ŚRODOWISKO
CZASOPISMO TECHNICZNE
TECHNICAL TRANSACTIONS
ENVIRONMENTAL ENGINEERING
WYDAWNICTWO POLITECHNIKI KRAKOWSKIEJ
2-Ś/2009
ZESZYT 2
ROK 106
ISSUE 2
YEAR 106

RYSZARDA IWANEJKO*

OCENA DOKŁADNOŚCI MIAR NIEZAWODNOŚCIOWYCH SYSTEMÓW WODOCIĄGOWYCH I KANALIZACYJNYCH

CZĘŚĆ II. OCENY STATYSTYCZNE

EVALUATION OF ACCURACY OF RELIABILITY PARAMETERS IN WATER AND SEWAGE SYSTEMS

PART II. STATISTICAL ESTIMATES

Streszczenie

Podstawą podejmowania decyzji dotyczących modernizacji obiektów wodociągowych i kanalizacyjnych są m.in. wartości parametrów niezawodnościowych wyznaczanych dla tych obiektów. Gromadzone przez przedsiębiorstwa wodociągowe dane o awariach są najczęściej wykorzystywane do wyznaczania wartości średnich, czyli do przeprowadzania estymacji punktowej. W artykule przedstawiono wybrane statystyczne metody estymacji przedziałowej wraz z ich aplikacjami. Wyznaczone przedziały ufności mogą stanowić ocenę dokładności uzyskanych oszacowań punktowych.

Słowa kluczowe: *wodociągi, kanalizacje, niezawodność, dokładność*

Abstract

Reliability parameters play a very important role and serve as one of the basic criteria during a decision making process on modernization of water and sewage facilities. The data on failures, collected by the water utilities, are frequently used to determine the average values, i.e. to perform point estimation. The paper presents some statistic methods of interval estimation and their applicatio`ns. The confidence interval determined in calculations may help to evaluate the accuracy of point estimations.

Keywords: *water supply systems, sewage systems, reliability, accuracy*

* Dr Ryszarda Iwanejko, Instytut Zaopatrzenia w Wodę i Ochrony Środowiska, Wydział Inżynierii Środowiska, Politechnika Krakowska.

1. Szacowanie statystyczne

Na potrzeby artykułu przyjęto następujące oznaczenia: $\boldsymbol{X}$ – zmienna losowa (np. $\boldsymbol{Tp}$ – zmienna losowa opisująca czas sprawności obiektu, $\boldsymbol{Tn}$ – zmienna losowa opisująca czas niesprawności obiektu, $\boldsymbol{U}$ – zmienna losowa opisująca liczbę uszkodzeń/awarii obiektu), X^0 – dokładna wartość szacowanego parametru będąca wartością oczekiwaną odpowiedniej zmiennej losowej $X^0 = E(\boldsymbol{X})$, (np. $Tp^0 = E(\boldsymbol{Tp})$, $Tn^0 = E(\boldsymbol{Tn})$, $\lambda^0 = E(\boldsymbol{U})$, gdzie E to symbol wartości oczekiwanej), X – punktowy estymator nieznanego parametru X^0.

Estymator punktowy X jest statystyką z próby (funkcją) służącą do szacowania nieznanej wartości parametru X^0. Estymator musi spełniać pewne warunki. Wymaga się, by estymator był zgodny i nieobciążony. Estymator X jest zgodny, jeśli $\forall\, \xi > 0$ $\lim\limits_{n \leftarrow \infty} P\left(\left|X^0 - X\right| < \xi\right) = 1.$ Inaczej: jest to estymator, który podlega działaniu prawa wielkich liczb, a wzrost liczebności próby zwiększa dokładność szacunku. Estymator jest nieobciążony, jeśli $E(X) = X^0.$ Zatem jest to estymator szacujący wartość parametru X^0 bez błędu systematycznego. W dalszej części artykułu wykorzystywane będą estymatory o powyższych własnościach.

Wyznaczona wartość estymatora punktowego służy do określenia tzw. przedziału ufności. Przedziałem ufności dla parametru X^0 jest losowy przedział wyznaczony za pomocą rozkładu estymatora X o tej własności, że z dużym z góry zadanym prawdopodobieństwem równym $(1 - \alpha)$ pokrywa wartość szacowanego parametru X^0, tzn.

$$P(X_d < X^0 < X_g) = 1 - \alpha \tag{1}$$

Prawdopodobieństwo $1 - \alpha$ nazywa się współczynnikiem (poziomem) ufności. Wartość α może być interpretowana jako ryzyko, że wyznaczony przedział ufności nie pokryje szacowanego parametru X^0. Wielkości X_d i X_g oznaczają – odpowiednio – dolną i górną granicę przedziału ufności dla parametru X^0. We wzorach stosowanych do wyznaczania granic przedziału ufności bardzo często występuje wielkość oznaczana jako u_x (np. u_α, $u_{1-\alpha/2}$). Przez u_x rozumie się dalej kwantyl rozkładu normalnego odczytywany z tablic dystrybuanty rozkładu $N(0, 1)$, tak aby spełniona była równość $\Phi(u_x) = u_x$, gdzie Φ to dystrybuanta rozkładu $N(0, 1)$. Technikę konstrukcji przedziału ufności dla nieznanej wartości X^0 badanej zmiennej losowej $\boldsymbol{X}$ dla przypadku ogólnego przedstawiono w pracy [18]. Aby ocenić dokładność szacunku na poziomie ufności $(1 - \alpha)$, wyznacza się błąd szacunku. Jeśli wyznaczany przedział jest symetryczny, to połowa jego szerokości umownie jest nazywana błędem bezwzględnym szacunku i wynosi $\varepsilon = \left(X_g - X_d\right)/2.$ Drugą oceną precyzji szacunku jest δ – błąd względny szacunku na poziomie ufności $(1 - \alpha)$ równy $\delta = \varepsilon / |X|$.

