Zadanie:
Rozkład μ jest zadany przez funkcję charakterystyczną:

$$\phi(t) = \frac{1}{2}\cos(\sqrt{2}t) + \frac{1}{2}\exp(-6t^2/2)$$

Niech $X_1, X_2, \dots$ będą niezależnymi zmiennymi losowymi o rozkładzie μ i $S_{49} = X_1 + \dots + X_{49}$. Wiedząc, że $\Phi(-2) = 0.0228$, gdzie $\Phi(x)$ jest dystrybuantą standardowego rozkładu normalnego, oszacować $P(S_{49} \in (-28, 28))$.

Możliwe rozwiązanie:

Mamy $E(X^2) = -\partial_t^2\phi(0)$. Jak ktoś nie pamięta wzoru to można go wyliczyć następująco: $\phi(t) = E(\exp(itX))$, różniczkując mamy $\partial_t\phi(t) = E(iX\exp(itX))$ i $\partial_t^2\phi(t) = E(-X^2\exp(itX))$. Dla $t = 0$ dostajemy $\partial_t^2\phi(0) = E(-X^2)$. Ten rachunek łatwo uzasadnić jeśli $E(X^2)$ istnieje (jest skończone), ale znane twierdzenie mówi że jeśli ϕ ma drugą pochodną w zerze to $E(X^2)$ jest skończone.

W naszym przypadku mamy $\partial_t^2\phi(0) = -4$, czyli $E(X^2) = 4$. Ponadto $\partial_t\phi(0) = 0$, czyli $E(X) = 0$. Z centralnego twierdzenia granicznego rozkład $\frac{S_n}{\sqrt{n}}$ gdzie $S_n = X_1 + \dots + X_n$ dąży przy n dążącym do nieskończoności do rozkładu normalnego o wariancji $\sigma^2 = 4$. Nasza funkcja charakterystyczna jest bardzo regularna, więc można liczyć że już dla $n = 49$ będziemy dostatecznie blisko granicy. Czyli $Y = \frac{S_{49}}{2\cdot 7}$ ma rozkład bliski standardowego rozkładu normalnego. Wtedy

$$P(Y \le -2) \approx \Phi(-2) \approx 0.0228.$$

Ze względu na symetrię rozkładu normalnego

$$P(Y \ge 2) \approx \Phi(-2) \approx 0.0228.$$

Czyli

$$\begin{aligned} P(S_{49} \in (-28, 28)) &= P(Y \in (-2, 2)) = 1 - P(Y \le -2) - P(Y \ge 2) \\ &\approx 1 - 2 \cdot 0.0228 = 0.9544. \end{aligned}$$

Inne rozwiązanie: Rozkład X_i to średnia (mieszanka) rozkładu normalnego o wartości oczekiwanej 0 i wariancji 6 z symetrycznym rozkładem dwupunktowym na $\{-\sqrt{2}, \sqrt{2}\}$. A więc rozkład S_{49} to

$$\frac{1}{2^{49}}\sum_{i=0}^{49}\binom{49}{i}D^{*i} * N_{0,6(49-i)}$$

gdzie $*$ oznacza splot, D^{*i} to potęga splotowa rozkładu dyskretnego wyżej (czyli przesunięty rozkład dwumianowy), zaś $N_{m,a}$ to rozkład normalny o wartości oczekiwanej m i wariancji a. Splot można rozpisać jako sumę:

$$D^{*i} * N_{0,6(49-i)} = \frac{1}{2^i}\sum_{j=0}^{i}\binom{i}{j}N_{\sqrt{2}*(i-2j),6(49-i)}.$$

Czyli łącznie mamy rozkład

$$\frac{1}{2^{49}}\sum_{i=0}^{49}\binom{49}{i}\frac{1}{2^i}\sum_{j=0}^{i}\binom{i}{j}N_{\sqrt{2}*(i-2j),6(49-i)}$$

Podwójną sumę wyżej można policzyć numerycznie na komputerze, co daje w przybliżeniu 0.9544103908, w więc podobnie jak przybliżenie normalne (używając dokładniejszą wartość $\Phi(-2)$ widać że przybliżenie normalne daje nieco większą wartość).

W sytuacji jak wyżej zwykle interesuje nas prawdopodobieństo tego że zmienna losowa jest poza podanym przedziałem (tzn. oszacowanie ogonów rozkładu). Prowadzi to do zagadnienia wielkich odchyleń. W zagadnieniu wielkich odchyleń przybliżenie normalne często prowadzi do znacznego błędu (powyżej odchylenie nie było zbyt duże i błąd względny dla ogona jest rzędu 0.0019, a więc rozsądnie mały).

