

Koncepcja zmiennej losowej o parametrach rozmytych jako narzędzia do szacowania ryzyka procesów budowlanych w warunkach ograniczonej ilości danych

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 [10]. W artykule autorzy przedstawili autorską koncepcję zmiennej losowej o parametrach rozmytych jako narzędzia łączącego zalety podejścia probabilistycznego i opartego o algebrę zbiorów rozmytych przy eliminacji podstawowych ograniczeń obu podejść.

1 Wstęp

W gospodarce rynkowej celem działalności jest osiągnięcie pozytywnego efektu finansowego. Efekt ten jest dodatni jeśli poniesione koszty będą mniejsze od maksymalnego zakładanego poziomu, ryzyko jest więc możliwością przekroczenia go. Budownictwo jest dziedziną, w której realizacja procesów jest szczególnie narażona na czynniki ryzyka, ponieważ realizacja robót w warunkach o charakterze zdeterminowanym praktycznie nie istnieje. Do czynników losowych należą czynniki atmosferyczne, płynność kadry, awaryjność stosowanych maszyn i urządzeń technicznych itp. Zakres oddziaływania jest różny dla każdego z czynników i jest niejednakowy dla różnych budów, ponieważ zależy od rzeczywistych warunków realizacyjnych oraz od określonej sytuacji decyzyjnej [6]. Obecnie w szacowaniu kosztów robót budowlanych na ogół stosuje się proste metody deterministyczne [13]. Główną barierą w stosowaniu bardziej zaawansowanych metod [3] jest brak danych wejściowych lub ich nieprecyzyjność [9,10].

2 Ryzyko w procesach budowlanych

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

$$K_{CZ} < K_C \quad (1)$$

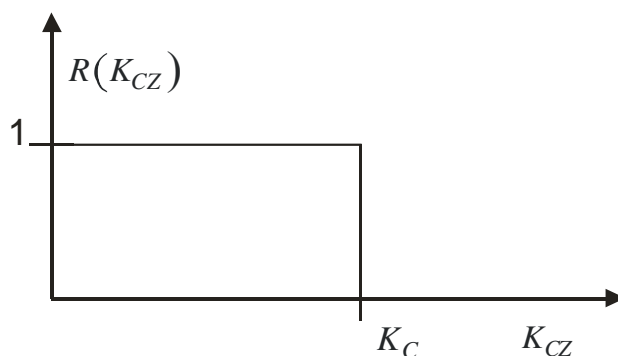
Zatem ryzyko można określić następująco

$$R(K_{CZ}) = \begin{cases} 1, & \text{gdy } K_{CZ} < K_C \\ 0, & \text{gdy } K_{CZ} \geq K_C \end{cases} \quad (2)$$

Jeśli traktować K_C jako zmienną losową o stałej wartości, $\forall_{\omega \in \Omega} K(\omega) = K_C = const$ to z probabilistycznego punktu widzenia można przyjąć, że:

$$R(K_{CZ}) = P\{K_{CZ} < K_C\} = \begin{cases} 1, & \text{gdy } K_{CZ} < K_C \\ 0, & \text{gdy } K_{CZ} \geq K_C \end{cases} \quad (3)$$

W przypadku, gdy wszystkie dane są ściśle określone, to ryzyko można zilustrować następującym rysunkiem.

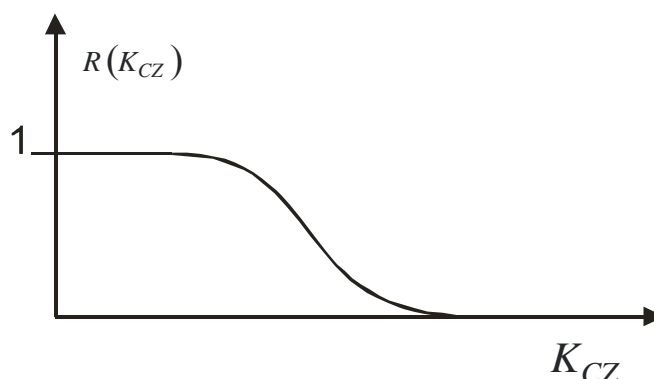


Rys. 1

Jeśli koszty składowe procesów K_i są zmiennymi losowymi $K_i : \Omega \ni \omega \rightarrow K_i(\omega) \in R$, to ryzyko można określić jako prawdopodobieństwo tego, że koszt całkowity K_C jest większa od założonego poziomu kosztów K_{CZ} .

$$R(K_{CZ}) = P\{K_{CZ} < K_C\} = P\{\omega : K_{CZ} < K_C(\omega)\} \quad (4)$$

Typowy kształt krzywej ryzyka przedstawiony jest na rysunku 2.



