1 Gaussowskie zmienne losowe

W tej serii rozwiążemy zadania dotyczące zmiennych o rozkładzie normalny. Wymagana jest wiedza na temat własności rozkładu normalnego, CTG oraz warunkowych wartości oczekiwanych.

1. (Eg 48/6) Niech $X_1, X_2, ..., X_n, ...$ będą niezależnymi zmiennymi losowymi o identycznym rozkładzie o gęstości $f(x) = \frac{1}{2\sqrt{x}} 1_{(0,1)}(x)$. Niech $U_n = (X_1 ... X_n)^{\frac{1}{n}}$. Wtedy: (asymptotyka $U_n - e^{-2}$)?
Odp: B-> $\lim_{n\to\infty} \mathbf{P}((U_n - e^{-2})\sqrt{n} < 4e^{-2}) = 0,977$.

   Rozwiązanie. Zadanie polega na umiejętnym zastosowaniu CTG. Zauważmy, że $Y_i = -\log X_i$, $i \in \{1, ..., n\}$ oraz
$$\mathbf{P}(Y_i > t) = \mathbf{P}(X_i < e^{-t}) = e^{-\frac{t}{2}}, \quad t \geqslant 0.$$
Czyli Y_i ma rozkład $Exp(\frac{1}{2})$, w szczególności $\mathbf{E}Y_i = 2$, $\mathbf{Var}Y_i = 4$. Zatem
$$(U_n - e^{-2})\sqrt{n} = e^{-2}(\exp(\frac{1}{n}\sum_{i=1}^{n}(Y_i - 2) - 1)\sqrt{n}.$$
Z mocnego prawa wielkich liczb wynika, że $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(Y_i - 2) = 0$. Zatem
$$(\exp(\frac{1}{n}\sum_{i=1}^{n}(Y_i - 2) - 1)\sqrt{n} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_i - 2)(1 + O(\frac{1}{n}\sum_{i=1}^{n}(Y_i - 2)).$$
Czyli w sensie rozkładu
$$\lim_{n\to\infty}(U_n - e^{-2})\sqrt{n} = e^{-2}\lim_{n\to\infty}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_i - 2) = 2e^{-2}Z,$$
gdzie Z ma rozkład $\mathcal{N}(0,1)$. Stąd
$$\lim_{n\to\infty}\mathbf{P}((U_n - e^{-2})\sqrt{n} < 4e^{-2}) = \mathbf{P}(Z < 2) \simeq 0,977.$$
■

2. (Eg 50/6) Rozważmy następujące zagadnienie testowania hipotez statystycznych. Dysponujemy próbką $X_1, ..., X_n$ z rozkładu normalnego o nieznanej średniej μ i znanej wariancji równej 4. Przeprowadzamy najmocniejszy test hipotezy $H_0 : \mu = 0$ przeciwko alternatywie $H_1 : \mu = -1$ na poziomie istotności $\alpha = 1/2$. Niech β_n oznacza prawdopodobieństwo błędu drugiego rodzaju, dla rozmiaru próbki n. Wybierz poprawne stwierdzenie: (asymptotyka β_n)
Odp: D-> $\lim_{n\to\infty} \beta_n \frac{e^{\frac{n}{8}}\sqrt{\pi n}}{\sqrt{2}} = 1$.

   Rozwiązanie. Najpierw znajdujemy błąd drugiego rodzaju czyli akceptacja hipotezy H_0 w sytuacji gdy zachodzi H_1. Test najmocniejszy oparty jest na konstrukcji obszaru krytycznego wynikającej z Twierdzenia Neymana-Pearsona (o porównywaniu gęstości). Niech $\mu_0 = 0$, $\mu_1 = -1$ oraz niech f_{μ_0}, f_{μ_1} będą gęstościami rozkładu wektora $(X_1, ..., X_n)$ niezależnych zmiennych losowych o tym samym rozkładzie co zmienna X, odpowiednio $X \sim \mathcal{N}(\mu_0, 4)$, $X \sim \mathcal{N}(\mu_1, 4)$. W metodzie Neymana-Pearsona badamy iloraz wiarygodności, to znaczy $f_{\mu_1}(x_1, ..., x_n)/f_{\mu_0}(x_1, ..., X_n)$. W przypadku rozkładów ciągłych nie potrzeba randomizacji, a obszar krytyczny ma postać
$$\mathcal{K} = \{(x_1, ..., x_n) \in \mathbb{R}^n : \ f_{\mu_1}(x_1, ..., x_n)/f_{\mu_0}(x_1, ..., x_n) > C\}$$
dla stałej C dobranej tak aby $\mathbf{P}_{\mu_0}((X_1, ..., X_n) \in \mathcal{K}) = \alpha$. W naszym przypadku oznacza to konieczność znalezienia stałej $\bar{C}$ takiej, że
$$\mathbf{P}_{\mu_0}(X_1 + ... + X_n < \bar{C}) = \frac{1}{2}.$$

