Das Problem der Alliierten mit deutschen Panzern oder: Wie schätzt man einen Populationsumfang?

NORBERT HENZE, KARLSRUHE, UND REIMUND VEHLING, HANNOVER

Zusammenfassung: *In diesem Aufsatz geht es um ein spannendes Problem, dem sich die Alliierten während des Zweiten Weltkriegs gegenübersahen: Wie schätzt man zuverlässig die Anzahl der in Nazi-Deutschland produzierten Panzer? Im Vergleich zu Geheimdienstinformationen, die diese Anzahl um den groben Faktor 6 überschätzt hatten, erwies sich ein statistisches Schätzverfahren als sehr genau. Wir zeigen, was man aus dieser nicht nur historisch bedeutsamen Fragestellung im Hinblick auf einen an echten Anwendungen orientierten Stochastikunterricht lernen kann.*

1 Einleitung

Wie viele Panzer hatte und produzierte die deutsche Wehrmacht? Dieser auch als *German Tank Problem* bezeichneten Fragestellung sahen sich die Westalliierten während des Zweiten Weltkriegs gegenüber. Die Kenntnis über die Stärke der deutschen Streitkräfte war unter anderem wichtig, um die als *Operation Overlord* bekannte und mit dem sog. *D-Day* beginnende Landung in der Normandie vorzubereiten. Das *German Tank Problem* war Gegenstand diverser Pressemitteilungen, s. z.B. Bischoff (2022), Dambeck (2010) oder Davies (2006). Im Hinblick auf eine gute Schätzung kam den Alliierten zugute, dass für verschiedene Panzerbauteile wie etwa Fahrgestell, Getriebe, Motor oder Räder fortlaufend aufsteigende Seriennummern vergeben wurden, s. z.B. Ruggles und Brodie (1947).

	Statist. Schätzung	Geheimdienst-Schätzung	deutsche Aufzeichn.
06. 1940	169	1000	122
06. 1941	244	1550	271
08. 1942	327	1550	342

Tab. 1: Anzahlen deutscher Panzer und Schätzungen

Die Ruggles und Brodie (1947) entnommenen Daten von Tab. 1 über monatliche Produktionszahlen zeigen eindrucksvoll, dass die statistischen Schätzungen wesentlich näher an der nach Ende des Krieges dokumentierten Realität lagen als die Schätzungen der involvierten Geheimdienste.

Die in diesem Aufsatz vorgestellte Fragestellung ist natürlich nicht auf Panzer beschränkt, sondern immer dann gegeben, wenn eine große Menge von nummerierten Subjekten oder Objekten vorliegt, wie etwa Marathonläufer oder gewisse elektronische Geräte, und die Gesamtzahl der Subjekte oder Objekte aufgrund einer Stichprobe geschätzt werden soll. So wurde etwa die Gesamtanzahl aller jemals verkauften Commodore-64-Computer anhand ihrer Seriennummern auf ca. 12,5 Millionen taxiert, s. Bischoff (2022).

Das Hauptziel dieses Aufsatzes besteht darin, anhand eines konkreten Szenarios eine Einführung in Prinzipien der Parameterschätzung zu geben und dabei zentrale Begriffe wie *Maximum Likelihood*, *Erwartungstreue (Unverzerrtheit)* und *Minimalvarianz* zu erläutern. Dabei werden mögliche Wege einer Umsetzung dieses Problems im Unterricht aufgezeigt. Der Aufsatz ist so aufgebaut, dass wir zunächst den mathematischen Kern herausschälen. Anschließend geht es um Simulationen, das Sammeln von Erfahrung und heuristische Überlegungen. Danach stellen wir das Maximum-Likelihood-Schätzprinzip vor und leiten das von den Alliierten verwendete und zu den statistischen Schätzungen in Tab. 1 führende Schätzverfahren in dessen einfachster Form her.

2 Mathematischer Kern

Der mathematische Kern des *German Tank Problems* lautet in seiner einfachsten Form: Es gibt eine unbekannte große Anzahl N von Objekten, die von 1 bis N durchnummeriert und mit diesen Zahlen identifiziert seien. Um N schätzen zu können, liegen k verschiedene der Zahlen $1,\ldots,N$ als Stichprobe vor. Ganz konkret könnten wir gefragt werden, einen Schätzwert für N abzugeben, wenn eine Stichprobe vom Umfang $k = 4$ die der Größe nach sortierten Werte 31, 43, 66 und 122 ergeben hat. Natürlich ist daraufhin N mindestens gleich 122, aber ganz allgemein wird der größte Wert in einer Stichprobe den tatsächlichen Wert von N vermutlich mehr oder weniger stark unterschätzen. Selbstverständlich möchten wir aber N nicht systematisch unterschätzen, sondern ein Schätzverfahren haben, dass bei oftmaliger Anwendung zumindest auf die Dauer im Mittel richtig schätzt. Wir werden unter anderem präzisieren, was ein Schätz*verfahren* überhaupt ist, und was es mathematisch bedeutet, „auf die Dauer im Mittel richtig zu schätzen“.

