PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1097 — 2005
Statystyka ekonomiczna

Witold Miszczak

ROZKŁADY ZWINIĘTE WZGLĘDEM USTALONEGO PUNKTU

Niech będzie dana zmienna losowa X o rozkładzie o gęstości $f(x)$. W wielu zastosowaniach zmienna losowa X przybiera wartości większe od pewnej stałej y, gdy tymczasem gęstość $f(x)$ jest określona również dla $x < y$. Wtedy pomimo zgodności rozkładu empirycznego w części dziedziny dla $x > y$ rozkład zmiennej X nie jest trafnie dobrany, gdyż nie spełnia warunku unormowania w odniesieniu do całego zakresu możliwych wartości X. Należy znaleźć taki rozkład, którego gęstość będzie dodatnia dla $x > y$ i dla tej dziedziny będzie unormowana, i – co więcej – będzie zgodna w dużej części z $f(x)$ dla odpowiednich y. Można to zrobić poprzez „zagięcie" tej części rozkładu $f(x)$, dla której $x < y$, oraz zsumowanie wartości części „zagiętej" rozkładu z częścią rozkładu po prawej stronie y. Innymi słowy, szukamy rozkładu następującego:

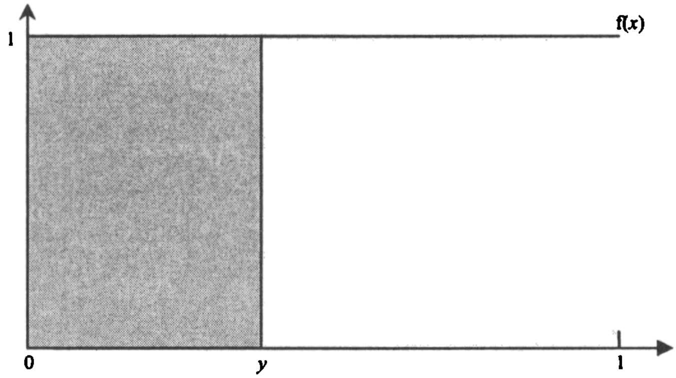
$$g(x) = \begin{cases} f(x) + f(2y - x) & \text{dla } x \geq y, \\ 0 & \text{dla } x < y. \end{cases} \tag{1}$$

Funkcja $g(x)$ jest gęstością rozkładu prawdopodobieństwa, co nietrudno sprawdzić.

$$\int_y^\infty g(x)dx = \int_y^\infty \big(f(2y-x) + f(x)\big)dx = \int_y^\infty f(2y-x)dx + \int_y^\infty f(x)dx = -\int_y^{-\infty} f(z)dz + \int_y^\infty f(x)dx$$

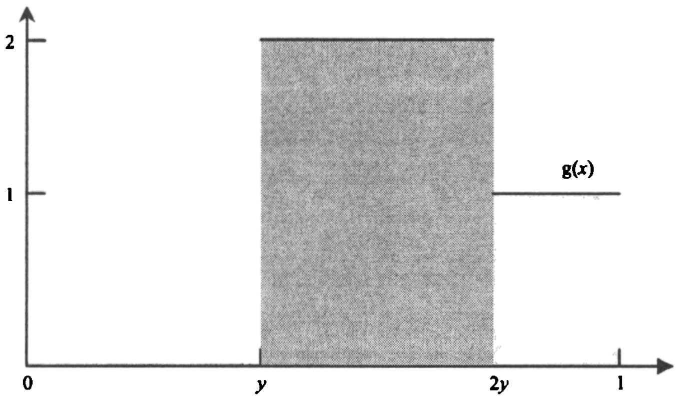
$$= \int_{-\infty}^{y} f(z)dz + \int_y^\infty f(x)dx = 1.$$

Przykładem może być zwinięty rozkład jednostajny $J(0,1)$ którego konstrukcję przedstawiają kolejne dwa rysunki.



Rys. 1. Rozkład jednostajny. Szara ciemniejsza część zostanie zagięta wzdłuż $x = y$

Źródło: opracowanie własne.



Rys. 2. Wykres zwiniętego rozkładu jednostajnego względem y

Źródło: opracowanie własne.

Zwinięty względem y ($0 \leq y < 1$) rozkład J(0,1) ma więc postać

$$g(x) = \begin{cases} 0 & \text{dla } x < y \text{ lub } x > 1, \\ 2 & \text{dla } y \leq x < 2y, \\ 1 & \text{dla } 2y \leq x \leq 1. \end{cases}$$

Obliczmy wartość oczekiwaną i drugi moment zwykły zwiniętego rozkładu w punkcie y.