Oczywiście $\bar{C} = 0$. Błąd drugiego rodzaju wynosi

$$\beta_n = \mathbf{P}_{\mu_1}(X_1 + ... + X_n > \bar{C}) = \mathbf{P}(-n + 2\sqrt{n}Z > 0),$$

gdzie Z ma rozkład $\mathcal{N}(0,1)$. Zatem zostaje zbadać asymptmpotykę

$$\beta_n = \mathbf{P}(Z > \frac{\sqrt{n}}{2}).$$

Oczywiście $\mathbf{P}(Z > t) \simeq \frac{1}{\sqrt{2\pi}t}e^{\frac{-t^2}{2}}$ stąd

$$\lim_{n\to\infty} \beta_n \frac{\sqrt{\pi n}}{\sqrt{2}} e^{\frac{n}{8}} = 1.$$

■

3. (Eg 51/4) Niech X i Y będą niezależnymi zmiennymi losowymi o rozkładach normalnych, przy tym $\mathbf{E}X = \mathbf{E}Y = 0$, $\mathbf{Var}X = 3$ i $\mathbf{Var}Y = 1$. Oblicz $\mathbf{P}(|X| < |Y|)$.
   Odp: A-> $\mathbf{P}(|X| < |Y|) = 0,3333$.

   Rozwiązanie. To zadanie ma czysto geometryczne rozwiązanie. Wystarczy wykorzystać rotacyjną niezmienniczość standardowego rozkładu normalnego nadto zauważyć, że $X = \sqrt{3}\bar{X}$, gdzie $\bar{X}$ ma rozkład $\mathcal{N}(0,1)$. Zatem $(\bar{X}, Y)$ ma standardowy rozkład normalny na $\mathbb{R}^2$ oraz

$$\mathbf{P}(|X| < |Y|) = \mathbf{P}(\sqrt{3}|\bar{X}| < |Y|) = \mu_{S^1}(\{\alpha \in S^1 : |\operatorname{tg}\alpha| < \frac{1}{\sqrt{3}}\}),$$

   gdzie μ_{S^1} jest miarą Lebesgue'a na okręgu jednostkowym unormowaną do 1. Oczywiście $\{\alpha \in S^1 : |\operatorname{tg}\alpha| < \frac{1}{\sqrt{3}}\} = \{\alpha \in S^1 : |\alpha| \leqslant \frac{\pi}{6} \text{ lub } |\pi - \alpha| \leqslant \frac{\pi}{6}\}$ Zatem

$$\mu_{S^1}(\{\alpha \in S^1 : |\operatorname{tg}\alpha| < \frac{1}{\sqrt{3}}\}) = \frac{1}{3}.$$

■

4. (Eg 52/5) Załóżmy, że X, Y są zmiennymi o łącznym rozkładzie normalnym, $\mathbf{E}X = \mathbf{E}Y = 0$, $\mathbf{Var}X = 2$, $\mathbf{Var}Y = 4$ i $\mathbf{Cov}(X,Y) = 1$. Oblicz $\mathbf{E}(XY|X - Y = t)$,
   Odp: C-> $\frac{7}{4} - \frac{3}{16}t^2$.

   Rozwiązanie. Potrzebujemy metody znajdowania bazy niezależnych liniowych funkcji od X, Y zawierającej $X - Y$. Szukamy α takiego, że $X - \alpha Y$ będzie niezależne od $X - Y$. Wystarczy sprawdzić kowariancję

$$\mathbf{Cov}(X - Y, X - \alpha Y) = 2 - \alpha - 1 + 4\alpha = 1 + 3\alpha.$$

   Stąd $\alpha = -\frac{1}{3}$. Wystarczy teraz rozpisać X, Y w znalezionej bazie. Mamy

$$\begin{cases} X &= \frac{1}{4}[(X - Y) + 3(X + \frac{1}{3}Y)] \\ Y &= \frac{3}{4}[-(X - Y) + (X + \frac{1}{3}Y)] \end{cases}$$

   Niech $Z = (X + \frac{1}{3}Y)$, zmienną $X - Y$ możemy traktować jako parametr t przy wyliczaniu warunkowej wartości oczekiwanej (bo jest niezależna od Z). Zatem

$$\mathbf{E}(XY|X - Y = t) = \mathbf{E}(\frac{1}{4}(t + 3Z)\frac{3}{4}(-t + Z)) = \frac{3}{16}(-t^2 + 3t\mathbf{E}Z + 3\mathbf{E}Z^2).$$

   Oczywiście $\mathbf{E}Z = 0$. Natomiast z dwuliniowości kowariancji i

$$\mathbf{E}Z^2 = \mathbf{Var}Z = 2 + \frac{2}{3} + \frac{4}{9} = \frac{28}{9}.$$

   Czyli $\mathbf{E}(XY|X - Y = t) = -\frac{3}{16}t^2 + \frac{7}{4}$.

■

5. (Eg 53/9) Mamy próbą prostą $((X_1, Y_1), (X_2, Y_2), ..., (X_{10}, Y_{10}))$ z rozkładu normalnego dwuwymiarowego o nieznanych parametrach

$$\mathbf{E}X_i = \mathbf{E}Y_i = \mu, \ \ \mathbf{Var}X_i = \mathbf{Var}Y_i = \sigma^2, \ \ \mathbf{Cov}(X_i, Y_i) = \sigma^2\rho.$$