Was hat dieses Problem mit Stochastik zu tun? Stochastik und damit Wahrscheinlichkeiten kommen dadurch ins Spiel, dass die mit s bezeichnete Stichprobe vom Umfang k als *Realisierung einer rein zufälligen k-Auswahl ohne Zurücklegen* aus der Menge $\{1,\ldots,N\}$ angesehen wird. In der Stichprobentheorie spricht man dann auch von einer *einfachen Stichprobe* vom Umfang k. Mit *rein zufällig* ist gemeint, dass jede der insgesamt $\binom{N}{k}$ möglichen k-elementigen Teilmengen von $\{1,\ldots,N\}$ die gleiche Wahrscheinlichkeit besitzt, als Stichprobe aufzutreten. Wir unterstellen also im Folgenden ein Laplace-Modell auf der Menge

$$\Omega := \big\{s : s \subset \{1,\ldots,N\} \text{ und } |s| = k\big\}$$

aller k-elementigen Teilmengen der Menge $\{1,\ldots,N\}$. Jede solche Menge s aus Ω besitzt somit die Wahrscheinlichkeit $1/\binom{N}{k}$. Es empfiehlt sich, an dieser Stelle eine Zufallsgröße einzuführen, deren mögliche Realisierungen die Stichproben s sind. Eine *reelle* Zufallsgröße ist eine Abbildung, die den *Ergebnisraum* oder *Grundraum* genannten möglichen Resultaten eines stochastischen Vorgangs *reelle Zahlen* zuordnet. In unserem Fall liefert der stochastische Vorgang, den wir uns gedanklich als blindes Ziehen von k verschiedenen der Zahlen $1,\ldots,N$ und damit als Ermittlung der Gewinnzahlen bei einem „k-aus-N-Lotto“ vorstellen können, eine Stichprobe s und damit ein Element von Ω. Die mit S bezeichnete Zufallsgröße ordnet jedem s aus Ω die Menge s selbst zu; formal gilt somit $S(s) := s$ für jedes s aus Ω. Die Realisierungen dieser Zufallsgröße sind also keine reellen Zahlen, sondern *Mengen* (genauer: k-elementige Teilmengen von Ω). Hiermit gilt

$$\mathbb{P}(S = s) = \frac{1}{\binom{N}{k}} \quad \text{für jedes } s \in \Omega.$$

Wenn man der Zufallsgröße S einen Namen geben sollte, würde man sie *rein zufällige Stichprobe* (vom Umfang k aus Ω) nennen, denn wir haben ja schon früher geschrieben, dass s die *Realisierung* einer rein zufälligen Stichprobe sei. Grundsätzlich ist es wichtig, konzeptionell zwischen Zufallsgrößen und deren Realisierungen zu unterscheiden. Diesbezüglich hat es sich auch eingebürgert, Zufallsgrößen mit Großbuchstaben und deren Realisierungen mit den entsprechenden Kleinbuchstaben zu bezeichnen.

Im Folgenden wird es nötig sein, die k Zahlenwerte einer Stichprobe s der Größe nach zu sortieren. Grundsätzlich können die Elemente einer Menge in beliebiger Reihenfolge aufgelistet werden. Für das obige Zahlenbeispiel mit $k = 4$ gilt also etwa $s = \{66,122,31,43\} = \{43,66,31,122\}$. Beide Mengen sind gleich, weil jede Zahl in der ersten Menge in der zweiten Menge auftritt und umgekehrt. Wir bezeichnen ab jetzt für jedes j von 1 bis k mit x_j die *j-kleinste* Zahl in der Menge s. Im obigen Beispiel gelten also $x_1 = 31$, $x_2 = 43$, $x_3 = 66$ und $x_4 = 122$. Allgemein sind bei einer Stichprobe s vom Umfang k die Elemente der Menge s definitionsgemäß in der Form $x_1 < \ldots < x_k$ angeordnet. Dabei müssten wir eigentlich $x_1(s) < \ldots < x_k(s)$ schreiben, denn die Zahlen $x_1,\ldots,x_k$ hängen ja von s ab. In der Deutung eines *k-aus-N-Lottos* sind $x_1,\ldots,x_k$ die Realisierungen der aufsteigend geordneten Gewinnzahlen.

An dieser Stelle sei gesagt, dass das *German Tank Problem* eng mit dem sog. *Taxiproblem* verwandt ist. Beim Taxiproblem ist die große unbekannte Anzahl N der von 1 bis N durchnummerierten Taxis in einer Stadt zu schätzen, aber im Unterschied zum German Tank Problem aus Realisierungen von k stochastisch unabhängigen Zufallsgrößen, die jeweils eine Gleichverteilung auf den Werten von 1 bis N besitzen. Man kann ja Taxinummern mehrfach beobachten, s. z.B. Henze (2019), S. 239. Ein deutlich anderes Szenario, einen unbekannten Populationsumfang schätzen zu müssen, führt auf die sog. *Capture–Recapture–Methode*, s. z.B. Engel (2000).

Nach diesen Vorbereitungen wenden wir uns der Frage zu, aufgrund einer Stichprobe s die unbekannte Anzahl N möglichst gut zu schätzen. Natürlich muss N mindestens gleich dem größten Stichprobenwert x_k sein, d.h., es muss $N \geq x_k$ gelten, aber um wieviel ist N größer als x_k? Intuitiv wird man erwarten, dass hierbei auch der Stichprobenumfang k eine Rolle spielt. Für Schülerinnen und Schüler dürfte hier das Szenario eines *k-aus-N-Lottos* leicht zugänglich sein: Man hat k Gewinnzahlen beobachtet und möchte daraus N schätzen.

3 Simulieren und schätzen

Als Einstieg bieten sich Simulationen an. Hierzu werden mehrfach Stichproben vom Umfang k ohne Wiederholung aus den Zahlen von 1 bis N gezogen, wobei die Lehrkraft den Wert von N nicht mitteilt. Solche Stichproben lassen sich mithilfe digitaler Werkzeuge wie z.B. GeoGebra leicht erzeugen. Abb. 1 zeigt das Ergebnis eines 5-aus-N-Lottos, bei dem sich die der Größe nach sortierten Gewinnzahlen 4, 9, 10, 15 und 16 ergeben haben.

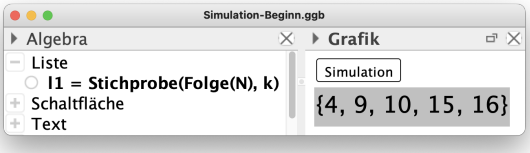

Abb. 1: Erste Simulation und Schätzungen

Der Befehl Folge(N) liefert die Zahlen $1,\ldots,N$. Als Ausgabe wird durch Drücken auf die Schaltfläche jeweils eine Stichprobe der Länge k gezogen. Die Lernenden können jetzt Schätzungen für N abgeben und so erste Erfahrungen mit dem Problem machen. Dieser Einstieg kann selbstverständlich auch als Ratespiel in Gruppen durchgeführt werden. Wenn der Wert von N mitgeteilt wird, kann jede(r) sehen, wie gut er oder sie geschätzt hat.

