Metody probabilistyczne
Rozwiązania zadań

10. Ciągłe zmienne losowe II

13.01.2021

Zadanie 1. *Rozważ zmienne (X, Y) o gęstości łącznej:*

$$f(x,y) = \begin{cases} c(x+y) & x, y \in [0,1] \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

Oblicz stałą c, gęstości brzegowe i warunkowe.

Odpowiedź: Korzystając z warunku normalizacji gęstości obliczamy stałą c:

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^1\int_0^1 c(x+y)\,\mathrm{d}x\,\mathrm{d}y \\ &= c\int_0^1\int_0^1 x\,\mathrm{d}x\,\mathrm{d}y + c\int_0^1\int_0^1 y\,\mathrm{d}x\,\mathrm{d}y \\ &= c\int_0^1 \underbrace{\left(\int_0^1 \mathrm{d}y\right)}_{=1} x\,\mathrm{d}x + c\int_0^1 \underbrace{\left(\int_0^1 \mathrm{d}x\right)}_{=1} y\,\mathrm{d}y \\ &= c\frac{1}{2}x^2\Big|_0^1 + c\frac{1}{2}y^2\Big|_0^1 = \frac{1}{2}c + \frac{1}{2}c = c, \end{aligned}$$

z czego wynika, że $c = 1$, a więc $f(x,y) = x + y$ dla $x, y \in [0,1]$. Wyznaczmy teraz gęstości brzegowe:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \int_0^1 (x+y)\,\mathrm{d}y = x\int_0^1 \mathrm{d}y + \int_0^1 y\,\mathrm{d}y = x + \frac{1}{2}y^2\Big|_0^1 = x + \frac{1}{2}.$$

Z symetrii gęstości ze względu na x i y otrzymujemy również $f_Y(y) = y + \frac{1}{2}$. Wyznaczamy na koniec gęstości warunkowe:

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{x+y}{x+\frac{1}{2}}.$$

Z symetrii zagadnienia mamy również $f_{X|Y}(x|y) = \frac{x+y}{y+\frac{1}{2}}$.

Zadanie 2. *Losujemy punkt jednostajnie z koła o promieniu 1. Innymi słowy mamy parę zmiennych (X, Y) o gęstości łącznej:*

$$f(x,y) = \begin{cases} c & x^2 + y^2 \leqslant 1 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

Oblicz stałą c i gęstości brzegowe.

Odpowiedź: Obliczenie stałej c nie jest trudne, jeśli zauważymy, że całka gęstości to po prostu c razy pole koła o promieniu 1, czyli πc. Ponieważ musi to być równe jeden, otrzymujemy $c = \frac{1}{\pi}$.

Ponieważ rozkład jest całkowicie symetryczny ze względu na x i y, wystarczy policzyć jeden z rozkładów brzegowych, np. po x. Z definicji:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y.$$

Dla zadanego x, funkcja gęstości jest niezerowa (i równa $\frac{1}{\pi}$) tylko wtedy, gdy $x^2 + y^2 \leqslant 1$. Nierówność ta jest spełniona dla y z przedziału $[-\sqrt{1-x^2}, \sqrt{1-x^2}]$, a więc:

$$f_X(x) \;=\; \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\,\mathrm{d}y \;=\; \frac{1}{\pi} y\Big|_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \;=\; \frac{2\sqrt{1-x^2}}{\pi}.$$

Podobnie $f_Y(y) = \frac{2\sqrt{1-y^2}}{\pi}$.

Dygresja: Zauważmy, że sprawdzenie normalizacji rozkładu brzegowe $f_X(x)$ nie jest łatwe! Trzeba wykonać całkę na przedziale $[-1, 1]$ (wartości, które może przyjąć x), tzn.:

$$\begin{aligned}
\int_{-1}^{1} \frac{2\sqrt{1-x^2}}{\pi}\,\mathrm{d}x \;&=\; \left|\begin{array}{c} x = \cos\alpha \\ \mathrm{d}x = -\sin\alpha\,\mathrm{d}\alpha \end{array}\right| \;=\; -\frac{2}{\pi}\int_{\pi}^{0} \sqrt{1-\cos^2\alpha}\sin\alpha\,\mathrm{d}\alpha \\
&=\; \frac{2}{\pi}\int_0^{\pi} |\sin\alpha|\sin\alpha\,\mathrm{d}\alpha \;=\; \frac{2}{\pi}\int_0^{\pi} \sin^2\alpha\,\mathrm{d}\alpha \\
&=\; \frac{2}{\pi}\int_0^{\pi} \frac{1-\cos(2\alpha)}{2}\,\mathrm{d}\alpha \;=\; \frac{2}{\pi}\int_0^{\pi}\frac{1}{2}\,\mathrm{d}\alpha + \frac{2}{\pi}\int_0^{\pi}\cos(2\alpha)\,\mathrm{d}\alpha \\
&=\; \left|\begin{array}{c} \beta = 2\alpha \\ \mathrm{d}\beta = 2\,\mathrm{d}\alpha \end{array}\right| \;=\; 1 + \frac{2}{\pi}\int_0^{2\pi}\cos(\beta)\frac{1}{2}\,\mathrm{d}\beta \\
&=\; 1 + \frac{1}{\pi}\sin(\beta)\Big|_0^{2\pi} \;=\; 1.
\end{aligned}$$

