Antoni DRAPELLA

O złej radzie dotyczącej testu *F* Snedecora

Streszczenie. *W polskojęzycznych pracach (wydanych w formie drukowanej, jak również dostępnych w Internecie) omawiających weryfikację hipotezy o równości wariancji w dwóch populacjach gaussowskich (test F) pojawia się rada, by oszacowania wariancji o większej wartości umieszczać w liczniku statystyki testowej, a jako granicę obszaru dopuszczalnego przyjąć stosowny kwantyl rozkładu F. Prowadząc w artykule rozważania analityczne i eksperymenty numeryczne pokazano, że jest to zła rada, ponieważ idąc za nią czyni się rzeczywisty błąd pierwszego rodzaju dwukrotnie większy od założonego.*

Słowa kluczowe: rozkład chi-kwadrat, rozkład *F*, test *F*, metoda Monte Carlo.

Sprawa, której dotyczy artykuł, miała początek, gdy natrafiałem w Internecie na materiały dydaktyczne omawiające m.in. test *F*. Autorzy tych materiałów radzili, aby liczyć statystykę testową umieszczając większe z oszacowań wariancji w liczniku, a mniejsze w mianowniku oraz przyjmując za górną granicę obszaru dopuszczalnego stosowny kwantyl rozkładu *F*. Okazało się, że podobne rady można znaleźć np. w pracach Brandta (1998), Starzyńskiej (2002) i Zielińskiego (1972). Ciekawe, w ilu jeszcze nowszych podręcznikach będzie zawarta ta sama rada, z jednoczesnym powołaniem się na poprzednie. W dalszej części artykułu przedstawiono eksperymenty numeryczne pokazujące, że jest to zła rada, bowiem postępując według niej uzyskuje się błąd pierwszego rodzaju dwukrotnie większy od założonego. Następnie przedstawiono rozważania analityczne wyjaśniające, dlaczego tak się dzieje.

By stwierdzić, że rada jest zła nie potrzebujemy ani eksperymentu, ani analizy. Przypomnijmy, że zmienna losowa *F* podlegająca rozkładowi Snedecora jest ilorazem niezależnych zmiennych losowych o rozkładzie chi-kwadrat:

$$F = \frac{\chi_1^2 / n_1}{\chi_2^2 / n_2} \tag{1a}$$

gdzie n_1, n_2 — liczba stopni swobody zmiennej losowej chi-kwadrat indeksowanych 1, 2.

Statystykę utworzoną według omawianej rady oznaczmy symbolem *WF*. Ma ona postać:

$$WF = \max(F, 1/F) \tag{1b}$$

i oczywiście nie podlega rozkładowi F choćby z tego powodu, że jej dystrybuanta jest z definicji tożsamościowo równa 0 w przedziale $< 0, 1)$. Skoro statystyka testowa nie podlega rozkładowi F, nie może on być użyty do wyznaczania granicy obszaru dopuszczalnego.

EKSPERYMENT NUMERYCZNY METODĄ MONTE CARLO

Eksperyment jest bardzo prosty i wykonano go w Excelu. Stosowny skoroszyt Excela można uzyskać od autora pisząc na adres adrastat.@hotmail.com. Wygenerowano dwa zestawy po 5000 próbek. Każda próbka to 30 realizacji zmiennej losowej o rozkładzie $N(0, 1)$. Z każdej próbki oszacowano wariancję. Wygenerowano też 5000 realizacji b_i^*, $i = 1, 2, ..., 5000$, binarnej zmiennej losowej B o rozkładzie $\Pr(B=0)=\Pr(B=1)=1/2$. Gdy okazywało się, że $b_i^* = 1$, to w liczniku formuły umieszczano wariancję oszacowaną z pierwszego zestawu, natomiast gdy $b_i^* = 0$ — z drugiego. Postępując tak uzyskiwano realizacje F_i^*, $i = 1, 2, ..., 5000$, realizacji zmiennej losowej z rozkładu zmiennej losowej F.

Korzystając z tych samych zestawów wyznaczono realizacje zmiennej losowej WF_i^*, $i = 1, 2, ..., 5000$, zawsze umieszczając większą wariancję w liczniku. Mając realizacje F_i^* oraz WF_i^* weryfikowano 5000 razy hipotezę H_0 mówiącą, że wariancje w populacjach gaussowskich, z których pochodziły próby są jednakowe. Konkurowała z nią hipoteza H_1 mówiąca, że wariancje są różne. Prawdziwa była oczywiście H_0. Przyjęto poziom istotności $\alpha = 0{,}1$. Realizowano dwa schematy weryfikacji:

Schemat S_1

Granice obszaru dopuszczalnego to kwantyle rozkładu zmiennej losowej F odpowiadające prawdopodobieństwu $\alpha/2$ i $1-\alpha/2$ przy 29 stopniach swobody dla licznika i mianownika. Dolna granica obszaru dopuszczalnego wynosi $F_d^{kr} = 0{,}537$, z kolei górna granica — $F_g^{kr} = 1{,}861$. Hipotezę H_0 odrzucano, jeśli $F_i^* < F_d^{kr}$ lub $F_i^* > F_g^{kr}$.

