JAKIE WNIOSKI UZNAJEMY ZA RACJONALNE? WPŁYW PRAWDOPODOBIEŃSTWA APRIORYCZNEGO NA PRAWDOPODOBIEŃSTWO APOSTERIORYCZNE

Anna Wójtowicz*
Uniwersytet Warszawski

Streszczenie: *Badania wydają się sugerować, że w wielu wypadkach ustalamy prawdopodobieństwo pewnych zdarzeń w oparciu o dostarczaną informację w sposób niezgodny z modelem normatywnym. Wpływa to bezpośrednio na wnioski, które jesteśmy gotowi uznać za racjonalne na podstawie danych przesłanek. W artykule wskazano czynniki mogące rzutować na przyjmowaną wielkość prawdopodobieństwa apriorycznego i – pośrednio – na obliczanie prawdopodobieństwa aposteriorycznego. Zbiorczo czynniki te będą określane jako parametr φ. Jego wielkość zależy od kontekstu, w jakim przebiega dane wnioskowanie, i od określonych cech wnioskującego podmiotu. Po uwzględnieniu jego wpływu okazuje się, że krytyczna ocena ludzkich wnioskowań nie zawsze jest uzasadniona.*

Słowa kluczowe: *prawdopodobieństwo bazowe, prawdopodobieństwo aprioryczne, czynnik bayesowski, wzór Bayesa, błąd pominięcia bazy, sofizmat prokuratora.*

WHAT CONCLUSIONS ARE RATIONAL? IMPACT OF PRIORI PROBABILITY ON A POSTERIORI PROBABILITY

Abstract: *Empirical data suggest, that in many situations people are not able to estimate the probability of events on the base of available information in a way consistent with the normative model. This fact influences the choice of conclusions which are considered to be rational. The article outlines the factors that may affect the assumed value of a priori probability and – indirectly – the value of a posteriori probability. All these factors will be collectively referred to as the parameter φ. Its value depends on the context in which the reasoning*

* Anna Wójtowicz, Instytut Filozofii Uniwersytetu Warszawskiego, Krakowskie Przedmieście 3, 00-927 Warszawa, e-mail: amwojtow@uw.edu.pl

 DOI: 10.7206/DEC.1733-0092.102

is made. In the article I show, that a critical assessment of our reasoning is not always justified.

Key words: *base rate, prior probability, Bayes factor, Bayes theorem, base-rate fallacy, prosecutor's fallacy.*

1. WSTĘP

Czy wnioskując na podstawie danych przesłanek, że zajdzie zdarzenie opisywane we wniosku, działamy racjonalnie? Odpowiedź na to pytanie nie zawsze jest prosta. Dysponując tylko niepełną informacją, gdy z posiadanych przez nas przesłanek E nie wynika logicznie dany wniosek H1, nie mamy gwarancji, że będzie on *prawdziwy*. Możemy jednak zapytać, *czy* i *na ile* przesłanki czynią go *prawdopodobnym*. W takim wypadku oceniamy dwa parametry wnioskowania:

- to, czy przesłanki uprawdopodabniają wniosek, zwiększając jego prawdopodobieństwo w stosunku do tego, ile wynosiło ono, *zanim* stwierdziliśmy, że zachodzi E. Innymi słowy pytamy o to, czy wnioskowanie z E do H1 uznamy za spełniające minimalne kryterium racjonalności dla wnioskowań zawodnych:

  (1) $P(H1) < P(H1|E)$;

- to, czy prawdopodobieństwo wniosku przy założeniu, że prawdziwe są przesłanki (czyli prawdopodobieństwo wniosku *a posteriori*) jest na tyle duże (przekracza pewien ustalony próg *t*, bliski 1), że racjonalnie jest – zgodnie z tzw. regułą Locke'a[1] – ten wniosek zaakceptować:

  (2) $P(H1|E) > t \approx 1$.

Również w literaturze przedmiotu można się spotkać z dwuznacznością pojęcia racjonalności wnioskowania; w tym kontekście przesłanki bądź *przyrostowo potwierdzają wniosek*, bądź *absolutnie potwierdzają wniosek* (odpowiednio: *incrementally confirm* i *absolute confirm* – por. np. Huber, 2008).

Wskazane wyżej cechy wnioskowania zawodnego są od siebie niezależne. Może bowiem zdarzyć się tak, że przesłanki wspierają wniosek (czyli spełnione jest (1)), ale jednocześnie prawdopodobieństwo, że wniosek ten jest prawdziwy – przy założeniu,

1 Zgodnie z regułą Locke'a jeśli jakiemuś przekonaniu przypisujemy prawdopodobieństwo wyższe niż pewien ustalony próg (bliski 1), to powinniśmy to przekonanie uznać za prawdziwe. Oczywiście w niektórych sytuacjach przyjęcie tej reguły może prowadzić do paradoksalnych konsekwencji – świadczy o tym tzw. paradoks loterii i paradoks przedmowy – por. np. Hawthorne i Bovens, 1999.

 DOI: 10.7206/DEC.1733-0092.102

że przesłanki są prawdziwe – jest na tyle niskie, że nie chcemy go przyjąć (nie jest spełniony warunek (2)). I odwrotnie: przesłanki mogą faktycznie nie wspierać wniosku, ale mimo to jego prawdopodobieństwo i tak może być bardzo wysokie[2].

Zastanawiając się, czy dany wniosek H1 można uznać za racjonalny na gruncie przesłanek E, często pytamy właściwie nie o to, czy P(H1|E) jest bliskie 1, ale o to, czy H1 jest uzasadniony *lepiej* niż pewien inny, konkurencyjny wniosek H2 (na ten temat por. np. Clarke, 2001). Pytamy więc, który z wniosków jest *bardziej racjonalny*. Interesuje nas jakość uzasadnienia H1 w porównaniu do jakości uzasadnienia H2, bo np. tylko te dwa wnioski zdecydowaliśmy się z jakichś powodów wziąć pod uwagę[3]. Istotne jest więc, czy

(3) $P(H1|E) > P(H2|E)$

dla danego H2, będącego wnioskiem konkurencyjnym względem H1.

Warto w tym miejscu odróżnić dwa przypadki: kiedy wnioski H1 i H2 są ze sobą sprzeczne (tzn. zarazem wykluczają się i dopełniają) i kiedy tylko się wykluczają[4]. W pierwszej sytuacji – jeśli H1 jest dużo lepiej uzasadnione niż H2 – będzie również spełniony warunek (2) i P(H1|E) będzie bliskie 1 – por. niżej oraz przypis 4), podczas gdy w drugim przypadku takiej zależności nie ma.

2. Wykorzystanie wzoru Bayesa

Obliczenie wielkości P(H1|E) sprowadza się do oszacowania wielkości: $P(H1 \wedge E)$, P(E), a to z kolei zwykle wymaga ustalenia, jakie są $P(E|{\sim}H1)$, P(E|H1) i P(H1). Zgodnie ze wzorem Bayesa otrzymujemy bowiem zależność:

$$(4)\ P(H1|E) = \frac{P(H1 \wedge E)}{P(E)} = \frac{P(H1) \times P(E|H1)}{P(H1) \times P(E|H1) + P({\sim}H1) \times P(E|{\sim}H1)} = \frac{P(H1) \times P(E|H1)}{P(E)}.$$

[2] Wyobraźmy sobie, że w pewnej rodzinie wszyscy – z wyjątkiem najmłodszego dziecka – zachorowali na bardzo zaraźliwą grypę. Po pewnym czasie najmłodsze dziecko zaczyna się skarżyć na ból głowy, ale nie ma podwyższonej temperatury. Brak gorączki zmniejsza prawdopodobieństwo wniosku, że zachorowało na tę samą chorobę, co pozostali członkowie rodziny (gdyż gorączka jest objawem dość charakterystycznym), ale mimo to prawdopodobieństwo, że dziecko zaraziło się od pozostałych domowników, jest na tyle wysokie, że należy uznać, że choruje na grypę.

[3] To, że oceniamy tylko relatywną racjonalność wniosków H1 i H2, może mieć różne powody. Po pierwsze, pozostałe wnioski (tzn. takie, które stanowią możliwe logiczne dopełnienie alternatywy $(H1 \vee H2)$) są na tyle mało prawdopodobne, że nie warto się nimi zajmować. Po drugie, może nas po prostu interesować, który z tej pary wniosków jest lepszy, a nie to, czy jest on najlepszy w ogóle. Założenie, że wnioski H1 i H2 są ze sobą sprzeczne (tzn. $H1 = {\sim}H2$), będzie istotne tylko wtedy, gdy będziemy chcieli skorzystać z zależności $P(H1) = 1 - P(H2)$. Oczywiście brak sprzeczności między H1 i H2 powoduje, że taka równość nie zachodzi.

[4] Dwa zdania wykluczają się, gdy nie mogą być jednocześnie prawdziwe (mogą być też jednocześnie fałszywe). Przykładem zdań wykluczających się, ale niebędących sprzecznymi, są zdania (p i q) oraz (p i nie-q).

 DOI: 10.7206/DEC.1733-0092.102

Ponieważ interesuje nas, czy na gruncie przesłanek E racjonalniejszy jest wniosek H1 niż wykluczający się z nim wniosek H2, dokonamy odpowiedniego przekształcenia tego wzoru, aby ułatwić porównanie:

$$(5)\ \frac{P(H1|E)}{P(H2|E)} = \frac{P(H1) \times P(E|H1)}{P(E)} \times \frac{P(E)}{P(H2) \times P(E|H2)}.$$

Stąd uzyskujemy:

$$(6)\ \frac{P(H1|E)}{P(H2|E)} = \frac{P(H1)}{P(H2)} \times \frac{P(E|H1)}{P(E|H2)}.$$

W szczególnej sytuacji, gdy wnioski H1 i H2 są ze sobą sprzeczne, tzn. H1 = ~H2, można wzór ten zapisać w postaci:

$$(7)\ \frac{P(H1|E)}{P(\sim H1|E)} = \frac{P(H1)}{P(\sim H1)} \times \frac{P(E|H1)}{P(E\sim|H1)}.$$

Korzystając dodatkowo z zależności P(H2) = 1 – P(H1), otrzymujemy:

$$(8)\ \frac{P(H1|E)}{1-P(H1|E)} = \frac{P(H1)}{1-P(H1)} \times \frac{P(E|H1)}{P(E\sim|H1)}.$$

Widać stąd, że jeśli H1 i H2 są wnioskami sprzecznymi i $\frac{P(H1|E)}{P(H2|E)}$ jest np. równe 19, to P(H1|E) wynosi 0,95[5] – spełniony jest więc warunek (2) dla wniosku H1[6].