Zadanie 3. *Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie opisanym dystrybuantą F_X. Wyznacz dystrybuanty zmiennych $Y = \max\{X_1, \ldots, X_n\}$ oraz $X = \min\{X_1, \ldots, X_n\}$.*

Odpowiedź:

1. Dystrybuanta F_Y zmienne Y:

$$\begin{aligned}
F_Y(y) \;&=\; P(Y \leqslant y) \;=\; P(\max\{X_1, \ldots, X_n\} \leqslant y) \\
&=\; P(X_1 \leqslant y, X_2 \leqslant y, \ldots, X_n \leqslant y) \\
&\overset{(*)}{=}\; P(X_1 \leqslant y) \cdot P(X_2 \leqslant y) \cdot \ldots \cdot P(X_n \leqslant y) \;=\; F_X(y)^n,
\end{aligned}$$

   gdzie w $(*)$ wykorzystaliśmy niezależność $X_1, \ldots, X_n$.

2. Dystrybuanta F_Z zmienne Z:

$$\begin{aligned}
F_Z(z) \;&=\; P(Z \leqslant z) \;=\; 1 - P(Z > z) \;=\; 1 - P(\min\{X_1, \ldots, X_n\} > z) \\
&=\; 1 - P(X_1 > z, X_2 > z, \ldots, X_n > z) \\
&\overset{(*)}{=}\; 1 - P(X_1 > z) \cdot P(X_2 > z) \cdot \ldots \cdot P(X_n > z) \\
&=\; 1 - \big(1 - P(X_1 \leqslant z)\big) \cdot \big(1 - P(X_2 \leqslant z)\big) \cdot \ldots \cdot \big(1 - P(X_n \leqslant z)\big) \\
&=\; 1 - \big(1 - F_X(z)\big)^n,
\end{aligned}$$

   gdzie w $(*)$ wykorzystaliśmy niezależność $X_1, \ldots, X_n$.

Zadanie 4. *Zmienne losowe X i Y są niezależne, o rozkładzie wykładniczym, tzn. $X \sim \mathrm{Exp}(\lambda_X)$ oraz $Y \sim \mathrm{Exp}(\lambda_Y)$. Wyznacz rozkład zmiennej $Z = \min\{X, Y\}$. Wskazówka: trzeba wpierw uogólnić wzór na minimum zmiennych niezależnych, ale o różnym rozkładzie.*

Odpowiedź: Rozpoczniemy od wyznaczenia wzoru na dystrybuantę minimum dwóch zmiennych losowych niezależnych (ale niekoniecznie o tym samym rozkładzie). Jest to bardzo podobne wyprowadzenie do tego w poprzednim zadaniu. Niech $Z = \min\{X, Y\}$, gdzie X i Y są niezależne. Mamy:

$$\begin{aligned}
P(Z > z) \;&=\; P(\min\{X, Y\} > z) \;=\; P(X > z, Y > z) \;=\; P(X > z)P(Y > z) \\
&=\; (1 - P(X \leqslant z))(1 - P(Y \leqslant z)) \;=\; (1 - F_X(z))(1 - F_Y(z)),
\end{aligned}$$

a stąd:

$$F_Z(z) \ = \ P(Z \leqslant z) \ = \ 1 - P(Z > z) = 1 - (1 - F_X(z))(1 - F_Y(z)).$$

Jeśli $X \sim \mathrm{Exp}(\lambda)$, to dystrybuanta F_X dana jest wzorem:

$$F_X(x) = 1 - e^{-\lambda_X x}.$$

Podobnie $F_Y(y) = 1 - e^{-\lambda_Y y}$, a więc:

$$F_Z(z) \ = \ 1 - e^{-\lambda_X z} e^{-\lambda_Y z} \ = \ 1 - e^{-(\lambda_X + \lambda_Y) z}.$$

Ale to jest dystrybuanta rozkładu wykładniczego o parametrze $\lambda_X + \lambda_Y$. Tym samym $Z \sim \mathrm{Exp}(\lambda_X + \lambda_Y)$.

