Statistical Models and the Benford Hypothesis: A unified framework

Supplement: Additional Simulation Results and Complementary Information

Lucio Barabesi
University of Siena, Department of Economics and Statistics, Siena, Italy
and
Andrea Cerioli
University of Parma, Department of Economics and Management
and University Centre "Robust Statistics Academy" (Ro.S.A.), Parma, Italy
and
Marco Di Marzio
"G. D'Annunzio" University, Department of Philosophical,
Pedagogical and Economic-Quantitative Sciences, Pescara, Italy

Supplement description

In this Supplement, we first provide additional information and simulation results that complement those given in the paper. Specifically, after a brief summary of our simulation algorithm and a link to a repository where our code and the data analyzed in Section 6 can be retrieved, we extend the power comparisons of Section 5.2 to sample sizes $n = 50, 200, 500$ for all the considered alternative models where power computation turns out to be interesting.

In the second part of this Supplement we display the full set of results obtained on the two star-distances databases analyzed in Section 6 of the paper. We also show the empirical effect on test sizes of a truncation scheme mimicking the features of database *Star 1*.

Summary of the simulation algorithm

The test statistics under comparison are

- χ^2: first-digit Pearson statistic (12).
- M: mean absolute deviation (MAD) statistic.
- $\chi^2_{\{1,2\}}$: two-digit Pearson statistic testing conformance to the first-two digit distribution implied by the Benford hypothesis (see, e.g., Nigrini, 2012; Barabesi et al., 2018).
- Q: Hotelling-type statistic of Barabesi et al. (2022), based on the (first-digit) sum-invariance property of Benford's law.
- K: Kolmogorov-Smirnov statistic (14).
- V: Kuiper statistic (15).
- $Z^2_{N,n}$: Score statistic of the Benford hypothesis with N non-negative trigonometric components in the model, for $N = 1, 2$.
- $Z^2_{\widehat{N},n}$: Score statistic of the Benford hypothesis with data-driven selection of the number of non-negative trigonometric components through the simplified BIC criterion (20).
- $\Lambda_{N,n}$: Likelihood Ratio statistic of the Benford hypothesis with N non-negative trigonometric components in the model, for $N = 1, 2$.
- $\Lambda_{\widehat{N},n}$: Likelihood Ratio statistic of the Benford hypothesis with data-driven selection of the number of non-negative trigonometric components through the BIC criterion (22).

The tests based on $\chi^2_{\{1,2\}}$ and Q are not detailed in the paper. However, they are reported in this Supplement for completeness.

When applying rule (21) for model selection, we define the upper bound of model dimension, say $N_{\max}$, to depend on n through a relationship that satisfies the theoretical bound of Inglot and Ledwina (1996) and that yields $N_{\max} = 10$ when $n = 100$, a good practical choice in our applications.

The number of simulations for each power estimate is 10^4. All the tests are performed by approximating the 0.95th quantile of their null distribution through 10^6 data sets of n observations simulated under the Benford hypothesis through stochastic representation (19), as in the paper. Each test is then exact, up to Monte Carlo error, with size $\gamma = 0.05$. Power values for different test sizes in the range $0.001 \leq \gamma \leq 0.1$ yield equivalent conclusions and are not reported.

Code and data

Simulations have been performed through a Fortran code which is available at the following address:

`https://anonymous.4open.science/r/Supplementary-Material-Benford-Test-1044`

Our code makes use of some IMSL functions and routines, including `DKSONE` for the computation of K and V. Since this routine does not admit ties, some mild form of jittering may be necessary under certain models involving truncation and in the analysis of real data where ties occur.

At the address reported above we also provide the star-distance samples analyzed in Section 6 of the paper. These samples are compressed in files `stars1.zip` and `stars2.zip`, which refer to the two databases that we consider in our experiment.

The hydrological data set analyzed in Section 7, called *streamflow.rda*, is instead available at the web address

`https://github.com/carloscinelli/benford.analysis`,

from which we have retrieved it. This data set comes with the R package *benford.analysis* (see also Cinelli, 2022).

Power results

Normal law

X follows the Normal law $N(\mu, \sigma^2)$ for different values of μ and σ.

