

Ilmenauer Beiträge zur Wirtschaftsinformatik

Herausgegeben von U. Bankhofer, V. Nissen
D. Stelzer und S. Straßburger

Dieter W. Joensen

**Zur Güte von Signifikanztests auf
Benfordverteilung unter Nutzung der ersten
zwei Ziffern**

Arbeitsbericht Nr. 05, Dezember 2013



Technische Universität Ilmenau
Fakultät für Wirtschaftswissenschaften
Institut für Wirtschaftsinformatik

Autor: Dieter W. Joensen

Titel: Zur Güte von Signifikanztests auf Benfordverteilung unter Nutzung der ersten zwei Ziffern

Ilmenauer Beiträge zur Wirtschaftsinformatik Nr. 2013-05, Technische Universität Ilmenau, Dezember 2013

ISSN 1861-9223

ISBN 978-3-938940-50-1

URN urn:nbn:de:gbv:ilm1-2013200249

© 2013 Institut für Wirtschaftsinformatik, TU Ilmenau

Anschrift: Technische Universität Ilmenau, Fakultät für Wirtschaftswissenschaften, Institut für Wirtschaftsinformatik, PF 100565, D-98684 Ilmenau.
<http://www.db-thueringen.de/servlets/DocumentServlet?id=5507>

Gliederung

Gliederung.....	iii
Abbildungsverzeichnis.....	iv
Tabellenverzeichnis.....	v
1 Einführung.....	1
2 Ein Monte Carlo Vergleich der Vorgehensweisen.....	4
2.1 Ausgewählte Tests.....	4
2.2 Resultate.....	5
3 Schlussfolgerungen.....	8
Literaturverzeichnis.....	9

Abbildungsverzeichnis

Abbildung 1: Abweichungen in der Güte für vier Anpassungstests 6

Tabellenverzeichnis

Tabelle 1: Gemeinsame Verteilung der ersten zwei Ziffern mit Randverteilungen (in %)..... 3

Tabelle 2: Empirische Perzentile für $\alpha = 5\%$, mittels 1000000 Stichproben berechnet..... 4

Zusammenfassung: Das benfordsche Gesetz wird nicht nur im Bereich der Wirtschafts- und Steuerprüfung, sondern auch bei Wahlen verwendet, um Datenmanipulation aufzudecken. Der prinzipielle Gedanke hinter Benfords Gesetz ist, dass die ersten Ziffern von Zahlen, die aus bestimmten Prozessen stammen, einer als Benfords Verteilung bekannten diskreten Verteilung entsprechen. Operativ werden statistische Anpassungstests verwendet, um zu prüfen, ob die ersten Ziffern von vorliegenden Daten dieser diskreten Verteilung entsprechen. Wenn die Daten Benfords Gesetz folgen sollten, legt eine Verwerfung der Null-Hypothese nahe, dass eine Manipulation der Daten stattgefunden hat. Jedoch tritt bei Anpassungstests, wie bei allen Signifikanztests, nicht nur der Fehler erster Art, welcher durch das gewählte Signifikanzniveau begrenzt wird, sondern auch der Fehler zweiter Art auf. Dieser Fehler sinkt nicht nur mit steigender Stichprobengröße, sondern ist von Natur aus für manche Testverfahren kleiner als für andere. Eine mögliche Variante ist es, nicht nur die erste Ziffer, sondern die gemeinsame Verteilung der ersten beiden Ziffern zu prüfen. Der Zuwachs an Güte wäre auf eine stärkere Nutzung vorhandener Informationen zurückzuführen, da unter der Null-Hypothese die Verteilungen der ersten und zweiten Ziffern nicht unabhängig sind. Dieser Aufsatz beschreibt, wie vier Anpassungstests erweitert werden können, um die gemeinsame Verteilung der ersten und zweiten Ziffer zu prüfen. Zusätzlich wird die Güte des ursprünglichen Ansatzes (Prüfen der ersten Ziffer) mit der Güte des neuen Ansatzes (Prüfen beider Ziffern) verglichen.

