

Prawo Benforda

Anastasiia Sichko

- Prawo Benforda (znane też jako prawo Newcomba–Benforda) określa rozkład cyfr w danych liczbowych
- Zaobserwowane przez Simona Newcomba (1881) i formalnie opisane przez Franka Benforda (1937)
- "Liczby nie są rozłożone równomiernie; zamiast tego, cyfry wiodące wykazują osobliwą, powtarzalną regularność."
- Stosowane w analizie statystycznej i wykrywaniu nadużyć w danych

Historia odkrycia

- Jak nuda w bibliotece doprowadziła Simona Newcomba do wielkiego odkrycia
- Artykuł jednak się nie spotkał z szerokim zainteresowaniem i zjawisko to zostało zapomniane na 57 lat
- To samo miejsce, inny czas
- Zafascynowany tym zjawiskiem, Benford zaczął sprawdzać, czy jego teoria znajduje potwierdzenie również w innych zbiorach danych

Historyczne Dane Benforda

PERCENTAGE OF TIMES THE NATURAL NUMBERS 1 TO 9 ARE USED AS FIRST DIGITS IN NUMBERS, AS DETERMINED BY 20,229 OBSERVATIONS

Group	Title	First Digit									Count
		1	2	3	4	5	6	7	8	9	
A	Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
B	Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
C	Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
D	Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
E	Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
F	Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
G	H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
H	Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
I	Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
J	Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
K	n^{-1} , $\sqrt{n}$, ...	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
L	Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
M	<i>Digest</i>	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
N	Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
O	X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
P	Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Q	Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
R	Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
S	$n^1, n^2 \dots n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
T	Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
Average		30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
Probable Error		± 0.8	± 0.4	± 0.4	± 0.3	± 0.2	± 0.2	± 0.2	± 0.2	± 0.3	—

Przykładowe Prawdopodobieństwa

Cyfra	$P(D = d)$
1	$\log_{10}(2) \approx 0.301$
2	$\log_{10}(1.5) \approx 0.176$
3	$\log_{10}(1.\overline{3}) \approx 0.125$
4	$\log_{10}(1.25) \approx 0.097$
5	$\log_{10}(1.2) \approx 0.079$
6	$\log_{10}(1.\overline{16}) \approx 0.067$
7	$\log_{10}(1.\overline{142857}) \approx 0.058$
8	$\log_{10}(1.125) \approx 0.051$
9	$\log_{10}(1.\overline{1}) \approx 0.046$

Formalne Sformułowanie

Prawo Benforda stwierdza, że prawdopodobieństwo, iż pierwszą cyfrą jest d (gdzie $d \in \{1, 2, \dots, 9\}$), dane jest:

$$P(D = d) = \log_{10}\left(1 + \frac{1}{d}\right)$$

Przykład

Prostym przykładem jest prawdopodobieństwo $P(31)$ – pierwszą cyfrą jest 3, natomiast drugą – 1. Jest to równoważne prawdopodobieństwu, że liczba znajduje się w przedziale $(31, 32)$, bądź dowolnej potędze 10 tego przedziału:

$$P_{31} = \frac{\int_{31}^{32} \frac{dx}{x}}{\int_{10}^{100} \frac{dx}{x}} = \log \left(1 + \frac{1}{31} \right)$$

Własności zbiorów podlegających Rozkładowi Benforda

Nie wszystkie duże, naturalne (niezmanipulowane) zbiory danych podlegają Prawu Benforda. Można wymienić kilka warunków koniecznych, które muszą być spełnione, aby wynik obserwacji był wiarygodny:

- Zbiór danych nie może posiadać górnej ani dolnej granicy przyjmowanych wartości
- Zestaw danych powinien posiadać więcej liczb o niższej wartości, niż wyższej
- Zbiór danych nie może posiadać żadnych przypisanych numerów. Służą one jedynie do identyfikacji i nie są wynikiem naturalnych procesów. Numery pracowników, numery kont bankowych, numery telefonów nie podlegają Prawu Benforda

Cechy charakterystyczne

- Niezmiennność skali – mnożąc wszystkie wartości w Zbiorze Benforda przez stałą otrzymujemy inny, nowy Zbiór Benforda. Cecha ta jest szczególnie istotna w odniesieniu do przeliczania kursów walut
- Niezmiennność bazy – Theodore P. Hill (1995) udowodnił, że Prawo Benforda dotyczy nie tylko zbiorów opartych na dziesiętnym systemie liczbowym, ale również wszystkich innych systemach
- Niezmiennność w operacjach matematycznych – potęgowanie, mnożenie, dzielenie, dodawanie oraz odejmowanie Zbiorów Benforda ze sobą powoduje powstanie nowego Zbioru Benforda

Przykłady z Rzeczywistości

- Dane finansowe (bilanse firm, dochody, dane księgowe).
- Dane demograficzne (populacje miast, wielkości rzek).
- Dane fizyczne (konstante fizyczne, wyniki pomiarów).

