Metody Probabilistyczne i Statystyka

Rafał Włodarczyk

INA 3, 2024

Contents

1 Lista 9

1.1 1L9

Pokażmy, że dla dowolnych $a, b, c, d \in \mathbb{R}$ takich, że $ac \neq 0$ współczynnik korelacji

$$\rho(X,Y) = \frac{\mathbf{cov}(X,Y)}{\sigma_X \sigma_Y}$$

spełnia $\rho(aX+b, cY+d) = \rho(X,Y)$.
Weźmy:

$$\rho(aX+b, cY+d) = \tag{1.1.1}$$
$$= \frac{\mathbf{cov}(aX+b, cY+d)}{\sigma_{aX+b}\sigma_{cX+d}} = \tag{1.1.2}$$
$$= \frac{\mathbf{E}((aX+b)(cY+d)) - \mathbf{E}(aX+b)\cdot\mathbf{E}(cY+d)}{\sqrt{\mathbf{var}(aX+b)}\sqrt{\mathbf{var}(cY+d)}} = \tag{1.1.3}$$
$$= \frac{\mathbf{E}(acXY + adX + bcY + bd) - (a\mathbf{E}(X)+b)(c\mathbf{E}(Y)+d)}{\sqrt{a^2\mathbf{var}(X)}\sqrt{c^2\mathbf{var}(Y)}} = \tag{1.1.4}$$
$$= \frac{ac\mathbf{E}(XY) + ad\mathbf{E}(X) + bc\mathbf{E}(Y) + bd - ac\mathbf{E}(X)\mathbf{E}(Y) - ad\mathbf{E}(X) - bc\mathbf{E}(Y) - bd}{ac\cdot\sigma_X\sigma_Y} = \tag{1.1.5}$$
$$= \frac{ac\left[\mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)\right]}{ac\cdot\sigma_X\sigma_Y} = \tag{1.1.6}$$
$$= \frac{\mathbf{cov}(X,Y)}{\sigma_X\sigma_Y} = \tag{1.1.7}$$
$$= \rho(X,Y) \tag{1.1.8}$$

1.2 2L9

Rzucamy dwiema szcześciennymi kostkami, niech X - mniejszy, Y - większy z rzutów. Obliczmy $\mathbf{cov}(X,Y)$ oraz $\rho(X,Y)$.
Z zadania 1L6 wiemy, że wartości

$$\mathbf{E}(X) = \frac{91}{36} \tag{1.2.1}$$
$$\mathbf{E}(Y) = \frac{161}{36} \tag{1.2.2}$$
$$\tag{1.2.3}$$

Ponadto z zadania 3L7 mamy

$$\mathbf{var}(X) = \frac{2555}{1296} \tag{1.2.4}$$
$$\mathbf{var}(Y) = \frac{2555}{1296} \tag{1.2.5}$$
$$\tag{1.2.6}$$

Wyznaczmy teraz

$$\mathbf{E}(XY) = \sum_x \sum_y xy \cdot P(X = x \wedge Y = y) = \dots \tag{1.2.7}$$

Znając rozkład łączny z zadania 1L6 możemy zsumować wszystkie kombinacje prawdopodobieństw:

$$\dots = \frac{1}{36} \cdot (1 + 4 + 9 + 16 + 25 + 36) + \tag{1.2.8}$$

$$+ \frac{2}{36} \cdot (2 + 3 + 6 + 4 + 8 + 12 + 5 + 10 + 15 + 20 + 6 + 12 + 18 + 24 + 30) = \tag{1.2.9}$$

$$= \frac{441}{36} \tag{1.2.10}$$

Zatem

$$\mathbf{cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X) \cdot \mathbf{E}(Y) = \frac{441}{36} - \frac{91}{36} \cdot \frac{161}{36} = \frac{1225}{1296} \approx 0.94 \tag{1.2.11}$$

oraz

$$\rho(X, Y) = \frac{\mathbf{cov}(X, Y)}{\sigma_X \sigma_Y} \approx \frac{\frac{1225}{1296}}{\frac{2555}{1296}} \sim 0.48 \tag{1.2.12}$$

1.3 3L9

Niech $X_1, X_2 \sim U([0, 1])$ - niezależne zmienne losowe. Niech $Y = \min\{X_1, X_2\}$, $Z = \max\{X_1, X_2\}$. Obliczmy $\mathbf{cov}(Y, Z)$ oraz $\rho(Y, Z)$.
TBD.

1.4 4L9

Niech $X \sim U([0, \pi])$ i niech $Y = \sin(X), Z = \cos(X)$ Sprawdźmy zależność i korelację.
Z zadania 2L8 pamiętamy następujące fakty:

$$\mathbf{E}(Y) = \frac{2}{\pi} \tag{1.4.1}$$

$$\mathbf{E}(Z) = 0 \tag{1.4.2}$$

Wyznaczmy $\mathbf{E}(YZ)$, wiedząc że $f_X(x) = \frac{1}{\pi}$

$$\mathbf{E}(YZ) = \mathbf{E}(\sin(X)\cos(X)) = \int_{-\infty}^{\infty} \sin(x)\cos(x) f_X(x) dx = \tag{1.4.3}$$

$$= \int_0^\pi \frac{1}{2} \sin(2x) \frac{1}{\pi} dx = \frac{1}{2\pi} \int_0^\pi \sin(2x) dx = \tag{1.4.4}$$

$$= \frac{1}{2\pi} \left[\frac{1}{2} \cos(2x) \right]_0^\pi = \frac{1}{4\pi} [\cos(2\pi) - \cos(0)] = 0 \tag{1.4.5}$$

