

**ZESZYTY NAUKOWE NR 2 (74)
AKADEMII MORSKIEJ
W SZCZECINIE**

EXPLO-SHIP 2004

Lech Kasyk

**Rozkład prawdopodobieństwa czasu oczekiwania
na zgłoszenie statku na torze wodnym Szczecin – Świnoujście**

Słowa kluczowe: zmienna losowa, rozkład wykładniczy, proces Poissona

Wykorzystując test „chi kwadrat” Pearsona zweryfikowano hipotezę o tym, że zmienna losowa, którą jest czas pomiędzy zgłoszeniami kolejnych jednostek płynących torem wodnym Szczecin – Świnoujście ma rozkład wykładniczy. Dane z 2002 roku analizowano w trzech zbiorach: dla dni świątecznych, okołoswiątecznych i dla dni roboczych. Ponadto zweryfikowano hipotezę o rozkładzie wykładniczym czasu oczekiwania dla zgłoszeń w poszczególnych porach dnia wyróżnionych okresów.

**Probability Distribution of the Waiting Times for a Ship
Report in the Szczecin – Świnoujście Fairway**

Key words: random variable, exponential distribution, Poisson process

The author has tested a hypothesis concerning the exponential distribution of the waiting times for the reporting of the next fairway vessel in the y Szczecin – Świnoujście fairway. The “chi square” Pearson test has been used. The relevant 2002 data have been examined in three sets: for weekdays, for holidays and for days immediately preceding or following holidays (“near-holidays”).

Wstęp

Czas oczekiwania na zgłoszenie kolejnej jednostki idącej torem wodnym jest zmienną losową. W wielu opracowaniach [2, 3, 6, 7, 9] zakłada się, że proces zgłoszeń statków jest procesem Poissona, a co za tym idzie czas oczekiwania na zgłoszenie kolejnej jednostki torowej jest zmienną losową o rozkładzie wykładniczym. Założenie to jest słuszne pod warunkiem, że proces zgłoszeń jest całkowicie losowy. Na torze wodnym są pewne regulacje, które tę losowość zaburzają (np. wymóg minimalnej, bezpiecznej odległości między jednostkami postępującymi kolejno po sobie). W związku z tym rozkład prawdopodobieństwa tej zmiennej może być różny w zależności od akwenu i warunków panujących na nim. Znajomość tego rozkładu jest niezbędna w wielu modelach matematycznych ruchu statków na akwenach ograniczonych [2, 4, 6]. Do weryfikacji hipotezy o rozkładzie wykładniczym zmiennej losowej, którą jest czas pomiędzy zgłoszeniami kolejnych jednostek na torze wodnym Szczecin – Świnoujście, na linii raportowej Police, zastosowano test χ^2 Pearsona [1, 3, 5]. W roku 2002 na punkcie raportowym Police zarejestrowano 7374 zgłoszenia jednostek idących do lub ze Szczecina (na podstawie danych VTS z Urzędu Morskiego w Szczecinie). W niniejszym artykule badano sam proces zgłoszeń jednostek torowych, nie analizując strumieni ruchu statków w poszczególnych kierunkach. Analiza szeregu czasowego liczby zgłoszeń w poszczególnych okresach, wykazała istotne różnice w rozkładzie liczby zgłoszeń w zależności od miesiąca, dnia tygodnia a także od pory dnia [4, 8].

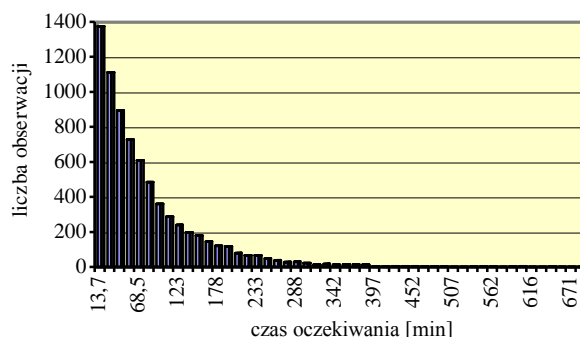
1. Model całoroczny

W całym roku 2002 zarejestrowano 7374 zgłoszenia, a co za tym idzie tyle samo wartości zmiennej losowej, które ułożono w szereg rozdzielczy o 86 klasach. Zastosowano tu wzór na ilość klas: $k \approx \sqrt{n}$, gdzie n oznacza licznosc próbki [5]. W związku z tym, że wiele klas posiadało mniej niż 5 elementów, połączono je i ostatecznie otrzymano 36 klas. Histogram częstości występowania poszczególnych wartości zmiennej przedstawia rysunek 1.