In Anschluss daran sollten erste Ideen gesammelt werden, wie man aus $x_1,\ldots,x_k$ eine möglichst gute Schätzung erhält. An dieser Stelle sollte noch nicht über Realisierungen, Zufallsgrößen und Schätzverfahren gesprochen werden. Solche Überlegungen ergeben sich zwangsläufig, wenn sich verschiedene Möglichkeiten für eine Schätzung herauskristallisiert haben. Der Klasse dürfte aber schnell klar werden, dass eine Schätzung von N aufgrund von $x_1,\ldots,x_k$ „nicht aus dem hohlen Bauch heraus gegeben werden sollte“, sondern einem „vernünftigen“ Algorithmus folgen sollte, der – ganz egal, wie $x_1,\ldots,x_k$ konkret aussehen – nach einer festen, für alle Stichproben gleichen Regel einen Schätzwert für N ausgibt. Dann dürfte auch der Sinn deutlich werden, warum Wahrscheinlichkeitsrechnung benötigt wird.

4 Heuristische Überlegungen

Die folgenden heuristischen Überlegungen sind durch ein *k-aus-N-Lotto* motiviert, bei dem $x_1,\ldots,x_k$ die nach aufsteigender Größe sortierten Gewinnzahlen sind. Diese Überlegungen müssen sicherlich von der Lehrkraft angeleitet werden, die Berechnungen stellen aber keine große Hürde dar.

4.1 Lücken – gleichmäßig verteilt

Die Zahlen $x_1,\ldots,x_k$ mit $x_1 < \ldots < x_k$ hinterlassen „Lücken“. Die erste dieser Lücken sind die $x_1 - 1$ Zahlen von 1 bis $x_1 - 1$. Die zweite Lücke bilden die $x_2 - x_1 - 1$ Zahlen zwischen x_1 und x_2, die dritte Lücke die $x_3 - x_2 - 1$ Zahlen zwischen x_2 und x_3 usw. Die letzte Lücke besteht aus den $x_k - x_{k-1} - 1$ Zahlen zwischen x_{k-1} und x_k. Es „klafft aber auch eine Lücke“ zwischen x_k und der größtmöglichen, unbekannten Zahl N, und diese Lücke umfasst $N - x_k$ Zahlen. Da der reine Zufall zu den Zahlen $x_1,\ldots,x_k$ geführt hat, sollten – in einer wohltuend vagen Formulierung – alle Lücken „aus stochastischer Sicht gleichwertig sein“ (für den mathematischen Hintergrund s. Henze (1995)). Eine alternative Formulierung ist, dass sich $x_1,\ldots,x_k$ „im Mittel gleichabständig anordnen sollten“. Insofern bietet es sich an, die Größe $N - x_k$ durch das arithmetische Mittel der anderen Lücken zu schätzen und den Ansatz

$$N - x_k \stackrel{!}{=} \frac{\big(x_1 - 1 + (x_2 - x_1 - 1) + \ldots + (x_k - x_{k-1} - 1)\big)}{k} = \frac{x_k - k}{k}$$

zu wählen. Dieser ist gleichbedeutend mit

$$N \stackrel{!}{=} \frac{k+1}{k} \cdot x_k - 1.$$

Die „im-Mittel-gleiche-Lücken-Heuristik“ führt uns also zu einem ersten Schätzwert für N, nämlich

$$\widehat{N}_1 := \frac{k+1}{k} \cdot x_k - 1. \tag{1}$$

An dieser Stelle wird deutlich, dass $\widehat{N}_1$ nicht nur ein konkreter, aufgrund von $x_1,\ldots,x_k$ gewonnener *Schätzwert* für N ist, sondern zugleich ein Schätz*verfahren* beschreibt, nämlich „ordne der Stichprobe $x_1,\ldots,x_k$ den Wert $\frac{k+1}{k}x_k - 1$ zu“. Das Schätzverfahren sieht $\widehat{N}_1$ also als Funktion an, deren Definitionsbereich die Menge Ω aller Stichproben vom Umfang k ist. Wir werden später sehen, dass dieses Schätzverfahren sogar in einem gewissen Sinn optimal ist. Zur „Dach-Notation“ $\widehat{N}_1$ sei angemerkt, dass es in der Statistik allgemein üblich ist, Schätzwerte und Schätzverfahren mit einem Dach zu versehen. Wir schließen uns dieser Notation an.

4.2 Lücken – erste und letzte gleichsetzen

Wird die erste Lücke $x_1 - 1$ mit der letzten, also mit $N - x_k$, gleichgesetzt, so ergibt sich der Ansatz

$$N \stackrel{!}{=} x_1 + x_k - 1$$

und damit der zweite Schätzwert

$$\widehat{N}_2 := x_1 + x_k - 1. \tag{2}$$

4.3 Mittelwerte gleich

Es bietet sich auch an, das arithmetische Mittel

$$\bar{x}_k := \frac{1}{k} \sum_{j=1}^{k} x_j$$

mit dem arithmetischen Mittel über $1,2,\ldots,N$, also mit $\frac{N+1}{2}$, gleichzusetzen. Dieser Ansatz, also

$$\frac{N+1}{2} \stackrel{!}{=} \bar{x}_k,$$

führt auf einen dritten Schätzwert für N, nämlich

$$\widehat{N}_3 := 2 \cdot \bar{x}_k - 1. \tag{3}$$

Genauso wie $\widehat{N}_1$ kann man auch $\widehat{N}_2$ und $\widehat{N}_3$ als Schätzverfahren ansehen, denn beide stellen ja eine Vorschrift dar, wie man von jeder Stichprobe s über die geordneten Werte $x_1, \ldots, x_k$ zu einem konkreten Schätzwert für N gelangt. Die Lehrkraft könnte an dieser Stelle fragen, welche Vor- und Nachteile diese drei Schätzverfahren besitzen. Man erkennt z.B., dass das dritte Verfahren Schätzwerte liefern kann, die kleiner als x_k sind, was natürlich unsinnig wäre, denn wir wissen ja, dass die Ungleichung $x_k \leq N$ erfüllt ist. Sowohl das erste als auch das dritte Verfahren können zudem Schätzwerte liefern, die keine ganzen Zahlen sind. In einem solchen Fall würde man die nächstgrößere ganze Zahl als konkreten Schätzwert nehmen.

