Prof. Dr. Daniel Tillich
ist seit Sommer 2020 Professor für Wirtschaftsmathematik am Standort Dresden. Im Oktober 2021 hat er die Leitung der Studienrichtung Finanzwirtschaft – Bank übernommen. Vor seinem Wechsel an die Berufsakademie Sachsen arbeitete Prof. Dr. Tillich knapp vier Jahre im Risikocontrolling der Sächsischen Aufbaubank (SAB). Dem Thema Kreditrisiko hat er sich bereits während seiner Tätigkeit am Lehrstuhl für Quantitative Methoden, insbesondere Statistik der TU Dresden gewidmet. Seit dieser Zeit begleitet ihn auch sein Interesse für Hochschuldidaktik, welches er im Zertifikatsprogramm des Hochschuldidaktischen Zentrums Sachsen erweitern und nachweisen konnte.

Kontakt: daniel.tillich@ba-sachsen.de

Statistisches Schätzen in Theorie und Praxis

Über kunstvolles Vermuten, die Tücken der Standardannahme und Statistiker im Bankwesen [1]

Daniel Tillich

Das Schätzen unbekannter Größen aus Daten gehört zu den Grundproblemen der Statistik. Konkurrieren mehrere Schätzverfahren miteinander, muss beurteilt werden, welches Verfahren das geeignetste ist. Der vorliegende Text beleuchtet mittels dreier Beispiele aus Theorie und Praxis, anhand welcher Kriterien über die Eignung entschieden werden kann. Dabei zeigt sich, dass häufig schon sehr einfache, intuitive Güteeigenschaften genügen, um ein Schätzverfahren zu beurteilen. Tiefer und kritischer muss dagegen darüber nachgedacht werden, ob die üblicherweise unterstellten Annahmen im konkreten Anwendungsfall erfüllt sind.

The estimation of unknown quantities from data is one of the fundamental problems of statistics. If different estimation methods compete with each other, it is necessary to decide which method is the most suitable one. Considering three examples from theory and practice, the present article discusses the criteria that can be used to decide on the suitability of an estimation method. It is shown that rather simple, intuitive quality properties are often sufficient to assess an estimation procedure. However, deeper and more critical reflections are required to determine whether the commonly presumed assumptions are met in the specific case of application.

Einführung

Statistische Informationen begegnen uns in allen Lebensbereichen: In der beruflichen Praxis, genau wie im Privaten. Deshalb nimmt das Fach Statistik auch in der Lehre an der Berufsakademie Sachsen einen wichtigen Platz ein. Anhand dreier Fälle, in denen jeweils Wahrscheinlichkeiten gesucht sind, genauer:

- eine klassische Eintrittswahrscheinlichkeit im Hörsaalexperiment,
- die Trefferwahrscheinlichkeit beim Strafwurf im Handball und
- die Ausfallwahrscheinlichkeit bei der Kreditrisikomessung,

werden folgende Fragen beantwortet:

- Wieso werden Wahrscheinlichkeiten eigentlich mit relativen Häufigkeiten geschätzt?
- Was macht ein Mathematiker, der sich mit Statistik beschäftigt?
- Welche Herausforderungen bestehen bei der Messung von Kreditrisiken in der Finanzwirtschaft?

1 *Dieser Text verschriftlicht die Antrittsvorlesung des Autors an der Berufsakademie Sachsen.*

Dieser – eher didaktisch als forschungsorientiert ausgelegte – Beitrag stellt dabei drei leicht verständliche Güteeigenschaften für Schätzmethoden vor, die sich neben die üblichen Lehrbuch-Güteeigenschaften einreihen. Außerdem beschreibt er praktische Probleme der Parameterschätzung, die die Standardannahmen der schließenden Statistik betreffen. Diese treten u. a. bei der Kreditrisikomessung in Banken auf. Schlussendlich soll der Beitrag dabei ganz im Sinne einer Antrittsvorlesung auch den Verfasser näher vorstellen.

