H. STEINHAUS (Wrocław)

PRAWDOPODOBIEŃSTWO, WIAROGODNOŚĆ I MOŻLIWOŚĆ

Przedmiot niniejszej pracy jest związany z nazwiskiem Tomasza Bayesa[1]). Związek jest istotny, bo Bayes nie tylko odkrył sposób obliczania prawdopodobieństw przyczyn, których skutki obserwujemy, lecz także był pierwszym, w którym to odkrycie wzbudziło wątpliwości.

1. Przypomnijmy klasyczny wzór Bayesa. Niech $A, B, C, \ldots, N$ będzie kompletem wyłączających się nawzajem ewentualności; nazywamy je *przyczynami* zjawiska Z, którego zajście zaobserwowaliśmy. Chodzi o prawdopodobieństwo *a posteriori* $P_Z(A)$ tego, iż to właśnie A (a nie $B, C, \ldots$ lub N) poprzedziło Z. Znane są z założenia tak zwane *prawdopodobieństwa aprioryczne* $P(A), P(B), \ldots$ $\ldots, P(N)$, czyli prawdopodobieństwa każdej z ewentualności $A, B, \ldots, N$ zanim zaobserwowano „skutek" Z, i znane są tak zwane *prawdopodobieństwa warunkowe* $P_A(Z), P_B(Z), \ldots, P_N(Z)$ — tu $P_A(Z)$ oznacza prawdopodobieństwo faktu, że po A nastąpi Z, $P_B(Z)$ prawdopodobieństwo, że po B nastąpi Z, *etc.* Reguła Bayesa wyraża $P_Z(A)$ przez znane prawdopodobieństwa aprioryczne i warunkowe:

$$P_Z(A) = \frac{P(A)P_A(Z)}{P(A)P_A(Z) + P(B)P_B(Z) + \ldots + P(N)P_N(Z)}. \tag{1}$$

Regule (1) nie można nic zarzucić z formalnego punktu widzenia, ale rzadko się zdarza sytuacja, w której można ją zastosować poprawnie do realnych zagadnień, bo zwykle nie są znane prawdopodobieństwa aprioryczne. Ciekawym wyjątkiem jest dochodzenie ojcostwa przed sądem: dzięki badaniom Hirszfelda obejmującym przeszło dwa tysiące spraw o orzeczenie ojcostwa udało się obliczyć wcale dokładnie prawdopodobieństwo aprioryczne, że pozwany

[1]) T. Bayes, *An essay towards solving a problem in the doctrine of chances*, Philosophical Transactions 53 (1763), p. 370.

o alimenta jest ojcem — wynosi ono (w Polsce przedwojennej) około 70%[2]. Tutaj prawdopodobieństwo aprioryczne odnosi się do momentu zarządzenia ekspertyzy serologicznej, ale przed poznaniem jej rezultatu Z; po ekspertyzie można obliczyć z wzoru (1) prawdopodobieństwo *a posteriori* $P_Z(A)$, że pozwany jest ojcem dziecka, w którego imieniu wniesiono pozew o alimenta — właśnie to prawdopodobieństwo *a posteriori* interesuje sędziego. Ale nie o wyjątkach mamy tu mówić, lecz o tych codziennie powstających sytuacjach, w których nie znamy prawdopodobieństw apriorycznych. Dostarcza takich sytuacyj niemal każde badanie przyrodnicze lub techniczne, tak, że zagadnienie uratowania wzoru (1) lub zastąpienie go przez inny wzór jest fundamentalnym problematem indukcji przyrodniczej i wnioskowania statystycznego. Takie typowe pytanie, ile jest chorych na gruźlicę w mieście, w którym zbadano 10 osób wybranych na chybił-trafił i znalezione 2 chore, albo ile jest w magazynie dobrych prętów do żelbetu, jeżeli zbadano wytrzymałość 10 z nich i 2 okazały się za słabe, różnią się tylko pozornie — ich treść matematyczna jest jednakowa i należy do problematyki niniejszej pracy. W żadnym z tych przykładów nie znamy prawdopodobieństw apriorycznych.

Bayes radził sobie w takich sytuacjach postulatem, który nazwiemy jego nazwiskiem i oznaczymy literą $\mathfrak{B}$. Przyjmuje on mianowicie, że rozkład prawdopodobieństw apriorycznych jest *równomierny*, to znaczy, że liczby $P(A), P(B), \ldots, P(N)$ są sobie równe. W przykładzie dochodzenia ojcostwa są tylko dwie ewentualności, A i B (pozwany jest ojcem, nie jest ojcem). Postulat Bayesa orzekałby zatem, że $P(A) = P(B) = 0{,}5$. Rzeczywiście, niektórzy antropologowie posługiwali się wzorem (1) zakładając $P(A) = 0{,}5$, lecz nie zdawali sobie sprawy z tego, że używają reguły Bayesa i jego postulatu — tym mniej można ich podejrzewać o świadomość, że dotknęli zagadnienia zasadniczego. W tym przykładzie postulat Bayesa jest wręcz fałszywy, gdyż jest $P(A) = 0{,}70$, jak wyżej wspomnieliśmy. W przykładzie prętów do żelbetu postulat Bayesa oznaczałby, że występują z równą częstością partie prętów o jakości 1%, 2%, 3%, ..., 100%; taka hipoteza ani nie da się uzasadnić teoretycznie, ani nie wskazuje na nią doświadczenie.

---

[2]) H. Steinhaus, *O dochodzeniu ojcostwa*, Zastosowania Matematyki 1 (1953), str. 67-82.

Powiedzmy od razu, że postulat Bayesa nie budziłby sprzeciwów, gdyby stosowano go do przypadków, w których równomierność ma uzasadnienie empiryczne lub teoretyczne, lub do takich, w których przyjmuje się go za hipotezę roboczą aż do czasu, gdy doświadczenie wykaże, jaki jest rzeczywisty rozkład aprioryczny. Tymczasem najczęściej przyjmuje się ten postulat w roli dogmatu — jest on najpotrzebniejszy właśnie wtedy, gdy nie ma wcale nadziei, że okaże się zgodny lub sprzeczny z rzeczywistością. Weźmy przykład nowego leku. Jaki sens ma tutaj przyjęcie równomiernego rozkładu siły leku? Znaczyłoby to, że równie często pojawiają się w laboratoriach nowe leki o sile od 10% do 15%, jak leki o sile od 55% do 60% etc. Ta hipoteza nie będzie nigdy ani zbita, ani sprawdzona, bo leków przeciw danej chorobie jest mało, każdy pochodzi z innej epoki naukowej i wnioskowanie ze starych leków surowicowych na nowy lek syntetyczny, który ma zupełnie inne oparcie teoretyczne, jest sztuczne i niezgodne z rozsądkiem.

2. Dogmat równomierności, czyli postulat Bayesa, obudził namiętny sprzeciw angielskiej szkoły statystycznej. Odnowiciel statystyki R. A. Fisher stworzył koncepcję „fiducial probability", a polski uczony Jerzy Spława-Neyman, który rozpowszechnił w Stanach Zjednoczonych nowsze metody statystyczne, wprowadził pojęcie „przedziału ufności" (zauważmy, że nazwa „confidence interval" powstała z przetłumaczenia terminu polskiego na angielski). Już przed drugą wojną światową przyjęło się powszechnie mniemanie, że tylko te nowe koncepcje, wolne od postulatu Bayesa, są poprawne. Toteż znakomity znawca rachunku prawdopodobieństwa W. Feller przyrównuje zwolenników reguły i postulatu Bayesa „którzy jej używają z racji jej logicznej dopuszczalności i zgodności z naszym sposobem myślenia" do „Platona, który używał argumentu tego samego typu, by udowodnić istnienie Atlantydy"...Konkluduje: „...współczesna teoria testów statystycznych i estymacji jest mniej intuitywna, ale bardziej realistyczna. Można ją nie tylko bronić, ale i stosować".[3]) Jest charakterystyczne w podręczniku Fellera zepchnięcie reguły Bayesa do petitu.