$$\begin{aligned}
\mathrm{E}_g(X) &= \int_y^{\infty} x\,\mathrm{g}(x)dx = \int_y^{\infty} x\big(\mathrm{f}(2y-x)+\mathrm{f}(x)\big)dx = \int_y^{\infty} x\,\mathrm{f}(2y-x)dx + \int_y^{\infty} x\,\mathrm{f}(x)dx \\
&= -\int_y^{-\infty}(2y-z)\,\mathrm{f}(z)dz + \int_y^{\infty} x\,\mathrm{f}(x)dx = 2y\int_{-\infty}^{y}\mathrm{f}(z)dz - \int_{-\infty}^{y} z\,\mathrm{f}(z)dz + \int_y^{\infty} x\,\mathrm{f}(x)dx \qquad (2)\\
&= 2\left(y\mathrm{P}_f(X<y) - \int_{-\infty}^{y} z\,\mathrm{f}(z)dz\right) + \mathrm{E}_f(X),
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}_g\left(X^2\right) &= \int_y^{\infty} x^2\,\mathrm{g}(x)dx = \int_y^{\infty} x^2\,\mathrm{f}(2y-x)dx + \int_y^{\infty} x^2\,\mathrm{f}(x)dx = \int_{-\infty}^{y}(2y-z)^2\,\mathrm{f}(z)dz \\
&+ \int_y^{\infty} x^2\,\mathrm{f}(x)dx = 4y^2\int_{-\infty}^{y}\mathrm{f}(z)dz - 4y\int_{-\infty}^{y} z\,\mathrm{f}(z)dz + \int_{-\infty}^{y} z^2\,\mathrm{f}(z)dz + \int_y^{\infty} x^2\,\mathrm{f}(x)dx \qquad (3)\\
&= 4y\left(y\mathrm{P}_f(X<y) - \int_{-\infty}^{y} z\,\mathrm{f}(z)dz\right) + \mathrm{E}_f\left(X^2\right),
\end{aligned}$$

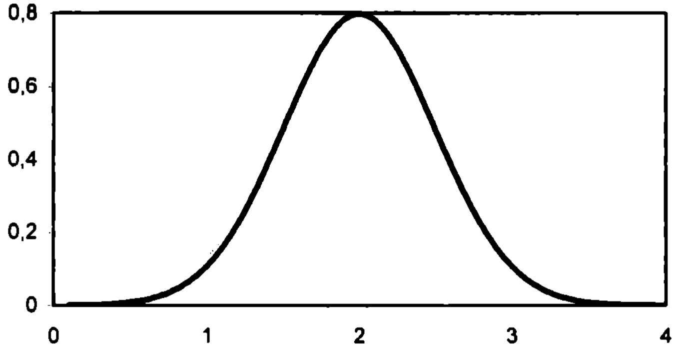
gdzie $\mathrm{P}_f(X<y)$ i $\mathrm{E}_f(\mathrm{h}(X))$ oznaczają odpowiednio prawdopodobieństwo i wartość oczekiwaną liczoną dla rozkładu o gęstości $\mathrm{f}(x)$. Dla zwiniętego rozkładu jednostajnego $\mathrm{J}(0,1)$ mamy

$$\mathrm{E}_g(X) = 2y\int_0^y dz - \int_0^y zdz + \int_y^1 xdx = 2y^2 - \frac{y^2}{2} + \frac{1}{2} - \frac{y^2}{2} = y^2 + \frac{1}{2},$$

$$\mathrm{E}_g\left(X^2\right) = 4y^2\int_0^y dz - 4y\int_0^y zdz + \mathrm{E}_f\left(X^2\right) = 4y^3 - 2y^3 + \frac{1}{3} = 2y^3 + \frac{1}{3},$$

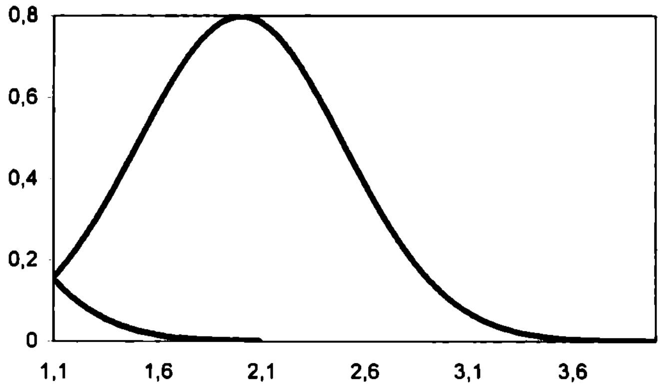
$$\mathrm{V}_g(X) = \mathrm{E}_g\left(X^2\right) - \mathrm{E}_g^2(X) = 2y^3 + \frac{1}{3} - \left(y^2 + \frac{1}{2}\right)^2 = -y^4 + 2y^3 - y^2 + \frac{1}{12}.$$

Rozważmy przypadek zmiennej losowej $X \sim \mathrm{N}(\mu, \sigma)$. Zwińmy ten rozkład względem punktu y, gdy $0 < y < \mu$. Oczywiście ten przedział dla y nie ma znaczenia i można wybrać punkt zwijania dowolnie, ale wydaje się, że ten przypadek jest ciekawszy. Przedstawiono to na kolejnych dwóch rysunkach.