Fall 1: Hörsaalexperiment – Bausteine im Beutel

In Abbildung 1 ist ein Beutel mit Bausteinen erkennbar. In dem Beutel befinden sich rote, grüne und gelbe Bausteine identischer Form und Größe. Uns interessiert der unbekannte Anteil (Prozentwert) der roten Bausteine an der Gesamtzahl aller Bausteine. Oder mit anderen Worten: Uns interessiert die Wahrscheinlichkeit dafür, dass beim Ziehen eines Bausteins ein roter Baustein gezogen wird.

Abbildung 1: Bausteine im Beutel, Handball und Geldschein als Sinnbild für die drei im Beitrag behandelten Fälle

Schätzen Sie mal!

Das umgangssprachliche Schätzen kommt einem willkürlichen Raten gleich. Statistisches Schätzen als Teil der schließenden Statistik ist dagegen „kunstvolles Vermuten“[2]. Kunstvoll ist das Schätzen in dem Sinne, dass Daten ausgewertet und daraus Informationen gezogen werden, um auf die Grundgesamtheit zu schließen. Beispielhaft sei beim wiederholten einmaligen Ziehen eines Bausteins mit Zurücklegen die in Abbildung 2 dargestellte konkrete Stichprobe vom

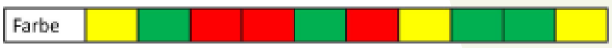

Abbildung 2: Stichprobe vom Umfang n=10

Umfang $n = 10$ entstanden.

Im Hörsaal wird nun fast sicher die **relative Häufigkeit** als Schätzwert für den Anteil der roten Bausteine im Beutel vorgeschlagen. Für die gegebene Stichprobe ergibt sich:

$$\frac{\text{Anzahl der roten Bausteine}}{\text{Anzahl der Versuche}} = \frac{3}{10} = 0{,}3.$$

Doch warum verwendet man die relative Häufigkeit und z. B. nicht eine der folgenden zwei Alternativen?

Alternative 1: Die Anzahl der roten Bausteine wird durch die Anzahl der Versuche minus 1 geteilt.[3] Für die gegebene Stichprobe ergibt sich:

$$\frac{\text{Anzahl der roten Bausteine}}{\text{Anzahl der Versuche} - 1} = \frac{3}{9} = \frac{1}{3} = 0{,}3\overline{3}.$$

Alternative 2: Bei der Berechnung der relativen Häufigkeit werden der erste und der letzte Versuch ignoriert. Im vorliegenden Fall entfallen zwei gelbe Bausteine und es ergibt sich:

$$\frac{\text{Anzahl der roten Bausteine}}{\text{Anzahl der Versuche}} = \frac{3}{8} = 0{,}375.$$

Es könnte doch sein, dass zumindest eine der beiden Alternativen einen besseren Schätzwert liefert als die klassische relative Häufigkeit.

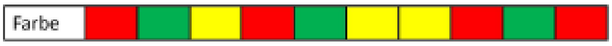

Abbildung 3: Eine weitere Stichprobe vom Umfang n=10

Die Suche nach der besten Alternative verkompliziert sich zusätzlich, wenn man bedenkt, dass die Stichprobe ja auch hätte ganz anders aussehen können. Für die in Abbildung 3 dargestellte Stichprobe hätten sich beispielsweise folgende Schätzwerte ergeben:

- relative Häufigkeit: $\frac{4}{10} = 0{,}4$
- Alternative 1 $\frac{4}{9} = 0{,}4\overline{4}$
- Alternative 2: $\frac{2}{8} = 0{,}25$

(Beachte: Die ignorierten Versuche 1 und 10 lieferten beide „rot“ als Ergebnis.)