Po prawej stronie równości we wzorze (6) mamy do czynienia z iloczynem dwóch ułamków, które w wielu wypadkach można opisywać niezależnie.

Wielkość $\frac{P(H1)}{P(H2)}$ określa, jaki jest wyjściowy stosunek prawdopodobieństw obu wniosków. W dalszej części artykułu będę oznaczać go przez A(H1, H2). Zwykle mianownik i licznik w tym ułamku (czyli P(H1) i P(H2)) nazywa się *prawdopodobieństwem a priori* danego wniosku (odpowiednio H1 i H2), choć można też określać je jako *prawdopodobieństwo bazowe* (*base rate*; do problemu tego rozróżnienia wrócimy później). Jeśli np. A(H1, H2) wynosi 1/100, to wyjściowo, czyli zanim uzyskaliśmy

---

5 Mamy bowiem $\frac{P(H1|E)}{P(H2|E)} = \frac{P(H1|E)}{1-P(H1|E)} > 19$, a więc P(H1|E) > 19/20, co daje P(H1|E) > 0,95.

6 Oddzielną kwestią jest to, czy do zmierzenia, czy wniosek H1 jest bardziej racjonalny niż wniosek H2 (wykluczający się z H1, ale niebędący po prostu jego negacją), najlepiej jest używać ułamka $\frac{P(H1|E)}{P(H2|E)}$. Można próbować bronić przekonania, że powinno się raczej pytać nie o to, *ile razy* wniosek H1 jest bardziej prawdopodobny od H2 w świetle przesłanek E, ale o to, *o ile* jest bardziej prawdopodobny, a więc mierzyć różnicę P(H1|E) – P(H2|E). Argumentem na rzecz tej tezy jest to, że jeśli różnica ta faktycznie jest bardzo mała, to – biorąc pod uwagę, jak bardzo nieprecyzyjne mogą być niektóre dane (por. dalej) – preferowanie H1 nad H2 (w przypadku wniosków wykluczających się, ale niesprzecznych) jest wątpliwe, mimo że $\frac{P(H1|E)}{P(H2|E)} > 20$. Jeśli np. P(H1|E) = 0,0001, P(H2|E) = 0,000004, to wprawdzie $\frac{P(H1|E)}{P(H2|E)} > 20$, ale P(H1|E) – P(H2|E) = 0,000096. Dokładna analiza tego zagadnienia wykracza jednak poza ramy niniejszego tekstu. Więcej na ten temat por. np. Kaye i Koehler, 2003, oraz Wegwarth i Gigerenzer, 2013.

 DOI: 10.7206/DEC.1733-0092.102

dostęp do przesłanek (czy innych danych), uznajemy prawdziwość wniosku H1 za 100 razy mniej prawdopodobną niż prawdziwość wniosku H2.

Wielkość $\frac{P(H1)}{P(H2)}$ jest nazywana czynnikiem bayesowskim (*Bayes factor*) lub ilorazem wiarygodności (*likelihood ratio*)[7] i będę ją oznaczać jako BF(H1, H2). Wielkość ta pokazuje, który z wniosków lepiej wyjaśnia posiadane przez nas dane E. Jeśli np. czynnik ten wynosi 100, oznacza to, że fakt, iż przesłanki E są prawdziwe, jest 100 razy bardziej prawdopodobny, jeśli założymy, że prawdziwy jest wniosek H1, niż jeśli założymy, że prawdziwy jest wniosek H2. Mówiąc trochę nieściśle, jest to z jednej strony miara tego, jak należy oceniać wnioskowanie pod kątem jego abdukcyjności[8], a z drugiej – jeśli E miałoby być jakimś standardowym argumentem na rzecz wniosku H1 – jest to miara skuteczności takiego argumentu. Jeśli np. E opisuje wynik testu, którego celem było ustalenie, czy pacjent choruje na pewną chorobę H1, to BF(H1, ~H1) opiera się na stosunku między liczbą wyników prawdziwie dodatnich tego testu (jego czułością) a liczbą wyników fałszywie dodatnich (nieswoistością testu). Wielkość ta jest w wielu wypadkach przypisana do testu w sposób w tym sensie trwały, że wynik ustalony na pewnej (uznanej za reprezentatywną) grupie jest traktowany jako wewnętrzna cecha testu, zachowywana przy jego zastosowaniu do każdego przedstawiciela populacji.

Ze wzoru (6) wynika w szczególności, że gdy A(H1, H2) = 1, tzn. kiedy wyjściowo traktujemy oba wnioski za równie prawdopodobne, o tym, który z nich uznamy za bardziej racjonalny, decyduje wyłącznie czynnik BF(H1, H2). I odwrotnie: jeśli BF(H1, H2) = 1, to o tym, który wniosek powinniśmy wybrać, decyduje czynnik A(H1, H2), czyli nasze pierwotne przekonanie na temat ich prawdopodobieństw. Wzór (6) pokazuje zatem, że modyfikacja naszych przekonań co do wniosków H1 i H2 jest zawsze pewnym kompromisem między tym, jak bardzo jesteśmy o nich wstępnie przekonani, a tym, jak istotnie nowych informacji dostarczają nam przesłanki E.

Jeśli wiemy, że P(H1) = 1 – P(H2) (czyli H1 i H2 są sprzeczne), i chcemy uznać wniosek za racjonalny, o ile jego prawdopodobieństwo przekroczy jakiś ustalony próg (np. 0,95), to możemy tak zrobić[9], gdy

(9) $A(H1, H2) \times BF(H1, H2) \geq 20$

lub gdy

(10) $A(H1, H2) \times BF(H1, H2) \leq 1/20$.

7 W sytuacji, gdy oba wnioski są proste (tzn. stwierdzają zachodzenie pewnego faktu czy zależności), pojęcia te można utożsamić. Ogólnie jednak należy je rozróżnić – por. np. dyskusja na ten temat: Nickerson, 2000, Kass i Raftery, 1995.

8 Wnioskowanie abdukcyjne to tzw. wnioskowanie „do najlepszego wyjaśnienia". Dany wniosek jest uzasadniony abdukcyjnie, gdy najlepiej ze wszystkich innych wniosków wyjaśnia prawdziwość posiadanych przesłanek.

9 Por. przypis 4.

 DOI: 10.7206/DEC.1733-0092.102

W pierwszym wypadku uznamy za racjonalne H1, a w drugim – H2.

Wydaje się, że każdy, kto posiada dane dotyczące A(H1, H2) i BF(H1, H2), powinien – korzystając z powyższych wzorów – obliczyć wartość proporcji $\frac{P(H1|E)}{P(H2|E)}$ i w związku z tym rozstrzygnąć, który wniosek (H1 czy H2) uznaje za lepszy. Dysponując odpowiednimi danymi liczbowymi, można również prowadzić rozważania *stricte* matematyczne, np. analizować, jak rozkład A(H1, H2), przy danym czynniku bayesowskim wpływa na ostateczny wynik $\frac{P(H1|E)}{P(H2|E)}$ (na ten temat por. np. Kass i Raftery, 1995, a także Domurat i Białek, 2016).

Analiza praktycznych, życiowych przykładów – czyli takich, gdzie wyrażeniom H1, H2 i E przypisuje się *konkretną treść* i umieszcza się je *w określonym kontekście* – pokazuje, że ludzie najczęściej w różny sposób szacują proporcję $\frac{P(H1|E)}{P(H2|E)}$. Znajduje to również potwierdzenie w serii eksperymentów, które przeprowadzono specjalnie po to, aby ocenić, czy zachowanie ludzi w sytuacjach z niepełną informacją można uznać za racjonalne. W klasycznych przykładach dotyczących taksówek (Taxi Cab Problem)[10], bibliotekarzy i rolników[11] czy w tzw. problemie diagnozy medycznej (por. dalej), wnioskujący podawali wielkość mającą odpowiadać proporcji $\frac{P(H1|E)}{P(H2|E)}$, istotnie odbiegającą jednak od tej, którą mogliby wyliczyć ze wzoru (6) na podstawie danych jawnie dostarczonych im przez eksperymentatorów lub znanych z innych źródeł[12]. Badania tego problemu, zapoczątkowane przez Tversky'ego i Kahnemana (Tversky i Kahneman, 1974), wydają się pokazywać, że ludzie popełniają w takich wypadkach systematyczne błędy. Zazwyczaj identyfikuje się je jako:

[10] W tym wypadku badanym podano następujące dane: w pewnym mieście jest 85% taksówek zielonych i 15% niebieskich. Jedna z taksówek spowodowała wypadek, a naoczny świadek, którego wiarygodność oceniono na 80%, twierdził, że w zdarzeniu brała udział niebieska taksówka. Na tej podstawie uczestnicy eksperymentu mieli stwierdzić, jakie jest prawdopodobieństwo wniosku, że wypadek rzeczywiście spowodowała niebieska taksówka. Większość badanych szacowała tę wartość na około 0,8, chociaż faktycznie wynosi ona 0,41. Zauważmy, że tym wypadku oba możliwe wnioski – *Taksówka była zielona* i *Taksówka była niebieska* – są sprzeczne.

[11] W tym wypadku badani dysponowali opisami osoby o cechach, które powszechnie przypisuje się bibliotekarzom (skrupulatność, zamiłowanie do książek itp.) i na tej podstawie mieli stwierdzić, jakie jest prawdopodobieństwo wniosku, że dana osoba rzeczywiście jest bibliotekarzem, a nie rolnikiem. Większość badanych szacowała to prawdopodobieństwo bardzo wysoko, chociaż faktycznie jest ono niskie: w populacji jest zdecydowanie mniej bibliotekarzy niż rolników. Zauważmy, że tym wypadku oba możliwe wnioski – *Opisana osoba jest bibliotekarzem* i *Opisana osoba jest rolnikiem* – nie są sprzeczne, a jedynie się wykluczają.