1.1. Szacowanie intensywności uszkodzeń

Dla obiektów wodociągowych i kanalizacyjnych, na podstawie przesłanek fizycznych lub weryfikacji prowadzonych wielokrotnie przez różnych badaczy, powszechnie przyjmuje się założenie, że przyczyny ich uszkodzeń są losowe i nie wynikają ze starzenia się obiektów i zużycia materiałów. Jest to równoważne z przyjęciem, że badania dotyczą tzw.

normalnego okresu eksploatacji tych obiektów, czyli że intensywność uszkodzeń jest stała w czasie [17]. Dla obiektów odnawialnych przyjmuje się [17, 10], że:

1) zmienne losowe $\boldsymbol{Tp}$ i $\boldsymbol{Tn}$ opisujące – odpowiednio – czasy sprawności i czasy niesprawności obiektu mają rozkłady wykładnicze,
2) średni czas sprawności $Tp^0 = E(\boldsymbol{Tp})$ jest znacznie większy (co najmniej o rząd lub dwa rzędy wielkości) od średniego czasu niesprawności $Tn^0 = E(\boldsymbol{Tn})$, gdzie E to symbol wartości oczekiwanej.

Jeśli można uznać, że czas niesprawności jest pomijalny ($Tn^0 \approx 0$), to zmienna losowa $\boldsymbol{U}$ opisująca liczbę uszkodzeń w przyjętej jednostce czasu (np. 1 rok) ma rozkład Poissona z parametrem λ^0 równym $\lambda^0 = E(\boldsymbol{U})$. Wówczas prawdopodobieństwo zajścia k awarii (w jednostce czasu) wynosi

$$P(U = k) = \frac{\lambda^{0\,k}}{k!} e^{-\lambda^0} \tag{2}$$

Zachodzi przy tym zależność

$$\lambda^0 = \frac{1}{Tp^0} \tag{3}$$

Parametr λ^0, nazywany tutaj intensywnością uszkodzeń, jest interpretowany jako średnia liczba uszkodzeń obiektu w jednostce czasu. Liczba zdarzeń w przedziale o długości czasu Δt jest zmienną losową o rozkładzie Poissona z parametrem $\Lambda^0 = \lambda^0 \cdot \Delta t$.

Punktowy estymator intensywności uszkodzeń obiektu w praktyce wyznacza się jako

$$\lambda = \frac{k}{t} \tag{4}$$

gdzie k to liczba uszkodzeń (awarii) obiektu, które zaszły w czasie t.

Ze względu na ergodyczność procesu uszkodzeń parametr można wyznaczać na podstawie informacji o uszkodzeniach n obiektów w czasie t jako średnią

$$\lambda = \frac{\sum_{i=1,..n} k_i}{n \cdot t} \tag{5}$$

gdzie:

k_i – liczba awarii i-tego elementu w czasie t,
n – liczba jednorodnych obiektów, pracujących w analogicznych warunkach.

Dla elementów liniowych (np. przewodów sieci wodociągowej) operuje się tzw. jednostkową intensywnością uszkodzeń (odniesioną do jednostki czasu i jednostki długości) oznaczoną tutaj jako λ_{jedn} i wyznaczaną jako

$$\lambda_{\text{jedn}} = \frac{k}{t \cdot L} \tag{6}$$

gdzie L to długość obiektu.

Aby w praktyce można było stosować rozkład Poissona, muszą być spełnione warunki:
1) próba jest liczna $n \geq 50$,
2) prawdopodobieństwo wystąpienia zdarzenia (np. awarii) równe $p = \lambda/n$ jest niewielkie $p < 0{,}1$.

Jak wiadomo, wyznaczenie estymatora punktowego nie zawsze jest wystarczające. Dlatego potrzebne jest dokonywanie estymacji przedziałowej, czyli wyznaczenie przedziału, który z wysokim prawdopodobieństwem (np. 0,95) będzie pokrywać nieznaną szacowaną wartość parametru λ^0.

Przedział ufności (λ_d, λ_g) dla parametru λ na poziomie ufności $1 - \alpha$ wyznaczają każde takie dwie wielkości λ_d oraz λ_g, dla których

$$P\left(\lambda_d < \lambda^0 < \lambda_g\right) = 1 - \alpha \tag{7}$$

Zazwyczaj przyjmuje się takie granice dwustronnego przedziału ufności, by

$$P\left\{\lambda_d < \lambda^0\right\} = P\left\{\lambda^0 < \lambda_g\right\} = 1 - \alpha/2 \tag{8}$$

Wówczas (0, λ_g) oraz (λ_d, ∞) są jednostronnymi przedziałami ufności dla parametru λ^0 na poziomie ufności $1 - \alpha/2$. Granice przedziału ufności można odczytywać z tablic ([19], tab. 20a dla $k \leq 50$, tab. 20b dla $k > 50$) lub wyznaczać za pomocą podanych poniżej wzorów.

Ze względu na następujący związek rozkładu Poissona i rozkładu χ^2

$$P(\mathbf{U} \leq k) = \sum_{i=0}^{k} p(i) = 1 - F_{2(k+1)}(2\lambda^0) \tag{9}$$

(gdzie $F_r(x)$ jest dystrybuantą rozkładu χ^2 o r stopniach swobody) granice przedziału ufności można wyznaczać według wzorów [19]

$$\lambda_d = \frac{1}{2}\chi^2\left(1 - \alpha/2; 2k\right) \quad \text{oraz} \quad \lambda_g = \frac{1}{2}\chi^2\left(\alpha/2; 2k\right) \tag{10}$$

gdzie $\chi^2(\alpha;\, r)$ jest wartością krytyczną rozkładu χ^2 o r stopniach swobody (odczytywaną z tablic [19]).

Dla dużych wartości k ($k > 50$) można korzystać z przybliżenia za pomocą rozkładu normalnego. Wówczas [19]

$$\lambda_d = k - \frac{1}{2} + \frac{1}{4}u_{1-\alpha/2}^2 - u_{1-\alpha/2} \cdot \sqrt{k - \frac{1}{2}} \quad \text{oraz} \quad \lambda_g = k + \frac{1}{2} + \frac{1}{4}u_{\alpha/2}^2 - u_{\alpha/2} \cdot \sqrt{k + \frac{1}{2}} \tag{11}$$

gdzie u_x to kwantyl rzędu x rozkładu normalnego $N(0, 1)$.