5 Simulationen

Wie können wir nun die Güte dieser Schätzungen beurteilen? Was heißt hier eigentlich *Güte*? An diese anspruchsvollen Fragen kann man zuerst mithilfe von Simulationen herangehen. Es sollte klar werden, dass es sinnvoll ist, möglichst viele Simulationsergebnisse zu betrachten. Wenn diese Ergebnisse gleichmäßig um den (unbekannten) Wert N schwanken und eine möglichst kleine Streuung aufweisen, dann sollte das Schätzverfahren gut sein. Wenn man also bei festem N sehr häufig Stichproben vom Umfang k zieht, sollten sich bei jedem der Schätzverfahren $\widehat{N}_j$ mit $j \in \{1,2,3\}$ die resultierenden Schätzwerte um den wahren Wert N „ausmitteln“, d.h., die jeweiligen Mittelwerte der Schätzungen sollten jeweils in der Nähe von N sein. Unter dieser Bedingung, die auf der theoretischen Ebene *Unverzerrtheit* oder *Erwartungstreue* eines Schätzverfahrens heißt, sollten die Schätzwerte möglichst wenig streuen. Die Stärke der jeweiligen Streuung kann über die empirische Varianz der Schätzwerte gemessen werden.

Abb. 2: Simulation mit GeoGebra

Abb. 2 zeigt die wesentlichen Befehle einer Simulation mit GeoGebra, um die drei Schätzverfahren $\widehat{N}_1$, $\widehat{N}_2$ und $\widehat{N}_3$ zu vergleichen.

Das Objekt *Daten* ist eine Matrix mit w Zeilen, in der jeweils eine Stichprobe der Länge k aus den Zahlen $1, \ldots, N$ erzeugt wird. Die Objekte N_1, N_2 und N_3 liefern Listen mit je w Schätzungen für die drei Schätzverfahren. Hiermit werden dann jeweils der Mittelwert sowie die empirische Standardabweichung berechnet. Es ist schon beeindruckend, wie elegant eine Umsetzung mit GeoGebra erfolgen kann. Die Simulationen ergeben Schätzwerte, die sich größtenteils in der Nähe von N befinden.

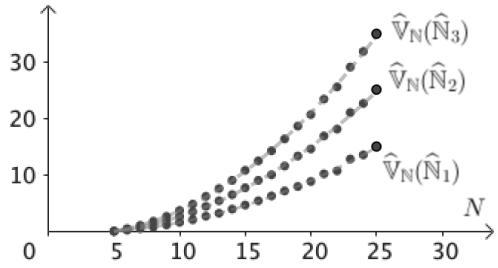

Abb. 3: Geschätzte Varianzen von $\widehat{N}_1$, $\widehat{N}_2$ und $\widehat{N}_3$ ($k = 5$, 10 000 Wiederholungen)

Für den Fall $k = 5$ und $N \in \{5, 6, \ldots, 25\}$ wurden für jedes der Schätzverfahren $\widehat{N}_1$, $\widehat{N}_2$ und $\widehat{N}_3$ jeweils 10 000 Simulationen durchgeführt. Abb. 3 zeigt in Form von Punkten die jeweiligen empirischen Varianzen der erhaltenen 10 000 Schätzwerte. Die in Abschn. 7 angegebenen theoretischen Varianzen liegen auf den gestrichelt eingezeichneten Kurven, die jeweils Parabeln bilden. Es ist klar zu erkennen, dass der Schätzer $\widehat{N}_1$ die kleinsten Varianzen aufweist und damit beiden anderen Schätzern überlegen ist. Dabei schneidet $\widehat{N}_3$ noch deutlich schlechter ab als $\widehat{N}_2$.

An dieser Stelle ist der Übergang zur Wahrscheinlichkeitsrechnung und damit zu Realisierungen, Zufallsgrößen, erwartungstreuen Schätzern und den beiden Kenngrößen Erwartungswert und Varianz angebracht. Dieser behutsame Weg zeigt sehr schön die Sinnhaftigkeit von Modellierungen mithife der Wahrscheinlichkeitsrechnung auf.

Der mathematische Kern wurde ja schon in Abschn. 2 dargelegt. Insbesondere kann die Bedeutung einer begrifflichen Trennung zwischen einer Zufallsgröße und deren Realisierungen nicht hoch genug eingeschätzt werden. Erst dadurch kann nämlich ein wirkliches Verständnis aufgebaut werden. Eine solche Trennung von Begriffen findet oft nicht genügend statt, was sich etwa besonders deutlich bei der Behandlung von Konfidenzintervallen zeigt (s. z.B. Callaert (2007)).

6 Maximum-Likelihood-Schätzung

In Abschn. 4 haben wir den Populationsumfang N aufgrund verschiedener heuristischer Überlegungen quasi „aus dem Bauch heraus“ geschätzt, und in unserem konkreten Kontext klangen die Vorschläge $\widehat{N}_1$, $\widehat{N}_2$ und $\widehat{N}_3$ durchaus plausibel. Erstrebenswert sind jedoch übergeordnete Prinzipien, nach denen sich in allgemeinen Situationen gute Schätzverfahren herleiten lassen. Ein solches Prinzip heißt *Maximum Likelihood*. Es wurde von dem britischen Statistiker (Sir) Ronald Aylmer Fisher (1890-1962) propagiert und mathematisch genauer untersucht. Das *Maximum-Likelihood-Schätzprinzip* besagt, bei vorliegenden Daten und verschiedenen, zur Auswahl stehenden stochastischen Modellen dasjenige Modell für das glaubwürdigste zu halten, welches den Daten die größte Auftretenswahrscheinlichkeit verleiht.