Rys. 2

Założymy teraz, że koszty procesów składowych są znane z dokładnością do przedziału liczbowego.

$$K_i \in [K_i^-, K_i^+] = \hat{K}_i \quad (5)$$

W takim przypadku koszt całkowity jest również przedziałem i może być obliczona z następującego wzoru.

$$\hat{K}_C = \sum_{i=1}^N \hat{K}_i \quad (6)$$

Przy obliczaniu przedziału $\hat{K}_C = [K_C^-, K_C^+]$ można wykorzystać arytmetykę przedziałową [12]. W tym przypadku na podstawie wzoru (4) można zdefiniować nieskończenie wiele krzywych ryzyka. Zbiór wszystkich krzywych ryzyka jest ograniczony z góry i z dołu przy pomocy funkcji $R^-(K_{CZ})$ (dolne ryzyko), $R^+(K_{CZ})$ (górne ryzyko). Funkcje $R^-(K_{CZ})$ $R^+(K_{CZ})$ można je zdefiniować następująco:

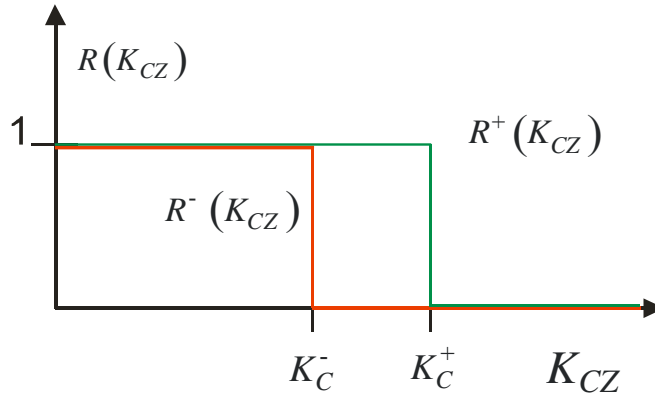
$$R^-(K_{CZ}) = \begin{cases} 1, & \text{gdy } \forall K_C \in \hat{K}_C : K_{CZ} < K_C \\ 0, & \text{w przeciwnym przypadku} \end{cases} \quad (7)$$

$$R^+(K_{CZ}) = \begin{cases} 1, & \text{gdy } \exists K_{CZ} \in \hat{K}_C : K_{CZ} < K_C \\ 0, & \text{w przeciwnym przypadku} \end{cases} \quad (8)$$

Można pokazać, że $R^-(K_{CZ}) \leq R^+(K_{CZ})$. Jeśli traktować K_C jako zmienną losową o stałej wartości przedziałowej $\forall \omega \in \Omega, \hat{K}_C(\omega) = \hat{K}_C = \widehat{const}$, to z probabilistycznego punktu widzenia można przyjąć, że:

$$R^-(K_{CZ}) = P\{\omega : \forall K_{CZ} \in \hat{K}_C(\omega), K_{CZ} < K_C\} = \begin{cases} 1, & \text{gdy } \forall K_C \in \hat{K}_C : K_{CZ} < K_C \\ 0, & \text{w przeciwnym przypadku} \end{cases} \quad (9)$$

$$R^+(K_{CZ}) = P\{\omega : \exists K_C \in \hat{K}_C(\omega), K_{CZ} < K_C\} = \begin{cases} 1, & \text{gdy } \exists K_C \in \hat{K}_C : K_{CZ} < K_C \\ 0, & \text{w przeciwnym przypadku} \end{cases} \quad (10)$$



Rys. 3

Jeśli mamy niepewne (np. przedziałowe) informacje na temat kosztów poszczególnych procesów K_i , w takim przypadku można opisać K_i przy pomocy liczb rozmytych.