Przedział ufności dla parametru rozkładu Poissona można wyznaczać również za pomocą innych wzorów [13]

$$\lambda_d = \lambda + \frac{u_\alpha}{2n} - R \quad \text{oraz} \quad \lambda_g = \lambda + \frac{u_\alpha}{2n} + R \tag{12}$$

gdzie:

n – liczność próby,

$R = \frac{u_\alpha^2}{\sqrt{n}}\sqrt{\lambda + \frac{u_\alpha^2}{4n}}$.

Wartość kwantyla u_α odczytuje się z tablic rozkładu t-Studenta dla prawdopodobieństwa α i przy nieskończonej liczbie stopni swobody (równoważnie z tablic rozkładu normalnego).

Przykład 1. W ciągu jednego roku zarejestrowano $a = 4$ awarie pewnego obiektu (np. pompy osadowej). Stąd na podstawie (4) mamy $\lambda = 4\,[1/\text{rok}]$. Na podstawie tablic ([19], tab. 20a dla $k \leq 50$) 95% przedział ufności dla średniej liczby awarii w roku wynosi (1,09; 10,24). Natomiast 90% przedział ufności dla średniej liczby awarii w roku wynosi (1,37; 9,15). Jak widać, w obu wypadkach przedział ufności jest dość szeroki, co wynika z małej liczby danych. Tak szeroki przedział jest nieprzydatny do analiz. Aby uzyskać bardziej przydatny wynik (węższy przedział), należałoby zwiększyć liczbę danych. Dla dłuższego okresu obserwacji (t = 5 lat) lub, równoważnie, obserwacji większej liczby jednorodnych obiektów pracujących w podobnych warunkach (n = 5) zanotowano $k_1 = 3$, $k_2 = 5$, $k_3 = 6$, $k_4 = 2$ i $k_5 = 4$ awarie (razem 20 awarii). Średnia liczba awarii w ciągu roku zgodnie z (5) wynosi $\lambda = \frac{20}{5} = 4$. Na podstawie tablic ([19], tab. 20a) 95% przedział ufności dla łącznej liczby awarii (dla jednego obiektu w ciągu 5 lat lub dla 5 obiektów w ciągu roku) wynosi (12,22; 30,89). Stąd 95% przedział ufności dla średniej liczby awarii jednego obiektu w ciągu roku wynosi (12,22/5; 30,89/5) = (2,444; 5,576). Zatem zwiększenie liczby danych (wydłużenie okresu badań lub zwiększenie liczby badanych obiektów) wpływa na skrócenie szerokości przedziału ufności.

Przykład 2. W ciągu roku zarejestrowano $k = 148$ awarii pewnej sieci rozdzielczej. Na podstawie tablic ([19], tab. 20b dla $k > 50$) 95% przedział ufności dla średniej liczby awarii w roku wynosi (125,125; 173,848). Stąd 95% przedział ufności dla średniej liczby awarii w ciągu doby wynosi (0,34; 0,47). Na tej podstawie można wnioskować, że z prawdopodobieństwem 95% awaria przewodów rozdzielczych wystąpi średnio raz na 2–3 doby.

Przykład 3. W pracy [14] analizowano zanotowane w latach 1979–2002 przypadki braku zasilania w Zakładzie Uzdatniania Wody Rudawa w Krakowie. W ciągu 24 lat było 19 takich istotnych zdarzeń (trwających co najmniej 1 h). Wykazano, że te istotne dla funkcjonowania zakładu przypadki można opisać za pomocą rozkładu Poissona o intensywności uszkodzeń równej $\lambda = \frac{19}{24} = 0{,}7917\,[1/\text{rok}]$. Poniżej wyznaczono przedział ufności dla parametru λ^0 na poziomie ufności 0,95 za pomocą trzech opisanych powyżej metod:

1) z tablic ([19], tab. 20a) odczytano granice przedziału ufności (dla 24 lat), czyli dla liczby zdarzeń $k = 19$; są one równe 11,44 oraz 29,67; natomiast przedział ufności dla parametru λ^0 odniesionego do jednego roku można wyznaczyć jako (11,44/24; 29,67/24); stąd $\lambda_d = 0{,}477$ oraz $\lambda_g = 1{,}236$,
2) wykorzystując związek z rozkładem χ^2 dla $k = 19$, otrzymujemy (dla 24 lat) $\lambda_d = 11{,}439$ i $\lambda_g = 29{,}671$. Stąd dla roku mamy $\lambda_d = 0{,}477$ i $\lambda_g = 1{,}236$,
3) na podstawie wzoru (12) cytowanego z pracy [13] po przyjęciu $n = 24$ i $k = 19$ wyznacza się kolejno: $u_{0,95} = 1{,}96$, następnie $R = 0{,}3649$ oraz $\lambda_d = 0{,}507$ i $\lambda_g = 1{,}237$.

Przykład 4. W pracy [14] analizowano zanotowane w latach 1997–2001 przypadki uszkodzeń przewodów sieci wodociągowej miasta Krakowa. Wykazano, że dla pewnych grup średnic (np. 80–150, 500 mm) liczbę uszkodzeń można opisać za pomocą rozkładu Poissona. Dla pozostałych średnic nie można było na podstawie posiadanych danych przyjąć takiego założenia. Powodem był fakt, że np.:

1) sumaryczna długość przewodów o danej średnicy była niewielka (np. dla 750 mm było to 2,09 km, dla 800 mm – 39,38 km), wówczas może nie być spełnione założenie stosowalności rozkładu Poissona (zbyt mała liczba elementów, przy czym jako element przyjmuje się odcinek sieci między zasuwami [17]),
2) nastąpił znaczny wzrost długości sieci w ciągu tych 5 lat (np. dla średnicy 1000 mm przyrost wyniósł prawie 90%),
3) w badanym okresie nie odnotowano ani jednej awarii przewodów o danej średnicy.

