Przegląd Filozoficzny – Nowa Seria
R. 23: 2014, Nr 4 (92), ISSN 1230–1493

Anna Wójtowicz

Falsyfikacjonizm a test hipotezy zerowej[1]

Słowa kluczowe: *falsyfikacjonizm, test hipotezy zerowej, prawdopodobieństwo, racjonalność wnioskowania*

1. Wstęp

Postęp nauki odbywa się dzięki walce między różnymi teoriami i hipotezami. W jaki sposób – korzystając z dostępnych danych empirycznych, wyników projektowanych eksperymentów czy obserwacji – rozstrzygać, które z nich są prawdziwe? Można zbierać dane, które potwierdzają hipotezę lub szukać danych, które będą ją obalać. Teoretyczną podstawę pierwszej metody stworzył Carnap w swojej logice indukcji, zwolennikiem drugiej był Popper, co znalazło wyraz w jego koncepcji falsyfikacjonizmu. Czy współcześnie idee Poppera mają praktyczne zastosowanie? Jaka jest ich wartość? Jak są one rozumiane? Celem artykułu jest próba przynajmniej częściowej odpowiedzi na te pytania.

2. Stosunek między hipotezami a danymi empirycznymi

Między hipotezą naukową H i zdaniem opisującym wynik obserwacji E mogą zachodzić następujące typy związków:

(a) Z H wynika logicznie E;
(b) Jeśli prawdziwe jest H, to z dużym prawdopodobieństwem E;
(c) Jeśli prawdziwe jest H, to zwiększa to prawdopodobieństwo E;
(d) Z H wynika logicznie nie-E;

[1] Praca powstała w ramach grantu NCN 2012/05/B/HS1/01711.

(e) Jeśli prawdziwe jest H, to z dużym prawdopodobieństwem nie-E;
(f) Jeśli prawdziwe jest H, to zwiększa to prawdopodobieństwo nie-E;
(g) Jeśli E, to zwiększa to prawdopodobieństwo H;
(h) Jeśli nie-E, to zwiększa to prawdopodobieństwo H;
(i) E jest niezależne od H[2].

W języku bardziej sformalizowanym można te związki odpowiednio uściślić i zapisać krótko jako:

(a’) H |= E;
(b’) P(E/H) jest bliskie 1;
(c’) P(E/H) > P(E);
(d’) H |= ~E;
(e’) P(~E/H) jest bliskie 1;
(f’) P(~E/H) > P(~E);
(g’) P(H/E) > P(H);
(h’) P(H/~E) > P(H);
(i’) P(E/H) = P(E);

gdzie przez |= oznaczamy standardową operację wynikania, P jest funkcją prawdopodobieństwa, spełniającą aksjomaty Kołmogorowa, a P(A/B) to prawdopodobieństwo zajścia A, o ile wiem, że zaszło B (definiowane klasycznie jako $P(A \wedge B)$ / P(B)).

Zauważmy od razu, że:

– (b) jest osłabioną wersją (a), podobnie jak (e) jest osłabioną wersją (d). Zależności (b) i (e) nie stwierdzają po prostu, że jeśli hipoteza jest prawdziwa, to musi zajść E (czy nie-E), ale coś ostrożniejszego – że zajście takiego zdarzenia jest wysoce prawdopodobne.

– Zamiast (b’) możemy – ze względu na podstawowe prawa rachunku prawdopodobieństwa[3] – stwierdzić, że P(~E/H) jest bliskie 0. Podobnie zamiast (e’) możemy stwierdzić, że P(E/H) jest bliskie 0.

Tak jak zostało powiedziane we wstępie, jedno z podstawowych pytań stawianych na gruncie metodologii brzmi:

Jak, mając do dyspozycji zdanie stwierdzające, czy zachodzi E, rozstrzygnąć, jaka jest wartość logiczna hipotezy H?

---

[2] Ponieważ zakładamy, że hipoteza H ma mieć charakter zdania ogólnego, a obserwacja E – zdania szczegółowego, więc nie jest możliwe, aby z E wynikało H. Z nie-E (lub z E) może jedynie wynikać nie-H – ale jest to zależność równoważna opisanej odpowiednio w (a) lub w (d).

[3] Dla dowolnego A: P(A/B) = 1 – P(~A/B).

I mamy tu dwie możliwości – za pomocą wiedzy na temat E możemy próbować potwierdzać H lub próbować obalać H.

Możliwości te są – ze względu na swoje własności logiczne – istotnie różne. Jeśli np. uzyskujemy E, które wynika z H (czyli mamy przypadek (a')), to H jest w pewnym stopniu potwierdzone: albo korzystamy z wnioskowania abdukcyjnego (uznając, że H jest najlepszym wyjaśnieniem zajścia E), albo możemy obliczyć, że E zwiększa prawdopodobieństwo zajścia (czyli że zachodzi również (g'))[4]. Nie daje nam to jednak pewności, że H jest prawdziwe, bo schemat wnioskowania:

(S1) *Przesłanki:*

1) Z tego, że hipoteza H jest prawdziwa, wynika, że zajdzie zdarzenie E ($H \models E$);
2) Zaszło zdarzenie E;

Wniosek:

Hipoteza H jest prawdziwa

nie jest schematem dedukcyjnym: mimo prawdziwości przesłanek wniosek może być fałszywy.