In unserem konkreten Fall liegen Daten in Form einer Stichprobe $s = \{x_1, \ldots, x_k\}$ vor, und die verschiedenen, zur Auswahl stehenden stochastischen Modelle sind durch den Populationsumfang N gegeben.

Wenn wir das für Wahrscheinlichkeit stehende Symbol $\mathbb{P}$ mit dem Index N versehen, um zu betonen, dass wir N spezifizieren müssen, um überhaupt erst Wahrscheinlichkeiten ausrechnen zu können (denn wir arbeiten ja mit einem Laplace-Modell auf allen k-elementigen Teilmengen der N-elementigen Menge $\{1,2,\ldots,N\}$!), gilt

$$\mathbb{P}_N(S = s) = \frac{1}{\binom{N}{k}}, \tag{4}$$

falls die Ungleichung $N \geq x_k$ erfüllt ist.

Das Maximum-Likelihood-Schätzprinzip fordert auf, dasjenige N als Schätzwert für den unbekannten Populationsumfang zu wählen, für das die in (4) stehende Wahrscheinlichkeit größtmöglich wird. Wenn wir diese Wahrscheinlichkeit maximieren wollen, müssen wir N möglichst klein wählen, denn N tritt ja auf der rechten Seite von (4) im Nenner auf. Der dort stehende Binomialkoeffizient wird minimal, wenn wir N möglichst klein wählen. Der kleinstmögliche Wert für N ist aber der maximale Wert x_k in der Stichprobe s. Wir haben also erhalten, dass

$$\widehat{N}_{ML}(s) := x_k$$

der sog. *Maximum-Likelihood-Schätzwert* für N zur Stichprobe s ist. Dabei steht *ML* für *Maximum Likelihood*. Wie schon bei $\widehat{N}_1$, $\widehat{N}_2$ und $\widehat{N}_3$ bemerkt, sieht man auch hier, was ein Schätz*verfahren* ist, nämlich eine Vorschrift (Abbildung, Funktion), die jeder Stichprobe s einen konkreten Zahlenwert als *Schätzwert* für N zuordnet. Die Abbildung $\widehat{N}_{ML}$ mit Definitionsbereich Ω, die jeder Stichprobe s aus Ω deren größten Wert zuordnet, ist also das Maximum-Likelihood-Schätz*verfahren*.

Wie gut ist das Maximum-Likelihood-Schätzverfahren in unserem speziellen Kontext? Um dessen Güte auf der theoretischen Ebene zu beurteilen, können wir uns nicht nur auf eine Stichprobe s beschränken. Diese Stichprobe hatten wir ja als Realisierung einer Zufallsgröße S mit einer Gleichverteilung auf allen k-elementigen Teilmengen der Menge $\{1,\ldots,N\}$ und damit auf Ω angesehen. Wir müssen uns jetzt wiederum auf die Ebene von Zufallsgrößen begeben und für jedes $j \in \{1,\ldots,k\}$ den Wert x_j – also den j-kleinsten Wert in der Stichprobe s – als Realisierung einer mit X_j bezeichneten Zufallsgröße ansehen. Da definitionsgemäß $x_1 < x_2 < \ldots < x_k$ gilt, gelten auch für die Zufallsgrößen $X_1, \ldots, X_k$ die Ungleichungen $X_1 < X_2 < \ldots < X_k$. Wir können für jedes $j \in \{1,\ldots,k\}$ die Zufallsgröße X_j ganz einfach in Worten beschreiben: Vor Beobachtung einer rein zufälligen k-Auswahl der Zahlen $1,\ldots,N$ ist X_j die j-kleinste dieser Zahlen. Insbesondere ist dann X_k die größte dieser Zahlen. In der gleichwertigen Vorstellung von einem *k-aus-N-Lotto* modelliert die Zufallsgröße X_j die j-kleinste der k Gewinnzahlen einer Ziehung bei einem solchen Lotto. Wir werden fortan sprachlich diesen Lotto-Kontext verwenden.

Was passiert, wenn wir das Maximum-Likelihood-Schätzverfahren oft anwenden? Wie streuen die Schätzwerte für N? Welcher Schätzwert ergibt sich „auf die Dauer im Mittel“? Die letzte, vage formulierte Frage betrifft den Erwartungswert von X_k. Da wir auch für die Berechnung des Erwartungswertes N spezifizieren müssen, machen wir diesen Umstand ebenfalls durch Indizierung mit N kenntlich. Nach der Formel „Summe aus Wert mal Wahrscheinlichkeit“ gilt dann

$$\mathbb{E}_N(X_k) = \sum_{\ell=k}^{N} \ell \cdot \mathbb{P}_N(X_k = \ell). \tag{5}$$

Dabei wurde verwendet, dass die größte Lottozahl bei einem *k-aus-N-Lotto* nur die Werte $k,\ldots,N$ annehmen kann. Die in (5) stehende Wahrscheinlichkeit lässt sich leicht ermitteln: Unter allen $\binom{N}{k}$ k-Auswahlen aus $\{1,\ldots,N\}$ gibt es $\binom{\ell-1}{k-1}$ Stück, bei denen die größte der ausgewählten Zahlen gleich ℓ ist, denn dann müssen die kleineren $k-1$ Zahlen aus den Zahlen von 1 bis $\ell-1$ ausgewählt werden. Es

folgt

$$\mathbb{P}_N(X_k = \ell) = \frac{\binom{\ell-1}{k-1}}{\binom{N}{k}}, \quad \ell = k, \ldots, N. \tag{6}$$

Abb. 4 zeigt das Stabdiagramm der Verteilung von X_5 für den Fall $N = 20$.

