Sensitivitätsanalysen im Statistikunterricht[1]

Lingyun Zhang und Kondaswamy Govindaraju, Palmerston North, Neuseeland

[1] Original ‚Sensitivity analysis in statistics teaching' in *Teaching Statistics* 34 (2012) 1, 38–40. Übersetzung und Bearbeitung: Manfred Borovcnik, Klagenfurt

Zusammenfassung: *In diesem Artikel wollen die Autoren zeigen, wie notwendig es ist, im Statistikunterricht „was wäre, wenn?"-Fragen zu stellen. Technisch kommen Wahrscheinlichkeitsnomogramme und die Exponentialverteilung vor.*

1 Einleitung

Wenn man mathematische Modellierung oder Entscheidungstheorie unterrichtet, so sind „was wäre, wenn"-Fragen durchaus üblich. Weniger trifft das für den Statistikunterricht zu. Sehr wenige einführende Bücher in Wahrscheinlichkeitsrechnung und Statistik thematisieren Fragen, wie sensitiv Ergebnisse auf Änderungen des Modells reagieren.

Hunt (1996) merkt an, dass Studierende wohl „was wäre, wenn"-Fragen stellen wollten, aber sich nicht trauen, weil sie glauben, sich damit vor der Klasse zu blamieren. Daher sollte man solche Fragen an die Klasse herantragen, um die Studierenden zu ermuntern, sich diesbezüglich Gedanken zu machen. Doane (2004) empfiehlt Simulationssoftware im Unterricht für die Beurteilung von Risiken, weil in diesem Rahmen solche Fragen natürlich erscheinen.

Der erste Autor gab seinen Studierenden im Bachelor-Kurs für Ingenieure als Teil der Beurteilung die Aufgabe, reale Daten zu suchen und analysieren, die einer Exponentialverteilung folgen. Eine Studentin bezog sich auf die Flugzeugunglücke in Neuseeland, welche sich in den Jahren 1948, 1954, 1963, 1979, 2001, 2003 und 2008 ereigneten. (The Dominion Post 2008). Sie berechnete die Zeit zwischen den Unglücken: 6 (= 1954–1948), 9, 16, 22, 2 und 5. Das Nomogramm in Abb. 1 zeigt, dass gegen die Annahme einer Exponentialverteilung kaum ein Einwand besteht. Die mittlere Zeit zwischen den Ereignissen (Unglücken) wird mit 10 geschätzt.

Die eingetragenen Punkte befinden sich im Band zwischen jenen beiden Modellen, die sich durch ein 95-%-Konfidenzintervall für den Parameter λ der Exponentialverteilung (siehe weiter unten) ergeben; ein weiteres Argument für das verwendete Modell. Auch der Anderson-Darling-Test kann die Hypothese der Exponentialverteilung nicht ablehnen (AD = 0,310; p-Wert 0,756). Auf der Basis dieses Modells berechnete die Studentin ihre Prognosen.

2 „Was wäre, wenn?"-Fragen

Die Studentin hat ihre Kenntnisse im Umgang mit dem Nomogramm deutlich gezeigt, den Fit der Exponentialverteilung überprüft und die Prognosen berechnet (Modell 1 in Tabelle 1).

[Die Wahrscheinlichkeit, dass sich die nächste Tragödie in Neuseeland (das nächste Ereignis) innerhalb der nächsten t Jahre ereignet, beträgt nach dem Exponentialmodell für die Zwischenankunftszeiten $F(t) = 1 - e^{-\lambda \cdot t}$. Für die erwartete Zeit zwischen den Ereignissen gilt $E = \frac{1}{\lambda}$. Für $t = 5$ ergibt sich daher für Modell 1 mit der Schätzung des Parameters $\hat{\lambda} = \frac{1}{\bar{x}} = \frac{1}{10}$ eine Wahrscheinlichkeit für das nächste Ereignis innerhalb der nächsten fünf Jahre von $F(t) = 1 - e^{-5/10} \approx 0{,}393$.

