The BFP (Benford-Fibonacci-Perez) method validates the consistency of COVID-19 epidemiological data in France and Italy

Abstract

The Benford method can be used to detect manipulation of epidemiological or trial data during the validation of new drugs. We extend here the Benford method after having detected particular properties for the Fibonacci values 1, 2, 3, 5 and 8 of the first decimal of 10 runs of official epidemiological data published in France and Italy (positive cases, intensive care, and deaths) for the periods of March 1 to May 30, 2020 and 2021, each with 91 raw data. This new method – called "BFP" for Benford-Fibonacci-Perez - is positive in all 10 cases (i.e. 910 values) with an average of favorable cases close to 80%, which, in our opinion, would validate the reliability of these basic data.

Keywords: Benford's law, DNA, SARS-CoV2, mRNA, Benford-Fibonacci-Perez

Jean Claude Perez
Department of Maths & Computer Science, Retired (IBM Artificial Intelligence European Lab), France

Correspondence: Jean Claude Perez, Retired (IBM Artificial Intelligence European Lab, Montpellier France), Tel 33 0781181112, Email Jeanclaudeperez2@gmail.com

Received: December 22, 2022 | **Published:** December 30, 2022

Introduction

On the one hand, there is Benford's law (http://www.fusioninvesting.com/2009/11/benfords-law-and-fibonacci-numbers/) which stipulates that the majority of series of measurements more or less linked to natural or biological phenomena are confirmed, if they are now, to this law which is defined as follows: In (http://www.fusioninvesting.com/2009/11/benfords-law-and-fibonacci-numbers/) we note**: « Benford's law,** also called the **first-digit law**, states that in lists of numbers from many real-life sources of data, the leading digit is distributed in a specific, non-uniform way. According to this law, the first digit is 1 almost one third of the time, and larger digits occur as the leading digit with lower and lower frequency, to the point where 9 as a first digit occurs less than one time in twenty.

This counter-intuitive result has been found to apply to a wide variety of data sets, like electricity bills, street addresses, stock prices, population numbers, death rates, lengths of rivers, physical and biological (which are very common in nature).

It is named after physicist Franck Benford, who stated it in 1938, although it had been previously stated by Simon Newcomb in 1881.

Particularly, in epidemiology and health drugs trials, this law permets to validate accuracy and réalité of basic data ».

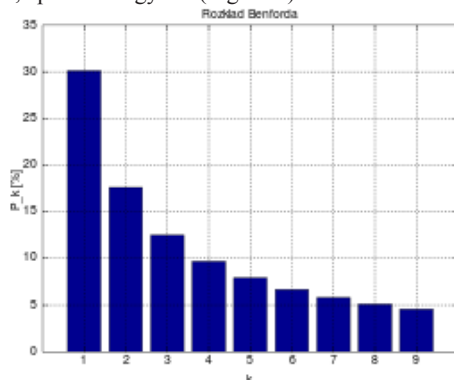This law is used in various areas like stock exchange, social phenomena, epidemiology etc (Figure 1).[1]

Figure 1 Percentages of Benford's law.

This can therefore help detect fraud in scientific publications as well as unintentional errors in these datasets. Often, we present the Fibonacci sequence as an example of a distribution obeying my Benford law fairly well.

On the other hand, there is, precisely, this Fibonacci law

Well known in natural forms: nautilus spiral, sunflower flowers, pineapple, palm trees or pine cones, Fibonacci numbers also control the relative proportions of TCAG nucleotides in DNA: we had already demonstrated this 30years.[2,3] More recently, we have shown that these same Fibonacci proportions of the genome of the mitochondria, the energy source of the human cell, are deteriorated by mutations associated with various cancers.[4] We also demonstrate how these same Fibonacci proportions of DNA make it possible to distinguish a genome of a real bacterium from its attempt at a synthetic chimera.[5]

In the field of SARS-CoV2, its mRNA vaccines, and its multiple variants, we have demonstrated since the start of the COVID-19 pandemic how these Fibonacci numbers offered a new angle for the analysis of mRNA sequences and mutations of SARS-CoV2: a biomathematic point of view of the genome,[6,7] mRNA vaccines or variants.[8] or the last Indian variant "Delta" B.1.617.2.[9]

The paradox which is at the source of our method

On the one hand, Benford's law is often illustrated by its "good correlation" when applied to the Fibonacci sequence, which everyone knows is at the root of many forms of nature.

