Analiza ryzyka kosztowego robót remontowo-budowlanych w warunkach niepełnej informacji

Mgr inż. Michał Bętkowski, dr inż. Andrzej Pownuk
Wydział Budownictwa
Politechnika Śląska w Gliwicach
Michal.Betkowski@polsl.pl, Andrzej.Pownuk@polsl.pl

Streszczenie

Jednym z kluczowych problemów w procesie budowlanym jest oszacowania ryzyka. Podstawową barierą jest brak wiarygodnych statystycznie danych. W artykule autorzy przedstawili autorską koncepcje zmiennej losowej o parametrach rozmytych jako narzędzia łączącego zalety podejścia probabilistycznego i opartego o algebrę zbiorów rozmytych przy eliminacji podstawowych ograniczeń obu podejść.

Matematyczny opis ryzyka

Z ryzykiem mamy do czynienia wtedy, gdy poniesione koszty realizacji projektu K_C będą większe od założonego poziomu kosztów maksymalnych K_{CZ}.

$$\begin{cases} K_C \le K_{CZ}, \text{ brak ryzyka} \\ K_C > K_{CZ}, \text{ ryzyko} \end{cases} \tag{1}$$

W przypadku, gdy dysponujemy odpowiednio dużą liczbą informacji, to możemy statystycznie opisać wartości poniesionych kosztów realizacji K_C przy pomocy zmiennej losowej.

$$K_C : \Omega \ni \omega \to K_C(\omega) \in R \tag{2}$$

Teraz można zdefiniować ryzyko jako prawdopodobieństwo przekroczenia założonego poziomu kosztów.

$$R(K_{CZ}) = P\{\omega : K_C(\omega) > K_{CZ}\} \tag{3}$$

Jak widać ryzyko jest funkcją założonego poziomu kosztów K_{CZ}. Jeśli znamy dystrybuantę zmiennej losowej K_C,

$$\Phi_{K_C}(K_{CZ}) = P\{\omega : K_C(\omega) < K_{CZ}\} = \int_{-\infty}^{K_{CZ}} f_{K_C}(x)dx \tag{4}$$

to ryzyko może zostać określone następująco

$$R(K_{CZ}) = P\{\omega : K_C(\omega) > K_{CZ}\} = 1 - P\{\omega : K_C(\omega) \le K_{CZ}\} = 1 - \Phi_{K_C}(K_{CZ}) \tag{5}$$

Przedziałowe charakterystyki ryzyka

Z uwagi na różnego rodzaju niepewności bardzo trudno precyzyjnie przewidzieć całkowity koszt inwestycji budowlanej. Niepewności, jakie należy uwzględnić w tym procesie można ogólnie rzecz biorąc podzielić na trzy grupy:

- niepewności przypadkowe – wynikające z losowego charakteru wielu zjawisk,
- niepewności systematyczne – wynikają z braku wiarygodnego źródła informacji,
- błędy grube – zwykle są to różnego rodzaju pomyłki.

W przypadku, gdy mamy wystarczającą ilość danych niepewności przypadkowe można opisywać przy pomocy teorii prawdopodobieństwa. W takim przypadku zakładamy, że dysponujemy pewnym zbiorem pomiarów danego kosztu składowego K_i :

$$K_{i,1}, K_{i,2}, ..., K_{i,n} \tag{6}$$

Liczby (6) można uznać za realizacje pewnej zmiennej losowej. W takim przypadku można konstruować probabilistyczne charakterystyki (statystki) opisujące zmienną losową $K_i : \Omega \ni \omega \to K_i(\omega) \in R$. Najczęściej obliczamy wartość średnią i/lub wariancję:

$$\bar{K}_i = \sum_{j=1}^{N} K_{i,j}\ , \quad s_{K_i}^2 = \frac{1}{N-1}\sum_{j=1}^{N}\left(K_{i,j} - \bar{K}_i\right)^2 \tag{7}$$

Przy pomocy obliczonych statystyk możemy dokonywać dalszej obróbki statystycznej otrzymanych danych.