Der Beitrag wurde 2009 veröffentlicht, so meint $t = 5$, dass sich das Ereignis vor dem Ende von 2013 ereignet; Jahreszahlen eignen sich besser, um Eindruck zu erwecken; Wartezeit in Jahren ist dagegen übersichtlicher. Gegenüber dem Original wurden die Prognosezeiten von 4, 9, … auf 5, 10, … Jahre abgeändert, weil sich die entsprechenden Wahrscheinlichkeiten aus Abb. 1 leichter ablesen lassen. Für ein Nomogramm siehe Borovcnik (o. D.)]

Jedoch wird gleichzeitig ein Mangel an statistischem Denken beim Erstellen der Prognosen sichtbar. Es sind ja nur sechs Daten und die Variabilität, die dadurch verursacht ist, dass der Parameter nur geschätzt ist, wurde nicht thematisiert. Konfidenz- und Vorhersageintervalle für die Exponentialverteilung waren im Kurs nicht eigens behandelt worden.

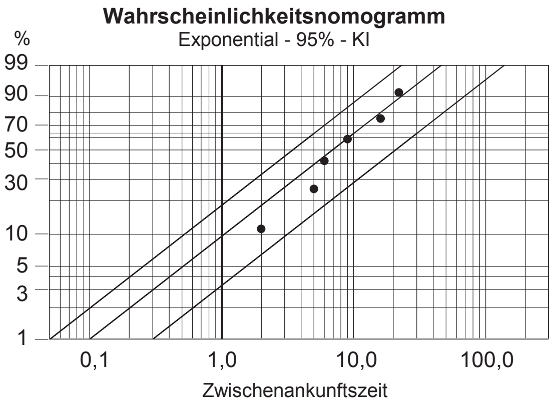

Abb. 1: Nomogramm für die Exponentialverteilung – Daten der Neuseeland-Flugunglücke

Dennoch können und sollten eine Reihe von „was wäre wenn"-Fragen, die sich indirekt mit Konfidenzintervallen befassen, behandelt werden. Solche Fragen sind im Kontext der Unglücke etwa:

Frage (a)

Was wäre, wenn die „wahre" mittlere Zeit zwischen den Ereignissen zwischen 5 und 30 Jahren läge?

Das sind ungefähr die Werte, die sich aus einem exakten 95-%-Konfidenzintervall ergeben:

$$\left(\frac{2n\,\overline{X}}{\chi^2_{0{,}975,\,2n}}, \frac{2n\,\overline{X}}{\chi^2_{0{,}025,\,2n}}\right) = (5 \text{ Jahre}, 27 \text{ Jahre}).$$

Frage (b)

Was wäre, wenn wir einen größeren Datensatz hätten, sagen wir 20 Daten, und der Mittelwert immer noch 10 betrüge?

Frage (c)

Was wäre, wenn der Fortschritt in der Technik den Flugverkehr sicherer macht und die Zeit zwischen den Unglücken dadurch auf 15 Jahre anstiege?

Die Fragen thematisieren Stichprobenvariabilität, Stichprobenumfang und Präzision der Information aus den Daten und Auswirkungen der Änderung von Parametern der Verteilung in der Population.

Frage (a) wird in Tabelle 1 mit den Modellen 2 und 3 beantwortet. Man beachte, dass sich dadurch anstelle einzelner Zahlen *Intervalle* für die gesuchten Prognosen ergeben.

Für Frage (b) beachte man, dass die berechneten Konfidenzintervalle kürzer werden, wenn mehr Daten vorliegen. Bei 20 Daten und demselben Mittelwert ergibt das 95-%-Konfidenzintervall von oben 7 bis 16 Jahre. Wir könnten die Analyse mit modifizierten Modellen 2 und 3 fortsetzen und daraus Intervalle für die Prognosen errechnen.

Für Frage (c) unterstellen wir, dass sich die Verbesserungen auf die mittlere Zeit zwischen den Unglücken so auswirken, dass sie nun 15 Jahre betrage. Die Frage an sich ist wertvoll, weil sie uns daran erinnert, dass sich die Welt verändert (hoffentlich zum Besseren!) und wir daher bei langfristigen Prognosen immer darauf Bedacht nehmen müssen.