On the other hand, when we observe this same histogram, taken as proof of Benford's law by the primes, I note, on the contrary, that the (Fibonacci) numbers 1 2 3 5 and 8 differ in this histogram other numbers 4 6 7 and 9 (Figures 2&3, Table 1). It is this observation which will be at the root of our method, then illustrated by this article.
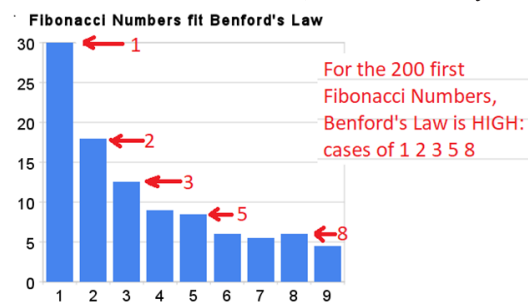

For the 200 first Fibonacci Numbers, Benford's Law is HIGH: cases of 1 2 3 5 8

Figure 2 Percentages of Benford's law over the first 200 Fibonacci numbers.

The BFP (Benford-Fibonacci-Perez) method validates the consistency of COVID-19 epidemiological data in France and Italy
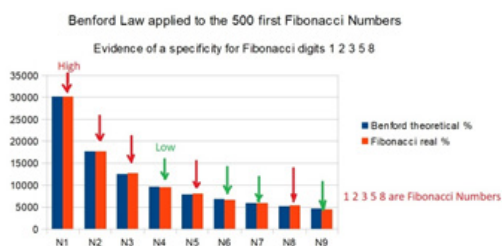
Copyright:
©2022 Perez   **19**

Figure 3 Percentages of Benford's law over the first 500 Fibonacci numbers.

Table 1 Percentages of Benford's law over the first 500 Fibonacci numbers

d	%Théorique	%Observé
1	30.100	**30.130**
2	**17.600**	**17.560**
3	12.490	**12.570**
4	09.691	09.381
5	07.918	**07.984**
6	06.694	06.586
7	05.799	05.788
8	05.115	**05.389**
9	04.575	04.391

What about the "BFP" method running on the firsts Fibonacci numbers? (Table 2). It seems that our "BFP" law is all the more clear that the Fibonacci numbers are small here 27 on the first 34=79.41%.

Table 2 2 clusters partition of the 34 firsts Fibonacci numbers and BFP digits (Benford-Fibonacci-Perez)

Fibonacci	BFP digit
1	1
1	1
2	2
3	3
5	5
8	8
13	1
21	2
34	3
55	5
89	1
144	1
233	2
377	3
610	6
987	9
1597	1
2584	2
4181	4
6765	6
10946	1
17711	1
28657	2
46368	4
75025	7
121393	1
196418	1
317811	3
514229	5
832040	6
1346269	1
2178309	2
3524578	3
5702887	5

Methods and data

Fibonacci numbers:

0 1 1 2 3 5 8 13 21 34 55 89 144 233 377 610 987 1597 2584 4181 6765 10946 17711 28657 46368 75025 121393 196418 317811 514229 832040 1346269 2178309 3524578 5702887.

For any whole number in the list, consider only its decimal with the highest weight decimal.

Example

13 ==> 1

3398 ===> 3

4765 ===> 4

If the selected decimal digit belongs to fibonacci 1 2 3 5 8 do +1

Otherwise 4 6 7 9 0 do +0

We then calculate the % of positives/total.

Basic datas

Main data sources from: For France

https://www.data.gouv.fr/fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19/

For Italy

https://www.sciencedirect.com/science/article/pii/S2352340920304200 (Tables 3&4)

Table 3 Italy: from 1 March to 30 May 2020 and 2021

Positive cases		Death		Intensive care	
2020	2021	2020	2021	2020	2021
566	13114	12	246	140	2289
342	17083	11	343	166	2327
466	20884	27	347	229	2411
587	22865	28	339	295	2475
769	24036	41	297	351	2525
778	23641	49	307	462	2571
1247	20765	36	207	567	2605
1492	13902	133	318	650	2700
1797	19749	97	376	733	2756
977	22409	168	332	877	2827
2313	25673	196	373	1028	2859
2651	26824	189	380	1153	2914
2547	26062	250	317	1328	2982
3497	21315	175	264	1518	3082
3590	15267	368	354	1672	3157
3233	20396	349	502	1851	3256
3526	23059	345	431	2060	3317
4207	24935	475	423	2257	3333
5322	25735	427	386	2498	3364
5986	23832	627	401	2655	3387
6557	20159	793	300	2857	3448
5560	13846	651	386	3009	3510
4789	18765	601	551	3204	3546
5249	21267	743	460	3390	3588
5210	23798	683	460	3489	3620
6203	23987	712	457	3612	3628
5909	23839	919	380	3732	3635
5974	19611	889	297	3856	3679
5217	12916	756	417	3906	3721
4050	16017	812	529	3981	3716

Table Continued...

