

# KRYTERIUM INFORMACYJNE BAYESA (BIC) W WERYFIKACJI HIPOTEZ STATYSTYCZNYCH DOTYCZĄCYCH PARAMETRÓW STRUMIENIA RUCHU STATKÓW

*W artykule omówiono rozbieżności wyników testów statystycznych, weryfikujących hipotezy dotyczące dopasowania rozkładów prawdopodobieństwa do wartości parametrów strumienia ruchu statków, powstałe w wyniku stosowania kryterium informacyjnego Bayesa i testu chi-kwadrat Pearsona lub Cramera von Misesa. Przeanalizowano testy dotyczące czasu pomiędzy zgłoszeniami kolejnych jednostek oraz prędkości statków na różnych torach wodnych.*

## WSTĘP

Znajomość rozkładu prawdopodobieństwa poszczególnych parametrów strumienia ruchu statków jest często wykorzystywana w modelowaniu strumienia statków, jak również w analizach bezpieczeństwa nawigacji. Weryfikacja hipotezy dotyczącej rozkładu prawdopodobieństwa badanego zjawiska, jest często wykonywana za pomocą różnych, dostępnych pakietów statystycznych. Najczęściej stosowanym testem statystycznym jest test chi kwadrat Pearsona [2, 3, 7, 9]. Jako uniwersalny i nieskomplikowany, często jest podstawowym narzędziem pakietów statystycznych weryfikujących zgodność danych empirycznych z rozkładem teoretycznym [9]. Jednak wielu autorów zwraca uwagę na niedoskonałości tego testu [7]. Stąd w wielu przypadkach stosuje się inne testy, jak np. test Cramera von Misesa, test Andersona-Darlinga [7, 8]. Oprócz większej różnorodności testów, można też wykorzystać kryteria informacyjne, które od wielu lat są stosowane w ocenie dopasowania różnych modeli statystycznych [6].

## 1. KRYTERIA INFORMACYJNE

Kryteria informacyjne służą do wyboru takiego modelu statystycznego, w którym kompromis między prostotą modelu a jakością dopasowania jest najlepszy. W przypadku bayesowskiego kryterium Schwarza, zwanego inaczej kryterium Bayesa, wybiera się model dla którego wartość  $BIC$  jest najmniejsza [6]. Wartość  $BIC$  oblicza się według następującego wzoru:

$$BIC = K \ln(N) - 2 \ln L(\theta) \quad (1)$$

gdzie

- $N$  – liczba obserwacji,
- $K$  – liczba parametrów modelu,
- $L(\theta)$  – funkcja wiarygodności dla oszacowanego wektora parametrów

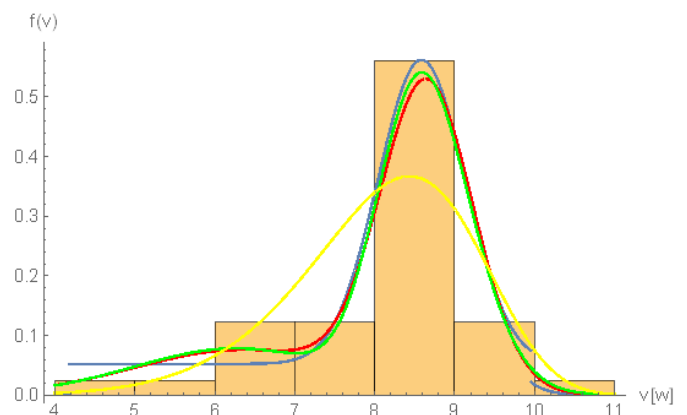
Niektóre pakiety obliczeń symbolicznych, jak np. *Mathematica* posiadają wbudowane procedury dopasowywania rozkładów, z uwzględnieniem kryterium informacyjnego Bayesa. W niniejszym artykule wykorzystano te procedury.

## 2. TESTOWANIE HIPOTEZ DOTYCZĄCYCH PRĘDKOŚCI STATKÓW

Prędkości statków na akwenach nieograniczonych, zwykle modeluje się rozkładem normalnym lub log-normalnym [1, 4]. Natomiast na akwenach ograniczonych, stosowane są różne rozkłady, w zależności od akwenu [2, 5].