Niech

$$Z_i = X_i + Y_i, \ \ R_i = X_i - Y_i, \ \ S_Z^2 = \frac{1}{9}\sum_{i=1}^{10}(Z_i - \bar{Z})^2, \ \ S_R^2 = \frac{1}{9}\sum_{i=1}^{10}(R_i - \bar{R})^2,$$

gdzie $\bar{Z}$ oraz $\bar{R}$ to odpowiednie średnie z próbki. Do testowania hipotezy $H_0 : \rho = \frac{1}{3}$ przeciwko alternatywie $H_1 : \rho \neq \frac{1}{3}$ możemy użyć testu o obszarze krytycznym postaci:

$$\frac{S_Z^2}{S_R^2} < k_1, \ \text{lub} \ \frac{S_Z^2}{S_R^2} > k_2,$$

przy czym liczby k_1 i k_2 dobrane są tak, aby przy założeniu, że H_0 jest prawdziwa

$$\mathbf{P}(\frac{S_Z^2}{S_R^2} < k_1) = \mathbf{P}(\frac{S_Z^2}{S_R^2} > k_2) = 0,05.$$

Liczby k_1 i k_2 są równe: ?
Odp: D-> $k_1 = 0,629$ i $k_2 = 6,358$.

Rozwiązanie. Zauważmy, że $(X_i + Y_i)$ jest niezależne od $(X_i - Y_i)$, istotnie

$$\mathbf{Cov}(X_i + Y_i, X_i - Y_i) = \mathbf{Var}X_i - \mathbf{Cov}(X_i, Y_i) + \mathbf{Cov}(Y_i, X_i) - \mathbf{Var}(Y_i) = 0.$$

To oznacza, że S_Z^2 i S_R^2 są niezależne. Wystarczy wyznaczyć ich rozkłady. Mamy

$$\mathbf{Var}(X_i + Y_i) = 2\sigma^2 + 2\rho\sigma^2 = 2(1 + \rho)\sigma^2.$$

Analogicznie

$$\mathbf{Var}(X_i - Y_i) = 2\sigma^2 - 2\rho\sigma^2 = 2(1 - \rho)\sigma^2.$$

To oznacza, że $Z_i = [2(1 + \rho)]^{\frac{1}{2}}\sigma\hat{Z}_i$, $R_i = [2(1 - \rho)]^{\frac{1}{2}}\hat{R}_i$, gdzie $\hat{Z}_i, \hat{R}_i$ są niezależne z rozkładu $\mathcal{N}(0, 1)$. Zatem

$$\frac{S_Z^2}{S_R^2} = \frac{1 + \rho}{1 - \rho}\frac{\hat{S}_Z^2}{\hat{S}_R^2}.$$

To oznacza, że aby wyznaczyć k_1, k_2 należy wziąć wartości dla $F_{9,9}$ i pomnożyć je przez $\frac{1+\rho}{1-\rho}$, które przy H_0 wynosi 2. Czyli $k_2 = 2 \cdot 3,1789 \simeq 6,358$ oraz $k_1 = 2 \cdot (1/3,1789) \simeq 0,629$

■

6. (Eg 54/9) Zmienne losowe X i Y są niezależne i zmienna X ma rozkład logarytmiczno-normalny $LN(\mu, \sigma^2)$, gdzie $\mu = 1$ i $\sigma = 2$, a zmienna Y ma rozkład wykładniczy o wartości oczekiwanej 2. Niech $S = X + Y$. Wtedy $\mathbf{E}(S|X > e)$ jest równa?
Odp: E-> 41, 26.

Rozwiązanie. Przypomnijmy definicję

$$\mathbf{E}(S|X > e) = \frac{\mathbf{E}S1_{X>e}}{\mathbf{P}(X > e)}.$$

Teraz zauważmy, że $X = e^Y$, gdzie Y ma rozkład $\mathcal{N}(1, 4)$, czyli $Y = 1 + 2Z$, gdzie $Z \sim \mathcal{N}(0, 1)$. Zatem

$$\mathbf{P}(X > e) = \mathbf{P}(1 + 2Z > 1) = \mathbf{P}(Z > 0) = \frac{1}{2}.$$

Obliczamy

$$\mathbf{E}S1_{X>e} = \mathbf{E}X1_{X>e} + \mathbf{P}(X>e)\mathbf{E}Y = \mathbf{E}e^{1+2Z}1_{Z>0} + \frac{1}{2}2 = 1 + e\mathbf{E}e^{2Z}1_{Z>0}.$$

Należy obliczyć

$$\mathbf{E}e^{2Z}1_{Z>0} = \int_0^\infty e^{2x}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx = e^2\int_0^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-2)^2}.$$

Zatem

$$\mathbf{E}e^{2Z}1_{Z>0} = e^2\int_{-2}^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx = e^2\mathbf{P}(Z>-2).$$

Ostatecznie

$$\mathbf{E}(S|X>e) = 2 + 2e^3\mathbf{P}(Z>-2) \simeq 41,26.$$

■

7. (Eg 55/4) Załóżmy, że zmienne losowe X, Y mają łączny rozkład normalny taki, że

$$\mathbf{E}X = 1,\ \mathbf{E}Y = 0,\ \mathbf{Var}(X) = 2,\ \mathbf{Var}(Y) = 9,\ \text{i } \mathbf{Cov}(X,Y) = 3.$$

Oblicz $\mathbf{Cov}(X^2, Y^2)$.
Odp: D-> 18.