Poniżej oszacowano jednostkową intensywność uszkodzeń przewodów sieci o średnicy 500 mm. Sumaryczne długości tych przewodów w kolejnych latach wynosiły, odpowiednio: 33,24 km; 33,82 km; 33,82 km; 34,04 km i 34,04 km. Przyrost długości przewodów Φ500 w analizowanym okresie był niewielki – wyniósł ok. 2,41%. W kolejnych latach omawianego okresu zanotowano 32, 29, 25, 33 i 23 uszkodzenia. W celu wyznaczenia punktowego estymatora jednostkowej intensywności uszkodzeń wykorzystano wzór (6) w dostosowanej postaci uwzględniającej przyrost długości przewodów

$$\lambda_{\text{jedn}} = \frac{\sum_{i=1}^{5} k_i}{\sum_{i=1}^{5} L_i} = \frac{32 + \ldots + 23}{33{,}24 + \ldots + 34{,}04} = \frac{142}{168{,}96} = 0{,}8404 \left[\frac{1}{\text{km} \cdot \text{rok}} \right]$$

Następnie za pomocą różnych metod wyznaczono przedział ufności:

1) za pomocą tablic (tab. 20b, [19]) dla całego okresu 5 lat mamy $k = 142$, stąd $\lambda_{\text{jedn}\ d} = 119{,}604$ i $\lambda_{\text{jedn}\ g} = 167{,}367$; dla roku i w przeliczeniu na 1 km przewodów jest więc $\lambda_{\text{jedn}\ d} = 0{,}708$ i $\lambda_{\text{jedn}\ g} = 0{,}991$,
2) ponieważ łączna liczba awarii jest znaczna ($k = 142$), więc przybliżone granice przedziału ufności można wyznaczyć z przybliżenia (11). Dla poziomu ufności $1 - \alpha = 0{,}95$ mamy $1 - \alpha/2 = 0{,}975$ oraz $\alpha/2 = 0{,}0{,}25$. Z tablic rozkładu normalnego odczytujemy $u_{0,975} = 1{,}96$ oraz $u_{0,025} = -1{,}96$. Po podstawieniu do (11) dla wszystkich przewodów o Φ500 i dla 5 lat mamy $\lambda_d = 118{,}969$ i $\lambda_g = 166{,}857$. Stąd dla roku oraz w przeliczeniu na 1 km przewodów jest więc $\lambda_{\text{jedn}\ d} = 0{,}704$, $\lambda_{\text{jedn}\ g} = 0{,}988$. Jak widać, uzyskano wynik zbliżony do otrzymanego poprzednio.

Powyżej przedstawiono metody wyznaczania przedziału ufności intensywności uszkodzeń dla przypadku, gdy są spełnione przedstawione wcześniej założenia stosowalności rozkładu Poissona. Jak wykazuje praktyka, nie wszystkie zdarzenia awaryjne spełniają ten warunek. Pozostaje więc pytanie, jak sobie radzić w pozostałych sytuacjach. Poniżej przedstawiono kilka możliwości.

Jeśli wielkość próby n jest znaczna, a rozkład liczby uszkodzeń jest dowolny o nieznanej wartości średniej i skończonej wariancji σ, to na podstawie centralnego twierdzenia granicznego statystyka $Y = \frac{\lambda - \lambda^0}{\sigma}\sqrt{k}$ ma asymptotyczny rozkład $N(0, 1)$. Wówczas przedział ufności dla średniej liczby uszkodzeń (nazywanej nadal intensywnością uszkodzeń) jest określony przez wartości

$$\lambda_d = \lambda - u_{1-\alpha/2} \cdot \frac{s^*}{\sqrt{n}} \quad \text{oraz} \quad \lambda_g = \lambda + u_{1-\alpha/2} \cdot \frac{s^*}{\sqrt{n}} \tag{13}$$

gdzie:

s^* – oszacowanie wariancji σ z próby,

$s^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \lambda)^2$ [9].

Stosowanie tej metody wymaga dużej liczebności próby. Warunek ten w literaturze ma różne znaczenie (np. $n > 120$ [16], $n \geq 100$ [9], $n > 50$ lub n conajmniej kilkadziesiąt [6]). Jednak gdy liczba danych jest stosunkowo niewielka, to tę trudność można czasem ominąć (np. przy dokładnej znajomości rozkładu liczby uszkodzeń w czasie zasadnicze obliczenia można przeprowadzać w odniesieniu do jednej doby zamiast do jednego roku).

Przykład 5. W ciągu roku w pewnej sieci dystrybucji zanotowano $k = 23$ awarie armatury określone jako „nie na przewodzie”. Są to np. awarie hydrantów czy zasuw. Łączna liczba tego typu obiektów jest znaczna, lecz nie można ich traktować jako jednorodnych. Nie można więc wyznaczać intensywności uszkodzeń pojedynczego obiektu. Można natomiast oszacować intensywność uszkodzeń całej grupy obiektów, co może być przydatne przy planowaniu optymalnej strategii eksploatacji sieci dystrybucji. Punktowym estymatorem intensywności uszkodzeń grupy obiektów zgodnie z (4) jest $\lambda = 23\,[1/\text{rok}] = 0{,}063\,[1/\text{d}]$. Przedziały ufności dla λ^0 można wyznaczyć następująco:

1) odczytany z tablic ([19], tab. 20a) przedział ufności na poziomie 95% wynosi (14,58; 34,51) [1/rok]. Szerokość niesymetrycznego przedziału ufności wynosi 19,33 [1/rok],
2) analizując próbkę o wielkości $n = 365$, w której danymi są liczby zanotowanych awarii w odpowiedniej dobie roku; przykładowo, jeśli wiadomo, że w ciągu tego roku tylko podczas jednej doby były dwie awarie, po jednej awarii zaszło w 21 dobach, a w ciągu 342 dób nie było żadnych awarii, wówczas oszacowanie wariancji intensywności uszkodzeń wynosi

$$s^{*2} = \frac{1}{365-1}\left((0-0{,}063)^2 + 21\cdot(1-0{,}063)^2 + 1\cdot(2-0{,}063)^2\right) = 0{,}06469$$

Odczytując z tablic kwantyle rozkładu normalnego $u_{0,025} = -1,96$ oraz $u_{0,975} = 1,96$ i stosując (13), uzyskuje się $\lambda_d = 0,0369$ [1/d] oraz $\lambda_g = 0,0891$ [1/d]. Po przeliczeniu uzyskuje się $\lambda_d = 13,48$ [1/rok] oraz $\lambda_g = 32,524$ [1/rok]. Szerokość symetrycznego przedziału ufności wynosi 19,05. Jest to wynik zbliżony do wyniku odczytanego na podstawie tablic.

