

## PRZEGLĄD NARZĘDZI MATEMATYCZNYCH SŁUŻĄCYCH DO UZYSKANIA CIĄGŁEGO ROZKŁADU PRAWDOPODOBIENSTWA CZASU TRWANIA ROBÓT BUDOWLANYCH NA PODSTAWIE ROZKŁADU DYSKRETNEGO

Krzysztof Kaczorek✉

Wydział Inżynierii Lądowej, Politechnika Warszawska, Warszawa

### STRESZCZENIE

Jednymi z narzędzi służących do prognozowania czasu realizacji przedsięwzięcia budowlanego są metody symulacyjne. W celu ich przeprowadzenia należy czasowi trwania każdej roboty budowlanej przyporządkować rozkład prawdopodobieństwa. Wadą takiego postępowania jest to, że rozkłady są przyjmowane w sposób bardzo subiektywny. Co więcej, kształt poszczególnych rozkładów nie pokazuje, na podstawie jakich przesłanek został ustalony. Jest to dość problematyczne, ponieważ oczywiste jest to, że na roboty budowlane oddziałują czynniki ryzyka, których wpływu po kształcie rozkładu prawdopodobieństwa ich czasu trwania nie widać. Co więcej, każdy czynnik ma takie parametry, jak: częstość występowania, wpływ na przebieg danej roboty budowlanej oraz wpływ na pozostałe czynniki ryzyka. W celu uwzględnienia oddziaływania czynników ryzyka na kształt rozkładu prawdopodobieństwa czasu trwania robót budowlanych możliwe jest zastosowanie wnioskowania rozmytego, które pozwala uzyskać wspomniany rozkład uzależniony od parametrów poszczególnych czynników ryzyka dla konkretnych robót. Będzie on w postaci dyskretnej. Chcąc uzyskać możliwie dobry wynik symulacji, należy na podstawie rozkładu dyskretnego uzyskać rozkład ciągły. W niniejszej publikacji opisano oraz wskazano wady i zalety estymacji punktowej, estymacji przedziałowej oraz estymacji nieparametrycznej, które służą do rozwiązania postawionego problemu.

**Słowa kluczowe:** estymacja parametryczna, estymacja nieparametryczna, harmonogramy budowlane, czas wykonania robót budowlanych

### WSTĘP

Jak podają Jakubowski i Marcinkowski (2014) oraz Hoła i Mrozowicz (2003), symulacje komputerowe, dzięki swoim zaletom, stały się jednymi z podstawowych narzędzi stosowanych do prognozowania czasów realizacji przedsięwzięć budowlanych. Wśród wspomnianych atutów Łatuszyńska (2011) wymienia możliwość „prześledzenia” nawet wieloletnich procesów w bardzo krótkim (często na poziomie kilku lub

kilkudziesięciu minut) czasie. Pozwala to planiście na szybkie przeprowadzanie analiz kolejnych rozważanych scenariuszy. Ponadto symulacje komputerowe umożliwiają rozszerzenie czasu działania systemu, gdyż mogą w nich zostać uwzględnione zdarzenia i okoliczności trudne do zaobserwowania w czasie rzeczywistym. Poszczególne eksperymenty są przeprowadzane w takich samych warunkach i można je powtarzać dowolną liczbę razy, dodatkowo wprowadzając pożądane modyfikacje. Otrzymane wyniki są

✉ k.kaczorek@il.pw.edu.pl

łatwe do przechowywania oraz porównywania. Niestety symulacje komputerowe nie są wolne od wad. Według Biruka (2015) jedną z nich jest dalece subiektywny charakter przyjmowanych rozkładów prawdopodobieństwa czasu wykonania poszczególnych robót budowlanych. Ponadto Kaczorek, Kulejewski i Ibadov (2018) twierdzą, iż rozkłady prawdopodobieństwa nie są budowane według wpływów poszczególnych czynników ryzyka, ale na podstawie doświadczenia planisty. Wobec tego przyjęty rozkład nie ukazuje wpływu okoliczności kreujących postępy danej roboty budowlanej, a jest przyjęty całkowicie dowolnie.