Należy jednak podkreślić, że każdy pomiar $K_{i,1}, K_{i,2}, ..., K_{i,n}$ obarczony jest również błędami systematycznymi, a czasem również grubymi. Zatem zamiast punktowej wartości pomiaru $K_{i,j}$ należy rozważyć cały przedział $\left[K_i^-, K_i^+\right]$ możliwych wartości, które może przyjmować

$$K_i \in \left[K_i^-, K_i^+\right] \tag{8}$$

Górną i dolną granice przedziału można zwykle oszacować na podstawie ogólnej wiedzy o danym koszcie K_i. Dowolna statystyka h zmiennej losowej K_i jest funkcją danych pomiarowych

$$h\left(K_i\right) = h\left(K_{i,1}, K_{i,2}, ..., K_{i,n}\right) \tag{9}$$

Zatem jeśli pomiary będą przedziałami $\hat{K}_{i,1}, \hat{K}_{i,2}, ..., \hat{K}_{i,n}$ zatem statystyka $h\left(K_i\right)$ też będzie przedziałem

$$h\left(\hat{K}_i\right) = h\left(\hat{K}_{i,1}, ..., \hat{K}_{i,n}\right) = \left\{h\left(K_{i,1}, ..., K_{i,n}\right) : K_{i,1} \in \hat{K}_{i,1}, ..., K_{i,n} \in \hat{K}_{i,n}\right\} \tag{10}$$

Przykładowo przedziałową wartość średnią można określić następująco:

$$\hat{\bar{K}}_i = \frac{1}{N}\sum_{j=1}^{N}\hat{K}_{i,j} = \left\{\frac{1}{N}\sum_{j=1}^{N}K_{i,j} : K_{i,j} \in \hat{K}_{i,j}\right\} \tag{11}$$

To samo dotyczy dowolnych innych probabilistycznych charakterystyk takich jak np. funkcji gęstości rozkładu prawdopodobieństwa $f_{K_i}(x)$. W tym przypadku otrzymujemy przedziałową funkcję gęstości rozkładu prawdopodobieństwa $\hat{f}_{K_i}(x)$.

$$\hat{f}_{K_i}(x) = \left\{ f_{K_i}\left(x, K_{i,1}, ..., K_{i,n}\right) : K_{i,j} \in \hat{K}_{i,j} \right\} \tag{12}$$

gdzie $f_{K_i}\left(x, K_{i,1}, ..., K_{i,n}\right)$ jest empiryczną funkcją gęstości otrzymaną na podstawie pomiarów $\hat{K}_{i,1}, \hat{K}_{i,2}, ..., \hat{K}_{i,n}$.

Ryzyko może zostać oszacowane jest na podstawie zbioru pewnych informacji dotyczących kosztów $K_1, ..., K_M$. Zbiór ten można przedstawić jako macierz o zmiennej liczbie elementów w każdym wierszu.

$$\mathbf{K} = \begin{bmatrix} K_{1,1}, ..., K_{1,N_1} \\ ... \\ K_{M,1}, ..., K_{M,N_N} \end{bmatrix}, \quad \hat{\mathbf{K}} = \begin{bmatrix} \hat{K}_{1,1}, ..., \hat{K}_{1,N_1} \\ ... \\ \hat{K}_{M,1}, ..., \hat{K}_{M,N_N} \end{bmatrix} \tag{13}$$

Zatem ryzyko jest funkcją zbioru danych **K**.

$$R\left(K_{CZ}\right) = R\left(K_{CZ}, \mathbf{K}\right) \tag{14}$$

gdzie K_{CZ} jest założonym poziomem kosztów oraz ryzyko określone jest przy pomocy wzoru (3). Zatem przedziałową wartość ryzyka $\hat{R}\left(K_{CZ}\right)$ otrzymujemy następująco:

$$\hat{R}\left(K_{CZ}\right) = \left[R^-\left(K_{CZ}\right), R^+\left(K_{CZ}\right)\right] = \left\{R\left(K_{CZ}, \mathbf{K}\right) : \mathbf{K} \in \hat{\mathbf{K}}\right\} \tag{15}$$