Und wieder lautet die Frage: Welche Variante ist hier am besten? Ohne Blick in den Beutel ist es unmöglich, diese Frage für den konkreten Fall zu beantworten. Doch dieser Blick auf die Wahrheit bleibt wie so häufig in der Praxis verwehrt. Insofern muss die Frage umformuliert werden: Welche Variante liefert über alle möglichen Stichproben hinweg die besten Schätzwerte?

Statistisches Schätzen

Hier kommt die mathematische Statistik ins Spiel. Diese stellt nicht die Frage, welcher Zahlenwert bei der konkret vorliegenden Stichprobe am besten ist. Sie beschäftigt sich vielmehr mit der Frage, welche Methodik über alle möglichen Versuchsausgänge am besten geeignet ist. Dazu untersuchen die Statistiker:innen das zugehörige Zufallsexperiment und die entstehenden

2 *In Anlehnung an das Werk „Ars conjectandi“ des Schweizer Mathematikers Jakob Bernoulli (1654-1705).*

3 *In Anlehnung an eine Variante der Varianzschätzung, bei der nicht durch den Stichprobenumfang , sondern durch n−1 geteilt wird.*

Zufallsvariablen auf theoretischer Ebene, denn die konkreten Ergebnisse der Stichprobe sind zufallsabhängig. Die Statistiker:innen prüfen dann, ob die Methode der Schätzung bestimmte Güteeigenschaften erfüllt.
Drei sehr einfache und sehr intuitive Gütekriterien werden im Folgenden vorgestellt. Sie sind trotz ihrer Verständlichkeit in der aktuellen Lehrbuchliteratur häufig nicht enthalten:[4]

1. **Plausible Schätzwerte:**
   Die Schätzmethode sollte keine unsinnigen Werte liefern. Wenn z. B. die Erfolgswahrscheinlichkeit geschätzt werden soll, sollten nur Schätzwerte im Intervall von 0 bis 1 möglich sein.
2. **Berücksichtigung aller Beobachtungen:**
   Die Schätzmethode sollte das Ergebnis jeder einzelnen Ziehung berücksichtigen, also keine Informationen vernachlässigen.
3. **Gleichgewichtung aller Beobachtungen:**
   Jede einzelne Ziehung sollte den gleichen Einfluss auf das Ergebnis der Schätzung haben, d. h. keine Ziehung ist wichtiger als die andere.

Die obige Alternative 2 erfüllt zwar das erste Kriterium, das zweite und dritte dagegen nicht. Offensichtlich werden in dieser Variante nicht alle Beobachtungen genutzt und dementsprechend Informationen verschenkt. Alternative 1 erfüllt die Kriterien 2 und 3. Im Widerspruch zu Kriterium 1 kann allerdings ein unplausibler Schätzwert entstehen, wenn alle Ziehungen „rot" liefern. Bei einem Stichprobenumfang von $n = 10$ ergäbe sich in diesem Fall $10/9 = 1{,}1\overline{1}$ als Schätzwert für eine Wahrscheinlichkeit, die jedoch höchstens 1 betragen kann. Die relative Häufigkeit erfüllt alle drei genannten Gütekriterien. Um die Überlegenheit der relativen Häufigkeit zu untermauern, können weitere Kriterien an die Schätzmethoden berücksichtigt werden.

Die folgenden drei Gütekriterien werden üblicherweise in Lehrbüchern zur Statistik behandelt. Sie werden hier nicht formelhaft und rigoros, sondern hinreichend anschaulich präsentiert.

4. **Erwartungstreue:**
   Die Schätzmethode hat keinen systematischen Schätzfehler. Sie liegt im Mittel richtig. Der Erwartungswert des Verfahrens ist gleich dem zu schätzenden Parameter. Im oben dargestellten Fall ist der zu schätzende Parameter die Wahrscheinlichkeit für „rot".
5. **Effizienz:**
   Die Schätzmethode soll eine möglichst kleine Streuung um den zu schätzenden Parameter besitzen oder – mit anderen Worten – eine möglichst große Präzision.
6. **Konsistenz:**
   Eine Erhöhung des Stichprobenumfangs führt regelmäßig zu einer stärkeren Annäherung des Schätzergebnisses an den gesuchten Parameter. Je mehr Beobachtungen also in das Ergebnis einfließen, umso verlässlicher ist der ermittelte Schätzwert.