Zadanie 5*. *Rozważ dwie niezależne zmienne losowe $X \sim N(\mu_X, \sigma_X^2)$ oraz $Y \sim N(\mu_Y, \sigma_Y^2)$. Wyznacz gęstość zmiennej $Z = X + Y$.*

Odpowiedź: Gęstość f_Z zmiennej Z jest splotem gęstości f_X i f_Y:

$$\begin{aligned} f_Z(z) \ &= \ \int_{-\infty}^{\infty} f_X(t) f_Y(z-t) \, \mathrm{d}t && (1) \\ &= \ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{ -\frac{(t-\mu_X)^2}{2\sigma_X^2} \right\} \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left\{ -\frac{(z-t-\mu_Y)^2}{2\sigma_Y^2} \right\} \mathrm{d}t \\ &= \ \frac{1}{\sqrt{2\pi\sigma_X^2}} \frac{1}{\sqrt{2\pi\sigma_Y^2}} \int_{-\infty}^{\infty} \exp\left\{ -\frac{(t-\mu_X)^2}{2\sigma_X^2} - \frac{(z-t-\mu_Y)^2}{2\sigma_Y^2} \right\} \mathrm{d}t. && (2) \end{aligned}$$

Aby policzyć tę całkę, spróbujemy uprościć wykładnik (wyrażenie w exp) w (2):

$$\begin{aligned} -\frac{(t-\mu_X)^2}{2\sigma_X^2} - \frac{(z-t-\mu_Y)^2}{2\sigma_Y^2} \ &= \ -\frac{(t-\mu_X)^2\sigma_Y^2 + ((z-\mu_Y)-t)^2\sigma_X^2}{2\sigma_X^2\sigma_Y^2} \\ &= \ -\frac{t^2(\sigma_X^2+\sigma_Y^2) - 2t\big(\mu_X\sigma_Y^2 + (z-\mu_Y)\sigma_X^2\big) + \mu_X^2\sigma_Y^2 + (z-\mu_Y)^2\sigma_X^2}{2\sigma_X^2\sigma_Y^2}. && (3) \end{aligned}$$

Zauważmy, że w liczniku w (3) mamy wyrażenie kwadratowe (ze względu na t) postaci:

$$at^2 - 2tb + c,$$

gdzie:

$$a \ = \ \sigma_X^2 + \sigma_Y^2, \quad b \ = \ \mu_X\sigma_Y^2 + (z-\mu_Y)\sigma_X^2, \quad c \ = \ \mu_X^2\sigma_Y^2 + (z-\mu_Y)^2\sigma_X^2 \qquad (4)$$

Chcielibyśmy je zamienić na bardziej zwartą postać $a(t-d)^2 + r$. Jak znaleźć d i r? Wystarczy oba wyrażenia przyrównać:

$$at^2 - 2tb + c \ = \ a(t-d)^2 + r \ = \ at^2 - 2tad + ad^2 + r.$$

Ponieważ wyrażenia przy kolejnych potęgach t muszą być sobie równe, dostaniemy:

$$b \ = \ ad, \ c \ = \ ad^2 + r \qquad \Longrightarrow \qquad d \ = \ \frac{b}{a}, \quad r \ = \ c - ad^2 \ = \ c - \frac{b^2}{a}.$$

Czyli:

$$\begin{aligned} d \ &= \ \frac{\mu_X\sigma_Y^2 + (z-\mu_Y)\sigma_X^2}{\sigma_X^2+\sigma_Y^2} \\ r \ &= \ \mu_X^2\sigma_Y^2 + (z-\mu_Y)^2\sigma_X^2 - \frac{\big(\mu_X\sigma_Y^2 + (z-\mu_Y)\sigma_X^2\big)^2}{\sigma_X^2+\sigma_Y^2} \\ &= \ \frac{\mu_X^2\sigma_Y^4 + (z-\mu_Y)^2\sigma_X^4 + \sigma_X^2\sigma_Y^2\left(\mu_X^2 + (z-\mu_Y)^2\right) - \big(\mu_X\sigma_Y^2 + (z-\mu_Y)\sigma_X^2\big)^2}{\sigma_X^2+\sigma_Y^2} \\ &= \ \frac{\sigma_X^2\sigma_Y^2\left(\mu_X^2 + (z-\mu_Y)^2\right) - 2\mu_X\sigma_Y^2(z-\mu_Y)\sigma_X^2}{\sigma_X^2+\sigma_Y^2} \\ &= \ \frac{\sigma_X^2\sigma_Y^2}{\sigma_X^2+\sigma_Y^2}\left(\mu_X^2 + (z-\mu_Y)^2 - 2\mu_X(z-\mu_Y)\right) \ = \ \frac{\sigma_X^2\sigma_Y^2}{\sigma_X^2+\sigma_Y^2}(z-\mu_X-\mu_Y)^2. && (5) \end{aligned}$$