Table 1: Estimated power when X is distributed as $N(\mu, \sigma^2)$, for different values of μ and σ, and $n = 50$. The exact test size is $\gamma = 0.05$.

μ	σ	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
0	1	0.163	0.181	0.101	0.164	0.155	0.272	0.305	0.233	0.259	0.307	0.233	0.237
0	5	0.213	0.231	0.177	0.193	0.301	0.270	0.303	0.226	0.250	0.303	0.223	0.233
3	5	0.236	0.220	0.219	0.222	0.348	0.302	0.337	0.261	0.286	0.339	0.254	0.267
3	10	0.160	0.188	0.087	0.165	0.162	0.267	0.302	0.230	0.249	0.302	0.225	0.236
5	5	0.315	0.241	0.288	0.322	0.409	0.480	0.528	0.421	0.462	0.527	0.416	0.436
5	10	0.161	0.206	0.072	0.164	0.189	0.289	0.316	0.251	0.267	0.321	0.248	0.249

Table 2: Estimated power when X is distributed as $N(\mu, \sigma^2)$, for different values of μ and σ, and $n = 200$. The exact test size is $\gamma = 0.05$.

μ	σ	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
0	1	0.588	0.658	0.246	0.589	0.461	0.821	0.865	0.785	0.857	0.867	0.787	0.860
0	5	0.649	0.713	0.377	0.624	0.841	0.815	0.860	0.776	0.851	0.861	0.780	0.853
3	5	0.689	0.729	0.464	0.687	0.886	0.867	0.903	0.838	0.897	0.907	0.843	0.900
3	10	0.603	0.693	0.212	0.604	0.495	0.824	0.869	0.792	0.861	0.872	0.793	0.863
5	5	0.847	0.810	0.683	0.886	0.934	0.983	0.991	0.978	0.990	0.992	0.979	0.991
5	10	0.640	0.737	0.182	0.634	0.590	0.849	0.887	0.814	0.879	0.888	0.813	0.880

Gamma law

X follows the Gamma law $G(\alpha)$ for different values of the shape parameter α. The scale parameter is fixed to 1.

Table 3: Estimated power when X is a Gamma random variable with scale parameter 1, for different values of the shape parameter α and $n = 50$. The exact test size is $\gamma = 0.05$.

α	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
1.0	0.050	0.063	0.040	0.050	0.069	0.069	0.073	0.063	0.065	0.074	0.060	0.063
1.5	0.060	0.087	0.026	0.065	0.156	0.139	0.152	0.119	0.129	0.153	0.116	0.118
2.0	0.083	0.121	0.012	0.095	0.271	0.270	0.300	0.233	0.253	0.298	0.229	0.233
2.5	0.160	0.184	0.015	0.171	0.319	0.458	0.506	0.401	0.441	0.511	0.399	0.422
3.0	0.316	0.367	0.031	0.332	0.274	0.664	0.722	0.605	0.662	0.731	0.604	0.644
4.0	0.762	0.837	0.246	0.746	0.702	0.930	0.956	0.909	0.937	0.960	0.912	0.932

Table 4: Estimated power when X is a Gamma random variable with scale parameter 1, for different values of the shape parameter α and $n = 200$. The exact test size is $\gamma = 0.05$.

α	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
1.0	0.087	0.108	0.050	0.089	0.130	0.143	0.159	0.127	0.154	0.160	0.127	0.154
1.5	0.218	0.295	0.038	0.219	0.461	0.429	0.479	0.367	0.464	0.483	0.371	0.466
2.0	0.525	0.571	0.043	0.531	0.810	0.816	0.859	0.770	0.851	0.860	0.773	0.851
2.5	0.851	0.825	0.129	0.862	0.932	0.980	0.989	0.972	0.988	0.990	0.973	0.989
3.0	0.982	0.980	0.407	0.985	0.961	1.000	1.000	0.999	1.000	1.000	0.999	1.000

Table 5: Estimated power when X is a Gamma random variable with scale parameter 1, for different values of the shape parameter α and $n = 500$. The exact test size is $\gamma = 0.05$.

