

**Witold Miszczak**

## SPLYTY ROZKŁADÓW JEDNOSTAJNYCH

Niech będzie dana dwuwymiarowa zmienna losowa  $\mathbf{X} = (X_1, X_2)$  o ciągłym rozkładzie o gęstości  $f_{\mathbf{X}}(\mathbf{x})$ . Niech będą dane ciągłe i wzajemnie jednoznaczne przekształcenia  $g_1(\mathbf{x})$  i  $g_2(\mathbf{x})$  o ciągłych pochodnych cząstkowych względem  $x_1$  i  $x_2$ :

$$Y_1 = g_1(\mathbf{X}) \quad \text{i} \quad Y_2 = g_2(\mathbf{X}).$$

Załóżmy, że jacobian przekształcenia przybiera niezerowe wartości w dwuwymiarowym przedziale ( $a \leq X_1 < b$ ,  $c \leq X_2 < d$ ).

Oznaczmy przez  $h_1$  i  $h_2$  przekształcenia odwrotne do  $g_1$  i  $g_2$ . Z założeń dla  $g_1$  i  $g_2$  wynika, że przekształcenia odwrotne są wzajemnie jednoznaczne i mają ciągłe pochodne cząstkowe względem  $y_1$  i  $y_2$ . Jaki jest rozkład dwuwymiarowej zmiennej losowej  $\mathbf{Y} = (Y_1, Y_2)$ ?

Gęstość rozkładu  $\mathbf{Y}$  oznaczmy przez  $f_{\mathbf{Y}}(\mathbf{y})$ , a dystrybuantę przez  $F_{\mathbf{Y}}(\mathbf{y})$ . Na podstawie znanych wzorów mamy:

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{dF_{\mathbf{Y}}(\mathbf{y})}{d\mathbf{y}} = \frac{dF_{\mathbf{X}}(h_1(\mathbf{y}), h_2(\mathbf{y}))}{d\mathbf{x}} \frac{d\mathbf{x}}{d\mathbf{y}} = f_{\mathbf{X}}(h_1(\mathbf{y}), h_2(\mathbf{y})) |J|,$$

gdzie  $J$  jest jacobianem przekształcenia odwrotnego:

$$J = \begin{vmatrix} \frac{dx_1}{dy_1} & \frac{dx_1}{dy_2} \\ \frac{dx_2}{dy_1} & \frac{dx_2}{dy_2} \end{vmatrix} = \begin{vmatrix} \frac{\partial h_1(\mathbf{y})}{\partial y_1} & \frac{\partial h_1(\mathbf{y})}{\partial y_2} \\ \frac{\partial h_2(\mathbf{y})}{\partial y_1} & \frac{\partial h_2(\mathbf{y})}{\partial y_2} \end{vmatrix}.$$

Niech  $\mathbf{Y} = (X_1, X_1 + X_2)$ . Wtedy  $\mathbf{X} = (X_1, Y_2 - X_1)$ . Stąd

$$J = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1.$$

Jakobian powyższego przekształcenia przybiera niezerowe wartości dla  $X_1 \in \mathbb{R}$ , gdzie  $\mathbb{R}$  jest zbiorem liczb rzeczywistych. Gęstość rozkładu  $Y$  jest równa

$$f_Y(y) = f_X(x_1, y_2 - x_1).$$

Dla uproszczenia zapisu opuścimy indeksy przy zmiennych  $X_1$  i  $Y_2$ . W celu wyznaczenia gęstości rozkładu zmiennej  $Y$  należy obliczyć gęstość rozkładu brzegowego  $f_X(x, y - x)$ . Mamy zatem

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(x, y - x) dx.$$

Założmy dodatkowo, że zmienne losowe  $X_1$  i  $X_2$  są niezależne o brzegowych gęstościach  $f_1(x)$  i  $f_2(x)$ . Zatem

$$f_Y(y) = f_1(y) * f_2(y) = \int_{-\infty}^{\infty} f_1(x) f_2(y - x) dx. \quad (1)$$

Ostatnia gęstość jest nazywana gęstością splotu dwu rozkładów brzegowych:  $f_1(x)$  i  $f_2(x)$ .