Mając na uwadze powyższe fakty, zasadnym jest opracowywanie modeli matematycznych, w których planiści określają parametry czynników ryzyka: częstość występowania i wpływ czynnika na czas wykonania roboty budowlanej, których wagę podkreślają Marcinkowski i Koper (2008), oraz zaproponowaną przez Kaczorka i in. (2018) podatność, czyli parametr, którego zadaniem jest określenie, na ile wpływ danego czynnika na czas wykonywania roboty budowlanej jest uzależniony od wpływu pozostałych czynników ryzyka. Dopiero na podstawie uzyskanych informacji o orientacyjnym poziomie wymienionych parametrów możliwe jest przeprowadzenie wnioskowania rozmytego, którego efektem jest dyskretny rozkład prawdopodobieństwa, określający wydłużenie czasu realizacji roboty budowlanej. Niemniej bardziej wiarygodne wyniki symulacji komputerowej są uzyskiwane dzięki zastosowaniu rozkładów ciągłych. Jest to niezwykle istotne dla zapewnienia odpowiedniej terminowości prac, co według Gicali i Sobotki (2017) jest kluczowe dla osiągnięcia sukcesu w branży budowlanej. W niniejszej pracy przedstawiono narzędzia matematyczne, które umożliwiają uzyskanie rozkładu ciągłego na podstawie dyskretnego rozkładu prawdopodobieństwa. Są to estymacje nieparametryczne oraz parametryczne: punktowa oraz przedziałowa. Dzięki przeglądowi oraz porównaniu narzędzi możliwe jest wybranie najbardziej odpowiedniego dla postawionego problemu.

## ESTYMACJA I ESTYMATORY

Jak podaje Gajek (1998), estymacja to „dział wnioskowania statystycznego będący zbiorem metod pozwalających na uogólnianie wyników badania próby

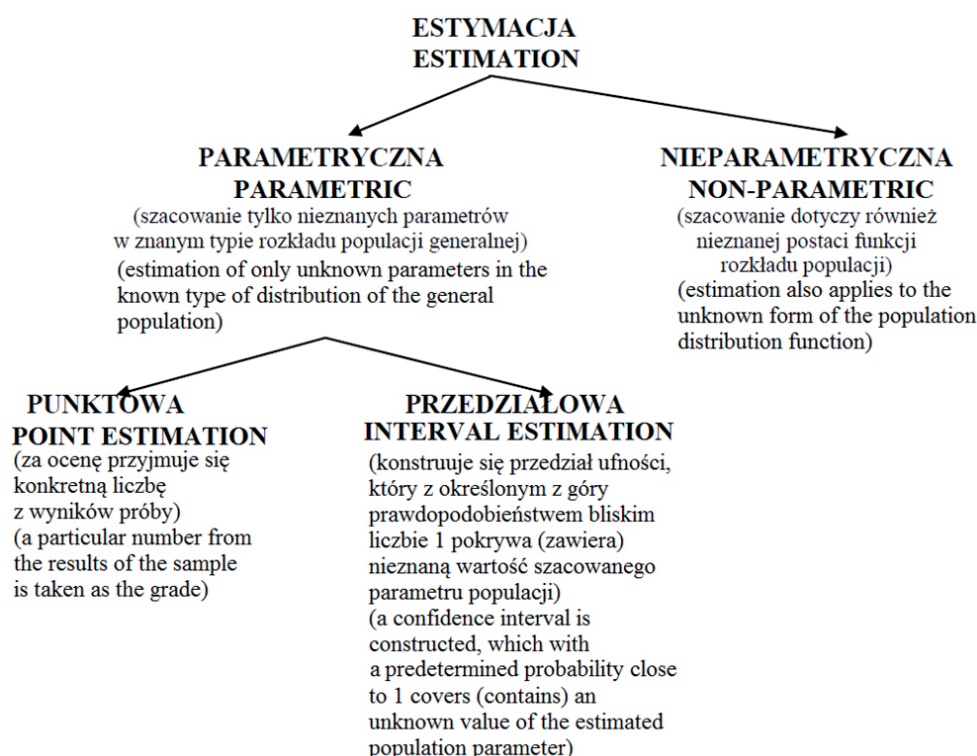
losowej na nieznaną postać i parametry rozkładu zmiennej losowej całej populacji oraz szacowanie błędów wynikających z tego uogólnienia”. Wyrażenie nieznaną postać jest podstawą do odróżnienia estymacji od drugiego działu wnioskowania statystycznego, jakim jest weryfikacja hipotez statystycznych. W weryfikacji hipotez statystycznych najpierw stawiane są przypuszczenia na temat rozkładu, a następnie sprawdzana jest ich poprawność. Rodzaje estymacji przedstawia rysunek 1.