Rozmyte charakterystyki ryzyka

Przedziałowe charakterystyki ryzyka dają najbardziej pesymistyczne oszacowanie wartości ryzyka.

$$R^-\left(K_{CZ}\right) = \inf\left\{R\left(K_{CZ}, \mathbf{K}\right) : \mathbf{K} \in \hat{\mathbf{K}}\right\}, \quad R^+\left(K_{CZ}\right) = \sup\left\{R\left(K_{CZ}, \mathbf{K}\right) : \mathbf{K} \in \hat{\mathbf{K}}\right\} \tag{16}$$

Ponadto oszacowanie nie uwzględnia zależności pomiędzy różnymi wartościami $K_{i,j}$. Przykładowo błędy systematyczne zwykle, aczkolwiek nie zawsze, mają podobny wpływ na część pomiarów i nie trzeba uwzględniać całego zakresu zmienności przedziału $\left[K_{i,j}^-, K_{i,j}^+\right]$. Ponadto przedziały $\left[K_{i,j}^-, K_{i,j}^+\right]$ są czasem zbyt szerokie.

Częściowym rozwiązaniem tych problemu jest koncepcja zbiorów rozmytych [1]. W tym ujęciu zbiór danych nie składa się tylko z przedziałów.
Składa się z zagnieżdżonej rodziny danych przedziałowych $\mathbf{K}_\alpha$ (tzw. α -przekrojów).

$$\hat{\mathbf{K}}_{\alpha_0} \supseteq \hat{\mathbf{K}}_{\alpha_1} \supseteq ... \supseteq \hat{\mathbf{K}}_{\alpha_N} \tag{17}$$

Zakładamy przy tym, że przedziałowe dane $\hat{\mathbf{K}}_\alpha$ zostały utworzone na pewnym poziomie istotności α, czyli

$$P\left\{K_{i,j} \cap \left(R \setminus \hat{K}_{i,j,\alpha}\right) \neq \varnothing\right\} = P\left\{K_{i,j} \cap \left[\left(-\infty, K^-_{i,j,\alpha}\right) \cup \left(K^+_{i,j,\alpha}, +\infty\right)\right] \neq \varnothing\right\} = \alpha \qquad (18)$$

W tym przypadku dla każdego α -przekroju możemy obliczyć przedziałowe ryzyko.

$$\hat{R}_\alpha\left(K_{CZ}\right) = \left[R^-_\alpha\left(K_{CZ}\right), R^+_\alpha\left(K_{CZ}\right)\right] = \left\{R\left(K_{CZ}, \mathbf{K}\right) : \mathbf{K} \in \hat{\mathbf{K}}_\alpha\right\} \qquad (19)$$

Teraz można zbudować funkcję przynależności rozmytego ryzyka $R_F\left(K_{CZ}\right)$.

$$\mu_{R_F(K_{CZ})}\left(x\right) = \sup\left\{\alpha : x \in \hat{R}_\alpha\left(K_{CZ}\right)\right\} \qquad (20)$$

Zmienne losowe z rozmytymi parametrami

W dotychczasowych rozważaniach nie uwzględnialiśmy informacji na temat charakteru rozkładu prawdopodobieństwa. Jednak w pewnych zastosowaniach nie trzeba rozważać wszystkich możliwych funkcji gęstości i można się ograniczyć do pewnej klasy funkcji. Najczęściej mamy do czynienia np. z rozkładami normalnymi, beta i beta-Pert.