Policzmy zatem $\mathbf{cov}\{Y, Z\}$

$$\mathbf{cov}(Y, Z) = \mathbf{E}(YZ) - \mathbf{E}(Y)\mathbf{E}(Z) = 0 - \frac{2}{\pi} \cdot 0 = 0 \tag{1.4.6}$$

Wobec tego mamy pewność, że zmienne są nieskorelowane. Sprawdźmy natomiast zależność zmiennych.

$$P(Y \leqslant \frac{1}{2} \wedge Z \leqslant \frac{1}{2}) = \frac{1}{6} \tag{1.4.7}$$

Obie funkcje są powyżej $y = \frac{1}{2}$ jedynie w przedziale $[\frac{\pi}{6}, \frac{\pi}{3}]$ Zobaczmy natomiast, że osobno, $Y \leqslant \frac{1}{2}$ w przedziale $[\frac{\pi}{2}, \pi]$, a $Z \leqslant \frac{1}{2}$ w przedziale $[\frac{\pi}{3}, \pi]$. Wtedy:

$$P(Y \leqslant \frac{1}{2}) \cdot P(Z \leqslant \frac{1}{2}) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9} \tag{1.4.8}$$

Widzimy, że warunek niezależności nie został spełniony, więc zmienne losowe Y, Z są zależne nieskorelowane.

1.5 5L9

Wyznaczmy funkcję tworzące prawdopodobieństwo oraz funkcje tworzące momenty zmiennej losowej $X \sim Geo(p)$

$$\varphi_X(z) = \mathbf{E}(z^x) = \sum_{n \geqslant 0} z^n P(X = n) = \sum_{n \geqslant 0} z^n p(1-p)^{n-1} = \tag{1.5.1}$$

$$= \sum_{n \geqslant 1} z^n p(1-p)^{n-1} = pz \sum_{n \geqslant 1} z^{n-1}(1-p)^{n-1} = pz \sum_{n \geqslant 1} (z(1-p))^{n-1} = \tag{1.5.2}$$

$$= pz \sum_{n \geqslant 0} (z(1-p))^n = pz \frac{1}{1 - z(1-p)} = \frac{pz}{1 - z(1-p)} \tag{1.5.3}$$

Oraz MGF:

$$M_X(t) = \mathbf{E}(e^{tX}) = \varphi_X(e^t) = \frac{pe^t}{1 - e^t(1-p)} \tag{1.5.4}$$

Wyznaczmy następnie wartość oczekiwaną i wariancję

$$\mathbf{E}(X) = \varphi'_X(1) = \frac{d}{dz}(\frac{pz}{1 - z(1-p)}) = \frac{p}{(1 - z(1-p))^2} = \frac{p}{(1 - 1 + p)^2} = \frac{1}{p} \tag{1.5.5}$$

$$\tag{1.5.6}$$

Wyznaczmy pomocniczo $\varphi''_X(1)$

$$\varphi''_X(1) = \frac{d}{dz}\varphi'_X(z) = \frac{p}{(1 - z(1-p))^2} = \frac{2(p-1)}{(1 - z(1-p))^2} = \frac{2(1-p)}{p^2} \tag{1.5.7}$$

Wobec tego

$$\mathbf{var}(X) = \varphi''_X(1) + \varphi'_X(1) - (\varphi'_X(1))^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2 - 2p + p - 1}{p^2} = \frac{1-p}{p^2} \tag{1.5.8}$$

Weźmy zmienną losową $Y \sim NB(k, p)$, możemy z poprzedniej części zapisać:

$$M_X(t) = \frac{pe^t}{1 - e^t 1 - p} \tag{1.5.9}$$

$$M_{\sum_{i=0}^k X_i}(t) = M_{X_1 + X_2 + \cdots + X_k}(t) = \prod_{i=0}^k M_X(t) = (M_X(t))^k \tag{1.5.10}$$

1.6 6L9

Mamy wyznaczyć MGF dla zmiennej losowej o standardowym rozkładzie normalnym $X \sim \mathcal{N}(0,1)$. Znając PMF dla rozkładu normalnego:

$$f_X(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$M_X(t) = \int_{-\infty}^{\infty} e^{xt} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{xt} \cdot e^{-\frac{x^2}{2}} dx = \tag{1.6.1}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2 + \frac{t^2}{2}} dx = \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} dx = \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}} \sqrt{2\pi} = e^{\frac{t^2}{2}} \tag{1.6.2}$$

Jeśli teraz $Y \sim \mathcal{N}(\mu, \sigma^2)$ to z własności MGF możemy napisać:

$$M_Y(t) = M_{\sigma X + \mu}(t) = \mathbf{E}(e^{t(\sigma X + \mu)}) = \mathbf{E}(e^{t\mu})\mathbf{E}(e^{t\sigma X}) = \tag{1.6.3}$$

$$= e^{t\mu} M_X(\sigma t) = e^{t\mu} \cdot e^{\frac{(\sigma t)^2}{2}} = e^{t\mu + \frac{(\sigma t)^2}{2}} \tag{1.6.4}$$

Weźmy dwie nowe niezależne zmienne losowe $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ oraz $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = \tag{1.6.5}$$

$$= e^{t\mu_1 + \frac{(\sigma_1 t)^2}{2} + t\mu_2 + \frac{(\sigma_2 t)^2}{2}} = \tag{1.6.6}$$

$$= e^{t(\mu_1+\mu_2) + \frac{1}{2}(t^2(\sigma_1^2+\sigma_2^2))} \tag{1.6.7}$$

Widzimy wobec tego że nasza nowa zmienna $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$