

## 2

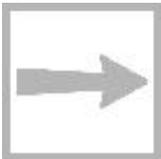
**Computersimulation des Qualitätstests**

In diesem Kapitel erreichen wir ein erstes entscheidendes Ziel: Wir ermitteln *näherungsweise* die *Wahrscheinlichkeiten*  $\alpha$  und  $\beta$  für die Fehler 1. und 2. Art – und zwar ohne eigentliche Rechnung!

Wie ist das möglich? Sie erinnern sich, dass die Wahrscheinlichkeit eines zufälligen Ereignisses die *idealisierte relative Häufigkeit* ist, mit welcher dieses vorkommt, wenn man das zugehörige Zufallsexperiment sehr oft wiederholt. Wenn wir beispielsweise zwei ideale Würfel sehr oft gleichzeitig werfen, so werden wir durchschnittlich unter je 18 Würfeln einmal die Summe 3 erhalten. Deshalb legen wir fest:

$$P(\text{„Summe 3“}) = \frac{1}{18} = 0.0555 = 5.55\%.$$

Mit dieser Idee können wir näherungsweise auch die Wahrscheinlichkeit für den Fehler 1. Art bestimmen. Wir betrachten eine Urne mit 100'000 Kugeln, von denen exakt 4000 schwarz sind. Nun simulieren wir den Qualitätstest hundertmal, indem wir hundertmal je 200 Kugeln aus der vollen Urne auswählen. Jedesmal prüfen wir nach, ob die Stichprobe mehr als 13 schwarze enthält, was ja einem Fehler 1. Art entsprechen würde. Wenn dies in z. B. vier der hundert Ziehungen vorkommt, hat der Fehler 1. Art die relative Häufigkeit 4% und wir würden 4% schätzen.



Bitte erschrecken Sie nicht vor dieser Aufgabe! Selbstverständlich werden wir nicht 100 mal unter 100'000 *realen* Kugeln 200 auswählen und die schwarzen zählen, sondern wir simulieren diesen Vorgang praktisch gleichwertig auf dem Computer.

**Der Stellenwert der Simulation**

Denken Sie nicht, dass es sich bei dieser spielerischen Methode um ein unseriöses oder unwissenschaftliches Verfahren handelt. Im Gegenteil: Immer mehr Statistikanwender entschliessen sich, ihr statistisches Problem nicht mit Hilfe einer abstrakten (Wahrscheinlichkeits-) Rechnung zu durchleuchten, sondern mittels der viel anschaulicheren Simulation. Ermöglicht wird dies durch die immensen Speicher- und Rechenkapazitäten der allgegenwärtigen Computer. Und man kann theoretisch zeigen (sog. Gesetz der grossen Zahl), dass die simulierten Resultate verlässlich sind, ja, man kann sogar die Grösse des Fehlers abschätzen (zentraler Grenzwertsatz). Bei komplizierten Problemen ist die Simulationsmethode manchmal sogar die einzige Möglichkeit, ein Resultat zu erhalten.

**Aufgabe 2.0.1**

*Eine Repetitionsaufgabe, die Sie wahrscheinlich früher schon einmal gelöst haben: Mit welcher Wahrscheinlichkeit würfelt man mit zwei idealen Würfeln die Summe 7?*



### Aufgabe 2.0.2

Wie gross schätzen Sie bei unserem Knallkörper–Qualitätstest  $\alpha$ ?

 1‰

 1%

 2%

 5%

## 2.1 Simulation mit Ikosaedern

Wir wollen, wie gesagt,  $\alpha$  durch Simulation ermitteln. Damit wir die Simulation auf dem Computer besser verstehen, führen wir zuvor eine Simulation „von Hand“ durch. Um die Übersicht zu behalten, beschränken wir uns dabei auf 10-er Stichproben.

Als erstes wollen wir erreichen, dass jede einzelne Kugel identifizierbar und unterscheidbar von den anderen ist. Dazu denken wir uns die 100'000 Kugeln numeriert mit den Nummern

0, 1, 2, ..., 99998, 99999.

Am übersichtlichsten wird es, wenn wir den 4000 schwarzen Kugeln gerade die Nummern 0, 1, 2, ..., 3998, 3999 zuordnen.