Szczególnym przypadkiem[5] takiego wnioskowania jest indukcja enumeracyjna niezupełna, czyli wnioskowanie o schemacie:

(S2) *Przesłanki:*

E_1) Stwierdzamy, że a_1 ma własność W;

...

E_n) Stwierdzamy, że a_n ma własność W;

Wniosek:

Wszystkie a mają własność W.

Z jednej strony jest oczywiste, że wnioskowanie takie jest rozsądne – prawdopodobieństwo hipotezy H zawartej we wniosku, przy założeniu, że prawdziwe są przesłanki, jest większe niż prawdopodobieństwo aprioryczne hipotezy:

$$p(H/\text{przesłanki}) > p(H).$$

---

[4] Analogiczna sytuacja ma miejsce, gdy uzyskamy nie-E – zachodzi wtedy (d') i ewentualnie również (g').

[5] To, że ze zdania „Wszystkie *a* mają własność *W*" wynikają poszczególne przesłanki, jest zakładane milcząco.

Co więcej, dodanie każdej kolejnej przesłanki mówiącej, że a_{n+1} ma własność W, wzmacnia hipotezę zawartą we wniosku. Z drugiej jednak strony, gromadzenie kolejnych przesłanek potwierdzających wniosek nigdy nie da nam pewności, że jest on prawdziwy. Nie widać również dobrego sposobu ustalania, w jakim stopniu (z jaką siłą) przesłanki potwierdzają H, aby można było np. porównywać wiarygodność różnych, alternatywnych hipotez. Próby zdefiniowania tak rozumianego pojęcia konfirmacji (potwierdzania hipotezy przez dane empiryczne) i ogólnie, uzasadnienia indukcji, podejmował m.in. R. Carnap (por. np. Carnap 1950).

Ponieważ stanowisko Carnapa ma wady (zawiera w szczególności wiele arbitralnych rozstrzygnięć), atrakcyjną alternatywą wydaje się druga możliwość – wykorzystywanie danych empirycznych nie tyle jako sposobu potwierdzania hipotezy, ale jako sposobu jej obalania. Wnioskowanie o schemacie:

(S3) *Przesłanki:*

1) Z tego, że hipoteza H jest prawdziwa, wynika, że zajdzie zdarzenie E (H |= E);
2) Zaszło zdarzenie nie-E;

Wniosek:

Hipoteza H jest fałszywa

jest wnioskowaniem dedukcyjnym i pozwala – na podstawie przyjętych przesłanek – w sposób jednoznaczny rozstrzygnąć status H.

Taką możliwość wybrał K. Popper, uznając, że dedukcyjność (pewność) wniosku jest jedną z najważniejszych cech, jaką powinny mieć rozumowania przeprowadzane w nauce. Wnioskowanie o schemacie (S3) jest podstawowym wnioskowaniem definiującym jego koncepcję falsyfikacjonizmu (Popper 2002).

Stanowisko Poppera było krytykowane (najbardziej klasyczne pozycje to: Feyerabend 1971; Lakatos 1974; Meehl 1978; Quine 2000). Wskazywano m.in. na następujące jego wady:

– To, że zajdzie zdarzenie E, zazwyczaj nie wynika z samej hipotezy H, ale również z pewnych dodatkowych założeń na temat świata jako takiego, a więc we wniosku możemy stwierdzić jedynie, że fałszywa jest hipoteza H lub któreś z dodatkowo przyjętych założeń.
– Sam opis E nie jest niezależny od hipotezy, którą ma obalać. Nie istnieje coś takiego, jak neutralne teoretycznie sprawozdanie z danych empirycznych.
– Zgodnie z zaproponowaną procedurą, hipotezy naukowe nigdy nie mogą zostać uznane za prawdziwe – mogą mieć tylko status hipotez nieobalonych (koroborujących, „dzielnych”). Jedyne jednoznaczne wnioski mają

więc charakter wyłącznie negatywny: pozwalają stwierdzić, że hipoteza jest fałszywa. Z punktu widzenia rozwoju nauki wniosek taki jest mniej korzystny niż wniosek pozwalający uznać jakąś hipotezę za prawdziwą (bo np. nie pozwala na generowanie nowych tez). Na jego podstawie możemy jedynie twierdzić, że należy szukać hipotezy lepszej – ma więc własność nie tyle praktyczną, co meta-teoretyczną.

– Rozumowanie o schemacie (S3) ma charakter idealizacyjny. Z realnie stawianych hipotez (nawet rozszerzonych o pewne dodatkowe założenia na temat świata i prawa pomostowe) wynikają najczęściej wnioski przybliżone (np. to, że uzyskana wielkość prawdopodobnie będzie należała do określonego przedziału, że z dużym prawdopodobieństwem zajdzie zdarzenie określonego typu itp.), a nie przewidywania empiryczne, które mają postać precyzyjnych stwierdzeń (np. ilościowych: uzyskana wielkość będzie równa x, tor ruchu obiektu będzie identyczny z krzywą φ itp.). Aby taki zarzut uzasadnić, nie trzeba się nawet odwoływać do poglądu, że to z samej struktury rzeczywistości wynika, iż wszystkie prawa mają charakter probabilistyczny. Wystarczy pamiętać o tym, że błąd pomiaru i niejednoznaczność wyników jest nieodłącznym elementem wszelkich badań empirycznych. A więc pierwsza przesłanka w praktyce naukowej w zasadzie nigdy nie jest prawdziwa.

