

H. S T E I N H A U S (Wrocław)

## O PROGNOZIE

Gdy rok temu na seminarium Grupy Zastosowań Przyrodniczych i Gospodarczych Instytutu Matematycznego PAN określiłem, jak należy kwalifikować pragnię, J. Łukaszewicz postawił zagadnienie najlepszej prognozy, na które odpowiedziałem niebawem. Zarówno definicja, jak i odpowiedź prowadzą do niebanalnych konsekwencji, które chcę tutaj przedstawić.

1. *Prognozę* rozumiem w niniejszym artykule ogólnie. To pojęcie obejmuje zarówno przepowiednię meteorologiczną, jak oszacowanie wartości kruszcu w podziemiu obszaru, które geolog daje w swej eksperytyzie. Dokładność przepowiedni pogody stwierdzimy jutro, dokładność oszacowania ilości cynku w podziemiu wtedy, gdy kopalnia zostanie bez reszty wyeksploatowana. Można mówić także o prognozie wstecznej: gdy mówię, że pociąg bezpośredni z Rzymu do Warszawy opuścił Rzym dziś rano, wtedy ta przepowiednia odnosi się do przeszłości, choć jej zgodność z faktami okaże się dopiero jutro. Ważna jest także praktycznie prognoza lekarska, nie tylko ta, która przepowiada dziś, czy chory przeżyje najbliższą dobę, ale i ta, którą nazywamy diagnozą, a która odnosi się do aktualnego stanu pacjenta: twierdzenie, że chory ma raka, jest w tym sensie prognozą, bo jest domniemaniem, którego zgodność z rzeczywistością kiedyś zostanie ustalona pozytywnie lub negatywnie.

Pomijamy tutaj pewną odmianę prognozy, którą nazwałbym *predykcją*. Przykładem predykcji są przepowiednie almanachów nautycznych. Dzisiejsza znajomość teorii ruchu ciał niebieskich oraz doskonałość instrumentów pozwala przepowiedzieć deklinację księżyca w momencie, który nadejdzie np. dopiero za rok, z taką precyzją, że różnica między pozycją przepowiedzianą a tą, która za rok się zrealizuje, nie będzie miała żadnego znaczenia praktycznego dla żeglarza, który od dziś za rok użyje almanachu do ustalenia swojego kursu. Nas właśnie interesują gorsze, niedoskonałe przepowiednie: różnica między temperaturą przepowiedzianą na jutrzejsze południe a tą, która naprawdę jutro o 12-tej zapanuje, ma znaczenie praktyczne nie tylko dla rolnika lub budowniczego, ale także dla właściciela mieszkania, który dziś napelnia koksem



piec centralnego ogrzewania na całą dobę; jak doświadczenie uczy, ta różnica często jest dość duża, żeby dała się odczuć pieniężnie w postaci zbędnego wydatku na koks (jeżeli temperatura będzie wyższa o 10 stopni, niż przewidziano) lub na remont grzejnika (jeżeli będzie o tyle niższa, że zaniechanie palenia spowoduje pęknięcie rur).

2. Jak widać z poprzedniego paragrafu, interesuje nas prognoza jako podstawa decyzji. Ograniczymy się do przepowiedni wartości pewnego parametru liczbowego  $s$ . Teoria i obserwacja dostarczają łącznie pewnej funkcji  $F(x)$ , która jest dystrybuantą zmiennej losowej  $s$  w tym sensie, że dziś nie wiemy, jakie się okaże  $s$ , gdy poznamy jego rzeczywistą wartość, ale na tle teorii i obserwacji określamy prawdopodobieństwo  $P(s < x)$ , że  $s$  okaże się mniejsze od  $x$ , i nazywamy je  $F(x)$ :

$$(1) \quad F(x) = P(s < x).$$

W naszym systemie pojęć nazwiemy  $F(x)$  *informacją* o parametrze  $s$ . Jak każda dystrybuanta, jest  $F(x)$  funkcją lewostronnie ciągłą, niemalejącą i jest  $F(-\infty) = 0$ ,  $F(\infty) = 1$ ; ciągłość odnosi się także do  $-\infty$ .