W teście χ^2 Pearsona otrzymano następujące wyniki:

- wartość sprawdzianu: $U_n = 169,22$;
- wartość krytyczna na poziomie ufności 0,95: $W_k = 49,8$.

Wartość sprawdzianu jest zdecydowanie większa od wartości krytycznej, więc należy odrzucić hipotezę o rozkładzie wykładniczym czasu oczekiwania na zgłoszenie kolejnej jednostki torowej. Średni czas oczekiwania wynosi 71,1 minut.



Rys. 1. Histogram częstości zmiennej losowej dla całego roku
 Fig. 1. Frequency histogram of the random variable for the whole year

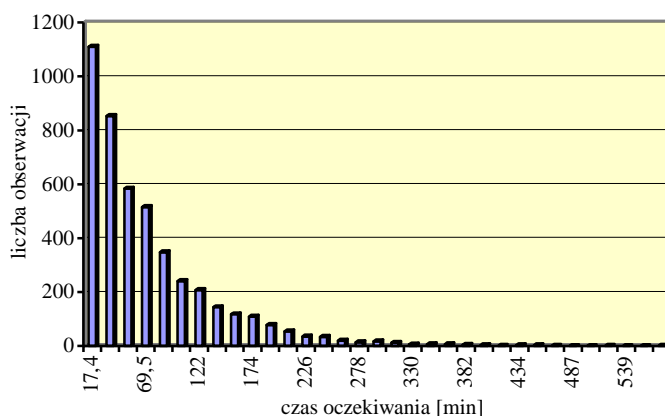
Wyniki testu mogą wydawać się zaskakujące, gdyż słupki histogramu układają się bardzo wyraźnie w kształcie krzywej wykładniczej. Jednak przy analizie wyników obliczeń cząstkowych w teście χ^2 Pearsona widać, w których klasach jest największa różnica między częstościami teoretycznymi i faktycznymi. Wartość testu różni się od wartości krytycznej o prawie 120 jednostek, z tego 116 przypada na 3 ostatnie klasy, które łącznie tworzą przedział $(534,16; \infty)$. Różnica wynika stąd, iż teoretycznie przy średniej 71,1 w powyższym przedziale powinno znajdować się najwyżej kilka wartości zmiennej losowej, a w rzeczywistości jest ich aż 22. Skąd taka częstość występowania zdarzeń teoretycznie mało prawdopodobnych? 17 spośród tych zdarzeń miało miejsce w dni świąteczne lub okołoswiąteczne (bezpośrednio przed lub bezpośrednio po dniu świątecznym), co wskazuje na zróżnicowanie rozkładu zmiennej losowej w dniach roboczych i świątecznych.

2. Model z podziałem na dni robocze, świąteczne i okołoswiąteczne

Szczegółowa analiza statystyczna [8] wykazała istotne różnice w rozkładzie liczby zgłoszeń w poszczególnych dniach tygodnia, okazało się, że w okolicach niedzieli jest najmniej zgłoszeń, a we wtorek, środę, czwartek i piątek jest ich zdecydowanie więcej. W związku z tym, hipotezę o typie rozkładu zmiennej losowej zweryfikowano oddzielnie dla dni roboczych – wtorek, środę, czwartek i piątek, oddzielnie dla dni sąsiadujących z niedzielą – sobota, poniedziałek oraz oddzielnie dla dni świątecznych.

2.1. Dni robocze

W wymienione dni robocze zanotowano 4535 danych, które ułożono ostatecznie w szereg rozdzielczy o 24 klasach. Histogram częstości występowania poszczególnych wartości zmiennej przedstawia rysunek 2.