[12] Analogiczne zjawisko można zaobserwować w przypadku tzw. błędu koniunkcji. Jeśli badani mają po prostu ocenić, co jest bardziej prawdopodobne – zajście dwóch bliżej niesprecyzowanych zdarzeń p i q czy samego p, wybierają oczywiście tę drugą ewentualność. Jeśli jednak są pytani o to, co – w oparciu o informację, w której opisuje się Lindę jako osobę inteligentną, wygadaną, aktywną uczestniczkę demonstracji i filozofkę – jest bardziej prawdopodobne: *Linda jest pracownikiem banku i feministką* czy też *Linda jest pracownikiem banku*, wybierają pierwszą ewentualność.

 DOI: 10.7206/DEC.1733-0092.102

- błąd odwracania prawdopodobieństw (*invers fallacy*), ogólnie polegający na tym, że niesłusznie utożsamia się wartość P(H1|E) z wartością P(E|H1) czy wartością 1 – P(E|~H1) (co w kontekstach wnioskowań prawniczych nazywane bywa również *sofizmatem prokuratora*). Jako argument na rzecz np. H1 często uznaje się to, że wielkość P(E|~H1) jest bardzo mała, przyjmując fałszywą zależność: P(H|E) = P(~E|~H);
- błąd pominięcia wartości bazowej (*base-rate fallacy, neglecting the prior odds*), czyli niewzięcia pod uwagę wartości P(H1) i jej wpływu na obliczanie P(H1|E).

W naszej terminologii opis powyższych błędów można sprowadzić do tego, że ludzie z jakichś powodów uznają, że A(H1, H2) jest bliskie 1, a więc, że P(H1) = P(H2) (por. np. Fischhoff i De Bruin, 1999) i o tym, który z konkurencyjnych wniosków należy uznać za lepszy, decyduje wyłącznie wielkość BF(H1, H2). W takim wypadku rzeczywiście powstaje złudzenie, że można utożsamić wartość P(H1|E) z wartością P(E|H1). Rozstrzygnięcie, z którym dokładnie błędem mamy do czynienia, jest utrudnione przez to, że zwykle w badaniach ocenia się tylko uzyskaną (oszacowaną) wartość ilorazu $\frac{P(H1|E)}{P(H2|E)}$, a nie sposób, w jaki została obliczona (por. tabelę i komentarz do niej w Koehler, 1996).

Zauważmy jednak, że podawanie wyniku różniącego się od uzyskanego na podstawie wzoru (6) niekoniecznie świadczy o nieznajomości wzoru Bayesa czy zapominaniu o wartości A(H1, H2) (czyli utożsamianiu jej z 1). Osoby badane mogły dodatkowo manipulować danymi i zamiast wzoru (6), zastosować taki o postaci

$$(11)\ \frac{P(H1|E)}{P(H2|E)} = \varphi \frac{P(H1)}{P(H2)} \times \frac{P(E|H1)}{P(E|H2)},$$

albo innym i słowy

$$(12)\ \frac{P(H1|E)}{P(H2|E)} = \varphi\ A(H1, H2) \times BF(H1, H2),$$

gdzie $\varphi \neq 1$ jest pewnym dodatkowym parametrem.

Używanie wzoru (12) może prowadzić do innych oszacowań niż tych wynikających z wykorzystania z wzoru (6), a w skrajnych wypadkach do tego, że $\frac{P(H1|E)}{P(H2|E)}$ osiągnie taką samą wartość, co BF(H1, H2). Tak się stanie, jeśli φ będzie bliskie $\frac{P(H2)}{P(H1)}$, gdyż wtedy j A(H1, H2) = 1.

Przy ocenie, który z dwóch wniosków (H1 czy H2) jest lepszy (bardziej racjonalny), wnioskujący inaczej traktują BF(H1, H2) i A(H1, H2). Ma to niewątpliwie związek z kontekstem, w jakim cały problem jest umieszczany, i wynika (przynajmniej

 DOI: 10.7206/DEC.1733-0092.102

częściowo) z widocznej różnicy w statusie metodologicznym obu tych wielkości. W realnych, życiowych sytuacjach, zdecydowanie łatwiej podać przekonującą wartość i odpowiednią metodologię ustalania BF(H1, H2) niż A(H1, H2)[13]. Dodatkowo, jeśli ludzie mają do czynienia z sytuacją w pewien sposób odwołującą się do ich codziennej praktyki, w której muszą oszacować A(H1, H2), trudno jest wymagać, aby całkowicie abstrahowali od swojego doświadczenia życiowego. Zarówno w sytuacji, gdy dane dotyczące A(H1, H2) są dostarczane w sposób jawny (jak w przykładzie taksówek) lub niejawny (jak w przykładzie bibliotekarzy i rolników), czy też gdy ich znajomość jest w pewnym sensie obowiązkiem badanych (jak w problemie diagnozy medycznej, gdzie znajomość częstości danej choroby w populacji należy do wiedzy zdobytej na studiach medycznych), ich ocena z różnych powodów jest krytyczna (por. na ten temat np. Kaye i Koehler, 2003). W konsekwencji, przynajmniej w niektórych wypadkach, można bronić stanowiska,

że wartość $\frac{P(H1|E)}{P(H2|E)}$, którą podają wnioskujący w realnych okolicznościach czy badani w projektowanych eksperymentach – mimo że na pierwszy rzut oka nie jest zgodna z tym, co można obliczyć ze wzoru Bayesa i „oficjalnych" danych liczbowych – jest jednak w jakiś sposób uzasadniona. Wynika ze specyfiki rozważanego problemu i indywidualnej wiedzy wnioskującego.

Aby ułatwić doprecyzowanie takiego poglądu i analizę zamieszczonych dalej przykładów, wprowadźmy (za Kayem i Koehlerem, 2003, s. 652, a także Welshem i Navarro, 2012) rozróżnienie terminologiczne dotyczące prawdopodobieństwa wyjściowego P(H) danego wniosku (tzn. takiego prawdopodobieństwa, które jest dane, zanim jeszcze znane są przesłanki E). Przez *prawdopodobieństwo bazowe wniosku* będziemy rozumieć prawdopodobieństwo wyjściowe o charakterze interpersonalnym, dostarczone badanym w sposób jawny lub zawarte w oficjalnych danych. Przez *prawdopodobieństwo aprioryczne* danego wniosku będziemy natomiast rozumieć prawdopodobieństwo wyjściowe o charakterze subiektywnym i zależnym od dodatkowych, być może bardzo specyficznych uwarunkowań rozumowania. Oczywiście takie rozróżnienie nie jest w pełni ostre, ale pozwala podkreślić fakt, że jeśli celem jest ustalenie, który z dwóch wniosków (H1 czy H2) dany podmiot uzna za bardziej racjonalny na gruncie przesłanek E, i dopuszczamy, że mogą na to wpływać jego subiektywne poglądy, związane ze specyfiką sytuacji, w której się znalazł, to musimy pozwolić, aby również mógł subiektywnie modyfikować A(H1, H2). Jeśli A(H1, H2) będziemy postrzegać jako stosunek *prawdopodobieństw bazowych wniosków H1 i H2*,

[13] Wyobraźmy sobie np., że rzucamy monetą i 10 razy wypadł orzeł. Jeśli mamy teraz np. ustalić, co jest bardziej prawdopodobne: to, że moneta, którą rzucamy, jest rzetelna (wniosek H1), czy to, że jest przeważona, i prawdopodobieństwo wyrzucenia orła wynosi 0,6 (wniosek H2), to wiadomo, jak obliczyć BF(H1, H2), ale nie wiadomo, jaką wartość przypisać A(H1, H2). Aby to zrobić, musielibyśmy bowiem wiedzieć, ile wśród wszystkich dostępnych monet jest tych rzetelnych, a ile tych przeważonych dokładnie o 0,6 na stronę orła.

 DOI: 10.7206/DEC.1733-0092.102

to wielkość φA(H1, H2) możemy utożsamiać z tym, co uzyskalibyśmy, dzieląc przez siebie *prawdopodobieństwa aprioryczne wniosków H1 i H2*, przyjmowane przez dany podmiot wnioskujący (w dalszym ciągu będziemy mówili odpowiednio o *czynniku bazowym i czynniku apriorycznym*). Zilustrujmy to rozróżnienie prostym przykładem.

Przykład

Powróćmy do sytuacji opisanej w przypisie 13. Załóżmy, że według oficjalnych danych wśród monet dwuzłotowych bitych przez Mennicę Państwową 99% stanowią monety rzetelne, 0,1% to monety przeważone na stronę orła w stopniu 0,6, a 0,9% to monety przeważone w stronę reszki w stopniu 0,6.

Pewien znudzony Jan rzucał monetą dwuzłotową i 10 razy z rzędu wypadł mu orzeł. Zastanawia się teraz, który z możliwych wniosków jest na tej podstawie bardziej racjonalny: H1: jego moneta jest rzetelna czy H2: jego moneta jest przeważona w stronę orła.

Potrafi oczywiście obliczyć BF(H1, H2):

$$\frac{\mathrm{P(E|H1)}}{\mathrm{P(E|H2)}} = \frac{0{,}5^{10}}{0{,}6^{10}} = \frac{9765625}{60466176},$$

co daje w przybliżeniu 0,161. Oznacza to, że wyrzucenie pod rząd 10 orłów ponad sześciokrotnie lepiej jest wytłumaczone przez to, że moneta jest przeważona, niż przez to, że jest rzetelna.

Gdyby bohater naszego przykładu posługiwał się oficjalnymi danymi (czyli, w naszej terminologii, opierał się na *prawdopodobieństwach bazowych* H1 i H2), to A(H1, H2) – *czynnik bazowy* – wynosiłoby[14]:

$$\frac{\mathrm{P(H1)}}{\mathrm{P(H2)}} = \frac{0{,}99}{0{,}001} = 990.$$

Obliczenie

$$\mathrm{A(H1,\ H2)} \times \mathrm{BF(H1,H2)} \approx 160$$

pokazuje, że prawie 160 razy bardziej rozsądnie jest uznać wniosek H1 niż H2.