α	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
1.0	0.170	0.236	0.063	0.166	0.253	0.305	0.338	0.262	0.335	0.339	0.263	0.336
1.5	0.608	0.701	0.104	0.593	0.845	0.829	0.876	0.790	0.873	0.876	0.791	0.873
2.0	0.960	0.969	0.313	0.961	0.995	0.996	0.998	0.995	0.998	0.998	0.995	0.998

Beta law

X follows the Beta law $Be(\alpha, \beta)$ for different values of the the shape parameters α and β.

Table 6: Estimated power when X is a Beta random variable, for different values of the shape parameters α and β, and $n = 50$. The exact test size is $\gamma = 0.05$.

α	β	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
0.5	1.0	0.405	0.372	0.382	0.382	0.564	0.365	0.352	0.327	0.324	0.339	0.312	0.296
1.0	1.0	0.941	0.932	0.868	0.933	0.987	0.936	0.894	0.900	0.883	0.886	0.890	0.862
1.5	1.0	1.000	0.999	0.992	1.000	1.000	1.000	0.998	1.000	0.998	0.997	0.999	0.997
0.5	2.0	0.085	0.089	0.070	0.083	0.082	0.105	0.112	0.091	0.096	0.111	0.095	0.092
1.0	2.0	0.291	0.338	0.182	0.271	0.361	0.397	0.425	0.355	0.374	0.427	0.358	0.358
1.5	2.0	0.743	0.783	0.509	0.721	0.874	0.860	0.862	0.824	0.829	0.865	0.825	0.818

Table 7: Estimated power when X is a Beta random variable, for different values of the shape parameters α and β, and $n = 200$. The exact test size is $\gamma = 0.05$.

α	β	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
0.5	1.0	0.940	0.945	0.758	0.934	0.989	0.952	0.915	0.922	0.918	0.913	0.918	0.917
0.5	2.0	0.194	0.223	0.111	0.185	0.216	0.301	0.325	0.262	0.317	0.327	0.268	0.318
1.0	2.0	0.891	0.925	0.529	0.888	0.942	0.966	0.970	0.954	0.969	0.973	0.956	0.972

Table 8: Estimated power when X is a Beta random variable, for different values of the shape parameters α and β, and $n = 500$. The exact test size is $\gamma = 0.05$.

α	β	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
0.5	2.0	0.457	0.531	0.180	0.462	0.493	0.657	0.695	0.615	0.691	0.700	0.623	0.696

Positive Stable law

X follows the Positive Stable law $PS(\alpha)$ for different values of the shape parameter α. The Positive Stable law considered in our work has Laplace transform $L_X(t) = e^{-t^\alpha}$, with $\Re(t) \in \mathbb{R}^+$. It is simulated through the Kanter algorithm (see Chambers et al., 1976, Equation (2.2)).

Table 9: Estimated power when X is a Positive Stable random variable, for different values of the shape parameter α and $n = 50$. The exact test size is $\gamma = 0.05$.

α	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
0.5	0.047	0.048	0.042	0.047	0.049	0.053	0.053	0.056	0.054	0.051	0.056	0.054
0.6	0.076	0.081	0.064	0.073	0.076	0.087	0.091	0.079	0.080	0.091	0.078	0.073
0.7	0.263	0.249	0.217	0.246	0.368	0.348	0.387	0.298	0.330	0.385	0.295	0.307
0.8	0.847	0.790	0.640	0.845	0.892	0.947	0.956	0.931	0.937	0.952	0.929	0.929

Table 10: Estimated power when X is a Positive Stable random variable, for different values of the shape parameter α and $n = 200$. The exact test size is $\gamma = 0.05$.

α	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
0.5	0.051	0.049	0.044	0.052	0.058	0.065	0.065	0.063	0.064	0.065	0.063	0.065
0.6	0.147	0.173	0.098	0.140	0.182	0.219	0.244	0.183	0.236	0.244	0.182	0.236
0.7	0.775	0.796	0.495	0.766	0.918	0.922	0.946	0.902	0.942	0.948	0.904	0.943

Table 11: Estimated power when X is a Positive Stable random variable, for different values of the shape parameter α and $n = 500$. The exact test size is $\gamma = 0.05$.