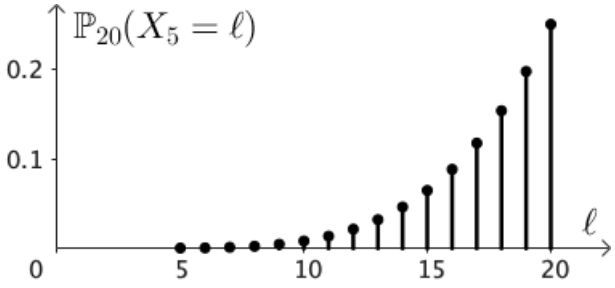

Abb. 4: Stabdiagramm der Verteilung von X_5 ($N = 20$)

An dieser Stelle kann mithilfe von GeoGebra-CAS sogar der Term für $\mathbb{E}_N(X_k)$ ermittelt werden, der noch durch die Befehle *Vereinfache* und *Faktorisiere* vereinfacht dargestellt wird (siehe Abb. 5).

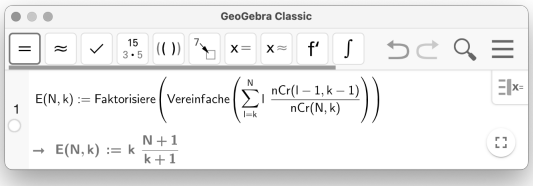

Abb. 5: Berechnung von $\mathbb{E}_N(X_k)$ mit GeoGebra-CAS

Vielleicht werden einige Lernende durch dieses Vorgehen motiviert, den Term ohne Rechnereinsatz zu erhalten. Mit (6) folgt

$$\mathbb{E}_N(X_k) = \frac{1}{\binom{N}{k}} \sum_{\ell=k}^{N} \ell \cdot \binom{\ell-1}{k-1} = \frac{1}{\binom{N}{k}} \cdot k \cdot \sum_{\ell=k}^{N} \binom{\ell}{k}.$$

Die hier auftretende Summe ist ein Binomialkoeffizient, den wir letztlich schon hergeleitet haben. Da die Summe aller in (6) stehenden Wahrscheinlichkeiten gleich eins ist, gilt

$$\sum_{\ell=k}^{N} \binom{\ell-1}{k-1} = \binom{N}{k},$$

und zwar für jede Wahl von k und N mit $1 \le k \le N$. Wir können also auch N durch $N+1$ und k durch $k+1$ ersetzen und erhalten das aufgrund der beteiligten Binomialkoeffizienten oft auch *Hockeyschlägerregel* genannte *Gesetz der oberen Summation*

$$\sum_{\ell=k}^{N} \binom{\ell}{k} = \binom{N+1}{k+1}, \tag{7}$$

siehe z.B. Henze (2023). Hiermit folgt

$$\begin{aligned} \mathbb{E}_N(X_k) &= \frac{1}{\binom{N}{k}} \cdot k \cdot \binom{N+1}{k+1} \\ &= \frac{k}{k+1} \cdot (N+1) \\ &< N. \end{aligned} \tag{8}$$

Die strikte Ungleichung ist gleichbedeutend mit $k < N$. Sie zeigt, dass das Maximum-Likelihood-Schätzverfahren das unbekannte N systematisch unterschätzt. Im nächsten Abschnitt lernen wir ein in einem gewissen Sinn bestes Schätzverfahren für N kennen. Dieses haben wir schon auf heuristischem Weg in Abschn. 4.1 gewonnen. Ein Schätzverfahren wird fortan kurz als *Schätzer* bezeichnet.

7 Der gleichmäßig beste erwartungstreue Schätzer für N

Das Gleichheitszeichen in (8) gibt uns einen Hinweis darauf, wie wir einen Schätzer erhalten können, der das unbekannte N „auf die Dauer im Mittel richtig schätzt." Es liefert nämlich die interessante Gleichung

$$N = \frac{k+1}{k}\mathbb{E}_N(X_k) - 1 = \mathbb{E}_N\Big(\frac{k+1}{k} \cdot X_k - 1\Big).$$

Dabei gilt das zweite Gleichheitszeichen aufgrund der Linearität der Erwartungswertbildung. Obige Gleichung besagt, dass wir einen sog. *erwartungstreuen Schätzer* erhalten, wenn wir den maximalen Wert aus der Stichprobe vergrößern, indem wir ihn mit dem Faktor $\frac{k+1}{k}$ multiplizieren und danach noch eins subtrahieren. Ganz egal, wie groß das unbekannte N ist: „Im Mittel" schätzen wir also auf diese Weise richtig. Setzen wir

$$\widehat{N}_1 := \frac{k+1}{k} \cdot X_k - 1, \tag{9}$$

so haben wir durch Modifikation des Maximum-Likelihood-Schätzers $\widehat{N}_{ML} = X_k$ einen erwartungstreuen Schätzer erhalten. Dieser Schätzer ist der gleiche wie in (1) (im Unterschied zu (1) ist hier nur die Ebene von Zufallsgrößen betont!), und er liefert stets Werte größer oder gleich X_k, denn es gilt

$$\frac{k+1}{k} \cdot X_k - 1 \ge X_k \Longleftrightarrow X_k \ge k.$$

Wenn wir diesen Schätzer auf die Daten $k = 4$ und $x_1 = 31$, $x_2 = 43$, $x_3 = 66$ sowie $x_4 = 122$ anwenden, ergibt sich der konkrete Schätzwert 151,5.