Równość (1) definiuje splot dowolnych funkcji, nie tylko gęstości rozkładów prawdopodobieństwa. Gdy splatamy funkcje nie będące funkcjami gęstości, to splot też nie będzie gęstością rozkładu prawdopodobieństwa. Zauważmy, że wartość funkcji splotu w punkcie  $y$  jest równa polu powierzchni pod funkcją  $f_1(x)$  ważoną przez wagi  $f_2(y - x)$ . Wagi  $f_2(y - x)$  otrzymuje się z  $f_2(x)$  poprzez odbicie symetryczne względem osi  $OY$ , co daje  $f_2(-x)$ . W kolejnym kroku przesuwamy wykres  $f_2(-x)$  o  $y$ . Postępowanie to ilustrują rys. 1-5. Na rys. 5 pokazano, jak mógłby wyglądać algorytm numeryczny obliczania splotów dwu dowolnych funkcji:

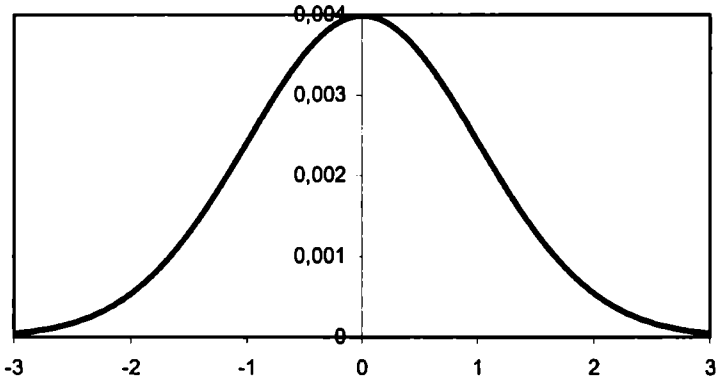
1. Dane są funkcje  $f_1(x)$  i  $f_2(x)$ , które mają być splatane.
2. Obliczamy dla ustalonego punktu  $y$  wartości funkcji  $f_2(y - x)$ .
3. Wyznaczamy

$$x_{\min}(y) = \min_{-\infty < x < \infty} (x : f_1(x) f_2(y - x) \neq 0) \text{ oraz } x_{\max}(y) = \max_{-\infty < x < \infty} (x : f_1(x) f_2(y - x) \neq 0).$$

4. Dziedzinę  $(x_{\min}, x_{\max})$  dzielimy na przedziały o długości  $dx$ . Założmy, że jest ich  $n$ .

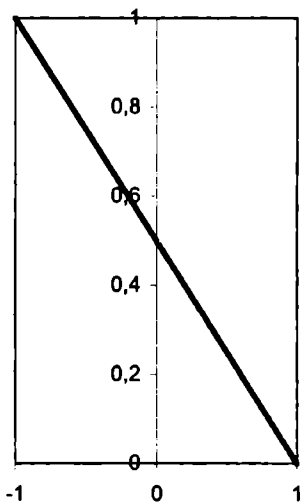
5. Obliczmy sumę pól

$$f_y(y) = dx \sum_{i=1}^n f_1(x_{\min}(y) + (i-0,5)dx) f_2(x_{\min}(y) + (i-0,5)dx).$$



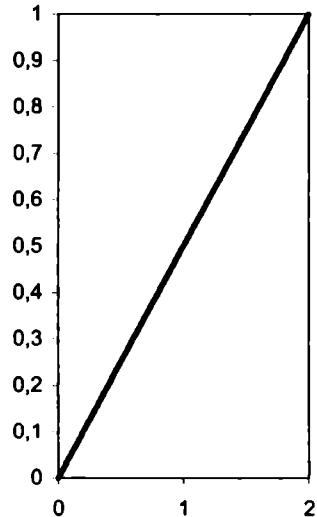
Rys. 1. Rozkład normalny  $N(0,1)$  jako  $f_1(x)$

Źródło: opracowanie własne



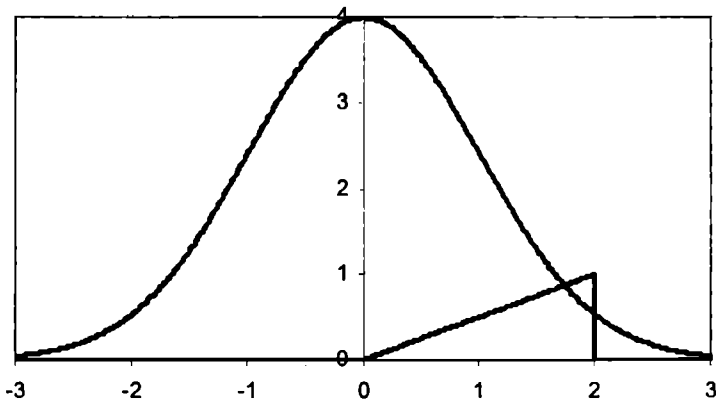
Rys. 2. Rozkład trójkątny jako  $f_2(x)$

Źródło: opracowanie własne.