Rozwiązanie. Przypomnijmy, że ogólny wzór na k-ty moment zmiennej V rozkładzie $\mathcal{N}(0,\sigma^2)$ ma postać

$$\begin{aligned}\mathbf{E}V^{2k} &= \sigma^{2k}(2k-1)(2k-3)...3\cdot 1,\\ \mathbf{E}V^{2k+1} &= 0\end{aligned}$$

Aby obliczyć $\mathbf{Cov}(X^2,Y^2)$ potrzebujemy policzyć $\mathbf{E}X^2Y^2$. Po raz kolejny należy posłużyć się bazą złożoną z liniowych względem X, Y zmiennych niezależnych. Ściślej szukamy α takiego, że $X-\alpha Y$ jest nieskorelowane z Y, a przez to niezależne bo (X,Y) tworzą wektor gaussowski. Mamy

$$\mathbf{Cov}(Y, X-\alpha Y) = 3 - 9\alpha,$$

stąd $\alpha = \frac{1}{3}$. Niech $Z = X - \frac{1}{3}Y$, zachodzi $\mathbf{E}Z = 1$, $\mathbf{Var}Z = 2 - \frac{2}{3}3 + \frac{1}{9}9 = 1$. Dalej mamy rozkład $X = Z + \frac{1}{3}Y$, stąd

$$\mathbf{E}X^2Y^2 = \mathbf{E}(Z+\frac{1}{3}Y)^2Y^2 = \mathbf{E}Z^2\mathbf{E}Y^2 + \frac{2}{3}\mathbf{E}Z\mathbf{E}Y^3 + \frac{1}{9}\mathbf{E}Y^4.$$

Drugi składnik powyżej jest zerem bo $\mathbf{E}Y = 0$, a stąd $\mathbf{E}Y^3 = 0$. Zauważmy jeszcze, że podobnie

$$\mathbf{E}X^2\mathbf{E}Y^2 = (\mathbf{E}Z^2 + \frac{1}{9}\mathbf{E}Y^2)\mathbf{E}Y^2.$$

Dlatego

$$\mathbf{Cov}(X^2,Y^2) = \frac{1}{9}(\mathbf{E}Y^4 - (\mathbf{E}Y^2)^2).$$

Pozostaje zauważyć, że $\mathbf{E}Y^4 = 3\cdot 9^2$ oraz $\mathbf{E}Y^2 = 9$. Zatem

$$\mathbf{Cov}(X^2,Y^2) = \frac{1}{9}(3\cdot 9^2 - 9^2) = 2\cdot 9 = 18.$$

■

8. (Eg 57/3) Niech $X_1, X_2, ..., X_n$ będą niezależnymi zmiennymi losowymi z rozkładu normalnego o wartości oczekiwanej 0 i nieznanej wariancji σ^2. Rozważmy estymatory odchylenia standardowego σ postaci $\hat{\sigma}_c = c\sum_{i=1}^n |X_i|$. Niech $\hat{\sigma}_{\bar{c}}$ oznacza estymator o najmniejszym błędzie średniokwadratowym w klasie rozważanych estymatorów. Wtedy $\bar{c}$ jest równe ?
Odp: D-> $\frac{\sqrt{2\pi}}{\pi+2n-2}$.

Rozwiązanie. Niech X ma rozkład $\mathcal{N}(0,\sigma^2)$. Oczywiście $\mathbf{E}|X| = \frac{\sqrt{2}}{\sqrt{\pi}}\sigma$, zatem

$$f(c) = \mathbf{E}(\hat{\sigma}_{\bar{c}} - \sigma)^2 = \mathbf{Var}(\hat{\sigma}) + (\mathbf{E}\hat{\sigma} - \sigma)^2 = nc^2\mathbf{Var}(|X|) + (cn\frac{\sqrt{2}}{\sqrt{\pi}} - 1)^2\sigma^2 =$$
$$= nc^2(1 - \frac{2}{\pi})\sigma^2 + (cn\frac{\sqrt{2}}{\sqrt{\pi}} - 1)^2.$$

Aby obliczyć punkt minimum znajdujemy z równania $f'(c) = 0$. Zachodzi

$$f'(c) = 2nc(1 - \frac{2}{\pi})\sigma^2 + 2(n^2c\frac{2}{\pi} - n\frac{\sqrt{2}}{\sqrt{\pi}})\sigma^2.$$

stąd $f'(c) = 0$ dla

$$n(\frac{2}{\pi}(n-1) + 1)c = n\frac{\sqrt{2}}{\sqrt{\pi}} \text{ czyli } c = \frac{\sqrt{2\pi}}{\pi + 2n - 2}.$$

■

9. (Eg 58/4) Niech $X_1, X_2, ..., X_n, ..., N$ będą zmiennymi losowymi. Zmienne $X_1, X_2, ..., X_n, ...$ mają rozkład logarytmiczno-normalny $LN(\mu, \sigma^2)$, gdzie $\mu = 2$, $\sigma^2 = 4$. Zmienna N ma rozkład Poissona o wartości oczekiwanej 2. Niech $S_N = \sum_{i=1}^N X_i$ dla $N > 0$ oraz $S_N = 0$ dla $N = 0$. Wtedy współczynnik asymetrii $\frac{\mathbf{E}(S_N - \mathbf{E}S_N)^3}{(\mathbf{Var}S_N)^{3/2}}$ jest równy?
Odp: D-> $\frac{e^6}{\sqrt{2}}$.