α	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
0.5	0.050	0.055	0.040	0.054	0.068	0.073	0.076	0.068	0.076	0.076	0.068	0.076
0.6	0.321	0.407	0.141	0.303	0.412	0.487	0.537	0.426	0.531	0.539	0.429	0.533
0.7	0.995	0.997	0.889	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Mixtures

X is distributed as a two-component Normal mixture $0.5N(\mu_1, \sigma_1^2) + 0.5N(\mu_2, \sigma_2^2)$, for different values of μ_1, μ_2, σ_1 and σ_2. We refer to Section 3.5 of the paper for the definition of cases (a)–(f).

Table 12: Estimated power when X is a two-component Normal mixture $0.5N(\mu_1, \sigma_1^2) + 0.5N(\mu_2, \sigma_2^2)$ and $n = 50$. The exact test size is $\gamma = 0.05$.

Case	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
(a)	0.109	0.107	0.053	0.105	0.074	0.106	0.075	0.199	0.172	0.071	0.192	0.185
(b)	0.310	0.329	0.090	0.339	0.130	0.244	0.070	0.625	0.610	0.068	0.631	0.637
(c)	0.186	0.196	0.143	0.182	0.124	0.121	0.063	0.286	0.258	0.067	0.283	0.273
(d)	0.170	0.220	0.067	0.203	0.148	0.198	0.098	0.408	0.368	0.109	0.411	0.383
(e)	0.117	0.179	0.035	0.160	0.168	0.174	0.085	0.380	0.342	0.074	0.366	0.352
(f)	0.553	0.511	0.191	0.531	0.319	0.586	0.403	0.757	0.672	0.356	0.729	0.649

Table 13: Estimated power when X is a two-component Normal mixture $0.5N(\mu_1, \sigma_1^2) + 0.5N(\mu_2, \sigma_2^2)$ and $n = 200$. The exact test size is $\gamma = 0.05$.

Case	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
(a)	0.416	0.411	0.130	0.394	0.176	0.363	0.154	0.684	0.501	0.144	0.686	0.507
(b)	0.957	0.962	0.599	0.973	0.439	0.902	0.125	0.999	0.996	0.134	0.999	0.996
(c)	0.630	0.681	0.385	0.644	0.432	0.478	0.098	0.880	0.757	0.112	0.884	0.769
(d)	0.734	0.804	0.338	0.805	0.477	0.788	0.263	0.977	0.915	0.313	0.978	0.921
(e)	0.597	0.691	0.203	0.701	0.578	0.691	0.226	0.954	0.871	0.196	0.953	0.872
(f)	0.998	0.995	0.864	0.996	0.931	0.998	0.943	1.000	0.999	0.925	1.000	0.998

Table 14: Estimated power when X is a two-component Normal mixture $0.5N(\mu_1, \sigma_1^2) + 0.5N(\mu_2, \sigma_2^2)$ and $n = 500$. The exact test size is $\gamma = 0.05$.

Case	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
(a)	0.875	0.862	0.444	0.862	0.445	0.824	0.310	0.987	0.919	0.291	0.988	0.920
(b)	1.000	1.000	0.998	1.000	0.964	1.000	0.251	1.000	1.000	0.277	1.000	1.000
(c)	0.976	0.987	0.820	0.984	0.866	0.943	0.184	1.000	0.996	0.214	1.000	0.996
(d)	0.994	0.998	0.924	0.999	0.904	0.999	0.590	1.000	1.000	0.665	1.000	1.000
(e)	0.980	0.989	0.825	0.993	0.950	0.993	0.492	1.000	0.999	0.443	1.000	0.999

Generalized Benford law

X follows a Generalized Benford law of parameter $\alpha \neq 0$. For the parameter values selected in the paper, all the test statistics under comparison have power approaching 1.0 when $n > 100$. Therefore, we only report the power results for $n = 50$.

Table 15: Estimated power when X is a Generalized Benford random variable of parameter α, for different values of α and $n = 50$. The exact test size is $\gamma = 0.05$.

α	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
-1.0	0.818	0.907	0.004	0.847	0.987	0.939	0.893	0.905	0.883	0.885	0.893	0.863
-0.8	0.568	0.712	0.003	0.598	0.925	0.786	0.723	0.725	0.701	0.712	0.709	0.665
-0.6	0.283	0.440	0.002	0.313	0.723	0.515	0.469	0.454	0.439	0.459	0.439	0.407
0.6	0.553	0.516	0.492	0.527	0.722	0.515	0.475	0.461	0.444	0.468	0.442	0.417
0.8	0.797	0.781	0.709	0.777	0.922	0.784	0.720	0.722	0.695	0.708	0.705	0.665
1.0	0.942	0.929	0.866	0.933	0.987	0.937	0.898	0.901	0.886	0.889	0.891	0.862

Table 16: Estimate of $E[\widehat{N}]$ when X is a Generalized Benford random variable of parameter α, for different values of α and $n = 50$.

