Górnictwo i Geoinżynieria • Rok 29 • Zeszyt 3 • 2005

*Ryszard Snopkowski**

FUNKCJE ZMIENNYCH LOSOWYCH — MOŻLIWOŚCI REDUKCJI MODELI STOCHASTYCZNYCH. CZĘŚĆ II

1. Wprowadzenie

Niniejsza praca stanowi część drugą publikacji [8]. W części pierwszej omówiono funkcje zmiennych losowych o następujących rozkładach: beta, chi-kwadrat, Cauchy'ego, *F*-Snedecora, gamma, jednostajny.

Ta część publikacji zawiera omówienie funkcji zmiennych losowych o rozkładach: normalnym, lognormalnym, *t*-Studenta oraz wykładniczym.

Wprowadzenie do zagadnienia jest analogiczne jak w części pierwszej publikacji, gdyż obie części dotyczą tego samego zagadnienia — możliwości redukcji modeli stochastycznych.

Metoda symulacji stochastycznej wykorzystywana jest do komputerowego modelowania dowolnych procesów (fizycznych, ekonomicznych, technologicznych itp.) lub ich fragmentów, których cechą charakterystyczną jest występowanie w ich opisie co najmniej jednej zmiennej losowej.

Metodę po raz pierwszy zastosowano w trakcie badań w ramach projektu Manhattan, mających na celu budowę amerykańskiej bomby atomowej. Opracowany wówczas model stochastyczny dotyczył analizy propagacji neutronów w reaktorze jądrowym. Opracowali go wspólnie John von Neumann oraz polski matematyk Adam Ulman.

Metoda symulacji stochastycznej wykorzystywana jest z powodzeniem także współcześnie. Możliwości tworzenia złożonych modeli stochastycznych, ich zapis w postaci programu komputerowego w języku zorientowanym na rozwiązywanie tego typu zagad-

* Wydział Górnictwa i Geoinżynierii, Akademia Górniczo-Hutnicza, Kraków

nień, a także wciąż szybsze komputery – to wszystko stanowi o częstym wyborze symulacji stochastycznej, jako metody rozwiązywania zagadnień opisywanych modelami o charakterze niezdeterminowanym.

Biorąc pod uwagę charakter procesów górniczych, także udział wielu czynników niezdeterminowanych w ich przebiegu, uzasadnione jest stosowanie tej metody także w górnictwie. W pracy [10] opisano model stochastyczny procesu produkcyjnego realizowanego w przodku ścianowym kopalni węgla kamiennego. Występujące w modelu zmienne losowe są charakteryzowane odpowiednimi funkcjami gęstości prawdopodobieństwa.

Niejednokrotnie, występujące zależności funkcyjne między zmiennymi losowymi mogą być zastąpione jedną funkcją gęstości prawdopodobieństwa (tzw. rozkładem wynikowym), co powoduje, iż opracowany model stochastyczny analizowanego procesu ulega uproszczeniu.

W dalszej części przedstawiono funkcje zmiennych losowych, których wykorzystanie umożliwia uzyskanie rozkładów wynikowych, wśród których znalazł się rozkład normalny, lognormalny, *t*-Studenta oraz wykładniczy. Charakterystyka każdego rozkładu wynikowego zawiera ponadto wzór funkcji gęstości prawdopodobieństwa, przykładowy wykres rozkładu, a także przykłady zmiennych losowych opisywanych danym rozkładem.