Jeśli problem dotyczy sieci z istotnymi przyrostami długości w analizowanym okresie, to dysponując dokładnymi informacjami o miejscach i przyczynach występowania awarii, zagadnienie niekiedy można sprowadzić do rozkładu Poissona. Uzasadnienie może być następujące: przy rozbudowie sieci i zastosowaniu przewodów dobrej jakości i przy profesjonalnym wykonawstwie przewody nowe mogłyby ulec uszkodzeniu tylko w wyjątkowych wypadkach (np. wskutek nieostrożnie prowadzonych innych prac budowlanych lub katastrofy kanalizacyjnej w pobliżu). Jeśli nie stwierdzono takich przypadków, to można założyć, że awarie odnotowane po rozbudowie sieci dotyczą tylko przewodów starych (co najmniej kilkunastoletnich) i odnosić je do długości tych przewodów z początkowego okresu badań.

Jeśli nie można pominąć czasów odnowy (tzn. są znaczące), to zamiast rozpatrywania pełnego okresu badań t można analizować liczbę awarii w czasie t' pomijającym niesprawności.

Przykład 6. W ciągu roku (t = 1 rok = 365 · 24 godz.) zarejestrowano $k = 5$ awarii pewnego obiektu. Kolejne czasy niesprawności wynosiły: 22 h, 35 h, 39 h, 32 h, 42 h Sumaryczny czas niesprawności obiektu wyniósł $\sum_{i=1}^{5} t_{n\,i} = 170$ h Pozostały czas sprawności wynosi $t' = 365 \cdot 24 - 170 = 8590$ h. Stąd $\lambda = 5\,[1/t'] = \;\; 5,1\,[1/\text{rok}]$. Na podstawie tablic ([19], tab. 20a dla $k \leq 50$) interpolowany liniowo 95% przedział ufności dla średniej liczby awarii w roku wynosi (1,678; 11,809).

1.2. Szacowanie średniego czasu pracy

W praktyce szacowanie parametru rozkładu wykładniczego zależy od rodzaju obiektów (nieodnawialne, odnawialne) i liczby danych (próba pełna, ucięta). Dla obiektów nieodnawialnych sposób estymacji zależy też od planu badań (np. [n, B, t], [n, B, r]). Szczegółowe informacje zawarto w [3] oraz [12]. W dalszej części artykułu, uwzględniając specyfikę większości obiektów wodociągowych i kanalizacyjnych, analizuje się obiekty odnawialne.

Jeśli próba jest dostatecznie duża (co najmniej kilkadziesiąt danych), to można wykorzystać wzory dla planu [n, B, n]. Wówczas punktowym estymatorem średniego czasu pracy między uszkodzeniami jest

$$Tp = \frac{\sum_{i=1}^{n} tp_i}{n} \tag{14}$$

gdzie:

n – wielkość próby,

tp_i – i-ta realizacja zmiennej losowej $\boldsymbol{Tp}$.

Wobec faktu, że $Tp^0 = \frac{1}{\lambda^0}$ (gdzie λ^0 to intensywność uszkodzeń) końce przedziału ufności dla parametru Tp^0 są równe [3]

$$Tp_d = \frac{1}{\lambda_g} \quad \text{oraz} \quad Tp_g = \frac{1}{\lambda_d} \tag{15}$$

W wypadku dużej liczby danych granice przedziału ufności Tp^0 można wyznaczyć wprost za pomocą modelu wykorzystującego graniczny rozkład normalny.

Przykład 7. W pewnym przedsiębiorstwie wodociągowym na podstawie ksiąg eksploatacji stwierdzono, że w okresie 5 lat i 8 miesięcy wystąpiło $k = 51$ awarii o łącznym czasie trwania 1539,6 h. Pod uwagę wzięto 50 odcinków pracy między awariami. Czasy te zawierały się między 60 h i 3200 h. Średni czas pracy oszacowano za pomocą wzoru (14) przez $Tp = 48100 / 50 = 962$ h. Ponieważ próba jest duża, więc granice przedziału ufności dla Tp^0 wyznaczono za pomocą modelu wykorzystującego jako graniczny rozkład normalny. Wyznaczono wariancję z próby $s^{*2} = 262277{,}9$. Stąd $s^* = 512{,}13$. Na podstawie wzorów analogicznych do (13) granice 95% przedziału ufności są równe: Tp_d = 823,76 i Tp_g = 1100,25 h. Dodatkowo dla porównania przedział ufności wyznaczono przez wyznaczanie przedziału ufności dla intensywności uszkodzeń. W tym celu wyznaczono wielkości pomocnicze $z = 1/(2\sqrt{k})$ oraz z tablic ([19], tab. 20b dla $a > 50$) dla tego samego poziomu ufności 95% wartości r_1 = 0,97; r_2 = 0,059 i ψ = 1,959964. Za pomocą wzorów $\lambda_d = k - \Psi \cdot \sqrt{k} + r_1$ oraz $\lambda_g = k + \Psi \cdot \sqrt{k} + r_2$ obliczono λ_d = 37,973 i λ_g = 65,056. Ten przedział dla parametru λ odnosi się do sumarycznego czasu pracy równego 48 100 h. Po przeliczeniu uzyskuje się granice przedziału ufności dla intensywności uszkodzeń w odniesieniu do roku równe λ_d = 0,000789 i λ_g = 0,001353. Na podstawie (15) granicami przedziału ufności średniego czasu pracy między uszkodzeniami są Tp_d = 739,37 oraz Tp_g = 1266,70. Jak widać, te dwie metody dają podobne wyniki.