Właściwie można by ogłaszać prognozę podając  $F(x)$  (graficznie lub tabelarycznie), gdyż niczego więcej nie wiemy (dziś) o tym, jakie będzie  $s$ . Ale prognoza ma służyć za podstawę działania. Mamy tu na myśli akcję, którą należy podjąć dziś tak, jak gdyby już dziś była wiadoma wartość  $s$  (zakładamy, że odroczenie decyzji do jutra wywoła większą szkodę, niż oparcie akcji dzisiejszej na fałszywej hipotezie co do jutrzejszej rzeczywistości).

Przypuśćmy, że działamy dziś tak, jak gdybyśmy wiedzieli, że jutro będzie  $s = y$ . Oznaczmy przez  $O$  zysk z działania opartego na trafnej hipotezie co do  $s$  i nazwijmy *szkodą* z błędnej hipotezy kwotę, o którą zmniejszy się  $O$  przez to, że prawdziwą wartością parametru  $s$  jest  $x$ , a nie  $y$  (co okaże się jutro, po szkodzie!). Przypuśćmy, że ta szkoda jest proporcjonalna do kwadratu błędu, a więc, że jest równa  $c(x-y)^2$ . Wobec tego oczekiwana szkoda z akcji opartej na hipotezie  $s = y$  jest

$$(2) \quad H(y) = c \int_{-\infty}^{\infty} (x-y)^2 dF(x).$$

Odbiorca informacji  $F(x)$ , który podejmuje dziś akcję, powinien oprzeć ją na takiej hipotezie  $s = y$ , która redukuje do minimum oczekiwaną szkodę  $H(y)$ . Optymalne  $y$  znajdujemy z równania  $dH(y)/dy = 0$ , to jest z równania

$$(3) \quad \int_{-\infty}^{\infty} (x-y) dF(x) = 0,$$

które daje

$$(4) \quad y = \int_{-\infty}^{\infty} x dF(x) = \bar{x}.$$

Tak więc przyjęcie za  $s$  wartości  $\bar{x}$ , czyli średniej wartości obliczonej z informacji  $F(x)$ , jest optymalną hipotezą, na której powinien oprzeć swoją akcję odbiorca informacji  $F(x)$ ; mamy tu na myśli sytuację wyżej opisaną. Ale założenie, że szkoda jest proporcjonalna do kwadratu błędu, jest dowolne. Łatwo podać przykłady, w których szkoda nie tylko nie jest proporcjonalna do  $(x-y)^2$ , ale nawet nie jest symetryczną funkcją zmiennych  $x, y$ .

Statystyka odjazdów pociągu, który według rozkładu jazdy ma odchodzić codziennie o 22<sup>h</sup>19<sup>m</sup>, daje informację  $F(x)$  o porze  $x$  dzisiejszego odjazdu. Gdybyśmy stąd obliczyli  $\bar{x}$ , jak każe wzór (4), otrzymalibyśmy porę późniejszą niż 22<sup>h</sup>19<sup>m</sup>, bo nieraz pociągi się spóźniają, ale niemal nigdy nie odchodzą przedwcześnie. Przypuśćmy, że  $\bar{x} = 22^h27^m$ . Ogłoszony oficjalny czas odjazdu 22<sup>h</sup>19<sup>m</sup> jest — praktycznie rzecz biorąc — niczym innym jak prognozą. Dlaczego zarząd kolei nie trzyma się naszej teorii i nie ogłasza 22<sup>h</sup>27<sup>m</sup> jako pory odjazdu? Dlatego, że szkoda nie jest tu funkcją symetryczną błędu. Jeżeli  $x_1 < x_2$  i pasażer działa tak, jak gdyby pociąg miał odejść w chwili  $x_1$ , a pociąg odejdzie w chwili  $x_2$ , to szkoda będzie inna niż wtedy, gdy pasażer uwierzy, że pociąg odejdzie w chwili  $x_2$ , a naprawdę odjazd nastąpi w chwili  $x_1$ : pierwszy błąd spowoduje szkodę równą wartości kilku minut straconych na stacji lub w wagonie, drugi może spowodować poważne szkody przez uniemożliwienie podróży, a co najmniej stratę pieniędzy wydanych na taksówkę. Wobec tego szkoda w tym zagadnieniu prognozy nie jest funkcją symetryczną, w szczególności nie jest równa  $c(x-y)^2$ . Stąd widzimy, że w praktyce mogą wystąpić inne funkcje  $g(x, y)$  wyrażające szkodę z działania pod supozycją  $s = y$ , gdy w rzeczywistości  $s = x$ . Ogólnie będzie zatem — zamiast (2) — wzór