Schemat S_2 (według wspomnianej rady)

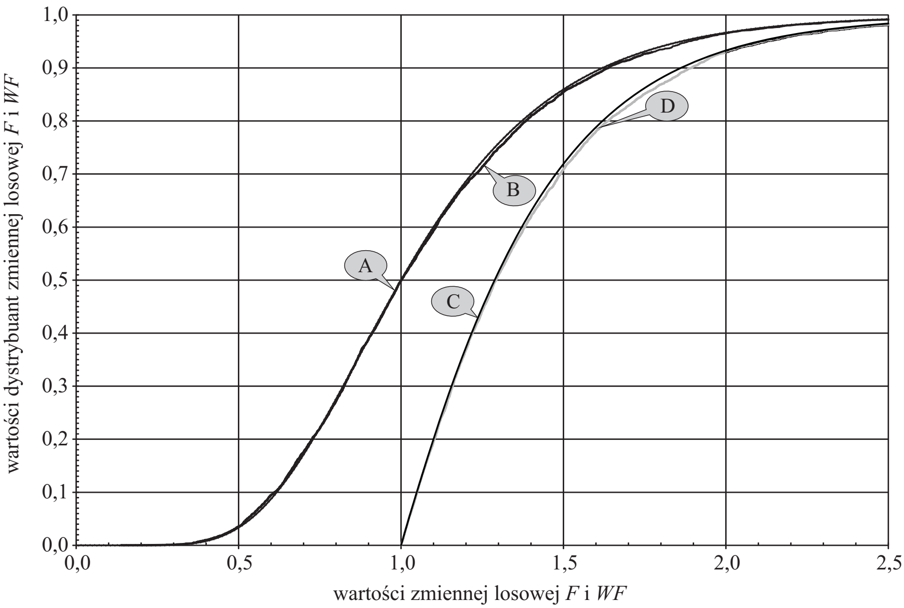
Górna granica obszaru dopuszczalnego to $F_g^{kr} = 1{,}620$. Jest to kwantyl rozkładu zmiennej losowej F odpowiadający prawdopodobieństwu $1-\alpha$. Hipotezę H_0 odrzucano, jeśli $WF_i^* > F_g^{kr}$.

W toku realizacji schematu S_1 miało miejsce 546 przypadków odrzucenia H_0. Zatem poziom istotności *ex post* wynosił 546/5000=0,1092, czyli był równy założonemu.

W toku realizacji schematu S_2 miało miejsce 1041 przypadków odrzucenia H_0, zatem poziom istotności *ex post* to 1041/5000=0,2082. Jest on więc dwukrotnie większy od założonego.

Na wykr. 1 przedstawiono cztery dystrybuanty oznaczone jako A, B, C, D.

Wykr. 1. TEORETYCZNE I DOŚWIADCZALNE DYSTRYBUANTY ZMIENNEJ LOSOWEJ *F* i *WF*



Ź r ó d ł o: opracowanie własne na podstawie przeprowadzonego badania.

Są to dystrybuanty:

A — rozkładu zmiennej losowej *F* dla 29 stopni swobody zarówno w liczniku, jak i w mianowniku;

B — doświadczalna rozkładu zmiennej losowej *F* uzyskana z „próby” powstałej przez 5000-krotną realizację schematu S_1;

C — rozkładu zmiennej losowej *WF*=max(*F*,1/*F*). Sposób jej wyliczenia podano w następnej części artykułu;

D — doświadczalna rozkładu zmiennej losowej *WF* uzyskana z „próby” powstałej przez 5000-krotną realizację schematu S_2.

Dystrybuanta B pokrywa się z dystrybuantą A, zatem schemat S_1 przynosi statystykę podlegającą rozkładowi *F*. Dystrybuanta D pokrywa się z dystrybuantą C, zatem schemat S_2 przynosi statystykę podlegającą rozkładowi *WF*. Eksperyment powtórzono, ale z jedną zmianą odnoszącą się do drugiego zestawu próby o liczebności 15. Uzyskano identyczny wynik. Schemat S_1 skutkował błędem pierwszego rodzaju identycznym z założonym, schemat S_2 — dwukrotnie większym od założonego.