Jan opiera się jednak na dodatkowych założeniach. Po pierwsze, nie wierzy, aby dokładność instytucji państwowej, jaką jest mennica, mogła być tak duża. Według niego monety rzetelne stanowią co najwyżej 90% wszystkich wybijanych. Co więcej, jego znajomy ma znajomego, który twierdzi, że ze względów konstrukcyjnych w przypadku dwuzłotówek dużo częściej powstaje moneta przeważona na stronę orła niż reszki. W oparciu o to wszystko *aprioryczne* (i subiektywne) prawdopodobieństwa obu wniosków ocenia odpowiednio na: P(H1) = 0,9 i P(H2) = 0,08.

14 Skoro 99% monet bitych przez mennicę stanowią monety rzetelne, P(H1) = 0,99. Wiemy również, że 0,1% to monety przeważone na stronę orła w stopniu 0,6, a więc P(H2) = 0,001.

DOI: 10.7206/DEC.1733-0092.102

Stosunek prawdopodobieństw apriorycznych wniosków H1 i H2 – czyli *czynnik aprioryczny* charakterystyczny dla Jana – wynosi więc

$\frac{0,9}{0,08} = 11,25.$

Stąd można wyliczyć, że przyjmowany przez niego parametr φ to

$\frac{11,25}{990} = 0,0113.$

Oznacza to, że przy powyższych założeniach wniosek H1 o rzetelności monety jest tylko niecałe dwa razy bardziej racjonalny niż wniosek H2, że jest ona przeważona na stronę orła, gdyż

$\varphi A(H1, H2) \times BF(H1, H2) = 0,0113 \times 990 \times 0,161 = 1,8.$

Podsumowując, w tym przykładzie mamy na początku następujące dane:

$P(H1) = 0,99$

$P(H2) = 0,001.$

Są to prawdopodobieństwa bazowe obu wniosków.

Mamy również wynik obserwacji E, który możemy traktować jako przesłankę do dalszego rozumowania. Uzyskanie takiego wyniku jest tłumaczone (wyjaśniane) przez prawdziwość H2 dużo lepiej niż przez prawdziwość H1 (sześciokrotnie lepiej).

Gdyby Jan bezkrytycznie akceptował dane dotyczące wyjściowych prawdopodobieństw wniosków H1 i H2, czyli gdyby traktował podane prawdopodobieństwa bazowe jako swoje prawdopodobieństwa aprioryczne, to powinien uznać, że na podstawie E wniosek H1 jest 160 razy bardziej racjonalny niż H2. W takim wypadku parametr $\varphi = 1$.

Jeśli jednak Jan opiera się na dodatkowych danych i przyjmuje prawdopodobieństwa aprioryczne różniące się od bazowych i wynoszące odpowiednio:

$P(H1) = 0,90$ i

$P(H2) = 0,08,$

to może zasadnie uznawać, że na podstawie E wniosek H1 jest tylko dwa razy bardziej racjonalny niż H2. W takim wypadku parametr $\varphi = 0,0113$.

Odpowiedź na pytanie, czy Jan słusznie ocenia, ile razy bardziej prawdopodobny – na podstawie przesłanek E – jest wniosek H1 niż wniosek H2, nie powinna sprowadzać się do stwierdzenia, czy podawana przez niego wielkość (dwa razy bardziej) jest zgodna z „oficjalną" wielkością (160 razy bardziej). Należy ocenić, czy zasadnie przyjął, że parametr $\varphi = 0,0113$.

 DOI: 10.7206/DEC.1733-0092.102

Ogólnie rzecz ujmując, równość dwóch czynników (bazowego i apriorycznego), a więc to, czy $\varphi = 1$, zależy w szczególności od tego, czy wnioskujący:

- w taki sam sposób, jak zostało to zrobione na potrzeby wyliczenia prawdopodobieństwa bazowego, ustala grupę odniesienia, dla której określane jest prawdopodobieństwo;
- odwołuje się do jakichś „ukrytych" parametrów, które w danym wnioskowaniu należy według niego rozważyć.

Jeśli potraktujemy problem oceny wnioskowań zawodnych w taki sposób, to uznanie, że ktoś wnioskuje racjonalnie, nie sprowadzi się jedynie do mechanicznego zbadania, czy zastosował wzór (6) i otrzymał odpowiedni wynik. Należy również rozważyć, czy ewentualne wprowadzenie parametru $\varphi \neq 1$ do wzoru (12) jest zabiegiem rozsądnym.

Spójrzmy pod tym kątem na znane z literatury przykłady i spróbujmy ocenić racjonalność pojawiających się w nich wniosków z punktu widzenia rozróżnienia między czynnikiem bazowym i czynnikiem apriorycznym. Innymi słowy, zastanówmy się, jaki czynnik φ był w nich przyjmowany i dlaczego, a poza tym osądźmy, czy taka modyfikacja ogólnego wzoru (6) jest w tych wypadkach uprawniona.

3. Diagnoza medyczna

Przyjrzyjmy się mammografii jako metodzie wykrywania raka piersi. Statystyczne prawdopodobieństwo wystąpienia tej choroby w grupie kobiet w wieku 40–50 lat wynosi 0,008 (oznaczmy je przez P(R)). Stąd wartość czynnika bazowego to:

(13) $A(R, \text{nie}R) = \frac{P(R)}{P(\text{nie}R)} = \frac{0{,}008}{0{,}992} = 0{,}00801 \approx 0{,}008.$

Znamy również właściwości mammografii jako badania przesiewowego – tzn. mamy dane $P(M(+)|R)$ i $P(M(-)|\text{nie}R)$, gdzie przez $M(+)$ i $M(-)$ oznaczamy odpowiednio wykrycie podejrzanej zmiany w badaniu mammograficznym i niestwierdzenie istnienia podejrzanej zmiany w badaniu mammograficznym. Innymi słowy, mamy prawdopodobieństwo wyników prawdziwie dodatnich i prawdziwie ujemnych mammografii. Przyjmuje się, że wynoszą one odpowiednio: $P(M(+)|R) = 0{,}90$ i $P(M(-)|\text{nie}R) = 0{,}93$. Stąd wartość czynnika bayesowskiego to:

(14) $BF(R, \text{nie}R) = \frac{0{,}90}{1-0{,}93} = 12{,}85.$

 DOI: 10.7206/DEC.1733-0092.102

Podstawiając uzyskane wyniki do interesującego nas wzoru (6), otrzymujemy:

(15) $\frac{P(R|M(+))}{P(nieR|M(+))} = 0{,}008 \times 12{,}85 = 0{,}1028.$

Lekarze, którym w ramach przeprowadzanego eksperymentu (por. Westover, Westover i Bianchi, 2011) zaprezentowane te dane, oszacowali wielkość $\frac{P(R|M(+))}{P(nieR|M(+))}$ zdecydowanie wyżej – na ok. 9. Miało to świadczyć o ich nieracjonalności (por. np. Kuklin, 2006) objawiającej się zbytnią koncentracją na BF(R, nieR) i niebraniu pod uwagę czynnika A(R, nieR). Uznano, że jest to typowy błąd pominięcia wartości bazowej. Czy jednak taka ocena zachowań lekarzy nie jest zbyt pochopna?

Lekarz przyjmujący pacjentkę z wykrytą w badaniu mammograficznym podejrzaną zmianą może rozumować w następujący sposób:

Po pierwsze, biorąc pod uwagę, że bardziej kosztowny jest błąd uznania osoby chorej za zdrową niż uznanie zdrowej za chorą, z punktu widzenia oczekiwanej użyteczności wstępnej diagnozy lepiej zawyżyć prawdopodobieństwo choroby niż go zaniżyć. Stąd – już nawet z tego powodu – należy zastosować wzór

(16) $\frac{P(R|M(+))}{P(nieR|M(+))} = \varphi\ 0{,}008 \times 12{,}85,$

gdzie $\varphi > 1$. (Właściwie każdy pacjent powinien mieć nadzieję, że jego lekarz rozumuje właśnie w taki sposób).

Po drugie, regularnie badania mammograficzne robią kobiety dwóch typów – bardzo poważnie traktujące zalecenia dotyczące kontrolnych badań przesiewowych i takie, które z jakichś powodów niepokoją się swoim zdrowiem. Wydaje się, że pacjentek pierwszego typu jest mniej niż drugiego typu. W związku z tym istotną cechą pozwalającą faktycznie oszacować prawdopodobieństwo, że dana kobieta ma raka, jest nie tylko jej wiek, ale również fakt, że poddała się badaniom mammograficznym (o ile faktycznie należy do drugiego typu). Wykonanie badania świadczy o jakimś niepokoju, a to z kolei może sygnalizować chorobę. Oznaczmy przez B zdarzenie polegające na tym, że pacjentka należy do typu drugiego i z tego powodu wykonała badanie. W związku z tym, obliczając aprioryczne prawdopodobieństwo, że dana pacjentka jest chora, należy dodatkowo oszacować, ile wynosi P(R|B) i obliczyć $\frac{P(R|B)}{P(nieR|B)}$ zamiast obliczenia zwykłego $\frac{P(R)}{P(nieR)}$. Również z tego powodu w miejsce wzoru (15) powinno się zastosować wzór (16) i założyć, że $\varphi > 1$.

W konkluzji lekarz może stwierdzić, że istnieją dwa niezależne powody, aby uznać, że czynnik aprioryczny jest inny niż czynnik bazowy, i przyjąć np., że $\varphi = 100$. Ostatecznie więc

 DOI: 10.7206/DEC.1733-0092.102

(17) $\frac{P(R|M(+))}{P(nieR|M(+))} = 10{,}28.$

Nie wydaje się, że takie rozumowanie jest nieracjonalne i wymaga napiętnowania (tak jak zrobiła to np. Cieślińska, 2014[15])[16].