Rys. 2. Histogram częstości zmiennej losowej dla dni roboczych
Fig. 2. Frequency histogram of the random variable for weekdays

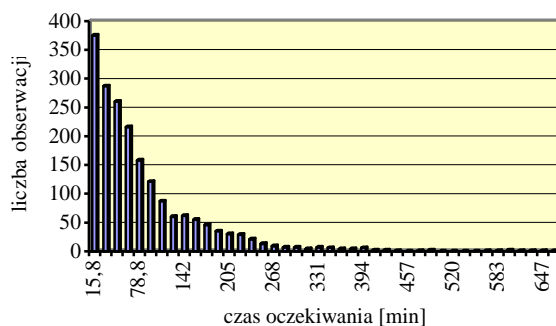
W teście χ^2 Pearsona otrzymano następujące wyniki:

- wartość statystyki testowej: $U_n = 26,36$;
- wartość krytyczna na poziomie ufności 0,95: $W_k = 35,17$.

Wartość statystyki testowej jest zdecydowanie mniejsza od wartości krytycznej, więc nie ma podstaw do odrzucenia hipotezy o rozkładzie wykładniczym czasu oczekiwania na zgłoszenie kolejnej jednostki torowej. Średni czas oczekiwania wynosi 63,78 minut. Parametr rozkładu wykładniczego λ , czyli średnia intensywność zgłoszeń na minutę wynosi 0,0157.

2.2. Dni okoloświąteczne

W soboty i poniedziałki zanotowano 1926 danych, które ułożono w szereg rozdzielczy o 43 klasach. W związku z tym, że wiele klas posiadało mniej niż 5 elementów, połączono je i ostatecznie otrzymano 26 klas. Histogram częstości występowania poszczególnych wartości zmiennej prezentuje rysunek 3.



Rys. 3. Histogram częstości zmiennej losowej dla dni okołoswiątecznych
 Fig. 3. Frequency histogram of the random variable for "near-holidays"

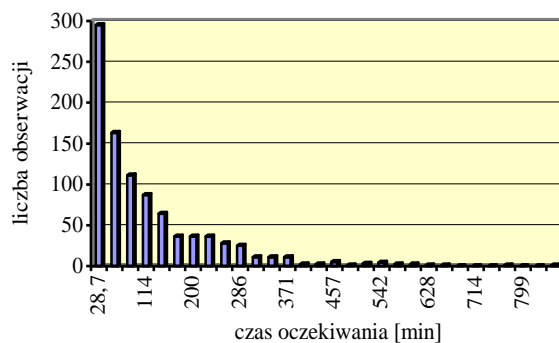
W teście χ^2 Pearsona otrzymano następujące wyniki:

- wartość sprawdzianu: $U_n = 35,89$;
- wartość krytyczna na poziomie ufności 0,95: $W_k = 37,65$.

Wartość sprawdzianu jest mniejsza od wartości krytycznej, więc nie ma podstaw do odrzucenia hipotezy o rozkładzie wykładniczym czasu oczekiwania na zgłoszenie kolejnej jednostki torowej. Średni czas oczekiwania wynosi 75,55 minut. Parametr rozkładu wykładniczego λ , czyli średnia intensywność zgłoszeń na minutę wynosi 0,0132.

2.3. Dni świąteczne

Dni świąteczne to w niniejszym opracowaniu niedziele i dni wolne od pracy, takie jak 1.01, 1.05, 3.05 itd. W niedziele zanotowano 815 wartości a w wymienione dni świąteczne 124, w sumie szereg rozdzielczy zawierał 939 danych, w 30 klasach.



Rys. 4. Histogram częstości zmiennej losowej dla dni świątecznych
 Fig. 4. Frequency histogram of the random variable for holidays

W teście χ^2 Pearsona otrzymano następujące wyniki:

- wartość statystyki testowej: $U_n = 44$;
- wartość krytyczna na poziomie ufności 0,95: $W_k = 25$.

Wartość statystyki testowej jest większa od wartości krytycznej, więc należy odrzucić hipotezę o rozkładzie wykładniczym czasu oczekiwania na zgłoszenie kolejnej jednostki torowej. Średni czas oczekiwania wynosi 98,6 minut.

