

DER MISSVERSTANDENE BENFORD

ODER AUCH: WIE SICH BENFORDSCHE FINANZKONTROLLEN ÜBERTÖLPELN LASSEN

Die Waffe der Datenforensik besitzt ein gefährliches Leck

In der modernen Wirtschaftsprüfung, der automatisierten Datenanalyse und bei vielen weiteren Anwendungen gilt das Benfordsche Gesetz als die unbestrittene Wunderwaffe zur Aufdeckung von Fälschungen. Die Logik scheint bestechend: Weicht die Häufigkeitsverteilung der führenden Ziffern (die Dominanz der Ziffer 1) in einem Datensatz signifikant von der logarithmischen Benford-Kurve ab, schlägt das System Alarm. Diagnose: Manipulationsverdacht.

Doch die forensische Praxis weist ein gewisses Interpretationsrisiko auf, denn eine rein statistische **Abweichung** ist noch lange nicht mit einer nachgewiesenen **Manipulation** gleichzusetzen.

Das latente Missverständnis: Abweichung ist nicht gleich Fälschung

Wenn eine Zahlenreihe die Benford-Kriterien bricht, ist damit exakt eines bewiesen: die rein mathematische Abweichung von einer globalen Erwartungswertkurve – mehr nicht. Die Gründe hierfür können völlig legal, strukturell oder prozessbedingt sein (z.B. feste Tarifgrenzen, gelistete Preisklassen, usw.).

Viel problematischer ist jedoch der Umkehrschluss: Das blinde Vertrauen darauf, dass ein Datensatz, der die Benford-Häufigkeiten perfekt oder stark angenähert erfüllt, automatisch „sicher“, d.h. nicht punktuell manipuliert oder gar global gefälscht ist.

Betrachten wir ein klassisches Beispiel: Das dokumentierte Bevölkerungswachstum oder die Zinseszins-Entwicklung über die Zeit. Diese Datenreihen folgen mathematisch dem Benfordschen Gesetz. Doch sie entstehen keineswegs aus einem ungeordneten, zufälligen Prozess. Die Werte erwachsen zwar einer zeitlichen Abfolge, sind aber **höchst abhängig vom jeweiligen Vorwert** und damit für eine Prüfung auf Manipulation völlig ungeeignet. Die vorausgesetzte „Zufälligkeit in der Abfolge“ (der gemessenen Absolutwerte!) ist hier per Definition überhaupt nicht gegeben.

Falls sich dagegen ein Prozess mit reiner Zufälligkeit abwickelt – wie dies beispielsweise bei Finanzdaten (Zahlungsabfolgen) erwartet wird und damit den Benfordschen Gesetzen über die Häufigkeitsverteilung der Erstziffern, der Zweitziffern, der Drittziffern, etc. unterliegen muss – bedeutet dies für die Praxis:

Ein geschickter Fälscher wird versuchen, einen Finanz- oder Datenstrom so zu frisieren, dass die jeweilige Verteilung einer Ziffernspalte (Makro-Ebene) die Benford-Prüfung fehlerfrei passiert.

Die Lösung: Zweidimensionale Datenprüfung

Um eine Datenreihe verlässlich als „NICHT - MANIPULIERT“ oder „UNGESTÖRT“ taxieren zu können, muss der Blick von der reinen Häufigkeit (Makro-Ebene) auf die stochastische Struktur ihrer Abfolge (Mikro-Ebene) gelenkt werden. Erst hier wird der verbotene Eingriff erkennbar.

Die von OEKOPRIORITY® entwickelte «**Modellanwendung zum Gesetz der grossen Zahlen (GGZ)**»^{***} definiert dazu ein mathematisches Referenzprofil für die Abfolge von Ziffern. Es wird geprüft, ob die Cluster, die Wiederholungsraten und die Nachbarschaften der Zahlen, die sich nach den Gesetzen stochastischer Unabhängigkeit zwangsläufig ergeben, bei der untersuchten Zahlenreihe eingehalten sind. ^{***} *Frühere Bezeichnung: «Gesetz zur Gruppenbildung».*

Daraus ergeben sich zwei unumstössliche Postulate für die Forensik:

1. Datenintegrität erfordert Strukturkonformität: Ein Datensatz ist erst dann nachweislich ungestört, wenn seine Abfolge dem GGZ entspricht – unabhängig davon, welche globale Häufigkeitsverteilung vorliegt.

2. Grenzen der Anwendung: Nicht jede Zahlenfolge, die der Benford-Häufigkeit entspricht, ist überhaupt für eine GGZ-Prüfung geeignet. Deterministisch gewachsene, in sich geordnete Datenreihen laufen bei einer stochastischen Sequenzanalyse mathematisch ins Leere.