Niech dane jest $\hat{K}_{i,1}, \hat{K}_{i,2}, \dots, \hat{K}_{i,M_i}$ informacji na temat kosztu K_i . Można teraz zbudować rodzinę przedziałów ufności zależnych od parametru $\alpha \in [0,1]$ takich, że [4, 11]:

$$P\{K_i \in (R \setminus \hat{K}_{i\alpha})\} \approx \frac{n(\{\hat{K}_{i,j} : \hat{K}_i \cap (R \setminus \hat{K}_{i\alpha}) \neq \emptyset\})}{M_i} = \alpha \quad (11)$$

$$\alpha_1 < \alpha_2 \Rightarrow \hat{K}_{i\alpha_1} \subseteq \hat{K}_{i\alpha_2} \quad (12)$$

gdzie $n(A)$ oznacza liczbę elementów zbioru A . Funkcję przynależności zbioru rozmytego można zdefiniować jako [15]:

$$\mu(x | K_{iF}) = \sup\{\alpha : x \in \hat{K}_{i\alpha}\} \quad (13)$$

gdzie K_{iF} oznacza zbiór rozmyty skojarzony z kosztem K_i oraz $\mu(x | K_{iF})$ jest jego funkcją przynależności. Jeśli wszystkie koszty są określone przy pomocy zbiorów rozmytych, wtedy można zdefiniować koszt całkowity przy wykorzystaniu algebry liczb rozmytych [8].

α -przekrojem liczby rozmytej K_{iF} nazywamy następujący przedział:

$$\hat{K}_{iF\alpha} = \{x : \mu(x | K_{iF}) \leq \alpha\} \quad (14)$$

Dla ustalonego stopnia przynależności α można obliczyć odpowiednie przedziały $\hat{K}_{i\alpha} = \hat{K}_{iF\alpha}$ do których mogą należeć poszczególne koszty K_i . Przedział, do którego należy koszt całkowity $\hat{K}_{C\alpha}$ można obliczyć następująco:

$$\hat{K}_{C\alpha} = \sum_{i=1}^N \hat{K}_{i\alpha} \quad (15)$$

Następnie można zdefiniować funkcje przynależności:

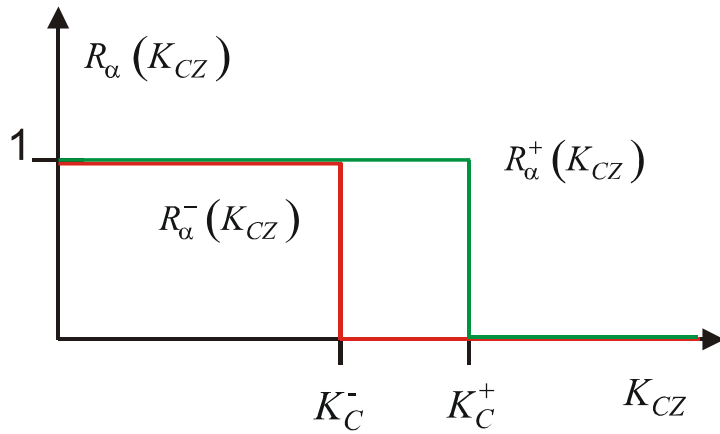
$$\mu(x | K_{CF}) = \sup \{ \alpha : x \in \hat{K}_{C\alpha} \} \quad (16)$$

Dla każdego α -przekroju liczby rozmyte są przedziałami liczbowymi, dlatego poprzez analogie do formuł (7, 8) górne i dolne rozmyte ryzyko można zdefiniować następująco:

$$R_{\alpha}^{-}(K_{CZ}) = \begin{cases} 1, & \text{gdy } \forall K_C \in \hat{K}_{C\alpha} : K_{CZ} < K_C \\ 0, & \text{w przeciwnym przypadku} \end{cases} \quad (17)$$