4. Uzasadnianie winy na podstawie testów identyfikujących

W sporach dotyczących interesującego nas problemu często spotykane są analizy przykładów mniej lub bardziej zaczerpniętych z praktyki kryminalistycznej (por. np. Gigerenzer, 1998 i 2002, oraz De Macedo, 2008). Dość standardowo rozważana jest sytuacja, w której ustala się winnego przy użyciu pewnego testu identyfikującego. Zwykle mówi się o znalezionym na miejscu przestępstwa materiale genetycznym lub odciskach palców (zakłada się, że należą one do sprawcy czynu, np. mordercy), które zostają następnie wprowadzone do policyjnej bazy danych. Pozwala to stwierdzić, że należą do osoby X. Jest to jedyny (na danym etapie śledztwa) dowód na rzecz wniosku o winie X. Powstaje pytanie, jakie jest prawdopodobieństwo, że to faktycznie X jest mordercą? Podejmując próbę oszacowania tego prawdopodobieństwa, wprowadźmy najpierw użyteczne oznaczenia:

– przez „X = M" będziemy oznaczać stwierdzenie, że osoba X jest rzeczywiście mordercą;
– przez „(X) ≈ (M)" będziemy oznaczać stwierdzenie, że profil DNA czy odciski palców osoby X *uznano* za identyczne z profilem czy odciskami mordercy. Innymi słowy, test ustalił identyczność osoby X i mordercy.

To, co chcemy obliczyć, to wartość $P(X = M \mid (X) \approx (M))$, a więc prawdopodobieństwo wniosku, że X jest mordercą, w oparciu o poszlakę w postaci pozytywnej identyfikacji za pomocą testu. Zastosowanie wzoru

[15] W powoływanym tekście po zaprezentowaniu takich samych danych, jak w naszym przykładzie, pada stwierdzenie: „Dlaczego więc lekarze straszą pacjentów? Bo nie rozumieją statystyki".

[16] W powyższym przykładzie warto zwrócić uwagę na pewnego rodzaju ekwiwokację związaną ze stwierdzeniem, że dana kobieta jest chora na raka. Badania pokazują, że 35% kobiet, u których dzięki mammografii wykryto tę chorobę we wczesnym stadium i które poddano leczeniu, nigdy nie dowiedziałoby się, że są chore, gdyż nowotwór nie dałby żadnych objawów (por. np. Jørgensen i Gøtzsche, 2009). Gdyby żadna kobieta nie robiła mammografii i P(R) szacowano wyłącznie na podstawie faktycznie stwierdzonych innych objawów nowotworu, byłoby ono znacząco niższe. (Takim „naddiagnozowaniem" tłumaczy się również współczesny wzrost zdiagnozowanych zachorowań na raka piersi). Mammografia jest więc badaniem nadwrażliwym nie tylko w sensie istnienia wyników fałszywie dodatnich. Z drugiej jednak strony, ze względu na samą istotę tego typu badań, pozwalających wykrywać potencjalnie bardzo groźne choroby we wczesnym, łatwiejszym do wyleczenia stadium, na taką nadwrażliwość chętnie się godzimy. Pośrednio wpływa to również na szacowanie prawdopodobieństwa *a priori* raka – właściwą (z punktu widzenia społeczeństwa) tendencją jest zawyżanie tego prawdopodobieństwa.

 DOI: 10.7206/DEC.1733-0092.102

$$(18)\ \frac{P(X = M|(X) \approx (M))}{P(X \neq M|(X) \approx (M))} = \frac{P(X = M)}{P(X \neq M)} \times \frac{P((X) \approx (M)|X = M)}{P((X) \approx (M)|X \neq M)}$$

wymaga od nas posiadania następujących danych:

– $P(X = M)$ – oszacowania prawdopodobieństwa wyjściowego, że dana osoba X jest mordercą;
– $P((X) \approx (M)|X = M)$ – stwierdzenia, jakie jest prawdopodobieństwo prawdziwie dodatniej identyfikacji;
– $P((X) \approx (M)|X \neq M)$ – stwierdzenia, jakie jest prawdopodobieństwo fałszywie dodatniej identyfikacji.

Dwie ostatnie wielkości charakteryzują łącznie skuteczność identyfikacji za pomocą danego testu (zakładamy, że są to ogólne własności tej metody i ich zastosowanie do osób M i X nie zmienia tych parametrów, ponieważ dostępne dane nie odbiegają od przyjętej normy). Ich iloraz daje nam wielkość $BF(X = M, X \neq M)$. Są one ustalone w oparciu o liczne analizy działania zarówno czysto automatycznych programów identyfikujących, jak i programów identyfikujących wspomaganych opiniami ekspertów. W niniejszym artykule skoncentrujemy się na teście identyfikującym za pomocą odcisków palców. Na podstawie Bradford, 2011 wiemy, że dla takiego testu $P((X) \approx (M)|X = M) = 0{,}925$, a $P((X) \approx (M)|X \neq M) = 0{,}01$. Stąd wartość czynnika bayesowskiego wynosi:

$$(19)\ BF(X = M, X \neq M) = 92{,}5.$$

Oznacza to, że odciski tej samej osoby są 92,5 razy częściej identyfikowane jako identyczne niż odciski pochodzące od dwóch różnych osób.

Podstawiając te dane do równania (18) uzyskujemy:

$$(20)\ \frac{P(X = M|(X) \approx (M))}{P(X \neq M|(X) \approx (M))} = \frac{P(X = M)}{P(X \neq M)} \times 92{,}5.$$

Powstaje jednak problem co do sposobu oszacowania tego, że osoba X jest mordercą, a więc co do wartości dla $A(X = M, X \neq M)$. Zauważmy, że nierozsądne byłoby zastosowanie w tym przypadku zasady nierozróżnialności (*indifference principle*), zgodnie z którą jeśli nie wiemy nic na temat tego, które z dopełniających się i wykluczających zdarzeń zajdzie, trzeba przypisać im równe prawdopodobieństwo. Oznaczałoby to bowiem, że X jest winny z prawdopodobieństwem 1/2, a czynnik $A(X = M, X \neq M) = 1$. Prawo nakazuje jednak zastosować w takiej sytuacji zasadę domniemania niewinności, która z kolei nie mówi po prostu, że trzeba uznać *a priori* niewinność osoby X (co przekładałoby się na to, że $A(X = M, X \neq M) = 0$, a w konsekwencji, że $P(X = M|(X) \approx (M)) = 0$), ale jedynie, że traktujemy osobę X jako nie

 DOI: 10.7206/DEC.1733-0092.102

bardziej winną niż dowolną inną osobę z *grupy jemu podobnych*. Problem przenosi się więc na ustalenie, co to za grupa i jakie prawdopodobieństwo bycia mordercą przypisujemy jej członkom[17]. Gigerenzer (2002), prowadząc rozważania na podobny temat i mówiąc o prawdopodobieństwie, że dana osoba jest winna, przyjmuje, że wynosi ono 1/100 000, ponieważ miasto, w którym popełniono przestępstwo, miało właśnie 100 000 mieszkańców. Oznaczałoby to, że $A(X = M, X \neq M) = 1/99\ 999$. We wcześniejszym tekście dotyczącym tych kwestii (1998) autor zmniejszył to prawdopodobieństwo, przyjmując, że wynosi ono 1/1 000 000, co daje $A(X = M, X \neq M) = 1/999\ 999$. Podobne oszacowania możemy znaleźć w pracy Donnelly'ego (2005) dotyczącej sprawy R v Adams (morderca miał zostać w tym przypadku zidentyfikowany na podstawie materiału DNA). Autor założył, że prawdopodobieństwo, że dana osoba popełniła przestępstwo w Wielkiej Brytanii, jest równe dla wszystkich jej mieszkańców i wynosi 1/60 000 000. Oszacowanie takie jest jednak ewidentnie zaniżone – przestępstwo wiązało się z gwałtem, a to wyklucza z grupy podejrzanych przynajmniej kobiety i osoby w określonym wieku. Jeśli przyjąć takie (i wyłącznie takie) ograniczenia (por. Sesardic, 2007), to grupa odniesienia powinna liczyć 10 000 000. W tym wypadku mamy więc $A(X = M, X \neq M) = 1/\ 9\ 999\ 999$.

Zachęcamy Czytelnika, by zanim przejdzie do dalszej części tekstu spróbował oszacować wartość $P(X = M | od(X) \approx od(M))$, przyjmując, że przestępstwa dokonano w miasteczku liczącym 10 000 mieszkańców, a więc że $A(X = M, X \neq M) = 1/9\ 999$[18].

Powstaje jednak pytanie, czy tak ustalony czynnik bazowy jesteśmy gotowi w danym kontekście uznać już za nasz ostateczny czynnik aprioryczny. Oczywiście, z jednej strony, ponieważ wnioskowanie ma rozstrzygnąć o winie X (a bardziej kosztowny jest błąd uznania osoby niewinnej za winną niż uznania osoby winnej za niewinną), to φ nie powinno być większe niż 1. W przeciwieństwie więc do diagnozy medycznej nie należy świadomie zawyżać czynnika apriorycznego. Przyjmijmy więc, że $\varphi = 0,5$. Z drugiej jednak strony wydaje się, że ustalając ten czynnik, pominęliśmy istotny fakt, że: *odciski palców osoby X były w policyjnej bazie danych* i że tylko dzięki temu mogliśmy (nie mając żadnych innych przesłanek) wskazać na osobę X jako na potencjalnego mordercę. Jest to policyjna baza danych, a więc nie zawiera zbioru osób wyróżnionych zupełnie losowo – są tam ci, którzy w przeszłości znaleźli się w polu zainteresowań policji. Wśród takich osób – ze względu na powszechność zjawiska

[17] Gdybyśmy np. wiedzieli jedynie, że X jest jedną z czterech osób, które mogły popełnić morderstwo, to domniemanie niewinności nakazywałaby nam uznać, że X jest winny z prawdopodobieństwem ¼.

[18] Podstawiając dane do wzoru ostatecznie otrzymujemy w zaokrągleniu

$$(21)\ \frac{P(X = M|(X) \approx (M))}{P(X \neq M|(X) \approx (M))} = \frac{1}{9\ 999} \times 91,5 = 0,009.$$

Jest to wielkość zaskakująco mała i na pewno nie skłania nas do przyjęcia wniosku o winie X (ponad 110 razy bardziej prawdopodobne jest to, że jest niewinny).

