Politechnika Wrocławska

Wydział Matematyki

Kierunek: Matematyka Stosowana
Specjalność: *nie dotyczy*

Praca dyplomowa — inżynierska

TESTOWANIE HIPOTEZ I ESTYMACJA W SYTUACJI POPULACJI SKOŃCZONEGO ROZMIARU

Kinga Kurowska

Słowa kluczowe:
testowanie hipotez
przedziały ufności
zastosowanie w medycynie

Krótkie streszczenie:
W pracy został rozważony problem testowania hipotez w sytuacji populacji skończonego rozmiaru. Przedstawiono i porównano dwa schematy pobierania obserwacji, dla nieskończonej oraz skończonej populacji. Zostały opisane trzy testy statystyczne oraz przeprowadzone badania symulacyjne w celu porównania testów. Ostatecznie został wybrany jeden test, który jest najlepszy dla małych populacji.

Opiekun pracy dyplomowej	dr inż. Andrzej Giniewicz		
	Stopień naukowy, imię i nazwisko	*Ocena*	*Podpis*

*Do celów archiwalnych pracę dyplomową zakwalifikowano do: **
a) kategorii A (akta wieczyste),
b) kategorii BE 50 (po 50 latach podlegające ekspertyzie).
** niepotrzebne skreślić*

pieczątka wydziałowa

Wrocław, rok 2017

Politechnika Wrocławska

Faculty of Pure and Applied Mathematics
Field of study: Applied Mathematics
Specialty: *not applicable*

Engineering Thesis

HYPOTHESIS TESTING AND ESTIMATION IN THE CASE OF FINITE POPULATION SIZE

Kinga Kurowska

keywords:
hypothesis testing
confidence intervals
medical applications

Short summary:
The problem of hypothesis testing in the case of finite population size was considered in the thesis. Two schemes of sampling, for infinite and finite population, were presented and compared. Three statistical tests were described. Simulations studies were carried out in order to compare tests. Finally, one test was chosen, which is the best for small population size.

Supervisor	dr inż. Andrzej Giniewicz		
	Title, degree, name and surname	*Grade*	*Signature*

*For the purposes of archival thesis qualified to: **
a) *Category A (perpetual files)*
b) *Category BE 50 (subject to expertise after 50 years)*
** Delete as appropriate*

stamp of the faculty

Wrocław, 2017

Spis treści

Wstęp

Początki teorii rachunku prawdopodobieństwa i statystyki sięgają XVI wieku. Zajmowano się wtedy analizą rzutu kostką oraz prawdopodobieństwem błędów pomiarowych. Już w XVII Blaise Pascal sformułował i dowiódł własności trójkąta arytmetycznego oraz użył pojęcia kombinacji [1]. Na początku XVIII wieku opublikowane zostały prace Jacoba Bernoullego, w których zawarł wiele swoich tez na temat prawdopodobieństwa. Przez te kilka wieków teoria rachunku prawdopodobieństwa i statystyki znacząco się wzbogaciła i rozwinęła. Rozpoczęto rozważania na temat estymacji i testowania hipotez, które są w naszych czasach zasadniczą domeną statystyki.

W przypadku dyskretnym najczęściej testowane są proporcje populacji [2]. Chcemy się przekonać czy dana próbka ma jakąś konkretną proporcję albo dwie próbki mają tę samą proporcję elementów z badaną cechą. Znana jest powszechnie teoria dotycząca testowania hipotez, gdy populacja jest nieskończona, a raczej na tyle duża, że możemy ją w przybliżeniu uznać za nieskończoną. Wtedy schemat próbkowania [3] jest opisany jako losowanie ze zwracaniem. Jednak przypadek nieskończonej populacji nie wyczerpuje tematu testowania proporcji. Gdy populacja jest bardzo mała albo, gdy próbka jest niewiele mniejsza od całej populacji, schemat próbkowania opiera się o losowanie bez zwracania.

Warto zająć się teorią testowania hipotez dla skończonej populacji, ponieważ w określonych przypadkach rozkład hipergeometryczny daje dużo dokładniejszą informację o badanym przypadku niż przybliżenie rozkładem dwumianowym. Ponadto zastosowanie tego typu testów ma duże znaczenie w medycynie, gdzie często rozważane populacje mają na tyle wyspecjalizowane cechy, że są uważane za małe.

W pierwszym rozdziale znajduje się opis schematu pobierania danych w przypadku nieskończonej i skończonej populacji. Następnie są przedstawione testy, bez skończonej poprawki oraz z jej uwzględnieniem. Rozdział 3 zawiera porównanie testów na podstawie prawdopodobieństwa błędu I rodzaju oraz mocy testu. Na końcu pracy znajduje się podsumowanie uzyskanych wyników.

Rozdział 1

Schemat pobierania obserwacji

W niniejszej pracy analizowane są testy, które sprawdzają, czy dwie populacje mają te same proporcje badanej cechy. Zatem, aby wykonać test, potrzebne są próbki z obu populacji. Proporcja jest liczona jako stosunek wartości obserwacji do wielkości próbki, gdzie Wartość obserwacji to ilość osobników z badaną cechą w próbce. Zakładamy również, że populacje są od siebie niezależne, tym samym próbki pochodzące z tych populacji także nie zależą od siebie.

1.1. Nieskończona populacja

Gdy populacja jest bardzo duża, możemy traktować ją jako nieskończoną. Wobec tego losowanie kolejnych elementów próbki jest niezależne, czyli jest to losowanie ze zwracaniem. Zakładamy, że pobierane obserwacje pochodzą z rozkładu Bernoulliego $\mathcal{B}(n,p)$, gdzie n to rozmiar próby z nieskończonej populacji, a p to proporcja zdarzeń sprzyjających w populacji. Funkcja prawdopodobieństwa zmiennej losowej X z rozkładu dwumianowego jest równa

$$b(k;n,p) = P(X=k) = \binom{n}{k} p^k (1-p)^{n-k},\ 0 \leqslant k \leqslant n. \tag{1.1}$$

1.2. Skończona populacja

Kiedy populację rozważamy jako skończoną, kolejne elementy próbki są losowane bez zwracania. Oznacza to, że prawdopodobieństwo sukcesu zmienia się w trakcie pobierania elementów obserwacji w zależności od wyboru poprzednich. W takiej sytuacji próbka pochodzi z rozkładu hipergeometrycznego $\mathcal{H}(n,M,N)$. Przy czym n jest rozmiarem próbki, M ilością osobników w populacji z daną cechą, a N rozmiarem populacji. Zmienna losowa X z rozkładu hipergeometrycznego ma funkcję prawdopodobieństwa określoną wzorem

$$h(k;n,M,N) = P(X=k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\ L \leqslant k \leqslant U, \tag{1.2}$$