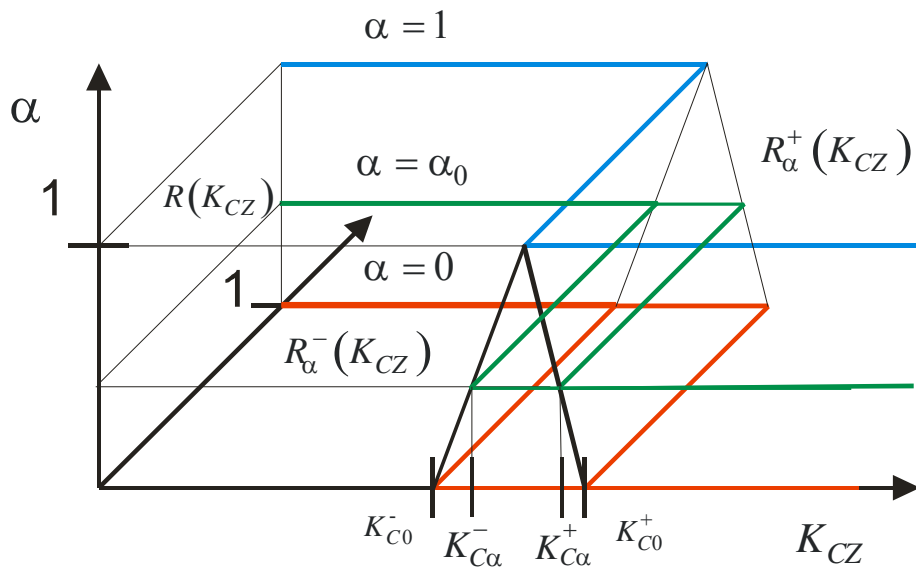
$$R_{\alpha}^{+}(K_{CZ}) = \begin{cases} 1, & \text{gdy } \exists K_C \in \hat{K}_{C\alpha} : K_{CZ} < K_C \\ 0, & \text{w przeciwnym przypadku} \end{cases} \quad (18)$$

Dla każdego α -przekroju krzywe ryzyka wyglądają jak na rys. 3.



Rys. 4

Zestawiając wszystkie krzywe ryzyka razem otrzymujemy bryłę przedstawioną na rysunku 5.



Rys. 5

3 Porównanie zastosowania zmiennych losowych oraz zmiennych rozmytych do modelowania niepewności kosztów

Każdy proces budowlany jest złożony z szeregu procesów składowych przyjmujących zmienne parametry. Charakter poszczególnych procesów możemy opisać przy pomocy zmiennych losowych, a w przypadku niedostatecznej liczby informacji możemy posłużyć się zmienną rozmytą. Każde z podejść ma szereg zalet i wad. Główną zaletą zmiennej rozmytej jest niewielka liczba informacji potrzebnych do jej skonstruowania. Niestety wady tego opisu ujawniają się wraz ze wzrostem liczby procesów składowych. Załóżmy, że zmienna losowa i rozmyta jest określona na tym samym przedziale $[a, b]$. Przeanalizujemy zmienną losową X o rozkładzie normalnym. Na podstawie przyjętych założeń można przyjąć, że

$$m = \frac{a+b}{2}, \sigma = \frac{b-a}{6} \quad (19)$$

gdzie m jest wartością średnią i σ jest odchyleniem standardowym rozkładu normalnego.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (20)$$

O zmiennej rozmytej F zakładamy, że jej nośnik pokrywa się z przedziałem $[a, b]$.

$$F_0 = [F_0^-, F_0^+] = [a, b] \quad (21)$$

Suma N zmiennych losowych X posiada również rozkład normalny o wartości średniej

$$m_N = N \cdot m = N \frac{a+b}{2} \quad (22)$$

i odchyleniu standardowym

$$\sigma_N = \sqrt{\sum_{i=1}^N \sigma^2} = \sqrt{N\sigma^2} = \sqrt{N}\sigma = \sqrt{N} \frac{b-a}{6} \quad (23)$$

Co odpowiada przedziałowi $m_N \pm 3\sigma_N$ o szerokości $L_{N,l}$

$$\frac{L_{N,l}}{6} = \sigma_N = \sqrt{N} \frac{b-a}{6}, \quad L_{N,l} = 6\sigma_N = \sqrt{N}(b-a) \quad (24)$$

Dodając N liczb rozmytych ich nośnik (zerowy α -przekrój) będzie miał szerokość $L_{N,r}$.

$$N \cdot F_0 = [N \cdot F_0^-, N \cdot F_0^+] = [N \cdot a, N \cdot b], \quad L_{N,r} = Nb - Na = N(b-a) \quad (25)$$

Porównując szerokość przedziału odpowiadającego sumie liczb rozmytych $L_{N,r}$ i losowych $L_{N,l}$ otrzymamy:

$$\frac{L_{N,r}}{L_{N,l}} = \frac{N(b-a)}{\sqrt{N}(b-a)} = \sqrt{N} \quad (26)$$

Iloraz szerokości przedziału otrzymanego na podstawie liczb rozmytych i losowych przedstawiona jest w tabeli 1. Jak widać, różnica pomiędzy największą i najmniejszą wartością w przypadku liczb rozmytych w sposób potęgowy ($N^{\frac{1}{2}}$).