Abstract: Benford's law has been used by auditors to help reveal data manipulation not only in the context of tax audits and corporate accounting, but also election fraud. The principle idea behind Benford's law is that the frequency distribution of the first digits from numerical data of certain processes should conform to a discrete distribution known as the Benford distribution. Goodness-of-fit tests have been used to assess if the data's first digits conform to said distribution. When data should conform to Benford's law, a null-hypothesis rejection suggests that some form of data manipulation has taken place. Goodness-of-fit tests, like all tests of statistical significance, are prone not only to the type I error, which is limited by the chosen level of significance, but also to the type II error which decreases not only with sample size but is also inherently lower for some testing procedures than others. One possible procedural change is not to test the distribution of the data's first digit, as is the current standard, but to test the joint distribution of the data's first two digits. The gain in power would be due to an increased utilization of information, because, given that the null hypothesis is true, the distributions of the first and second digits are not independent. This paper describes how four goodness-of-fit tests can be extended to test the joint distribution of the first and second digit for conformity to Benford's law. Additionally, a comparison of power yielded by the original (one-digit) as well as the proposed (two-digit) analysis is provided.

Schlüsselworte: Benfordverteilung, Gesetz von Benford, Anpassungstests, Güte

1 Einführung

Wirtschaftsprüfer verwenden Benfords Gesetz zur Aufdeckung von Datenmanipulation, indem sie die ersten Ziffern von Daten auf Abweichungen von der sogenannten Benfordverteilung prüfen. Dieses Gesetz der ungewöhnlichen Zahlen, wie Frank Benford (1938) es nannte, bestimmt, dass der Logarithmus der ersten Ziffern von bestimmten Daten gleichverteilt ist. In anderen Worten, eine zufällige, stetige Variable X mit der Realisierung x , folgt Benfords Gesetz, wenn die Anwendung der ersten- k -Ziffern Funktion, $D_k(x)$, eine Benford-verteilte diskrete Variable erzeugt, d.h.

$$p_{d_k} = P(D_k(X) = d_k) = \log_{10}(1 + d_k^{-1}) \forall k \in \mathbb{N}^+$$

mit

$$D_k(x) = \lfloor |x| \cdot 10^{-1 \cdot \lceil \log_{10}|x| + k - 1} \rfloor$$

und $d_k \in \{10^{k-1}, 10^{k-1} + 1, \dots, 10^k - 1\}$ (vgl. Hill, 1995, S. 354).

Zwar zeigten Leemis et al. (2000), dass diese Wahrscheinlichkeiten für etliche theoretischen Verteilungen gelten, aber wahrlich interessant ist es, dass viele echte Datensätze auch Benfords Gesetz folgen (für Beispiele siehe Hill (1995, S. 355)). Da Daten, die aus einem geometrischen Wachstumsprozess kommen, per Definition, dem benfordschen Gesetz folgen, ist dessen Verwendung, um Betrug dort aufzudecken, wo exponentielles Wachstum herrscht, natürlich. Dies erklärt die weite Verbreitung bei Steuer- und Wirtschaftsprüfung (vgl. Nigrini, 1996; Nigrini and Mittermaier, 1997; Swanson et al., 2003; Watrin et al., 2008; Rauch et al., 2011).

Wurde im Vornherein festgestellt, dass Benfords Gesetz für die unmanipulierten Daten wahr sein muss, so ist die Verwendung von Anpassungstests und die Interpretation der Resultate einfach. Mittels eines statistischen Tests werden die beobachteten relativen Häufigkeiten der ersten Ziffern auf Übereinstimmung zu den erwarteten relativen Häufigkeiten, p_{d_k} , geprüft. Wird die Null-Hypothese verworfen, wird angenommen, dass eine Datenmanipulation stattgefunden hat, und eine stringenter Prüfung der Daten wird vorgenommen.