Jak już mówiliśmy, dylemat „pro czy contra Bayes?" nie jest błahy, bo obejmuje szerokie pole zastosowań od geodezji do medy-

[3]) W. Feller, *An Introduction to Probability Theory and its Applications*, New York 1950, str. 85.

cyny; nie zapominajmy o tym, że zarówno zatwierdzenie zdjęcia geodezyjnego obszaru, na którym ma stanąć miasto, jak dopuszczenie do użytku krajowej penicyliny wymagają wnioskowania o prawdziwych wielkościach ze spostrzeżeń obarczonych błędami — w pierwszym przykładzie są takimi wielkościami prawdziwe współrzędne punktów orientacyjnych, w drugim siła penicyliny. Zamęt na tym polu jest tym większy, że zanim przyrodnicy, lekarze i inżynierowie nauczyli się reguły Bayesa, już zakazano im nią się posługiwać; wielu z nich sądzi, że chodzi tu o spory terminologiczne matematyków bez praktycznego znaczenia. W zakresie statystycznej kontroli jakości powstała Walda analiza sekwencyjna omijająca regułę Bayesa; korzyści, jakie daje zasada sekwencyjna, przypisano odrzuceniu tej reguły i uznano za nowy argument przeciw niej [4]). Z obozu jej obrońców uszli niemal wszyscy poważni matematycy [5]), a pozostali w nim niemal sami tacy, co nie zrozumieli zarzutów nowej szkoły. Dla krótkości nazwiemy nową doktrynę *teorią wiarogodności*. Jeden tylko Norbert Wiener [6]) ośmielił się nazwać ją „trikiem terminologicznym". Niezależnie od niego J. Oderfeld zwrócił uwagę na podobieństwo przepisów postępowania w statystycznej kontroli jakości wynikających z reguły i postulatu Bayesa do przepisów wynikających z nowej teorii [7]). To zachęciło mnie do przestudiowania podstaw statystycznej kontroli jakości: okazało się, że ze znanych sposobów żaden nie ma zasadniczego prymatu, a różnią się od siebie tylko mniej lub więcej skutecznym zamaskowaniem arbitralnych hipotez [8]). Obecnie chcę zająć się wzajemnym stosunkiem między regułą i postulatem Bayesa a odpowiednikiem tych sposobów w nowej teorii.

Tu powstaje trudność przez to, że w podręcznikach spotykamy różne terminy, jak „likelihood", „verisimilitude", „fiducial probability", „confidence interval", z których żaden nie jest szukanym

---

[4]) H. Steinhaus, *Quality control by sampling*, Colloquium Mathematicum 2 (1951), str. 98-108. Por. § 7 na str. 105-7.

[5]) Pozostał znany astronom H. Jeffreys, autor książki *Theory of Probability*, Oxford 1948.

[6]) Zdanie Wienera zacytowane w [4]) na str. 104 z jego książki *Cybernetics*, New York 1948, ze str. 109-110.

[7]) J. Oderfeld, *On the dual aspect of sampling plans*, Colloquium Mathematicum 2 (1951), str. 89-97.

[8]) H. Steinhaus, *Podstawy kontroli statystycznej*, Zastosowania Matematyki 1 (1953), str. 4-27.

odpowiednikiem. Nie troszcząc się o wierność tłumaczenia ani o zgodność z przyjętą terminologią, będziemy musieli określić sami koncepcję wiarogodności, która to wielkość zastępuje w teorii Fishera prawdopodobieństwo przyczyn i stanowi fundamentalne pojęcie nowej doktryny. Wyjaśnimy je na przykładzie.

Niech pewna typowa reakcja zwierzęcia doświadczalnego zależy od nieznanej zawartości x pewnej substancji w 1 cm^3 krwi pobranym z badanego człowieka i zastrzykniętym zwierzęciu i niech prawdopodobieństwo, że reakcja zajdzie, rośnie wraz z x. Jeżeli na n zwierząt doświadczalnych m wykazało reakcję (a $n-m$ nie), to *ex definitione* nazwiemy *wiarogodnością* hipotezy, iż u badanego pacjenta jest $x<a$, prawdopodobieństwo tego, że zawartość a tej substancji u pacjenta da w n próbach *więcej* niż m razy ową reakcję. Oto treść wyłuskana z różnych tekstów panującej szkoły — nie łatwo ją znaleźć.

3. Przedyskutujemy najpierw pojęcie wiarogodności na innym przykładzie, w którym obserwacja może dawać wszelkie wartości rzeczywiste (a nie tylko skończenie wiele, jak w doświadczeniu wyżej opisanym). Takim przykładem jest położenie punktu materialnego na prostej materialnej zaopatrzonej w skalę. Prawdziwe położenie punktu odpowiada cesze x skali, lecz x nie jest znane, obserwacja zaś daje rezultat ξ; $\xi-x$ jest błędem obserwacji. Znamy prawdopodobieństwa błędów, a mianowicie funkcję $p(x,\xi)$, która pozwala obliczyć prawdopodobieństwo, iż położenie x da obserwację ξ należącą do przedziału $\langle\xi_0, \xi_0+\Delta\xi\rangle$; to prawdopodobieństwo jest

$$p(x,\xi_0)\Delta\xi+o(\Delta\xi)\ ^{9}).$$

Moglibyśmy uwypuklić przykład przez specjalizację funkcji $p(x,\xi)$, przyjmując np.

$$p(x,\xi)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\xi)^2/2\sigma^2}, \tag{2}$$

co odpowiada normalnemu rozkładowi błędów. Ta specjalizacja jest zbędna; do naszych celów wystarczy założenie, że gęstość p jest nieujemną i ciągłą funkcją błędu $\xi-x$,

$$p(x,\xi)=f(\xi-x), \tag{3}$$

[9]) Symbol $o(x)$ oznacza wielkość, która podzielona przez x dąży wraz z x do zera.

i że

$$\int_{-\infty}^{\infty} p(x,\xi)\,d\xi = 1 \tag{4}$$

dla każdego x. Założenie (4) jest matematycznym odpowiednikiem pewności, że dla każdej wielkości x otrzymamy jakąś obserwację ξ, co jest oczywiste.

Zauważmy, że gęstość (2) odpowiadająca normalnemu rozkładowi błędów (z błędem średnim niezależnym od mierzonego x) spełnia powyższe założenia; spełnia je także gęstość odpowiadająca obserwacji $\bar{\xi}$ będącej średnią z k niezależnych obserwacyj ξ_i $(i=1,2,\ldots,k)$, jeżeli każdej z nich odpowiada jedno i to samo p spełniające założenia.

Typowe zagadnienie brzmi teraz: Jeżeli obserwacja iksa dała $\xi=b$, jakie jest prawdopodobieństwo, że $x<a$? To prawdopodobieństwo oznaczymy przez $P(x<a;\xi=b)$. Aby tu zastosować regułę Bayesa (1), trzeba znać aprioryczne prawdopodobieństwo $F(a)$, że jest $x<a$, czyli prawdopodobieństwo tej nierówności sprzed obserwacji. $F(a)$ można nazwać *apriorycznym rozkładem* zmiennej losowej x.

Przeciwnicy reguły Bayesa mówią, że

1° nie znamy $F(a)$, a nie wolno za F wziąć jakiejś dowolnej funkcji, np. rozkładu równomiernego, który zresztą nie istnieje, gdy zmienna losowa x nie jest ograniczona;

2° postulat Bayesa rozkładu równomiernego prowadzi do sprzeczności;

3° ponieważ x nie jest zmienną losową, więc wzór (1) nie ma odpowiednika w prawie wielkich liczb, który by nadawał mu sens statystyczny;

4° reguły Bayesa można używać tylko wtedy, gdy się zna prawdopodobieństwa warunkowe $P_A(Z), P_B(Z), \ldots$, co wymaga — według klasycznej definicji prawdopodobieństwa warunkowego — znajomości prawdopodobieństw $P(AZ)$, $P(BZ)$ itd., a więc znajomości łącznego rozkładu pary zmiennych losowych (X,Y), z których X przebiega wszystkie przyczyny, takie jak $A,B,C\ldots$ itd., Y zaś wszystkie skutki (takie jak Z,Z',Z'' *etc*); wtedy jednak wzór (1) jest zbędny, bo szukana lewa strona $P_Z(A)$ jest po prostu ilorazem $P(AZ)/\sum P(XZ)$;

5° wzór Bayesa jest błędny, bo gdy zaszedł skutek Z, to albo zaszła była przyczyna A, albo nie, więc jej prawdopodobieństwo *a posteriori* $P_Z(A)$ może być tylko albo 1, albo 0, wbrew wzorowi (1);

6° reguła Bayesa jest zbędna, bo są inne metody uniwersalne i wolne od zarzutów.