O ile trzy pierwsze zarzuty są związane z samą naturą schematu (S3) i nie widać, jak się przed nimi bronić, to wydaje się, że czwarty zarzut możemy uchylić, modyfikując schemat (S3) do następującej wersji probabilistycznej:

(S4) *Przesłanki:*
1) Jeśli hipoteza H jest prawdziwa, to z dużym prawdopodobieństwem zajdzie zdarzenie E;
2) Zaszło zdarzenie nie-E;

Wniosek:
Hipoteza H z dużym prawdopodobieństwem jest fałszywa;

co formalnie możemy zapisać jako:

(S5) *Przesłanki:*
1) P(E/H) jest bliskie 1;
2) Zaszło zdarzenie nie-E.

Wniosek:
$P_{\text{nie-E}}(\sim H)$ jest bliskie 0;[6]

---

[6] Przez $P_{\text{nie-E}}(H)$ oznaczamy prawdopodobieństwo hipotezy H *a posteriori* – czyli po uzyskaniu informacji, że zaszło nie-E.

albo jeszcze prościej[7] jako:

(S6) *Przesłanka:*
P(E/H) jest bliskie 1;
Wniosek:
P(~H/~E) jest bliskie 1.

Rozumowanie o powyższej strukturze wydaje się być naturalne: skoro bowiem następnik implikacji w pierwszej przesłance nie jest pewny, ale tylko wysoce prawdopodobny (zachodzi (b'), a nie (a')), to wniosek o fałszywości hipotezy też musi zostać osłabiony. W efekcie wprawdzie otrzymujemy wniosek o mniej radykalnym charakterze, ale za to schemat ma szansę być częściej stosowany. Czy jednak takie wnioskowanie można uznać za równie niezawodne co wnioskowanie o schemacie (S3), a jeśli nie, to czy przynajmniej jest to wnioskowanie racjonalne? Innymi słowy, czy tak zmodyfikowany schemat falsyfikacjonizmu będzie miał cechę, którą za podstawową dla wnioskowań naukowych uznawał Popper – będzie prowadził do prawdziwych (czy bliskich prawdzie) wniosków?

3. Osłabione schematy dedukcyjne

Aby zrozumieć, jak działają schematy o strukturze podobnej do (S4), w których występuje pojęcie prawdopodobieństwa, rozważmy następujące wnioskowania:

I. *Przesłanki:* 1) Jeśli α, to β; 2) Zajdzie α;
Wniosek: Zajdzie β.

II. *Przesłanki:* 1) Jeśli α, to β; 2) Zajdzie ~β.
Wniosek: Zajdzie ~α.

III. *Przesłanki:* 1) Jeśli α, to β; 2) α zajdzie z dużym prawdopodobieństwem;
Wniosek: β zajdzie z dużym prawdopodobieństwem.

[7] Przyjmujemy tu dosyć oczywistą i niebudzącą kontrowersji zasadę warunkowania: Jeśli wyjściowo prawdopodobieństwo A miało wartość P(A), to po tym, jak zaszło zdarzenie B, prawdopodobieństwo A wynosi P(A/B). Uzasadnieniem dla tej zasady jest tzw. Diachroniczny Zakład Holenderski.

IV. *Przesłanki:* 1) Jeśli α, to z dużym prawdopodobieństwem β; 2) α zajdzie;
Wniosek: β zajdzie z dużym prawdopodobieństwem.

V. *Przesłanki:* 1) Jeśli α, to β; 2) $\sim\beta$ zajdzie z dużym prawdopodobieństwem;
Wniosek: $\sim\alpha$ zajdzie z dużym prawdopodobieństwem.

VI. *Przesłanki:* 1) Jeśli α, to z dużym prawdopodobieństwem β; 2) $\sim\beta$;
Wniosek: $\sim\alpha$ zajdzie z dużym prawdopodobieństwem.

VII. *Przesłanki:* 1) Jeśli α, to z małym prawdopodobieństwem β; 2) β;
Wniosek: $\sim\alpha$ zajdzie z dużym prawdopodobieństwem.

Wnioskowania (I) i (II) są niezawodne – prawdziwość przesłanek gwarantuje prawdziwość wniosku. Ich schematy są po prostu prawami logiki klasycznej. We wnioskowaniach (III)–(VII) mamy odwołanie do pojęcia prawdopodobieństwa, więc należy oceniać je ostrożniej i ewentualnie wykorzystywać naszą wiedzę na temat rachunku prawdopodobieństwa.