Aby zilustrować możliwy błąd rozpatrzmy zadanie które wygląda podobnie do poprzedniego zadania: rozkład μ jest skoncentrowany w trzech punktach: -28, 0 i 28. $P(\{0\}) = \frac{195}{196}$, $P(\{-28\}) = P(\{28\}) = \frac{1}{392}$. Łatwo sprawdzić że gdy X_i mają rozkład μ to $E(X_i) = 0$ zaś $E(X_i^2) = 4$. A więc używając przybliżenie normalne by oszacować $P(S_{49} \in (-28, 28))$ dostaniemy ten sam wynik czyli 0.9544, zaś dla $P(S_{49} \notin (-28, 28))$ wynik 0.0456. Jednakże, μ można zapisać jako mieszankę dwu rozkładów, rozkładu δ_0 skoncentrowanego w 0 i rozkładu dwupunktowego skoncentrowanego na $\{-28, 28\}$ takiego że $P(\{-28\}) = P(\{28\}) = \frac{1}{2}$. Oznaczając ten ostatni rozkład przez ν można napisać

$$\mu = (1-c)\delta_0 + c\nu$$

gdzie $c = \frac{1}{196}$. Rozkład sumy jest zadany przez splot, czyli potrzebujemy obliczyć μ^{*49}, tzn. 49-tą potęgę splotową μ. Można to zrobić używając wzór dwumianowy Newtona:

$$\mu^{*49} = ((1-c)\delta_0 + c\nu)^{*49} = \sum_{i=0}^{49} \binom{49}{i} (1-c)^{49-i} \delta_0^{*(49-i)} c^i \nu^{*i}$$

$$= \sum_{i=0}^{49} \binom{49}{i} (1-c)^{49-i} c^i \nu^{*i}$$

gdzie ostatnia równość wynika z tego że δ_0 jest jedynką względem splotu. Popatrzmy na wyraz z $i = 1$:

$$\binom{49}{i} (1-c)^{49-1} c\nu = 49 * (\frac{195}{196})^{48} \frac{1}{196} \nu \approx 0.195573\nu.$$

Zauważmy że $\nu(\mathbb{R} - (-28, 28)) = 1$. We wzorze dwumianowym wyżej wszystkie składnik są dodatnie, a więc

$$P(S_{49} \notin (-28, 28)) = \mu^{*49}(\mathbb{R} - (-28, 28)) \geq 49 * (\frac{195}{196})^{48} \nu(\mathbb{R} - (-28, 28))$$
$$\approx 0.195573$$

co oznacza że naprawdę prawdopodobieństwo ogona jest ponad 4 razy większe niż oszacowanie z przybliżenia normalnego.

Najprostsze oszacowanie z góry wielkich odchyleń daje nierówność Czebyszewa. Niestety, oszacowanie przy pomocy nierówności Czebyszewa jest zwykle bardzo pesymistyczne (zbyt duże). W naszym "złym" przykładzie mamy $E(S_{49}^2) = 49E(X_i^2) = 196$ i nierówność Czebyszewa daje

$$P(S_{49}^2 \notin (-28, 28)) \leq \frac{E(S_{49}^2)}{28^2} = \frac{1}{4}.$$

Porównując to oszacowaniem z dołu widzimy że nierówność Czebyszewa w tym przypadku jest znacznie bliższa prawdy niż oszacowanie normalne.

Dla rozkładów z szybko malejącymi ogonami lepiej działają momenty wykładnicze (równoważnie transformacja Laplace'a):

$$P(S > a) \leq \exp(-ta)E(\exp(tS)).$$

W naszym przypadku $E(\exp(tS))$ łatwo otrzymać z funkcji charakterystycznej przez przedłużenie zespolone (podstawiając it zamiast t), co daje

$$E(\exp(tS_{49})) = (\frac{1}{2}\cosh(\sqrt{2}t) + \frac{1}{2}\exp(6t^2/2))^{49}.$$

W oszacowaniu wyżej należy wybrać t tak by uzyskać najmniejszą możliwą wartość. Dokładna minimalizacja jest trochę kłopotliwa ale pamiętając o przybliżeniu normalnym można dobrać t tak by uzyskać minimalną wartość dla rozkładu normalnego. Biorąc $a = 28$ (zgodnie z treścią zadania) dla rozkładu normalnego mamy wyrażenie:

$$\exp(-28t)\exp(98t^2)$$

z pochodną

$$(196t - 28)\exp(-28t)\exp(98t^2)$$

która zeruje się dla $t = \frac{1}{7}$. Biorąc takie t w oszacowniu wyżej dostaniemy

$$P(S_{49} > 28) \leq 0.136$$

co jest gorsze od przybliżenia normalnego. Zaletą oszacowania wykładniczego jest to że przy znanej funkcji charakterystycznej nierówność jest zawsze spełniona (nie ma niesprawdzonych założeń).

Zadanie: Niech (X, Y) będzie wektorem losowym z wektorem średniej $(2, 1)$ i macierzą kowariancji

$$\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \tag{1}$$

oraz $Z = X+2Y$. Ciąg $Z_1, Z_2, \ldots, Z_{100}$ jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie jak Z. Oszacować

$$p = P(Z_1 + \cdots + Z_{100} \in (400 - 10\sqrt{10}, 400 + 10\sqrt{10})).$$

Możliwe rozwiązanie: Oznaczmy $S_{100} = \sum_{i=1}^{100} Z_i$. Wartość oczekiwana Z to 4. Wariancja Z to $\langle\Sigma(1,2),(1,2)\rangle = 10$. Na mocy centralnego twierdzenia granicznego

$$p = P(\frac{S_{100}}{10} - 40 \in A) \approx P(N_{10} \in A) = P(N_1 \in \frac{1}{\sqrt{10}}A)$$

gdzie N_a to rozkład normalny o wartości oczekiwanej 0 i wariancji a. U nas $A = (-\sqrt{10}, \sqrt{10})$, czyli dostajemy $1 - 2\Phi(-1) = \mathrm{erf}(1/\sqrt{2}) \approx 0.6827$.