 DOI: 10.7206/DEC.1733-0092.102

recydywy[19] – niewątpliwie szansa na popełnienie morderstwa jest wyższa niż w całej populacji. Oznaczając przez D fakt, że odciski palców osoby X były w policyjnej bazie danych, uzyskujemy w sposób oczywisty następującą prawidłowość:

(22) $P(X = M|D) > P(X = M)$.

Trzeba teraz oszacować, na ile $P(X = M|D)$ jest większe od $P(X = M)$. Załóżmy, że osoby już raz notowane przez policję 100 razy częściej popełniają przestępstwo niż przeciętny obywatel[20]. Oznacza to, że przy wykorzystaniu informacji o tym, że odciski palców osoby X były w policyjnej bazie danych, otrzymujemy $\varphi = 100$.

Zakładając, że czynnik $BF(X = M, X \neq M)$ nie uległ zmianie (czyli skuteczność identyfikacji za pomocą danego testu jest taka sama w grupie osób figurujących w bazie danych jak w całej populacji) i ustalając ostatecznie $\varphi = (0{,}5 \times 100)$, po podstawieniu do wzoru (21) uzyskujemy:

(23) $\varphi \frac{P(X = M|(X) \approx (M))}{P(X \neq M|((X) \approx (M)))} = 50 \times \frac{1}{9\,999} \times 91{,}5 = 0{,}455,$

co daje nam wynik niewątpliwie dużo wyższy od tego, który obliczyliśmy pierwotnie i chociaż nie wystarcza do uznania osoby X winnym, to wzmacnia ten wniosek i jest np. dobrym argumentem do wniesienia przeciwko tej osobie oskarżenia[21].

Całe rozumowanie wydaje się rozsądne, mimo że uzyskany ostatecznie we wzorze (23) wynik odbiega od tego, który obliczyliśmy, podstawiając mechanicznie oficjalne dane do wzoru (21).

Podsumowując, w obu przeanalizowanych przykładach istnieją rozsądne powody, aby odróżniać czynnik związany z prawdopodobieństwem bazowym od czynnika apriorycznego albo, innymi słowy, aby przyjmować, że $\varphi \neq 1$.Wynika to po pierwsze stąd, że ze względu na specyfikę problemu i maksymalizowanie oczekiwanej użyteczności przyjęcia wniosku (związanej z oceną kosztów popełnienia błędu) już na wstępie zawyżamy lub zaniżamy prawdopodobieństwo bazowe. Po drugie zaś wynika stąd, że w obu wypadkach – przyjmując dość oczywiste założenia dodatkowe (odpowiednio o tym, że kobieta należy do drugiego typu osób robiących badania i o tym, że umieszczenie w policyjnej bazie danych zwiększa prawdopodobieństwo bycia przestępcą) – powinniśmy dokonać dalszej modyfikacji czynnika apriorycznego. Oczywi-

[19] Według danych dostępnych np. pod adresem *https://isws.ms.gov.pl/pl/baza-statystyczna/publikacje* dotyczących powrotności do przestępstwa w Polsce, wynosi ono 25%. Oznacza to, że co czwarta osoba skazana prawomocnym wyrokiem sądu popełnia przestępstwo w ciągu pierwszych pięciu lat po wyjściu na wolność.

[20] Przykładowo w przypadku przestępstw seksualnych szacuje się, że osoby, które już raz popełniły takie przestępstwo, popełnią je ponownie z prawdopodobieństwem 80-250 razy większym niż przeciętny obywatel (por. np. Redmayne, 2002) – cyt. za De Macedo, 2008.

[21] Co ciekawe, tego typu rozumowanie – wykorzystujące fakt, że X popełnił już wcześniej jakieś przestępstwo – jest niedopuszczalne w niektórych stanach USA jako argument wzmacniający wniosek o winie osoby X (por. np. przypis 16 w pracy Koehlera, 1996).

 DOI: 10.7206/DEC.1733-0092.102

ście zaproponowane tu oszacowania wielkości j są bardzo zgrubne, ale niewątpliwie nie można im zarzucić całkowitej arbitralności. Uznanie, że $\varphi \neq 1$, i kierunek jego zmian są wyznaczone przez specyficzne własności sytuacji, której dotyczy analizowane rozumowanie.

Trzeci omówiony przykład jest najbardziej skomplikowany i niejednoznaczny. Dotyczy bardzo głośnej swego czasu sprawy i boleśnie pokazuje, jak trudno – w realnych okolicznościach – ocenić, czy przeprowadzone wnioskowanie można uznać za racjonalne i jak łatwo wygłaszać na ten temat pochopne opinie.

5. SIDS *VS* SERYJNE MORDERSTWO

Opisany casus jest jednym z najgłośniejszych i najbardziej kontrowersyjnych przypadków użycia wnioskowania opartego na rachunku prawdopobieństwa w praktyce prawniczej. Co ciekawe, funkcjonuje on obecnie w literaturze w podwójnej roli. Z jednej strony (zdecydowanie częściej) podaje się go jako przykład błędnego użycia narzędzi, jakie dostarcza rachunek prawdopodobieństwa przy ocenie poprawności wnioskowania. Oskarżycielom w procesie Sally Clark przypisuje się zastosowanie modelowej wersji sofizmatu prokuratora[22] (por. np. Royal Statistical Society, 2001, i Dawid, 2002). Z drugiej jednak strony służy jako egzemplifikacja tego, jak trudno takie narzędzie w sposób odpowiedzialny stosować (por. np. Sesardic, 2007).

Najistotniejsze fakty dotyczące tej sprawy są następujące. Sally Clark między 1996 a 1998 rokiem straciła dwoje dzieci. Pierwsze z nich zmarło w wieku 11 tygodni. Ponieważ nie znaleziono żadnej przyczyny śmierci, stwierdzono, że jest to przypadek tzw. zespołu nagłej śmierci łóżeczkowej (ang. SIDS). Drugie dziecko Sally Clark zmarło w wieku 8 tygodni i ponieważ również nie znaleziono żadnej przyczyny, także w tym wypadku stwierdzono SIDS. Na wniosek znanego psychiatry Roya Meadowa z Uniwersytetu w Leeds, zajmującego się m.in. badaniem zespołu zaburzeń MSbP (Munchausen Syndrome by Proxy)[23], sprawą zajęła się prokuratura. Zdaniem oskarżyciela, który opierał się na danych dostarczonych przez Meadowa, szansa, że dwoje dzieci w rodzinie umrze z powodu SIDS, jest nikła, i zachodzi podejrzenie, iż to matka dokonała podwójnego morderstwa. Argumentacja przedstawiona w sądzie była na tyle przekonująca, że Sally Clark została skazana i trafiła do więzienia w 1999 roku.

[22] Sofizmat prokuratora w tym wypadku polegałby na utożsamieniu dwóch prawdopodobieństw warunkowych: *Sally Clark nie zabiła swoich dzieci pod warunkiem, że miały miejsce wszystkie przedstawione w sądzie poszlaki* i *miały miejsce wszystkie przedstawione w sądzie poszlaki pod warunkiem, że Sally Clark nie zabiła swoich dzieci*.

[23] Jest to zaburzenie polegające na tym, że matka wywołuje u swojego dziecka choroby, by zyskać uwagę i współczucie otoczenia.

 DOI: 10.7206/DEC.1733-0092.102

Sprawa wywołała duże poruszenie. Brytyjskie Królewskie Towarzystwo Statystyczne (Royal Statistical Society, RSS) krytykowało Meadowa, eksperta oskarżenia, twierdząc, że dokonał nieuprawnionej manipulacji danymi statystycznymi, a prokuratora ganiło za to, że przeprowadził w oparciu o te dane wadliwe rozumowanie.

Meadowowi zarzucono, że obliczając prawdopodobieństwo śmierci dwojga dzieci w danej rodzinie, podniósł po prostu prawdopodobieństwo śmierci jednego dziecka do kwadratu (uzyskując w przybliżeniu wynik równy 1/73 mln), traktując te zdarzenia jako niezależne. Według członków RSS jest to błąd absolutnie dyskwalifikujący go jako naukowca. Twierdzono również, że zasugerował sądowi wnioskowanie opierające się na sofizmacie prokuratora, podając dane dotyczące prawdopodobieństwa śmierci dwojga dzieci przy założeniu, że oskarżona jest niewinna. Dodatkowo niesłusznie przyjął wartość prawdopodobieństwa śmierci dziecka z powodu SIDS charakterystyczną dla zamożnych, niepalących rodzin, gdyż można by się zastanawiać, czy jest to właściwa grupa odniesienia dla danego przypadku (chociaż rzeczywiście rodzina Clarków była zamożna i niepaląca). W efekcie w 2005 roku Meadowa pozbawiono prawa wykonywania zawodu. Sąd apelacyjny uchylił jednak to postanowienie, nie dopatrując się w działaniu Meadowa żadnych rażących uchybień i stwierdzając, że werdykt sądu skazujący Sally Clark oparty był na innych poszlakach – w szczególności urazach, jakie odkryto na ciałach dzieci. Ustalono, że podając prawdopodobieństwo śmierci dwojga dzieci w danej rodzinie, Meadow rozumiał, jak należy obliczać prawdopodobieństwo iloczynu zdarzeń i świadomie traktował te zdarzenia jako niezależne (świadczą o tym jego wcześniejsze publikacje naukowe na ten temat). Co najwyżej można by go obwiniać za to, że nie miał w tej kwestii racji, ale taki zarzut należałoby postawić również innym (licznym) pediatrom i patologom sądowym, którzy podzielają głoszony przez niego pogląd.

Sally Clark dwukrotnie wnosiła apelację (w latach 2000 i 2003) i ostatecznie została zwolniona z więzienia po trzech latach. Zaważyły na tym dodatkowe dowody dotyczące możliwej infekcji, na jaką cierpiał jej drugi syn przed śmiercią. Jej uniewinnienie jest traktowane często jako zwycięstwo kompanii prowadzonej przez RSS, a cała sprawa służy jako przykład błędnej sądowej argumentacji – sofizmatu prokuratora (por. np. https://en.wikipedia.org/wiki/Sally_Clark). Sally Clark zmarła w 2007 roku w wyniku choroby alkoholowej.