Wie schon früher betont, würde man in einem solchen Fall eines nicht ganzzahligen Schätzwertes auf

die nächstgrößere ganze Zahl aufrunden und als Schätzwert 152 angeben. In Johnson (1994) wird die Varianz des Schätzers $\widehat{N}_1$ mit

$$\mathbb{V}_N(\widehat{N}_1) = \frac{1}{k} \cdot \frac{(N-k)(N+1)}{k+2} \qquad (10)$$

angegeben. Man beachte, dass auch das Symbol $\mathbb{V}$ für Varianz mit dem Index N versehen wurde, um zu betonen, dass diese Varianz unter Zugrundelegung des Wertes N berechnet wurde.

Durch Anwendung der Definition der Varianz und Einsatz von GeoGebra-CAS haben Lernende auch eine Chance, den Term für diese Varianz zu ermitteln (s. Abb. 6).

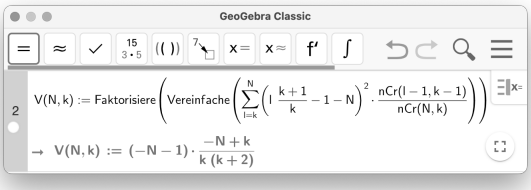

Abb. 6: Berechnung von $\mathbb{V}_N(\widehat{N}_1)$ mit GeoGebra-CAS

Die Herleitung von (10) ist kein Hexenwerk: Allgemein gilt für eine Zufallsgröße Y und reelle Zahlen a,b die Gleichung $\mathbb{V}(aY+b) = a^2\mathbb{V}(Y)$, und somit folgt aus (9): $\mathbb{V}_N(\widehat{N}_1) = \frac{(k+1)^2}{k^2} \cdot \mathbb{V}_N\big(X_k\big)$.

Weiter gilt: $\mathbb{V}_N(X_k) = \mathbb{E}_N\big(X_k^2\big) - \Big(\mathbb{E}_N\big(X_k\big)\Big)^2$.

Der Subtrahend ist das Quadrat des in (8) stehenden Ausdrucks, und es gilt

$$\mathbb{E}_N\big(X_k^2\big) = \sum_{\ell=k}^{n} \ell^2 \cdot \mathbb{P}_N(X_k = \ell).$$

Mit (6) folgt: $\mathbb{E}_N\big(X_k^2\big) = \frac{k}{\binom{N}{k}} \sum_{\ell=k}^{N} \ell \binom{\ell}{k}$.

Hier kommt man weiter, indem man das ℓ vor dem Binomialkoeffizienten als $(\ell+1)-1$ schreibt. Es gilt

$$\begin{aligned} \sum_{\ell=k}^{N} (\ell+1)\binom{\ell}{k} &= (k+1)\sum_{\ell=k}^{N}\binom{\ell+1}{k+1} \\ &= (k+1)\binom{N+2}{k+2}. \end{aligned}$$

Dabei folgt das letzte Gleichheitszeichen aus dem Gesetz der oberen Summation (7). Berücksichtigt man noch die Gültigkeit von

$$\frac{k}{\binom{N}{k}} \sum_{\ell=k}^{N} \binom{\ell}{k} = \frac{k(N+1)}{k+1} \quad \Big(= \mathbb{E}_N(X_k)\Big),$$

so kann man alles zusammensetzen, und (10) folgt mithilfe direkter Rechnung unter Verwendung der die Binomialkoeffizienten bildenden Fakultäten und Zusammenfassen. Man beachte, dass die Varianz von $\widehat{N}_1$ gleich null ist, wenn $k = N$ gilt, denn in diesem Fall nimmt ja X_k immer den Wert N an.

Unter allen erwartungstreuen Schätzern für N ist $\widehat{N}_1$ derjenige mit der (gleichmäßig in N) kleinsten Varianz, d.h., für jeden anderen erwartungstreuen Schätzer $\widetilde{N}$ für N gilt $\mathbb{V}_N(\widehat{N}) \le \mathbb{V}_N(\widetilde{N})$ für jedes $N \ge 1$. Hintergrund für diesen Sachverhalt ist, dass die Zuordnung $\{X_1,\ldots,X_k\} \mapsto X_k$ *suffizient* (*erschöpfend*) für N ist. Dieser in der Mathematischen Statistik wichtige Fachbegriff bedeutet, dass die bedingte Verteilung der gesamten Stichprobe $\{X_1,\ldots,X_k\}$ unter der Bedingung, dass man (eine Realisierung von) X_k kennt, nicht vom unbekannten N abhängt. Bei Kenntnis von X_k bilden nämlich $X_1,\ldots,X_{k-1}$ eine rein zufällige $(k-1)$-Auswahl von $1,2,\ldots,X_k-1$. Nach einem berühmten Satz von Rao-Blackwell (s. z.B. Czado und Schmidt (2011), S. 109) hat ein erwartungstreuer Schätzer, der auf einer suffizienten Statistik basiert, gleichmäßig kleinste Varianz. Im Fachjargon ist er dann ein *UMVUE-Schätzer* (*uniformly minimum variance unbiased estimator*).

Wir überlegen uns jetzt, dass auch der durch Gleichsetzen der ersten und letzten Lücke entstehende Schätzer $\widehat{N}_2$ aus (2) erwartungstreu ist. Auf der Ebene von Zufallsgrößen gilt $\widehat{N}_2 := X_1 + X_k - 1$ und damit

$$\mathbb{E}_N\big(\widehat{N}_2\big) = \mathbb{E}_N(X_1) + \mathbb{E}_N(X_k) - 1.$$

Die Verteilung von X_1 ist wie diejenige von X_k recht schnell bestimmt: Es gilt

$$\mathbb{P}_N(X_1 = j) = \frac{\binom{N-j}{k-1}}{\binom{N}{k}}, \quad j = 1,\ldots,N+1-k,$$

denn wenn die kleinste der k ausgewählten Zahlen gleich j ist, müssen aus den verbleibenden $N-j$ größeren Zahlen $k-1$ Zahlen ausgewählt werden. Nun gilt nach (6) (mit $\ell = N-j+1$)

$$\mathbb{P}_N(X_k = N-j+1) = \frac{\binom{N-j}{k-1}}{\binom{N}{k}}, \quad j = 1,\ldots,N+1-k.$$