Rozwiązanie. Niech X ma rozkład $LN(\mu, \sigma^2)$, najpierw zauważmy, że

$$\mathbf{E}(S_N - \mathbf{E}S_N)^3 = \mathbf{E}\mathbf{E}[(S_N - \mathbf{E}S_N)^3|N] = \mathbf{E}\mathbf{E}[(S_N - N\mathbf{E}X + (N - \mathbf{E}N)\mathbf{E}X)^3|N] =$$
$$= \mathbf{E}N\mathbf{E}(X - \mathbf{E}X)^3 + 3\mathbf{E}N(N - \mathbf{E}N)\mathbf{E}X\mathbf{E}(X - \mathbf{E}X)^2 + \mathbf{E}(N - \mathbf{E}N)^3(\mathbf{E}X)^3.$$

Zauważmy, że

$$2 = \mathbf{E}N = \mathbf{Var}N = \mathbf{E}(N - \mathbf{E}N)^2 = \mathbf{E}N(N - \mathbf{E}N) = \mathbf{E}(N - \mathbf{E}N)^3$$

(scentrowane momenty drugi i trzeci dla rozkładu Poissona są równe wartości oczekiwanej). Zatem

$$\mathbf{E}(S_N - \mathbf{E}S_N)^3 = 2(\mathbf{E}(X - \mathbf{E}X)^3 + 3\mathbf{E}X\mathbf{E}(X - \mathbf{E}X)^2 + (\mathbf{E}X)^3) = 2\mathbf{E}X^3.$$

Z definicji $X = \exp(Y)$, gdzie Y ma rozkład $\mathcal{N}(\mu, \sigma^2)$, a dalej $Y = \sigma Z + \mu$, gdzie Z ma rozkład $\mathcal{N}(0,1)$. Przypomnijmy wzór n transformatę Laplace'cea dla rozkładu $\mathcal{N}(0,1)$

$$\mathbf{E}\exp(\lambda Z) = \exp(\frac{\lambda^2}{2}).$$

Stąd

$$\mathbf{E}X^3 = \mathbf{E}\exp(3\sigma Z + 3\mu) = e^{\frac{9\sigma^2}{2} + 3\mu} = e^{24}.$$

Zatem $\mathbf{E}(S_N - \mathbf{E}S_N)^3 = 2e^{24}$. Analogicznie pokazujemy

$$\mathbf{Var}S_N = \mathbf{E}(S_N - \mathbf{E}S_N)^2 = \mathbf{E}N\mathbf{E}(X - \mathbf{E}X)^2 + \mathbf{Var}N(\mathbf{E}X)^2 = 2\mathbf{E}X^2.$$

Mamy

$$\mathbf{E}X^2 = \mathbf{E}\exp(2\sigma Z + 2\mu) = e^{2\sigma^2+2\mu} = e^{12}.$$

czyli $\mathbf{Var}S_N = \mathbf{E}(S_N - \mathbf{E}S_N)^2 = 2e^{12}$. Obliczamy

$$\frac{\mathbf{E}(S_N - \mathbf{E}S_N)^3}{(\mathbf{Var}S_N)^{3/2}} = \frac{2e^{24}}{(2e^{12})^{\frac{3}{2}}} = \frac{e^6}{\sqrt{2}}.$$

■

10. (Eg 59/1) Zmienna losowa X rozkład logarytmiczno-normalny $LN(\mu, \sigma^2)$, gdzie $\mu = 1$ i $\sigma^2 = 4$. Wyznacz $\frac{\mathbf{E}(X-e|X>e)}{\mathbf{E}X}$.
Odp: C-> $1,82$.

Rozwiązanie. Zauważmy, że $X = e^Y$, gdzie Y ma rozkład $\mathcal{N}(\mu, \sigma^2)$, nadto $Y = \sigma Z + \mu = 2Z + 1$, gdzie Z pochodzi z rozkładu $\mathcal{N}(0, 1)$. Z definicji

$$\mathbf{E}(X - e|X > e) = \frac{\mathbf{E}1_{X>e}(X - e)}{\mathbf{P}(X > e)}.$$

Obliczamy

$$\mathbf{P}(X > e) = \mathbf{P}(2Z + 1 > 1) = \mathbf{P}(Z > 0) = \frac{1}{2}.$$

Zatem

$$\mathbf{E}1_{X>e}(X - e) = e\mathbf{E}1_{Z>0}(e^{2Z} - 1) = e(\mathbf{E}1_{Z>0}e^{2Z} - \frac{1}{2}).$$

Nadto

$$\mathbf{E}1_{Z>0}e^{2Z} = \int_0^\infty e^{2x}e^{-\frac{x^2}{2}}\frac{1}{\sqrt{2\pi}}dx = e^2\int_0^\infty e^{-\frac{(x-2)^2}{2}}\frac{1}{\sqrt{2\pi}}dx$$

a stąd

$$\mathbf{E}1_{Z>0}e^{2Z} = e^2\int_{-2}^\infty e^{-\frac{x^2}{2}}\frac{1}{\sqrt{2\pi}}dx = e^2\mathbf{P}(Z > -2).$$

Otrzymujemy

$$\mathbf{E}(X - e|X > e) = 2e^3\mathbf{P}(Z > -2) - e$$

z drugiej strony z transformaty Laplace'a

$$\mathbf{E}X = \mathbf{E}\exp(2Z + 1) = e^3.$$

Czyli

$$\frac{\mathbf{E}(X - e|X > e)}{\mathbf{E}X} = 2\mathbf{P}(Z > -2) - e^{-2} \simeq 1,82.$$

■

11. (Eg 60/6) Zmienne losowe $X_1, X_2, ..., X_{20}$ są niezależne o jednakowym rozkładzie normalnym o wartości oczekiwanej 1 i wariancji 4. Niech $S_5 = \sum_{i=1}^5 X_i$ i $\mathbf{E}(S_5^2|S_{20} = 24)$ jest równa ?
Odp: D-> 51.