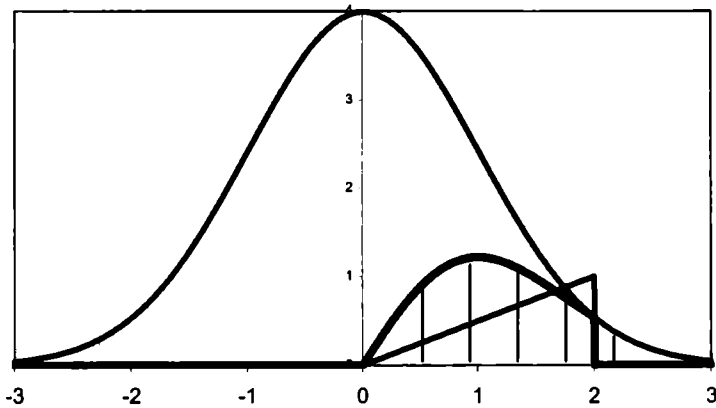


Rys. 3. Wykres  $f_2(1-x)$

Źródło: opracowanie własne.

Rys. 4. Wykres  $1000f_1(x)$  i  $f_2(1-x)$ 

Źródło: opracowanie własne.

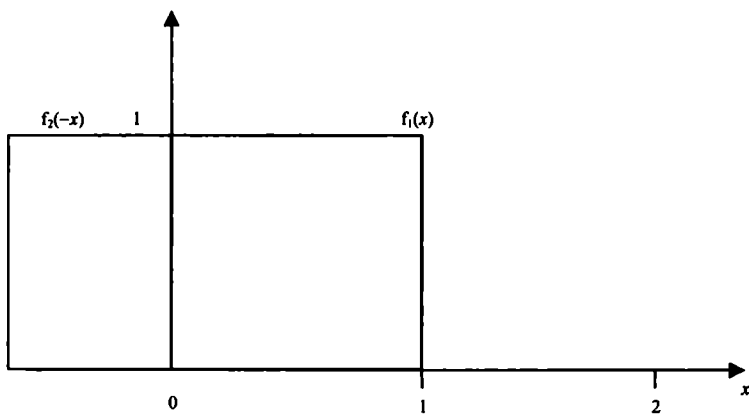
Rys. 5. Pole zakreskowanego obszaru jest wartością splotu  $1000f_V(1)$ 

Źródło: opracowanie własne.

Im drobniejszy podział  $dx$ , tym dokładniej wyznaczona jest wartość splotu  $f_V(y)$ . Funkcję splotu otrzymamy po obliczeniu jego wartości dla  $y$  należących do  $(y_{\min}, y_{\max})$ , gdzie

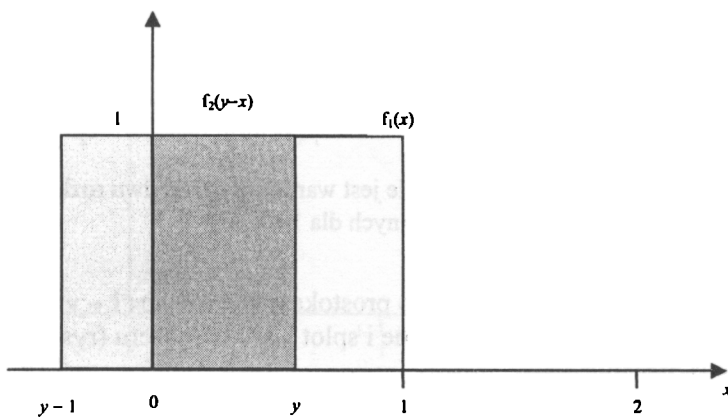
$$y_{\min} = \min_{-\infty < y < \infty} (x_{\min}(y)) \quad \text{oraz} \quad y_{\max} = \max_{-\infty < y < \infty} (x_{\max}(y)).$$

Postępowanie przedstawione na rys. 1-5 bardzo ułatwia obliczanie splotów rozkładów jednostajnych.



Rys. 6. Splot funkcji  $f_1(0)*f_2(0)$

Źródło: opracowanie własne.



