art*

Beim geschilderten Entscheidungsschema nimmt der Einkäufer (und vor allem der Lieferant!) in Kauf, dass die Sendung *irrtümlicherweise* abgelehnt wird, weil 14 oder mehr schwarze Kugeln in die Stichprobe geraten, *obwohl* *die Sendung in Ordnung* ist (nicht mehr als 4000 schwarze Kugeln enthält). Man bezeichnet diesen Fehlentscheid als Fehler 1. Art.

Es hängt vom Zufall ab, ob es zu einem solchen Fehler kommt. Deshalb kann man einem Fehler 1. Art eine Wahrscheinlichkeit zuordnen. Diese hängt natürlich von der *tatsächlichen* Anzahl schwarzer Kugeln in der Urne ab. Sie ist am grössten, wenn die maximale Anzahl von 4000 schwarzen Kugeln vorhanden ist. Der Maximalwert dieser Wahrscheinlichkeit heisst traditionsgemäss $\alpha$.

Zusammenfassend können wir sagen: Wenn die Sendung in Ordnung ist, sind höchstens 4000 schwarze Kugeln in der Urne und es ist

$$P(\text{ Fehler 1. Art }) \leq \alpha.$$

In den nächsten Kapiteln werden wir uns näher mit der Bestimmung von $\alpha$ befassen.

### *Der Fehler 2. Art*

Der Fehler 1. Art trifft den Lieferanten. Bei unserem Qualitätstest geht aber auch der Einkäufer ein (beachtliches) Risiko ein: Nehmen wir an, die Urne enthalte weit mehr als 4000, z. B. 8000 schwarze Kugeln. Trotzdem könnte es in der Stichprobe nicht mehr als 13 schwarze Kugeln haben. Der Einkäufer würde in diesem Fall also nicht realisieren, dass die Sendung *nicht* in Ordnung ist und diese akzeptieren. Man spricht hier vom Fehler 2. Art. Wieder hängt es vom Zufall ab, ob dieser Fehler gemacht wird, und man kann diesem eine Wahrscheinlichkeit zuordnen. Diese wird üblicherweise mit $\beta$ bezeichnet.

Die Wahrscheinlichkeit

$$\beta = P(\text{Fehler 2. Art})$$

ist aber erst festgelegt, wenn wir über die Anzahl schwarzer Kugeln eine Annahme machen. Diese Annahme muss nicht notwendigerweise "$\geq 4000$ schwarze Kugeln", sondern kann z. B. "8000 schwarze Kugeln" sein.

**_Fehler 1. und 2. Art (Zusammenfassung)_**

Bei unserem Qualitätstest sind zwei mögliche *Fehlentscheidungen* grundsätzlich unvermeidlich:

• Wir finden mehr als 13 defekte Knallkörper in der Stichprobe, obwohl die Sendung in Ordnung ist (Ausschussanteil $\leq$ 4%). Dann *lehnen wir sie zu Unrecht ab (Fehler 1. Art) und zwar mit der Wahrscheinlichkeit* $\alpha$.

• Wir finden höchstens 13 defekte Knallkörper in der Stichprobe, obwohl die Sendung *nicht* in Ordnung ist (mehr als 4% Ausschuss). Wir *akzeptieren sie zu Unrecht (Fehler 2. Art) und zwar mit der Wahrscheinlichkeit* $\beta$.

Die Darstellung der Situation im Baumdiagramm ist instruktiv (Fig. 1.4):