Die obige Alternative 1 ist nicht erwartungstreu, sondern liefert systematisch zu große Schätzwerte. Auch ihre Streuung ist im Vergleich zur relativen Häufigkeit größer, damit schätzt Alternative 1 weniger präzise. Insofern zeigt sich, dass Alternative 1 weniger gut geeignet ist, um die gesuchte Wahrscheinlichkeit zu schätzen, auch wenn die Güteeigenschaft der Konsistenz erfüllt ist.
Dass Erwartungstreue als alleiniges Gütekriterium unzureichend ist, zeigt Alternative 2: Das zugehörige Schätzverfahren ist erwartungstreu, obwohl es schon mithilfe der ersten drei Kriterien als ungenügend identifiziert wurde. Die Nichtberücksichtigung von Information aus der Stichprobe führt dazu, dass Alternative 2 weniger effizient ist als die relative Häufigkeit. Konsistenz liegt dagegen auch hier vor.
Die relative Häufigkeit erfüllt im obigen Kontext alle sechs genannten Güteeigenschaften. Sie schlägt alle anderen Alternativen. Daher ist die relative Häufigkeit die Standardmethode zur Schätzung unbekannter Wahrscheinlichkeiten, wenn man – wie oben der Fall – im Standardmodell der schließenden Statistik unterwegs ist.
Im Standardmodell wird angenommen, dass die einzelnen Zufallsversuche stochastisch unabhängig und identisch verteilt sind, im Englischen: **independent and identically distributed** oder kurz: **i.i.d.** Die stochastische Unabhängigkeit steht dafür, dass sich die einzelnen Versuche nicht gegenseitig beeinflussen. Die identische Verteilung sorgt dafür, dass sich die Wahrscheinlichkeiten zwischen den einzelnen Versuchen nicht unterscheiden. Würde man dagegen in obigem Hörsaalexperiment die Bausteine ohne Zurücklegen ziehen, so wären beide Teile der Standardannahme nicht erfüllt.

Fall 2: Strafwurf beim Handball – Trefferwahrscheinlichkeit

Um das Standardmodell tiefergehend zu besprechen, wird ein zweiter Fall betrachtet. Gesucht ist die **Trefferwahrscheinlichkeit** des Autors beim Strafwurf im Handball (Siebenmeter). Dazu stehen die in Tabelle 1 dargestellten Daten zur Verfügung.
Sind die einzelnen Versuche unabhängig und identisch verteilt? Das kann über alle Versuche hinweg sicherlich verneint werden. Grund hierfür ist, dass die Versuche in verschiedenen Spielzeiten bei jeweils mehreren Spielen mit verschiedenen gegnerischen Torhütern bei verschiedenen Spielständen stattfanden. Gäbe es z. B. eine identische Verteilung, dann wäre die Trefferwahrscheinlichkeit in jeder Situation gleich. Das ist aus Sicht des Sportlers unrealistisch, denn die Trefferwahrscheinlichkeit hängt u. a. ab

- vom Trainings- und mentalen Zustand des Schützen,
- von der Routine des Werfers und
- vom gegnerischen Torhüter, der vielleicht als „Siebenmeterkiller" bekannt und gefürchtet ist.

Als Modell, also als vereinfachtes Abbild der Realität, können die stochastische Unabhängigkeit und die identische Verteilung der Versuche hier ggf. noch gerechtfertigt werden. Das bedarf allerdings – wie in wissenschaftlichen Arbeiten üblich – der kritischen Würdigung des Modells und seiner Annahmen. Und es sollte auch keine augenscheinliche Alternative geben, die die Realität besser abbildet.