Rys. 7. Ciemniejsze szare pole jest wartością splotu dwu rozkładów jednostajnych dla  $0 < y \leq 1$

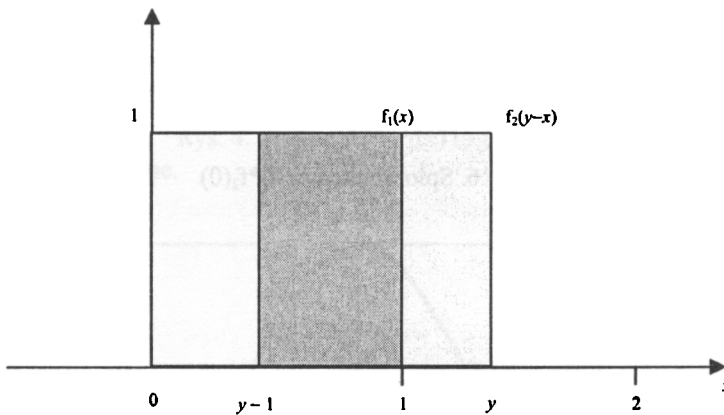
Źródło: opracowanie własne.

Niech

$$f_1(x) = f_2(x) = \begin{cases} 1 & \text{dla } 0 \leq x \leq 1, \\ 0 & \text{dla } x < 0 \text{ lub } x > 1. \end{cases}$$

Znajdźmy graficznie spłot  $f_1(y) * f_2(y)$ . Spłot dla wszystkich  $y \leq 0$  przyjmuje wartość równą zero, gdyż oba wykresy funkcji są rozłączne.

Dla  $0 < y \leq 1$  mamy rys. 7. Wartość spłotu dla podanego zakresu  $y$  jest równa polu ciemnoszarego prostokąta, tzn.  $y$ . Przesuńmy teraz  $f_2(y-x)$  do miejsca, dla którego  $1 < y \leq 2$ .



Rys. 8. Ciemniejsze szare pole jest wartością spłotu dwu rozkładów jednostajnych dla  $1 < y \leq 2$

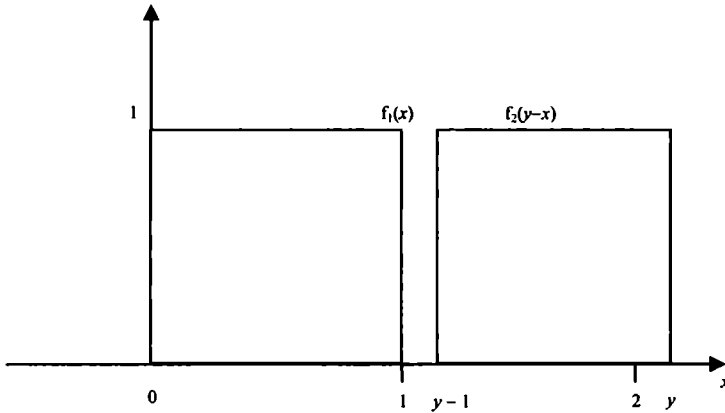
Źródło: opracowanie własne.

W tym przypadku pole wyróżnionego prostokąta jest równe  $(1 - y + 1) = 2 - y$ .

Dla  $y > 2$  oba wykresy są rozłączne i spłot jest równy zero (rys. 9).

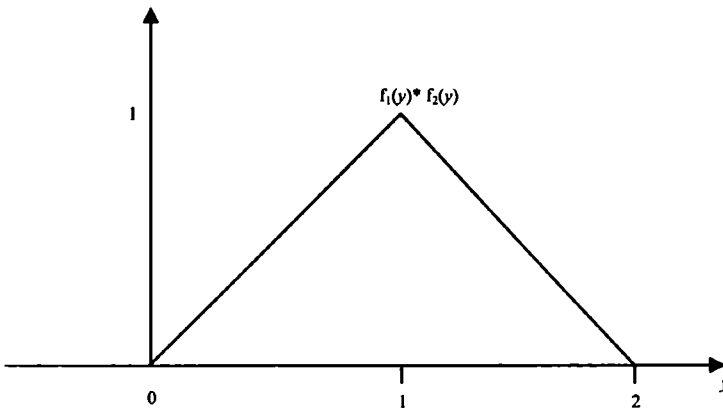
Zatem ostatecznie

$$f_1(y) * f_2(y) = \begin{cases} 0 & \text{dla } y \leq 0 \text{ lub } y > 2, \\ y & \text{dla } 0 < y \leq 1, \\ 2 - y & \text{dla } 1 < y \leq 2. \end{cases}$$

