PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 103
Prognozowanie w zarządzaniu firmą 2010

Jacek W. Mercik
Politechnika Wrocławska

PROGNOZOWANIE, PRZEWIDYWANIE, WIESZCZENIE – KOLEJNE PODEJŚCIE

Niepewność jest wątkiem i ściegiem
tkanki naszej codzienności

Leszek Kołakowski

Streszczenie: W wielu sytuacjach obserwowana nieskuteczność prognozowania stawia przed nami zasadnicze pytanie o przyczyny takiego wyniku odnoszącego się do przyszłości badanych zjawisk. Występujące ograniczenia wywołane są zarówno przez charakter samych dostępnych danych, jak i przez nasze rozumienie tego, czym jest przyszłość. W artykule wprowadzono trzy różne pojęcia odnoszące się do sposobów określania przyszłości: prognozowanie, przewidywanie i wieszczenie, oraz przeanalizowano ich stosowalność.

Słowa kluczowe: prawo potęgowe, wymiar topologiczny, prognozowanie.

1. Wstęp

Występowanie zjawisk kryzysowych jest najbardziej dobitnym przykładem nieskuteczności technik prognostycznych. Ich *post factum* wyjaśnialność jest tym bardziej zdumiewająca. Prowokuje ona do stawiania pytań, dlaczego pomimo wielu wysiłków dobrze motywowanych możliwością osiągnięcia z tego tytułu znacznych zysków większość wynalazków technicznych, odkryć naukowych, zjawisk społecznych czy ekonomicznych nie została w porę przewidziana? Przyczyn upatrywać można w (rozumianej bardzo szeroko) nieregularności zjawisk, które badamy, w ograniczeniach stosowanych metod czy może w fundamentalnej nieprzewidywalności przyszłości, która z samej natury jest po prostu nieprzewidywalna.

W rozważaniach przedstawionych w tej pracy wychodzimy z następujących przesłanek odnoszących się do badanej rzeczywistości:

– więcej niż 80% danych rzeczywistych [Taleb 2008b] ma inne cechy niż normalność, w szczególności tzw. grube ogony rozkładów, podleganie tzw. prawom potęgowym,
– rzeczywistość raczej nie płynie, ale skacze (decydują zdarzenia rzadkie, których statystyczna przewidywalność jest niewielka).

Jakie zatem są możliwości stosowania prognozowania (przewidywania, wieszczenia)? W pierwszej części artykułu przedstawione będą tzw. prawa potęgowe jako ogólne narzędzie do modelowania zjawisk losowych. Następnie powiązano to ze strukturą topologiczną badanych zjawisk. W konsekwencji można było zarysować granice stosowalności rozważań odnoszących się do przewidywania przyszłości.

2. Prawa potęgowe

Nazywa się tak związki między zmiennymi, w których zależność ma charakter potęgowy. Formalnie rzecz biorąc, dwie takie zmienne połączone są związkiem $f(x) = ax^{-k} + o(x^k)$, gdzie a i k to stałe, $k > 0$, a $o(x^k)$, to nieskończenie mała rzędu wyższego niż x^k. Wykładnik potęgi k zwany jest wykładnikiem skalującym, co oznacza, że funkcja zachowuje swoją postać niezależnie od skali, w której jest stosowana[1].

Przykłady takich praw są liczne, np.: prawo Stefana-Boltzmanna, prawo śmiertelności Gompertza, prawo stresu Ramberga-Osgooda, prawo odwróconego pierwiastka w grawitacji Newtona, funkcja masy początkowej, korekcja gamma intensywności światła przy podanym napięciu, prawo Kleibera opisujące zależność zwierzęcej masy i metabolizmu, krzywa efektów wynikających z doświadczenia, prawo odwróconego pierwiastka, fraktale, zasada Pareto (zwana też zasadą 80-20), prawo Zipfa (populacja miasta jest odwrotnie proporcjonalna do rangi danego miasta względem populacji), prawo Lotki, prawo wielkości trzęsień ziemi (Gutenberga-Richtera), prawo Hortona opisujące systemy rzeczne, prawo Richardsona opisujące liczbę ofiar konfliktów wojennych i terroryzmu, liczbę wyznawców danej religii, dochody netto poszczególnych osób, częstość występowania słów w tekście, liczbę odwiedzin stron webowych i wiele innych[2].