### Zufällig eine Kugel ziehen

Zufällig eine Kugel wählen ist nun gleichbedeutend mit zufällig eine der obgenannten Nummern würfeln. Am einfachsten geht das mit dem Ikosaeder, dem regulären Zwanzigflächner (Fig. 2.1), aus welchem wir wie folgt einen „Zehnerwürfel“ konstruieren: Wir malen auf ein erstes beliebig gewähltes gleichseitiges Dreieck eine 0 und gleich anschliessend eine 0 auch auf das diametral gegenüberliegende Dreieck. In analoger Weise beschriften wir je Paare gegenüberliegender Seitenflächen mit den Ziffern 1, 2, ..., 9 (man kann solche Würfel sogar fixfertig in Spielwarengeschäften kaufen).

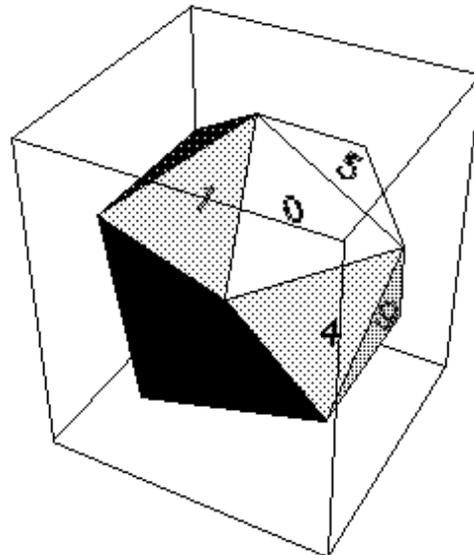


Fig. 2.1: Das mit den Ziffern 0, 1, ..., 9 beschriftete Ikosaeder.

Bei jedem Wurf eines solchen idealen Ikosaeders erscheint mit gleicher Wahrscheinlichkeit eine der zehn Ziffern auf der zuoberst liegenden Seitenfläche. Werfen wir den Ikosaeder fünfmal nacheinander und notieren die gewürfelten Ziffern hintereinander, so erhalten wir gerade eine der obgenannten 100'000 Zahlen, z. B. 01033.

Die Kugel mit der Nummer 1033 wurde gewürfelt; sie ist zufällig gerade schwarz.

### **Zufällig eine Stichprobe ziehen**

Diese umständliche Prozedur wurde 10-mal durchgeführt und so ergab sich die folgende 10-er Stichprobe:

1033 - 45726 - 93952 - 14152 - 3070 - 64939 - 81283 - 92019 - 30218 - 90733

Bei jeder Nummer wurde überprüft, ob sie nicht vorher schon einmal vorgekommen ist. Man hätte sie allenfalls einfach weggelassen.

Nur nebenbei: Die zehn gewürfelten Zahlen sind mit der Wahrscheinlichkeit 0.99945 verschieden, d. h. die gleiche Nummer wird sehr selten mehr als einmal gewürfelt.

### **Das 0-1-Protokoll**

Unsere Stichprobe enthält zwei schwarze Kugeln, nämlich die Nummern 1033 und 3070. Da uns eigentlich jeweils nur interessiert, ob die gezogene Kugel schwarz oder weiss ist, genügt es, das sog. 0-1-Protokoll der Stichprobe zu notieren:

1	0	0	0	1	0	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---	---

Dabei haben wir schwarze Kugeln durch eine 1, weisse durch eine 0 gekennzeichnet. Diese Darstellung hat den Vorteil, dass wir die Anzahl der schwarzen Kugeln einfach als Summe aller 10 gewürfelten 0-er und 1-er erhalten.

### **25 Simulationen einer 10-er Stichprobe**

In der folgenden Tabelle sind 25 simulierte 10-er Stichproben zusammengestellt und in der letzten Spalte wurde je die Anzahl schwarzer Kugeln notiert.

	1	2	3	4	5	6	7	8	9	10	Anzahl schwarze
Simulation 1	0	0	0	0	0	0	0	1	0	0	1
Simulation 2	0	0	0	0	0	0	1	0	0	0	1
Simulation 3	0	0	0	0	1	0	0	0	0	0	1
Simulation 4	0	0	0	0	0	0	0	0	0	0	0
Simulation 5	0	0	0	0	0	0	0	0	0	0	0
Simulation 6	0	0	0	0	0	0	0	0	0	0	0
Simulation 7	0	0	0	0	0	0	0	0	0	0	0
Simulation 8	0	0	0	0	0	0	0	0	0	0	0
Simulation 9	0	0	0	0	0	0	0	1	0	0	1
Simulation 10	0	0	0	0	0	0	0	0	0	0	0
Simulation 11	0	0	1	0	0	0	0	0	0	0	1
Simulation 12	1	0	0	0	0	0	0	0	0	0	1
Simulation 13	0	0	0	0	0	0	0	0	0	0	0