Rys. 9. Splot dla  $y > 2$ 

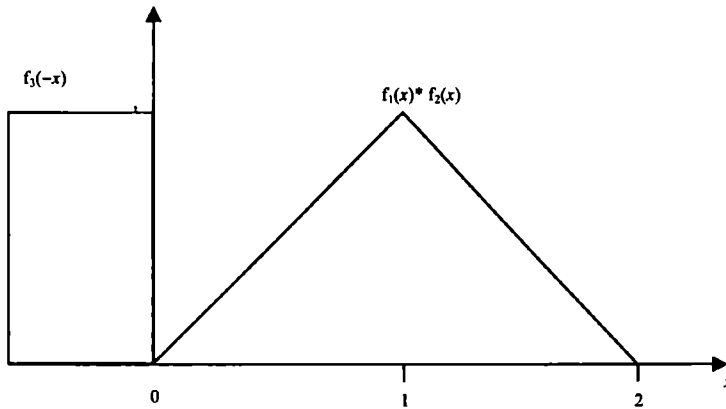
Źródło: opracowanie własne.

W wyniku otrzymujemy rozkład trójkątny na odcinku  $\langle 0, 2 \rangle$ .

Rys. 10. Splot dwu rozkładów jednostajnych  $J(0, 1)$ 

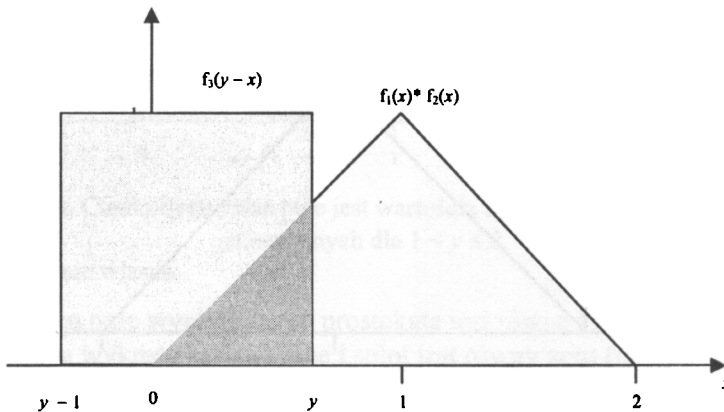
Źródło: opracowanie własne.

Splot trzech rozkładów jednostajnych  $J(0, 1)$  jest równoważny splotowi dwu rozkładów jednostajnego  $J(0, 1)$  i trójkątnego na odcinku  $\langle 0, 2 \rangle$ . Wynika to z własności łączności splotu. Ten splot też można łatwo wyznaczyć graficznie. Rysunki 11-14 przedstawiają kolejne kroki.

Rys. 11. Splot dla  $y \leq 0$ 

Źródło: opracowanie własne.

Jak widać z rysunku, wartość splotu trzech rozkładów jednostajnych  $J(0, 1)$  dla  $y \leq 0$  jest równa zero. Przyjmijmy teraz, że  $0 < y \leq 1$ . Mamy wtedy

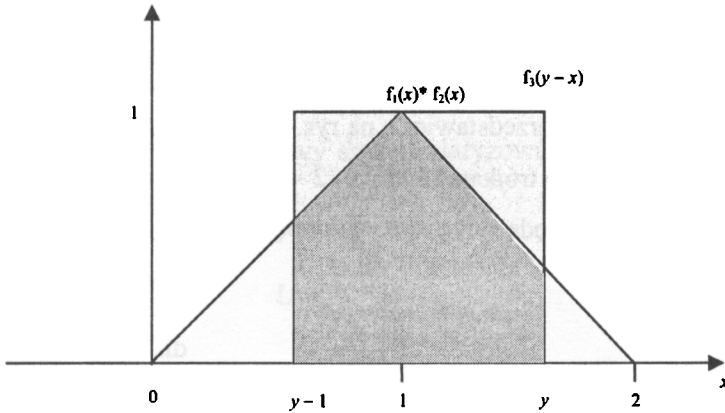
Rys. 12. Splot dla  $0 < y \leq 1$ 

Źródło: opracowanie własne.

W tym przypadku pole określające wartość splotu jest polem trójkąta. Wynosi ono  $y^2/2$ .