W schemacie (III) z pierwszej przesłanki wiemy, że prawdopodobieństwo tego, że zajdzie β, o ile zaszło α, jest równe 1:

$$P(\beta/\alpha) = 1.$$

Zgodnie natomiast z drugą przesłanką, prawdopodobieństwo α jest bliskie 1:

$$P(\alpha) \approx 1.$$

Na mocy definicji prawdopodobieństwa warunkowego mamy:

$$P(\beta/\alpha) = P(\beta\wedge\alpha) \ / \ P(\alpha),$$

co po podstawieniu danych z przesłanek daje nam:

$$P(\beta) \approx 1.$$

Schemat (IV) zawiera pojęcie prawdopodobieństwa w sposób nieistotny i jest *de facto* podstawieniem schematu (I), gdzie zamiast „β” mamy: „z dużym prawdopodobieństwem β”.

Schemat (V) ma strukturę taką samą jak schemat (III) – wystarczy zauważyć, że pierwsza przesłanka jest równoważna zdaniu: Jeśli $\sim\beta$, to $\sim\alpha$.

Niestety analogicznego usprawiedliwienia nie można znaleźć dla schematów (VI) i (VII) (mają one dokładnie taką samą strukturę)[8]. Nie stanowią dla nich uzasadnienia ani prawa czystej logiki, ani rachunek prawdopodobieństwa. A są to właśnie interesujące nas schematy falsyfikacjonizmu w wersji probabilistycznej (są to po prostu, różniące się tylko notacją, wersje schematów (S4)–(S6)).

W dalszej części artykułu pokażemy, że:

– schematy (VI) i (VII) często występują w praktyce naukowej.
– schematy (VI) i (VII) są nie tylko zawodne, ale również nie spełniają słabszych warunków nakładanych na wnioskowania niededukcyjne – nie są nawet racjonalne.
– stosujący je naukowcy powszechnie błędnie interpretują uzyskane przy ich pomocy wyniki.

4. Test hipotezy zerowej i jego krytyka

W latach trzydziestych XX wieku R. Fisher i niezależnie E. Parson i J. Spława-Neyman analizowali sposób, w jaki można testować hipotezy statystyczne. Ich prace pozwoliły sformułować procedurę nazywaną obecnie testem hipotezy zerowej (NHST – *Null Hypothesis Significance Testing*). Co ciekawe – chociaż schemat tej procedury jest praktycznie identyczny ze schematem (S4), badania statystyków i Poppera rozwijały się niezależnie od siebie. Można powiedzieć, że o ile Popper zajmował się problemem testowania hipotez na poziomie meta-naukowym i starał się zdyskredytować opracowywaną przez Carnapa metodę indukcyjną, to wspomniani wyżej autorzy poszukiwali po prostu użytecznego narzędzia do opisu wyników eksperymentów. Różne motywacje doprowadziły do tego samego efektu. Obecnie zwolenników NHST opisuje się jako współczesnych popperystów (por. np. Lakatos 1978, s. 25; Gigerenzer i in. 1989).

Procedura określana w literaturze skrótem NHST ma oryginalnie następującą postać:

- Niech dana będzie pewna hipoteza prosta H_0, dotycząca określonej klasy zjawisk.
- Wykonujemy eksperyment związany z tą klasą zjawisk, który daje nam wynik E.

[8] W tym celu wystarczy zauważyć, że „z małym prawdopodobieństwem β” znaczy to samo, co „z dużym prawdopodobieństwem ~β”. Wtedy schemat (VI) różni się od (VII) tylko tym, że w pierwszym mówimy o zajściu ~β, a w drugim – o zajściu β.

- Liczymy, jakie jest prawdopodobieństwo uzyskania takiego wyniku przy założeniu, że H_0 jest prawdziwa.
- Jeśli to prawdopodobieństwo jest mniejsze niż pewna ustalona wartość p (zwykle przyjmuje się, że wynosi ona 0,05), uznajemy, że hipoteza H_0 nie przeszła testu.
- Jeśli dysponujemy hipotezą H konkurencyjną do H_0 (taką, że H ma postać nie-H_0), to należy ją przyjąć (jest ona bardzo prawdopodobna).

Zapisując to rozumowanie schematycznie mamy:

(S7) *Przesłanki*:
1) Uzyskaliśmy wynik E;
2) $P(E/H_0) < p = 0,05$;
Wniosek:
H_0 nie przeszła testu (jest bardzo prawdopodobne, że H_0 jest fałszywe, czy innymi słowy: jest bardzo prawdopodobne, że nie-H_0 jest prawdziwe).

Albo jeszcze inaczej – żeby podkreślić jego podobieństwo do schematu (S4):

(S8) *Przesłanki:*
1) Jeśli hipoteza H_0 jest prawdziwa, to z dużym prawdopodobieństwem pewien parametr t będzie miał wartość większą niż ustalone p;
2) Parametr t ma wartość mniejszą niż p;
Wniosek: Hipoteza H_0 jest z dużym prawdopodobieństwem fałszywa.

Zilustrujmy działanie tej procedury przykładem.

PRZYKŁAD 1
Mamy monetę dwuzłotową.
- Nasza hipoteza H_0 brzmi: jest to moneta rzetelna, a więc szansa na wyrzucenie orła i reszki jest taka sama i wynosi 1/2.
- Testujemy tę hipotezę, rzucając monetą 6 razy. Za każdym razem wypadł orzeł.
- Obliczamy, jaka jest szansa, że taki wynik uzyskamy przy założeniu, że moneta jest rzetelna, a więc ile wynosi P(6 razy orzeł/moneta rzetelna). Wynosi ono $(1/2)^6$ czyli 0,01.
- Ponieważ jest to wynik niższy niż zakładany próg 0,05, uznajemy, że hipoteza o rzetelności monety nie przeszła testu i należy ją odrzucić.
- Przyjmujemy zatem hipotezę konkurencyjną – moneta jest nierzetelna.

