O STOPNIU DOKŁADNOŚCI SPÓŁCZYNNIKA ROZBIEŻNOŚCI

napisał

Wł. Bortkiewicz.

Na polu badania naukowego, które Laplace nazywa zastosowaniem teoryi prawdopodobieństwa do życia cywilnego (l'application de la théorie des probabilités à la vie civile), wybitne miejsce zajmują rozważania nad stałością względną szeregów statystycznych. Wskazanie znaczenia zasadniczego pytania, dotyczącego tej stałości dla teoryi statystyki, jest zasługą Lexisa [1]). Przed niedawnym czasem różni autorowie [2]) zwrócili uwagę na to, że rozważania te mają też doniosłość dla sprawy uzasadnienia naukowego oraz praktyki ubezpieczeń życiowych. Niechaj mi tu będzie wolno poświęcić słów kilka pytaniu o wydoskonaleniu stosowanej w tem zagadnieniu metody.

Wiadomo, że dla scharakteryzowania względnej stałości szeregu statystycznego rozważa się zwykle stosunek dwóch pierwiastków kwadratowych, z których jeden przedstawia miarodajny dla danego szeregu statystycznego błąd średni kwadratowy, obliczony według metody bezpośredniej („fizykalnej"), drugi zaś takiż

---

[1]) Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft. Freiburg 1877. Porówn. E. Dormoy, Théorie mathématique des assurances sur la vie. Paryż 1878. Tom I № 39—5). E. Czuber, Die Entwickelung der Wahrscheinlichkeitstheorie und ihrer Anwendungen. Lipsk 1899. № 71—73. Wł. Bortkiewicz, W wydawnictwie: Encyclopaedie der mathem. Wissenschaften. I D4a № 4-6, gdzie podana jest dalsza literatura.

[2]) J. H. Peek, Das Problem vom Risico in der Lebensversicherung (Zeitschr. für Versicherungsrecht und Wissenschaft. t. V. Strassburg 1899). G. Bohlmann w książce Kleina i Rieckego: „Ueber angewandte Mathematik" etc. Lipsk 1900, str. 137—145. E. Blaschke, Die Anwendbarkeit der Wahrscheinlichkeitslehre im Versicherungswesen (Statistische Monatschrift. Wiedeń 1901).

błąd, obliczony według metody pośredniej („kombinatoryjnej“). Jeżeli tedy

$$p'_1, p'_2, \ldots, p'_\sigma$$

jest szeregiem statystycznym, złożonym z σ elementów, gdzie zakłada się, że każdy pojedyńczy wyraz tego szeregu może być pod względem formalnym uważany jako wyrażenie przybliżone pewnego prawdopodobieństwa, i nadto, że każdy pojedyńczy wyraz p'_i otrzymano z tej samej liczby s pojedyńczych spostrzeżeń, to wspomnianemi dwoma pierwiastkami kwadratowemi są:

$$\sqrt{\sum \frac{(p'_i - p'_0)^2}{\sigma - 1}}, \quad \sqrt{\frac{p'_0 q'_0}{s}},$$

gdzie

$$p'_0 = \sum \frac{p'_i}{\sigma}, \quad q'_0 = 1 - p'_0.$$

Jeżeli mamy większą liczbę analogicznych szeregów statystycznych, których stałość ma być zbadana, wtedy to wyznaczywszy osobno dla każdego szeregu iloraz:

$$D = \sqrt{\sum \frac{(p'_i - p'_0)^2}{\sigma - 1}} : \sqrt{\frac{p'_0 q'_0}{s}},$$

zwany spółczynnikiem rozbieżności [1]), bierzemy średnią arytmetyczną wszystkich w ten sposób otrzymanych wartości na D i otrzymujemy wyrażenie stosunków dyspersyjnych, panujących w rozważanej dziedzinie. Jeżeli przy tem na spółczynnik rozbieżności znajdziemy wartość średnią bliską jedności, stanowić to będzie kryteryum dyspersyi normalnej, i wtedy zagadnienie dotyczące dyspersyi, zwykle jak dotąd, uważano już za załatwione.

---

[1]) Nazwę tę wprowadził E. Dormoy, ale przez „coefficient de divergence“ rozumiał stosunek w rozmaity sposób obliczonych wartości nie błędu średniego kwadratowego lecz średniego błędu arytmetycznego. I dla tego nie jest to ściśle zgodne z prawdą, że Blaschke na str. 15 cytowanej rozprawy przypisuje ten sposób wyrażenia wprost E. Dormoy'owi. Porów. G. Bohlmann w książce Kleina i Rieckego, str. 138.