gdzie

$$L = \max\{0, M-N+n\}, \quad U = \min\{n, M\}. \tag{1.3}$$

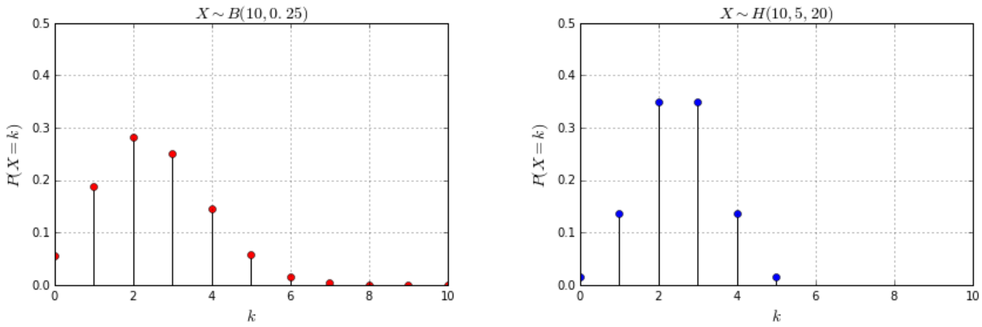
Zauważmy, że wzór (1.2) ma prostą interpretację. Klasyczna definicja prawdopodobieństwa określa szansę zajścia zdarzenia A jako iloraz liczby zdarzeń elementarnych w A przez liczbę zdarzeń elementarnych w Ω, czyli $P(A) = |A|/|\Omega|$. W rozważanym przypadku zdarzeniu A odpowiada sytuacja, w której próbka będzie zawierać k osobników z daną cechą. Zatem ilość zdarzeń elementarnych w A to iloczyn dwóch kombinacji. Wybór k osobników z M posiadających daną cechę $\binom{M}{k}$ mnożymy przez możliwość wyborów pozostałych osobników z reszty populacji $\binom{N-M}{n-k}$. Natomiast zbiór wszystkich zdarzeń elementarnych w Ω to wybór losowej próbki n osobników z N-elementowej populacji $\binom{N}{n}$. Ograniczenia nałożone na k są również naturalne. Dolne ograniczenie L jest równe maksimum z 0 i $M - N + n$. Będzie ono niezerowe, gdy $M - N + n > 0$. Przekształcając nierówność, otrzymujemy $n > N - M$, czyli rozmiar próbki jest większy od ilości osobników w populacji bez badanej cechy. W konsekwencji czego mamy pewność, że w próbce będzie przynajmniej tyle osobników z daną cechą, ile wynosi różnica $n - (N - M)$. Ograniczenie górne jest równe minimum z n i M, co wynika z faktu, że nie może być więcej osób w próbce z daną cechą niż w całej populacji. Analiza wzoru (1.2) pokazuje, że rozkład hipergeometryczny jest ściśle związany z rozmiarem populacji.

1.3. Porównanie rozkładów

Rozkład hipergeometryczny daje bardzo podobne wyniki do rozkładu Bernoulliego, gdy populacja jest duża, albo próbka stosunkowo mała. Natomiast przy małej populacji i dużej próbce różnica między tymi rozkładami jest znaczna.

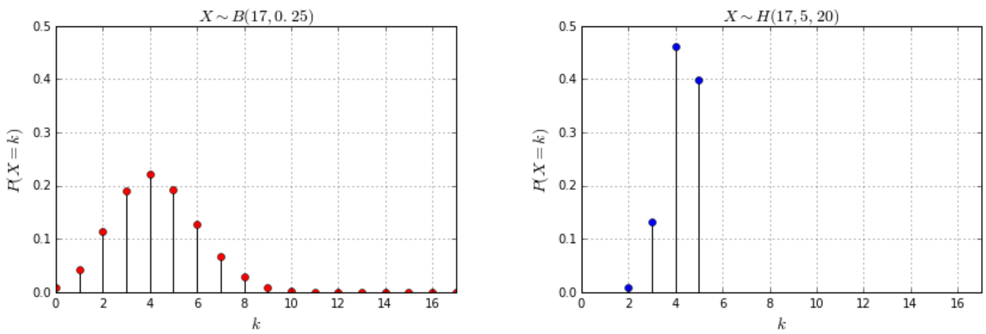
Rozważmy to na medycznym przykładzie. Załóżmy, że jest grupa 20 osób, które są chore na jakąś bardzo rzadką chorobę oraz że 25% z nich ma szanse na wyzdrowienie. Chcemy dowiedzieć się, ile osób spośród przebadanych może wyzdrowieć. Weźmy 3 różne próbki o wielkościach n równych odpowiednio 10, 17, 20. Możemy tę sytuację opisać za pomocą rozkładu Bernoulliego, wtedy badana zmienna losowa jest z rozkładu $\mathcal{B}(n, 0.25)$. Drugim sposobem jest rozkład hipergeometryczny, wtedy zmienna losowa $X \sim \mathcal{H}(n, 5, 20)$. Rysunki 1.1 – 1.3 przedstawiają funkcję prawdopodobieństwa dla wymienionych przypadków.

Przeanalizujmy jakie wartości mogą przyjmować zmienne losowe z obu rozkładów. W rozkładzie dwumianowym zmienna w każdym z przypadków przyjmuje wartości od 0 do n. Na przykład na wykresie 1.1a argumenty funkcji to zbiór $\{0, 1, \ldots, 9, 10\}$. Mimo, że w populacji maksymalnie może wyzdrowieć 5 osób, dla rozkładu Bernoulliego prawdopodobieństwo $X = 6$ wynosi 0.09. Przedstawiona rozbieżność wynika z założenia, że pobieranie próbki to losowanie ze zwracaniem, czyli w próbce mogą znaleźć się dwie te same osoby. Natomiast w rozkładzie hipergeometrycznym zmienną losową ograniczają L i U, zdefiniowane we wzorze (1.3), które uwzględniają wielkość próbki. Porównując, na wykresie 1.1b zmienna losowa nie przyjmuje wartości większej niż 5. Dodatkowo na wykresie 1.2b najmniejszą wartością jest 2, ponieważ populacja zawiera 15 osób, które nie wyzdrowieją, zatem w 17-to osobowej próbce jest pewne, że przynajmniej dwie będą zdrowe. Kolejne losowania są od siebie zależne, więc nie możemy drugi raz wylosować tej samej osoby.