Positive cases		Death		Intensive care	
4053	23904	837	467	4023	3710
4782	23649	727	501	4035	3681
4668	21932	760	481	4053	3704
4585	21261	766	376	4068	3714
4805	18025	681	326	3994	3703
4316	10680	525	296	3977	3737
3599	7767	636	421	3898	3743
3039	13708	604	627	3792	3683
3836	17221	542	487	3693	3663
4204	18938	610	718	3605	3603
3951	17567	570	344	3497	3588
4694	15746	619	331	3381	3585
4092	9789	431	358	3343	3593
3153	13447	566	476	3260	3526
2972	16168	602	469	3186	3490
2667	16974	578	380	3079	3417
3786	15943	525	429	2936	3366
3493	15370	575	310	2812	3340
3491	12694	482	251	2733	3311
3047	8864	433	316	2635	3244
2256	12074	454	390	2573	3151
2729	13844	534	364	2471	3076
3370	16050	437	360	2384	3021
2646	14761	464	342	2267	2979
3021	13817	420	322	2173	2894
2357	13158	415	217	2102	2862
2324	8444	260	301	2009	2849
1739	10404	333	373	1956	2748
2091	13385	382	344	1863	2711
2086	14320	323	288	1795	2640
1872	13446	285	263	1694	2583
1965	12965	269	226	1578	2522
1900	9148	474	144	1539	2524
1389	5948	174	256	1501	2490
1221	9116	195	305	1479	2423
1075	10585	236	267	1427	2368
1444	11807	369	258	1333	2308
1401	10554	274	207	1311	2253
1327	10176	243	224	1168	2211
1083	8292	194	139	1034	2192
802	5080	198	165	1027	2158
744	6946	251	179	999	2056
1402	7852	172	262	952	1992
888	8085	195	201	893	1893
992	7567	262	182	855	1860
789	6659	242	136	808	1805
875	5753	153	93	775	1779
675	3455	145	140	762	1754
451	4452	99	201	749	1689
813	5506	162	149	716	1643
665	5741	161	164	676	1544
642	5218	156	218	640	1469
652	4717	130	125	595	1430
669	3995	119	72	572	1410
531	2490	50	110	553	1382
300	3224	92	166	541	1323
397	3937	78	121	521	1278
584	4147	70	171	505	1206
593	3738	87	126	475	1142
516	3351	111	83	489	1095
416	2949	75	44	450	1061

Table 4 France: from 1 March to 30 May 2020 and 2021

Positive cases		Death	
2020	2021	2020	2021
43	20412	0	114
23	20453	0	375
48	19786	1	410
34	21912	1	322
73	13157	0	278
138	2364	3	405
179	29327	2	196
103	23466	1	127
410	23706	9	358
286	23945	11	356
371	26255	3	264
497	17026	15	265
586	4135	13	290
770	38276	18	168
818	30555	12	132
923	30439	36	333
1198	30375	21	392
1070	33123	27	236
1377	20670	69	269
1846	4986	128	267
1595	46270	78	184
1847	38088	112	138
1529	37136	112	343
3794	37079	186	269
2410	39932	240	245
2895	24890	230	223
3866	5122	365	897
3748	52323	297	156
4611	43554	319	158
2491	39629	291	360
4341	38379	415	361
7500	38483	497	299
4784	24320	505	304
2066	3775	471	318
5209	8602	2003	185
4221	53843	1053	157
1850	39110	518	228
3881	35899	832	412
3737	37967	1417	433
3869	23785	540	331
4256	3728	1339	319
4333	46288	987	207
3104	34343	635	176
1595	30754	560	373
2669	31275	572	345
5483	32128	762	295
2621	20291	1438	297
2623	4515	753	302
385	39723	760	177
2558	30755	636	169
742	29142	395	446
2050	26869	544	381
2638	28363	531	313
1810	17221	544	283
1623	3476	516	331
1755	32465	389	166
1537	24812	368	174
404	21077	242	391
1160	20966	437	315
1025	22575	366	344
1601	3828	427	306
1055	3376	288	290
601	27245	218	164

Citation: Perez JC. The BFP (Benford-Fibonacci-Perez) method validates the consistency of COVID-19 epidemiological data in France and Italy. *Int J Vaccines Vaccin.* 2022;7(1):18–22. DOI: 10.15406/ijvv.2022.07.00115

Table Continued...