$$(5) \quad H(y) = \int_{-\infty}^{\infty} g(x, y) dF(x)$$

wyrażał oczekiwaną szkodę, co zamiast (4) da, jako  $y$  optymalne, pierwiastek  $y$  równania

$$(6) \quad \int_{-\infty}^{\infty} \frac{\partial g(x, y)}{\partial y} dF(x) = 0.$$

Może zaciekać czytelnika fakt, że swego czasu pruskie koleje przesuwały o minutę w przód zegary stacyjne przeznaczone dla publiczności

w poczekalniach i restauracjach dworcowych, kierując się (intuicyjnie) względami, które nasza analiza uzasadnia.

Tak więc każdy plan akcji wymaga innej supozycji co do wartości parametru  $s$ . Oblicza się tę wartość z informacji uniwersalnej  $F(x)$  i z funkcji  $g(x, y)$ . Ta funkcja jest specyficzna, to znaczy zależna od rodzaju akcji. Dla każdej akcji z osobna wynika „ostra” supozycja  $s = y$  co do wartości parametru i plan należy tak ułożyć, jak gdyby było zupełnie pewne, że ta właśnie wartość się zrealizuje. Ten rezultat jest pozornie sprzeczny z zasadami rozsądnego sceptycyzmu, ale tylko pozornie. Zasady te odpowiadają mianowicie faktowi, że każdy rodzaj akcji  $A_1, A_2, \dots, A_n$  wymaga innej (ostrej) supozycji  $y_1, y_2, \dots, y_n$ ; intuicyjne wyczucie tego faktu wywołuje sceptyczną postawę. Ta postawa staje się źródłem mylnych postanowień, gdy chce się ją zająć względem jednej akcji: tak zwane „uwzględnianie różnych możliwości” jest wtedy błędne; jest ono jednak możliwe i wskazane, gdy można akcję rozszcześcić na kilka niezależnych od siebie akcji różnego typu.

3. Pokażemy teraz bezcelowość rozszczepiania akcji jednorodnej, to jest podlegającej tej samej funkcji określającej szkodę z błędu. Konkretny przykład niech ułatwi zrozumienie naszej tezy. Rozszczepia jednorodną akcję rolnik, który niektóre części łąki nawadnia obficie, inne skąpiej, chociaż te części niczym się nie różnią: czyni to, bo nie wie na pewno, czy ma się spodziewać posuchy czy deszczu; mniema, że powinien zmniejszyć ryzyko szkody z błędnej decyzji obejmując akcją opartą na hipotezie bardziej prawdopodobnej większy obszar niż ten, do którego zastosuje hipotezę mniej prawdopodobną... podobnie postąpiłby gracz na wyścigach, który by podzielił stawkę pomiędzy kilka koni. Matematycznie rzecz się tak przedstawia:

Każde „rozkładanie ryzyka” da się opisać funkcją  $G(y)$ , która ma własności dystrybuanty (podane w § 2). Na przykładzie łąki supozycja, że  $y \leq s \leq y + dy$ , będzie podstawą akcji, którą zastosujemy do pewnej części łąki, a ta część będzie co do rozmiaru frakcją całości równą  $dG(y)$ . Ten ogólny schemat obejmuje wszelkie „rozszczepianie akcji”, a każdy konkretny przykład otrzymamy przez podstawienie za  $G(y)$  konkretnej funkcji. Teraz wzór na oczekiwaną szkodę  $S$  przybierze postać funkcjonału (przedtem szkoda była funkcją)

$$(7) \quad S(G) = c \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-y)^2 dF(x) dG(y).$$

Najlepszy plan otrzymamy przez zminimalizowanie  $S(G)$ . Powtórzmy tu definicję (2):

$$(8) \quad H(y) = c \int_{-\infty}^{\infty} (x-y)^2 dF(x).$$

Z (7) i (8) wynika

$$(9) \quad S(G) = \int_{-\infty}^{\infty} H(y) dG(y).$$

Niech  $y_0$  oznacza miejsce, w którym  $H(y)$  przyjmuje najmniejszą wartość. Wykażemy, że dystrybuanta  $G_0(y)$ , określona wzorem

$$(10) \quad G_0(y) = \begin{cases} 0 & \text{dla } y \leq y_0, \\ 1 & \text{dla } y > y_0, \end{cases}$$

nadaje całce (9) najmniejszą wartość. Rzeczywiście, definicja (10) i wzór (9) dają natychmiast

$$(11) \quad S(G_0) = H(y_0),$$

a twierdzenie o wartości średniej i (9) dają

$$(12) \quad S(G) = H(\eta),$$

gdzie  $\eta$  jest pewną liczbą zależną od  $G$  — ponieważ  $H$  ma minimum w punkcie  $y_0$ , więc (11) i (12) pokazują, że  $G_0$  ma zapowiadaną własność.

Ale z definicji (10) wynika bezpośrednio, że plan działania określony przez  $G_0$  — a udowodniliśmy właśnie, że to jest plan najlepszy — przypisuje supozycji  $y \leq s \leq y + dy$  frakcję 0 powierzchni łąki, jeżeli przedział  $(y, y + dy)$  nie obejmuje  $y_0$ , a całą łakę, jeżeli obejmuje. Tak więc najlepszy plan  $G_0$  każe zastosować do całej łąki prognozę ostrą, którą znaleźliśmy już w § 2, chociaż teraz nie zakładamy z góry, że musimy się zdecydować na jedną prognozę dla całości — ta konieczność wynikła dopiero z dyskusji matematycznej. Ciekawe jest tutaj, że dystrybuanta  $G$ , równa informacji  $F$  też jest gorsza jako ewentualny plan rozkładu decyzji, niż ostra prognoza uniwersalna.

Wprawdzie w niniejszym paragrafie użyliśmy  $(x-y)^2$  jako miary szkody z błędu, ale widać od razu, że dla dowolnej miary  $g(x, y)$  rozumowanie zachowuje się bez zmiany.

4. Jakie jest zadanie instytucji, która dysponuje teorią i zbiera obserwacje prowadzące do informacji  $F(x)$ ? Jeżeli ogłosi  $F(x)$ , to większość konsumentów prognozy nie zdoła obliczyć najlepszej supozycji dla siebie z tej informacji i ze specyficznego charakteru szkody  $g(x, y)$ . Jeżeli zamiast  $F(x)$  da ostrą prognozę  $\bar{x}$ , to dostarczy optymalnej supozycji tym klientom, których szkoda wyraża się przez  $c(x-y)^2$ ; inni nie będą mogli skorzystać z takiej prognozy. Błędne byłoby mniemanie, jakoby znajomość średniej  $\bar{x}$  i funkcji  $g(x, y)$  pozwalały konsumentowi obliczyć specyficzną ostrą prognozę do każdej akcji. Wzory (5) i (6) wykazują, że to jest niemożliwe. Gdyby instytucja — np. instytut

meteorologiczny — chciała ogłaszać prognozy specyficzne, to natknęłaby się na trudności nie do przewyciężenia, w postaci nieznajomości funkcji  $g(x, y)$ , których jest tyle, ile różnych akcji; otrzymanie od konsumentów danych co do postaci tych funkcji byłoby możliwe tylko w wyjątkowych przypadkach, a zdefiniowanie ich bez pomocy konsumentów zadaniem z góry skazanym na fiasko.