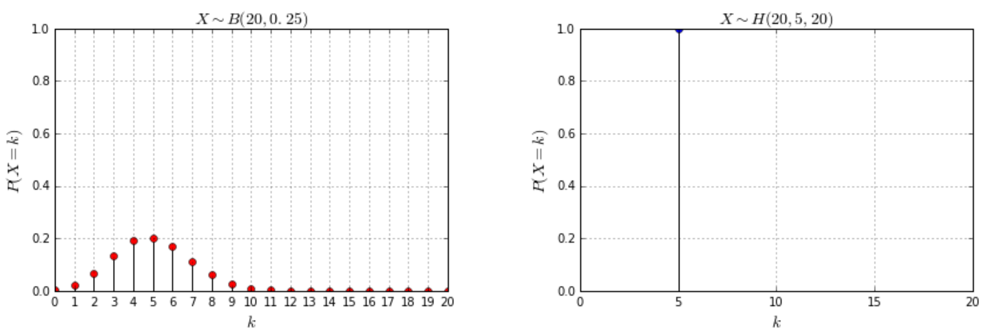
(a) (b)

Rysunek 1.1: Funkcje prawdopodobieństwa $b(k; 10, 0.25)$ oraz $h(k; 10, 5, 20)$



(a) (b)

Rysunek 1.2: Funkcje prawdopodobieństwa $b(k; 17, 0.25)$ oraz $h(k; 17, 5, 20)$



(a) (b)

Rysunek 1.3: Funkcje prawdopodobieństwa $b(k; 20, 0.25)$ oraz $h(k; 20, 5, 20)$

Różnica między funkcjami obu rozkładów rośnie wraz ze wzrostem próbki. Skrajny przypadek przedstawia rysunek 1.3, na którym próbka równa się populacji. Wykres 1.3b idealnie obrazuje przypadek $n = N$. Pewne jest to, że w 20 osobach będzie dokładnie 5, które mogą wyzdrowieć. Natomiast na wykresie 1.3a prawdopodobieństwo, że $X = 5$ wynosi jedynie 0.2, ponieważ funkcja prawdopodobieństwa dla rozkładu dwumianowego nie uwzględnia wielkości populacji. Warto również zaznaczyć, że im większa próbka, tym $P(X = 5)$ dla rozkładu Bernoulliego jest coraz mniejsze, ponieważ mamy więcej elementów w obserwacji. Tymczasem dla rozkładu hipergeometrycznego jest odwrotnie, funkcja prawdopodobieństwa dla $k = 5$ rośnie, aż w końcu osiąga wartość 1. Gdy przebadamy więcej osobników, wzrasta nasza wiedza o próbce oraz jest bardziej prawdopodobne to, że znajdzie się w niej aż 5 zdrowych pacjentów.

Spójrzmy jeszcze, jak wyglądają średnia i wariancja dla rozważanych rozkładów.

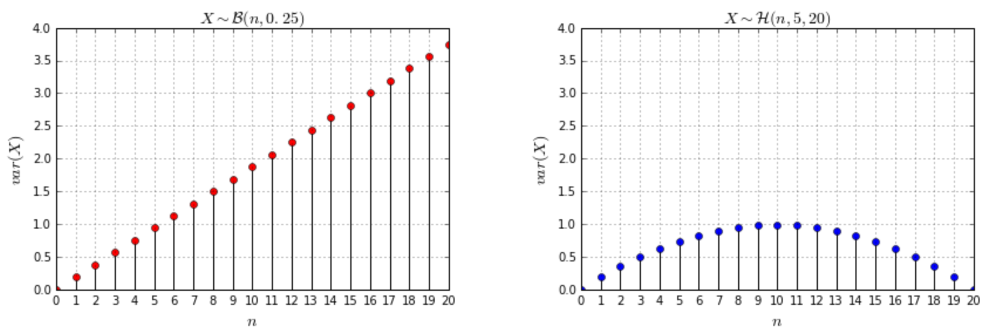
Gdy $X \sim \mathcal{B}(n, p)$, to

$$E(X) = np, \quad Var(X) = np(1 - p). \tag{1.4}$$

Podczas gdy $X \sim \mathcal{H}(n, M, N)$, to

$$E(X) = n\frac{M}{N}, \quad Var(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N - n}{N - 1}. \tag{1.5}$$

Parametr p rozkładu dwumianowego odpowiada ilorazowi parametrów M/N w rozkładzie hipergeometrycznym. Wobec tego średnie obu rozkładów są sobie równe, ale wariancje różni dodatkowy składnik w rozkładzie hipergeometrycznym $(N - n)/(N - 1)$. Przeanalizujmy, jak ten czynnik wpływa na zróżnicowanie rozkładów. Rysunek 1.4 przedstawia wykresy wariancji dla rozważanego przykładu w zależności od wielkości obserwacji. W przypadku rozkładu dwumianowego wariancja stale rośnie wraz ze wzrostem próbki. Ostatecznie, gdy $n = 20$ jest ona największa. Jednakże, gdy przebadamy całą populację, oczekiwanym rezultatem jest wartość zerowa wariancji, ponieważ nie ma wtedy losowości. Taki wynik daje nam wykres 1.4b. Funkcja na początku rośnie, ale gdy wielkość próbki przekroczy połowę rozmiaru populacji, wariancja zaczyna maleć aż do zera. Odzwierciedla to fakt, że gdy coraz więcej wiemy o populacji, losowość uzyskanych wyników maleje.



(a) (b)

Rysunek 1.4: Wariancja rozkładów Bernoulliego i hipergeometrycznego w zależności od rozmiaru próbki

Rozdział 2

Przedstawienie testów

W niniejszym rozdziale znajduje się opis trzech testów. Dwa testy ze skończoną poprawką wykorzystują rozkład hipergeometryczny, a trzeci, test bez skończonej poprawki, opiera się o rozkład dwumianowy. Następnie omówiony jest sposób liczenia mocy dla wymienionych testów.

2.1. Sformułowanie problemu

Załóżmy, że X_1 i X_2 są niezależnymi zmiennymi losowymi. Zaobserwowane wartości X_1 i X_2 oznaczmy odpowiednio k_1 i k_2 oraz proporcje w obserwacjach p_1 i p_2. Będziemy testować

$$H_0\colon p_1 = p_2 \quad \text{przeciwko} \quad H_1\colon p_1 \neq p_2, \tag{2.1}$$

na podstawie wartości obserwacji i znanych parametrów populacji. Rozważmy unormowaną statystykę

$$Z_{X_1,X_2} = \frac{X_1/n_1 - X_2/n_2}{\sqrt{V_{X_1,X_2}}}, \tag{2.2}$$

gdzie V_{X_1,X_2} to estymator wariancji rozkładu zmiennej losowej $X_1/n_1 - X_2/n_2$, pod warunkiem prawdziwości H_0, w połączonej próbie. Jego wzór zależy od rozkładu, z którego pochodzą zmienne losowe X_1 i X_2. Wartość statystyki Z_{X_1,X_2} oznaczmy jako Z_{k_1,k_2}. Jest ona wyliczana według powyższych wzorów, poprzez zamienienie zmiennych losowych X_1 i X_2 odpowiednio ich wartościami k_1 i k_2.

2.2. Testy ze skończoną poprawką

Jak już było wspomniane w podrozdziale 1.2, próbki w przypadku skończonej populacji pochodzą z rozkładu hipergeometrycznego. Zatem $X_1 \sim \mathcal{H}(n_1, M_1, N_1)$, $X_2 \sim \mathcal{H}(n_2, M_2, N_2)$ oraz proporcje są równe $p_1 = M_1/N_1$, $p_2 = M_2/N_2$. Znane parametry to rozmiary próbek n_1 i n_2 i wielkości populacji N_1 i N_2.