Positive cases		Death	
736	20203	164	113
250	17478	135	279
534	15889	306	255
1050	17918	326	273
4176	3888	278	219
512	1562	177	206
598	23884	242	205
430	16618	80	86
65	15762	70	319
863	1571	262	202
977	17164	348	182
985	8576	81	160
213	2262	349	173
73	18820	130	100
1110	15432	88	81
963	13220	66	195
1027	11514	186	192
131	11986	124	169
761	4817	109	133
226	12572	83	121
101	11821	74	75
692	9704	43	70
689	1908	34	61
649	1749	64	178
587	12438	98	172
552	13704	66	141
316	11090	65	94
126	10007	52	95

Results and discussion

Basic results

Here is the expected result on the public data covid19 in France and in Italy between March 1 and May 30 in 2020 and in 2021. Then, having the RELIABILITY of these basic data, we will illustrate an example of application: bravais correlations pearson in France (data smoothed over 7 sliding days) on time between positive test and death.

Synthetic results: Test "BFP" method to validate SARS-CoV2 epidemiologic data.

Italy

Positive cases 2020: 65/91=71.4%

Death 2020: 53/91 =58.2%

Positive cases 2021: 80/91=87.9%

Death 2021: 71/91=78.02%

Intensive care 2020: 73/91=80.2%

Intensive care 2021: 91/91=100%

France

Positive cases 2020: 63/91=69.2%

Death 2020: 83/91 =91.2%

Positive cases 2021: 65/91=71.4%

Death 2021: 81/91 =89%

Average 725/910=79.67% for 10 batches with 91 cases each, then a total of 910 cases. It seems that "BFP" law is all the more clear that the Fibonacci numbers are small here 27 on the first 34=79.41%. We notice that everything is > in 2021 than in 2020.

2020: 64 + 53 + 73 + 63 + 65=318/455=69.89%

2021: 80 + 71 + 91 + 83 + 81=406/455=89.23%

How to explain?

It may be because the 2021 values are > the 2020 values. So the method would prefer larger values?

Comparing with random values

The results obtained here, that is to say nearly 80% success for 910 real values cumulating 10 races of 91 values each coming from epidemiological measurements in France and Italy, are they GREATER than what would be produced by CHANCE? To answer this question, we performed 100 random batches, each simulating 910 representative random values, for a total of 91,000 random tests. Here are the results:

While the number of successes of real cases is 725 favorable cases (first significant number=1,2,3,5 or 8), the 100 batches produce an average number of successes of 667.96 with random values between 641 and 697 (Figure 4).
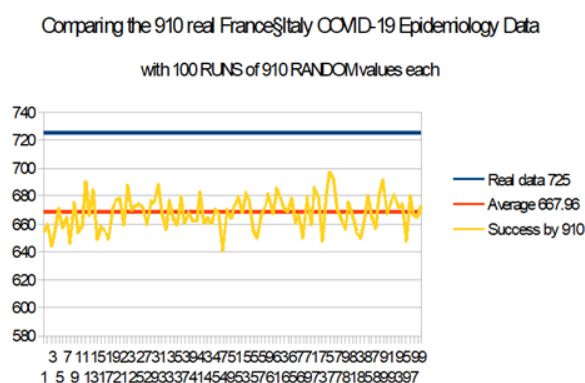

Figure 4 Comparing the 910 real France§Italy results with 100 RANDOM RUNS, each simulating 910 random values between 1 and max value France§Italy (i.e. [1, 53843].

Out of curiosity we tested the same technique, no longer on the first but on the last digit: nothing happens which confirms the strong meaning of the first digit when it takes the values 1 2 3 5 8.

See example last digit here:

last digit out of the 910 France§ Italy values …

Result

456 (against 725 for the first digit) 100 random last digit test simulations with 910 cases each:

Positive results

435 443 465 458 463 440 478 452 479 457 446 432 483 450 465 440 455 463 465 450 468 483 443 466 480 457 441 469 449 435 469 449 485 447 432 453 449 477 448 453 460 471 456 446 457 446 408 468 476 452 471 442 472 447 447 482 428 466 484 435 444 455 460 460 452 460 442 431 461 455 444 448 462 447 459 439 433 463 439 476 478 447 442 443 463 456 472 477 446 455 459 460 448 476 428 483 443 460 427 443

Average result score : 455.22

We bring here the proof that this remarkable property of the first digit disappears completely when considering the last digit.

Example of application: (Table 5 & Figure 5).