2. Rozkłady wynikowe, funkcje zmiennych losowych

2.1. Rozkład wynikowy normalny

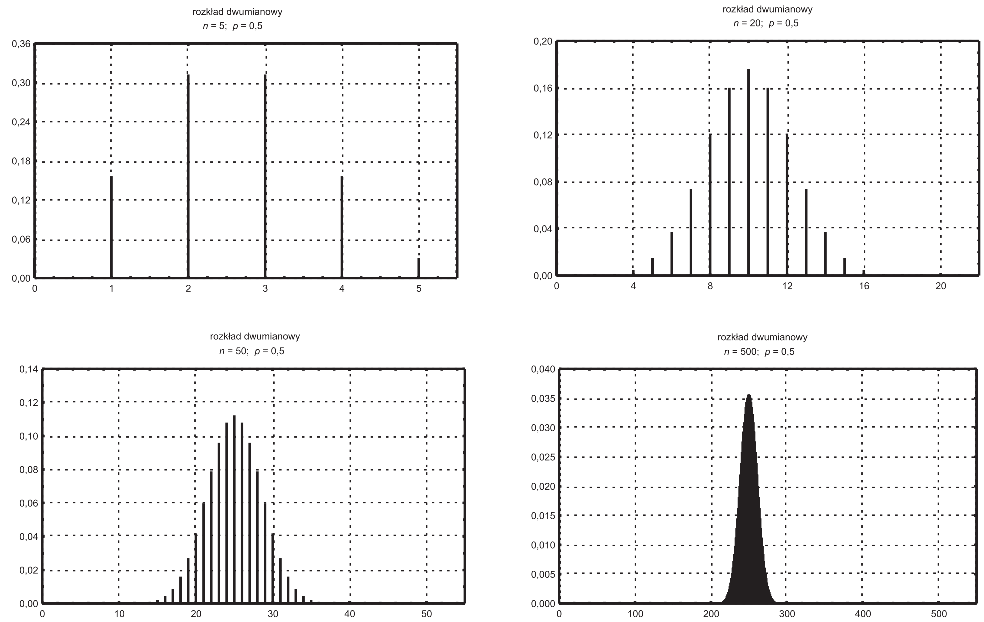
Rozkład normalny należy do rozkładów, które często wykorzystywane są w modelach stochastycznych. Ze względu na osobę jego odkrywcy, która nie jest powszechnie znana (z rozkładem kojarzeni są raczej Gauss oraz Laplace), przedstawiono poniżej krótką historię jego odkrycia.

12 listopada 1733 roku francuski matematyk Abraham de Moivre (1667–1754) opublikował broszurę, w której przedstawił zbieżność rozkładu dwumianowego do rozkładu normalnego, gdy n rośnie (zbieżność tę zilustrowano graficznie na rys. 1), a także podał wzór na funkcję gęstości prawdopodobieństwa rozkładu normalnego.

Abraham de Moivre był członkiem The Royal Society, nauczycielem matematyki oraz doradcą uczestników gier losowych w karczmach.

Autorstwo rozkładu normalnego przypisano jednak dwóm innym naukowcom. Byli to Pieerre Simon de Laplace (1749–1827) oraz Carl Friedrich Gauss (1777–1855), którzy prawie cały wiek później, w sposób niezależny od siebie, zaczęli posługiwać się tym rozkładem.

Dopiero w roku 1924 angielski statystyk Karl Pearson przypadkowo trafił na publikację de Moivre'a z 1733 roku. W tym samym roku, w czasopiśmie „Biometrika” [6] Pearson podał do wiadomości, że odkrywcą rozkładu normalnego był Abraham de Moivre. Rozkład normalny był kiedyś nazywany prawem (rozkładem) błędów (por. [6]).



Rys. 1. Zbieżność rozkładu dwumianowego do rozkładu normalnego, gdy *n* rośnie

Źródło: opracowanie własne

Funkcja gęstości prawdopodobieństwa rozkładu normalnego (rys. 2) określona jest wzorem

$$f(x) = \frac{1}{\delta\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\delta^2}} \tag{1}$$