Rozwiązanie. Niech $S_{15} = \sum_{i=6}^{20} X_i$. Z jednej strony $S_{20} = S_5 + S_{15}$ nadto szukamy α takiego, że

$$0 = \mathbf{Cov}(S_5 + S_{15}, S_5 - \alpha S_{15}) = \mathbf{Var}(S_5) - \alpha\mathbf{Var}(S_{15}) = 5 \cdot 4 - \alpha 15 \cdot 4 = 20(1 - \alpha 3).$$

Stąd $\alpha = 3$. Oczywiście

$$S_5 = \frac{1}{4}[(S_5 + S_{15}) + 3(S_5 - \frac{1}{3}S_{15})] = \frac{1}{4}[S_{20} + 3(S_5 - \frac{1}{3}S_{15})].$$

Zatem

$$\mathbf{E}(S_5^2|S_{20}=24)=\mathbf{E}\left(\frac{1}{16}\left[S_{20}+3(S_5-\frac{1}{3}S_{15})\right]^2|S_{20}\right)=$$

$$=\frac{1}{16}S_{20}^2+\frac{3}{8}S_{20}\mathbf{E}(S_5-\frac{1}{3}S_{15})+\frac{9}{16}\mathbf{E}(S_5-\frac{1}{3}S_{15})^2.$$

Sprawdzamy

$$\mathbf{E}(S_5-\frac{1}{3}S_{15})=0,\ \ \mathbf{E}(S_5-\frac{1}{3}S_{15})^2=5\cdot 4+\frac{15}{9}\cdot 4=\frac{80}{3}.$$

Zatem

$$\mathbf{E}(S_5^2|S_{20}=24)=\frac{1}{16}S_{20}^2+\frac{9}{16}\frac{80}{3}=36+15=51.$$

■

12. (Eg 61/6) Rozważmy następujące zagadnienie testowania hipotez statystycznych. Dysponujemy próbką $X_1,...,X_n$ z rozkładu normalnego o nieznanej średniej μ i znanej wariancji równej 2. Przeprowadzamy najmocniejszy test hipotezy $H_0:\mu=0$ przeciwko alternatywie $H_1:\mu=2$ na poziomie istotności $\alpha=\frac{1}{2}$. Niech β_n oznacza prawdopodobieństwo błędu drugiego rodzaju, dla rozmiaru próbki n. Wybierz poprawne stwierdzenie: (asymptotyka β_n).
Odp: E-> $\lim_{n\to\infty}\frac{\beta_n}{e^{-n}/\sqrt{4\pi n}}=1$.

Rozwiązanie. Niech f_0,f_2 będą gęstościami odpowiednio $\mathcal{N}(0,2),\mathcal{N}(2,2)$. Przypomnijmy, że test najmocniejszy Neymana Pearsona polega na porównaniu gęstości, czyli zbiór krytyczny ma postać

$$\mathcal{K}=\{(x_1,...,x_n)\in\mathbb{R}^n:\ \frac{f_2(x_1,...,x_n)}{f_0(x_1,...,x_n)}>C\},$$

gdzie C jest stałą taką, że

$$\mathbf{P}_{\mu=0}((X_1,...,X_n)\in\mathcal{K})=\frac{1}{2}.$$

Łatwo zauważyć, że

$$\mathcal{K}=\{(x_1,...,x_n):\ \sum_{i=1}^n x_i>\bar{C}\}.$$

Stąd szukamy stałej $\bar{C}$ takiej, że

$$\mathbf{P}_{\mu=0}(X_1+...+X_n>\bar{C})=\frac{1}{2}$$

stąd $\bar{C}=0$. Błąd drugiego rodzaju to akceptacja hipotezy H_0 podczas, gdy zachodzi H_1. Oznacza to, że

$$\beta_n=\mathbf{P}_{\mu=2}(X_1+...+X_n\leqslant 0).$$

Przy $\mu=2$ zachodzi $X_i=\sqrt{2}Z_i+2$, gdzie Z,Z_i, $i=1,2,\ldots$ będą niezależne z rozkładu $\mathcal{N}(0,1)$. Stąd

$$\beta_n=\mathbf{P}(\sqrt{2}(Z_1+...+Z_n)\leqslant -2n)=\mathbf{P}(\sqrt{2n}Z\leqslant -2n)=\mathbf{P}(Z\leqslant -\sqrt{2n}).$$

Ponieważ $\mathbf{P}(Z>t)\simeq\frac{1}{\sqrt{2\pi}t}e^{-\frac{t^2}{2}}$, więc

$$\lim_{n\to\infty}\frac{\beta_n}{e^{-n}/\sqrt{4\pi n}}=1.$$

■

13. (Eg 62/2) Niech $X_1, X_2, X_3, ...$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie logarytmiczno normalnym parametrami $\mu \in \mathbb{R}$ i $\sigma > 0$. Niech T_n oznacza estymator największej wiarygodności wariancji V^2 w tym modelu w oparciu o próbę $X_1, X_2, ..., X_n$. Niech $\mu = -0,5$ i $\sigma = 1$. Wtedy

$$\mathbf{P}(|T_n - V^2|\sqrt{n} > 10,73) = ?$$

Odp: A-> $0,134$.