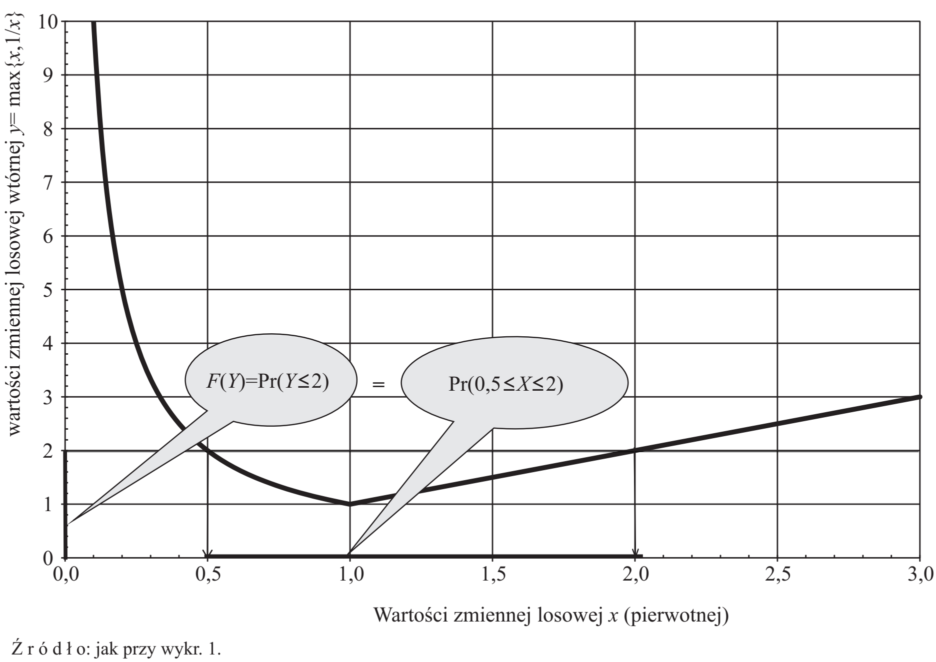
ROZWAŻANIA ANALITYCZNE

Litera *F* przypisana rozkładowi Snedecora jest też powszechnie stosowana do oznaczania dystrybuanty. Dlatego teraz zastąpimy *F* przez *X*, zaś *WF* — przez *Y*. Funkcja przekształcająca wartość *x* zmiennej losowej *X* w wartość *y* zmiennej losowej *Y* ma postać:

$$y = \max(x, 1/x) = \begin{cases} 1/x \text{ gdy } x < 1 \\ x \text{ gdy } x \geq 1 \end{cases} \tag{2}$$

Funkcję przekształcającą pokazano na wykr. 2.

Wykr. 2. ILUSTRACJA TECHNIKI WYZNACZANIA DYSTRYBUANTY ZMIENNEJ LOSOWEJ *WF*



Ź r ó d ł o: jak przy wykr. 1.

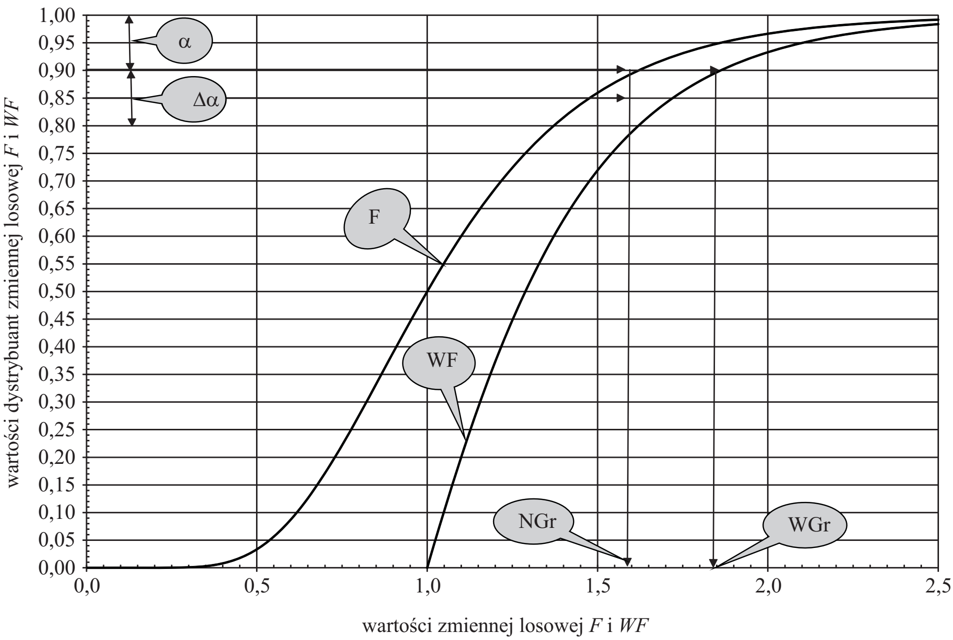
Kierując się nim otrzymujemy:

$$F_Y(y) \stackrel{def}{=} \Pr(Y \leq y) = \Pr(1/y \leq X \leq y) = F_X(y) - F_X(1/y) \tag{3}$$