Wydaje się wszakże rzeczą, nie pozbawioną znaczenia, bliższe zbadanie odchyleń pojedyńczych wartości spółczynnika rozbieżności od 1, mianowicie pod tym względem, czy wielkości tych odchyleń odpowiadają przewidywaniom teoryi.

Poniższe wywody mają właśnie za zadanie pozyskanie odpowiedniej miary tych przewidywań.

Oprócz oznaczeń powyższych, wprowadzimy jeszcze następujące:

1. Liczbę zdarzeń, odpowiadającą pojedyńczej wartości p'_i, oznaczmy przez x_i, tak że $p'_i = \frac{x_i}{s}$.

2. Prawdopodobieństwo oderwane, którego wartościami przybliżonemi są wielkości p'_i, oznaczmy przez p, i niechaj $1 - p = q$.

3. Prawdopodobieństwem, że uważane zdarzenie zachodzi x razy przy s spostrzeżeniach pojedyńczych, niechaj będzie $P_{s,x}$, tak że:

$$P_{s,x} = \frac{s(s-1)\ldots(s-x+1)}{1.2\ldots x} p^x q^{s-x}.$$

4. Niechaj będzie:

$$\sum_{x=0}^{x=s} P_{s,x} (x - sp)^r = u_r.$$

5. $F(z)$ niechaj oznacza błąd średni kwadratowy dowolnej wielkości z.

Przedewszystkiem winniśmy wyznaczyć średni błąd kwadratowy wyrażenia $\frac{p'_0 q'_0}{s}$. Najlepiej w tym celu wyjść ze średniego błędu kwadratowego wyrażenia $sp'_i q'_i$, gdzie $q'_i = 1 - p'_i$. Kładąc $p'_i = p + a$, otrzymujemy:

$$sp'_i q'_i = s(p + a)(q - a),$$

albo też z powodu, że $sa = x_i - sp$,

$$sp'_i q'_i = spq + (x_i - sp)(q - p) - \frac{(x_i - sp)^2}{s}.$$

Z określenia pojęcia średniego błędu kwadratowego wypływa:

$$F(sp'_i q'_i) = \sqrt{\sum_{x=0}^{x=s} \left\{ (x_i - sp)(q-p) - \frac{(x_i - sp)^2}{s} \right\}^2 P_{s,x}}\,,$$

a przy zastosowaniu oznaczenia, podanego wyżej pod 4:

$$F(sp'_i q'_i) = \sqrt{(q-p)^2 u_2 - 2(q-p)\frac{u_3}{s} + \frac{u_4}{s^2}}\,.$$

Uwzględniając tu znane związki [1]:

$$u_2 = spq \quad ; \quad u_3 = spq(q-p) \quad , \quad u_4 = 3(spq)^2 + spq - 6sp^2q^2\,,$$

znajdziemy:

$$F(sp'_i q'_i) = \sqrt{(q-p)^2 spq - 2(q-p)^2 pq + 3p^2q^2 + \frac{pq}{s} - \frac{6p^2q^2}{s}}\,.$$

Z tego równania otrzymamy $F(\sigma s p'_0 q'_0)$, jeżeli po stronie prawej napiszemy wszędzie σs zamiast s. Ale mamy związek:

$$F(\sigma s p'_0 q'_0) : \sigma s^2 = F\left(\frac{p'_0 q'_0}{s}\right),$$

stąd więc:

$$(\alpha)\quad F\left(\frac{p'_0 q'_0}{s}\right) = \frac{1}{s}\sqrt{\frac{(q-p)^2 pq}{\sigma s} - \frac{2(q-p)^2 pq}{(\sigma s)^2} + \frac{3p^2q^2}{(\sigma s)^2} + \frac{pq}{(\sigma s)^3} - \frac{6p^2q^2}{(\sigma s)^3}}\,.$$

Wyznaczymy teraz błąd średni kwadratowy wyrażeń $\sum \frac{(p'_i - p)^2}{\sigma}$, $\sum \frac{(p_i - p'_0)^2}{\sigma - 1}$.