dla: $\delta > 0$ i $-\infty < x < \infty$

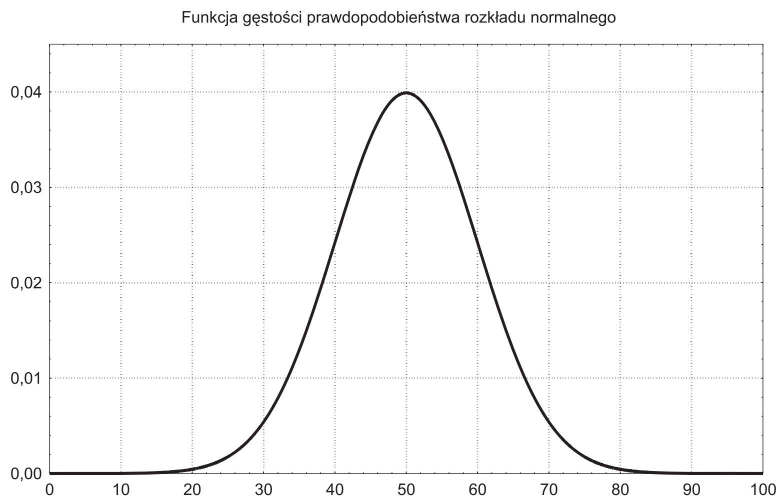
gdzie:

μ — wartość oczekiwana zmiennej losowej *X*,
δ — odchylenie standardowe zmiennej losowej *X*.

Postać rozkładu po standaryzacji wyrażają wzory:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \tag{2}$$

$$Z = \frac{X-\mu}{\delta} \tag{3}$$



Rys. 2. Przykładowy wykres rozkładu normalnego

Źródło: opracowanie własne

Przykłady zmiennych losowych (lub procesów), do których opisu wykorzystuje się rozkład normalny

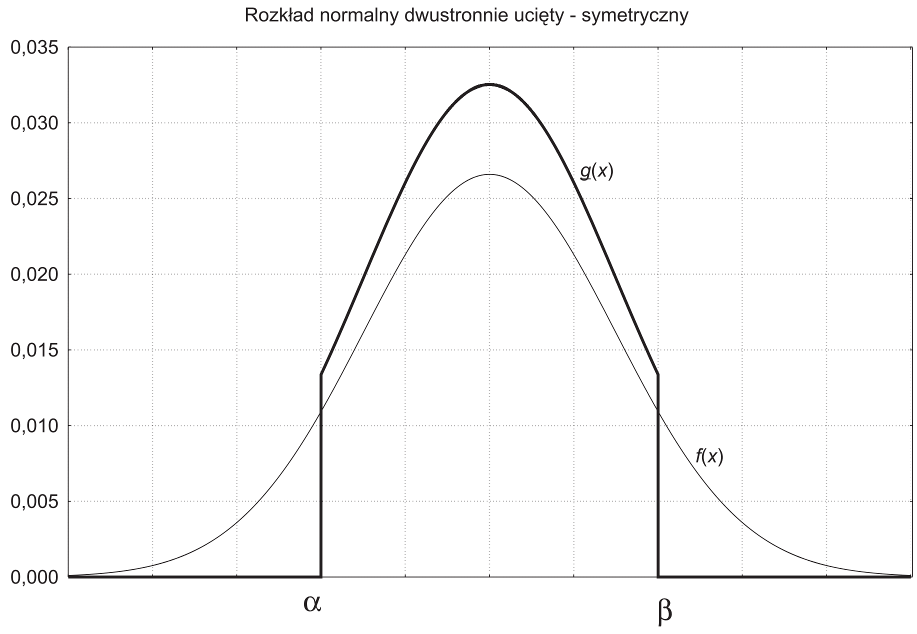
— Rozkład normalny jest modelem dla losowych błędów pomiarów. Jeżeli błąd pomiaru jest sumą wielu losowych błędów (zarówno dodatnich, jak i ujemnych), to ich suma ma rozkład bliski rozkładowi normalnemu. Stwierdzenie to wynika z tzw. centralnego twierdzenia granicznego, którego jedna z wersji brzmi: „Suma dużej ilości niezależnych zmiennych losowych ma w przybliżeniu (asymptotycznie) rozkład normalny”.