Inna metoda, o której mowa w 6°, polega na koncepcji wiarogodności. Nie daje ona wcale odpowiedzi na pytanie, ile wynosi $P(x<a;\xi=b)$, lecz określa wiarogodność $W(x<a;\xi=b)$ faktu, że jest $x<a$, gdy $\xi=b$, przez relację

$$W(x<a;\xi=b)=P(\xi>b;x=a). \tag{5}$$

Obliczanie prawej strony (5) nie wymaga znajomości funkcji $F(a)$, gdyż

$$P(\xi>b;x=a)=\int_b^\infty p(a,t)\,dt, \tag{6}$$

co wynika bezpośrednio z definicji funkcji p. Nowa metoda określa więc jednoznacznie pewien parametr statystyczny W jako funkcję a i b, jak wynika z (5) i (6),

$$W(x<a;\xi=b)=\int_b^\infty p(a,t)\,dt, \tag{7}$$

który to parametr nazwaliśmy wiarogodnością faktu, że $x<a$, gdy $\xi=b$. Nie jest on wcale prawdopodobieństwem tego faktu, lecz prawdopodobieństwem innego faktu (który nie zaszedł) przy innym warunku (o którym nie twierdzi się, że jest spełniony). Jest to więc parametr konwencjonalny, mierzący stopień zaufania, na które zasługuje hipoteza, iż jest $x<a$, gdy zaobserwowano $\xi=b$. Ta konwencja może tylko wtedy pretendować do użyteczności, gdy funkcja $f(a,b)=P(\xi>b\,;\,x=a)$ dla każdego b jest niemalejącą funkcją zmiennej a — chcemy bowiem, żeby wiarogodność konsekwencji była co najmniej taka jak wiarogodność racji; ponieważ dla $a_1<a_2$ nierówność $x<a_1$ implikuje $x<a_2$, chcemy, żeby było $W(x<a_2;\xi=b)\geqslant W(x<a_1;\xi=b)$, a to i (5) zmusza do przyjęcia wyżej podanego warunku dotyczącego $f(a,b)$; ten warunek wcale nie wynika automatycznie z tego, że P jest prawdopodobieństwem.

Obliczmy teraz $P(x<a;\xi=b)$ według klasycznej reguły Bayesa przy postulacie $\mathcal{B}$, to jest przy równomiernym rozkładzie apriorycznym zmiennej losowej x. Wobec nieskończoności przedziału użyjemy aproksymacji; założymy wpierw równomierność w prze-

dziale $|x| \leqslant T$ a potem we wzorze Bayesa przejdziemy z T do ∞. Niech zatem prawdopodobieństwo aprioryczne, że x leży w przedziale $\langle x_0, x_0+dx \rangle$ będzie $g(x)dx$ i niech

$$g(x) = \begin{cases} 1/2T & \text{dla} \quad |x| \leqslant T; \\ 0 & \text{dla} \quad |x| > T. \end{cases} \tag{8}$$

Reguła Bayesa daje

$$P(x<a\,;\xi=b) = \frac{\int\limits_{-\infty}^{a} \cdot g(x)p(x,b)db \cdot dx}{\int\limits_{-\infty}^{\infty} \cdot g(x)p(x,b)db \cdot dx} = \frac{\int\limits_{-\infty}^{a} g(x)p(x,b)dx}{\int\limits_{-\infty}^{\infty} g(x)p(x,b)dx}. \tag{9}$$

Granicą ułamka (9) dla $T \to \infty$ jest — wobec (8) —

$$\frac{\int\limits_{-\infty}^{a} p(t,b)dt}{\int\limits_{-\infty}^{\infty} p(t,b)dt},$$

a zatem — wobec (4) — wyrażenie

$$\int\limits_{-\infty}^{a} p(t,b)dt.$$

Oznaczmy przez P_T lewą stronę wzoru (9) a przez $P_{\mathcal{B}}$ graniczną jej wartość dla $T \to \infty$, a będzie, w myśl poprzedniego zdania,

$$P_{\mathcal{B}}(x<a\,;\xi=b) = \int\limits_{-\infty}^{a} p(t,b)dt,$$

z czego przez podstawienie $u=a+b-t$ wyniknie dzięki (3)

$$P_{\mathcal{B}}(x<a\,;\xi=b) = \int\limits_{b}^{\infty} p(a,u)du. \tag{10}$$

Porównując relacje (7) i (10) otrzymujemy

$$W(x<a\,;\xi=b) = P_{\mathcal{B}}(x<a\,;\xi=b). \tag{11}$$

Nazwijmy *możliwością* prawdopodobieństwo $P_{\mathcal{B}}$ po prawej stronie (11); ten nowy termin ma przypominać, że nie jest to prawdopodobieństwo obliczone bez hipotez, lecz przy postulacie $\mathcal{B}$. Zamiast symbolu $P_{\mathcal{B}}$ użyjmy litery M, a będzie

(12) $$M(x<a;\xi=b)=W(x<a;\xi=b).$$

A więc w przykładzie tu traktowanym *możliwość* tego, że $x<a$, gdy $\xi=b$, jest równa *wiarogodności* tego, że $x<a$, gdy $\xi=b$.

Określiliśmy nowe pojęcie, możliwość $M(x<a;\xi=b)$, przez

(13) $$M(x<a;\xi=b)\overset{\text{df}}{=}P_{\mathcal{B}}(x<a;\xi=b).$$

W naszej teorii jest to odpowiednik wiarogodności, którą określono relacją (5). „Możliwość" jest tylko terminem, ale wprowadzenie go stawia na równi obie teorie, klasyczną teorię Bayesa i nową jego przeciwników. Obecnie sytuacja jest taka: Rezygnujemy z obliczenia $P(x<a;\xi=b)$, gdyż mają rację przeciwnicy klasycznej teorii, gdy mówią, że bez znajomości rozkładu apriorycznego to prawdopodobieństwo P jest nieobliczalne, a my nie chcemy uciekać się do arbitralnego postulatu $\mathcal{B}$. Ale termin „możliwość" wyprowadza nas z dylematu, gdyż wolno nam nazwać P (obliczone tak, jak gdyby był spełniony postulat $\mathcal{B}$) możliwością $M(x<a;\xi=b)$ tego, że jest $x<a$, gdy $\xi=b$; nie wolno nam tylko twierdzić, że jest to szukane prawdopodobieństwo tego, że $x<a$, gdy $\xi=b$. Jak postępują przeciwnicy reguły Bayesa i twórcy „fiducial probability"? Obliczają inne prawdopodobieństwo niż to, o które pyta ich praktyk, i nazywają je wiarogodnością, $W(x<a;\xi=b)$, tego, że $x<a$, gdy $\xi=b$. Otóż relacja (12), $M\equiv W$, rozbija mit o uwolnieniu rachunku prawdopodobieństwa od trudności przez wprowadzenie pojęcia wiarogodności i rehabilituje regułę i postulat Bayesa, gdyż pokazuje (choć na razie tylko na przykładzie), że wiarogodność jest po prostu prawdopodobieństwem obliczonym przy użyciu tej reguły i postulatu, z pozbyciem się skrupułów przez nadanie tej wielkości specjalnej nazwy. Tak upada zarzut 1° w zagadnieniu z § 3.

Powstaje pytanie, dlaczego wprowadzamy nową nazwę możliwości, skoro relacja $M\equiv W$ dowodzi, że można by poprzestać na „wiarogodności". Czynimy to dlatego, że choć w przykładzie tu rozważanym zachodzi owa relacja, to można podać przykłady, w których ona nie zachodzi — będą to takie zagadnienia, w których prawdopodobieństwa warunkowe nie spełniają założeń (3) i (4). Wtedy jednak można trzymać się równie dobrze metody możliwości, jak metody wiarogodności: paralelizm obu metod jest oczywisty, a w pewnej obszernej klasie zagadnień okazaliśmy ich identyczność. W tej klasie metoda wiarogodności zachowa swoją użyteczność w przypadkach, w których rachunek za pomocą wzoru (5) jest łatwiejszy niż za pomocą wzoru Bayesa, co zwykle zachodzi.