Zgodnie z tym, co pisze Sobczyk (2007), estymator to „wielkość (statystyka, charakterystyka) wyznaczona na podstawie próby losowej, służąca do oceny wartości nieznanymi parametrów populacji generalnej”. Niemniej należy pamiętać, że nie każda statystyka obliczona z próby (czyli nie każda wielkość będąca funkcją wartości w próbie) może być estymatorem konkretnego parametru. Estymatorem parametru  $Q$  rozkładu zmiennej losowej  $X$  nazywana jest taka statystyka  $Z_n = f(X_1, X_2, \dots, X_n)$ , której rozkład prawdopodobieństwa jest zależny od szacowanego parametru  $Q$ . Przykładowo dla populacji generalnej o rozkładzie normalnym  $N(m, \sigma)$  z nieznaną średnią  $m$  i wariancją  $\sigma$  statystyki  $\bar{X}$  (średniej arytmetycznej) i  $S^2$  (wariancja z próby) są estymatorami. Wynika to z tego, że ich rozkłady zależą od odpowiednich parametrów populacji.

Oceną parametru  $Q$  nazywana jest konkretna wartość liczbowa  $z_n = f(x_1, x_2, \dots, x_n)$ , którą przyjmuje estymator  $Z_n$  parametru  $Q$  dla realizacji próby  $(x_1, x_2, \dots, x_n)$ .

Estymacja pewnego parametru z użyciem określonego estymatora  $Z_n$  jest dokonywana na podstawie wyników próby losowej, zachodzi więc możliwość popełnienia błędu. Błędem szacunku (estymacji) parametru  $Q$  nazywana jest różnica między uzyskaną liczbą oceną parametru oraz jego faktyczną wartością, tj.  $z_n - Q$ . Chcąc uzyskać możliwie najwyższą precyzję szacunku (czyli możliwie mały błąd estymacji), należy zapewnić zarówno prawidłowe losowanie próby, jak i dobór najbardziej odpowiedniego estymatora  $Z_n$  dla szacowanego parametru  $Q$ . Wobec tego wprowadzono własności, którymi powinien wykazywać się dobry estymator. Są to: nieobciążoność, zgodność, efektywność, i dostateczność (wystarczalność).

Nieobciążoność estymatora oznacza, że przy wielokrotnym losowaniu próby średnia wartości



**Rys. 1.** Rodzaje estymacji według Borowskiej (2016)

**Fig. 1.** Types of estimation according to Borowska (2016)

przyjmowanych przez estymator równa się wartości szacowanego parametru. Upraszczając, własność nieobciążoności estymatora zapewnia otrzymywanie przy jego zastosowaniu ocen wolnych od błędu systematycznego.

Zgodność estymatora  $Z_n$  parametru  $Q$  oznacza, że jest on stochastycznie zbieżny (zbieżny według prawdopodobieństwa) do szacowanego parametru  $Q$ . Wynika z tego, że zwiększając liczebność próby, zwiększa się również prawdopodobieństwo tego, że estymator będzie przyjmował wartości coraz bliższego do szacowanego parametru  $Q$ .

Najefektywniejszy estymator parametru  $Q$  to ten, który ma największą precyzję. Odwrotnością wariancji estymatora jest właśnie precyzja.