Wie bei jeden Signifikanztest gibt das gewählte Signifikanzniveau den maximal zu erwartenden Anteil an falschen Positiven an. Daher stellt das gewählte Signifikanzniveau den Anteil an Datensätzen dar, die einer strengeren Überprüfung unterzogen werden, obwohl keine Datenmanipulation festgestellt werden sollte. In diesem Sinne quantifiziert der α -Fehler

den Teil der Kosten, die aufgrund der Überprüfung der Daten, die nicht manipuliert wurden, aufgewendet werden müssen. Der β -Fehler quantifiziert die Menge an Datensätzen, die manipuliert sind, aber keiner weiteren Überprüfung unterzogen werden. Demzufolge sind Anpassungstest mit einer höheren Güte in der Erkennung von Betrug effizienter.

Möglichkeiten zur Reduktion des Anteils der falschen Negativen bei statistischem Testen beinhalten eine Erhöhung der Stichprobengröße, eine Reduzierung des Konfidenzniveaus oder die Wahl einer besseren Vorgehensweise. Ceteris paribus steigt die Güte aller statistischen Tests monoton mit der Stichprobengröße, aber die Größe mancher Datensätze lässt sich wegen prohibitiv hoher Kosten nicht erhöhen. Auch eine Reduktion des Konfidenzniveaus kann den β -Fehler senken, wenn die Ressourcen, die für eine gründlichere Prüfung notwendig sind, ausreichen, um die erhöhte Menge an Datensätzen zu bearbeiten, bei denen keine Manipulation stattgefunden hat. Eine verbesserte Vorgehensweise kann beinhalten, einen Test zu verwenden, der eine höhere inhärente Macht hat, oder mehr Informationen aus den vorhandenen Daten zu verwenden. Sicherlich können neue, mächtigere Tests speziell für die Benford Verteilung entwickelt werden. Jedoch kann die Güte aller bereits existierenden Tests, durch die Verwendung von mehr Informationen, die in den vorliegenden Daten vorhanden sind, erhöht werden.

Normalerweise wird zum Testen lediglich die erste Ziffer verwendet (vgl. Nigrini, 1996; Nigrini and Mittermaier, 1997; Leemis et al., 2000; Cho and Gaines, 2007; Morrow, 2010; Rauch et al., 2011). Diekmann (2007) schlägt jedoch vor, dass die Entscheidung, ob Daten dem benfordschen Gesetz entsprechen, nicht die erste Ziffer, sondern die zweite Ziffer berücksichtigen sollte. Die Wahrscheinlichkeiten, dass die zweite Ziffer einen bestimmten Wert annimmt, sind eine Randverteilung von Benfords Vorhersage für die Verteilung der ersten zwei Ziffern (vgl. Tabelle 1), und können durch einfache Summation bestimmt werden. Basierend auf den Resultaten eines Experiments argumentiert Diekmann (2007), dass bei manipulierten Daten die Abweichungen zwischen den erwarteten und beobachteten Häufigkeiten bei der zweiten Ziffer größer sind als bei der ersten. Festzuhalten bleibt jedoch, dass diese Empfehlung lediglich andere und nicht zusätzliche Informationen in die Analyse mit einbezieht. Bei Beschränkung der Analyse auf die zweite Ziffer kann keine Manipulation erkannt werden, die sich lediglich in der ersten Ziffer äußert. Dieses Problem könnte umgangen werden, indem die Häufigkeiten der ersten und zweiten Ziffern unabhängig voneinander mit einem angepassten, gemeinsamen Signifikanzniveau geprüft werden; aber die erwarteten Häufigkeiten der ersten und zweiten Ziffern sind nicht unabhängig. Dies wurde

bereits von Hill (1995, S. 355) gezeigt und kann mittels der Tabelle 1 berechnet werden. Das Testen beider Ziffern auf diese Art und Weise, ob unter der Verwendung einer Bonferroni oder Sidák Anpassung des Signifikanzniveaus, wird immer zu einem konservativen, und daher weniger mächtigen Test führen, obwohl mehr Information verwendet wird. Daher sollte das Testen der gemeinsamen Verteilung der ersten beiden Ziffern zu einer höheren Güte für aktuell verfügbare Anpassungstests für diskrete Verteilungen führen.