Takie rozumowanie wydaje się jak najbardziej racjonalne. Co więcej, nie widać innego prostego sposobu uzasadnienia lub obalenia hipotezy o rzetelności monety (jeżeli np. nie wiemy wcześniej, jaki jest rozkład wszystkich typów monet nierzetelnych). Można tylko rzucać monetą, obserwować wyniki i próbować wyciągnąć stąd jakieś wnioski.

Ten przykład można przerobić również tak, że będzie miał strukturę schematu (S4):

Przesłanki:

1) Jeśli teoria mówiąca, że moneta jest rzetelna, jest prawdziwa, to z dużym prawdopodobieństwem ilość orłów w 6 rzutach będzie należała do przedziału od 2 do 5;
2) Uzyskano 6 orłów w 6 rzutach;

Wniosek:

Teoria mówiąca, że moneta jest rzetelna, jest z dużym prawdopodobieństwem fałszywa.

Procedura NHST jest wykorzystywana przede wszystkim w takich naukach jak medycyna, socjologia, ekonomia czy psychologia. Ich wspólną cechą jest to, że w przypadku wielu głoszonych przez te nauki tez nie istnieje inny prosty sposób ich weryfikacji. To, czy dany lek faktycznie działa, albo to, czy na człowieka mają wpływ określone cechy środowiska, w którym człowiek się wychowuje, jest problemem bardzo złożonym. Zamiast szukać koniecznych związków przyczynowych między określonymi cechami obiektów[9], łatwiej jest przeprowadzić badanie wybranej próbki obiektów, które te cechy posiadają. Musimy jednak zdawać sobie sprawę, że taka procedura ma swoje ograniczenia.

5. Krytyka procedury testu hipotezy zerowej i falsyfikacjonizmu w wersji probabilistycznej

Oba schematy rozumowania: (S4) i charakterystyczny dla NHST (S8) mają taką samą strukturę, a więc widać, że probabilistyczna modyfikacja schematu falsyfikacjonizmu rzeczywiście występuje w praktyce naukowej (wszędzie tam, gdzie stosowany jest test hipotezy zerowej). Stąd wniosek, że po pierwsze, pomysł, aby stosować schemat (S4), jest jak najbardziej naturalny, po

[9] W przypadku monety występującej w przykładzie 1, zamiast nią rzucać, możemy przeprowadzić jej dokładne pomiary fizyczne i próbować ustalić, czy jest ona symetryczna. Rzetelność monety przekłada się bowiem na jej określone, mierzalne własności.

drugie, że ocena jego racjonalności jest z punktu widzenia metodologii nauki istotna, a po trzecie, że tę ocenę możemy wzorować na ocenie schematu NHST.

Procedurę NHST krytykuję się formułując pod jej adresem następujące zarzuty (por. np. Gill 1999; Nickerson 2000):

a) Wybór poziomu istotności p jest arbitralny.
   Zgodnie z procedurą, jeśli P(E/H) = 0,049, to hipotezę H należy odrzucić, a jeśli P(E/H) = 0,05 – to należy ją utrzymać (powód, aby ją odrzucić, nie jest wystarczający). Wydaje się to bardzo arbitralnym rozstrzygnięciem, przypisującym w efekcie istotnie inny status hipotezom niewiele się od siebie różniącym.
b) Wybór badanej próbki, która dostarcza nam wyniku E, jest arbitralny.
   Sam eksperyment, w efekcie którego uzyskujemy wynik E, często nie jest neutralny wobec celu, jakim ma być podważenie hipotezy H. W naszym przykładzie z kostką istotne znaczenie ma ilość wykonanych rzutów (można przypuszczać, że zwiększenie liczby rzutów działałoby na korzyść hipotezy H_0).
c) Schemat wnioskowania NHST nie spełnia kryteriów racjonalności.

Oczywiście z punktu widzenia oceny schematu (S4) najbardziej istotny jest zarzut (c). Falsyfikacjonizm miał różne wady (przy okazji widać podobieństwo zarzutu drugiego pod jego adresem i zarzutu (b)), ale jego przewaga nad indukcjonizmem miała polegać na wyższości logicznej samego schematu wnioskowania. Skoro schematu (S4) nie można nazwać racjonalnym, to nie widać, dlaczego falsyfikacjonizm (w wersji probabilistycznej) miałby być lepszy niż indukcjonizm. Jak jednak uzasadnia się (c)?

Najbardziej podstawowym warunkiem, jaki stawia się wnioskowaniom zawodnym (a wiemy już, że schemat (S4), a więc i (S8), jest właśnie takim wnioskowaniem), jest to, że przyjęte w nim przesłanki muszą uprawdopodobniać wniosek. Formalnie oznacza to tyle, że prawdopodobieństwo wniosku *a priori* jest ostro mniejsze niż prawdopodobieństwo wniosku, o ile założymy, że prawdziwe są przesłanki[10]. Można zapisać to następująco:

$$P(\text{wniosek}/\text{przesłanki}) > P(\text{wniosek})$$

W literaturze poświęconej krytyce testu hipotezy zerowej spotyka się przykłady wnioskowań przebiegających według schematu (S4), które ewidentnie nie spełniają tego warunku.