### 2.1. Prędkości statków na różnych akwenach

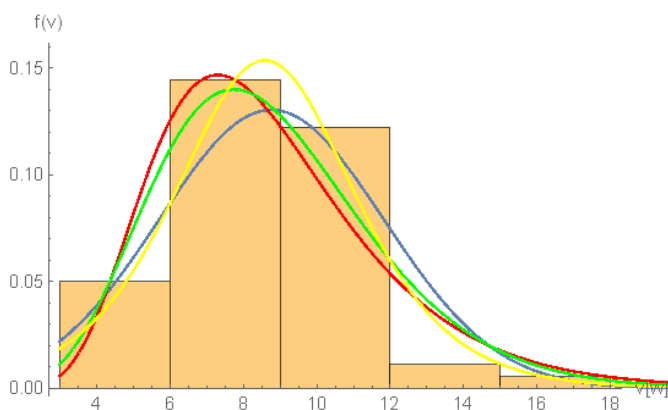
14.11.2014 r. zarejestrowano 41 statków płynących Kanalem Kilońskim. Rozkład średnich prędkości okazał się zgodny z rozkładem, będącym mieszaniną rozkładów normalnych i jednostajnych. Najlepsze dopasowanie w teście chi kwadrat wykazywała mieszanina dwóch rozkładów normalnych:  $N(6,35; 1,35)$  i  $N(8,65; 0,58)$ , w proporcjach 26% na 74% (prawdopodobieństwo testowe  $p$  wyniosło 0,67). Natomiast najmniejszy wskaźnik  $BIC$  miał rozkład mieszany rozkładów jednostajnego  $U(4,2; 9,9)$  (30%) i normalnego  $N(8,6; 0,54)$  (70%), przy prawdopodobieństwie testowym  $p$  równym 0,39. Interesującą rzeczą jest to, że na podstawie kryterium informacyjnego Bayesa równie dobry do opisu prędkości średnich okazał się rozkład Weibulla  $W(8,48; 8,56)$ , którego wartość  $BIC$  niewiele różniła się od pozostałych rozkładów, natomiast wartość prawdopodobieństwa testowego w teście chi kwadrat wyniosła 0,01, co pozwalało odrzucić hipotezę o zgodności z rozkładem Weibulla.



**Rys. 1.** Wykresy funkcji gęstości rozkładów opisujących średnie prędkości statków na Kanale Kilońskim

Na powyższym wykresie wyraźnie widać, że w kwestii dopasowania do danych empirycznych rozkład Weibulla odstaje od pozostałych. Jednak prostota modelu opartego na tym rozkładzie, w porównaniu ze złożonością modeli opartych na mieszaninach rozkładów, pozwala brać go pod uwagę, przy modelowaniu prędkości statków na Kanale Kilońskim.

W lutym 2014 roku na torze podejściowym do Antwerpii badano prędkości chwilowe statków. W większości przypadków rozkład prędkości był zgodny z rozkładem normalnym [5], a w niektórych momentach prędkości chwilowe były zgodne z rozkładem log-normalnym. 10 lutego o godzinie 14.45 zanotowano 62 statki płynące tym torem. Test chi kwadrat Pearsona potwierdził hipotezę, że rozkład prędkości chwilowych w tym momencie jest zgodny z rozkładem log-normalnym LN(2,11;0,35) (prawdopodobieństwo testowe p wyniosło 0,35). Stosując kryterium informacyjne BIC, za najlepiej dopasowany do danych, należy uznać rozkład Gamma G(8,67; 1,01). Drugi w kolejności jest rozkład log-normalny LN(2,11;0,35). Natomiast dużym zaskoczeniem jest bardzo dobry wynik BIC rozkładu normalnego. Test normalności Shapiro-Wilka, wykazał, że rozkład prędkości chwilowych nie jest normalny (p=0,01), jednak wartość kryterium BIC była jedną z najlepszych właśnie dla rozkładu normalnego N(8,76; 3,04).

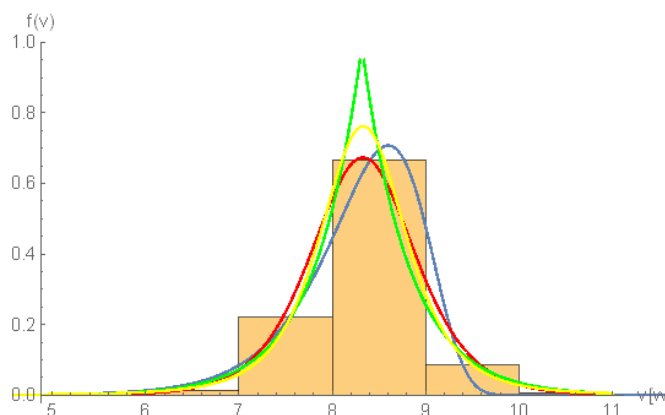


**Rys. 2.** Wykres funkcji gęstości opisujących prędkości chwilowe statków na torze do Antwerpii

Na rysunku 2 przedstawiono wykresy funkcji gęstości rozkładów normalnego, log-normalnego, gamma i logistycznego, który miał również dobry wynik BIC.