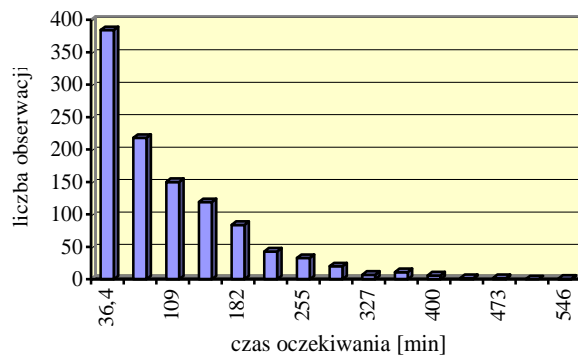
3. Model z wyróżnionymi porami doby w dni robocze

Analiza danych w rozbiciu na poszczególne pory dnia roboczego (okresy dwugodzinne) również wykazała istotną zmienność liczby zgłoszeń [8].

Wyróżniono dwa okresy o zwiększonej intensywności ruchu ($8^{00} - 10^{00}$ i $16^{00} - 18^{00}$), dwa o średniej ($10^{00} - 16^{00}$ i $18^{00} - 24^{00}$) oraz okres „nocny” o zmniejszonej intensywności ruchu (od godziny 0^{00} do 8^{00}).

3.1. Okres nocny

W okresie nocnym (0^{00} do 8^{00}) zanotowano 1081 danych, które ułożono ostatecznie w szereg rozdzielczy o 12 klasach (brano pod uwagę czas, który upłynął pomiędzy zgłoszeniami dwóch kolejnych jednostek, z których druga zgłosiła się już w badanym okresie). Histogram częstości występowania poszczególnych wartości zmiennej przedstawia rysunek 5.



Rys. 5. Histogram częstości zmiennej losowej dla okresu nocnego
 Fig. 5. Frequency histogram of the random variable for the night time

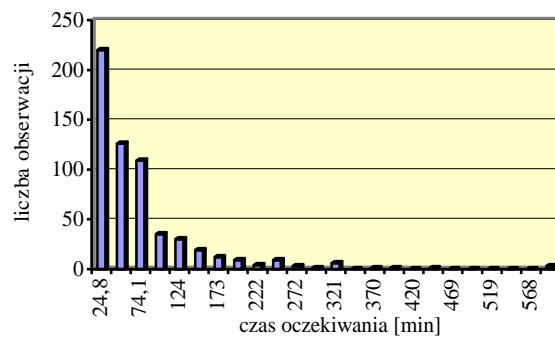
W teście χ^2 Pearsona otrzymano następujące wyniki:

- wartość statystyki testowej: $U_n = 14,85$;
- wartość krytyczna na poziomie ufności 0,95: $W_k = 19,675$.

Wartość statystyki testowej jest mniejsza od wartości krytycznej, więc nie ma podstaw do odrzucenia hipotezy o rozkładzie wykładniczym czasu oczekiwania na zgłoszenie kolejnej jednostki torowej. Średni czas oczekiwania wynosi 86,33 minut. Parametr rozkładu wykładniczego λ , czyli średnia intensywność zgłoszeń na minutę wynosi 0,0116.

3.2. Okres poranny

W okresie od 8⁰⁰ do 10⁰⁰ zanotowano 589 danych, które ułożono w szereg rozdzielczy o 11 klasach. Histogram częstości występowania poszczególnych wartości zmiennej przedstawia rysunek 6.



Rys. 6. Histogram częstości zmiennej losowej dla okresu 8⁰⁰ – 10⁰⁰
Fig. 6. Frequency histogram of the random variable for the 8⁰⁰ – 10⁰⁰ am period

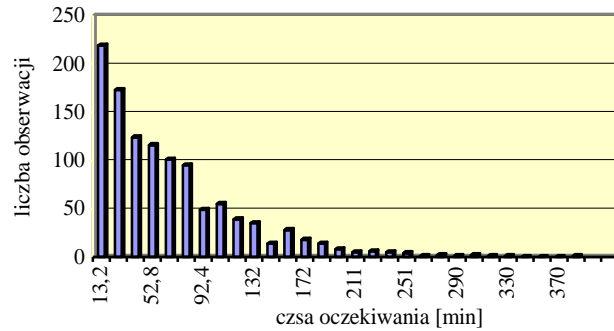
W teście χ^2 Pearsona otrzymano następujące wyniki:

- wartość statystyki testowej: $U_n = 30,92$;
- wartość krytyczna na poziomie ufności 0,95: $W_k = 18,31$.