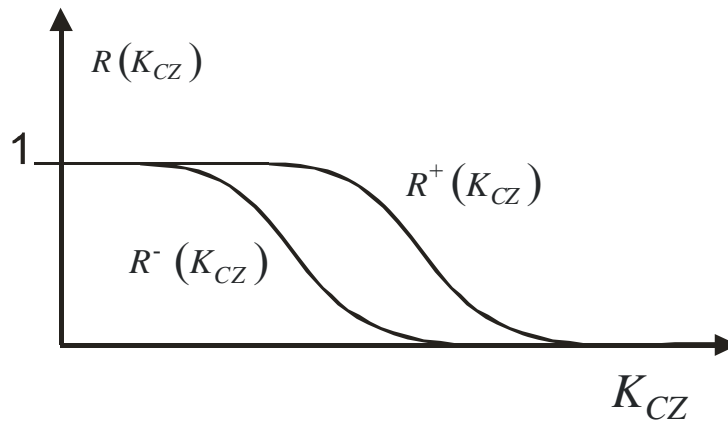
Tabela 1

Lp.	Liczba dodawanych przedziałów	$\frac{L_{n,p}}{L_{n,l}}$
1	1	1
2	2	1,414214
3	3	1,732051
4	4	2
5	10	3,162278
6	100	10
7	1000	31,62278
8	10000	100

Jeśli zmienna losowa określona jest na podstawie niepewnych danych wtedy nie jesteśmy w stanie określić dokładnej wartości prawdopodobieństwa (4). Realizacje zmiennej losowej $\hat{K}_i: \Omega \ni \omega \rightarrow \hat{K}_i(\omega) \in I(R)$ ($I(R)$ oznacza zbiór wszystkich przedziałów określonych na zbiorze liczb rzeczywistych). Można wtedy obliczyć górne i dolne prawdopodobieństwo:

$$R^-(K_{CZ}) = P \left\{ \omega: \forall_{K_C \in \hat{K}_C(\omega)} K_{CZ} < K_C \right\} \quad (27)$$

$$R^+(K_{CZ}) = P \left\{ \omega: \exists_{K_C \in \hat{K}_C(\omega)} K_{CZ} < K_C \right\} \quad (28)$$



Rys. 6

Przy pomocy górnej i dolnej dystrybuanty można zdefiniować odpowiednie funkcje gęstości prawdopodobieństwa $f_{K_C}^-(x)$ i $f_{K_C}^+(x)$, które odpowiadają górnemu i dolnemu prawdopodobieństwu.

$$R^+(K_{CZ}) = 1 - \int_{-\infty}^{K_{CZ}} f_{K_C}^+(x) dx \quad (29)$$

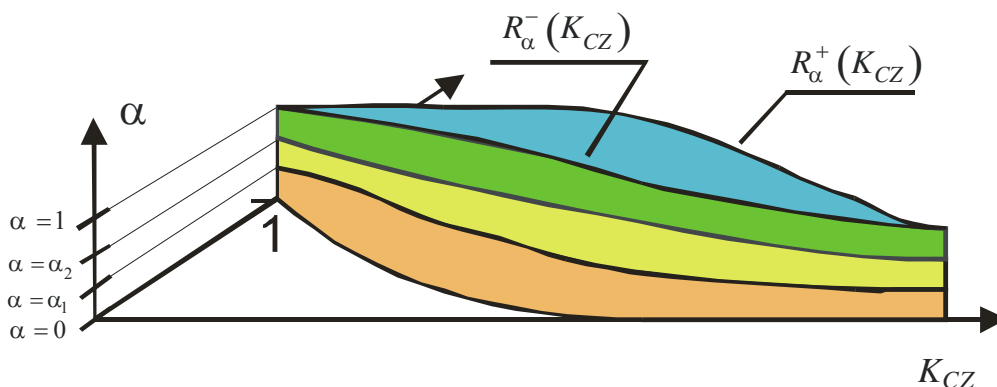
$$R^-(K_{CZ}) = 1 - \int_{-\infty}^{K_{CZ}} f_{K_C}^-(x) dx \quad (30)$$

Jeśli dane dotyczące kosztów K_i są liczbami rozmytymi wtedy funkcja $K_i: \Omega \ni \omega \rightarrow K_i(\omega) \in F(R)$ jest rozmytą zmienną losową [14]. Zbiór $F(R)$ zawiera wszystkie liczby rozmyte określone na osi rzeczywistej R . Górne i dolne ryzyko można obliczyć dla każdego α -przekroju.