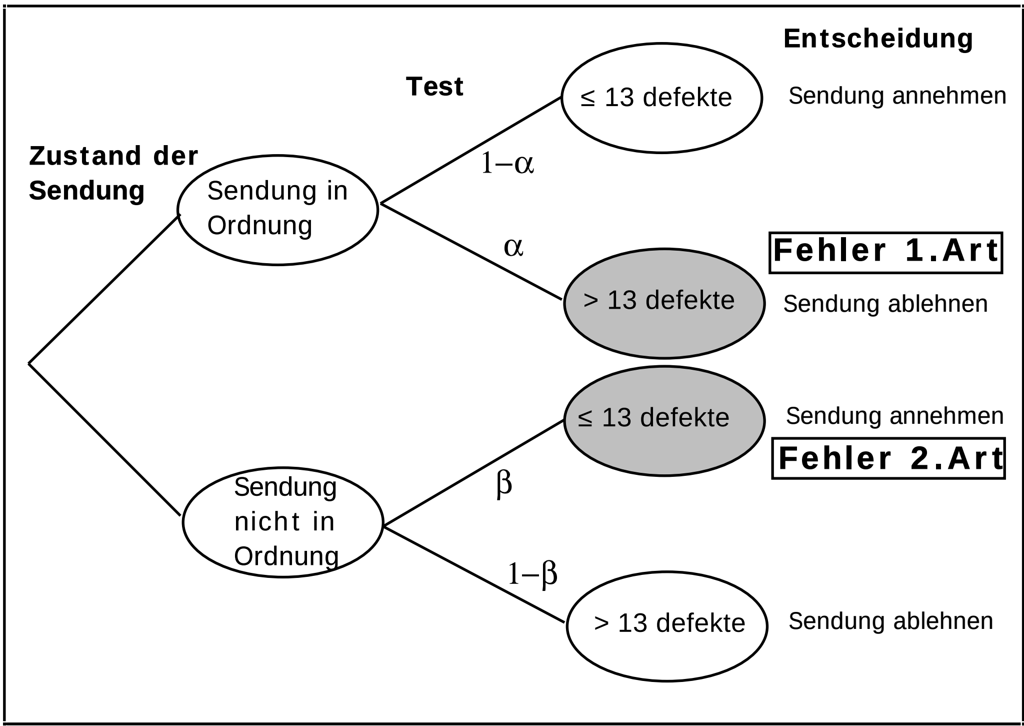



Fig. 1.4: Die Fehler 1. Art und 2. Art und ihre Wahrscheinlichkeiten α und β im Baumdiagramm

### *Beurteilende Statistik*

Der Qualitätstest ist ein typisches Beispiel aus der sog. *beurteilenden* Statistik: Eine Grundgesamtheit soll aufgrund einer Stichprobe, d. h. aufgrund einer *zufälligen* Auswahl, beurteilt werden. In unserem Fall soll von der (relativen) Häufigkeit der schwarzen Kugeln in der Stichprobe auf diejenige in der Grundgesamtheit geschlossen werden (Fig. 1.5).

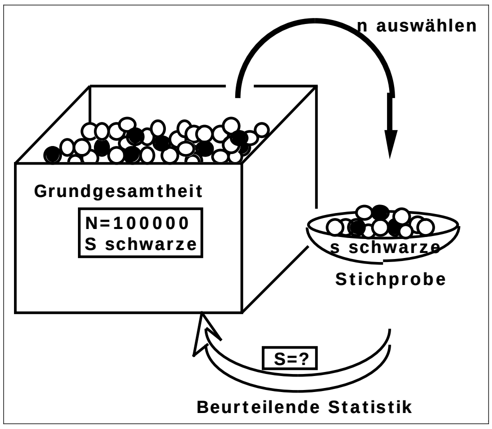



Fig. 1.5: Die beurteilende Statistik im Urnenmodell

Es ist unvermeidlich, dass der Schluss mit einer gewissen Wahrscheinlichkeit falsch ist. Wir werden diese Wahrscheinlichkeit in den nächsten Kapiteln auf zwei ganz verschiedene Arten ermitteln: 1. Durch Simulation des Qualitätstests auf dem Computer und 2. mit Hilfe der Wahrscheinlichkeitsrechnung.

### *Zusatzinformation*:

Intuitiv ist Ihnen sicher klar, dass ein Qualitätstest um so zuverlässiger ist, je grösser der Umfang n der Stichprobe ist. Aber je grösser n ist, um so grösser sind die Testkosten. Eine weitere Aufgabe der Statistik ist es, n so klein als möglich und so gross wie nötig festzulegen. Dieses wichtige und interessante Problem steht aber nicht im Zentrum dieses Leitprogramms.

***Haben Sie die Lernziele erreicht?***

Sie sind bereits am Ende des ersten Kapitels angelangt. In der folgenden *Lernkontrolle* geht es um eine neue statistische Fragestellung. Sie sollen versuchen, hier die Bedeutung der eben eingeführten Begriffe und Variablen zu erkennen.

Wenn Ihnen die Aufgabe Schwierigkeiten bereitet, ist es angezeigt, den vorangehenden Abschnitt nochmals durchzuarbeiten. Schlagen Sie nicht zu früh im Lösungsteil auf der nächsten Seite nach. Wenn Sie aber das Gefühl haben, dass Sie die Sache beherrschen, so melden Sie sich gleich anschliessend bei Ihrer Lehrperson zum Kapiteltest.