— Wiele zjawisk fizycznych może być opisanych rozkładem normalnym, mimo że dziedzina funkcji gęstości prawdopodobieństwa rozkładu to przedział od minus do plus nieskończoności. Przykład rozwiązania, polegającego na modelowaniu rozkładem normalnym zmiennej losowej opisanej na przedziale skończonym, zamieszczono w publikacji autora [9]. Rozwiązanie polega na zastosowaniu operacji obustronnego symetrycznego ucięcia rozkładu (ucięcie może być oczywiście niesymetryczne, jak również jednostronne, prawo- lub lewostronne), tak by był on opisany na przedziale skończonym, wydzielonym z obszaru liczb rzeczywistych dodatnich. Jeśli założymy, że przedział ten zawierać się będzie w granicach od α do β oraz oznaczymy funkcję gęstości rozkładu normalnego opisanego na przedziale od minus do plus nieskończoności jako $f(x)$, wówczas gęstość $g(x)$ rozkładu uciętego dwustronnie będzie można wyznaczyć z zależności

$$g(x) = \frac{f(x)}{F(\beta) - F(\alpha)} \tag{4}$$

gdzie $F(\alpha)$ oraz $F(\beta)$ są wartościami dystrybuanty $F(x)$ w punktach α oraz β.

Na rysunku 3 zamieszczono przykład funkcji $g(x)$, opisanej na przedziale $\langle \alpha, \beta \rangle$.



Rys. 3. Funkcja gęstości $g(x)$ rozkładu normalnego dwustronnie uciętego

Źródło: opracowanie własne

Funkcje zmiennych losowych, których rozkładem wynikowym jest rozkład normalny [3, 5]

— Jeśli $X_1, X_2, ..., X_n$ są niezależnymi zmiennymi losowymi o rozkładach normalnych odpowiednio $N(\mu_1, \delta_1)$, $N(\mu_2, \delta_2)$, ..., $N(\mu_n, \delta_n)$ oraz $a_1, a_2, ..., a_n$ są dowolnymi liczbami, to zmienna losowa $a_1X_1 + a_2X_2 + ... + a_nX_n$ ma rozkład normalny $N(\mu, \delta)$, gdzie $\mu = a_1\mu_1 + a_2\mu_2 + ... + a_n\mu_n$ oraz $\delta = \sqrt{a_1^2\delta_1^2 + a_2^2\delta_2^2 + ... + a_n^2\delta_n^2}$.

— Jeśli X jest zmienną losową o rozkładzie normalnym $N(\mu, \delta)$, to zmienna losowa $Y = aX + b$ dla $a \neq 0$ ma rozkład normalny $N(a\mu + b, a\delta)$.

— Jeśli $X_1, X_2, ..., X_n$ są niezależnymi zmiennymi losowymi o rozkładzie normalnym o parametrach $\mu_1, \delta_1; \mu_2, \delta_2, ..., \mu_n, \delta_n$, to zmienna losowa $X_1 + X_2 + ... + X_n$ ma rozkład normalny $N(\mu, \delta)$, gdzie $\mu = \mu_1 + \mu_2 + ... + \mu_n$ oraz $\delta = \sqrt{\delta_1^2 + \delta_2^2 + ... + \delta_n^2}$.

— Jeśli $\overline{X}$ jest średnią arytmetyczną niezależnych zmiennych losowych $X_1, X_2, ..., X_{n_1}$ o rozkładzie normalnym $N(\mu_1, \delta_1)$ oraz jeśli $\overline{Y}$ jest średnią arytmetyczną niezależnych zmiennych losowych $Y_1, Y_2, ..., Y_{n_2}$ o rozkładzie normalnym $N(\mu_2, \delta_2)$, to zmienna losowa $\overline{X} - \overline{Y}$ ma rozkład **normalny** $N\left(\mu_1 - \mu_2, \sqrt{\frac{\delta_1^{\,2}}{n_1} + \frac{\delta_2^{\,2}}{n_2}}\right)$.