Dostateczność (wystarczalność) estymatora  $Z_n$  parametru  $Q$  istnieje wtedy, kiedy zawiera on wszystkie informacje, jakie na temat parametru  $Q$  występują w próbie. Żaden inny estymator nie może już dać dodatkowych informacji o szacowanym parametrze.

## ESTYMACJA PUNKTOWA

Według Sobczyka (2007) „w estymacji punktowej za ocenę nieznannej wartości parametru  $Q$  w populacji przyjmuje się jedną konkretną liczbę otrzymaną – przy zachowaniu odpowiednich reguł postępowania – z wyników próby losowej”. Zazwyczaj podczas szacowania danego parametru, obok jego oceny, podaje się również średni błąd szacunku  $D(Z_n)$  estymatora nieobciążonego. Służy on do mierzenia przeciętnej wielkości błędów szacunku, które by popełniono, gdyby z wyróżnionej populacji wielokrotnie pobierano  $n$ -elementowe próby i na ich podstawie szacowano parametry przy użyciu określonego estymatora. Prawdopodobieństwo zmaterializowania się sytuacji, w której estymator przyjmie wartość równą wartości szacowanego parametru, przynajmniej w przypadku populacji ciągłych, jest równe zero. Płyne z tego wniosek, że popełnienie błędu w ocenie wartości parametru  $Q$  przy zastosowaniu estymacji punktowej jest pewne. Jest to

jeden z powodów, dla których stosowana jest estymacja przedziałowa.

Wśród metod, które służą do znajdowania estymatorów o pożądanym własnościach, można wymienić: metodę momentów (MM), metodę największej wiarygodności (MNW) oraz metodę najmniejszych kwadratów (MNK).

Za autora podstaw najstarszej (z przełomu wieków XIX i XX) z wyżej wymienionych metod, czyli metody momentów, uznawany jest Karla Pearson. Istotą metody jest to, że wartość momentu z próby jest przyjmowana za oszacowanie odpowiedniego momentu populacji generalnej. Wobec tego w celu oszacowania momentu rzędu  $k$  populacji wykorzystywany jest moment rzędu  $k$  z próby. Estymatory uzyskiwane dzięki zastosowaniu MM są zazwyczaj obciążone i mało efektywne.

Lepsze estymatory niż dzięki MM uzyskiwane są dzięki metodzie największej wiarygodności. Koncepcja tej metody została opracowana w latach 20. XX wieku przez Ronalda Aylmera Fishera. Kluczowym zagadnieniem dla MNW jest pojęcie wiarygodności próby (funkcji wiarygodności). Jak zapisał Sobczyk (2007), wiarygodność próby prostej to „łączne prawdopodobieństwo (dla rozkładu ciągłego populacji) wyników  $(x_1, x_2, \dots, x_n)$ , jakie dała próba, przy czym wiarygodność ta zależy od prawdziwej wartości szacowanego parametru  $Q$ ”. Wiarygodność próby oznaczana jest literą  $L$  (od angielskiego słowa *likelihood*). Zapis sformalizowany wiarygodności  $n$ -elementowej próby prostej można przedstawić w postaci wyrażenia:

$$L(X_n; Q) = \begin{cases} \prod_{i=1}^n f(x_i; Q) & \text{dla populacji o rozkładzie ciągłym} \\ \prod_{i=1}^n p(x_i; Q_i) & \text{dla populacji o rozkładzie skokowym} \end{cases}$$

gdzie:

$f(x_i; Q)$  – funkcja gęstości prawdopodobieństwa,  
 $p(x_i; Q_i)$  – funkcja rozkładu prawdopodobieństwa uzyskania zbioru realizacji  $x_n = \{x_1, x_2, \dots, x_n\}$ , przy czym  $Q$  może być pojedynczym parametrem lub wektorem parametrów.