W takim przypadku znamy rozkład prawdopodobieństwa, a nie znamy jego parametrów. Przykładowo, parametry rozkładu normalnego określamy na podstawie wartości średniej z próby, oraz wariancji z próby. Na skutek występowania bardzo trudnych do usunięcia błędów grubych i systematycznych dane na temat kosztów $K_{i,j}$ są obarczone niepewnościami. Można je jednak oszacować przy pomocy przedziałów $\hat{K}_{i,j}$ lub bardziej dokładnie, przy pomocy liczb rozmytych o α -przekrojach $\hat{K}_{i,j,\alpha}$. Dlatego parametry rozkładu normalnego będą odpowiednio rozmytą (przedziałową) wartością średnią:

$$\hat{\bar{K}}_{i,\alpha} = \left\{\frac{1}{N}\sum_{j=1}^{N} K_{i,j} : K_{i,j} \in \hat{K}_{i,j,\alpha}\right\} \qquad (21)$$

oraz rozmytą (przedziałową) wariancją

$$\hat{s}^2_{i,\alpha} = \left\{\frac{1}{N-1}\sum_{j=1}^{N}\left(K_{i,j} - \bar{K}_{i,j}\right)^2 : K_{i,j} \in \hat{K}_{i,j,\alpha}\right\} \qquad (22)$$

W tak opisany sposób można otrzymać funkcję przynależności rozmytej wartości średniej $\bar{K}_{i,F}$ i wariancji $s^2_{i,F}$.

$$\mu_{\bar{K}_{i,F}}\left(x\right) = \sup\left\{\alpha : x \in \hat{\bar{K}}_{i,\alpha}\right\} \qquad (23)$$

$$\mu_{s^2_{i,F}}\left(x\right) = \sup\left\{\alpha : x \in \hat{s}^2_{i,\alpha}\right\} \qquad (24)$$

W podobny sposób można zdefiniować rozmyte charakterystyki dowolnych rozkładów prawdopodobieństwa.
Poprzez uwzględnienie dodatkowej informacji o charakterze rozkładu prawdopodobieństwa uzyskujemy zwykle węższe oszacowanie niż w przypadku zastosowania metody opisanej w poprzednim punkcie.
Jeśli mamy dany rozkład prawdopodobieństwa $f\left(x,\mathbf{h}\right)$ zależny od wektora parametrów $\mathbf{h}$, który charakteryzuje rozkład kosztu całkowitego inwestycji budowlanej, to ryzyko (zależne od parametru **h**) można obliczyć następująco

$$R\left(K_{CZ},\mathbf{h}\right)=\int_{K_{CZ}}^{\infty} f_{K_C}\left(x,\mathbf{h}\right)dx \qquad (25)$$

Teraz można obliczyć ryzyko jako liczbę rozmytą określoną przy pomocy wzoru (25) oraz metody α -przekrojów.

$$\hat{R}_{\alpha}\left(K_{CZ}\right)=\left\{R\left(K_{CZ},\mathbf{h}\right):\mathbf{h}\in\hat{\mathbf{h}}_{\alpha}\right\} \qquad (26)$$

$$\mu_{R_F\left(K_{CZ}\right)}\left(x\right)=\sup\left\{\alpha : x\in\hat{R}_{\alpha}\left(K_{CZ}\right)\right\} \qquad (27)$$

Tak określone rozmyte ryzyko $R_F\left(K_{CZ}\right)$ może zostać obliczone np. przy wykorzystaniu zmodyfikowanej metody Monte Carlo [2].

Literatura

[1] Neumaier A., Clouds, fuzzy sets and probability intervals, Reliable Computing 10: 2004.
[2] Będkowski M, Pownuk A., Szacowanie ryzyka kosztowego procesu budowlanego z wykorzystaniem zmiennej losowej o parametrach rozmytych w oparciu o metodę Monte-Carlo. Konferencja naukowo-techniczna pt. BUDOWNICTWO POLSKIE W ROK PO WSTAPIENIU DO UNII EUROPEJSKIEJ Gdańsk 9 - 11 czerwca 2005