		Zweite Ziffer										$\approx \Sigma$
		0	1	2	3	4	5	6	7	8	9	
Erste Ziffer	1	4,14	3,78	3,48	3,22	3,00	2,80	2,63	2,48	2,35	2,23	30,10
	2	2,12	2,02	1,93	1,85	1,77	1,70	1,64	1,58	1,52	1,47	17,61
	3	1,42	1,38	1,34	1,30	1,26	1,22	1,19	1,16	1,13	1,10	12,49
	4	1,07	1,05	1,02	1,00	0,98	0,95	0,93	0,91	0,90	0,88	9,69
	5	0,86	0,84	0,83	0,81	0,80	0,78	0,77	0,76	0,74	0,73	7,92
	6	0,72	0,71	0,69	0,68	0,67	0,66	0,65	0,64	0,63	0,62	6,69
	7	0,62	0,61	0,60	0,59	0,58	0,58	0,57	0,56	0,55	0,55	5,80
	8	0,54	0,53	0,53	0,52	0,51	0,51	0,50	0,50	0,49	0,49	5,12
	9	0,48	0,47	0,47	0,46	0,46	0,45	0,45	0,45	0,44	0,44	4,58
	$\approx \Sigma$	11,97	11,39	10,88	10,43	10,03	9,67	9,34	9,04	8,76	8,50	

Tabelle 1: Gemeinsame Verteilung der ersten zwei Ziffern mit Randverteilungen (in %)

Abschnitt 2 beschreibt einen Vergleich zwischen der neuen und der konventionellen Vorgehensweise, die lediglich eine Ziffer betrachtet. Die Differenzen in Güte werden für vier Tests und fünf Alternativverteilungen verglichen. Abschnitt 3 fasst die Resultate zusammen, bewertet die vorgeschlagene Vorgehensweise kritisch und diskutiert mögliche zukünftige Forschung.

2 Ein Monte Carlo Vergleich der Vorgehensweisen

Im folgenden Abschnitt werden die Unterscheide ermittelt, die sich ergeben, wenn nicht nur die erste Ziffer überprüft wird, sondern die ersten beiden Ziffern gemeinsam überprüft werden. Die erforderliche Monte Carlo Simulation wurde dem Framework für statistische Berechnungen R (R Core Team, 2013) implementiert. Abschnitt 2.1 beschreibt kurz die vier Anpassungstests und die fünf alternativen Verteilungen, für die der Methodenvergleich durchgeführt wird, sowie weitere Details der Monte Carlo Simulation. Die Resultate der Simulation werden in Abschnitt 2.2 präsentiert.

2.1 Ausgewählte Tests

Vier Anpassungstests wurden zum Gütevergleich ausgewählt. Diese vier Tests sind der Pearson χ^2 -Test (Pearson, 1900), der Kolmogorov-Smirnov Test (Kolmogorov, 1933), die Freedman modification des Watson U_n^2 für diskrete Verteilungen (Freedman, 1981) und der J_P^2 Korrelationstest, eine Art Shapiro-Francia Test (Shapiro and Francia, 1972). Alle Tests sind im R-Paket **BenfordTests** (Joenssen, 2013) verfügbar. Getestet wurden die Alternativverteilungen zu einem Signifikanzniveau von $\alpha = 5\%$. Die hierfür erforderlichen kritischen Werte wurden mittels einer Million Datensätze für jeden der zwölf betrachteten Stichprobengrößen bestimmt (vgl. Tabelle 2).