In vielen realen Situationen werden die Zwischenankunftszeiten von einem nicht-homogenen Poisson-Prozess erzeugt, während wir im Modell mit unserer Schätzung davon ausgegangen sind, dass die Wartezeiten auf das nächste Ereignis über 60 Jahre hinweg unverändert geblieben sind (also einem homogenen Poisson-Prozess entstammen).

		www – „was wäre, wenn …?"		
	1. Analyse	F r a g e (a)		Frage (c)
	Modell 1	Modell 2	Modell 3	Modell 4
	Parameter	Worst case	Best case	Plausibel
	aus Daten geschätzt	Unterer Rand des Konfidenzint.	Oberer	Szenario
E	10,00	30,00	5,00	15,00
λ	0,100	0,033	0,200	0,067
Wahrscheinlichkeit fürs nächste Ereignis *innerhalb von …*				
Jahren	Modell 1	Modell 2	Modell 3	Modell 4
5	0,393	0,154	0,632	0,283
10	0,632	0,283	0,865	0,487
20	0,865	0,487	0,982	0,736
30	0,950	0,632	0,998	0,865

Tab. 1: Zeit bis zum nächsten Unglück mit verschiedenen Exponentialmodellen

Mit jeder der „was wäre, wenn"-Fragen werden die Studierenden einsehen, dass die Wahrscheinlichkeit des nächsten Unglücks bis vor dem Ende eines gegebenen Jahrs nur sehr ungenau angegeben werden kann. Die Auswirkung einer kleinen Stichprobe auf die Schätzung des oberen Ausläufers einer Verteilung ist viel größer als auf die Schätzung des Mittelwerts. Sensitivitätsanalysen, wie sehr sich solche Ungenauigkeiten im Modell auf das Ergebnis auswirken, sind ein nützliches Werkzeug, statistisches Denken anzuregen, das leider nur wenig Aufmerksamkeit im Statistikunterricht gefunden hat.

3 Schlussfolgerungen

Ein geeignetes statistisches Modell zu identifizieren ist schwierig. So etwa sind bei einer Stichprobe von 50 die Chancen, eine stetige Gleichverteilung oder eine Betaverteilung fälschlich als Normalverteilung zu klassifizieren sehr groß. Modellvalidierungsfragen tauchen auf jedem Niveau des Statistikunterrichts auf und können teilweise mit geeigneten „was wäre, wenn"-Fragen thematisiert werden.

4 Modellvalidierung und Wahrscheinlichkeitsdiagramme

[Für die Modellvalidierung sind auch Wahrscheinlichkeits- und Q-Q-Diagramme geeignet; erstere wurden im Beitrag auch eingesetzt. Wahrscheinlichkeitsplots zeichnen die Verteilungsfunktion in einem Koordinatensystem, das sie als Gerade darstellt. Dazu sind Skalentransformationen notwendig. Techniker müssen mit solchen Nomogrammen umgehen können.

Nicht nur die visuelle Prüfung der empirischen Verteilungsfunktion wird dadurch ermöglicht (die Punkte müssen, wie in Abb. 1, ungefähr auf einer Geraden liegen). Man kann alle Wahrscheinlichkeiten aus diesem Nomogramm ablesen. Ein zusätzlicher Test, ob das verwendete Modell – die Exponentialverteilung – passt, ergibt sich so: Verschiedene Parameter verschieben die Gerade. Wenn man nun für den Parameter der Exponentialverteilung ein Konfidenzintervall berechnet, so kann man eine worst- und best-case-Gerade einzeichnen und prüfen, ob die Punkte der empirischen Verteilung dazwischen liegen.

Das ist besonders wichtig, weil eigentlich alle Modelle „falsch" sind und eine Überprüfung, ob ein unterstelltes Modell doch ausreicht, um ein Phänomen zu beschreiben (weil es ganz brauchbar an die Daten angepasst ist), eigentlich ein statistischer Test ist, der darauf hinaus läuft, die Nullhypothese zu bestätigen – was ja methodisch gar nicht geht.

Hier wird zusätzlich der Anderson-Darling-Test eingesetzt. Details kann man in Nist/Sematech (o. D.) nachlesen. Es ist ein modifizierter Kolmogorow-Smirnow-Test, der statt dem gewöhnlichen „supremalen" Abstand zwischen empirischer und theoretischer Verteilungsfunktion einen gewichteten Abstand heranzieht, der die Ränder der Verteilung besonders hoch gewichtet. Damit erhält dieser Test eine besondere Sensitivität (Macht), Alternativen zur getesteten Null-Verteilung zu erkennen, die sich in den Ausläufern unterscheiden. Und Lebensdauerverteilungen oder Zwischenankunftsverteilungen sind ganz besonders dadurch gekennzeichnet, wie sie extreme Ereignisse modellieren.