Przedstawione wyrażenie ma sens tylko w sytuacji, kiedy elementy zbiorowości generalnej są wybierane do próby w drodze losowania niezależnego. Tylko przy spełnieniu tego wymogu prawdopodobieństwo uzyskania zbioru realizacji  $x_n = \{x_1, x_2, \dots, x_n\}$  jest

równe iloczynowi prawdopodobieństw otrzymania poszczególnych realizacji  $x_i \in X_n$ . Efektem zastosowania metody jest uzyskanie estymatora najwiarygodniejszego. Jest to taki estymator  $Z_n$  parametru  $Q$ , który maksymalizuje wiarygodność próby losowej, tzn. dla którego zachodzi poniższa równość:

$$L(Z_n) = \max_Q (X_n; Q)$$

Poszukiwania estymatorów MNW są prowadzone z wykorzystaniem rachunku różniczkowego. Istnieje możliwość znacznego uproszczenia tego rachunku, jeśli zamiast maksymalizować funkcję  $L$ , poszukuje się maksimum  $\ln L$ . Wynika to z tego, że  $\ln L$  osiąga maksimum w tym samym punkcie co  $L$ , a jest nieporównywalnie łatwiejsza do różniczkowania (dla prób prostych  $L$  jest iloczynem a  $\ln L$  sumą).

Wadą MNW jest istotne ograniczenie wynikające z konieczności znajomości postaci funkcyjnej rozkładu populacji. Wyłącznie wtedy możliwym jest wyznaczenie funkcji wiarygodności  $L$ .

Ostatnią omawianą w niniejszym artykule metodą otrzymywania estymatorów w estymacji punktowej jest metoda najmniejszych kwadratów. Metoda pochodzi z początku XIX wieku i za jej ojca uważany jest Carl Friedrich Gauss, a rozwinięcie metody zawdzięczane jest przede wszystkim Andrejowi Markowowi. Metoda ta umożliwia estymację parametrów funkcji wyrażających zależności pomiędzy zmiennymi losowymi (np. funkcji regresji, funkcji trendu). Istotą metody jest taki dobór ocen szacowanych parametrów, aby suma kwadratów odchyłeń (różnic) empirycznych wartości danej funkcji od jej wartości teoretycznych była najmniejsza. Dla funkcji liniowych lub możliwych do sprowadzenia do postaci liniowej (np. przez logarytmowanie) estymatory uzyskane dzięki MNK są zgodne, nieobciążone i najefektywniejsze w klasie estymatorów liniowych.

## ESTYMACJA PRZEDZIAŁOWA

Zgodnie z tym, co podaje Sobczyk (2007), metody estymacji punktowej umożliwiają uzyskiwanie oceny wartości nieznanymi parametrów rozkładu, ale nie dają informacji, jaka jest dokładność uzyskanej oceny. W celu otrzymania odpowiedzi na postawione wyżej



pytanie, należy zastosować estymację przedziałową. Wynikiem oceny parametru  $Q$  nie jest (jak w przypadku estymacji punktowej) konkretna wartość, ale przedział, w którym z określonym prawdopodobieństwem mieści się szacowana wartość parametru  $Q$ . Przedział ten określany jest mianem przedziału ufności, a ogólny zapis jest następujący:

$$P\{g_1(Z_n) < Q < g_2(Z_n)\} = 1 - \alpha$$

Twórcą metody jest Jerzy Sława-Neyman. W jego podejściu parametr  $Q$  jest wielkością stałą (nie-losową), a końce przedziału  $g_1(Z_n)$  oraz  $g_2(Z_n)$ , zwane odpowiednio dolną i górną granicą przedziału, są zmiennymi losowymi zależnymi od  $Q$ . Wobec tego wynik estymacji przedziałowej należy interpretować, wskazując na zmienność granic przedziału, a nie samego parametru  $Q$ . Ustalone odgórnie przez statystyka prawdopodobieństwo  $1 - \alpha$ , z jakim przedział ufności pokrywa nieznaną wartość parametru  $Q$ , nazywane jest współczynnikiem ufności (rzadziej poziomem ufności). Najczęściej w miejsce  $1 - \alpha$  przyjmowane są prawdopodobieństwa: 0,90; 0,95; 0,99. Różnica między górną a dolną granicą przedziału ufności nazywana jest długością przedziału. Im krótszy przedział ufności, tym precyzyjniejsza (dokładniejsza) jest przeprowadzana estymacja przedziałowa. Warto zwrócić uwagę na zależność między współczynnikiem ufności a precyzją estymacji przedziałowej. Im współczynnik ufności bliższy 1, tym długość przedziału większa, a co za tym idzie – precyzja estymacji jest mniejsza. Wobec tego należy pamiętać, aby współczynnik ufności dobierać do potrzeb i nie doprowadzić do jego niepotrzebnego zawyżenia, ponieważ przełoży się to wtedy negatywnie na precyzję przeprowadzanej estymacji.