	$\alpha = -1.0$	$\alpha = -0.8$	$\alpha = -0.6$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1.0$
Simplified BIC	1.25	1.14	1.08	1.08	1.14	1.26
Likelihood BIC	1.24	1.15	1.09	1.09	1.15	1.24

Table 17: Estimate of $E[\widehat{N}]$ when X is a Generalized Benford random variable of parameter α, for different values of α and $n = 200$.

	$\alpha = -1.0$	$\alpha = -0.8$	$\alpha = -0.6$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1.0$
Simplified BIC	1.92	1.50	1.20	1.20	1.49	1.91
Likelihood BIC	1.83	1.46	1.19	1.20	1.46	1.82

Table 18: Estimate of $E[\widehat{N}]$ when X is a Generalized Benford random variable of parameter α, for different values of α and $n = 500$.

	$\alpha = -1.0$	$\alpha = -0.8$	$\alpha = -0.6$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1.0$
Simplified BIC	2.92	2.29	1.62	1.63	2.30	2.94
Likelihood BIC	2.83	2.23	1.60	1.60	2.23	2.79

Analysis of astronomical distances

We provide the results of the analysis described in Section 6.2 of the paper.

Table 19: Proportion of rejections of the Benford hypothesis in 1,000 samples of size $n = 200$ drawn with replacement from each star-distance database, at exact test size γ

γ	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
						Stars 1						
0.10	0.332	0.378	0.065	0.323	0.579	0.574	0.605	0.496	0.597	0.606	0.498	0.596
0.05	0.213	0.238	0.033	0.210	0.456	0.429	0.487	0.367	0.477	0.487	0.368	0.468
0.01	0.065	0.077	0.006	0.078	0.236	0.236	0.270	0.182	0.212	0.267	0.185	0.196
						Stars 2						
0.10	0.291	0.297	0.109	0.278	0.405	0.558	0.600	0.502	0.590	0.587	0.493	0.585
0.05	0.191	0.183	0.056	0.188	0.275	0.427	0.466	0.358	0.463	0.462	0.342	0.452
0.01	0.064	0.060	0.009	0.055	0.104	0.198	0.224	0.156	0.174	0.222	0.156	0.166

Table 20: Proportion of rejections of the Benford hypothesis in 1,000 samples of size $n = 300$ drawn with replacement from each star-distance database, at exact test size γ

γ	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
						Stars 1						
0.10	0.464	0.508	0.120	0.462	0.747	0.728	0.775	0.677	0.774	0.781	0.678	0.775
0.05	0.335	0.373	0.051	0.325	0.634	0.604	0.668	0.565	0.660	0.669	0.566	0.658
0.01	0.133	0.150	0.010	0.131	0.387	0.378	0.438	0.336	0.387	0.439	0.326	0.385
						Stars 2						
0.10	0.399	0.398	0.124	0.413	0.539	0.716	0.757	0.665	0.751	0.751	0.661	0.748
0.05	0.270	0.289	0.062	0.281	0.409	0.599	0.658	0.540	0.649	0.647	0.536	0.637
0.01	0.106	0.109	0.011	0.106	0.205	0.356	0.403	0.288	0.345	0.398	0.281	0.337

Table 21: Proportion of rejections of the Benford hypothesis in 1,000 samples of size $n = 500$ drawn with replacement from each star-distance database, at exact test size γ

γ	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
						Stars 1						
0.10	0.718	0.729	0.207	0.702	0.903	0.907	0.942	0.889	0.941	0.941	0.895	0.94
0.05	0.576	0.587	0.116	0.556	0.837	0.839	0.887	0.800	0.882	0.885	0.804	0.883
0.01	0.330	0.325	0.032	0.306	0.654	0.64	0.717	0.601	0.692	0.722	0.601	0.693
						Stars 2						
0.10	0.631	0.617	0.248	0.647	0.744	0.89	0.921	0.851	0.919	0.919	0.849	0.919
0.05	0.500	0.490	0.150	0.524	0.610	0.819	0.856	0.783	0.852	0.856	0.781	0.853
0.01	0.269	0.248	0.035	0.285	0.377	0.634	0.691	0.574	0.668	0.683	0.576	0.659