W celu wyprowadzenia wariancji rozkładu $X_1/n_1 - X_2/n_2$ pod warunkiem $p_1 = p_2$ w połączonej próbie, zapiszmy wariancję rozważanej zmiennej losowej w łącznej próbie, korzystając z własności wariancji oraz tego, że $Cov(X_1, X_2) = 0$ z niezależności X_1 i X_2

$$\begin{aligned} Var(X_1/n_1 - X_2/n_2) =& Var(X_1/n_1) + Var(X_2/n_2) = \\ =& Var(X_1)/n_1^2 + Var(X_2)/n_2^2. \end{aligned} \tag{2.3}$$

Wariancje X_1 i X_2 są równe

$$Var(X_1) = n_1p_1(1-p_1)(N_1-n_1)/(N_1-1), \tag{2.4}$$

$$Var(X_2) = n_2p_2(1-p_2)(N_2-n_2)/(N_2-1). \tag{2.5}$$

Pamiętając, że zakładamy równość $p_1 = p_2$ zastąpmy oba parametry jednym równym p. Po podstawieniu otrzymujemy

$$\begin{aligned} V_{X_1,X_2} &= \frac{1}{n_1}p(1-p)\frac{N_1-n_1}{N_1-1} + \frac{1}{n_2}p(1-p)\frac{N_2-n_2}{N_2-1} = \\ &= p(1-p)\left(\frac{N_1-n_1}{n_1(N_1-1)} + \frac{N_2-n_2}{n_2(N_2-1)}\right), \end{aligned} \tag{2.6}$$

przy czym p to proporcja liczby osobników z daną cechą do całości populacji w rozkładzie łącznym. Wobec czego $p = (X_1 + X_2)/(n_1 + n_2)$. Ostatecznie otrzymujemy

$$V_{X_1,X_2} = \left(\frac{N_1-n_1}{n_1(N_1-1)} + \frac{N_2-n_2}{n_2(N_2-1)}\right)\left(\frac{X_1+X_2}{n_1+n_2}\right)\left(1 - \frac{X_1+X_2}{n_1+n_2}\right). \tag{2.7}$$

2.2.1. Test Z

Test Z jest oparty na centralnym twierdzeniu granicznym, według którego, rozważana statystyka Z_{X_1,X_2}, pod warunkiem prawdziwości H_0, jest w przybliżeniu z rozkładu standardowego normalnego $\mathcal{N}(0,1)$. Wtedy p-wartość wyraża się wzorem

$$P(|Z_{X_1,X_2}| \geqslant |Z_{k_1,k_2}|\ |H_0) \approx 2(1 - \Phi(|Z_{k_1,k_2}|)), \tag{2.8}$$

gdzie Φ oznacza dystrybuantę rozkładu $N(0,1)$. Test Z odrzuca hipotezę zerową, gdy p-wartość jest mniejsza od poziomu istotności α.

2.2.2. Test E

Test E opiera się o rzeczywistą p-wartość, która, według artykułu K. Krishnamoorthy i J. Thomson z 2002 roku, jest równa [4]

$$\begin{aligned} P(|Z_{X_1,X_2}| \geqslant |Z_{k_1,k_2}|\ |H_0) &= E_{X_1,X_2}(\mathbb{1}\left(|Z_{X_1,X_2}| \geqslant |Z_{k_1,k_2}|\right)\ |H_0) = \\ &= \sum_{x_1=L_1}^{U_1}\sum_{x_2=L_2}^{U_2} h(x_1; n_1, N_1p, N_1)h(x_2; n_2, N_2p, N_2)\mathbb{1}\left(|Z_{X_1,X_2}| \geqslant |Z_{k_1,k_2}|\right), \end{aligned} \tag{2.9}$$

gdzie E_{X_1,X_2} to wartość oczekiwana łącznego rozkładu (X_1, X_2), a p jest nieznaną wspólną proporcją pod warunkiem H_0. Nie jest możliwe policzenie

p-wartości wprost ze wzoru (2.9), ponieważ nie znamy parametru proporcji p. Krishnamoorthy i Thomson (2002) zaproponowali estymator p-wartości [4]

$$P(|Z_{X_1,X_2}| \geqslant |Z_{k_1,k_2}|\ |H_0) \approx$$
$$= \sum_{x_1=L_{x_1}}^{U_{x_1}} \sum_{x_2=L_{x_2}}^{U_{x_2}} h(x_1; n_1, \hat{M}_1, N_1) h(x_2; n_2, \hat{M}_2, N_2) \mathbb{1}\left(|Z_{X_1,X_2}| \geqslant |Z_{k_1,k_2}|\right), \tag{2.10}$$

przy czym $\hat{p} = (k_1+k_2)/(n_1+n_2)$, $\hat{M}_i = [N_i\hat{p}]$ oraz $L_{x_i} = \max\{0, \hat{M}_i - N_i + n_i\}$, $U_{x_i} = \min\{n_i, \hat{M}_i\}$, $i = 1, 2$. Test odrzuca H_0 wtedy, gdy p-wartość wyliczona według wzoru (2.10) jest mniejsza od poziomu istotności α.

2.2.3. Moc testu

Moc testu to prawdopodobieństwo odrzucenia hipotezy zerowej, gdy jest ona nieprawdziwa. Wobec tego, spośród testów na zadanym poziomie istotności, interesują nas te o najwyższej mocy.

Moc obu testów można wyliczyć, korzystając z funkcji prawdopodobieństwa rozkładu hipergeometrycznego. Dla testu Z pod warunkiem hipotezy alternatywnej H_1 moc, zgodnie z rozważaniami Krishnamoorthy i Thomson (2002), jest równa [4]

$$\sum_{k_1=L_1}^{U_1} \sum_{k_2=L_2}^{U_2} h(k_1; n_1, M_1, N_1) h(k_2; n_2, M_2, N_2) \mathbb{1}\left(|Z_{k_1,k_2}| > z_{1-\alpha/2}\right), \tag{2.11}$$

gdzie $L_i = \max\{0, M_i - N_i + n_i\}$ i $U_i = \min\{n_i, M_i\}$, a $z_{1-\alpha/2}$ oznacza kwantyl rozkładu normalnego standardowego rzędu $1 - \alpha/2$.

Natomiast dla testu E, według Krishnamoorthy i Thomson (2002), moc zdefiniowana jest następująco [4]

$$\sum_{k_1=L_1}^{U_1} \sum_{k_2=L_2}^{U_2} h(k_1; n_1, M_1, N_1) h(k_2; n_2, M_2, N_2) \times$$
$$\times \mathbb{1}\left(\sum_{x_1=L_{x_1}}^{U_{x_1}} \sum_{x_2=L_{x_2}}^{U_{x_2}} h(x_1; n_1, \hat{M}_1, N_1) h(x_2; n_2, \hat{M}_2, N_2) \mathbb{1}\left(|Z_{x_1,x_2}| \geqslant |Z_{k_1,k_2}|\right) \leqslant \alpha \right), \tag{2.12}$$

gdzie parametry są takie same jak we wzorach (2.10) i (2.11).

2.3. Test bez skończonej poprawki

Dla testu bez poprawki na skończony rozmiar populacji, zamiast rozkładu hipergeometrycznego stosujemy dwumianowy, więc X_1 i X_2 są niezależnymi zmiennymi losowymi o rozkładzie Bernoulliego $X_1 \sim \mathcal{B}(n_1, p_1)$, $X_2 \sim \mathcal{B}(n_2, p_2)$. Znane parametry to rozmiary próbek n_1 i n_2.