— Jeśli X oraz Y są niezależnymi zmiennymi losowymi o rozkładzie jednostajnym na przedziale (0,1), to zmienne losowe $T = \sqrt{-2\ln X}\cos(2\pi Y)$ oraz $Z = \sqrt{-2\ln X}\sin(2\pi Y)$ są niezależnymi zmiennymi losowymi o rozkładzie **normalnym** $N(0,1)$.

2.2. Rozkład wynikowy logarytmiczno-normalny (lognormalny)

Tak jak rozkład normalny jest rozkładem sum wielu czynników losowych (model sum), tak rozkład logarytmiczno-normalny (lognormalny) jest określany mianem modelu iloczynów. Modelem iloczynów można wyjaśnić zjawisko rozdrabniania kruszywa lub transport osadów w rzece. W procesach tych końcowa wielkość ziarna zależy od liczby wcześniejszych zderzeń z innymi ziarnami, przy czym każde zderzenie zmniejsza proporcjonalnie rozmiar ziarna.

Oznaczając końcowy rozmiar ziarna jako X, rozmiar początkowy ziarna jako X_0, można zapisać

$$X = X_0 \cdot W_1 \cdot W_2 \cdot ... \cdot W_n \tag{5}$$

gdzie $W_1, W_2, ..., W_n$ są losowymi czynnikami powodującymi zmniejszanie się wymiaru ziarna w trakcie kolejnych zderzeń (są to zmienne losowe).

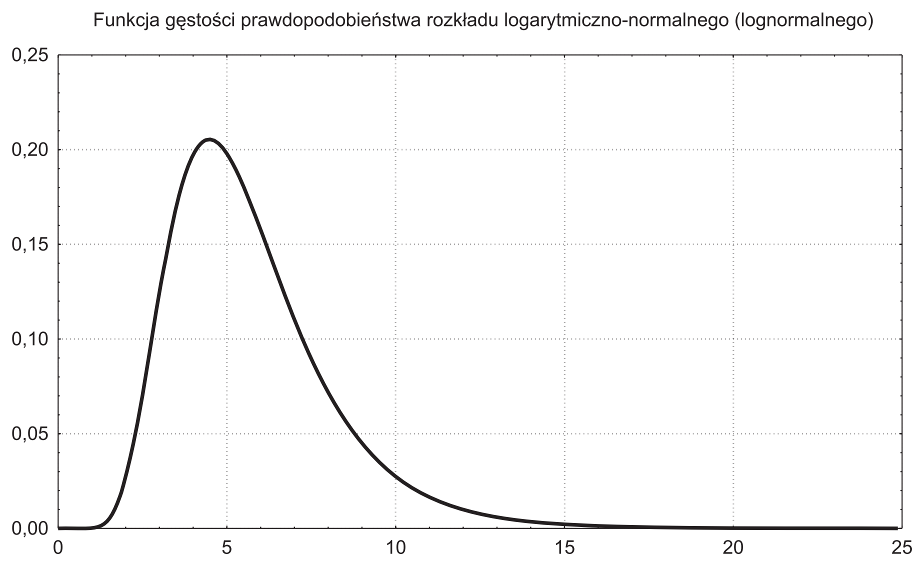
Logarytm obu stron równania jest następujący

$$\ln X = \ln X_0 + \ln W_1 + \ln W_2 + ... + \ln W_n \tag{6}$$

Czynniki $W_1, W_2, ..., W_n$ są zmiennymi losowymi, a ponieważ ich logarytmy są także zmiennymi losowymi, stąd na podstawie centralnego twierdzenia granicznego można wnioskować, że suma tych zmiennych będzie miała w przybliżeniu rozkład normalny. Oznaczając $Y = \ln X$ oraz wiedząc, że Y ma rozkład normalny $N(\mu, \delta)$, można wyznaczyć rozkład zmiennej X jako $X = e^Y$.