Jeśli próba jest niewielka, to estymacja średniego czasu pracy nie jest prosta. Należy zaznaczyć, że pewne obiekty wodociągowe i kanalizacyjne są wysoce niezawodne. Dla nich pomimo długiego okresu obserwacji (równoważnie okresu rejestrowania awarii w księgach eksploatacji odpowiednich służb) próby nie są duże, w skrajnym wypadku mogło nawet nie zajść ani jedno uszkodzenie. Wówczas problem jest bardziej złożony. Poniżej krótko przedstawiono idee dwóch metod, a zainteresowanym poleca się prace [1, 4] oraz dokładne przestudiowanie zamieszczonych tam przykładów.

Pierwsza metoda ma zastosowanie tyko do badań niepełnych. Znana jest wówczas wyłącznie część realizacji zmiennej losowej Tp. Realizacje te są równe tp_1, tp_2, ..., tp_m. O pozostałych realizacjach wiadomo tylko tyle, że spełniają pewną nierówność (np. $tp_j > TB$ lub $tp_j > \tau_j$) oznaczającą, że obiekt był sprawny co najmniej przez całkowity czas badania TB (w wypadku, gdy w ogóle nie uległ uszkodzeniu) lub przez czas pozostały do końca badań (gdy obiekt przynajmniej raz uległ uszkodzeniu). Wówczas granicą jednostronnego przedziału ufności dla Tp^0 jest taka najmniejsza wartość Tp_d, która spełnia nierówność

$$T = \frac{1}{\lambda} \geq -\frac{t_\Sigma}{k \cdot \ln \alpha} \tag{16}$$

gdzie t_Σ to suma niepełnych czasów pracy.

Metoda może być stosowana nawet wtedy, gdy nie odnotowano ani jednego uszkodzenia, lecz wówczas przedział ufności jest bardzo szeroki i w praktyce mało przydatny.

Druga metoda może być zastosowana zarówno do badań pełnych, jak i niepełnych. Analizuje się zmienną losową $\boldsymbol{T}_{\Sigma}(k)$ opisującą sumaryczny czas do k-tego uszkodzenia. Wiadomo, że zmienna losowa $\boldsymbol{T}_{\Sigma}(k)$ będąca sumą k zmiennych losowych o rozkładzie wykładniczym ma rozkład Erlanga, czyli rozkład gamma z parametrem k naturalnym, o gęstości [4]

$$f(t,\lambda)=\frac{\lambda^k t^{k-1}\exp(-\lambda t)}{\Gamma^{(k)}}=\frac{\lambda^k t^{k-1}\exp(-\lambda t)}{(k-1)!} \tag{17}$$

Granice dwustronnego przedziału ufności dla czasu sumarycznego wyznacza się (np. numerycznie) z dwóch nierówności określających, że prawdopodobieństwo nieobjęcia przedziałem prawdziwej wartości czasu na poziomie ufności 1 – α wynosi $\alpha/2$, czyli

$$P(t_1 \le T)=\int_0^{t_1} f(t,\lambda)\,dt=\frac{\alpha}{2} \qquad \text{oraz} \qquad P(T \le t_2)=\int_{t_2}^{+\infty} f(t,\lambda)\,dt=\frac{\alpha}{2} \tag{18}$$

Następnie wyznacza się przedział ufności dla T_p^0. Obie metody, o ile to możliwe, należy stosować łącznie.

1.3. Szacowanie średniego czasu niesprawności

Przy wyznaczaniu miar niezawodności obiektów wodociągowych i kanalizacyjnych powszechnie przyjmuje się założenie, że zmienna losowa $\boldsymbol{Tn}$ opisująca czas niesprawności ma rozkład wykładniczy. Analiza rzeczywistych czasów niesprawności na ogół nie potwierdza takiej własności. Ponieważ jednak średni czas niesprawności jest dużo krótszy niż średni czas sprawności ($Tn \ll Tp$) najczęściej co najmniej o rząd lub dwa rzędy wielkości, a przyjęcie wspomnianego założenia bardzo upraszcza przeprowadzanie obliczeń, więc w praktyce założenie to przyjmuje się. Wówczas do szacowania $Tn^0 = E(\boldsymbol{Tn})$ możliwe jest stosowanie metod przedstawionych w rozdz. 1.4.

1.4. Szacowanie wskaźnika struktury

W rozdziałach 1.1–1.3 przedstawiono przypadki, gdy badana cecha miała charakter mierzalny (ilościowy), a wartość szacowanego parametru wyrażano w odpowiednich jednostkach (np. [1/rok], [d]). Czasem jednak badana cecha ma charakter niemierzalny (jakościowy). Wówczas z pojedynczego badania otrzymuje się jedynie informację o tym, czy dany obiekt ma wyróżnioną cechę jakościową, czy nie. Obiekty populacji ze względu na cechę jakościową mają rozkład dwupunktowy, a cała populacja ma rozkład dwumianowy. Parametrem, który charakteryzuje populację ze względu na daną cechę jest tzw. wskaźnik struktury w populacji lub inaczej frakcja. Parametr ten określa, jaka część populacji ma daną cechę niemierzalną. Po pomnożeniu przez 100 jest to procent (odsetek) obiektów wyróżniających się daną cechą w populacji. Wskaźnik struktury jest zwykle oznaczany przez p. Jego estymatorem punktowym jest [6]

$$p = \frac{m}{n} \tag{19}$$

gdzie:

m – liczba obiektów mających daną cechę,

n – liczebność próby.

Wskaźnik struktury może być interpretowany jako prawdopodobieństwo sukcesu, że wybrany losowo obiekt będzie miał daną cechę. Przedział ufności dla wskaźnika struktury wyznacza się różnie w zależności od wielkości próby [6]: dla próby małej (n < 100) za pomocą tablic, dla próby dużej ($n \geq 100$) za pomocą wzorów.