Fazit: Wer Datenmanipulationen sicher aufdecken will, darf sich nicht länger hinter dem „missverstandenen Benford“ verstecken. Die Forensik muss zweidimensional werden und neben der Ziffernhäufigkeit zwingend die mikro-strukturelle Ordnung der Abfolge prüfen.

Praxisbezug

Idealfall: Folgen die Erstziffern einer Zahlenreihe exakt dem Benfordschen Gesetz, bildet sich die gemäss **BILD 1** dargestellte Makro – Struktur (Tabellenkopf) und die sich für die «Erstziffernspalte» bildende Mikrostruktur (Verteilungszahlen).

Erkennbare Fälschung: Der Zahlenblock (1) **BILD 2** zeigt eine Reihe von 500 Zahlen, welche für die Erstziffern und alle Folgeziffern die (fast exakten) Häufigkeitsverteilungen (wie laut Tabellenkopf BILD 1) aufweisen. Eine Überprüfung allein «nach Benford» wird diesen Zahlenblock somit glatt durchwinken.

Die Kontrolle mittels «Modellanwendung zum GGZ» jedoch, weist in **BILD 3** einen ausgewiesenen tiefen PLAUSIBILITÄTSGRADES von **nur 48.5 Prozent** für die Erstziffern auf. Dies bedeutet, dass die erste Ziffernreihe offensichtlich manipuliert wurde.

Erstes Fazit: Bei Einhaltung der Häufigkeitsverteilungen je nach Ziffernspalte wird eine Fälschung allein „mittels Benford“ nicht entlarvt – wohl aber von der „MODELLANWENDUNG ZUM GGZ“. Denn: Es wird Dritten kaum gelingen, den Zahlenblock – nebst den richtigen Häufigkeitsverteilungen je nach Ziffernspalte – **auch noch** entsprechend der sich in der Abfolge bildenden Cluster, Wiederholungsraten und Nachbarschaften nach stochastischen Gesetzmässigkeiten auf Mikro – Ebene richtig zusammenzustellen.

Nicht erkennbare Fälschung: Ein weiterer Zahlenblock (2) **BILD 4** mit wiederum 500 Daten treibt das Thema auf die Spitze. Auch hier erfüllt die Zahlenreihe bezüglich der Häufigkeitsverteilungen (Makro – Ebene) alle Bedingungen nach Benford fast perfekt – **und auch** die Bedingungen für die Mikro – Ebene. So weist die gefälschte Erstziffernreihe einen sehr hohen PLAUSIBILITÄTSGRAD von **90.3%** auf, siehe **BILD 5**. Damit würde die Totalfälschung des gesamten Zahlenblocks (Bild 4) auch unter Zuhilfenahme weiterer mathematischer Prüfmethode sicher nicht als solche erkannt!

Was heisst das nun:

In Kenntnis des mathematischen Zusammenhangs bei der «MODELLANWENDUNG ZUM GGZ» besteht tatsächlich die Möglichkeit,

ENTWEDER beliebige Zahlenreihen (nicht bloss solche, die Benford folgen müssen) auf Zufälligkeit, und damit auf «Ungestörtheit», und damit auf «NICHT – EINWIRKUNG» zu kontrollieren – was die eigentliche Zielsetzung der «Modellanwendung OEKOPRIORITY®» ist,

ODER aber, durch Umkehrung des Berechnungsmodells Manipulationen an, bzw. Fälschungen von **beliebigen** Zahlenreihen vorzunehmen, **welche auch mit erweitertem forensischen Aufwand nicht entlarvt werden können.**

Zweites Fazit: Bei der Verbreitung der «MODELLANWENDUNG ZUM GGZ» ist darauf zu achten, dass keine Details zum Berechnungsmodell in falsche Hände gelangen.