Zmienna losowa X, której logarytm naturalny podlega rozkładowi normalnemu, ma rozkład logarytmiczno-normalny (lognormalny), o funkcji gęstości (dla $x > 0$) (rys. 4)

$$f(x) = \frac{1}{x\delta\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\delta^2}} \tag{7}$$



Rys. 4. Przykładowy wykres rozkładu logarytmiczno-normalnego (lognormalnego)

Źródło: opracowanie własne

Przykłady zmiennych losowych (lub procesów), do których opisu wykorzystuje się rozkład logarytmiczno-normalny (lognormalny)

Rozkład logarytmiczno-normalny stosuje się:

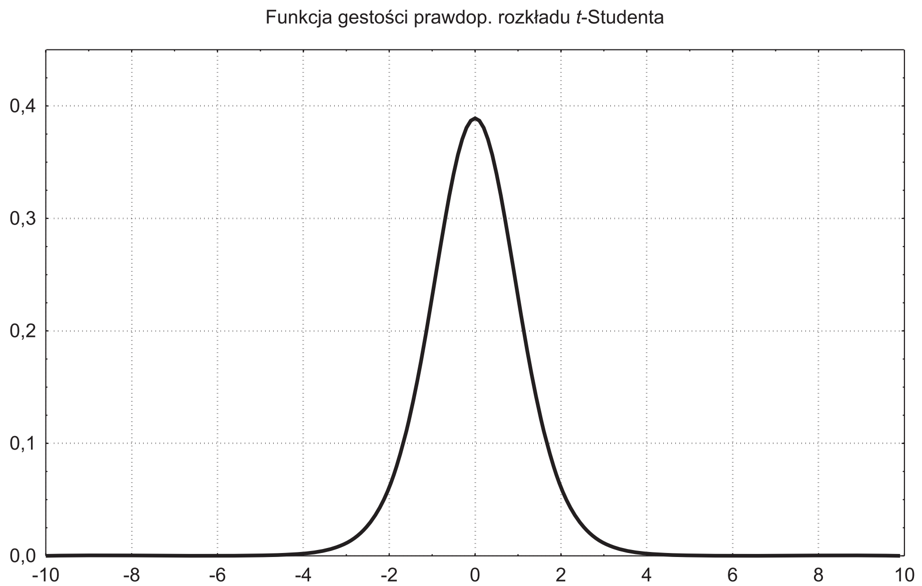
— w procesie rozdrabniania kruszywa, a także w procesie transportu osadów w rzece;
— w badaniach ekonomicznych, w których występują zmienne o wartościach dodatnich rozłożone asymetrycznie w taki sposób, że wartości mniejsze od dominanty są bardziej skupione, natomiast wartości większe do dominanty są bardziej rozproszone;
— w tych przypadkach, w których stosuje się rozkład Pareto;
— w modelowaniu procesów zmęczeniowych.

Funkcje zmiennych losowych, których rozkładem wynikowym jest rozkład logarytmiczno-normalny (lognormalny) [2]

Jeśli zmienna losowa Y ma rozkład normalny $N(\mu, \delta)$, to zmienna losowa $X = e^Y$ ma rozkład logarytmiczno-normalny (lognormalny).

2.3. Rozkład wynikowy *t*-Studenta

Autorem rozkładu *t*-Studenta (rys. 5) jest statystyk angielski W. Gosset publikujący pod pseudonimem Student.