W pierwszym wypadku ($n \leq 100$) stosowanie efektywnych wzorów jest kłopotliwe, więc granice przedziałów ufności wyznacza się z tablic ([19], tab. 19) zawierających wartości $p_1 = p_1(1-\alpha, k, n-k)$ zależne od poziomu ufności ($1 - \alpha$), liczby elementów w próbie mających daną cechę (k) oraz liczby elementów niemających tej cechy ($n - k$). Jeśli na poziomie ufności $1 - \alpha$ wyznacza się przedział dwustronny symetryczny, to należy przyjąć [19]

$$p_d = p_1(1-\alpha, k, n-k) \quad \text{oraz} \quad p_g = 1 - p_1(1-\alpha, n-k, k) \tag{20}$$

Natomiast przedziały (p_d, 1) i (0, p_g) są jednostronnymi przedziałami ufności dla parametru p na poziomie ufności $1 - \alpha/2$.

W drugim wypadku (n > 100) statystyka $p = \frac{m}{n}$ ma rozkład asymptotycznie normalny $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$. Wówczas dla współczynnika ufności $1 - \alpha$ granice są określone jako

$$p_d = p - u_\alpha \cdot \sqrt{\frac{p(1-p)}{n}} \quad \text{oraz} \quad p_g = p + u_\alpha \cdot \sqrt{\frac{p(1-p)}{n}} \tag{21}$$

gdzie u_α odczytuje się z tablic rozkładu $N(0, 1)$.

Przykład 8. Stwierdzono, że spośród n = 12 jednorodnych agregatów pompowych pracujących w analogicznych warunkach k = 3 nie spełnia pewnego kryterium (np. zbyt głośna praca, zbyt duża liczba awarii). Wówczas estymator punktowy wskaźnika p wynosi $p = 3/12 = 0{,}25$. Ponieważ próba jest mała (n = 12), więc przeprowadzenie estymacji przedziałowej wymaga skorzystania z tablic. Przyjmując poziom ufności $1 - \alpha = 0{,}95$, odczytujemy ([19], tab. 19) wartości $p_1(0{,}95; 3; 9) = 0{,}055$ oraz $p_1(0{,}95; 9; 3) = 0{,}428$. Stąd na podstawie (1) mamy $p_d = 0{,}055$ oraz $p_g = 1 - 0{,}428 = 0{,}572$. Zatem z ufnością 95% przedział liczbowy (0,055; 0,572) obejmuje nieznaną frakcję elementów spełniających przyjęte kryterium.

Przykład 9. Przedsiębiorstwo Wodociągów i Kanalizacji (PWiK) obsługuje kilkanaście tysięcy odbiorców indywidualnych. Do przeprowadzenia ankiety nt. jakości usług świadczonych przez PWiK wylosowano próbę liczącą n = 250 odbiorców. Okazało się, że było $k = 20$ negatywnych ocen (np. zbyt częste wyłączenia dostawy wody, zły smak wody). Stąd $p = {}^{20}\!/_{250} = 0{,}08$. Ponieważ próba była duża (n = 250), więc przedział ufności wyznaczono na podstawie wzorów (21), przyjęto poziom ufności 90%. Stąd $u_\alpha = 1{,}64$. Granice przedziału ufności wynoszą: $p_d = 0{,}0519$ oraz $p_g = 0{,}1081$. Na poziomie ufności 0,90 można powiedzieć, że odsetek niezadowolonych odbiorców jest nie mniejszy niż 5,19%, a nie większy niż 10,81%. Szerokość przedziału ufności wynosi $p_g - p_d = 5{,}62\%$. Natomiast bezwzględny błąd szacunku, wyrażony przez połowę przedziału ufności wynosi $\varepsilon = (p_g - p_d)/2 = 2{,}81\%$.

Przykład 10. Spośród zainstalowanych na pewnej sieci wodociągowej n = 150 obiektów pewnego typu (np. zasuwy, hydranty) w ciągu 3 lat uszkodzeniu nie uległ ani jeden taki obiekt. Dla poziomu ufności 95% odczytane z tablic ([19], tab. 19) mamy m.in. $p_1(0{,}95; 100; 0) = 0{,}964$ oraz $p_1(0{,}95; 200; 0) = 0{,}982$. Stąd po liniowej interpolacji $p_1(0{,}95; 150; 0) = 0{,}973$. Zadanie można sformułować i rozwiązać na dwa sposoby:

1) liczba obiektów, które mają daną cechę (tj. uległy uszkodzeniu) wynosi $k = 0$; stąd na podstawie tablic ([19], tab. 19) oraz wzoru (21) mamy $p_d = p_1(0{,}95; 0; 150) = 0$ oraz $p_g = 1 - p_1(0{,}95; 150; 0) = 1 - 0{,}973 = 0{,}027$; interpretacja: z wiarygodnością 95% nie więcej niż 2,7% obiektów tego typu może ulec uszkodzeniu w ciągu kolejnych 3 lat,
2) liczba obiektów, które mają daną cechę (tj. nie uległy uszkodzeniu) wynosi $k = 150$; stąd na podstawie tablic ([19], tab. 19) i wzoru (20) mamy $p_d = p_1(0{,}95; 150; 0) =$ $= 0{,}973$ oraz $p_g = 1 - p_1(0{,}95; 0; 1500) = 1 - 0 = 0$; interpretacja: z wiarygodnością 95% co najmniej 97,3% obiektów tego typu nie powinno ulec uszkodzeniu w ciągu kolejnych 3 lat.

Jak wiadomo, dokładność szacunku na danym poziomie ufności zależy od liczebności próby. Jeśli przy badaniu cechy niemierzalnej względy czasowe i finansowe nie stanowią istotnych ograniczeń, to możliwe jest pozyskanie próby wystarczająco dużej, aby zapewnić niewielki, określony bezwzględny błąd szacunku $\varepsilon = (p_g - p_d)/2$. Wówczas uzyskana z przekształcenia wzorów (21) minimalna liczebność próby wynosi

$$n = \frac{u_\alpha{}^2 \cdot p \cdot (1 - p)}{\varepsilon^2} \tag{22}$$

gdzie:

- ε – bezwzględny błąd szacunku (połowa przedziału ufności),
- p – wstępne oszacowanie frakcji elementów w próbie (przyjęte na podstawie próby pilotażowej; jeśli brak jest jakichkolwiek informacji, należy przyjąć wartość $p = 0{,}5$ maksymalizującą licznik wyrażenia (22)).