## ESTYMACJA NIEPARAMETRYCZNA

Opisane wcześniej w artykule estymacje parametryczne mogą być stosowane, jeśli znana jest postać funkcyjna (typ) rozkładu populacji. Jest to poważne ograniczenie, które zamyka proces estymacji jedynie w obrębie kilkunastu dostępnych typów rozkładów. Jeśli zachodzi sytuacja, w której brak jest takich in-

formacji i szacowanie dotyczy także nieznannej postaci funkcyjnej rozkładu populacji, to należy zastosować estymację nieparametryczną. Wzrost popularności tego rodzaju estymacji według Łukasika (2008), oprócz zniwelowania wymienionego ograniczenia, jest zawdzięczany w dużej mierze gwałtownemu wzrostowi mocy obliczeniowej współczesnych komputerów. Dzięki temu możliwe jest, aby coraz szybciej rozwiązywać coraz bardziej złożone problemy. Jednym z najchętniej stosowanych estymatorów nieparametrycznych jest jądrowy estymator gęstości.

Estymator jądrowy funkcji gęstości rozkładu prawdopodobieństwa zmiennej losowej  $X$  definiuje się jako:

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^m K\left(\frac{x - x_i}{h}\right)$$

gdzie:

$\hat{f}(x)$  – estymator funkcji gęstości rozkładu skonstruowany na podstawie próby  $X$ ,

$X$  – rozważana zmienna losowa,

$n$  – wymiar zmiennej losowej  $X$ ,

$m$  – liczność elementów próby,

$h > 0$  – tzw. współczynnik wygładzenia,

$K: \mathbb{R}^n \rightarrow [0, \infty)$  – funkcja spełniająca trzy poniższe warunki:

1)  $\int_{\mathbb{R}^n} K(x) dx = 1.$

2)  $K(x) = K(-x)$  dla dowolnego  $x \in \mathbb{R}^n$ .

3)  $K(0) \geq K(x)$  dla dowolnego  $x \in \mathbb{R}^n$ .

Odwzorowanie  $K$  jest nazywane jądrem poprzez analogię do jądra operatora całkowego. Jakość estymacji w dużej mierze zależy od wyboru postaci jądra oraz zastosowania odpowiedniej wartości parametru wygładzania.

Wadą metody jest bardzo duże zapotrzebowanie oprogramowania na moc obliczeniową komputera. Przekłada się to na długie, a czasem nawet bardzo długie przeprowadzanie estymacji.

Na rysunku 2 w sposób syntetyczny zebrano informacje przedstawione w poprzednich rozdziałach niniejszego artykułu.