Tak jak określiliśmy możliwość, że jest $x<a$, gdy $\xi=b$, możemy — przez analogię — określić możliwość, że jest $a_1 \leqslant x < a_2$, gdy $\xi=b$, a mianowicie

$$(14) \qquad M(a_1 \leqslant x < a_2; \xi=b) \overset{\mathrm{df}}{=} P_{\mathcal{B}}(a_1 \leqslant x < a_2; \xi=b) \qquad (a_1<a_2);$$

tu znakowanie łatwo się samo tłumaczy: po lewej figuruje możliwość podwójnej nierówności pod warunkiem $\xi=b$, po prawej prawdopodobieństwo tejże nierówności pod tymże warunkiem, obliczone z reguły i postulatu $\mathcal{B}$. Klasyczny rachunek prawdopodobieństwa daje bezpośrednio

$$(15) \qquad P_{\mathcal{B}}(a_1 \leqslant x < a_2; \xi=b) = P(x<a_2; \xi=b) - P_{\mathcal{B}}(x<a_1; \xi=b),$$

a z (13), (14) i (15) wynika

$$(16) \qquad M(a_1 \leqslant x < a_2; \xi=b) = M(x<a_2; \xi=b) - M(x<a_1; \xi=b).$$

Własność (16) można wziąć za *definicję* lewej strony zamiast (14). To nasuwa analogiczną definicję w teorii wiarogodności:

$$(17) \qquad W(a_1 \leqslant x < a_2; \xi=b) \overset{\mathrm{df}}{=} W(x<a_2; \xi=b) - W(x<a_1; \xi=b)$$

dla $a_1 \leqslant a_2$. To pożyteczne uogólnienie pojęcia wiarogodności pozwala określać np. wiarogodność faktu, że przy pewnym wyniku próbki jakość partii towaru leży w granicach (a_1, a_2); bez stwierdzenia, że wiarogodność jest prawdopodobieństwem (chociaż innego faktu), to uogólnienie nie byłoby się nam nasunęło.

Przejdziemy teraz do innego przykładu, obejmującego olbrzymi zakres badań przyrodniczych i technicznych.

4. Niech wielkość x nie podlega bezpośredniemu pomiarowi, a obserwacje niech polegają na niezależnych próbach; przy każdej próbie niech x będzie prawdopodobieństwem, że zdarzy się zjawisko Z. Tak np. x może być charakterystyką leku, a Z testem tego leku. Zauważmy, że w tym przykładzie termin „prawdopodobieństwo" występuje w jeszcze jednej roli: wielkość mierzona sama jest prawdopodobieństwem. Rezultat obserwacji zapisujemy w formie $(\xi=m)/n$;

ma to znaczyć, że było n niezależnych prób, wśród których m razy zaszło Z (a $n-m$ razy nie); liczby całkowite m, n spełniają nierówności $0 \leqslant m \leqslant n$.

Pytanie brzmi: Jak obliczyć $P(x<a; (\xi=m)/n)$ czyli prawdopodobieństwo, że x jest mniejsze od a, gdy na n prób m razy zaobserwowano Z?

Na to pytanie daje teoria wiarogodności wymijającą odpowiedź, obliczając zwykłe prawdopodobieństwo P innego faktu przy innym warunku, i nazywając je wiarogodnością W pierwotnego faktu przy pierwotnym warunku, podobnie jak to wyjaśniliśmy na pierwszym przykładzie. Jest więc

$$W\left(x<a; \frac{\xi=m}{n}\right) \overset{\mathrm{df}}{=} P\left(\frac{\xi>m}{n}; x=a\right); \tag{18}$$

po prawej stronie P jest prawdopodobieństwem, że gdy $x=a$, zajdzie Z więcej niż m razy w n próbach. Definicji (18) można zarzucić, że dla $m=n$ i dla każdego a jest $W=0$. W szczególności jest wtedy $W=0$ dla $a=1$. Gdy więc np. na 5 prób leku wszystkie będą pomyślne, wiarogodność hipotezy $x<1$, czyli hipotezy, że lek nie jest pewny, będzie zero. To orzeczenie będzie jeszcze dziwniejsze przy wykonaniu jednej tylko próby, która wypadnie pomyślnie. Można by uniknąć tego paradoksu pisząc w definiensie po prawej stronie (18) zamiast $(\xi>m)/n$ nierówność $(\xi \geqslant m)/n$, ale wtedy pojawi się inny: zmieniona definicja da $W=1$ dla $m=0$ przy każdym a, nawet bardzo małym, gdyż dla $m=0$ poprawiona nierówność zachodzi na pewno dla $n=1$; dla $n=100$ otrzymamy stąd, że wiarogodność hipotezy, iż lek próbowany raz bez skutku nic nie jest wart, jest 1, taka sama jak wiarogodność tejże hipotezy co do leku 100 razy próbowanego bez skutku!

Łatwym wyjściem z tej niedogodności jest zmiana mianownika w definiensie (18) z n na $n+1$:

$$W\left(x<a; \frac{\xi=m}{n}\right) \overset{\mathrm{df}}{=} P\left(\frac{\xi>m}{n+1}; x=a\right) \quad (m=0,1,\ldots,n; 0\leqslant a\leqslant 1). \tag{19}$$

Tu już nie ma paradoksu ani dla $m=0$, ani dla $m=n$. W przyjmuje teraz wartość 0 tylko i zawsze dla $a=0$, co jest zgodne z naturalnym postulatem zerowej wiarogodności niemożliwej relacji $x<a$, gdy $a=0$; W jest równe 1 tylko i zawsze dla $a=1$, co wyraża niemniej naturalny postulat jedynkowej (a więc maksymalnej) wiarogodności

relacji $x<a$, gdy $a=1$: jakikolwiek byłby wynik skończonej liczby prób, możemy z maksymalną wiarogodnością twierdzić, że lek nie jest absolutnie skuteczny. Dla $0<a<1$ jest zawsze $0<W<1$.

Zauważmy mimochodem, że nawet praktyczni lekarze spostrzegli trudności związane ze skrajnymi eksperymentami $m=0$, $m=n$, które uniemożliwiają naiwne przyjęcie siły leku za równą m/n.

Określmy teraz *możliwość* M, że jest $x<a$, gdy Z zaszło m razy w n próbach; postąpimy analogicznie do (13), gdy napiszemy, że

$$(20) \qquad M\left(x<a;\frac{\xi=m}{n}\right)\overset{\mathrm{df}}{=}P\left(x<a;\frac{\xi=m}{n}\right).$$

Chcąc otrzymać relację $M=W$ analogiczną do (12) trzeba udowodnić równość prawych stron w (19) i (20). Jest — według reguły Bayesa —

$$P\left(x<a;\frac{\xi=m}{n}\right)=\int_0^a x^m(1-x)^{n-m}dx\Big/\int_0^1 x^m(1-x)^{n-m}dx,$$

$$P\left(\frac{\xi>m}{n+1};x=a\right)=\sum_{k=m+1}^{n+1}P\left(\frac{\xi=k}{n+1};x=a\right)=\sum_{k=m+1}^{n+1}\binom{n+1}{k}a^k(1-a)^{n+1-k}=$$

$$=\int_0^a x^m(1-x)^{n-m}dx\Big/\int_0^1 x^m(1-x)^{n-m}dx\,^{10}),$$

co właśnie daje żądaną równość i pozwala napisać twierdzenie:

$$(21) \qquad M\left(x<a;\frac{\xi=m}{n}\right)=W\left(x<a;\frac{\xi=m}{n}\right).$$

Zasadnicza relacja (21), podobnie jak w przypadku ciągłym relacja (12), ratuje regułę Bayesa, pokazując, że ta reguła prowadzi do tej samej liczby, co wiarogodność [10]).

5. Pozostaje jednak poważny zarzut 2°, który odnosi się zarówno do przypadku ciągłego jak do ostatnich wywodów. Ten zarzut orzeka, że $P_{\mathfrak{B}}$, a więc i M, nie spełnia warunku inwariancji. Mówiąc wyraźniej: Gdy obserwujemy x i otrzymujemy ξ, to inny obserwator, którego interesuje np. x^3, odczytuje, równocześnie z naszym odczytaniem ξ, liczbę na innej skali związanej ze skalą ξ tak, jak dwie skale obok siebie na tej samej linijce. Wprowadźmy oznaczenia: $X=x^3$, $\Xi=\xi^3$, $A=a^3$, $B=b^3$. Pytanie drugiego obserwatora brzmi:

$$P(X<A;\Xi=B)=?$$

[10]) Por. [7]). Niektóre wzory tej pracy skontrolował B. Kowalczyk.