Jako ciekawostkę można podać tzw. prawo Bradforda[3] − zasadę głoszącą, że w każdej dziedzinie nauki istnieje pewien stały, dość nieliczny zestaw najważniejszych czasopism, w których są drukowane niemal wszystkie wartościowe publikacje z danej dziedziny, pozostałe wydawnictwa nie wnoszą zaś prawie nic do jej rozwoju. Swoistą odmianą tego systemu w skali globalnej jest lista Impact Factor prowadzona przez Instytut Filadelfijski. Bradford stwierdził, że jeżeli zbierzemy czasopisma na jakiś temat i uporządkujemy je w kolejności malejącej pod względem artykułów

---

[1] Zauważmy, że logarytm funkcji $f(x) = ax^{-k} + o(x^k)$ to $\log f(x) = -k \log x + \log a$, a zatem zmiana k przesunie tylko to liniowe wyrażenie na osiach logarytmicznych, nie zmieniając jego kształtu.

[2] Przegląd (a także dowód występowania wśród firm europejskich) znaleźć można np. w [Solomon, Richmond 2001; 2002] lub [Fujiwara i in. 2008].

[3] Samuel C. Bradford, bibliotekarz Science Museum w Londynie. W ramach oszczędności zaproponował on dyrekcji Science Museum, aby analizować rejestr czytelnictwa czasopism naukowych w jego bibliotece i na tej podstawie co roku decydować, które czasopisma prenumerować dalej, a których prenumerat zaprzestać.

w nich zawartych, to możemy wyróżnić jądro oraz strefy. Przykładowe wartości wykładnika potęgowego dla wybranych zjawisk przedstawiono w tab. 1.

Tabela 1. Wartości wykładnika w prawie potęgowym dla wybranych zjawisk

Zjawisko	Wykładnik
Częstotliwość użycia słów w języku angielskim	1,2
Liczba kliknięć w sieci	1,4
Liczba książek sprzedanych w USA	1,5
Rozmowy telefoniczne odebrane	1,22
Siła trzęsienia ziemi	2,8
Średnica kraterów na księżycu	2,14
Intensywność rozbłysków słońca	0,8
Dochód netto na osobę w USA	1,1
Liczba osób noszących nazwisko rodowe	1
Ruchy na rynku	3
Wielkość firm	1,5

Źródło: na podstawie [Newman 2005; Taleb 2008a].

Rozkład prawdopodobieństwa typu prawa potęgowego to dowolna funkcja prawdopodobieństwa, która najogólniej rzecz ujmując, ma postać: $p(x) \approx K(x)x^{-a}$, gdzie $\alpha > 1$ oraz $\lim_{x\to\infty} \dfrac{K(tx)}{K(x)} = 1$ dla dowolnej stałej t.

Ponieważ żądamy, aby $p(x)$ była asymptotycznie inwariantna względem skali, forma $K(x)$ ma być taka, aby móc kontrolować kształt i skończone rozszerzenie dolnego ogona rozkładu. W wielu przypadkach tak nie jest, więc wygodnie jest określić dolną wartość ($x_{\min}$), od której dane prawo potęgowe zachodzi. Otrzymujemy stąd:

$$p(x) = K(x)x^{-\alpha} = (\alpha-1)x_{\min}^{\alpha-1}x^{-\alpha} = \frac{\alpha-1}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\alpha},$$

gdzie: $(\alpha-1)x_{\min}^{\alpha-1}$ to tzw. stała normująca [Fisz 1969], czyli

$$\int_{x_{\min}}^{\infty} p(x)dx = (\alpha-1)x_{\min}^{\alpha-1}\int_{x_{\min}}^{\infty} x^{-\alpha}dx = 1.$$

Rozważmy zwykły moment rzędu k, czyli m_k:

$$m_k = \int_{x_{\min}}^{\infty} x^k p(x)dx = \frac{\alpha-1}{\alpha-1-k}x_{\min}^k.$$

Moment ten jest zbieżny tylko dla $k < \alpha - 1$. Wszystkie momenty rzędu $k \geq \alpha - 1$ są rozbieżne[4]. Stąd dla $\alpha < 2$ zarówno średnia, jak i wszystkie momenty wyższego rzędu są nieskończone; dla $2 < \alpha < 3$ istnieje tylko średnia itd. Zatem dla skończonej próbki pobranej z takiej populacji, gdzie $\alpha < 3$, nie zachodzi centralne twierdzenie graniczne. Co więcej, im próbka większa, tym wartość momentu się zwiększa.