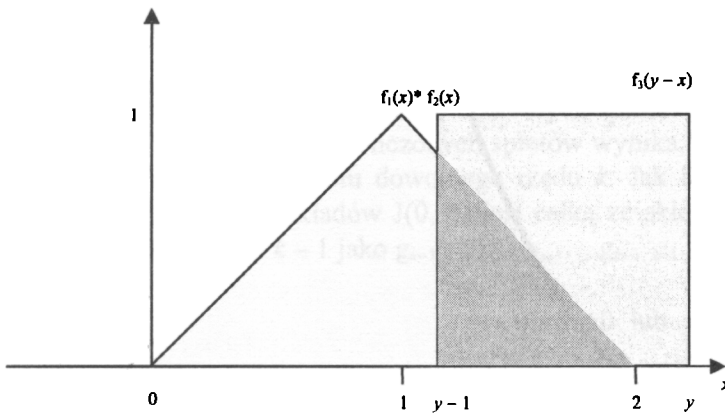
Przenieśmy  $y$  do przedziału  $\langle 1, 2 \rangle$ . Wykres będzie wyglądał następująco:



Rys. 13. Splot dla  $1 < y \leq 2$ 

Źródło: opracowanie własne.

Tym razem trzeba policzyć pole pięciokąta, który można podzielić na dwa trapezy prostokątne:

Rys. 14. Splot dla  $2 < y \leq 3$ 

Źródło: opracowanie własne.

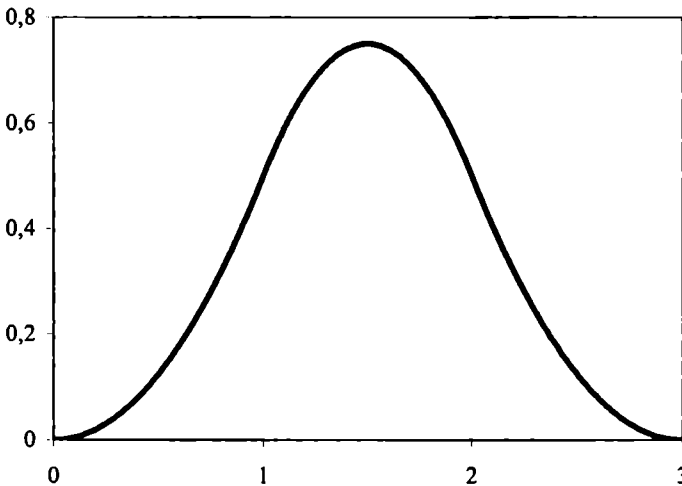
1. Wysokość pierwszego trapezu jest równa  $1 - y + 1 = 2 - y$ , zaś długość podstaw odpowiednio  $y - 1$  i  $1$ . Stąd pole tego trapezu jest równe  $y(2 - y)/2$ .

2. Wysokość drugiego trapezu jest równa  $y - 1$ , a długości podstaw  $2 - y$  i  $1$ . Pole tego trapezu jest równe  $(y - 1)(3 - y)/2$ . Wartość splotu jest więc równa  $(2y - y^2 + 3y - y^2 - 3 + y)/2 = (-2y^2 + 6y - 3)/2$ .

Dla  $2 < y \leq 3$  wykres przedstawiono na rys. 14. Teraz wartość splotu dla zadanego przedziału jest polem trójkąta równym  $(2 - y + 1)^2/2 = (3 - y)^2/2$ .

Dla  $y > 3$  obie figury będą rozłączne i splot będzie równy zeru. Zatem

$$f^{**}(y) = f_1(y) * f_2(y) * f_3(y) = \begin{cases} 0 & \text{dla } y \leq 0 \text{ i } y > 3, \\ \frac{y^2}{2} & \text{dla } 0 < y \leq 1, \\ \frac{-2y^2 + 6y - 3}{2} & \text{dla } 1 < y \leq 2, \\ \frac{(3 - y)^2}{2} & \text{dla } 2 < y \leq 3. \end{cases}$$



Rys. 15. Splot  $f^{**}(y)$  rozkładów  $J(0, 1)$

Źródło: opracowanie własne.

Wykres splotu składa się z trzech skłerek w punktach 1 i 2 wielomianów stopnia drugiego.