Tak więc podana tu teoria prognozy na razie może mieć znaczenie tylko przy akcjach jednolitych, zakrojonych na wielką skalę, w których jest możliwe — i opłaca się — zarówno zebranie danych co do funkcji  $g(x, y)$ , jak i rozwiązywanie równania (6) tylekroć, ilekroć trzeba. Być może, że dałoby się to zrobić np. dla kolejnictwa z prognozą pogody. Ale lepszym przykładem jest tu szacowanie zawartości złóż mineralnych: gdy wiadomo, do czego mają służyć wydobyte kruszce, można by się pokusić o obliczenie właściwej supozycji co do ich zawartości. Wydaje się paradoksalne, że geolog (wraz z matematykiem) nie mogą orzec bez pomocy ekonomisty, jaką supozycję co do zawartości podziemia należy przyjąć za prawdziwą. Zwykle żądano tzw. obiektywnej oceny od geologów i statystyków, a dopiero potem pytano ekonomistów, czy opłaca się eksploatować zbadany obszar; teraz okazuje się, że ekonomiści muszą interweniować już w pierwszej fazie zagadnienia: tak zwane bezinteresowne, czyli czysto naukowe oszacowanie zawartości złóż nie da się bez niej przeprowadzić konsekwentnie, da się jedynie obliczyć informacja  $F(x)$ . Ale i to ostatecznie zadanie podlega krytyce związanej z tak zwanym prawdopodobieństwem *a priori*; to już jednak wychodzi poza ramy tego artykułu.

INSTYTUT MATEMATYCZNY POLSKIEJ AKADEMII NAUK

*Praca wpłynęła 5. 11. 1955*

Г. ШТАЙНХАУЗ (Вроцлав)

## О ПРОГНОЗЕ

### РЕЗЮМЕ

Сообщение о значении, которое примет завтра параметр  $x$ , дает функция  $F(a)$  = вероятность того, что  $x$  будет меньше  $a$ . Для того, чтобы на основании этого сообщения построить последовательный план действия, необходимо сделать его строгим прогнозом,  $x = x_0$ . Это возможно тогда (и только тогда), когда известна функция  $g(x, y)$  определяющая потерю вытекающую из действия основанного на прогнозе  $x$ , если произойдет  $y$ .

И так, когда напр.  $x$  является температурой,  $g(x, y)$  не обозначает для крестьянина того же, что для строителя. Поэтому искомый прогноз  $x_0$  будет различен в обоих случаях. Вследствие того, чисто теоретическое понятие „наилучшего прогноза“ не имеет смысла.

---

H. STEINHAUS (Wrocław)

### ON PROGNOSIS

#### SUMMARY

The information about the value which the parameter  $x$  will take to-morrow is given by the function  $F(\alpha)$  = the probability that  $x$  will be less than  $\alpha$ . To base a consistent course of action on such information one must transform it into a precise prediction  $x = x_0$ . This can be done if (and only if) the damage-function  $g(x, y)$  is known:  $g(x, y)$  = the loss resulting from action based on the prognosis  $x$  if  $y$  proves to be the case. Now, if  $x$  is, for instance, the temperature,  $g(x, y)$  will not be the same for farming and for housebuilding. Thus the sought-for prediction  $x_0$  will be different in the two cases. Consequently, the concept of a purely theoretical best prognosis is meaningless.

---