4 *Beispielhaft seien Mosler & Schmid (2006), Auer & Rottmann (2010), Mittag & Schüller (2020) oder Kohn & Öztürk (2022) genannt.*

Saison	Spiele	Versuche	Tore	Trefferquote
2018/2019	11	15	9	60,0 %
2019/2020	16	46	38	82,6 %
2020/2021	3	7	5	71,4 %
2021/2022	4	9	8	88,9 %
Gesamt	**34**	**77**	**60**	**77,9 %**

Tabelle 1: Daten zur Anzahl der Spiele, an denen der Verfasser teilgenommen hat, zur Anzahl an Strafwürfen durch den Verfasser (=Versuche), der dabei erzielten Tore und zur jeweils resultierenden Trefferquote in den Spielzeiten 2018/19 bis 2021/22 der Handball-Ostsachsenliga Männer; Quelle: https://hvs-handball.de/ligen-pokale/archiv

Fall 3: Kreditrisikomessung in Banken – Ausfallwahrscheinlichkeit

Im dritten und letzten behandelten Fall aus der Kreditrisikomessung interessiert sich die Bank für die Ausfallwahrscheinlichkeit eines Kreditnehmers, d. h. für die Wahrscheinlichkeit, dass der Kreditnehmer seinen Zahlungsverpflichtungen gegenüber der Bank nicht nachkommen kann. Bei der Schätzung dieser **Ausfallwahrscheinlichkeit** aus beobachteten Daten ist zu beachten, dass die Standardannahmen der schließenden Statistik nicht erfüllt sind: Stochastische Unabhängigkeit und identische Verteilung liegen nicht vor.

Das beobachtete **Ausfallverhalten** eines Kreditnehmers *i* mit den zwei Zuständen

- y_i=1: Ausfall,
- y_i=0: kein Ausfall

ist abhängig von seiner Bonität. Die Bonität wiederum ist abhängig von

- systematischen Faktoren, die auf alle Kreditnehmer wirken (Konjunktur, Energiepreise) oder die sich auf Gruppen von Kreditnehmern beziehen (Region, Branche), und
- individuellen Faktoren, die nur für den einzelnen Kreditnehmer gelten (z. B. Vermögen, Managementfähigkeiten).

Formelhaft sei die **Bonität** als Funktion f dieser Faktoren wie folgt dargestellt:

$$x_i = f\text{ (systematische Faktoren, individuelle Faktoren)}$$

Die erste Aufgabe der mathematischen Statistik besteht nun darin, die Zusammenhänge zwischen Ausfallverhalten und Bonität sowie deren Einflussfaktoren zu modellieren. Die zweite Aufgabe besteht darin, geeignete Methoden zur Schätzung der Ausfallwahrscheinlichkeit bereitzustellen. Um die Eignung zu bestätigen, müssen im jeweiligen Modell die Güteeigenschaften der Schätzer untersucht werden. Beispielsweise sind dann folgende Fragen zu beantworten: Ist die relative Häufigkeit des Ausfallzustands erwartungstreu für die Ausfallwahrscheinlichkeit? Ist sie konsistent?

Wie eine entsprechende Datenlage bei zwei Perioden aussehen könnte, illustriert Abbildung 4: Ausfälle (y = 1) beobachtet man eher bei niedrigen Bonitäten; bei hoher Bonität sind Ausfälle selten, vielmehr sind in diesem rechten Bereich Nicht-Ausfälle (y = 0) vorherrschend. Zwischen den Perioden kann sich u.a. wegen konjunktureller Einflüsse die Bonitätsverteilung

- nach links in Richtung schlechter Bonitäten oder
- nach rechts in Richtung guter Bonitäten verschieben.