Simulation 14	0	1	0	0	0	0	0	0	0	0	1
Simulation 15	0	0	0	0	0	0	0	0	0	0	0
Simulation 16	0	0	0	0	0	0	1	1	0	0	2
Simulation 17	0	0	0	0	0	0	0	0	0	0	0
Simulation 18	0	0	0	0	0	0	0	0	0	0	0
Simulation 19	0	0	0	0	0	0	0	0	0	0	0
Simulation 20	1	0	1	0	0	0	0	0	0	0	2
Simulation 21	0	0	0	0	0	0	0	0	0	0	0
Simulation 22	0	0	0	0	0	0	0	0	0	0	0
Simulation 23	0	0	0	1	0	0	0	0	0	0	1
Simulation 24	0	0	0	0	0	0	0	0	0	1	1
Simulation 25	0	0	0	0	0	0	0	0	0	0	0

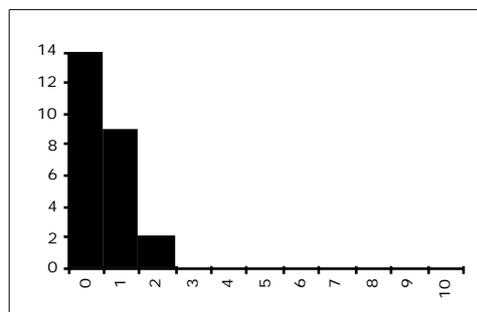
### Absolute Häufigkeiten

Theoretisch kann eine 10-er Stichprobe 0 bis 10 schwarze Kugeln enthalten. In den obigen 25 Simulationen sind aber nur die Werte 0, 1 und 2 vorgekommen und zwar mit den in der nachstehenden Tabelle zusammengestellten absoluten Häufigkeiten:

Anzahl schwarze Kugeln	0	1	2	3	4	5	6	7	8	9	10
Anzahl Stichproben	14	9	2	-	-	-	-	-	-	-	-

### Das Histogramm

Die nebenstehende Figur 2.2 zeigt das Histogramm (Säulendiagramm) dieser Häufigkeitstabelle. Auf der waagrechten Achse ist die mögliche Anzahl der schwarzen Kugeln aufgetragen und auf der senkrechten die zugehörigen absoluten Häufigkeiten.



Das Histogramm ist anschaulicher als die Häufigkeitstabelle. Wir sehen z. B. auf einen Blick, dass in den meisten Fällen keine oder genau eine schwarze Kugel gezogen wurde.

Fig. 2.2: Histogramm der absoluten Häufigkeiten von Stichproben mit 0, 1, ..., 10 schwarzen Kugeln.

### Relative Häufigkeiten

Aus den absoluten können wir die *relativen* Häufigkeiten berechnen. Das Ereignis „Keine schwarze Kugel“ hat z. B. die relative Häufigkeit:

$$\frac{\text{absolute Häufigkeit}}{\text{Anzahl Simulationen}} = \frac{14}{25} = 0.56 = 56\%.$$

Analog ergeben sich die relativen Häufigkeiten der anderen Ereignisse.



### Aufgabe 2.1.1

Berechnen Sie die restlichen relativen Häufigkeiten und ergänzen Sie die nachfolgende Tabelle.

Anzahl defekte	0	1	2	3	4	5	6	7	8	9	10
Relative Häufigkeit (%)	56										

### Wahrscheinlichkeiten

Gute und zuverlässige Schätzwerte für die Wahrscheinlichkeiten erhalten wir, wenn wir das zugehörige Zufallsexperiment sehr oft wiederholen. Obwohl 25 in der Statistik eine eher kleine Zahl ist, wollen wir im vorliegenden Fall doch annehmen, dass das Ereignis

„Keine schwarze Kugel in einer Stichprobe von 10 Kugeln“

ungefähr die Wahrscheinlichkeit 56% hat:

$$P(\text{„Keine schwarze Kugel“}) = 56\%$$

Analoges gilt für die Ereignisse „Genau 1 schwarze“, „Genau 2 schwarze“ etc.



### Aufgabe 2.1.2

Stellen Sie sich vor, dass sich Einkäufer und Lieferant den folgenden Testplan vereinbart hätten:

Es werden zufällig 10 Knallkörper ausgewählt und gezündet. Wenn diese Stichprobe mehr als einen defekten Knallkörper enthält, wird die Sendung nicht akzeptiert.