Otóż zdania $X<A$ i $x<a$ są ze sobą równoważne, a tak samo są ze sobą równoważne zdania $\Xi=B$ i $\xi=b$. Wobec tego nierówność $X<A$ przy warunku $\Xi=B$ zdarza się wtedy i tylko wtedy, gdy zdarza się nierówność $x<a$ przy warunku $\xi=b$, i możliwości obu zdarzeń warunkowych powinny być równe:

$$M(X<A;\Xi=B)=M(x<a;\xi=b); \tag{22}$$

gdy obydwaj obserwatorzy obliczą możliwości relacyj, które ich interesują, powinni otrzymać to samo; definicja (13) i to żądanie daje

$$P_{\mathfrak{B}}(X<A;\Xi=B)=P_{\mathfrak{B}}(x<a;\xi=b). \tag{23}$$

Postulat $\mathfrak{B}$ dla X oznacza jednak co innego niż postulat $\mathfrak{B}$ dla x, gdyż równomierny rozkład zmiennej x jest sprzeczny z równomiernym rozkładem x^3. Przy przejściu od zmiennych x,ξ do X,Ξ zmienią się także prawdopodobieństwa warunkowe tak, że relacja (23) może się utrzymać lub nie. Łatwo jednak wykazać, że główną rolę grają tu prawdopodobieństwa warunkowe. Jeżeli mianowicie oznaczymy przez $P(X,\Xi)$ funkcję analogiczną do tej, którą oznaczyliśmy przez $p(x,\xi)$, a którą można obliczyć z wyrażenia $p(x,\xi)d\xi$, przez podstawienie $x=\gamma(X)$, $\xi=\gamma(\Xi)$ $\left(\text{tutaj } \gamma(X)=\sqrt[3]{X}\right)$, to wystarczy sprawdzić, czy $P(X,\Xi)$ ma kształt $P=F(X-\Xi)$. Jeżeli tak, to zachodzi relacja (23), a więc także (22). Rzeczywiście, mamy wtedy dzięki (11) i (5)

$$\begin{aligned} P_{\mathfrak{B}}(X<A;\Xi=B)&=P(\Xi>B;X=A),\\ P_{\mathfrak{B}}(x<a;\xi=b)&=P(\xi>b;x=a), \end{aligned} \tag{24}$$

gdzie prawe strony są równe, bo są prawdopodobieństwami relacyj równoważnych; stąd już wynika (23). Wniosek z tego następujący:

Obliczanie możliwości według naszej definicji (13) prowadzi zawsze do tych samych wyników, gdy używamy takich zmiennych jako przedmiotu obserwacji, których prawo błędu spełnia relację (23). Ponieważ to prawo nie jest hipotetyczne, lecz da się z dowolną precyzją określić za pomocą dostatecznie licznych obserwacyj, a także da się obliczyć matematycznie dla zmiennej przetransformowanej, gdy jest znane dla pierwotnej, więc nigdy nie ma wątpliwości, czy używamy właściwej zmiennej. Może się jednak zdarzyć, że ani sama zmienna, ani jej transformata nie spełnia relacji (23); wtedy pojęcie możliwości traci swój właściwy sens.

Tutaj zwolennicy wiarogodności mogą podnieść argument, że wiarogodność można obliczyć zawsze bez badania charakteru zmiennej obserwowanej i że transformacja zmiennych jej nie zmienia. Na to odpowiemy, że w tych przypadkach, w których nie można użyć pojęcia możliwości, zdarzają się takie funkcje błędu, do których można dobrać odpowiedni rozkład aprioryczny tak, by reguła Bayesa wraz z tym rozkładem dawała prawdopodobieństwo *a posteriori* równe wiarogodności. Tak więc poleganie na absolutnej inwariancji wiarogodności (która to inwariancja jest niezaprzeczona) jest nieświadomym narzucaniem zmiennej obserwowanej za każdym razem innego rozkładu apriorycznego, nieraz zupełnie fantastycznego. Jeżeli postulat Bayesa został potępiony za to, że w każdym przypadku narzuca równomierny rozkład, nieraz niezgodny z naturą zagadnienia, to nieświadome używanie reguły Bayesa z różnymi rozkładami apriorycznymi, i to bez ich podawania *explicite* i porównywania z naturą, jest chyba jeszcze ryzykowniejsze.

W przypadku ciągłym pouczający jest przykład zmiennej x, która *a prori* jest ograniczona do przedziału skończonego (c, d), a w tym przedziale ma aprioryczny rozkład równomierny. Funkcja $p(x, \xi)$ niech będzie dana przez (2). Jeżeli znamy rozkład *a priori*, to możemy użyć klasycznej reguły Bayesa i obliczyć np. prawdopodobieństwo, że gdy pomiar dał $\xi = c$, x jest mniejsze od c; to prawdopodobieństwo aposterioryczne jest zero i reguła Bayesa da je bez żadnych hipotez. Jeżeli jednak nie znamy rozkładu apriorycznego (który naprawdę jest taki, jak wyżej ustalono), to musimy się uciec do pojęcia wiarogodności lub do pojęcia możliwości, żeby rozwiązać to samo zadanie. Jak wykazaliśmy w § 3, wiarogodność i możliwość są sobie równe w tym przypadku. Wystarczy zatem obliczyć wiarogodność $W(x<c; \xi=c)$; możemy użyć do tego wzoru (7), który da

$$W(x<c; \xi=c) = \int_c^\infty p(c, t)\, dt = \frac{1}{\sigma\sqrt{2\pi}} \int_c^\infty e^{-(c-t)^2/2\sigma^2} dt = 0{,}5.$$

Obydwie metody dają 50%, co wobec faktycznej niemożliwości relacji $x<c$ jest fałszywą informacją. Różnica jest tylko ta, że zwolennik wiarogodności dał ową informację śmielej niż zwolennik możliwości, bo ten drugi musiał przedtem sprawdzić, czy funkcja (2) ma postać (3); w naszym przykładzie tak jest.

Zarzut 2⁰ jest znacznie łatwiejszy do odparcia w przypadku nieciągłym. Tutaj bowiem zjawisko Z lub jakiekolwiek zjawisko z nim równoważne jest bezpośrednio dane, x zaś jest prawdopodobieństwem, że Z zajdzie. Przez to jest jednoznacznie określona zmienna x, której szukamy na podstawie eksperymentów, których wynik może być jedynie „Z zaszło”, „Z nie zaszło”. Żadna zmienna X wyrażająca się przez x (np. $X=x^3$) (z wyjątkiem $X\equiv x$) nie jest prawdopodobieństwem i dlatego x jest wyróżnione i wskazane przez samo zagadnienie. Tym samym odpada zarzut 5⁰ w olbrzymiej klasie wyrywkowych badań jakości względem cechy alternatywnej.

6. W tej klasie możemy — podobnie jak w przykładzie wyniku ciągłego — określić możliwość relacji dwustronnej $a_1\leqslant x<a_2$. Otrzymujemy wzory analogiczne do (14)-(17), zastępując w nich wszędzie $\xi=b$ przez $(\xi=m)/n$:

$$M\left(a_1\leqslant x<a_2;\frac{\xi=m}{n}\right)\overset{\mathrm{df}}{=}P\left(a_1\leqslant x<a_2;\frac{\xi=m}{n}\right),\tag{25}$$

$$M\left(a_1\leqslant x<a_2;\frac{\xi=m}{n}\right)=M\left(x<a_2;\frac{\xi=m}{n}\right)-M\left(x<a_1;\frac{\xi=m}{n}\right),\tag{26}$$

$$W\left(a_1\leqslant x<a_2;\frac{\xi=m}{n}\right)\overset{\mathrm{df}}{=}W\left(x<a_2;\frac{\xi=m}{n}\right)-W\left(x<a_1;\frac{\xi=m}{n}\right),\tag{27}$$

$$W\left(a_1\leqslant x<a_2;\frac{\xi=m}{n}\right)=M\left(a_1\leqslant x<a_2;\frac{\xi=m}{n}\right).\tag{28}$$

Z definicji (25) i (20) wynika twierdzenie (26), a stąd i z definicji (27) — twierdzenie (28).

Te wzory ułatwiaja zrozumienie zagadnień statystycznej kontroli jakości. Można np. określić sekwencyjny plan odbioru, gdy dane są dwie jakości graniczne a_1, a_2 $(a_1<a_2)$ przez przepis nakazujący przyjąć partię, gdy pojawi się taki rezultat $(\xi=m)/n$, że

$$M\left(a_1\leqslant x<1;\frac{\xi=m}{n}\right)\geqslant 95\%,\tag{I}$$

a odrzucić ją, gdy się pojawi taki rezultat $\dfrac{\xi=m}{n}$, że

$$M\left(0\leqslant x<a_2;\frac{\xi=m}{n}\right)\geqslant 95\%,\tag{II}$$

badając dopóty, dopóki nie spełni się (I) lub (II). Taki plan różni się od analizy sekwencyjnej A. Walda tym, że daje zawsze rozstrzygnięcie w skończonej liczbie kroków, gdy $a_1 < a_2$. Wzór (28) pozwala określić nasz plan sekwencyjny także w terminach teorii wiarogodności, ale powstał on z myślenia retrospektywnego, a mianowicie z przepisu badania aż do chwili, gdy wynik próbki zredukuje prawdopodobieństwo *a posteriori*, iż partia przyjęta (odrzucona), jest gorsza (lepsza) niż a_1 (a_2), poniżej 5%.