Przebieg dystrybuant zmiennej losowej X i Y pokazano na wykr. 3, gdzie:

α — błąd pierwszego rodzaju, czyli poziom istotności testu;
$\Delta\alpha$ — przyrost błędu pierwszego rodzaju spowodowany zastosowaniem niewłaściwego rozkładu statystyki testowej w schemacie S_2, czyli rozkładu *F*;
F — krzywa oznaczająca dystrybuantę rozkładu zmiennej losowej *F*;
WF — krzywa oznaczająca dystrybuantę zmiennej losowej *WF*;
NGr — niewłaściwa granica obszaru dopuszczalnego;
WGr — właściwa granica obszaru dopuszczalnego.

Wykr. 3. ILUSTRACJA PRZYCZYN WZROSTU PRAWDOPODOBIEŃSTWA BŁĘDU I RODZAJU



Ź r ó d ł o: jak przy wykr. 1.

Wykres pomaga w wyjaśnieniu, dlaczego postępując według S_2 popełnia się błąd pierwszego rodzaju znacznie większy od założonego. Rozkład *F* ma tę szczególną właściwość, że ok. 50% prawdopodobieństwa leży w przedziale $\langle 0,1 \rangle$ i, co bardzo ważne, zmiana liczby stopni swobody niewiele tę frakcję zmienia. Zatem połowa (!) prawdopodobieństwa musi być przy przejściu od *X* do *Y* przemieszczona powyżej jedynki.

Niech α będzie przyjętym poziomem istotności testu. Kwantyl rozkładu zmiennej losowej *Y* (czyli *WF*) odpowiadający prawdopodobieństwu $1-\alpha$ będzie z opisanego powodu znacznie większy od odpowiadającego temu samemu prawdopodobieństwu kwantyla rozkładu zmiennej losowej X (czyli F). Nawet jeśli H_0 jest słuszna, to 12% realizacji statystyki testowej powstałej w toku obliczeń prowadzonych według schematu S_2 wpadnie między NGr i WGr. Wzrost błędu pierwszego rodzaju i tak jest mały, jak na to, co mogłoby być, gdyby powstający w wyniku przekształcenia (2) rozkład zmiennej losowej Y (WF) nie wypiętrzał się tuż za jedynką, a rósłby mu tzw. „ogon".

Podsumowanie

Weryfikując testem F hipotezę o równości wariancji w dwóch populacjach gaussowskich, można postępować dwojako:

- losowo przydzielać oszacowania wariancji do licznika i mianownika, czyli zrealizować schemat S_1. Generatorem liczb losowych binarnych o rozkładzie równomiernym może być chociażby moneta lub kostka;
- większe co do wartości oszacowanie umieszczać w liczniku, mniejsze w mianowniku, czyli realizować S_2. Z kolei jako górną granicę obszaru dopuszczalnego przyjmować kwantyl rozkładu F odpowiadający prawdopodobieństwu $1-\alpha$ i odrzucać słuszną H_0 dwa razy częściej niż się zamierzało.

---

prof. dr hab. Antoni Drapella — *Akademia Marynarki Wojennej w Gdyni*

LITERATURA

Brandt S. (1998), *Analiza danych*, PWN, Warszawa.
Starzyńska W. (2002), *Statystyka praktyczna*, PWN, Warszawa.
Zieliński R. (1972), *Tablice statystyczne*, PWN, Warszawa.

Summary. *Readers of some domestic statistical textbooks and Internet publications related to F test are advised to accomplish the following test scheme: After having sample variances calculated use quotient of greater to smaller of them as the test statistics. Then take $1-\alpha$ quantile of the F distribution as the critical value. This paper identifies this advice to be wrong and gives reason for it: test statistics in question definitely does not follow the F distribution. So, derivation of the proper test statistics named WF as well as the method of calculating $WF's$ cumulative distribution function is given. Analytical considerations are confirmed by two Monte Carlo experiments. These show that following the advice one makes first type error two times greater than wanted.*

Keywords: the F distribution, the chi-square distribution, the Snedecor test, the Monte Carlo method.

Резюме. *В разработках на польском языке (опубликованных в печатном виде и доступных в Интернете) обсуждающих проверку гипотезы равенства дисперсии в двух гауссовских популяциях (критерий F) появляется совет, чтобы оценки дисперсии с большим значением помещать в счетчике тестовой статистики, а в качестве границы допустимой площади принять соответствующий квантиль распределения F. Представляя в статье аналитические соображения и численные эксперименты было показано, что это плохой совет, так как следуя ему делается реальная ошибка первого вида два раза больше чем планированная.*

Ключевые слова: распределение хи-квадрат, распределение *F*, критерий *F*, метод Монте Карло.