Rozpatrując ten przykład w stylistyce poprzednich, postaramy się oszacować poszczególne wartości konieczne do ustalenia, który wniosek jest bardziej racjonalny – taki, że Sally Clark zabiła swoje dzieci, czy taki, że zmarły one z powodu SIDS.

Przyjmijmy najpierw następujące oznaczenia:

– przez „2S” będziemy oznaczać stwierdzenie, że dwoje dzieci Sally Clark zmarło na SIDS;

 DOI: 10.7206/DEC.1733-0092.102

– przez „2M” będziemy oznaczać stwierdzenie, że Sally Clark dokonała podwójnego morderstwa;
– przez „E1” będziemy oznaczać stwierdzenie, że pierwsze dziecko Sally Clark zmarło;
– przez „E2” będziemy oznaczać stwierdzenie, że drugie dziecko Sally Clark zmarło;
– przez „E” będziemy oznaczać stwierdzenie dotyczące pozostałych dowodów w sprawie (istnienie obrażeń na ciałach dzieci, takich jak wylewy, siniaki, pęknięcia kości, zachowanie Sally Clark po śmierci dzieci);
– przez „S1” będziemy oznaczać stwierdzenie, że pierwsze dziecko Sally Clark zmarło na SIDS;
– przez „S2” będziemy oznaczać stwierdzenie, że drugie dziecko Sally Clark zmarło na SIDS;
– przez „M1” będziemy oznaczać stwierdzenie, że pierwsze dziecko Sally Clark zostało zamordowane;
– przez „M2” będziemy oznaczać stwierdzenie, że drugie dziecko Sally Clark zostało zamordowane.

Naszym celem jest obliczenie:

$$(24)\ \frac{P(2S|E1 \wedge E2 \wedge E)}{P(2M|E1 \wedge E2 \wedge E)} = \frac{P(2S)}{P(2M)} \times \frac{P(E1 \wedge E2 \wedge E|2S)}{P(E1 \wedge E2 \wedge E|2M)}$$

albo, innymi słowy,

$$(25)\ \frac{P(2S|E1 \wedge E2 \wedge E)}{P(2M|E1 \wedge E2 \wedge E)} = A(2S, 2M) \times BF(2S, 2M).$$

Zastanówmy się najpierw nad wielkością czynnika bayesowskiego. Gdybyśmy jako jedyne dowody w sprawie mieli informacje o śmierci obydwojga dzieci (czyli informacje E1 i E2), to oczywiście

$$(26)\ P(E1 \wedge E2 | 2M)) = P(E1 \wedge E2 | 2S) = 1.$$

Mamy jednak jeszcze do dyspozycji informację E, która, sądząc z protokołu rozprawy, odegrała bardzo ważną rolę w zapadnięciu wyroku skazującego. Nie było bowiem tak, że jedynym argumentem na rzecz skazania Sally Clark było rozumowanie oparte na rachunku prawdopodobieństwa (por. argumenty na ten temat przedstawione przez Sesardica, 2007). W odróżnieniu od poprzednich dwóch przykładów nie możemy tu jednak podać jakiejś ugruntowanej badaniami wartości interesującego nas czynnika BF(2M, 2S). Z przebiegu rozprawy należy wnosić, że sędziowie uznali, że $P(E|2M) > P(E|2S)$, tzn. że posiadane dowody lepiej uzasadniają wniosek o podwójnym morderstwie niż o podwójnej śmierci w wyniku SIDS. Na ile lepiej? Można

 DOI: 10.7206/DEC.1733-0092.102

jedynie to jakoś oszacować, opierając się na uzasadnieniu wyroku. Przyjmijmy (za Sesardiciem, 2007, i Dawidem, 2002), że

(27) $\mathrm{BF}(2S, 2M) = 0{,}2$.

Ze względu na problem poruszany w niniejszym artykule interesuje nas jednak przede wszystkim to, jak oszacować czynnik:

(28) $A(2S, 2M) = \frac{P(2S)}{P(2M)}$.

Przede wszystkim zauważmy, że ułamek ten można rozłożyć na mniejsze składniki. Mamy bowiem

(29) $P(2S) = P(S1 \wedge S2) = P(S1) \times P(S2|S1)$

i

(30) $P(2M) = P(M1 \wedge M2) = P(M1) \times P(M2|M1)$.

Stąd

(31) $A(2S,2M) = \frac{P(S1)}{P(M1)} \times \frac{P(S2|S1)}{P(M2|M1)}$.

Pierwszy z tych ułamków ma nam powiedzieć, co jest częstszą przyczyną śmierci dzieci w pierwszym roku życia – morderstwo popełnione przez matkę czy SIDS. Oficjalnie w Wielkiej Brytanii zdarza się rocznie średnio 360 przypadków śmierci z powodu SIDS i 30 przypadków orzeczonych morderstw dokonanych na dzieciach w ich pierwszym roku życia. Gdybyśmy opierali się wyłącznie na takiej statystyce, mielibyśmy (por. dane dotyczące liczby urodzin przedstawione w Sesardic, 2007)) następujące wartości składników czynnika bazowego:

(32) $\frac{P(S1)}{P(M1)} = \frac{0{,}000768}{0{,}000047} = 16{,}4$.

Jeśli teraz czysto mechanicznie (argumentując np. tak, jak zrobił to Dawid, 2002) będziemy obliczali $\frac{P(S2|S1)}{P(M2|M1)}$, traktując jako zdarzenia niezależne zarówno kolejne przypadki śmierci w wyniku SIDS, jak i kolejne przypadki morderstwa w danej rodzinie, to otrzymamy:

(33) $\frac{P(S2|S1)}{P(M2|M1)} = \frac{0{,}000768}{0{,}000047} = 16{,}4$,

a więc ostatecznie czynnik bazowy będzie wynosił

(34) $A(2S, 2M) = 16{,}4 \times 16{,}4 = 268{,}96$.

Podstawiając te dane do wzoru (22), otrzymujemy

(35) $\frac{P(2S|E1 \wedge E2 \wedge E)}{P(2M|E1 \wedge E2 \wedge E)} = 268{,}96 \times 0{,}2 = 53{,}792$.

 DOI: 10.7206/DEC.1733-0092.102

Oznacza to, że ponad 53 razy bardziej prawdopodobne jest to, że dzieci Sally Clark zmarły w wyniku SIDS, niż to, że zamordowała je matka. Jest to silny argument na rzecz niewinności Sally Clark i słuszności kampanii na rzecz oczyszczenia jej z zarzutów, którą zapoczątkowało RSS.

Czy jednak rzeczywiście gotowi jesteśmy czynnik bazowy utożsamić z naszym czynnikiem apriorycznym? Czy chcemy całkowicie zaufać danym statystycznym?

Niezależnie od tego, skąd zaczerpniemy informacje na temat śmierci dzieci do pierwszego roku życia w wyniku SIDS i w wyniku morderstwa, rzucają się w oczy następujące ich cechy:

- oszacowania opierają się na dość małych próbkach;
- istnieje duża asymetria przy stwierdzaniu, czy śmierć dziecka nastąpiła w wyniku morderstwa, czy z powodu SIDS. Rodzic, który zabije swoje dziecko, będzie twierdził, że zmarło na SIDS, ale w sytuacji, kiedy dziecko umrze z powodu SIDS, żaden rodzic nie przyzna się do morderstwa. Ma to związek z samą definicją SIDS – jest to przypadek śmierci, *w którym również nie można wskazać żadnej istotnej przyczyny*. Jeśli nie uda się udowodnić, że zdrowe dziecko zostało zamordowane, orzeka się, że śmierć nastąpiła w wyniku SIDS[24].

W literaturze toczy się spór o to, jaki procent przypadków śmierci dzieci zakwalifikowanych jako SIDS było faktycznie efektem morderstwa. Niektóre szacowania mówią o tym, że jest to 5%, inne, że 10%, a jeszcze inne, że waha się on między 20% a 40% (na ten temat por. np. Koehler i Applegate, 2013, oraz cytowane tam źródła). Przyjmijmy tak jak Sesardic (2007), że błędnych kwalifikacji jest 10%. Oznacza to, że w oparciu o dane statystyczne, ale po dokonaniu korekty (czyli uznając, że było 327 przypadków SIDS i 66 morderstw), uzyskujemy już inną wartość składnika czynnika apriorycznego:

$$(36)\ \varphi\frac{\mathrm{P(S1)}}{\mathrm{P(M1)}} = \frac{0{,}000691}{0{,}000124} = 5{,}6.$$

Rozważmy teraz iloraz $\frac{\mathrm{P(S2|S1)}}{\mathrm{P(M2|M1)}}$.

Wbrew zarzutom stawianym Meadowi przez RSS, współczesne opracowania na ten temat (por. np. American Academy of Pediatrics, 2001, s. 439) wskazują, że ryzyko śmierci drugiego dziecka w wyniku SIDS w rodzinie, w której jedno zmarło już z tego powodu, nie zwiększa się[25]. Przyjmijmy zatem, że

[24] Jak brutalnie stwierdził jeden z pediatrów zajmujących się takimi sprawami, jedynym, co faktycznie odróżnia śmierć dziecka w wyniku zabójstwa od śmierci w wyniku SIDS, jest przyznanie się matki do popełnienia tego czynu (por. też Koehler i Applegate, 2013, s. 19).

[25] Klasyczną pracą, w której udowadniana była teza o istnieniu nieznanego czynnika genetycznego mogącego zwiększać prawdopodobieństwo kolejnej śmierci w wyniki SIDS w danej rodzinie, jest publikacja Steinsch-

 DOI: 10.7206/DEC.1733-0092.102

(37) $P(S2|S1) = P(S1) = 0{,}000691$,

zakładając takie same błędne kwalifikacje 10% przypadków, jak poprzednio.

Nie wydaje się jednak, że analogiczną równość można przyjąć w przypadku oceny prawdopodobieństwa drugiego morderstwa w rodzinie. Jest oczywiste, że prawdopodobieństwo dokonania drugiego morderstwa przez osobę, która już raz zabiła, jest wyższe niż dokonanie pierwszego morderstwa przez losową osobę:

(38) $P(M2|M1) > P(M1)$.