Wariancję rozkładu $X_1/n_1 - X_2/n_2$, pod warunkiem $p_1 = p_2$, w łącznej próbie możemy wyprowadzić analogicznie jak w podrozdziale 2.2, wychodząc od wariancji rozważanej zmiennej losowej

$$Var(X_1/n_1 - X_2/n_2) = Var(X_1)/n_1^2 + Var(X_2)/n_2^2. \tag{2.13}$$

Wariancje X_1 i X_2 są równe

$$Var(X_1) = n_1 p_1 (1 - p_1), \tag{2.14}$$
$$Var(X_2) = n_2 p_2 (1 - p_2). \tag{2.15}$$

Zastępując p_1 i p_2 jednym parametrem równym p, otrzymujemy

$$V_{X_1,X_2} = p(1-p)/n_1 + p(1-p)/n_2 = p(1-p)(1/n_1 + 1/n_2) \tag{2.16}$$

przy czym $p = (X_1 + X_2)/(n_1 + n_2)$.

2.3.1. Test Zb

Test Zb jest, podobnie jak omówiony wcześniej test E, oparty o estymator p-wartości, który, zgodnie z artykułem Storer i Kim z 1990 roku, jest równy [5]

$$\begin{aligned} P(|Z_{X_1,X_2}| \geqslant |Z_{k_1,k_2}| \; |H_0) &\approx \\ &\approx \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} b(x_1; n_1, \hat{p_1}) b(x_2; n_2, \hat{p_2}) \mathbb{1} \left(|Z_{X_1,X_2}| \geqslant |Z_{k_1,k_2}| \right), \end{aligned} \tag{2.17}$$

gdzie $\hat{p} = (k_1 + k_2)/(n_1 + n_2)$. Test odrzuca hipotezę zerową, gdy p-wartość jest mniejsza od poziomu istotności α.

Moc testu wyraża się wzorem

$$\begin{aligned} &\sum_{k_1=0}^{n} \sum_{k_2=0}^{n} b(k_1; n_1, p_1) b(k_2; n_2, p_2) \times \\ &\qquad \times \mathbb{1} \left(\sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} b(x_1; n_1, \hat{p_1}) b(x_2; n_2, \hat{p_2}) \mathbb{1} \left(|Z_{x_1,x_2}| \geqslant |Z_{k_1,k_2}| \right) \leqslant \alpha \right). \end{aligned} \tag{2.18}$$

Rozdział 3

Analiza testów

W celu porównania testów opisanych w rozdziale 2, napisano programy, które wyliczają prawdopodobieństwo błędu I rodzaju i moc testu oraz generują wykresy dla różnych parametrów.

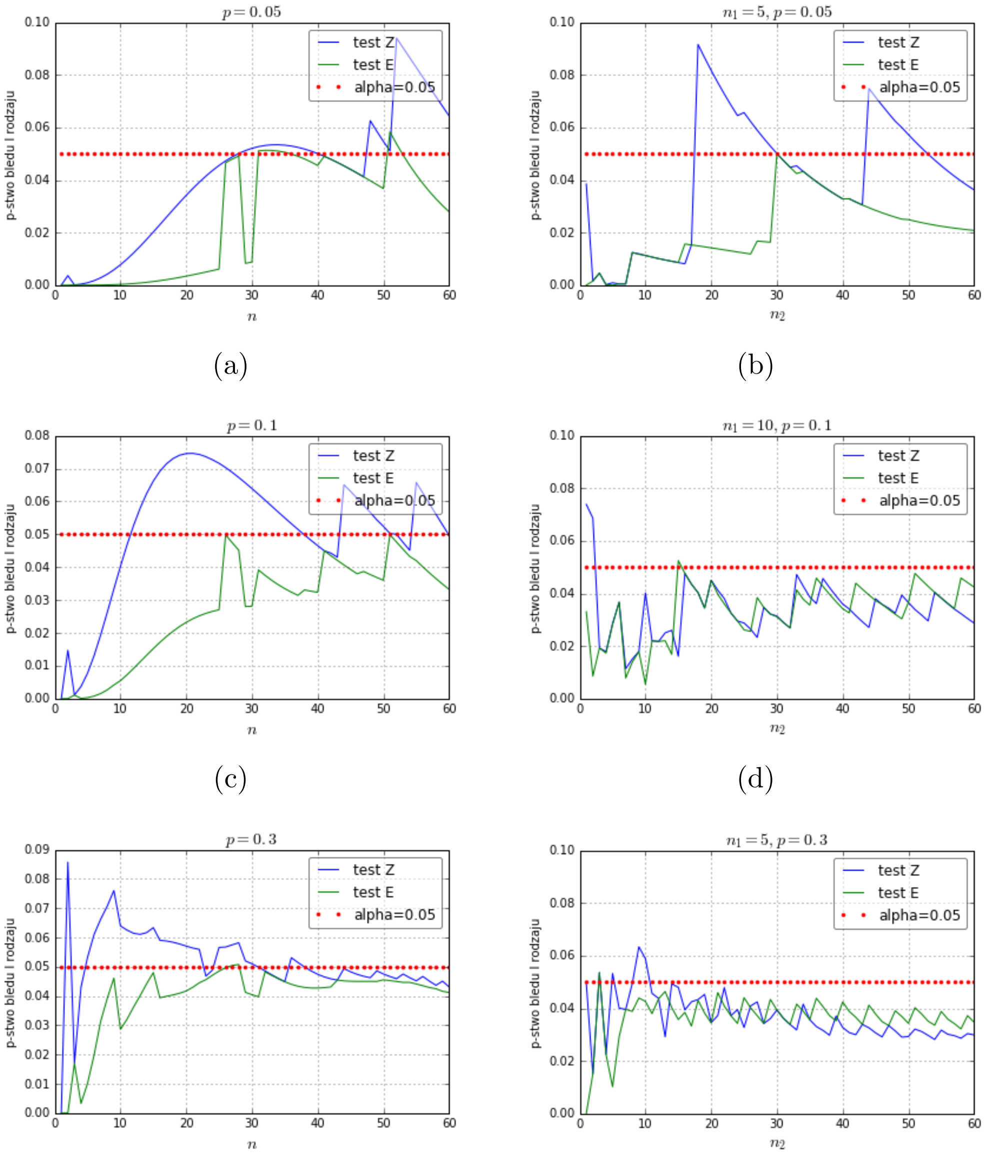
3.1. Porównanie testów ze skończoną poprawką

Na początku zostały porównane testy ze skończoną poprawką. Rysunek 3.1 przedstawia funkcję prawdopodobieństwa błędu I rodzaju w zależności od rozmiaru próbki dla różnych proporcji p. Po lewej stronie obserwacje z obu populacji są tej samej wielkości $n_1 = n_2 = n$, natomiast po prawej różnią się, rozmiar pierwszej próbki ma ustaloną wartość. Pierwszym spostrzeżeniem jest to, że funkcja dla testu E w większości przypadków jest mniejsza niż dla testu Z. Dodatkowo dla testu E prawdopodobieństwo błędu przekracza poziom istotności jedynie kilka razy i to bardzo nieznacznie. Tymczasem dla testu Z funkcja częściej przyjmuje wartości powyżej α, dochodząc nawet do 0.95, jak na wykresie 3.1a. Warto wspomnieć także, że dla większych p (0.1, 0.3) oraz wraz ze wzrostem wielkości próbki funkcje są bliżej siebie i bardziej skupiają się przy poziomie istotności α.