The BFP (Benford-Fibonacci-Perez) method validates the consistency of COVID-19 epidemiological data in France and Italy

Copyright:
©2022 Perez   **22**

Table 5 Comparing France Covid-19 March to May years 2020 and 2021, distances between positive case and death using Bravais-Pearson method on 7 days average splines values

TEST TO DEA	2020	2021
7days	7514	6625
8days	8063	7154
9days	8577	7647
10days	9067	7883
11days	9408	8042
12days	9658	8212
13days	9832	8364
14days	9836	8419
15days	9731	8458
16days	9530	8482
17days	9187	8506
18days	8795	8496
19days	8308	8422
20days	7708	8338
21 days	7103	8195
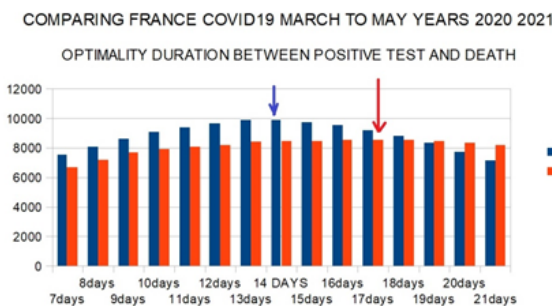
Figure 5 Comparing France COVID19 distance between positive test and death for both periods March-May in 2020 and 2021.

Conclusion

Benford's law already makes it possible to validate or doubt the relevance, reliability and non-manipulation of batches of natural or medical data. What we are proposing today is beyond this Benford law, it is a PARTITION of the first 9 digits (or 10 when, as here, there is also some null data) in 2 clusters: Fibonacci cluster (1 2 3 5 8) and non-Fibonacci cluster (0 4 6 7 9). We suggest that the Fibonacci numbers cluster are all the more in the majority the more the data set is reliable and real. This constitutes a breakthrough in the analysis of natural, social and medical data. This method and the prospects that it should now be consolidated and deepened.

Finally, we have demonstrated by 91,000 random values draws that the "BFP" law applied to the 910 COVID-19 epidemiological values of France and Italy studied here produces results which cannot result from mere chance.

Conflicts of interest

Author declare there are no conflicts of interest

References

1. Sarkar T. *What is Benford's Law and why is it important for data science?* Canada: Towards Data Science; 2018.

2. Perez JC. Chaos DNA and Neuro–computers: A Golden Link". *Speculations in Science and Technology*. 1991;14(4):0155–7785.

3. Perez JC. DNA decrypted. Belgium: *Marco Pietteur*; 1997. 400p.

4. Perez JC. Sapiens Mitochondrial DNA Genome Circular Long Range Numerical Meta Structures are Highly Correlated with Cancers and Genetic Diseases mtDNA Mutations. *J Cancer Sci The*. 2017;9(6):512–527.

5. Perez JC. Epigenetics Theoretical Limits of Synthetic Genomes: The Cases of Artificials Caulobacter (C. eth–2.0), Mycoplasma Mycoides (JCVI–Syn 1.0, JCVI–Syn 3.0 and JCVI_3A), E–coli and YEAST chr XII. *Global Journal of Science Frontier Research: GBio–Tech & Genetics*. 2022;22(2):35–47

6. Perez JC. Wuhan COVID–19 Synthetic Origins And Evolution. *International Journal of Research – Granthaalayah*. 2020;8(2): 285–324.

7. Perez JC, Montagnier L. COVID–19, SARS and Bats Coronaviruses Genomes unexpected Exogeneous RNA Sequences. OSF Preprints. 2020.

8. Perez J C, Montagnier L. Covid–19, Sars And Bats Coronaviruses Genomes Peculiar Homologous RNA Sequences. *International Journal Of Research –GRANTHAALAYAH*. 2020;8(7):217–263.

9. Perez J C, SARS–CoV2 Variants and Vaccines mRNA Spikes Fibonacci Numerical UA/CG Metastructures. Preprints. 2021.

10. Perez J C, DUF1220 Homo Sapiens and Neanderthal fractal periods architectures breakthrough. *SDRP Journal of Cellular and Molecular Physiology*. 2017;1(1):25–49.

11. Perez J.C, Codex biogenesis – Les 13 codes de. 2009.

12. (Perez, 2018), Perez, J.C. Six Fractal Codes of Biological Life:perspectives in Exobiology, Cancers Basic Research and Artificial Intelligence Biomimetism Decisions Making. *Journal of Medical Informatics and Decision Making*. 2018;1(4):22–79.