7. Zarzut 3°, że x nie jest zmienną losową, należy rozumieć w ten sposób, że wartości x nie tworzą kolektywu, w którym można by określić rozkład. Wystarczy jednak, by każde orzeczenie dotyczące prawdopodobieństwa dopuszczało weryfikację za pomocą — choćby tylko pomyślanego — eksperymentu i policzenia relatywnej frekwencji udałych prób. W naszych przykładach jest jednak taka weryfikacja właściwie zbędna, bo — jak wykazaliśmy — jest w przykładzie ciągłym wobec (5) i (11)

$$(29) \qquad P_{\mathfrak{B}}(x < a; \xi = b) = P(\xi > b; x = a),$$

a więc wystarczy dla porównania wielkości $P_{\mathfrak{B}}$ z doświadczeniem porównać prawą stronę wzoru (29), czyli obliczyć frekwencję wyników $\xi > b$ dla ustalonego $x = a$; będzie to po prostu weryfikacja przez zwykłe prawo wielkich liczb, które nie podlega krytyce. Jest jednak możliwa interpretacja frekwencyjna bezpośrednia lewej strony wzoru (29), zwłaszcza w przykładzie z § 4:

Nadaje się zmiennej x kolejno takie wartości $x_1, x_2, \ldots, x_k, \ldots$, żeby ciąg $\{x_n\}$ miał ekwipartycję w przedziale $(0,1)$, to znaczy, żeby relatywna frekwencja wyrazów ciągu mniejszych od t była równa t dla każdego t z przedziału $(0,1)$. Dla każdego x_k wyznacza się rezultat $(\xi = m_k)/n$, który wyraża się liczbą m_k pojawień się zjawiska Z w n niezależnych kolejnych próbach (n ustalone raz na zawsze), z których każda ma prawdopodobieństwo sukcesu równe x_k. Z ciągu $\{m_k\}$ wyjmuje się ciąg częściowy $\{m_{k_j}\}$ określony przez warunek $m_{k_j} = m$. Dla skrócenia nazwijmy x'_j liczbę x_{k_j} i obliczmy w ciągu $\{x'_j\}$ relatywną frekwencję wyrazów, dla których $x'_j < a$. Ta frekwencja powinna być równa $P_{\mathfrak{B}}(x < a; (\xi = m)/n)$. Jest to specjalna forma prawa wielkich liczb, którego to sformułowania nie znajdujemy zwykle w podręcznikach rachunku prawdopodobieństwa; jedną z przyczyn — może nieuświadomionych — awersji do reguły

Bayesa jest zapewne przeświadczenie, że bądź to sformułowanie jest niemożliwe, bądź też, że jego dowód wykracza poza aksjomatykę rachunku prawdopodobieństwa. Że tak nie jest, najłatwiej sprawdzić na elementarnym wzorze (1). Wzór ten można mianowicie interpretować poza rachunkiem prawdopodobieństwa jako twierdzenie o frekwencjach relatywnych w ciągu par (,), w których na pierwszym miejscu stoi $A, B, \ldots$ lub N, na drugim zaś Z lub non-Z. W tej interpretacji należy czytać np. $P(A)$ jako frekwencję relatywną w całym ciągu tych par, które mają A na pierwszym miejscu, $P_Z(A)$ zaś jako frekwencję relatywną tych samych par wśród wszystkich mających Z na drugim miejscu etc. Według zwykłego prawa wielkich liczb można wszędzie po prawej zastąpić frekwencje przez prawdopodobieństwa; frekwencja ta jest równa prawdopodobieństwu *a posteriori* $P_Z(A)$ obliczonemu z wzoru Bayesa (1), co należało wykazać.

Ta weryfikacja operuje fikcją ciągu $\{x_n\}$ obdarzonego ekwipartycją. Jest ona zatem odparciem zarzutu, jakoby orzeczenia o prawdopodobieństwach *a posteriori* były statystycznie niesprawdzalne nawet pomyślanym eksperymentem, ale naraża na dwa nowe zarzuty. Pierwszy, to wprowadzenie w innej formie postulatu $\mathcal{B}$, który ukrywa się pod założeniem ekwipartycji, drugi, to konieczność liczenia frekwencji pewnego podciągu w innym podciągu prób, co czyni wszelki realny eksperyment niemal iluzorycznym. Dlatego poddamy teraz nasze pojęcie możliwości innej weryfikacji i na innym przykładzie, a mianowicie na przykładzie z § 3. Tam możliwość określono przez (13). Niech w j-tym pomiarze ($j=1, 2, 3, \ldots$) X_j będzie prawdziwą wartością wielkości mierzonej x_j, tak że dopuszczamy zupełną różnorodność obserwacyj, a nawet różność funkcyj p_j grających rolę gęstości p (określonej w § 3) w kolejnych pomiarach, byle utrzymany był postulat wzajemnej niezależności pomiarów. Niech ξ_j oznacza ogólnie wynik j-tej obserwacji. Przypuśćmy, że wyniki w konkretnej serii były $\xi_j = b_j$. Zadajmy sobie z góry liczbę P ($0 \leqslant P \leqslant 1$; np. $P=0{,}95$). Za każdym razem obliczamy a_j z warunku

$$M(x_j < a_j; \xi_j = b_j) = P \qquad \text{(np. } M=0{,}95\text{).} \tag{30}$$

To obliczenie nie wymaga oczywiście znajomości X_j. Relacja $X_j \leqslant a_j$ jest jednak obiektywnie prawdziwa lub fałszywa.

Teza. *Frekwencja tych pomiarów, w których $X_j \leqslant a_j$, jest — względem ciągu wszystkich pomiarów — równa P (w przykładzie konkretnym 95%) z prawdopodobieństwem jeden.*

Zanim je udowodnimy, zauważmy, że twierdzenie powyższe uzasadnia pojęcie możliwości z praktycznego punktu widzenia w sposób doskonały, gdyż w ogóle niczego nie zakłada o ciągu prawdziwych wartości X_j, ani nie żąda, żeby były one wylosowane z jakiegoś kolektywu. Pomimo to pozwala ono wypowiadać na podstawie pomiarów sądy o tych wartościach z określoną z góry frekwencją omyłek (tu ta frekwencja jest $1-P$; np. 5%).

Dowód. Wobec (5) i (11) relacja (30) jest równoważna z relacją

$$\mathcal{P}(\xi_j > b_j; x_j = a_j) = P; \tag{31}$$

używamy tu litery $\mathcal{P}$ na prawdopodobieństwa warunkowe oznaczane dotąd przez P. Wobec nieujemności funkcji p jest $\mathcal{P}$ niemalejącą funkcją a_j, a stąd i z (31) wynika, że dla $X_j \leqslant a_j$ (i tylko dla takich X_j) zachodzi

$$\mathcal{P}(\xi_j > b_j; x_j = X_j) \leqslant P. \tag{32}$$

Ale teza głosi, że frekwencja przypadków $X_j \leqslant a_j$ jest P (z § 1). Ponieważ w tych i tylko w tych przypadkach zachodzi (32), więc teza jest równoznaczna z tym, że (z prawdopodobieństwem 1) frekwencja przypadków $\mathcal{P} \leqslant P$ jest P. Innymi słowy, twierdzenie sprowadza się do tego, że *ciąg* $\mathcal{P}\{\xi_j > b_j; x_j = X_j\}$ *ma ekwipartycję.* Jest to lemat zupełnie ogólny, który wystarcza zrozumieć, żeby go natychmiast sprawdzić. Jego sens jest następujący: z pomiarów wielkości x_j (których prawdziwych wartości X_j nie znamy) otrzymujemy rezultaty b_j, a później, dowiedziawszy się jakie są X_j, obliczamy $\mathcal{P}_j = \mathcal{P}(\xi > b_j; x_j = X_j)$; $\mathcal{P}_j$ jest zatem prawdopodobieństwem, iż powtórzenie pomiaru wielkości x_j da wynik większy niż aktualny wynik b_j; prawdopodobieństwo to obliczono dzięki znajomości prawdziwej wartości X_j mierzonej wielkości x_j. Wobec tego $\{\mathcal{P}_j\}$ jest ciągiem zmiennych losowych. O tym ciągu twierdzimy, że ma ekwipartycję.