Daher ist es nicht eine „Attitüde" der Statistiker, immer neue Tests zu erfinden, wo doch das Problem mit dem Kolmogorow-Smirnow-Test schon erledigt ist, sondern eine Antwort auf die schwierige Frage der Modellevaluation.

Gerade in der Zuverlässigkeitsanalyse (der Analyse von technischen Risiken) ist die Überprüfung von Modellen anhand solcher Nomogramme wichtig, weil die Antworten sehr sensitiv darauf reagieren, wie man die Ausläufer („tails") einer Verteilung modelliert. Der verwendete Anderson-Darling-Test hat durch die „Vergrößerung" der Abweichungen in den Ausläufern auch schon bei sechs Daten eine brauchbare Chance (statistische Macht), alternative Modelle anzuzeigen.

Mit dem Aufkommen der explorativen Datenanalyse sind auch so genannte Q-Q-plots modern geworden, dann aber auch wieder aus dem Blickfeld verschwunden. Diese Diagramme zeigen den Zusammenhang zwischen den empirischen und theoretischen Quantilen. Wenn die Verteilung passen soll, ist dieser Zusammenhang annähernd linear. (Die Diagramme zeigen gewissermaßen beobachtete und erwartete Häufigkeiten.) Zur Orientierung, ob eine Datentransformation eine Schiefe herausnehmen kann (und eine bessere Anpassung an die Normalverteilung erreicht), ist ein Quantil-Diagramm gut zu gebrauchen.

Schiefe Verteilungen machen immer ein Problem bezüglich der Aussagefähigkeit von statistischen Tests: der berechnete obere und untere Ablehnungsbereich reagiert sehr sensitiv darauf, ob die unterstellte Symmetrie der Verteilung des untersuchten Merkmals durch Schiefe verletzt wird. (Bei allen üblichen Tests wie Gauß-, *t*-Test, der Varianzanalyse, auch in den nicht-parametrischen Alternativen wie dem Wilcoxon-Test basiert die Berechnung auf symmetrischen Modellen).

Bei Wartezeiten zwischen Ereignissen ergeben sich schiefe Verteilungen. Für die Methoden siehe Nist/Semantech (o. D.) oder ReliaWiki (o. D.). Diese sind besonders in der Zuverlässigkeitstheorie eminent wichtig und dienen zur Bewertung von Risiken.]

Literatur

Borovcnik, M. (o. D.): Wahrscheinlichkeitsnetz zur Anpassung an die Exponentialverteilumg mit logarithmisch unterteilter Merkmalsachse. wwwg.uni-klu.ac.at/stochastik.schule/Boro/index_inhalt.

Fisher, N. I.; Nair, V. N. (2009): Quality management and quality practice: Perspectives on their history and their future. In: *Applied Stochastic Models in Business and Industry* 25(1), S. 1–28, DOI: 10.1002/asmb.756.

Doane, D. P. (2004): Using simulation to teach distributions. In: *Journal of Statistics Education* 12(1). www.amstat.org/publications/jse/v12n1/doane.html. (Zugriff: 8.8.2014).

Hunt, N. (1996): Teaching statistical concepts using spreadsheets. In: *Teaching Statistics* 17 (issue supplement s1), S. 1–3.

Nichols, L.; Chalmers, A.; Broun, B. (2008): Search for answers. The Dominion Post, Nov. 29, 2008. A1.

Nist/Sematech (o. D.): e-Handbook of Statistical Methods. Probability plotting. www.itl.nist.gov/div898/handbook/apr/section2/apr221.htm (Zugriff: 8.8.2014).

ReliaWiki (o. D.): The Exponential Distribution. reliawiki.org/index.php/The_Exponential_Distribution (Zugriff: 8.8.2014).

Anschrift der Verfasser

Lingyun Zhang und Kondaswamy Govindaraju
Massey University
Science Tower, Turitea
Palmerston North, 4474, Neuseeland
L.Y.Zhang@massey.ac.nz
K.Govindaraju@massey.ac.nz