Wartość statystyki testowej jest zdecydowanie większa od wartości krytycznej, więc należy odrzucić hipotezę o rozkładzie wykładniczym czasu oczekiwania na zgłoszenie kolejnej jednostki torowej. Średni czas oczekiwania wynosi 59,21 minut.

3.3. Okres południowy

W okresie od 10⁰⁰ do 16⁰⁰ zanotowano 1119 dane, które ułożono w szereg rozdzielczy o 22 klasach. Histogram częstości występowania poszczególnych wartości zmiennej prezentuje rysunek 7.



Rys. 7. Histogram częstości zmiennej losowej dla okresu 10⁰⁰ – 16⁰⁰
 Fig. 7. Frequency histogram of the random variable for the 10⁰⁰ am – 4⁰⁰ pm period

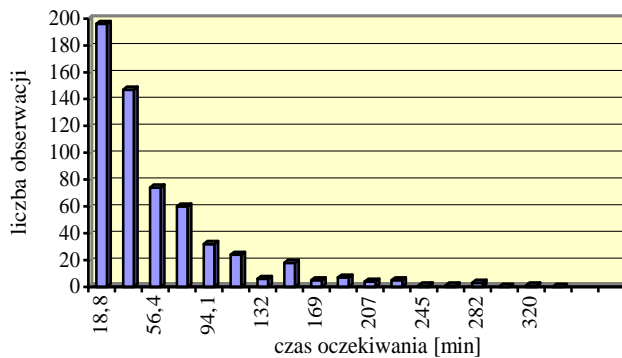
W teście χ^2 Pearsona otrzymano następujące wyniki:

- wartość statystyki testowej: $U_n = 25,57$;
- wartość krytyczna na poziomie ufności 0,95: $W_k = 32,67$.

Wartość statystyki testowej jest mniejsza od wartości krytycznej, więc nie ma podstaw do odrzucenia hipotezy o rozkładzie wykładniczym czasu oczekiwania na zgłoszenie kolejnej jednostki torowej. Średni czas oczekiwania wynosi 59,52 minut. Parametr rozkładu wykładniczego λ , czyli średnia intensywność zgłoszeń na minutę wynosi 0,0168.

3.4. Okres popołudniowy

W okresie od 16⁰⁰ do 18⁰⁰ zanotowano 585 danych, które ułożono w szereg rozdzielczy o 12 klasach. Histogram częstości występowania poszczególnych wartości zmiennej przedstawia rysunek 8.



Rys. 8. Histogram częstości zmiennej losowej dla okresu 16⁰⁰ – 18⁰⁰
 Fig. 8. Frequency histogram of the random variable for the 4⁰⁰ – 6⁰⁰ pm period

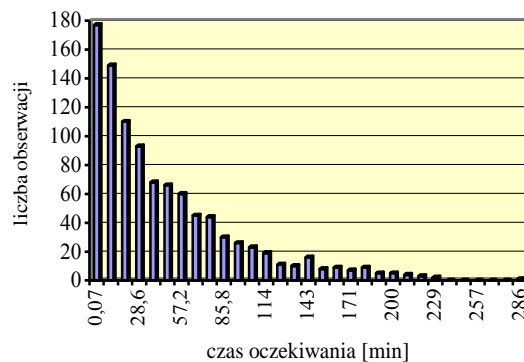
W teście χ^2 Pearsona otrzymano następujące wyniki:

- wartość statystyki testowej: $U_n = 22,52$;
- wartość krytyczna na poziomie ufności 0,95: $W_k = 19,68$.

Wartość statystyki testowej jest większa od wartości krytycznej, więc należy odrzucić hipotezę o rozkładzie wykładniczym czasu oczekiwania na zgłoszenie kolejnej jednostki torowej. Średni czas oczekiwania wynosi 47,34 minut.

3.5. Okres wieczorny

W okresie wieczornym ($18^{00} - 24^{00}$) zanotowano 1161 danych, które ułożono w szereg rozdzielczy o 24 klasach. Histogram częstości występowania poszczególnych wartości zmiennej przedstawia rysunek 9.



Rys. 9. Histogram częstości zmiennej losowej dla okresu $18^{00} - 24^{00}$

Fig. 9. Frequency histogram of the random variable for the $6^{00} - 12^{00}$ pm period

W teście χ^2 Pearsona otrzymano następujące wyniki:

- wartość statystyki testowej: $U_n = 21,78$;
- wartość krytyczna na poziomie ufności 0,95: $W_k = 35,17$.