[10] Taki warunek spełnia np. wnioskowanie enumeracyjne (S2).

PRZYKŁAD 2

Przesłanki:

1) Jeśli osoba O jest obywatelem USA, to jest bardzo prawdopodobne, że osoba O nie jest obecnym Prezydentem USA;
2) Osoba O jest obecnym Prezydentem USA;

Wniosek:

Jest bardzo prawdopodobne, że O nie jest obywatelem USA.

Pierwsza przesłanka jest prawdziwa, ponieważ wśród 314 milionów obywateli USA tylko jeden jest obecnym prezydentem. Jeśli założymy, że prawdziwa jest druga przesłanka, to wniosek będzie ewidentnie fałszywy. Nie sama fałszywość jest jednak tutaj problemem[11], ale to, że prawdopodobieństwo *a priori* tego wniosku (że pewna osoba O nie jest obywatelem USA) jest samo z siebie dość wysokie. Założenie, że przesłanki są prawdziwe, redukuje je do zera. Mamy więc:

P(wniosek) > P(wniosek/przesłanki),

czyli niespełniony jest warunek racjonalności wnioskowania.

Ten przykład pokazuje, że sam fakt zajścia jakiegoś mało prawdopodobnego zdarzenia E nie wpływa na to, że hipoteza H jest fałszywa. Zdarzenie mało prawdopodobne również co jakiś czas zachodzi. Schemat nakazuje nam jednak uznanie na tej podstawie, że bardzo prawdopodobna jest fałszywość hipotezy H, występującej w poprzedniku implikacji pierwszej przesłanki. Można się bronić przed używaniem przykładu o podobnej strukturze co przykład 2, twierdząc, że pierwsza przesłanka wprawdzie jest literalnie prawdziwa, ale jej przyjęcie sugeruje, że istnieje pozytywna korelacja między byciem obywatelem USA a niebyciem prezydentem USA. Faktycznie jednak tak nie jest:

P(osoba O nie jest obecnym Prezydentem USA / osoba O jest obywatelem USA) < P(osoba O nie jest obecnym Prezydentem USA).

Wśród wszystkich możliwych osób jest bowiem więcej nie-prezydentów USA niż wśród samych Amerykanów.

Wprawdzie nałożenie odpowiedniego warunku na pierwszą przesłankę (żądanie pozytywnej korelacji między faktem opisanym w następniku a hipotezą w poprzedniku) eliminuje przykłady podobne do powyższego, ale nie

[11] Cechą charakterystyczną wnioskowań zawodnych jest to, że mogą one dawać fałszywy wniosek nawet wtedy, gdy przesłanki są prawdziwe.

eliminuje wad samego wnioskowania. Aby je zilustrować, rozważmy kolejny przykład:

PRZYKŁAD 3
Przesłanki:
1) Jeśli Jan jest zdrowy (nie ma HIV), to jest bardzo prawdopodobne, że przeprowadzony przez niego test na HIV da wynik negatywny;
2) Test na HIV dał w przypadku Jana wynik pozytywny;
Wniosek:
Jest bardzo prawdopodobne, że Jan nie jest zdrowy (ma HIV).

Co można zapisać skrótowo jako:

Przesłanki:
1) P(test negatywny/Jan jest zdrowy) jest bliskie 1;
2) Test dał wynik pozytywny;
Wniosek:
P(Jan jest chory/test pozytywny) jest bliskie 1.

Chcąc ocenić to wnioskowanie, sprawdzimy najpierw jego działanie na (zbliżonych do prawdziwych) danych dotyczących własności testu na HIV:

P(test pozytywny/osoba chora) = 1 (innymi słowy, nie ma wyników fałszywie ujemnych i test wychodzi pozytywnie u wszystkich osób chorych)[12].
P(test pozytywny/osoba zdrowa) = 10/99.990 (innymi słowy, z takim prawdopodobieństwem test daje wynik fałszywie dodatni, wskazując na chorobę u osoby faktycznie zdrowej).

Jest to równoważne temu, że:

P(test negatywny/osoba zdrowa) = 99.980/99.990.

Załóżmy teraz, że Jan nie należy do grupy ryzyka, i jego szansa *a priori*, że jest chory, wynosi 1/10.000, co zapiszemy jako

P(Jan jest chory) = 1/10.000

[12] Jest to idealizacja, służąca wygodzie obliczeń. Faktycznie żaden realny test nie ma tak wysokiej czułości.

lub równoważnie:

P(Jan jest zdrowy) = 9.999/10.000.

Obliczamy interesującą nas wielkość P(Jan jest chory/test pozytywny), stosując twierdzenie Bayesa, które w tym przypadku ma postać:

$$\frac{\text{P(test pozytywny/Jan jest chory) x P(Jan jest chory)}}{\text{P(test pozytywny/Jan jest chory) x P(Jan jest chory) + P(test pozytywny/Jan jest zdrowy) x P(Jan jest zdrowy)}}$$

Po podstawieniu danych do tego wzoru otrzymujemy:

P(Jan jest chory/test pozytywny) = ½.