Mamy najprzód:

$$F\{(p'_i - p)^2\} = \sqrt{\sum_{x=0}^{x=s} \left\{ \left(\frac{x}{s} - p\right)^2 - \frac{pq}{s} \right\}^2 P_{s,x}}\,,$$

---

[1]) Wywód elementarny tych związków znaleść można w pracy mojej: „Das Gesetz der kleinen Zahlen“. Lipsk 1898, str. 41.

skąd:

$$F\{(p'_i-p)^2\}=\sqrt{\frac{u_4}{s^4}-\frac{2u_2pq}{s^3}+\frac{p^2q^2}{s^2}},$$

lub też:

$$F\{(p'_i-p)^2\}=\frac{1}{s}\sqrt{2p^2q^2+\frac{pq}{s}-\frac{6p^2q^2}{s}},$$

Ostatecznie dochodzimy do wzoru:

$$(\beta)\qquad F\left\{\sum\frac{(p'_i-p)^2}{\sigma}\right\}=\frac{1}{s}\sqrt{\frac{2p^2q^2}{\sigma}+\frac{pq}{\sigma s}-\frac{6p^2q^2}{\sigma s}}.$$

Z zestawienia równań (α) i (β) widać, że gdy liczba spostrzeżeń s rośnie, błąd średni kwadratowy wyrażenia $\frac{p'_0q'_0}{s}$ maleje nieograniczenie nie tylko co do wielkości bezwzględnej lecz i w stosunku do błędu średniego kwadratowego wyrażenia $\sum\frac{(p'_i-p)^2}{\sigma}$; można przeto w przypadku, gdy s jest dużą liczbą, przyjąć bezpiecznie, że błąd średni kwadratowy ilorazu

$$\sum\frac{(p'_i-p)^2}{\sigma}:\frac{p'_0q'_0}{s}$$

równa się błędowi średniemu kwadratowemu wyrażenia $\sum\frac{(p_i-p)^2}{\sigma}$, podzielonemu przez $\frac{pq}{s}$. Gdy równocześnie σ nie jest zbyt małe i zgodnie z tem $\sum\frac{(p'_i-p'_0)^2}{\sigma-1}$ może być uważane za przybliżenie dostateczne dla wyrażenia $\sum\frac{(p_i-p)^2}{\sigma}$, to z równania ($\beta$) można wyprowadzić wprost:

$$F(D^2)=\sqrt{\frac{2}{\sigma}+\frac{1}{\sigma spq}-\frac{6}{\sigma s}},$$

a ponieważ zatrzymanie wyrazów, mających s w mianowniku, byłoby tu bez znaczenia, przeto:

$$F(D^2) = \sqrt{\frac{2}{\sigma}},$$

skąd ostatecznie według znanej metody przybliżenia:

$$F(D) = \frac{1}{\sqrt{2\sigma}}$$

stanowi wyrażenie średniego błędu kwadratowego spółczynnika rozbieżności.

Jeżeli więc mamy szereg:

$$D_1, D_2, \ldots, D_n,$$

wartości spółczynników rozbieżności, to przy dyspersyi normalnej należałoby oczekiwać, że równanie

$$\sqrt{\sum_{i=1}^{i=n} \frac{(1-D_i)^2}{n}} = \frac{1}{\sqrt{2\sigma}},$$

jest przybliżenie spełnionem.

Zobaczmy najprzód na wziętym z Lexisa przykładzie, dotyczącym stosunku płci nowonarodzonych, czy wzór ten daje się stwierdzić przy pomocy danych z obserwacyj statystycznych.

W Nr. 43 swej pracy: „Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft“ Lexis wyniki swych badań, dotyczących powyższego przykładu, zestawił w tablicy, podającej pod literą *C* dokładności odpowiednich prawdopodobieństw [1]), obliczone według metody „kombinatoryjnej“, pod literą zaś *Q* dokładności, obliczone według metody „fizykalnej“, a to oddzielnie dla okresu 1868—69 i dla okresu 1870—72. Lexis oparł się przy tem na liczbach urodzeń miesięcznych. Według tego, wiel-

[1]) Właściwie mówiąc, u Lexisa mowa jest o wartościach empirycznych prawdopodobieństwa względnego, jakiem jest stosunek liczby urodzonych chłopców do liczby urodzonych dziewcząt. Nie zwracamy tu na to uwagi i rachujemy z wielkościami *C* i *Q* tak, jak gdyby odnosiły się do prawdopodobieństw prawdziwych. Należy przyznać, że postępowanie to zawiera w sobie pewną niedokładność.