Wartość statystyki testowej jest zdecydowanie mniejsza od wartości krytycznej, więc nie ma podstaw do odrzucenia hipotezy o rozkładzie wykładniczym czasu oczekiwania na zgłoszenie kolejnej jednostki torowej. Średni czas oczekiwania wynosi 57,51 minut. Parametr rozkładu wykładniczego λ , czyli średnia intensywność zgłoszeń na minutę wynosi 0,0174.

Podsumowanie

Weryfikacja hipotezy o rozkładzie wykładniczym zmiennej losowej, którą jest czas pomiędzy zgłoszeniami kolejnych jednostek na torze wodnym Szczecin – Świnoujście, w okolicach Polic, potwierdziła przypuszczenia, tylko w niektórych badanych okresach, że proces zgłoszeń jednostek torowych jest procesem Poissona. Nie ma podstaw do odrzucenia hipotezy o rozkładzie wykładniczym, gdy rozpatrujemy zgłoszenia globalnie w dni robocze (wtorek, środa, czwartek i piątek) lub w dni okołoswiąteczne (sobota i poniedziałek). Natomiast dla dni świątecznych powyższą hipotezę należy zdecydowanie odrzucić.

W okresach zwiększonej intensywności ruchu statków, proces zgłoszeń nie jest procesem Poissona i do jego opisu trzeba zastosować empiryczne rozkłady prawdopodobieństwa. Związane jest to z tym, że system transportowy znajduje się lub zmierza do stanu nadkrytycznego [10], to znaczy, że przy zwiększonej intensywności ruchu zwiększa się wpływ wszelkich regulacji obowiązujących na danym akwenie, co zaburza losowy charakter procesu zgłoszeń.

Literatura

1. Aczel A.D., *Statystyka w zarządzaniu*, Wydawnictwo Naukowe PWN, Warszawa 2000.
2. *Badania symulacyjne ruchu promów na przeprawie centralnej w Świnoujściu w aspekcie jej optymalnej modernizacji*, praca zbiorowa pod kierunkiem S. Gucmy (sprawozdanie z badań), nie publikowana, INM WSM, Szczecin 1997.
3. Gajek L., Kałużka M., *Wnioskowanie statystyczne*, WNT Warszawa 1996.
4. Gucma L., *Intensywność ruchu statków w porcie Szczecin i na torze wodnym Szczecin – Świnoujście*, Zeszyty Naukowe nr 70, WSM Szczecin 2003.
5. Józwiak J., Podgórski J., *Statystyka od podstaw*, PWE Warszawa 1995.
6. Kasyk L., *Probability model of a ferry crossing*, Materiały z VIII Międzynarodowej Konferencji Naukowo-Technicznej Inżynieria Ruchu Morskiego, WSM Szczecin 1999.
7. Kasyk L., *Prawdopodobieństwo uniknięcia sytuacji kolizyjnej na projektowanej przeprawie promowej Police – Święta, w różnych modelach probabilistycznych*, Materiały z X Międzynarodowej Konferencji Naukowo-Technicznej Inżynieria Ruchu Morskiego, WSM Szczecin 2003.
8. Kasyk L., *Statystyczna analiza liczby zgłoszeń statków na torze wodnym Szczecin – Świnoujście w roku 2002*, Zeszyty Naukowe nr 2(74), AM Szczecin 2004, s. 193 – 204.
9. Montgomery D.C., Runger G.C., *Applied statistics and probability for engineers*, John Wiley & Sons, Inc., New York 1994.

10. Piszczek W., *Modele miar systemu inżynierii ruchu morskiego*, Studia nr 14, WSM Szczecin 1990.

Wpłynęło do redakcji w lutym 2004 r.

Recenzenci

dr inż. kpt.ż.w. Zbigniew Burciu, prof. AM
dr hab. inż. Roman Śmierzchalski, prof. AM

Adres Autora

dr Lech Kasyk
Akademia Morska w Szczecinie
Zakład Matematyki
ul. Wały Chrobrego 1-2, 70-500 Szczecin
tel. 48 09 370