Właściwość ta ma kardynalne znaczenie ze względu na następujące fakty:

– powiązanie wykładnika α z wymiarem przestrzeni zjawisk,
– rozbieżną kurtozę będącą miarą odstawania od normalności (w tym przypadku rozbieżność momentu rzędu 4 oznacza całkowity brak przystawalności do rozkładu normalnego).

Przykłady rozkładów typu prawa potęgowego to: rozkład Pareto (ciągły), rozkład Yule'a-Simona (dyskretny), rozkład t-Studenta (rozkład Cauchy'ego to jego przypadek specjalny).

3. Wymiar przestrzeni zjawisk

Przejdźmy teraz do pojęcia wymiaru. Otóż zakładamy, że każde badane zjawisko mierzone jest w pewnej przestrzeni, której punkty pozwalają wyznaczyć stany danego zjawiska. Poszukiwanie regularności w rozkładach tych punktów jest istotą prognozowania: dopiero ich występowanie pozwala na stosowanie metod prognostycznych. Ich brak oznacza, że przyszłości w takim klasycznym rozumieniu prognozować się nie da.

Analizując przestrzeń stanów, z wielu względów zainteresowani jesteśmy wyznaczeniem minimalnej liczby współrzędnych takiej przestrzeni w taki sposób, aby z jednej strony móc zbudować w miarę dobry model danego zjawiska (teoretycznie im więcej współrzędnych, tym opis dokładniejszy), z drugiej zaś, by móc dokonywać jego interpretacji w dającej się ogarnąć wizualnie przestrzeni[5] (a zatem wymiar powinien być nie większy niż 3). Intuicyjnie wyczuwamy, że wypełnienie danej przestrzeni stanami (trajektoriami) danego zjawiska nie musi być gęste. Zasadne wtedy jest określenie poziomu wypełnienia przestrzeni o danej liczbie wymiarów. Stąd już krótka droga do pojęcia wymiaru przestrzeni, który niekoniecznie musi być całkowitoliczbowy.

W literaturze spotkać można m.in. następujące rodzaje metryk (miar):

- wymiar topologiczny,
- wymiar Hausdorffa,
- wymiar fraktalny,
- wymiar samopodobieństwa,
- wymiar korelacyjny,
- wymiar pudełkowy,
- wymiar pojemnościowy,

[4] Tylko wtedy górna granica całki $-\lim_{x\to\infty} x^{1-(a-k)} = 0$.

[5] Takie podejście właściwe jest dla wielu technik statystycznych, np. dla analizy czynnikowej.

- wymiar informacyjny,
- wymiar Lapunowa,
- wymiar euklidesowy,
- wymiar cyrklowy,

wśród których nie wszystkie są równoważne: ogólnie należy przyjąć, że funkcje wymiaru dają różne wyniki. Prześledźmy to na przykładzie trójkąta Sierpińskiego[6].

Niech T będzie trójkątem ABC.

- Dzieląc T na cztery mniejsze trójkąty: T_1, T_2, T_3 i S, gdzie środki krawędzi są wierzchołkami trójkąta S, traktując S jako zbiór otwarty, a trójkąty T_{i} jako zbiory domknięte, otrzymuje się zbiory rozłączne: S i $T_1 \cup T_2 \cup T_3$. Środki krawędzi leżą w dwóch małych trójkątach (np. $T_1 \cup T_2$ zawiera dokładnie jeden punkt – środek odpowiedniej krawędzi).
- Każdy trójkąt T_i dzieli się w podobny sposób na cztery mniejsze trójkąty: $T_{i,1}$, $T_{i,2}$, $T_{i,3}$ i S_i.
- Każdy trójkąt $T_{i,j}$ dzieli się na cztery mniejsze trójkąty $T_{i,j,1}$, $T_{i,j,2}$, $T_{i,j,3}$ i $S_{i,j}$ itd. (początkowe efekty tego algorytmu pokazano na rys. 1).