Dowód lematu. Lemat wyniknie z prawa wielkich liczb, gdy wykażemy, że każda zmienna losowa $\mathcal{P}_j$ z osobna ma rozkład równomierny. Opuszczając wskaźnik j mamy wykazać, że zmienna losowa $\mathcal{P}$ określona przez

$$\mathcal{P} = \mathcal{P}(\xi > b; x = X) \tag{33}$$

ma rozkład równomierny. Z (33) widać, że

$$\mathcal{P}=\int_{b}^{\infty}p(\xi,X)\,d\xi. \tag{34}$$

Niech u,v należą do przedziału $(0,1)$ i niech będzie $u\leqslant v$. Obliczmy liczby b_u,b_v z relacyj

$$\int_{b_u}^{\infty}p(\xi,X)\,d\xi=u,\qquad \int_{b_v}^{\infty}p(\xi,X)\,d\xi=v. \tag{35}$$

Wobec (34) i (35) jest oczywiste, że relacja $u\leqslant\mathcal{P}\leqslant v$ jest równoważna z relacją $b_u\geqslant b\geqslant b_v$, wobec czego prawdopodobieństwo pierwszej relacji jest równe prawdopodobieństwu drugiej. Ale to ostatnie prawdopodobieństwo $\Gamma(b_u,b_v)$ jest

$$\int_{b_v}^{b_u}p(\xi,X)\,d\xi,$$

a więc mamy według (35)

$$\Gamma(b_u,b_v)=v-u.$$

Z tego wynika, że prawdopodobieństwo $\Pi(u,v)$ relacji $u\leqslant\mathcal{P}\leqslant v$ też jest $v-u$, a to właśnie jest treścią lematu.

Chcemy tutaj podkreślić, że są zatem dwie weryfikacje reguły możliwości: pierwsza, od której odwraca się statystyk, bo wymaga w przypadku odbioru partyj sztucznego założenia ekwipartycji nadchodzących jakości prawdziwych, w przypadku zaś ciągłym w ogóle jest niemożliwa, i druga, która jest wolna od tych trudności i jest praktycznie wykonalna. Zwykle tę drugą łączy się z koncepcją wiarogodności. Trzeba jednak pamiętać o tym, że gdy eksperyment dał pewien rezultat R, z którego wnosimy, że zachodzi nierówność N z prawdopodobieństwem P, to pytanie praktyka o sens liczby P wolno nam rozumieć nie tylko w ten sposób, że ustalamy R i N (jak każe pierwsza weryfikacja) i badamy, czy rzeczywiście skutkowi R towarzyszy przyczyna N we frakcji P ciągu eksperymentów kolejnych, ale także i w ten sposób, że w każdym eksperymencie do rezultatu R_i dobieramy — według teoretycznego rachunku — takie N_i, żeby odpowiednie P_i było równe P i badamy, biorąc pod uwagę cały ciąg eksperymentów, czy daje on frakcję P przypadków, w których rzeczywiście R_i kojarzy się z obliczonym teoretycznie N_i. Ta właśnie druga weryfikacja interesuje przede wszystkim praktyka, bo daje mu środki do dowolnego zredukowania frakcji błędnych orzeczeń.

8. Wyłania się kwestia, czy można weryfikacji podobnej do powyższej poddać przykład z § 4. Odpowiedź jest twierdząca, ale dla małych n tezy analogiczne sprawdzają się tylko w przybliżeniu. Ta wada planów pojedynczych typu $m//n$ bierze się stąd, że wysnuwamy wnioski z dyskretnej zmiennej losowej, jaką jest m, na ciągłą zmienną x. Można przypuszczać, że gdyby nie zmieniając klasycznego wzoru na prawdopodobieństwo wyniku m/n dopuszczono na m wszystkie wartości od 0 do n (a więc i ułamkowe), owa niedogodność dałaby się skasować, jednak autor tego nie sprawdzał. Takie rozszerzenie pojęcia liczby dobrych sztuk (np. „jest 9,7 dobrych sztuk w próbce zbadanej, a złożonej z 10 sztuk”) miałoby także praktyczny sens, bo można — na przykład — uważać za dobry tylko taki drut, który wytrzymuje obciążenie 100 kg lub wyższe, a drut, który zrywa się przy 70 kg obciążenia uważać za 0,7 sztuki dobrej, a 0,3 wadliwej.

Drugie pytanie dotyczyłoby tego, czy można przykład z § 3 zweryfikować tak jak przykład z § 4, to jest używając ciągu $\{x_n\}$ obdarzonego ekwipartycją. Chcąc to uczynić natrafilibyśmy na dwie trudności. Pierwsza, mniej groźna, to brak definicji ciągu ekwipartycyjnego w przedziale nieskończonym (bo taki jest przedział na x). Groźniejsza jest trudność wynikająca stąd, że z ciągu pomiarów musielibyśmy wyjąć podciąg złożony z tych pomiarów, które dały wynik b. Ponieważ w każdym pomiarze prawdopodobieństwo uzyskania wyniku b jest zero, więc jest prawdopodobieństwo 1, że takich pomiarów w ogóle nie będzie. Być może, że ta druga trudność jest jedną z mniej lub więcej świadomie odczuwanych przeszkód, które odstręczają matematyków od reguły Bayesa. Jak widzieliśmy jednak, w przykładzie z § 3 można się powołać na znacznie skuteczniejszą weryfikację, wolną od zarzutów teoretycznych i praktycznych. Ale trudność, o której wspomnieliśmy przed chwilą, nie ogranicza się do weryfikacji statystycznej. Ma ona swój odpowiednik w interpretacji samego wzoru Bayesa, a mianowicie wariantu ciągłego owego wzoru; ten wariant znajdujemy w § 3 oznaczony przez (9); funkcję $g(x)$ należy teraz rozumieć jako dowolną funkcję nieujemną o całce 1 w przedziale $(-\infty,\infty)$; w tej chwili nie zakładamy wcale postulatu $\mathcal{B}$. Klasyczny rachunek prawdopodobieństwa każe obliczać prawdopodobieństwa warunkowe $p(x,b)$ jako ilorazy: $p(x,b)$ jest tam mianowicie ilorazem apriorycznego prawdopodobieństwa, że „x i b” przez aprioryczne prawdopodobieństwo, że „x”. Ponieważ te obydwa prawdopodobieństwa są zwykle zerami, więc $p(x,b)=0/0$. Pomijamy tu sposób wyjścia z trudności przez przejście graniczne. Ważniejsze jest to, że znajomość prawdopodobieństwa, że „x i b" (albo ściślej, że x leży między x_0 a x_0+dx, a zarazem b między b_0 a b_0+db) presumuje znajomość dystrybuanty łącznej pary zmiennych (x,b), a wtedy wzór Bayesa w ogóle jest zbędny. Jest to zarzut 4°. Ale w praktyce jest wiele przykładów na to, że prawdopodobieństwa warunkowe dane są bezpośrednio — tak jest właśnie w zagadnieniu pomiarów w § 3, a także zagadnienie dochodzenia ojcostwa można ująć w taki sposób (choć nie jest to konieczne). Te uwagi wystarczają do wyjaśnienia wątpliwości, które się nasuwają przy czytaniu poprzednich paragrafów.

9. Pozostaje jeszcze kwestia, jaki jest stosunek niniejszego artykułu do tego, co o regule i postulacie Bayesa mówią kompetentni współcześni znawcy rachunku prawdopodobieństwa i statystyki matematycznej. Niewątpliwie wiele z tego, o czym tu może po raz pierwszy dowiaduje się polski czytelnik nie będący specjalistą, można znaleźć w poważniejszych podręcznikach zagranicznych. Autor zebrał, na próbę, kilka takich podręczników. Niestety, nie pozwalają one na wyrobienie sobie sądu o tym, jakie jest stanowisko wiedzy współczesnej, a to nie tylko z powodu sprzecznych poglądów, ale także z powodu niejasności wywodów. Bierze się ona stąd, że większość autorów traktuje nasze zagadnienie na tle specjalnych zadań. Jedni zauważają, że przy licznych próbach rozkład *a priori* gra małą rolę, inni, że w typowych zadaniach z urną nie można mówić o takim rozkładzie. Ale i zbytnia ogólność jest nie na miejscu, gdy np. chce się wyznaczać *a posteriori* od razu dwa nieznane parametry; to utrudnia zrozumienie wywodu, który powinien przede wszystkim wyjaśnić zagadnienie zasadnicze.