Test Ziffern	χ^2		K-S		U_n^2		J_P^2	
	1	2	1	2	1	2	1	2
25	15,643	120,752	1,142	1,287	0,177	0,173	0,222	0,024
50	15,493	116,921	1,139	1,285	0,178	0,173	0,515	0,090
75	15,504	115,502	1,146	1,285	0,178	0,174	0,659	0,158
100	15,492	114,684	1,146	1,288	0,178	0,173	0,740	0,221
250	15,480	113,042	1,147	1,286	0,178	0,174	0,894	0,478
500	15,480	112,537	1,146	1,287	0,178	0,174	0,947	0,669
750	15,465	112,398	1,147	1,288	0,179	0,174	0,964	0,759
1000	15,491	112,261	1,148	1,289	0,178	0,174	0,973	0,810
2500	15,501	112,125	1,148	1,287	0,179	0,174	0,989	0,916
5000	15,489	112,093	1,147	1,289	0,178	0,174	0,994	0,956
7500	15,504	112,126	1,148	1,288	0,178	0,174	0,996	0,970
10000	15,485	112,097	1,146	1,288	0,178	0,174	0,996	0,977

Tabelle 2: Empirische Perzentile für $\alpha = 5\%$, mittels 1000000 Stichproben berechnet

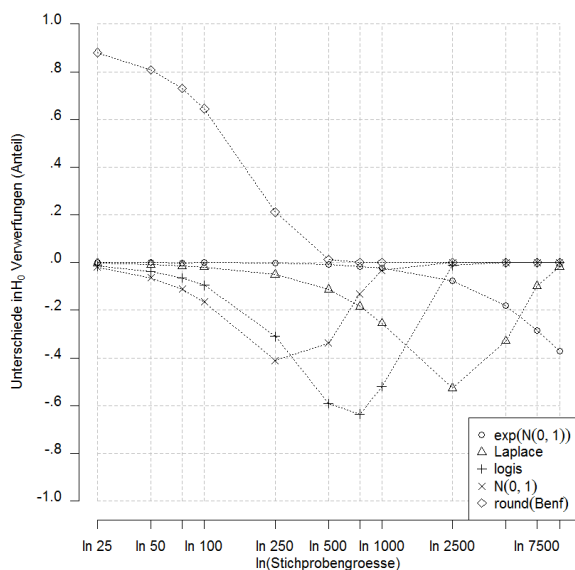
Die entsprechenden, vorberechneten kritischen Werte wurden verwendet, um die fünf Alternativverteilungen zu testen. Getestet wurden die standard Normal, Log-Normal, und

Laplace Verteilungen sowie eine weitere Verteilung. Diese weitere Verteilung, fortan gerundete Benfordverteilung genannt, ist so konstruiert, dass ein kaufmännisches Runden zur Beeinflussung der ersten oder zweiten Ziffer führt. Wird nur die erste Ziffer geprüft, wird so gerundet, dass diese beeinflusst wird (bsp. Rundung von 6,66 auf 7 und 3,14 auf 3). Werden die erste und zweite Ziffer geprüft, wird so gerundet, dass die zweite Ziffer beeinflusst wird (bsp. Rundung von 6,66 auf 6,7 und 3,14 auf 3,1). Alle Verteilungen werden für Stichprobengrößen von 25, 50, 75, 100, 250, 500, 750, 1000, 2500, 5000, 7500 und 10000 verglichen. Null-Hypothesenverwerfungen für jede Kombination von Anzahl der betrachteten Ziffern, Stichprobengröße, Alternativverteilung und Anpassungstest wurden für je 1000000 Datensätze gezählt. Die gezählten Null-Hypothesenverwerfungen werden zwischen den zwei Vorgehensweisen verglichen.