Rysunek 3.2 obrazuje również prawdopodobieństwo błędu I rodzaju, ale w zależności od proporcji p. Analizując wykresy funkcji, można dojść do analogicznych wniosków, jak dla rysunku 3.1. W każdym przypadku prawdopodobieństwo błędu I rodzaju testu Z jest większe od prawdopodobieństwa błędu testu E. Oprócz tego funkcja dla testu E przekracza poziom istotności jedynie na wykresie 3.2f, podczas gdy funkcja błędu testu Z tylko w jednym przypadku (3.2b) pozostaje całkowicie poniżej α.

Podsumowując, dla małych rozmiarów próbek test Z nie jest na poziomie istotności α, co wynika z zastosowania centralnego twierdzenia granicznego w wyliczaniu p-wartości. Przybliżenie rozkładem normalnym działa dobrze tylko dla dużych prób. Wobec czego test Z zbyt często odrzuca hipotezę zerową, gdy jest ona prawdziwa.

Rysunek 3.3 zawiera wykresy mocy obu testów w zależności od rozmiaru próbki dla różnych proporcji oraz rozmiarów populacji. Druga populacja ma ustaloną proporcję p_2, a pierwsza przyjmuje trzy wartości różne od p_2. Dla obu testów zauważalny jest wzrost mocy wraz ze zwiększaniem się rozmiaru próbki. Ponadto moc testu wzrasta także dla proporcji bardziej oddalonych od siebie. W każdym przypadku, dla próby $n = 50$ i różnicy między proporcjami 0.3, moce mają wartość bliską 1.



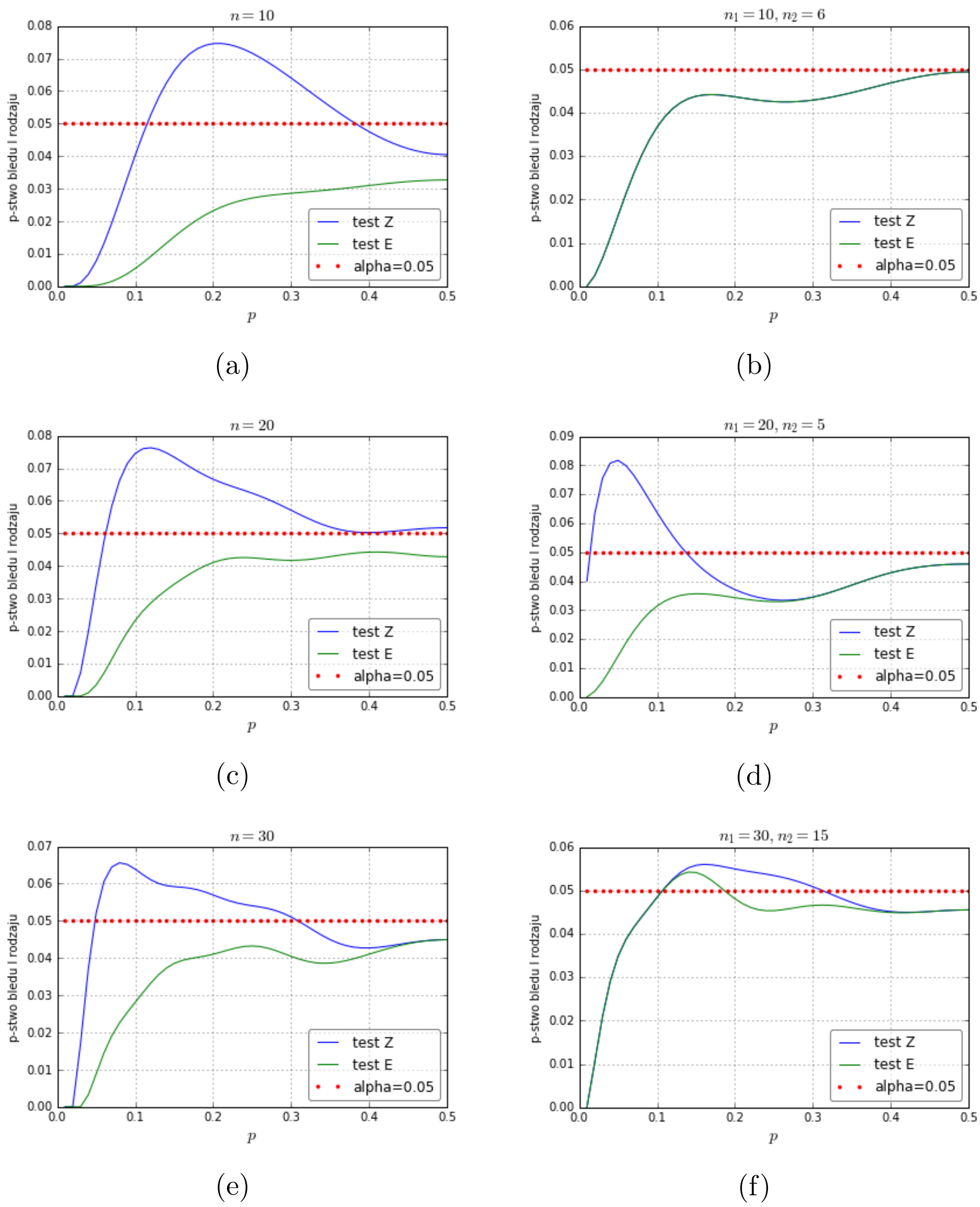
(a) (b)

(c) (d)

(e) (f)

Rysunek 3.1: Prawdopodobieństwo błędu I rodzaju testów Z i E jako funkcja rozmiaru próbki n, przy zadanym poziomie istotności $\alpha = 0.05$; $N_1 = N_2 = 100$

Porównując, moc testu Z jest większa od mocy testu E, jednakże różnice nie są znaczne. Ponadto, z analizy błędu I rodzaju, wiemy, że test Z zbyt często odrzuca H_0, co wpływa na wyższą moc testu. Natomiast test E utrzymuje poziom istotności α, wobec czego niewiele mniejsza moc testu nie oznacza, że test E jest gorszy. Wprost przeciwnie, patrząc na wartości prawdopodobieństwa błędu I rodzaju, test E jest poprawny, podczas gdy test Z nie spełnia swojej roli.