J. V. Uspensky mniema, że brak rozkładu apriorycznego w zadaniach urnowych tylko na pierwszy rzut oka jest oczywisty. J. L. Coolidge na przykładzie Bertranda procesu o złą ruletkę stwierdza, że choć formuła Bayesa jest wadliwa, nie ma nic lepszego. Astronom angielski H. Jeffreys znalazł równomierny rozkład dla przedziału nieskończonego (por. § 3, wzory (8)-(12)) i relację (11), jednak dla specjalnego p, a mianowicie dla p z wzoru (2), a nie dla ogólnego, danego przez (3). Zajmuje się też inwariancją, o której traktuje u nas § 5, jednak nie zauważa kryterium (23). Twierdzi z naciskiem, że żadne prawdopodobieństwo aprioryczne, warunkowe, czy też aposterioryczne, nie jest po prostu frekwencją. Należy to chyba rozumieć jako rezygnację z weryfikacji statystycznej, którą niniejsza praca uważa za jedyne rozsądne kryterium poprawności orzeczeń. Że nie widzi możliwości takiej weryfikacji, wynika z tego, że wyraźnie oświadcza, iż nie rozumie, co mogłaby oznaczać laplace'owa „jednakowa możliwość" w sensie frekwencyjnym. I rzeczywiście, u żadnego z tych autorów nie znajdujemy naszego modelu ekwipartycyjnego. Już nawet taki szczegół historyczny jak kwestia zgodności lub niezgodności Jeffreysa z R. A. Fisherem budzi w przedstawieniu pierwszego z tych autorów wątpliwości czytelnika. Według niego Karol Pearson był jedynym człowiekiem, który wierzył w regułę Bayesa wraz z frekwencyjną definicją prawdopodobieństwa. Nawet taki poważny autor jak M. G. Kendall mówi, że assercja, którą my oznaczamy przez $(x<a;\xi=b)$ nie może mieć innego prawdopodobieństwa jak 0 lub 1, bo x jest mniejsze od a lub nie jest i *tertium non datur*. Autor niniejszego artykułu uważa ten zarzut (5^0 w spisie § 3) za niedorzeczny, bo odnosi się on do każdego niemal zadania z rachunku prawdopodobieństwa, gdy przyjąć determinizm zjawisk fizycznych (któremu nie przeczy żaden z autorów cytowanych): gdy zapytujemy, czy karta wyciągnięta z talii będzie asem pik, to — wobec tego, że już przed wyciągnięciem rzecz jest przesądzona — prawdopodobieństwo byłoby 1 lub 0. Jeszcze lepszym kontrprzykładem jest karta już wyciągnięta, ale zasłonięta. Każdy gracz stosuje tu rachunek prawdopodobieństwa, chociaż assercja „karta jest asem pik" nie różni się wcale od „$x<a$", bo obie dotyczą faktów już zaistniałych. U żadnego ze wspomnianych autorów nie ma naszej koncepcji możliwości. Jak wiadomo, pisma R. A. Fishera nie odznaczają się jasnością, tak że po dziś dzień różnice między jego teorią „fiducial probability" (którą nazwaliśmy wiarogodnością) a teorią Jerzego Neymana „przedziału ufności" nie dla wszystkich matematyków są zrozumiałe bez reszty.

Polemikę, którą wciąż na nowo wznieca kontrowersja związana z regułą Bayesa i teorią wnioskowania indukcyjnego, charakteryzują cytaty, jakie można znaleźć w tekście u autorów przed chwilą wspomnianych lub też w nagłówkach rozdziałów jako motta. „Hamlet" Szekspira, „Waleczny Kapitan" Kiplinga, „Poprzez zwierciadło" L. Carrolla (autora „Alicji w krainie czarów"), „Bajki z tysiąca i jednej nocy" i inne dzieła z teorii nieprawdopodobieństw muszą świadczyć za tezami poważnych i zasłużonych znawców rachunku prawdopodobieństwa. Ta uwaga niech uchroni autora niniejszego artykułu przed zarzutem, że wszystko, co napisał, dotyczy spraw, o których nauka wypowiedziała ostatnie słowo i które przestały być problematami, a stały się rozdziałami programów szkolnych i podręczników powszechnie uznanych.

Instytut Matematyczny Polskiej Akademii Nauk

(Praca wpłynęła dnia 2.1.1953 r.)

Г. ШТЕЙНГАУЗ (Вроцлав)

ВЕРОЯТНОСТЬ, ПРАВДОПОДОБИЕ, ВОЗМОЖНОСТЬ

РЕЗЮМЕ

Предметом настоящей работы является противоречие, связанное с правилом Бейеса. Обычно нельзя применять это правило потому, что неизвестны априорные вероятности. Постулат о равномерном распределении *a priori* (постулат Бейеса) встречается с возражением, что он произвольный и с еще более тяжелым упреком, что при одновременном наблюдении переменной и ее функции (например куба) неизвестно, к которой серии наблюдений отнести постулат. Английская школа (Р. А. Фишер, К. Пирсон) распространила метод, который я называю *теорией правдоподобия*. Вместо того, чтобы вычислять *вероятность* $P(x<a;\ y=b)$ того, что неизвестная величина x меньше чем a, когда ее измерение y дало b (что невозможно без правила Бейеса), вычисляется *правдоподобие* $W(x<a;\ y=b)$ того, что $x<a$, когда $y=b$, причем W определяется посредством формулы $W(x<a;y=b)=P(y>b;x=a)$, где по правой стороне фигурирует условная вероятность наблюдения $y>b$, которая известна. Можно показать, что когда условные вероятности зависят только от абсолютной погрешности, то так определенное правдоподобие дает точно то же самое, что правило Бейеса при предположении равномерного распределения величины x *a priori* (постулат Бейеса). И так, если вероятность $P\mathfrak{B}(x<a;y=b)$ назовем *возможностью* $M(x<a;y=b)$ того, что $x<a$, когда $y=b$, то получим то же самое число, которое дает теория правдоподобия как заменитель вероятности: $M=W$. Термин „возможность" освобождает от оправдывания постулата Бейеса, так как термин „правдоподобие" использованный для определения не искомой вероятности а иной, оправдывает смещение вопроса в новой теории. Возражение против того, что выбор переменной, распределенной равномерно является произвольным устраняется замечением, что все переменные, при которых вероятность погрешности наблюдения зависит только от величины погрешности, дают те же самые возможности: каждую из них можно принять как равномерную, и никакой другой нельзя.

Аналогичные рассуждения справедливы в случае, когда x является неизвестной вероятностью события Z, а наблюдения заключаются в констатировании в n пробах, сколько раз имело место Z. Здесь x однозначно указаная вероятность события и поэтому отпадает возражение, что гипотеза о равномерном распределении неконсистентна.

H. STEINHAUS (Wrocław)

PROBABILITY, VERISIMILITUDE, CREDIBILITY

SUMMARY

The subject of the paper is the controversy connected with Bayes' rule. Usually, the rule cannot be applied because the prior probabilities are not known. The assumption of a uniform prior distribution (the postulate of Bayes) incurs the charge of arbitrariness, and it may be pointed out that in observing simultaneously a variable and its function (*e. g.* a cube) we do not know to which series of observations the postulate should be related, which is a still more serious objection. The English school (R. A. Fisher, K. Pearson) has propagated a method which I have called the *theory of verisimilitude*. Instead of calculating the *probability* $P(x<a; y=b)$ of an unknown quantity being smaller than a, if its measurement y has given b, which calculation is not possible without Bayes' rule, we calculate the *verisimilitude* $W(x<a; y=b)$ of $x<a$ when $y=b$, W being defined by means of the relation $W(x<a; y=b)$ $=P(y>b; x=a)$ — on the right side we have the conditional probability of the observation $y>b$, which is known. It can be shown that if the conditional probabilities depend only on absolute error, then the verisimilitude defined above gives exactly the same result as Bayes' rule does when we assume the uniform prior distribution of x (the postulate of Bayes). Therefore, if the probability $P_{\mathfrak{B}}(x<a; y=b)$ is called the *credibility* $M(x<a; y=b)$ of $x<a$ when $y=b$, the same number will be obtained as that given by the theory of verisimilitude as a substitute for the probability: $M=W$. The term "credibility" allows us to dispense with the justification of the postulate of Bayes, just as the term "verisimilitude", used to define a different probability from that which has been sought, justifies the shifting of the problem in the new theory. We refute the objection that the choice of a uniformly distributed variable has been arbitrary by pointing out that all random variables for which the probability of an error of observation depends only on the size of the error present the same credibility: each of them may be regarded as uniform, and no other can be regarded as such.

Similar considerations relate to the case when x is an unknown probability of a phenomenon Z, and the observations consist in determining in n trials how many times Z has occurred. Here x is recognized as the probability of the phenomenon, and the objection of the equipartition being arbitrary is refuted on that account.