Z przedstawionych rozważań daje się zauważyć następujące fakty:

1. Splot  $f^{*k}(y)$  rozkładów jednostajnych jest funkcją sklejaną w punktach całkowitych  $1, 2, \dots, k-1$ .
2. Sklejkami splotu  $f^{*k}(y)$  rozkładów jednostajnych są wielomiany  $k-1$  stopnia.
3. Z twierdzenia Lindeberga-Levy'ego standaryzowany splot  $k$  rozkładów jednostajnych  $J(0, 1)$

$$U = \frac{\sum_{i=1}^k X_i - \frac{k}{2}}{\sqrt{\frac{k}{12}}},$$

gdy  $k \rightarrow \infty$  staje się rozkładem normalnym  $N(0, 1)$ . Wynika stąd, że rozkład normalny jest sklejką  $k$  wielomianów  $k-1$  stopnia, gdy  $k$  zmierza do nieskończoności.

Z punktu widzenia numerycznego gęstość rozkładu  $N(0, 1)$  lub  $N\left(\frac{k}{2}, \sqrt{\frac{k}{12}}\right)$  daje

się aproksymować za pomocą wielomianów  $k-1$  stopnia sklejanym w punktach całkowitych dla odpowiednio dużych  $k$ . Często do generowania wartości zmiennych losowych z rozkładu  $N(0, 1)$  stosuje się sumę 12 niezależnych zmiennych losowych o rozkładzie  $J(0, 1)$ . Jakość tego generatora jest zależna od jakości generatora  $J(0, 1)$ .

Analiza rys. 6-14 pozwala na konstrukcję dowolnych splotów rozkładów jednostajnych  $J(0, 1)$ . Z konstrukcji już obliczonych splotów wynika, że łatwo można napisać wzór rekurencyjny dla splotu dowolnego rzędu  $k$ . Jak łatwo zauważyć, każda sklejka splotu rzędu  $k$  rozkładów  $J(0, 1)$  jest całką ze sklejki rzędu  $k-1$ . Oznaczmy sklejki splotu rzędu  $k-1$  jako  $g_{k-1,1}, g_{k-1,2}, \dots, g_{k-1,k-1}$ . Wtedy dla splotu rzędu  $k$  mamy:

$$f^{*k}(y) = \begin{cases} 0 & \text{dla } y \leq 0 \text{ lub } y > k, \\ \int_0^y g_{k-1,1}(x) dx & \text{dla } 0 < y \leq 1, \\ \int_{y-1}^j g_{k-1,j}(x) dx + \int_j^y g_{k-1,j+1}(x) dx & \text{dla } j < y \leq j+1, \quad j = \overline{1, k-2}, \\ \int_{y-1}^{k-1} g_{k-1,k-1}(x) dx & \text{dla } k-1 < y \leq k, \end{cases}$$

gdzie  $g_{11}(x) = 1$  dla  $0 \leq x \leq 1$ .

Wykorzystajmy powyższy wzór rekurencyjny do wyznaczenia splotu  $f^{*4}(y)$ .

Dla  $f^{*3}(y)$  mamy następujące sklejki:

$$g_{3,1}(y) = \frac{y^2}{2} \quad \text{dla } 0 < y \leq 1,$$

$$g_{3,2}(y) = \frac{-2y^2 + 6y - 3}{2} \quad \text{dla } 1 < y \leq 2,$$

$$g_{3,3}(y) = \frac{(3-y)^2}{2} \quad \text{dla } 2 < y \leq 3.$$

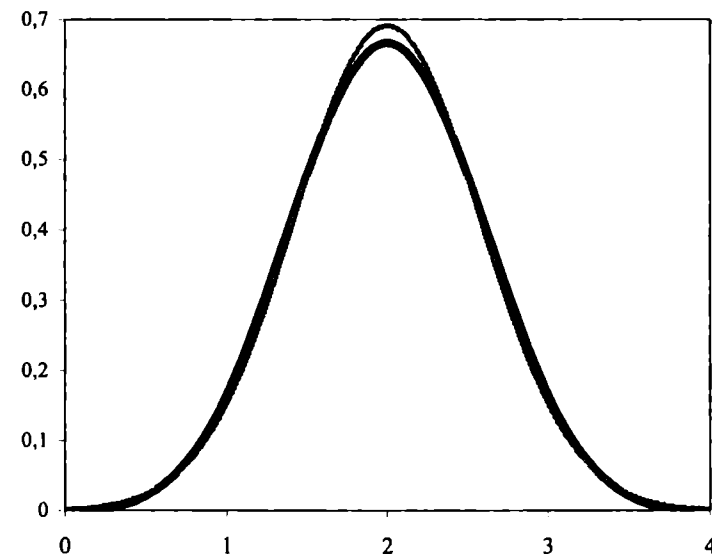
Splot  $f^{*4}(y)$  ma postać:

$$f^{*4}(y) = \begin{cases} 0 & \text{dla } y \leq 0 \text{ lub } y > 4, \\ \frac{1}{2} \int_0^y x^2 dx = \frac{y^3}{6} & \text{dla } 0 < y \leq 1, \\ \frac{1}{2} \int_{y-1}^1 x^2 dx + \frac{1}{2} \int_1^y (-2x^2 + 6x - 3) dx \\ = -\frac{y^3}{2} + 2y^2 - 2y + \frac{2}{3} & \text{dla } 1 < y \leq 2, \\ \frac{1}{2} \int_{y-1}^2 (-2x^2 + 6x - 3) dx + \frac{1}{2} \int_2^y (3-x)^2 dx \\ = \frac{y^3}{2} - 4y^2 + 10y - \frac{22}{3} & \text{dla } 2 < y \leq 3, \\ \frac{1}{2} \int_{y-1}^3 (3-x)^2 dx = -\frac{y^3}{6} + 2y^2 - 8y + \frac{32}{3} & \text{dla } 3 < y \leq 4. \end{cases}$$

Splot  $f^{*k}(y)$  jest rozkładem sumy  $k$  zmiennych losowych niezależnych o rozkładach  $J(0, 1)$ . Rozkład ten nosi nazwę rozkładu Irwina-Halla. Rozkład  $\bar{X}$  obliczony na podstawie  $k$  zmiennych losowych niezależnych o rozkładach  $J(0, 1)$  nosi nazwę rozkładu Batesa. Znajdźmy jego gęstość:

$$F_{\bar{X}}(z) = P(\bar{X} \leq z) = P\left(\sum_{i=1}^k X_i \leq kz\right) = P(Y \leq kz) = F_Y(kz),$$

$$f_{\bar{X}}(z) = \frac{dF_{\bar{X}}(z)}{dz} = \frac{dF_Y(kz)}{dz} = k f_Y(kz) = k f^{*k}(kz).$$



Rys. 16. Wykres splotu  $f^{*4}(y)$  i odpowiadającego mu rozkładu normalnego  $N(2, 0,577)$   
 Źródło: opracowanie własne.

Parametry obu rozkładów są łatwe do wyliczenia na podstawie własności wartości oczekiwanych i wariancji: dla rozkładu Irwina-Halla  $E(Y) = 0,5k$  i  $V(Y) = k/12$ . W rozkładzie Batesa  $E(\bar{X}) = 0,5$  i  $V(\bar{X}) = \frac{1}{12k}$ .

Rozkład Batesa dla średniej obliczonej z czterech obserwacji z rozkładu jednostajnego  $J(0, 1)$  ma postać

$$f_{\bar{x}}(z) = \begin{cases} 0 & \text{dla } z \leq 0 \text{ lub } z > 1, \\ \frac{128z^3}{3} & \text{dla } 0 < z \leq \frac{1}{4}, \\ -128z^3 + 128z^2 - 32z + \frac{2}{3} & \text{dla } \frac{1}{4} < z \leq \frac{1}{2}, \\ 128z^3 - 256z^2 + 160z - \frac{22}{3} & \text{dla } \frac{1}{2} < z \leq \frac{3}{4}, \\ -\frac{128z^3}{3} + 128z^2 - 128z + \frac{32}{3} & \text{dla } \frac{3}{4} < z \leq 1. \end{cases}$$

Wydaje się, że taki sposób przedstawienia splotu rozkładów ułatwia zrozumienie istoty splotów i przyczyni się do popularyzacji tej metody tworzenia nowych rozkładów.

## Literatura

- [1] Bracewell R., *Przekształcenie Fouriera i jego zastosowania*, WNT, Warszawa 1968.  
 [2] Fisz M., *Rachunek prawdopodobieństwa i statystyka matematyczna*, Biblioteka Matematyczna, tom 18, PWN, Warszawa 1967.

## CONVOLUTION OF UNIFORM DISTRIBUTIONS

### Summary

In this paper is given a recursive formula for convolution of  $k$  uniform distributions. Besides its novelty, it seems that the proposed formula is very intuitive. It could be useful for the popularization of the convolution method for creating new distributions.