- Worin besteht der Fehler 1. Art bei diesem Test?
- Wie oft wird der Fehler 1. Art in den obigen 25 Simulationen gemacht?
- Wie gross ist etwa die Wahrscheinlichkeit  $\alpha$  für den Fehler 1. Art bei diesem einfachen Qualitätstest?

## 2.2

### Die Fehler 1. und 2. Art simuliert

Sie haben im letzten Abschnitt erfahren, wie *im Prinzip* die Wahrscheinlichkeit für die Fehler 1. Art über Simulation ermittelt werden kann. Bei einer Computersimulation verwendet man anstelle des Ikosaeders den Zufalls-generator eines Computers. Das ist ein Programm, das Zahlen erzeugt, die man von "echten" Zufallszahlen praktisch nicht unterscheiden kann. In diesem Abschnitt präsentieren wir Ihnen nun in Form von EXCEL-Tabellen und Histogrammen die Auswertung von hundert Computersimulationen einer 200-er Stichprobe. Es ist Ihnen wohl klar, dass jede neue Serie von Simulationen leicht verschiedene Resultate liefert.

### Häufigkeitstabelle für den Fehler 1. Art

Hier rechnet die Simulation mit genau 4000 schwarzen unter den 100'000 Kugeln in der Urne. Die nachstehende Tabelle ist wie diejenige auf Seite 2.4 aufgebaut: In der 2. Zeile stehen die absoluten Häufigkeiten von Stichproben mit  $s$  schwarzen Kugeln ( $s=0, 1, \dots, 20$ ; diese Werte stehen in der 1. Zeile). Beispielsweise enthielten 18 der 100 simulierten Stichproben 8 schwarze Kugeln.

		Tabelle 3																				
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	Y
1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
2	0	0	1	0	6	8	18	14	18	6	8	10	2	6	2	1	0	0	0	0	0	
3																						
4																						
5																						

Fig. 2.2: Häufigkeitstabelle für die Anzahl schwarzer Kugeln in einer 200-er Stichprobe.



#### Aufgabe 2.2.1

Wie viele der 100 Stichproben enthielten mehr als 10 schwarze Kugeln?



#### Wahrscheinlichkeit des Fehlers 1. Art

Der Fehler 1. Art ergibt sich jedesmal, wenn eine Stichprobe mehr als 13 schwarze Kugeln enthält. Im ganzen gibt es gemäss Häufigkeitstabelle nur 3 Stichproben mit mehr als 13 schwarzen Kugeln, nämlich zwei mit 14 und eine mit 15. Der Fehler 1. Art hat bei dieser Simulation also die relative Häufigkeit von 3%. Wir dürfen diesen Wert als Schätzwert für die Wahrscheinlichkeit des Fehlers 1. Art betrachten:

3%

#### Was bedeutet $\alpha = 3\%$ für die Praxis?

Für den Qualitätstest bei den Knallkörpern bedeutet 3%: Ein Einkäufer, der diesen oder einen ähnlichen Einkauf häufig macht, wird in durchschnittlich 3 von 100 Einkäufen die Knallkörpersendung irrtümlicherweise ablehnen!

#### Der Fehler 1. Art im Histogramm

Im nebenstehenden Histogramm der Häufigkeitstabelle (Fig. 2.3) sind die Stichproben, die zu einem Fehler 1. Art führen, durch punktierte Säulen unterschieden.

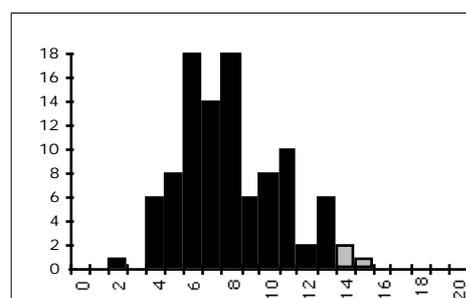


Fig. 2.3: Histogramm der Häufigkeit schwarzer Kugeln in Hundert 200-er Stichproben.

Damit haben wir auch schon das Ende dieses Kapitels erreicht. Wenn Sie die folgende Lernkontrolle zufriedenstellend gelöst haben, sind Sie reif für den Kapiteltest. Zunächst jedoch noch ein Hinweis:

### ***Die Technik der Computersimulation***

Wenn Sie erfahren möchten, wie eine Computersimulation funktioniert und wie man sie programmiert, müssen Sie im Anhang S "Simulationen" das Kap. I studieren. Ihre Lehrperson hat Sie orientiert, ob und wie weit Sie sich in diese technische Seite des Problems einarbeiten sollen.