$$R_{\alpha}^-(K_{CZ}) = P \left\{ \omega: \forall_{K_C \in \hat{K}_{C\alpha}(\omega)} K_{CZ} < K_C \right\} \quad (31)$$

$$R_{\alpha}^+(K_{CZ}) = P \left\{ \omega: \exists_{K_C \in \hat{K}_{C\alpha}(\omega)} K_{CZ} < K_C \right\} \quad (32)$$

Wykres górnego i dolnego ryzyka zdefiniowanego przy pomocy rozmytych zmiennych losowych można dla każdego α -przekroju przedstawić na rysunku analogicznego do rysunku 6. Zestawiając wszystkie α -przekroje otrzymujemy następującą bryłę.



Rys. 7

Dla każdego α -przekroju mamy krzywą górnego i dolnego ryzyka (porównaj rys. 6, 7) z tego wynika, że pomiędzy nimi może znajdować się teoretycznie dowolna krzywa ryzyka.

Rozpatrzmy teraz ważny z praktycznego punktu widzenia przypadek zmiennych losowych rozmytych, które można opisać przy pomocy zwykłych zmiennych losowych o parametrach rozmytych.

W analizie ryzyka zwykle mamy do czynienia z przypadkami, gdy nie musimy brać pod uwagę dowolnie możliwych dystrybuant $\Phi_\alpha(K_{CZ})$, które spełniają warunek

$$R_\alpha^-(K_{CZ}) \leq 1 - \Phi_\alpha(K_{CZ}) \leq R_\alpha^+(K_{CZ}) \quad (33)$$

ponieważ na podstawie pewnych dodatkowych informacji (obserwacji statystycznych wybranych procesów, analogii etc) wiadomo, że dystrybuanty $\Phi_\alpha(K_{CZ})$ są określonego typu. Z matematycznego punktu widzenia mamy tutaj do czynienia ze zmiennymi losowymi o określonym rozkładzie i z parametrami rozmytymi.

$$K_i : \Omega \times F(R) \ni (\omega, h) \rightarrow K_i(\omega, h) \in F(R) \quad (34)$$

Zmienne losowe można opisywać przy pomocy funkcji gęstości prawdopodobieństwa z rozmytymi parametrami $f_{K_i}(x, h)$ lub dystrybuant z rozmytymi parametrami $\Phi_{K_i}(x, h)$. Wynik obliczeń ryzyka w tym przypadku wygląda podobnie jak na rysunku 7.

4 Identyfikacja funkcji gęstości prawdopodobieństwa z rozmytymi parametrami

W analizie ryzyka kosztowego obszar rozważań możemy ograniczyć do pewnej grupy typowych rozkładów [1,7,] Często jednak nie dysponujemy informacjami na temat dokładnej wartości probabilistycznych charakterystyk danych na temat kosztu K takich jak wartość średnia, wariancja czy inne statystyki z próby. Często dostępne dane pozwalają jedynie na wyznaczenie przedziału możliwych wartości parametrów. Jednym ze stosowanych w analizie ryzyka rozkładem jest rozkład Beta-Pert. Posiada on dość intuicyjne uzasadnienie. Na jego przykładzie można zilustrować konstrukcję zmiennej losowej o parametrach rozmytych. Do jego charakterystyki wystarczy znajomość wartości: pesymistycznej (K_p), optymistycznej (K_o) i najbardziej prawdopodobnej (K_m) poniesionego kosztu K .

Rozważmy funkcje gęstości rozkładu beta

$$f_{\alpha, \beta}(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} (1-x)^{\beta-1} x^{\alpha-1}, \quad x \in [0, 1] \quad (35)$$

Rozkład beta Pert jest szczególnym przypadkiem rozkładu beta (35) przy założeniach

$$\alpha = \frac{4 \cdot (K_m - K_p)}{K_o - K_p}, \quad \beta = 4 - \alpha, \quad (36)$$

$$E(X) = \frac{K_o + 4K_m + K_p}{6}, \quad K_o \leq K_m \leq K_p. \quad (37)$$

gdzie $E(X)$ jest wartością oczekiwaną zmiennej losowej o rozkładzie beta Pert (35). Ponieważ parametry K_o, K_p i K_m często są szacowane na podstawie szacunkowych danych lub ocen eksperckich można je zdefiniować za pomocą liczb rozmytych, zatem parametry α i β jako funkcje ich funkcje są również liczbami rozmytymi.