Rys. 5. Przykładowy wykres rozkładu *t*-Studenta

Źródło: opracowanie własne

Dla małych prób ($n \leq 30$) Gosset stwierdził, że dla ciągu $X_1, X_2, ..., X_n$ niezależnych zmiennych losowych, z których każda posiada rozkład $N(\mu, \delta)$, ich średnia $\overline{X}$ posiada również rozkład normalny, zmienna losowa T postaci

$$T = \frac{\sqrt{n-1}}{S}\left(\overline{X} - \mu\right) \tag{8}$$

gdzie S oznacza odchylenie standardowe ciągu $\{X_n\}$, ma funkcje gęstości rozkładu *t*-Studenta

$$f(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}}\left(1 + \frac{t^2}{k}\right)^{-(k+1)/2} \tag{9}$$

gdzie k oznacza liczbę stopni swobody $(k = n-1)$.

Przykłady zmiennych losowych (lub procesów), do których opisu wykorzystuje się rozkład *t*- Studenta

Rozkład *t*-Studenta wykorzystuje się:

— jako model opisu wytrzymałości konstrukcji montażowych;

— do opisu zmiennych losowych, które z większym prawdopodobieństwem niż zmienne normalne przybierają wartości nietypowe, odbiegające od średniej (rozkład *t*-Studenta ma grubsze „ogony”).

Funkcje zmiennych losowych,
których rozkładem wynikowym jest rozkład *t*-Studenta [11]

Jeśli X oraz Y są niezależnymi zmiennymi losowymi, X o rozkładzie normalnym $N(0,1)$, Y o rozkładzie chi-kwadrat o n stopniach swobody, to zmienna losowa

$$Z = \frac{X}{\sqrt{\frac{Y}{n}}} \tag{10}$$

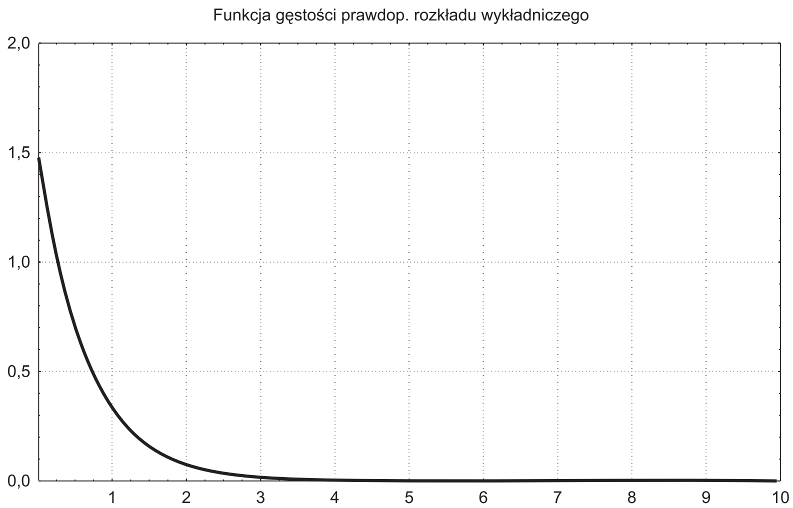
ma rozkład ***t*-Studenta** o n stopniach swobody.

2.4. Rozkład wynikowy wykładniczy

Rozkład wykładniczy jest ciągłym odpowiednikiem rozkładu geometrycznego. Funkcja gęstości prawdopodobieństwa rozkładu wykładniczego (rys. 6) jest określona wzorem

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{dla } x > 0 \\ 0 & \text{dla } x \le 0 \end{cases} \tag{11}$$

$\lambda > 0$ (λ — intensywność uszkodzeń, średni czas między zdarzeniami).