## 2.2. Prędkości na torze Szczecin - Świnoujście

W pracy [2] wykazano, że dla różnych typów statków, ich prędkości dobrze opisuje rozkład Gumbela. Zastosowanie procedur, wykorzystujących kryterium informacyjne Bayesa, pozwoliło wyznaczyć inne rozkłady, które mają mniejsze BIC niż rozkład Gumbela. Najlepszym wynikiem BIC charakteryzuje się rozkład Laplace'a z parametrami 8,32 i 0,37. Niewiele większy wynik dają rozkłady: logistyczny z parametrami 8,32 i 0,37 oraz t-Studenta z parametrami 8,33, 0,48 i 3,32. Na rysunku poniżej na tle histogramu przedstawiono wykresy funkcji gęstości wszystkich czterech rozpatrywanych rozkładów.



**Rys. 3.** Wykresy funkcji gęstości rozkładów opisujących prędkości statków na odcinku 11km - I Brama

Niewielkie różnice pomiędzy krzywymi, nie zmieniają faktu, że wszystkie rozkłady dobrze opisują dane empiryczne.

## 3. TESTOWANIE HIPOTEZ DOTYCZĄCYCH CZASU POMIĘDZY ZGŁOSZENIAMI STATKÓW

W przypadku ruchu niezaburzonego, czas pomiędzy kolejnymi jednostkami trawersującymi dany punkt jest zmienną losową wykładniczą. Natomiast w przypadkach, gdzie ruch statków jest zaburzony regulacjami lub ograniczeniami na drodze wodnej zmienna ta nie jest wykładnicza [2, 3].

### 3.1. Ruch niezaburzony

Rozpatrując ruch statków na torze Szczecin - Świnoujście, w kierunku do Szczecina, stwierdzono, że istniejące regulacje na tym akwenie, przy takiej intensywności ruchu, nie są na tyle silne, by czas pomiędzy kolejnymi statkami w tym strumieniu, nie był modelowany rozkładem wykładniczym. Na przykład w punkcie raportowym 11 km, zanotowano w drugim półroczu 2009 roku 1320 statków. Test chi kwadrat Pearsona zdecydowanie potwierdził hipotezę o rozkładzie wykładniczym czasu pomiędzy kolejnymi jednostkami trawersującymi ten punkt raportowy (prawdopodobieństwo testowe p wyniosło 0,48). Również test Cramera von Misesa potwierdził tę hipotezę (p = 0,2). Z pozostałych rozkładów dobrze dopasowanych do danych empirycznych były: rozkłady Weibulla (p = 0,17 oraz p = 0,19) i rozkład Pareto (p = 0,4 oraz p = 0,23). Najniższy BIC miał w tym przypadku rozkład wykładniczy. A więc w tym przypadku jak i w innych, gdzie hipoteza o rozkładzie wykładniczym, była spełniona, nie zanotowano istotnych rozbieżności między wynikami standardowych testów, a zastosowaniem BIC.

### 3.2. Ruch zaburzony

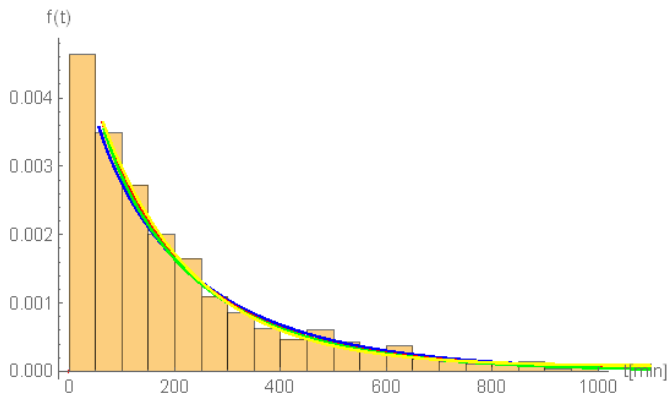
Inaczej sprawy wyglądały w przypadkach, gdy hipoteza o rozkładzie wykładniczym nie była spełniona. Wtedy bardzo dobre dopasowania dawały mieszaniny rozkładów, wyznaczone metodą analityczną - graficzną opracowaną przez autora [4]. I tak np. dla strumienia statków płynących ze Szczecina do Świnoujścia, na punkcie raportowym Inoujście test chi kwadrat Pearsona wykazał, że czas pomiędzy zgłoszeniami kolejnych statków ma rozkład o gęstości:

$$f(x) = \begin{cases} 0.005157e^{-0.0057x} + 0.000077 & \text{for } 0 \leq x \leq 1300 \\ 0.005157e^{-0.0057x} & \text{for } x > 1300 \end{cases} \quad (2)$$

Rozkład ten jest mieszaniną rozkładu wykładniczego (90%) i rozkładu jednostajnego (10%). Prawdopodobieństwo testowe p wyniosło w tym teście 0,55, a w teście Cramera von Misesa aż 0,94.