		parametryczna – parametric			nieparametryczna non-parametric	
		punktowa – point		przedziałowa interval	jądrowa nuclear	
		+ MM	+ MNW			+ MNK
Zalety – ADVANTAGES		<ul style="list-style-type: none"> <li>nieskomplikowana – uncomplicated</li> <li>nie wymaga dużej mocy obliczeniowej komputera – it does not require high computing power of the computer</li> <li>szybko przeprowadzane obliczenia – quickly carried out calculations</li> </ul>	<ul style="list-style-type: none"> <li>nieskomplikowana – uncomplicated</li> <li>lepsze estymatory niż w MM – better estimators than in MM</li> <li>nie wymaga dużej mocy obliczeniowej komputera – it does not require high computing power of the computer</li> <li>szybko przeprowadzane obliczenia – quickly carried out calculations</li> </ul>	<ul style="list-style-type: none"> <li>nieskomplikowana – uncomplicated</li> <li>lepsze estymatory niż w MM – better estimators than in MM</li> <li>najlepsze estymatory w klasie estymatorów liniowych – the best estimators in the class of linear estimators</li> <li>nie wymaga dużej mocy obliczeniowej komputera – it does not require high computing power of the computer</li> <li>szybko przeprowadzane obliczenia – quickly carried out calculations</li> </ul>	<ul style="list-style-type: none"> <li>pozwala ocenić dokładność uzyskanej oceny – allows you to assess the accuracy of the obtained rating</li> <li>nieskomplikowana – uncomplicated</li> <li>nie wymaga dużej mocy obliczeniowej komputera – it does not require high computing power of the computer</li> <li>szybko przeprowadzane obliczenia – quickly carried out calculations</li> </ul>	<ul style="list-style-type: none"> <li>brak konieczności ograniczania się do dostępnych typów rozkładów – no necessity to limit to available types of distributions</li> <li>możliwość rozwiązywania bardziej skomplikowanych problemów – ability to solve more complicated problems</li> </ul>
	Wady – DISADVANTAGES		<ul style="list-style-type: none"> <li>konieczność ograniczania się do dostępnych typów rozkładów – necessity to limit to available types of distributions</li> <li>estymatory zazwyczaj obciążone i mało efektywne – estimators usually laden and ineffective</li> </ul>	<ul style="list-style-type: none"> <li>konieczność ograniczania się do dostępnych typów rozkładów – necessity to limit to available types of distributions</li> </ul>	<ul style="list-style-type: none"> <li>konieczność ograniczania się do dostępnych typów rozkładów – necessity to limit to available types of distributions.</li> </ul>	<ul style="list-style-type: none"> <li>konieczność ograniczania się do dostępnych typów rozkładów – necessity to limit to available types of distributions</li> </ul>

**Rys. 2.** Zestawienie wad i zalet poszczególnych estymacji (opracowanie własne)

**Fig. 2.** List of advantages and disadvantages of each estimation (own study)

## PODSUMOWANIE, WNIOSKI I DALSZE KIERUNKI PRACY NAUKOWO-BADAWCZEJ

Jak podają Leśniak oraz Plebankiewicz (2010), jednym z podstawowych problemów w planowaniu przedsięwzięcia inwestycyjno-budowlanego jest prawidłowe oszacowanie skali opóźnień poszczególnych robót budowlanych. To oszacowanie pozwala następnie przyjąć stosowne bufor, dzięki czemu termin zakończenia robót jest ustanowiony w sposób wiarygodny (Rogalska i Hejducki, 2005). W celu określenia wspomnianych opóźnień planiści chętnie stosują metody symulacyjne. Niemniej w obliczu coraz większych wymagań zleńców oraz coraz bardziej skomplikowanych realizacji osoby planujące budowę przestały zadowalać

się przyjmowanymi *ad hoc* rozkładami prawdopodobieństwa. Naprzeciw ich oczekiwaniom wychodzą modele wnioskujące, które pozwalają oszacować rozkład prawdopodobieństwa na podstawie chociażby parametrów czynników ryzyka. Niektóre modele dają wynik w postaci dyskretnej, wskazane jest więc uzyskanie postaci ciągłej, dzięki czemu końcowy wynik symulacji będzie bardziej wiarygodny. Do uzyskiwania rozkładu ciągłego prawdopodobieństwa służą metody estymacyjne. W artykule omówiono oraz porównano estymację punktową, estymację przedziałową oraz estymację jądrową. Dzięki przeprowadzonej analizie możliwe jest sprawne dobranie właściwej metody do przeprowadzenia estymacji na podstawie uzyskanego z badań dyskretnego rozkładu prawdopodobieństwa

czasu trwania robót budowlanych. W dalszej kolejności wskazane jest wykonanie wielu studiów przypadków, które pozwolą potwierdzić zasadność dopasowania poszczególnych rodzajów estymacji do konkretnych modeli wnioskujących.