Truncation effect on test size

We examine the effect of truncation on a Benford random variable X, when its reported values only contain an integer – and possibly small – number η of significant digits. This is obtained by first simulating X through stochastic representation (19) and then retaining the first η significant digits of the generated significand $S(x)$. The three-digit distances available in database *Stars 1* thus correspond to the case $\eta = 3$. The results shown below parallel the findings obtained by Barabesi et al. (2022, Table 7) in the context of international trade.

Table 22: Estimated size when X is a Benford random variable truncated to have η significant digits, for decreasing values of η and $n = 50$. The exact test size is $\gamma = 0.05$.

η	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
6	0.050	0.049	0.049	0.050	0.048	0.051	0.051	0.051	0.049	0.052	0.052	0.050
5	0.046	0.052	0.050	0.049	0.046	0.050	0.048	0.052	0.050	0.051	0.051	0.051
4	0.052	0.051	0.049	0.051	0.051	0.049	0.050	0.049	0.048	0.048	0.049	0.046
3	0.050	0.050	0.049	0.051	0.048	0.048	0.048	0.051	0.051	0.047	0.049	0.049

Table 23: Estimated size when X is a Benford random variable truncated to have η significant digits, for decreasing values of η and $n = 200$. The exact test size is $\gamma = 0.05$.

η	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
6	0.053	0.051	0.050	0.052	0.050	0.056	0.055	0.051	0.055	0.056	0.050	0.054
5	0.048	0.048	0.052	0.049	0.054	0.053	0.052	0.051	0.051	0.052	0.051	0.052
4	0.047	0.049	0.048	0.046	0.051	0.054	0.053	0.053	0.052	0.054	0.053	0.053
3	0.053	0.049	0.051	0.053	0.049	0.048	0.049	0.049	0.051	0.050	0.049	0.052

Table 24: Estimated size when X is a Benford random variable truncated to have η significant digits, for decreasing values of η and $n = 500$. The exact test size is $\gamma = 0.05$.

η	χ^2	M	$\chi^2_{\{1,2\}}$	Q	K	V	$Z^2_{1,n}$	$Z^2_{2,n}$	$Z^2_{\widehat{N},n}$	$\Lambda_{1,n}$	$\Lambda_{2,n}$	$\Lambda_{\widehat{N},n}$
6	0.051	0.053	0.054	0.053	0.052	0.054	0.053	0.052	0.052	0.052	0.052	0.053
5	0.050	0.054	0.051	0.049	0.048	0.049	0.052	0.050	0.051	0.052	0.049	0.051
4	0.050	0.047	0.049	0.050	0.051	0.047	0.052	0.053	0.051	0.050	0.051	0.051
3	0.051	0.053	0.050	0.050	0.054	0.049	0.048	0.048	0.047	0.049	0.048	0.047

References

Barabesi, L., Cerasa, A., Cerioli, A., and Perrotta, D. (2018). Goodness-of-fit testing for the Newcomb-Benford law with application to the detection of customs fraud. *Journal of Business and Economic Statistics*, 36:346–358.

Barabesi, L., Cerasa, A., Cerioli, A., and Perrotta, D. (2022). On characterizations and tests of Benford's law. *Journal of the American Statistical Association*, 117:1887–1903.

Chambers, J. M., Mallows, C. L., and Stuck, B. W. (1976). A method for simulating stable random variables. *Journal of the American Statistical Association*, 71:340–344.

Cinelli, C. (2022). Package 'benford.analysis'. `https://cran.r-project.org/web/packages/benford.analysis/benford.analysis.pdf`. Last accessed: 2023-04-13.

Inglot, T. and Ledwina, T. (1996). Asymptotic optimality of data-driven Neyman's tests for uniformity. *The Annals of Statistics*, 24:1982–2019.

Nigrini, M. J. (2012). *Benford's Law*. Wiley, Hoboken.