EIGENE MISCHUNG (WAHRES R. - PROFIL) (%/100)	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046		PAKETSUMME	
Gruppenbildung "m"	Zahlenband 1	Zahlenband 2	Zahlenband 3	Zahlenband 4	Zahlenband 5	Zahlenband 6	Zahlenband 7	Zahlenband 8	Zahlenband 9	Zahlenband 10	ALLER Zahlenbänder 1 bis 10	
m = 1	147.068901	119.499776	95.703125	79.094673	67.011039	58.322763	51.467112	45.930651	41.865336	0	705.963376	
m = 2	44.2677392	21.03196058	11.96289063	7.672183281	5.293872081	3.907625121	2.985092496	2.342463201	1.925805456	0	101.389632	
m = 3	13.3245895	3.701625061	1.495361328	0.744201778	0.418215894	0.261810883	0.173135365	0.119465623	0.088587051	0	20.32699248	
m = 4	4.010701439	0.651486011	0.186920166	0.072187572	0.033039056	0.017541329	0.010041851	0.006092747	0.004075004	0	4.992085176	
m = 5	1.207221133	0.114661538	0.023365021	0.007002195	0.002610085	0.001175269	0.000582427	0.00031073	0.00018745	0	1.357115849	
m = 6	0.363373561	0.020180431	0.002920628	0.000679213	0.000206197	7.8743E-05	3.37808E-05	1.58472E-05	8.62271E-06	0	0.387497023	
m = 7	0.109375442	0.003551756	0.000365078	6.58836E-05	1.62895E-05	5.27578E-06	1.95929E-06	8.08209E-07	3.96645E-07	0	0.113382889	
m = 8	0.032922008	0.000625109	4.56348E-05	6.39071E-06	1.28687E-06	3.53477E-07	1.13639E-07	4.12187E-08	1.82457E-08	0	0.033600956	
m = 9	0.009909524	0.000110019	5.70435E-06	6.19899E-07	1.01663E-07	2.3683E-08	6.59104E-09	2.10215E-09	8.393E-10	0	0.010026003	
m= 10	0.002982767	1.93634E-05	7.13044E-07	6.01302E-08	8.03138E-09	1.58676E-09	3.8228E-10	1.0721E-10	3.86078E-11	0	0.003002914	
											SUMME ALLER PAKETE	834.5767113
											PROZENTUALER ANTEIL ALLER ERFASTEN DATEN	99.99852706
ERFASSTE ANZAHL DATEN alle bis m ≤ 10; 2 ≤ n ≤10	300.9853182	175.9999536	124.9999989	96.99999993	78.99999999	67	58	51	46	0	999.9852706	
ERFASSTE ANZAHL PAKETE alle bis m ≤ 10; 2 ≤ n ≤10	210.3977156	145.0239959	109.3749999	87.59099999	72.759	62.511	54.636	48.399	43.884	0	834.5767113	
											SUMME ALLER ERFASTEN DATEN	999.9852706

BILD 1: HÄUFIGKEITSVERTEILUNG UND MIKROSTRUKTUR DER ERSTZIFFERN EINER BENFORD – REIHE

80862	54024	22212	41581	22728	40624	77514	38634	35922	75249	84732	18819	65884	72720	79005	50133	20199	30052	27696	62397
64833	56656	27103	4948	2399	54778	14490	93756	7103	15878	20595	16149	92103	65221	63951	81000	29933	6035	3981	4003
74656	91250	79240	30262	5358	71298	38348	75809	15306	22556	19493	74006	41746	6119	44591	44249	51070	3061	55257	85808
81787	95919	72601	45651	61815	17965	29413	51272	40507	41661	34781	46134	74946	88455	40278	84961	90236	25081	84232	24503
6394	38584	99019	57030	39891	60022	59680	2297	70067	23601	89412	97255	79140	37826	57447	14099	15822	50681	16889	70570
42569	50103	55926	36584	9733	44949	95507	47427	99895	55102	9544	89682	47883	92220	63724	71624	18399	37403	32606	31584
82485	95962	29060	74895	80247	80719	50157	77957	93747	39589	92451	98779	77538	86237	52271	30268	43252	91339	40703	71359
31176	52296	99915	8250	92577	71605	26724	3365	90947	89449	35509	45584	22598	7158	59172	93209	70794	95550	71283	60198
89523	60711	28301	81255	15661	84840	14772	29032	10045	82866	1701	10810	15614	8843	63373	14469	15471	39539	52719	73979
10193	40181	14013	14458	14520	14816	15288	14019	1780	12863	16891	10199	13941	17681	15320	12551	14612	70239	17760	12766
26488	13309	49143	90001	11344	1373	18184	50146	17092	19289	14555	10652	10088	98578	13311	10892	11672	1795	11597	18569
17218	17661	13953	18540	1212	19635	17903	14950	14971	11938	14748	10189	10016	18975	80039	11561	17057	15532	11410	16238
10540	11672	14814	15567	15352	1516	14166	17275	63288	17343	13028	1070	12373	43497	15982	13549	17029	10337	14690	15109
24737	3511	14339	14443	9205	30831	12951	17123	16170	58717	13194	16940	16571	1383	12517	1390	12522	1046	14400	58723
16643	64855	17997	12127	10189	14248	19952	17761	11210	16207	12409	10	13933	85210	2498	10766	13100	54729	40888	11860
17664	11446	18111	16825	12294	14551	15271	78173	11420	15892	10778	1083	1088	31638	63569	18888	12434	18381	10714	14592
13328	16033	10696	8282	2421	55088	49605	82907	18079	1022	63882	57622	69848	88541	60370	35200	2463	92447	7475	8698
74714	25962	72350	93089	12980	21793	50002	6393	67292	44445	42819	31303	75251	89015	72542	31834	32431	21318	32509	70279
781	48265	1719	6950	61130	98517	87264	4014	61308	69485	3744	2300	34509	46479	47396	24150	20846	57841	5870	90562
80398	73897	81211	16095	90139	83577	55309	64813	74845	34010	13148	67332	71565	55056	15814	10802	72226	38931	50570	37909
10283	26946	99589	70143	295	17130	64416	61381	38708	73633	53174	47588	53673	92736	34067	47581	70761	56778	70559	58375
60589	366	24120	81613	47028	68920	38162	56254	13778	91839	49016	81870	74568	14239	34648	64789	61498	33957	30895	39026
67935	26312	45529	5954	17018	76684	54343	32278	52893	63664	17023	87381	8554	34346	65988	86016	88337	10281	27106	40914
15764	64564	38800	76779	92899	52324	67787	93715	75979	5750	60401	30962	19609	21981	62834	90465	93986	82223	11192	42191
99880	82897	7530	83466	51626	69907	21543	77451	21550	51714	23320	66822	23312	56551	89760	75150	43597	11398	85420	72826