Rys. 1. Trójkąt Sierpińskiego otrzymywany w pięciu kolejnych krokach

Wiadomo, że stosunek pól płaskich (wymiaru 2) figur podobnych równa się kwadratowi skali ich podobieństwa. Na przykład figura podobna do innej w skali 3 ma dziewięć razy większe pole od tamtej ($9 = 3^2$ albo $2 = \log_3 9$). Umożliwia to wprowadzenie pojęcia samopodobieństwa, kiedy wymiar samopodobieństwa figury daje się określić jako logarytm o podstawie równej skali podobieństwa i liczbie logarytmowej wskazującej, ile razy większa od figury wyjściowej (jaką częścią figury wyjściowej) jest figura podobna do niej w tej skali. Dla fraktali liczba ta może nie być całkowita. W związku z tym trójkąt Sierpińskiego jest podobny do swoich trzech części w skali 2, a jego wymiar Hausdorffa jest równy $d = \log 3/\log 2 = 1{,}584962501$[7]. Widzimy więc, że wymiar topologiczny trójkąta Sierpińskiego wynosi 1 (zbiór daje się rozciąć pojedynczymi punktami), miara Lebesque'a

[6] Wacław Sierpiński pokazał ten trójkąt w roku 1915. Trójkąt Sierpińskiego jest fraktalem, jednak pojawił się na długo przed spopularyzowaniem tego pojęcia przez Benoita Mandelbrota. Trójkąt Sierpińskiego jest tzw. fraktalem deterministycznym (matematycznym) w odróżnieniu od fraktali naturalnych (czyli losowych).

[7] Ogólnie rzecz biorąc, jeśli założymy, że każda część jest podobna do całości w innej skali ri, $i = 1, 2, ..., N$, to wymiar Hausdorffa jest rozwiązaniem równania z niewiadomą s: $\sum_{i=1}^{N} r_i^s = 1$.

(czyli pole) wynosi zero, a miara Hausdorffa − 1,5849. Oznacza to zatem, że nie ma jednoznacznego określenia miary pewnych obiektów, w tym fraktali. Najczęściej stosowana w wymiarowaniu fraktali jest obecnie miara Hausdorffa, choć używa się także czasami wymiaru pudełkowego. W dalszych rozważaniach stosować będziemy metrykę Hausdorffa jako miarę określającą wymiar fraktali.

Miarę Hausdorffa obiektu można zdefiniować ogólnie jako

$$d = \lim_{a \to 0} \frac{\log N(a)}{\log\left(\frac{1}{a}\right)},$$

gdzie: $N(a) = \frac{L(a)}{a}$, $L(a)$ to długość danego obiektu (w tym i fraktala).

Zauważmy, że w odniesieniu do fraktali liniowych z prawa potęgowego wynika, że $L(a) \approx a^{-\varepsilon}$. Stąd otrzymujemy

$$d = 1 + \varepsilon$$

lub równoważnie

$$d = \lim_{a \to 0} \left(1 + \frac{\log L(a)}{|\log a|}\right).$$

Zależność ta pozwala empirycznie wyznaczać wymiar fraktalny obiektu liniowego[8].

Powróćmy teraz do prognozowania. Ogólna zasada jest taka: na podstawie całej wiedzy (głównie historycznej) staramy się przewidywać stany przyszłe. Dotychczasowe obserwacje są punktami w pewnej przestrzeni. Z jednej strony im więcej wymiarów, tym bardziej całościową informacją dysponujemy. Z drugiej strony im więcej wymiarów, tym większy błąd ewentualnych estymacji. Co więcej, znaczna część (większość, jak twierdzi Taleb [Taleb 2008a][9]) zjawisk nie ma charakteru regularnego, którego synonimem jest normalność, pod postacią rozkładu normalnego (Gaussa). Czy zatem do wszystkich zjawisk możemy stosować (nawet wyrafinowane) techniki prognostyczne?

Dokonajmy następującego rozróżnienia:

– jeśli jedynym źródłem wiedzy są zgromadzone obserwacje, to niezależnie od zastosowanej metody sądy dotyczące przyszłości nazywamy prognozowaniem,

---

[8] Na przykład dla obiektów powierzchniowych lub objętościowych (wymiar topologiczny d_t równy 2 lub 3) $d = d_t + \varepsilon$, gdzie ε wiąże obwód z gęstością [Stauffer, Stanley 1995].

[9] Autor twierdzi, że wspólnie z Pallopem Angsupunem przebadał więcej niż 20 milionów zbiorów danych (w tym ponad 98% występujących danych makroekonomicznych) odnoszących się do danych z ostatnich 40 lat i stwierdził, że ponad 80% z nich nie ma charakteru normalnego.