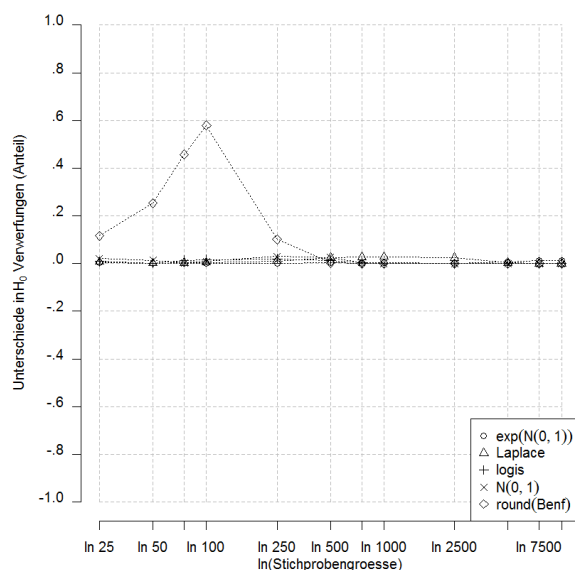
2.2 Resultate

Die Resultate werden im folgenden Abschnitt für jeden Test einzeln diskutiert, indem auf die Abbildungen 1a bis 1d Bezug genommen wird. Die Abbildungen zeigen die Differenzen in Güte zwischen der neuen Vorgehensweise, die die zwei ersten Ziffern berücksichtigt, und der alten die lediglich die erste Ziffer berücksichtigt. Positive Werte deuten daher darauf hin, dass die neue Vorgehensweise überlegen ist, während negative Werte bedeuten, dass lediglich die erste Ziffer zu testen dominant ist.

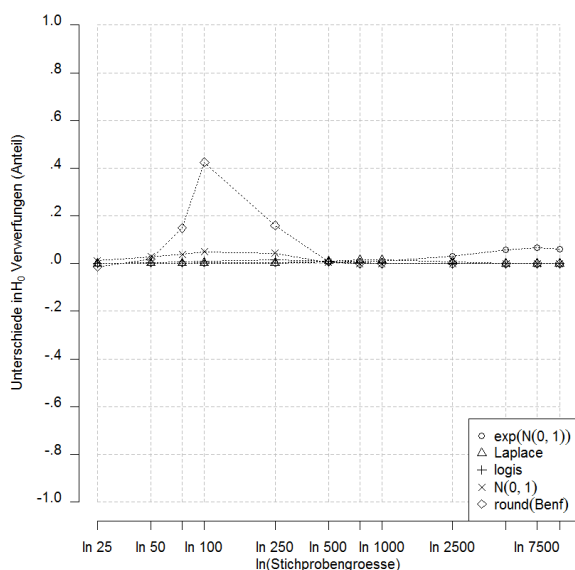
Abbildung 1a zeigt die Resultate für den χ^2 -Test. Für alle alternativen Verteilungen, mit Ausnahme der gerundeten-Benfordverteilung, ist die alte Vorgehensweise überlegen. Für kleine Stichprobengrößen sind die Unterschiede klein, da die Güte dieses Tests im Allgemeinen gering ist. So wie die Stichprobengröße erhöht wird, so steigt auch der Vorteil, nur die erste Ziffer zu betrachten. Für große Stichprobengrößen verschwinden die Unterschiede, da beide Vorgehensweisen auf 100% Güte konvergieren. Vorteile für das Testen lediglich der ersten Ziffer sind daher für die mittleren Stichprobengrößen am stärksten. Abweichend von diesen Ausführungen verhält sich der χ^2 -Test bei der gerundeten Benfordverteilung. Bei dieser Verteilung ist die Überprüfung der ersten beiden Ziffern, insbesondere für kleine Stichprobengrößen, überlegen. Da jedoch beide Verfahren auch für diese Alternativverteilung zu größtmöglicher Macht tendieren, tendiert der Vorteil der neuen Methode für große Stichprobengrößen gegen Null. Dieses letzte Ergebnis ist besonders überraschend, da es bekannt ist, dass die Güte des χ^2 -Test wesentlich von der Anzahl der verwendeten Klassen abhängt.