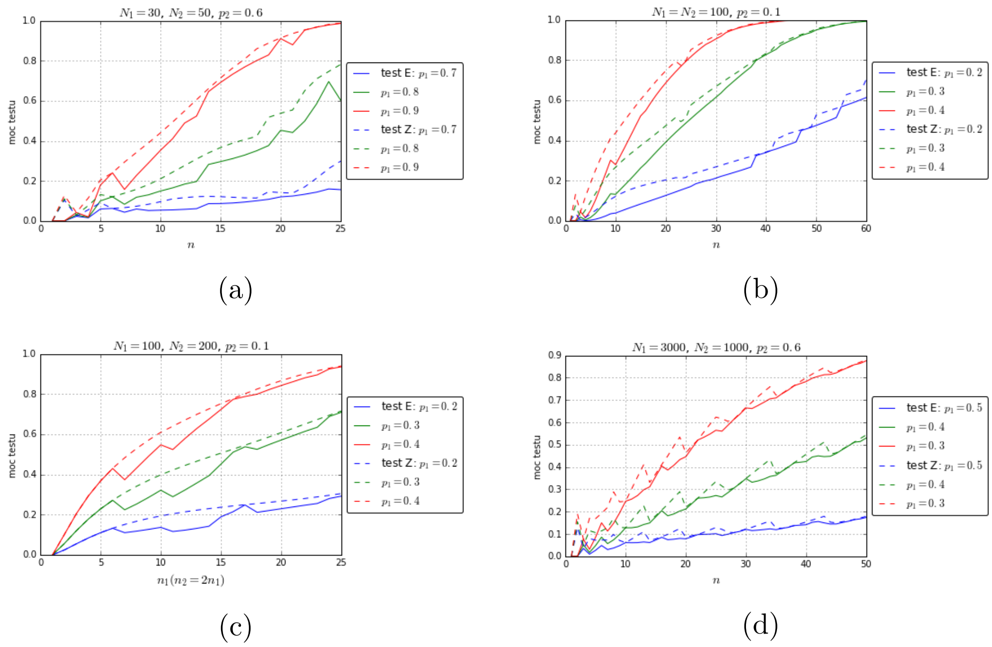


Rysunek 3.2: Prawdopodobieństwo błędu I rodzaju testów Z i E jako funkcja proporcji $p = M_1/N_1 = M_2/N_2$, przy zadanym poziomie istotności $\alpha = 0.05$; $N_1 = N_2 = 100$

3.2. Porównanie testu bez skończonej poprawki Zb z testem E

W niniejszym podrozdziale porównano test Zb z testem E. Nie włączono do analizy testu Z, z powodu wniosków opisanych w podrozdziale 3.1. Niniejsze rozważania zostały przeprowadzone, aby potwierdzić zasadność stosowania testu ze skończoną poprawką dla małych populacji lub o stosunkowo dużej próbie.

Na rysunku 3.4 znajdują się wykresy prawdopodobieństwa błędu I rodzaju analizowanych testów. Test E, jak stwierdzono w podrozdziale 3.1, nie



(a) (b) (c) (d)

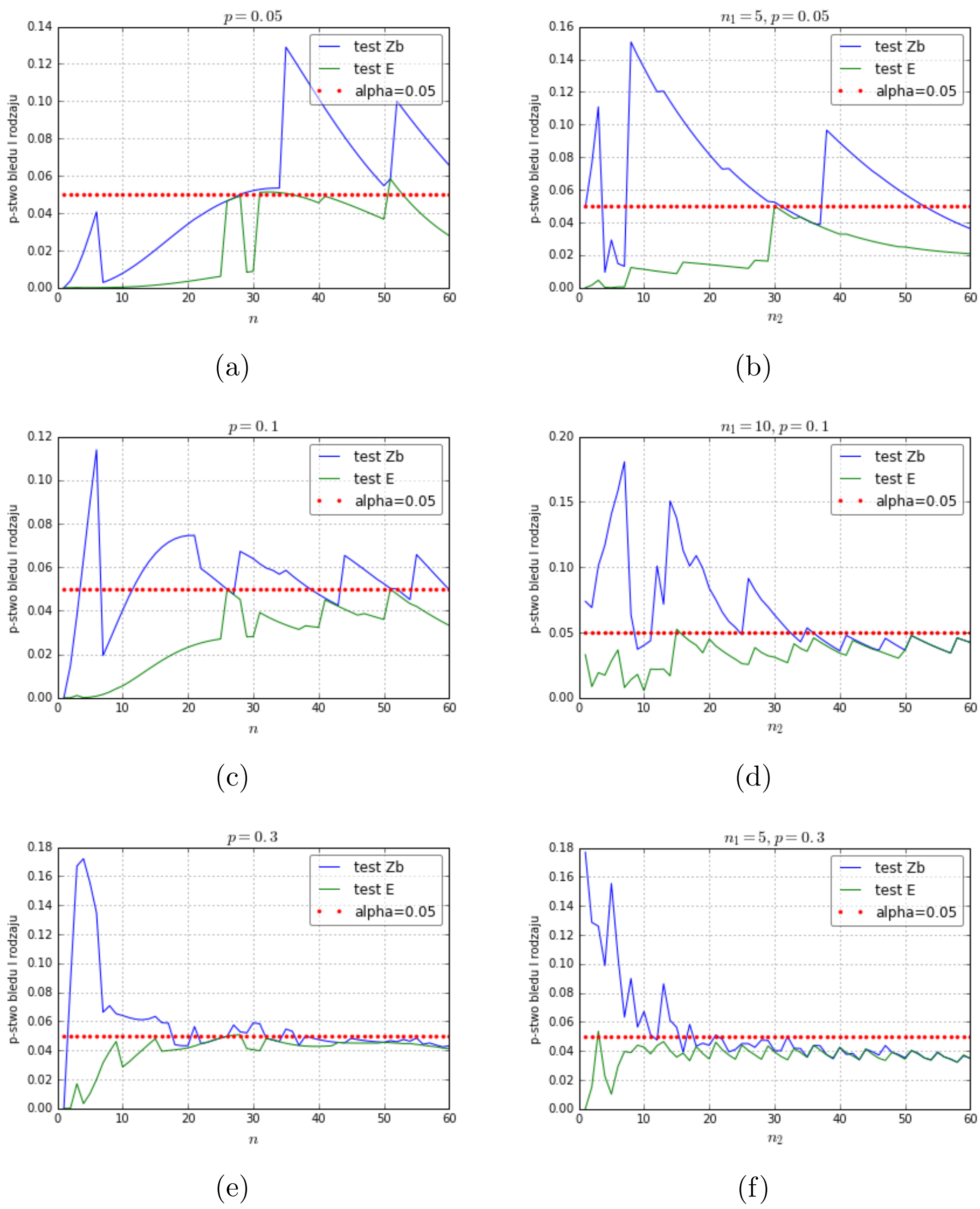
Rysunek 3.3: Moc testów Z i E jako funkcja rozmiaru próbki n

przekracza poziomu istotności α. Jednakże dla testu Zb prawdopodobieństwo błędu I rodzaju znacząco przekracza poziom istotności. Funkcja błędu Zb przyjmuje wartości bliskie 0.2, gdzie największa znajduje się na wykresie 3.4d. Warto wspomnieć także, że funkcja jest największa w przypadku małych próbek. Na wykresach 3.4e i 3.4f widać, że mniej więcej do $n = 20$ test Zb jest powyżej poziomu α, następnie oscyluje, podobnie jak test E, w okolicach poziomu istotności.