Rys. 6. Przykładowy wykres rozkładu wykładniczego

Źródło: opracowanie własne

Przykłady zmiennych losowych (lub procesów), do których opisu wykorzystuje się rozkład wykładniczy

Rozkład wykładniczy wykorzystuje się:

— do opisu upływającego czasu między pojazdami mijającymi określony punkt na drodze;
— do opisu czasu trwania rozmowy telefonicznej;
— do opisu czasu wykonywania określonej operacji na obrabiarce;
— do opisu czasu między kolejnymi powodziami;
— w teorii masowej obsługi, jako rozkład interwałów czasu między zgłoszeniami [4];
— w teorii niezawodności (trwałość elementów elektronicznych, mechanicznych); stosowanie rozkładu ma uzasadnienie wtedy, gdy pojawiające się uszkodzenia mają charakter awarii występujących na skutek zadziałania przyczyn zewnętrznych pojawiających się przypadkowo i ze stałym natężeniem.

Funkcje zmiennych losowych, których rozkładem wynikowym jest rozkład wykładniczy [3]

Jeśli X jest zmienną losową o rozkładzie jednostajnym na przedziale (0,1), to zmienna losowa $Y = -\frac{1}{2}\ln X$ ma rozkład **wykładniczy** z parametrem $\lambda = 2$.

3. Wnioski końcowe

W celu badania i analizy zjawisk i procesów rzeczywistych (ekonomicznych, technicznych, także z zakresu górnictwa — np. [10]), tworzone są modele stochastyczne. Występujące w tych modelach zmienne losowe, opisywane są odpowiednimi rozkładami prawdopodobieństwa. Niejednokrotnie w modelach tych występują także zależności funkcyjne między zmiennymi losowymi.

Zamieszczone i omówione w obu częściach publikacji funkcje zmiennych losowych wraz z opisem ich rozkładów (przyjęto tu termin „rozkłady wynikowe"), mogą zostać wykorzystane w celu uproszczenia tworzonego modelu stochastycznego lub jego fragmentu. W ten sposób przeprowadzona redukcja modelu ma korzystny wpływ na dalsze etapy jego wykorzystania, a więc upraszcza zapis w postaci programu komputerowego i skraca sam proces symulacji stochastycznej.

LITERATURA

[1] *Aczel Amir D.*: Statystyka w zarządzaniu. Wydawnictwo Naukowe PWN, Warszawa 2000 (dane oryginału: Complete Business Statistics" Richard D. Irwin Inc., Boston Sydney, 1993)
[2] *Brandt S.*: Analiza danych. Wydawnictwo Naukowe PWN, Warszawa 1998, (dane oryginału: Statistical and Computational Methods in Data Analysis" Springer Verlag New York 1997)

[3] *Klonecki W.*: Statystyka dla inżynierów. Wydawnictwo Naukowe PWN, Warszawa Wrocław 1999
[4] *Morse P.M.*: Queues, Inventories and Maintenance. New York, Wiley, 1958
[5] *Pacut A.*: Prawdopodobieństwo, teoria, modelowanie probabilistyczna w technice. Wydawnictwo Naukowo--Techniczne, Warszawa 1985
[6] *Pearson K.*: Historical Note on the Origin of the Normal Curve of Errors. Biometrika, 1924, nr XVI, 402–404
[7] *Pawłowski Z.*: Statystyka matematyczna. Państwowe Wydawnictwo Naukowe, Warszawa 1980
[8] *Snopkowski R.*: Funkcje zmiennych losowych — możliwości redukcji modeli stochastycznych (część I). Górnictwo i Geoinżynieria, z. 2, Kraków 2005
[9] *Snopkowski R.*: Wskaźniki efektywności układu kombajn — obudowa — przenośnik. Konferencja pn. „Szkoła Ekonomiki i Zarządzania w Górnictwie" AGH, Komitet Górnictwa PAN, Krynica 2004
[10] *Snopkowski R.*: Metoda identyfikacji rozkładu prawdopodobieństwa wydobycia uzyskiwanego z przodków ścianowych kopalń węgla kamiennego. Wydawnictwa AGH, Rozprawy i Monografie, nr 85, Kraków 2000
[11] *Zeigler B.*: Teoria modelowania i symulacji. Państwowe Wydawnictwo Naukowe, Warszawa 1984