Rozwiązanie. Zachodzi $X_i = e^{Y_i}$, $i = 1, 2, ..., n$, gdzie $Y, Y_1, Y_2, ..., Y_n$ są niezależne i pochodzą z rozkładu $\mathcal{N}(\mu, \sigma^2)$. Trzeba przypomnieć estymatory największej wiarygodności dla rozkładu normalnego

$$\bar{\mu}_n = \bar{Y} = \frac{1}{n}(Y_1 + ... + Y_n), \quad \bar{\sigma}_n^2 = \bar{Y^2} - (\bar{Y})^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

Powyższe estymatory wykorzystujemy aby znaleźć estymator wariancji V^2. Mamy

$$V^2 = \mathbf{Var}Y = \mathbf{E}Y^2 - (\mathbf{E}Y)^2 = e^{2\sigma^2+2\mu} - e^{\sigma^2+2\mu} = e - 1.$$

Nadto z powyższego wzory wynika, że

$$T_n = e^{2\bar{\mu}_n}(e^{2\bar{\sigma}_n^2} - e^{\bar{\sigma}_n^2}).$$

Należy pamiętać, że $\bar{\mu}_n$ i $\bar{\sigma}_n^2$ są niezależne. Mamy

$$\begin{aligned} T_n - V^2 &= (e^{2\bar{\mu}_n} - e^{-1})(e^{2\bar{\sigma}_n^2} - e^{\bar{\sigma}_n^2}) + \\ &+ e^{-1}(e^{2\bar{\sigma}_n^2} - e^2 - e^{\bar{\sigma}_n^2} - e). \end{aligned}$$

Z mocnego prawa wielkich liczb wynika, że $\bar{\mu}_n \to \mu$, $\bar{\sigma}_n^2 \to \sigma^2$ prawie na pewno. Nadto

$$\sqrt{n}(e^{2\bar{\mu}_n} - e^{-1}) = \sqrt{n}e^{-1}(e^{2\bar{\mu}_n+1} - 1) = \sqrt{n}e^{-1}(2\bar{\mu}_n + 1)(1 + O(2\bar{\mu}_n + 1)).$$

Oczywiście w sensie słabej zbieżności

$$\lim_{n\to\infty} \sqrt{n}(2\bar{\mu}_n + 1) = \mathcal{N}(0, 4).$$

Stąd

$$\lim_{n\to\infty} \sqrt{n}(e^{2\bar{\mu}_n} - e^{-1})(e^{2\bar{\sigma}_n^2} - e^{\bar{\sigma}_n^2}) = A \simeq \mathcal{N}(0, [2(e-1)]^2)$$

Ściślej korzystając z $\bar{\sigma}_n^2 \to \sigma^2$

$$\lim_{n\to\infty} \sqrt{n}(e^{2\bar{\mu}_n} - e^{-1})(e^{2\bar{\sigma}_n^2} - e^{\bar{\sigma}_n^2}) = \lim_{n\to\infty} \sqrt{n}(e^{2\bar{\mu}_n} - e^{-1})(e^2 - e) = A.$$

Analogicznie

$$e^{2\bar{\sigma}_n^2} - e^2 + e^{\sigma_n^2} - e = (2e^2 - e)\sqrt{n}(\bar{\sigma}_n^2 - 1)(1 + O(\bar{\sigma}_n^2 - 1)).$$

W sensie słabej zbieżności

$$\lim_{n\to\infty} \sqrt{n}(\bar{\sigma}_n^2 - 1) = \mathcal{N}(0, 2).$$

Zatem

$$\lim_{n\to\infty} e^{-1}(e^{2\bar{\sigma}_n^2} - e^2 - e^{\bar{\sigma}_n^2} - e) = B \simeq \mathcal{N}(0, 2[2e-1]^2).$$

Ostatecznie korzystając z niezależności $\bar{\mu}_n$ oraz $\bar{\sigma}_n^2$ dostajemy dla niezależnych A i B

$$\lim_{n\to\infty} \sqrt{n}(T_n - V^2) = A + B \simeq \mathcal{N}(0, [2(e-1)^2] + 2[2e-1)]^2).$$

Niech Z będzie z $\mathcal{N}(0, 1)$ otrzymujemy

$$\lim_{n\to\infty} \mathbf{P}(|T_n - V^2|\sqrt{n} > 10,73) = \mathbf{P}(|Z| > \frac{10,73}{\sqrt{[2(e-1)^2] + 2[2(e-2)]^2}}) = 0,134.$$

Przedstawiona metoda ma swoją nazwę jako metoda delta. Powyższy przypadek szczególny można zebrać w ogólne twierdzenie.

Twierdzenie 1 *(Metoda delta) Jeżeli dla ciągu zmiennych T_n mamy $\sqrt{n}(T_n - \mu) \to \mathcal{N}(0, \sigma^2)$ przy $n \to \infty$ i $h : \mathbb{R} \to \mathbb{R}$ jest funkcją różniczkowalną w punkcie μ, to*

$$\sqrt{n}(h(T_n) - h(\mu)) \to \mathcal{N}(0, \sigma^2(h'(\mu))^2),$$

w sensie zbieżności według rozkładu.