Wegen $\mathbb{P}_N(X_k = N-j+1) = \mathbb{P}_N(N-X_k+1 = j)$ haben also die Zufallsgrößen X_1 und $N-X_k+1$ dieselbe Verteilung und damit auch denselben Erwartungswert. Es folgt wie behauptet

$$\mathbb{E}_N\big(\widehat{N}_2\big) = \mathbb{E}_N\big(N-X_k+1\big) + \mathbb{E}_N\big(X_k\big) - 1 = N.$$

Auch der durch Gleichsetzen des Stichprobenmittels

mit $\frac{N+1}{2}$ gewonnene Schätzer $\widehat{N}_3$ aus (3) ist erwartungstreu für N. Auf der Ebene von Zufallsgrößen gilt

$$\widehat{N}_3 = 2 \cdot \frac{1}{k} \sum_{j=1}^{k} X_j - 1.$$

Das obige arithmetische Mittel $\overline{X}_k$ hat den gleichen Erwartungswert wie eine Zufallsgröße, die auf den Zahlen von 1 bis N gleichverteilt ist. Somit folgt

$$\mathbb{E}_N(\overline{X}_k) = \frac{1}{N} \sum_{j=1}^{N} j = \frac{N+1}{2},$$

und es ergibt sich $\mathbb{E}_N\big(2 \cdot \overline{X}_k - 1\big) = N$. Also ist auch $2 \cdot \overline{X}_k - 1$ ein erwartungstreuer Schätzer für N.

Abschließend seien noch die in Johnson (1994) aufgeführten Varianzen der Schätzer $\widehat{N}_2$ und $\widehat{N}_3$ angegeben. In Ergänzung zu (10) gelten

$$\mathbb{V}_N\big(\widehat{N}_2\big) = \frac{2}{k+1} \cdot \frac{(N-k)(N+1)}{k+2}, \qquad (11)$$

$$\mathbb{V}_N\big(\widehat{N}_3\big) = \frac{k+2}{3k} \cdot \frac{(N-k)(N+1)}{k+2}. \qquad (12)$$

Jede der Varianzen $\mathbb{V}_N\big(\widehat{N}_j\big)$, $j \in \{1,2,3\}$, ist also eine quadratische Funktion in N. Man beachte, dass (10), (11) und (12) den gleichen Faktor $(N-k)(N+1)/(k+2)$ aufweisen, sodass man die unterschiedlichen Streckfaktoren direkt vergleichen kann.

8 Fazit

Das vorgestellte Problem ist ein motivierendes Beispiel mit viel Potential, und zwar besonders durch die unterschiedlichen Herangehensweisen. Dabei ist jeder Weg (probieren, simulieren, CAS-Einsatz, Theorie) wichtig. Es wird deutlich, warum stochastische Modellierungen bedeutsam sind, denn nur die Wahrscheinlichkeitsrechnung liefert tragfähige Antworten auf die Frage nach der Güte der verschiedenen Schätzverfahren. Auch Elemente der in einigen neuen Lehrplänen endlich wieder mehr Gewicht erhaltenden Kombinatorik werden benötigt. Des Weiteren werden Begriffe wie *Zufallsgröße*, *Realisierungen* und *erwartungstreuer Schätzer* mit Leben gefüllt, und auch der Unterschied zwischen einem *Schätzwert* und einem *Schätzer* als Schätz*verfahren* wird deutlich. Besonders hervorheben möchten wir, dass auch die wichtige Philosophie der *Maximum-Likelihood-Schätzung* an einem historisch und praktisch wichtigen Fall thematisiert wird.

Fußnote: Die verwendeten GeoGebra-Programme können vom zweiten Autor erhalten werden.

Danksagung: Wir danken den Gutachtern für wertvolle Hinweise.

Literatur

Bischoff, M. (2022): Wie Mathematik den Alliierten half, den Zweiten Weltkrieg zu gewinnen. In: Spektrum.de 28.10.2022, abgerufen am 20.6.2023.

Callaert, H. (2007): Understanding confidence intervals. Proceedings of the 5th Congress of the European Society for Research in Mathematics Education, 692–701.

Czado, C., und Schmidt, Th. (2011): Mathematische Statistik. Berlin, Heidelberg. Springer-Verlag.

Dambeck, H. (2010): Rechentricks der Alliierten. Wie Seriennummern die Nazi-Industrie verrieten. In: spiegel.de 22.11.2010, abgerufen am 20.6.2023.

Davies, G. (2006): Gavin Davies does the maths – how a statistical formula won the war. In: The Guardian, July 20^{th}, 2006, abgerufen am 20.6.2023.

Engel, J. (2000): Markieren–Einfangen–Schätzen: Wie viele wilde Tiere? *Stochastik in der Schule, 20*(2), 17–24.

Henze, N. (1995): The distribution of spaces on lottery tickets. *Fibonacci Quarterly, 33*(5), 426–431.

Henze, N. (2019): Stochastik: Eine Einführung mit Grundzügen der Maßtheorie. Heidelberg, Springer Spektrum.

Henze, N. (2023): Binomialkoeffizienten – verstehen oder rechnen? *Stochastik in der Schule, 43*(1), 13–18.

Johnson, Roger W. (1994): Estimating the size of a population. *Teaching Statistics, 16*(2), 50–52.

Ruggles, R., und Brodie, H. (1947): An empirical approach to economic intelligence in World War II. *Journal of the American Statistical Association, 42*(237), 72–91.

Anschrift der Verfasser:

Prof. i.R. Dr. Norbert Henze
KIT Distinguished Senior Fellow
Institut für Stochastik
Karlsruher Institut für Technologie (KIT)
Englerstr. 2
76131 Karlsruhe
Henze@kit.edu

Reimund Vehling
vehling@icloud.com