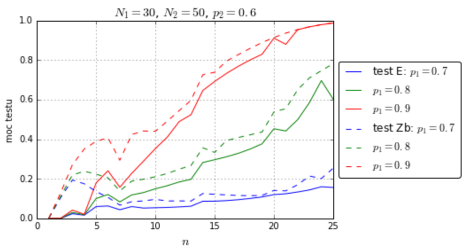
Podsumowując, test Zb nie jest adekwatny do rozważanego przypadku gdy wartość n jest mała. Odrzuca hipotezę zerową dużo częściej, niż powinien test na poziomie α. Jest to spowodowane użyciem rozkładu dwumianowego w formułowaniu statystyki testowej. Faktycznie dla dużych próbek różnica jest nieznaczna, natomiast w sytuacji małej próbki rozbieżności są bardzo widoczne.

Rysunek 3.5 przedstawia wykresy mocy rozważanych testów w zależności od rozmiaru próbki dla różnych rozmiarów populacji oraz proporcji p. Moc testu Zb jest niewiele większa od mocy testu E. Jednakże, podobnie jak w przypadku testu Z, wynika to ze zbyt częstego odrzucania H_0. Biorąc pod uwagę ten fakt wyższa wartość testu Zb nie świadczy o jego poprawności.

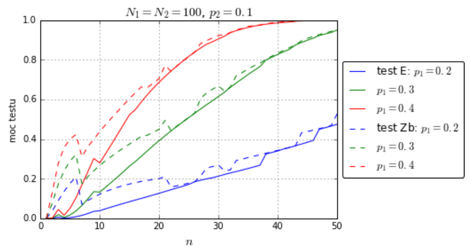
Ostatecznie pokazano, że test bez skończonej poprawki nie powinien być stosowany w przypadku małej populacji, ponieważ wtedy znacznie przekracza poziom istotności α. Najlepszym testem ze skończoną poprawką okazał się test E, który dla różnych przypadków utrzymuje poziom α.



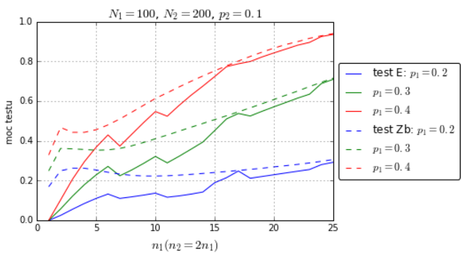
Rysunek 3.4: Prawdopodobieństwo błędu I rodzaju testów Zb i E jako funkcja rozmiaru próbki n, przy zadanym poziomie istotności $\alpha = 0.05$; $N_1 = N_2 = 100$



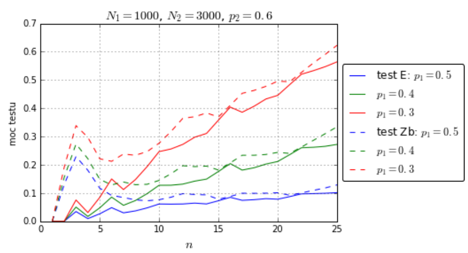
(a)



(b)



(c)



(d)

Rysunek 3.5: Moc testów Zb i E jako funkcja rozmiaru próbki n

Podsumowanie

W pracy został rozważony problem testowania hipotez w sytuacji populacji skończonego rozmiaru. W rozdziale 1 przedstawione są schematy pobierania obserwacji. W sytuacji nieskończonej populacji schemat opiera się o rozkład dwumianowy i próbka jest losowana ze zwracaniem. Natomiast, gdy populacja jest skończona, pobieranie obserwacji modeluje rozkład hipergeometryczny, przy czym elementy próbki są losowane bez zwracania. Po analizie obu rozkładów na konkretnym przykładzie sformułowano wnioski, że rozważane rozkłady dają podobne wyniki, jeśli populacja jest duża albo próbka stosunkowo mała. Jednakże, gdy obserwacja stanowi znaczną część populacji, albo sama populacja jest mała, to różnica między rozkładami staje się bardzo widoczna. Dzieje się tak, ponieważ rozkład hipergeometryczny uwzględnia wielkość populacji, a rozkład Bernoulliego nie przechowuje takiej informacji.

Rozdział 3 zawiera analizę porównawczą testów opisanych w rozdzialne 2 na podstawie wykresów prawdopodobieństwa błędu I rodzaju i mocy testu dla różnych parametrów testowanych populacji. Porównanie dwóch testów ze skończoną poprawką wskazało na to, że test Z nie utrzymuje poziomu istotności α, ze względu na zastosowanie centralnego twierdzenia granicznego do statystyki testowej. Natomiast test E nie wykracza znacząco powyżej poziomu istotności oraz jego moc jest niewiele mniejsza od mocy testu Z. Tym samym, test E możemy uznać za dobry w przypadku małej populacji.

W celu pokazania zasadności stosowania testu ze skończoną poprawką do małych populacji został także porównany test E z testem bez skończonej poprawki. Na wykresach przedstawiających błąd I rodzaju możemy zauważyć, że test Zb znacząco przekracza poziom istotności α, szczególnie dla małych próbek. Oznacza to, że zbyt często odrzuca hipotezę zerową. Wobec czego stosowanie testu bez skończonej poprawki w sytuacji małej populacji wiąże się z dużymi błędami.

Problem skończonej populacji często występuje w medycynie. Lekarze badają przypadki, w których populacja jest mała, ze względu na konkretne cechy badanych osób. We Wrocławiu w Ośrodku Badawczo-Rozwojowym przy Wojewódzkim Szpitalu Specjalistycznym jeden z kardiologów przeprowadza badania na populacji dzieci z chorym sercem po operacji. Chce on sprawdzić jakie czynniki wpływają na wystąpienie powikłań pooperacyjnych u pacjentów. Niestety nie można na obecną chwilę przeprowadzić analizy danych, ponieważ badanie nie zostało jeszcze ukończone. Jednakże, aby sprawdzić, który czynnik jest najbardziej powiązany z powikłaniami, można zastosować opisany w pracy test E. Powinien dać on najlepsze wyniki spośród trzech testów przedstawionych w pracy.

Spis rysunków

Bibliografia

[1] A. Hald. *A History of Probability and Statistics and Their Applications before 1750*. Wiley Interscience, 2003.

[2] E. L. Lehmann. *Testowanie hipotez statystycznych*. PWN, 1968.

[3] F. Y. Edgeworth. On the value of a mean as calculated from a sample. *Journal of the Royal Statistical Society*, 81(4):624–632, Jul. 1918.

[4] K. Krishnamoorthy i Jessica Thomson. Hypothesis testing about proportions in two finite populations. *The American Statistician*, 56(1):215–222, 2002.

[5] Barry E. Storer i Choongrak Kim. Exact properties of some exact test statistics for comparing two binomial proportions. *Journal of the American Statistical Association*, 85:146–155, 1990.