■

14. (Eg 63/7) Zmienna losowa (X, Y, Z) ma rozkład normalny z wartością oczekiwaną $\mathbf{E}X = 0$, $\mathbf{E}Y = \mathbf{E}Z = 1$ i macierzą kowariancji

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 4 & 2 \\ 0 & 2 & 4 \end{bmatrix}.$$

Obliczyć $\mathbf{Var}(X(Y + Z))$.
Odp: D-> 17.

Rozwiązanie. Stosujemy metodę z uniezależnianiem zmiennych. Z założenia wynika, że $\mathbf{Cov}(X, Y) = 1$ oraz $\mathbf{Cov}(X, Z) = 0$, $\mathbf{Cov}(Y, Z) = 2$. Zatem zmienne X i Z są niezależne, Wystarczy dobrać α i β tak aby $Y - \alpha X - \beta Z$ było niezależne (czyli równoważnie nieskorelowane) ze zmiennymi X i Z. Sprawdzamy, że

$$\mathbf{Cov}(Y - \alpha X - \beta Z, X) = 1 - \alpha, \text{ czyli } \alpha = 1$$

nadto

$$\mathbf{Cov}(Y - \alpha X - \beta Z, Z) = 2 - 4\beta, \text{ czyli } \beta = \frac{1}{2}.$$

Zatem baza liniowa składa się ze zmiennych niezależnych $X, \bar{Y} = Y - X - \frac{1}{2}Z, Z$, gdzie X ma rozkład $\mathcal{N}(0, 1)$, zmienna $\bar{Y}$ ma rozkład $\mathcal{N}(\frac{1}{2}, 2)$, a zmienna Z rozkład $\mathcal{N}(1, 4)$. Obliczamy

$$\begin{aligned}
&\mathbf{Var}(X(Y + Z)) = \mathbf{Var}(XY) + 2\mathbf{Cov}(XY, XZ) + \mathbf{Var}(XZ) = \\
&= \mathbf{Var}(X\bar{Y}) + 2\mathbf{Cov}(X\bar{Y}, X(X + \frac{1}{2}Z)) + \mathbf{Var}(X(X + \frac{1}{2}Z)) + \\
&+ 2\mathbf{Cov}(X\bar{Y}, XZ) + 2\mathbf{Cov}(X(X + \frac{1}{2}Z), XZ) + \mathbf{Var}(XZ) = \\
&= \mathbf{E}X^2\mathbf{E}\bar{Y}^2 + \mathbf{E}X^2\mathbf{E}\bar{Y}\mathbf{E}Z + \mathbf{E}X^4 - (\mathbf{E}X^2)^2 + \frac{1}{4}\mathbf{E}X^2\mathbf{E}Z^2 + \\
&+ 2\mathbf{E}X^2\mathbf{E}\bar{Y}\mathbf{E}Z + 2\mathbf{E}X^2\mathbf{E}Z^2 = (2 + \frac{1}{4}) + \frac{1}{2} + 2 + \frac{5}{4} + 1 + 10 = 17.
\end{aligned}$$

■

15. (Eg 64/9) Niech $Y_1, Y_2, ..., Y_n$ będą niezależnymi zmiennymi losowymi, przy czym zmienna Y_i, $i = 1, 2, ..., n$, ma rozkład logarytmiczno-normalny $LN(bx_i, 1)$, gdzie $x_1, x_2, ..., x_n$ są znanymi liczbami, a b jest nieznanym parametrem. Załóżmy, że $\sum_{i=1}^n x_i^2 = 4$. Niech $\bar{b}$ będzie estymatorem największej wiarogodności parametru b, a $\bar{g} = \exp(2\bar{b})$ estymatorem funkcji $g(b) = \exp(2b)$. Wtedy obciążenie estymatora $\bar{g}$

$$\mathbf{E}_b\bar{g} - g(b)$$

jest równe
Odp: B-> $e^{2b}(\sqrt{e} - 1)$.

Rozwiązanie. Najpierw obliczamy wiarygodność dla zmiennych $Z_1, Z_2, ..., Z_n$, gdzie $Z_i = \ln Y_i$, czyli Z_i ma postać $\mathcal{N}(bx_i, 1)$. Obliczamy wiarygodność dla $Z_1, ..., Z_n$

$$L(b, z) = (2\pi)^{-\frac{n}{2}} \exp(-\frac{1}{2}\sum_{i=1}^n (z_i - bx_i)^2).$$

Rozwiązujemy równanie $f'(b) = 0$ dla $f(b) = \ln L(b, z)$. Wówczas

$$\sum_{i=1}^{n}(z_i - bx_i)x_i = 0, \quad \text{czyli} \quad b = \frac{\langle x, z\rangle}{\langle x, x\rangle} = \frac{1}{4}\langle x, z\rangle.$$

Stąd

$$\bar{b} = \frac{1}{4}\sum_{i=1}^{n} x_i Z_i.$$

Pozostaje zauważyć, że $\bar{b}$ ma rozkład $\mathcal{N}(b, \frac{1}{4})$. Stąd

$$\mathbf{E}_b \bar{g} - g(b) = \mathbf{E}_b \exp(2\bar{b}) - e^{2b} = e^{2b}(e^{\frac{1}{2}} - 1).$$

■