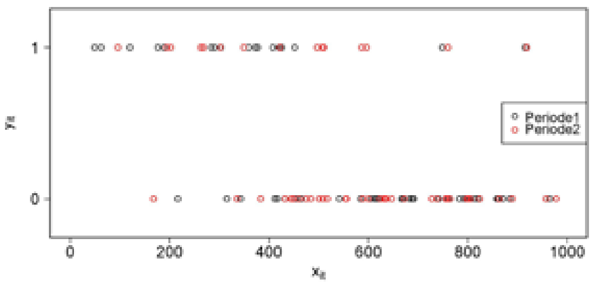

Abbildung 4: Streudiagramm mit Daten zu Bonität x_{it} und Ausfallverhalten y_{it} verschiedener Individuen *i* aus zwei Perioden *t*=1,2. In Anlehnung an Tillich (2016).

Der Modellierung des Ausfallverhaltens und der Schätzung von Ausfallwahrscheinlichkeiten haben sich neben dem Verfasser dieses Beitrags (u. a. in Tillich (2013), Tillich & Ferger (2015) und Tillich (2016)) viele weitere Autoren gewidmet (vgl. z. B. Henking et al. (2006) und McNeil et al. (2015)). Diese thematisieren auch den Einfluss der Ausfallwahrscheinlichkeit bei der Kreditportfoliomodellierung und der Kreditrisikomessung, welche in der Bankpraxis von besonderer Bedeutung sind.

Fazit

Der vorliegende Beitrag zeigt anhand dreier Beispiele aus Theorie und Praxis die hervorgehobene Rolle der relativen Häufigkeit zur Schätzung unbekannter Wahrscheinlichkeiten aus Daten. Er führt neben den drei klassischen Güteeigenschaften (Erwartungstreue, Effizienz und Konsistenz) drei weitere einfache und intuitive Güteeigenschaften ein, die in keiner einführenden Lehrveranstaltung zur Statistik fehlen sollten.

Der Beitrag stellt zudem dar, dass neben Daten insbesondere auch die Modellierung realer Sachverhalte die inhaltliche Arbeit des mathematischen Statistikers prägen. Beispielhaft wird dabei auf die Kreditrisikomessung im Risikomanagement von Banken als einem typischen Berufsfeld eingegangen. In diesem Bereich, wie auch in vielen anderen praktischen Anwendungsgebieten, muss dabei ein alternatives Modell zu der in der schließenden Statistik oft verwendeten i.i.d.-Annahme gefunden werden.

Literatur

Auer, B. & Rottmann, H., 2010. *Statistik und Ökonometrie für Wirtschaftswissenschaftler.* Wiesbaden: Gabler.

Henking, A., Bluhm, C. & Fahrmeir, L., 2006. *Kreditrisikomessung.* Berlin, Heidelberg, New York: Springer.

Kohn, W. & Öztürk, R., 2022. *Statistik für Ökonomen.* Berlin: Springer.

McNeil, A., Frey, R. & Embrechts, P., 2015. *Quantitative Risk Management.* Princeton, NJ: Princeton University Press.

Mittag, H.-J. & Schüller, K., 2020. *Statistik.* Berlin: Springer.

Mosler, K. & Schmid, F., 2006. *Wahrscheinlichkeitsrechnung und schließende Statistik.* Berlin, Heidelberg, New York: Springer.

Tillich, D., 2013. *Bruchpunktschätzung bei der Ratingklassenbildung.* Dissertation. Dresden: https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa-130581.

Tillich, D., 2016. Generalized Modeling and Estimation of Rating Classes and Default Probabilities Considering Dependencies in Cross and Longitudinal Section. *Dresdner Beiträge zu Quantitativen Verfahren,* Band 67, S. 1-30.

Tillich, D. & Ferger, D., 2015. Estimation of Rating Classes and Default Probabilities in Credit Risk Models with Dependencies. *Applied Stochastic Models in Business and Industry,* 31(6), S. 762-781.