Podobnie jak poprzednio, problem nie polega na tym, że wniosek z przykładu 3 jest po prostu fałszywy, ale że informacje zawarte w przesłankach nie są wystarczające, aby obliczyć, jaka jest szansa, że Jan jest chory. Nie wdając się w szczegóły techniczne, wystarczy zauważyć, że jest to wnioskowanie o strukturze zbliżonej do następującej[13]:

Przesłanka: Liczba a jest bliska 0

Wniosek: Liczba $\frac{b}{b+a}$ jest bliska 1.

To, czy wniosek będzie prawdziwy, zależy nie tyle od informacji o wielkości liczby a, zawartej w przesłance, ale od tego, jaki jest stosunek między liczbą a i liczbą b (jeśli np. $a = b$, to ułamek występujący we wniosku będzie równy ½).

Skoro więc wnioskowania o strukturze (S4) są nieracjonalne, to skąd ich duża popularność? Problem ten analizowany jest w wielu pracach – m.in. w: Gigerenzer, Krauss, Vitouch 2004; Westover, Westover, Bianchi 2011. Wydaje się, że odpowiedzialne jest za to pewnego typu złudzenie, któremu ulegamy przechodząc od wnioskowania niezawodnego (II) do wnioskowania typu (VI) czy (VII). Za w pełni rozsądne uznajemy osłabienie wniosku w sposób analogiczny do tego, w jaki osłabiliśmy przesłankę. Okazuje się jednak, że nie do

[13] Złośliwi krytycy testu hipotezy zerowej porównują wnioskowania o schematach w rodzaju (S4) do wnioskowania o tym, ile lat ma kierowca autobusu, jeśli z przesłanek wiemy, że na pierwszym przystanku wsiadło 9 osób, a na drugim przystanku 7 (por. np. Yudkowsky 2010). Przykład w tekście ma z oryginalnym wnioskowaniem dużo więcej wspólnego.

końca rozumiemy, jaka informacja jest faktycznie w tej przesłance zawarta. Znając jedynie prawdopodobieństwo zajścia zdarzenia E w sytuacji, w której prawdziwa byłaby hipoteza H, mamy pokusę, aby wnioskować coś na temat samej hipotezy.

Wskazują na to badania przeprowadzone i opisane w pracy: Gigerenzer, Krauss, Vitouch 2004. Autorzy prosili swoich respondentów, aby na podstawie dwóch informacji:

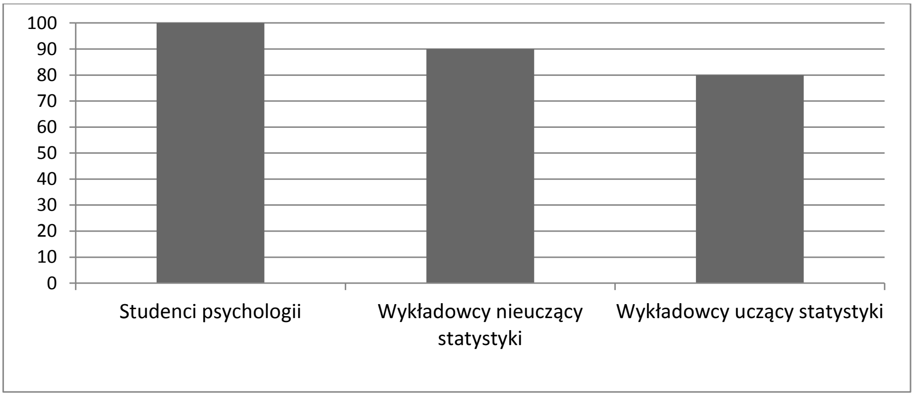
- zaszło zdarzenie nie-E;
- P(nie-E/H) = 0,01;

rozstrzygnęli, które z poniższych zdań są prawdziwe:

(1) Hipoteza H jest fałszywa.
(2) Możemy rozstrzygnąć, z jakim prawdopodobieństwem hipoteza H jest fałszywa.
(3) Hipoteza konkurencyjna do H jest prawdziwa.
(4) Możemy rozstrzygnąć, z jakim prawdopodobieństwem hipoteza konkurencyjna do H jest prawdziwa.
(5) Możemy ustalić, z jakim prawdopodobieństwem popełnimy błąd, przyjmując, że H jest fałszywa.
(6) Wynik: nie-E, znaczący z punktu widzenia testowanej hipotezy, otrzymamy w 99% przypadków, jeśli będziemy eksperyment powtarzać.

Badania przeprowadzono na trzech grupach: 40 studentów psychologii, 39 wykładowców nieuczących statystyki i 30 wykładowców uczących statystyki, z trzech różnych uniwersytetów niemieckich. W wykresie pokazane jest, jaki procent badanych uznał za prawdziwe przynajmniej jedno z powyższych 6 zdań[14].