Czy można stwierdzić, że liczba jest błędna, po prostu na nią patrząc?

Date of Check	Amount
October 9, 1992	\$1,927.48
	27,902.31
October 14, 1992	86,241.90
	72,117.46
	81,321.75
	97,473.96
October 19, 1992	93,249.11
	89,658.17
	87,776.89
	92,105.83
	79,949.16
	87,602.93
	96,879.27
	91,806.47
	90,231.83
	93,766.67
	88,336.72
	96,412.21
	71,552.16
TOTAL	\$1,878,687.58

Czy można stwierdzić, że liczba jest błędna, po prostu na nią patrząc?

- Podświadomie menedżer powtarzał niektóre cyfry i kombinacje cyfr
- Wśród pierwszych dwóch cyfr wymyślonych kwot, 87, 88, 93 i 96 zostały użyte dwukrotnie. W przypadku dwóch ostatnich cyfr, 16, 67 i 83 zostały zduplikowane
- Cyfry od 7 do 9 były najczęściej używane, w przeciwieństwie do prawa Benforda
- Kontroler finansowy zaznajomiony z Prawem Benforda badając te liczby od razu spostrzeże, że nie pokrywają się one z modelem, a tym samym zasługują na głębszą uwagę.

Prawo Benforda a ściąganie

- Czy studenci biorący udział w teście mogliby poradzić sobie lepiej niż z przypadkowym zgadywaniem, gdyby zamiast tego postępowali zgodnie z prawem Benforda? Teoria sugeruje, że gdyby zastosowali prawo Benforda, powinni być w stanie uzyskać co najmniej 51 procent poprawnych odpowiedzi
- Aby się o tym przekonać, został stworzony własny egzamin wielokrotnego wyboru składający się z 500 pytań (każde z trzema opcjami) wraz z zestawem danych rzeczywistych odpowiedzi z prawdziwego testu. Niepoprawne odpowiedzi były reprezentowane losowo.
- Aby przeprowadzić test, naukowcy wybrali najniższą możliwą liczbę dla każdej odpowiedzi. Analizując wyniki, odkryli, że rzeczywiście mogli uzyskać 51 proc. punktów w teście, bez znajomości podstawowych pojęć fizyki

Prawo Benforda a ściąganie

- Był jednak pewien haczyk: kiedy naukowcy zastosowali tę samą naukę do prawdziwego testu z fizyki, odkryli, że nie byli w stanie osiągnąć dużo lepszego wyniku porównując ze zgadywaniem
- Odkryli bowiem, że błędne odpowiedzi również były zgodne z prawem Benforda
- Zespół badawczy nadal nie jest w stanie wyjaśnić, dlaczego błędne odpowiedzi były zgodne z prawem Benforda

Ograniczenia i Warunki Stosowalności

- Prawo Benforda nie dotyczy wszystkich danych.
- Dane generowane sztucznie lub mające ograniczony zakres mogą nie wykazywać tego rozkładu.
- Stosowalne głównie w przypadku szeroko rozproszonych danych wielkościowych.

Metody Testowania Zgodności z Prawem Benforda

- Test chi-kwadrat
- Test Kolmogorowa-Smirnowa
- Analiza odchyień procentowych od oczekiwanego rozkładu

Przykłady z Fizyki

Dane z mierzonych stałych fizycznych, które są liczbami wielkogabarytowymi, często podlegają Prawu Benforda:

- Pomiary wielkości gwiazd
- Rozkład mas planet i księżyców w Układzie Słonecznym,
- Energia cząsteczek w skali nano
- Natężenie prądów i napięć w obwodach elektrycznych o szerokim zakresie wartości

Różnorodne zjawiska przyrodnicze generują dane rozproszone na wielu rzędach wielkości.

Podsumowanie

- Prawo Benforda jest fascynującą regularnością spotykaną w danych empirycznych.
- Ma zastosowanie w audycie, kontroli danych, oraz naukach przyrodniczych.
- Jego solidne teoretyczne podstawy i praktyczne zastosowania czynią go cennym narzędziem statystyka
- Niestety nie potrafi pomóc w napisaniu egzaminu

Bibliografia

Benford, Frank, *The Law of Anomalous Numbers*, 1937

Prawo Benforda

Bob Yirka, *Researchers apply Benford's law to physics exams to see if they can do better than chance*, 2013

Hill P. Theodore, *Base-invariance implies Benford's law*, 1995

Mark J. Nigrini, *I've Got Your Number*, 1999