BILD 2: ZAHLENBLOCK (1) – ERKANNT FÄLSCHUNG (VON LINKS OBEN NACH RECHTS UNTEN LESEN)

OBWOHL: DIE ZAHL 1 IN DER ERSTZIFFERN – REIHE WEIST EINE HÄUFIGKEIT VON 30.2% AUF (≡ BENFORD)

ZUSAMMENSTELLUNG:	
ANZAHL GELIEFERTE DATEN --> HANDEINTRAG	500
DAVON ANZAHL AUSREISSER LAUT GRUBBS -> HANDEINTRAG	0
ANZAHL ERFASSTE DATEN INNERHALB $m \leq 10$ UND OHNE AUSREISSER	441
ZUFALLSANTEIL "P*ZF" DER ERFASSTEN DATEN (%)	55.01
ZUFALLSANTEIL BEZ. GELIEFERTER ANZAHL (%)	48.52
KOMPLEMENTÄR - ANTEIL CHAOSVERHALTEN (%)	32.58
CHAOS - ANTEIL AUSREISSER (%)	0
KOMPLEMENTÄRER CHAOS - ANTEIL BEZ. GELIEFERTER ANZAHL (%)	28.74
KOMPLEMENTÄR - ANTEIL HERDENTRIEB (%)	12.4
HERDENTRIEB - ANTEIL AUS $m > 10$ (%)	11.8
KOMPLEMENTÄRER HERDENTRIEB - ANTEIL BEZ. GELIEFERTER ANZAHL (%)	22.74
KONTROLLSUMME BEZ. GELIEFERTER ANZAHL DATEN (%)	100

BILD 3: PLAUSIBILITÄTSGRAD 48.5% DER GEFÄLSCHTEN ERSTZIFFER – REIHE AUS BILD 2 → FÄLSCHUNG MIT GGZ ERKANNT!