## PIŚMIENNICTWO

- Biruk, S. (2015). Analiza sieci PERT metodą symulacji komputerowej. W T. Kasprówicz (red.), *Inżynieria przedsięwzięć budowlanych. Rekomendowane metody i techniki* (strony 157–171). Warszawa: Wydawnictwo Polskiej Akademii Nauk.
- Borowska, M. (2016). *Statystyka. Materiały pomocnicze dla studentów do nauki statystyki*. Sandomierz: Wydawnictwo Diecezjalne i Drukarnia w Sandomierzu.
- Gajek, L. (1998). *Wnioskowanie statystyczne dla studentów. Modele i metody*. Warszawa: Wydawnictwo Naukowe Techniczne.
- Gicala, M. i Sobotka, A. (2017). Jaką strategię obrać? Kluczowe czynniki sukcesu według badań – część 1. *Builder*, 7, 44–45.
- Hoła, B. i Mrozowicz, J. (2003). *Modelowanie procesów budowlanych o charakterze losowym*. Wrocław: Dolnośląskie Wydawnictwo Naukowe.
- Jakubowski, G. i Marcinkowski, R. (2014). Estimating the time of building processes with probabilistic models. *Technical Transaction Civil Engineering*, 6, 273–281.
- Kaczorek, K., Kulejewski, J. i Ibadov, N. (2018). Szacowanie czasów wykonania robót budowlanych z uwzględnieniem wzajemnego oddziaływania czynników ryzyka. *Materiały Budowlane*, 10, 55–58.
- Leśniak, A. i Plebankiewicz, E. (2010). Delays in construction works. *Scientific Journal of the Military University of Land Forces*, 157 (3), 332–339.
- Łatuszyńska, M. (2011). Metody symulacji komputerowej – próba klasyfikacji logicznej. *Studies & Proceedings of Polish Association for Knowledge Management*, 41, 163–176.
- Łukasik, S. (2008). Identyfikacja rozkładu w systemach rzeczywistych za pomocą estymatorów jądrowych. *Czasopismo Techniczne. Elektrotechnika*, 105, 3–13.
- Marcinkowski, R. i Koper, A. (2008). Ocena ryzyka czasu i kosztów w planowaniu produkcji budowlanej. *Przeгляд Budowlany*, 79, 70–75.
- Rogalska, M. i Hejducki, Z. (2005). Zastosowanie buforów czasu w harmonogramowaniu procesów budowlanych. *Przeгляд Budowlany*, 6, 36–41.
- Sobczyk, M. (2007). *Statystyka*. Warszawa: Wydawnictwo Naukowe PWN.

## REVIEW OF MATHEMATICAL TOOLS FOR THE CREATING OF CONTINUOUS DISTRIBUTION OF PROBABILITY OF THE DURATION OF CONSTRUCTION WORKS ON THE BASIS OF DISCRETE DISTRIBUTION

### ABSTRACT

The simulation methods are one of the tools used to forecast the duration of a construction project. In order to carry them out, a distribution of probability of the duration of construction works should be assigned to each construction work. The disadvantage of such a procedure is that distributions are received in a very subjective way. What is more, a shape of distribution of probability of the duration of construction works does not show on the basis of which premises was established. It is quite problematic, because it is obvious that the construction works are affected by risk factors whose impact after the shape of the distribution cannot be seen. What's more, each factor has parameters such as incidence, impact on the course of a given construction work and impact on other risk factors. In order to take into account the impact of risk factors on the shape of the probability distribution, it is possible to use fuzzy inference, which allows to obtain a distribution of probability of the duration of construction works depending on the parameters of particular risk factors for specific works. It will be in a discrete form. In order to obtain the best possible simulation result, a continuous distribution should be obtained on the basis of a discrete distribution. This publication describes and shows advantages and disadvantages of the point estimate, interval estimation and non-parametric estimation that are used to solve the problem.

**Key words:** parametric estimation, non-parametric estimation, construction schedules, execution time of construction works