$$\alpha_F = \alpha(K_{p,F}, K_{m,F}, K_{o,F}), \quad \beta_F = \beta(K_{p,F}, K_{m,F}, K_{o,F}) \quad (38)$$

Zatem przy pomocy wzorów (38) można zdefiniować funkcję gęstości rozkładu beta pert z rozmytymi parametrami.

$$f_{\alpha_F, \beta_F}(x) = \frac{\Gamma(\alpha_F + \beta_F)}{\Gamma(\alpha_F)\Gamma(\beta_F)} (1-x)^{\beta_F-1} x^{\alpha_F-1}, \quad x \in [0, 1] \quad (39)$$

W przypadku zastosowania innych typowych rozkładów można zastosować podobną procedurę.

5 Wnioski

Oszacowanie ryzyka kosztowego jest szczególnie istotne w gospodarce rynkowej. Podstawowym problemem przy szacowaniu jego poziomu w przypadku procesów budowlanych jest brak wiarygodnych statystycznie danych, oraz identyfikacja wszystkich warunków mogących wpływać na koszty realizacji.

Sztucznie wprowadzanie dokładności do problemów zawierających nieprecyzyjne lub nie jednoznaczne dane i relacje, zmniejszają wiarygodność wyników[8].

Często stosowany probabilistyczny opis traci wiarygodność wobec niepewności danych. Podejście oparte czysto o algebrę rozmytą wprowadza jednak nadmierną niedokładność do oszacowań nie wykorzystując w pełni posiadanej informacji. Autorska koncepcja zmiennej losowej o parametrach rozmytych pozwala połączyć zalety probabilistycznego opisu kosztów procesów - ograniczenie rozważań do konkretnego rozkładu – z zaletami stosowania algebry rozmytej. Uzyskane oszacowanie ryzyka ujmuje w maksymalnym stopniu dostępną informację jednocześnie ilustrując poziom jej niepewnego charakteru będącego stopniem zaufania do posiadanych danych.

Literatura

- [1] AbouRikz, S.M., Halpin, D. and Wilson, J.: Fitting beta distribution based simple data, Journal of Construction Engineering and Management, 120, 1993.
- [2] Barry R. Cobb and Prakash P. Shenoy, On the Plausibility Transformation for Transforming Belief Function Models to Probability Models. <http://lark.cc.ku.edu/~pshenoy>
- [3] Bętkowski.M ; Wyznaczanie charakterystyki ryzyka kosztowego procesu budowlanego. Zeszyty Naukowe Politechniki Śląskiej, Budownictwo Z.95 Gliwice 2002
- [4] Dubois D., Foulloy L., Mauris G., Probability-Possibility Transformations, Triangular Fuzzy Sets and Probabilistic Inequalities. Reliable Computing 10, 2004.
- [5] Idzikiewicz A.Z , PERT – Metody analizy sieciowej PWN Warszawa 1967
- [6] Jaworski K.M ; Metodyka projektowania realizacji budowy- PWN Warszawa 1999.
- [7] Jorion P. Financial, Risk Manager Handbook John Wiley & Sons, New Jersey 2003
- [8] Kacprzyk J., Zbiory rozmyte w analizie systemowej. PWN, Warszawa 1986
- [9] Kapliński.O; Problematyka ryzyka w zarządzaniu w budownictwie na konferencji Malta'2001. III Ogólnopolska konferencja naukowa „Ryzyko-Zarządzanie ryzykiem w przedsiębiorstwie” Bydgoszcz 2001
- [10] Mok C.K, Rao Tummala V.M. and Leung H.M., Practices, barriers and benefits of risk management process in building services cost estimation, Construction Management and Economics, 15 1997..
- [11] Neumaier A., Clouds, fuzzy sets and probability intervals, Reliable Computing 10: 2004.
- [12] Neumaier A.: Interval methods for system of equations. Cambridge University, Press, 1990.
- [13] Środowiskowe Metody Kosztorysowania – Stowarzyszenie Kosztorysantów Budowlanych Zrzeszenie Biur Kosztorysowania Budowlanego - Warszawa – Grudzień 2001
- [14] Puri, M.L., and Ralescu, D.A., 1986, Fuzzy Random Variables, J. Math. Analysis and Applications, 114:
- [15] Zadeh, L., Fuzzy sets. Information and Control, 8(3), 1965