Ułamek po prawej stronie równania (3) ma więc postać:

$$-\frac{a(t-d)^2+r}{2\sigma_X^2\sigma_Y^2},$$

a tym samym całka w (2) wygląda następująco:

$$\begin{aligned}\int_{-\infty}^{\infty}\exp\left\{-\frac{a(t-d)^2+r}{2\sigma_X^2\sigma_Y^2}\right\}\mathrm{d}t &= \int_{-\infty}^{\infty}\exp\left\{-\frac{a(t-d)^2}{2\sigma_X^2\sigma_Y^2}\right\}\exp\left\{-\frac{r}{2\sigma_X^2\sigma_Y^2}\right\}\mathrm{d}t\\ &= \exp\left\{-\frac{r}{2\sigma_X^2\sigma_Y^2}\right\}\int_{-\infty}^{\infty}\exp\left\{-\frac{a(t-d)^2}{2\sigma_X^2\sigma_Y^2}\right\}\mathrm{d}t. \qquad (6)\end{aligned}$$

Mając dowolny rozkład normalny o parametrach μ i σ^2, gęstość rozkładu musi się normalizować do 1, tzn:

$$\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\}\mathrm{d}t \;=\; 1,$$

z czego wynika:

$$\int_{-\infty}^{\infty}\exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\}\mathrm{d}t \;=\; \sqrt{2\pi\sigma^2} \qquad (7)$$

Zauważmy teraz, że całka po prawej stronie (6) jest dokładnie postaci (7), jeśli przyrównamy $\mu = d$ oraz $\sigma^2 = \frac{\sigma_X^2\sigma_Y^2}{a}$. Tym samym dostajemy:

$$\int_{-\infty}^{\infty}\exp\left\{-\frac{a(t-d)^2}{2\sigma_X^2\sigma_Y^2}\right\}\mathrm{d}t \;=\; \sqrt{2\pi\frac{\sigma_X^2\sigma_Y^2}{a}} \;=\; \sqrt{2\pi\frac{\sigma_X^2\sigma_Y^2}{\sigma_X^2+\sigma_Y^2}},$$

gdzie w ostatniej równości podstawiliśmy wartość a z (4). Wykorzystując definicję r z (5), dostajemy:

$$\exp\left\{-\frac{r}{2\sigma_X^2\sigma_Y^2}\right\} \;=\; \exp\left\{-\frac{\cancel{\sigma_X^2\sigma_Y^2}(z-\mu_X-\mu_Y)^2}{2\cancel{\sigma_X^2\sigma_Y^2}(\sigma_X^2+\sigma_Y^2)}\right\} \;=\; \exp\left\{-\frac{(z-\mu_X-\mu_Y)^2}{2(\sigma_X^2+\sigma_Y^2)}\right\}.$$

Możemy więc zapisać prawą stronę (6) jako:

$$\exp\left\{-\frac{(z-\mu_X-\mu_Y)^2}{2(\sigma_X^2+\sigma_Y^2)}\right\}\sqrt{2\pi\frac{\sigma_X^2\sigma_Y^2}{\sigma_X^2+\sigma_Y^2}}.$$

Podstawiając to do (2) dostajemy:

$$\begin{aligned}f_Z(z) &= \frac{1}{\sqrt{2\pi\sigma_X^2}}\frac{1}{\sqrt{2\pi\sigma_Y^2}}\exp\left\{-\frac{(z-\mu_X-\mu_Y)^2}{2(\sigma_X^2+\sigma_Y^2)}\right\}\sqrt{2\pi\frac{\sigma_X^2\sigma_Y^2}{\sigma_X^2+\sigma_Y^2}}\\ &= \sqrt{\cancel{2\pi}\frac{\cancel{\sigma_X^2}\cancel{\sigma_Y^2}}{\sigma_X^2+\sigma_Y^2}\frac{1}{\cancel{2\pi}\cancel{\sigma_X^2}}\frac{1}{2\pi\cancel{\sigma_Y^2}}}\\ &= \frac{1}{\sqrt{2\pi(\sigma_X^2+\sigma_Y^2)}}\exp\left\{-\frac{(z-\mu_X-\mu_Y)^2}{2(\sigma_X^2+\sigma_Y^2)}\right\}.\end{aligned}$$

Ale to ostatnie wyrażenie ma postać gęstości rozkładu normalnego z parametrami $\mu = \mu_X + \mu_Y$ i $\sigma^2 = \sigma_X^2 + \sigma_Y^2$. Czyli:

$$Z \sim N\left(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2\right).$$