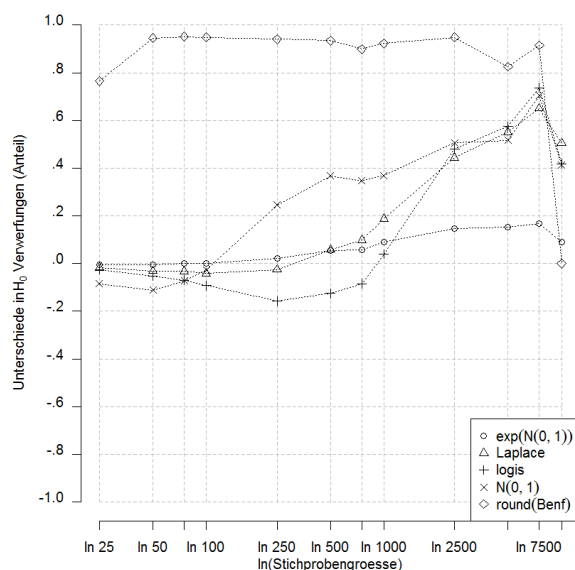
(a) Pearson χ^2 -Test



(b) Kolmogorov-Smirnov Test



(c) U_n^2 -Test



(d) J_p^2 Korrelationstest

Abbildung 1: Abweichungen in der Güte für vier Anpassungstests

Bei der Betrachtung der Abbildung 1b fällt zunächst auf, dass die Überprüfung der ersten beiden Ziffern niemals eine schlechtere Vorgehensweise darstellt, wenn der Kolmogorov-Smirnov Test verwendet wird. Zwar ist der Vorteil meist von 1 bis 3 Prozentpunkten klein, aber dennoch bei der verwendeten Anzahl an Wiederholungen statistisch signifikant. Jedoch ist auch hier die gerundete Benfordverteilung die Ausnahme. Der Vorteil, die gemeinsame

Verteilung der ersten beiden Ziffern zu betrachten, kann bei kleinen Stichprobengrößen bis zu 60 Prozentpunkte betragen.

Die Resultate von Freedmans Erweiterung des Watson U_n^2 zeigen ein ähnliches Verhalten wie der vorherige Test. Deutlich in der Abbildung 1c zu erkennen ist, dass die Differenzen in Güte niemals weniger als Null sind und daher die ersten zwei Ziffern zu prüfen, deutlich die bessere Strategie darstellt. Die Vorteile sind zwar mit maximal 5 Prozentpunkten für die meisten Verteilungen eher verhalten, aber wieder zeigen sich größere Vorteile bei der Verwendung des zwei-Ziffer-Ansatzes wenn die gerundete Benfordverteilung betrachtet wird. Hier erreichen die Unterschiede bei mittleren Stichprobengrößen bis zu 40 Prozentpunkte.

Das heterogenste Verhalten ist bei dem J_p^2 -Korrelationstest zu erkennen (vgl. Abbildung 1d). Bei diesem Test ist die Prüfung von lediglich der ersten Ziffer nur für kleinere Stichprobengrößen besser. Bei größeren Stichprobengrößen ist wiederum die neue Vorgehensweise mächtiger. Welche Vorgehensweise besser ist, unterscheidet sich für die Alternativverteilungen. Bei der logistischen Verteilung ändert sich, dass nur die erste Ziffer zu prüfen besser ist, bei einer Stichprobengröße von 1000, bei der Laplace-Verteilung bei 300 und bei der Normalverteilung bei ungefähr 100. Bei der Log-Normalverteilung und der gerundeten Benfordverteilung ist es wiederum immer besser, die beiden ersten Ziffern zu betrachten. Unterschiede können mit bis zu 20 Prozentpunkte Vorteil für die konventionelle Vorgehensweise, und bis zu 95 Prozentpunkte Vorteil für das Testen der ersten beiden Ziffern, variieren.

3 Schlussfolgerungen

Die Vermutung, dass die Verwendung von mehr vorhandener Information zu einer höheren Güte bei Anpassungstests für das benfordsche Gesetz führen kann, hat sich bestätigt. Während die vorgeschlagene Vorgehensweise nicht universal, d.h. unabhängig vom verwendeten Test, besser ist als die konventionelle Methode, ist die neue Methode mindestens für eine Alternativverteilung bei jedem Test überlegen. Bei zwei der Tests, nämlich dem Kolmogorov-Smirnov und dem Freedman-Watson Test, konnte für jede betrachtete Situation eine Steigerung der Güte durch den Einbezug der zweiten Ziffer erlangt werden. Zwar sind die Vorteile mit erwarteten 4 bis 5% klein, aber diese Vorteile können in manchen Zusammenhängen, beispielsweise bei Steuerprüfungen, zu deutlichen Zuwächsen in Effizienz führen. Ferner ist anzumerken, dass diese Vorteile durch die Nutzung des in dem R-Paket **BenfordTests** (Joenssen, 2013) bereitgestellten Programmcode leicht zu erzielen sind. Die Resultate für den χ^2 -Test und den J_p^2 -Korrelationstest zeigen auf, dass eine differenzierte Strategie bei diesen Tests erforderlich ist. Sollen diese Tests optimal verwendet werden, so muss basierend auf der vorhandenen Stichprobengröße und vermuteten Art der Abweichung von der Benford Verteilung eine Vorgehensweise gewählt werden.