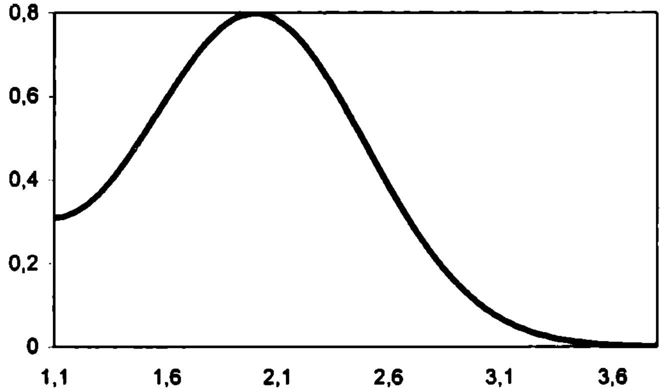
Rys. 3. Wykres rozkładu N(2, 0,5)
Źródło: opracowanie własne.

Zwiniemy ten rozkład w punkcie $y = 1,1$.



Rys. 4. Zagięty lewy ogon rozkładu N(2, 0,5)
Źródło: opracowanie własne.

Po zagięciu sumujemy obie funkcje i otrzymujemy zwinięty rozkład N(2, 0,5).



Rys. 5. Zwinięty rozkład N(2, 0,5) w punkcie 1,1
Źródło: opracowanie własne.

Funkcja gęstości zwiniętego w punkcie y rozkładu N(μ, σ) ma postać

$$g(x)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\sigma}\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}}+e^{-\frac{(2y-x-\mu)^2}{2\sigma^2}}\right) & \text{dla } x\geq y,\\ 0 & \text{dla } x<y.\end{cases} \tag{4}$$

Dla rozkładu z rys. 5 gęstość ma postać

$$g(x)=\begin{cases}\sqrt{\dfrac{2}{\pi}}\left(e^{-2(x-2)^2}+e^{-2(0,2-x)^2}\right) & \text{dla } x\geq 1,1,\\ 0 & \text{dla } x<1,1.\end{cases}$$

Szczególnym przypadkiem (4) jest rozkład dla $y = 0$

$$g(x)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\sigma}\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}}+e^{-\frac{(x+\mu)^2}{2\sigma^2}}\right) & \text{dla } x\geq 0,\\ 0 & \text{dla } x<0.\end{cases} \tag{5}$$

Parametry rozkładu (5) obliczone na podstawie (2) i (3) wynoszą:

$$\mathrm{E}_g(X) = e^{-\frac{\mu^2}{2\sigma^2}}\sqrt{\frac{2}{\pi}}\sigma + \mu\,\mathrm{Erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right),$$

$$\mathrm{V}_g(X) = \mu^2 + \left(1 - \frac{2e^{-\frac{\mu^2}{\sigma^2}}}{\pi}\right)\sigma^2 - \mu\,\mathrm{Erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right)\left(2e^{-\frac{\mu^2}{2\sigma^2}}\sqrt{\frac{2}{\pi}}\sigma + \mu\,\mathrm{Erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right)\right),$$

gdzie

$$\mathrm{Erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt.$$

Dla rozkładu (4) autor nie zna postaci analitycznej powyższych parametrów. Być może w przyszłości uda się je znaleźć. Na podstawie wzorów (2) i (3) można wyznaczyć numerycznie wartości $\mathrm{E}_g(X)$ i $\mathrm{E}_g(X^2)$ dla ustalonych parametrów μ, σ i y. Na przykład dla rozkładu z rys. 5 obliczamy najpierw wartość

$$W = y\mathrm{P}_f(X < y) - \int_{-\infty}^{y} x\,\mathrm{f}(x)dx = 1{,}1\cdot\mathrm{P}_f(X < 1{,}1) - \sqrt{\frac{2}{\pi}}\int_{-\infty}^{1,1} xe^{-2(x-2)^2}dx$$
$$= 1{,}1\cdot 0{,}035930319 - 0{,}0323856 = 0{,}007137751.$$

Wartość całki w powyższym wzorze została wyznaczona numerycznie. Parametry $\mathrm{E}_g(X)$ i $\mathrm{E}_g(X^2)$ przybierają wartości

$$\mathrm{E}_g(X) = 2W + \mathrm{E}_f(X) = 2\cdot 0{,}007137751 + 2 = 2{,}014275502$$
$$\mathrm{E}_g(X^2) = 4yW + \mathrm{E}_f(X^2) = 4{,}4\cdot 0{,}007137751 + 4{,}25$$
$$= 0{,}031406105 + 4{,}25 = 4{,}281406105.$$

Wariancja $\mathrm{V}_g(X)$ jest równa 0,2241. W zwiniętym rozkładzie N(2, 0,5) w punkcie 1,1 wartość oczekiwana jest nieco większa niż w oryginale, a wariancja nieco mniejsza.

DISTRIBUTIONS ROLLED UP TOWARDS THE FIXED POINT

Summary

In the paper is presented a distribution of any random variable X with a density function given by:

$$g(x)=\begin{cases} f(x)+f(2y-x) & \text{for } x \geq y, \\ 0 & \text{for } x < y, \end{cases}$$

where f(x) is a basic density function. This density function characterizes a distribution that we have named as the „rolled up" the point y. Examples of such distributions are given for uniform and normal distributions. For g(x) such parameters are calculated as expected value and variance.