– jeśli do zgromadzonych obserwacji dodamy wiedzę pozaobserwacyjną (np. analogie z innych dziedzin, prawa fizyki przeniesione do ekonomii itp.), to sądy o przyszłości nazywamy przewidywaniem,
– jeśli sąd o przyszłości czynimy nie na podstawie zgromadzonych informacji, to mamy do czynienia z wieszczeniem (np. przewidywanie przyszłości na podstawie wnętrzności zabitego rytualnie zwierzęcia).

Tabela 2. Zakres stosowalności prognozowania, przewidywania i wieszczenia w zależności od wiedzy posiadanej o danym zjawisku

Wyszczególnienie	Istnieją struktury danych	Brak wiedzy na temat struktury danych
Rozkłady, dla których $m_4 < \infty$ lub istnieje co najmniej wykładnik skali	Prognozowanie	Prognozowanie
Inne niż wymienione rozkłady	Prognozowanie/przewidywanie	Wieszczenie

W tabeli 2 przedstawiono, jak wiedza na temat rozkładów prawdopodobieństwa dotyczących badanego zjawiska oraz ewentualna wiedza na temat samego zjawiska odnoszą się do prognozowania, przewidywania lub wieszczenia. Zauważmy, że znajomość struktury danych może być odnajdywana na wiele sposobów, w tym i tak skomplikowanych lub niekoniecznie intuicyjnych jak algorytmy genetyczne czy inne wyrafinowane techniki komputerowe.

4. Podsumowanie

Analizując możliwości stosowania prognozowania, stwierdza się, że w niektórych sytuacjach, nie tak rzadkich jakby się mogło wydawać (ponad 80% danych rzeczywistych nie ma charakteru normalnego!), to, co możemy zastosować, należy nazwać przewidywaniem. Stąd raczej nie jest możliwe określanie prawdopodobieństw występowania określonych sytuacji i mówić należy jedynie o określonych scenariuszach przyszłości. Ich wybór będzie kształtował przyszłość, a więc w sposób paradoksalny tę przyszłość będzie jednak przewidywał.

Wieszczenie, rozumiane jako wypowiadanie się o przyszłości nie na podstawie dotychczasowej historii (dotychczasowych danych), też ma miejsce i swoją rolę w rozważaniach dotyczących przyszłości. Podobnie jak przewidywanie, tak i wieszczenie może mieć charakter samospełniających się przepowiedni, które nakierowując ludzi na określone rozwiązania, spowodują, że przyszłość się zrealizuje. A więc efektywność wieszczenia mierzona *post factum* może się okazywać całkiem wysoka.

Literatura

Fisz M., *Rachunek prawdopodobieństwa i statystyka matematyczna*, PWN, Warszawa 1969.
Fujiwara Y., Di Guilmi C., Aoyama H., Gallegati M., Souma W., *Do Pareto-Zipf and Gibrat Laws Hold True? An Analysis with European Firms*, pre-print submitted to Elsevier Science, 2.02.2008, http://arxiv.org/PS_cache/cond-mat/pdf/0310/0310061v2.pdf, 12.06.2009.
Makridakis S., Taleb N., *Living in a world of low levels of predictability*, ,,International Journal of Forecasting”, in printing, doi:10.1016/j.ijforecast, 2009.05.008.
Newman M.E.J., *Power laws, Pareto distributions and Zipf's law*, ,,Complexity Digest“ 2005, 02, 1-27.
Pareto V., *Le Cours d'´Economie Politique,* Macmillan, London 1897.
Solomon S., Richmond P., *Power laws of wealth, market order volumes and market returns*, ,,Elsevier Science B.V” 2001.
Solomon S., Richmond P., *Stable power laws in variable economies: Lotka-Volterra implies Pareto--Zipf*, The European Physical Journal B 2002, 27, 257-261.
Stauffer D., Stanley H.E., *From Newton to Mandelbrod,* 2 wyd., Springer, 1995.
Taleb N.N., *The Black Swan*, Penguin Books, 2008a.
Taleb N.N., *The fourth quadrant: a map of the limits of statistics*, 2008b, http://www.edge.org/3rd_culture/taleb08/taleb08_index.html, 22.06.2009.

FORECASTING, FORESIGHTING, DIVINING – REVISITED

Summary: The observed inefficiency of forecasting leads to considerations connected with character of our believes about future events. Existing limits are caused by data itself as well as by our understanding of the future. Three different descriptions of the future are introduced: forecasting, foresighting, and divining. The applicability of them is analyzed too.