Przykład 11. Do jakiej liczby indywidualnych odbiorców wody należy skierować ankietę (zob. przykład 9), aby na poziomie ufności 90% zmniejszyć błąd szacunku do wartości 2% populacji. Na podstawie wzoru (22) mamy $n = \dfrac{1{,}64^2 \cdot 0{,}08 \cdot (1 - 0{,}08)}{0{,}02^2} = 494{,}88$.

Stąd ankietę należy skierować co najmniej do 495 osób. Gdyby nie było wstępnego szacunku p, minimalna liczebność próby wyniosłaby $n = \frac{1{,}64^2 \cdot 0{,}5^2}{0{,}02^2} = 1681$. Jak widać, na etapie wstępnym jakikolwiek szacunek wstępny jest pomocny.

2. Wnioski

W obu zamieszczonych w niniejszym zeszycie artykułach (cz. I i II) przedstawiono metody rachunku przybliżonego (w tym klasyczne prawo propagacji błędów) oraz wybrane metody statystyczne mogące posłużyć do oceny dokładności szacowanych parametrów niezawodnościowych. Choć różnorodność metod jest znaczna, to w każdym wypadku, aby uzyskać wyniki bardziej przydatne w praktycznym zastosowaniu (np. węższy przedział ufności), należy wykorzystać wszystkie możliwe informacje z maksymalnie długiego czasu. Jeśli to możliwe, to poleca się stosowanie kilku różnych metod i przeprowadzenie analizy uzyskanych wyników. W wypadku, gdy:

1) nie są spełnione założenia stosowalności podanych metod,
2) szacujemy parametry nieznanych lub trudnych do wyznaczenia rozkładów,

to wnioskowanie statystyczne może być ograniczone, mało wiarygodne, a nawet błędne.

Wówczas można stosować np. metody bootstrapowe należące do metod statystyki nieklasycznej [5]. Metody te zostały zaproponowane przez Efrona w 1979 r., a współcześnie wraz ze wzrostem możliwości obliczeniowych komputerów daje się zauważyć ich szybki rozwój. Są to metody symulacyjne, w których z oryginalnej próby empirycznej o wielkości n losuje się ze zwracaniem N sztucznych próbek n-elementowych ($N \geq 1000$). Jedna oryginalna próbka jest więc źródłem wielu innych wirtualnych próbek (*resampling*). Po skompletowaniu j-tej wirtualnej próbki wyznacza się estymator parametru X^0 równy X_j. Następnie, w celu punktowego oszacowania X^0 oblicza się średnią z tych estymatorów. Jest to tzw. bootstrapowy estymator parametru X^0. *Resampling* umożliwia również przeprowadzenie estymacji przedziałowej.

Praca została wykonana w ramach umowy Ś-3/314/BW/2008 pt. „Ocena dokładności parametrów niezawodnościowych systemów wodociągowych i kanalizacyjnych".

Literatura

[1] Bajer J., Iwanejko R., *Eksploatacyjne badania niezawodności podstawowych elementów uzbrojenia pompowni wodociągowych*, Instal (228) 10/2008.
[2] Björck A., Dahlaquist G., *Metody numeryczne*, PWN, Warszawa 1983.
[3] Bobrowski D., *Modele i metody matematyczne teorii niezawodności w przykładach i zadaniach*, WNT, Warszawa 1985.
[4] Dawidowicz A.L., Iwanejko R., Bajer J., *Metody wyznaczania oczekiwanego czasu poprawnej pracy między uszkodzeniami wysoce niezawodnych elementów pompowni*, Gospodarka Wodna 4(532), 1993.

[5] D o m a ń s k i Cz., P r u s k a K., *Nieklasyczne metody statystyczne*, Polskie Wydawnictwo Ekonomiczne, Warszawa 2000.

[6] G r e ń J., *Statystyka matematyczna. Modele i zadania*, PWN, Warszawa 1982.

[7] I w a n e j k o R., B u d z i ł o B., *Uwagi do dwuparametrycznej metody wyznaczania niezawodności obiektów wodociągowych*, Czasopismo Techniczne, z. 7-Ś/2003, Wyd. PK, Kraków 2003.

[8] I w a n e j k o R., *Program obliczeniowy „Błędy”*, materiały niepublikowane, Kraków 2008.

[9] K r y s i c k i W. i in., *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach. Część II – Statystyka matematyczna*, Wydawnictwo Naukowe PWN, Warszawa 1994.

[10] K w i e t n i e w s k i M., R o m a n M., K ł o s s - T r ę b a c z k i e w i c z H., *Niezawodność wodociągów i kanalizacji*, Arkady, Warszawa 1993.

[11] K w i e t n i e w s k i M., *Metodyka badań eksploatacyjnych sieci wodociągowych pod kątem niezawodności dostawy wody do odbiorców*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 1999.

[12] M i g d a l s k i J. (red.), *Poradnik niezawodności: podstawy matematyczne*, Wydawnictwo WEMA, Warszawa 1982.

[13] O k t a b a W., *Elementy statystyki matematycznej i metodyka doświadczalnictwa*, PWN, Warszawa 1980.

[14] P i o t r o w s k a M., *Ocena możliwości opisu zdarzeń losowych zachodzących w systemie zaopatrzenia w wodę za pomocą rozkładu Poissona*, praca dyplomowa na Wydziale Inżynierii Środowiska Politechniki Krakowskiej, Kraków 2002.

[15] P o ł o ż y G.N. i in., *Metody przybliżonych obliczeń*, WNT, Warszawa 1966.

[16] S o b c z y k M., *Statystyka*, PWN, Warszawa 1996.

[17] W i e c z y s t y A. i in., *Metody oceny i podnoszenia niezawodności działania komunalnych systemów zaopatrzenia w wodę*, Monografie Komitetu Inżynierii Środowiska Polskiej Akademii Nauk, Vol. 2, Kraków 2001.

[18] W ę g l a r c z y k S., *Metody statystyczne*, Wyd. PK, Kraków 1993.

[19] Z i e l i ń s k i R., *Tablice statystyczne*, PWN, Warszawa 1972.