20389	23930	75128	21332	26670	28538	20669	25900	68131	32164	70934	82913	15782	10568	47358	98240	79114	40061	78359	81353
15331	22285	81527	56011	47758	17329	45587	146	52263	45514	38963	14410	11691	27159	1593	47927	169	48955	36597	29211
16200	16218	21479	36121	19400	25994	79789	23353	79458	20276	17181	12636	27138	42800	44213	33647	82568	62162	98151	11488
20832	45701	16571	18684	40216	68385	16294	12324	1671	13224	18797	40622	69856	1426	46090	1370	1295	12337	42523	97267
63768	78988	1939	31296	74002	42881	25366	21919	56180	1204	56311	52376	17941	26736	12853	81053	75561	75903	62036	30078
41903	25571	77038	1543	35893	68929	54852	97638	17898	13884	47457	1157	16354	74973	42938	29660	17227	22058	11949	39510
12532	52149	21289	18888	40523	60293	51425	40823	82125	1823	1522	22307	76166	61853	18112	61432	30420	61627	18157	12069
75810	24292	23282	36618	38079	1509	25739	30965	29378	17464	37106	10310	13871	22657	55701	44422	60050	80686	20623	51552
16293	25710	12345	25337	75435	74762	69703	70383	1845	33948	68167	26201	30506	41861	1058	43736	1105	24732	76353	25783
49014	99422	50400	68945	42480	56249	22357	18104	1320	39376	11763	16406	17979	1462	35926	29252	44605	55987	1346	15548
43686	11000	56839	25605	60648	17920	13959	63739	84870	1225	1552	37564	21145	53756	62470	37928	63726	22025	1993	74654
84442	10782	1714	31300	24483	10299	40841	1857	28823	54491	80584	16194	68647	22673	82733	18526	11807	26807	31199	44875
95941	24766	37214	13977	42822	1771	12336	13178	17771	79210	84809	97359	23782	78109	33074	79797	22435	12220	23127	50246
1777	74016	30764	29431	23020	23556	79017	39198	68931	78042	28669	79374	37237	15685	82352	17689	21792	39820	16668	1257
33733	70442	10973	25852	80849	61858	56111	11615	40971	15883	28556	14014	28094	78713	15331	1867	39863	46743	44429	44673
28049	23000	20856	63467	10084	15932	69739	26988	15436	55801	25243	1482	15658	39782	50165	10574	78385	57033	95254	16831
1412	22829	35188	1410	1270	20673	25926	16717	19338	31167	74208	16520	98429	14601	97002	31251	36961	24655	15357	1786
11846	13038	15034	12063	27902	24329	41822	20671	69651	35421	1224	30970	11560	15000	41753	16681	80119	1098	18880	44734
15136	79915	62949	18283	26455	14433	16355	1647	1887	21710	38133	31141	37162	24320	25556	10616	15956	54920	62587	25394
1825	40668	10588	22805	30063	76886	51395	1681	20474	46180	40709	12699	95485	20790	27345	49824	95492	50069	26663	19242
30662	1419	40762	81563	21818	33281	1852	39696	1284	41118	52855	26088	75085	29807	43504	24473	27655	99995	139	10364
42198	1229	16828	35819	76704	1397	50811	81290	97617	24221	31874	60211	52928	54717	47161	84056	63550	50311	75668	51007
11780	11595	62350	50560	60765	133	22966	40458	24913	16901	67453	11430	36957	61105	36763	20141	35946	81044	51186	33930
61103	24668	38258	26945	68546	176	46784	1425	63741	54188	1042	15028	29545	19989	98017	16217	17310	83783	78012	79654
56278	15636	54521	12177	13703	81120	76544	1300	1128	17981	53036	1190	32290	33850	19782	15791	44880	50632	17408	14015

BILD 4: ZAHLENBLOCK (2) – NICHT ERKENNBARE FÄLSCHUNG (VON LINKS OBEN NACH RECHTS UNTEN LESEN)

UNTER ANDEREM WEIL: DIE ZAHL 1 IN DER ERSTZIFFERN – REIHE EINE HÄUFIGKEIT VON 31.0% AUFWEIST (≈ BENFORD)

ZUSAMMENSTELLUNG:	
ANZAHL GELIEFERTE DATEN --> HANDEINTRAG	500
DAVON ANZAHL AUSREISSER LAUT GRUBBS -> HANDEINTRAG	0
ANZAHL ERFASSTE DATEN INNERHALB $m \leq 10$ UND OHNE AUSREISSER	500
ZUFALLSANTEIL "P*ZF" DER ERFASSTEN DATEN (%)	90.27
ZUFALLSANTEIL BEZ. GELIEFERTER ANZAHL (%)	90.27
KOMPLEMENTÄR - ANTEIL CHAOSVERHALTEN (%)	4.86
CHAOS - ANTEIL AUSREISSER (%)	0
KOMPLEMENTÄRER CHAOS - ANTEIL BEZ. GELIEFERTER ANZAHL (%)	4.86
KOMPLEMENTÄR - ANTEIL HERDENTRIEB (%)	4.87
HERDENTRIEB - ANTEIL AUS $m > 10$ (%)	0
KOMPLEMENTÄRER HERDENTRIEB - ANTEIL BEZ. GELIEFERTER ANZAHL (%)	4.87
KONTROLLSUMME BEZ. GELIEFERTER ANZAHL DATEN (%)	100

BILD 5: PLAUSIBILITÄTSGRAD 90.3 % DER GEFÄLSCHTEN ERSTZIFFER – REIHE AUS BILD 4 → FÄLSCHUNG NICHT ERKENNBAR

AUGUST 2026 /Ba.