kość, oznaczoną wyżej przez σ, należy przyjąć dla okresu pierwszego jako równą 24, dla drugiego — 36. Spółczynnikowi rozbieżności D odpowiada u Lexisa iloraz $\frac{C}{Q}$, albowiem dokładności są w stosunku odwrotnym do średnich błędów kwadratowych. Stąd, według wzoru wyżej podanego, należałoby oczekiwać, że wielkość

$$\sqrt{\frac{1}{68}\sum\left(\frac{C}{Q}-1\right)^2}$$

równa się przybliżenie wielkości

$$\sqrt{\frac{1}{68}\left(34\cdot\frac{1}{48}+34\cdot\frac{1}{72}\right)}.$$

Jako wartość pierwszej z tych wielkości otrzymujemy 0,145, jako wartość drugiej 0,132. Wynik jest jeszcze bardziej zadawalający, jeżeli zamiast spółczynnika rozbieżności weźmiemy dokładność Q, obliczoną według metody fizykalnej i jej błąd średni kwadratowy $F(Q)$. Przy pomocy powyższego łatwo otrzymujemy $C\sqrt{\frac{1}{2\sigma}}$, jako wyrażenie przybliżone na $F(Q)$, gdzie C jest dokładnością, obliczoną według metody kombinatoryjnej. Tym sposobem w przypadku rozważanym należy porównać ze sobą wielkości

$$\sqrt{\sum\frac{(Q-C)^2}{68}} \text{ i } \sqrt{\sum\frac{C^2}{68\,.\,2\sigma}},$$

gdzie σ w 34 przypadkach równa się 24 i w 34 równa się 36. Rachunek daje dla pierwszego z tych pierwiastków kwadratowych 0,002115, dla drugiego 0,002110.

Daleko mniej zadawalającemi są otrzymane z przyjętego tu punktu widzenia wyniki, odnoszące się do wahań prawdopodobieństwa śmierci, ogłoszonych przez J. H. Peeka [1]). Co się najprzód tyczy wyników, osiągniętych ze statystyki ludności-

[1]) W rozprawie wyżej przytoczonej, str. 180—182 i 191.

wej Niderlandów za czas 1880—89, to — odwracając nawet uwagę od pierwszych 11 klas wieku, dla których spółczynnik rozbieżności D dość znacznie przewyższa 1 — otrzymujemy:

$$\sqrt{\sum \frac{(1-D)^2}{n}} = 0{,}3577\,,$$

gdy tymczasem odpowiednia wartość teoretyczna wynosi $\sqrt{\frac{1}{2\sigma}} = 0{,}2236$. (Tu $n = 80$, $\sigma = 10$). Drugi zaś przykład P e e k a, oparty na obserwacyach śmiertelności urzędników niderlandzkich w czasie 1878—1894, daje wyniki:

$$\sqrt{\sum \frac{(1-D)^2}{n}} = 0{,}1284 \qquad \sqrt{\frac{1}{2\sigma}} = 0{,}1715\,.$$

(W tym przypadku $n = 12$, $\sigma = 17$).

W tym drugim przykładzie uwzględnił P e e k nierówność liczb spostrzeżeń, będących podstawą wartości pojedyńczych wyrazów odpowiednich szeregów, a to w ten sposób, że dla każdego pojedyńczego wyrazu p'_i obliczył iloraz $\delta_i^2 = (p_i - p'_0)^2 : \frac{p'_0 q'_0}{s_i}$ (gdzie s_i jest liczbą spostrzeżeń, odpowiadającą danej wartości p'_i, p'_0 zaś równa się $\frac{\Sigma p'_i s_i}{\Sigma s_i}$), a następnie za spółczynnik rozbieżności przyjął [1]) $\sqrt{\sum \frac{\delta_i^2}{\sigma - 1}}$. Najdogodniej obliczyć tę ostatnią wielkość przy pomocy wzoru

$$\sqrt{\frac{1}{(\sigma-1)p'_0 q'_0} \sum \frac{(x_i - s_i p'_0)^2}{s_i}}\,.$$

Łatwo się przekonać, że błąd średni kwadratowy, obliczonego i w ten sposób spółczynnika rozbieżności równa się przybliżenie $\frac{1}{\sqrt{2\sigma}}$.

[1]) Metodę tę, która, że tak powiemy, nasuwa się sama przez się, zaproponowałem w artykule: „Kritische Betrachtungen zur theoretischen Statistik“. (Jahrbücher für Nationaloekonomie und Statistik. Neue Folge, t. VIII. 1894, str. 673, przypisek).