Natomiast stosując kryterium informacyjne  $BIC$ , za najlepiej dopasowany do danych, należy uznać rozkład Weibulla  $W(0,89; 213,8; 0,6)$ . Drugi w kolejności jest rozkład Pareto  $P(1664,39; 8,26; 1,03; 0,60)$ , a trzeci znów rozkład Weibulla  $W(0,89; 192,99)$ .

Poniżej przedstawiono, na tle histogramu, wykresy funkcji gęstości wspomnianych czterech rozkładów.



**Rys. 4.** Wykresy funkcji gęstości rozkładów opisujących czas pomiędzy zgłoszeniami kolejnych jednostek w punkcie raportowym Inoujście

Jak można zauważyć różnice pomiędzy krzywymi są niewielkie, wszystkie bardzo dobrze odzwierciedlają zmienność danych.

## PODSUMOWANIE

Kryterium informacyjne Bayesa, ze swej natury preferuje prostsze modele, stąd wszędzie tam, gdzie test chi kwadrat potwierdzał zgodność z rozkładem mieszanym, występowały wyraźne rozbieżności pomiędzy dopasowanymi rozkładami wyznaczonymi na podstawie kryterium informacyjnego  $BIC$  i na podstawie testu zgodności chi kwadrat Pearsona. W niektórych wypadkach, wręcz testy wykazywały odrzucenie hipotezy o zgodności z danym rozkładem, a kryterium  $BIC$  uznawało taki rozkład za bardzo dobre dopasowanie do danych empirycznych.

Natomiast tam, gdzie testy potwierdzały zgodność z rozkładem wykładniczym nie zanotowano istotnych rozbieżności wyników otrzymanych różnymi metodami.

Wynika stąd, że należy z pewną ostrożnością podchodzić do metod wyznaczających bardzo dobrze dopasowane rozkłady, gdyż mogą one wykazywać silne właściwości przeuczenia, co znacznie ogranicza możliwość ich szerszego zastosowania.

## BIBLIOGRAFIA

1. Gućma L., Schefs S., *Studium prędkości statków na torze wodnym Świnoujście – Szczecin*. Wydawnictwo AM, Szczecin 2007.
2. Kasyk L., Kijewska M., *Gumbel Distribution in Analysis of Vessel Speed on the Świnoujście – Szczecin Fairway*, Zeszyty Naukowe AM nr , Szczecin 2013.
3. Kasyk L., *Non-Poisson character of vessel traffic on the Szczecin - Świnoujście fairway*, Zeszyty Naukowe, Akademia Morska w Szczecinie nr 32/2012.
4. Kasyk L., *Probabilistyczne metody modelowania parametrów strumienia ruchu statków na akwenach ograniczonych*. Wydawnictwo UTH, Radom 2012.
5. Kasyk L., *Rozkład prędkości chwilowych statków na wybranych torach wodnych Północnej Europy*, Logistyka 3/2014.
6. Piłatowska M., *Porównanie kryteriów informacyjnych i predykcyjnych w wyborze modelu*, „Prace i Materiały Wydziału Zarządzania Uniwersytetu Gdańskiego” 2011, nr 4/8.
7. Rogowski A., *Podstawy metod probabilistycznych w transporcie*, Wydawnictwo UTH w Radomiu, Radom 2012.
8. Rogowski A., *Testowanie hipotezy o rozkładzie Poissona w oparciu o statystykę Cramera – von Misesa*, Logistyka 3/2015.
9. Stanisz A., *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*, StatSoft, Kraków 2006.

### A Bayesian information criterion in statistical hypothesis testing concerning parameters of vessel traffic flow

*In this paper differences between using Bayesian information criterion and the most popular goodness-of-fit tests in hypothesis testing about parameters of vessel traffic flow have been presented. Statistical tests concerning intensity and velocity of vessel traffic flow on different fairways have been analysed. Among goodness-of-fit tests, Pearson Chi Square and Cramer - von Misses test have been considered.*

Wyniki badań powstałe w ramach realizacji pracy badawczej pt. „Modelowanie ryzyka eksploatacji statku manewrującego na akwencie ograniczonym” nr 4/S/INM/15, finansowanej z dotacji Ministerstwa Nauki i Szkolnictwa Wyższego na finansowanie działalności statutowej.

Autorzy:

dr hab. **Lech Kasyk** – Akademia Morska w Szczecinie, Instytut Technologii Morskich, l.kasyk@am.szczecin.pl