[14] Zdanie (1) za prawdziwe uznało 34% badanych w pierwszej grupie, 15% w drugiej grupie i 10% w trzeciej grupie. Zdanie (2) za prawdziwe uznało 32% badanych w pierwszej grupie, 26% w drugiej grupie i 17% w trzeciej grupie. Zdanie (3) za prawdziwe uznało 20% badanych w pierwszej grupie, 13% w drugiej grupie i 10% w trzeciej grupie. Zdanie (4) za prawdziwe uznało 59% badanych w pierwszej grupie, 33% w drugiej grupie i 33% w trzeciej grupie. Zdanie (5) za prawdziwe uznało 68% badanych w pierwszej grupie, 67% w drugiej grupie i 73% w trzeciej grupie. Zdanie (6) za prawdziwe uznało 41% badanych w pierwszej grupie, 49% w drugiej grupie i 37% w trzeciej grupie.



Jak już wiemy, wszystkie zdania (1)–(6) są fałszywe!

6. Wnioski

Modyfikacja charakterystycznego dla falsyfikacjonizmu schematu wnioskowania wydaje się kuszącą alternatywą. Dzięki temu można uniknąć zarzutu jego ograniczonej stosowalności w praktyce naukowej. Takiej pokusie ulega wielu naukowców. Okazuje się jednak, że jest ona szkodliwa – uzyskane w ten sposób wyniki (odrzucenie lub akceptacja hipotez naukowych) nie mają wystarczającego uzasadnienia. Jeśli więc z jakichś powodów nie podoba nam się rozumowanie indukcyjne, powinniśmy pozostać wierni Popperowi wyłącznie w jego „dedukcyjnej" wersji (a więc posługiwać się tylko schematem (S3)), mając nadzieję, że trafi nam się kiedyś hipoteza naukowa, z której logicznie wynika, że zajdzie określone zjawisko E.

Bibliografia

Carnap R. (1950), *Logical Foundations of Probability*, London: Routledge & Kegan Paul Ltd.

Cohen J. (1994), *The Earth is Round (p < .05)*, „American Psychologist" 12, s. 997–1003.

Feyerabend P.K. (1971), *Problems of empiricism. Part II*, w: R.G. Colodny (red.), *The nature and function of scientific theories*, Pittsburgh: University of Pittsburgh Press.

Gigerenzer G., Krauss S., Vitouch O. (2004), *The Null Ritual. What You Always Wanted to Know About Significance Testing but Were Afraid to Ask*, w:

D. Kaplan (red.), *The Sage handbook of quantitative methodology for the social sciences*, Thousand Oaks, CA: Sage, s. 391–408.

Gigerenzer G., Swijtink Z., Porter T., Daston L., Beatty J., Kruger L. (1989), *The empire of chance: How probability changed science and everyday life*, Cambridge, UK: Cambridge University Press.

Gill J. (1999), *The Insignificance of Null Hypothesis Significance Testing*, „Political Research Quarterly" 3 (52), s. 647–674.

Lakatos I. (1974) *Popper on demarcation and induction*, w: P.A. Schilpp (red.), *The philosophy of Karl Popper*, Vol. 1, LaSalle, Ill.: Open Court.

Lakatos I. (1978), *The Methodology of Scientific Research Programmes*, Philosophical Volume Papers 1, Cambridge: Cambridge University Press.

Meehl P.E. (1978), *Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology*, „Journal of Consulting and Clinical Psychology" 46, s. 806–834.

Nickerson R.S. (2000), *Null Hypothesis Significance Testing: A Review of an Old and Continuing Controversy*, „Psychological Methods" 5 (2), s. 241–301, DOI: I0.1037//1082-989X.S.2.241.

Popper K.R. (2002), *Logika odkrycia naukowego*, przeł. U. Niklas, Warszawa: Fundacja Aletheia.

Quine W.V.O. (2000), *Dwa dogmaty empiryzmu*, w: tenże, *Z punktu widzenia logiki*, przeł. B. Stanosz, Warszawa: Fundacja Aletheia, s. 49–75.

Yudkowsky E (2010), *An Intuitive Explanation of Bayes' Theorem*, http://yudkowsky.net/rational/bayes.

Westover M.B., Westover K.D., Bianchi M.T. (2011), *Significance testing as perverse probabilistic reasoning*, „BMC Medicine", Feb 28, s. 9–20, DOI:10.1186/1741-7015-9-20.

Streszczenie

Podstawowy dla falsyfikacjonizmu Poppera jest następujący schemat wnioskowania: *Przesłanki:* 1) Z tego, że hipoteza H jest prawdziwa, wynika, że zajdzie zdarzenie E; 2) Zaszło zdarzenie nie-E; *Wniosek:* Hipoteza H jest fałszywa. W pracy analizowana jest próba jego modyfikacji do schematu, w którym przyjmujemy słabsze założenie o związku między hipotezą H a zajściem zdarzenia E i w związku z tym – słabszy wniosek: *Przesłanki:* 1) Jeśli hipoteza H jest prawdziwa, to z dużym prawdopodobieństwem zajdzie zdarzenie E; 2) Zaszło zdarzenie nie-E; *Wniosek:* Hipoteza H z dużym prawdopodobieństwem jest fałszywa. Okazuje się, że probabilistyczna postać falsyfikacjonizmu występuje w praktyce naukowej, ale niestety wnioskowań przebiegających według tego zmodyfikowanego schematu nie można uznać za racjonalne.