To, ile jednak wynosi takie prawdopodobieństwo, trudno stwierdzić w oparciu o bardzo ubogie i niejednoznaczne dane (nie wiadomo np., jak odróżnić zwykły nieszczęśliwy wypadek, który powoduje śmierć dziecka, od zamierzonego działania rodziców).

Zauważmy również, że – podobnie jak we wcześniej rozważanych przykładach – powinniśmy uwzględnić pewną istotną informację, która tkwi w całym problemie niejako *implicite*. A mianowicie, że śmierć pierwszego dziecka Sally Clark *została na początku uznana za SIDS*. Jeśli oznaczymy to zdarzenie przez U, to faktycznie należy oszacować wielkość $P(M2|M1 \wedge U)$, a więc to, jaka jest szansa, że w rodzinie zostanie zabite drugie dziecko, *o ile śmierć pierwszego dziecka została błędnie zakwalifikowana i sprawca nie poniósł za swój czyn żadnej kary*. Przekonanie o bezkarności jest niewątpliwie dodatkowym czynnikiem zwiększającym prawdopodobieństwo popełnienia po raz kolejny czynu tego samego rodzaju[26].

Joyce (2002), która zresztą była zaangażowana w obronę Sally Clark, a za nią Sesardic, oszacowali, że

(39) $P(M2|M1) = 0{,}1$.

Stąd

(40) $\varphi\frac{P(S2|S1)}{P(M2|M1)} = \frac{0{,}000691}{0{,}1} = 0{,}00691$.

Ostatecznie więc czynnik aprioryczny wynosi

(41) $\varphi A(2S, 2M) = 5{,}6 \times 0{,}00691 = 0{,}038696$,

neidera (1972). Artykuł ten miał duży wpływ na utrwalenie się takiego poglądu w literaturze medycznej. Okazało się jednak, że w rodzinie badanej przez autora matka – Waneta Hoyt – ostatecznie przyznała się do zabicia pięciorga swoich dzieci.

26 Dyskusję na ten temat dobrze podsumowuje przytoczona w artykule Koehlera i Applegate (2013, s. 18) uwaga na temat przekonania funkcjonującego w środowiskach patologów sądowych: nieoczekiwana śmierć pierwszego dziecka w danej rodzinie może być zakwalifikowana jako SIDS, nieoczekiwana śmierć drugiego dziecka powinna być bardzo dokładnie przeanalizowana, natomiast nieoczekiwana śmierć trzeciego dziecka powinna być uznana za morderstwo, a poglądy dotyczące śmierci dwojga poprzednich dzieci powinny zostać zweryfikowane.

 DOI: 10.7206/DEC.1733-0092.102

co oznacza, że *podwójne morderstwo jest a priori ponad 25 razy bardziej prawdopodobne niż podwójna śmierć z powodu SIDS*.

Podstawiając uzyskane wyniki do wzoru (25) otrzymujemy

$$(42)\ \frac{P(2S|E1 \wedge E2 \wedge E)}{P(2M|E1 \wedge E2 \wedge E)} = 0{,}03869 \times 0{,}2 = 0{,}0077392.$$

Na podstawie powyższych wyliczeń należy stwierdzić, że wniosek o tym, że Sally Clark popełniła podwójne morderstwo, jest ponad 125 razy bardziej prawdopodobny niż wniosek, że jej dzieci zmarły w wyniku SIDS. A to oznacza, że kampania na rzecz oczyszczenia Sally Clark z zarzutów, którą zapoczątkowało RSS, była niczym nieuzasadniona i niewłaściwa.

6. Podsumowanie

W realnych, życiowych sytuacjach stwierdzenie, który z dwóch konkurencyjnych wniosków jest lepiej uzasadniony na podstawie posiadanych przez nas przesłanek, bardzo rzadko jest proste. Jedynie w szczególnych okolicznościach wystarczy w tym celu podstawić do wzoru jednoznaczne dane i dokonać czasochłonnego wyliczenia. Najczęściej kontekst, w którym wnioskowanie jest zanurzone, zmusza nas do modyfikacji dostępnych danych liczbowych. Dotyczy to szczególnie czynnika bazowego. Kierunek takiej modyfikacji wynika czasami stąd, że staramy się uniknąć popełnienia błędu, który jest bardziej kosztowny z punktu widzenia użyteczności przypisywanej różnym konsekwencjom przyjęcia danego wniosku. Zdarza się również, że po prostu nie ufamy istniejącym danym – temu, że zgłaszająca się do lekarza pacjentka ma dokładnie taką samą szansę bycia chorą jak ktoś, kto do lekarza nie przyszedł, że osoba figurująca w policyjnej bazie danych jest równie niewinna, jak nasz uroczy sąsiad, i temu, że jeśli w rodzinie jedno dziecko zostało już bezkarnie zamordowane, to nie zwiększa to ryzyka śmierci drugiego dziecka. Oczywiście oparte na takich modyfikacjach rozstrzygnięcia nie muszą zawsze prowadzić do akceptacji wniosku, który jest prawdziwy. Celem artykułu nie było w szczególności wykazanie, że Sally Clark popełniła podwójne morderstwo. Chodziło o zwrócenie uwagi na to, że aby ocenić, czy ktoś wnioskuje racjonalnie, nie zawsze wystarcza sprawdzenie, czy wynik jego wyliczeń zgadza się z tym wynikiem, który sami otrzymaliśmy. Wydaje się, że dobrym narzędziem jest tu odróżnienie czynnika bazowego (który często ma charakter obiektywny i interpersonalny) i czynnika apriorycznego, który jest funkcją specyficznych, przyjmowanych przez wnioskujący podmiot założeń. Dopiero ocena tych założeń jako racjonalnych bądź nie powinna decydować o ocenie całego wnioskowania. Jeśli przyjmiemy taką perspektywę, to należy ostrożniej podchodzić do ferowanych przez niektórych badaczy wyroków na temat mizernej ludzkiej racjonalności.

Bibliografia

__ (2001). *News Release: Royal Statistical Society Concerned by Issues Raised in Sally Clark Case*. Royal Statistical Society, October 23.

American Academy of Pediatrics (2001). Distinguishing Sudden Infant Death Syndrome from Child Abuse Fatalities. *Pediatrics*, *107*, 437–441.

Bradford, T. (2011). Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, t. *108*(19), 7733–7738.

Cieślińska, I., (2014). Nieznajomość matematyki zabija, wyborcza.pl/.../1,160474,17000749,.html

Clarke, K. (2001). *The effect of priors on approximate Bayes factors from MCMC output*. Mimeograph, University of Rochester, Rochester, NY.

Dawid, A.P. (2002). Bayes's Theorem and Weighing Evidence by Juries, [w:] R. Swinburne (red.), *Bayes's Theorem*. Oxford: Oxford University Press, 71–90.

De Macedo, C. (2008). Guilt by Statistical Association: Revisiting the Prosecutor's Fallacy and the Interrogator's Fallacy. *The Journal of Philosophy,105*(6), 320–332.

Domurat, A. Białek, M. (2016). Dowodzenie hipotez za pomocą czynnika bayesowskiego (Bayes factor): przykłady użycia w badaniach empirycznych. *Decyzje*, *26*, 109–141.

Donnelly, P. (2005). Appealing statistics. *Significance*, *2*, 46–48.

Gigerenzer, G. (1998). Ecological Intelligence: An Adaptation for Frequencies. W: D.D.

Cummins, C. Allen, (red.), *The Evolution of Mind*. New York: Oxford, 9–29.

Gigerenzer, G. (2002). *Reckoning with Risk: Learning to Live with Uncertainty*. New York: Penguin, 156–58.

Fischhoff, B., De Bruin, W. (1999). Fifty–Fifty = 50%? *Journal of Behavioral Decision Making*, *12*, 149–163.

Hawthorne, J., & Bovens, L. (1999). The Preface, the Lottery, and the Logic of Belief. *Mind,108*(430), 241–264. Retrieved from http://www.jstor.org/stable/2660219.

Huber, F. (2008). Assessing theories, Bayesian style. *Synthese*, *161*, 89–118.

Joyce, H. (2002): Beyond Reasonable Doubt, *Plus*, *Issue 21*, <plus.maths.org/issue21/features/clark/>.

Jørgensen, K.J., Gøtzsche, P.C. (2009). Over diagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ 2009*, 339:b2587.

Kass, R.E., & Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical. Association*, 90, 773–795.

Kaye, D.H., Koehler, J.J (2003). The Misquantification of Probative Value. *Law and Human Behavior*, *27*(6), 645–659.

Koehler, J.J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, *19*(1), 1–53.

Koehler, S.A., Applegate, K.M (2013). Covert homicide: when sids is not sids, reasons for the missed identification. *Pediatrics Today*, *9*(1), 13–23.

Kuklin, B. (2006). Probability Misestimates in Medical Care. *Arkansas Law Review*, *59*, 527–554.

Levene, S. Bacon, C. J. (2004). Sudden Unexpected Death and Covert Homicidein Infancy. *Archives of Disease in Childhood*, *89*, 443–447.

Nickerson R.S. (2000). Null Hypothesis Significance Testing: A Review of an Old and Continuing Controversy. *Psychological Methods*, *5*(2), 241–301.

Redmayne M. (2002). *The Relevance of Bad Character*. Cambridge Law Journal, l xi: 695

Sesardic, N. (2007). Sudden infant death or murder? A royal confusion about probabilities. *British Journal for the Philosophy of Science*, *58*(2), 299–329.

Steinschneider, A. (1972). Prolonged Apnea and the Sudden Infant Death Syndrome: Clinical and Laboratory Observations. *Pediatrics*, *50*, 646–54.

Tversky, A. Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185,* 1124–1131.

Westover, M.B., Westover, K.D., Bianchi, M.T. (2011). Significance testing as perverse probabilistic reasoning. *BMC Medicine*, *2011 Feb 28*, 9–20.

Wegwarth O, Gigerenzer G. (2013). Over diagnosis and Over treatment Evaluation of What Physicians Tell Their Patients About Screening Harms. *JAMA Intern Med.*, *173*(22), 2086–2088.

Welsh, M.B., Navarro, D.J. (2012). Seeing is believing: Priors, trust, and base rate neglect. *Organizational Behavior and Human Decision Processes*, *119*(1), 1–14.

DOI: 10.7206/DEC.1733-0092.102