Trotzdem diese Ergebnisse eindeutig sind, bleiben etliche Fragen noch offen. Während die große Anzahl an Stichprobengrößen diese Studie in diesem Aspekt umfassend macht, existieren weitere durchaus noch mehr potenzielle Alternativverteilungen und Anpassungstests. Des Weiteren könnte eine Betrachtung, welche Kombination von Test und geprüfte Anzahl von Ziffern die insgesamt beste Vorgehensweise darstellt. Zudem weisen die Resultate daraufhin, dass es alternative Verteilungen gibt, bei denen das Testen mehrerer Ziffern immer zu besseren Ergebnissen führt. Auch könnte eine optimale Anzahl an Ziffern, die geprüft wird, existieren. Die Verwendung der dritten und vierten Ziffer könnte zu weiteren Verbesserungen führen.

Literaturverzeichnis

- Benford, F. (1938):** The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78, pp. 551-572.
- Cho, W.K.T. and Gaines, B.J. (2007):** Breaking the (Benford) law: Statistical fraud detection in campaign finance. *The American Statistician*, 61, pp. 218-223.
- Diekmann, A. (2007):** Not the first digit! Using Benford's law to detect fraudulent scientific data. *Journal of Applied Statistics*, 34, pp. 321-329.
- Freedman, L.S. (1981):** Watson's U^2 statistic for a discrete distribution. *Biometrika*, 68, pp. 708-711.
- Hill, T.P. (1995):** A statistical derivation of the significant digit law. *Statistical Science*, 10, pp. 354-363.
- Joenssen, D.W. (2013):** BenfordTests: Statistical Tests for Evaluating Conformity to Benford's Law. *R package version 1.0*.
<http://cran.r-project.org/web/packages/BenfordTests/>
- Kolmogorov, A.N. (1933):** Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 4, pp. 83-91.
- Leemis, L.M., Schmeiser, B.W. and Evans, D.L.(2000):** Survival distributions satisfying Benford's law. *The American Statistician*, 54, pp. 236-241.
- Morrow, J. (2010):** Benford's law, families of distributions and a test basis.
<http://www.johnmorrow.info/projects/benford/benfordMain.pdf>.
- Nigrini, M.J. (1996):** A taxpayer compliance application of Benford's law: Tests and statistics for auditors. *Journal of the American Taxation Association*, 18, pp. 72-91.
- Nigrini, M.J. and Mittermaier, L.J. (1997):** The use of Benford's law as an aid in analytical procedures. *Auditing: A Journal of Practice and Theory*, 16, pp. 52-67.
- Pearson, K. (1900):** On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 50, pp. 157-175.
- R Core Team (2012):** *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, ISBN 3-900051-07-0.

- Rauch, B., Götsche, M., Brähler, G. and Engel, S. (2011):** Fact and fiction in EU-governmental economic data. *German Economic Review*, 12, pp. 243–255.
- Shapiro, S. and Francia, R. (1972):** An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67, pp. 215–216.
- Swanson, D., Cho, M.J. and Eltinge, J. (2003):** Detecting possibly fraudulent or error-prone survey data: Using Benford’s law. *Proceedings of the Section Research Methods*, American Statistical Association, pp. 4172–4177.
- Watrin, C., Struffert, R. and Ullmann, R. (2008):** Benford’s law: An instrument for selecting tax audit targets? *Review of Managerial Science*, 2, pp. 219-237.